Thèse de Doctorat de l'Université Pierre et Marie Curie

Spécialité: Astronomie & Astrophysique

Ecole Doctorale 127 "Astronomie et Astrophysique d'Île de France"
Institut d'Astrophysique de Paris

Presentée par

Clotilde Laigle

Pour obtenir le grade de

Docteur de l'Université Pierre et Marie Curie

Observational and theoretical constraints on galaxy evolution at high redshift

soutenue le 22 septembre 2016 devant le jury composé de

M.	Henry-Joy McCracken	Directeur de thèse
M.	Christophe Pichon	Co-directeur de thèse
M.	Oliver Hahn	Rapporteur
M.	Piero Madau	Rapporteur
M.	Jeremy Blaizot	Examinateur
M.	Chris Martin	Examinateur
M.	Benoit Semelin	Examinateur
Mme	Adrianne Slyz	Examinateur
M.	Stéphane Arnouts	Membre invité

Résumé

Les propriétés des galaxies sont façonnées par l'interaction complexe entre des processus non-linéaires à la fois internes et externes, induits par la gravité à grande échelle et la physique baryonique à petite échelle. Si le scénario géneral de la formation des galaxies est relativement bien établi, les détails de sa mise en oeuvre sont encore intensément débattus. Je présente dans cette thèse de nouvelles contraintes sur la formation et l'évolution des galaxies, en étudiant leur croissance en masse et leur évolution au sein de la toile cosmique depuis l'époque de leur formation jusqu'à maintenant.

Pour cela j'ai développé des techniques pour interpréter les grands relevés de galaxies. J'ai créé un nouveau catalogue photométrique sur le champ COSMOS incluant la mesure précise des redshifts photométriques. Ce catalogue permet de sonder avec précision l'Univers à haut redshift (par exemple, nous pouvons construire des échantillons complets en masse jusqu'à redshift 4 avec une masse limite de 10^{10} $M_{\odot}$). J'analyse ce relevé observé à l'aide de relevés virtuels, produits à partir de simulations hydrodynamiques cosmologiques. Ces simulations implémentent l'état de l'art de nos connaissances sur la formation et l'évolution des galaxies. Les processus d'observations et de détection peuvent être modélisés dans les relevés virtuels. L'analyse de mêmes quantités, soit observées par un télescope, soit générées par un supercalculateur calibré pour avoir les mêmes limitations que ce télescope, m'assure que la comparaison est significative.

Je présente le résultat de cette analyse dans la première partie de cette thèse. Je montre que l'évolution en redshift des propriétés des galaxies est relativement bien comprise en invoquant des processus qui dépendent essentiellement de la masse, tels que le feedback des étoiles et des AGN. Je souligne également comment nos méthodes observationnelles génèrent des biais dans les propriétés physiques des galaxies calculées à partir de la photométrie.

Galaxies et halos sont plongés dans la structure filamentaire appelée "toile cosmique". Dans un deuxième temps, je montre comment la dynamique des flots de matière à grande échelle gouverne l'acquisition du moment angulaire des galaxies et halos de matière noire, ce qui implique que certaines propriétés des galaxies sont supposées dépendre de leur environnement anisotrope. J'ai extrait la structure filamentaire du catalogue photométrique que j'ai créé sur le champ COSMOS et j'ai mesuré cette dépendance. Je trouve des gradients de masse et de couleurs dans la direction du filament. Il apparaît que la masse et le moment angulaire des galaxies sont deux quantités inter-dépendantes et tous deux impactés par leur environement anisotrope.
La question de l'influence de l'environnement à grande échelle sur les propriétés des galaxies est incontournable pour affiner notre modèle d'évolution de formation et d'évolution galaxies.

Mots Clés: cosmologie, grandes structures de l'Univers, grands relevés photometriques, astrophysique numérique, évolution des galaxies, haut redshift.

Observational and theoretical constraints on galaxy evolution at high redshift

Abstract

Galaxy properties arise from a complex interplay between internal and externally-induced non-linear processes, driven by gravity at large scales and baryonic physics at small scales. While their basic mass build-up scenario is qualitatively well established, its exact quantitative implementation is still under intense investigation. I present in this thesis new constraints on galaxy formation and evolution while studying the galaxy mass growth and the co-evolution of the cosmic web and the embedded galaxies, from the epoch of cosmic dawn to today.

To do so, I developed techniques to interpret galaxy surveys. I first created a new photometric catalog on the COSMOS field with precise photometric redshifts allowing to probe accurately the high-redshift Universe (for instance we can make complete samples up to $z \sim 4$ with a mass limit of 10^{10} $M_{\odot}$). I analyze this survey while relying heavily on comparisons with virtual galaxy surveys produced from state-of-the-art cosmological hydrodynamical simulations, which capture all our current knowledge of galaxy formation and evolution together with a detailed modelling of all aspects of the process of observation and detection. They allow me to carry out meaningful statistical measurements of the differences between models and observations. In other words, analysing galaxy observables, whether captured by a telescope or generated with a supercomputer tuned to have the same limitations as this telescope, ensures that I am comparing (millions of) apples to apples.

From this comparative analysis, in the first part of my thesis I show that the redshift evolution of galaxy properties is reasonably well understood when invoking mass-dependent processes (AGN and stellar feedback). I highlight also the effect of simplifying assumptions inherent to our observational methods, which bias the physical properties computed from galaxy photometry.

Galaxies and haloes are embedded in the cosmic web, an intricate large-scale structure of walls, filaments and nodes. In the second part of my thesis, I show how galaxies and dark haloes gain their angular momentum from the large-scale flow, implying that some of their properties depend on their anisotropic filamentary environment. I then extract the filamentary structure from the observed photometric catalog and measure the dependence of galaxy properties to the anisotropic environment. I find mass and colour gradients towards the filaments. In turn it emerges that galaxy masses and angular momenta are two dependent quantities impacted by their anisotropic environment.
Determining the full extent of the influence of the large-scale environment on galaxy properties is clearly one of the next frontier in our understanding of galaxy formation and evolution.

Keywords: cosmology, large-scale structure of the Universe, large photometric surveys, computational astrophysics, galaxy evolution, high redshift.

Contents

Acronyms

Abbreviations

ΛCDM	Λ-cold dark matter
AGB	Asymptotic giant branch
AGN	Active galactic nuclei
BC03	Bruzual and Charlot (2003)
BH	Black hole
CCD	Charge-coupled device
CGM	Circumgalactic medium
DEC	Declination
FIR	Far-infrared
FOF	Friends-of-friends
FUV	Far-ultraviolet
FWHM	Full width at half maximum
HOD	Halo occupation distribution
IGM	Intergalactic medium
IMF	Initial mass function
IR	Infrared
ISM	Interstellar medium
LSS	Large-scale structure
MIR	Mid-infrared
MLCR	Mass-to-light-colour relation
NIR	Near-infrared
NUV	Near-ultraviolet
PAH	Poly-cyclic aromatic hydrocarbons
PDF	Probability density function
PSF	Point spread function
QSO	Quasi stellar objects
RA	Right ascension
SED	Spectral energy distribution
SFH	Star formation history
SFR	Star formation rate
SFRD	Star formation rate density
SPS	Stellar population synthesis
SSP	Single stellar population
TTT	Tidal torque theory
UV	Ultraviolet

Instruments & surveys

2dFGRS	2 degree Field Galaxy Redshift Survey
2MASS	2 Micron All Sky Survey
ACS	Advanced Camera for Surveys
CANDELS	Cosmic Assembly NIR Deep Extragalactic Legacy Survey
CFHT	Canada France Hawaii Telescope
COSMOS	Cosmic Evolution Survey
DR1	Data release 1
DR2	Data release 2

E-ELT	European Extremely Large Telescope
EUVE	Extreme UV Explorer
GALEX	Galaxy Evolution Explorer
GAMA	Galaxy And Mass Assembly
HSC	Hyper Suprime Cam
HST	Hubble Space Telescope
IRAC	IR Array Camera
IRAS	IR Astronomical Satellite
J-PAS	Javalambre physics of the Accelerating Universe Astrophysical Survey
JWST	James Webb Space Telescope
K-CWI	Keck Cosmic Web Imager
LSST	Large Synoptic Survey Telescope
MIPS	Multiband Imaging Photometer for *Spitzer*
MUSE	Multi Unit Spectroscopic Explorer
PACS	Photodetector Array Camera and Spectrometer
PFS	Prime Focus Spectrograph
SDSS	Sloan Digital Sky Survey
SKA	Square Kilometer Array
SPIRE	Spectral and Photometric Imaging Receiver
SPLASH	*Spitzer* Large Area Survey with HSC
SXDS	Subaru/XMM-Newton Deep Survey
UKIRT	United Kingdom IR Telescope
VIMOS	VIsible MultiObject Spectrograph
VIPERS	VIMOS Public Extragalactic Redshift Survey
VIRCAM	Vista IR Camera
VISTA	Visible and IR Survey Telescope for Astronomy
VLA	Very Large Array
VLT	Very Large Telescope
WFIRST	Wide-Field IR Survey Telescope
WIRCAM	Wide-field IR Camera
XMM	X-ray multi-mirrors

Part I.

Introduction

Cosmology aims to propose a self-consistent model for understanding the origin, dynamics and eventual fate of the Universe on large scale. The particular status of cosmology among other sciences lies in the specificity of the available data to test the theory: the validity of the model can in fact only be assessed through the confrontation with observations of our own past lightcone. Hence cosmological observations rely primarily (although not only) on galaxies, the visible bricks of the Universe.
As a matter of fact, rapid progress has been made once the first observations of the recession of distant galaxies (Hubble, 1929) were able to confirm the theory of an expanding Universe (Lemaître, 1927) in the framework of the General Relativity (Einstein, 1915). This discovery of the expansion of the Universe implied a primordial singularity (called the "Big Bang" by Hoyle, 1949) and the existence of a cosmic microwave background radiation (Alpher et al., 1948), lately confirmed "unwittingly" by observations (the "excess antenna temperature at 4080 Mc/s", Penzias and Wilson, 1965). The standard model of cosmology describes an Universe for which the mass-energy density is dominated by dark matter and dark energy, in which structures have grown by gravitational amplification of seed primordial fluctuations: cosmology is thus "encoded" in the masses and distribution of structures. From this solid basis, cosmology has progressed from a purely theoretical pursuit to one which is capable of making quantitative predictions, many of which have been found consistent with observations. With the explosion of available data during the last decades, cosmology is now believed by some to have entered in a precision area. Among the useful observational data, large multi-wavelength galaxy catalogues are indispensable for mapping the three-dimensional distribution of matter, converting light into mass, and improving our knowledge of the cosmological parameters.
However, as galaxy observations accumulated, galaxy populations have proved to be very diverse and the theory of galaxy evolution has emerged as a complex topic. The situation as it stands is that galaxy surveys are essential to make progresses in cosmology, but they are also highly biased tracers of the underlying matter. That is why there is no doubt that, to constrain cosmology, the assembly of visible matter has to be understood in detail by building simultaneously a theory of galaxy formation and evolution.

Observationally, galaxies can be described in terms of colours and morphologies.This description highlights clear distinctions between galaxy subpopulations, gathered under the name of "Hubble sequence" (for a mainly morphology-based classification) or "bimodality" (for a mainly colour-based classification). The question of what are the main processes driving this classification is a long-standing issue and still a matter of debate.
Galaxies form from gas collapsing into the center of dark matter halos. Galaxies grow in mass by gas accretion from the outside, the cold gas being the fuel for galaxies to form stars (White and Rees, 1978). The later evolution of galaxies is determined all together by internal mechanisms (stellar winds, supernovae explosions, active nuclei feedback), the rate and mode of gas accretion (depending of the mass of the host dark halo, the amount of gas available in galaxy environment and the galaxy physics itself), and interactions with other galaxies (mergers). Galaxies are thus systems shaped by a large range of processes, occurring at various spatial and time scales. Building a self-consistent model of galaxy evolution requires to adequately describe physical processes at all the different scales, to determine the level of independence of each scale in relation to the others, and finally to understand the links between galaxy formation history and their properties. Galaxy colours and luminosities are related to their star formation activity, which, in turn, depends on the availability of cold gas. Following the gas, how it is accreted, turned into stars and possibly removed from the galaxy is thus the most convenient strategy to understand galaxy properties and it could be critical to pin down the role played by the environment.

The amount of gas available, the rate at which it is turned into stars and the gas outflow mechanisms are not only described but also essentially quantified by the global build up of **galaxy mass** over cosmic time. In this context, constraining galaxy evolution and properties implies tracking the galaxy (and host dark halo) mass growth over cosmic time through simple statistical quantities such as the mass function. It is in particular important to adjust our galaxy evolution model on the full history of the Universe to understand

when galaxy build-up proceeded most efficiently. This is possible especially with large deep near-infrared and infrared surveys, which allow to extend this study up to $z \sim 6$. Comparing such surveys with *ab initio* cosmological simulations (which can now include most of the known baryonic physics processes down to relatively small scales) over this redshift range provides a solid basis to disentangle the relative importance of the processes at work in stellar mass growth. The first aspect of my thesis is thus concerned with answering this question: *what shapes galaxy stellar mass growth and regulates star formation?* Reliable conclusions can be drawn from such studies only if a large enough galaxy sample is available over the full redshift range of interest. That is why I spent the necessary time to build such a multi-wavelength photometric catalogue from observations and its counterpart on hydrodynamical simulations. Their joint analysis can uniquely allows us to assess the reliability of our galaxy evolution model, as implemented in the simulations. Importantly, this approach is meaningful only if the uncertainties and biases which are inherently linked to the observable estimations are carefully taken into account.
Galaxies and haloes are embedded in the cosmic web, an intricate large-scale structure of walls, filaments and nodes. On one hand the cold gas closely follows the filamentary dark matter at large scales and the gas infall onto galaxies is anisotropic. On the other hand the dynamics of the large-scale flows is expected to impact the dynamics of the galaxies, especially their **angular momentum** distribution. Hence some galaxy properties (such as their morphology) are expected to dynamically depend on this anisotropic environment. The second part of my thesis is thus dedicated to answer this question: *How galaxies acquire their angular momentum and how it is related to their observable properties?* Determining the extent of the influence of the large-scale environment is clearly one of the next frontier in our understanding of galaxy mass evolution.

This thesis presents the work that I have carried out during my PhD between october 2013 and september 2016 at the Institut d'Astrophysique de Paris, under the supervision of Henry-Joy McCracken and Christophe Pichon. More specifically, this thesis is organised as follows.
The introductory **Part** I provides the context of my research. In this first Part, **Chapter** 1 presents a panorama of the broad field of research concerned with galaxy evolution by recalling the milestones of the history of the field and the new challenges. **Chapter** 2 reports on our current knowledge about the growth of structures in our current cosmological framework and the galaxy life-cycle.
Part II presents the tools used in this thesis, both on the observational and numerical sides. **Chapter** 3 summarises the stellar evolution physics required to interpret galaxy light in terms of stellar contents, presents the basis of the modelling of galaxy light through stellar population synthesis codes and gives keys to understand the galaxy spectrum. **Chapter** 4 introduces the Horizon-AGN hydrodynamical simulation and the virtual catalogue from the simulation. **Chapter** 5 presents the computation of the cosmos2015 photometric catalogue on COSMOS field (Laigle, McCracken et al. 2016, ApJ Sup., 224:24).
Part III presents the results obtained during this thesis. In **Chapter** 6, the virtual catalogue from Horizon-AGN is compared with an array of observational data through global quantities (Kaviraj, Laigle et al. 2016, submitted to ApJ). This comparison assesses the reliability of the model implemented in the simulation and emphasises the required improvements by highlighting the point of tensions. The comparison is followed by a discussion about the biases inherent to our observational strategies and their impacts on statistical estimators (Laigle, Davidzon et al. in preparation). **Chapter** 7 reviews the acquisition of angular momentum in the cosmic web. The main results of this Chapter are based on dark matter simulations and have been published in Laigle, Pichon et al. 2015, MNRAS, 446:2744-2759. This Chapter then extents the measure of angular momentum alignment with filaments to galaxies using the virtual catalogue from the Horizon-AGN simulation (Dubois et al. 2014, MNRAS, 444:1453-1468). **Chapter** 8 reports on the measurement of the dependencies of galaxy properties on their distance to the filaments, relying on projected two-dimensional slices computed with the cosmos2015 catalogue (Laigle, Pichon, McCracken et al. in preparation).
Appendix A describes a method to survey the cosmic web with tomography. This on-going work is based on the Horizon-AGN lightcone and aims at making predictions for upcoming spectroscopic sur-

veys (Laigle, Pichon, Arnouts et al. in preparation). **Appendix** B summarizes other results obtained in collaborations for which my contribution has been less significant.

Chapter 1

Galaxy evolution: overview

Extragalactic astrophysics is now reaching a certain level of maturity. The reason, no doubt, lies in the significant technical and computational progresses which have been made in the 20th century. Our knowledge about the Universe has exponentially increased, not only because of our ability to image the sky in different wavelengths and to correctly extract from the images the quantities of interest (photometry, morphology), but also because of the possibility to confront these observations with complex models, especially with the help of numerical simulations. An exhaustive description of sky survey and numerical simulation history is beyond the scope of this Chapter. Nevertheless, this thesis is intended to bring new insights on galaxy evolution while relying both on deep surveys and simulations. Therefore, it seems appropriate to recall the main milestones of this story, and how they have transformed the way astrophysics is done, for these historical facts are a witness to the motivation and challenges of the study of galaxy evolution. A more complete overview can be find in Djorgovski et al. (2013).

1.1 Sky surveys and multi-wavelength studies of galaxy evolution

1.1.1 Historical background

We often remember Hubble's work on galaxy distance measurements (Hubble, 1925) as the beginning of extragalactic astrophysics, though really the field started implicitly long before that. The 18th and 19th centuries saw the construction of the precursors of the sky catalogues in the modern sense, established by naked eye. Many deep-sky diffuse objects ("Nebulæ") included in these catalogues were in fact nearby bright galaxies, though they were not really counted as such (*e.g.* Messier, 1781; Dreyer, 1888), for it was commonly thought that Milky Way comprised the entire Universe. At this time, studying the sky was mainly studying the stars. Among other interesting contributions, the Herschel family played an important role in introducing stellar statistics, and used it to explore the three-dimensional structure of the Milky Way (Herschel, 1785). It is fair to consider that these works, classifying stars and deep-sky objects and making systematic analysis, laid the foundations of the statistical study of populations of objects, so laborious though they may seem, and prepared the way for our modern astrophysics.

At the beginning of the 20th century, as soon as photography was used in combination with large reflective telescopes, astronomers have been able to record much more information than possible using the unaided eye. Moreover, it became also possible to reach fainter magnitudes as size of telescopes increased. Soon

afterwards Hubble's works on "Nebulæ" (Hubble, 1925) brought a conceptual revolution concerning our view of the Universe by assessing the existence of distant galaxies, the so-called "islands Universe"[1].

From this point, extragalactic astrophysics became a rapidly evolving field. Once having confirmed their nature, Hubble proposed a classification of galaxies in the local Universe, mainly based on their morphology (The "Hubble Tuning Fork", or "Hubble sequence": Hubble, 1936; Sandage, 1961). Galaxies could thus be classified in lenticular, spiral (divided themselves into spiral types with and without bars), elliptical and irregular types. Such a classification became rapidly a widely-used way to distinguish galaxies, and has since been improved and refined (*e.g.* de Vaucouleurs, 1959; van den Bergh, 1960). Importantly, this categorisation is much more general than morphology-based, for it was also quickly established an overall connection between galaxy shapes and other physical properties, in particular their mass or their colours (*e.g.* Holmberg, 1958; Roberts, 1963). Elliptical galaxies were found to be generally more massive, redder and less star-forming than spirals. In this sense, a modern transposition of the Hubble classification generalised in terms of colours is the popular "bimodality" concept: the colour distribution of galaxies clearly highlights a division between a "red sequence" and a "blue sequence" of galaxies (*e.g.* Strateva et al., 2001, in the SDSS). As this colour bimodality can be traced not only with colours but also with stellar masses and star formation rates, a large number of works have tried to explain its origin by introducing the concept of "quenching", describing the shut down of star formation in certain galaxies. The main processes regulating star formation in galaxies is still a highly debated topic (see Section 2.3).
In some ways, the Hubble diagram condenses schematically the observable part of the "galaxy evolution" iceberg, and in so doing, questions the nature of the main processes driving this evolution. Generally speaking, this evolution occurs at two -connected- levels: the evolution of galaxy populations *within* the diagram in a given range of redshift (the related question is what are the processes which drive galaxy transformation from one type to another), and the evolution of the classification itself *as a function of redshift* (in other terms how evolve the relative counts of different galaxy types). The difficulty to address this question of evolution resides in the fact that the galaxy evolution timescale is much larger than a human lifetime. Obviously it is not possible to follow directly the evolution of individual galaxies. A convenient way to circumvent this situation is to draw statistical conclusions on galaxy populations taken as a whole as a function of redshift. Galaxy populations are thus studied with the help of "scaling relations" or very simple one- or two-point statistics. Scaling relations define strong trends within these populations between galaxy physical properties, indicating similar properties for galaxies within a similar type. As an example, the "main sequence" describes the linear relationship between galaxy star formation rates and stellar masses for star-forming galaxies. The counts, luminosity and mass functions are one-point statistics describing simply the number of galaxies per bin of apparent magnitude, rest-frame magnitude or mass. These quantities are related to the general build up of the galaxy stellar mass through cosmic time. Finally two-point statistics, which measure the spatial correlation of galaxy positions, enable to draw conclusions on their environment or to infer on host halo properties via simplifying assumptions such as the so-called halo model. More generally, higher order statistics are of interest, but also non parametric estimators such as topological tools or filament tracers are sometimes better suited to describe distributions which are highly non-gaussian at low redshift (see below Section 1.1.3).

Measuring these quantities requires to disentangle between a real physical evolution from an evolution purely driven by selection effects. This question has started to be satisfactorily addressed in the 1990's, shortly after the emergence of wide-field and high sensitive digital imaging detectors and the advent of new generations large telescopes, which facilitated an easier and systematic exploration of the sky. Fur-

[1] The term "island Universe" was first used by Immanuel Kant in 1755 (Kant, 1755) to describe these distant galaxies, whose existence has been postulated by Thomas Wright in 1750 (Wright, 1750). The latter was writing in 1750:
"[...] *The many cloudy spots, just perceivable by us, as far by without our starry Regions, in which tho' visibly luminous Spaces, no one Star or particular constituent Body can possibly be distinguished; those in all likelihood may be external Creation, bordering upon the known one, too remote for even our telescopes to reach.*"

thermore, they permitted the first glimpse of the Universe at high redshift. This decade has also seen the first very large spectroscopic surveys able to image the filamentary cosmic web.

1.1.2 The advent of large multi-wavelength photometric surveys

The second half of the 20th century has also seen the opening of new regimes of observations in the electromagnetic spectrum. In particular, space telescopes have made possible imaging the sky in the ultra-violet (UV) and mid/far infrared (IR) passbands, with the first major missions being EUVE (1992-2001, Sirk et al., 1997) followed by GALEX (2003-2011, Bianchi and GALEX Team, 1999; Martin and GALEX Science Team, 2003) in the UV, and IRAS (1983, Neugebauer et al., 1984) in the IR domain, allowing in particular to better determine galaxy redshifts, stellar masses and star formation rates (Chapter 5 presents in more detail why these passbands are important for galaxy evolution). Today a main concern in astrophysics is to observe the build-up of the galaxies in the first 1-2 billions years of the Universe, which is why a lot of effort are invested in the IR domain (*e.g.* with the IRAC camera onboard of the *Spitzer* telescope), for it enables to study the rest-frame optical fluxes at very high redshift.
Surveys now extensively image same parts of the sky in different wavelengths, and enable us to build efficiently multi-wavelength galaxy catalogues. The development of automatic pipeline to post-process the images, to extract galaxies (*e.g.* DAOPHOT or SExtractor, see Stetson, 1987; Bertin and Arnouts, 1996, respectively) and to estimate photometric redshifts based on these multi-wavelengths extractions (from color-color diagrams first, see *e.g.* Koo (1985); then with template fitting techniques, see *e.g.* Loh and Spillar (1986) among the first codes) has considerably increased our knowledge of distant galaxies and the accuracy of their measured properties. Photometric redshift software opened the doors of the statistical study of the distant Universe, as it became possible to follow galaxy populations as a function of redshift instead of a function of apparent magnitudes. These photometric surveys have to be coupled with spectroscopic ones for a better accuracy in the determination of the photometric redshift (see Chapter 3). Additionally, the combination of panchromatic surveys with radio or x-ray data revealed many new types of objects, such as the active galactic nuclei (AGN).

Of particular interest in these panchromatic surveys are the near-infrared (NIR) and IR passbands. The first studies of galaxy populations were based on galaxy samples selected at different redshift in different ways, implying that the measured evolutionary effect could possibly be due only to selection effects. To select uniformly galaxy samples at different redshifts and to understand selection effects, it was needed to extend the initially optical wavelength coverage to the NIR (and to the IR for higher redshift). In order to follow the evolution of galaxy populations over a certain range of redshift, we have to be sure that galaxies of the same intrinsic luminosity are visible both at high and low redshifts. In particular, galaxy samples should be selected based on a part of the spectrum which is a good proxy for the galaxy mass. NIR are crucial for this study at redshift above $z \sim 1.3$, for most of the galaxy light produced by the stellar continuum is redshifted into the NIR bandpasses (see Chapter 5).

Today, the strategy of current and upcoming multi-wavelength photometric surveys (coupled with sparser spectroscopic surveys) is two-fold. Firstly, much efforts are spent to enlarge the field of observations at low and intermediate redshift to get large statistical samples and to limit the effect of the cosmic variance in the statistical measurements. Secondly, smaller and selected areas are covered by much deeper surveys to image the high-redshift Universe. My work on COSMOS2015 presented in this thesis (Chapter 5) is in line with this program. In addition, tools to optimise the automated extraction of sources from the images, computing more accurate redshifts at lower cost or optimising the spectroscopic coverage (which is required to help the computation of photometric redshifts) are going through flourishing developments (in particular with the development of machine learning), and are indispensable to fully take advantage of the considerable amount of data provided by large surveys.

1.1.3 The anisotropic galaxy distribution revealed by large spectroscopic surveys

The question of whether galaxies were uniformly distributed or not arose right after the existence of distant galaxies was realised. Early works on this topic described a rather uniform distribution at large scales, while rather clumpy at small scales (Hubble, 1926; Shapley and Ames, 1932; Hubble, 1934). The first turning point was reached by Seldner et al. (1977) and followed by Peebles (1975) who showed clearly that the galaxy distribution is non-uniform, and that the galaxy two-point correlation function could be approximated by a power law.
These results were confronted by the first large spectroscopic surveys, which where able to measure quite accurately the three-dimensional positions of a large number of galaxies. In the 1980's the CfA redshift surveys play an important part in the discovery of the filamentary large-scale structure (*e.g.* Davis et al., 1982; Geller and Huchra, 1989). Galaxies were clearly seen to be distributed along an intricate structure of nodes, filaments, and walls separating voids, the so-called "cosmic web".
Lately, with the arrival of digital detectors, multi-object spectrographs and larger telescopes, mapping the distribution of the galaxies on larger scales and with a less sparse sampling has revealed systematically and quite admirably the filamentary cosmic web, in particular with thanks to the Las Campanas Redshift Survey (Shectman et al., 1996), SDSS survey (York et al., 2000), the 2dFGRS survey (Colless et al., 2001; Colless, 2004) and the 2MASS redshift survey (Huchra et al., 2005).
The distribution of galaxies on large scales was show to be consistent with numerical dark matter simulations (*e.g.* Davis et al., 1985; Jenkins et al., 1998) and a solid theoretical understanding of the cosmic web has been developed at the same time (*e.g.* Bond et al., 1996a; Pogosyan et al., 1998, see Chapter 2).
The idea that environment (first defined in terms of local density) could be one of the driver of the Hubble sequence or bimodality emerged also quicky after the first galaxy classifications (*e.g.* Dressler, 1980; Baldry et al., 2006). Clearly, this question has to be readdressed today by considering the environment in terms of structures (filaments, walls, nodes) rather than in terms of local density. It is starting to be possible with the help of specific tools developed for the purpose of automatically extracting the filamentary patterns (*e.g.* Sousbie, 2011, see Chapter 7 for a description of this tool). Nowadays, the cosmic web has become an active field of research. Nevertheless, despite of the recent progresses in tracing and understanding the structures, this exploration is still in its early stages in particular in the observation side, for we are still missing large and complete surveys with precise redshifts in the distant Universe. Yet, this field of research has certainly a promising future with very recent or upcoming large surveys and facilities, such as VIPERS (Vimos Public Extragalactic Redshift Survey, Garilli et al., 2014; Guzzo et al., 2014), LSST (Large Synoptic Survey Telescope, Ivezic et al., 2008), PFS (Prime Focus Spectrograph, Takada et al., 2014a). Studying the cosmic web "in projection" with photometric redshifts is also an emerging new and promising idea, as it potentially allows to reach more easily higher redshift. This will be further discussed in Chapter 8.

1.2 Numerical simulations

While in recent years observations have been revolutionised by charge-coupled devices (CCDs) and multiplexing, computational cosmology has emerged in parallel as a fruitful field of research. The benefit of numerical simulations for astrophysics is three-fold. First of all, simulations provide a way to implement our current knowledge about galaxy evolution and cosmological model, to create mock observations and to confront them with the real one. The outcome of this comparison allow then to qualify the reliability of the model. Secondly, a well-calibrated simulation, for which virtual observables reproduce correctly the reality, may help to disentangle the processes at works in shaping these observables, and to connect observed galaxies with their past formation history. Finally, predictions for future observations can be successfully

done from simulations and allow instrumentalists to design optimal data gathering techniques.

The evolution of collisionless dark matter has been successfully implemented numerically from the introduction of the first N^2 particle-particle codes in the 1970's (*e.g.* Aarseth et al., 1979; Peebles, 1970; Springel, 2005; Klypin et al., 2011). Modelling the evolution of baryonic matter raises more difficulties, as the physical processes at work are much more complex and more uncertain. Nowadays, making simulations from initial conditions to galaxy formation and evolution can be achieved with two methods: semi-analytical models or hydrodynamical simulations (see *e.g.* Somerville and Davé, 2015, for a review and comparison of these techniques).
Semi-analytical models rely on dark-matter-only simulations. Galaxies are then added in the simulation by the mean of scaling relations, empirically calibrated on physical processes occuring at the scale of the galaxy and on the halo mass (see Baugh, 2006, for a review). They have been quite successful in reproducing the bulk of galaxy properties over a large fraction of cosmic time (*e.g.* White and Frenk, 1991; Baugh, 2006; Benson, 2012), demonstrating that mass is the main underlying parameter driving galactic properties. In addition, they allow to investigate relatively quickly the parameter space (*e.g.* Henriques et al., 2009). However, because of the large number of simplifying assumptions, disentangling the relative importance of the different physical processes at work in galaxy evolution is a degenerate problem, which can reduce the usefulness of these tools. By construction, such models also tend to underestimate the spread in the diversity of observed galaxy properties for a given mass, since they only account for the merging tree of dark halos, ignoring processes within the intergalactic medium (IGM) which do not impact the mass of halos but could impact galaxies.
In contrast to semi-analytical models, hydrodynamical simulations are much more computationaly expensive (*e.g.* more than 6 million CPU hours for the Horizon-AGN (Dubois et al., 2014) simulation). This is because they model consistently from the beginning the co-evolution of baryonic and dark matter, by following both gas and dark matter dynamics. Consequently, they uniquely connect observed galaxy properties with their gas accretion and merger history. Hydrodynamical simulations are able to address a wide range of scales. However, because of their finite resolution, physical processes have also to be modelled below a certain scale (the "subgrid" scale, typically the kiloparsec scale for a cosmological simulation). As such, they inherit some of the very same defects of semi-analytical models, and typically reproduce some scaling laws but not always fully the observed diversity at a given mass. I will provide more details in Chapter 4 concerning the implementation of hydrodynamical simulations when describing the Horizon-AGN simulation.

Despite of their intrinsic limitations, simulations can however now reproduce reasonably well the rich diversity of observed colours, morphologies and star formation activity though the complex combination of internal and external processes (such as feedback, turbulence, smooth accretion, dry minor mergers, mergers) occurring at different scales and times. Yet the exact balance between all these processes and how they affect galaxy evolution and shape galaxy properties is still actively debated (see Section 2.3).

Chapter 2

Galaxy formation and evolution in the ΛCDM model

The distribution of matter in the Universe is determined by cosmology. Galaxies are built on cosmological scales and galaxy populations are thus encoding cosmology in their mass distribution (one-point statistics) and also clustering properties (N-point statistics). In the following I will first for completeness briefly describe the basis of this cosmological background (more details could be found in cosmology or galaxy evolution textbooks, *e.g.* Mo et al., 2010). After a general description of the standard model of cosmology (ΛCDM) in Section 2.1, The growth of structures is sketched in Section 2.2. The last Section 2.3 will describe in more detail the galaxy life-cycle and other debated questions.

2.1 The ΛCDM concordance model of cosmology

An expanding Universe – The two pillars of the standard model of cosmology are the cosmological principle and Einstein's theory of general relativity. The cosmological principle specifies that Universe is homogenous at large scale and isotropic, while Einstein's theory of relativity predicts the structure of the space-time based on its mass-energy distribution. Together these two pillars lead to the Friedmann equations. The resolution of these equations gives the dynamics of the Universe, and in particular outline three successive periods in the history of the Universe: the radiation-dominated era in the primordial Universe, the matter-dominated era, and the contemporary dark-energy-dominated era.
Thus defined, our standard model of cosmology predicts an expanding Universe to be composed of essentially three components: the baryonic matter, the dark matter and the dark energy. Their respective densities are characterised by the cosmological parameters $\mathbf{\Omega_m}$, $\mathbf{\Omega_b}$, $\mathbf{\Omega_\lambda}$. The radiation component has a very low density $\mathbf{\Omega_r}$, and the curvature term $\mathbf{\Omega_K}$ is measured to be consistent with zero, indicating an almost flat Universe. Dark energy refers to the component Λ introduced in Einstein equations to explain the acceleration of cosmic expansion. Dark matter has been introduced as an unknown component to supplement the luminous matter: indeed luminous observed structures do not provide enough material to explain galaxy rotation curves and more dramatically the growth of structures. An additional matter component is thus required.
The cosmic expansion rate is characterised by the Hubble parameter $\mathbf{H} = \dot{a}/a$ ($\mathbf{H_0}$ being the present-day Hubble parameter) where a is the scale factor (involved in the Friedmann equations, this factor quantifies the relative expansion of the Universe). In particular, from this expansion arises a shift (called "redshift",

z) towards the redder wavelengths of the light emitted by distant objects:

$$\frac{\lambda_{\text{observed}}}{\lambda_{\text{emitted}}} = 1 + z = \frac{a(t_{\text{observed}})}{a(t_{\text{emitted}})} \tag{2.1}$$

This redshift, as directly connected to the time at which the light has been emitted, is our method to measure the galaxy recession velocity and in turn to estimate their distance.
As the theory predicts that the Universe is in expansion, it predicts also its past thermal history, up to a certain time at which Universe should have been much more dense and hot: the primordial singularity called the "Big-Bang". Structures are formed from the gravitational growth on cosmological timescales of primordial density fluctuations. These initial fluctuations arises from quantum fluctuations shortly after the Big-Bang. At that time, primordial Universe was so dense that what happened at quantum level was determinant. These fluctuations are Gaussian, meaning that their spatial correlation is entirely encoded in their power spectrum (characterised by its amplitude $\sigma_{\mathbf{8}}$ and shape $\mathbf{n_s}$). The nature of the dark matter is essentially encoded in the shape of this power spectrum.

Cosmological Parameters – In the light of the above, our standard model of cosmology can be finally constrained by a set of cosmological parameters, which describe how the Universe evolves as a whole. Those parameters are: three parameters for the matter-energy content, two for the power spectrum of initial fluctuations (scalar spectral index n_s and curvature fluctuation amplitude σ_8) and additionally the Hubble parameter at present time. The determination of cosmological parameters can be made observationally, either from the pattern of cosmic microwave background anisotropies, or from present observations of galaxy distributions (*e.g.* current matter power spectrum, weak gravitational lensing, Minkowski functionals, void analysis, *etc.*). In this thesis, unless specified otherwise, I will adopt the following values for these parameters:

- matter-energy content: $\mathbf{\Omega_b} = 0.045$, $\mathbf{\Omega_m} = 0.272$, $\mathbf{\Omega}_\lambda = 0.728$
- Hubble parameter $\mathbf{H_0} = 70.4$ km.s^{-1}.Mpc^{-1}
- Power spectrum: $\sigma_{\mathbf{8}} = 0.81$, $\mathbf{n_s}$=0.967.

These values are compatible with the Wilkinson Microwave Anisotropy Probe 7 cosmology (Komatsu et al., 2011) and with those of the recent Planck results within a 10 per cent relative variation (Planck Collaboration et al., 2014).

Dark matter – The cold dark matter is the most accepted model for dark matter in our scenario of structure formation, as it allows to explain the observed galaxy distribution on large scale. The qualifier "cold" is employed because dark matter has initially negligible thermal velocity. The minimal scale at which structures collapse functions as an indicator of the temperature of the dark matter particles at decoupling: cold, warm or hot.
In contrast to cold dark matter, the depreciated warm dark matter model postulates that the initial perturbations below a certain mass cannot collapse because of the high-velocity of dark matter particles, which has as a consequence that the smallest structures are fairly large. A visual result is that the high frequencies in the matter distribution are smoothed. Because of that and for visualisation purpose only, I nonetheless use in Chapter 7 warm dark matter simulation as a proxy for cold dark matter for it allows us to visualise what is happening on large scales.
What is the nature of dark matter is still an open question in particle physics. Dark matter escapes indeed

direct detection for it is collisionless: the probability that two dark matter particles interact is close to zero.

2.2 Structure formation

Spherical collapse – Structures form from the gravitational amplification of tiny primordial fluctuations in the density field. An overdensity is essentially characterised by its density contrast δ with respect to the critical density of the Universe ρ_c. While the density contrast is small, the growth occurs linearly, and can be described from the Friedmann equations. This description can be extent further with the help of the perturbation theory until the homogeneity becomes apparent, for $\delta \sim 1$. When the initial critical density is reached, overdensities decouple from the expansion of the Universe and collapse to self-gravitating virialized halos, entering thus in a highly non-linear regime. These steps in structure formation can be described by simple and quite successful models: the spherical collapse model (Gunn and Gott, 1972) describes this collapse in a spherical symmetry, solving the Friedmann equations by considering the overdensity as a small spherical overdense universe. With the formalism of Press and Schechter (1974) it is possible to go one step further and estimate the growth of structures as a function of the mass of halos, and then to get the coresponding mass function of collapsed objects. After collapse and virialization, halos continue to grow in mass by accretion and merging. Yet it is not possible within this simple model to predict the growth of structures in details beyond that point: cosmological simulations are essential (see Chapter 4).

Lagrangian versus Eulerian description – The growth of structures may be followed either in a Lagrangian or in a Eulerian framework. In the Lagrangian framework, the observer follows the fluid particles (belonging to a small region of the initial dark matter fluid) through space and time. In the space-fixed Eulerian frame, the observer follows the flow of particles through a volume fixed in space.
The advantage of the Lagrangian description is that it keeps tracks of the particle trajectories. In this sense, it allows us to follow processes occurring close to the initial conditions, such as initial angular momentum acquisition. The Eulerian description is closer to the observer's point of view when following highly non-linear processes today. It studies populations of particles which are at the same place at a given time.
Both descriptions are complementary and are useful theoretically and phenomenologically, to either describe the time evolution of a given population or the observed population at given redshift. At early time the mapping between Eulerian and Lagrangian description may be done through the Zel'dovich approximation (see below). However this mapping is limited to weakly non linear regime and breaks down when particle trajectories cross. In this thesis, I will make use of both descriptions, in particular in Chapter 7, in which I investigate angular momentum acquisition in the cosmic web.

Cosmic web – At the end of the 20th century, large spectroscopic surveys have revealed a large scale filamentary pattern in the galaxy distribution: an intricate structure of nodes, filaments, and walls separating voids (see Chapter 1). This anisotropic environment is the natural metric in which smaller structures form and evolve, and, as such, is expected to play an important role in shaping their properties (in particular in terms of dynamics).
The cosmic web constitutes a transitional stage between linear and non-linear structures (Cautun et al., 2014) arising from the gravitational collapse of initial anisotropies. Cosmic web growth is well described in the Zel'dovich approximation (the so-called "pancake model" Zel'dovich, 1970), which applies on relatively large scales (long wavelength fluctuation mode), while the above described spherical collapse applies bet-

ter on the particular behaviour of isolated spherical density perturbations at the halo scale (Bond et al., 1996b).

The Zel'dovich approximation relies on a Lagrangian description of the growth of structures. It maps the initial positions $\mathbf{q}$ of the particles in the dark matter fluid with their positions $\mathbf{x}$ at time t through the displacement field $\mathbf{s}$:

$$\mathbf{x}(\mathbf{q}, D_+(t)) = \mathbf{q} + D_+(t)\mathbf{s}(\mathbf{q}). \tag{2.2}$$

In this equation, $D_+(t)$ is the linear density growth factor which depends on the cosmological parameters. The displacement field is potential and can be expressed as $\mathbf{s}(\mathbf{q}) = \nabla\mathbf{\Psi}(\mathbf{q})$. The motion of particles is caused by their mutual self-gravity and the displacement potential Ψ is directly proportional to the gravitational potential ϕ. In turn, the deformation tensor, which is defined as $d_{ij} = \partial s_i/\partial q_j$, is proportional to the Hessian of the gravitational potential. The number of positive eigenvalues of the tidal tensor corresponds directly to the number of axes along which matter will contract.

Hence, the main idea of the Zel'dovich approximation is that the initial perturbation will enter in the non-linear regime faster along one of the three directions of the deformation tensor, collapse amplifying initial anisotropies and leading to highly anisotropic structures. The compression of matter occurs first along one direction to form walls ("pancakes"), then along a second direction to form filaments, and finally along the third to form nodes, depending on the number of positive eigenvalues and their respective amplitudes.

Note that if two trajectories cross, the mapping $\mathbf{q} \to \mathbf{x}$ becomes ambiguous: this is a shell-crossing and the Zel'dovich approximation is in principle no longer valid.

Vorticity is defined as the curl of the velocity field. An important point to be noted is that in the linear regime primordial vorticity in the cosmic web decays as $\omega \propto a^{-1}$ where a is the scale factor, implying that the flows are laminar and curl-free before shell-crossing. However, using the Zel'dovich approximation, Pichon and Bernardeau (1999) showed how vorticity is generated in the cosmic web during the first shell-crossing, namely in the walls. This vorticity is predicted to lie in the plane of the walls. We can expect that the second shell-crossing (to form filaments) will generate vorticity aligned with that filaments. This point will be investigated in Chapter 7 in relation with the angular momentum acquisition.

Tracing the cosmic web with the skeleton – To study the distribution of matter, one has to robustly extract the filamentary structure of the matter, the so-called "skeleton". The extraction method, to be rigorous, should be topologically motivated so as to preserve the connectivity of the web.

One of the simplest way to extract the filamentary structure is to classify the different components (voids, walls, filaments and nodes) based either on the Hessian of the density field or on the tidal tensor (Pogosyan et al., 1996; Novikov et al., 2006; Aragón-Calvo et al., 2007; Hahn et al., 2007b; Sousbie et al., 2008; Forero-Romero et al., 2009). As the number of positive eigenvalues of these tensors corresponds to the number of axes along which matter collapses, it is directly connected to the morphology of structures.

Morse theory (Jost et al., 1995) provides an other simple way to extract the skeleton (Novikov et al., 2006; Sousbie et al., 2008), by defining walls at the interface of voids and filaments at the interface of walls. Let us consider first a two-dimensional density field. Intuitively, the filaments of this field are the lines connecting maxima through saddle points along the "crest lines". Let us define the "void patch" as the regions of the density field converging towards a same minimum along the gradient field lines. The filaments are then rigorously the borders of these patches. These borders connect maxima passing through saddle points. Hence the skeleton can then be drawn from saddle points towards maxima following the gradient of the density field $\nabla\rho$. Critical points are defined by $\nabla\rho = 0$. The number of positive eigenvalues of the Hessian matrix defines the type of the critical point. Maxima have all their eigenvalues negatives and minima all their eigenvalues positives. In two dimension saddle points have one positive and one negative eigenvalues. When adding a third dimension of the density field, one can define the filament-type saddle points and the wall-type saddle points, which have respectively one and two negative eigenvalues.

Nevertheless, important difficulties are specific to the computation of the skeleton in galaxy catalogues, which are discrete, sparse and noisy. To overcome these difficulties, Sousbie (2011) proposed a robust

algorithm based on the concept of persistence and called DisPerSE. I used it in this thesis both on observations and simulations and I will describe it in Chapter 4.

2.3 The galaxy life-cycle and associated open questions

The growth of dark matter halos in our cosmological framework is now reasonably well understood and can be successfully implemented in simulations. However galaxy formation and evolution involves more complex processes and up to now leaves many unanswered questions and shady areas. Without aspiring to completeness, I attempt below to summarise our current knowledge about the galaxy life-cycle, while emphasising current questions such as the role of quenching, the nature of gas accretion, the role of feedback and the role of environment on small and larger scales. These are key questions to understand how many parameters effectively drive galaxy mass and angular momentum acquisition.

Star formation and quenching – Galaxies form most of their stellar mass through gas consumption (White and Rees, 1978), during the first half of cosmic history. Observational data converge towards a model in which the cosmic star formation history peaked at $z \sim 2$ and then decreases exponentially (Madau and Dickinson, 2014). It is now clear that by $z \sim 1$ most of the mass has already been assembled into galaxies. At high redshift, star formation occurs vigorously in blue massive galaxies and with cosmic time the peak of star formation activity shifts to progressively low-mass objects at low redshift (*e.g.* Cowie et al., 1996; Arnouts et al., 2007; Pozzetti et al., 2007; Noeske et al., 2007). In the local Universe, the observed "bimodality" (Kauffmann et al., 2003) divides galaxy populations between a "red sequence" of massive and passive galaxies and a "blue sequence" of active and star-forming objects, while the gap between the "red" and the "blue" sequence becomes more pronounced with time. This observational fact would suggest that some processes are slowing down star formation activity in massive systems, sustaining the argument of quenching. However, despite the success of phenomenological models in reproducing at least some of these observational trends (Peng et al., 2010), the precise physical mechanisms of this quenching process remain a topic of debate.
Cold gas is is the basic fuel for galaxies to form stars. Hence, a better understanding of how gas accretion feeds galaxies and of the effect of possible outflows – which could stop the gas supply in galaxies – are crucial to explain both the peak of star formation at high redshift and its quenching at lower redshift.

The modes of gas accretion – The observed tight relation between star formation rate (SFR) and galaxy mass for star-forming galaxies ("main sequence") found at $0 < z < 2$ (*e.g.* Brinchmann et al., 2004; Daddi et al., 2007) is reproduced in semi-analytical models and partly in hydrodynamical simulations (see *e.g.* Chapter 6) and is now shown to exist up to $z \sim 6.5$ (*e.g.* Steinhardt et al., 2014; Salmon et al., 2015) and down to $\log M/\mathrm{M}_\odot \sim 9.4$ (Kochiashvili et al., 2015), although the sample selection effects, in addition to the different methods used to compute stellar masses and star formation rates (SFR) and the dust uncertainties may still produce partially different results at high redshift (Lee et al., 2012). This SFR-stellar mass relation nonetheless could suggest that most of the mass assembly occurs essentially in a smooth mode (*e.g.* Elbaz et al., 2011; Rodighiero et al., 2011), compared to a clumpy accretion mainly driven by mergers (see also L'Huillier et al., 2012). However the privileged mode of smooth gas accretion remains unclear.
The conventional model relied on the "hot mode" accretion scenario, in which the infalling gas is shock-heated at the virial radius and then radiatively cools starting from the central part and forming centrifugally supported disks (*e.g.* Rees and Ostriker, 1977; White and Rees, 1978; Maller and Bullock, 2004). However, recent hydrodynamical simulations now suggest that most of the gas is accreted directly from

cold dense filaments without being shock-heated (Binney, 1977; Katz et al., 2003a; Kereš et al., 2005a; Dekel et al., 2009; Faucher-Giguère and Kereš, 2011; Hobbs et al., 2015) at least for low-mass haloes or at high redshift. In this view, the anisotropic large-scale environment of galaxies is therefore likely to play an important role as it literally drives such cold flow accretion within the galaxy. The mode of gas accretion has important consequences in terms of star formation activity but also for angular momentum build-up. In the hot-mode, the time required for gas to cool implies a delay in the star formation, while in the cold-mode, star formation is expected to proceed more rapidly and more efficiently (Dekel and Birnboim, 2006). Furthermore, in the cold-mode paradigm, infalling gas tends to have very high specific angular momentum, which should strongly impacts the angular momentum acquisition in galaxies (*e.g.* Pichon et al., 2011), independently of the small-scale baryonic physics (*e.g.* Stewart et al., 2016, for a comparison on hydrodynamical simulations using different subgrid physics and code implementations).

Observationally the infall of cold gas tends to escape direct detection (e.g. Kimm et al., 2011b). However Rosdahl and Blaizot (2012) showed that cold accretion inflows should be observable via direct Lyman-α emission, which becomes now possible with integral-field spectrographs such as the Multi Unit Spectroscopic Explorer (MUSE, Bacon et al., 2010) or the Keck Cosmic Web Imager (K-CWI, Martin et al., 2010; Morrissey et al., 2012). A growing number of individual detections of inflows in several galaxies have been made thanks to absorption (*e.g.* Giavalisco et al., 2011; Ribaudo et al., 2011; Kacprzak et al., 2012; Crighton et al., 2013) or emission features (*e.g.* Cantalupo et al., 2014; Martin et al., 2016), often compatible with cold-mode accretion. Detectability should increase with redshift as the transition from hot- to cold- mode of accretion is crossed. All these phenomena are most intense in the 100-300 physical kpc region around galaxies – the circumgalactic medium feed by cold flows.

The role of feedback in regulating star formation – The basic comparison of the galaxy luminosity function to the dark matter halo mass function highlights a shallower faint-end slope and sharper cut-off at the massive end. This suggests that other mass-dependent effects are responsible for the lack of low-mass and high-mass galaxies compared to the number of dark haloes in the same mass ranges (*e.g.* Benson et al., 2003). To explain this discrepancy, feedback processes are expected to play a role in regulating star formation by perturbing, heating or removing the cold gas from the galaxies.

In low-mass galaxies, stellar feedback strongly impacts the star formation history (*e.g.* Dekel and Silk, 1986; Larson, 1974), explaining why galaxies are so inefficient to turn the gas into stars (*e.g.* Hopkins et al., 2014; Kimm et al., 2015). In addition, feedback participates in the ejection of metal-rich gas outside the galaxy (*e.g.* Shen et al., 2012), which likely shapes the evolution of the galaxy mass-metallicity relation (*e.g.* Maiolino et al., 2008; Zahid et al., 2013) and the enrichment of the intergalatic medium (*e.g.* Aguirre et al., 2001).

In high mass galaxies, active galactic nuclei (AGN) are probably the most dominant mechanism of star formation quenching (*e.g.* Croton et al., 2006; Hopkins et al., 2006; Cattaneo et al., 2007). However, some recent studies have investigated possible mechanisms of triggering of the star formation by AGN feedback (*e.g.* Gaibler et al., 2012; Bieri et al., 2015).

Stellar and AGN feedback are thus expected to directly shape galaxy stellar mass growth and in turn their observable properties (see Chapter 6). The complex physics involving feedback is still not fully understood in detail, and modelling it correctly remains difficult at the subgrid scales in cosmological simulations. As such, stellar feedback is often an adjustment parameter in the simulations in order to reproduce observable properties and its simulated strength is typically much stronger than direct observations of winds suggest.

The role of environment at the scale of the dark halo – Investigating the relationship between the integrated stellar properties of galaxies (in particular, stellar mass, star formation rate, and star formation history) and their dark matter environment has been an important subject of research during the last decade. In fact, the gas accretion mode is expected to be intimately connected to the halo mass and,

depending on the dominant scenario, the galaxy star formation history (SFH) will be different due to the cooling delay implied by the hot-mode accretion. In practice, the stellar-to-halo mass relation is derived statistically by comparing galaxy clustering measurements with predictions from the phenomenological halo model (*e.g.* Cooray and Sheth, 2002). The halo model makes prediction on the expected two-point function of galaxies of a given mass, while relying on perturbation theory at the large separation end, and on the autocorrelation of dark halo profiles (Navarro-Frenk-White profile) calibrated on simulations. This relation has already been extensively studied (*e.g.* Béthermin et al., 2014; McCracken et al., 2015; Coupon et al., 2015). These studies show that the stellar-to-halo-mass relation is well described by a double power law with a transition mass for which the star formation is the most efficient. Interestingly, this transition mass evolves very little with time up to $z \sim 2$ (*e.g.* McCracken et al., 2015). This relationship is worth extending at higher redshift before the peak of star formation activity, which requires sufficiently large and deep data sets, such as the COSMOS2015 catalogue presented in Chapter 5.

The role of environment at larger scales – At larger scale, some observational evidence suggest that environment play a role in shaping galaxy properties, though its relative importance compared to other internal processes is still not clear. Briefly, galaxies are found to be more massive and much less star-forming in high-density regions relative to low-density regions (*e.g.* Dressler, 1980; Kauffmann et al., 2004) which is consistent with the clustering measurements of UV-selected galaxies (Heinis et al., 2007; Milliard et al., 2007). Using local samples, Peng et al. (2010) have demonstrated that quenching of star formation activity can be separated into environmental (density dependent) and internal (galaxy mass related) effects, suggesting that nature and nurture both act in shaping galaxy properties.
However, the distinction between mass-dependent and density-dependent effects is not completely understood, for there is also a connection between galaxy mass and density. In any case, environmental effects should be classified as "local" and "large-scale" effects. While the local environmental effects include all the processes enabling quenching of gas infall or gas removal from galaxies (strangulation, stripping, *etc.*), the "large-scale" effect is concerned with the intrinsic anisotropic nature of the gas accretion following the filaments. This large-scale effect is tightly connected with the cold-mode paradigm presented above. Recent theoretical works have predicted that there is a significant connection between the dynamics within the intrinsically anisotropic large-scale structure on the one hand, which drives the cold gas accretion, and the physical properties of the galaxies embedded in them on the other hand (Pichon et al., 2011; Codis et al., 2012; Dubois et al., 2014; Welker et al., 2014; Laigle et al., 2015; Codis et al., 2015). This will be the topic of my work presented in Chapters 7 and 8.

Part II.

Methods

Chapter 3

From light to galaxy properties and vice-versa

Part of the work presented in this thesis involves the comparison of galaxy properties from observations with galaxy properties from simulations to draw lessons for our galaxy evolution model (Chapter 6). However this comparison is challenging, because observations trace light (or ideally three-dimensional spectroscopy), while simulations trace mass (and more generally phase space distribution), which means that both tracers are not directly comparable. Let us focus on light and mass. To achieve this comparison, I could either compare the mass or star formation rate (a post-processed quantity in the observations but an intrinsic one in the simulations), or the light through, for instance, colour-colour diagram or luminosity functions (which is a post-processed quantity in the simulations, but an intrinsic one in the observations[1]). These comparisons are however not equivalent. For physical interpretation, the former is more convenient. However, the operation which consists in computing the mass and star formation history from the light is an inverse ill-conditionned and degenerate problem: similar spectral profiles can lead to dramatically different mass and star formation history solutions. That is why from the point of view of sound data analysis, the latter is always preferred. In any case, I need to be able either to *interpret* the light *in terms of* stellar content, or to *model* the light *from* the stellar and gas contents.

The galaxy SED encodes information in terms of stellar mass, stellar formation history, dust, and chemical abundance which are not trivially distinguishable. Once extracted the observed light of a galaxy (either the photometry or the full Spectral Energy Distribution, SED in the following), one of the main challenges for an observer is to interpret it in terms of mass, which requires especially a good modelling of stellar evolution, of the initial mass function and of dust extinction.

I present below the basic outline of this model, paying particular attention to its limitations and weaknesses. Section 3.1 recalls the required notions of stellar evolution (which is also a good introduction for Section 4.1), as well as the working of evolutionary stellar population synthesis codes. Section 3.2 dissects the galaxy spectrum to give a broad idea on how to interpret it. Finally, the computation of redshifts and physical parameters from the photometry by SED fitting is detailed in Section 3.3.

Further details about star formation and stellar evolution, spectral synthesis models galaxy photometry and physical property determinations are found *e.g.* in Mo et al. (2010), Madau and Dickinson (2014) and Courteau et al. (2014).

[1] The reality is slightly more complex; For example, absolute magnitudes and luminosities are already a "post-processed" quantity in the observations, in the sense they require already the calculation of the redshift by, *e.g.*, SED fitting.

3.1 Internal galaxy physics and spectral synthesis

3.1.1 Stellar evolution

In order to understand how to compute galaxy properties from their light, one must first become acquainted with the stellar evolution processes. This Section is required to understand Section 3.1, in which I will address the question of the modelling of stellar evolution with spectral synthesis, and Section 4.2, in which I will present the implementation of stellar evolution in the particular case of the Horizon-AGN hydrodynamical simulation.

Star formation Stars form as star clusters in the dense core of giant molecular clouds, which are clumpy structures of cold gas and dust, forming through shocks and instabilities in the interstellar medium (ISM). Initial collapse of these gas overdensities (of typical masses greater than 10^5 $\mathrm{M}_\odot$, Blitz, 1993) occurs as soon as gravity overcomes pressure (Jeans, 1902). Some additional processes, acting at different scales, such as filamentary accretion of cold gas in the ISM, winds from protostars, active galactic nuclei (AGN) and supernovae-driven turbulence can accelerate or slow down star formation (see below). The resulting star formation efficiency is of about 2% (Kennicutt, 1998), meaning that there is a (quite low) constant fraction of molecular gas that is turned into stars per free-fall timescale, although recent studies show it can vary and be much higher in the densest molecular clouds (up to 30%, *e.g.* Murray, 2011). Empirically, star formation rate obeys to the empirical Kennicutt-Schmidt law (Kennicutt, 1998):

$$\dot{\rho}_* = \epsilon_* \rho_\mathrm{g} / t_\mathrm{ff} \,, \tag{3.1}$$

where $\dot{\rho}_*$ is the star formation rate mass density, ρ_g the gas mass density, ϵ_* the star formation efficiency, and t_ff the local free-fall time of the gas.

The Initial mass function – The initial mass function (IMF) describes the distribution of star masses in a given region following a starbust[2]. Low-mass stars are much more numerous and are an important part of the mass budget, but massive stars are much more luminous and consequently dominate the light budget. Correctly modelling the shape of the IMF is therefore important: it allows to infer the relative abundance of stars in the different mass ranges knowing only the amount of bright massive stars.
The functional form of the IMF $\phi(m)$ describes the number of stars born with a mass in the range $[m, m + \mathrm{d}m]$. There are few commonly admitted functional forms for the IMF. The oldest is the Salpeter (1955) IMF defined in the range $0.4 < M_*/\mathrm{M}_\odot < 50$:

$$\phi(m) = A M^{-\alpha} \,, \tag{3.2}$$

where $\alpha = 2.35$ and A a constant. Kroupa et al. (2013) proposed a broken power law, with $\alpha = 2.7$ in $1 < M_*/\mathrm{M}_\odot < 100$, $\alpha = 2.3$ in $0.5 < M_*/\mathrm{M}_\odot < 1$, $\alpha = 1.3$ in $0.08 < M_*/\mathrm{M}_\odot < 0.5$, and $\alpha = 0.3$ in $0.01 < M_*/\mathrm{M}_\odot < 0.08$.
Chabrier (2003) introduced a power law with $\alpha = 2.3$ above 1 $\mathrm{M}_\odot$ but found a lognormal form below 1 $\mathrm{M}_\odot$:

$$\phi(m) = A \exp^{-(\log m - \log m_c)^2 / 2\sigma^2} / m \,, \tag{3.3}$$

with $m_c = 0.08$ $\mathrm{M}_\odot$ and $\sigma = 0.69$.
There is generally an agreement about the power-law form for stars above 1 $\mathrm{M}_\odot$, however both the low-mass end and the limits of the validity regime are uncertain. For this reason, when comparing a set of

[2] All these stars are born at the same time.

observational data in terms of stellar content (e.g stellar mass), particular care has to be taken on which IMF has been chosen for the computation of the mass, for it can lead to a significant offset.
Note that the IMF is classically considered as universal (the same within a galaxy, and from one galaxy to another, whatever the redshift), though this assumption is an oversimplification (see *e.g.* the review by Bastian et al., 2010). In this thesis I will use both Salpeter (1955) and Chabrier (2003) IMF.

Stellar evolution – Stellar evolution is driven by a balance between gravity and pressure from nuclear reactions. Light elements, starting from hydrogen, are progressively transformed into heavier elements (up to iron for the most massive stars). Once formed, stars evolve along the so-called "stellar main sequence" (recognisable in the the Hertzprung-Russell diagram, see Figure 3.1) corresponding to the time during the hydrogen core is burnt in helium. Massive stars burn their core much quickly, and their lifetime is consequently shorter than low-mass stars. The next steps of star evolution after the "main sequence" (i.e the production or not of heavier elements) are mainly governed by their initial mass. During their different phases of evolution, stellar surface temperatures (hence their colours), luminosity and sizes change. Mass, age, and metallicity[3] distributions of stars shape the total SED of the galaxy in which they are embedded. Depending on their initial mass, the final fate of the stars can be either planetary nebulae or white dwarves for low-mass stars, or supernovae, neutron stars or black holes for more massive stars.

Stellar mass losses and feedback – Stars are not closed systems. Some feedback processes are responsible for the ejection of a large amount of energy, stellar mass and metals in the ISM. They impact it thermally and mechanically at different scales. This feedback includes protostellar outflows, stellar winds (occurring mainly during the main-sequence evolution), UV radiations emitted by supergiants, and Type Ia and Type II supernovae explosions in the final phase of massive stars evolution. Gravitational interactions of binary systems can also be the cause for mass transfer inducing type Ia supernovae explosions (*e.g.* Woosley et al., 2002, and references herein).
This feedback may slow down star formation (*e.g.* Kim et al., 2013; Kimm et al., 2015) locally by heating and/or dispersing the dense ambient gas. This is a possible explanation for the relatively low star formation efficiency in galaxies (see above). They may also possibly trigger star formation (*e.g.* Whitworth and Francis, 2002; Palouš, 2014; Liu et al., 2015) on relatively short timescales, although in observations it is not easy to distinguish between enhanced and spontaneous star formation.
Low mass stars (< 8 $M_\odot$) primarily lose mass through stellar winds in their main-sequence evolution. However most of the mass loss occurs in the late phases of their evolution, in particular during the asymptotic giant branch[4] (ABG) phase and during the formation of planetary nebulae. Correctly modelling the AGB phase is important despite its short timescale because their high luminosity in the near-infrared (NIR) may contribute significantly to the SED of the galaxy (see below).
Massive stars (those which are massive enough to explode as type II supernovae, typically with $M > 8M_\odot$ and generally less than 1% of the star populations) encounter much more important mass losses through fast winds during the main-sequence evolution and lately through supernovae explosions (probably the dominant stellar feedback processes). Supernovae explosions create a hot over-pressurised bubble expanding in the ambient medium, driving a shock wave and injecting energy, mass and momentum in the ISM. These explosions deeply impact their surroundings in terms of turbulence but also because heavier elements are released in the ISM. Taking into account this chemical enrichment is important to model the evolution of the generation of stellar populations in the galaxy. Indeed, a population born in a metal-rich environment compared to a low metallicity environment will be redder (*e.g.* Conroy et al., 2009, for the effect on the galaxy SED). The mass losses will be generally higher. Furthermore, a new stellar population

[3] Metals are defined as elements heavier than helium. Metallicity is the mass abundance of these elements.
[4] The AGB phase is a late period of low-mass stellar evolution preceding the planetary nebulae formation. This period can be divided in the early AGB (E-AGB) and thermally pulsing AGB (TP-AGB) phases.

born in an enriched ISM could have a modified IMF (Kroupa, 2008), but the dependence of the IMF on metallicity of the ISM is poorly constrained.

When modelling stellar evolution at the scale of a stellar population (through evolutionary stellar population synthesis models) or even larger scale (*e.g.* in semi-analytical models or low-resolution cosmological simulations), we use simple recipes to take into account the collective effect of mass losses and feedback. The impact of mass loss on stellar evolution in massive stars is complex. Mass loss rates implemented in stellar synthesis models are quite uncertain, however, this mass loss is expected to affect properties and final fate of the stars (see the review by Smith, 2014), and, thus, impact the ISM and more largely the properties of galaxies.

Interstellar dust – Dust production from aging star populations is an important element and source of uncertainties in modelling the galactic spectrum. Correctly accounting for dust is essential for two -connected- reasons: 1) dust plays a crucial role in the interstellar chemistry and therefore the star formation in galaxies, and 2) it shapes the galaxy spectrum, determining in turn what the galaxy looks like (*e.g.* see the review by Draine, 2003, 2004). In Chapter 6 I show that taking into account (or not) dust extinction in simulated galaxies significantly impacts how well their properties agree with observational data.
These microscopic dust grains constitute less than 1% of the interstellar matter. Dust is formed in the late stages of stellar evolution and transported in the ISM through stellar feedback. This feedback, together with grain collisions and thermal sputtering, are also responsible for grain erosion and destruction. The grain growth depends on the ambient gas properties (density, temperature). Dust grains are solid particles predominantly composed of graphite and silicate, and maybe diamond, iron or ice, although there are still large uncertainties on their chemical composition.
Our knowledge about dust is mainly provided by the information which can be inferred from starlight. Dust absorbs and scatters more easily stellar light at short wavelengths, essentially in the UV, and reradiates the absorbed energy in the FIR (*e.g.* Mathis, 1990, for a review). By reducing the UV radiation, dust limits the molecular dissociation and catalyses heavily the formation of H_2 molecules in the site of star forming regions. Dust is also responsible for the depletion of chemical elements, by binding metals in dust grains. Finally it regulates the temperature of the ISM.
Our understanding of the effect of dust is still incomplete. When modelling galaxy light with spectral synthesis models, in general only the absorption effect of dust is taken into account through attenuation laws empirically calibrated, for instance on the Large Magellanic Cloud (*e.g.* Howarth, 1983), the Small Magellanic Cloud (*e.g.* Prevot et al., 1984), the Milky Way (*e.g.* Cardelli et al., 1989) or nearby galaxies (*e.g.* Calzetti et al., 2000). These attenuation laws do not account for the geometrical distribution of the dust. They are therefore limited, though useful because very simple. Very promising codes manage to treat dust self-consistently, either performing a full radiative transfer or mixing radiative transfer and simple prescriptions (Chevallard et al., 2013; Chevallard and Charlot, 2016). On the hydrodynamical simulation side (to compute the photometry of simulated galaxies), sophisticated radiative transfer codes are physically consistent (*e.g.* Sunrise, Jonsson, 2006) but computationally expensive, that is why dust attenuation is generally calibrated from the gas metallicity distribution. In Chapter 4 I present such implementation for the computation of a virtual photometric catalog from the Horizon-AGN simulation. Some very recent works also propose interesting self-consistent ("end-to-end") implementations of the effect of dust in hydrodynamical simulations (Bekki, 2015; McKinnon et al., 2016a,b).

Stellar light – The light of stars depends on their masses and the stages of their evolution (age and metallicity). Some stars do not contribute at all to the light but still to the stellar mass budget (this is the case for white dwarfs, neutron stars and black holes), and others are excessively bright in certain

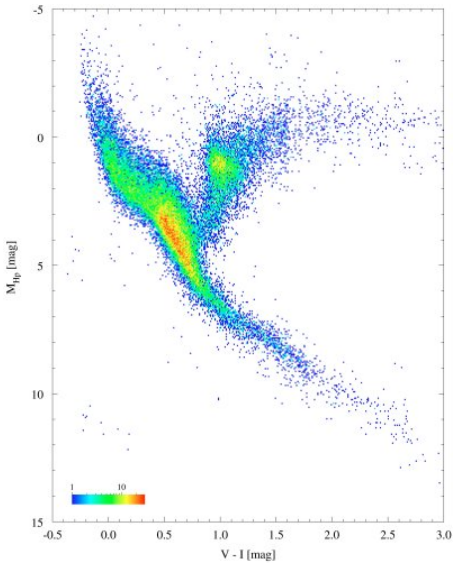

Figure 3.1: The Hertzprung-Russell diagram displays the stars luminosity as a function of their temperatures. Here the diagram is based on Hipparcos data (ESA) and generated using 41 453 stars. The colour scale reflects the number of stars. The V-I colour traces the temperature. In this diagram, the different phases of stellar evolution are clearly visible. The "main sequence" is populated by young stars. Giants and supergiants lie in the above branch. Isolated stars below the "main sequence" are white dwarfs. Figure from the esa website (see also Perryman et al., 1995).

wavelengths without contributing much to the mass budget (for instance the TP-AGB stars which are very bright in the NIR).
Young massive stars (on the "main sequence") strongly dominate the blue/optical spectral range. Red giant stars and AGB phase stars are mostly brighter in the red part of the spectrum.
Metal-rich stars are cooler and fainter, due to increasing opacity in the stellar photosphere, meaning their mass-to-light ratio M_*/L will be higher. These metal-rich stars will be bluer than metal-poor stars in the early-phase of their evolution, but redder in the red giant phase. Since age and metallicity contribute both to make the galaxy SED redder, it leads to the "age-metallicity degeneracy" (Worthey, 1994). Finally, as emphasised above, the role of dust is not negligible: it attenuates the light in blue bands, but re-emits it in IR.

Active galactic nuclei – Active galactic nuclei (AGN) are found in the centre of most massive galaxies and are characterised by a strong emission over all the electromagnetic spectrum. Although AGN constitute a whole topic by themselves, they are mentioned here for completeness, for they are expected to play an important role in the regulation of star formation either in a positive or negative sense (see Section 2.3). AGN feedback is thought to quench star formation by heating and dispersing the gas. Depending on the accretion rate, they emit radiation (during the quasar-mode corresponding to high accretion rates) or jets (during the radio-mode, with low accretion rates). It is still not clear which mode dominates at which time of the AGN evolution. The negative role of AGN feedback in regulating star formation has been known for a long time now (*e.g.* Bower et al., 2006; Sijacki et al., 2007; Dubois et al., 2010). In Chapter 6, I illustrate this role by comparing the mass functions in the Horizon-AGN (with AGN feedback) and Horizon-noAGN (without AGN feedback) simulations.
More recent works investigate conversely how AGN feedback enhances star formation in galaxies. Analytically studied (Silk and Norman, 2009), the possible triggering of star formation by AGN jets have been also shown in simulations (*e.g.* Gaibler et al., 2012; Bieri et al., 2015), and are sustained by few observational facts (Salomé et al., 2015). AGN activity could explain the very high star formation rate in high-redshift galaxies (Bieri et al., 2016).

3.1.2 Modelling stellar evolution

A single stellar population (SSP) is defined as a population of stars born at the same time from the same gas cloud (and therefore sharing the same initial metallicity). SSP models are considered as a quite valid approximation of star clusters (Kuncarayakti et al., 2016) although there is not a one-to-one equivalence.
A galaxy can be modelled as a set of single stellar populations of different ages and metallicities embedded in the ISM. As a result, the galaxy SED will be the superposition of the SED of individual SSPs. Modelling galaxy SEDs requires an evolutionary stellar population synthesis model for *the SED of individual SSP*, but also models for the *star formation history* and the *chemical enrichment* of the galaxy. I detail below the main aspects of this modelling. I will present in Chapter 6 how simplifying assumptions related to star formation history and chemical enrichment are responsible for biases in the resulting masses and star formation rates computed by SED fitting. Hydrodynamical simulations which implement stellar evolution at the subgrid scale also use stellar population synthesis as described below. However they do not need to make assumptions on the star formation history as it naturally arises in the simulation.
A review of stellar population synthesis techniques can be found in Conroy (2013).

Stellar population synthesis Modelling the evolution of a given SSP through stellar population synthesis involves essentially predicting at a given age the mass loss and the emitted SED, knowing the mass, initial metallicity and initial mass function of the stellar population. Simply put, this can be achieved by building up isochrones[5] in following the stellar evolution tracks in the Hertzprung-Russel diagram[6] (Figure 3.1) of a SSP with a given IMF and initial metallicity (Bruzual A. and Charlot, 1993). The main ingredients for the stellar population synthesis are then:
i) the initial mass function;
ii) the stellar evolution prescriptions (tracks and "isochrones"). Various prescriptions exist such as the Padova (Bertelli et al., 1994; Marigo and Girardi, 2007; Marigo et al., 2008) or Geneva (Lejeune and Schaerer, 2001) tracks;
iii) the parametrisation for the associated stellar mass losses which affect the different stages of stellar evolution;
iv) the stellar spectral libraries allowing to associate a SED to a position in the Herzprung-Russel diagram. These libraries are either composed of synthetic spectra (*e.g.* the BaSeL library, recalibrated using photometric data Lejeune et al., 1998) or composed of real spectra (*e.g.* the elodie library Prugniel and Soubiran, 2001);
v) stellar spectra and stellar evolution prescriptions are both dependent on the α-enhancement (excess of chemical elements O, Mg, Si, S, Ca, and Ti relative to Iron at low metallicity compared to the solar ratio Wallerstein et al., 1963), but this dependence and this excess are difficult to model properly.

Composite stellar populations vs evolutive spectral synthesis While galaxies are not SSP, their stellar content can be viewed at time τ as a composition of coeval stellar populations and described as:

$$\int_{t=0}^{\tau} \mathrm{SFR}(t) \times \mathrm{SSP}(\tau - t, m, Z(t), \mathrm{IMF}) \mathrm{d}t \,, \tag{3.4}$$

with $Z(t)$ the metallicity of the gas (the initial metallicity of an SSP). SFR is the star formation history and describes the distribution in time of the SSPs. To compute the spectral evolution of a single galaxy (a composite stellar population), choices have to be made concerning the functional form of the star formation history, as it is *a priori* unknown, and depends on gas accretion, mergers, internal galaxy physics. We can parametrise it in a simply way. The "closed box" model describes a configuration where the galaxy has no

[5] An isochrone corresponds to the locus occupied by stars of the same age in the Hertzprung-Russel diagram.
[6] Or more precisely the temperature-luminosity-surface gravity diagram.

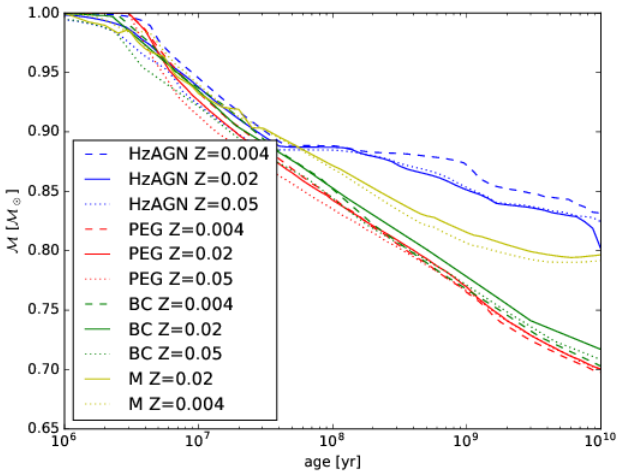

Figure 3.2: Evolution of the stellar mass as a function of time and metallicity for a stellar particle in the Horizon-AGN simulation (blue curves, Dubois et al., 2014), for a SSP using Bruzual and Charlot (2003) models (dark green curves), Maraston (2005) (light green curves) and pegase (Fioc and Rocca-Volmerange, 1997, 1999; Le Borgne et al., 2004) models (red curves) with a Salpeter IMF and different metallicities: $Z = 0.004$ (dashed lines), $Z = 0.02$ (solid lines), $Z = 0.05$ (dotted lines).

late gas infall. Usually in a "closed box" model, star formation rate decreases with time, as it is supposed that with time more and more gas is already turned into stars, so no more available to form new stars. The simplest model for the star formation history in this configuration is exponentially declining, i.e.

$$SFR(t) \propto e^{-t/\tau} . \tag{3.5}$$

An alternative definition is the "delayed" star formation history, more suitable to model a galaxy with gas infall:

$$SFR(t) \propto \frac{t}{\tau^2} e^{-t/\tau} . \tag{3.6}$$

In both cases τ is the e-folding time, varying between 0.1 and several Gyr. In the latter case, τ also represents the galaxy age (since its formation) at which the SFR peaks. I used these two simple models for the photometric redshift estimation of the cosmos2015 catalogue in Chapter 5.

As discussed in the previous Section 3.1.1, it is important to take into account the chemical enrichment of the ISM to correctly model the SED of the galaxy. At the time of galaxy birth, SSPs form in metal-poor gas clouds. But as stellar populations evolve and release feedback in the ISM, the newly born SSPs are formed from more metal-rich gas clouds, and will have different stellar evolution and SED. If this enrichment is not taken into account, it can be a source of biases for the lately computed galaxy properties. As a matter of fact, for the sake of simplicity in stellar composite populations, the metallicity is often assumed constant within the galaxy (Bruzual and Charlot, 2003, BC03 in the following).

An alternative way to model galaxy SED is to use evolutionary spectral synthesis codes. In this case, the shape of the star formation history is not imposed by a model but is computed consistently from the gas content at each time step with an evolutionary scenario. This scenario has to be parametrised. It incorporates the gas accretion, the quantity of gas, the stellar feedback, the star formation efficiency, and importantly the chemical enrichment. Examples of evolutionary spectral synthesis are the pegase code (Fioc and Rocca-Volmerange, 1997, 1999; Le Borgne et al., 2004), or the Maraston (2005) models.

Taking the dust extinction into account in the galaxy spectrum modelling can be achieved self-consistently with radiative transfer (*e.g.* with the evolutionary spectral synthesis pegase code: Fioc and Rocca-Volmerange, 1999), however this is computationally costly. An empirical law for dust extinction is often preferred (see above), which will add a supplementary uncertainty.
Finally, differences in the stellar evolution calculations can deeply affect the galaxy spectrum. For instance a source of uncertainty is the way AGB phase stars are taken into account. AGB phase stars are the dominant source of NIR fluxes after $\sim$ 1 Gyr. Nevertheless, the correct accounting of this phase in stellar

evolution modelling is challenging because the key physical processes at work are still not well understood (*e.g.* Rosenfield et al., 2014, for a study of the AGB-phase). The weight given to the TP-AGB stars in the stellar evolution models can significantly change the stellar mass loss of the whole population. As AGB stars strongly emit in the NIR, if not taken into account the model has to compensate with a higher contribution of old massive stars to match the observed NIR fluxes, leading to an overestimation of the mass of the galaxy. Furthermore, when leaving the AGB phase (and entering the post-AGB phase), the mass loss rate of stars decreases significantly before the stars enters in the planetary nebulae phase. The timescale of this transition is also relatively short (a few thousand years), but if not taken into account in stellar evolution models it can lead to excessive mass losses. Figure 3.2 presents the mass losses for a single SSP with the Bruzual and Charlot (2003) models, the Maraston (2005) models and the STARBUST99 (Leitherer et al., 1999) model (used in the Horizon-AGN simulation, see Chapter 4). Maraston (2005) uses the "fuel consumption" approach instead of evolutionary isochrones for aging stars which enables to better resolve the rapid evolutionary phases such as the AGB phase. In the Bruzual and Charlot (2003) model, the energy contribution from the AGB phase[7] is significantly lower. To reach the correct NIR flux, it overestimates the contribution of the old population, for which the stellar mass losses are more important (Maraston et al., 2006). This point is probably one of the main cause of the difference in the stellar mass loss as a function of time for a single stellar population highlighted in Figure 3.2.

To summarise, a non-exhaustive list of uncertainties in the modelling of galaxy spectra includes the IMF, the chosen stellar templates library, the stellar mass loss prescription (and in particular the weight given to the less well understood phase stars such as the AGB, post-AGB, Wolf-Rayet stars and so on), the SFH and the way chemical enrichment is taken into account. Finally, dust is one of the principal source of uncertainties (see this very detailed study of uncertainties: Conroy et al., 2009; Conroy et al., 2010; Conroy and Gunn, 2010). Spectral synthesis is in fact, a complex topic and each model has it own assets and weaknesses. Entering more deeply in the details of the comparison between the spectral synthesis models is beyond the scope of this thesis. As it has proven to give consistent results for observations in the optical wavelength range, I have used the Bruzual and Charlot (2003) models for the work presented in this thesis, and, when necessary, I will discuss the eventual uncertainties connected to this choice.

Most of the SED-fitting codes, which interpret galaxy photometry by fitting with templates of composite stellar populations use a single metallicity value for the whole galaxy, a simple star formation history, and an empirical law for taking dust into account. The effects of these oversimplifications will be discussed in Section 6.3 using mock photometry from simulated galaxies.

3.2 Observing galaxies

To determine galaxy redshifts one relies on characteristic features ("breaks") in the SED of galaxies (Section 3.3). To determine their properties (mass, metallicity, age, *etc.*), one relies both on characteristic features and on the galaxy continuum. I present below some keys to interpret the galaxy spectrum.

3.2.1 Dissecting the galaxy spectrum

Starlight shapes the galaxy continuum. This continuum is then disturbed by absorption and emission lines from the gas in stellar atmosphere or in the ISM and by dust in the ISM. If present in the galaxy, the active nuclei modifies the galaxy spectrum.

[7] More precisely the TP-AGB phase.

The galaxy continuum in the range $300\,\text{Å} < \lambda < 22000\,\text{Å}$ is dominated by emission from stellar populations. Spectral features originate from the interaction between photons produced by the stellar population and the surrounding gas. Photons emitted by stars ionise the gas (absorption lines). When the gas recombines, a photon is emitted (emission lines). The number, types, and strength of emission and absorption lines are useful for identifying galaxy type and redshifts.

UV continuum – The UV continuum is dominated by young massive stars in the "main sequence" phase. For a given single stellar population, after $\sim$ 100 Myr of its life, the flux below 1000 Å diminishes significantly due to the short timescale of the life of massive stars. For this reason, the UV continuum traces the very recent star formation in the galaxy and can be used to infer the instantaneous star formation rate (but taking into account that this UV emission may be strongly attenuated by dust, *e.g.* Bouwens et al., 2012).
UV emission of a single stellar population life can increase again after $\sim$ 3 Gyr, due to stars in certain phases of the post "main-sequence" evolution (horizontal branch, white dwarfs).

Optical and NIR continuum – At the same time UV emission decreases in the early-phase of galaxy formation, the optical and NIR light increases and is dominated by intermediate and old stars. As these stars contribute the most to the galaxy stellar mass, this part of the spectrum traces the mass assembly over several Gyr. After 1 Gyr, red giant stars and AGB phase stars account for most of the NIR production.
The "age-metallicity" degeneracy may be broken by the study of emission lines (see below) using either narrow band photometry or spectroscopy.

Effect of dust and IR emission – Because of the absorption in the UV, dust is responsible for the reddening of the stellar light and participates to the "age-metallicity degeneracy". The reddening by dust is estimated through the colour excess measured between the B and V bands:

$$E(B-V) = A_B - A_V\,, \tag{3.7}$$

where A_B and A_V are the change in magnitude in the B and V bands due to extinction. It requires the assumption of an attenuation curve (see above), which, in turn, depends on the grain composition.
A number of other spectroscopic features in emission (*e.g.* due to poly-cyclic aromatic hydrocarbons: PAH) and absorption (mainly because of silicates) in the galaxy spectrum are also tracers of dust. Dust re-emits the absorbed energy in the MIR and FIR, making by consequent the IR emission another tracer of star formation.

Absorption lines – Absorption lines occur from absorption by chemical elements in the atmosphere of cool, old, low-mass, low-luminosity stars, or by cold gas in the ISM. They are a good indicator of the metal content of star-forming galaxies. They are more numerous in elliptical and spiral bulges.
The IGM is also responsible for absorption lines in the galaxy spectrum. Consider a galaxy at redshift z_s. A neutral hydrogen cloud at redshift z_c on the line-of-sight of the galaxy will cause an absorption line in the galaxy spectrum at wavelength $\lambda(1+\Delta z) = 1215.7$ Å, where $\Delta z = z_s - z_c$. I will discuss this in Appendix A.

Emission lines – Emission lines originate from hot ionised gas in HII (ionised hydrogen) regions, heated by hot young stars and re-radiating at specific wavelengths, and depend on the chemical composition of the gas. Emission lines can be used as diagnostics for the star formation rate (in particular recombination

lines of hydrogen) and for the metallicity when spectroscopy or a good photometry is available (*e.g.* Faisst et al., 2016, on COSMOS2015. see Appendix B). These emission lines, when not taken into account, are a source of contamination of the observed broad band fluxes.

Breaks – Breaks are spectral discontinuities. They are easily identifiable without the need of high resolution spectroscopy: photometry with well-chosen filters can trace them, which enables to constrain quite precisely the redshift of the galaxy.
The **Lyman break** is the most important feature, occurring at 912 Å. It corresponds to the limit of the Lyman alpha series (ionisation of an hydrogen atom from its ground state). It is due to the intrinsic opacity of the galaxy to Lyman continuum photons: UV photons with $\lambda < 912\,\text{Å}$ are completely absorbed by hydrogen atoms around star-forming regions.
The **Balmer break** occurs at 3646 Å and corresponds to the limit of the Balmer series (ionisation of an excited hydrogen atoms from the $n = 2$ first excited state). It is mostly prominent in stars of intermediate age.
The **4000Å break** originates from the accumulation of ionised metal absorption lines of high energy radiation in the stellar atmospheres. The break increases as a function of the age of the stellar populations (which becomes more metal rich with time).
From a photometric point of view, Balmer and 4000 Å breaks are often associated in a single feature since they are very close but their physical meaning is completely different.

Contamination from the AGN emission – AGN emit in the large wavelength range from radio to gamma rays. This large spectral coverage implies that a large range of processes are involved in emission, and these processes are different from a galaxy without active nuclei. I explain below the most salient aspects of the AGN spectrum, how it can affect the galaxy spectrum and be a reliable signature of a active nuclei in the galaxy (see Section 5.1).
The bulk of the accretion radiation occurs in the x-ray passbands and is mainly due to the inverse Compton up-scattering of UV photons from the thermal accretion disk. At these wavelengths, the contribution from non-AGN activity (such as star formation) is negligible. In the UV passbands, emission from AGN is mainly due to thermal emission from the accretion disk. In the FIR, the AGN dust torus re-emits the radiation absorbed in the UV. Finally the synchrotron jets for radio-loud AGN are responsible for the emission in the radio passbands. Generally, strong emission lines are also a good indicator of the presence of AGN.
AGN are usually classified in two types based on the amount of emitted radio light ("quasar-mode" or "radio-mode", see above) but they are many more classification schemes according to their emission in different wavelength ranges. Only a panchromatic study of AGN would reliably allow to distinguish between AGN types.

3.2.2 Spectroscopic and photometric observations

Spectroscopy – Galaxies can be observed either through spectroscopy or photometry. A spectrograph disperses the galaxy light according to wavelength and provides one spectrum for each object. The precision of the properties deduced from the galaxy SED will depend on the resolution of the spectrum, the exposure time and on the observational strategy (*e.g.* positioning of the slit).
When possible, we prefer to observe objects through spectroscopy. However, despite significant improvements in instrumentation, it remains costly (in terms of telescope time) to obtain spectra for a large number of objects. Hence, in the case of large surveys, photometry is in general preferred.

Photometry – Photometry is a method to measure the flux in some chosen wavebands on a given area of the sky. A photometric survey allows therefore to improve completeness[8] down to a flux limit fainter than the spectroscopic limit at lower price. Nevertheless, photometric surveys have to be combined with (sparser) spectroscopic ones for a better calibration of the galaxy redshifts (see below).
As presented above, different wavelengths of the spectrum are sensitive to different evolutionary stages of the stellar evolution. For this reason, following consistently galaxy properties throughout cosmic time implies following the same part of the galaxy spectrum and requires thus panchromatic surveys.
The choice of the filter passbands, the number of filters, the exposure time, the homogeneity of the data (point spread function, pixel scale) determines the precision of the properties deduced from galaxy photometry.

3.3 Inversion problem: from light to redshift and mass

Given an observed galaxy SED (broadly speaking, it can be also photometric measurements), one wants to infer the redshift and the stellar content. Despite our knowledge of stellar evolution (and still we have shown the uncertainties of this modelling), this is difficult because the overall population generations (SFH) in this galaxy are not known *a priori*.
Generally speaking, the redshift determination is based on observed colours[9] to estimate the shift of the spectrum using characteristic features, while the determination of mass is done through the estimation of the amount of light and is considerably more model dependent.

3.3.1 Inversion techniques

When the medium or high-resolution SED is available ($\lambda/\Delta\lambda > 200$) – Determining the redshift when the full SED is available is quite easy by identifying spectral features and their shift. For other physical parameters, a **full inversion approach** would consist in recovering the galaxy star formation history, mass, metallicity and dust distributions from the galaxy SED (*e.g.* Tojeiro et al., 2007; Ocvirk et al., 2006). This inverse problem is non-parametric and no *a priori* model is required to derive the star formation history and metallicity distribution. However this problem is also ill-conditionned, suffering from instabilities (errors in the observational space can lead to discontinuities in the solution) and degeneracies, such as the age-metallicity degeneracy (see *e.g.* Moultaka and Pelat, 2000; Moultaka et al., 2004). This is still partly true with very high resolution spectra and in a simplifying framework (*e.g.* in the case of a constant metallicity distribution that makes the problem linear), as shown in Ocvirk et al. (2006). Although this method is very useful when high-resolution spectra are available, in particular to estimate the errors associated to the solution, it is not appropriate for recovering the galaxy stellar content when only photometry is available.

When only apparent magnitudes are available – When the only observables are apparent magnitudes in some passbands, other methods are available to determine redshift and masses, which rely either on colour-colour diagrams (in a manner similar to the "ultra-violet dropout" technique Steidel et al., 1996; Madau et al., 1996), SED fitting (introduced by Baum, 1962), or on training on the colours of a subsample of objects with a very good redshift or mass estimate ("machine learning"). Today, other very promising techniques are being developed to compute the redshifts, using in particular clustering information in addition to photometry (*e.g.* the "clustering redshift" technique or the "Photo-Web" techniques, see Scottez

[8] Completeness is a flux limit giving an estimation of the probability that a galaxy of apparent flux f is observable.
[9] A colour is the difference between the magnitudes in two bands.

et al., 2016; Aragon-Calvo et al., 2015, respectively).

SED-fitting codes are routinely used. These codes use pre-computed libraries of galaxy templates. The computation of these templates requires assumptions, generally concerning the star formation histories, the metallicity enrichment and distribution and the dust extinction. The observables to fit (in our case apparent magnitudes) are extracted from the templates. Comparison between real and model observables gives the likelihood of the data given the model, and the probability distribution of the parameters is obtained via marginalisation. A wealth of existing codes allow to compute photometric redshits and masses with different methods, leading to different accuracies on different datasets (see Hildebrandt et al., 2010, for a comparison).

In this thesis I used the LePhare code (Arnouts et al., 2002; Ilbert et al., 2006) which relies on a SED-fitting technique. The code is used to compute photometric redshifts and stellar masses and can give an estimation of ages, metallicities, star formation rates. Importantly, I do not aim at estimate the star formation history by SED fitting. I present an overview of this code below. The particular configuration used for the computation of the cosmos2015 catalog (Laigle et al., 2016) will be explained in Chapter 5. LePhare is also used in Chapter 6 on a virtual photometric catalog computed from the Horizon-AGN simulation (see Chapter 4 for the description of this catalog).

3.3.2 Redshift determination with LePhare

Method – LePhare (Arnouts et al., 2002; Ilbert et al., 2006) is based on a χ^2-template fitting procedure. Photometric redshifts are obtained by comparing the modelled fluxes (or magnitudes) derived from templates and the observed fluxes (or magnitudes) with a χ^2 merit function. At each redshift, $z_{\rm step}$, and for each template of the library, the χ^2 is computed as

$$\chi^2(z_{\rm step}) = \sum_{\rm filters\,i} \frac{(F_{\rm obs\,i} - \alpha F_{\rm SED\,i}(z_{\rm step}, T))^2}{\sigma^2_{\rm obs\,i}}\,, \tag{3.8}$$

where $F_{\rm SED\,i}(z_{\rm step}, T)$ is the flux predicted for a template T in the filter i at $z_{\rm step}$ and α is the normalization factor. $F_{\rm obs\,i}$ is the observed flux in the filter i and $\sigma_{\rm obs\,i}$ the associated uncertainty. Then, the χ^2 is converted to a probability: $p = \exp^{-\chi^2/2}$. All of the probability values are summed up at each redshift $z_{\rm step}$ to produce the probability distribution function (PDF). The photometric redshift solution is then determined from the median of this distribution. The 1σ uncertainties given in the catalog are derived directly from the PDF and enclose 68% of the area around the median.

Spectral coverage – The photometric redshift accuracy depends not only on the method used, but also on the spectral coverage and sampling resolution. Basically, the redshift is measured from the shift of Balmer or Lyman break. If these breaks are not well sampled by the chosen filters, the accuracy is degraded and it is difficult to constrain the position of these features.

As an example, I present on Figure 3.3 the 1σ-redshift errors as a function of redshift and apparent magnitude in the i^+ band for the cosmos2015 catalog computed with LePhare in Laigle et al. (2016). On this Figure photometric redshift accuracy decreases at $z > 1.5$ at fixed apparent magnitudes (hence roughly fixed signal-to-noise ratio). The redshift uncertainty is roughly larger by a factor of 2 on the redshift range $[1.5, 2.5]$. Such a trend is expected. More precisely, at $z > 1.5$, the Balmer break moves outside the medium optical band coverage into the NIR range. The photometric accuracy is poorer in the broad NIR brands than in the optical. Furthermore the position of the Balmer break is less constrained in NIR bands because these passbands are broader than the medium bands. This is naturally reflected in the redshift uncertainty that becomes larger[10]. At $z \sim 2.5$, the Lyman break enters in optical bands and helps

[10] In addition, the upper and lower panels of Figure 3.3 show the measure for different part of the field. In the $\rm A^{UD}$ part the limiting magnitude in the NIR is better than in the $\rm A^{Deep}$ part: as a consequence, the redshift errors are lower in $\rm A^{UD}$.

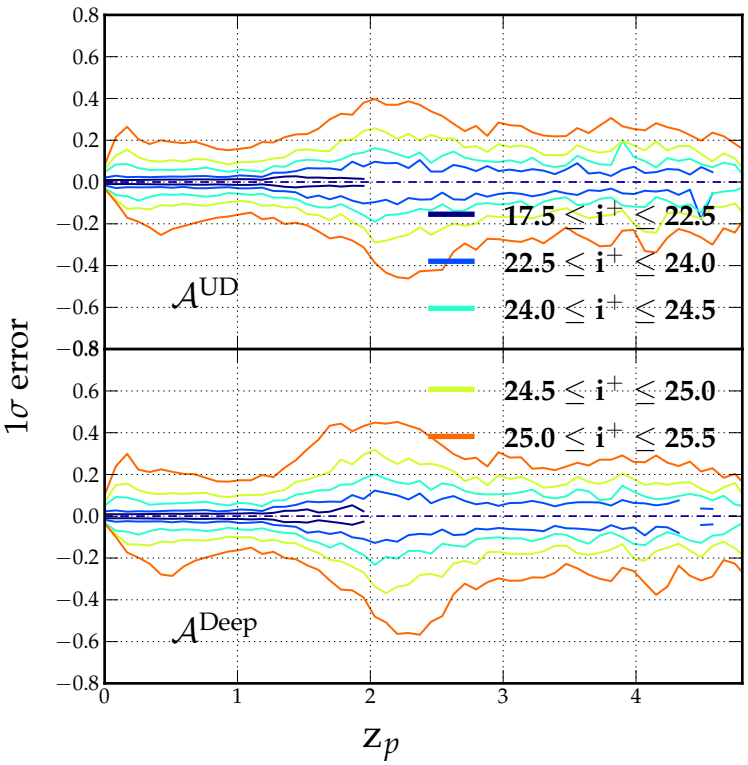

Figure 3.3: The 1σ-redshift error in the COSMOS2015 catalog (on COSMOS field, described in Chapter 5), from the photometric redshift computation with LePhare. Upper and Lower panel show different part of the field for which the limiting flux in the NIR bands are different (the limit is lower for the upper than for the lower panel).

to better constrain the redshift, while it was not covered by any bands at $1.5 < z < 2.5$. That is why the redshift errors decrease again at $z > 2.5$. This simple example illustrates how important is the spectral coverage (both the chosen filters than the signal-to-noise ratio) for the quality of the photometric redshifts (see also Bolzonella et al., 2000).

Calibration using spectroscopic redshifts – An important aspect of the method implemented in LePhare is the automatic calibration of the zero-points using a spectroscopic subsample available on the same field. Zero-points are offsets applied to match the predicted magnitudes and the observed ones (Ilbert et al., 2006). It enables to identify and correct systematics in the photometry or in the template library. For each spectroscopic object, we search for the template which minimizes the χ^2 at fixed redshift. Then, we measure the systematic offset which minimizes the difference between the predicted and observed magnitudes. This procedure iterates until convergence. In order to remove the systematics in the photometry, the spectroscopic sample should be representative of the different galaxy populations.

Choice of the templates – In LePhare, a set of several templates is usually used to improve the fit at different redshifts and for different galaxy types. Computation time scales with the number of templates: it is therefore crucial to optimise the number of chosen templates. Galaxy templates (*e.g.* Polletta et al., 2007; Bruzual and Charlot, 2003) but also AGN (*e.g.* Polletta et al., 2007) and star (*e.g.* Pickles, 1998; Charlot and Fall, 2000) templates are used in the fitting procedure. The computation of the fluxes also takes into account the contribution of emission lines using an empirical relation between the UV light and the emission line fluxes as described in Ilbert et al. (2009).
Internal extinction ($E(B-V)$, see Section 3.2 above) is added as a free parameter to each galaxy and several extinction laws can be considered (*e.g.* Calzetti et al., 2000).
The choice of templates is important because if it is not adapted to the dataset it can lead to systematics in the redshift derivation (and computation of physical parameters). As underlined in Dahlen et al. (2013), a too coarse set of templates compared to the SED distribution may lead to the underestimation of redshift

errors, and this is often the case whatever the photometric redshift code used (In Chapter 5, I present this problem in the specific case of COSMOS2015). Testing the representativeness of galaxy template sets is in fact one major direction to improve LePhare code. One possible could be provided by the concept of self-organizing maps, used to map the colour-redshift relation in a high-dimensional space and to project it in a two-dimensional representation (*e.g.* as suggested in Masters et al., 2015). Another reason for this redshift error underestimation could be the fact that the intrinsic template uncertainties are not included in the computation.

3.3.3 Absolute magnitudes and stellar masses with LePhare

Once photometric redshifts are determined, we would like to obtain galaxy masses and absolute magnitudes.

Absolute magnitudes – Computing absolute magnitudes requires determining the k-correction. The k-correction is applied to take into account the fact that in a given passband, rest-frame fluxes of galaxies at different redshifts are in fact seen through different observed passbands. In this context, even two identical galaxies at different redshifts would have different absolute magnitudes. An estimate of the k-correction term (Oke and Sandage, 1968) relies on the best-fitting template. This component is one of the main sources of systematic error in the absolute magnitude and rest-frame colour estimate. The method used to estimate these quantities is presented in Appendix A of Ilbert et al. (2005). In order to minimize the k-correction-induced uncertainties, the rest-frame luminosity at a given wavelength λ is derived from the apparent magnitude $m_{\rm obs}$ observed at the nearest filter to $\lambda(1+z)$. Using this procedure, the absolute magnitudes are less dependent on the best-fit SED, but are more dependent on any observational problem affecting $m_{\rm obs}$. Therefore, the code is constrained to consider only the broad bands for $m_{\rm obs}$ and those bands with a systematic offset lower than 0.1 mag derived for the photometric redshift. Absolute magnitudes are not directly used for the stellar mass estimation.

Stellar masses – Galaxy stellar masses are derived in the same way as photometric redshifts, but independently in a second pass. A library of synthetic spectra is generated using stellar population synthesis with a given IMF, metallicity and models for the star formation history (as presented in Section 3.1.2) and shifted to the galaxy's estimated redshift. Masses are obtained by comparing the modelled fluxes with the observed ones in the same way as redshifts (with a χ^2 merit function, and taken as the median of the probability distribution).
In this derivation, photometric redshift uncertainties are not taken into account. A possible way to include them is is to do a Monte Carlo simulation, repeating the SED-fitting operation a certain number of time, each time re-extracting the photometric redshift according to the probability density function (*e.g.* Grazian et al., 2015).

Other methods exist to estimate stellar masses directly based on the tight relation between mass-to-light ratio and photometric colours (Bell and de Jong, 2001). For instance, Taylor et al. (2011) present an estimation of the masses from the M_*/L_i versus $(g-i)$ relation for galaxies at low redshift ($z<0.65$) in the GAMA survey and assess the reliability of their masses by comparison with SED fitting.
Since the mass-to-light-colours relation (MLCR) method seems considerably simpler and less computationally expensive than SED-fitting, it is worthwhile to examine its reliability. The MLCR requires the calibration from models and is therefore not free of systematics. However, if we ignore the uncertainty due to the model, it is possible to constrain the mass from the MLCR within 0.1 dex (Taylor et al., 2011), although the results of the comparison with SED-fitting algorithms differ from on study to another (*e.g.*

Vulcani et al., 2014; Roediger and Courteau, 2015). Nevertheless, it should be acknowledged that some galaxy populations have their mass significantly biased by the MLCR estimate, in particular galaxies with bursty star formation histories (Gallazzi and Bell, 2009) or low-mass galaxies (Ilbert et al., 2010, using K-band).

Chapter 4

Numerical methods: mocking the Universe with Horizon-AGN

The ambition of cosmological simulations is to model the Universe in its entirety, both spatially and throughout cosmic time. As a matter of fact, on one hand we know quite well the initial conditions of our Universe (with the results from the *Planck collaboration*) and on the other hand we observe the luminous galaxies, which we know are the results of the gravitational amplification of the tiny density fluctuations in the primordial Universe. Although it is analytically possible to follow the growth of primordial dark matter perturbations with the help of linear perturbation theory (*e.g.* Zel'dovich, 1970) to a certain extent, this task becomes impossible once the structures enter the non-linear regime (for density contrasts higher than one). In addition, the evolution of the baryonic matter further complicates the problem by adding new levels of non-linearity. Hence, the full growth of structures can only be solved numerically. To this purpose, cosmological simulations aim to build a bridge between the initial conditions of the Universe and the observed galaxies. State-of-the-art simulations are now able to implement numerically our best knowledge about the initial conditions and the formation of structures up to the stellar formation. Hence they offer the possibility to confront our galaxy evolution model with the real Universe.

A wealth of different codes exist to implement hydrodynamical simulations. They are either based on particles, which means discretising the mass (namely the Lagrangian codes, *e.g.* the GADGET or GASOLINE codes, see Springel, 2005; Wadsley et al., 2004, respectively), or grid-based, which means discretising the space (namely the Eulerian codes, e.g the RAMSES or AREPO codes, see Teyssier, 2002; Springel, 2010, respectively). No code can be definitively said to be superior to the other; each code has its own strength and weaknesses (see *e.g.* Agertz et al., 2007, for a comparison of different codes). Since numerical methods are approximate solutions to differential equations, they suffer from truncation errors under different forms, depending on the technique. I do not intend to discuss this point here. One must examine the codes critically to be able to distinguish between the failures related to the computational implementation (and consequently to the use of a specific code) and those related to the physical model itself.

In the following, I will present the state-of-the-art Horizon-AGN cosmological hydrodynamical simulation (Dubois et al., 2014), which I have used during my work, either through a statistical confrontation with observations (Chapters 6) or to make predictions for observations and future surveys (Chapters 8 and Appendix A). One of the great asset of Horizon-AGN compared to other existing cosmological hydrodynamical simulations (*e.g.* Vogelsberger et al., 2014; Khandai et al., 2015; Schaye et al., 2015) is that a lightcone has been extracted on the fly, meaning that I have at my disposal a virtual survey with the

same geometry as in observed surveys. In Section 4.1, I will explain the main aspects of the simulation, and how dark matter, gas, and star evolutions are implemented. From this simulation, I have extracted a galaxy catalogue and computed fully consistently the photometry of the galaxies, which I will present in the Section 4.2. An important aspect of the virtual photometry is how dust extinction is taken into account. I will address this point in Section 4.3.

4.1 The Horizon-AGN simulation

The cosmological Horizon-AGN[1] simulation (Dubois et al., 2014) has been run with the ramses code (Teyssier, 2002) by Yohan Dubois, Christophe Pichon and Julien Devriendt. The periodic box is large enough ($L_{\rm box} = 100\,h^{-1}$ comoving Mpc on a side) to offer a reliable statistics of the matter distribution, comparable with existing large galaxy surveys. Dark matter is modelled as particles ($M_{\rm DM,res} = 8 \times 10^7\,{\rm M}_\odot$), while the gas evolution is followed on an adaptative grid (finer gas cell: 1 proper kpc). The simulation models star formation, stellar evolution and AGN feedback at the subgrid scale, implementing our current knowledge about internal galaxy physics. Hence, the simulation can be directly compared to results from galaxy surveys. Through these comparisons we may improve the model implemented in the simulations. It is important to note, for a fair comparison, that the subgrid physics is *not* calibrated on the local Universe, except for the black hole-galaxy relations (see below).
A twin simulation has been run with exactly the same initial conditions, but turning off the AGN feedback. The comparison of the simulations with and without AGN will allow us to lean the importance of AGN feedback in regulating galaxy mass growth and star formation (see Chapter 6).

4.1.1 Running the simulation

Generating initial conditions – Initial conditions of a cosmological simulation are usually generated assuming a primordial Gaussian random field. All information about the field is contained in the power spectrum describing the spatial correlation of the random fluctuations. Gaussian random fluctuations can be modelled as white noise convolved by a transfer function: this transfer function encodes the power spectrum.
In Horizon-AGN, initial conditions are generated with the mpgrafic software (Prunet et al., 2008). A standard Λ-cold dark matter cosmology is adopted, with the cosmological parameters values given in Chapter 2.

Dark matter evolution – As explained in the introductory Chapter 2, dark matter can be modelled as a collisionless fluid. Concretely, this means that the dark matter distribution function $f(\mathbf{x}, \mathbf{v}, t)$ in phase-space follows the collisionless Boltzman Equation (specifically the Vlasov equation):

$$\frac{\partial f}{\partial t} + \mathbf{v}\frac{\partial f}{\partial x} - \nabla\phi\frac{\partial f}{\partial v} = 0\,. \tag{4.1}$$

Here ϕ is the gravitational potential, which obeys the Poisson equation:

$$\nabla^2\phi = 4\pi G m \int f {\rm d}^3 v\,, \tag{4.2}$$

where m is the dark matter particle mass, and $\int f {\rm d}^3 v$ is the density.
In Horizon-AGN, the dark matter distribution is sampled by massive particles. Those particles are macro

[1] http://www.horizon-simulation.org/

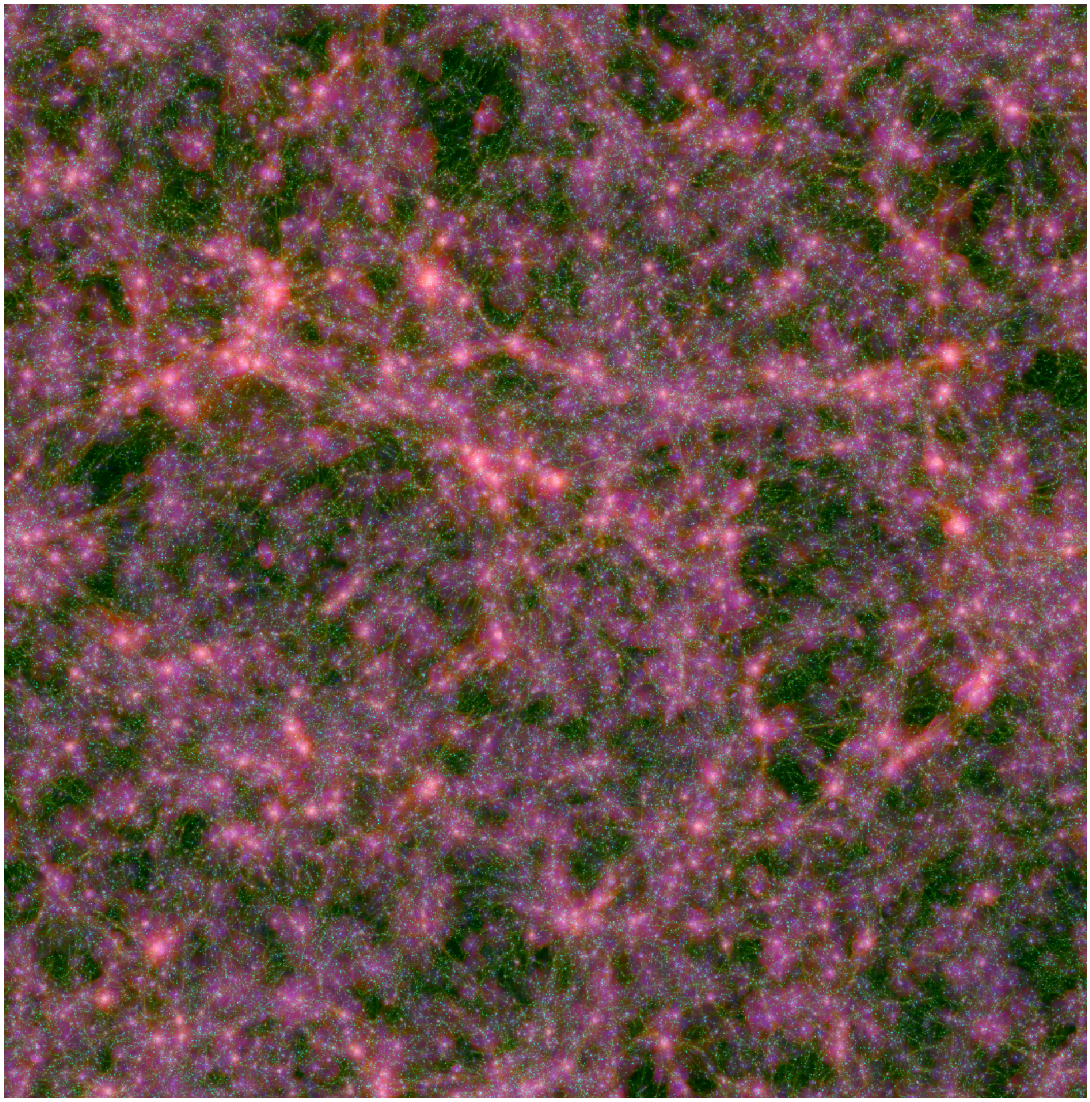

Figure 4.1.: The projection of the entire volume of the Horizon-AGN simulation at $z \sim 1.2$ of the gas density (green), the gas temperature (red), and the gas metallicity (blue). The comoving side of the box is $100/h$ cMpc.

particles of mass $M_{\mathrm{DM,res}} = 8 \times 10^7\ \mathrm{M}_\odot$. They follow trajectories governed by their mutual self-gravity. So at this stage, the core part of the work consists essentially in the computation of the force $\mathbf{F} = -\nabla\phi$. The most direct way to do this is to compute directly the particle-particle summation. However, this operation is very computationally demanding as the algorithm scale as N^2 where N is the number of particles. The principle is used in the tree-codes, which reduce the number of operations required by relying on a tree structure (e.g the GADGET code Springel et al., 2001; Springel, 2005).

ramses uses an alternative way to get the force, by computing it on a non-uniform grid using the adaptive particle-mesh method (Teyssier, 2002). Each cell of the grid is associated to a density field, by projecting the particle distribution onto the grid. To avoid density discontinuities when a particle will move from one cell to an other, a Cloud-in-Cell technique assigns to each particle a spatial extension with a cloud size equal to that of the host cell. The method is said to be "adaptive" as the grid is refined according to the density, in order to correctly resolve the force in the highly non-linear regions. Once the density field is computed on the grid, the gravitational potential is derived by solving the Poisson Equation (conjugate gradient method), and the force calculated and interpolated at each particle position.

Hydrodynamics – In contrast with dark matter, baryonic gas is a collisional fluid. From the collisional Boltzman equation, one can derive the system of Euler equations which describe the motion of the gas, specifically the conservation laws:

$$\frac{\partial \rho}{\partial t} + \nabla.\rho\mathbf{u} = 0 \ : \text{ conservation of mass} \tag{4.3}$$

$$\frac{\partial \rho\mathbf{u}}{\partial t} + \nabla P = -\rho\nabla\phi \ : \text{ conservation of momentum} \tag{4.4}$$

$$\frac{\partial \rho E}{\partial t} + \nabla.\rho\mathbf{u}(E+P) = -\rho\mathbf{u}.\nabla\phi \ : \text{ conservation of energy} \tag{4.5}$$

In this set of equations, ρ, $\mathbf{u}$ and E are density, velocity and total energy (derived from first, second and third moments of the distribution function). P is the thermal pressure. Note that the energy equation includes in reality heating and cooling source terms (see below for the implementation of the heating and cooling of the gas).
In RAMSES, the dynamics of the gas is discretised on the adaptively refined mesh: the Euler system is solved by computing the flux of mass, momentum and energy across the cell boundaries. The Riemann problem is solved at the cell interfaces, leading to a flux at each cell face and giving the force on the fluid across the cell. The gas and its properties are then advected across the cell face. The algorithm used here is the unsplit Godunov scheme (Godunov, 1959) with second-order accuracy.
The grid is said adaptively refined, as the cells are of different sizes according to their density. In Horizon-AGN, the initially uniform 1024^3 grid is refined when 8 times the initial total matter resolution is reached in a cell. The initial volume element can be split in 2 up to 7 times, down to a minimum cell size of 1kpc in proper units. This criterion is said "quasi-Lagrangian" for it refines the grid based on the mass within the cell. In order to keep the minimum cell size approximately constant in proper units, a new maximum level of refinement is introduced each time the expansion scale factor doubles.

4.1.2 Subgrid recipes

The fundamental equations describing dark matter and gas evolution presented above are not sufficient to correctly model galaxy formation. A prescription for radiative processes is needed, as gas cooling is crucial for galaxy formation. Additionally, one needs prescriptions describing how to form stars from gas in dense and cold regions and how to follow their evolution and feedback processes on the ambient gas. These prescriptions are presented below, and encompass most of the processes involved in stellar evolution which have been presented in Chapter 3.

Radiative cooling and heating – Baryons reduce their internal energy through radiative losses. It is crucial to correctly model gas cooling, for cool gas is the fuel for galaxies to form stars. To take into account this dissipation, in Horizon-AGN, gas can cool down to 10^4 K through the collision of hydrogen and helium atoms with a contribution from metals unsing rates tabulated by Sutherland and Dopita (1993).
Heating of the gas is implemented assuming all the gas is in ionization equilibrium with a uniform UV background, which takes place after redshift $z_{\rm reion} = 10$ following Haardt and Madau (1996)

Star formation – Star particles are created using the standard 2% efficiency per free fall time (Kennicutt, 1998), and following the Schmidt-Kennicutt law (Kennicutt, 1998) described in Chapter 3. Stars are only allowed to form when the hydrogen gas density is sufficiently high, above a critical threshold of $n_0 = 0.1\,\mathrm{H\,cm^{-3}}$. In star-forming regions, N equal mass star particles are formed, with N drawn from

a Poissonian random process (Rasera and Teyssier, 2006) that has a stellar mass resolution of $\rho_0 \Delta x^3 \sim 2 \times 10^6 M_\odot$. The mass of gas in the parent gas cell is reduced accordingly. Stars are then assigned the position and velocity of their parent cell, and treated as collisionless particles in the same way than dark matter.

Stellar evolution – In Horizon-AGN stellar evolution is implemented using a subgrid model for stellar feedback that probes all processes that may impart thermal and kinetic feedback on the ambient gas, as described in Chapter 3. Each star particle is assumed to behave like a single stellar population. Many previous works that implement stellar feedback employ a single supernova explosion per star particle to minimize computational cost (e.g Dubois and Teyssier, 2008). However, this is an oversimplification, particularly from the point of view of chemical enrichment. As explained in Chapter 3, a significant fraction of stellar mass is lost through various phases of stellar evolution. Thus, stellar feedback should be modelled realistically by taking into account stellar winds and both Type II and Type Ia supernovae (SNe; *e.g.* Kobayashi and Nakasato, 2011; Hopkins et al., 2012).
Horizon-AGN implements continuous stellar feedback that includes momentum, mechanical energy and metals from Type II SNe, stellar winds, and Type Ia SNe. For stellar winds and Type II SNe, Starburst99 (Leitherer et al., 1999, 2010) is used to generate look-up tables as a function of metallicity and age. Specifically, the Padova model (Girardi et al., 2000) is used with TP-AGB stars (Vassiliadis and Wood, 1993), with the kinetic energy of stellar winds calculated via the evolution model of Leitherer et al. (1992). Type Ia SNe are implemented following Matteucci and Greggio (1986), assuming a binary fraction of 5% (Matteucci and Recchi, 2001). The chemical yields for Type Ia explosions are taken from the W7 model of Nomoto et al. (2007). Although the energy input from this source is minor compared to that of Type II SNe (∼10% of the total kinetic energy), they provide a significant fraction (∼50%) of the iron for the chosen parameters.
In order to mimic the propagation of bubbles as realistically as possible, the injection of energy, mass, and momentum is allowed only if a blast wave from star particles in each cell propagates to $r_{\rm B} \geq 2\Delta x$, where Δx is the size of the host cell and $r_{\rm B}$ is the radius of the shock front at Δt:

$$r_{\rm B} \sim 44{\rm pc} \left[\frac{E}{10^{47}{\rm erg}} \frac{0.1 H/{\rm cm}^3}{n_{\rm H}} \right]^{0.2} \left(\frac{\Delta t}{10^7 {\rm yr}} \right)^{0.4} \tag{4.6}$$

If the energy released from each cell at $\Delta t = t_{\rm last} - t_{\rm now}$ is not large enough to push the blast wave to $2\Delta x$, the energy, momentum, and metals are accumulated until the next time step, where $t_{\rm last}$ is the time at which the last blast wave is launched. This produces a more realistic evolution of bubbles and prevents them expanding too rapidly.
To reduce computational cost, stellar feedback is modelled as a heat source after 50 Myr, while the energy liberated before 50 Myr from star particles is modelled as kinetic feedback, as described above. This is a reasonable choice, given that, after 50 Myr, almost all of the energy is liberated via Type Ia SNe that have time delays between several hundred Myrs to a few Gyrs (*e.g.* Maoz et al., 2012). These systems are less prone to excessive radiative losses, as stars are likely to disrupt or move away from their dense birth clouds after around a few tens of Myr (Blitz and Shu, 1980; Hartmann et al., 2001).

Feedback from black holes – The simulation also follows the formation of black holes (BHs). Seed black holes with a mass of $10^5 M_\odot$ are assumed to form in dense star forming regions where both the gas and stellar densities are above ρ_0, and where the stellar velocity dispersion is larger than 100km.s^{-1}. The growth of the black hole is tracked self consistently, based on a modified Bondi accretion rate at high gas densities (Booth and Schaye, 2009). The accretion rate is capped at the Eddington rate (maximal growth rate for radiative efficient accretion), with a standard radiative efficiency of 0.1.
The central black hole impacts ambient gas in two possible ways, depending on the gas accretion rate (as

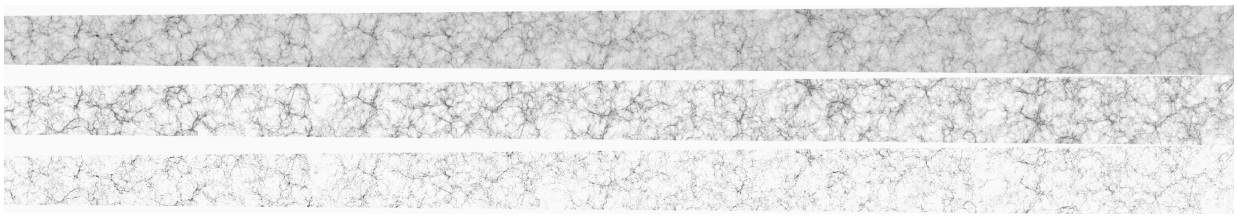

Figure 4.2.: The lightcone extracted from the Horizon-AGN simulation. The gray scale encodes the gas density (*top*), the stellar density (*middle*) and the gas temperature (*bottom*).

described in Chapter 3). For Eddington ratios > 0.01 (high accretion rates), 1.5% of the accretion energy is injected as thermal energy (a quasar-like feedback mode), whilst for Eddington ratios < 0.01 (low accretion rates), bipolar jets are employed with a 10% efficiency. As described in Dubois et al. (2012a), the parameters are chosen to produce agreement with the local cosmic black-hole mass density, and the $M_{\rm BH} - M_*$ and $M_{\rm BH} - \sigma_*$ relations (BH mass-to-bulge mass and BH mass-to-bulge velocity dispersion relations, *e.g.* Magorrian et al., 1998). An explicit dynamical drag force is exerted from the gas onto the black holes (Ostriker, 1999; Chapon et al., 2013) in order to stabilize black hole motions into galaxies and suppress limited resolution effects (Dubois et al., 2013). Finally, black holes are allowed to merge when they are closer than 4 kpc and when their relative velocity is smaller than the escape velocity of the binary. We note that, apart from choosing the BH-feedback parameters to match these relations at $z = 0$, Horizon-AGN is not otherwise tuned to reproduce the bulk properties of galaxies such as stellar mass and luminosity functions, galaxy sizes and so on.

4.1.3 Extraction of the lightcone

Lightcone extraction – The lightcone is extracted on-the-fly, following Pichon et al. (2010). The idea is to extract at each coarse time-step those particles which belong to the slice between this time-step (z_1) and the following (z_2). Assuming that the observer is at the origin of the box, one would like to select the particles such as $d(z_2) < d < d(z_1)$, where d is the line-of-sight comoving distance to the observer, defined as:

$$d = \int c \frac{\mathrm{d}t}{a(t)} \,, \tag{4.7}$$

with a the scale factor and c the speed of light. For this extraction, gas cells are treated as particles.

Lightcone characteristics – The angular opening of the lightcone is one degree at redshift above 1, and 2.25 degrees at redshift below 1, which makes its area comparable to the cosmos field. Note that because of the periodicity of the box, the lightcone presents at large temporal scales a periodic pattern: structures in the lightcone at a high redshift will be recovered at a lower redshift in a more evolved state.
The lightcone is displayed in Figure 4.2.

4.2 A virtual galaxy catalogue

Let me know describe the mock catalogue I have computed which is designed to mimic an observational dataset starting from the information available in the Horizon-AGN simulation. Two catalogues are available, one computed on the simulation box at each fixed redshift step and the other in the lightcone.

For each catalogue, different levels of realism are available (*e.g.* with or without redshift distortion, with or without dust extinction, ...etc.). The goal here is to assign to each simulated galaxy a set of apparent or absolute magnitudes coherent with its intrinsic physical properties.

4.2.1 Extraction of the haloes

Once the simulation has been run, one needs a good way to extract halo and galaxy catalogues from this simulation, to take fully advantage of its quality. The first step is to extract the gravitationally bound systems from the particle distribution, with the help of a halo finder. The most basic halo finder and historically the first one is the friends-of-friends algorithm (FOF) which relies on a linking length threshold and a minimal number of particles to select the structures. In this work, I mostly used the AdaptaHOP halo finder (Aubert et al., 2004). Both dark matter haloes and galaxies are identified from the dark matter and star particle distribution respectively, using a similar procedure.

Identifying galaxies and haloes with AdaptaHOP – To identify galaxies from the stellar particle distribution in the lightcone, I run the AdaptaHOP halo finder (Aubert et al., 2004) in its updated version (Tweed et al., 2009). The algorithm isolates haloes[2] based on prescriptions identifying local maxima of the density field. The halo is thus primarily identified in real space: particles are grouped based on their positions, according to the following steps. For each particle in the simulation, the local density is computed from a total of 20 neighbours. Local density maxima are identified, and a first density threshold is applied on the density field. It allows isolating structures as being the set of particles around their local maxima above this density threshold. This density threshold sets the halo boundary. For dark matter the density threshold is chosen to be $\rho_t = 80$ times the average matter density, and $\rho_t = 178$ for the star particles. For dark matter, $\rho_t = 80$ corresponds to a standard linking length of $b = 0.2$ in the classical friends-of-friends algorithm. The value of 178 corresponds to the ratio between the density in a sphere at virial equilibrium and the background density in the context of the spherical collapse model. It is a good proxy to define virialized structures. Only structures with more than 50 particles are kept in the catalogue. To identify sub-structures, saddle points between the groups are determined in the density field, and used to connect group together. Groups for which saddle points are above the density threshold ρ_t are linked together, and a hierarchy level is assigned to each substructure within the group: A level of 1 is given to the structure with the highest maximum, and a higher level is given to each substructure according to their rank. The minimum size below which structures are considered as irrelevant is $\sim$ 2 kpc (force softening length).
One can safely consider that this level hierarchy has a physical meaning. For dark matter, the level-1 halo would correspond to the main halo, while higher level haloes would be sub-haloes within this main halo. Concerning galaxies, the level-1 galaxy can be considered as the central galaxies, and substructures should be understood as satellite galaxies around the central (see Chisari et al., 2016, Figure 2). In this sense, satellite galaxies are taken out of the mass of the central galaxy.
AdaptaHOP is run at each time step in the simulation box at fixed redshifts. Dark matter halo and galaxy catalogues are identified separately.

The case of the lightcone – The identification of galaxies is based on a density threshold depending on the average matter density. It is thus redshift dependent, which implies that haloes cannot be extracted all at the same time in the full lightcone. To work around, I run AdaptaHOP iteratively on thin overlapping slices of the lightcone, each slice being buffered by half its thickness from its neighbouring slice. Then I

[2] In this section what is referred to as "haloes" is a gravitationally bound system of particles, without distinguishing between stars and dark matter particles.

keep in each slice only those galaxies with $x_s + s/4 < x < x_s + 3s/4$, where s is the thickness of the slice, x is the comoving coordinate along the line-of-sight (decreasing with the redshift) and x_s is the coordinate of the slice along the line-of-sight the closest to the observer.

Weaknesses of AdaptaHOP – There are a plethora of existing halo finders. The main question is how halo and galaxy properties are affected by the choice of the algorithm. However, comparative studies have shown that whatever the chosen algorithm, it recovers reasonably well the spatial location of the structures, as well as their fundamental physical properties (*e.g.* Knebe et al., 2011). One could think that the main weakness of the AdaptaHOP finder arises from the fact that it relies only on the positions of particles to determine the structures. A more precise algorithm would use indeed simultaneously positions and velocities, hence detecting the structures in the 6-dimensional space (*e.g.* Maciejewski et al., 2009). Nevertheless it has also been shown that working either in position or in phase-space does not really impact the recovered halo if the number of particles threshold is sufficiently high (Knebe et al., 2011). The only concern is about the identification of subhaloes. As AdaptaHOP relies on particle positions only, close structures may be not very well separated and particles from the main halo may pollute particles from the sub-haloes. A phase-space based algorithm will naturally disentangle particles from both haloes. However, this probably does not affect the comparison with observations in terms of global quantities (*e.g.* mass function), as galaxies extracted from observations potentially suffer similar problem (close galaxies are not separated with a specific procedure).

4.2.2 Producing mock photometry and spectra

Virtual galaxy photometry – Once galaxies are identified in the simulation, their photometry has to be computed. The method is similar in the simulation box or in the lightcone, although, in the simulation box, I do not compute apparent magnitudes.

I use the BC03 single stellar population templates (Bruzual and Charlot, 2003) with a Salpeter IMF (Salpeter, 1955) to compute the photometry of the galaxy. A BC03 single stellar population (SSP) is attached to each particle in each galaxy of the Horizon-AGN simulation. To select the correct SSP, its characteristics (in terms of age and metallicity) should be as close as possible to those of the linked star particle. However, this task is not trivial: in the BC03 model there are only 6 metallicity steps from $Z = 0.0001$ to 0.05 and 221 time steps from 2 to 10^{10} yr, whereas in Horizon-AGN, Z and age vary across a finer grid of values. Thus, I logarithmically interpolate between SSPs to reproduce the desired values. The SSP is also rescaled to match the initial stellar mass of the particle. This point is important, as it allows to consider that the particle loses its mass according to the BC03 templates. In fact, stellar mass losses for a SSP in the BC03 models are different than stellar mass losses in the Horizon-AGN simulation, as discussed in Section 3.1. If the initial mass of the SSP is not rescaled, the resulting photometrically derived mass will be too low compared to the intrinsic mass. Furthermore, it is also important that the chosen IMF to compute the photometry is the same IMF as the one used for the subgrid recipes in the Horizon-AGN simulation.
At this stage, dust can be added as a screen in front of each particle (see Section 4.3 below for the exact computation of dust attenuation). By summing the SEDs of all the SSPs within a given galaxy, I obtain the spectrum of that object from 91 to 16×10^5 Å, with a resolution of 3 Å from 3200 Å to 9500 Å and a lower resolution outside this range.
If apparent magnitudes are computed, the galaxy spectrum is shifted according to its spectroscopic redshift (including peculiar velocities, *e.g.* finger-of-god distortion). Eventually, the photometric baseline is obtained by convolving the spectrum with the filter passbands. The photometry is computed in the following 13 bands: NUV, u, B, V, r, i^+, z^{++}, Y, J, H, K_s, 3.6μm, 4.5μm, using the same filter passbands

than used for the COSMOS2015 catalogue (see Chapter 5 and Figure. 5.2).

Re-normalisation of the metallicity – It should be noted that the simulated gas-phase metallicity in HORIZON-AGN is underestimated by a factor of ~ 2 to 4 compared to observations (e.g Maiolino et al., 2008). One reason could be that the HORIZON-AGN blast wave model allows for the propagation of energy and metals only when it reaches $2\Delta x$, where Δx is the size of a host cell of SNe. Given the relatively low spatial resolution adopted in HORIZON-AGN, this tends to delay the metal enrichment of star-forming clouds in the simulation, particularly when the specific star formation rate is low (i.e. at lower redshift). However, the main reason is probably that the simulation fails at producing the correct star formation rate evolution for low-mass galaxies: stars are generally formed too early in the galaxy life and are thus less metal-enriched (see Chapter 6). To correct for these lower metallicities, the gas-phase and stellar metallicities are calibrated by multiplying a redshift-dependent renormalisation factor ($f_{\rm no}$) that brings the simulated metallicity in agreement with the observed mass-metallicity relations at $z = 0$, 0.7, 2.5 and 3.5, where $f_{\rm no} = 4.08 - 0.21z - 0.11z^2$. This correction is not perfect. Indeed the metallicity discrepancy is expected to increase with the age of the galaxy. Here the computed factor is the same for all galaxies at a fixed redshift, while a finer choice will be to derive different factors for different galaxy populations.

Producing mock images – A major asset of the HORIZON-AGN lightcone is that it allows me to conduct a end-to-end comparison of galaxy properties in the simulation with galaxy properties in observations, derived exactly as in observations (see Chapter 6). For this purpose a prerequisite is to compute virtual "observation-like" images from the lightcone.
Photometric images are computed in sky coordinates at the COSMOS pixel scale ($0.15''$/pixel). All particles in the volume $0.1 < z < 5.8$ are considered. Each particle is expected to contribute to its four neighbour pixels with a certain weight depending on its position. At each pixel, the observed SEDs of the particles contributing to the pixel are stacked and then convolved by the filter passbands, to get the flux of the pixel. At this stage dust attenuation is not taken into account. Figure 4.5 displays a composite image of a region of the HORIZON-AGN lightcone.
To make it realistic, I chose to convolve the images with the same PSF as in COSMOS2015 (A Moffat profile with $\mathcal{M}[\theta, \beta] = \mathcal{M}[0.8'', 2.5]$, as explained in Chapter 5). I also added noise in each band: I perturbed the value of the pixel magnitudes assuming a gaussian noise with 0 average and rms values corresponding to the RMS_MAP available on COSMOS field in the band.

The initial motivation for having these images is to be able to quantify how much the photometry extracted from an image is different from the photometry computed directly from the star particle distribution. The main sources of uncertainties and biases intervening here are blending, the effect of the PSF, and the intrinsic limiting magnitudes. Working on a simulated lightcone is interesting as the true galaxy clustering is available, which is especially important to test the effect of blending and confusion. Although this work is still an on-going project, it will be discussed in Chapter 6.
These images are calibrated for the comparison with the COSMOS field, but can be obviously redone to match the requirements of other surveys. In the same vein, I produced from HORIZON-AGN mock images similar to those obtained in the CANDELS field for morphology deep-learning algorithm calibration (Huertas-Company et al., 2015). A sub-sample of these images is presented in Figure 4.4.

Figure 4.3.: A 14 arcmin^2 simulated composite image from the Horizon-AGN ligthcone, in the u, r and z filters. The resolution is $0.15''$/pixel and the image is computed using star particles in the redshift range $0.1 < z < 5.8$. Dust extinction is not taken into account.

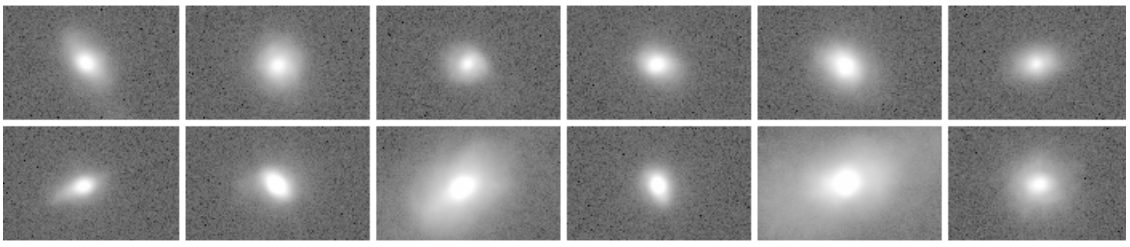

Figure 4.4.: An example of virtual images of galaxies produced from the Horizon-AGN lightcone (including PSF and gaussian noise) in the $F160W$ filter passband and classified as early disks by the deep learning algorithm (Huertas-Company et al., 2015).

4.3 Accounting for dust attenuation

As discussed previously (see Chapter 3), dust is the main source of uncertainties when extracting physical information from galaxy spectrum or photometry. It is thus of prime importance to have a reliable model for dust when projecting data from the simulation in observational colour space. I computed the attenuation by dust on the simulated galaxies using either the SUNSET code (dust is modelled as a screen in front of each star particle) or the SUNRISE code (effect of dust is accounted from a full radiative transfer, which is more accurate but much more computationally expensive). I present below these two codes and their differences.

4.3.1 The SUNSET code

The easiest way to consider dust attenuation is to model it as a constant screen in front of the galaxy (left panel on Figure 4.5). This is basically how dust is taken into account when computing galaxy properties by SED fitting (see Chapter 3). Nevertheless this method is also the least accurate one, as it does not consider the relative distribution of dust and stars which modifies the line-of-sight view of the galaxy.
To model the dust attenuation for the HORIZON-AGN galaxies, I use a more elaborate method (firstly implemented by Romain Teyssier under the name of SUNSET) by using the gas metallicity distribution as a tracer for the dust (middle panel on Figure 4.5). This assumption is justified, for it is known that dust is formed in the late stage of stellar evolution together with metals (see Chapter 3). In turn, we assume that 40% of metals in gas are locked in dust grains.
I evaluated the gas density and metallicity in a cube of 138^3 comoving kpc^3 around the galaxy, with a resolution of 128^3. I assumed that the dust mass scales with the gas metal mass, and I used an empirical dust-to-metal ratio of 0.4 (Dwek, 1998; Jonsson, 2006). Furthermore for the sake of simplicity I took this value as a constant, though some works have shown that it could vary with redshift or within a same galaxy as a function of the metallicity (*e.g.* Galametz et al., 2011; Mattsson et al., 2012; De Cia et al., 2013; Fisher et al., 2014). Nevertheless, there is an overall agreement that, in many case, the variation of this ratio is relatively small (Zafar and Watson, 2013).
I computed the dust column density and the optical depth along the line of sight for each stellar particle in the galaxy, using the $R = 3.1$ Milky Way dust grain model by Weingartner and Draine (2001). I assumed that the gas is transparent beyond one virial radius. I checked that relaxing this assumption does not modify sensitively the photometry. Finally, the model allows to compute the extinction in each passband.

4.3.2 The SUNRISE code

An alternative and more accurate way to take dust into account is to perform a radiative transfer to self-consistently track absorption, scattering and radiation of light by dust (right panel on Figure 4.5). The SUNRISE code (Jonsson, 2006; Jonsson et al., 2010) is a free software[3] which has been written for this purpose. As the code performs a full radiative transfer (and despite optimisations used to speed up the analysis), it is very time consuming to run it on one galaxy. Therefore it is not feasible to run it on all galaxies in the simulation. However, as it is expected to give a more accurate answer than the SUNSET code, it can be used as a proxy on few galaxies to estimate how good is the answer with SUNSET. I have done this test on few galaxies at redshift $z = 0.3$ in the simulation HORIZON-AGN.

[3] https://bitbucket.org/lutorm/sunrise/

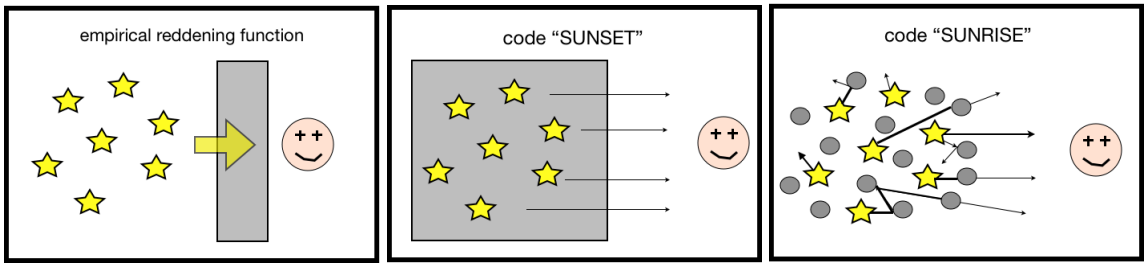

Figure 4.5.: A cartoon explaining the different ways to model dust presented in this Chapter. *Left*: In the most simple case, the effect of dust is modelled through an empirical reddening function. The extinction is therefore the same for all particles. *Middle*: The SUNSET code estimates the dust distribution from the gas metallicity distribution and computes the dust extinction in front of each particle. *Right:* The SUNRISE code performs the full radiative transfer and computes both attenuation and scattering.

Running SUNRISE – SUNRISE has been specifically designed to compute images and spectra from a hydrodynamical simulation. The finite resolution of the simulation (1 kpc in the case of Horizon-AGN) which is insufficient to correctly resolve the physics of dust in star-forming regions (Jonsson et al., 2010) requires a specific subgrid modelling to correctly perform the radiative transfer calculation. That is why an important aspect of SUNRISE is that it treats specifically the emission from the heavily star-forming regions (HII+photodissociation regions, PDR in the following) using the photoionisation code MAPPINGIII (Dopita et al., 2005; Groves et al., 2008). An other important point is that SUNRISE propagates the light using the polychromatic algorithm, which speeds up considerably the radiative transfer calculation.
The code is designed as a suite of three programs that have to be run sequentially (SFRHIST, MCRX and BROADBAND). I detail below these programs but more details are given in Jonsson (2006) and Jonsson et al. (2010).

SFRHIST loads the star and gas distributions of the simulated galaxy. It computes the initial spectrum of the galaxy (from the stellar particles as described above) using the STARBUST99 model. The MAPPINGIII model is used if the particle is younger than 10 Myr. The ray tracing is performed on a grid. The adaptive grid is generated at this stage to store the dust density distribution according to the geometry of the gas distribution. Dust density is estimated using a dust-to-metal ratio of 0.4 (see above). The Milky Way Weingartner and Draine (2001) dust model is used for the grain cross-section and the Draine and Li (2007) one for modelling the PAH emission.

The core part of the work is done by the code MCRX which performs the radiative transfer calculation. The code takes in input a list of camera positions at a reasonable distance of the galaxy (with a specific field of view and pixel size), together with the gas and metallicity distributions, a model for the dust as well as various parameters which define the way to compute the transfer.
Solving the radiative transfer means calculating the propagation of the ray through the ISM which can absorb, scatter and emit light. The most difficult point is the computation of the scattering. The propagation is accounted with a Monte Carlo method. The ray tracing is performed by propagating a range of wavelengths simultaneously, which speeds up the code for only one random walk has to be performed for all wavelengths. However, to correct the fact that cross sections are different depending on the wavelength, the intensity of each wavelength is rescaled before each interaction (scattering and absorption) by a biasing factor to compensate the difference in the probability of interaction.
The output rays in the direction of the camera are finally projected onto an image plane. The output of this program are thus data cubes, giving the spectra at each pixel of the images seen by the camera. There is one data cube for the un-attenuated spectra, one for the attenuated spectra and one for the spectra emitted by the heated dust grains. In the following, I do not analyse the emitted light in the IR.

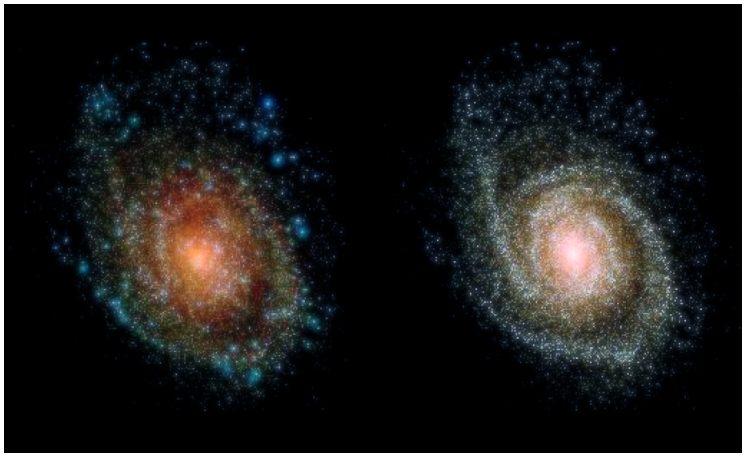

Figure 4.6: A 3-colour image in the NUV, R, K bands of a galaxy in the Horizon-AGN simulation, with dust (*Left*) and without (*Right*). This picture emphasises the effect of the dust on galaxies: large scattering in the spiral arms, very strong attenuation in the centre of the galaxy.

broadband convolves the output spectra of mcrx with the chosen filter passbands to compute the fluxes. The output is a fits file which contains the photometric images in the given bands, for the given camera positions, as well as the total fluxes of the galaxy in each direction.

A result is given in Figure 4.6, showing a simulated galaxy in the NUV-r-K bands, with and without dust. This picture emphasises the effect of the dust on galaxy: large scattering in the spiral arms and very strong attenuation in the centre of the galaxy. Figure 4.8 presents the non-attenuated, attenuated and re-emitted spectra of an other simulated galaxy.

Limitations of dust modelling with sunrise in Horizon-AGN – The main limitation which should be noted (related to the simulation and not to the sunrise code) is the low resolution of Horizon-AGN. This is a limitation because clumping on scale below the limit scale in the simulation (1 kpc) will change the opacity of the dust at larger scale and therefore the attenuated galaxy spectrum.
In addition, I used sunrise mainly with the default parameters. To test the effects of changing these parameters on the dust attenuation, I slightly modified their values around the recommended ones, without noticing dramatic changes on the result. However a more careful and systematic exploration of the parameter space could allow a better calibration of the model. In particular, the choice of the dust-to-metal ratio, the chosen dust extinction curves, and the PDR covering fractions are clearly expected to affect the attenuated galaxy spectrum.
Finally an other point already outlined in Chapter 3 is that dust affects also the dynamics of the gas and the star formation itself. As such the radiative transfer should be treat in conjunction with the evolution of the simulation (dynamics of the gas and star formation), which is not done here.

Comparison sunrise-sunset – As noted above, processing all simulated galaxies using the sunrise approach is prohibitively time-consuming. Therefore, I explored the potential differences between sunrise and sunset for a small random sample of galaxies at $z \sim 0.3$. This comparison is shown in Figure 4.7.
I found that, while the optical magnitudes are almost identical using the two approaches, the predicted UV magnitudes are somewhat fainter when they are calculated via full radiative transfer and show a dependence on the UV magnitude itself. First of all the spectral population synthesis models used in sunrise and sunset are different (starbust99 and BC03 respectively) but this difference is not the cause for the observed discrepancy in the NUV, for this fact should mainly impact the NIR light (see Chapter 3). The difference between sunrise and sunset in the NUV outlines the importance to have a good dust modelling and allows to quantify the uncertainties connected to simplified models. Nevertheless this difference is mainly observed in the NUV band, because the cross section for scattering is much smaller in optical and NIR bands than in UV bands. Hence in optical and NIR sunset is good enough to correctly

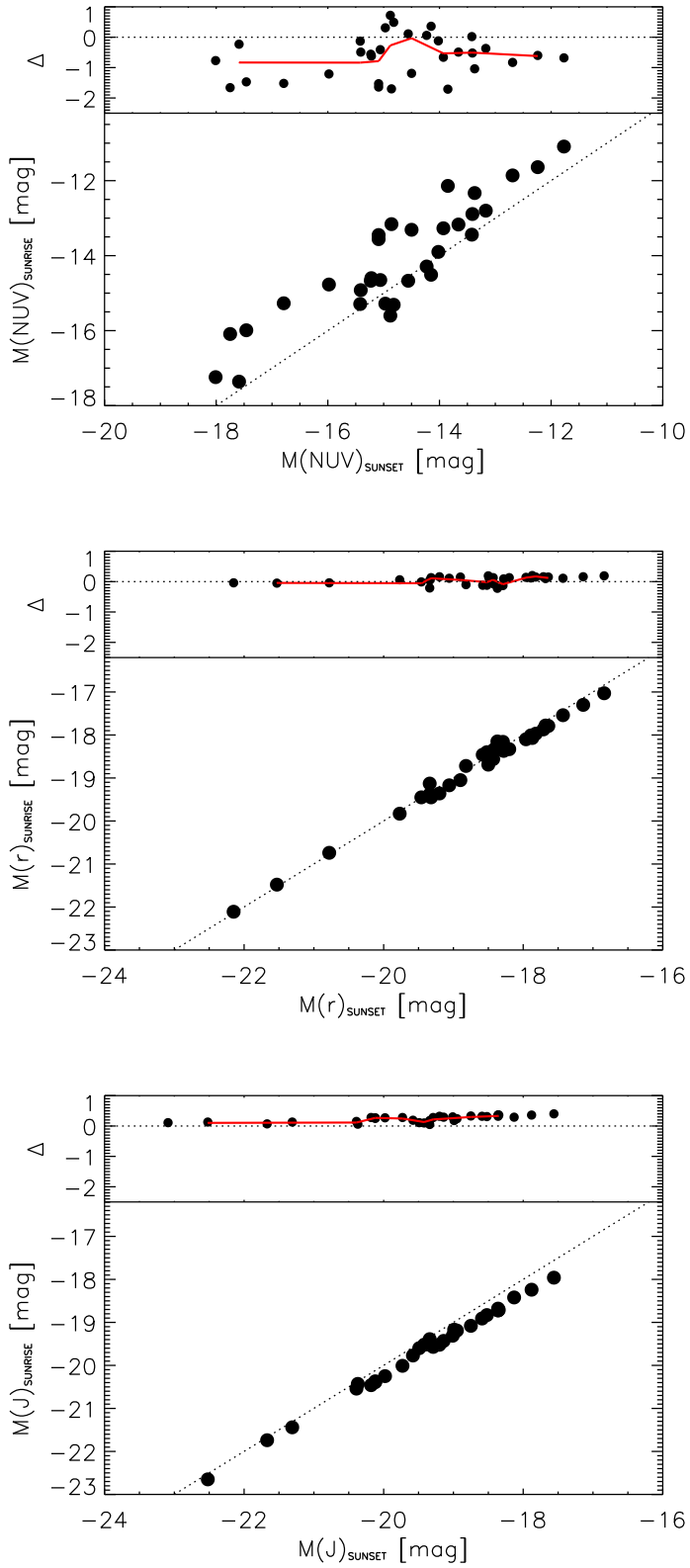

Figure 4.7.: Comparison of the galaxy photometry for a few galaxies when taking into account the effect of dust with either SUNSET code (x-axis) or SUNRISE (y-axis) code in the NUV band (*top*), r band (*middle*) and J band (*bottom*). While the difference is negligible in optical or NIR bands, the UV light is sensitively different.

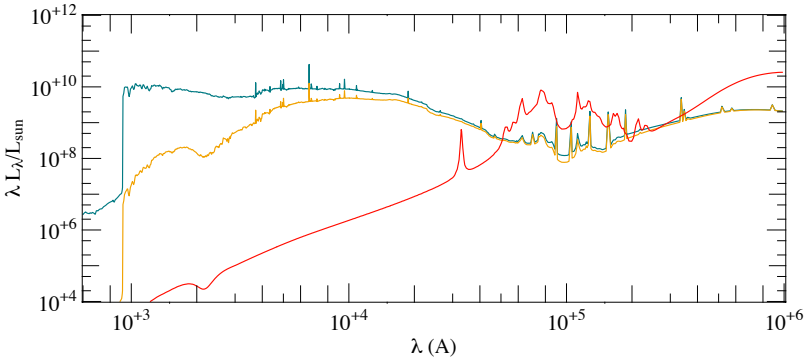

Figure 4.8: The non-attenuated (blue), attenuated (yellow) and re-emitted (red) spectra computed by SUNRISE on a galaxy in HORIZON-AGN.

account for dust, but it fails in the UV because it simply neglects dust scattering.
It is worth noting here that there are additional uncertainties in the treatment of dust that may further complicate the comparison between theory and observation. As already acknowledged above, the dust-to-metal ratio is assumed to be a fixed (Milky-Way-like) value in my analysis but could in fact varies as a function of metallicity and redshift. In a similar vein, the extinction law (assumed in this study to be Milky-Way-like), may also vary as a function of galaxy properties like age and metallicity.

4.4 Tracing the skeleton with DisPerSE

To identify the cosmic network from the density, I use the persistence based filament tracing algorithm (DisPerSE, Sousbie, 2011), a theoretically and numerically robust tool which enables to extract the topological structures: walls, filaments, nodes, from a density field. I detail below the main aspects of the algorithm.

The Delaunay tessellation – DisPerSE[4] operates directly from the particle distribution. This is an important asset to work with discrete datasets such as galaxy catalogue. It computes the density field from the Delaunay tessellation. The Delaunay tessellation of a set of points is a triangulation such that for each triangle in the tessellation no points of the distribution lies in the circumcircle of this triangle (or, in three dimensions, in the circumsphere of the tetrahedron). Once computed the triangulation, the density in the centre of the basis cell is defined as being proportional to the inverse of the volume of the cell. Computing this triangulation enables to have a density field for which the resolution depends on the density: high density regions are very well sampled, while the density in low-density regions is never set to zero.

Identification of the structures – The method of filaments and walls identification is based on Morse theory. The Morse-Small complex is defined as the set of integral lines and critical points of the density field. Critical points are points where the gradient of the density is null. In two dimensions, critical points are extrema, saddle, and minima. In three dimensions, there are two types of saddle points (see Chapter 2). Integral lines start (or end) at critical points, and are the curves tangent to the gradient in every point. They determine a tessellation of the space, each cell (manifold) of the tessellation being covered by critical lines originating (in the case of the ascending skeleton) or leading (in the case of the descending skeleton) to the same critical point. In the case of the ascending skeleton (simply called "skeleton" in the following) in three dimensions, filaments are ascending 1-manifolds (one-dimensional regions) defined as the special lines

[4] http://www2.iap.fr/users/sousbie/web/html/indexd41d.html?

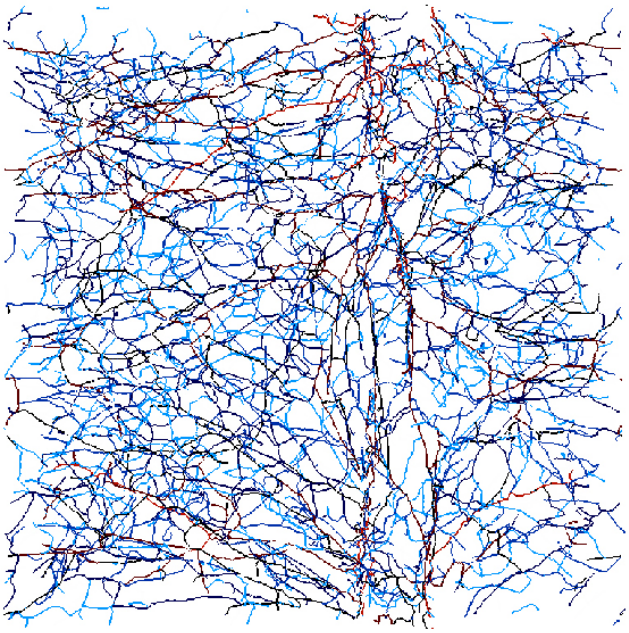

Figure 4.9: The skeleton measured in a simulation (boxsize: 20 Mpc) for increasing persistence threshold, 0.06, 0.12, ..., 2, from light blue to red; the skeleton has tree-like structure where the main branches correspond to the most persistent ones.

following the gradient and originating from saddle points of type 1 (filament-type saddle points). Walls are ascending 2-manifolds (two-dimensional regions) defined as the patch originating from the saddle points of type 2 (wall-type saddle points). Voids are ascending 3-manifolds (three-dimensional regions) defined as the volume originating from minima.

Identification of the filaments and persistence – Each filament is defined as a set of small segments connecting neigbouring points together. The robustness of a topological pair of critical points (maximum-saddle, saddle-saddle or saddle-minimum) is estimated through persistence.
Persistence is defined as the ratio of the density value at the two critical points of the pair. This ratio quantifies the robustness of the underlying topological feature characterized by the pair (the fact that the connected critical points are responsible for the appearance of a new topology, connected component, tube, ball, in the corresponding excursion set).
Expressed in terms of numbers of σ, persistence quantifies the significance of the critical points pair in the Delaunay tessellation of a random discrete Poisson distribution. For instance, a skeleton with persistence level $n_\sigma = 6$ is a skeleton for which all the persistence pairs which have their probability to be found in a random discrete Poisson distribution below 6σ of the mean have been removed.
Removing low-persistence pairs is a multi-scale non-local method to filter noise/ low significance filaments. This method is particularly well adapted to noisy datasets such as redshift catalogues (which will be useful for the work presented in Chapter 8). In particular, Sousbie (2011) shows that even with a very noisy dataset (*e.g.* considering that one particle over two is noise) it is still possible to use persistence to recover the underlying structures. Probing filaments with different persistences corresponds to probing different structure scales.
Figure 4.9 gives visual impression of the corresponding structure of the skeleton as a function of these persistence levels: the skeleton has a tree-like structure, for which each level of lower persistence contributes to smaller branches.

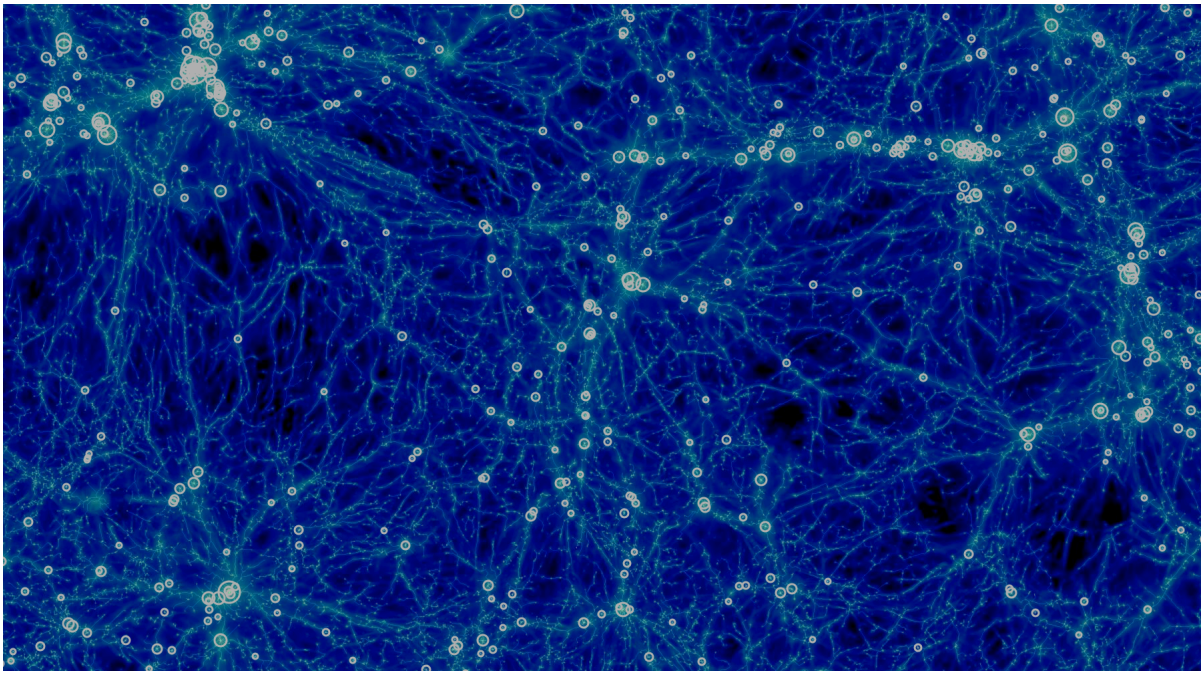

Figure 4.10.: Gas density (dark blue), persistent skeleton (green) and dark haloes (white circles) in a thin slice of the HORIZON-AGN simulation at $z \sim 0.1$.

4.5 Summary

The hydrodynamical cosmological HORIZON-AGN simulation presented here allows a comprehensive analysis of galaxy evolution both statistically (thanks to the large cosmological volume) and down to relatively small scales.

- The simulation models jointly the gas, dark matter and stellar evolution. The consistent implementation of stellar and AGN feedback at the subgrid scale help to understand in detail which processes are shaping galaxy stellar mass growth and regulating star formation;
- The lightcone extracted from the simulation, for which the geometry mimic observed surveys, enables a better comparison with observational datasets. Virtual photometry has been computed for the simulated galaxies and realistic synthetic images have been created. This step is important to build a bridge between theory and observations which are otherwise not directly comparable;
- The DISPERSE tool to extract the filamentary structures of the cosmic web allows for a robust study of the anisotropic environment and its possible effects on galaxy properties.

In turn, HORIZON-AGN is a powerfull tool to understand in detail which processes are contributing to galaxy stellar mass growth, regulating star formation and shaping galaxy properties. The virtual photometric catalogue provides a solid basis to be confronted with the observations and in particular with the photometric COSMOS2015 catalogue which I present in the following Chapter.

Chapter 5

Observational methods: COSMOS2015

This Chapter presents the computation of the new photometric multi-wavelength catalogue COSMOS2015 on the COSMOS field, as described in Laigle et al. (2016). This catalogue combines more than 30 photometric bands from NUV to IR. Particular care has been given to the maximisation of the number of detected sources (from the $z^{++}YJHK_s$ χ^2-image), the homogenisation of the data and the computation of upper limits[1]. This careful extraction is evident in particular in terms of redshift accuracy. The ambition of COSMOS2015 is to provide photometry, redshift measurements and stellar masses for all the objects such as $K_s < 24$ ($K_s < 24.7$ in the ultra-deep regions, at 3σ in a $3''$ aperture) up to $z = 6$ over the 2 deg^2 in the COSMOS field. In other terms, it is primarily intended to constitute a reliable basis to follow consistently galaxy populations above a certain mass on the redshift range $0.5 < z < 6$. But COSMOS2015 is much more than that. Because of the relatively large area and the exquisite photometric redshift accuracy, COSMOS2015 can also be applied to the study of the large-scale structure. I will develop this study from two-dimensional photometric slices in Chapter 8. In this sense, COSMOS2015 may be also considered as a pilot study for such large-scale structure analysis from upcoming photometric surveys. Additionaly, the recent observational programs in x-ray (the Chandra Cosmos Legacy Survey, Civano et al., 2016; Marchesi et al., 2016) and radio (the VLA-COSMOS Survey Smolčić et al., 2014) on the COSMOS field combined with the COSMOS2015 catalogue allow to look at populations of particular objects, such as intensively star forming galaxies, active galactic nuclei and radio galaxies, and to constrain uniquely the co-evolution of galaxies and AGN (see *e.g.* Magliocchetti et al. accepted, Delveccio et al. in prep). For all these reasons, COSMOS2015 has become the reference photometric catalogue on COSMOS field and is used by quite a fews teams since it has been made public[2].

In Section 5.1 I introduce the COSMOS field and its multi-wavelength coverage. I do not aim to make an exhaustive description of all the bandpasses available on COSMOS, but I report on all the datasets which are included in the COSMOS2015 catalogue. If the photometry has been extracted independently of COSMOS2015, I specify how the match with the COSMOS2015 sources has been done. The extraction of the photometry for the optical and NIR bandpasses is then presented in Section 5.2. I explain the computation of the physical parameters through SED fitting in Section 5.2.4. In Section 5.3 is presented a critical assessment of the performances of the catalogue. Finally, in Section 5.4, I detail the selection of galaxies populations.

[1] Upper limits are given in each photometric band as the mean magnitude brighter than which objects have more than 99.6% of chance to be real and not noise.

[2] The COSMOS2015 catalogue is distributed via anonymous ftp: `ftp://ftp.iap.fr/pub/from_users/hjmcc/COSMOS2015/` and through the usual astronomical archive systems (CDS, ESO, IRSA).

Table 5.1.: Summary of available galaxy catalogues of photometry and photometric redshifts on COSMOS field. The given redshift range is approximative. In the table, "main data" refers to the main dataset used for the photometry extraction, which drives essentially the performance of the catalogue. "lim. (band)" refers to the limiting magnitude in a 3″ aperture at 3σ and the band on which this limiting magnitude is computed. "comments" refers to the main observables given in the catalogue. Additionally to these catalogue, AGN catalogues and spectroscopic catalogues are available on the field.

Name	Reference	Main Data	Selection	Lim. (Band)	z	Comments
ACS catalogue	Leauthaud et al. (2007)	ACS HST data	$F814W$	26.5	[0,1]	photometry & photo-z
i-band v2.1	Capak et al. (2007)	30 bands	$i_{\rm AB}$	26.2 (i)	-	photometry
i-band v2.1	Ilbert et al. (2009)	30 bands	$i_{\rm AB}$	26.2 (i)	[0,2]	photo-z
NIR-selected	McCracken et al. (2012)	UVISTA-DR1	$YJHK_{\rm s}$	23.9 ($K_{\rm s}$)	-	photometry
NIR-selected	Ilbert et al. (2013)	UVISTA-DR1	$YJHK_{\rm s}$	23.9 ($K_{\rm s}$)	[0,3]	photo-z
cosmos2015	Laigle et al. (2016)	UVISTA-DR2	$YJHK_{\rm s}$+z^{++}	24.7 ($K_{\rm s}$)	[0,4]	photometry & photo-z

5.1 The Cosmic Evolution Survey: Imaging the high-z Universe

5.1.1 The COSMOS field

The Cosmic Evolution Survey (COSMOS[3] Scoville et al., 2007) probes the evolution of galaxies and AGN as a function of cosmic time and environment.
The COSMOS survey is centred at [+150.11916667,+2.20583333] and is accessible by all current and future facilities. The main datasets covers a 2 deg^2 region around this centre. However, some particular bands are slightly smaller or larger (*e.g.* 1.4 deg^2 for the NIR region from UltraVISTA). To get an idea, 1.4 degree a side at $z = 0.5$ represents 54 comoving Mpc, 82 Mpc at $z = 1$, 130 Mpc at $z = 2$ and 159 Mpc at $z = 3$, which provides a reasonable surface to envisage the study of the large-scale structure.
COSMOS includes deep observations from a number of ground-based (Keck, Subaru, VLA, ESO-VLT, UKIRT, CFHT, *etc.*) and space-based (Hubble, *Spitzer*, GALEX, XMM, Chandra, Herschel, NuStar) facilities. In fact, the field has been observed in photometry by all accessible wavelengths from x-ray to radio. In top of that, there are several large spectroscopic programs on COSMOS. From the point of view of photometry, spectroscopic redshifts are a great asset to calibrate the SED-fitting procedure and then to get more precise photometric redshift estimation (cf. Section 3.3).
In the previous years, a number of multi-wavelength catalogues have been already extracted on the field as new data releases progressed. The catalogue cosmos2015 presented is in line with these previous efforts aiming at systematically extract the best information from available data. Existing catalogues, their specificities and performances are summarised in Table 5.1.

5.1.2 The vital role of NIR: the UltraVISTA program

Importance of the NIR for studying the high redshift Universe – I have recalled in Chapter 1 the essential role of NIR to consistently follow galaxy populations over a large range of redshift. Firstly, spectral features which are essential to compute galaxy redshifts are moved towards the redder wavelengths as the galaxy redshift increases (the Balmer break enters into the NIR bandpasses at $z \sim 1.5$ and up to $z \sim 5.2$). NIR bandpasses play then an important role for accurately determining the redshifts and minimising the

[3] http://cosmos.astro.caltech.edu

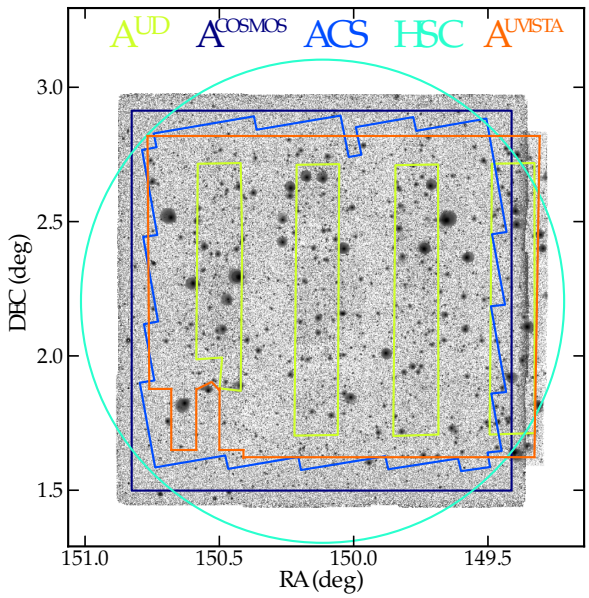

Figure 5.1: Schematic of the COSMOS field showing all of the optical (dark blue and turquoise) and NIR (green and orange) observations used in COSMOS2015. The background image corresponds to the χ^2 YKHK$_s$-z^{++} detection image. For reference, the region covered by the COSMOS-Advanced Camera for Surveys (ACS) HST data (Koekemoer et al., 2007) is shown in cyan. $\mathcal{A}^{\rm COSMOS}$ defines the 2 deg^2 COSMOS square (dark blue). $\mathcal{A}^{\rm Uvista}$ (orange area) is the region covered by the UltraVISTA-DR2 observations. I define $\mathcal{A}^{\rm UD}$ as the light green area, corresponding to the ultra-deep stripes in the UltraVISTA-DR2 observations.

number of catastrophic failures. Secondly, the selection of the samples should rely on a part of the galaxy spectrum for which most of the light is emitted by stars, and consequently which is a good proxy for galaxy stellar mass, specifically the optical and NIR continuum. More precisely, the optical and NIR continuum is the part of the spectrum reflecting the best the long-lived star light in the galaxy. It is well covered by the NIR wavelength range ($[1\ \mu\mathrm{m}, 2.5\ \mu\mathrm{m}]$) from $z = 0.5$ to $z = 4$. At higher redshifts, IR photometry would allow a better sample selection. In all the cases, a purely optical (apparent) selection will dramatically miss some galaxies at redshift above 1, as it would detect galaxies on a rest-frame near-ultraviolet light basis, which may be heavily obscured by dust in young galaxies. Obtaining comparable galaxy samples in terms of mass necessitates to select them at progressively redder wavelengths at higher redshifts. That is why at high redshift, NIR and IR data become determinant to select objects in order to complement optically selected objects at lower redshift. Aware of this vital role of NIR previously realised (Lilly and Longair, 1984; Cowie et al., 1990), a number of observational programs have made a particular effort on obtaining deeper NIR data during the last decade (*e.g.* among others UKIDSS, Lawrence et al. (2007); WIRDS, Bielby et al. (2012); McCracken et al. (2010), and now UltraVISTA McCracken et al. (2012) on COSMOS, as described below). The advent of NIR observations has been made possible thanks to improved detector technology and data collection methods: the sky background in the NIR is extremely bright. This requires that only short exposures are taken to avoid detector saturation, and to estimate in turn the sky background which must be subtracted from all images.

The NIR coverage used in COSMOS2015 – The $YJHK_s$-band data used for the computation of COSMOS2015 in this thesis were taken between 2009 December and 2012 May with the VIRCAM instrument on the *VISTA* telescope as part of the UltraVISTA survey program[4]. The layout of UltraVISTA observations is composed of deep stripes alternated with ultra-deep stripes. The UltraVISTA-DR2 processing steps are the same as those in the first release (DR1, McCracken et al., 2012). The final images are stacks of short-exposure images. A particular care has been taken for an uniform sky background removal to allow the detection of very faint NIR objects. Compared to the first UltraVISTA release (DR1), the exposure time has been increased significantly in the ultra-deep stripes, as shown in yellow in Figure 5.1; they cover a discontinuous area of 0.62 deg^2. An important consequence of this is that the signal-to-noise ratio for an object of a given magnitude is not constant across the image due to the variable depth. The

[4] http://www.eso.org/sci/observing/phase3/data_releases/uvista_dr2.pdf

alternating between ultra-deep ($\mathcal{A}^{\rm UD}$) and deep ($\mathcal{A}^{\rm Deep}$) stripes is also visible in the variation of the seeing as a function of right ascension (RA). However, this variation is of order of $\sim 0.05''$ and can be considered as negligible.
The UltraVISTA area is slightly smaller than the rest of the field. To provide NIR photometry in zones non covered by UltraVISTA, I include H and K WIRCAM data (from the WIRDS program, see McCracken et al., 2010) in my photometric catalogue (see Figure 5.1). Nevertheless, I do not assess the performance of the catalogue in these zones.

The extraction of the NIR photometry for the cosmos2015 catalogue will be detailed in section 5.2.

5.1.3 The IR dataset as part of the SPLASH program

Importance of the IR for the high-redshift study – IR observations are complementary of NIR observations to determine galaxy masses up to redshift 4, and became clearly crucial at higher redshift where all the rest-frame optical light is shifted in the IR. Furthermore, at low redshift IR datasets are very useful to properly account for the dust emission in nearby galaxies and thus to better determine the SFR.

The SPLASH program – The *Spitzer* Large Area Survey with Hyper-Suprime-Cam (SPLASH) is a program aiming to collect conjointly ultra-deep IR data from the *Spitzer* telescope (IRAC imaging) and optical data from the Hyper-Suprime-Cam (HSC) imaging on Subaru (see the description of this dataset below) on the two fields COSMOS and SXDS. The purpose of this program is to improve stellar mass estimates in the redshift range $4 < z < 6$.

IR data used in cosmos2015 – The 3.6 μm, 4.5 μm, 5.8 μm and 8.0 μm (respectively, channel 1, 2, 3, and 4) IRAC data used in cosmos2015 consist of the first two-thirds of the SPLASH COSMOS dataset (see Steinhardt et al., 2014) together with S-COSMOS (Sanders et al., 2007), the *Spitzer* Extended Mission Deep Survey, the *Spitzer*-CANDELS survey data, along with a several smaller programs that observed the COSMOS field. The final processing is described in Capak et al. in preparation. The average exposure time per pixel is 3.8 hr, increasing to 50hr in the central S-CANDELS coverage. Before processing, a median image was created for each AOR (observing block) and subtracted from the frames to remove residual bias in the frames and persistence from previous observations. For the S-CANDELS data, a secondary median was subtracted from the observations taken with repeats to remove the "first frame effect" residual bias. The resulting median-subtracted images have a mean background near zero, so no overlap correction was applied. The median subtracted frames were then combined with the MOPEX mosaic pipeline[5]. The outlier and box-outlier modules were used to reject cosmic rays, transients, and moving objects. The data were then drizzled onto a $0.6''$ pixel scale using a "pixfrac" of 0.65 and combined with an exposure time weighted mean combination. Mean, median, coverage, uncertainty, standard-deviation, and colour-term mosaics were also created. Obviously, this variation as a function of position can be expected to have an influence on the precision of the photometric redshifts and stellar masses for the very high redshift ($z > 4$) objects.
The extraction of the IR photometry for the cosmos2015 catalogue will be detailed in section 5.2.

[5] http://irsa.ipac.caltech.edu/data/SPITZER/docs/dataanalysistools/tools/mopex/

5.1.4 NUV and optical dataset

Importance of the UV-optical coverage at low redshift – UV photometry is very important at low redshift, and especially valuable to determine instantaneous star formation rate. Progressively as the spectrum is redshifted, optical bands start to be useful to image the rest-frame UV. At redshift $z > 2.5$, the Lyman break enters in the optical bands, which start thus to play an important role to improve the redshift accuracy by constraining its position (see Section 3.3).
Among optical bands, intermediate bands are very useful to increase the spectral resolution of the measured SED in the apparent optical wavelength range, contributing in this way to the accuracy of photometric redshifts especially at low redshift. For example, Cardamone et al. (2010) noted a factor of 3 improvement in the photometric redshift accuracy at $z < 1.2$, when the Balmer break falls into the intermediate bands coverage. They also reduce the fraction of catastrophic failures. The improvement is less important at higher redshifts. Intermediate bands are also important to estimate the emission lines.

The UV-optical coverage used in COSMOS2015 **–** The UV-optical dataset used in this thesis for the COSMOS2015 catalogue is similar to the one used in previous releases and is described in Capak et al. (2007); Ilbert et al. (2009).
NUV ($0.23\mu m$) observations come from the spatial telescope GALEX (Zamojski et al., 2007). In the catalogue, FUV ($0.15\mu m$) observations are also included, but not used in the SED-fitting procedure.
u^*-band data is taken from the Canada-France Hawaii Telescope (CFHT/MegaCam) and covers the entire COSMOS field, obtained with 5 pointings of MegaCam. Additional optical bands (z, g, Y, i, r) from CFHT are existing on COSMOS field, but as they cover only 1 deg^2 on the field (1 pointing of MegaCam), they are not used in the final catalogue computation.
The COSMOS-20 survey provides most of the optical coverage: six broad bands (B, V, g, r, i, z^+), twelve medium bands ($IA427$, $IA464$, $IA484$, $IA505$, $IA527$, $IA574$, $IA624$, $IA679$, $IA709$, $IA738$, $IA767$, and $IA827$), and two narrow bands ($NB711$, $NB816$), taken with Subaru Suprime-Cam (Taniguchi et al., 2007, 2015). I have discarded poor seeing ($\sim 1.3''$) g-band data. Finally, the initial COSMOS z-band data were replaced by deeper z^{++} band data taken with thinned upgraded CCDs and a slightly different filter. At this stage, in each band, image point-spread functions[6] (PSFs) were homogenized to minimize tile-to-tile variations (Capak et al., 2007, and see the discussion in Section 5.2.2). RMS_MAP (root-mean-square map, estimating for each pixel in the image the precision of the measurement) and FLAG_MAP (map in which bad pixels are tagged) images were also generated for each of the bands, and saturated pixels and bad areas were flagged.
The extraction of the NUV and optical photometry for the COSMOS2015 catalogue will be also detailed in section 5.2.

5.1.5 Hyper-Suprime-Cam optical observations as part of the SPLASH Program

This catalogue also contains new Y-band data taken with Hyper-Suprime-Cam (HSC) Subaru (Miyazaki et al., 2012). The average exposure time per pixel is 2.1 hr. This data set is described fully in Hasinger G. et al. in preparation. The addition of the Y-band data is intended to improve stellar mass and redshift estimates in the important $1 < z < 1.5$ range because it is slightly bluer than the Y filter from VIRCAM, but it is also intended to serve as a "pilot program" to assess the utility of HSC data and to prepare for future COSMOS data sets which will include much more HSC imaging in the following bandpasses: g, i, z and Y. When the SPLASH program will be completed, the Y-band from HSC is expected to be the

[6]The point spread function describes the response of the camera to a point source, such as a star. It is often assumed to be well represented by a Gaussian function, but is generally much more complicated.

deepest ever taken in the COSMOS field ($\sim$ 1 magnitude deeper than Y from UltraVISTA data).

5.1.6 X-ray data with the Chandra COSMOS-Legacy survey

The role of x-ray data in the AGN selection – Photometric redshift estimation in COSMOS2015 does not rely directly on x-ray data, except for the AGN classification. Combined with optical and radio counterpart, x-ray data are indispensable to select AGN and investigate their properties (see Section 3.2). X-ray emission in galaxies is mostly due to the inverse Compton up-scattering of UV photons from the black hole thermal accretion disk, and the contamination from non-nuclear emission (star formation processes) is insignificant at these energies. The COSMOS field has been first imaged in the x-ray with XMM-COSMOS (Hasinger et al., 2007; Cappelluti et al., 2007; Brusa et al., 2010), and then by Chandra observations (C-COSMOS, in the 0.9 deg^2 central part, Elvis et al., 2009; Civano et al., 2012).

The x-ray coverage matched with COSMOS2015 – The new *Chandra COSMOS-Legacy* Survey (Civano et al., 2016; Marchesi et al., 2016) is a combination of the old C-COSMOS survey with new Chandra observations. It covers the 1.7 deg^2 of the COSMOS/HST field. The released catalogue contains 4016 x-ray sources down to a flux limit of $f_X \simeq 2 \times 10^{-16}$ erg s^{-1} cm^{-2} in the 0.5-2 keV band: 3755 of these sources lie inside the UltraVISTA field of view. The *Chandra COSMOS-Legacy* catalogue was matched with the COSMOS2015 catalogue using the likelihood ratio technique (Sutherland and Saunders, 1992). This method provides a much more statistically accurate result than a simple positional match, takes account the following: (i) the separation between the x-ray source and the candidate UltraVISTA counterpart; (ii) the counterpart K_s-band magnitude with respect to the overall magnitude distribution of sources in the field. Of the 3755 *Chandra COSMOS-Legacy* sources, 3459 ($\simeq$92%) have an UltraVISTA counterpart.

5.1.7 Radio data

The role of radio data in the AGN selection – As shown in Section 3.2, radio emission are important to classify AGN between radiatively efficient and radiatively inefficient as a signature of the underlying AGN activity. Radio emission may be also due to star forming galaxies. The combination with multi-wavelength data allows to distinguish the nature of the radio sources. Radio data are important to complement x-ray and UV-to-IR data, as selections in different bands may lead to selections of different populations: a multi-wavelength selection of AGN is essential for an exhaustive investigation (see Delvecchio I. et al. in preparation for a multi-wavelength study of AGN in COSMOS).

The radio coverage matched with COSMOS2015 – A large compilation of radio data is available on COSMOS field, at wavelength of 20 cm (1.4 GHz, Schinnerer et al., 2004; Schinnerer et al., 2007, 2010), 50 cm (610MHz, Kloeckner et al. in preparation) and 90 cm (320MHz, Smolčić et al., 2014). The VLA-COSMOS 3 GHz Large Project (P.I V. Smolcic) provides the most recent radio data on COSMOS field (10 830 sources, Smolcic et al. in preparation). The cross-match of the radio catalogue with the COSMOS2015 multi-wavelength catalogue is described in Baran et al. (in preparation). They combined a likelihood ratio technique and a nearest-neighbour matching within 1.2″ radius. With this technique, the fraction of false association is very low on average (0.6%).

5.1.8 Additional data

Far-IR photometry – Photometry at 24 μm was obtained for a total of 42633 sources using an updated version of the COSMOS MIPS-selected band-merged catalogue published by Le Floc'h et al. (2009). In this catalogue, 90% of the 24 μm-selected sources were securely matched to their K_s-band counterpart using the WIRCAM COSMOS map of McCracken et al. (2010), assuming a matching radius of 2 $''$. Counterparts to another 5% of the sample were found using the IRAC-3.6 μm COSMOS catalogue of Sanders et al. (2007), while the rest of the 24 μm source population remained unidentified at shorter wavelengths. The coordinates of the WIRCAM K-band or IRAC counterparts (or the initial 24 μm coordinates for the unidentified MIPS sources) are thus considered, and cross-correlated these positions with the VISTA catalogue using a matching radius of 1$''$. VISTA counterparts were found for all of the previously-identified 24 μm sources and for an additional set of 117 objects detected by MIPS which had no previous identification.

In COSMOS2015 is also provided FIR photometry obtained at 100, 160, 250, 350, and 500 μm using the PACS (Poglitsch et al., 2010) and SPIRE (Griffin et al., 2010) observations of the COSMOS field with the *Herschel* Space Observatory. The PACS data were obtained as part of the PEP guaranteed time program (Lutz et al., 2011), while the SPIRE observations were carried out by the HERMES consortium (Oliver et al., 2012). For each band observed with *Herschel*, source extraction was performed by a PSF fitting algorithm and using the 24 μm source catalogue as priors. Hence, far-IR matches to VISTA were unambiguously obtained from the 24 μm source counterparts described above, leading to a total of 6608 sources with a PACS detection and 17923 sources detected with SPIRE. Total uncertainties in the SPIRE bands include the contribution from confusion. Flux density measurements with a signal to noise smaller than 3 in the initial SPIRE COSMOS catalogue published by Oliver et al. (2012) are not considered in this work.

5.2 Computation of the COSMOS2015 catalogue

I present below how I computed the COSMOS2015 catalogue, starting from the extraction of the photometry to the assessment of the reliability of the catalogue. When the work has been done in collaboration and not specifically by myself, I mention it clearly.

5.2.1 Strategy for the photometry extraction

Dataset – NUV, optical, NIR and IR data which compose the core of the photometric catalogue are presented in Table 5.2, and the associated filter passbands are presented in Figure 5.2. For each filter, the quoted central wavelength is the median wavelength weighted by transmission and the widths are defined using the half-maximum transmission points. The computation of the 3σ depth[7] is detailed in the following. This value is given for 3$''$ and 2$''$ apertures for the NUV, optical, NIR dataset, but the value given for the IR dataset corresponds to total magnitudes.

The χ^2-image – Objects are identified based on a panchromatic detection. The multicolour detection from a composite image is presented for the first time in Szalay et al. (1999). In the previous Section 5.1, I explained how useful are the different passbands for detecting galaxies in different redshift ranges, or of different types. If one chooses only one passband to detect galaxies, one will surely miss a large number of objects. To overcome this difficulty, Szalay et al. (1999) proposes to build a "χ^2-image", where each pixel

[7]The level of magnitude above which one considers an object as part of the background.

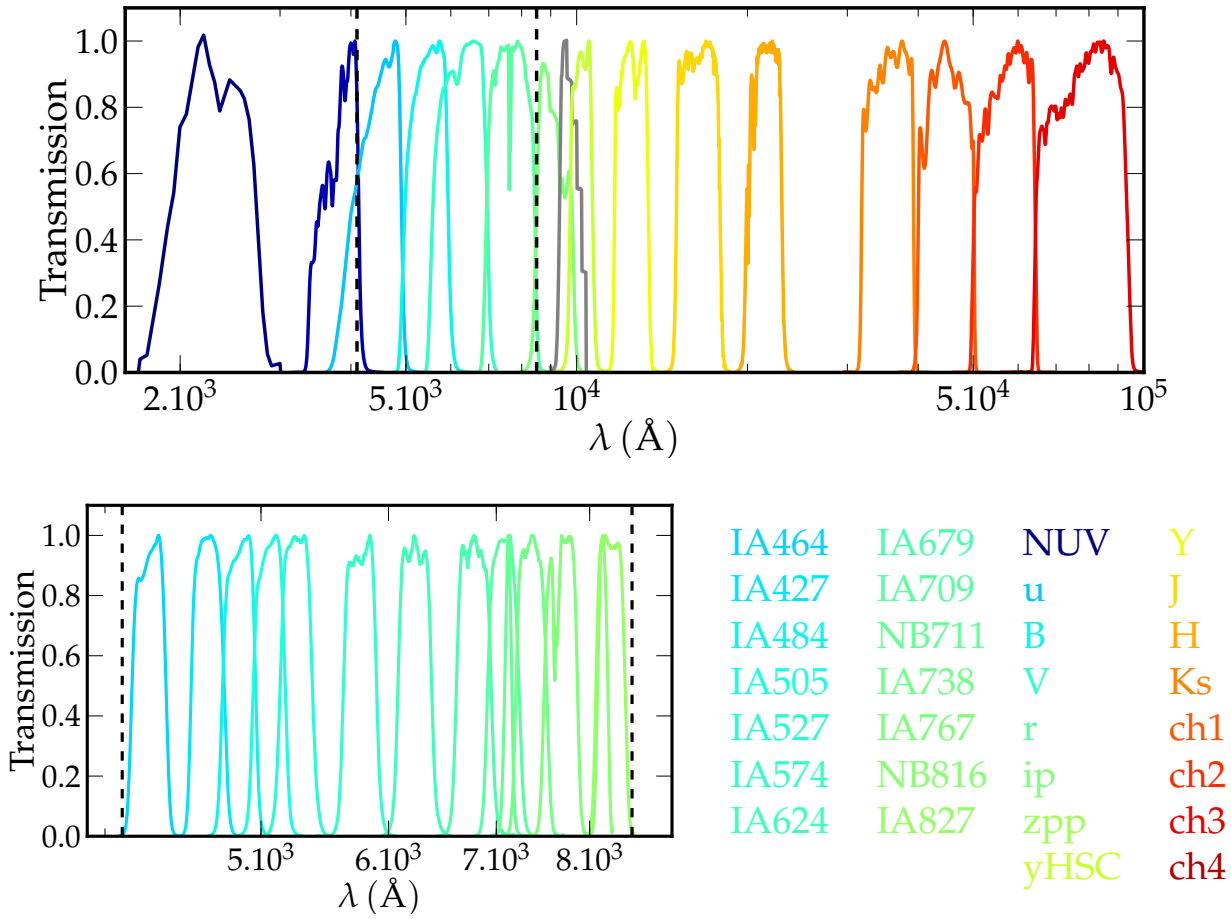

Figure 5.2.: Transmission curves for the photometric bands used. The effect of atmosphere, telescope, camera optics, filter, and the detector are included. Note that for clarity the profiles are normalized to a maximum throughput of one: therefore, the relative efficiencies of each telescope and detector system are *not* shown. Intermediate and narrow bands are represented in the *bottom* panel, but the region of the spectrum covered by these bands is marked by dashed lines in the *top* panel.

value is defined as:

$$F = \sqrt{\frac{\sum_i w_i f_i^2}{n_{\neq 0}}}\,, \tag{5.1}$$

where the summation is done on the different passbands used, f_i is the pixel flux value in each passband, w_i is proportional to the inverse of the variance at this position and $n_{\neq 0}$ is the number of non-zero pixels. This combination ensures us that if an object is visible in one passband but not in the other, it may still be detected. The χ^2 detection requires that the noise is uncorrelated between the images. As discussed in the previous Section, it is essential to proceed with a NIR selection to extract reasonably complete galaxy sample at high redshift. To maximise the number of detected objects, I built a detection χ^2-image from the 4 NIR images of UltraVISTA and the optical z^{++} image from Subaru. Including this optical band allows to select also low-redshift sources and UV-luminous sources at $z > 2$. The optical z^{++}-band data is preferred compared to the i^+-band because compact objects in the i^+ image from Subaru saturate around $i = 21$. The χ^2 $z^{++}YJHK_\mathrm{s}$ detection image is produced using SWarp (Bertin et al., 2002) starting with the non-homogenized images.

5.2.2 In-homogenous datasets

The primarily difficulty in the photometry extraction operation arises from the intrinsic in-homogeneity of the data. From one band to another, the shape and the width of the PSF vary, but also the depth. On top of that, considering a particular band, the PSF and the depth may vary *across* the field. As I intend to make an homogenous catalogue over a large redshift range by extracting photometry at fixed apertures, it is crucial to operate the homogenisation of the data before extracting the photometry.

PSF homogenisation – In this work, the variation of the PSF across individual images in a given band is neglected. This is reasonable because band-to-band variations are almost always greater than the variation within a single band. In the following, I however study how the variation across the field can affect the photometry.

From u to K_s the full width at half maximum (FWHM) of the PSF has a range of values between $\sim 0.5''$ and $1.02''$ (when fitted with a Moffat profile, see below). Therefore, the fraction of the total flux falling in a fixed aperture is band-dependent. One way to address this problem is to "homogenize" the PSF so that it is the same in all of the bands. GALEX and IRAC bands are not homogenized, because their photometry are extracted with a source-fitting technique, as detailed in the following.
The PSF is estimated from the point sources. In the first step in my homogenization process, SExtractor (Bertin and Arnouts, 1996) is used to extract a catalogue of bright objects. Stars are identified by cross-matching with point sources in the COSMOS-Advanced Camera for Surveys (ACS) HST catalogue (Koekemoer et al., 2007; Leauthaud et al., 2007). Saturated or faint stars are removed by considering the position of each object in the FWHM versus $\mathrm{m_{AB}}$ diagram, as shown on Figure 5.3. For each star I extracted a postage stamp using SExtractor.

The PSF is modelled in pixel space using PSFEx (Bertin, 2013) as a linear combination of a limited number of known basis functions:

$$\Psi_c = \sum_b c_b \psi_b\,, \tag{5.2}$$

where the c index reflects the dependance of Ψ on the set of coefficients c_b. Given a basis, this PSF model can be entirely determined knowing the coefficients c_b of the linear combination. The pixel basis is the most "natural" basis but requires as many coefficients as the number of pixels on the image postage stamp. I can then make some assumptions to simplify the basis and to reduce the number of coefficients. The

Table 5.2.: Available data and their average limiting magnitudes computed from variance map in 2 and 3″ diameter apertures.

Instrument /Telescope (Survey)	**Filter**	**Central λ (Å)**	**Width (Å)**	**3σ depth (3″/2″) ±0.1**
GALEX	NUV	2313.9	748	25.5
MegaCam/*CFHT*	u^*	3823.3	670	26.6/ 27.2
Suprime-Cam	B	4458.3	946	27.0/ 27.6
/Subaru	V	5477.8	955	26.2/ 26.9
	r	6288.7	1382	26.5/ 27.0
	i^+	7683.9	1497	26.2/ 26.9
	z^{++}	9105.7	1370	25.9/ 26.4
	$IA427$	4263.4	206.5	25.9/ 26.5
	$IA464$	4635.1	218.0	25.9 / 26.5
	$IA484$	4849.2	228.5	25.9/ 26.5
	$IA505$	5062.5	230.5	25.7/ 26.2
	$IA527$	5261.1	242.0	26.1/ 26.6
	$IA574$	5764.8	271.5	25.5/ 26.0
	$IA624$	6233.1	300.5	25.9/ 26.4
	$IA679$	6781.1	336.0	25.4/ 26.0
	$IA709$	7073.6	315.5	25.7/ 26.2
	$IA738$	7361.6	323.5	25.6/ 26.1
	$IA767$	7684.9	364.0	25.3/ 25.8
	$IA827$	8244.5	343.5	25.2/ 25.8
	$NB711$	7119.9	72.5	25.1/ 25.7
	$NB816$	8149.4	119.5	25.2/ 25.8
HSC/Subaru	Y	9791.4	820	24.4/ 24.9
VIRCAM	$Y^{\rm UD}$	10214.2	970	25.3/ 25.8
/VISTA	$Y^{\rm Deep}$			24.8/ 25.3
(UltraVISTA-DR2)	$J^{\rm UD}$	12534.6	1720	24.9/ 25.4
	$J^{\rm Deep}$			24.7/ 25.2
	$H^{\rm UD}$	16453.4	2900	24.6/ 25.0
	$H^{\rm Deep}$			24.3/ 24.9
	$K_{\rm s}^{\rm UD}$	21539.9	3090	24.7/25.2
	$K_{\rm s}^{\rm Deep}$			24.0/ 24.5
WIRCam	$K_{\rm s}$	21590.4	3120	23.4/ 23.9
/CFHT	H	16311.4	3000	23.5/ 24.1
IRAC/*Spitzer*	ch1	35634.3	7460	25.5/ o
(SPLASH)	ch2	45110.1	10110	25.5/ o
	ch3	57593.4	14140	23.0/ o
	ch4	79594.9	28760	22.9/ o

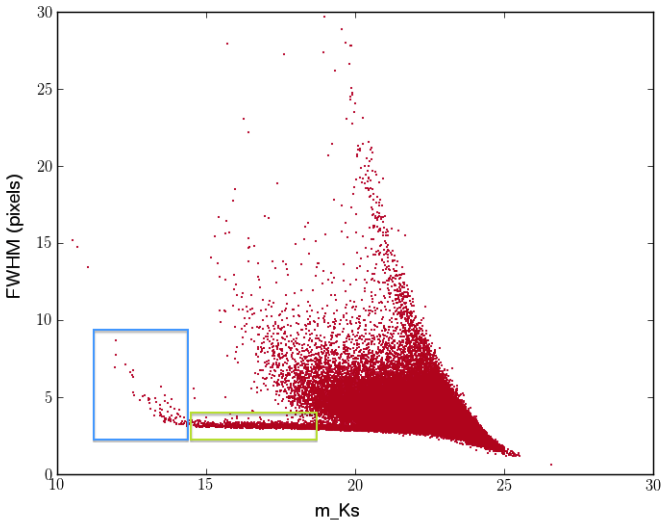

Figure 5.3: FWHM versus magnitude in the K_s band. On this figure, points in the green frame correspond to point sources objects. The blue frame contains saturated point sources objects. This diagram offers a convenient way to identify stars in the field.

adopted basis is the "polar shapelet" basis (Massey and Refregier, 2005), for which the components have useful explicit rotational symmetries. I assumed that the PSF is constant over the field. The global PSF of one band is then expressed as a function of the coefficients c_b at each pixel position x_i on the postage stamp image, which are derived by minimizing the χ^2 sum over all of the sources:

$$\chi^2(c) = \sum_{\text{sources } s} \sum_{\text{pixels } i} \frac{(p_s(x_i) - f_s \Psi_c(x_i))^2}{\sigma_i^2}\,, \tag{5.3}$$

where f_s is the total flux of the source s, σ_i is the variance estimate of pixel i of the source s, $p_s(x_i)$ is the intensity of the pixel i, and c refers to the set of PSF coefficients. Once the global PSF has been determined in each band, I then decided on the "target PSF", corresponding to the desired PSF of all of the bands after homogenization. This is chosen so as to minimize the applied convolutions. I used a Moffat profile to represent the PSF (Moffat, 1969); this provides a better description of the inner and outer regions of the profile than a simple Gaussian. The stellar radial light profile is

$$I_r = I_0[1 + (r/\alpha)^2]^{-\beta}\,, \tag{5.4}$$

with $\alpha = \theta/(2\sqrt{2^{1/\beta} - 1})$, $I_0 = (\beta - 1)(\pi\alpha^2)^{-1}$ and θ the FWHM. My target PSF is defined as a Moffat profile with $\mathcal{M}\,[\theta, \beta] = \mathcal{M}\,[0.8'', 2.5]\,.$

The required convolution kernel is calculated in each band by finding the kernel that minimizes the difference between the target PSF and the convolution product of this kernel with the current PSF. The images are then convolved with this kernel.

To estimate the precision of my PSF matching procedure, the photometry of the stars is extracted at 14 fixed apertures of radii r_k, logarithmically spaced between $0.25''$ and $2.5''$. In each band, the difference between the magnitude of the stars extracted in the aperture r_k and the total magnitude (computed from the $4''$ diameter aperture) is plotted in Figure 5.4 as a function of aperture. For comparison, the difference that would be obtained with the target profile $\mathcal{M}\,[0.8'', 2.5]$ is overplotted as a red dashed line. The agreement is excellent up to a $2''$ radius on the plot.

The flux obtained with the best-fitting PSF in each band is normalized to the target profile and also plotted in Figure 5.4 (left panel), before and after homogenization. For perfect homogenization, this ratio should be one, independent of aperture. For the $3''$ diameter aperture, the relative photometric error for point sources objects after homogenization is below 5% (or equivalently a difference of ∼0.05 in magnitude).

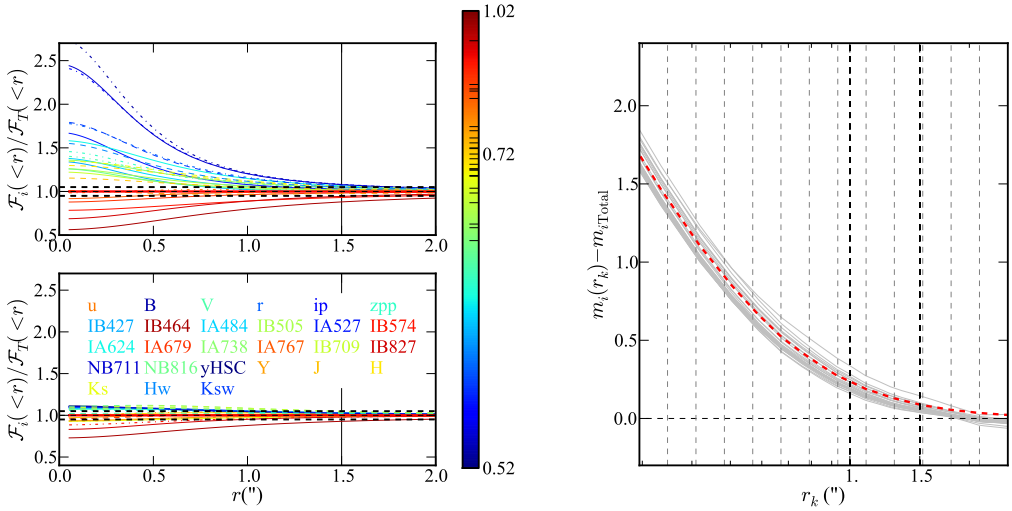

Figure 5.4.: *Left*: best-fitting stellar PSFs for all bands before and after homogenization (*left upper* and *lower* panels respectively), normalized to the target PSF $\mathcal{F}_T$. The vertical black solid line corresponds to the 3″ diameter aperture used for photometric redshifts. The horizontal dashed lines show the 5% relative error. The colour map reflects the increase in seeing before homogenization. *Right*: Median curves of growth (the difference between the magnitude in the kth aperture r_k with the total magnitude for point-like sources, estimated from the 4″ diameter aperture) as a function of aperture after homogenization for each band. The target Moffat profile with (θ,β)=(0.8,2.5) is shown in red. The vertical dark dashed lines are the apertures provided in the final public catalogue (2″ and 3″ diameters).

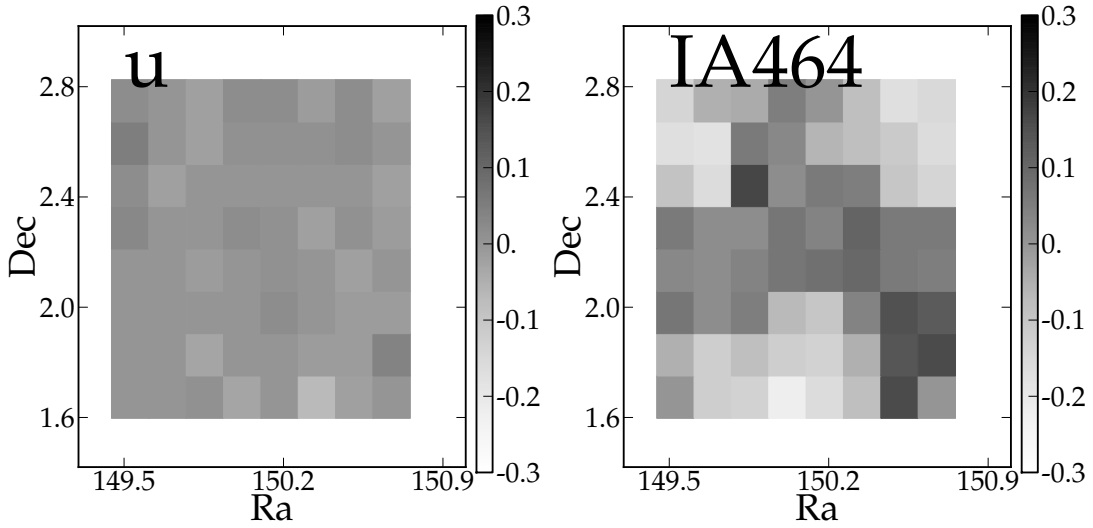

Figure 5.5.: Distribution of the difference local seeing and the median image seeing for the selected stars in the u and $IA464$ bands as a function of the position. Note that for these seeing estimations, I did not fit each star *individually* with a Moffat profile, but I used the Gaussian-profile-based FWHM_WORLD parameter from SExtractor. While the u band is relatively homogenous across the field, $IA464$ shows large positional variations. This is the most extreme case in my catalogue.

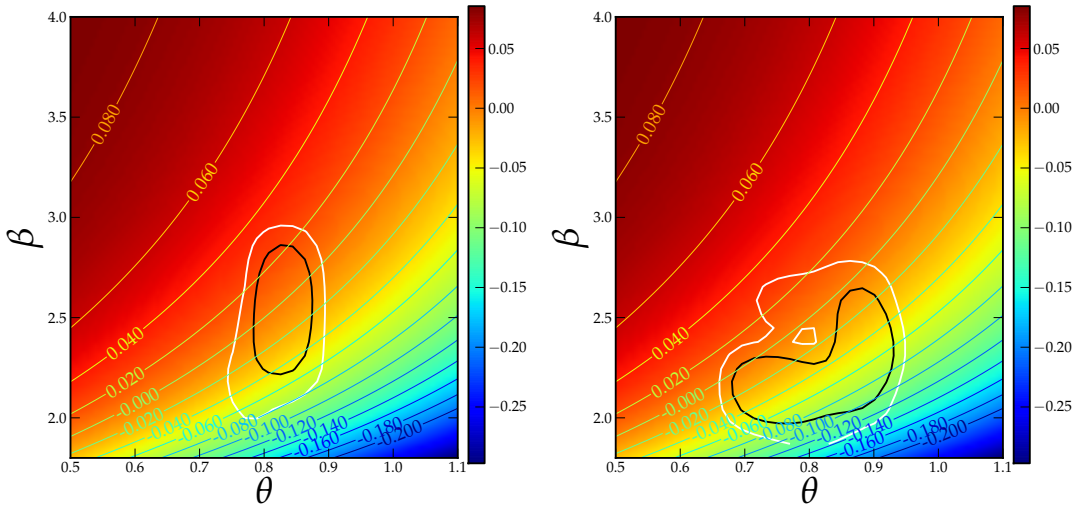

Figure 5.6.: Magnitude difference (measured in 3″ diameter apertures) between a point-like object convolved with the target PSF $\mathcal{M}\,[0.8'', 2.5]$ and with $\mathcal{M}\,[\theta, \beta]$ in the 2D parameter space $[\theta'', \beta]$. The black and white contours represent the regions which enclose 68% and 95% of the β-θ stellar distribution for two representative bands: u (*left*), which is relatively homogenous across the field, and $IA464$ (*right*) which is not.

Unfortunately, despite previous attempts at PSF homogenization inside each field (Capak et al., 2007; McCracken et al., 2012), residual variations remain across the field. These are shown in Figure 5.5, which displays the distribution of the stellar FWHM and the median FWHM for two representative bands. While the PSF is relatively homogenous across the field for most of the bands (e.g u-band), there is larger scatter for some bands (*e.g.* $IA464$). the aperture magnitude.

Effect of the seeing on the aperture magnitude – As discussed previously, there is a variation of the PSF within the field which is not taken into account in my homogenization. For this reason, it is important to estimate the magnitude differences arising from this variation. To achieve this, I present here a toy model to estimate the effect of the seeing variation on the aperture magnitude for point-like objects.

I denote $D_{\rm stars}(\theta,\beta,r)$ as the difference of the aperture magnitudes for a PSF represented by a Moffat profile $\mathcal{M}\,[\theta, \beta]$ and with a PSF $\mathcal{M}\,[0.8'', 2.5]$. $D_{\rm stars}$ is a function of θ and β, the two parameters which define the Moffat profile, and r which is the aperture diameter. I present in Figure 5.6 $D_{\rm stars}(\theta,\beta,3'')$ in the two-dimensional (2D) parameter space $[\theta, \beta]$. I overplotted on this 2D distribution the contours which enclose 68% and 95% of the $[\theta,\beta]$ distribution for the two bands u and $IA464$. For the purpose of this figure, each star seeing is individually computed from a fit with a Moffat profile on the PSF-homogenized star profiles (reconstructed from the flux extracted at 14 fixed apertures, logarithmically spaced between 0.25″ and 2.5″). Note that since the Moffat Profile is fitted on individual stars from 14 discrete apertures and not on all of the point sources at the same time, the precision of the fit is limited. However, this immediately provides a qualitative insight as to the bias generated by internal PSF variation when extracting the star photometry within a 3″ aperture. For the worst band $IA464$, this bias is expected to remain below 0.1 mag. I also estimate that the median of the magnitude difference is below 0.05 mag, which is in agreement with Figure 5.4.
I then estimate this bias in the photometry for extended objects. Let's chose two different galaxy luminos-

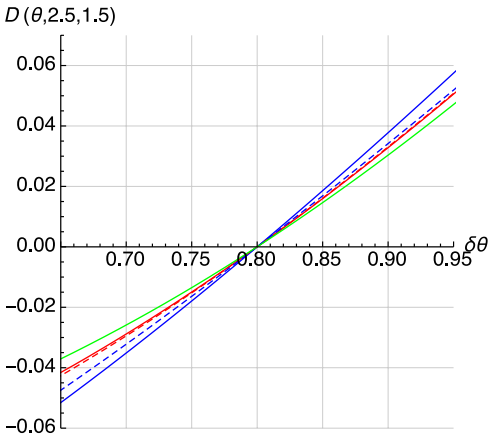

Figure 5.7: Magnitude difference for a point-like object (green), an elliptical galaxy (red), and a spiral galaxy (blue) convolved with a PSF $\mathcal{M}[\theta, 2.5]$ as a function of the seeing θ and for two different effective radii, 0.5 ″ (solid line) and 0.8 ″ (dashed line), when the extraction is performed in 3 ″ diameter apertures.

ity profiles, namely, a de Vaucouleurs profile (1948, 1959) to model a typical elliptical galaxy profile,

$$F_{\text{elliptical}}(R_e, r) \propto \exp\left[-7.67\left(\frac{r}{R_e}\right)^{\frac{1}{4}}\right], \tag{5.5}$$

and an exponential profile to model a spiral galaxy profile,

$$F_{\text{spiral}}(R_e, r) \propto \exp\left[-\frac{r}{R_e}\right]. \tag{5.6}$$

Here, R_e is the effective radius such that half of the total flux is within R_e. I then convolve the luminosity profiles with the Moffat profile, and integrate them in a circular aperture of 3″. For this exercise, I keep β constant and equal to 2.5 and I allow θ to vary. In Figure 5.7, I present the difference $D_{\text{spiral}}(\theta,2.5,1.5'')$ and $D_{\text{elliptical}}(\theta,2.5,1.5'')$ for two effective radii (R_e = 0.5 and 0.8 ″). It can be noted that for FWHM differences below 0.1″ the induced magnitude discrepancies are always lower than 0.05, regardless of the galaxy profile.

Concerning the cosmetic quality of the image, the convolution operation produces several undesirable effects. First, it induces a covariance in the background noise which can lead photometric errors being underestimated. Secondly, since the homogenization process acts both on the FWHM and the profile slope (α and β parameters), the convolution kernel may contains negative components. In some bands it can lead to artefacts (such as rings) around saturated objects. These saturated objects are masked in the final catalogue. I deal with the correlation of the background by multiplying in each band the photometric errors derived from SExtractor by a correction factor. The computation of these factors is reported below.

Setting upper limits – In some bands, a source may be below the measurement threshold while at the same time be detected in the combined $z^{++}YJHK_s$ χ^2 image. In this case, in the measurement band, SExtractor may not report consistent magnitudes or magnitude errors. To be able to identify these sources as beyond the detection limit, I report upper limits[8] instead of the source magnitudes in each band where they are too faint to be detected.
To compute the magnitude limits, I run SExtractor on each individual image using the same detection parameters as used for the detection of the objects (see below and Table 5.3). All of the pixels belonging to objects are flagged. Fluxes are measured from PSF-homogenized images in empty apertures of 2 and 3″, discarding all of the apertures containing an object. The magnitude limit is then computed from the standard deviation of fluxes in each aperture.

[8] equivalently "limiting magnitudes" or "magnitude limits".

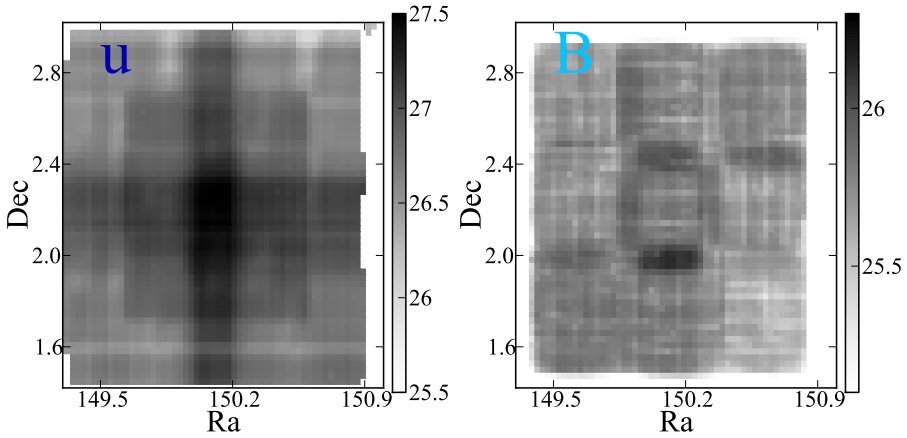

Figure 5.8.: The locally computed magnitude limits, estimated from the variance map as a function of right ascension and declination for u, B and i^+ bands in a $2''$ diameter aperture.

This method is not always appropriate since the values of the upper limits may vary over the field, as shown in Figure 5.8. That is why I used a local estimate for the upper limits in the six broad bands of optical data (u, B, V, r, i^+, z^{++}). In these bands, upper limits are calculated for each object from the variance map and are defined as being the square root of the variance per pixel integrated over the aperture. The magnitude of the object is set to the 3σ magnitude limit if the flux is below the 3σ flux limit, or if the flux is below the flux error. The averaged values of these upper limits are consistent with the value computed with the first method and are displayed in Table 5.2. The upper limits in these bands are important because young, star-forming objects at high redshift will have apparent magnitudes in the optical bands of the order of the limiting magnitude. The computation of photometric redshifts uses fluxes and so does not use the upper limits which are only applied to the magnitudes, but it may be useful when working with magnitudes to know whether or not the object is within the upper limit.

5.2.3 Photometric measurements

Optical and NIR photometry – Object photometry is carried out using SExtractor in "dual image" mode: object are detected in the χ^2 image, and then fluxes are extracted from 2 and $3''$ diameter apertures on PSF-homogenized images in each band.
The well-known difficulty in source extraction from astronomical images is that objects have ill-defined, potentially overlapping boundaries, making flux measurements challenging. The two main parameters that control extraction are the deblending threshold and the flux threshold. Therefore, a reasonable balance must be found on one hand between deblending too much (splitting objects) and not deblending enough (leading to merging). Simliar problems occur with the choice of the detection threshold: a low detection threshold can create too many spurious objects, and one that is too high may miss objects. This can be mitigated in part by a judicious choice of detection threshold and the minimal number of contiguous pixels which constitute an object. The solution I adopted is to set a low deblending and detection threshold while increasing the number of contiguous pixels to reject false detections. I validated this choice through careful inspection of catalogues superimposed on the detection and measurement images, which is feasible in the case of a single-field survey like COSMOS.
The background is estimated locally within a rectangular annulus (30 pixels thick) around the objects, delimited by their isophotal limits. Additionally, object mask flags indicating bad regions in the optical and NIR bands were included and saturated pixels in the optical bands were flagged by using the appropriate

FLAG_MAPs. My chosen parameters are given in Table 5.3.

In the last step, catalogues from each band are merged together into a single FITS table and galactic extinction values are computed at each object position using the Schlegel et al. (1998) values. These reddening values have to be multiplied by a factor computed for each band, derived from the filter response function and integrated against the galactic extinction curve (Bolzonella et al., 2000; Allen, 1976).

IR photometry extraction – The IR photometry extraction has been performed by Paul Hsieh from the χ^2-image. I summarise below the main lines of this extraction.
The IRAC data have a very extended PSF and exceed the confusion limits in channel 1 and 2. Better deblending techniques are necessary to estimate fluxes in these crowded, low-resolution images. Many methods have been developed to extract accurate fluxes when images are convolved with a large PSF. A technique is to use the object morphological and positional information from a higher quality image of the same field. IRACLEAN (Hsieh et al., 2012) is the method used to derive the SPLASH photometry from the χ^2-image. Unlike other similar methods assuming that the intrinsic morphology of an object is identical in the two wavebands (i.e., there is no morphological k-correction), IRACLEAN is essentially the same as CLEAN deconvolution in radio imaging with nearly no morphological restrictions, except for the locations where CLEAN can opperate. This scheme minimizes the effect of morphological k-correction when a high-resolution image (i.e., the prior) and its low-resolution counterpart are taken in very different wavebands. This is in particular the case for cosmos2015, for the χ^2-image is computed from the combination of 5 passbands: the morphological information of the objects in the detection image is in any case comparable to the IRAC one.
However, there is a limitation for this scheme. If the separation of two objects is less than ~ 1 FWHM, then the flux of the brighter object can be overestimated while the flux of the fainter one can be underestimated, as discussed in Hsieh et al. (2012). To solve this issue, IRACLEAN has been improved by taking the surface brightness information in the prior into account. The new IRACLEAN uses the surface brightness of the prior to weight where CLEAN occurs for each object. The weighting strength is determined by the power of the surface brightness, i.e., (surface brightness of the prior)^n, where n is the weighting parameter. If n is zero, then the surface brightness of the prior is ignored, and so the new IRACLEAN behaves like the original IRACLEAN. When n is greater than zero, the higher the n is, the more heavily weighted the surface brightness is. If the wavebands of the prior and the target images are very different, then n can be set to a lower value, *e.g.*, 0.1 - 0.3. If the wavebands of the prior and the target images are very similar, then n can be set to a higher value, *e.g.*, 0.3 - 0.5. In general, $n = 0.3$ is sufficiently good for most cases.
To accelerate the process, both the UltraVISTA $z^{++}YJHK_s$ chi-square image and the SPLASH images are broken up into the 144 tiles that are used for the COSMOS Subaru/ACS data, making parallel processing easier. The tiles overlap by $14.4''$ around the edges to avoid flux underestimation for those objects close to the edges of the tiles. The SPLASH PSFs in each tile are generated using point sources in that tile. The aperture size used to measure the flux ratios between sources and PSFs for the CLEAN procedure is $1.8'' \times 1.8''$, and the weighting parameter n is 0.3. After the CLEAN procedure is completed, a residual map is generated which is used toe stimate the flux errors. The flux error of each object is estimated based on the fluctuations in the local area around that object in the residual map. The IRACLEAN procedure is described fully in Hsieh et al. (2012).

NUV extraction – As in Ilbert et al. (2009), GALEX photometry (Zamojski et al., 2007) for each object was derived by cross-matching my catalogue with the publicly available photometric i^+-selected catalogue described in Capak et al. (2007)[9]. GALEX fluxes were measured using a PSF fitting method with the

[9]The version of the catalogue used is available at http://irsa.ipac.caltech.edu/data/COSMOS/tables/photometry/. This catalogue supersedes that of Capak et al. (2007) with improved source detection and photometry.

Table 5.3.: SExtractor parameters used for dual-mode χ^2 detection and photometry.

Name	**Value**
ANALYS_THRESH	1.5
FILTER_NAME	gauss_4.0_7×7.conv
CATALOG_TYPE	FITS_1.0
DETECT_TYPE	CCD
THRESH_TYPE	ABSOLUTE
DETECT_MINAREA	10
DETECT_MAXAREA	100 000
DETECT_THRESH	1.51
FILTER	Y
DEBLEND_NTHRESH	32
DEBLEND_MINCONT	0.00001
CLEAN	Y
CLEAN_PARAM	1.0
MASK_TYPE	CORRECT
PHOT_APERTURES	13.33, 20.00, 47.33
PHOT_AUTOPARAMS	2.5, 3.5
PHOT_FLUXFRAC	0.2, 0.5, 0.8
PHOT_AUTOAPERS	13.3, 13.3
SATUR_LEVEL	30000.
MAG_ZEROPOINT	depends on the band
GAIN	depends on the band
PIXEL_SCALE	4.16666×10^{-5}
BACK_SIZE	128
BACK_FILTERSIZE	3
BACKPHOTO_TYPE	LOCAL
BACKPHOTO_THICK	30
WEIGHT_GAIN	N
RESCALE_WEIGHTS	N
WEIGHT_TYPE	depends on the band
GAIN_KEY	DUMMY

u^*-band image used as a prior.

Rescaling photometric errors – Precise photometric error measurements are essential for accurate photometric redshifts. For each Subaru band, I used effective gain values (Capak et al., 2007) for the non-convolved data[10] to compute the magnitude errors. This is particularly important for the Subaru bands because of the long exposure times used for each individual exposure. The magnitude errors are thus given by (Bertin, 2010):

$$\Delta m = 1.0857 \frac{\sqrt{A\sigma^2 + F/g}}{F}, \tag{5.7}$$

where A is the area, over which the flux F is summed, σ the standard deviation of noise given from the RMS_MAPS, and g the gain.
However, because SExtractor errors are underestimated in data with correlated noise, they have to be rescaled. I multiplied the magnitude and flux errors with a correction factor computed for each band from empty apertures: the segmentation map[11] provided me a way to discard apertures containing an object. Following Bielby et al. (2012), this factor is computed in each band for the 2 and 3″ apertures and taken as the ratio between the standard deviation of the flux extracted in empty apertures on the field and the median of the SExtractor errors. For UltraVISTA, I computed separate values for the Ultra-deep ($\mathcal{A}^{\rm UD}$) and deep ($\mathcal{A}^{\rm Deep}$) regions. The corrections are given in Table 5.4.

5.2.4 Photometric redshifts and physical parameters computation

The photometric redshifts and physical parameters have been computed with the help of O. Ilbert and I. Davidzon (LAM) with the SED-fitting code LePhare described in Chapter 3. I report below the main aspects of the computation in the configuration specific to cosmos2015.

Catalogue preparation – I used fluxes rather than magnitudes for my photometric measurements to deal robustly with faint or non-detected objects. Faint objects may have physically meaningful flux measurements, whereas their magnitudes and magnitude errors may be undetermined (for example, if the flux is negative, the magnitude will be set to an extreme value by SExtractor). Consequently, when using magnitudes, I must set an upper limit: for flux measurements with correct flux errors, this is no longer necessary. There is no loss of information when using flux measurements. This leads to a better determination of the photometric redshifts and a lower number of catastrophic failures at $z > 2$.

Photometric redshifts are computed using 3″ aperture fluxes. The fixed-aperture magnitude estimate is expected to be less noisy for faint sources than the pseudo-total Kron (Kron, 1980) magnitude mag_auto. This is because mag_auto's variable aperture is derived from the detection image, which means that fainter objects can potentially have noisier colours (Hildebrandt et al., 2012; Moutard et al., 2016). This magnitude measurement is also susceptible to blended sources. We found that the 3″ aperture photometry gives slightly better photometric redshifts than the 2″ aperture at low redshift (below $z \leq 1$) and we adopted this aperture over the entire redshift range of the survey. We suspected that the photometric redshift precision is lower in the 2″ apertures due to small-scale residual astrometric errors. This is being investigated for the upcoming DR3 UltraVISTA release.

Photometric redshift computations use colours, and, consequently, should not be sensitive to a systematic magnitude calibration offset. However, in contrast to optical and NIR data, GALEX and IRAC data

[10] The gain values, which are computed from the whole images, are not supposed to vary through convolution.
[11] The segmentation map is computed after the detection of the objects in the χ^2-image. On this map, pixels unlikely to be sky are tagged.

Table 5.4.: Photometric corrections, including multiplicative error factors for SExtractor, systematic offsets (s_f) derived from the spectroscopic sample and factors F for the foreground extinction (Allen, 1976). s_f values have to be subtracted to the apparent magnitudes.

	band	**error fact. (2″)**	**error fact. (3″)**	s_f	F
HSC	Y	2.2	2.7	-0.014	1.298
	Y^{UD}	3.2	3.7	0.001	1.211
	Y^{Deep}	2.8	3.2	0.001	1.211
UVista	J^{UD}	3.0	3.3	0.017	0.871
UVista	J^{Deep}	2.6	2.9	0.017	0.871
	H^{UD}	2.9	3.1	0.055	0.563
	H^{Deep}	2.4	2.9	0.055	0.563
	$K_{\mathrm{s}}^{\mathrm{UD}}$	2.7	3.1	-0.001	0.364
	$K_{\mathrm{s}}^{\mathrm{Deep}}$	2.3	2.6	-0.001	0.364
WIRCam	Ks	2.1	3.4	0.068	0.364
	H	2.1	3.2	-0.031	0.563
CFHT	u	2.3	3.3	0.010	4.660
	B	1.6	1.8	0.146	4.020
	V	1.7	1.9	-0.117	3.117
	r	1.4	1.7	-0.012	2.660
	i^{+}	1.3	1.7	0.020	1.991
	z^{++}	2.0	2.9	-0.084	1.461
	$IA427$	1.7	2.5	0.050	4.260
	$IA464$	1.7	2.4	-0.014	3.843
	$IA484$	1.7	2.5	-0.002	3.621
	$IA505$	1.6	2.3	-0.013	3.425
	$IA527$	1.5	2.2	0.025	3.264
SUBARU	$IA574$	1.9	2.8	0.065	2.937
	$IA624$	1.4	2.0	-0.010	2.694
	$IA679$	2.0	2.8	-0.194	2.430
	$IA709$	1.7	2.4	0.017	2.289
	$IA738$	1.5	2.1	0.020	2.150
	$IA767$	1.8	2.6	0.024	1.996
	$IA827$	2.2	3.1	-0.005	1.747
	$NB711$	1.2	1.8	0.040	2.268
	$NB816$	2.5	3.5	-0.035	1.787
	ch1	-	-	-0.025	0.162
IRAC	ch2	-	-	-0.005	0.111
	ch3	-	-	-0.061	0.075
	ch4	-	-	-0.025	0.045
GALEX	NUV	-	-	0.128	8.621

provide total magnitudes or fluxes, which requires an estimate of the total flux from the corrected 3″ aperture fluxes to be consistent over the full wavelength range. This is also needed to derive stellar masses. For each object, I computed a single offset o (the same for all the bands) which allows for the conversion from aperture to total magnitude. The offset is computed following Moutard et al. (2016):

$$o = \frac{1}{\sum_{\text{filters i}} \mathrm{w}_i} \times \sum_{\text{filters i}} (\mathrm{MAG_{AUTO}} - \mathrm{MAG_{APER}})_i \times \mathrm{w}_i \,, \tag{5.8}$$

where we have:

$$\mathrm{w}_i = \frac{1}{(\sigma^2_{\mathrm{AUTO}} + \sigma^2_{\mathrm{APER}})_i} \,. \tag{5.9}$$

This leads to the assumption that the PSF profile is the same in all of the bands. As it is averaged over all of the broad bands, i.e, u, B, V, r, i^+, z^{++}, Y, J, H, and K_s, this offset is more robust than the one which would have been computed by band. These offsets are given in the final catalogue (for which the description of the columns is given in Appendix).

Photometric redshift computation – LePhare (Arnouts et al., 2002; Ilbert et al., 2006, and see Section 3.3) is used with the same method as in Ilbert et al. (2013). For the photometric redshift computation, we used a set of 31 templates including spiral and elliptical galaxies from Polletta et al. (2007) and a set of 12 templates of young blue star-forming galaxies using Bruzual and Charlot (2003) models. We consider the extinction laws of Calzetti et al. (2000), Prevot et al. (1984) and a modified version of the Calzetti laws including a "bump" at 2175Å (Fitzpatrick and Massa, 1986). Using a spectroscopic sample of quiescent galaxies, Onodera et al. (2012) showed that the estimate of photometric redshifts for the quiescent galaxies in Ilbert et al. (2009) were underestimated at $1.5 < z < 2$. Following Ilbert et al. (2013), we improved the photometric redshifts for this specific population by adding two new Bruzual and Charlot (2003) templates assuming an exponentially declining SFR with a short timescale $\tau = 0.3\,\mathrm{Gyr}$ and extinction-free templates. We take also into account the contribution of the emission lines.

We use the large spectroscopic sample available on COSMOS field to measure in each band the systematic offsets for calibration of the photometric redshifts (see Section 3.3). These offsets are given in Table 5.4.

Absolute magnitudes and stellar masses – Stellar masses using LePhare are derived following exactly the same method as in Ilbert et al. (2015). We derived the galaxy stellar masses using a library of synthetic spectra generated using the Stellar Population Synthesis (SPS) model of Bruzual and Charlot (2003), assuming a Chabrier (2003) Initial Mass Function (IMF). We combined the exponentially declining SFH and delayed SFH ($\tau^{-2} t e^{-t/\tau}$) presented in Section 3.1.2. Two metallicities (solar and half-solar) are considered. Emission lines are also added and two attenuation curves are considered: the starburst curve of Calzetti et al. (2000) and a curve with a slope $\lambda^{0.9}$ (Appendix A of Arnouts et al., 2013). The $E(B-V)$ values are allowed to take values as high as 0.7. Given the uncertainties on the SFR based on template fitting (Ilbert et al., 2015) Ido not include the SFRs estimated from template fitting in my distributed catalogues.

5.3 Catalogue quality assessment

In this Section, I assess the reliability of the catalogue by comparing with previous catalogues when the redshift and magnitude ranges overlap. I show that the agreement with the literature is excellent, and that

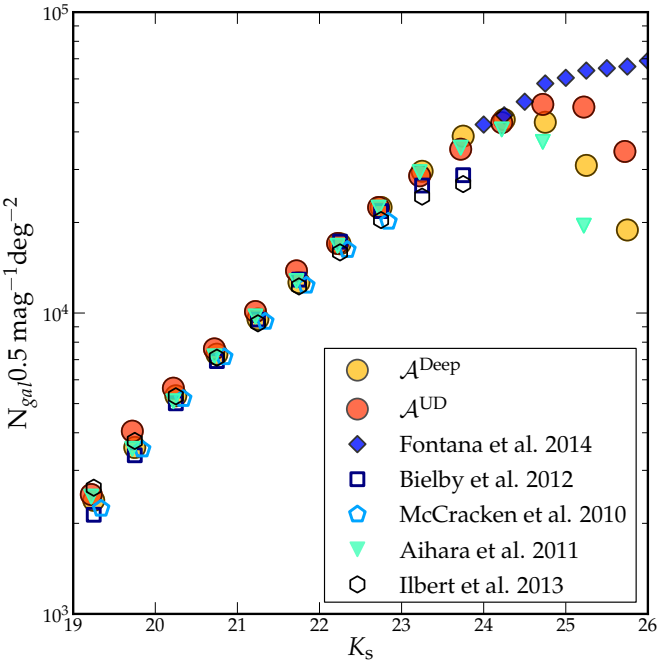

Figure 5.9: K_s-band-selected galaxy number counts of the $YJHK_sz^{++}$-detected galaxies in $\mathcal{A}^{\mathrm{UD}}$ (yellow circles) and $\mathcal{A}^{\mathrm{Deep}}$ (pink circles), compared to a selection of literature measurements. The Ilbert et al. (2013) and McCracken et al. (2010) points show previous measurements in COSMOS.

this COSMOS2015 catalogue allows to probe fainter magnitude and higher redshift ranges with a better accuracy.

5.3.1 Galaxy counts

In Figure 5.9 is plotted the number of galaxies per square degree per magnitude as a function of K_s magnitude for objects in both the $\mathcal{A}^{\mathrm{Deep}}$ and $\mathcal{A}^{\mathrm{UD}}$ regions (details of the stars-galaxy separation can be found in Section 5.4). The corresponding values are presented in Table 5.5.

The counts are in excellent agreement with the literature. We reach more than one magnitude deeper compared to the previous UltraVISTA-DR1 (McCracken et al., 2012). In addition, these counts are in good agreement with the much deeper Hawk-I survey (Fontana et al., 2014) up to at least $K_s \sim 24.5$.

At the 3σ limit in K_s, almost twice as many objects per square degree are detected in $\mathcal{A}^{\mathrm{UD}}$ than in $\mathcal{A}^{\mathrm{Deep}}$. Furthermore, my catalogue contains $\sim 1.5\times10^5$ objects with $K_s < 24.7$ in the $\mathcal{A}^{\mathrm{UD}}$ compared to $\sim 0.8\times10^5$ found with UltraVista-DR1 (McCracken et al., 2012) in the same region at the detection limit in K_s. In $\mathcal{A}^{\mathrm{Deep}}$, the difference is less significant since the depths are comparable, with $\sim 0.9\times10^5$ objects compared to $\sim 0.7\times10^5$ found in UltraVista-DR1.

5.3.2 Photometric redshift distribution

The photometric redshift distribution for the i^+- and K_s-selected samples is given in Figure 5.10. Magnitudes are measured in corrected 3" aperture magnitudes (equivalent to total magnitudes) with the derived systematic offset applied. Several interesting trends are apparent. In general, the median redshift of the K_s-sample is higher than the i^+-selected samples. Also, the fraction of sources at higher redshifts is greater for the NIR-selected samples. These effects are largely due to the well-known positive evolutionary corrections and k-corrections for NIR-selected samples. Optically selected samples at higher redshifts move progressively to shorter rest-frame UV wavelengths, which are strongly attenuated by dust and the inter-

Table 5.5.: K_s-band-selected galaxy number counts. Logarithmic galaxy number counts are in units of galaxies per 0.5 magnitude per square degree.

Magnitude Bin	**Deep Regions**	**Ultra-Deep Regions**
		regions
[16.0 – 16.5]	2.24	2.19
[16.5 – 17.0]	2.28	2.33
[17.0 – 17.5]	2.58	2.60
[17.5 – 18.0]	2.72	2.80
[18.0 – 18.5]	2.94	3.00
[18.5 – 19.0]	3.19	3.24
[19.0 – 19.5]	3.39	3.42
[19.5 – 20.0]	3.57	3.62
[20.0 – 20.5]	3.73	3.76
[20.5 – 21.0]	3.87	3.89
[21.0 – 21.5]	3.99	4.01
[21.5 – 22.0]	4.11	4.14
[22.0 – 22.5]	4.23	4.23
[22.5 – 23.0]	4.35	4.36
[23.0 – 23.5]	4.48	4.46
[23.5 – 24.0]	4.59	4.55
[24.0 – 24.5]	4.65	4.64
[24.5 – 25.0]	4.63	4.69
[25.0 – 25.5]	4.49	4.68
[25.5 – 26.0]	4.27	4.54

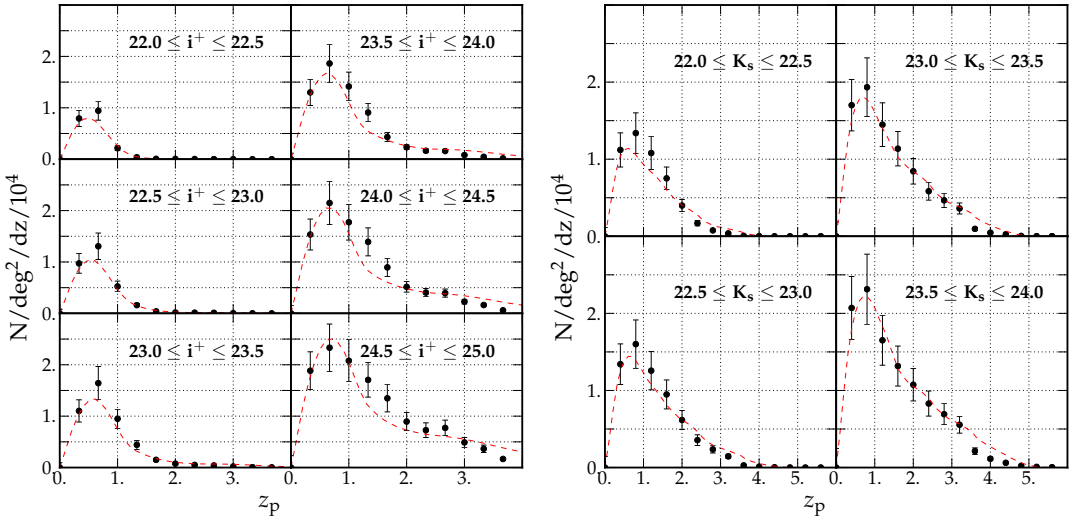

Figure 5.10.: Photometric redshift distributions for i^+- (*left*) and K_s(*right*)-selected samples for the full sample, compared with a model prediction (red dashed line) from PEGASE.2 (Fioc and Rocca-Volmerange, 1997, 1999). Plotted errorbars are uncertainties estimated from jacknife errors, splitting the field into 25 sub-fields.

galactic medium.

We compared these distributions with a simple three-components galaxy population model generated with the PEGASE.2 code (Fioc and Rocca-Volmerange, 1997, 1999, and see Section 3.1.2). Each population starts forming at $z = 8$ via the infall of pristine gas on a specific timescale and gas is converted into stars at a specific rate. The corresponding star formation histories peak at $z \sim 4$, 2 and 0, and the $z = 0$ predicted optical-NIR colours correspond to those of local Sa (ordinary spiral type), Sbc (barred spiral type), and Sd (spiral type with diffuse arms) galaxies, respectively. The total baryonic mass (gas, stars, and hot halo-gas) of each galaxy is assumed to be constant, and the mass function of each population is tuned so that the sum of the three populations matches simultaneously the local luminosity function in the B band, the deep galaxy counts in the B, V, I, and K_s bands, as well as the cosmic star formation rate density and the stellar mass density observed at $z = 0 - 6$.

The agreement between the data and this simple three-component model is quite good. This success lies in the differential contributions of the three galaxy populations to the counts. Indeed, these modelled counts at $K_s = 24$ are the sums of the almost equal contributions of Sd progenitors at $z \sim 0.7$, of Sb progenitors at $z \sim 1.2$, and Sab progenitors at $z \sim 2$. In contrast, a simpler modelling of the galaxy populations using a single scenario and a unique mass function leads to very good agreement between the integrated counts in i^+ and K_s bands, as well as a good match between the star formation rate densities, but it completely overshoots the mean redshift of the $K_s \sim 24$ or $i^+ \sim 24.5$ sources ($z \sim 2$, whereas the COSMOS data shows it is peaked at $z \leq 1$). Other choices of modelling which have been explored also lead to a higher levels of tension.

5.3.3 Photometric redshift accuracy

Photometric redshift accuracy measured using spectroscopic samples – The COSMOS field is unique in its unparalleled spectroscopic dataset. These spectroscopic samples, derived from many hundreds of hours of telescope time in many different observing programs, are a key ingredient in allowing us to characterise the precision of our photometric redshifts.

i^+	Star-forming galaxies				Quiescent galaxies			
	σ_{15}	η_{15} (%)	σ_{13}	η_{13} (%)	σ_{15}	η_{15} (%)	σ_{13}	η_{13} (%)
[16,21]	0.007	0.5	0.008	0.5	0.005	0.0	0.005	0.0
[21,22]	0.008	0.6	0.008	0.6	0.007	0.3	0.006	0.4
[22,23]	0.01	1.7	0.01	1.9	0.01	0.6	0.011	0.6
[23,24]	0.022	6.7	0.022	7.2	0.027	6.0	0.030	4.4
[24,25]	0.034	10.2	0.037	15.0	0.054	18.9	0.062	16.7
[25,26]	0.057	22.0	0.058	24.2				

Table 5.6: Photometric redshift precision of the catalogue as a function of magnitude and galaxy types for galaxies detected both in cosmos2015 and Ilbert et al. (2013) compared to spectroscopic samples. In almost all cases, photometric redshift precision (σ_{13} and σ_{15}) increases and the number of catastrophic failures (η_{13} and η_{15}) fall compared to Ilbert et al. (2013) *for a selection common to both catalogues.*

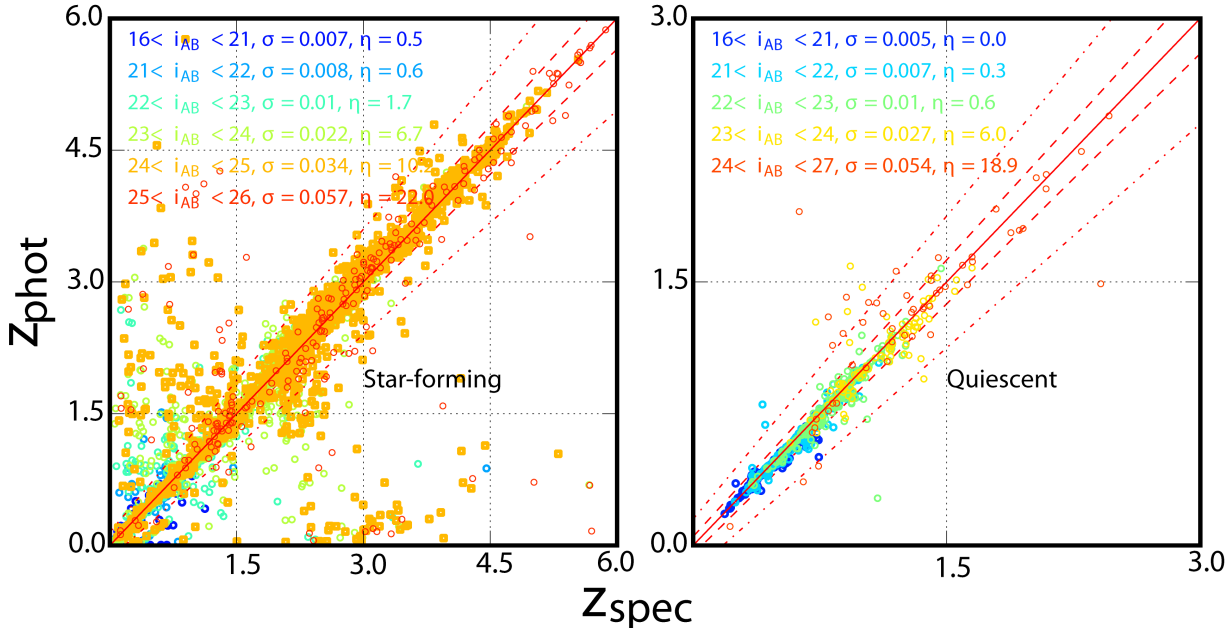

Figure 5.11.: Comparison between photometric and spectroscopic redshifts as a function of $i_{\rm AB}$ magnitudes and types: star-forming galaxies(*left*) and quiescent galaxies (*right*), keeping only non-flagged galaxies. The dashed and dashed-dot lines show $z_{\rm p} = z_{\rm s} \pm 0.05(1+z_{\rm s})$ and $z_{\rm p} = z_{\rm s} \pm 0.15(1+z_{\rm s})$, respectively.

From the COSMOS spectroscopic master catalogue (Salvato M. et al. in preparation.), I retain only the highly reliable 97% confidence-level spectroscopic redshifts (Lilly et al., 2007). I estimated the precision of the photometric redshifts using the normalized median absolute deviation (Hoaglin et al., 1983) defined as $1.48 \times \text{median}\left(|z_{\rm p} - z_{\rm s}|/(1+z_{\rm s})\right)$. This dispersion measurement, denoted by σ, is not affected by the fraction of catastrophic errors (denoted by η), *i.e.* objects with $|z_{\rm p} - z_{\rm s}|/(1+z_{\rm s}) > 0.15$.

The photometric redshift precision of the cosmos2015 catalogue is described in Table 5.6 and Figure 5.11. In Table 5.6, I compare the photometric redshift precision in cosmos2015 with that of the catalogue of Ilbert et al. (2013) by cross-matching the two catalogues and considering the same sources in both cases. Compared to Ilbert et al. (2013), the number of catastrophic failures are reduced and the photometric redshift precision is either increased or is unchanged. It should be recalled, however, that the main gain of cosmos2015 is the considerable increase in catalogue size compared to Ilbert et al. (2013).
The left and right panels of Figure 5.11 show the photometric redshift precision as a function of the i-band magnitude and for star-forming and quiescent galaxies, respectively (classified using NUV$-r/r-J$ diagram, Figure 5.15). Very bright, low-redshift, star-forming galaxies have the most precise photometric redshifts ($\sigma = 0.007$, $\eta = 0.5\%$ for $16 < i^+ < 21$). Moreover, even at $z > 3$, the accuracy is still very good (0.021), with only 13.2% of catastrophic failures.

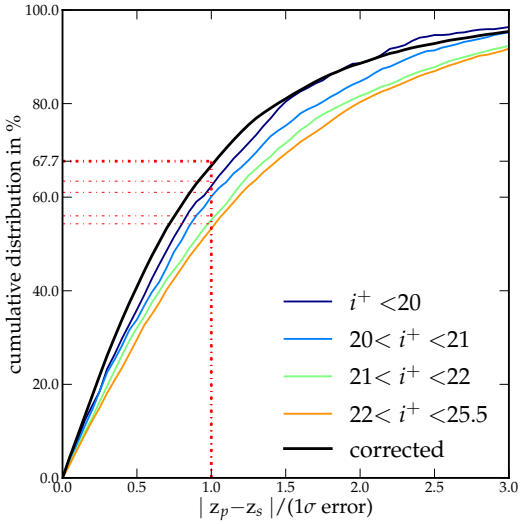

Figure 5.12: *Left:* Cumulative distribution of $|z_{\rm phot} - z_{\rm spec}|/1\sigma$. Of the spectroscopic redshifts, 58% have their photometric redshifts within the 1σ error; this implies that photometric errors are slightly underestimated. This plot is made with the high-confidence spectroscopic redshift catalogue. *Right:* 1σ photometric redshift error as a function of redshift for different magnitude bins on $\mathcal{A}^{\rm Deep}$ (*Bottom*) and on $\mathcal{A}^{\rm UD}$ (*Top*).

Photometric redshift accuracy based on the probability distribution function – I also assess the photometric redshift accuracy using the 1σ uncertainty derived from the photometric redshift probability distribution function (PDFz[12]). The advantage of this method is that one can investigate the photometric redshift accuracy in any redshift-magnitude range. However, it requires an accurate estimate of the PDFz.

In Figure 5.12, I show the cumulative distribution of the ratio $|z_p - z_s|/1\sigma$ (see the definition of the 1σ error in Chapter 3). Assuming that $z_{\rm s}$ is the true redshift, 68% of the times $z_{\rm p}$ should fall within the 1σ error. This comparison shows that the 1σ-uncertainties enclose less than the 68% of the expected value. This is confirmed when I split the spectroscopic sample per magnitude and redshift bin. It appears that our errors on photometric redshifts are underestimated by a factor which depends on the magnitude (see Chapter 3 for a discussion about this common problem). I consequently chose to correct these errors by applying the following magnitude-dependent correction: errors are multiplied by a factor of 1.2 for bright objects ($i^+ < 20$) and by a factor of $(0.1 \times i^+ - 0.8)$ for faint objects ($i^+ > 20$). With this magnitude-dependent correction, there is no consequence on the computation of the physical parameters. However, the PDFz remains systematically too peaky around the median values.

5.3.4 Clustering measurements

I estimated the projected galaxy clustering in my sample by computing the angular two-point auto-correlation function $w(\theta)$. The angular correlation function $w(\theta)$ measures the excess probability of finding two objects separated by an angle θ compared to a random distribution in a series of angular bins. This measurement is an excellent test of the uniformity of my photometric catalogue as w is very sensitive to large-scale photometric systematic errors. Adding cuts in stellar mass and photometric redshift allows for an independent check of our photometric redshift procedures. I computed w using ATHENA[13], which uses the usual Landy and Szalay (1993) estimator:

$$w(\theta) = \frac{1}{RR} \times \left(\frac{N_r(N_r-1)}{N_d(N_d-1)} DD - 2\frac{N_r}{N_d} DR + RR, \right) \tag{5.10}$$

where N_r and N_d are the number of points in the random and the galaxy sample, and RR, RD and DD are the number of pairs in the random catalogue, between the random and galaxy catalogue, and in the

[12] see Chapitre 3
[13] www.cosmostat.org/software/athena/

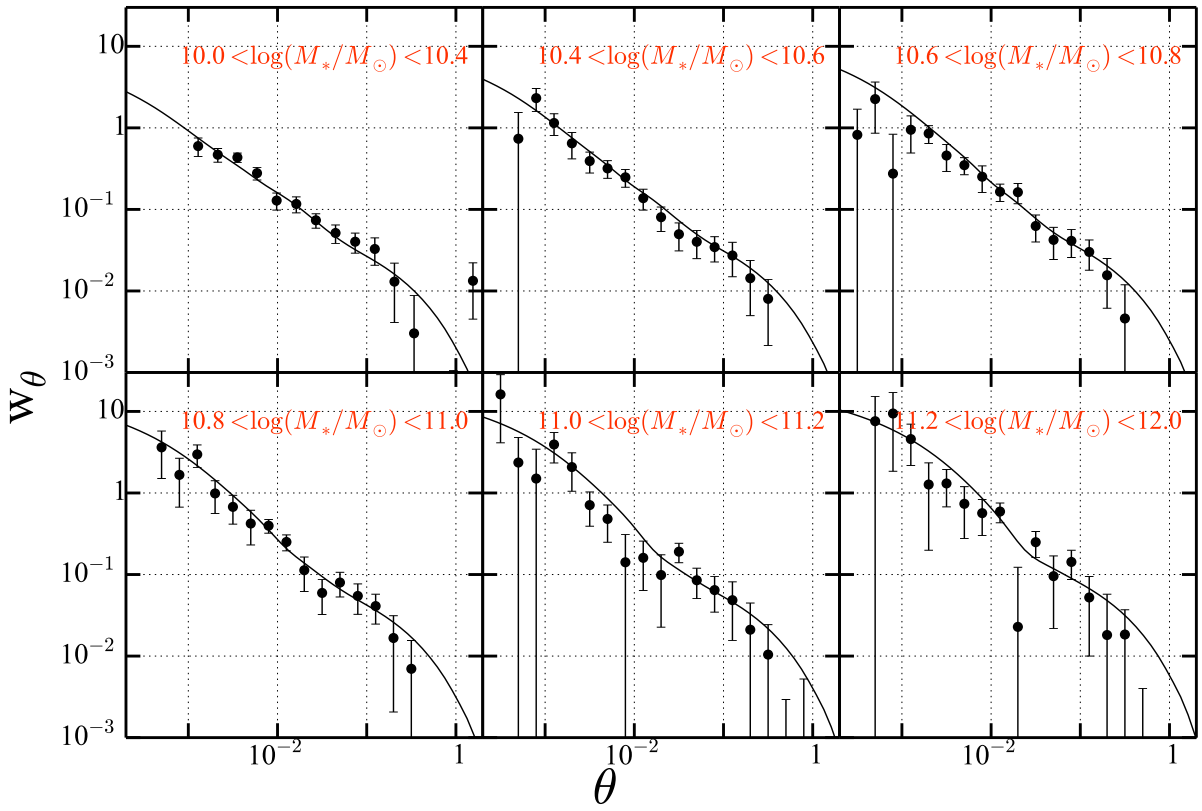

Figure 5.13.: Angular correlation function w as a function of angular scale θ in degrees at $0.5 < z < 1$ for several bins in stellar mass. The solid line shows the analytic predictions of Coupon et al. (2015).

galaxy catalogue. The random catalogue contains 500,000 objects. The measurements are corrected for the "integral constraint" (Groth and Peebles, 1977), a systematic effect arising from using a clustered sample to estimate the mean background density in a finite area.

Figure 5.13 shows w in $0.5 < z < 1$ in six mass bins compared to the best-fitting occupation distribution (HOD) model derived by Coupon et al. (2015) in the MIRACLES/CFHTLS field. My measurements are in excellent agreement with the predictions of Coupon et al.'s best-fitting HOD model, computed from a larger 25 deg^2 field. This suggests that, at these redshift ranges and masses, cosmic variance is not an important issue in the COSMOS field. Only at high stellar masses and small scales there is a systematic offset from the models, which may indicate the limitations of the halo model in this mass regime.

In contrast to this result, some works have noted that there is a clear excess in the number of galaxies in COSMOS compared to other fields (see, *e.g.*, Figure 33 in Molino et al., 2014) due to the presence of large structures at $z \sim 1$ and below, which could influence my correlation function measurements (McCracken et al., 2007). The measurements presented here cover quite a large redshift range and consequently probe a large volume, and are therefore less susceptible to the effects of cosmic variance. In smaller redshift slices and at higher redshifts, the effect of cosmic variance becomes more pronounced, especially when these redshift ranges overlap with several of the large structures known to exist in the COSMOS field, for example, at $1 < z < 1.3$ (see also the discussion in McCracken et al., 2015).

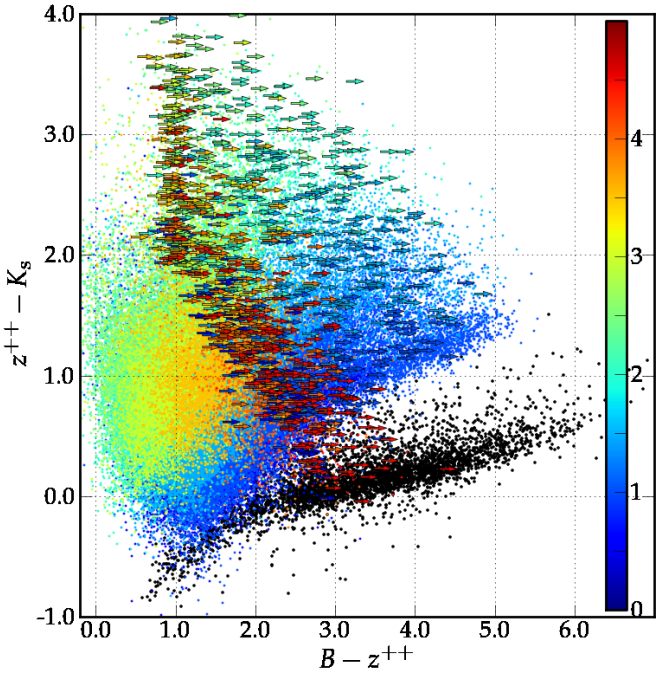

Figure 5.14: Colour-colour diagram $z^{++} - K_s/B - K_s$ for all sources. Sources classified as stars are represented in black. Colours represent redshifts. Right-pointing arrows are the upper limits in the B band.

5.4 Selecting specific populations of objects

In this Section I present how the selection of different object populations can be achieved, based on their absolute magnitudes properties (which allow to distinguish easily between stars and galaxies, or star forming and quiescent galaxies). Furthermore, I detail how the stellar mass limit is estimated.

5.4.1 Star/galaxy classification

LePhare has been used with both galaxy and stellar templates. To determine star/galaxy classification, the best-fitting χ^2 for the galaxy templates $\chi^2_{\rm gal}$ is compared with the one derived for the stellar templates $\chi^2_{\rm stars}$. I flagged as stars all objects for which $\chi^2_{\rm gal} - \chi^2_{\rm stars} > 0$ but only if the object is detected in NIR or IRAC ($m_{3.6\mu m} \leq 25.5$ or $K_s < 24.7$) and is not too far from the BzK stellar sequence ($z^{++} - K_s < (B - z^{++}) * 0.3 - 0.2$).

Figure 5.14 shows a BzK colour-colour diagram (Daddi et al., 2004) for all of the sources including stars and galaxies. The BzK colour-colour diagram is convenient to separate galaxies from stars and to classify galaxies into passive and star-forming ones. Symbols are coloured according to their photometric redshifts. As expected, B-drop-outs occur predominately at $z > 4$, and galaxies with bluer $z^{++} - K$ colour are at lower redshifts. Stars selected using the above classification are shown in black. In the $\mathcal{A}^{\rm UVISTA}$ region, 24,074 objects are classified as stars. A cross-match with the ACS stellar catalogue Leauthaud et al. (2007) shows that 77% of the stars with $i^+ < 24$ from ACS are classified as stars with this method. However, 15% are misclassified as galaxies but are in masked areas. Finally, 0.6% of the extended sources are misclassified as stars.

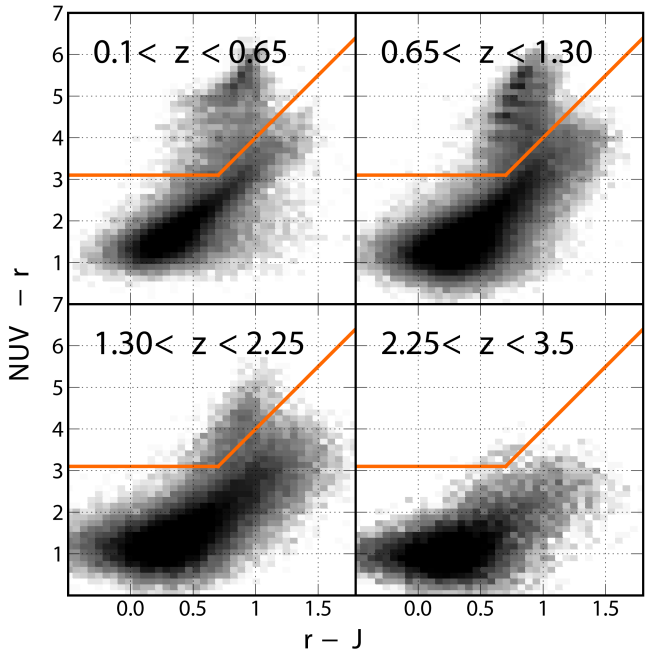

Figure 5.15: NUV$-r/r-J$ galaxy distributions. Quiescent galaxies lie in the top-left corner. The objects fainter than limiting magnitudes are not used.

5.4.2 The NRJ diagram

Quiescent galaxies can be identified using their locations in the colour-colour plane NUV-r/r-J (Williams et al., 2009). Quiescent objects are identified as those with $M_{NUV} - M_r > 3(M_r - M_J) + 1$ and $M_{NUV} - M_r > 3.1$. This technique is described in more detail in Ilbert et al. (2013); in particular, this technique avoids mixing the red dusty galaxies and quiescent galaxies. In my catalogue, galaxies with a flag of 0 are quiescent galaxies and the others are star-forming galaxies. The redshift-dependent evolution of this distribution is presented in Figure 5.15. The rapid build-up of quiescent galaxies at low redshift inside the box is evident, as is the relative decrease in bright, star-forming galaxies outside the box.

5.4.3 Stellar mass limits

I empirically estimate the stellar mass completeness (Pozzetti et al., 2010; Ilbert et al., 2013; Moustakas et al., 2013; Davidzon et al., 2013). I first determine the magnitude limit $K_{\mathrm{s\,lim}}$. For each galaxy, I then determine the mass it would need to have to be observed, at that redshift, at the magnitude limit (given in Table 5.2):

$$\log M_{\mathrm{lim}} = \log M - 0.4(K_{\mathrm{s\,lim}} - K_{\mathrm{s}})\,. \tag{5.11}$$

Next, in each redshift bin, I estimate the stellar mass completeness M_{lim} within which 90% of the galaxies lie. I estimate independently the mass limits on $\mathcal{A}^{\mathrm{Deep}}$ and $\mathcal{A}^{\mathrm{UD}}$. I compute these mass limits using the 3σ limiting magnitude, which is 24.0 for $\mathcal{A}^{\mathrm{Deep}}$ and 24.7 for $\mathcal{A}^{\mathrm{UD}}$. These mass limits are given in Table 5.7 and are shown in Figure 5.16. In $\mathcal{A}^{\mathrm{UD}}$, the mass limits reach a factor of two lower in mass compared Ilbert et al. (2013).

As expected, the mass limit is lower in $\mathcal{A}^{\mathrm{UD}}$ compared to $\mathcal{A}^{\mathrm{Deep}}$, because $\mathcal{A}^{\mathrm{UD}}$ reaches 0.7 magnitudes fainter in the K_{s} band. This estimate is robust to $z \sim 4$ because the observed K_{s} magnitude correlates well with stellar mass in this redshift range. However, these estimates should be treated cautiously at $z > 4$. Above this redshift, the observed K_{s} band lies below the Balmer break and the K_{s} flux does not

Table 5.7.: Mass limits of COSMOS2015 for the full, quiescent and the star-forming samples in the $\mathcal{A}^{\rm Deep}$ and $\mathcal{A}^{\rm UD}$ regions. $N_{\rm gal}^{\rm full}$, $N_{\rm gal}^{\rm quies}$, $N_{\rm gal}^{\rm SF}$ are the percentage of galaxies in each redshift bins for the full, quiescent, and star forming populations. $M_{\rm lim}^{\rm full}$, $M_{\rm lim}^{\rm quies}$, $M_{\rm lim}^{\rm SF}$ are the logarithm of the limiting mass in units of solar masses.

	$\mathcal{A}^{\rm Deep}$						$\mathcal{A}^{\rm UD}$					
bin	$N_{\rm gal}^{\rm full}$	$M_{\rm lim}^{\rm full}$	$N_{\rm gal}^{\rm quies}$	$M_{\rm lim}^{\rm quies}$	$N_{\rm gal}^{\rm SF}$	$M_{\rm lim}^{\rm SF}$	$N_{\rm gal}^{\rm full}$	$M_{\rm lim}^{\rm full}$	$N_{\rm gal}^{\rm quies}$	$M_{\rm lim}^{\rm quies}$	$N_{\rm gal}^{\rm SF}$	$M_{\rm lim}^{\rm SF}$
$0.00< z <0.35$	8.6	8.1	11.3	8.4	8.3	8.1	9.0	7.9	13.8	8.1	8.7	7.8
$0.35< z <0.65$	14.1	8.7	18.4	9.0	13.7	8.6	13.5	8.4	19.2	8.7	13.0	8.4
$0.65< z <0.95$	17.4	9.1	27.4	9.4	16.7	9.0	17.5	8.9	27.4	9.1	16.7	8.7
$0.95< z <1.30$	16.4	9.3	20.5	9.6	16.1	9.2	16.3	9.1	18.9	9.3	16.1	9.0
$1.30< z <1.75$	14.2	9.7	12.5	9.9	14.4	9.6	14.9	9.4	11.8	9.6	15.2	9.3
$1.75< z <2.25$	12.0	9.9	4.9	10.1	12.5	9.8	11.0	9.6	4.0	9.8	11.5	9.6
$2.25< z <2.75$	6.8	10.0	2.4	10.3	7.1	10.0	6.5	9.8	2.2	10.0	6.8	9.8
$2.75< z <3.50$	6.4	10.1	1.4	10.4	6.8	10.1	7.1	9.9	1.5	10.2	7.5	9.9
$3.50< z <4.00$	1.9	10.1	0.5	10.5	2.0	10.5	2.0	10.0	0.4	10.3	2.1	10.0
$4.00< z <4.80$	1.4	10.1	-	-	1.5	10.8	1.5	10.2	-	-	1.6	10.1

correspond precisely to the stellar mass. It is then better traced by mid-IR bands. We will estimate the mass limit at high redshift for an IRAC-selected sample in a future work (Davidzon et al. in preparation).

5.5 Summary

I have now an homogenous galaxy sample from $z = 0$ to $z = 6$, on a relatively large field. Galaxy properties are accurately derived, and I will rely on this catalogue to study the stellar mass growth and the connection of galaxy properties with their large-scale environment in the following (see Chapters 4 and 8).
COSMOS2015 contains more than half a million secure objects over two square degrees. Including new $YJHK_{\rm s}$ images from the UltraVISTA-DR2 survey, Y-band images from Hyper Suprime-Cam and IR data from the SPLASH *Spitzer* legacy program, this NIR-selected catalogue is highly optimized for the study of galaxy evolution and environment in the early Universe.
Detailed comparisons of the colour distributions, number counts and clustering show good agreement with the literature in the mass ranges where these previous studies overlap with ours. In particular, my mass-selected clustering measurements at $0.5 < z < 1$ are in excellent agreement with Coupon et al.'s halo model calibrated using 25 deg^2 of the CFHTLS. Finally, the main improvements of the catalogue compared with previous versions are given below:

- A great number of sources thanks to the combination of deeper data (UltraVISTA-DR2) and an improved extraction image. This image now contains the bluer z^{++} band in addition to the redder NIR bands. There are now $\sim 6 \times 10^5$ objects in the 1.5 deg^2 UltraVISTA-DR2 area and $\sim 1.5 \times 10^5$ in the "ultra-deep stripes" sub-region at the limiting magnitude in $K_{\rm s}$. This represents more than twice as many objects per square degree compared to Ilbert et al. (2013).

- More precise photometric redshifts. Based on comparisons with the unique spectroscopic redshift sample in the COSMOS field, I measure $\sigma_{\Delta z/(1+z_s)} = 0.021$ for $3 < z < 6$ with 13.2% of outliers. At lower redshifts, the precision is better than 0.01, only a few percent of catastrophic failures. The precision at low redshifts is consistent with Ilbert et al. (2013), while it improves significantly at high redshift.

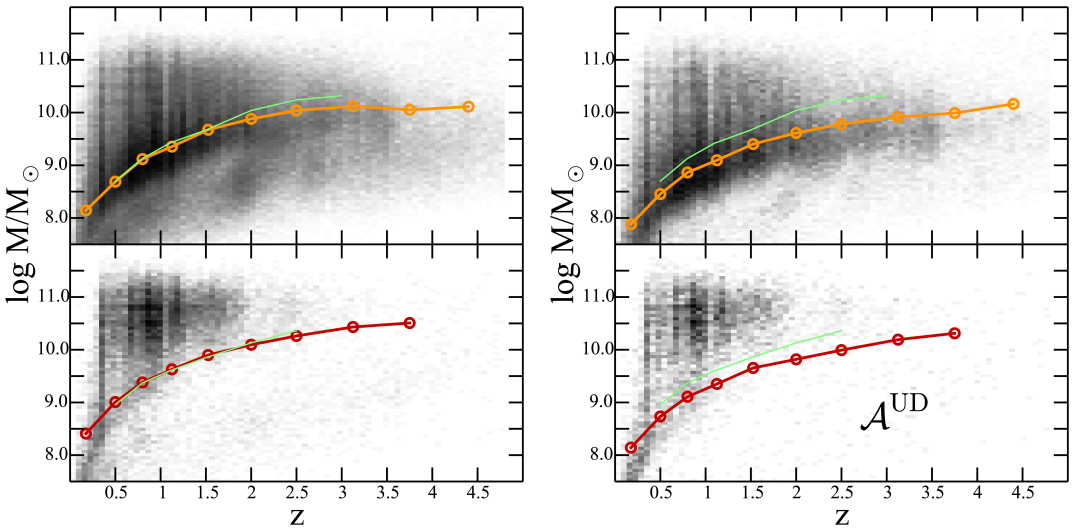

Figure 5.16.: Stellar mass-redshift histogram (the grayscale corresponds to the the number of galaxies in each cell) on $\mathcal{A}^{\rm Deep}$ (*left*) and $\mathcal{A}^{\rm UD}$ (*right*) regions for the full catalogue (top) and for the quiescent sample (bottom). Orange lines shows the mass limit for the full catalogue and red lines for the quiescent ones. These values are reported in Table 5.7. The solid green line is the mass limit in COSMOS as given in Ilbert et al. (2013).

- The characteristic mass limits are much lower. The deepest regions reach a completeness limit of $10^{10}\mathrm{M}_\odot$ to $z = 4$, which is more than 0.3 dex better compared to Ilbert et al. (2013) for the full sample.

Part III.

Results

Chapter 6

Galaxy mass growth from $z \sim 0$ to $z \sim 6$: comparison with HORIZON-AGN

Galaxy properties are shaped by a complex interplay between internal and externally-induced non-linear processes, essentially driven by gravity and baryonic physics over a vast range of scales (described in more details in Chapter 2). On the one hand, the mode of gas inflow ("cold" or "hot"), the strength of gas outflows and the transformation of gas into stars within the galaxies are expected to be mostly controlled and determined by the build-up of the galaxy stellar mass. On the other hand the gas reservoir is primarily that of the intergalactic medium (IGM), which closely follows the cosmic filamentary structure of the dark matter on large scales. Gas inflows onto galaxies are mainly driven by the cosmic filaments, implying a possible dependency of galaxy properties on this environment. This Chapter presents the study of stellar mass growth and galaxy properties through one-point statistics and the comparison with the HORIZON-AGN simulation. Hence it focuses essentially on the impact of galaxy mass and mass-related processes in shaping galaxy properties. The impact of the environment on galaxy evolution will be the focus of the next two Chapters using higher order statistics.

While the basic mass build-up scenario is qualitatively well established, its exact quantitative interpretation is still under intense investigation. We know that gas is accreted cold onto galaxies in the dark matter potential well of low-mass haloes (e.g White and Rees, 1978; Kereš et al., 2005b). In low-mass galaxies, supernovae can eject the dense star-forming gas of galaxies, and regulate star formation (Dekel and Silk, 1986). However, feedback from supernovae is less efficient in the deeper gravitational well of massive galaxies. In this regime, AGN offer a plausible source of regulation. Both theoretical and observational work indicates that AGN are likely to play an important role in shaping galaxy evolution (*e.g.* Silk and Rees, 1998; Bower et al., 2006; Kimm et al., 2012; Dubois et al., 2012a).

Today, both semi-analytical models and hydrodynamical simulations can reproduce the bulk of galaxy properties over a large redshift range. Hydrodynamical simulations, while computationally more demanding, model *ab initio* the co-evolution of dark matter and the non-linear baryonic physics with less simplifying assumptions than the semi-analytical models (see Chapters 2 and 4). In this sense, they offer a detailed survey-scale prediction that can be compared to contemporary datasets across a large fraction of cosmic time. In conjunction with current and forthcoming observational data, these simulations will play a central role in disentangling the key processes that drive galaxy evolution, particularly in the (still poorly understood) high-redshift Universe.

As our description of the Universe is necessarily statistical, the comparison between observations and simulations should primarily be statistical and the agreement should be assessed over wide area and redshift range. This first step is crucial to make the best use of simulations as a mean to identify the key processes

that shape galaxies and build a robust scenario for their formation and evolution.

Such comparisons have already been carried by others (*e.g.* Somerville and Davé, 2015, for a review) and seem already very encouraging. They show that galaxies build-up their mass gradually with time, and that mass is the quantity which primarily shapes observables. However, the agreement between observations and models remains controversial in certain ranges of redshift and mass. As an example, consider the comparison of the galaxy mass function. At low redshift, the overproduction of galaxies in simulations at both the high-mass end and the low-mass end is a long-standing issue. Nowadays, it is quite clear that implementing AGN feedback in simulations is important to model the steep decline in the star formation rate, the shift of the bulk of star formation from massive to low-mass galaxies and consequently to limit the number of massive galaxies, which solves at least partially the high-mass end issue. The overestimation of the low-mass end (*e.g.* Bielby et al., 2012; Weinmann et al., 2012) is more problematic. It is probably partly due to the calibration of stellar feedback (*e.g.* Henriques et al., 2013), but seems to be a persistent problem in state-of-the art simulations (Torrey et al., 2014, and see the work presented below). At high redshift, the problem is different. The measured number of massive galaxies is higher in observations with respect to both hydrodynamical simulations and semi-analytical models (Caputi et al., 2015; Steinhardt et al., 2016). It is still unclear if this discrepancy is due to uncertainties in observations, computational difficulties, or missing piece in the puzzle of our galaxy evolution model.

The comparison between observations and simulations should be made over the largest redshift range possible. Indeed, significantly different initial assumptions (for instance feedback strength) can yield similar results for some fraction of the time (*e.g.* $z \sim 0$) or some scales (*e.g.* large or intermediate scales). Hence in order to assess the reliability of our galaxy evolution model, the consistency with observations should be probed over the full redshift range concerned by galaxy formation and evolution. This is in particular true for simulations which have been calibrated on the local Universe. Finally, attention should be paid to ensuring that simulations reproduce both the aggregate stellar mass growth with cosmic time, but also the observed *diversity* of galaxy populations.
Yet, these discrepancies between observations and simulations may in fact not solely be due to failures in our galaxy evolution model. First of all, it is important to note that these statistical investigations should be carried out over a sufficiently large volume. "Cosmic variance" effects, due to the large-scale density variations, can indeed significantly affect the normalisation and the shape of statistical quantities such as counts, luminosity and mass functions (Somerville et al., 2004; Trenti and Stiavelli, 2008), in particular for low-mass objects (Moster et al., 2011) if the field is too small.
Furthermore, to ensure a fair and consistent comparison between observations and simulations, predicted quantities should be recast as observables. In Chapter 3, I explained that essentially observations and simulations do not trace the same quantities (light or mass). Assumptions in the models (as detailed in Chapter 3) to derive light from mass, or to extract mass from light are the cause for inherent *uncertainties* but also *biases* in the computed properties (*e.g.* the mass), which will inevitably impact the statistical quantities (*e.g.* the mass function). A possible way to ensure that we are comparing like with like could be to generate from the simulations images of galaxies with the same observational biases as real data, and to analyse these images in the same way as observations.

In this Chapter, I present the comparison of statistical quantities from observations and from the Horizon-AGN lightcone (Sections 6.1 and 6.2). Our best knowledge of the physical processes relevant for galaxy formation has been implemented in the Horizon-AGN simulation. Nevertheless, because of the finite resolution of the simulation, some processes are implemented at the subgrid scale, requiring fine-tuning of certain parameters. This work is described in full details in Kaviraj et al. (2016). The objectives of this work are two-fold: first, the comparison allows us to assess which processes truly regulate galaxy evolution when the match is good; conversely, disagreements between model predictions and observations suggest possible avenues of improvement for the model. Taking these considerations into account, I pay attention

to making the comparison at all redshifts consistently while at the same time robustly deriving observable quantities from the simulation. I address in particular the question of bias and uncertainty estimation. While deriving galaxy properties from the simulation with the same tools as observers, I aim at tracking specific biases which are generally not taken into account. Finally, my ultimate aim is to make comparisons with COSMOS2015, which provides a unique galaxy sample to make a consistent comparison over the redshift range $0 < z < 6$ (around 95% of the lifetime of the Universe). This will be described in a future work (Laigle, Davidzon et al. 2016 in preparation), which will include also an end-to-end analysis of the observational biases, partly presented in Section 6.3.

6.1 Galaxy stellar mass growth

In this Section I show that the aggregate stellar mass growth is generally well reproduced in the Horizon-AGN simulation, by relying on one-point statistics and comparing the cosmic star formation history, the mass and luminosity function and the so-called "main sequence". The following Section will investigate the problem in more details by comparing the division of galaxies into star-forming and quiescent populations.

6.1.1 Cosmic star formation history

The cosmic star formation history is one of the most fundamental observables of galaxy evolution. It represents the history of star formation as a whole, and does not distinguish the complex evolution of galaxy subpopulations (Madau and Dickinson, 2014). As such, it is the first diagnostic one can use to assess the reliability of the simulation. That is why I started by comparing the cosmic star formation history predicted by Horizon-AGN to a set of observational data (Hopkins and Beacom, 2006). The quantity is straightforward to compute in Horizon-AGN as the simulation tracks directly the creation of star particles. Observationally, the star formation rate can be derived from x-ray, UV, IR, radio and submillimeter emission as well as emission lines (Madau and Dickinson, 2014). The star formation history uncertainties are then essentially those connected to the measurements of the star formation rate from these passbands. Hopkins and Beacom (2006) showed that the assumed IMF plays a key role in the normalisation. Madau and Dickinson (2014) showed that under the assumption of a universal IMF, there is a good agreement between observations. The result of the comparison is displayed on Figure 6.1. Horizon-AGN is in good agreement with the observations within the errorbars, following both the shape and normalization.
However, this quantity is too simplistic to reveal the nature of internal galaxy physics. For this, we need to study in more detail galaxy populations, through mass and luminosity functions.

6.1.2 Aggregate mass growth of galaxies and the role of AGN

Stellar mass is expected to play a major role in shaping galaxy properties, and as such stellar mass function is a key tracer of galaxy evolution. A robust derivation of the high redshift, but also low-mass end of the stellar mass function are of prime importance to constrain the stellar feedback.

Galaxy luminosity and stellar mass functions are simply the number density of galaxies per bin of mass or luminosity. A Schechter function (Schechter, 1976) is the simplest way to fit them:

$$\frac{\mathrm{d}n}{\mathrm{d}L} = \phi * e^{-L/L_*} \left(\frac{L}{L_*}\right)^{\alpha} \tag{6.1}$$

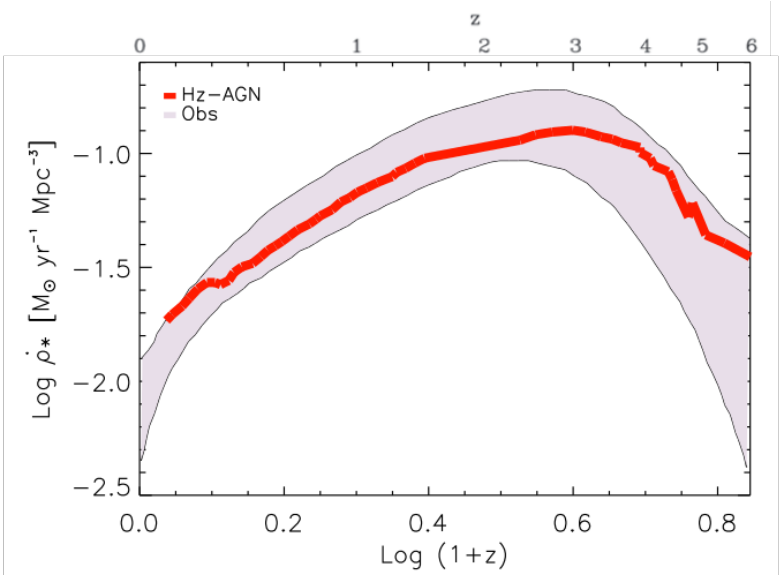

Figure 6.1: Comparison of the predicted cosmic star formation history in Horizon-AGN to observational data (Hopkins and Beacom, 2006) in the redshift range $0 < z < 6$. The red curve shows predictions from Horizon-AGN while the grey shaded region indicates the parameter space covered by the observational data.

However a double-Schechter function is usually preferred to correctly reproduce the faint end:

$$\frac{\mathrm{d}n}{\mathrm{d}L} = e^{-L/L_*}\left[\phi_1 * \left(\frac{L}{L_*}\right)^{\alpha_1} + \phi_2 * \left(\frac{L}{L_*}\right)^{\alpha_2}\right] \tag{6.2}$$

Note that in case of a flux-limited sample, corrections have to be applied to the luminosity and mass functions to account for the fact that faint galaxies cannot be observed. As an example, the $1/V_{\mathrm{max}}$ technique (Schmidt, 1968; Felten, 1976) weights each galaxy i with the inverse of the volume $V_{\mathrm{max},i}$. $V_{\mathrm{max},i}$ corresponds to the maximum volume the galaxy i could be observed while still being included in the flux-limited sample. The computation of $V_{\mathrm{max},i}$ should include properly all the selection effects. In the simulation however this problem is not relevant.
A good match of the luminosity and mass function should thus reproduce the normalisation, the cutoff luminosity or mass, the "knee" L_* and the correct slopes at the faint-end.

Deriving stellar masses from photometric measurements implies some assumptions (IMF, choice of templates, SFH, dust modelling) which can affect the mass functions, while simulated mass functions do not suffer from these uncertainties. On the contrary, luminosity functions require only a few assumptions in observations. However, it implies a post-processing of the simulation (see Chapter 4) to obtain galaxy photometry. To draw conclusions independent of these assumptions, the best approach is to consider both mass and luminosity functions as complementary quantities. This is presented in the following paragraphs.

Mass function – I compare in Figure 6.2 the predicted stellar mass functions in Horizon-AGN to a set of observational data at $0 < z < 6$. The grey shaded region in this plot indicates the Horizon-AGN predictions (with the width of the region indicating Poisson uncertainties). The pink dashed line indicates predictions from Horizon-noAGN, the twin simulation without black hole feedback. Not unexpectedly, low-mass galaxies are largely unaffected by black hole feedback. However, agreement between theory and observation at the high-mass end of the mass function depends strongly on how this feedback is implemented.
The role of black holes is most important after the epoch of peak cosmic star formation ($z \sim 2$), as the black holes reduce star formation rates in massive galaxies (*e.g.* by maintaining the temperature of their hot gas reservoirs) and prevent them from becoming too massive.
Horizon-AGN agrees with observed mass functions across our redshift range of interest. The agreement is particularly good for galaxies that are more massive than the knee of the mass function at all epochs.

However, two points of tension between theory and observation are worth noting here: the overproduction of galaxies at the low-mass end and the general underproduction of galaxies regardless of galaxy stellar mass at $z \sim 5$. Horizon-AGN tends to overproduce galaxies less massive than the knee of the stellar mass function at low and intermediate redshift. While the predictions are consistent with observations if uncertainties due to cosmic variance are also taken into account, the systematic nature of the overproduction at all epochs indicates that this disagreement may be due to missing physics and that the subgrid supernova feedback prescription employed by Horizon-AGN is not strong enough to regulate star formation in smaller haloes, making the galaxies embedded in them too massive.

Luminosity functions – I now compare the predictions of Horizon-AGN with observed luminosity functions, since these are one of the basic quantities delivered by observations. Galaxy luminosities can be derived from the simulation using the virtual photometry built in Chapter 4. I recall here that one of the main uncertainties in the predicted luminosity is the derivation of the dust mass from the gas and metal content of the model galaxies (e.g Devriendt et al., 1999).

In Figure 6.3, I compare the predicted rest-frame K_s and r-band luminosity functions in Horizon-AGN to observational data in the redshift range $0 < z < 2$. Observational estimates of rest-frame K_s-band luminosities are largely restricted to this redshift range because observational data currently extends into the observed mid-infrared, from facilities like Herschel. Rest-frame K_s corresponds to the longest wavelengths at which stellar light still dominates the SED, the contribution of the interstellar medium becoming increasingly important longward of this filter. In addition, its negligible sensitivity to dust and young stars makes the K_s-band a good tracer of the underlying stellar mass of the galaxy. The r-band, being a shorter wavelength filter, is more sensitive to the mass-to-light ratio of the galaxy, which in turn depends on its star formation history.
Given its low sensitivity to dust, the rest-frame K_s-band luminosity function with and without extinction are almost identical. Therefore I only show the dust-reddened K_s-band luminosities. The shaded regions indicate the uncertainties in the observed luminosity functions, based on the errors in the fitted Schechter function parameters (but not including the effect of cosmic variance).
Horizon-AGN agrees well with the observed K_s-band luminosity function across the redshift range studied here. This agreement indicates that, on average, the aggregate mass growth of galaxies over cosmic time is well reproduced by the simulation (this is also borne out by the mass function analysis presented above). Similarly, the model also reproduces the rest-frame r-band luminosity function, to the same extent as its K_s-band counterpart. While the overall agreement of the luminosity functions is good, I note that at all redshifts, the simulation tends to overpredict the number of low-luminosity galaxies. A similar trend has been seen in the mass function analysis presented above. The fact that this overproduction is seen both in mass and luminosity functions suggest that it is not an effect purely driven by inherent bias in the observations. It arises more likely from the calibration of the subgrid physics in the simulation. I discuss this point later.

Star formation "main sequence" – Figure 6.4 presents the star formation "main sequence" predicted by Horizon-AGN compared to observational data in the redshift range $0 < z < 6$. Comparison to the observed "main sequence" probes whether the instantaneous star formation activity in the simulation is consistent with observations. Performing this exercise over a large redshift range therefore offers insights into how well Horizon-AGN predicts stellar mass growth over cosmic time compared to observations. This is complementary to the stellar mass function.
This Figure indicates that, within the dispersion between the observed main sequences, the simulation produces good agreement with the observations, over the redshift range of interest ($0 < z < 6$). The observed locii shown are median values and the observed main sequences typically have spreads of 0.5

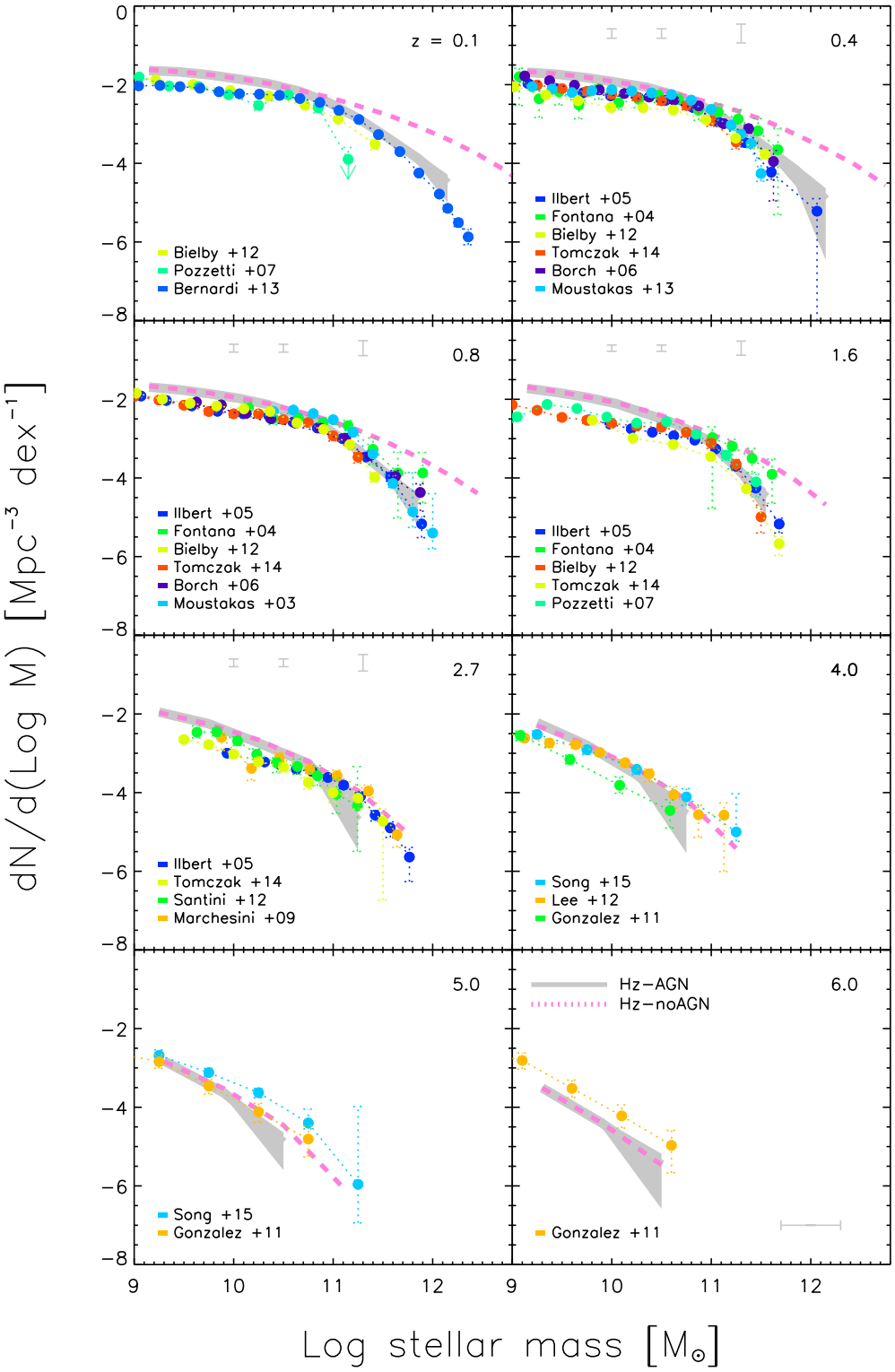

Figure 6.2.: Comparison of the predicted stellar mass function to observational data in the redshift range $0 < z < 6$. The grey shaded region shows the prediction from Horizon-AGN (with the width of the region indicating Poisson uncertainties). The pink dashed curves indicate predictions from Horizon-noAGN, a twin simulation without black hole feedback. Vertical errorbars indicate observational uncertainties due to cosmic variance (Ilbert et al., 2013). The horizontal errorbar (0.3 dex) indicates observational uncertainties in stellar masses derived from SED fitting, as found below. Note however that this value should be refined. The study presented below (Section 6.3) highlights that these uncertainties are mass and redshift dependent.

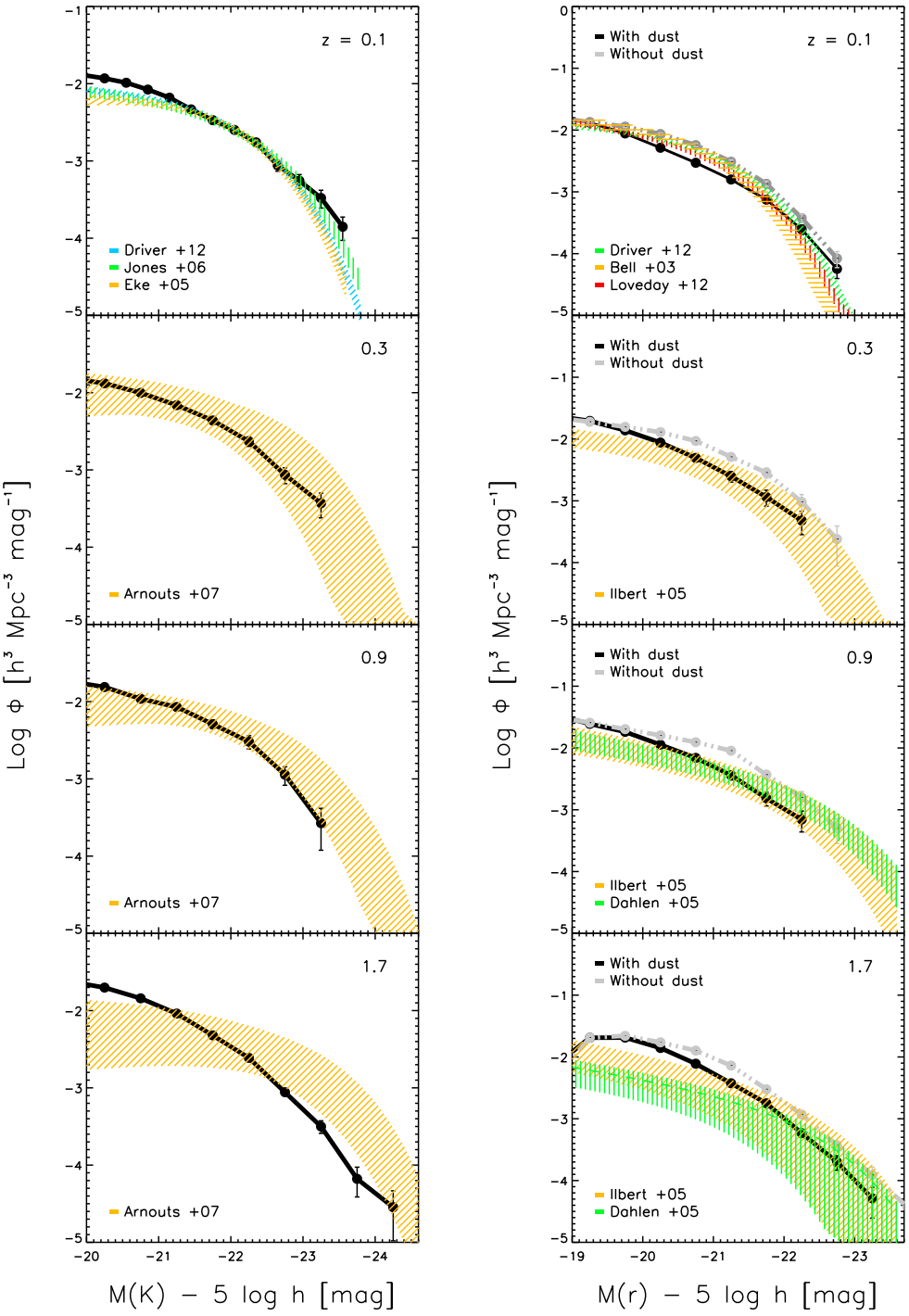

Figure 6.3.: Comparison of the predicted K_s (left-hand column) and r-band (right-hand column) luminosity functions from Horizon-AGN to observational data. The solid lines show dust-attenuated luminosity functions predicted by Horizon-AGN, while the grey- dotted curves show their unattenuated counterparts. Note that, given its low sensitivity to dust, the rest-frame K-band luminosity function with and without extinction are almost identical - we only show the dust-reddened K_s-band luminosities here. Observational data are shown using the coloured hatched regions (see legend for the individual datasets used).

dex, suggesting that the overlap between the theoretical and observational values is also reasonable at this redshift. Overall, these results indicate that Horizon-AGN predicts the observed "main sequence" with good accuracy between $z = 0$ and $z = 6$, suggesting that the aggregate stellar mass growth of galaxies with stellar masses greater than 10^9 $M_\odot$ is generally well reproduced by the simulation. However, it has to be noted that the simulated "main sequence" appears to fall slightly below the observed ones at $z < 1.7$, in particular in the low-mass range, while it is slightly above at very high redshift.
We can note that the trend observed in the "main sequence" is similar to the one observed in the cosmic star formation history in Figure 6.1. Although the cosmic star formation history tends to lie slightly above the observations at very high redshift and slightly below at low redshift, it is consistent with them within the errorbars. From the "main sequence" we learn that the agreement of the star formation rate is still good as a function of mass and not only as a function of redshift.

6.2 Galaxy populations and bimodality

6.2.1 Galaxy bimodality

I now investigate if the "bimodality" of galaxy population is reproduced in the simulation by comparing the predicted evolution of rest-frame colours in Horizon-AGN to observational data. Specifically I am comparing observed and predicted NUV-R-J diagrams. The UV (NUV-R) colour traces very recent star formation (stars with ages < 0.5 Gyr), with even residual ($<1\%$) mass fractions of young stars capable of driving galaxies into the UV blue cloud (Kaviraj et al., 2007a,b, and Chapter 3). The optical (R-J) colour, on the other hand, traces stellar mass growth over several Gyrs in the past (Kaviraj et al., 2007a). Taken together, these colours probe the formation history of the galaxy population over the last few Gyrs.
To perform this exercise in a consistent way at all redshifts, I use observational data from my cosmos2015 catalogue described in Chapter 5. The comparison is presented in Figure 6.5.
The Figure indicates that the model galaxies occupy similar parts of the NUV-R-J colour space as their observed counterparts. While the simulated galaxies agree well with the observed optical colours, the predicted bimodality in the UV colour is less pronounced than in observations, in particular at low redshift.
The good reproduction of the star formation "main sequence" and the optical colours indicates that this is probably partly due to small amounts of residual star formation in the model galaxies (recall that even negligible mass fractions of young stars can produce blue UV colours). While this residual star formation has little impact on the bulk stellar-mass growth of the galaxy population as a whole, the weakness of the bimodality in the predicted UV colours suggests that the feedback recipes employed by the model do not quench star formation as completely as is the case in real galaxies at low redshift.
To summarise, the simulation reproduces the general trend in the evolution of the rest-frame colour distribution. Nevertheless, and this is particularly visible at $z > 0.8$, the model does not reproduce the scatter in the observable properties, as was discussed in Chapter 1. There are several reasons that this may be the case. First of all, it could be due to the SSP model that has been chosen to compute the photometry in the simulation, the choice of the IMF or the implementation of dust. The effect of dust on the galaxy distribution in this diagram is discussed below. Another possible explanation is the calibration of the feedback at the subgrid scale, as already suggested and confirmed by the comparison of "main sequence", mass and luminosity functions. Finally, both the weaker bimodality and the narrower locus in the simulation could also suggest that, even if feedback plays an essential role in regulating star formation and creating a fraction of massive passive galaxies (as seen in the comparison of mass function in Figure 6.2), there could be additional processes regulating star formation which are not completely captured by the simulation. One could imagine that the local environment at the unresolved subgrid scale induces a spread in net outcome for the physical processes within that cell that is not properly accounted for in the simulation. Such

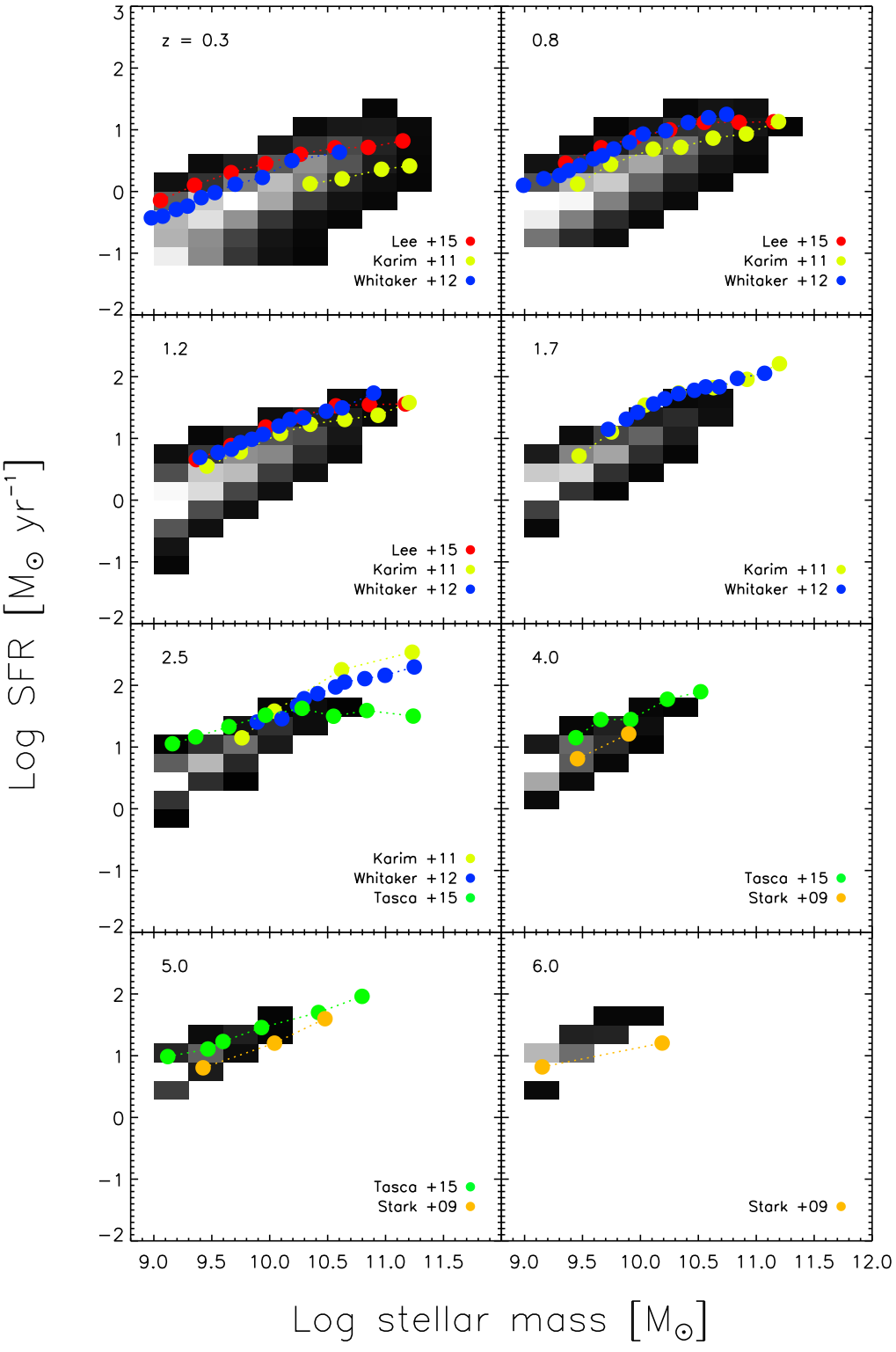

Figure 6.4.: Comparison of the predicted star formation "main sequence" in Horizon-AGN to observations in the redshift range $0 < z < 6$. Observational data are taken from Karim et al. (2011), who estimate SFRs via 1.4 GHz fluxes, and Lee et al. (2015), Whitaker et al. (2012), Tasca et al. (2015) and Stark et al. (2009), who estimate SFRs via multi-wavelength SED fitting. The Hess diagram indicates the predicted star formation "main sequence" from Horizon-AGN (darker shades indicating lower galaxy density).

diversity could *e.g.* be triggered by interactions of the larger scale environment which is not modelled in Horizon-AGN.

6.2.2 Effect of dust computation in the NUV-R-J diagram

Chapter 4 presented two methods to implement dust attenuation in the simulation. In preceding Sections, I used the sunset code to compute dust attenuation (assuming dust is a screen in front of each particle) instead of sunrise (performing the full radiative transfer treatment), as it is quicker and more adapted to large galaxy catalogues. In light of Figure 4.7, it should not affect the conclusions on the luminosity functions in the r and K_s band, for the choice of the method has less impact on optical and NIR luminosities. However, the NUV band is more strongly modified when performing the full radiative transfer. It is therefore worth exploring how it affects the rest-frame NUV-R-J diagram. In Figure 6.6 I show how the offsets between magnitudes computed with sunset and sunrise derived in Figure 4.7 modify the diagram. Applying these offsets enhances the bimodality slightly, but not enough to achieve reasonable agreement with the observations. The above conclusions therefore remain unchanged. The feedback recipes in the model appear unable to quench star formation to the extent that is required to produce the bimodality in rest-frame colours.

6.2.3 Going beyond current subgrid recipes in the simulation

Summing up together the results of the comparison between observations and simulations, it is possible to draw conclusions concerning the implementation of the stellar evolution and feedback. In fact, the overestimation of the mass function at the low-mass end at low redshift suggests that stellar feedback is not strong enough in the Horizon-AGN simulation. Star formation in low-mass galaxies is not regulated enough, and they consequently become too massive. They are then violently quenched at low redshift (as suggested by the "main sequence" and Figure 6 in Dubois et al., 2016) when most of the available cold gas has been consumed.

Recent works (*e.g.* Kimm et al., 2015) have shown that a more realistic treatment of the (inherently clumpy) interstellar medium and the momentum injection from supernova populations in the snowplow phase[1] can produce strong supernova-driven outflows that regulate star formation more efficiently than classical treatments. Alternatively, the simulated galaxy stellar masses can be reduced if the star formation efficiency is locally enhanced to the level that we observe in star clusters (10% per free fall time), as clustered star formation drives stronger winds (Agertz and Kravtsov, 2015) - enhanced efficiencies are indeed plausible, given the large fraction of field stars that are thought to come from the dissolution of star clusters. Such improved prescriptions for star formation are able to improve agreement with key observables, such as the stellar-to-halo mass relation and the mass-metallicity relation. While their implementation in Horizon-AGN is beyond the scope of this work, such revised recipes are likely to suppress the disagreement observed at the low-mass end, and will be explored in forthcoming works.

I now turn to the second point of tension between theory and observation. While the stellar mass function is well reproduced across cosmic time, agreement is poor at $z \sim 5$. At this epoch, model galaxies appear to be less massive than their observed counterparts across the entire mass range probed by our study. First of all it has to be noted that the observed stellar masses are more uncertain at these redshifts. Nevertheless, the mass and spatial resolutions of the simulation can also play an important role in biasing the predicted stellar mass growth of galaxies (Kimm et al., 2012). Adopting a higher resolution would permit resolving smaller haloes in the early universe ($z > 5$), leading to more star formation at these redshifts.

[1] The snowplow phase refers to the third and last phase of the supernova remnant evolution. During this momentum-conserved phase, the matter in the interstellar medium is swept up by the expanding shell.

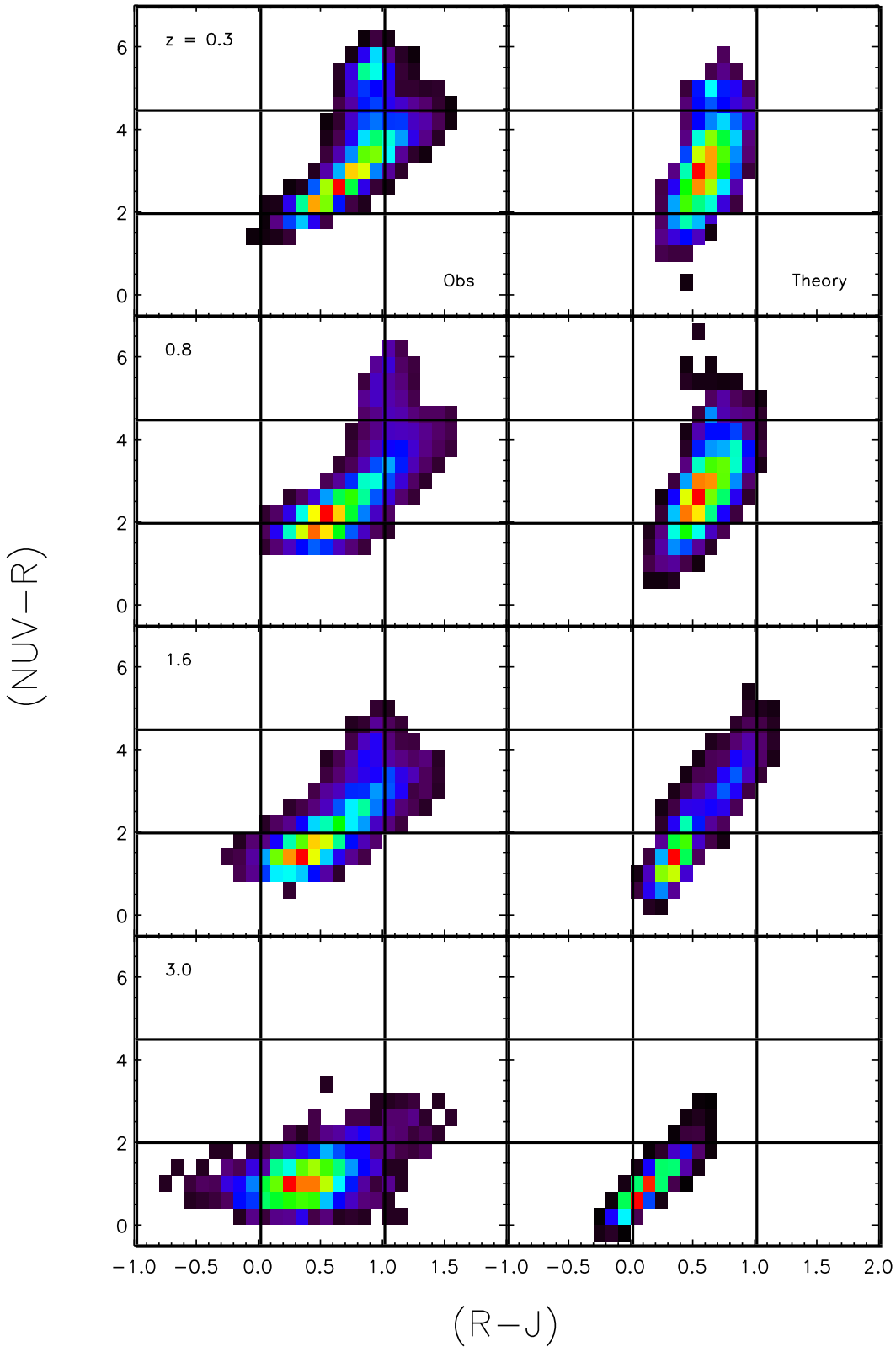

Figure 6.5.: Comparison of the predicted evolution of rest-frame *NUV*-*R*-*J* colours in Horizon-AGN to observational data from the COSMOS2015 catalogue (Laigle et al., 2016). The left-hand column shows observational data, while the right-hand column shows the predictions from Horizon-AGN. The Hess diagrams indicate the galaxy density (red = highest density, black = lowest density).

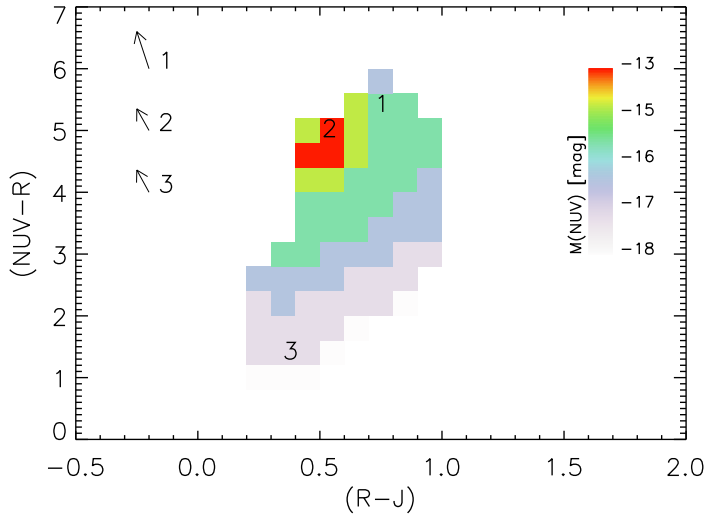

Figure 6.6: The predicted rest-frame *NUV*-*R*-*J* colours in Horizon-AGN at $z = 0.3$, colour-coded by the absolute *NUV* magnitude. The arrows indicate how the predicted colours are likely to move if the effect of dust were estimated using sunrise (full radiative transfer) rather than sunset (dust screen in front of each star particle).

This can result in an order of magnitude enhancement in the star formation activity in typical galaxies at this epoch (e.g Rasera and Teyssier, 2006). It is therefore probable that a higher-resolution simulation can improve the agreement between the theoretical and observed mass functions, without changing the baryonic physics currently implemented in Horizon-AGN.

6.3 Understanding observational biases through an end-to-end analysis

In this Section, I use the virtual photometric catalogue computed from the synthetic lightcone of Horizon-AGN to estimate some of the uncertainties and biases associated with the inference of physical properties from light (see Figure 6.7). The idea is to reproduce the observational uncertainties in the simulated dataset and estimate for the computed properties (essentially redshift, mass, age, metallicity) the effect of the degradation of the photometry (noise, PSF, blended objects), the method of extraction (fixed aperture) and the assumptions done at the SED-fitting stage (on the star formation history, metallicity distribution, dust extinction). In turn, one wants to correctly account for their effects in the above-presented comparison between observations and simulations. It is therefore important to estimate how the statistical estimators are affected: luminosity and mass functions, "main sequence", and the separation between passive and star-forming galaxies. These effects are usually modelled by convolving the simulated statistical quantities with the typical uncertainties inherent in the observational estimates. However in addition to uncertainties, biases are also driven by our observational methods and assumptions, especially at the SED-fitting stage. These biases can affect for instance the shape of the stellar mass function. The level of precision we have now reached in the comparison between observations and simulations means that these effects must now be carefully taken into account. For instance, the debated evolution in the shape of the high-mass end of the mass function (Ilbert et al., 2013; Marchesini et al., 2009) or the still controversial steepening of the low-mass end at high redshift (Ilbert et al., 2013; Santini et al., 2012) cannot be reasonably solved if these effects are not understood.

The performance of SED-fitting codes in recovering redshift and masses from photometry has been already investigated by comparing the physical quantities in the simulations with the one derived through SED fitting (Hildebrandt et al., 2010; Mobasher et al., 2015). Sorba and Sawicki (2015) showed that masses based on unresolved photometry may be underestimated. The importance of the dependance on SFH of mass and age estimation has been already assessed (Pforr et al., 2012). The dust implementation has also been shown to be an important ingredient (Pforr et al., 2012; Hayward and Smith, 2015; Guidi et al., 2016)

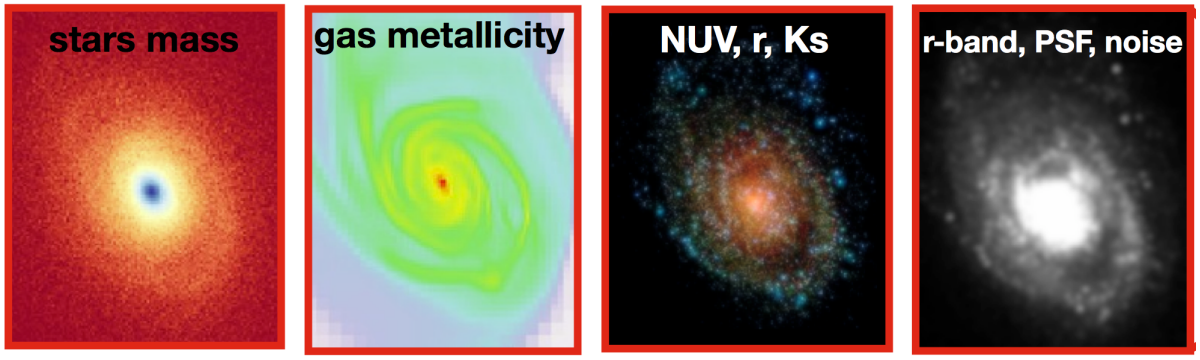

Figure 6.7.: A galaxy in Horizon-AGN. From *left* to *right*: its stellar particle distribution, its gas metallicity, the three-colours NUV-R-K_s image and the r-band including noise and PSF. The purpose of the study presented in 6.3 is to understand the biases associated to the inference of the former (stellar distribution) from the latter (degraded photometry).

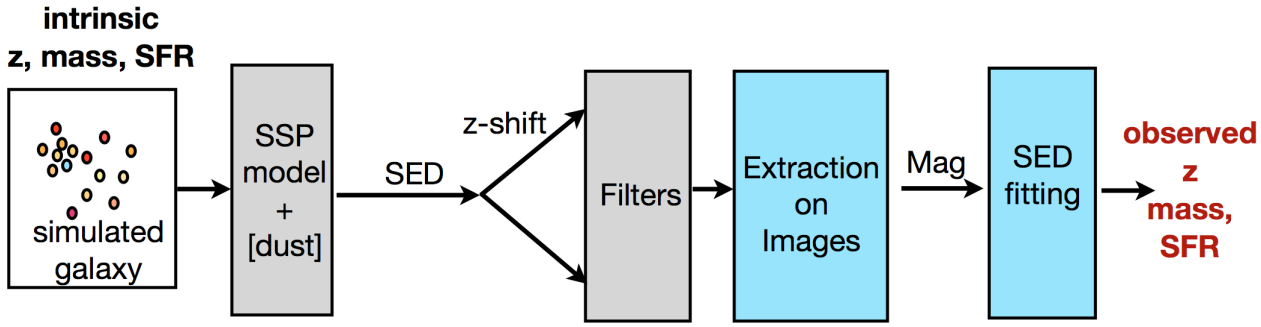

Figure 6.8.: The proposed "end-to-end" pipeline used to study biases inherent in the observational estimates. The scope of the study is to compare one- and two-point statistics on the quantities in input with the ones on the quantities in output.

in the estimation of galaxy properties. Finally, Guidi et al. (2016) explored how both observational biases and subgrid recipes may affect the comparison with observed galaxies, but with only a very small sample of 15 galaxies at redshift $z = 0$. All these works underline the importance of accounting for these biases, but no clear conclusions have been drawn on how they affect the evolution of the statistical quantities as a function of galaxy populations and redshift. The Horizon-AGN lightcone offers the ideal dataset for this analysis, for it is in broad overall agreement with observations as was shown above. Consequently, I plan to revisit the estimation of the biases using the lightcone and the comparison of the statistical estimators from this simulation with cosmos2015 *while fully taking into account these biases.*

The general outline is presented in Figure 6.8. The principle is to compare one- and two-point statistics on quantities in input with those in output. As this work is still preliminary, I will only focus in this Section on the biases associated to the estimation of the stellar mass through SED fitting. It is for now limited to low redshift ($0 < z < 1$) and the redshift is assumed to be the true galaxy redshift (no account of photometric redshift error). In addition, I do not implement dust attenuation. In fact in this simplified configuration, I test essentially the effect of the assumption on star formation history, and on the metallicity distribution. Note also that investigating uncertainties or systematics connected to the spectral synthesis are beyond the scope of this study (for such an investigation, see Conroy et al., 2009; Conroy et al., 2010; Conroy and Gunn, 2010). That is why I will be careful to use the same SSP models when computing the photometry from the simulation as when interpreting the photometry through SED fitting. The results presented below are part of a publication in preparation: Laigle, Davidzon, et al. 2016.

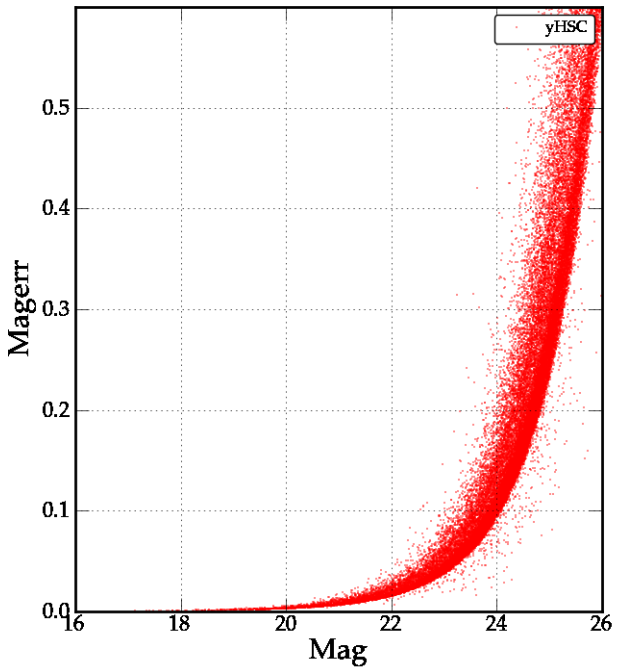

Figure 6.9: Magnitude errors as a function of magnitudes for the Y band from the HSC/*Subaru* in COSMOS2015. In each band, magnitude errors in the virtual catalogue from Horizon-AGN are calibrated on those of COSMOS2015.

6.3.1 Dataset preparation and SED fitting

Virtual catalogue – A set of apparent magnitudes coherent with its intrinsic physical properties has been assigned to each simulated galaxy (Chapter 4) using BC03 models with a Salpeter IMF in the following bands: NUV, u, B, V, r, i^+, z^{++}, Y, J, H, K_s, 3.6 μm, 4.5 μm, using the same filters passbands as used for the COSMOS2015 catalogue (see Chapter 5). To mitigate the effect due to the different stellar evolution models (BC03 for the photometry computation and Starbust99 for the stellar evolution in the simulation) in particular in the stellar mass loss (as shown in Figure 3.2), each SSP is rescaled to the initial mass of the particle (Chapter 4).
I then reduce photometric precision in the simulated catalogue to correspond to that in the COSMOS2015 catalogue. To better mimic observed data, magnitude errors are added to the catalogue by calibration on the COSMOS2015 magnitudes. In each band, the observed and simulated catalogues are divided in sub-samples of redshift width $\Delta z = 0.1$ and of magnitude width $\Delta m = 0.5$. Then the magnitude errors attributed to the objects in the simulated sub-samples are the median of the magnitude errors in the corresponding observed sub-samples (see Figure 6.9).

Stellar masses and star formation rate computation – Stellar masses are computed from the virtual photometry with the exact same tools as those used for the observed COSMOS2015 dataset. In this preliminary configuration, the redshift of each galaxy is fixed at its true value, and only the stellar mass is computed through SED fitting.
The virtual magnitudes included in the mock catalogue are given in input to LePhare (see Chapter 3 and 5). The following configuration is adopted:
i) The SED library contains BC03 templates with a Salpeter IMF;
ii) The models of star formation histories are exponentially declining and "delayed" (see Chapters 3). They cannot reproduce multiple bursts of star formation;
iii) BC03 models are computed with two metallicities (solar and half-solar);

iv) The dust attenuation is turned off since the virtual photometry does not take dust into account.
In LePhare, the star formation rate is directly computed from the star formation history and rescaled to the mass of the galaxy. This quantity is therefore very sensitive to the chosen star formation history, while this is less important for the stellar mass which corresponds in turn to an integration over time of the star formation history. For this reason the star formation rate value from SED fitting is in general not used, and other methods based on the photometry are preferred. However in my case it provides a good diagnostic of the success of the SED-fitting procedure.
Since both the virtual photometry and the SED fitting use BC03 templates with the same IMF, I do not expect that uncertainties and biases come from the choice of the templates. However, the star formation histories of simulated galaxies are complex compared to their simple parametrisation in the SED-fitting algorithm. In addition, the metallicity distribution within the galaxy is not constant in contrast to what is assumed in the SED-fitting code. In this simple exercise, I essentially test the effect of these two assumptions.

6.3.2 Impact on the computed stellar mass

Figure 6.10 shows how the masses and star formation rates of the simulated galaxies computed from the virtual photometry through SED fitting are biased compared to the intrinsic quantities in the simulation. To visualise a potential evolutionary trend, the redshift range $0.2 < z < 1$ is divided into four subranges.

Stellar mass comparison – At first glance, the mass recovered by the SED fitting is in good agreement with the intrinsic mass in Horizon-AGN within ± 0.3 dex. Nevertheless, some interesting redshift-dependent trends are apparent. The SED fitting underestimates the mass of all low-mass galaxies in the redshift range $0.6 < z < 1.0$. Below redshift 0.6, a certain population of galaxies in all mass ranges have their masses overestimated by the SED-fitting algorithm while a certain population of low-mass galaxies have their mass underestimated.

Star formation rate comparison – The star formation rate is globally well recovered by SED fitting within ± 0.3 dex. However a deviation in the distribution is noticeable. The star formation rate of low-SFR galaxies is overestimated compared to the star formation rate of high-SFR galaxies. A certain population of low-SFR galaxies have their star formation rate underestimated by SED fitting.
I also compared the age estimated by LePhare from the oldest SSP in the galaxy and the intrinsic age in the simulation. Let us now try to interpret these observed trends, focusing only on the effect of the simplifying star formation histories.

- I find that galaxies for which the masses are overestimated have also their ages overestimated and their star formation rates underestimated by LePhare. These populations have probably a complex star formation history for which the primary burst is predominant. The SED-fitting algorithm, which cannot fit multiple star formation bursts (regarding the simplistic models of star formation histories) will better recover the light from SSPs born in the first burst. It will probably overestimate the width of this burst trying to fit also the younger SSPs in the subsequent bursts. The mass will be therefore overestimated and the star formation rate underestimated (due to the exponential declin of the star formation history).

- In contrast, galaxies for which the mass is underestimated have their star formation rates overestimated and the ages underestimated by LePhare. These populations have probably a complex star formation history for which later bursts are predominant (Figure 6.11). The SED-fitting algorithm

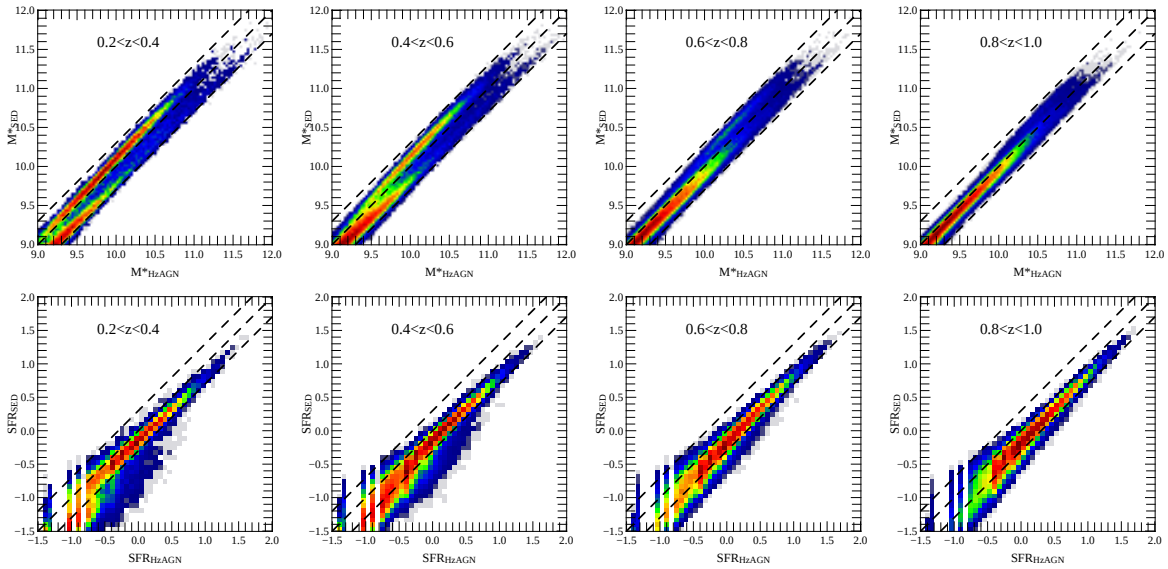

Figure 6.10.: Comparison between the logarithm of the masses (*top*) and SFR (*bottom*) in Horizon-AGN (x-axis) and from SED fitting with LePhare (y-axis). The colour indicates the density. From *left* to *right*, the following redshfit ranges are shown: $0.2 < z < 0.4$, $0.4 < z < 0.6$, $0.6 < z < 0.8$, $0.8 < z < 1.0$. The dashed lines shows $\log M_{\rm SED}/{\rm M}_\odot = \log M_{\rm Hz-AGN}/{\rm M}_\odot \pm 0.3$ and $\log {\rm SFR}_{\rm SED} = \log {\rm SFR}_{\rm Hz-AGN} \pm 0.3$.

will be biased towards young SSPs by the light emitted from recent bursts, and it will underestimate the light from the oldest SSPs. The galaxy will be therefore seen younger and less massive than it is really (see also *e.g.* Papovich et al., 2001, for a discussion on this topic).

Of course, this discussion is speculative and the results should not be over-interpreted. Importantly, these biases are population-dependent, in the sense that they affect differently galaxies with different star formation histories. To systematise the analysis, we could classify galaxies in sub-populations according to their star formation histories. In each subpopulation, we could then determine the mean star formation history and analyse how the output quantities are biased for this sub-population.

6.3.3 Discussion

To build robust statistical quantities, current photometric catalogues contain nowadays millions of objects. A rapid extraction of these objects and analysis of their photometries in terms of stellar content is possible only at the price of some assumptions at the different stages of the observational pipeline. Figure 6.12 shows an image of the COSMOS field (left panel) in the Y band from HSC/Subaru. The extraction of the photometry from such images is often done at fixed aperture (see Chapter 5). In this context, the total magnitude is inferred from this aperture magnitude, without taking into account the specific morphology of the object on which the light extracted at fixed aperture depends. The PSF and noise on the images, if not properly modelled, are another limitation of a right interpretation of the photometry. In addition, because we are seeing the sky in projection, objects at different redshifts are sometimes superimposed, which makes difficulty to properly extract their respective photometries. Although some recent developments of source extraction codes start to include a better accounting for these aspects, a comprehensive understanding of these limitations should be carry out for a fair analysis of the photometry. The mock photometric images (see Chapter 4 and right panel of Figure 6.12) built from the Horizon-AGN lightcone are the ideal tool for such an investigation. In turn, they would allow to make an "end-to-end" analysis of the biases (at

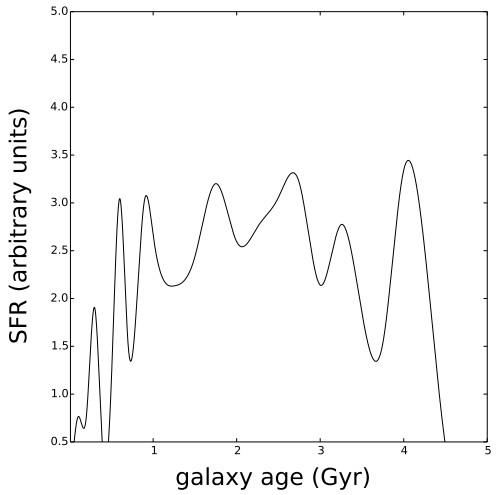

Figure 6.11: The complex star formation history of a low-mass galaxy in HORIZON-AGN. This galaxy has its mass underestimated by the SED-fitting algorithm.

the pixel level) starting from the extraction of the photometry. In this Chapter, I have only presented an "end-to-end" analysis (at the galaxy level) starting from the SED fitting but with simplified procedure, neglecting the effects of dust and the uncertainties on the redshift.

Unsurprisingly, I highlighted however that the computed mass is already biased compared to the intrinsic one due to over simplistic assumptions on the star formation history and the metallicity enrichment. Adopting more complex templates for the star formation history could help to reduce this bias (*e.g.* a two-component template).

Among the other biases, the effect of dust will be probably the most difficult to address, for our dust modelling on the simulation side is still quite uncertain and do not reproduce exactly the reality. Dust implementation with SUNRISE could nonetheless allow to test the effect of the assumption of a simple extinction law (as done in the SED-fitting algorithm) to reproduce a dust attenuation which in fact depends on the geometry of the galaxy.

Note that my end-to-end pipeline cannot mimic all the observational biases. For instance the SSP models that I used do not include emission lines, which are important to determine galaxy redshifts, and which could highly bias the massive end of the mass function if not taken into account. In addition, the choice of the IMF and stellar evolution models are indubitably impacting the computed quantities (as highlighted in Chapter 3). Finally, as discussed previously (Chapters 3 and 5) these quantities are also expected to be dependent on the wavelength coverage (the broader the wavelength coverage, the better the mass recovery).

6.4 Summary

In this Chapter, I have compared the predictions of HORIZON-AGN to an extensive array of observational data across cosmic time. This study has focussed on the temporal evolution of one-point statistics, while confronting the predicted evolution of luminosity functions, stellar mass functions, star formation "main sequence", rest-frame UV-optical colours, and cosmic star formation history with observational data in the redshift range $0 < z < 6$. These observables, which are functions of the evolving stellar mass growth of the galaxy population, strongly constrain our galaxy evolution model as implemented in the simulation. Recall that choosing black hole feedback parameters that reproduce the local black hole-bulge relations is the only explicit calibration of HORIZON-AGN to the local Universe. Nevertheless it has to be noted that some subgrid parameters such as the IMF and the star formation efficiency are by default calibrated on the local Universe for the only accessible observations to constrain them are done in the Milky Way.

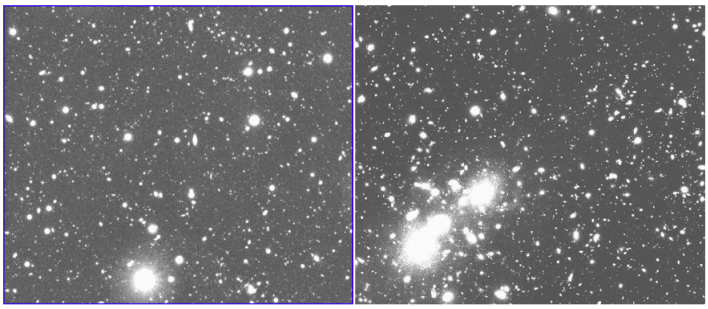

Figure 6.12: *Left*: A stamp from the COSMOS Y band from HSC/Subaru. *Right*: A mock image built from the Horizon-AGN lightcone, including the same PSF and level of noise as in COSMOS.

Since they are sensitive to the aggregate star formation history of galaxies, comparing the simulation to these quantities indicates how well Horizon-AGN captures the evolutionary trends of observed galaxies over cosmic time. This is also a measure of its usefulness as a tool for understanding galaxy evolution. This analysis shows that Horizon-AGN is broadly consistent with observations over our chosen redshift range, from the present day to the epoch when the Universe was ∼5% of its present age.

Despite this broad agreement, two points of tension are worth noting. First, the model tends to overproduce galaxies that are less massive than the knee of the mass function at all epochs. While this disagreement may be due to cosmic variance, it is more likely that the supernova feedback prescriptions in the model do not sufficiently quench star formation in small haloes. This is also highlighted by the too low galaxy metallicity. Because of insufficient stellar feedback, a too early peak of star formation in small galaxies restricts the formation of metals. More accurate modelling of the clumpy interstellar medium, combined with higher star formation efficiencies corresponding to star cluster formation may offer a solution to this disagreement. Secondly, at very early epochs the predicted galaxy stellar masses are too low across the mass range of interest in this study. Higher resolution simulations which are able to resolve smaller haloes in the early Universe may reduce the disagreement between theory and observation at these epochs.
Finally, a more accurate way to take into account observational biases should reduce the discrepancy. A preliminary study using the virtual photometric catalogue that I computed from Horizon-AGN has already highlighted systematic effects which can alter the mass computation. These systematics are due to overly simplistic assumptions made at the SED-fitting stage. Extending this study to other observables is essential. The ultimate goal is to understand how such biases affect mass and luminosity functions. Understanding them should allow to mitigate their effects on the comparison between observations and models.
Note in closing that in this Chapter I restricted myself to studying the cosmic evolution of one-point statistics, hence ignoring issues involving non local processes, which could strongly impact observables such as morphology or spin. This is the topic of the next Chapter.

Chapter 7

Angular momentum acquistion in the cosmic web

The cosmic web is the place in which small structures (haloes and galaxies) form and evolve. In the current paradigm of structure formation, matter is expelled from voids towards walls, then walls contract along two dimensions to form filaments and finally filaments contract along their three dimensions to form nodes (see Chapter 2). Haloes and galaxies form within this intricate web and one can expect that some of their properties (and at least their dynamics) are driven by the dynamics of the large-scale flow in which they are embedded.

However the dynamical relevance of the anisotropy of the cosmic web for galaxy formation and evolution may have been partially underestimated given the small amount of mass involved (in contrast to the mass in peaks). Spherical collapse and Press–Schechter theory have in fact been quite successful at explaining the mass function of galaxies (Press and Schechter, 1974). The galaxy mass distribution itself is sufficient to reasonably explain most of the observed galaxy properties, such as colours and star formation rates. Hence a fair fraction of the building of the observed "bimodality" (see Chapter 2) and the regulation of star formation ("quenching") appear to be mainly driven by the galaxy and host dark halo mass growth and processes controlled by their mass. Most of the baryonic processes involved in the regulation of star formation are indeed essentially mass dependent, for instance the growth of AGNs (for which feedback is one of the most important ingredients to regulate star formation, see Chapter 6). This mass growth occurs through mergers or gas infall, but at face value the important processes for driving galaxy properties are the mass and the energy involved, without further considerations about the intrinsic anisotropy of such processes. Indeed Chapter 6 demonstrated that the redshift evolution of one-point statistics of galaxy properties is well described by a one parameter facility set of functions driven by mass.
Nevertheless, this framework is not sufficient to explain in full certain observational properties. A well-confirmed, but still not fully understood, link exists between galaxy types and the density of their local environment (*e.g.* Dressler, 1980; Kauffmann et al., 2004; Blanton and Moustakas, 2009): red and passive galaxies dominate in high-density environment (clusters, which are in turn the nodes of the cosmic web) while blue star-forming disk galaxies dominate in low or intermediate density environment (corresponding to the density range covered by filaments). The connection between morphology and star formation activity on one side (*e.g.* Lee et al., 2013) and between morphology and angular momentum on the other side (*e.g.* Dubois et al., 2012b; Obreschkow and Glazebrook, 2014) suggests that observable galaxy properties depends at second order on their angular momentum, which is very likely driven by later infall of angular-momentum rich gas (Pichon et al., 2011; Danovich et al., 2012; Stewart et al., 2011; Danovich et al., 2015; Stewart et al., 2016).
In fact over the last 10 years, numerical simulations as well as theoretical considerations (Binney, 1977;

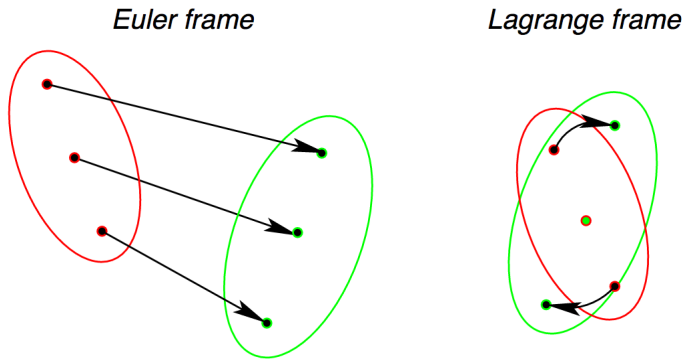

Figure 7.1: Tidal torque theory in a nutshell: a linear variation of the tide across the proto-object in the comoving Lagrange-frame corresponds to a rotation in the space-fixed Euler-frame after collapse.Figure from Schaefer (2009).

Katz et al., 2003b; Birnboim and Dekel, 2003; Kereš et al., 2005b; Ocvirk et al., 2008) have accumulated evidence that the intricate cosmic web may play an important role in the process of forming high-redshift galaxies. An important ingredient must therefore be the anisotropy of the infall of gas flows, driven by its dynamics within the cosmic web. Hence the properties of galaxies should be fully understood as driven by three fundamental quantities: mass, energy and angular momentum. Interpreting galaxy observations while omitting one of these underlying parameters is likely to be too simplistic.

In this Chapter, I review the acquisition of the angular momentum within the cosmic web. Chapter 2 presented the two possible descriptions of the structure growth: the *Lagrangian* and the *Eulerian* frameworks. Eulerian and Lagrangian descriptions brought together provide us the complementary framework to understand angular momentum acquisition. The main results reported in this Chapter give a particular emphasis to the Eulerian description, via the vorticity.
Section 7.1 reports on the measurement of the correlation between the angular momentum and the large-scale structure: low mass structures are aligned with the filament up to a certain transition mass before becoming perpendicular. It then presents the current paradigm for angular momentum acquisition in the Lagrangian framework of the tidal torque theory (TTT). This spin-flip is not fully understood in this classical *unconstrained* description of the TTT.
Sections 7.2 and 7.3 adopt a complementary point of view and present the Eulerian acquisition of angular momentum by studying its connection with the vorticity in the large-scale flows. This description explains satisfactorily the alignment up to the measured transition mass but does not explain the spin-flip. Those results have been presented in Laigle et al. (2015) and rely on dark matter simulations.
In view of these results, Section 7.4 connects the Eulerian vorticity description and the Lagrangian tidal torque theory. The latter can describe the angular momentum alignment (and flip) if augmented such as to take into account of a non local anisotropy (a filament embedded in a wall).
Finally Section 7.5 extends the measurement to baryonic matter, gas and galaxies, as presented in Dubois et al. (2014). In particular, it emphasises the consequences in terms of observable properties.

7.1 The current paradigm for angular momentum acquisition

7.1.1 The tidal torque theory

In the current paradigm, haloes and galaxies acquire angular momentum[1] by tidal torquing from the ambient matter distribution (Hoyle, 1949; Peebles, 1969; Doroshkevich, 1970; White, 1984) because of a misalignment between inertia tensor and the gravitational tidal tensor at the time of maximum expansion.

[1]In the following, I will use equivalently "angular momentum" or "spin".

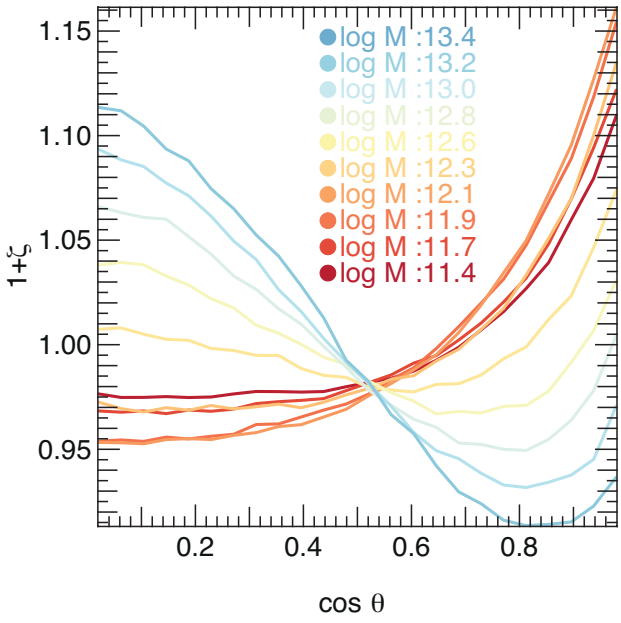

Figure 7.2: The probability distribution of the cosine of the angle between the spin of dark haloes and the direction of the closest filament as a function of mass in the $\mathcal{S}_{2000}^{\rm CDM}$ simulation. The smoothing length over which filaments are defined is 5 h^{-1}Mpc.

Their spin should be consequently initially correlated with the axes of the tidal tensor and thus the large-scale structures (see *e.g.* the review by Schaefer, 2009, and Figure 7.1).

Intuitively, imagine a dancer spinning on herself. A change in her angular momentum is equal to the torque applied on her, times the duration the torque was acting. If she maintains her arms extended a certain duration while she is exposed to such torques, she will gain a certain amount of angular momentum. Once torquing stops, she conserves her angular momentum which allows her to spin on herself when she folds her arms.
The idea is the same with proto-haloes: while exposed to tide (up to the maximum extension), each proto-halo is distorted and acquires a spin. The angular momentum $\mathbf{L}$, which is conserved during the collapse, can be approximated as (White, 1984):

$$L_i = \sum_{j,k,l} a^2(t)\dot{D}_+(t)\epsilon_{ijk}I_{jl}T_{lk}\,, \tag{7.1}$$

where $a(t)$ is the scale factor, D_+ the growth factor, T_{ij} the tidal tensor (detraced Hessian of the gravitational potential) and I_{ij} the proto-halo inertia tensor (only its traceless part, $\overline{I}_{ij}$ contributes to the spin). In its classical form, this description considers the environment through tidal fields but assumes they are unconstrained. In the presence of the cosmic web, this assumption is too simplistic, in particular because the larger scale environment is subject for instance to tides from filaments and walls in the vicinity. Such an extension of TTT allows us to explain the alignment of the angular momentum of low-mass haloes with the filaments as seen in simulations (*e.g.* Codis et al., 2012), and also explain the mass dependent spin-flip. In the following Section, I explain this point in more detail.

7.1.2 Evidence for a mass dependent halo spin-filament alignment

The link between dark matter halo spins and the cosmic web has been investigated for a few years in simulations (Aragón-Calvo et al., 2007; Hahn et al., 2007a; Sousbie et al., 2008; Paz et al., 2008; Zhang et al., 2009; Codis et al., 2012; Libeskind et al., 2013a). Despite initial apparent contradictions between some of these investigations due to the fact that they were probing different mass scales, a clear model

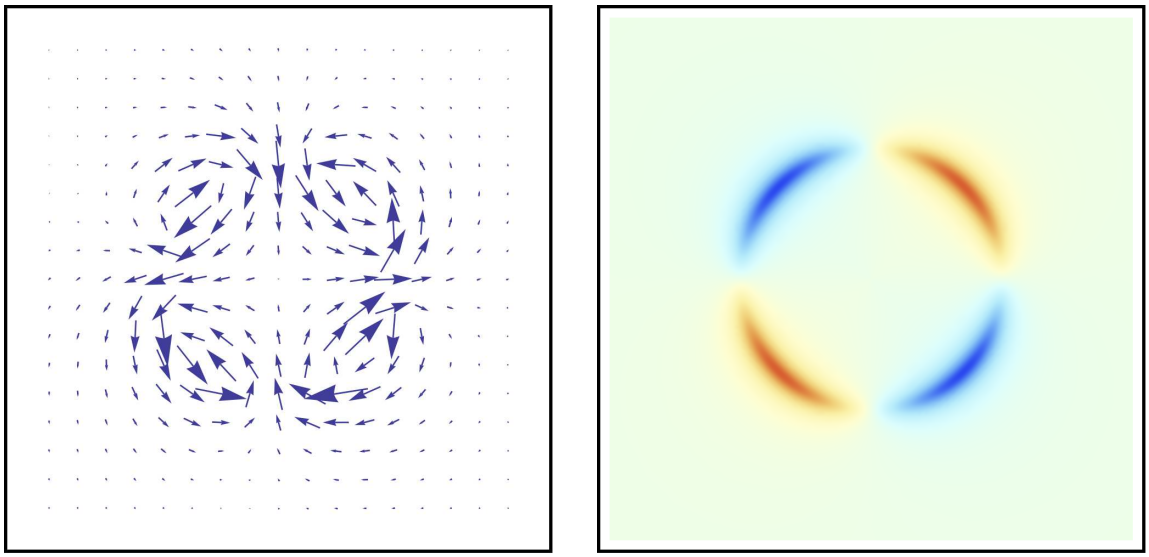

Figure 7.3.: A velocity field (*left*) and the associated map of the vorticity (*right*) projected along $\mathbf{e_z}$ (orthogonal to the sheet plane). Vorticity is defined as the curl of the velocity field. Hence it captures how the fluid rotates.

has finally emerged: smaller haloes have their angular momentum parallel to the direction to the filament, while more massive haloes have their angular momentum perpendicular. A redshift-dependent mass transition $M_{\rm crit}$ has thus been detected, varying with redshift or scale in proportion to one tenth of the mass of non-linearity (or equivalently with the hierarchical level of the cosmic structure in which the halo is embedded; see Aragon-Calvo and Yang, 2014).
This transition is presented in Figure 7.2. The measurement was made on the Horizon-4π simulation[2] which is a 2000 h^{-1} Mpc box with 4096^3 dark matter particles (Teyssier et al., 2009). The simulation allowed us to identify over 34 million haloes. In this figure, one observes that the probability to have a small angle between the halo spin and the filament direction first *increases* (in red) as mass grows to $\log M/\mathrm{M}_\odot \sim 12.1$. At larger masses (from orange to blue) the statistical spin–filament alignment quickly decays, with a critical mass (in yellow) corresponding to a transition to predominately orthogonal orientations (in blue) at $\log M_{\rm crit}/\mathrm{M}_\odot \approx 12.7$ (at $z \sim 0$) as defined by Codis et al. (2012).
Interestingly, Tempel et al. (2013) and Zhang et al. (2013) have recently found evidence of such alignment in the Sloan Digital Sky Survey: an orthogonal aligment for S0 galaxies and a weak alignment for late-type spirals.

7.1.3 Vorticity: the Eulerian framework for this acquisition

Vorticity in the cosmic web – Vorticity is defined as the curl of the velocity field and captures how a fluid element rotates, as illustrated on Figure 7.3. The spin of the dark halo represents, in essence, the vortical motion of matter and therefore can be expected to correlate with the vorticity in the surrounding protogalactic patch of a forming halo. Hence the vorticity of the flow should be the relevant property of the large-scale flow to understand angular momentum alignment. As mentioned in Chapter 2, in the Universe primordial vorticity decays very rapidly and the large-scale flows are laminar and curl-free in the linear regime. However vorticity is generated in multi-steam regions after shell-crossing. In fact Pichon and Bernardeau (1999) theoretically demonstrated that in the simplest pancake-like collapse (as briefly

[2] called $\mathcal{S}^{\rm CDM}_{2000}$ in the following.

described in Chapter 2) the level of vorticity generated is of the order of Hubble constant at collapse at the scale defined by the thickness of the forming structures. In particular, an important consequence of this work is that vorticity is confined to filaments. The cross-section of the vorticity with a plane perpendicular to the filament should also displays a quadripolar geometry.
More recent works have confirmed this vorticity generation and distribution, in particular by measuring one– and two–point statistics on the vorticity (Pueblas and Scoccimarro, 2009; Hahn et al., 2015). Wang et al. (2014) revisited this description by introducing three invariants of the velocity gradient tensor and concluded that vorticity generation is highly correlated with large-scale structure before and after shell-crossing, in a way which depends on the flow geometry.

Vorticity: a clue to explain angular momentum alignment – The above-mentioned spin alignment and mass transition could then be understood because of mass advection in an anisotropic, multi-flow environment. Codis et al. (2012) interpreted the alignment in terms of large-scale cosmic flows: high-mass haloes have their spins perpendicular to the filament because they are the results of major mergers (see also Peirani et al., 2004); low-mass haloes acquire their mass from incoming gas originating from embedding walls and flowing into filaments, which explains that their spins are parallel to the filament (see also Aubert et al., 2004; Bailin and Steinmetz, 2005; Pichon et al., 2011). Hence Codis et al. (2012) speculated that secondary shell-crossing could lead to the formation of vortices aligned with the forming filament. In turn, these vortices would build up the spin of these haloes. In this context, it could be appropriate to study this alignment while quantifying the rotational motion of the large-scale flow within filaments. Jointly with Laigle et al. (2015), Libeskind et al. (2013b) already have argued that the local vorticity was more relevant than the original tidal field in setting up the direction of dark halo spins. They explored the link between vorticity in halo environment and the origin of haloes spin and found a strong alignment between both. Vorticity tends to be perpendicular to the axis along which material is collapsing fastest.
Taking these considerations into account, my focus in this Chapter will be in revisiting these findings with an emphasis on *where* (tracing the filaments) and *why* (studying the origin of the vorticity and its orientation) these alignments are detected. I will also tentatively explain the origin of the mass transition for halo–spin alignment with the large-scale filaments. The main question addressed in the following are: is there statistical evidence that swirling filaments are responsible for spinning up dark haloes and gaseous discs? The focus will be mostly on lower mass haloes ($M_\odot < 5 \times 10^{12}$) for which angular momentum is essentially driven by the dynamics of the anisotropic infall.

7.2 Vorticity within the cosmic web

Let us first investigate the locus and the geometry of the vorticity within the cosmic web. This study is based on dark matter only simulations, which are presented below.

7.2.1 Simulations and tools

Dark matter simulations – The statistical results presented here rely on a set of dark matter (ΛCDM) simulations (see Chapter 4 for a brief description of the implementation of the dark matter dynamics in simulations).
Different box sizes are used here and adapted to different purposes. I used a 100 h^{-1} Mpc box with an initial mean spatial resolution of 390 h^{-1} kpc (256^3 dark matter particles) in order to build a statistical sample of haloes and filaments, several 50 h^{-1} Mpc boxes with a mean spatial resolution of 190 h^{-1} kpc (256^3 particles), and a 20 h^{-1} Mpc box with a mean spatial resolution of 39 h^{-1} kpc (512^3 particles). All

Table 7.1.: The set of simulations used in this Chapter. The so-called ΛHDM subset corresponds to simulations, the initial conditions of which have been smoothed over 2.3 h^{-1} Mpc and 0.23 h^{-1} Mpc. The velocity field, density field, initial conditions were smoothed with Gaussian filter with a kernel length of $\mathbf{R}_{\rm xxx}$.

Name	Type	Box size	Resolution	$\mathbf{R}_{\rm velocity}$	$\mathbf{R}_{\rm density}$	$\mathbf{R}_{\rm Lagrangian}$	Minimum halo mass
		h^{-1} Mpc		h^{-1} Mpc	h^{-1} Mpc	h^{-1} Mpc	$10^{10}\mathrm{M}_\odot$
$\mathcal{S}_{100}^{\rm CDM}$	ΛCDM	100	256^3	0.39	2.3	–	44
$\mathcal{S}_{100}^{\rm HDM}$	ΛHDM	100	256^3	0.39	2.3	2.3	–
$\mathcal{S}_{50}^{\rm CDM}$	ΛCDM	50	$256^3 \times (20)$	0.78	1.2	–	6.2
$\mathcal{S}_{20}^{\rm CDM}$	ΛCDM	20	512^3	0.039	0.23	–	0.044
$\mathcal{S}_{20}^{\rm HDM}$	ΛHDM	20	512^3	0.039	0.23	0.23	–

these simulations were run with GADGET (Springel et al., 2001) using a softening length of $1/20^{\rm th}$ of the mean inter-particle distance. The simulations are presented in Table 7.1.
As explained in Chapter 2, warm dark matter simulations allow to better visualise what happens at large-scale because small-scale modes are damped *ab initio*. Here the ΛHDM subset $\mathcal{S}_{100}^{\rm HDM}$ corresponds to simulations with initial conditions that have been smoothed with a Gaussian filtering on scales of 2.3 and $0.23h^{-1}$ Mpc sharing the same phase than the ΛCDM subset $\mathcal{S}_{100}^{\rm CDM}$. I will use it for visualisation and interpretation only.
All the simulations are studied at redshift $z = 0$ and are characterized by the following cosmological parameters: $\Omega_{\rm m} = 0.24$, $\Omega_\Lambda = 0.76$, $n = 0.958$, $H_0 = 73\,\mathrm{km.s^{-1}\,Mpc^{-1}}$ and $\sigma_8 = 0.77$ within one standard deviation of 3 year Wilkinson Microwave Anisostropy Probe results (Spergel et al., 2003).

Dark matter halo definition and spin – Dark matter haloes are defined using the Friend-of-Friend Algorithm (or FOF; Huchra and Geller, 1982). The used linking length is $0.2(L_{\rm box}^3/N_{\rm part})^{1/3}$. I only consider haloes with more than 100 particles, which corresponds to a minimum halo mass of 62×10^{10} $\mathrm{M}_\odot$ in $\mathcal{S}_{50}^{\rm CDM}$. The spin of a halo is defined as the sum over its particles i:

$$\sum_i (\mathbf{r}_i - \overline{\mathbf{r}}) \times (\mathbf{v}_i - \overline{\mathbf{v}})\,, \tag{7.2}$$

where $\overline{\mathbf{r}}$ is the centre of mass of the FOF and $\overline{\mathbf{v}}$ its mean velocity.

The FOF is prone to spuriously link neighbouring structures with tenuous bridges of particles, leading to artificial objects with a very high velocity dispersion, which could eventually bias the measure of the spin and consequently the alignment of the spin and the vorticity measured in the following. An additional criterion is therefore required to produce a trustworthy catalogue of haloes. Following Bett et al. (2007), I proceed using the distribution of the spin parameter defined by Peebles (1969):

$$\lambda = J|E|^{1/2}/\mathrm{G}M_h^{5/2}\,, \tag{7.3}$$

where J is the magnitude of the spin, E is the total energy of the halo, G is the gravitational constant and M_h is the halo mass.
Figure 7.4 shows the average normalized histogram of the logarithm of the spin parameter for the haloes in the simulations set $\mathcal{S}_{50}^{\rm CDM}$. At high spin we clearly see a long tail, up to $\lambda = 238.2$, due to spurious linking of the structures. I use the analytical model proposed in Bett et al. (2007) to fit the $\log\lambda$-distribution:

$$P(\log\lambda) = \mathrm{A}(\lambda/\lambda_0)^3 \exp\left[-\alpha(\lambda/\lambda_0)^{3/\alpha}\right]\,, \tag{7.4}$$

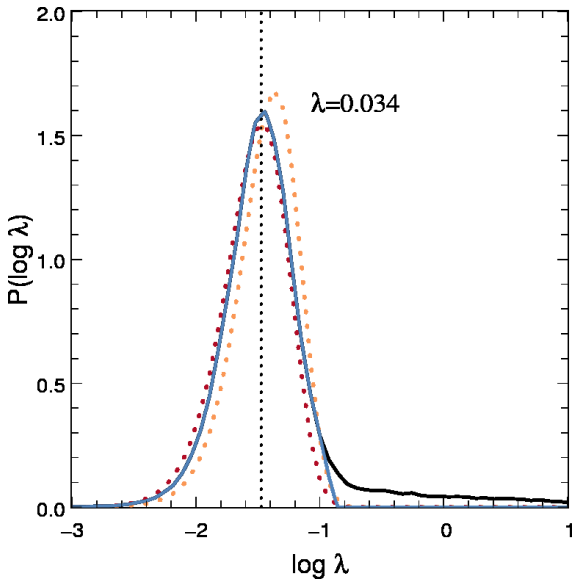

Figure 7.4: The normalized histogram of the logarithm of the spin parameter in the simulations set $\mathcal{S}_{50}^{\rm CDM}$. The median value of λ is 0.034 $\pm$ 0.0005 for the cleaned catalogue. The solid black line is the normalized distribution for all haloes. The solid blue line is the distribution for haloes with $\lambda \leq 0.12$, with the same normalization as for all haloes. The dotted lines are the analytical fit described in the text, red is my best fit to the distribution, and orange is the best fit found by Bett et al. (2007).

where $A = 3 \ln 10\, \alpha^{\alpha-1}/\Gamma(\alpha)$, with the values $\lambda_0 = 0.0341$ and $\alpha = 2.98$ which are providing the best fit. These values are in good agreement with those found by Bett et al. (2007) ($\lambda_0 = 0.043$ and $\alpha = 2.51$), though their way to clean their catalogue (TREEall) is more sophisticated, in particular by taking into account an additional condition on energy. They showed also that the minimal number of particles per halo, N_p, clearly affects the λ-distribution only for N_p lower than 100. Above this threshold, the change in the median value of λ is less than 10%. Consequently, I keep in my catalogue only haloes with more than 100 particles. These haloes are then selected through a cut in λ. I find that removing haloes with $\lambda \geq 0.12$ best fits the adopted analytical model. Removed haloes represent 9.4 $\pm$ 1.2 % of the total population. Inspection of some of these removed haloes shows that they generally are multiple objects. Finally they are around 5000 haloes in each 50 h^{-1} Mpc box of the $\mathcal{S}_{50}^{\rm CDM}$ simulations set[3].

Velocity and vorticity fields – In these dark matter simulations, the velocity is only sampled at the discrete location of the particles. A unfortunate consequence of a naive estimation of the velocity from the particle distribution would be that in regions where no particle is present, the velocity would be set to zero, while in practice the velocity field is undetermined in these regions.
As discussed in Pueblas and Scoccimarro (2009), I sampled optimally the velocity field using a Delaunay tessellation. To this end, I first computed the Delaunay tessellation on the particle distribution. For each cell of the three-dimensional grid, I computed the volume V_j and the mean velocity u_j for each Delaunay tetrahedrons contained in the cell. Each tetrahedron is defined by the three vectors v_k determining its geometry, and the four velocities u_j^i at the vertices. The volume of the tetrahedron is defined as

$$V_j = |\det(\mathbf{A})|\,, \tag{7.5}$$

with $\mathbf{A}$ =(v_1,v_2,v_3). The mean velocity is given by:

$$u_j = \frac{1}{4} \sum_{i=0,3} u_j^{(i)}\,. \tag{7.6}$$

[3]I quantified how the cut in λ affects the vorticity–spin alignment results presented below. Considering three different catalogues with three different cuts in λ ($\lambda < 0.08$, $\lambda < 0.12$, $\lambda < 0.2$) we look for each catalogue and for each bin of mass at the quantity $(\zeta_{\rm tot} - \zeta_{\rm cut})/(1 + \zeta_{\rm tot})$ where $1 + \zeta_{\rm cut}$ is the excess of alignment in the reduced catalogue and $1 + \zeta_{\rm tot}$ in the full catalogue. This difference is always $< 5\%$. I concluded that including or not the misidentified structures does not significantly change the measure of the spin alignment with the vorticity.

Finally, the velocity on each cell of the grid is given by:

$$\frac{\sum_j V_j u_j}{\sum_j V_j} \,. \tag{7.7}$$

The vorticity of the velocity is then measured from the resampled velocity at each point of the grid as the curl of the velocity field

$$\omega \,=\, \nabla \times \mathbf{v} \,, \tag{7.8}$$

after Gaussian smoothing of the velocity field (The kernel lengths are given in Table 7.1 and depend on the simulation)[4].

A comparison between vorticity maps in $\mathcal{S}_{20}^{\rm CDM}$ and in $\mathcal{S}_{20}^{\rm HDM}$ is shown in Figure 7.5. Vorticity along the normal to the section is plotted in the right panels of this figure. In $\mathcal{S}_{20}^{\rm HDM}$, high frequency features are suppressed but the low-frequency vorticity remains consistent with that of the more realistic $\mathcal{S}_{20}^{\rm CDM}$. In $\mathcal{S}_{100}^{\rm HDM}$, the smoothing is chosen such that in high-vorticity regions (defined here as being regions where the vorticity is greater than 20 per cent of the maximum vorticity), the mean vorticity is of the order of 90 $h\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$, i.e. it corresponds more or less to one revolution per Hubble time, in agreement with the theoretical predictions of Pichon and Bernardeau (1999). The orders of magnitude are similar in $\mathcal{S}_{100}^{\rm CDM}$, $\mathcal{S}_{20}^{\rm HDM}$ and $\mathcal{S}_{20}^{\rm CDM}$.

This resampling of the velocity using a Delaunay tessellation provides much more accurate results than a fixed kernel smoothing (Jennings, 2012). However an intrinsic small-scale noise is associated to the method. In fact this method interpolates between velocity which are close in configuration (position) space, but not necessarily in phase-space. Hahn et al. (2015) showed that the properties of the volume-averaged velocity field (as obtained with the resampling of the velocity on a Delaunay tessellation) are not exactly identical to the properties of the mean velocity field (see also Shandarin et al., 2012). To solve this issue, they propose a new and more accurate method which involves computing the mean velocity from a Lagrangian tessellation in phase space, and in doing so they improve the estimate of the cosmic velocity and vorticity. Although resampling the velocity on the Delaunay tessellation is sufficient for the purpose of my work, further investigations on the vorticity especially in regions having shell-crossed more than twice (see Section 7.4) could require the use of this more accurate method (Abel et al., 2012).

Skeleton computation – The cosmic network is identified with rSeX and DisPerSE, the filament tracing algorithms based on either watersheding (Sousbie et al., 2009) or persistence (Sousbie, 2011; Sousbie et al., 2011) without significant difference for the purpose of this investigation. I explained the DisPerSE code in Chapter 4.
In this Chapter, the scale at which the filaments are traced (6 pixels Gaussian for each simulation) corresponds to large enough scales so that we are investigating the flow relative to the large-scale structures. Filaments are defined as a set of small segments linking neighbours pixels together. The mean size of the segments is 0.6 pixels, which means 234 h^{-1} kpc in $\mathcal{S}_{100}^{\rm CDM}$.
For comparison with previous studies (*e.g.* Libeskind et al., 2013b), walls are defined according to the density Hessian. Given λ_i the eigenvalues of the Hessian $\mathcal{H} = \partial^2\rho/\partial \mathrm{r}_i \partial \mathrm{r}_j$ where ρ is the density field, with $\lambda_i > \lambda_j$ if $i < j$, walls are identified as being the region of space where $\lambda_1 > \lambda_2 > 0$ and $\lambda_3 < 0$. The normal of a wall is given by the direction of the eigenvector associated with λ_3. To obtain the Hessian, the density field of $\mathcal{S}_{100}^{\rm CDM}$ is smoothed with a Gaussian filter of 1.6 h^{-1}Mpc and differentiation of the density field is performed in Fourier space.

[4] I checked that the results of the alignments do not qualitatively depend on the smoothing scale. If I change the smoothing scale, the main conclusion remains unchanged, even if the magnitude of the signal varies slightly (but not monotically) according to the scale.

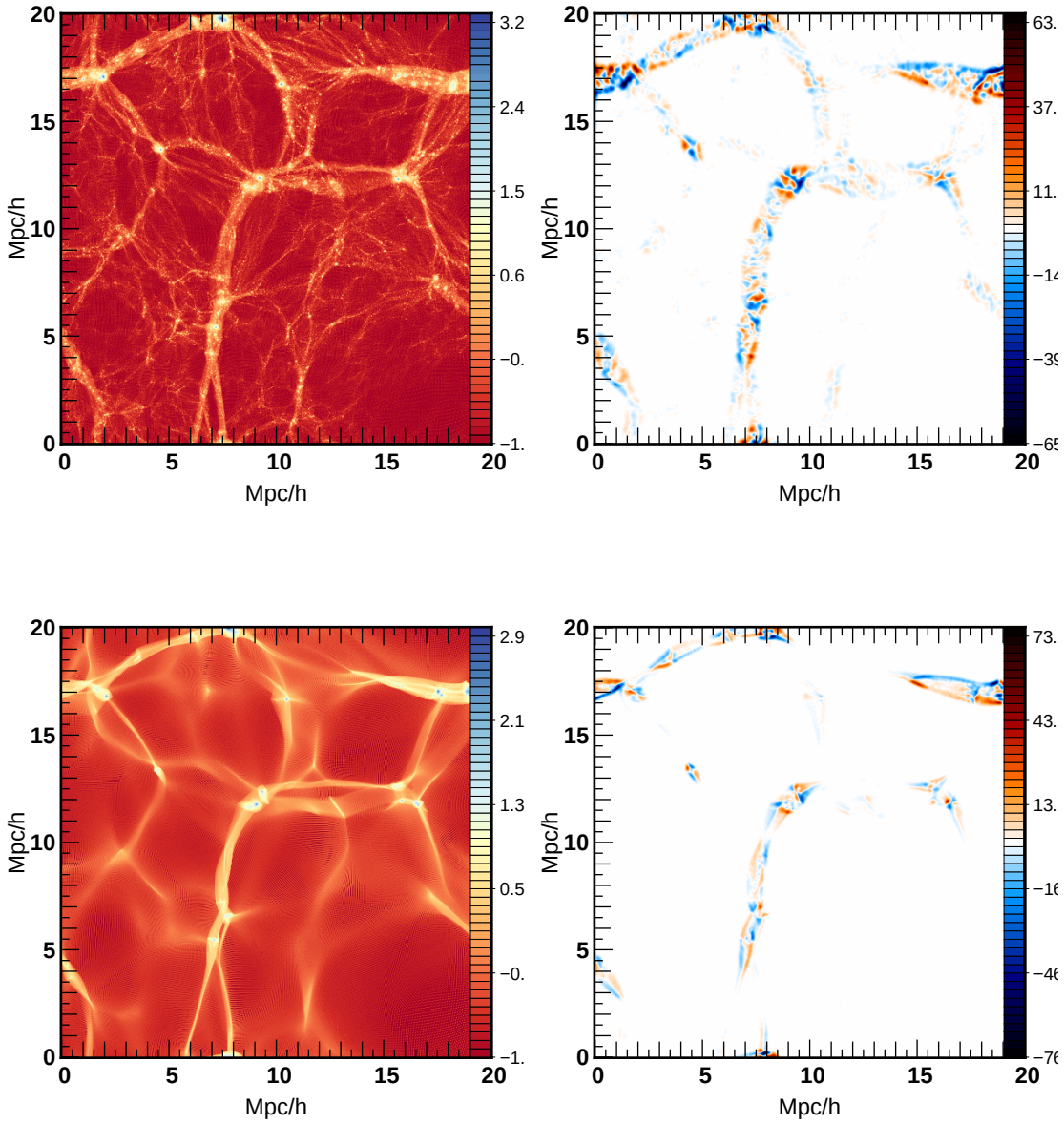

Figure 7.5.: A thin slice ($2\,h^{-1}$ Mpc thickness) of the projected dark matter density (*left* panels) and the projected vorticity along the normal to the slice in unit of $h\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$ (*right* panels). Dark matter density is plotted using a logarithmic scale. Vorticity is computed after smoothing of the velocity field with a Gaussian filter of 160 h^{-1} kpc for this figure only. The geometry of the vorticity closely follows the LSS, but switches polarity across the walls/filaments (recalling that walls appear as filaments and filaments as peaks in such a cross-section). Note also how the vorticity is localized around filaments (the 2D peaks, as exemplified in Figure 7.11). The two panels allow for a comparison between a section of $\mathcal{S}_{20}^{\mathrm{CDM}}$ (*top*) and $\mathcal{S}_{20}^{\mathrm{HDM}}$ (*bottom*). In $\mathcal{S}_{20}^{\mathrm{HDM}}$, high-frequency modes are suppressed but the low-frequency vorticity is qualitatively consistent with that found in the realistic $\mathcal{S}_{20}^{\mathrm{CDM}}$. On the bottom left panel, the density caustics are quite visible and correspond to the outer edge of the multi-flow region in the bottom right panel.

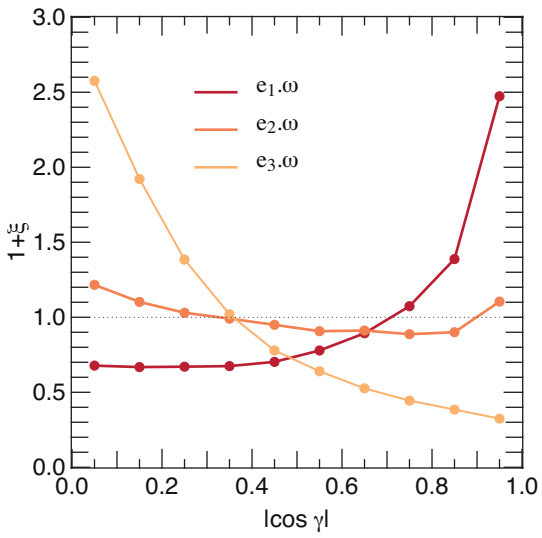

Figure 7.6: The probability distribution of the cosine of the angle between the vorticity and the eigenvectors of the tidal tensor, measured in $\mathcal{S}_{100}^{\rm CDM}$. The vorticity tends to be perpendicular to the minor axis (e_3) of the tidal tensor. The vorticity tends also to be aligned with the major axis (e_1). e_3 corresponds to the axis along which material is collapsing fastest.

7.2.2 Correlation between vorticity and filaments

Vorticity and tidal field alignment Libeskind et al. (2013b) investigated the alignment of the vorticity with the shear tensor. Figure 7.6 displays the probability distribution of the cosine of the angle between the vorticity and the eigenvectors of the tidal field tensor, $\cos\gamma$. The vorticity tends to be perpendicular to the minor axis (e_3) of the tidal tensor which corresponds to the axis along which material is collapsing fastest. The vorticity tends also to be aligned with the major axis (e_1). In the following, all the statistical alignments are quantified in Table 7.2. These results are in qualitative agreement with those of Libeskind et al. (2013b).

Alignment of vorticity with filament The alignment of vorticity with the direction of the filaments is then examined in $\mathcal{S}_{100}^{\rm CDM}$ and in $\mathcal{S}_{100}^{\rm HDM}$. The angle μ_1 between the direction of the vorticity and the direction of the filament is measured along each segment of the skeleton, and μ_2 between the direction of the vorticity and the direction of the normal of the wall. The probability distribution function (PDF) of the absolute value of the cosine of these angles is shown in Figure 7.7. This PDF is normalized for $\cos\mu$ between 0 and 1. A strong correlation is measured. The signal is stronger in $\mathcal{S}_{100}^{\rm HDM}$ (because high frequencies are smoothed) but a clear signal is also detected in $\mathcal{S}_{100}^{\rm CDM}$. As a check, the alignment between vorticity and shuffled segment directions is then measured: no signal is detected.
In the filaments I find an excess of probability of 20 per cent to have $|\cos\mu_1|$ in $[0.5, 1]$ (that is $0 \leq \mu_1 \leq 60^{\rm o}$) relative to random orientations. In the walls, I find an excess of probability of 45 per cent to have $|\cos\mu_2|$ in $[0, 0.5]$ (that is $60^{\rm o} \leq \mu_2 \leq 90^{\rm o}$) relative to random orientations, which means a strong signal for the vorticity to be aligned with the filament, and perpendicular to the normal of the surrounding wall. In conclusion, in the neighbourhood of filaments, vorticity is preferentially aligned with the filament axis and perpendicular to the normal of walls. In other words, vorticity tends to be perpendicular to the axis along which material is collapsing fastest. This is fully consistent with the conjectures of Codis et al. (2012) (which relied on merger-tree to trace back the origin of the progenitors of forming dark haloes).

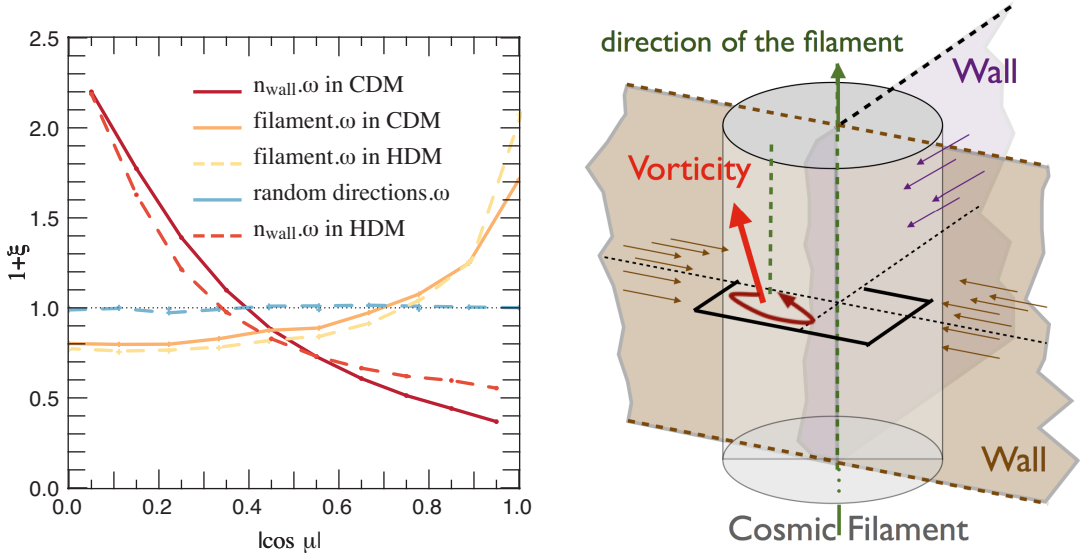

Figure 7.7.: *Left:* The PDF of $\cos\mu$, the cosine of the angle between the vorticity and the direction of the filament (orange) and the angle between the vorticity and the normal of wall (red) measured in the simulations $\mathcal{S}_{100}^{\mathrm{CDM}}$ (solid) and $\mathcal{S}_{100}^{\mathrm{HDM}}$ (dashed). The black dotted line corresponds to zero excess probability for reference. The large-scale total vorticity is preferentially aligned with filament axis. This description is in agreement with Figure 7.6, which shows the statistical alignment with the tidal tensor. This alignment is illustrated in the *right* panel.

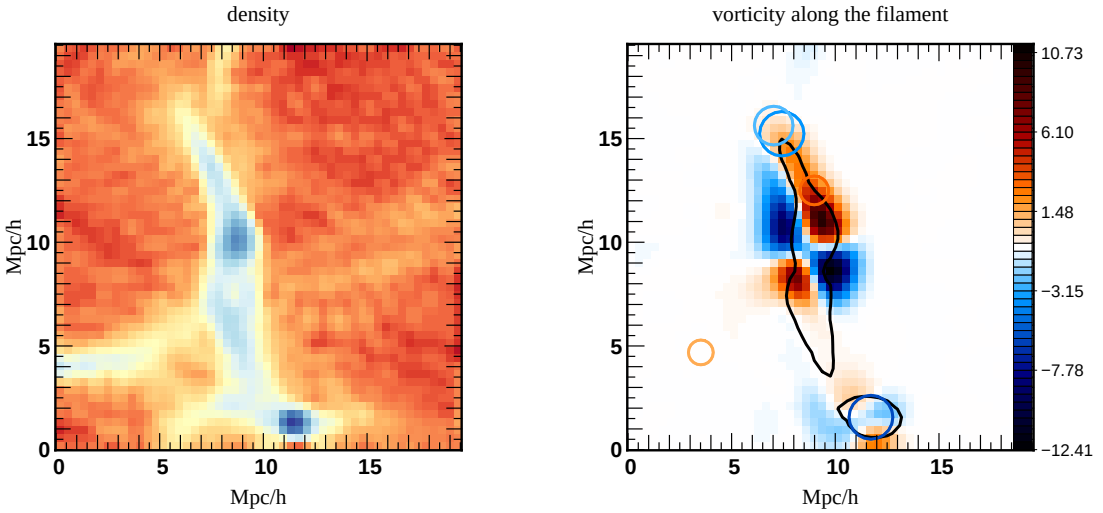

Figure 7.8.: Geometry/kinematics of a typical multi-flow region across a filament. *Left:* density map of a section perpendicular to a given filament in logarithmic scale. *Right*: projected vorticity along the filament within that section (towards observer in red and away from the observer in blue) in units of $h\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$ on which is plotted in dark a contour of the density. Circles are haloes with their corresponding virial radius. Circles colour matches to the values of $\cos\theta$ between the halo spins and the normal of the section, positively oriented towards the observer. $\mathcal{S}_{100}^{\mathrm{HDM}}$ is used here, and for this figure only, vorticity is computed after smoothing the velocity field with a Gaussian filter of $1.6\,h^{-1}\,\mathrm{Mpc}$.

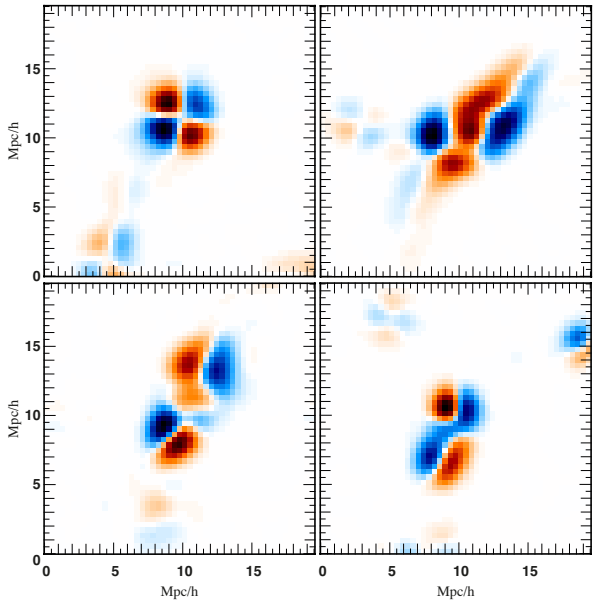

Figure 7.9.: Different kinds of vorticity cross-sections.

7.2.3 Geometry of the multi-flow region

Section 7.2.2 showed that vorticity tends to be aligned with the filamentary features of the cosmic web. I am now naturally led to focus on the structure of high-vorticity regions. The kinematics of the cross-sections of the filaments is therefore examined, by cutting the simulation with a plane perpendicular to the direction of each filament. I represent in this plane the projected vorticity along the filament. Results are shown in Figure 7.8 and Figure 7.9. These figures illustrate that vorticity is null outside the multi-flow region, and it is therefore confined to filaments (and to walls but the signal is weaker) which is consistent with the assumption that cosmic flows are irrotational before shell-crossing. In addition, from Figure 7.8 one can see that high-vorticity resides in the low-density regions of filaments: vorticity is mainly located at the edge of the multi-flow region *on* the caustic. Vorticity is in fact typically null *at* the peak of density. I will measure it formally later (see Figure 7.15).
Furthermore, I estimated the number and the radius of quadrants per vorticity cross-sections. To do so, I followed every filaments in the simulation $\mathcal{S}_{100}^{\mathrm{HDM}}$ and I regularly cut them with a plane perpendicular to their main direction. The number of multi-flow regions is given by the number of regions of positive and negative projected vorticity along this direction (with a given threshold), counted in a small window centred on the filament. To obtain the size of the regions with a given polarity, the area where the absolute projected vorticity along the normal is greater than 10 per cent of the maximum vorticity is measured, and this area is divided by the number of quadrants. Assuming that these regions are quarter of discs, it yields the corresponding radius. These histograms are displayed on Figure 7.10. From these plots, we see that the cross-sections of the filaments are partitioned into typically quadripolar multi-flow regions where the vorticity is symmetric with respect to the centre of the (density) caustic such that the global vorticity within that caustic is null (as expected). The typical size of each quadrant is of the order of a smoothing scale. The radius of the quadrant is also very important to explain the mass transition of the spin alignment. I will explain it in more detail in Section 7.3.

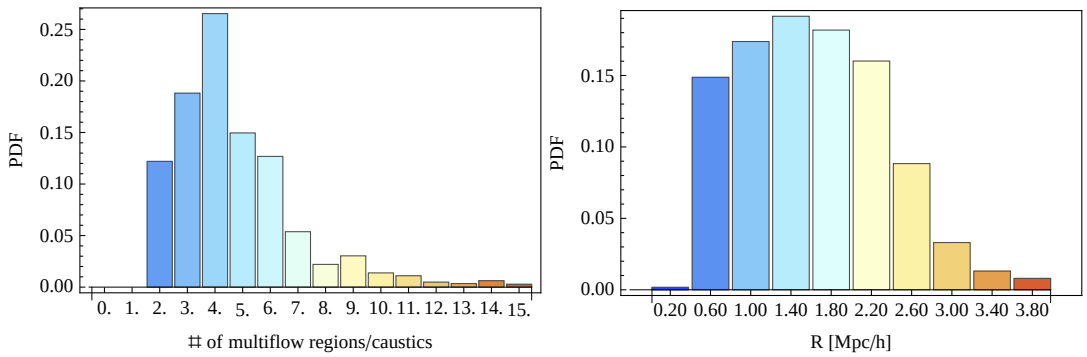

Figure 7.10.: *Top:* normalized histogram of the number of multi-flow regions with different polarity around a filament measured in the simulation $\mathcal{S}_{100}^{\mathrm{HDM}}$. The mean corresponds to $\langle n_{\mathrm{multiflow}} \rangle = 4.6$, the median is 4.25. On large scales, the multi-flow region is therefore typically quadrupolar. *Bottom:* normalized histogram of the size of a region in $\mathcal{S}_{100}^{\mathrm{HDM}}$ with a given polarity. The mean size of such region is $\langle R \rangle = 1.6\,h^{-1}\,\mathrm{Mpc}$, somewhat below the smoothing length of the initial conditions, $R_s = 2.3\,h^{-1}\,\mathrm{Mpc}$. It was checked on $\mathcal{S}_{20}^{\mathrm{HDM}}$that a similar scaling applies.

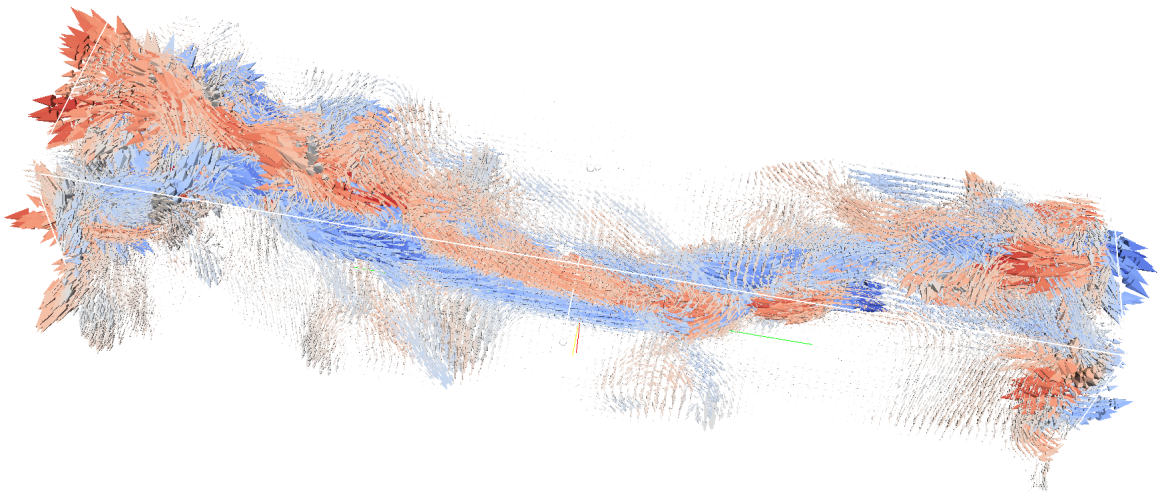

Figure 7.11.: Vorticity field in the neighbourhood of the main filament of the idealized warm dark matter simulation, $\mathcal{S}_{20}^{\mathrm{HDM}}$ colour coded through its 'z' component. The vorticity is clearly aligned with the direction of the filament, strongest within its multi-flow core region, while its polarity is twisted around it. Helicity measurements are consistent with the observed level of twisting.

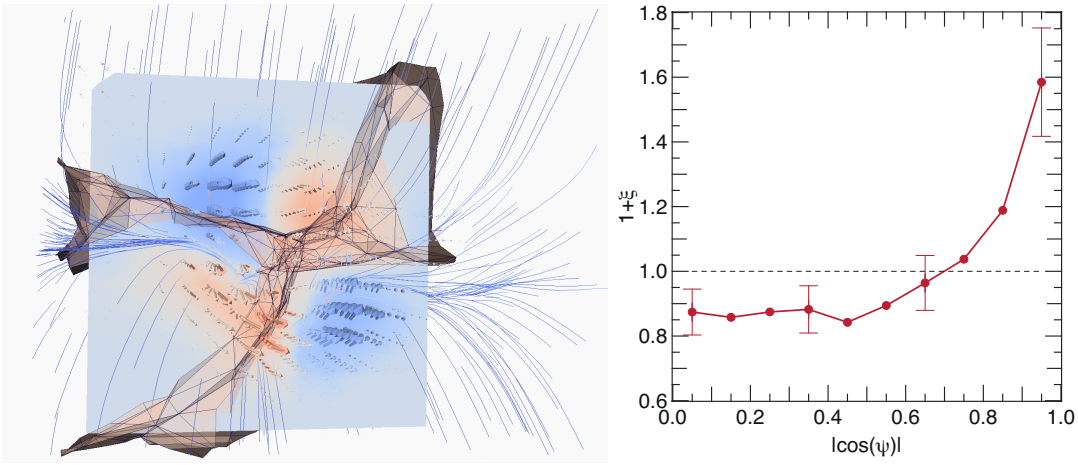

Figure 7.12.: *Left:* the cross-section of the vorticity perpendicular to the main filament shown in Figure 7.11. The colour coding in the section corresponds to the vorticity towards the observer (in blue) and away from us (in red) as shown by the corresponding arrows. The thin blue lines correspond to velocity field lines. The brown surfaces represent the local walls. The field lines converge towards the local walls and are in agreement with the vorticity field which is partitioned by these walls. *Right:* the probability distribution as a function of the cosine angle between the normal to the zero vorticity walls and the normal to the density walls, $\cos\psi$, computed on the simulation $\mathcal{S}_{100}^{\mathrm{CDM}}$. The simulation is divided into eight 50 h^{-1} Mpc sub-boxes. The plotted signal corresponds to the average of the PDFs for the eight sub-boxes. The displayed error bars are 1σ standard deviation on the mean.

7.2.4 Building up vorticity from LSS flow

Figure 7.11 displays the vorticity vector field in the neighbourhood of the main filament of the idealized HDM simulation $\mathcal{S}_{20}^{\mathrm{HDM}}$. The vorticity bundle is clearly coherent on large scales and aligned with the direction of the filament. It is strongest within its multi-flow core region while its essentially quadrupolarity is twisted around it. Figure 7.12 displays the cross-section of the vorticity perpendicular to the main filament shown in Figure 7.11. The velocity field lines (in blue) converge towards the local walls (in brown) and are visually in agreement with the vorticity field which is partitioned by these walls. This picture is qualitatively consistent with the scenario presented in Codis et al. (2012), as it shows that the filaments are fed via their embedding walls, while the geometry of the flow generates vorticity within their core. This vorticity is only significant in the local environment in which dark matter haloes form with a spin aligned with that vorticity.

The alignment between the contours of minimal vorticity and the density walls which is visually observed in Figure 7.12 (left panel) is then quantitatively examined. The probability distribution of the cosine of the angle between the zero vorticity contour and the wall within the caustic is plotted on the right panel of Figure 7.12.

I used DisPerSE (Sousbie, 2011) to define the density walls and the contours of minimal vorticity. The density walls are computed as being the "ascending two–manifolds" (see Chapter 4) of the skeleton calculated on the density field. The contours of minimal vorticity are defined as being the "descending two–manifolds" of the skeleton calculated on the norm of the vorticity field. Since the vorticity is really well defined only on the neighbourhood of caustics, a mask is applied when the walls are computed, which covers all the regions of space where the density is lower than 10 per cent of the maximum density and the vorticity lower than 10 per cent of the maximum vorticity. The results of the computation of the density walls and minimal vorticity contours are tessellations, i.e. sets of triangles. For each triangle in the minimal vorticity tessellation I find its nearest neighbours in the density tessellation. Smoothing is achieved by averaging the position of each vertex with that of its direct neighbours. A smoothing coefficient $S = N$

means that this operation is repeated N times. The cosine between the normals of both triangles is then calculated.
An excess of probability of 15 per cent is observed for $\cos\psi$ in $[0.5, 1]$ relative to random distribution, that is for the alignment of the walls with the minimal vorticity contours. This alignment increases with the smoothing of the tessellations, as expected.

7.3 Large-scale structure vortices are spinning up dark haloes

7.3.1 Correlation between vorticity and spin

The alignment of vorticity with filaments on the one hand, and previous results about alignment (or orthogonality) of the low-mass (high-mass) haloes spin with the filament on the other hand, suggests to analyse in depth the correlation between vorticity and angular momentum. The measurement is done by computing the vorticity at the positions of the haloes and the projection, $\cos\theta$, between the normalized vectors. First note that haloes typically lie within one quadrant of the vorticity within filaments and not at the intersection of these quadrants, which is why the spin/vorticity alignment is strong.
The resulting PDF of $\cos\theta$ is displayed in Figure 7.13. Here the set of simulations $\mathcal{S}_{50}^{\rm CDM}$ are used to compute error bars on the correlation between spin and vorticity. The measured correlations are noisier as only a finite number of dark haloes are found within the simulation volume [5]. I find an excess probability of 25 per cent relative to random orientations to have $\cos\theta$ in $[0.5, 1]$ for haloes with $10 \leq \log(M/{\rm M}_\odot) \leq 11$, 55 per cent for $11 \leq \log(M/{\rm M}_\odot) \leq 12$ and 165 per cent for $12 \leq \log(M/{\rm M}_\odot) \leq 13$. Note importantly that the intricate geometry of the multi-flow region (see also Figures 7.11 and 7.9) strongly suggests retrospectively that the alignment (including polarity) between the spin of dark matter haloes and the vorticity of the flow within that region cannot be coincidental.

Figure 7.2, which presented PDF of the cosine of the angle between the spin of 43 million dark haloes and the direction of the closest filament identified in the $\mathcal{S}_{2000}^{\rm CDM}$ simulation, displays an interesting feature at low mass. For the range of mass $\log M/{\rm M}_\odot \sim 11.5 - 12.5$, the actual alignment between the spin and the direction of the filament *increases* with mass, before it becomes abruptly perpendicular around $5 \times 10^{12}\,{\rm M}_\odot$. This is fully consistent with the corresponding increase in vorticity shown in Figure 7.13, and will be discussed further in the next section.

7.3.2 Progenitors of the multi-flow region

In a dark matter Lagrangian simulation, it is straightforward to identify the origin of particles within the multi-flow region. Figure 7.14 traces back in time dark matter particles ending up within a quadrant of the multi-flow region. The quadrant is fed by three flows of particles. The flow is irrotational in the initial phase of structure formation until the crossing of three flows in the vicinity of the filaments generates shear and vorticity close to the caustic.

Note that the sharp rise near the edge of the multi-flow region at the caustic is consistent with the picture proposed by Pichon and Bernardeau (1999). This is illustrated in Figure 7.15. To obtain this profile, a filament is cut in slices where each slice corresponds to a plane perpendicular to the direction of the

[5] I checked that the correlation is not dominated by the intrinsic vorticity of the haloes themselves by computing the alignment between the spin and the vorticity of the field *after extruding* the FOF haloes, which led to no significant difference in the amplitude of the correlation.

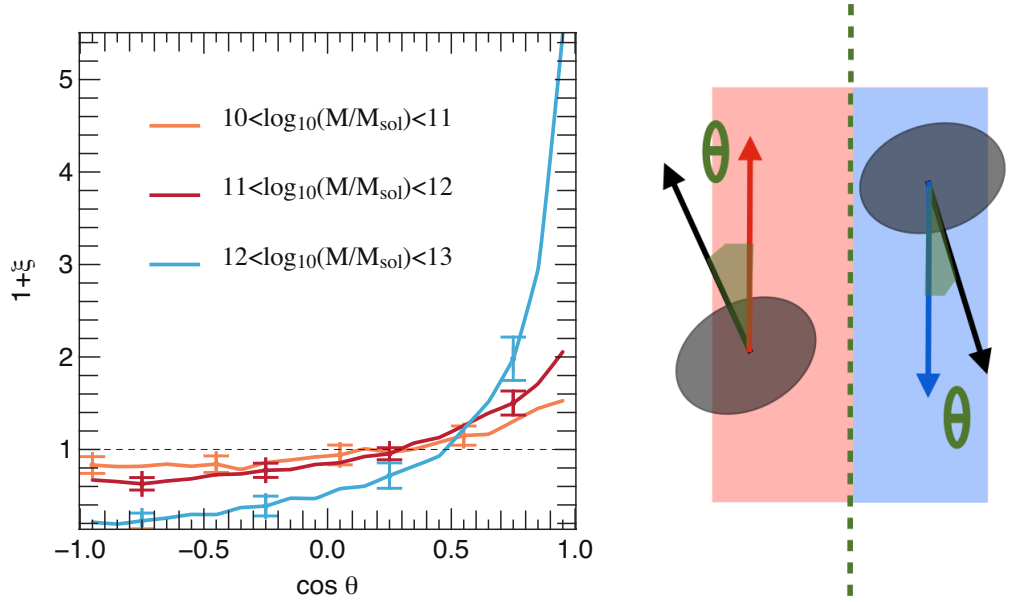

Figure 7.13.: *Left:* The PDF of the angle between the vorticity and the spin measured in 20 simulations of the $\mathcal{S}_{50}^{\mathrm{CDM}}$ set. Haloes are binned by mass as labeled. The displayed error bars are 1-σ standard deviation on the mean. The typical configuration is illustrated on the *right* panel.

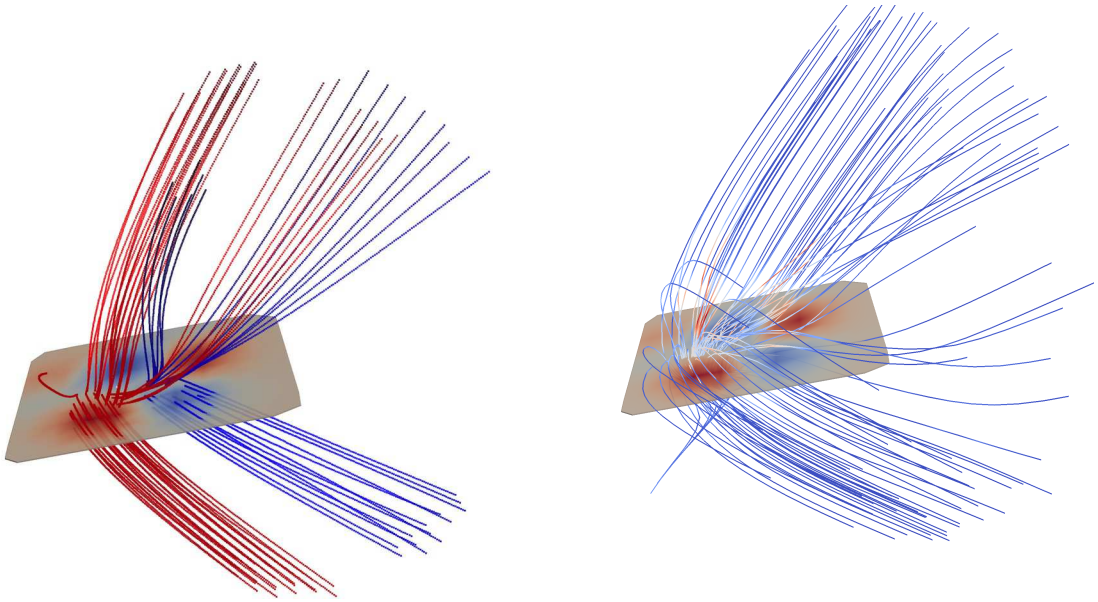

Figure 7.14.: *Left*: individual dark matter particle trajectories ending in a given quadrant of the vorticity multi-flow region. In blue are particles ending in a region of positive projected vorticity along the filament, and in red are the particles ending in the negative vorticity region. The quadrant is fed by at least three flows of particles. *Right:* The associated velocity field color-coded by the vorticity). The $\mathcal{S}_{100}^{\mathrm{HDM}}$ simulation is used for this figure.

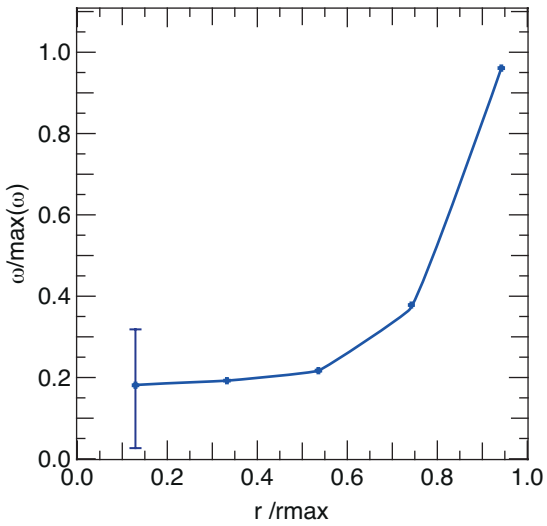

Figure 7.15: Azimuthal average of the radial profile of the vorticity. The profile is obtained by averaging on the sections of a complete filament (each section is associated to a filament segment, to which the section is perpendicular). Vorticity is clearly larger towards the caustic, and would theoretically become singular (as $1/\sqrt{1-r/r_{\rm max}}$) *at* the caustic for a Zel'dovich mapping, as shown in Pichon and Bernardeau (1999). Here the profile is convolved by shape variations from one caustic to another and by the azimuthal average. The indicative error bar was computed as the average over a larger stack.

segment. Local vorticity is measured within that plane and stacked. The amount of vorticity is greater near the caustic. These results are qualitatively consistent with the above-mentioned theoretical predictions which characterize the size and shape of the multi-flow regions after first shell-crossing, and estimate their vorticity content as a function of cosmic time. It is important to understand the mass dependence in the alignment of the angular momentum.

In short, looking in detail at the set of (Lagrangian-) smoothed simulations allows us to conclude that streaming motion of dark matter *away* from minima and wall-saddle points of the field, and *along* the walls of the density field is responsible for generating the multi-flow region in which vorticity arises. In turn, this vorticity defines the environment in which lower mass haloes collapse. Such haloes inherit their spin from this environment, as quantified by Figure 7.13.

7.3.3 Mass transition for spin–filament alignment

Up to now, I have not considered the mass of the forming halo within the multi-flow region. The assumption has been that the Lagrangian extension of the progenitor of the dark halo was small compared to the antecedents of a given vorticity quadrant, so that the collapse occurs within a quadrant of a given polarization, and leads to the formation of haloes with a spin parallel to that vorticity. For more massive objets (of the order of the transition mass), we can anticipate that their progenitor patch overlaps future vorticity quadrants of opposite polarity, hence that they will mostly cancel the component of their vorticity aligned with the filament as they form. The above-mentioned observed transition mass between aligned and anti-aligned spins relative to filaments would then typically correspond to the mass associated with the width of the quadrant of each caustics. In fact, as argued in Pichon and Bernardeau (1999, fig 7) and shown on Figure 7.15, the vorticity within the multi-flow region is mostly distributed near the caustic, on the outer edge of the multi-flow region. It is therefore expected that, as the size of the collapsing halo increases, but remains *below* that of the quadrant, its vorticity should increase (as it collects more and more coherent rotating flow as secondary inflow), as shown in Figure 7.13. As it reaches sizes *above* that of the quadrant, its spin should start to diminish significantly as it advects flows with opposite vorticity.

Let us return to Figure 7.2. For the range of mass $\log M/\mathrm{M}_\odot \approx 11.4-12.1$, the alignment between the spin and the direction of the filament *increases* with mass, peaking at $M_{max} \approx 10^{12}\,\mathrm{M}_\odot$, before it rapidly decreases and changes to preferably perpendicular one for $\log M > \log M_{\rm crit} \approx 12.7$, i.e $M_{\rm crit} \approx 5\times10^{12}\,\mathrm{M}_\odot$. This is fully consistent with the corresponding increase in vorticity shown in Figure 7.13.

The origin of these characteristic masses can be roughly understood by conjecturing that the highest alignment occurs for haloes which zone of advection is of the size of vortices in the caustic regions. The measured caustic structure depends on the chosen smoothing scale. A recently formed filament corresponds to the vortex that shows a basic four quadrant structure and, following Pichon and Bernardeau (1999), which has vorticity close to the Hubble value H. From our simulations, the typical Lagrangian radius of such vortex is $\approx 1.5\,h^{-1}\,\mathrm{Mpc}$, which if taken as the top-hat scale gives a mass estimate $\mathrm{M_{max}} \approx 1.5 \times 10^{12}\,\mathrm{M_\odot}$ for the mass of haloes with maximum spin/filament alignment. The transition to misalignment will happen at $M_{\rm crit} \approx 8 \times M_{\rm max}$ where the whole width of the filament is encompassed. Of course the quantitative accuracy of such argument should not be over-emphasized. For instance, if we took the Lagrangian radius of the vortex to be $1.3\,h^{-1}\,\mathrm{Mpc}$, I would get $M_{\rm max} = 10^{12}\,\mathrm{M_\odot}$, which would fit the transition of Figure 7.2 even better.

Let me qualitatively illustrate with a simple toy model this mass transition for the spin-filament alignment. In this toy problem, I consider an isolated infinite filament aligned along $\mathbf{e}_\mathrm{z}$. I define the corresponding idealized vorticity field as

$$\mathbf{\Omega}(\mathbf{r},\theta) = \mathrm{C}\epsilon \sin(2\theta) \frac{1}{(\epsilon^2 + (r-R)^2)}\,\mathbf{e}_\mathrm{z}\,, \tag{7.9}$$

with C a constant, R the radius of the caustic and ϵ a small number. The vorticity thus defined is largest along the caustic, point reflection symmetric and tends rapidly to 0 outside the caustic. Should ϵ tend to 0, vorticity would become singular on the edge of the caustic ($r \to R$). The map of the vorticity is displayed in Figure 7.3 (right panel).
By application of the Helmoltz-Hodge theorem[6], I find that the curl component of the velocity field consistent with that vorticity (i.e. such that $\mathbf{\Omega} = \nabla \times \mathbf{v}$) obeys

$$\mathbf{v}(\mathbf{r},\theta) = \frac{1}{4\pi}\int_{\mathcal{V}} \nabla \times \mathbf{\Omega}(\mathbf{r}',\theta') \frac{1}{|\mathbf{r}-\mathbf{r}'|}\mathrm{d}\mathcal{V}\,. \tag{7.10}$$

where $\mathcal{V}$ is the volume. I assume here that the shear part of the curl free component of the velocity flow is smaller on scales comparable to the halo.
I now consider a proto-spherical halo of radius r_h embedded in one of the four quadrants of the caustic, centred on $C_\mathrm{h}(x_\mathrm{h}, y_\mathrm{h})$ with $x_\mathrm{h}^2 + y_\mathrm{h}^2 \leq R$. From equation (7.10) I can simply compute its angular momentum $\mathcal{J}(r_\mathrm{h}, x_\mathrm{h}, y_\mathrm{h})$, and look at the variation of $\mathcal{J}$ as a function of its position at fixed radius, or as a function of its radius at fixed position. Figure 7.16 shows the magnitude of the angular momentum along the z-axis for a halo centred on $C_\mathrm{h}(0.5\ \mathrm{R}/\sqrt{2}, 0.5\ \mathrm{R}/\sqrt{2})$ as a function of the radius. We observe that the alignment increases until the size of the halo encompasses the whole quadrant. At a given radius, the position of the halo which maximizes the angular momentum is the one for which the edge of the halo coincides with the edge of the caustic, since the vorticity peaks close to the caustic.
This simple toy model illustrates the increase of the alignment with mass up to a certain mass, and then the decrease for massive haloes whose the size encompasses the size of the quadrant.

7.3.4 Variation of the signal with scale and redshift

Given the characteristics of ΛCDM hierarchical clustering, one can anticipate that this process of the alignment from vorticity occurs on several nested scales at various epochs – and arguably on various scales at the same epoch. The scenario I propose for the origin of vorticity and spin alignment is, like the signal itself, relative to the linear scale involved in defining the filaments and therefore multi-scale. The two sets of simulations, $\mathcal{S}^{\Lambda\mathrm{CDM}}$ and $\mathcal{S}^{\Lambda\mathrm{HDM}}$, allowed me to probe different scales of the vorticity field. I will

[6] The Helmoltz-Hodge theorem states that a sufficiently smooth vector field can be decomposed in two parts: an irrotational part expressed by the gradient of a scalar function, and a rotational part expressed by the curl of a vector function. In this toy-model, the vorticity field is assumed sufficiently smooth for this derivation.

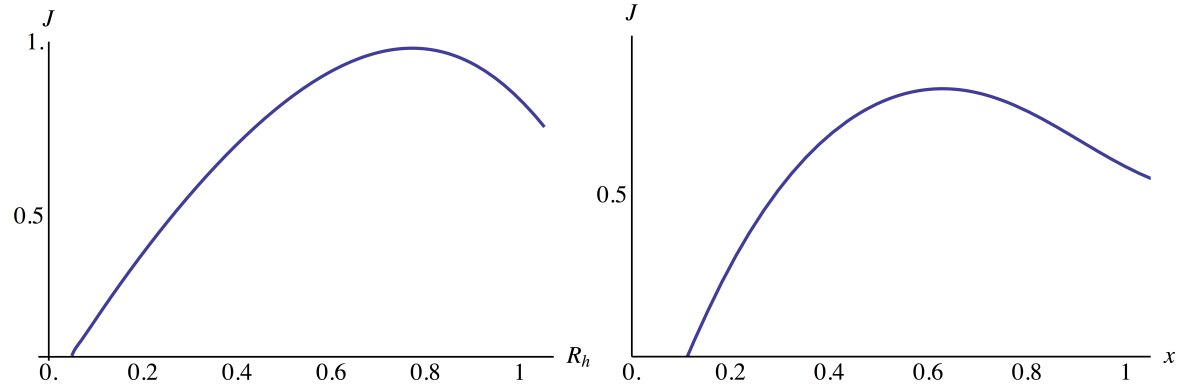

Figure 7.16.: Magnitude of the angular momentum along $\mathbf{e_z}$ for a halo embedded in one of the four quadrants. The *left* panel shows it varies as a function of the radius of the halo, the position of this latter being fixed $(x_{\rm h},y_{\rm h})=(0.5/\sqrt{2}\ ,\ 0.5/\sqrt{2})$. The alignment of the angular momentum of the halo with the vorticity increases until the halo size becomes comparable to that of the vorticity quadrant. I study then the magnitude of the angular momentum along $\mathbf{e_z}$ at fixed radius ($R_h = 0.3$) as a function of the position x along the diagonal (*right*). In this case, the alignment increases up to the point where the halo boundary coincides with the vorticity caustic, with the halo still being fully contained within the vorticity quadrant $(x_{\rm h},y_{\rm h})\sim(0.7/\sqrt{2},0.7/\sqrt{2})$.

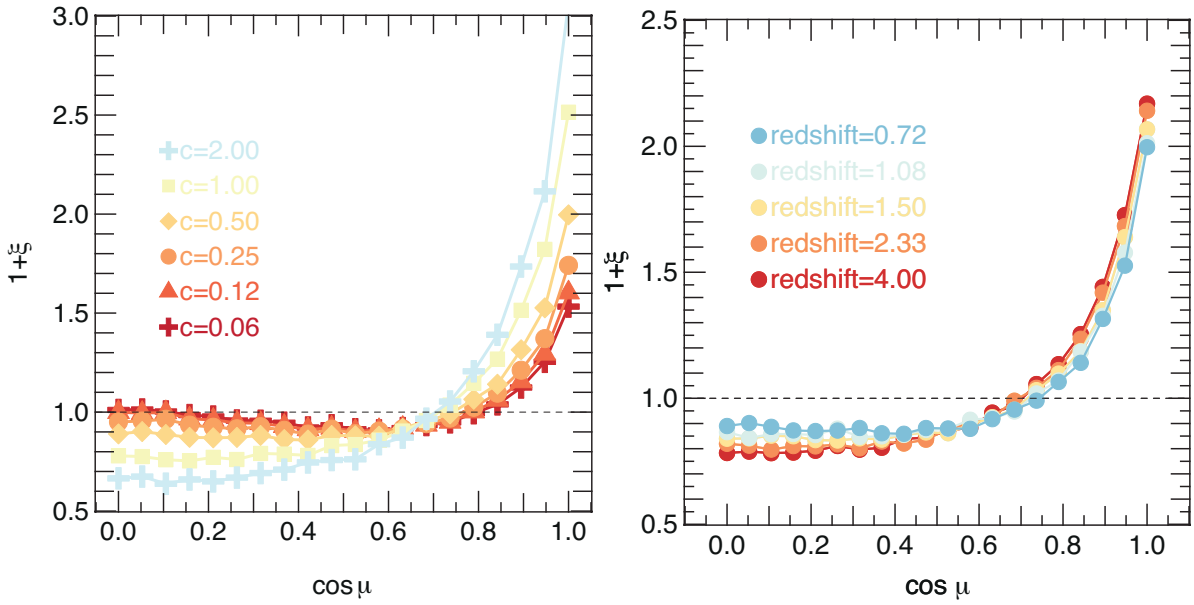

Figure 7.17.: *Left*: The probability distribution of the cosine of the angle between the vorticity and the direction of the filament measured in $\mathcal{S}_{20}^{\rm cool}(z = 0.7)$ for different persistence threshold. The level of persistence of the main text corresponds to $c = 0.5$. *Right*: The probability distribution of the cosine of the angle between the vorticity and the direction of the skeleton, measured in $\mathcal{S}_{20}^{\rm cool}(0.7)$ for various redshifts as labeled. The amplitude of the correlation decreases with cosmic time.

Table 7.2.: The median and mean cosine values for the set of studied alignments. In parenthesis are the expected values for random distributions.

Definition	Name	Mean	Median
Alignment between vorticity and cosmic web			
DM: vorticity/filaments		0.58 (0.5)	0.62
Hydro: vorticity/filaments	$\lvert\cos\mu\rvert$	0.58 (0.5)	0.63
DM: vorticity/walls		0.34 (0.5)	0.27
Alignment between vorticity and haloes spin			
$10 \leq \log(M/\mathrm{M}_\odot) \leq 11$		0.09 (0.0)	0.14
$10 \leq \log(M/\mathrm{M}_\odot) \leq 12$	$\cos\theta$	0.19 (0.0)	0.29
$12 \leq \log(M/\mathrm{M}_\odot) \leq 13$		0.53 (0.0)	0.72
Alignment between density walls and 0-vorticity walls			
	$\lvert\cos\psi\rvert$	0.54 (0.5)	0.56
Alignment between vorticity and tidal tensor eigenvectors			
vorticity / e1		0.62 (0.5)	0.69
vorticity / e2	$\lvert\cos\gamma\rvert$	0.48 (0.5)	0.47
vorticity / e3		0.31 (0.5)	0.23

now briefly explore the effects of probing different scales of the LSS via the skeleton level of persistence. Figure 7.17 shows the excess alignment probability as a function of the cosine of the angle between the vorticity and the filaments as a function of the persistence level for a range of values. The alignment is strongest with the largest scale filamentary structure corresponding to the least dynamically evolved features of the field.
Figure 7.17 displays the probability distribution of the cosine of the angle between the vorticity and the direction of the skeleton for a range of redshifts. The correlation between the direction of the filament and the vorticity is significant. As expected, this correlation decreases with cosmic time (at a fixed smoothing scale). All the results of statistical alignments are reported in Table 7.2.

7.4 Theory of the spin-flip: unifying Eulerian and Lagrangian views

The findings detailed above emphasise that the shell-crossing occurring during the later formation of the filament generates vorticity predominantly aligned with that filament. For low-mass haloes which are forming in large-scale filamentary regions, the secondary infall arrives along this marked preferred directions and is typically multi-flow and vorticity rich. That is why this infall contributes extra spin-up along the filament direction. Hence the global geometry of the inflow is consistent with a spin acquisition for haloes induced by the large-scale dynamics within the cosmic web, and in particular within its multi flow vortices.
This Eulerian scenario is in fact not fundamentally different of the Lagrangian tidal torque theory if the latter is augmented so as to account for the larger scale anisotropic environment of walls and filaments responsible for secondary infall. Codis et al. (2015) showed that if the critical condition that the filament is embedded into a given wall is added, the observed spin (and importantly its polarity) will correlate with the polarity of the vorticity quadrant the halo ends up. I now describe these results.

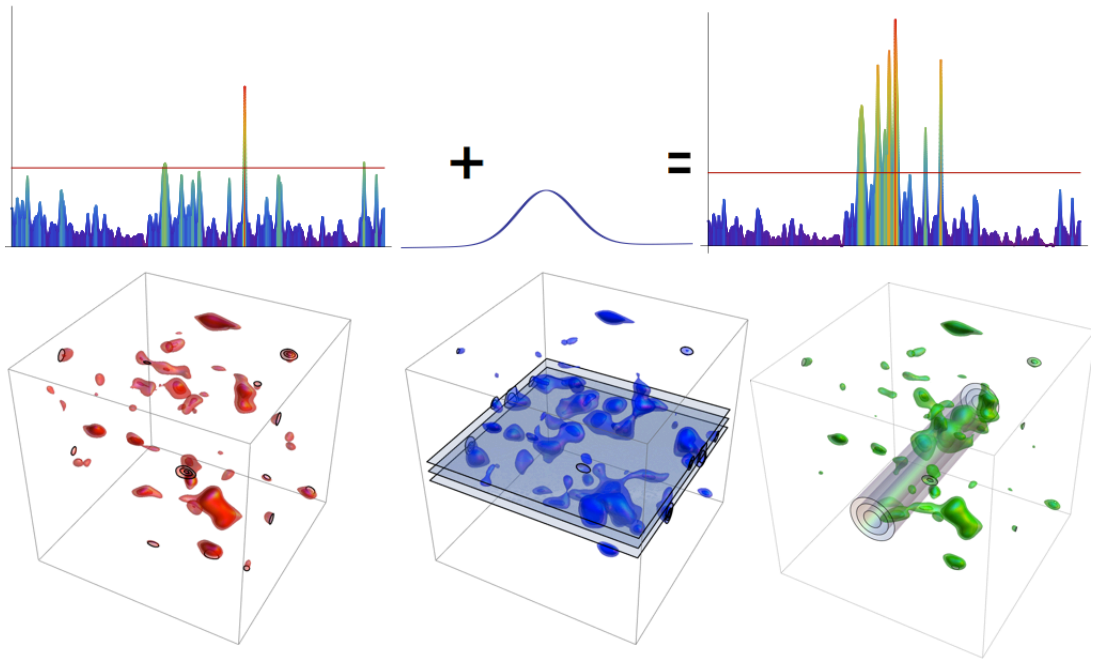

Figure 7.18.: *Top*: the long-wavelength modes induce a boost in density. *Bottom*: the anisotropic Kaiser bias. *Left*: isocontours of a Gaussian random field without underlying constraints (in red). *Middle*: with a wall-like low frequency boost added (in blue). *Right*: with a filament-like low frequency boost (in green). Which such boost, more proto-halos pass the threshold of collapse earlier, so that the cosmic web biases the mass function in its vicinity.

7.4.1 The Lagrangian view of spin-flip with the *constrained* tidal torque theory

Here for completeness I report on the results presented in Codis et al. (2015). In this paper, the authors revisited the tidal torque theory while taking into account the three-dimensional anisotropic geometry of the environment in which haloes form, due to the presence of filaments embedded in walls. In this way, they were able to predict both primordial and secondary spin acquistion.

Angular momentum in the vicinity of the saddle point/node – Haloes form preferentially in the vicinity of filaments, where the long-wavelength modes (large-scale fluctuations leading to walls and filaments) induce an anisotropic boost in density, allowing the proto-haloes to decouple from the expansion of the Universe and collapse (see Figure 7.18).

Solving the equation of the spin acquisition in the context of the tidal torque theory (Equation 7.1 in Section 7.1.1) implies knowledges of the inertia tensor I_{jl} of the proto-halo, hence the boundary of the Lagrangian area of the halo. The sensible approach is thus to consider that proto-haloes form around (elliptical) peaks in the initial density. In this context, Codis et al. (2015) shows that the Hessian of the density field H_{jl} can be used to approximate the inertia tensor. In this description, equation 7.1 is then approximated by:

$$s_i = \sum_{j,k,l} a^2(t)\dot{D}_+(t)\epsilon_{ijk}H_{jl}T_{lk}\,. \tag{7.11}$$

The motivation for this approximation is that Equation 7.11 only involves quadratic products of derivatives of the potential field, hence its expectation can be computed exactly for Gaussian random fields.
The theory developed by Codis et al. (2015) relies subsequently on the fact that the Hessian and the tidal field probes different scales: the Hessian probes its direct neighbourood while the tidal field a larger scale.

Let us now consider a low-mass proto-halo, embedded in a wall in the vicinity of a filament. In the plane of the filament-type saddle point, perpendicular to the filament, the dominant closer wall will re-orient more the Hessian than the tidal field, while the tidal field will feel the denser but more distant filament. The misalignment will induce an angular momentum aligned with that filament. One very interesting outcome is that the distribution of the longitudinal (along the filament) component of the angular momentum at the saddle-point is typically quadri-polar (and higher in boundaries of the cross-section), as predicted by the Eulerian view (quadri-polar vorticity cross-section). Figure 7.19 displays the intuitive representation of this process.
Now consider a more massive proto-halo which forms in a filament in the vicinity of a node. In the plane containing the filament, the process is the same: the Hessian is more tilted by the filament than the tidal field, which feel the denser but further-away node. The result of this misalignment is an angular momentum in the azimuthal direction (perpendicular to the filament). Figure 7.20 displays both configurations.
Hence the tidal torque theory, if constrained by the anisotropic geometry of the environment, predicts that low-mass haloes, forming in the vicinity of filament-type saddle point, would have a primoridal spin aligned with that filament, while more massive haloes, forming in the vicinity of node, would have a primordial spin in the azimuthal direction. The theory predicts also a quadripolar geometry of the angular momentum distribution, in perfect agreement with the quadripolar vorticity caustic. In addition, the transition mass is related to the size of the quadrants and can be estimated from first principle (typically 1/8th of the mass of the nonlinearity in relation of the size of the quadrant, Codis et al., 2015). The Lagrangian progenitor of a low-mass halo will cover a small Lagrangian patch and will fall within one quadrant, with a spin aligned with the filament. But the larger the extent of the proto-halo, the more sensitive it will be to the tide of the other quadrant, and/or to the azimuthal tide due to the node, inducing a misalignment. This description coincides with the toy-model presented above (see Figure 7.16).

Lagrangian dynamics of spin-flip – Let us consider again a proto-halo in the vicinity of the filament-type saddle point, *i.e.* a point proxy for a flattened filament embedded in a wall. As explained previously, it will acquire a spin aligned with that filament. Let us consider now a larger shell, which contains the Lagrangian extent of the proto-halo. This larger shell contains the material for secondary infall and upcoming mergers. The larger the shell the more sensitive it will be to the azimuthal tide due to the node. At some point and when the halo is moving towards node embedded in the large-scale flow, these mergers will reorient the angular momentum of the halo in a direction perpendicular to the filament. This Lagrangian view should have its Eulerian counterpart relying on vorticity. Let us explore below what could be the corresponding theory for the vorticity.

7.4.2 The Eulerian view of spin-flip along the cosmic web

Extending the vorticity study – The alignment of angular momentum with vorticity is the Eulerian counterpart of the above presented Lagrangian theory. The Lagrangian theory clearly suggests that the frame (a cross-section perpendicular to the filament) in which I studied the vorticity is not sufficient to capture the spin-flip. Up to know I focused only on a second shell-crossing, corresponding to the formation of filaments, and I ignored subsequent shell-crossings.
I expect that the vorticity flip (hence the spin-flip) should be measurable when considering regions of the cosmic web having shell-crossed three times (towards nodes). Following Hahn et al. (2015), it is indeed possible to easily identify such regions and compute the corresponding vorticity and its orientation with respect to the neighbouring filament. The correct framework for this study would be thus the longitudinal cross-section, in a plane which contain the node. In turn, this measurement would allow the unification of the Lagrangian and Eulerian view, and the development of a complete model for the spin acquisition of the haloes within the cosmic web.

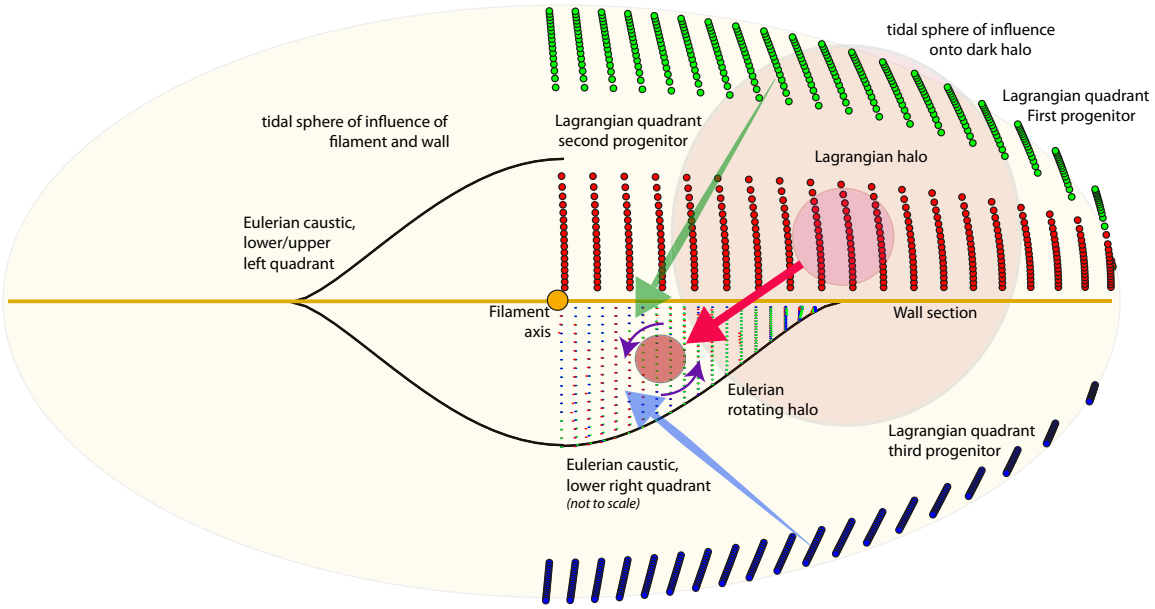

Figure 7.19.: Sketch of the dynamics of a low mass halo formation and spin-up within a wall near a filament, which are perpendicular to the plane of the image (*in yellow*). The tidal sphere of influence of this structure is represented by the pale yellow ellipse. The three bundles of large dots (*in green, red, and blue*) represent Lagrangian points (at high redshift) which image, after shell-crossing, will end up sampling regularly the lower right quadrant of the Eulerian multi-flow region; the three progenitor bundles are computed here in the Zel'dovich approximation (see Pichon and Bernardeau, 1999, for details; note that this Eulerian quadrant is not up to scale). Each pair of dots (one large, one small) represents the same dark matter particle in the initial condition and final condition. In black, three quarter of the Eulerian caustic. In light, resp. dark *pink*, the locus of the Lagrangian and Eulerian position of the halo, which has moved by a distance displayed by the *red* arrow, and spun up following the purple arrows while entering the quadrant. The *blue* and *green* arrows represent the path of fly-by dark matter particles originating from the other two bundles, which will contribute to torquing up the halo (following Codis et al., 2012). Given the geometry of the flow imposed by the wall and swirling filament, the spin of the dark matter halo will be parallel to the direction of the filament (Codis et al., 2015) *and* to the vorticity in that quadrant. In the language of TTT, the tidal field imposed on to the Lagrangian patch of the halo (*very light pink*, corresponding to secondary infall) should be evaluated subject to the constraint that the halo will *move* into the *anisotropic multi-flow* region (each emphasis imposing a constraint of its own); these constraints will in turn impose that the corresponding spin-up will be aligned with the vortex. Given the geometry of the caustic, this is fully equivalent to evaluating TTT subject to the presence of the filament and the wall.

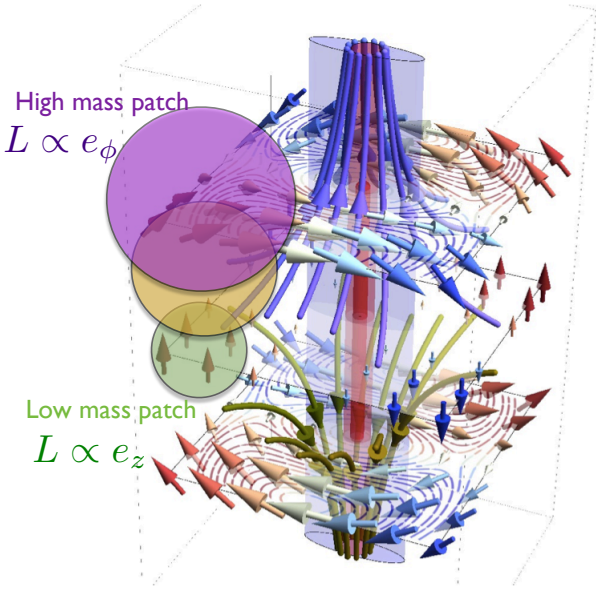

Figure 7.20: The velocity and spin flow near a vertical filament (in red) embedded in a (purple) wall. The purple and green flow lines trace the (Lagrangian) three-dimensional velocities (upwards and downwards, respectively). The red and blue arrows show the spin three-dimensional distribution, while the three horizontal cross-sections show spin flow lines in the corresponding plane. Note that the spin is along e_z in the mid-plane and along e_ϕ away from it, and that it rotates in opposite direction above and below the mid-plane. Figure from Codis et al. (2015).

The spin-flip as traced by mergers – In fact, the spin-flip of structures flowing along the cosmic web should be best traced by mergers. As anticipated by Codis et al. (2012), mergers of two low-mass structures (galaxies or haloes) with initial angular momenta aligned with the filaments induce merger remnant with a spin perpendicular with the direction of the filament, while orbital momentum is converted into spin during that merger. This result has been investigated by Welker et al. (2014) who found that the alignment and the magnitude of the spin of a galaxy depends on its merger history, and that the correlation is stronger when quantified in terms of merger rates and strength instead of mass.
From an Eulerian point of view, galaxies are moving along the filaments towards the nodes while embedded in the large scale filamentary flows. The third shell-crossing of the large-scale flows should clearly correspond to mergers of galaxies embedded in the flows. Distinguishing between minor and major mergers, Welker et al. (2014) show in particular that smooth accretion and minor mergers (contributing to less than 1% to the mass of the galaxy) increases the spin of the galaxy (which corresponds to the vorticity-rich secondary inflow) while major mergers are responsible for the perpendicular spin-flip along the filament (and is likely driven by the third shell-crossing of the large-scale flow).

Predictions for galaxy properties – An additional prediction connected with the Lagrangian anisotropic tidal torque theory are the corresponding mass gradients (anisotropic Kaiser bias, see Figure 7.18) in the direction of the filament, and within the filament in the direction of the node: more massive haloes are more confined in the filament, as displayed in Figure 7.21 and computed in Codis et al. (2015). This gradient should be measurable in observable datasets. I will present this investigation in Chapter 8.

Finally, one could expect that the orientation of the angular momentum is correlated with the efficiency of gas accretion. Indeed, cold gas follows closely the filaments, and the efficiency of gas accretion should depends on the alignment of the galaxy with the filament, provided that violent feedback processes do not disrupt completely the geometry of the infall. Some preliminary investigations of the dependency of galaxy properties as a function of their distance to the filament will be presented in Chapter 8 (through the segregation star-forming/passive galaxies away from filaments).
A preliminary step is however to investigate if gas and galaxies follow the same scheme of vorticity and angular momentum alignment as dark matter. I now consider this scenario.

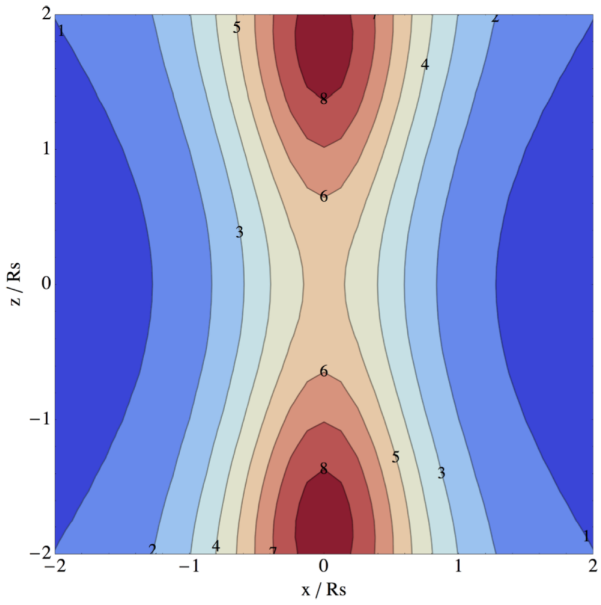

Figure 7.21: Mass distribution of the structures in a longitudinal cross-section of a filament. The saddle point is in the center and nodes are on top and bottom of the figure. The halo mass increases towards the filaments, and in the filaments towards nodes. Figure from Codis et al. (2015).

7.5 Intrinsic alignments of galaxies

The foregoing described angular momentum acquisition for dark matter only. It is interesting to connect these findings to the actual process of *galactic* alignment. Results in the SDSS from Tempel and Libeskind (2013) and Zhang et al. (2013) suggest that alignments are significant for galaxies, despite of the redistribution of the cosmic pristine gas falling on galaxies and the possible disruption of the geometry of the infall induced by feedback. These results highlights a color and morphology dependency for the alignment. Investigating the alignment of angular momentum with the filaments should allow us to determine how much of the dynamical evolution of galaxies is driven by the anisotropic environment and, in turn, to connect galaxy properties to this environment.
I first investigated the connection of the gas vorticity and the filaments. In a second step, I briefly present the alignment of galaxy angular momentum with the filaments as a function of their observable properties in Horizon-AGN. This study was presented in Dubois et al. (2014).

7.5.1 The vorticity of the gas

The vorticity of the gas is quantified as was that of the dark matter. I used three hydrodynamical simulations $\mathcal{S}_{100}^{\rm HA}$, $\mathcal{S}_{100}^{\rm HC}$ and $\mathcal{S}_{20}^{\rm cool}(0.7)$ (see also Table 7.3), carried out with the Eulerian hydrodynamic code ramses (Teyssier, 2002, see Chapter 4 for the description of this technique). The initial mesh is refined up to $\Delta x = 1.7\,$kpc. Gas is allowed to cool by H and He cooling for the cooling runs $\mathcal{S}_{100}^{\rm cool}$, and $\mathcal{S}_{20}^{\rm cool}$.

On large scales (as probed by the smoothed sets of simulations) the vorticity of gas shows the same correlations with the filaments as dark matter does. Figure 7.22 displays the probability distribution of the cosine of the angle between the vorticity and the direction of the filament for the dark matter field (in red), the adiabatic gas (in blue) cooling run (in yellow). These three simulations quantitatively show the same preference for their vorticity to be aligned with the filamentary structure. In a nutshell, differences between the adiabatic and the cooling run only appear on kpc scales, so that on large scales, the dark matter, adiabatic and cooling runs have the same velocity field structure.
As long as we consider large enough scales, the alignment pervades and is consistent with that of the dark

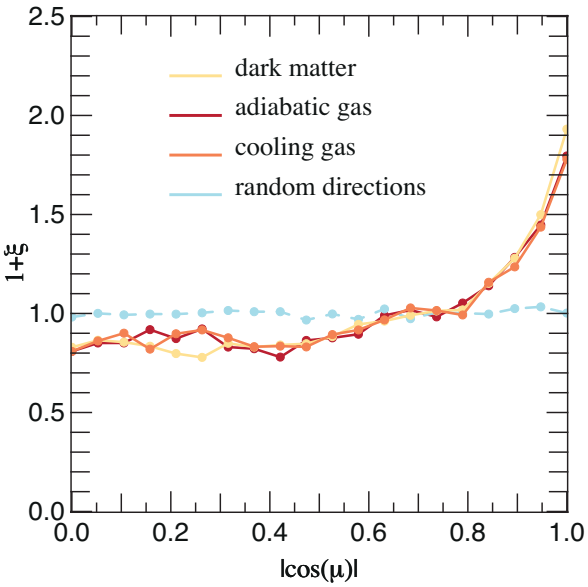

Figure 7.22: The probability distribution of the cosine of the angle between the vorticity in the smoothed dark matter and hydrodynamical simulations and the direction of the filament (solid line). The same measure is done for random directions of u (dashed blue line), plotted for the dark matter in $\mathcal{S}_{100}^{\rm HDM}$ and for the gas in $\mathcal{S}_{100}^{\rm HA}$ (adiabatic gas) and $\mathcal{S}_{100}^{\rm HC}$ (cooling run).Three profiles are very similar, which shows that large-scale modes dominate.

Name	**Type**	**Box size** h^{-1} Mpc	**Resolution**
$\mathcal{S}_{100}^{\rm HA}$	ΛHDM adiabatic	100	256^3
$\mathcal{S}_{100}^{\rm HC}$	ΛHDM cool	100	256^3
$\mathcal{S}_{20}^{\rm cool}(0.7)$	ΛCDM cool	20	1024^3

Table 7.3: The set of hydrodynamical simulations used here. The hydro runs come in two categories: adiabatic and cooling, including one high resolution run which was stopped at redshift 0.7.

matter. On smaller scales, the gas is dense enough to allow cooling to operate and re-structure the velocity flow. High-resolution cooling runs show that the small-scale structure of the velocity flow around forming galaxies does not destroy this larger scale coherence (see Figure 7.23).

7.5.2 Angular momentum alignments in Horizon-AGN

In Dubois et al. (2014), we have investigated from the Horizon-AGN simulation (described in Chapter 4) how the angular momentum alignment is preserved on galactic scales. We quantify the alignment by segregating galaxies according to their properties: colours, morphologies, mass, ages, metallicity. I briefly present below how the angular momentum alignment is correlated with some of these properties.
The spin of the galaxies is taken as the total angular momentum of the star particles in the galaxies, with respect to the center of the galaxy (particle of maximum density). The alignment of the galaxy angular momentum is first measured with respect to the vorticity of the gas and displayed in Figure 7.24. Galaxies present a strong alignment of their angular momentum with the direction of the vorticity, and this alignment respects the polarity. From this result we confirm that in despite of the complex physics of the gas at the galactic scale, the overall alignment induced by the rotational large-scale flow in the filament is preserved.

Let us now investigate if this alignment can be traced by observable galaxy properties. In Figure 7.25 is shown the alignment of the angular momentum with the direction of the filament as a function of the galaxy mass, V/σ and g-r colours. The V/σ estimator is defined as the stellar rotation over the dispersion, measured from the three-dimensional distribution of velocities. It is a proxy for galaxy morphology. Low values correspond to elliptical galaxies pressure-supported by random stellar motion while disc galaxies

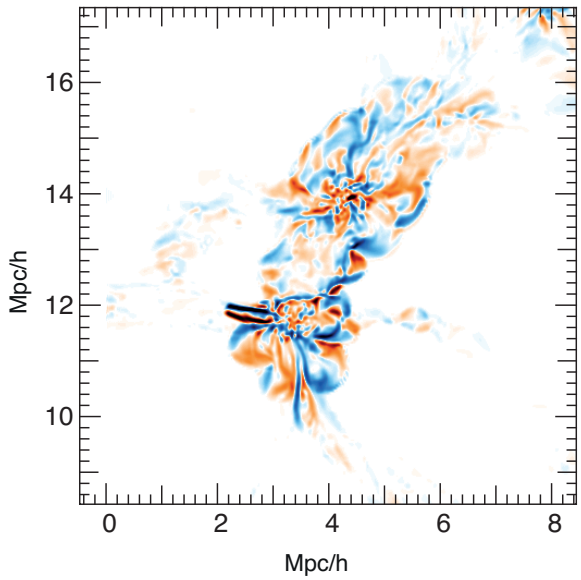

Figure 7.23: The vorticity projection along the normal of the slice in a high-resolution cooling run. The large-scale coherence is preserved, despite the small-scale structure of the velocity flow.

have high V/σ values. g and r colours are estimated through SDSS filters.

More massive galaxies tend to have their spin preferentially perpendicular to their filament and less massive ones have their spin preferentially parallel, which is fully consistent with the alignment of the galaxy spin with the vorticity. The stellar mass transition is $\sim M_{\rm tr,s} = 3\times10^{10}{\rm M}_\odot$, which is roughly in agreement with the mass transition for dark matter haloes found above.
Interestingly, this transition is traceable with the morphological V/σ estimator as well as with the observable colours. Massive, dispersion-dominated, red, galaxies tend to have a spin perpendicular or randomly oriented with the direction of their filament. Low-mass, centrifugally supported, blue, galaxies tend to align with the direction of their closest filament. However the misalignment signal for more massive haloes is somehow weaker than found in dark matter only simulations. One could anticipate that the reason for this effect is that angular momentum is redistributed within the inner region of the dark haloes in the vicinity of the galaxy (*e.g.* Kimm et al., 2011a), possibly inducing a misalignment of the galaxy spin with the halo spin, while for low-mass galaxies the direct fuelling through cold flows penetrating deep inside the haloes connect their angular momentum more tightly with the cosmic filaments (*e.g.* Pichon et al., 2011). In addition, strong stellar and AGN feedback could play a role in disrupting the geometry of the infall and thus the angular momentum alignment (Dubois et al., 2012b, 2016). Measuring precisely the magnitude of these intrinsic alignments is essential as it has potentially consequences for weak gravitational lensing studies. They can be a source of contaminations, for instance in the measure of cosmological constraints on dark energy (*e.g.* Chisari et al., 2016, see Appendix B).

7.6 Summary

The *Lagrangian* and *Eulerian* view are two complementary descriptions of the growth of structures. In the Lagrangian view, the angular momentum alignment of low-mass haloes with the filament and the perpendicular alignment for the more massive one can be predicted from first principle, with the help of *constrained anisotropic* tidal torque theory (Codis et al., 2015). The theory can also explain the spin-flip of the haloes along the cosmic web.
The Eulerian counterpart of this description relies on the vorticity of the large scale flow to quantify the spin-flip. The two formalisms are connected via non-linear clustering in simulations, or through the Zeldovich mapping in theory. They lead to consistent conclusions:

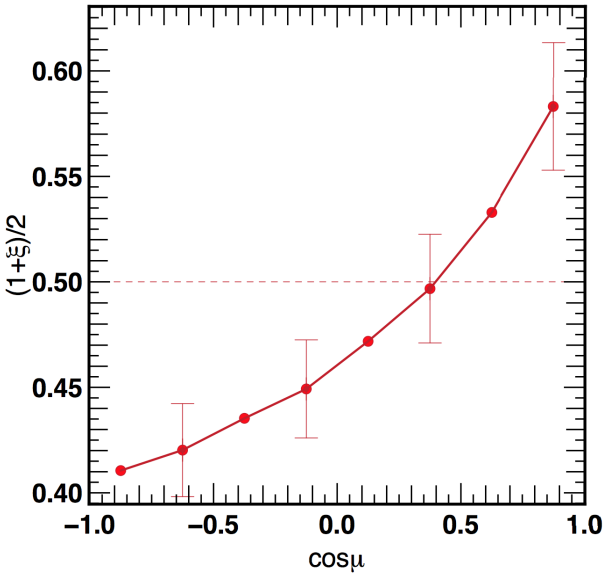

Figure 7.24: Excess probability of the cosine of the angle μ between the vorticity of the gas and the direction of the spin of the galaxies at $z = 1.83$. Figure from Dubois et al. (2014)

- The vorticity in large-scale structures is confined to, and aligned with its filaments with an excess of probability of 20 per cent relative to random orientations. It is perpendicular to the normal of the dominant walls at a similar level;

- At these scales, the cross-sections of these filaments are typically partitioned into quadripolar caustics, with opposite vorticity parallel to their filament, arising from multiple flows originating from neighbouring walls, as would secondary shell-crossing along these walls imply. The radial vorticity profile within the multi-flow region displays a sharp rise near the caustic. This quadripolar geometry is also predicted by the Lagrangian anisotropic tidal torque theory;

- The spins of embedded haloes within these filaments are consistently aligned with the vorticity of their host vorticity quadrant. The progenitor of lighter haloes within the multi-flow region can be traced back to three flows or more originating from the neighbouring walls, and form within the filament;

- The mass transition for spin–filament alignment is set by the size of sub-caustics with a given polarization. The alignment is strongest for Lagrangian patch commensurable with the sub-caustic as vorticity is strongest on the edge of the multi-flow region. Once the collapsed halo has a size larger than any such sub-caustic, it cancels out most of the vorticity within the caustics;

- this Eulerian description of the angular momentum acquisition is in good agreement with its Lagrangian counterpart, in the context of the constrained anisotropic tidal torque theory. It also predicts a perpendicular alignment of the angular momentum for more massive haloes. This will be measured in a future work from the vorticity in regions of the cosmic web which have shell-crossed three times;

- This correlation, found in dark matter simulation, is still observable on galactic scales, although it is somehow weaker.

In the previous Chapter, I have underlined the importance of following the mass growth to understand galaxy properties. Theoretical results from this Chapter suggest that the large-scale environment has a direct impact on galactic spin. This motivates once more studying galaxy properties as a function of their anisotropic environment. This is investigated in the next Chapter.

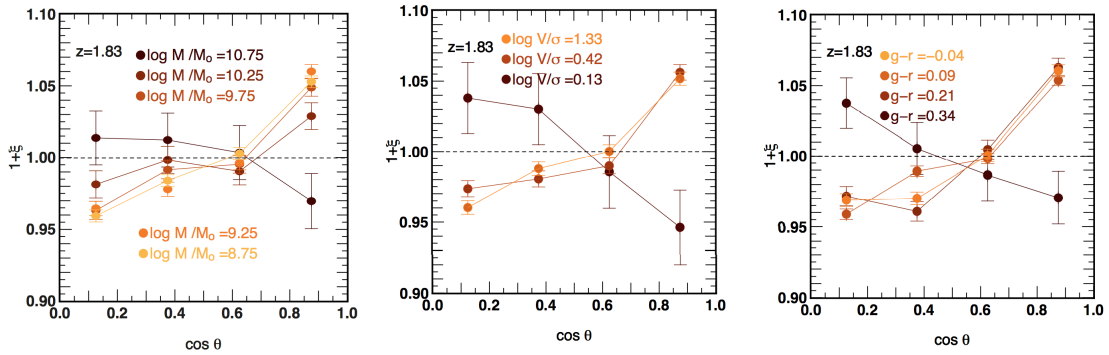

Figure 7.25.: Excess probability of the alignment between the spin of galaxies and their closest filament is shown as a function of some galaxy properties at $z = 1.83$: M_* (*left*), V/σ (*middle*), g - r (*right*). Half-sigma error bars are shown for readability. Dashed line is uniform PDF (null excess probability). Massive, dispersion-dominated, red, galaxies tend to have a spin perpendicular or randomly oriented with the direction of their filament. Low-mass, centrifugally supported, blue, galaxies tend to align with the direction of their closest filament. Figures from Dubois et al. (2014)

Chapter 8

Galaxy properties in filaments using COSMOS2015

Numerous studies have investigated what is the contribution of environment, if any, in regulating (or even shutting down) star formation in galaxies or impacting their morphologies. The easiest and most natural way to characterise environment is to define it in terms of local density, which can be estimated from the galaxy distribution. Environment can therefore be inferred from galaxy clustering or well-defined structures (clusters/groups/pairs and field galaxies, e.g. Lin et al., 2016) or using isotropic galaxy-density estimators, such as the nearest neighbour method (e.g. Dressler, 1980; Elbaz et al., 2007), the fixed aperture method (e.g. Gallazzi et al., 2009) or the annulus method (Wilman et al., 2010). An alternative approach is to determine environment from density maps, computed for instance using Voronoi tesselation or adaptative smoothing (e.g. Scoville et al., 2013; Darvish et al., 2015). Estimating environment from the local overdensity requires a complete sample. For this reason these estimators may be easier to compute from photometric redshift surveys above $z \sim 1.5$ (Lai et al., 2016).
From these investigations, there is now ample evidence that galaxy properties are segregated depending on their local density environment on typical scales of few Mpc (Oemler, 1974; Davis and Geller, 1976; Balogh et al., 1997). At low redshift it has long been established that galaxies are found to be more massive and much less star-forming in high-density regions relative to low-density regions (e.g. Dressler, 1980; Kauffmann et al., 2004; Davidzon et al., 2016). At higher redshift, the trend seems to be weaker (Darvish et al., 2016) and in all the cases debated (Lin et al., 2016). Some investigations even find the opposite trend (e.g. Cooper et al., 2008). This could be partly a consequence of the poor redshift accuracy of the surveys at high redshift and consequently the difficulty to recover correctly the environment from the density (Malavasi et al., 2016). This effect may be mitigated if a (sparser) spectroscopic sample is available on the same field (Cucciati et al., 2016). In fact, the variety of estimators, probing different range of density, can also be a reason for this discrepancy (Sobral et al., 2011).

To explain the predominance of passive galaxies in high density regions, at least at low redshift, group or cluster-specific processes are postulated. These mechanisms must explain the depletion or heating of the cold gas in the quenched galaxies. They can be galaxy harassment (Moore et al., 1996), ram-pressure stripping of the gas (Gunn and Gott, 1972; Quilis et al., 2000), or strangulation of low-mass satellite galaxies (Larson et al., 1980; Balogh et al., 2000). Galaxy mergers may also halt star-formation. In addition, it is not clear if the colour-density relation is a trend specifically related to being in a group or cluster (Peng et al., 2011; Carollo et al., 2014), or due to the central-satellite dichotomy (e.g. Kovač et al., 2014). Other effects of the environment have been highlighted via the measured so-called "galactic conformity" (Weinmann et al., 2006). It quantifies a correlation between the quenching of the central and the quenching of its satellite galaxies: the fraction of quiescent satellites is higher around quiescent

centrals than around star forming centrals. This effect has been detected to redshift 2.5 for both low and high mass satellite galaxies (Kawinwanichakij et al., 2016). It suggests that galactic property depend not only on halo mass but also on the larger scales, and is likely connected to the halo assembly history ("assembly bias" Croton et al., 2007). Hearin et al. (2015) observed this "galactic conformity" up to very large separation (4 Mpc): this effect is difficult to explain in semi-analytical models where galaxy properties depend essentially on halo mass. Attempting to make sense of it, they use the term of "halo accretion conformity" and connect this effect to the mutual evolution of haloes in the same large-scale tidal field, in other words arguing that these haloes are embedded in the same large-scale flow.

When investigating environmental effects using local density estimators, one essentially measures local effects biased towards clusters which are at the nodes of the cosmic web. "Galactic conformity" and assembly bias highlight the existence of tides at large scales which cannot be captured by local isotropic density measurements (see also Figure 7.18). In turn, the dependency of galaxy properties on their environment cannot be entirely explained at the cluster scale alone by halo mass-related processes. It seems therefore natural to consider the effect of coherent large-scale flows imposed by the filaments of the cosmic web. Defining environment in terms of structures (walls, filaments, nodes) allows on the one hand to probe a larger variety of environment than by the simple high-low density discrimination. It also provides the theoretical framework to understand the origin of the galaxy-environment connection as outlined by "galactic conformity".
In light of the results presented in Chapter 7, we expect the dynamics within the intrinsically anisotropic large-scale structures to correlate with the physical properties of the galaxies embedded in them. The vorticity-rich large-scale filaments (Libeskind et al., 2013a; Laigle et al., 2015) are the locus where low-mass galaxies steadily grow in mass via quasi-polar cold gas accretion, while their angular momentum is aligned with the host filament (Codis et al., 2012, 2015). The connection between galaxy angular momentum and observable properties such as colours is not obvious. Nevertheless mass gradients towards the filaments and along them are a simple outcome of the theory (Codis et al., 2015) and these gradients should be measurable in observations. Additionally, we can expect that galaxies accrete more efficiently cold gas when their angular momentum is aligned with the preferential direction of the gas infall (Pichon et al., 2011; Codis et al., 2012), *i.e.* aligned with the filament. This maximal alignment occurs in the highest vorticity regions, i.e at the edge of the filament (see Figure 7.15) whereas in the filament core, haloes terminate their mass assembly through mergers when moving towards nodes as driven by the large-scale flow (see Section 7.4.2).

The dependency of galaxy properties on their large-scale environment has only recently been investigated from observed datasets (e.g. Tempel et al., 2013; Alpaslan et al., 2016; Martínez et al., 2016, Malavasi et al. in preparation). These first investigations confirm the influence of filaments on some galactic properties. They rely on spectroscopic samples that allow reconstructing the three-dimensional distribution of the matter. The potential of photometric surveys may have been underestimated because of their insufficient precision along the line-of-sight, making it seemingly too difficult to reconstruct the three-dimensional field at the required scale to trace the cosmic web. Yet, the accuracy of current photometric redshifts opens the prospect of working with projected two-dimensional slices which are sufficiently thin to unambiguously study the projection of filaments. The great advantage of photometric surveys is their ability to probe different epochs of cosmic evolution to leverage their relative importance in building up galaxies, with complete sample and at much lower cost than spectroscopic surveys.

In this Chapter, I first investigate the feasibility of this reconstruction using my virtual photometric catalogue built from the Horizon-AGN simulation. I show that the structures (filaments, nodes) extracted from the projected density effectively trace their three-dimensional counterparts with a sufficiently high precision to study the effect of an anisotropic environment on galaxy properties. I then use cosmos2015 (Chapter 5) to quantify the variation of these properties with the distance to filaments. For now, I only

Table 8.1.: Limiting mass, comoving transverse width of the slice, redshift errors and corresponding thickness of the slices for the sample selected at $K_s < 24$. The limiting mass is estimated as in Chapter 5. The comoving transverse distance corresponds to 1.4 deg (side of the COSMOS field) at the upper limit of each redshift bin. $M_* > 10$ is the conservative limiting mass that I choose in this work to build a mass-selected sample at $0.5 < z < 0.9$. $\Delta z\,[M_* \sim 10]$ corresponds to two times the median 1σ redshift error determined by LePhare (value enclosing 68% of the redshift PDF) for the galaxies such as $10 < M_* < 10.5$ (the faintest galaxies in the bin, for which the redshift uncertainties are the highest). $D_{\Delta z}\,[M_* \sim 10]$ is the associated thickness in comoving Mpc at the upper limit of the redshift bin.

bin	$M_{\rm lim}$	$D_{\rm trans}$ (cMpc)	$\Delta z\,[M_* \sim 10]$ $(2 \times 1\sigma)$	$D_{\Delta z}\,[M_* \sim 10]$ (cMpc)
$0.5< z <0.6$	8.72	54	0.019	58.3
$0.6< z <0.7$	8.92	62	0.021	61.6
$0.7< z <0.8$	8.95	70	0.024	68.0
$0.8< z <0.9$	9.17	75	0.029	75.2
$0.9< z <1.0$	9.19	82	0.032	80.4
$1.0< z <1.1$	9.23	87	0.045	105.1
$1.1< z <1.2$	9.37	93	0.059	131.4
$1.2< z <1.3$	9.44	98	0.059	128.1

focus on transverse effects. In a later work I will extend this study to longitudinal effects (evolution of galaxy properties along the filament).
Section 8.1 presents the preparation of the catalogue and the extraction of the skeleton. Using a virtual photometric catalogue with the photometric redshift precision of COSMOS2015, it quantifies how well can one trace the cosmic web using the skeleton in projected two-dimensional slices with the photometric redshift precision of COSMOS2015, and how this skeleton relates statistically to its three-dimensional counterpart. Section 8.2 describes the statistical measure of the variation of galaxy properties as a function of their distance to the filament. Here I look for an effect specific to the filaments. Hence I explicitly remove the contribution from nodes to the signal. In this Chapter I consider only the stacked signal in the redshift range $0.5 < z < 0.9$. However, the ultimate aim of this study is to investigate its redshift evolution, which will be presented in Laigle, Pichon, McCracken et al. in preparation.

8.1 Tracing the filaments in projected two-dimensional slices with photo-z

8.1.1 Catalogue preparation and filament identification

Catalogue preparation – I use COSMOS2015 (presented in Chapter 5) to trace the cosmic web in projected two-dimensional slices. I only consider galaxies for which photometric redshifts are between 0.5 and 0.9. At these redshifts, the limiting mass (see Figure 5.16) computed from the K_s band allows me to estimate that the sample is complete down to $M_* \sim 8.9$ M$_\odot$ ($K_s < 24$, $3''$, 3σ).
I remove also from the catalogue all the objects which are flagged as belonging to a bad area, or for which the photometry is possibly contaminated by the light of saturated stars. Galaxy separation between active and passive has been performed from the NUV-R, R-J rest-frame colours (see Figure 5.15).

Thickness of the slices – It is important to carefully chose the thickness of the slices on which the projected density will be computed and the skeleton extracted. The optimal slice thickness has already been investigated for environment estimation from density maps, computed from Voronoi tessellation or adap-

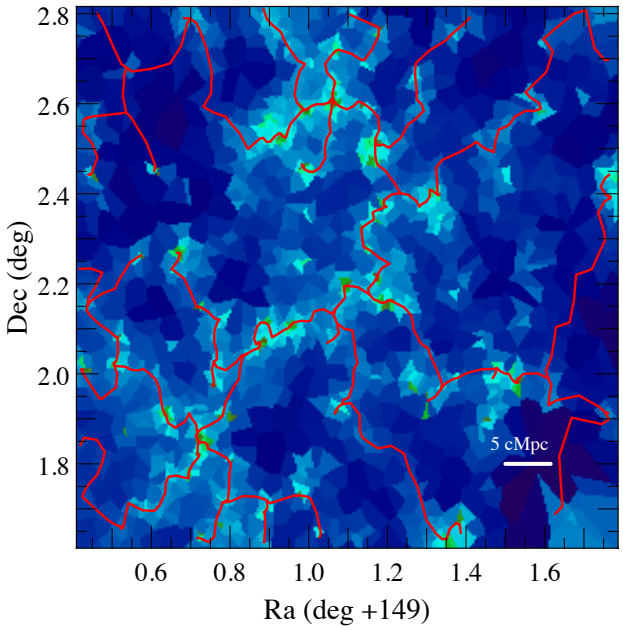

Figure 8.1: The density field estimated from the Delaunay tesselation at $z \sim 0.67$ in a slice of thickness 75 cMpc in the COSMOS field. The red skeleton has been computed with a persistence threshold of 2σ. The horizontal white bar in the lower right indicates the comoving scale length of 5 Mpc.

tive smoothing (Scoville et al., 2013; Darvish et al., 2015, 2016).
In two dimensions, there can be confusion between nodes and projected three-dimensional filaments, or filaments and projected three-dimensional walls. Furthermore, projected filaments may link some overdensities which are in fact not truly linked in three dimensions. Hence if the slices are too thick, the skeleton derived from the projected density will have little correspondence to its three-dimensional counterpart. However, if the chosen thickness of the slices is much below the redshift accuracy, the density will be completely blurred because most of the galaxies will have little probability to lie in the slices. A reasonable balance has to be struck, knowing that the redshift accuracy is a function of the galaxy photometry. Hence fainter galaxies have higher redshift uncertainties. Choosing to work in very thin slices implies removing too faint galaxies from the sample. However if reconstructed with only bright galaxies, the skeleton is expected to be more biased than using the full sample.
To optimally trace the filaments with the same accuracy at $z \sim 0.5$ and at $z \sim 0.9$, I choose to work with a mass-limited rather than a flux-limited sample. I set the limiting mass to 10^{10} $M_\odot$. This choice is not necessarily optimal and could be further investigated from simulations. However, tests on the virtual photometric catalog detailed below confirm that it is good enough to reasonably reconstruct the skeleton.
The redshift uncertainty in cosmos2015 is estimated from the 1σ error given by LePhare for each galaxy. The 1σ error is defined as the value enclosing half of 68% of the probability distribution function of the photometric redshift of the galaxy (see Chapter 3). In the following, the thickness of the slices $D_{\Delta z}$ $[M_* \sim 10]$ is calibrated on the median of the redshift uncertainties ($2 \times 1\sigma$) of the faintest galaxies (such as $10 < \log M_*/M_\odot < 10.5$) in the concerned redshift range.
Table 8.1 gives the evolution of $D_{\Delta z}$ $[M_* \sim 10]$ as a function of redshift. For the sake of simplicity I choose to work with slices of constant thickness (in comoving unit) at $0.5 < z < 0.9$: the appropriated slice thickness is therefore 75 cMpc. The adjacent redshift slices are overlapping and spaced by half of the thickness of the slices. On $0.5 < z < 0.9$ there are 31 slices. The transverse width of the slices is comprised between 54 cMpc and 75 cMpc.

Filament identification The filament identification with DisPerSE operates on the Delaunay tessellation of the particles. The first step is therefore to compute the two-dimensional Delaunay tessellation

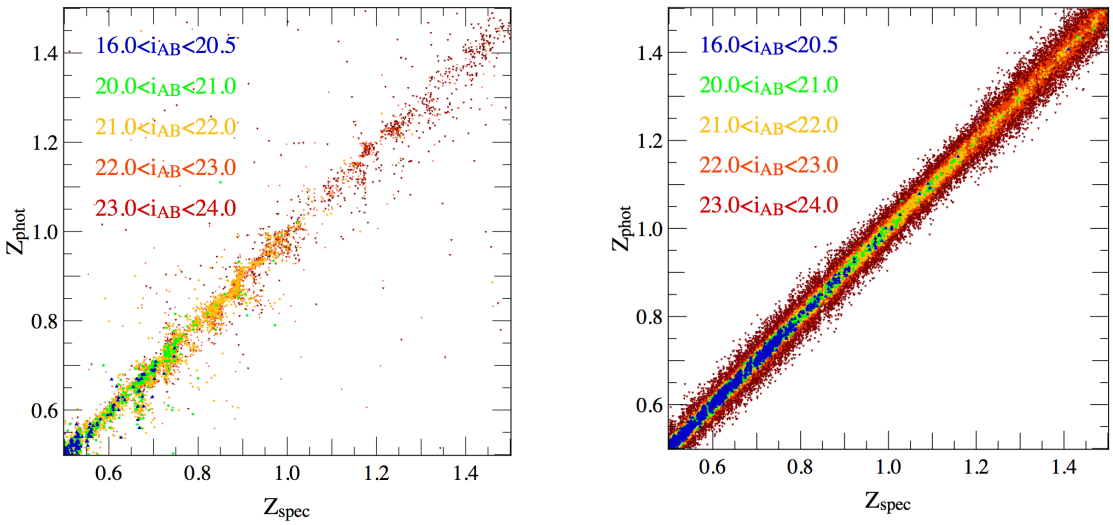

Figure 8.2.: Photometric redshifts versus spectroscopic redshifts in COSMOS2015 (*left*) and Horizon-AGN (*right*). In Horizon-AGN, spectroscopic redshifts are the true redshifts and photometric ones have been computed by calibration on COSMOS2015. Photometric redshift errors are magnitude– and redshift–dependent.

in every slices from the galaxy distribution. To take into account the errors on photometric redshifts in this calculation, I associate to each galaxy in each slice a weight $p_{\mathrm{gal},i}$ which is the probability for this galaxy to be in the slice i given the probability function distribution of its redshift $P_{\mathrm{gal}}(z)$ (as computed by LePhare, see Chapter 3):

$$p_{\mathrm{gal},i} = \int_{z_{1,i}}^{z_{2,i}} P_{\mathrm{gal}}(z)dz \,/\, \int P_{\mathrm{gal}}(z)dz \,, \tag{8.1}$$

where $z_{1,i}$ and $z_{2,i}$ are the lower and upper redshifts of the slice. This weighting allows to reduce the pollution effects of foreground and background galaxies for which the probability to be in the slices is low. To correctly estimate the density and the topology close to boundaries, a surface of "guard" particles is added outside the bounding box and new particles are added by interpolating the estimated density computed on the boundary of the distribution. Note that the Delaunay tessellation is by definition a robust way to reconstruct reliably the density in a field which contains masks. The triangulation automatically connects galaxies from across the masked regions and thus interpolate the density (*e.g.* Aragon-Calvo et al., 2015).
Filaments are then identified with DisPerSE from the Delaunay tessellation (see Chapitre 4). The chosen persistence threshold is 2σ, which guarantees less than 3% of critical pairs to be spurious (Sousbie, 2011). This choice is justified below on simulation by comparison with the three-dimensional skeleton. Figure 8.1 shows the density in a slice at $z \sim 0.67$ and the corresponding skeleton.

8.1.2 Validation of the reconstruction with the Horizon-AGN lightcone

Virtual catalogue preparation – To test our ability to reconstruct the filaments in two dimensions, I use the virtual photometric catalogue computed on the Horizon-AGN lightcone (Chapter 4). I prepared a mock catalogue which has the same characteristics (in terms of photometry and photometric redshift accuracy) as COSMOS2015. In particular, I divided the simulated galaxies in bins of redshift and magnitude, and I perturbed their redshift using a Gaussian error corresponding to the observed one in the same redshift

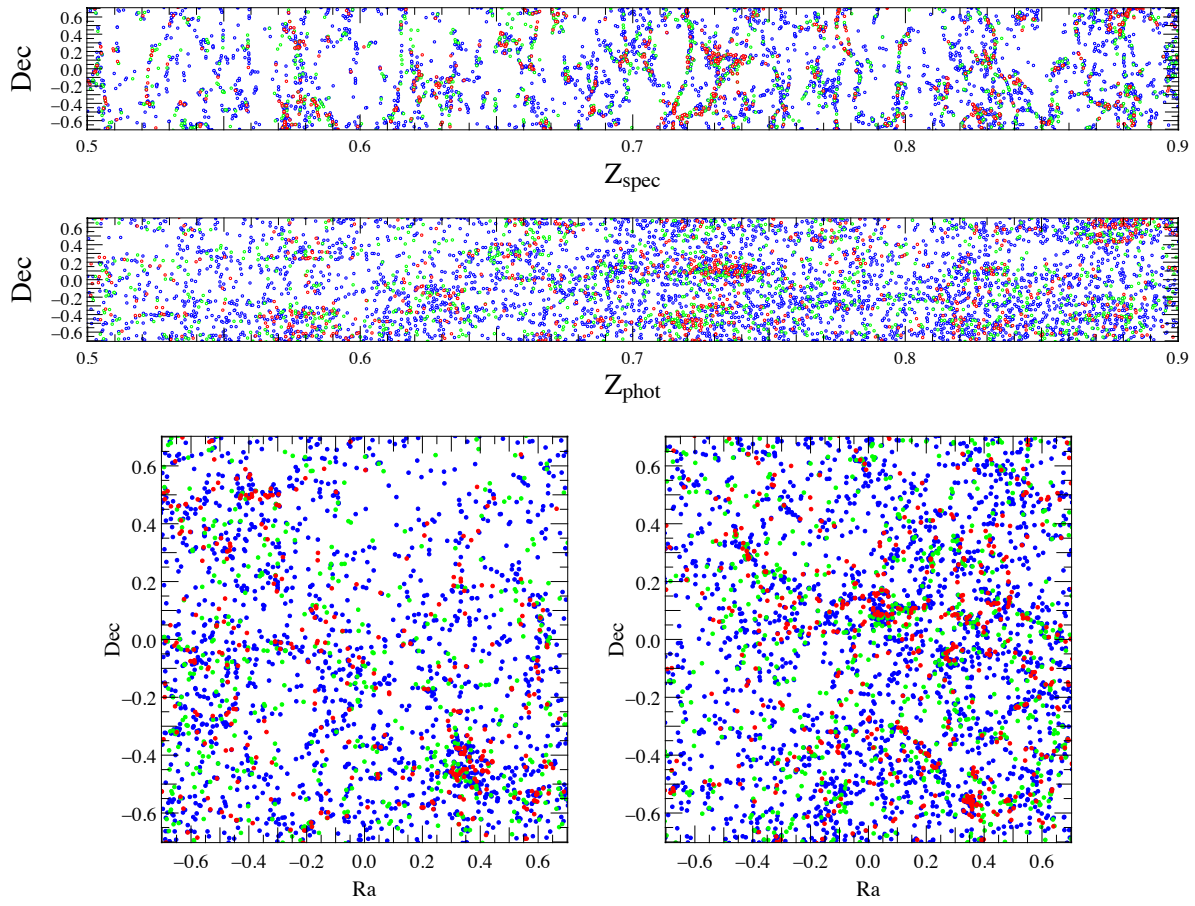

Figure 8.3.: The galaxy distribution from the Horizon-AGN lightcone in a thin slice along the line-of-sight with the true redshifts (*top* panel) and the photometric redshifts (*middle* panel). *Bottom* panels: the projected galaxy distribution in two slices of thickness 75 cMpc. The transverse width of the slices (1.4 deg) is chosen such as to be comparable to the COSMOS field. The colormap encodes the mass of the galaxies. Red: $10.8 < M_* < 12.0$, green: $10.5 < M_* < 10.8$ and blue: $10.0 < M_* < 10.5$.

and magnitude bins. Figure 8.2 displays the comparison between photometric and spectroscopic redshift for the observed sample and the simulated one. To get an idea of the effect of the redshift uncertainty on the three-dimensional galaxy distribution, Figure 8.3 shows the galaxy distribution in the virtual lightcone with their true and perturbed (photometric) redshifts.

Two-dimensional versus real skeleton – Let us now estimate the quality of the skeleton reconstructed with photometric redshifts using the simulated catalogue. I define the three-dimensional skeleton computed from dark matter particles ($\mathrm{SKL_{3D}}$) as the reference skeleton. For the qualification of the reconstruction, I compare its projection $\mathrm{SKL_{3D}^{proj}}$ to the two-dimensional skeleton $\mathrm{SKL_{2D}^{Phot}}$ computed from the observed galaxy distribution. $\mathrm{SKL_{2D}^{Phot}}$ is naturally degraded or biased compared to $\mathrm{SKL_{3D}^{proj}}$ for three reasons.
First, the photometric skeleton $\mathrm{SKL_{2D}^{Phot}}$ will be noisier than the spectroscopic one $\mathrm{SKL_{2D}^{Spec}}$ due to the galaxy redshift uncertainties. Secondly, even with exact redshifts, $\mathrm{SKL_{2D}^{Spec}}$ is a biased reconstruction of the projected three-dimensional skeleton computed from galaxies ($\mathrm{SKL_{3D}^{gal,proj}}$). Finally, it should be acknowledged that the three-dimensional galaxy skeleton ($\mathrm{SKL_{3D}^{gal}}$) is biased compared to the three-dimensional dark matter skeleton ($\mathrm{SKL_{3D}}$). This bias is dependent on the limiting mass of the galaxy catalogue.

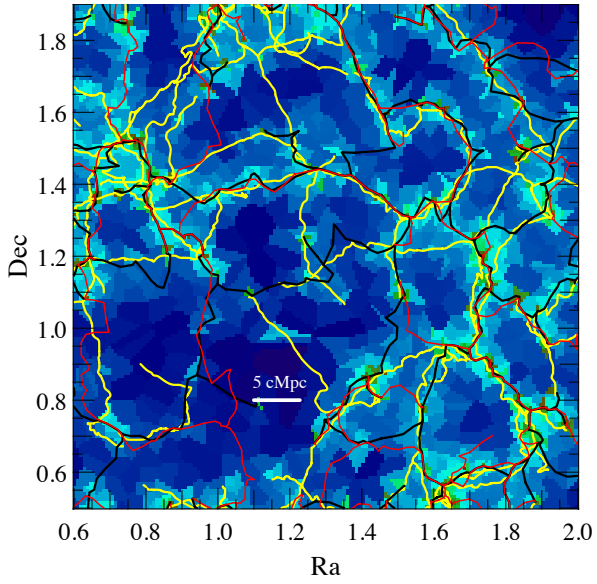

Figure 8.4: A slice of thickness 75 cMpc at redshift $z \sim 0.59$ from the Horizon-AGN lightcone. The yellow skeleton is SKL_{3D}^{proj} computed from the full dark matter particles distribution. The black skeleton is SKL_{2D}^{Spec} and the red skeleton is SKL_{2D}^{Phot} computed with a 2σ persistence threshold. The transverse width of the slices (1.4 deg), is comparable to the COSMOS field. The background is the density computed from the Delaunay tesselation with exact redshifts. Most of the filaments in SKL_{2D}^{Phot} have a close counterpart in SKL_{3D}^{proj}. The horizontal white bar in the lower middle indicates the comoving scale length of 5 Mpc.

These three effects are illustrated in Figure 8.4. This figure displays a slice of thickness 75 cMpc at redshift $z \sim 0.59$. The yellow skeleton corresponds to SKL_{3D}^{proj}, the black skeleton to SKL_{2D}^{Spec} while the red skeleton is SKL_{2D}^{Phot}. Black and red skeletons are reconstructed with a persistence threshold of 2σ. They are in general in good agreement in the densest regions of the field, which confirms that photometric redshift are sufficient to reconstruct the two-dimension skeleton.
Most filaments of SKL_{2D}^{Phot} have a reliable counterpart in SKL_{3D}^{proj}. However some filaments of the former are not recovered in the latter and some filaments of the latter have no counterpart in the former. The choice of the persistence threshold is important to mitigate the number of filaments in SKL_{2D}^{Phot} which have no counterpart in SKL_{3D}^{proj} and which are thus potentially noisy detection. In addition, filaments detected in SKL_{2D}^{Phot} but not in SKL_{3D}^{proj} are potentially walls seen in projection.

By eye, the reconstruction in two dimensions is quite successful. Let us quantify it statistically by measuring the distance between SKL_{2D}^{Phot} and SKL_{3D}^{proj}. For each segment in SKL_{2D} I measure the distance $d_{2D\to3D}$ to the nearest neighbour in SKL_{3D}^{proj}. Then for each segment in SKL_{3D}^{proj} I measure the distance $d_{3D\to2D}$ to the nearest neigbour in SKL_{2D}.
This measure is displayed in Figure 8.5. Solid lines are the distributions of the distances $d_{2D\to3D}$ and dashed lines are the distributions of the distances $d_{3D\to2D}$. Light blue line refers to photometric redshifts and dark blue line to spectroscopic redshifts for the computation of SKL_{2D}. The light and dark blue curves are quite similar, which suggests that the photometric redshift uncertainties have little impact on our ability to recover the three-dimensional structure, at least under the chosen conditions (persistence: 2σ, thickness: 75 cMpc, $M_{\mathrm{lim}} \sim 10^{10}$ $\mathrm{M}_{\odot}$). The fact that the dashed distribution is wider suggest that there are more three-dimensional filaments than two-dimensional ones. The solid distribution is well peaked, meaning that when a two-dimensional filament has a three-dimensional counterpart, both are very close. However, the tail of the solid distribution implies that a little number of two-dimensional filaments have no three-dimensional counterpart at all. Nevertheless, the overall agreement is good.

Recovering a three-dimensional signal in two dimensions at the COSMOS2015 accuracy – Although the filaments can be reliably traced from two-dimensional slices at the COSMOS2015 redshift accuracy, it is not obvious that a distance to filament trend will be statistically recovered in two dimensions. As a matter of fact, the projection will inevitably blur the signal: galaxies far from the filament but on the same line-of-sight will be counted as very close of it in projection.
Figure 8.6 presents the natural and cumulative distributions of the distances to SKL_{3D} for simulated galax-

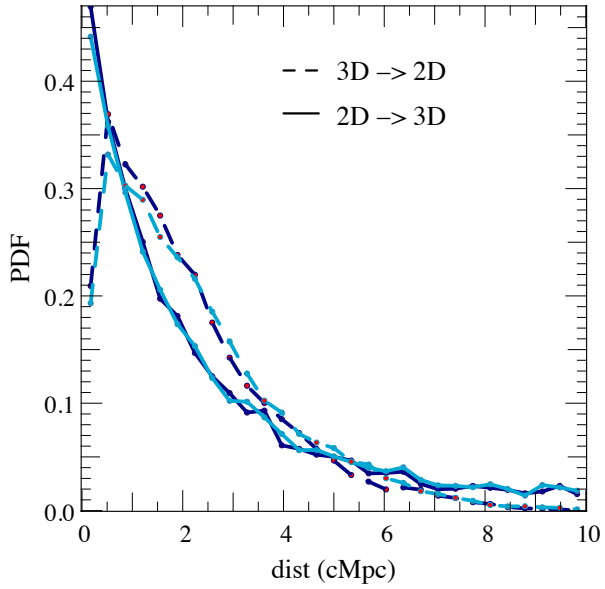

Figure 8.5: *Left:* PDF of the distances between $\mathrm{SKL_{2D}}$ and $\mathrm{SKL_{3D}^{proj}}$. For each segment in $\mathrm{SKL_{3D}^{proj}}$ I measured the distance to the closest segment in $\mathrm{SKL_{2D}}$ (dashed line) and for each segment in $\mathrm{SKL_{2D}}$ I measured the distance to the closest segment in $\mathrm{SKL_{3D}^{proj}}$ (solid line). The light blue curves refer to $\mathrm{SKL_{2D}^{Phot}}$ while the dark blue curves refer to $\mathrm{SKL_{2D}^{Spec}}$.

ies in four different bins of masses. The signal is measured here at fixed redshift in a snapshot of Horizon-AGN. Let me stress once again that I am looking for environmental effects driven specifically by filaments. Indeed, as galaxies are found more massive and passive in clusters than in the field, there exists mass gradients towards nodes that are purely driven by the effect of nodes. To down-weight the contribution of nodes in the analysis, I weight the distance of each galaxy by the inverse of the smoothed density field using a Gaussian filter with $\sigma = 3$ Mpc (Gay et al., 2010, Malavasi et al. in preparation).
We find mass gradients towards the filaments. More massive galaxies are more confined to filaments than less massive ones. Then the same measurement is carried in two dimensions with $\mathrm{SKL_{2D}^{Phot}}$ on the Horizon-AGN lightcone in every slices in $0.5 < z < 0.9$. The projected transverse distance to the filament, measured in degree, is converted in comoving Mpc. The cumulative distributions in every slices are stacked and the mean distribution is presented in the left panel of Figure 8.7. Errorbars represent the standard deviation to the mean. In two dimensions with photometric redshifts, the galaxy mass segregation is recovered.

8.2 Galaxy mass segregation towards filaments in COSMOS

Having confirmed the feasibility to trace the filaments in two-dimensional slices with a thickness of 75 cMpc and to recover the corresponding three-dimensional signal, I now study the distribution of galaxies as a function of their distance to filaments using the cosmos2015 catalogue.
Figure 8.7 presents the cumulative distributions of the distance to the filaments for different bins of mass in the Horizon-AGN lightcone and in cosmos2015. The signal measured in all slices in the redshift range $0.5 < z < 0.9$ has been stacked. The errorbars show the standard deviation to the mean. The contribution of each galaxy is weighted by its probability to be in the slice computed from its redshift uncertainty. In addition, I down-weight the contribution of the nodes as presented below. Note that the lowest mass bin $9.5 < \log M_*/\mathrm{M}_\odot < 10.0$ should be considered with caution. Those galaxies are not used for computing the skeleton, and their redshift uncertainties are generally higher than the thickness of the slices at redshift $z > 0.7$. However, their distribution is not consistent with the random signal (dashed lines). In COSMOS, the distance distributions show that massive galaxies are more confined to filaments compared to less massive ones, in agreement with our theoretical expectations and with the measurement in Horizon-

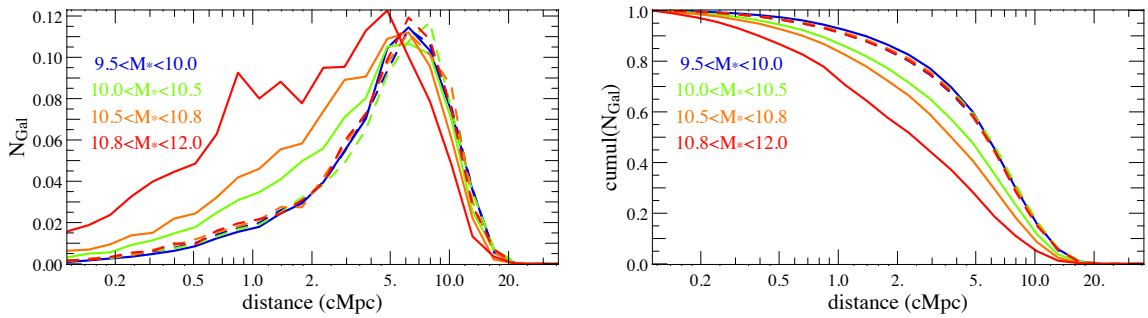

Figure 8.6.: Differential (*left*) and cumulative (*right*) distributions of the distances to the three-dimensional skeleton for galaxies as a function of their masses in HORIZON-AGN. As expected, more massive galaxies are found closer to the centre of the filaments with respect to less massive. The dashed lines correspond to a random signal: galaxy masses in the catalogue have been randomised with respect to their position, then the same mass cuts have been applied. Unsurprisingly, all mass bins are superimposed for the random signal. This measure is not made in the lightcone but in a snapshot at fixed redshift $z \sim 0.5$.

AGN. The galaxy mass segregation is stronger in the HORIZON-AGN lightcone than in COSMOS2015. A possible explanation is that uncertainties on galaxy masses are not yet modelled in the simulation, and not taken into account in the measurement in COSMOS2015. These uncertainties are likely to dilute the segregation signal.

Figure 8.8 presents the cumulative distributions of the distance to the filaments for star-forming and passive galaxies in COSMOS2015. All galaxies with $M_* > 10^{10}$ $M_\odot$ are considered in the measure. This Figure shows that passive galaxies are statistically closer to the filaments than star-forming. As I minimised the contribution of galaxies in nodes, this effect is specific to the filaments.

These preliminary measurements are quite promising. On the one hand they confirm the role played by the large-scale environment in shaping observable galaxy properties such as their masses and their colours. On the other hand they demonstrate our ability to carry out this study using photometric redshifts in two dimensions. A recent study that I co-sign on the VIPERS field from spectroscopic redshifts in three dimensions at $z \sim 0.7$ (Malavasi et al. in preparation) has highlighted the same mass segregation towards the filaments. These convergent results can be interpreted in the framework of Chapter 7. The anisotropic Kaiser bias (Figure 7.18) naturally explains the observed mass gradients. Low-mass galaxies formed in the neighbourhood of vorticity rich filaments with a spin aligned with that vorticity: they accrete gas coherently and are therefore strongly star-forming. More massive galaxies are predominantly found in the heart of the filaments, and encounter mergers when migrating towards nodes as driven by the large-scale flows. Their spins are misaligned with or perpendicular aligned to the filaments. Quenching of massive galaxies occurs already in the filaments and not only in nodes. However, it is not clear if this quenching is directly connected to to the large-scale environment or indirectly as driven by the mass which is itself connected to the environment.

8.3 Summary

Using the photometric catalogue COSMOS2015, I have investigated the anisotropic dependency of galaxy properties on their environment. The virtual photometric catalogue extracted from the HORIZON-AGN

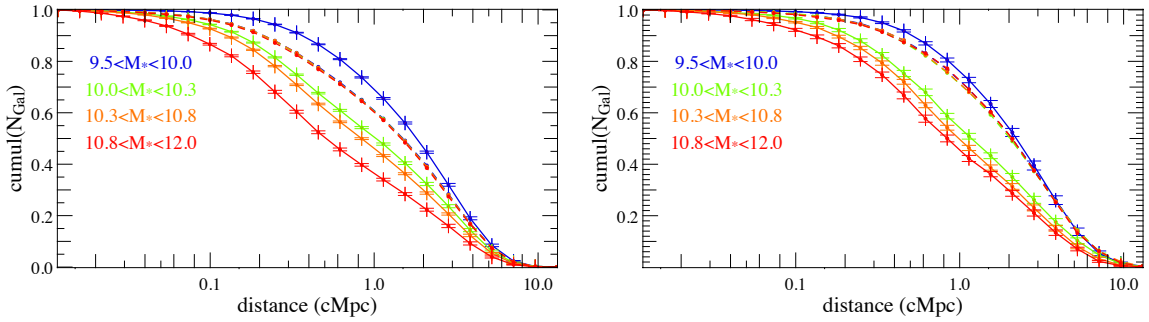

Figure 8.7.: Cumulative distribution of the distances to the two-dimensional photometric skeleton for galaxies as a function of their masses in Horizon-AGN (*left*) and cosmos2015 (*right*). The dashed lines correspond to a random signal: galaxy masses in the catalogue have been randomised with respect to their positions, then the same mass cuts have been applied. Note that the lowest mass bin $9.5 < \log M_*/\mathrm{M}_\odot < 10.0$ should be considered with caution because the thickness of the slices has been calibrated for galaxies with a mass $\log M_*/\mathrm{M}_\odot > 10$. Here, the signal measured in every slices between $z \sim 0.5$ and $z \sim 0.9$ is stacked. Errorbars correspond to 1σ deviation to the mean.

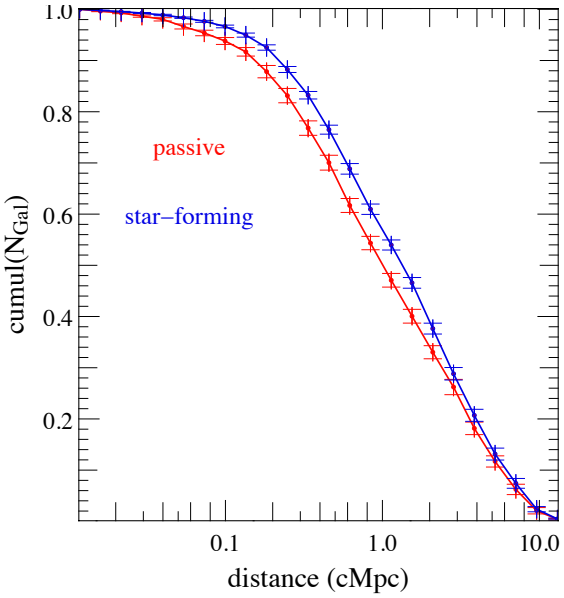

Figure 8.8: Cumulative distribution of the distances to the two-dimensional photometric skeleton for passive (red) and star-forming (blue) galaxies in cosmos2015. Only galaxies with $\log M_*/\mathrm{M}_\odot > 10$ are considered here. The signal measured in every slices between $z \sim 0.5$ and $z \sim 0.9$ is stacked. Errorbars correspond to 1σ deviation to the mean. Contribution of galaxies in nodes is down-weighted by the inverse of the density, such as to highlight an effect specifically related to the filaments.

hydrodynamical simulation has been used to demonstrate our ability to recover the filamentary structures in two-dimensional slices at the photometric redshift precision of COSMOS2015. The main findings of this Chapter are the following:

- Photometric surveys with the photometric precision of COSMOS2015 offer the prospect to reliably study the cosmic web from projected two-dimensional slices;
- We observe a significant trend for massive galaxies to be closer to filaments than less massive ones. The signal persists when down-weighting the contribution of nodes;
- We also observe segregation of passive and star-forming galaxies near the filaments. Passive galaxies are more confined in the heart of the filament than star-forming ones, suggesting than quenching processes occur already in the filaments and not only in the nodes;
- These findings are in agreement with an independent study in VIPERS using spectroscopic redshifts (Malavasi et al. in preparation).

The great advantage of photometric redshifts is their ability to probe a large redshift baseline. The next step is to extend this work to higher redshifts. This will be especially challenging as we will need to carefully disentangle the evolutionary effects and selection biases. My simulated catalogue will be a key ingredient in achieving this.

Conclusion

In this thesis I have presented new constraints on galaxy formation and evolution by studying the galaxy mass growth and the co-evolution of the cosmic web and embedded galaxies, from the epoch of cosmic dawn to today.
Galaxy light is the only observable at our disposal to study their mass evolution. Galaxy colours and luminosities are directly connected to their present and past star formation histories, and in turn to the way they accrete gas, transform it into stars, and outflow it through feedback processes. Understanding galaxy evolution implies therefore to follow the cycle of the gas, which involves different scales. The gas reservoir at *large scales* is the intergalactic medium. Gas flows along the filamentary structure of the dark matter which constitutes the natural environment for galaxy formation and evolution. This gas penetrates the circumgalactic medium and is accreted into galaxies. At these *intermediate scales*, the gas inflow depends both on the large-scale environment which drives it within the galaxy and on the galaxy and halo masses which can regulate it. At *small scales* within the galaxy, newly accreted gas is turned into stars. Stellar evolution and AGN feedback perturb the gas, heat it remove it from the galaxy while possibly disrupting the gas infall.
The first part of my thesis has been devoted to understanding the connection between galaxy stellar mass growth and their properties deduced from their photometry. The small-scale processes, which shape galaxy colours and luminosities, are indeed closely related to how and at which rate the gas is accreted, hence to the global (intermediate-scale) galaxy properties, especially their masses. As they also impact gas infall or remove gas from galaxies, these processes shape in return stellar mass growth. In this first part, I have shown that galaxy evolution is therefore driven at first order by mass-dependent processes.
However, galaxy mass growth itself is conditioned by the proximity to the large-scale filamentary gas reservoir. In the second part of my thesis, I have shown how the dynamics of infall builds up coherently the angular momentum of halos and galaxies. Hence it emerges that galaxy masses and angular momenta are two dependent quantities impacted by their anisotropic environment.
I detail below my main results and I suggest further developments.

Following galaxy stellar mass growth up to $z \sim 6$ through one-point statistics – To constrain processes involved in galaxy mass assembly which act gradually over cosmic time, the build-up has to be followed on a large redshift range covering the complete cosmic history of galaxy formation. Looking far is looking back in time: observations of our past lightcone allow us to follow globally this process by statistically studying galaxy populations at different epochs. Observationally, looking far is also more difficult. As spectra of distant galaxies are redshifted, panchromatic surveys with deep NIR and IR data are required to follow the mass and star formation rate of galaxies with cosmic time. In Chapter 5, I have built such a sample in the COSMOS field. I have used the best available data to create the new 30-band photometric catalog COSMOS2015 including the newly released NIR and IR data (UVISTA-DR2 and IRAC). The depth of the data, the improved method of PSF homogenisation and the accurate photometry extraction that I implemented make this catalog a unique resource, allowing to explore for the first time on a relatively

large field the low-mass end of the high-redshift galaxy population. The catalog reaches a completeness limit of 10^{10} $M_\odot$ to $z = 4$ and is presented in Laigle et al. 2016.

While observations provide an Eulerian view of the Universe by looking at galaxy populations at fixed redshifts, simulations which model *ab initio* the co-evolution of baryonic and dark matter, allow for a Lagrangian one: from the simulations it is possible to follow the same population as a function of redshift and therefore to understand how non-linear processes progressively shape galaxy properties. The comparison of simulations to observations is thus enlightening and strongly constrains our galaxy evolution model. In order to provide a solid comparison, I have built mock observations from the hydrodynamical simulation Horizon-AGN. These synthetic images and the virtual photometric catalog are presented in Chapter 4.

In Chapter 6 I have compared the one-point statistics of galaxy properties in observations and simulations over the $1 < z < 6$ redshift range. These results are presented in Kaviraj et al. 2016. Specifically I have compared the luminosity functions, mass functions, specific star formation rates and colours from Horizon-AGN with a set of observational data. These observables capture the evolutionary trends of observed galaxies throughout cosmic time. I found an overall agreement with the observations. The comparison with the Horizon-noAGN simulation attests that the regulation of the star formation is mainly driven by AGN feedback for massive galaxies. For low mass galaxies however, the tension between observed and simulated mass functions suggests that the importance of supernovae feedback is underestimated in Horizon-AGN.

Inferring mass and star formation rate from the photometry in observations can be done only at the cost of some assumptions which may partially bias the computed quantities and therefore the comparison with simulations. A way to understand and quantify these biases is to project the simulations in the observation frame and to process them with the exact same tool as observations. In Chapter 6 I have discussed some preliminary results of this strategy, highlighting the effect of simplifying assumptions at the SED-fitting stage. Once these biases are quantified the mismatch in the derived quantities between the mocks and the observed surveys allows one to (in)validate unambiguously the assumptions made by galaxy evolution models.

I have demonstrated that the redshift evolution of one-point statistics of galaxy properties is reasonably well understood when invoking mass-dependent processes and relying on merger-trees alone as the main driver of galactic evolution. Nevertheless, though the simulation reproduces the general observed trends, it does not fully reproduce the scatter in the observed properties at a given mass, suggesting that there could be additional physical processes not properly modelled in the simulation. Since I focused on one-point statistics I did not quantify non-local environmental processes in this Chapter.

Dependencies of galaxy properties on their large-scale environment – The infalling cold gas is the fuel from which galaxies form stars. Angular momentum rich gas flows anisotropically along the cosmic web. In Chapter 7 I have shown how haloes acquire their angular momentum from this large-scale flow. These results are presented in Laigle et al. 2015. I measured the orientation of vorticity within the cosmic web in simulations and found that the flow-crossing from walls generates multipolar high-vorticity quadrants in filaments. Structures which are forming in these flows will end up with a spin preferentially aligned with that vorticity and consequently with the filament. They will accrete matter through secondary infall with a coherent rotational motion, up to a specific transition mass corresponding to the Eulerian size of the quadrant, beyond which inflow will start to advect misaligned angular momentum. By revisiting Lagrangian tidal torque theory in the vicinity of a filament saddle point, it has recently been shown that less massive haloes would have a spin parallel to the filament, while more massive ones would have a spin in the perpendicular direction. Both this trend and the transition mass are in good agreement with my Eulerian view: the vorticity of the large-scale flow successfully explains the alignment of low-mass halo spin with the filament and the transition mass from alignment to mis-alignment.

This investigation has been extended to virtual galaxies in the Horizon-AGN hydrodynamical simulation. As we have shown in Dubois et al. 2014, spins of galaxies tend to be preferentially parallel to their

neighbouring filaments, for low-mass, young, blue galaxies, and perpendicular for higher mass, higher velocity dispersion, red, old galaxies. A simple outcome of this theory is that halo mass become a function of the distance to filaments and to the nodes.
While this is predicted for dark matter halos, my results from Chapter 8 confirm this for galaxies. Using the remarkable accuracy of my photometric redshifts I have studied this dependency with COSMOS2015. I have reconstructed the anisotropic environment from the galaxy density in projected slices taking into account the errors on photometric redshifts. I used the HORIZON-AGN lightcone to quantify the accuracy of the large-scale structure reconstruction from photometric redshifts before proceeding to the reconstruction in COSMOS. I found mass gradients towards the filaments. Interestingly, I also found a segregation between star-forming and passive galaxies towards the filaments.

To summarise, this thesis has first presented evidence that the galaxy mass growth and star formation regulation are at first order dependent on the mass of the galaxy. Secondly, I have shown that the halo angular momentum acquisition is driven by the vorticity of the large scale flow, implying that some halo properties (mass, angular momentum) also depend on their anisotropic filamentary environment. This dependency can be extended to galaxies in simulations, although it was *a priori* less obvious how this large scale effect translates at the very small galaxy scale despite of the baryonic physics. Nevertheless, as predicted by the theory, I have indeed found evidence that galaxy masses are segregated toward the filaments. I also found a segregation in terms of the bimodal split of the galaxy population, suggesting that quenching of massive galaxies occurs already in the filaments and not only in nodes. Hence the question which naturally arises is the following: are the angular momentum alignment with the filament and the star formation efficiency in the galaxy related?
This question is not fully addressed, but I can anticipate the following answer. Consider that low-mass galaxies formed in the neighbourhood of vorticity rich filaments with a spin consistent with the preferential direction of the gas infall, *i.e.* aligned with the filament. Cold flows are driven by the filament and penetrate directly in the core of dark haloes. Galaxies accrete gas coherently and the effects are two-fold: the direct gas fuelling makes them strongly star forming and the coherent rotational motion of the gas implies that the alignment of their angular momentum increases in the direction of the filament. The maximal galaxy alignment occurs in the highest vorticity regions, *i*.e. at the edge of the filament, which is also the locus where galaxies form more efficiently stars. On the other hand, more massive galaxies are predominantly found in the heart of the filaments where they terminate their mass assembly through mergers when migrating towards nodes driven by the large-scale flows. Their spins are misaligned or perpendicular to the filaments and they are less efficient to form stars probably because of a less efficient gas accretion and/or quenching processes (feedback-related or density/groups-related processes). This general scenario clearly highlights a global dependency of galaxy properties on their environment. The cornerstone of this scenario is the close relation between mass, angular momentum and cosmic environment.
This trend is not in competition with the results of Chapter 6, but complementary. Indeed, star formation within the galaxy from newly accreted gas is subsequently regulated by processes more directly connected to mass such as supernovae feedback at the low mass end and AGN feedback at the high mass end. On the one hand when measuring one-point statistics of masses, luminosities or colours on galaxy populations, I cannot account simultaneously for galactic environment, and I therefore highlight internal processes. On the other hand when I study the dynamical dependency of galaxies on the anisotropic large-scale environment, I average on the galaxy populations without considering the effect of their internal mass-dependent processes. A comprehensive model of galaxy evolution has to take both of these aspects into account.

Prospects – My thesis stresses the importance of the large-scale environment jointly with mass-dependent and density-dependent effects in shaping galaxy mass growth and regulating star formation. Let me now present some possible extensions of this work.

The COSMOS2015 catalog presented in Chapter 5 provides a reliable dataset up to $z \sim 4$. It is also a basis for higher redshift investigations. At $z > 4$, galaxy masses selected on COSMOS2015 could be re-computed by SED fitting with high-redshift specific templates while robust limiting mass could be estimated from the IR images. In turn this will allow us to reliably extend the mass function up to $z \sim 6$ on COSMOS, consistently with the low redshift data. Constraining the high-mass end of the mass function at high redshift is in particular important to probe the earliest stage of cosmic history and estimate the birth time of primordial massive galaxies. In particular, the shape of the mass function, if well constrained, should tell us if the privileged mode of galaxy mass growth at these redshift is fuelling by cold flows.

For this analysis, observational bias estimation through the "end-to-end" pipeline presented in Chapter 6 would considerably help to interpret mass functions. It is therefore worthwhile to extend this analysis to the direct photometric extraction on mock images. Furthermore, there are many other long term goals for this "end-to-end" analysis. For instance, it should allow us to test galaxy morphology and weak lensing estimators at the precision of upcoming surveys such as LSST.

Finally the next frontier for spectroscopic but also photometric large surveys is to study the large-scale dependency of galaxy properties *e.g.* on their distance to filaments presented in Chapter 8. In particular we should exploit the large redshift range probed by photometric surveys to carry out a study of the redshift evolution of the signal. For instance the exquisite accuracy that the J-PAS photometric survey will reach should be ideal for such analysis.

In addition to the identified galaxy segregation, my study on COSMOS2015 has demonstrated the possibility to extract a reliable topological information from the two-dimensional density field computed from photometric surveys if redshift accuracy is high enough. It would be interesting to investigate if this study could be extended to statistical quantities which can provide constraints on the cosmology (*e.g.* genus, peaks clustering, count-in-cells) which could be achieved with upcoming surveys such as J-PAS, Euclid or LSST. The Horizon-AGN lightcone is the ideal tool for making predictions on the reliability of the two-dimensional signal.

Bibliography

Aarseth, S. J., Turner, E. L., and Gott, III, J. R. (1979). *ApJ*, 228:664–683.
Abel, T., Hahn, O., and Kaehler, R. (2012). *MNRAS*, 427:61–76.
Agertz, O. and Kravtsov, A. V. (2015). *ApJ*, 804:18.
Agertz, O., Moore, B., Stadel, J., et al. (2007). *MNRAS*, 380:963–978.
Aguirre, A., Hernquist, L., Schaye, J., et al. (2001). *ApJ*, 561:521–549.
Allen, C. W. (1976). *Astrophysical Quantities*. Addison-Wesley Professional, 4 edition.
Alpaslan, M., Grootes, M., Marcum, P. M., et al. (2016). *MNRAS*, 457:2287–2300.
Alpher, R. A., Bethe, H., and Gamow, G. (1948). *Physical Review*, 73:803–804.
Aragón-Calvo, M. A., Jones, B. J. T., van de Weygaert, R., and van der Hulst, J. M. (2007). *A&A*, 474:315–338.
Aragon-Calvo, M. A., Weygaert, R. v. d., Jones, B. J. T., and Mobasher, B. (2015). *MNRAS*, 454:463–477.
Aragon-Calvo, M. A. and Yang, L. F. (2014). *MNRAS*, 440:L46–L50.
Arnouts, S., Le Floc'h, E., Chevallard, J., et al. (2013). *A&A*, 558:A67.
Arnouts, S., Moscardini, L., Vanzella, E., et al. (2002). *MNRAS*, 329:355.
Arnouts, S., Walcher, C. J., Le Fèvre, O., et al. (2007). *A&A*, 476(1):137–150.
Aubert, D., Pichon, C., and Colombi, S. (2004). *MNRAS*, 352(2):376–398.
Bacon, R., Accardo, M., Adjali, L., et al. (2010). The MUSE second-generation VLT instrument. In *Ground-based and Airborne Instrumentation for Astronomy III*, volume 7735 of *Proc. SPIE*, page 773508.
Bailin, J. and Steinmetz, M. (2005). *ApJ*, 627.
Baldry, I. K., Balogh, M. L., Bower, R. G., et al. (2006). *MNRAS*, 373:469–483.
Balogh, M. L., Morris, S. L., Yee, H. K. C., Carlberg, R. G., and Ellingson, E. (1997). *ApJ Let.*, 488:L75–L78.
Balogh, M. L., Navarro, J. F., and Morris, S. L. (2000). *ApJ*, 540:113–121.
Bastian, N., Covey, K. R., and Meyer, M. R. (2010). *ARA&A*, 48:339–389.
Baugh, C. M. (2006). *Reports on Progress in Physics*, 69:3101–3156.
Baum, W. A. (1962). *Problems of Extra-Galactic Research*, 15:390.
Bekki, K. (2015). *MNRAS*, 449:1625–1649.
Bell, E. F. and de Jong, R. S. (2001). *ApJ*, 550:212–229.
Benson, A. J. (2012). *New Astronomy*, 17:175–197.
Benson, A. J., Bower, R. G., Frenk, C. S., et al. (2003). *ApJ*, 599:38–49.
Bertelli, G., Bressan, A., Chiosi, C., Fagotto, F., and Nasi, E. (1994). *A&AS*, 106.
Bertin, E. (2010). pages 1–44.
Bertin, E. (2013). PSFEx: Point Spread Function Extractor. Astrophysics Source Code Library.
Bertin, E. and Arnouts, S. (1996). *ApJ Sup.*, 117:393.
Bertin, E., Mellier, Y., Radovich, M., et al. (2002). *Astronomical Data Analysis Software and Systems XI*, 281:228.
Béthermin, M., Kilbinger, M., Daddi, E., et al. (2014). *A&A*, 567:A103.
Bett, P., Eke, V., Frenk, C. S., et al. (2007). *Monthly Notices of the Royal Astronomical Society*, 376(1):215–232.
Bianchi, L. and GALEX Team (1999). *Mem. Soc. Astr. It.*, 70.
Bielby, R., Hudelot, P., McCracken, H. J., et al. (2012). *A&A*, 545:23.
Bieri, R., Dubois, Y., Silk, J., and Mamon, G. A. (2015). *ApJ Let.*, 812:L36.
Bieri, R., Dubois, Y., Silk, J., Mamon, G. A., and Gaibler, V. (2016). *MNRAS*, 455:4166–4182.
Binney, J. (1977). *ApJ*, 215:483–491.
Birnboim, Y. and Dekel, A. (2003). *MNRAS*, 345:349–364.
Blanton, M. R. and Moustakas, J. (2009). *ARA&A*, 47:159–210.
Blitz, L. (1993). Giant molecular clouds. In Levy, E. H. and Lunine, J. I., editors, *Protostars and Planets III*, pages 125–161.
Blitz, L. and Shu, F. H. (1980). *ApJ*, 238:148–157.
Bolzonella, M., Miralles, J.-M., and Pelló, R. (2000). *A&A*, 363:476–492.
Bond, J. R., Kofman, L., and Pogosyan, D. (1996a). *Nature*, 380:603–+.
Bond, J. R., Kofman, L., and Pogosyan, D. (1996b). *Nature*, 380:603–606.

Booth, C. M. and Schaye, J. (2009). *MNRAS*, 398:53–74.
Bouwens, R. J., Illingworth, G. D., Oesch, P. A., et al. (2012). *ApJ*, 754:83.
Bower, R. G., Benson, A. J., Malbon, R., et al. (2006). *MNRAS*, 370:645–655.
Brinchmann, J., Charlot, S., White, S. D. M., et al. (2004). *MNRAS*, 351:1151–1179.
Brusa, M., Civano, F., Comastri, A., et al. (2010). *ApJ*, 716:348–369.
Bruzual, G. and Charlot, S. (2003). *MNRAS*, 344:1000–1028.
Bruzual A., G. and Charlot, S. (1993). *ApJ*, 405:538–553.
Calzetti, D., Armus, L., Bohlin, R. C., et al. (2000). *ApJ*, 533:682.
Cantalupo, S., Arrigoni-Battaia, F., Prochaska, J. X., Hennawi, J. F., and Madau, P. (2014). *Nature*, 506:63–66.
Capak, P., Aussel, H., Ajiki, M., et al. (2007). *ApJ Sup.*, 172(1):99–116.
Cappelluti, N., Hasinger, G., Brusa, M., et al. (2007). *ApJ Sup.*, 172:341–352.
Caputi, K. I., Ilbert, O., Laigle, C., et al. (2015). *ApJ*, 810:73.
Cardamone, C. N., van Dokkum, P. G., Urry, C. M., et al. (2010). *ApJ Sup.*, 189:270–285.
Cardelli, J. A., Clayton, G. C., and Mathis, J. S. (1989). *ApJ*, 345:245–256.
Carollo, C. M., Cibinel, A., Lilly, S. J., et al. (2014). *ArXiv e-prints*.
Cattaneo, A., Blaizot, J., Weinberg, D. H., et al. (2007). *MNRAS*, 377:63–76.
Caucci, S., Colombi, S., Pichon, C., Rollinde, E., and Petitjean, P. e. a. (2008). *MNRAS*, 386:211–229.
Cautun, M., van de Weygaert, R., Jones, B. J. T., and Frenk, C. S. (2014). *MNRAS*, 441:2923–2973.
Chabrier, G. (2003). *PASP*, 115:763–795.
Chapon, D., Mayer, L., and Teyssier, R. (2013). *MNRAS*, 429:3114–3122.
Charlot, S. and Fall, S. M. (2000). *ApJ*, 539:718–731.
Chevallard, J. and Charlot, S. (2016). *ArXiv e-prints*.
Chevallard, J., Charlot, S., Wandelt, B., and Wild, V. (2013). *MNRAS*, 432:2061–2091.
Chisari, N. E., Laigle, C., Codis, S., et al. (2016). *ArXiv e-prints*.
Cisewski, J., Croft, R. A. C., Freeman, P. E., and Genovese, C. R. e. a. (2014). *MNRAS*, 440:2599–2609.
Civano, F., Elvis, M., Brusa, M., et al. (2012). *ApJ Sup.*, 201:30.
Civano, F., Marchesi, S., Comastri, A., et al. (2016). *ArXiv e-prints*.
Codis, S., Pichon, C., Devriendt, J., et al. (2012). *MNRAS*, 427:3320–3336.
Codis, S., Pichon, C., and Pogosyan, D. (2015). *MNRAS*, 452:3369–3393.
Colless, M. (2004). *Measuring and Modeling the Universe*, page 196.
Colless, M., Dalton, G., Maddox, S., et al. (2001). *MNRAS*, 328:1039.
Conroy, C. (2013). *ARA&A*, 51:393–455.
Conroy, C. and Gunn, J. E. (2010). *ApJ*, 712:833–857.
Conroy, C., Gunn, J. E., and White, M. (2009). *ApJ*, 699:486.
Conroy, C., White, M., and Gunn, J. E. (2010). *ApJ*, 708:58–70.
Cooper, M. C., Newman, J. A., Weiner, B. J., et al. (2008). *MNRAS*, 383:1058.
Cooray, A. and Sheth, R. (2002). *Phys. Rep.*, 372:1.
Coupon, J., Arnouts, S., van Waerbeke, L., et al. (2015). *MNRAS*, 449:1352–1379.
Courteau, S., Cappellari, M., de Jong, R. S., et al. (2014). *Reviews of Modern Physics*, 86:47–119.
Cowie, L. L., Gardner, J. P., Lilly, S. J., and McLean, I. (1990). *ApJ*, 360:L1.
Cowie, L. L., Songaila, A., Hu, E. M., and Cohen, J. G. (1996). *Astronomical Journal v.112*, 112:839.
Crighton, N. H. M., Hennawi, J. F., and Prochaska, J. X. (2013). *ApJ Let.*, 776:L18.
Croton, D. J., Gao, L., and White, S. D. M. (2007). *MNRAS*, 374:1303.
Croton, D. J., Springel, V., White, S. D. M., et al. (2006). *MNRAS*, 367:864–864.
Cucciati, O., Marulli, F., Cimatti, A., et al. (2016). *ArXiv e-prints*.
Daddi, E., Cimatti, A., Renzini, A., et al. (2004). *ApJ*, 617:746–764.
Daddi, E., Dickinson, M., Morrison, G., et al. (2007). *ApJ*, 670:156.
Dahlen, T., Mobasher, B., Faber, S. M., et al. (2013). *ApJ*, 775:93.
Danovich, M., Dekel, A., Hahn, O., Ceverino, D., and Primack, J. (2015). *MNRAS*, 449:2087–2111.
Danovich, M., Dekel, A., Hahn, O., and Teyssier, R. (2012). *MNRAS*, 422:1732–1749.
Darvish, B., Mobasher, B., Sobral, D., et al. (2016). *ApJ*, 825:113.
Darvish, B., Mobasher, B., Sobral, D., Scoville, N., and Aragon-Calvo, M. (2015). *ApJ*, 805:121.
Davidzon, I., Bolzonella, M., Coupon, J., et al. (2013). *A&A*, 558:A23.
Davidzon, I., Cucciati, O., Bolzonella, M., et al. (2016). *A&A*, 586:A23.
Davis, M., Efstathiou, G., Frenk, C. S., and White, S. D. M. (1985). *ApJ*, 292:371–394.
Davis, M. and Geller, M. J. (1976). *ApJ*, 208:13–19.
Davis, M., Huchra, J., Latham, D. W., and Tonry, J. (1982). *ApJ*, 253:423–445.
De Cia, A., Ledoux, C., Savaglio, S., Schady, P., and Vreeswijk, P. M. (2013). *A&A*, 560:A88.
de Vaucouleurs, G. (1959). *Handbuch der Physik*, 53:275.
Dekel, A. and Birnboim, Y. (2006). *MNRAS*, 368:2–20.
Dekel, A., Birnboim, Y., Engel, G., et al. (2009). *Nature*, 457:451–454.
Dekel, A. and Silk, J. (1986). *ApJ*, 303:39–55.

Devriendt, J. E. G., Guiderdoni, B., and Sadat, R. (1999). *A&A*, 350:381–398.
Djorgovski, S. G., Mahabal, A., Drake, A., Graham, M., and Donalek, C. (2013). *Sky Surveys*, page 223.
Dopita, M. A., Groves, B. A., Fischera, J., et al. (2005). *ApJ*, 619:755–778.
Doroshkevich, A. G. (1970). *Astrophysics*, 6:320–330.
Draine, B. T. (2003). *ARA&A*, 41:241–289.
Draine, B. T. (2004). Astrophysics of Dust in Cold Clouds. In Blain, A. W., Combes, F., Draine, B. T., Pfenniger, D., and Revaz, Y., editors, *The Cold Universe, Saas-Fee Advanced Course 32, Springer-Verlag, 308 pages, 129 figures, Lecture Notes 2002 of the Swiss Society for Astronomy and Astrophysics (SSAA), Springer, 2004. Edited by A.W. Blain, F. Combes, B.T. Draine, D. Pfenniger and Y. Revaz, ISBN 354040838x, p. 213*, page 213.
Draine, B. T. and Li, A. (2007). *ApJ*, 657:810–837.
Dressler, A. (1980). *ApJ*, 236:351–365.
Dreyer, J. L. E. (1888). *MNRAS*, 49:1.
Dubois, Y., Devriendt, J., Slyz, A., and Teyssier, R. (2010). *MNRAS*, 409:985–1001.
Dubois, Y., Devriendt, J., Slyz, A., and Teyssier, R. (2012a). *MNRAS*, 420:2662–2683.
Dubois, Y., Peirani, S., Pichon, C., et al. (2016). *ArXiv e-prints*.
Dubois, Y., Pichon, C., Devriendt, J., et al. (2013). *MNRAS*, 428:2885–2900.
Dubois, Y., Pichon, C., Haehnelt, M., et al. (2012b). *MNRAS*, 423:3616–3630.
Dubois, Y., Pichon, C., Welker, C., et al. (2014). *MNRAS*, 444:1453–1468.
Dubois, Y. and Teyssier, R. (2008). *A&A*, 477:79–94.
Dwek, E. (1998). *ApJ*, 501:643.
Einstein, A. (1915). *Sitzungsberichte der Königlich Preußischen Akademie der Wissenschaften (Berlin), Seite 844-847*.
Elbaz, D., Daddi, E., Le Borgne, D., et al. (2007). *A&A*, 468(1):33–48.
Elbaz, D., Dickinson, M., Hwang, H. S., et al. (2011). *A&A*, 533:A119.
Elvis, M., Civano, F., Vignali, C., et al. (2009). *ApJ Sup.*, 184:158–171.
Evans, C. J., Barbuy, B., Bonifacio, P., Chemla, F., and al. (2012). Multi-object spectroscopy with the European ELT: scientific synergies between EAGLE and EVE. In *Ground-based and Airborne Instrumentation for Astronomy IV*, volume 8446 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, page 84467K.
Faisst, A. L., Capak, P., Hsieh, B. C., et al. (2016). *ApJ*, 821:122.
Faucher-Giguère, C.-A. and Kereš, D. (2011). *MNRAS*, 412:L118–L122.
Felten, J. E. (1976). *ApJ*, 207:700–709.
Fioc, M. and Rocca-Volmerange, B. (1997). *A&A*, 326:950–962.
Fioc, M. and Rocca-Volmerange, B. (1999). *ArXiv Astrophysics e-prints*.
Fisher, D. B., Bolatto, A. D., Herrera-Camus, R., et al. (2014). *Nature*, 505:186–189.
Fitzpatrick, E. L. and Massa, D. (1986). *ApJ*, 307:286–294.
Fontana, A., Dunlop, J. S., Paris, D., et al. (2014). *A&A*, 570:A11.
Forero-Romero, J. E., Hoffman, Y., Gottlöber, S., Klypin, A., and Yepes, G. (2009). *MNRAS*, 396:1815–1824.
Gaibler, V., Khochfar, S., Krause, M., and Silk, J. (2012). *MNRAS*, 425:438–449.
Galametz, M., Madden, S. C., Galliano, F., et al. (2011). *A&A*, 532:A56.
Gallazzi, A. and Bell, E. F. (2009). *ApJ Sup.*, 185:253.
Gallazzi, A., Bell, E. F., Wolf, C., et al. (2009). *ApJ*, 690:1883–1900.
Garilli, B., Guzzo, L., Scodeggio, M., et al. (2014). *A&A*, 562:A23.
Gay, C., Pichon, C., Le Borgne, D., et al. (2010). *MNRAS*, 404:1801–1816.
Geller, M. J. and Huchra, J. P. (1989). *Science*, 246:897–903.
Giavalisco, M., Vanzella, E., Salimbeni, S., et al. (2011). *ApJ*, 743:95.
Girardi, L., Bressan, A., Bertelli, G., and Chiosi, C. (2000). *A&AS*, 141:371–383.
Grazian, A., Fontana, A., Santini, P., et al. (2015). *A&A*, 575:A96.
Griffin, M. J., Abergel, A., Abreu, A., et al. (2010). *A&A*, 518:L3.
Groth, E. J. and Peebles, P. J. E. (1977). *ApJ*, 217:385.
Groves, B., Dopita, M. A., Sutherland, R. S., et al. (2008). *ApJ Sup.*, 176:438–456.
Guidi, G., Scannapieco, C., Walcher, C. J., and Gallazzi, A. (2016). *ArXiv e-prints*.
Gunn, J. E. and Gott, III, J. R. (1972). *ApJ*, 176:1.
Guzzo, L., Scodeggio, M., Garilli, B., et al. (2014). *A&A*, 566:A108.
Haardt, F. and Madau, P. (1996). *ApJ*, 461:20–+.
Hahn, O., Angulo, R. E., and Abel, T. (2015). *MNRAS*, 454:3920–3937.
Hahn, O., Carollo, C. M., Porciani, C., and Dekel, A. (2007a). *MNRAS*, 381:41–51.
Hahn, O., Porciani, C., Carollo, C. M., and Dekel, A. (2007b). *MNRAS*, 375:489–499.
Hartmann, L., Ballesteros-Paredes, J., and Bergin, E. A. (2001). *ApJ*, 562:852–868.
Hasinger, G., Cappelluti, N., Brunner, H., et al. (2007). *ApJ Sup.*, 172:29–37.
Hayward, C. C. and Smith, D. J. B. (2015). *MNRAS*, 446:1512–1535.
Hearin, A. P., Behroozi, P. S., and van den Bosch, F. C. (2015). *ArXiv e-prints*.
Heinis, S., Milliard, B., Arnouts, S., et al. (2007). *ApJ Sup.*, 173:503–511.
Henriques, B. M. B., Thomas, P. A., Oliver, S., and Roseboom, I. (2009). *MNRAS*, 396:535–547.

Henriques, B. M. B., White, S. D. M., Thomas, P. A., et al. (2013). *MNRAS*, 431:3373–3395.
Herschel, W. (1785). *Philosophical Transactions of the Royal Society of London Series I*, 75:213–266.
Hildebrandt, H., Arnouts, S., Capak, P., et al. (2010). *A&A*, 523:A31.
Hildebrandt, H., Erben, T., Kuijken, K., et al. (2012). *MNRAS*, 421:2355–2367.
Hoaglin, D. C., Mosteller, F., and Tukey, J. W. (1983). *Understanding robust and exploratory data anlysis*.
Hobbs, A., Read, J., and Nicola, A. (2015). *MNRAS*, 452:3593–3609.
Holmberg, E. (1958). *Meddelanden fran Lunds Astronomiska Observatorium Serie II*, 136:1.
Hopkins, A. M. and Beacom, J. F. (2006). *ApJ*, 651:142–154.
Hopkins, P. F., Hernquist, L., Cox, T. J., et al. (2006). *ApJ Sup.*, 163:1–49.
Hopkins, P. F., Kereš, D., Murray, N., Quataert, E., and Hernquist, L. (2012). *MNRAS*, 427:968–978.
Hopkins, P. F., Kereš, D., Oñorbe, J., et al. (2014). *MNRAS*, 445:581–603.
Howarth, I. D. (1983). *MNRAS*, 203:301–304.
Hoyle, F. (1949). *Problems of Cosmical Aerodynamics, Central Air Documents, Office, Dayton, OH*. Central Air Documents Office, Dayton, OH.
Hsieh, B.-C., Wang, W.-H., Hsieh, C.-C., et al. (2012). *ApJ Sup.*, 203:23.
Hubble, E. (1925). *Contributions from the Mount Wilson Observatory / Carnegie Institution of Washington*, 304:1–25.
Hubble, E. (1929). *Proceedings of the National Academy of Science*, 15:168–173.
Hubble, E. (1934). *ApJ*, 79:8.
Hubble, E. P. (1926). *ApJ*, 64.
Hubble, E. P. (1936). *Realm of the Nebulae*.
Huchra, J., Jarrett, T., Skrutskie, M., et al. (2005). The 2MASS Redshift Survey and Low Galactic Latitude Large-Scale Structure. In Fairall, A. P. and Woudt, P. A., editors, *Nearby Large-Scale Structures and the Zone of Avoidance*, volume 329 of *Astronomical Society of the Pacific Conference Series*, page Fairall.
Huchra, J. P. and Geller, M. J. (1982). *ApJ*, 257:423–437.
Huertas-Company, M., Gravet, R., Cabrera-Vives, G., et al. (2015). *ApJ Sup.*, 221:8.
Hui, L., Gnedin, N. Y., and Zhang, Y. (1997). *ApJ*, 486:599–622.
Ilbert, O., Arnouts, S., Le Floc'h, E., et al. (2015). *A&A*, 579:A2.
Ilbert, O., Arnouts, S., McCracken, H. J., et al. (2006). *A&A*, 457(3):841–856.
Ilbert, O., Capak, P., Salvato, M., et al. (2009). *ApJ*, 690(2):1236–1249.
Ilbert, O., McCracken, H. J., Le Fèvre, O., et al. (2013). *A&A*, 556:A55.
Ilbert, O., Salvato, M., Le Floc'h, E., et al. (2010). *ApJ*, 709(2):644–663.
Ilbert, O., Tresse, L., Zucca, E., et al. (2005). *A&A*, 439:863–876.
Ivezic, Z., Tyson, J. A., Abel, B., et al. (2008). *ArXiv e-prints*.
Jeans, J. H. (1902). *Philosophical Transactions of the Royal Society of London Series A*, 199:1–53.
Jenkins, A., Frenk, C. S., Pearce, F. R., et al. (1998). *ApJ*, 499:20–40.
Jennings, E. (2012). *MNRAS*, 427:L25–L29.
Jonsson, P. (2006). *MNRAS*, 372:2–20.
Jonsson, P., Groves, B. A., and Cox, T. J. (2010). *MNRAS*, 403:17–44.
Jost, J., Peng, X., and Wang, G. (1995). Variational Aspects of the Seiberg-Witten Functional. In *eprint arXiv:dg-ga/9504003*.
Kacprzak, G. G., Churchill, C. W., Steidel, C. C., Spitler, L. R., and Holtzman, J. A. (2012). *MNRAS*, 427:3029–3043.
Kant, I. (1755). *Allgemeine Naturgeschichte und Theorie des Himmels*.
Karim, A., Schinnerer, E., Martínez-Sansigre, A., et al. (2011). *ApJ*, 730:61.
Katz, N., Keres, D., Dave, R., and Weinberg, D. H. (2003a). How Do Galaxies Get Their Gas? In Rosenberg, J. L. and Putman, M. E., editors, *The IGM/Galaxy Connection. The Distribution of Baryons at z=0*, volume 281 of *Astrophysics and Space Science Library*, page 185.
Katz, N., Keres, D., Dave, R., and Weinberg, D. H. (2003b). How Do Galaxies Get Their Gas? In J. L. Rosenberg & M. E. Putman, editor, *The IGM/Galaxy Connection. The Distribution of Baryons at z=0*, volume 281 of *Astrophysics and Space Science Library*, pages 185–191.
Kauffmann, G., Heckman, T. M., White, S. D. M., et al. (2003). *MNRAS*, 341:33–53.
Kauffmann, G., White, S. D. M., Heckman, T. M., et al. (2004). *MNRAS*, 353:713–731.
Kaviraj, S., Laigle, C., Kimm, T., et al. (2016). *ArXiv e-prints*.
Kaviraj, S., Schawinski, K., Devriendt, J. E. G., et al. (2007a). *ApJ Sup.*, 173:619–642.
Kaviraj, S., Sohn, S. T., O'Connell, R. W., et al. (2007b). *MNRAS*, 377:987–996.
Kawinwanichakij, L., Quadri, R. F., Papovich, C., et al. (2016). *ApJ*, 817:9.
Kennicutt, Jr., R. C. (1998). *ApJ*, 498:541–+.
Kereš, D., Katz, N., Weinberg, D. H., and Davé, R. (2005a). *MNRAS*, 363:2–28.
Kereš, D., Katz, N., Weinberg, D. H., and Davé, R. (2005b). *MNRAS*, 363:2–28.
Khandai, N., Di Matteo, T., Croft, R., et al. (2015). *MNRAS*, 450:1349–1374.
Kim, C.-G., Ostriker, E. C., and Kim, W.-T. (2013). *ApJ*, 776:1.
Kimm, T., Cen, R., Devriendt, J., Dubois, Y., and Slyz, A. (2015). *MNRAS*, 451:2900–2921.
Kimm, T., Devriendt, J., Slyz, A., et al. (2011a). *ArXiv e-prints*.
Kimm, T., Kaviraj, S., Devriendt, J. E. G., et al. (2012). *MNRAS*, 425:L96–L100.

Kimm, T., Slyz, A., Devriendt, J., and Pichon, C. (2011b). *MNRAS*, 413:L51–L55.
Klypin, A. A., Trujillo-Gomez, S., and Primack, J. (2011). *ApJ*, 740:102.
Knebe, A., Knollmann, S. R., Muldrew, S. I., et al. (2011). *MNRAS*, 415:2293–2318.
Kobayashi, C. and Nakasato, N. (2011). *ApJ*, 729:16.
Kochiashvili, I., Møller, P., Milvang-Jensen, B., et al. (2015). *A&A*, 580:A42.
Koekemoer, A. M., Aussel, H., Calzetti, D., et al. (2007). *ApJ Sup.*, 172(1):196–202.
Komatsu, E., Smith, K. M., Dunkley, J., and et al. (2011). *ApJ Sup.*, 192:18–+.
Koo, D. C. (1985). *AJ*, 90:418–440.
Kovač, K., Lilly, S. J., Knobel, C., et al. (2014). *MNRAS*, 438:717–738.
Kron, R. G. (1980). *ApJ Sup.*, 43:305–325.
Kroupa, P. (2008). The stellar initial mass function of metal-rich populations. In Israelian, G. and Meynet, G., editors, *The Metal-Rich Universe*, page 227.
Kroupa, P., Weidner, C., Pflamm-Altenburg, J., et al. (2013). *The Stellar and Sub-Stellar Initial Mass Function of Simple and Composite Populations*, page 115.
Kuncarayakti, H., Galbany, L., Anderson, J. P., Krühler, T., and Hamuy, M. (2016). *ArXiv e-prints*.
Lai, C.-C., Lin, L., Jian, H.-Y., et al. (2016). *ApJ*, 825:40.
Laigle, C., McCracken, H. J., Ilbert, O., et al. (2016). *ApJ Sup.*, 224:24.
Laigle, C., Pichon, C., Codis, S., et al. (2015). *MNRAS*, 446:2744–2759.
Landy, S. D. and Szalay, A. S. (1993). *ApJ*, 412:64.
Larson, R. B. (1974). *MNRAS*, 169:229–246.
Larson, R. B., Tinsley, B. M., and Caldwell, C. N. (1980). *ApJ*, 237:692.
Lawrence, A., Warren, S. J., Almaini, O., et al. (2007). *MNRAS*, 379(4):1599–1617.
Le Borgne, D., Rocca-Volmerange, B., Prugniel, P., et al. (2004). *A&A*, 425:881–897.
Le Floc'h, E., Aussel, H., Ilbert, O., et al. (2009). *ApJ*, 703:222–239.
Leauthaud, A., Massey, R., Kneib, J.-P., et al. (2007). *ApJ Sup.*, 172(1):219–238.
Lee, B., Giavalisco, M., Williams, C. C., Guo, Y., and et al. (2013). *ApJ*, 774(1):47.
Lee, K.-G., Hennawi, J. F., Stark, C., and Prochaska, e. a. (2014a). *ApJ*, 795:L12.
Lee, K.-G., Hennawi, J. F., White, M., Croft, R. A. C., and Ozbek, M. (2014b). *ApJ*, 788:49.
Lee, K.-S., Alberts, S., Atlee, D., et al. (2012). *ApJ Let.*, 758:L31.
Lee, N., Sanders, D. B., Casey, C. M., et al. (2015). *ApJ*, 801:80.
Leitherer, C., Ortiz Otálvaro, P. A., Bresolin, F., et al. (2010). *ApJ Sup.*, 189:309–335.
Leitherer, C., Robert, C., and Drissen, L. (1992). *ApJ*, 401:596–617.
Leitherer, C., Schaerer, D., Goldader, J. D., et al. (1999). *ApJ Sup.*, 123:3–40.
Lejeune, T., Cuisinier, F., and Buser, R. (1998). *A&AS*, 130:65–75.
Lejeune, T. and Schaerer, D. (2001). *A&A*, 366:538–546.
Lemaître, G. (1927). *Annales de la Société Scientifique de Bruxelles*, 47:49–59.
L'Huillier, B., Combes, F., and Semelin, B. (2012). *A&A*, 544:A68.
Libeskind, N. I., Hoffman, Y., Forero-Romero, J., et al. (2013a). *MNRAS*, 428:2489–2499.
Libeskind, N. I., Hoffman, Y., Steinmetz, M., et al. (2013b). *ApJ Let.*, 766:L15.
Lilly, S. J., Le Fèvre, O., Renzini, A., et al. (2007). *ApJ Sup.*, 172:70–85.
Lilly, S. J. and Longair, M. S. (1984). *MNRAS*, 211:833.
Lin, L., Capak, P. L., Laigle, C., et al. (2016). *ApJ*, 817:97.
Liu, H.-L., Wu, Y., Li, J., et al. (2015). *ApJ*, 798:30.
Loh, E. D. and Spillar, E. J. (1986). *ApJ*, 303:154–161.
Lutz, D., Poglitsch, A., Altieri, B., et al. (2011). *A&A*, 532:A90.
Maciejewski, M., Colombi, S., Springel, V., Alard, C., and Bouchet, F. R. (2009). *MNRAS*, 396:1329–1348.
Madau, P. and Dickinson, M. (2014). *ARA&A*, 52:415–486.
Madau, P., Ferguson, H. C., Dickinson, M. E., et al. (1996). *MNRAS*, 283:1388.
Magorrian, J., Tremaine, S., Richstone, D., et al. (1998). *AJ*, 115:2285–2305.
Maiolino, R., Nagao, T., Grazian, A., et al. (2008). *A&A*, 488:463–479.
Malavasi, N., Pozzetti, L., Cucciati, O., Bardelli, S., and Cimatti, A. (2016). *A&A*, 585:A116.
Maller, A. H. and Bullock, J. S. (2004). *MNRAS*, 355:694–712.
Maoz, D., Mannucci, F., and Brandt, T. D. (2012). *MNRAS*, 426:3282–3294.
Maraston, C. (2005). *MNRAS*, 362:799.
Maraston, C., Daddi, E., Renzini, A., et al. (2006). *ApJ*, 652:85–96.
Marchesi, S., Civano, F., Elvis, M., et al. (2016). *ApJ*, 817:34.
Marchesini, D., van Dokkum, P. G., Förster Schreiber, N. M., et al. (2009). *ApJ*, 701:1765.
Marigo, P. and Girardi, L. (2007). *A&A*, 469:239–263.
Marigo, P., Girardi, L., Bressan, A., et al. (2008). *A&A*, 482:883–905.
Martin, C. and GALEX Science Team (2003). The Galaxy Evolution Explorer (GALEX). In *American Astronomical Society Meeting Abstracts*, volume 35 of *Bulletin of the American Astronomical Society*, page 1363.

Martin, C., Moore, A., Morrissey, P., et al. (2010). The Keck Cosmic Web Imager. In *Ground-based and Airborne Instrumentation for Astronomy III*, volume 7735 of *Proc. SPIE*, page 77350M.
Martin, D. C., Matuszewski, M., Morrissey, P., et al. (2016). *ApJ Let.*, 824:L5.
Martínez, H. J., Muriel, H., and Coenda, V. (2016). *MNRAS*, 455:127–135.
Massey, R. and Refregier, A. (2005). *MNRAS*, 363:197–210.
Masters, D., Capak, P., Stern, D., et al. (2015). *ApJ*, 813:53.
Mathis, J. S. (1990). *ARA&A*, 28:37–70.
Matteucci, F. and Greggio, L. (1986). *A&A*, 154:279–287.
Matteucci, F. and Recchi, S. (2001). *ApJ*, 558:351–358.
Mattsson, L., Andersen, A. C., and Munkhammar, J. D. (2012). *MNRAS*, 423:26–37.
McCracken, H. J., Capak, P., Salvato, M., et al. (2010). *ApJ*, 708(1):202–217.
McCracken, H. J., Milvang-Jensen, B., Dunlop, J., et al. (2012). *A&A*, 544:A156.
McCracken, H. J., Peacock, J. A., Guzzo, L., et al. (2007). *ApJ Sup.*, 172(1):314–319.
McCracken, H. J., Wolk, M., Colombi, S., et al. (2015). *MNRAS*, 449:901–916.
McDonald, P., Miralda-Escudé, J., and Rauch, M. e. a. (2000). *ApJ*, 543:1–23.
McKinnon, R., Torrey, P., and Vogelsberger, M. (2016a). *MNRAS*, 457:3775–3800.
McKinnon, R., Torrey, P., Vogelsberger, M., Hayward, C. C., and Marinacci, F. (2016b). *ArXiv e-prints*.
Messier, C. (1781). Catalogue des Nébuleuses et des amas d'Étoiles (Catalog of Nebulae and Star Clusters). Technical report.
Milliard, B., Heinis, S., Blaizot, J., et al. (2007). *ApJ Sup.*, 173:494–502.
Miyazaki, S., Komiyama, Y., Nakaya, H., et al. (2012). Hyper Suprime-Cam. In *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, volume 8446 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, page 0.
Mo, H., van den Bosch, F. C., and White, S. (2010). *Galaxy Formation and Evolution*.
Mobasher, B., Dahlen, T., Ferguson, H. C., et al. (2015). *ApJ*, 808:101.
Moffat, A. F. J. (1969). *A&A*, 3:455.
Molino, A., Benítez, N., Moles, M., et al. (2014). *MNRAS*, 441:2891–2922.
Moore, B., Katz, N., Lake, G., Dressler, A., and Oemler, A. (1996). *Nature*, 379:613–616.
Morrissey, P., Matuszewski, M., Martin, C., et al. (2012). The Keck Cosmic Web Imager: a capable new integral field spectrograph for the W. M. Keck Observatory. In *Ground-based and Airborne Instrumentation for Astronomy IV*, volume 8446 of *Proc. SPIE*, page 844613.
Moster, B. P., Somerville, R. S., Newman, J. A., and Rix, H.-W. (2011). *ApJ*, 731:113.
Moultaka, J., Boisson, C., Joly, M., and Pelat, D. (2004). *A&A*, 420:459–466.
Moultaka, J. and Pelat, D. (2000). *MNRAS*, 314:409–419.
Moustakas, J., Coil, A. L., Aird, J., et al. (2013). *ApJ*, 767:50.
Moutard, T., Arnouts, S., Ilbert, O., et al. (2016). *A&A*, 590:A102.
Murray, N. (2011). *ApJ*, 729:133.
Neugebauer, G., Habing, H. J., van Duinen, R., et al. (1984). *ApJ Let.*, 278:L1–L6.
Noeske, K. G., Weiner, B. J., Faber, S. M., et al. (2007). *ApJ*, 660:L43.
Nomoto, K., Saio, H., Kato, M., and Hachisu, I. (2007). *ApJ*, 663:1269–1276.
Novikov, D., Colombi, S., and Doré, O. (2006). *MNRAS*, 366:1201–1216.
Obreschkow, D. and Glazebrook, K. (2014). *ApJ*, 784:26.
Ocvirk, P., Pichon, C., Lançon, A., and Thiébaut, E. (2006). *MNRAS*, 365:46–73.
Ocvirk, P., Pichon, C., and Teyssier, R. (2008). *MNRAS*, 390:1326–1338.
Oemler, Jr., A. (1974). *ApJ*, 194:1–20.
Oke, J. B. and Sandage, A. (1968). *ApJ*, 154:21.
Oliver, S. J., Bock, J., Altieri, B., et al. (2012). *MNRAS*, 424:1614–1635.
Onodera, M., Renzini, A., Carollo, M., et al. (2012). *ApJ*, 755:26.
Ostriker, E. C. (1999). *ApJ*, 513:252–258.
Ozbek, M., Croft, R. A. C., and Khandai, N. (2016). *MNRAS*, 456:3610–3623.
Palouš, J. (2014). *Astrophysics and Space Science Proceedings*, 36:181.
Papovich, C., Dickinson, M., and Ferguson, H. C. (2001). *ApJ*, 559:620–653.
Paz, D. J., Stasyszyn, F., and Padilla, N. D. (2008). *MNRAS*, 389:1127–1136.
Peebles, P. J. E. (1969). *ApJ*, 155:393–+.
Peebles, P. J. E. (1970). *AJ*, 75:13.
Peebles, P. J. E. (1975). *ApJ*, 196:647–651.
Peirani, S., Mohayaee, R., and de Freitas Pacheco, J. A. (2004). *MNRAS*, 348:921–931.
Peng, Y., Lilly, S. J., Renzini, A., and Carollo, M. (2011). *eprint arXiv*, 1106:2546.
Peng, Y.-j., Lilly, S. J., Kovač, K., et al. (2010). *ApJ*, 721(1):193–221.
Penzias, A. A. and Wilson, R. W. (1965). *ApJ*, 142:419–421.
Perryman, M. A. C., Lindegren, L., Kovalevsky, J., et al. (1995). *A&A*, 304.
Pforr, J., Maraston, C., and Tonini, C. (2012). *MNRAS*, 422:3285–3326.
Pichon, C. and Bernardeau, F. (1999). *A&A*, 343:663–681.
Pichon, C., Pogosyan, D., Kimm, T., et al. (2011). *MNRAS*, pages 1739–+.

Pichon, C., Thiébaut, E., Prunet, S., et al. (2010). *MNRAS*, 401:705–726.
Pichon, C., Vergely, J. L., Rollinde, E., Colombi, S., and Petitjean, P. (2001). *MNRAS*, 326:597–620.
Pickles, A. J. (1998). *PASP*, 110:863–878.
Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. (2014). *A&A*, 571:A16.
Poglitsch, A., Waelkens, C., Geis, N., et al. (2010). *A&A*, 518:L2.
Pogosyan, D., Bond, J. R., Kofman, L., and Wadsley, J. (1996). The Cosmic Web and Filaments in Cluster Patches. In *American Astronomical Society Meeting Abstracts*, volume 28 of *Bulletin of the American Astronomical Society*, page 1289.
Pogosyan, D., Bond, J. R., Kofman, L., and Wadsley, J. (1998). Cosmic Web: Origin and Observables. In S. Colombi, Y. Mellier, & B. Raban, editor, *Wide Field Surveys in Cosmology*, page 61.
Polletta, M., Tajer, M., Maraschi, L., et al. (2007). *ApJ*, 663:81.
Pozzetti, L., Bolzonella, M., Lamareille, F., et al. (2007). *A&A*, 474(2):443–459.
Pozzetti, L., Bolzonella, M., Zucca, E., et al. (2010). *A&A*, 523:A13.
Press, W. H. and Schechter, P. (1974). *ApJ*, 187:425–438.
Prevot, M. L., Lequeux, J., Prevot, L., Maurice, E., and Rocca-Volmerange, B. (1984). *Astronomy and Astrophysics (ISSN 0004-6361)*, 132:389.
Prugniel, P. and Soubiran, C. (2001). *A&A*, 369:1048–1057.
Prunet, S., Pichon, C., Aubert, D., et al. (2008). *ApJ Sup.*, 178:179–188.
Pueblas, S. and Scoccimarro, R. (2009). *Phys. Rev. D*, 80(4):043504.
Quilis, V., Moore, B., and Bower, R. (2000). *Science*, 288:1617–1620.
Rasera, Y. and Teyssier, R. (2006). *A&A*, 445:1–27.
Rees, M. J. and Ostriker, J. P. (1977). *MNRAS*, 179:541–559.
Ribaudo, J., Lehner, N., Howk, J. C., et al. (2011). *ApJ*, 743:207.
Roberts, M. S. (1963). *ARA&A*, 1:149.
Rodighiero, G., Daddi, E., Baronchelli, I., et al. (2011). *ApJ Let.*, 739:L40.
Roediger, J. C. and Courteau, S. (2015). *MNRAS*, 452:3209–3225.
Rosdahl, J. and Blaizot, J. (2012). *MNRAS*, 423:344–366.
Rosenfield, P., Marigo, P., Girardi, L., et al. (2014). *ApJ*, 790:22.
Salmon, B., Papovich, C., Finkelstein, S. L., et al. (2015). *ApJ*, 799:183.
Salomé, Q., Salomé, P., and Combes, F. (2015). *A&A*, 574:A34.
Salpeter, E. E. (1955). *ApJ*, 121:161.
Sandage, A. (1961). *The Hubble atlas of galaxies*.
Sanders, D. B., Salvato, M., Aussel, H., et al. (2007). *ApJ Sup.*, 172(1):86–98.
Santini, P., Fontana, A., Grazian, A., et al. (2012). *A&A*, 538:A33.
Scannapieco, E., Pichon, C., Aracil, B., et al. (2006). *MNRAS*, 365:615–637.
Schaefer, B. M. (2009). *International Journal of Modern Physics D*, 18:173–222.
Schaye, J., Crain, R. A., Bower, R. G., et al. (2015). *MNRAS*, 446:521–554.
Schechter, P. (1976). *ApJ*, 203:297–306.
Schinnerer, E., Carilli, C. L., Scoville, N. Z., et al. (2004). *AJ*, 128(5):1974–1989.
Schinnerer, E., Sargent, M. T., Bondi, M., et al. (2010). *ApJ Sup.*, 188:384–404.
Schinnerer, E., Smolčić, V., Carilli, C. L., et al. (2007). *ApJ Sup.*, 172:46–69.
Schlegel, D. J., Finkbeiner, D. P., and Davis, M. (1998). *ApJ*, 500:525.
Schmidt, M. (1968). *ApJ*, 151:393.
Scottez, V., Mellier, Y., Granett, B. R., et al. (2016). *ArXiv e-prints*.
Scoville, N., Abraham, R. G., Aussel, H., et al. (2007). *ApJ Sup.*, 172(1):38–45.
Scoville, N., Arnouts, S., Aussel, H., et al. (2013). *ApJ Sup.*, 206:3.
Seldner, M., Siebers, B., Groth, E. J., and Peebles, P. J. E. (1977). *AJ*, 82:249–256.
Shandarin, S., Habib, S., and Heitmann, K. (2012). *Phys. Rev. D*, 85(8):083005.
Shapley, H. and Ames, A. (1932). *Annals of Harvard College Observatory*, 88:41–76.
Shectman, S. A., Landy, S. D., Oemler, A., et al. (1996). *Astrophysical Journal v.470*, 470:172.
Shen, S., Madau, P., Aguirre, A., et al. (2012). *ApJ*, 760:50.
Sijacki, D., Springel, V., Di Matteo, T., and Hernquist, L. (2007). *MNRAS*, 380:877–900.
Silk, J. and Norman, C. (2009). *ApJ*, 700:262–275.
Silk, J. and Rees, M. J. (1998). *A&A*, 331:L1–L4.
Sirk, M. M., Vallerga, J. V., Finley, D. S., Jelinsky, P., and Malina, R. F. (1997). *The Astrophysical Journal Supplement Series*, 110(2):347.
Smith, N. (2014). *ARA&A*, 52:487–528.
Smolčić, V., Ciliegi, P., Jelić, V., et al. (2014). *MNRAS*, 443:2590–2598.
Sobral, D., Best, P. N., Smail, I., et al. (2011). *MNRAS*, 411:675–692.
Somerville, R. S. and Davé, R. (2015). *ARA&A*, 53:51–113.
Somerville, R. S., Lee, K., Ferguson, H. C., et al. (2004). *ApJ Let.*, 600:L171–L174.
Sorba, R. and Sawicki, M. (2015). *MNRAS*, 452:235–245.
Sousbie, T. (2011). *MNRAS*, 414:350–383.

Sousbie, T., Colombi, S., and Pichon, C. (2009). *MNRAS*, 393:457–477.
Sousbie, T., Pichon, C., Colombi, S., and Pogosyan, D. (2008). *MNRAS*, 383(4):1655–1670.
Sousbie, T., Pichon, C., and Kawahara, H. (2011). *MNRAS*, 414:384–403.
Spergel, D. N., Verde, L., Peiris, H. V., et al. (2003). *ApJ Sup.*, 148:175–194.
Springel, V. (2005). *MNRAS*, 364:1105–1134.
Springel, V. (2010). *MNRAS*, 401:791–851.
Springel, V., Yoshida, N., and White, S. D. M. (2001). *New Astronomy*, 6:79–117.
Stark, D. P., Ellis, R. S., Bunker, A., et al. (2009). *ApJ*, 697:1493–1511.
Steidel, C. C., Giavalisco, M., Pettini, M., Dickinson, M., and Adelberger, K. L. (1996). *Astrophysical Journal Letters*, 462:L17.
Steinhardt, C. L., Capak, P., Masters, D., and Speagle, J. S. (2016). *ApJ*, 824:21.
Steinhardt, C. L., Speagle, J. S., Capak, P., et al. (2014). *ApJ Let.*, 791:L25.
Stetson, P. B. (1987). *PASP*, 99:191–222.
Stewart, K., Maller, A., Oñorbe, J., et al. (2016). *ArXiv e-prints*.
Stewart, K. R., Kaufmann, T., Bullock, J. S., et al. (2011). *ApJ*, 738:39.
Strateva, I., Ivezić, Ž., Knapp, G. R., et al. (2001). *AJ*, 122:1861–1874.
Sutherland, R. S. and Dopita, M. A. (1993). *ApJ Sup.*, 88:253–327.
Sutherland, W. and Saunders, W. (1992). *MNRAS*, 259:413–420.
Szalay, A. S., Connolly, A. J., and Szokoly, G. P. (1999). *AJ*, 117:68.
Takada, M., Ellis, R. S., Chiba, M., et al. (2014a). *PASJ*, 66:R1.
Takada, M., Ellis, R. S., Chiba, M., and Greene, J. E. e. a. (2014b). *PASJ*, 66:R1.
Taniguchi, Y., Kajisawa, M., Kobayashi, M. A. R., et al. (2015). *PASJ*, 67:104.
Taniguchi, Y., Scoville, N., Murayama, T., et al. (2007). *ApJ Sup.*, 172:9–28.
Tasca, L. A. M., Le Fèvre, O., Hathi, N. P., et al. (2015). *A&A*, 581:A54.
Taylor, E. N., Hopkins, A. M., Baldry, I. K., et al. (2011). *MNRAS*, 418:1587–1620.
Tempel, E. and Libeskind, N. I. (2013). *ApJ Let.*, 775:L42.
Tempel, E., Stoica, R. S., and Saar, E. (2013). *MNRAS*, 428(2):1827–1836.
Teyssier, R. (2002). *A&A*, 385:337–364.
Teyssier, R., Pires, S., Prunet, S., et al. (2009). *A&A*, 497:335–341.
Tojeiro, R., Heavens, A. F., Jimenez, R., and Panter, B. (2007). *MNRAS*, 381:1252–1266.
Torrey, P., Vogelsberger, M., Genel, S., et al. (2014). *MNRAS*, 438:1985–2004.
Trenti, M. and Stiavelli, M. (2008). *ApJ*, 676:767–780.
Tweed, D., Devriendt, J., Blaizot, J., Colombi, S., and Slyz, A. (2009). *A&A*, 506:647–660.
van den Bergh, S. (1960). *ApJ*, 131:215.
Vassiliadis, E. and Wood, P. R. (1993). *ApJ*, 413:641–657.
Vogelsberger, M., Genel, S., Springel, V., et al. (2014). *Nature*, 509:177–182.
Vulcani, B., De Lucia, G., Poggianti, B. M., et al. (2014). *ApJ*, 788:57.
Wadsley, J. W., Stadel, J., and Quinn, T. (2004). *New Astronomy*, 9:137–158.
Wallerstein, G., Greenstein, J. L., Parker, R., Helfer, H. L., and Aller, L. H. (1963). *ApJ*, 137:280.
Wang, X., Szalay, A., Aragón-Calvo, M. A., Neyrinck, M. C., and Eyink, G. L. (2014). *ApJ*, 793:58.
Weingartner, J. C. and Draine, B. T. (2001). *ApJ*, 548:296–309.
Weinmann, S. M., Pasquali, A., Oppenheimer, B. D., et al. (2012). *MNRAS*, 426:2797–2812.
Weinmann, S. M., van den Bosch, F. C., Yang, X., and Mo, H. J. (2006). *MNRAS*, 366:2.
Welker, C., Devriendt, J., Dubois, Y., Pichon, C., and Peirani, S. (2014). *MNRAS*, 445:L46–L50.
Whitaker, K. E., van Dokkum, P. G., Brammer, G., and Franx, M. (2012). *ApJ Let.*, 754:L29.
White, S. D. M. (1984). *ApJ*, 286:38–41.
White, S. D. M. and Frenk, C. S. (1991). *ApJ*, 379:52.
White, S. D. M. and Rees, M. J. (1978). *Royal Astronomical Society*, 183:341.
White, S. D. M. and Rees, M. J. (1978). *MNRAS*, 183:341–358.
Whitworth, A. P. and Francis, N. (2002). *MNRAS*, 329:641–646.
Williams, R., Quadri, R., Franx, M., van Dokkum, P., and Labbé, I. (2009). *ApJ*, 691:1879.
Wilman, D. J., Zibetti, S., and Budavári, T. (2010). *MNRAS*, 406:1701–1720.
Woosley, S. E., Heger, A., and Weaver, T. A. (2002). *Reviews of Modern Physics*, 74:1015–1071.
Worthey, G. (1994). *ApJ Sup.*, 95:107–149.
Wright, T. (1750). *An original theory or new hypothesis of the universe : founded upon general phaenomena of the visible creation; and particularly the Via the laws of nature, and solving by mathematical principles : the Lactea ...compris'd in nine familiar letters from the author to his friendand : illustrated with upward of thirty graven and mezzotinto plates ...*
York, D. G., Adelman, J., Anderson, J. E., et al. (2000). *AJ*, 120:1579.
Zafar, T. and Watson, D. (2013). *A&A*, 560:A26.
Zahid, H. J., Yates, R. M., Kewley, L. J., and Kudritzki, R. P. (2013). *ApJ*, 763:92.
Zamojski, M. A., Schiminovich, D., Rich, R. M., et al. (2007). *ApJ Sup.*, 172:468–493.
Zel'dovich, Y. B. (1970). *A&A*, 5:84–89.
Zhang, Y., Yang, X., Faltenbacher, A., et al. (2009). *ApJ*, 706:747–761.

Zhang, Y., Yang, X., Wang, H., et al. (2013). *ApJ*, 779:160.

Appendices

Appendix A

Cosmic web reconstruction through tomography

One robust way to probe the circumgalactic and the intergalactic medium at intermediate redshifts is to take advantage of the hydrogen Lyman-α forest observed in the spectra of quasars or bright star-forming galaxies. Indeed the residual neutral hydrogen in the intergalactic medium closely follows the dark matter distribution, and analysing absorption lines in spectra of background objects allows us to map the three-dimensional distribution of matter, with a precision that scales with the number of objects. Hence tomography of the IGM/CGM is a unique tool to characterise in details gas inflows and how they affect galaxy properties. It is also a complementary way to reconstruct the cosmic web at high redshift (typically between 2 and 5) where reconstruction of the three-dimensional density is less easy from the galaxy distribution itself.

Although the theoretical method of cosmic tomography is already well defined and has been tested on simulations (Pichon et al., 2001; Caucci et al., 2008; Lee et al., 2014b), only few attempts of reconstruction have been applied to real datasets (Lee et al., 2014a; Cisewski et al., 2014; Ozbek et al., 2016). The line-of-sight resolution of the reconstruction depends on the spectral one, while the transverse resolution depends on the number of targeted objects. As obtaining spectra for a lot of background sources is very expensive in terms of telescope time, such reconstructions should be therefore optimised according to the pursued objectives. For instance inflows around galaxies or clusters can be inferred by ponctually observing at high resolution the spectra of background bright objects. On the contrary tracing the cosmic web at larger scales requires a more uniform distribution of background spectra. This can be obtained including in the background sample star forming galaxies in addition to quasars, the former having a more complex spectra and being fainter than the latter (Lee et al., 2014a). A reasonable balance has to be found between the required quality of spectra and the number of them.

The upcoming class of instruments on the next generation of telescopes will allow astronomers to probe the geometry of such anisotropic environment on galaxy, cluster and filament scales. In order to prepare the collection of data for future facilities such as E-ELT (Evans et al., 2012) or PFS (Takada et al., 2014b) with the purpose of tomography, I propose to carefully quantify the expected accuracy which will be reached by these probes relying on hydrodynamical simulations. The lightcone of the Horizon-AGN simulation presented in Chapter 4 should allow me to carry out such an investigation, as it mimics observational surveys, including realistic galaxy clustering and spectra. This Appendix aims at such a preliminary investigation. Section A.1 presents the Lyman-α forest modelling from the Horizon-AGN lightcone. Section A.2 explains the method of reconstruction using Wiener filtering and a first attempt in the Horizon-AGN lightcone using line-of-sight of star-forming galaxies selected in magnitude.

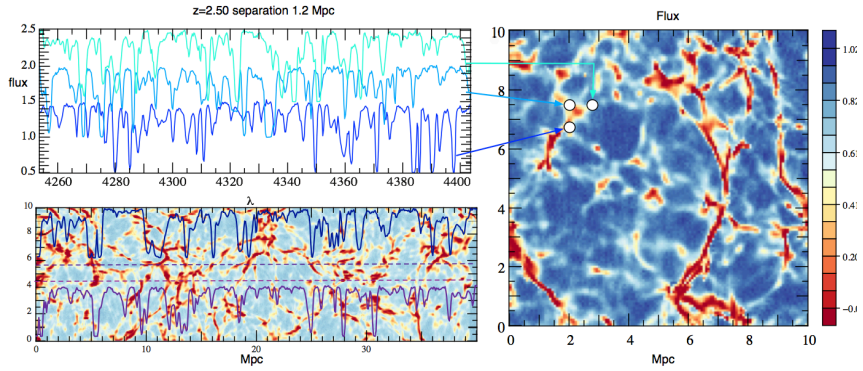

Figure A.1: The absorption lines on spectra of a triplet of background objects (*left*) should allow to infer the filamentary distribution of matter at the cluster scale (*right*).

A.1 Lyman-α forest modeling

The wavelength $\lambda_\alpha = 1215.7$ Å corresponds to the transition from the ground state to the first excited state of the hydrogen atom. Let us consider a background source, emitting at the wavelength λ_0 bluer than λ_α. The wavelength of the photons emitted by the background source and travelling toward us is redshifted by a factor of $(1+z)$. At some point the emitted light will be redshifted at λ_α. At this point, it may be absorbed by neutral hydrogen, which will cause an absorption line in the spectrum of the background source.
The probability of transmission of the light at the wavelength λ_0 (the "flux") will be given by:

$$F(\lambda_0) = e^{-\tau_\alpha(\lambda_0)}\,, \tag{A.1}$$

Where $\tau_\alpha(\lambda_0)$ is the Lyman-α optical depth at the observed wavelength λ_0. The optical depth depends on the neutral hydrogen density on the line-of-sight but also on the temperature and the peculiar velocity of the gas (see *e.g.* Hui et al., 1997; Scannapieco et al., 2006, for the full detail of the calculation). The flux is equal to 1 when there is no absorption and to 0 when all the background light is absorbed.
The neutral hydrogen fraction $x_{\rm HI} = n_{\rm HI}/n_{\rm H}$ is a balance between photo-ionization, collisional ionization and recombination. At equilibrium with the cosmic UV background field, we have:

$$\alpha(T) n_{\rm e}(1-x_{\rm HI}) = \gamma(T) n_{\rm e} x_{\rm HI} + \Gamma x_{\rm HI}\,, \tag{A.2}$$

where α and γ are the collisional recombination and ionisation rates, Γ is the photoionisation rate, and n_e is the free electron number density. Considering a uniform radiation field as implemented in the simulation and assuming that the chemical composition of the IGM is close to primordial (the radiative processes involve thus only hydrogen and helium) one can compute the neutral hydrogen density.
To compute numerically the transmitted flux from the Horizon-AGN lightcone, I work in comoving space. I extract the baryon density, temperature and radial velocity of the gas along the line-of-sight of each background source. I compute the neutral hydrogen density along the line-of-sight, the Ly-α optical depth and then the transmitted flux following Equation A.1. Figure A.1 shows the spectrum along three lines-of-sight through the hydrodynamical ligthcone.
The tomographic method allows me to reconstruct the gas density in three dimensions using the transmitted flux as a proxy of the density on the line-of-sight.

Assessing the realism of the Lyman-α modelling in the Horizon-AGN lightcone can be done by comparing flux statistics with real observations. The simplest statistic is the PDF of the transmitted flux which reflects the amplitude of the density fluctuation. The two-point correlation function of the one-dimensional transmitted flux allows to probe in addition the spatial clustering of the density fluctuations.

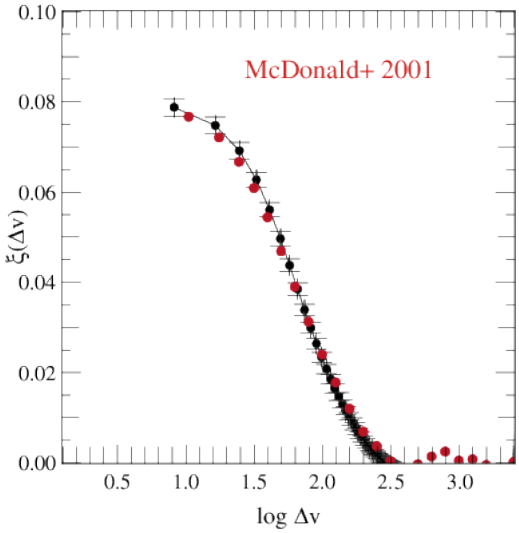

Figure A.2: Comparison of the correlation function of the transmitted flux in the Horizon-AGN lightcone as a function of velocity separation along the line-of-sight with McDonald et al. (2000) at $z = 2.4$. The correlation function is defined as $\xi(\Delta v) = \langle \delta F(v) \delta F(v + \Delta v) \rangle$ where $\delta F = F/\overline{F} - 1$ is the transmitted flux fraction perturbation.

On Figure A.2, I compare the correlation function of the simulated transmitted flux in Horizon-AGN with observational data from McDonald et al. (2000) and find a good agreement.

A.2 Tomographic reconstruction of the cosmic web in Horizon-AGN

Having obtained the transmitted flux on the lines-of-sight of background sources, the tomographic method interpolates between them to infer the three-dimensional density. I use Wiener filtering for this interpolation (Pichon et al., 2001).
Let us assume that $\mathbf{D}$ is a one-dimensional array representing the dataset and $\mathbf{M}$ is the three-dimensional array of the density that has to be estimated from the data.

$$\mathbf{M} = \mathbf{C}_{\delta^{3\mathrm{d}}\delta} (\mathbf{C}_{\delta\delta} + \mathbf{N})^{-1} \mathbf{D} \,, \tag{A.3}$$

Where $\mathbf{C}_{\delta^{3\mathrm{d}}\delta}$ is the mixed parameter-data covariance matrix, and $\mathbf{C}_{\delta\delta}$ is the data covariance matrix. I assume that the noise is uncorrelated. The noise covariance matrix can therefore be expressed as $\mathbf{N} = n^2 \mathbf{I}$. We assume that covariances have a gaussian shape and can both be expressed as:

$$\mathbf{C}_{\delta\delta}(x_1, x_2, \boldsymbol{x}_{1\mathrm{T}}, \boldsymbol{x}_{2\mathrm{T}}) = \sigma^2 e^{-\frac{|x_1 - x_2|^2}{2L_x^2}} e^{-\frac{|\boldsymbol{x}_{1T} - \boldsymbol{x}_{2T}|^2}{2L_\mathrm{T}^2}} \,, \tag{A.4}$$

where $(x_i, \boldsymbol{x}_{iT})$ are the coordinates of the points along and perpendicular to the line-of-sight.
The reconstruction depends therefore on 3 tuning parameters: the typical signal-to-noise ratio σ/n, and the correlation lengths along L_x and perpendicular L_T to the line-of-sight. The transverse correlation length L_T is naturally taken as the mean inter-LOS distance. The longitudinal correlation length is taken as the comoving scale corresponding roughly to the spectral resolution. The signal-to-noise ratio is estimated from the simulation. Due to the large volume of the Horizon-AGN ligthcone, the box is partitioned in sub-boxes and the reconstruction is performed on each block individually. Blocks are overlapping (over $\sim 2L_\mathrm{T}$) to avoid edge effects. The prior insures continuity of the reconstructed field.

To estimate the resolution which can be reached with a magnitude-limited sample of background galaxies, I use the virtual photometric catalog computed on the Horizon-AGN lightcone in Chapter 4. I focus on the reconstruction between redshift 2 and 2.5 and I consider two magnitude-limited samples selected in the observed r-band. The observed r-band traces indeed the rest-frame UV light below λ_α in this

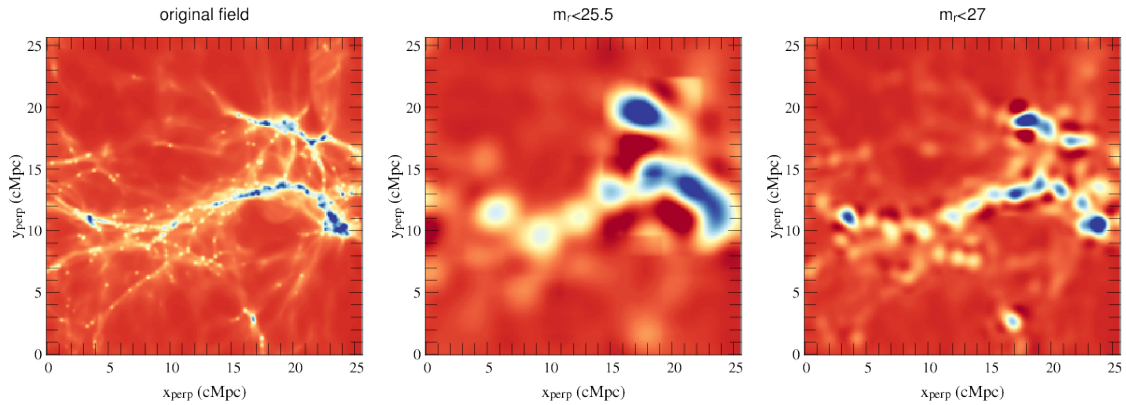

Figure A.3.: Visualisation of the reconstruction at $2 < z < 2.07$. From *left* to *right*: original field (no smoothing), reconstructed field with $r_{\rm lim} < 25.5$, and with $r_{\rm lim} < 27$.

redshift range which is the part of the spectrum which will contain the Lyman-α forest. The first thresold, $r_{\rm lim} < 25.5$, corresponds to a realistic threshold which could be reached by future probe such as E-ELT. The second threshold, $r_{\rm lim} < 27$, is more optimistic. Figure A.3 displays the transverse cross sections of the original and reconstructed density fields for these different thresholds at redshift $z \sim 2$. By definition, the reconstructed field is smoothed over a volume of size $L_x \times L_{\rm T} \times L_{\rm T}$. Note that in this reconstruction, I used only star-forming galaxies. Hovewer, quasars should be added to the sample for all the objects at $r < 22$ at $z \sim 2.3$ are essentially quasars (Lee et al., 2014b).

As shown in this Figure, the reconstruction of the density field at the E-ELT magnitude threshold will allow to trace the filaments at large scales but cannot reach the scales small enough to probe the transition region between the galactic halos and the cosmic web (on scales $\sim$2 arcmin). A complementary approach would be to select triplets of background sources (typically separated by less than $\sim$ 3 arcmin in the sky) and to reconstruct the density in the volume delimited by these lines-of-sight in the purpose to constrain the geometry of the filaments flowing into the clusters.

Alternatively, we could improve the tomographic reconstruction using as a prior density maps computed from photometric redshift distribution (see Chapter 8). Combining the high sampling of photometric surveys with the high longitudinal resolution from the Lyman-α forest in the spectra of background sources should help to reconstruct the three-dimensional density of matter with more accuracy.

Appendix B

Other publications in collaboration

B.1 Intrinsic alignments in the HORIZON-AGN simulation

The results presented in Chapter 7 about galaxy angular momentum alignment with the filaments have implications for weak lensing studies. Quantifying the contamination of these intrinsic alignments to the lensing signal is essential for the accuracy of future dark energy surveys, such as Euclid, LSST, WFIRST. In these two papers, we have used the HORIZON-AGN simulation to quantify the signal and its evolution as a function of redshift and luminosity. My contribution to the work has been to provide realistic galaxy fluxes from the virtual photometric catalog presented in Chapter 4.

Redshift and luminosity evolution of the intrinsic alignments of galaxies in HORIZON-AGN,
Chisari, N. E., Laigle, C., Codis, S., et al. (2016). ArXiv e-prints. Accepted to MNRAS.

Abstract – Intrinsic galaxy shape and angular momentum alignments can arise in cosmological large-scale structure due to tidal interactions or galaxy formation processes. Cosmological hydrodynamical simulations have recently come of age as a tool to study these alignments and their contamination to weak gravitational lensing. We probe the redshift and luminosity evolution of intrinsic alignments in HORIZON-AGN between $z = 0$ and $z = 3$ for galaxies with an r-band absolute magnitude of < -20. Alignments transition from being radial at low redshifts and high luminosities, dominated by the contribution of ellipticals, to being tangential at high redshift and low luminosities, where discs dominate the signal. This cannot be explained by the evolution of the fraction of ellipticals and discs alone: intrinsic evolution in the amplitude of alignments is necessary. The alignment amplitude of elliptical galaxies alone is smaller in amplitude by a factor of ~ 2, but has similar luminosity and redshift evolution as in current observations and in the nonlinear tidal alignment model at projected separations of > 1 Mpc. Alignments of discs are null in projection and consistent with current low redshift observations. The combination of the two populations yields an overall amplitude a factor of ~ 4 lower than observed alignments of luminous red galaxies with a steeper luminosity dependence. The restriction on accurate galaxy shapes implies that the galaxy population in the simulation is complete only to an r-band absolute magnitude of < -20. Higher resolution simulations will be necessary to avoid extrapolation of the intrinsic alignment predictions to the range of luminosities probed by future surveys.

Intrinsic alignments of galaxies in the HORIZON-AGN cosmological hydrodynamical simulation
Chisari, N. E., Codis, S., Laigle, C., et al. (2015). MNRAS, 454:2736-2753

Abstract – The intrinsic alignments of galaxies are recognized as a contaminant to weak gravitational lensing measurements. In this work, we study the alignment of galaxy shapes and spins at low redshift ($z \sim 0.5$) in Horizon-AGN, an adaptive-mesh-refinement hydrodynamical cosmological simulation box of 100 h^{-1} Mpc a side with AGN feedback implementation. We find that spheroidal galaxies in the simulation show a tendency to be aligned radially towards overdensities in the dark matter density field and other spheroidals. This trend is in agreement with observations, but the amplitude of the signal depends strongly on how shapes are measured and how galaxies are selected in the simulation. Disc galaxies show a tendency to be oriented tangentially around spheroidals in three dimensions. While this signal seems suppressed in projection, this does not guarantee that disc alignments can be safely ignored in future weak lensing surveys. The shape alignments of luminous galaxies in Horizon-AGN are in agreement with observations and other simulation works, but we find less alignment for lower luminosity populations. We also characterize the systematics of galaxy shapes in the simulation and show that they can be safely neglected when measuring the correlation of the density field and galaxy ellipticities.

B.2 Galaxy evolution study in COSMOS

A Coherent Study of Emission Lines from Broadband Photometry: Specific Star Formation Rates and [Oiii] /Hb Ratio at $3 > z > 6$ Faisst, A. L., Capak, P., Hsieh, B. C., Laigle, C. et al. (2016). ApJ, 821:122

In this work we make use of the accurate broad band photometry of the cosmos2015 catalog presented in Chapter 5 to study emission line properties up to redshift $z \sim 6$ and deduce the specific star formation rate. My contribution to this work has been to provide the photometric measurements.

Abstract – We measure the Hα and [Oiii] emission line properties as well as specific star formation rates (sSFRs) of spectroscopically confirmed $3 < z < 6$ galaxies in COSMOS from their observed colors versus redshift evolution. Our model describes consistently the ensemble of galaxies including intrinsic properties (age, metallicity, star formation history), dust attenuation, and optical emission lines. We forward-model the measured Hα equivalent widths (EW) to obtain the sSFR out to $z \sim 6$ without stellar mass fitting. We find a strongly increasing rest-frame Hα EW that is flattening off above $z \sim 2.5$ with average EWs of 300-600 Å at $z \sim 6$. The sSFR is increasing proportionally to $(1+z)^{2.4}$ at $z < 2.2$ and to $(1+z)^{1.5}$ at higher redshifts, indicative of a fast build-up of mass in high-z galaxies within e-folding times of 100-200 Myr at $z \sim 6$. The redshift evolution at $z > 3$ cannot be fully explained in a picture of growth driven by cold accretion. We find a progressively increasing [Oiii] $\lambda 5007/H\beta$ ratio out to $z \sim 6$, consistent with the ratios in local galaxies selected by increasing Hα EW (i.e., sSFR). This demonstrates the potential of using "local high-z analogs" to investigate the spectroscopic properties and relations of galaxies in the reionization epoch.

The SPLASH Survey: Quiescent Galaxies Are More Strongly Clustered but Are Not Necessarily Located in High-density Environments Lin, L., Capak, P. L., Laigle, C. (2016). ApJ:917,97

In this paper we use the cosmos2015 catalog to study the environments of galaxies via galaxy density and clustering analysis up to $z \sim 2.5$. My contribution to this work has been to provide the photometric measurements.

Abstract – We use the stellar-mass-selected catalog from the Spitzer Large Area Survey with Hyper-Suprime-Cam (SPLASH) in the COSMOS field to study the environments of galaxies via galaxy density and clustering analyses up to $z \sim 2.5$. The clustering strength of quiescent galaxies exceeds that of star-forming galaxies, implying that quiescent galaxies are preferentially located in more massive halos. When

using local density measurement, we find a clear positive quiescent fraction-density relation at $z < 1$, consistent with earlier results. However, the quiescent fraction-density relation reverses its trend at intermediate redshifts ($1 < z < 1.5$) with marginal significance ($< 1.8\sigma$) and is found to be scale dependent (1.6σ). The lower fraction of quiescent galaxies seen in large-scale dense environments, if confirmed to be true, may be associated with the fact that the star formation can be more easily sustained via cold stream accretion in "large-scale" high-density regions, preventing galaxies from permanent quenching. Finally, at $z > 1.5$, the quiescent fraction depends little on the local density, even though clustering shows that quiescent galaxies are in more massive halos. We argue that at high redshift the typical halo size falls below 10^13, where intrinsically the local density measurements are so varied that they do not trace the halo mass. Our results thus suggest that in the high-redshift universe, halo mass may be the key in quenching the star formation in galaxies, rather than the conventionally measured galaxy density.

Spitzer Bright, UltraVISTA Faint Sources in COSMOS: The Contribution to the Overall Population of Massive Galaxies at $z = 3 - 7$ Caputi, K. I., Ilbert, O., Laigle, C. (2015). ApJ:810,73

In this paper, we estimated the birth time of primordial massive galaxies, by measuring the galaxy stellar mass function at redshift above four on a sufficiently large area for the first time. This has been made possible thanks to a careful photometry extraction and photometric redshift computation on COSMOS. This paper does not rely on the COSMOS2015 catalog but on previous photometric measurements. My contribution to the work has been to confirm the redshift of some of the detected sources by cross-correlation with my COSMOS2015 catalog. This paper has been presented in a ESO press release.

Abstract – We have analyzed a sample of 574 Spitzer 4.5 μm selected galaxies with $[4.5] < 23$ and $K_s^{auto} > 24$ (AB) over the UltraVISTA ultradeep COSMOS field. Our aim is to investigate whether these mid-infrared (mid-IR) bright, near-infrared (near-IR) faint sources contribute significantly to the overall population of massive galaxies at redshifts $z > 3$. By performing a spectral energy distribution (SED) analysis using up to 30 photometric bands, we have determined that the redshift distribution of our sample peaks at redshifts $z \sim 2.5 - 3.0$, and $\sim 32\%$ of the galaxies lie at $z > 3$. We have studied the contribution of these sources to the galaxy stellar mass function (GSMF) at high redshifts. We found that the $[4.5] < 23$, $K_s^{auto} > 24$ galaxies produce a negligible change to the GSMF previously determined for $K_s^{auto} < 24$ sources at $3 < z < 4$, but their contribution is more important at $4 < z < 5$, accounting for >50% of the galaxies with stellar masses $M_* > 6 \times 10^10 M_\odot$. We also constrained the GSMF at the highest-mass end ($M_* > 2 \times 10^{11} M_\odot$) at $z > 5$. From their presence at $5 < z < 6$ and virtual absence at higher redshifts, we can pinpoint quite precisely the moment of appearance of the first most massive galaxies as taking place in the ~ 0.2 Gyr of elapsed time between $z \sim 6$ and $z \sim 5$. Alternatively, if very massive galaxies existed earlier in cosmic time, they should have been significantly dust-obscured to lie beyond the detection limits of current, large-area, deep near-IR surveys.

B.3 Cluster mass-richness relation in CFHTLS

CFHTLenS: a weak lensing shear analysis of the 3D-Matched-Filter galaxy clusters
Ford, J., Van Waerbeke, L., Milkeraitis, M.,Laigle, C., et al. (2015). MNRAS:447,1304

In this paper we present the cluster mass-richness relation from weak lensing shear analysis. I helped for the computing of the cluster richness.

Abstract – We present the cluster mass-richness scaling relation calibrated by a weak lensing analysis of > 18 000 galaxy cluster candidates in the Canada-France-Hawaii Telescope Lensing Survey (CFHTLenS).

Detected using the 3D-Matched-Filter (MF) cluster-finder of Milkeraitis et al., these cluster candidates span a wide range of masses, from the small group scale up to $\sim 10^{15}\mathrm{M}_{\odot}$, and redshifts $0.2 < z < 0.9$. The total significance of the stacked shear measurement amounts to 54σ. We compare cluster masses determined using weak lensing shear and magnification, finding the measurements in individual richness bins to yield 1σ compatibility, but with magnification estimates biased low. This first direct mass comparison yields important insights for improving the systematics handling of future lensing magnification work. In addition, we confirm analyses that suggest cluster miscentring has an important effect on the observed 3D-MF halo profiles, and we quantify this by fitting for projected cluster centroid offsets, which are typically ~ 0.4 arcmin. We bin the cluster candidates as a function of redshift, finding similar cluster masses and richness across the full range up to $z \sim 0.9$. We measure the 3D-MF mass-richness scaling relation $M_200 = M_0(N_{200}/20)^{\beta}$. We find a normalization $M_0 \sim (2.7^{+0.5}_{-0.4}) \times 10^{13}\mathrm{M}_{\odot}$, and a logarithmic slope of $\sim 1.4 \pm 0.1$, both of which are in 1σ agreement with results from the magnification analysis. We find no evidence for a redshift dependence of the normalization. The CFHTLenS 3D-MF cluster catalogue is now available at cfhtlens.org.

B.4 Other publications

The clustering properties of radio-selected AGN and star-forming galaxies up to redshifts z∼3
Magliocchetti, M., Popesso, P., Brusa, M., Salvato, M., Laigle, C., et al. (2016). ArXiv e-prints. Accepted to MNRAS.

The Chandra COSMOS Legacy survey: optical/IR identifications Marchesi, S., Civano, F., Elvis, M., Salvato, M., Brusa, M., Comastri, A., Gilli, R., Hasinger, G., Lanzuisi, G., Miyaji, T., Treister, E., Urry, C. M., Vignali, C., Zamorani, G., Allevato, V., Cappelluti, N., Cardamone, C., Finoguenov, A., Griffiths, R. E., Karim, A., Laigle, C., et al. (2016). ApJ:817:34

Rest-UV Absorption Lines as Metallicity Estimator: The Metal Content of Star-forming Galaxies at z ∼ 5
Faisst, A. L., Capak, P. L., Davidzon, I., Salvato, M., Laigle, C., et al. (2016). ApJ, 822:29

ISM Masses and the Star formation Law at $z = 1$ to 6: ALMA Observations of Dust Continuum in 145 Galaxies in the COSMOS Survey Field
Scoville, N., Sheth, K., Aussel, H., Vanden Bout, P., Capak, P., Bongiorno, A., Casey, C. M., Murchikova, L., Koda, J., Alvarez-Marquez, J., Lee, N., Laigle, C. et al. (2016). ApJ,820:83

(Sub)millimetre interferometric imaging of a sample of COSMOS/AzTEC submillimetre galaxies III. Environments Smolcic, V., Miettinen, O., Tomicic, N., Zamorani, G., Finoguenov, A., Lemaux, B. C., Aravena, M., Capak, P., Chiang, Y-K., Civano, F., Delvecchio, I., Ilbert, O., Jurlin, N., Karim, A., Laigle, C., et al. (2015). arXiv e-prints.

Can we use weak lensing to measure total mass profiles of galaxies on 20 kpc scales?
Kobayashi, M. I. N., Leauthaud, Al., More, S., Okabe, N., Laigle, C. et al. (2015). MNRAS:449:2128

Physical properties of $z > 4$ submillimeter galaxies in the COSMOS field
Smolcic, V., Karim, A., Miettinen, O., Novak, M., Magnelli, B., Riechers, D. A., Schinnerer, E., Capak, P., Bondi, M., Ciliegi, P., Aravena, M., Bertoldi, F., Bourke, S., Banfield, J., Carilli, C. L., Civano, F., Ilbert, O., Intema, H. T., Le Fèvre, O., Finoguenov, A., Hallinan, G., Klockner, H.-R., Koekemoer, A., Laigle, C, et al. (2015). AAP:576A:127S

Cluster magnification and the mass-richness relation in CFHTLenS
Ford, J., Hildebrandt, H., Van Waerbeke, L., Erben, T., Laigle, C., et al. (2014). MNRAS,439:3755