Magnifying the galaxy-halo connection in the Euclid Deep fields

Marko Shuntov

To cite this version:

Marko Shuntov. Magnifying the galaxy-halo connection in the Euclid Deep fields. Astrophysics [astro-ph]. Sorbonne Université, 2022. English. NNT : 2022SORUS549 . tel-04107442

HAL Id: tel-04107442
https://theses.hal.science/tel-04107442v1

Submitted on 26 May 2023

Magnifying the galaxy-halo connection in the Euclid Deep fields

---

Sorbonne Université
Institut d'Astrohysique de Paris
École Doctorale Astronomie et Astrophysique d'Ile de France 127

Presentée par :

Marko SHUNTOV

Pour obtenir le grade de
DOCTEUR DE SORBONNE UNIVERSITÉ
Spécialité doctorale : Astronomie & Astrophysique

Soutenu le 17 novembre 2022
Devant le cury composé de

M. Emanuele Daddi	Président
M. David Weinberg	Rapporteur
Mme Rozer Pelló	Rapporteure
Mme Micol Bolzonella	Examinatrice
M. Jean-Paul Kneib	Examinateur
M. Henry Joy McCracken	Directeur de thèse
M. Raphaël Gavazzi	Co-directeur de thèse

Paris, France
November 2022

Magnifying the galaxy-halo connection in the Euclid Deep fields, © October 2022

Author:
Marko Shuntov

Supervisors:
Dr. Henry Joy McCracken
Dr. Raphaël Gavazzi

Institute:
Institut d’Astrophysique de Paris, Paris, France

CONTENTS

ABSTRACT

The aim of this Thesis is to study the intricate relation between galaxies and dark matter halos in near-infrared surveys within the context of the *Euclid* mission. In particular, it focuses on studying the relationship between galaxy stellar mass and halo mass (called the stellar-to-halo mass relation, or SHMR) throughout cosmic history. The SHMR, which also expresses the star-formation efficiency of a galaxy integrated over the halo's lifetime, is key to understanding how different physical processes such as mergers, gas accretion, cold streams, supernovae and AGN feedback, shape galaxy properties. While various aspects of the SHMR have been extensively studied in the literature, the conclusions are hampered by heterogeneous data sets and there is still no consensus on its evolution with redshift throughout the majority of cosmic history, as well as what is the contribution from central and satellite galaxies.

This Thesis presents new measurements of the cosmic evolution of the SHMR, divided by central and satellite galaxies, since $z \sim 5$. The novelty of this work is the use of a single dataset – the COSMOS2020 catalog – to consistently measure the SHMR to $z \sim 5$. COSMOS2020 is a near-infrared selected catalog with photometry in over 30 bands complete down to low stellar mass limit and with accurate photometric redshifts. This was, in part, made possible by the deeper *Spitzer*/IRAC photometry that was extracted during this Thesis for the COSMOS2020 catalog. As such, the catalog allows measurements of galaxy clustering and abundance for complete samples in ten redshift bins. These measurements are described using a phenomenological model that parametrizes the SHMR for central galaxies and the number of galaxies inside each halo. The resulting SHMR in ten redshift bins out to $z = 5.5$ for both central and satellite galaxies is compared against hydrodynamical simulations to disentangle the effects of various physical mechanisms.

This Thesis also investigates the prospects of an additional weak lensing probe based on the magnification bias in the context of the upcoming deep survey of the *Euclid* mission. Magnification bias refers to the modification of the number densities of background sources due to lensing from foreground matter. As such, it does not require shape measurements, which means it can be used at higher redshifts than shear-based probes. Since COSMOS will be part of the Euclid Deep and Auxiliary fields, the COSMOS2020 catalog is used in this first study. Magnification is measured with the two-point angular cross-correlation function between stellar mass-selected foreground samples at $z < 1.5$, and apparent magnitude selected background samples at $2.7 < z < 4$. These measurements are fitted with a model based on the halo occupation distribution (HOD) that parametrizes the SHMR. Magnification can provide complementary constraints on HOD parameters, but is limited by its low signal-to-noise in COSMOS.

A second study of magnification relies on the Euclid FLAGSHIP2.0 simulation. Magnification is measured with the two-point angular cross-correlation for stellar mass-selected

foreground (at $z < 2$) and magnitude selected background samples (at $2.5 < z < 2.9$). Simulations of photometric errors and estimations of photometric redshifts are used to investigate the impact of systematic errors from clustering in the magnification measurements. This analysis shows that the Euclid Deep Survey's near-infrared data, combined with ground-based optical data from the Cosmic Dawn Survey, can provide photometric redshifts with high enough quality to control the systematic errors to measure unbiased magnification. However, the sample selection will have to be done carefully to minimize outliers in the true redshift vs. photo-z space. This analysis focuses on the redshift regime ($z > 1$) in which magnification can successfully be used to probe halo masses, where shear-based probes will have unacceptably large uncertainties.

Finally, this Thesis carries out a search of galaxy clusters at $z > 3$ using data from the first long-wavelength wide-field survey of the Euclid Deep fields. This is done using infrared Spitzer catalogs, as part of the Cosmic Dawn Survey, prepared during the Thesis. Clusters at $z > 3$ are rarely found in the literature, and their contribution to the cosmic star-formation rate density as well as their role in shaping galaxy properties remains poorly understood. Using this homogeneous wide-field data in the infrared, Cluster candidates were identified by searching for galaxy overdensities with red *Spitzer* colors that correspond to red peaks in *Herschel* maps. Coupled with NOEMA radio observations in the 3 mm band, this project reveals one of the most distant star-bursting protoclusters at $z = 4.2$ by detecting a CO(4-3) and a tentative $H_2O(4_{2,3} - 3_{3,0})$ emission line. This protocluster is a subject of follow-up proposal for NOEMA observations in the 2 mm band, to detect more lines and cluster members and characterize in detail the protocluster.

RÉSUMÉ

L'objectif de cette thèse est d'étudier la relation complexe entre les galaxies et les halos de matière noire dans le cadre des relevés en proche infrarouge de la mission *Euclid*. En particulier, elle se concentre sur l'étude de la relation entre la masse stellaire des galaxies et la masse des halos (appelée relation masse stellaire-masse halo, ou en anglais stellar-to-halo mass relation – SHMR) à travers l'histoire cosmique. La SHMR, qui exprime également l'efficacité de formation d'étoiles d'une galaxie intégrée sur la durée de vie du halo, est essentielle pour comprendre comment différents processus physiques tels que les fusions galactiques, l'accrétion de gaz, les flux froids, les supernovae et la rétroaction des noyaux actifs de galaxies (en anglais active galactic nucleii AGN) façonnent les propriétés des galaxies. Bien que divers aspects de la SHMR aient été largement étudiés, les conclusions dans la littérature divergent. En effet, la large variété de sets de données de sources différents arrive à de différentes conclusions en ce qui concerne son évolution avec le décalage vers le rouge (z), c'est-à-dire au cours de l'histoire cosmique, ainsi que son évolution avec la contribution des galaxies centrales et satellites.

Cette thèse présente de nouvelles mesures de l'évolution cosmique de la SHMR, depuis $z \sim 5$ et séparant l'analyse en deux sous-groupes: les galaxies centrales et les galaxies satellites. La nouveauté de ce travail réside dans l'utilisation d'un seul ensemble de données - le catalogue COSMOS2020 est un catalogue dont les galaxies sont sélectionnées dans les bandes en proche infrarouge et avec des mesures de la photométrie dans plus de 30 bandes. Ce catalogue est complet jusqu'à une limite de faible masse stellaire et avec des décalages vers le rouge photométriques précis. Cela a été rendu possible en partie par la photométrie plus profonde de *Spitzer*/IRAC qui a été extraite lors de cette thèse pour le catalogue COSMOS2020. Ainsi, le catalogue permet des mesures du regroupement et de l'abondance des galaxies pour des échantillons complets dans dix intervalles de décalage vers le rouge. Ces mesures sont décrites à l'aide d'un modèle phénoménologique qui paramètre la SHMR pour les galaxies centrales et le nombre de galaxies à l'intérieur de chaque halo de matière noire. La SHMR résultante dans dix intervalles de décalage vers le rouge jusqu'à $z = 5.5$ pour les galaxies centrales et satellites est comparée à des simulations hydrodynamiques pour démêler les effets de différents mécanismes physiques.

Cette thèse examine également les perspectives d'une sonde supplémentaire de lentilles gravitationnelles faibles basée sur le biais de magnification dans le contexte de l'étude approfondie à venir de la mission *Euclid*. Le biais de magnification fait référence à la modification des densités de nombre de sources en arrière-plan en raison de l'effet de lentille provenant de la matière en premier plan. Ainsi, cela ne nécessite pas de mesures de forme, ce qui signifie qu'il peut être utilisé à des décalages vers le rouge plus élevés que les sondes basées sur le cisaillement. Étant donné que COSMOS fera partie des champs profonds et auxiliaires d'Euclid, le catalogue COSMOS2020 est utilisé dans cette première

étude. La magnification est mesurée à l'aide de la fonction de corrélation angulaire à deux points entre des échantillons avant-plan sélectionnés en fonction de la masse stellaire à $z < 1.5$, et des échantillons arrière-plan sélectionnés en fonction de la magnitude apparente à $2.7 < z < 4$. Ces mesures sont ajustées avec un modèle basé sur la distribution d'occupation des halos (HOD, en anglais halo occupation distribution) qui paramètre la SHMR. La magnification peut fournir des contraintes complémentaires sur les paramètres HOD, mais est limitée par son faible rapport signal/bruit dans COSMOS.

Une deuxième étude de la magnification s'appuie sur la simulation FLAGSHIP2.0 d'Euclid. La magnification est mesurée à l'aide de la corrélation angulaire à deux points pour des échantillons en avant-plan sélectionnés en fonction de la masse stellaire (à $z < 2$) et des échantillons en arrière-plan sélectionnés en fonction de la magnitude (à $2.5 < z < 2.9$). Des simulations d'erreurs photométriques et des estimations de décalages vers le rouge photométriques sont utilisées pour étudier l'impact des erreurs systématiques provenant du regroupement dans les mesures de magnification. Cette analyse montre que les données en proche infrarouge de l'étude approfondie d'Euclid, combinées aux données optiques terrestres de l'étude Cosmic Dawn Survey, peuvent fournir des décalages vers le rouge photométriques d'une qualité suffisante pour contrôler les erreurs systématiques et mesurer une magnification non biaisée. Cependant, la sélection des échantillons devra être effectuée avec soin pour minimiser les valeurs aberrantes dans l'espace des décalages vers le rouge réels par rapport aux décalages vers le rouge photométrique. Cette analyse se concentre sur la gamme de décalages vers le rouge ($z > 1$) dans laquelle la magnification peut être utilisée avec succès pour sonder les masses des halos, là où les sondes basées sur le cisaillement présenteront des incertitudes trop importantes.

Enfin, cette thèse mène une recherche sur des groupes de galaxies à $z > 3$ en utilisant des données provenant de la première étude à grand champ et à longueurs d'ondes élevées des champs profonds d'Euclid. Cela est réalisé à l'aide de catalogues infrarouges de Spitzer, dans le cadre de l'étude Cosmic Dawn Survey, préparée pendant la thèse. Les groupes à $z > 3$ sont rarement trouvés dans la littérature, et leur contribution à la densité du taux de formation d'étoiles cosmiques ainsi que leur rôle dans la formation des propriétés des galaxies restent mal compris. En utilisant ces données homogènes à grand champ dans l'infrarouge, des candidats de groupes ont été identifiés en recherchant des surdensités de galaxies avec des couleurs rouges *Spitzer* correspondant à des pics rouges dans les cartes *Herschel*. Associé aux observations radio NOEMA dans la bande des 3 mm, ce projet révèle l'un des proto-amas les plus lointains en pleine formation à $z = 4.2$ en détectant une ligne d'émission CO(4-3) et une ligne d'émission $H_2O(4_{2,3} - 3_{3,0})$ provisoire. Ce proto-amas fait l'objet d'une proposition de suivi pour des observations NOEMA dans la bande des 2 mm, afin de détecter plus de lignes et de membres du groupe pour caractériser en détail le proto-amas.

ACKNOWLEDGMENTS

This manuscript represents, undoubtedly, my life's biggest intellectual achievement. All of this would not have been possible and the thorny journey would not have been worth it without all the wonderful people to whom I am endlessly grateful.

On the first row in my academic support, there were always and consistently my supervisors, Henry and Raphaël. They have guided me throughout my thesis and have given me the necessary knowledge, advice, and resources to develop into an independent and competent researcher. I have been lucky to be surrounded by extraordinary scientists and persons who have helped me immensely during my career so far, among which I am particularly grateful to Clotilde, Iary, Olivier, Stéphane, Sune, et al.

Ph.D. life at IAP has been an extraordinary experience, both academically and in terms of friendships with some of the best colleagues one can have: Nai, Amaury, Louis (see also the poster in 13A), Aline, Axel, Simon, Virginia, Chi An, Warren, Étienne, Lukas, Pierre, Martin, the IAP outsider but best flatmate ever Dani, and many more. All those nights and days, regular and clandestine, in the office, music room, apèros, restaurants, quais de la Seine, and the legendary Bar Damesme. All those unforgettable and endless moments spent searching for the best coffees in Paris, perfecting coffee brews, mastering authentic pizza and carbonara recipes, running the clandestine lockdown speak-easy cocktail bar at rue Damesme 52, 2000+ km ran on the streets of Paris, 100+ km long weekend bike rides in Vallée de Chevreuse, and many more moments, they all make these grueling journey to a Ph.D. the best three years of my life. Finally, I deeply appreciate and hold dear all people who appeared in my life at any point and made this journey more enjoyable.

Last but not least, I can only have endless gratitude for the most supportive parents one can have. They have encouraged me at every turn and have and, together with me relived, every step of the way. To my brother, who being my role model sparked my interest in astronomy in the first place and set the standards sky-high. I owe them so much of this success and it brings my heart joy to know how proud this has made them.

1

INTRODUCTION

Physical cosmology studies the origin, fate, dynamics and emergence of the large-scale structure of the Universe. Its development into the modern science it is today was made on the pillars of the laws of science and the scientific method. The uniqueness of cosmology is that its theoretical predictions, derived on the foundation of the Einstein's general theory of relativity, cannot be studied in a laboratory. Instead, it relies on observations of objects and phenomena on the sky that constitute our past light-cone. One of the key observations that led to the establishment of our modern standard cosmological model are those of the light emitted by objects, both in the near and far Universe – galaxies. The second key observation is of an electromagnetic radiation (called the cosmic microwave background radiation) from a particular event in the early history of the Universe (called the recombination). These led to the discovery that the Universe is in expansion from measurements of the recessional velocities of distant galaxies, that matter is homogeneously distributed on large scales from the detection of the uniform cosmic microwave background radiation. These two also allowed the inference that the Universe started from an initial small, dense and hot state – the Big Bang, which led to the primordial nucleosynthesis of the first chemical elements. These observations, along with the general theory of relativity, are the fundamental principles of the modern standard model of cosmology.

Numerous studies throughout the twentieth century, including theoretical, observational and numerical, have converged towards a picture in which the mass-energy density of the Universe is dominated by dark energy ($\sim 70\%$) and dark matter ($\sim 25\%$) with only a $\sim 5\%$ contribution from ordinary baryonic matter. Dark energy is thought to be responsible for the accelerated expansion of the Universe, as evidenced by the redshift-distance relation of Type Ia supernovae. Dark matter is thought to be made of non-baryonic massive particles with low thermal velocities that provide the gravitational potential in which structures can grow; it has an imprint on rotational velocities of galaxies and on the way they are distributed in the Universe.

Galaxies are sufficiently bright and numerous that can be observed across a large cosmological volume and that closely trace the overall matter distribution in the Universe. The advent of photographic plates near the end of the 19th century allowed astronomers to start making maps of the night sky. The first maps of the large-scale galaxy distribution were made in the 1930s by counting galaxies (by eye) in photographic plates. Such galaxy surveys using photographic plates resulted in the first statistical description of the large-scale structure by measuring two-point correlation functions, which in turn provided a quantitative way to compare observations with theoretical predictions. The advent of charge-coupled device detectors near the end of the 20th century made it possible to rapidly image large areas of the extragalactic sky and accurately quantify the position and brightness of distant and faint galaxies. Along with the development of multi-object spectrographs, such photometric surveys allowed the construction of multi-wavelength galaxy catalogs of galaxies with measurements of key galaxy properties such as colors, redshift, stellar mass, star formation rates, morphology etc. This made possible the selection of complete galaxy samples down to faint magnitudes and low stellar masses, and characterize the great diversity of galaxy populations. In addition, galaxies do not perfectly mirror the clustering of the overall matter in the Universe. Instead, they are biased tracers of the matter distribution, where the more luminous ones are typically being more clustered than the dark matter. This biasing also depends on galaxy properties such as luminosity, morphology and color. Therefore, understanding galaxy properties was indispensable in mapping the three-dimensional matter distribution in the Universe and helped usher in the era of precision cosmology.

Within the current paradigm of structure and galaxy formation, galaxies form by cooling and condensation of baryonic gas within the potential wells of dark matter halos. The way galaxies grow in mass is regulated by processes that happen at various spatial and timescales and are closely related to the dark matter environment. These processes include the hierarchical merging, the accretion and cooling of baryonic gas and formation of stars that make galaxies grow. These are in competition with processes that quench growth by heating the gas and/or blowing it away, such as stellar, supernovae and active galactic nuclei feedback. Furthermore, due to merging, a massive halo can end up hosting one central and many other satellite galaxies. These satellites can have their growth hampered by environmental quenching effects like harassment, stripping etc. The strength of these processes is closely related to the properties of the host dark matter halo.

The relationship between the mass of the dark matter halo and the stellar mass of the galaxy it hosts is referred to as the stellar-to-halo mass relation (SHMR). The SHMR, which expresses the efficiency of the stellar mass assembly of a galaxy integrated over the halo's lifetime, is at first-order a function of the halo mass and is shaped by the physical mechanisms of galaxy formation. As such, the SHMR is an important quantity in order to understand how galaxies have assembled their mass throughout the cosmic history.

Numerous studies throughout the years have measured the SHMR with many different methods. They have revealed that galaxy formation is an inefficient process – the SHMR is quite low. It is a strong function of halo mass and peaks at a characteristic peak halo mass, suggesting that at most only 20% of all the available baryons in the halo have turned into stars. At low redshifts, they have also revealed that in halos of low and intermediate the central galaxy dominates the stellar mass budget, but satellites start to contribute significantly and even dominate in massive halos.

Many studies in the literature rely on the phenomenological abundance matching (AM) method. This method matches the abundance of galaxies of a given mass to the abundance of dark matter halo. AM techniques are a powerful tool, but suffer from some shortcomings and are limited in their predictive power. For example, they can only estimate the SHMR of central galaxies, and can not inform us about how satellite galaxies contribute to the stellar mass budget of a halo.

There exist other methods to study the SHMR that are also sensitive to the contribution of satellite galaxies. These typically rely on measurements of galaxy clustering and galaxy-galaxy lensing, and on statistical models of the way galaxies occupy halos, such as the halo occupation distribution (HOD). They have been successful in measuring the SHMR for both central and satellite galaxies, but have some key limitations. Lensing studies are mostly limited to $z < 1$ and are thus unable to probe the cosmic evolution of the SHMR. Studies based on galaxy clustering, abundance and other statistical observables, have not provided a strong consensus on the cosmic evolution of the SHMR, especially at progressively higher redshifts. This is because of the difficulty of assembling a homogeneous data set to measure all observables from the local to the distant Universe. As such, most studies have relied on heterogeneous datasets. Different data sets can have different selection functions and methods of estimating galaxies' physical parameters. These can propagate various systematic biases that muddle the interpretations.

An important aspect in understanding the cosmic stellar mass assembly in halos is measuring the redshift evolution of the SHMR across the majority of cosmic history. Currently, there is no consensus that the SHMR, and consequently the star-formation efficiency (SFE), evolves at all. From the numerous studies using different data sets, some find that the SFE evolves, while others find no evolution. The debate on the SFE evolution, especially as we enter the redshifts of the epoch of reionization ($z > 6$) is, a crucial one that needs to be settled.

To address the questions on the cosmic evolution of the SHMR and its contribution from central and satellite galaxies, it is necessary to have galaxy catalogs complete down to low stellar mass with accurate measurements of redshifts and other physical properties. In addition, these catalogs need to provide sufficient numbers of galaxies in both the local and distant Universe with properties measured consistently. Such catalogs can only be constructed from deep and wide multi-wavelength surveys. Deep observations in the observer frame near-infrared are key to select complete samples to large redshifts ($z \sim 6$ and beyond).

Perhaps one of the best examples of such a deep and multi-wavelength survey is COSMOS. Its $\sim 2\,\mathrm{deg}^2$ field has been covered by all major facilities on Earth and in space at all wavelengths of interest. Thanks to the depth in near-infrared bands (AB ~ 26 at 3σ), COSMOS galaxy catalogs are near-infrared selected, allowing them to be complete down to low stellar mass and to high redshifts. As such, it offers the necessary conditions to study the relation between galaxies and halos consistently in a large range of redshift and stellar mass.

The main interest of this Thesis is the upcoming Euclid Deep Survey. *Euclid* is a mission of the European Space Agency, designed to constrain the properties of dark energy and gravity via weak gravitational lensing and galaxy clustering. Along with its core Wide survey of $\sim 15,000\,\mathrm{deg}^2$, it will also deeply image $\sim 50\,\mathrm{deg}^2$ down to AB ~ 26 (at 5σ) in the near-infrared bands. In synergy with ground based observations in the optical, the Euclid

Deep Survey promises to provide the ideal combination of depth and area to study the galaxy-halo connection for a large redshift range and for a variety of galaxy populations. The large area will also enable weak lensing probes like magnification, that does not rely on shape measurements, to be measured with high signal-to-noise ratio to redshifts higher than what is possible with shear.

This Thesis aims to study the evolution of the SHMR across a large redshift range (out to $z \sim 5$). The novel aspect of this work is the use of a single dataset, the COSMOS2020 catalog, to measure galaxy clustering and abundance, and the use of an HOD model to describe the measurements. As such, it constrains both the central and satellite galaxy contribution to the stellar mass budget in a dark matter halo to $z \sim 5$. The results are also interpreted by comparing them to hydrodynamical simulations, that helps to disentangle the effects of different physical mechanisms.

This Thesis also investigates the prospects of an additional weak lensing probe based on the magnification bias, that directly probes dark matter halos around normal galaxies. Magnification bias refers to the modification of the number densities of background sources due to lensing from foreground matter. As such, it does not require shape measurements, which allows its use at higher redshifts than what is possible with shear based probes. The prospects of magnification are investigated in the context of the upcoming deep survey of the *Euclid* mission, that will provide the best combination of depth and area for magnification studies.

To enable the scientific studies of this Thesis, a significant effort was spent in preparation of stellar mass-limited catalogs from multi-wavelength surveys. This consisted of photometry extraction in the infrared for the COSMOS2020 catalog and the Cosmic Dawn Survey, which is part of the Euclid Deep Survey.

Finally, this Thesis carries out a search of galaxy clusters at $z > 3$ using deep and wide infrared data, as well as targeted observations in radio. Assembling representative samples of clusters beyond the peak of cosmic star-formation is important in order to study their transformational impact on the properties of their member galaxies, measure their contribution to the total star-formation in the Universe and study how cold streams in hot halos fuel galaxy growth.

1.1 THESIS OUTLINE

This Thesis is organized in three parts and ten chapters, which we outline in the following.

Part I: Background

Chapter 2 : Formation and evolution of galaxies within the standard cosmological model, provides the background information relevant to this Thesis. It gives a broad overview of the standard cosmological model, structure formation in the Universe and formation and evolution of galaxies.

Chapter 3 : Theoretical framework to study galaxies in relation to the large-scale structure, lays down the theoretical framework that was employed throughout the thesis to study galaxies in relation to the large-scale structure. It presents the halo model and weak gravitational lensing.

Chapter 4 : **Galaxy-halo connection**, presents the main topic of interest in this Thesis – the relation between galaxies and halos. It details the halo occupation distribution approach to study the stellar-to-halo mass relation. It motivates the main study of this Thesis by giving an overview of the methods to study it, current limitations and open questions.

Chapter 5 : **Numerical simulations**, gives a brief overview of numerical simulations. By comparing them to our measurements in photometric surveys, numerical simulations serve to disentangle the contributions of different physical mechanisms shaping the galaxy-halo connection.

Part II: Unveiling the galaxy-halo connection with photometric surveys

Chapter 6 : **Photometric surveys**, gives a brief overview of the historical development of photometric surveys and their importance in our understanding of galaxy formation and evolution. This chapter also presents the work carried out during this thesis on photometric surveys such as COSMOS and the Cosmic Dawn Survey. This includes photometry extraction in and validation in *Spitzer*/IRAC, construction of survey masks and COSMOS2020 catalog validation from measurements of the angular correlation function. These data were used to carry out the main scientific studies of this Thesis.

Chapter 7 : **Stellar-to-halo mass relation in COSMOS2020**, presents the study of the relation between galaxy stellar mass and its host halo mass out to $z = 5.5$ in the COSMOS2020 catalog.

Chapter 8 : **Studying the galaxy-halo connection with magnification**, presents the measurements of weak lensing magnification in the COSMOS2020 catalog. We measured magnification with the two-point angular cross-correlation function and investigated how well these measurements can constrain the parameters of a model based on the halo occupation distribution (HOD).

Chapter 9 : **Prospects of magnification in the Euclid Deep Survey**, presents the prospects of measuring weak lensing magnification in the Euclid Deep Survey. In this chapter, we investigated the effects of clustering systematic errors and photometric redshifts in the measurements of magnification in the Euclid Deep Survey.

Chapter 10 : **Galaxy clusters at $z > 3$**, presents the search for galaxy protoclusters at $z > 3$ in the Euclid Deep Fields. In this work, we searched for overdensities of galaxies with red IRAC colors (at $3.6 - 4.5$ μm) that correspond to red peaks in Herschel maps (at $250 - 500$ μm). NOEMA observations were carried out for some of the most robust candidates, that revealed a new protocluster at $z = 4.2$, one of the most distant ones known to date.

Part III: Conclusions and future outlooks

Chapter 11 concludes this Thesis. First, we summarize the work in this Thesis, and then we discuss future prospects in the light of the upcoming galaxy surveys.

Part I

Background

2

FORMATION AND EVOLUTION OF GALAXIES WITHIN THE STANDARD COSMOLOGICAL MODEL

The study of the formation and evolution of galaxies can only be done within the framework of the cosmological model that describes our Universe as a whole. The dominant matter component in the Universe is dark matter, which only interacts gravitationally, and the way it is distributed is governed by the laws and principles that make the cosmological model. The large-scale distribution of dark matter represents the cosmic scaffolding in which baryonic processes occur that form galaxies. Galaxies are then distributed over cosmological scales, and they closely trace the large-scale structure of the Universe, shaped by the dark matter. Furthermore, the way they first formed, through gravitational collapse of matter in an expanding Universe, and the way they grow for the rest of their lives, is governed by the cosmological context. Some of the key observables that allow us to study galaxies, such as redshift, are stemming directly from cosmological effects. Therefore, to pursue the study of the properties of galaxies, especially in relation to the large-scale structure, it is necessary to lay down the essentials of the cosmological framework.

However, the formation and subsequent evolution of galaxies is a far more complex process than what the cosmological model can formally describe. Galaxies, being made of baryons, conform to the panoply of physical laws that govern how they interact – and this is a very complicated process. Galaxies, being made of matter in various different phases that all interact between themselves, therefore exhibit an extraordinary range of different properties. Some of them are related with scaling relations that facilitate their observations and interpretations, while others defy our models and predictions. In this chapter, we will give a brief and non-exhaustive overview of some galaxy properties, mainly related to their mass assembly in relation to the larger scale dark matter environment, that have taught us about the galaxy formation and evolution processes.

2.1 STANDARD COSMOLOGICAL MODEL

Cosmology is the study of the Universe as a whole – how it began, how it evolved throughout history, its present state and its fate. Cosmology has been persistently studied since the dawn of civilization by almost every culture. However, it was only after the firm establishment of the scientific method, that the study of cosmology became a science and started to provide accurate descriptions and predictions about our Universe. The modern study of cosmology began with the formulation of the theory of general relativity (GR) by

Einstein, which postulates that the geometry and the structure of the Universe is defined by its matter/energy content. The general theory of relativity allowed the formulation of the Einstein field equations that describe the geometry of the space-time in relation to the matter/energy distribution. The framework that the GR provided was used by to describe the Universe as a whole. However, the initial considerations were hardly based on any observational evidence, nor was it confirmed by observations around the time of their formulation, and it took several decades of mounting observational evidence to converge towards a *standard model of cosmology*.

The modern standard model of cosmology has emerged from a number of key observations, mainly from objects far in space. It is based on a few fundamental principles that were established based on mounting evidence from theory and observations across several decades. In the following, we list the fundamental principles of modern cosmology and briefly present their historical development.

General relativity. The laws of governing the dynamics and expansion of the Universe are described by the general relativity. In this framework, space-time is the "fabric" of the Universe that is defined by a metric. The curvature is shaped by the content of energy, momentum and stress within the space-time. This is written with the Einstein field equations

$$R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu} - \Lambda g_{\mu\nu} = 8\pi G T_{\mu\nu}, \tag{2.1}$$

where $T_{\mu\nu}$ is the stress-energy tensor, $g_{\mu\nu}$ is the metric tensor, $R_{\mu\nu}$ is the Ricci tensor which is a function of metric tensor derivatives, R is its trace, and Λ is the cosmological constant.

Originally, Einstein wrote these equations without the cosmological constant. Such a set of field equations would result in an expanding Universe. In the second decade of the twentieth century, an expanding Universe was defying the evidence and beliefs that the Universe is static. As a consequence, Einstein introduced the cosmological constant in order to counteract the expansion and maintain a static Universe. Later observational evidence that the Universe is expanding prompted Einstein to remedy his 'biggest blunder' by removing the cosmological constant. Later still, near the end of the twentieth century, observational evidence of a Universe in an accelerated expansion introduced back the cosmological constant.

Expanding Universe. The first observational evidence that the Universe is in expansion came from spectroscopic observations of galaxies that had their spectrum redshifted – an indication that their velocities are receding. Vesto Slipher made the first observations of redshifted galaxies in 1912 from which the interpretation that they must be receding emerged. More precise measurements were made by the Swedish astronomer Knut Lundmark in 1924, who, based on these receding velocities, found observational evidence for expansion within 1% accuracy compared to the best measurements today. In 1929, Edwin Hubble used measurements of Cepheid variable stars outside the Milky Way to determine distances, and in compilation with Slipher's measurements used them to establish a positive linear relation between distance and recessional velocity of galaxies. This resulted in what is known as the Hubble-Lemaître law

$$\vec{v} = H\vec{r} \tag{2.2}$$

Cosmological principle. The Universe, seen on large enough scales, is homogeneous and isotropic. This principle has been firmly established by the discovery of the cosmic microwave background (CMB) – a homogeneous and isotropic blackbody radiation at ~ 2.725 K coming from the epoch of the matter-radiation decoupling in the very early Universe (when it was 380 000 years old). However, this uniform radiation is not completely homogeneous. Small fluctuations on the order of 10^{-5} are present, which serve as the seeds of the structure formation in the later universe. An additional evidence in support of the homogeneity on cosmological scales came from precise measurements of the fluctuations in the large-scale distribution of galaxies.

The presence of the CMB is also consistent with the expansion of the Universe – if the Universe is expanding, then in its past it must have been in a dense and hot state. As George Gamow suggested in the 1940s, this hot state responsible for the nucleosynthesis of the first elements, would bathe the Universe in radiation, which would still be detectable today. The first observational evidence for the uniform CMB came unintentionally, when in 1964 Arno Penzias and Robert Wilson detected a persistent excess of antenna temperature in their radiometric measurements that could only come from cosmic origins. The huge cosmological implications of such a detection prompted the deployment of ever more sensitive instruments to map the CMB sky in detail, such as COBE, WMAP and Planck, among a plethora of others.

Big Bang. The measurement of the receding velocities of galaxies indicates that the Universe had to expand from an initial small, dense and hot state. The CMB provided a clear snapshot of the hot early state. If the Universe has been expanding and cooling, then it must have started from an initial state of high density and temperature – the Big Bang. The Big Bang theory therefore successfully explains the receding galaxies, the CMB as well as the primordial nucleosynthesis, since it is in agreement with measurements of the abundance of primordial chemical elements.

The metric of the Universe

The Universe within this framework is described by a metric, that in essence defines the distance between two points in a given space (the line element $\mathrm{d}s$). It emerges as a solution of the Einstein field equations. In the 1920s and 1930s, four scientists – Alexander Friedmann, Georges Lemaître, Howard P. Robertson and Arthur Geoffrey Walker, independently solved the Einstein field equations for a homogeneous isotropic and expanding Universe. Their solutions led to the most established metric for our Universe, the Friedmann-Lemaître–Robertson–Walker (FLRW) metric

$$\mathrm{d}s^2 = c^2\mathrm{d}t^2 - a(t)^2\left(\frac{\mathrm{d}r^2}{1-kr^2} + r^2\mathrm{d}\theta^2 + r^2\sin^2\theta\mathrm{d}\phi^2\right). \tag{2.3}$$

The FLRW metric is completely specified by the curvature of the universe k, and the scale factor $a(t)$ that describes the change of length scale with time. The coordinates r, θ, ϕ are *comoving* coordinates and t is the proper time. The line element, or space-time interval ds, is real/imaginary for two events with a time-like/space-like separation and is zero for two events on the same light path (null geodesic). The curvature parameter k defines the 'shape' of the universe and can take values $1, 0$ and -1, that correspond to a positively curved (closed), flat and negatively curved (open) Universe.

Content of the Universe and cosmological parameters

In the standard model of cosmology, the energy of the Universe is modelled as a perfect fluid described by an equation of state relating the pressure and density $p = w\,\rho\,c^2$, where w is the equation of state parameter. Using this assumption and inserting the FLRW metric (Eq. 2.3) in the Einstein field equations (Eq. 2.1) one obtains the equations that govern the expansion of the Universe: the Friedman-Lemaître equation

$$\begin{aligned}\left(\frac{\dot a}{a}\right)^2 &= \frac{8\pi G}{3}\rho - \frac{k\,c^2}{a^2} + \frac{\Lambda\,c^2}{3},\\ \frac{\ddot a}{a} &= -\frac{4\pi G}{3}\left(\rho + \frac{3p}{c^2}\right) + \frac{\Lambda\,c^2}{3}.\end{aligned} \tag{2.4}$$

They describe the time evolution of the key properties of the Universe such as scale parameter $a(t)$, density ρ and pressure p. In terms of the scale parameter and its derivative, one can define the Hubble parameter as

$$H(t) = \frac{\dot a(t)}{a(t)}, \tag{2.5}$$

which sets the expansion rate. The first Friedman-Lemaître equation can be rewritten in a simpler form by defining the so-called cosmological parameters

$$\Omega_\mathrm{m} = \frac{8\pi G}{3H}\rho_\mathrm{m}, \quad \Omega_\mathrm{r} = \frac{8\pi G}{3H}\rho_\mathrm{r}, \quad \Omega_k = \frac{k}{a^2H^2}, \quad \Omega_\Lambda = \frac{\Lambda}{3H^2}, \tag{2.6}$$

where the density was split into matter and radiation component. Then the first Friedman-Lemaître equation is written in terms of these density parameters as $\Omega_\mathrm{m} + \Omega_\mathrm{r} + \Omega_\mathrm{k} + \Omega_\Lambda = 1$; it defines the energy budget of the Universe. The matter density can be further split into a sum of a baryonic and dark matter component $\Omega_\mathrm{m} = \Omega_\mathrm{b} + \Omega_\mathrm{dm}$. Ω_r is the radiation density which is ≈ 0 today, Ω_k is called curvature density, and Ω_Λ is the cosmological

Table 2.1: Value of some cosmological parameters as estimated by Planck Collaboration et al. (2020). $h = H_0/100$ is the reduced Hubble constant.

Symbol	H_0	$\Omega_b h^2$	$\Omega_{\mathrm{dm}} h^2$
Value	$67.66 \pm 0.42 \frac{\mathrm{km}}{\mathrm{s\ Mpc}}$	0.02242 ± 0.00014	0.11933 ± 0.00091
Symbol	Ω_k	Ω_Λ	
Value	0.0007 ± 0.0019	0.6889 ± 0.0056	

constant density. These parameters are a function of time, and the reference time at which they are expressed is today.

One of the principal goals in cosmology is to measure the values of the cosmological parameters which determine the structure, history and fate of the Universe; based on their values different scenarios about the evolution of the Universe are possible. Cosmological studies have converged to a narrow region of the cosmological parameter space that defines the so-called Lambda-Cold Dark Matter model (ΛCDM). In the ΛCDM model, the Universe is flat and is composed of three main components: baryonic matter, cold dark matter (CDM) and a cosmological constant Λ. The cosmological constant is responsible for the accelerated expansion of the Universe and has an equation of state parameter $w = -1$. This energy density dominates the energy budget of the Universe with $\sim 70\%$, and is also called *dark energy*. The remaining part of the energy budget is matter, of which $\sim 25\%$ is due to cold dark matter and the rest $\sim 5\%$ is due to baryonic matter. Measuring the cosmological parameters to ever-increasing precision is at the heart of all the major observatories on Earth and in space such as Dark Energy Survey (DES) (Rodríguez-Monroy et al., 2021), *Euclid* (Laureijs et al., 2011), Vera C. Rubin Observatory (Ivezić et al., 2019) etc. At present, an unprecedented precision on the cosmological parameters has been set by the European Space Agency's *Planck* mission, from CMB measurements. A table with the values of some cosmological parameters is given in Table 2.1 (Planck Collaboration et al., 2020).

Redshift

Because of the expansion of the Universe, specified by $a(t)$, the lengths increase with time. A photon emitted with a wavelength λ_e at time t_e will consequently have its wavelength redshifted by the expansion to λ at some later time t. The ratio of this change in scales is called redshift

$$1 + z = \frac{\lambda}{\lambda_e} = \frac{a(t)}{a(t_e)}. \tag{2.7}$$

The redshift that is due to the cosmological expansion is called *cosmological* redshift and always increases the wavelength of a distant source. The fact that the sources can have their peculiar velocities also results in *Doppler* shift of the wavelength that can be red-wards or blue-wards depending on whether the source is moving away or approaching the observer.

The cosmological redshift of an object is a crucial quantity in extragalactic studies because it provides a means to measure distances. This is also made possible by detailed understanding of the spectral energy distribution (SED) of the galaxies. Redshift can be estimated *spectroscopically*, by identifying a shifted spectral line or *photometrically* by identifying broad features (such as Lyman and Balmer breaks) and colors in the spectral energy distribution (see e.g., Salvato, Ilbert and Hoyle, 2019).

Comoving coordinates and cosmological distances

In extragalactic astronomy and cosmology, distances are often expressed in terms of *comoving* distances, which we will denote as χ. These factor out the expansion of the Universe, and stay constant with time (provided the two points are fixed). *Proper* distance, on the other hand, measures the distance to an object at the specific cosmological time and changes with the expansion. The two are related with the scale factor as

$$r_{\mathrm{prop}} = a(t)\,\chi. \tag{2.8}$$

In these definitions we assume that the Universe is flat. Other distance measures typically used are: 1) *angular diameter distance*, which gives the comoving distance at which a source of a given transverse comoving size will be seen under a given angular size (in radians). It is used to convert angular separations in telescope images to transverse separations in the source plane. It scales with the comoving distance as

$$d_{\mathrm{A}} = \frac{\chi}{1+z}; \tag{2.9}$$

2) *luminosity distance*, which relates the bolometric luminosity of a source to its observed bolometric flux. It is related to the comoving distance and angular diameter distance as

$$d_{\mathrm{L}} = (1+z)\,\chi = (1+z)^2\,d_{\mathrm{A}}. \tag{2.10}$$

The development of the modern cold dark matter paradigm

The first considerations of the content and structure of the Universe were based on the assumption that all the matter is of the same ordinary composition that we can study in a laboratory on Earth and emits some kind of radiation (i.e., baryonic matter). The first indications that there might be a considerable amount of matter that does not interact electromagnetically came in the 1930s. Fritz Zwicky showed in 1933 that in order to explain the velocities that he measured for the galaxies in the Coma cluster, it requires a total mass several hundred times larger than that of the luminous. Several decades later, studies of the velocities of the Milky Way satellite galaxies (Ostriker, Peebles and Yahil, 1974; Einasto et al., 1974) also showed that they need to be enveloped by massive halos in order to have the observed velocities. Measurements of the rotational curves of spiral galaxies out to large radii, made possible by radio observations of the 21 cm line, reinforced the case that galaxies are embedded in massive halos of dark matter that does not emit light (Roberts and Rots, 1973; Rubin, Ford and Thonnard, 1978).

The main question that was imposed was what is this missing (or dark) matter? One of the earliest candidates to account for the dark matter were neutrinos, after particle physics experiments showed that they have a non-negligible mass. However, they were proven to be unlikely to account for the dark matter because they could not explain and reproduce the observed galaxy distribution in the Universe. That dark matter must be of non-baryonic nature was indicated by numerous studies in the 1980s. For example, non-baryonic particles were required to produce a flat Universe and the observed large scale structure in the time since decoupling As the favored dark matter particles of non-baryonic nature emerged the so-called cold dark matter (CDM) – particles with low thermal velocities and masses much higher than keV. Studies based on numerical simulations in the 1980s, made possible by the advent of significant computing power, showed that CDM can explain the observed clustering of galaxies, provided they are biased tracers of the underlying dark matter density field (White, Davis and Frenk, 1984; Davis et al., 1985). In the early 1990s, clustering measurements in large scale galaxy surveys (Maddox et al., 1990) favored a CDM model coupled with a cosmological constant – ΛCDM, as an explanation for the observed clustering at large scales. The ΛCDM was solidified as the concordance model of cosmology when measurements of the redshift-distance relation of Type Ia supernovae showed that the Universe was in an accelerated expansion (Garnavich et al., 1998; Perlmutter et al., 1999). Presently, our best experiments have shown that the energy budget of the Universe is made out of $\sim 70\%$ of dark energy that is thought to be responsible for the accelerated expansion, $\sim 25\%$ of cold dark matter and $\sim 5\%$ of baryonic matter.

2.2 STRUCTURE FORMATION AND THE LARGE-SCALE STRUCTURE

The very early Universe was not completely homogeneous – small perturbations in the density existed, as expected from quantum fluctuations in the very early and dense stages and as evidenced by the CMB, served as the seeds out of which structures formed in the Universe. The inhomogeneities can be described by a density contrast written as

$$\delta(\vec{r},t) = \frac{\rho(\vec{r},t) - \bar{\rho}(t)}{\bar{\rho}(t)}, \tag{2.11}$$

where $\rho(\vec{r},t)$ is the matter density at a coordinate $\vec{r}$ and time t and $\bar{\rho}(t)$ is the mean cosmic matter density at the same time. The initial density field, continued its spatial expansion along with the Hubble flow, with the mean density decreasing with as $\propto a^{-3}\rho_0$, where ρ_0 is some initial mean density. However, the small overdensities with $\delta > 0$ had a stronger gravitational field than the average, so they expanded more slowly. The slower expansion of the overdense region means that their relative density increased. Therefore, small initial overdensities attracted more matter, thus δ increased with time, while the underdense regions decreased their density contrast. This effect of increase in the absolute value of density contrast $|\delta|$ is called *gravitational instability* and is responsible for the formation of matter structures.

The evolution of the density field and subsequently the emergence and evolution of the large scale structure (LSS) is described by the *perturbation theory* (see e.g., Bernardeau et al., 2002, for a detailed review). This evolution is described in the fluid approximation (pressure-free, dust only Universe) by the equations of motions

$$\frac{\partial \delta}{\partial t} + \frac{1}{a}\nabla[(1+\delta)\vec{u}] = 0, \tag{2.12a}$$

$$\frac{\partial \vec{u}}{\partial t} + H\vec{u} + \frac{1}{a}(\vec{u}\cdot\nabla)\vec{u} = -\frac{1}{\bar{\rho}\,a}\nabla p - \frac{1}{a}\nabla\phi, \tag{2.12b}$$

$$\nabla^2\phi(\vec{x},t) = 4\pi G\bar{\rho}a^2\delta(\vec{x},t), \tag{2.12c}$$

This set of three partial differential equations is written in terms of the comoving coordinates $\vec{x}$, therefore $\vec{u}(\vec{x},t)$ is the velocity field from which the Hubble expansion is factored out and represents the *peculiar velocity* $\vec{u}(\vec{x},t) = \vec{v}(\vec{r},t) - H\vec{r}$. The first equation is the continuity equation, describing the conservation of mass, or the evolution in density given a velocity field of the matter; the second is the Euler equation of motion, describing the conservation of momentum, or how the fluid moves under the influence of the forces of pressure and gravity; the third is the Poisson equation relating the gravitational potential to the matter density.

Linear growth

While the overdensities were still relatively small ($\delta \ll 1$), their physical size continued to grow due to the expansion of the Universe, while the density contrast increased as power law with time $\delta \propto t^\alpha$ in comoving coordinates. This is called the *linear regime* in which the continuity and Euler equations can be linearized (by keeping only the first-order terms) and be combined in a second-order differential equation

$$\frac{\partial^2\delta}{\partial t^2} + 2H\frac{\partial\delta}{\partial t} - 4\pi G\bar{\rho}\delta = 0. \tag{2.13}$$

It has two solutions, a *growing* and a *decaying* mode. The growing mode describes the growth of structures and can be written as

$$\delta(\vec{x},t) = D_+\delta_0(\vec{x}), \tag{2.14}$$

where D_+ is the *growth factor* and governs the amplitude of the overdensities and is a function of the density cosmological parameters. During this linear growth regime, the comoving radius of an overdensity changes negligibly, while the amplitude increases.

The linear regime describes well the evolution of the density field at large scales and the linear perturbation theory is successfully employed to describe the observed structure of the Universe at large scales, typically at > 10 Mpc. However, it is only valid for $\delta \ll 1$ and it breaks down as soon as the overdensities reach $\delta \sim 1$. Beyond this point, the structure evolves non-linearly and another formalism is needed to model it.

Non-linear structure evolution, the Cosmic Web and the spherical collapse model

If the initial overdensity was sufficiently large, at some point it will reach $\delta \sim 1$, after which it decouples from the expansion and its comoving radius starts to shrink – this is called a *turn-around*. The theory that describes the collapse of the first structure at intermediate scales using an analytical approach was first developed by Zel'dovich (1970). According to the Zel'dovich approximation, the collapse of the first structures happened first along one dimension, creating *sheet*-like structures, often called 'pancakes'. The subsequent collapse along a second dimension creates *filament* structures, while a collapse along a third dimension creates *nodes*. Such an evolution on megaparsec scales creates a web-like pattern of the matter distribution in the Universe called the *cosmic web* (see e.g., Bond, Kofman and Pogosyan, 1996; Cautun et al., 2014).

On a subparsec scale, the collapse of the overdensities is typically described by the *spherical collapse model* (Gunn and Gott, 1972). This model considers spheres of overdensities that reach $\delta \sim 1$, decouple from the Hubble flow and collapse to a certain radius and density. At this point, a system undergoes a *virial relaxation*, where its radius is supported by the motions of the particles. Such virialized mass concentrations are called *halos*, or *dark matter halos*, when considering the dark matter component. The average density at which this virialization of halos occurs is roughly 200 times the critical density of the Universe at the time of collapse. This makes one of the most common definitions of dark matter halos. From this formalism, where halos are considered as elemental building blocks of the LSS, a very powerful framework has been developed – termed the *halo model* (see e.g., Cooray and Sheth, 2002, for a review)– that successfully models various observables in both the linear and non-linear regime. The halo model is one of the central theoretical frameworks used in this Thesis and will be more formally presented in Chapter 3.

The inhomogeneous distribution of matter within the spheres considered in this approach cause a fragmentation during collapse and formation of lower mass halos first. These later merge and grow to form larger and more massive halos, that can be traced with a so-called *merger-tree*. Apart from merging, halos can further grow by *accretion* of more mass from the surrounding. Furthermore, even after merging of, e.g., smaller and larger halos, the smaller one can remain gravitationally bound and remain within the larger halo as a *subhalo*. This description is referred to as the *hierarchical*, or *bottom-up* scenario, and is the leading model of structure formation.

However, these descriptions based on analytical approaches and approximations are limited in their realistic description of the evolution of the LSS. The complicated gravitational dynamics that give rise to the matter structures are more realistically studied with numerical simulations (see.e.g., Klypin, 2000; Angulo and Hahn, 2022, for a description of simulation methods and the historical development). Finally, when decreasing the scale, the complexity in modeling increases. At even smaller scales (< 1 Mpc) the baryonic physics of galaxy formation also plays a role in shaping the matter structure and its corresponding power spectrum, and is necessary to be taken into account for precision studies (see e.g., van Daalen et al., 2011; Chisari et al., 2018).

Formation of galaxies

Each initial perturbation contained both collisionless dark matter and baryonic matter in their universal proportions set by *the universal baryonic fraction* f_b

$$f_\mathrm{b} = \frac{\Omega_\mathrm{b}}{\Omega_\mathrm{m}} = 0.158. \tag{2.15}$$

When an object collapses, the dark matter undergoes violent relaxation to form a dark matter halo, governed by the laws of gravity. The gas of baryons shocks to the virial temperature of the halo $T_\mathrm{vir} \geq 10^7$ K. White and Rees (1978) in their seminal paper proposed that the galaxies are formed by cooling and condensation of the baryonic gas within the potential wells provided by the halos. Due to the action of gravity, the gas fragments, collapses catastrophically and triggers nuclear reactions. In this way, stars are born, and along with the remaining gas and dust make a *galaxy*. The formation and subsequent evolution of galaxies is a complicated process, and it will receive further attention in the § 2.3. Its relation to underlying dark matter structures is the main focus of this Thesis.

Statistical description of the large scale structure

The study of the large scale structure of the Universe is done using statistical measures and observations of galaxies. Galaxies are sufficiently bright and numerous that can be observed across a large cosmological volume and that closely trace the large scale structure. One of the earliest ideas that the distribution of galaxies can be thought of as a realization of a random Poisson process came in 1952 by Neyman and Scott. They were among the first to propose that galaxies are residing in randomly distributed halos, and the position and the number of galaxies within a halo is random variable. It's striking how this idea is at the heart of the modern and widely successful framework – the halo occupation distribution (see Chapter 3). Such a random galaxy distribution can then be described statistically. The most widely used and enduring measure for this is the two-point correlation function and its Fourier transform, the power spectrum (described in more detail below). The first characterizations of the large scale galaxy distribution date back to 1934 when Hubble counted galaxies (by eye) in photographic plates. He made some of the first maps of the galaxy distribution by counting them in cells and in interval of apparent brightness. Two decades later, Shane and Wirtanen (1954) produced a detailed map of the large scale distribution of galaxies in the Lick catalog by doing counts in cells. These data were used to measure the galaxy two-point correlation function by Totsuji and Kihara (1969) and Groth and Peebles (1977), who showed that it is close to a power law with a characteristic scale length of $r_0 \sim 5$ Mpc. Groth and Peebles (1977) established the 'scaling relation' of the amplitude of the correlation function with apparent magnitude – progressively fainter galaxy samples have lower amplitudes at fixed angular separation. This served as a confirmation of the cosmological origin of the clustering signal that is not due to galactic obscuration. By the 1990s, the advent of computational power made possible the calculation of the correlation function to large angular scales of millions of galaxies covering large portions of the sky (Maddox et al., 1990).

Measurements of correlation functions and power spectra have been the central tool in cosmological studies, historically and still today. However, as we will see in this Thesis, they offer much more than that – a method to also study the properties of galaxies in relation to the underlying dark matter.

Galaxies do not perfectly trace the matter density distribution – instead, they are biased tracers (Kaiser, 1984; Bardeen et al., 1986). The *galaxy bias* can be written in terms of the density contrasts of galaxies and matter as a linear relation

$$\delta_g = b_g \, \delta_m, \tag{2.16}$$

which is referred to as *linear bias* and is found to be correct on large scales. This biasing of galaxies is largely due to the bias in the distribution of dark matter halos with respect to the dark matter field – massive objects form at the sites of high peaks in the initial density field. Additionally, the bias is dependent on both galaxy properties and on scale (see Desjacques, Jeong and Schmidt, 2018, for a review on the large scale galaxy bias). Here we will give a brief overview of the statistical description of galaxies, since it is of main interest in this Thesis, and mention how it is related to the matter distribution through the biasing.

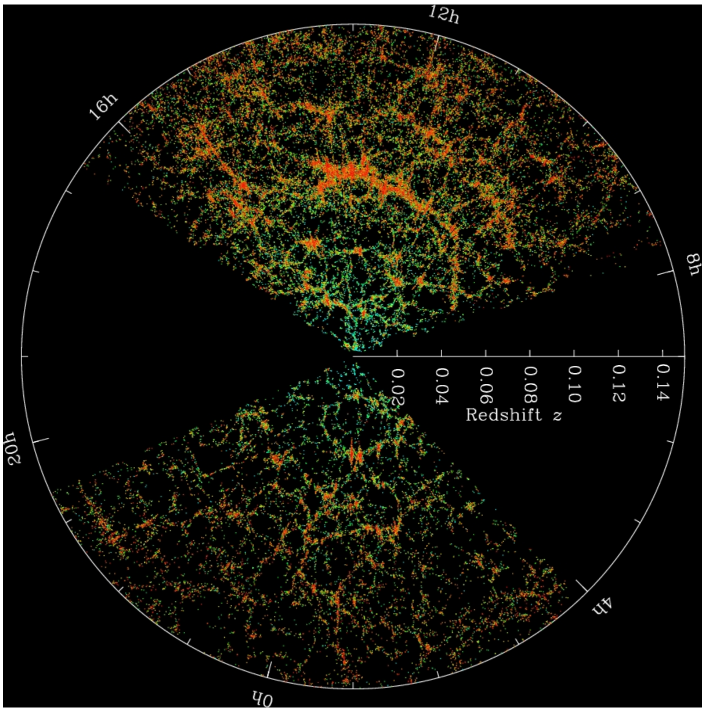


Figure 2.1: Large scale structure of the nearby universe as seen by the Sloan Digital Sky Survey (SDSS), using redshift measurements of galaxies. The distribution of galaxies traces that of the overall matter and reveals the web-like structure of the LSS Credit: SDSS

Correlation functions

One of the most widely used statistical measures to study the large scale structure using galaxies as tracers is the two-point correlation function (2PCF). The 2PCF is a measure of galaxy clustering – it quantifies how much galaxies are clustered together as a function of scale. The correlation function essentially measures the excess probability of finding a pair of galaxies in volumes $\mathrm{d}V_1$ and $\mathrm{d}V_2$ at a given separation r, compared to a completely random distribution (Peebles, 1975)

$$\mathrm{d}P = n^2[1 + \xi(r)]\mathrm{d}V_1\mathrm{d}V_2, \tag{2.17}$$

where n is the average density. More formally, a two-point correlation function can be defined using the density contrast $\delta_i(\vec{r}) = n_i(\vec{r})/\bar{n}_i(\vec{r})$ -1

$$\langle \delta_i(\vec{r})\, \delta_j(\vec{y}) \rangle = \xi_{ij}(r). \tag{2.18}$$

This represents an average of the product of the density contrast of all pairs separated by r. Here, $r = |\vec{x} - \vec{y}|$ which follows from the principles of homogeneity and isotropy. We keep here the most general form using the indices i, j that designate the density contrast of any tracer – this can be galaxy number density contrast δ_g or matter δ_m. If $i = j$ then what is measured is the *auto-correlation* and if $i \neq j$ then it is the *cross-correlation* function.

Practically, the correlation function is computed with the (Landy and Szalay, 1993) estimator, using random catalogs

$$\xi(r) = \frac{DD - 2\,DR + RR}{RR}. \tag{2.19}$$

DD is the number of data-data pairs in a given separation bin $[r, r + \delta r]$, RR is the number of pairs in the random catalog in the same bin, and DR is the number of pairs between the data and the random catalog. The data and random pairs are normalized by the total number of galaxies and random objects. The random catalog here has to be constructed with the same survey geometry[1] as the data catalog with, typically, N_{random} more than 10 times the number of galaxies in the data; such a large number ensures that there will be no additional noise added to the measurement, which should only contain the contribution from the finite number of galaxies.

The scale r can be three-dimensional comoving separation r, for which a redshift measurement is required. In this case, $\xi(r)$ is called a *real-space* correlation function. However, redshift measurements are not exact measures of distances to galaxies – instead, they are distorted due to the effects of peculiar velocities. A distance to a galaxy inferred from a redshift measurement thus contains two components, the true distance and the peculiar velocity distance. These peculiar velocities give rise to the so-called *redshift space distortions*, and since they are directly related to the gravitational potential, can be used to study the large-scale structure. In redshift space, these are manifested in two effects. 1) *Fingers of God* effect refers to the elongation of the galaxy distribution in the line-of-sight due to the

---

[1] Typically done with a mask that marks the region in the sky that is being analysed

random peculiar velocities of galaxies – their radial component causes red/blue shifting that results in this elongation. 2) *Kaiser effect* is a more subtle effect evident on larger scales, that results from the coherent inward motions of galaxies in clusters – this causes a flattening in the galaxy distribution in redshift-space. Because of the redshift distortions, the 2PCF correlation function will be anisotropic. Typically, it is decomposed into projections parallel (π) and perpendicular (p) to the line-of-sight $\xi(r_\pi, r_p)$, which is what is usually measured observationally.

Alternatively, where redshift information is unavailable, or only good enough to only roughly select galaxies in a broad redshift bins (like, for example, photometric redshifts), one can measure the *angular correlation function* $w(\theta)$. In this case, the scale is the angular separation between sources on the sky θ.

Finally, as mentioned earlier, galaxies are biased tracers of the dark matter distribution, and their respective 2PCF can be related, using the galaxy bias term introduced in Eqn. 2.16

$$\xi_g(r) = b_g^2 \, \xi_m(r) \tag{2.20}$$

Figure Fig. 2.2 (left) presents the clustering of galaxies as measured by the two-point projected correlation function in Sloan Digital Sky Survey (SDSS) by Zehavi et al. (2011). Measurements are shown for galaxies selected in different absolute magnitude thresholds – from less luminous in blue to more luminous in red and purple. More luminous galaxies are more strongly clustered, which is an evidence that galaxies are a biased tracer of the underlying matter distribution, and this bias is dependent on the galaxy properties. Apart from luminosity, the bias is also a function of stellar mass, color, morphology etc (see e.g., Zehavi et al., 2002, and rerefences therein). The correlation function can be well described by a power law, especially at low redshifts. However, ever more precise measurements have shown deviations from a power law, especially on scales of $\sim 1-2\,h^{-1}$ Mpc (Zehavi et al., 2004), and more prominently at $z > 1$ (Kravtsov et al., 2004). This posed a limitation for the simple power-law models, and raised the necessity to properly model the non-linear evolution at intermediate and small scales. One of the most successful frameworks that were developed to model this is the halo occupation distribution (Berlind and Weinberg, 2002; Kravtsov et al., 2004; Zehavi et al., 2004, and references therein). The halo occupation distribution is one of the central tools of this Thesis and will be more thoroughly presented in Chapter 3.

Power spectrum

A convenient way to express the distribution of matter in the Universe, as measured by the correlation function, is its Fourier transform – *the power spectrum*. The two of them are related as

$$P(k) = 2\pi \int_0^\infty \mathrm{d}r \, r^2 \frac{\sin(kr)}{kr} \xi(r) \tag{2.21}$$

The power spectrum is a function of the wave number k, which is related to a length-scale as $L \sim 2\pi/k$. In modeling terms, the power spectra are easier to work with, since convolutions in real space become multiplications in Fourier space. Additionally, they also

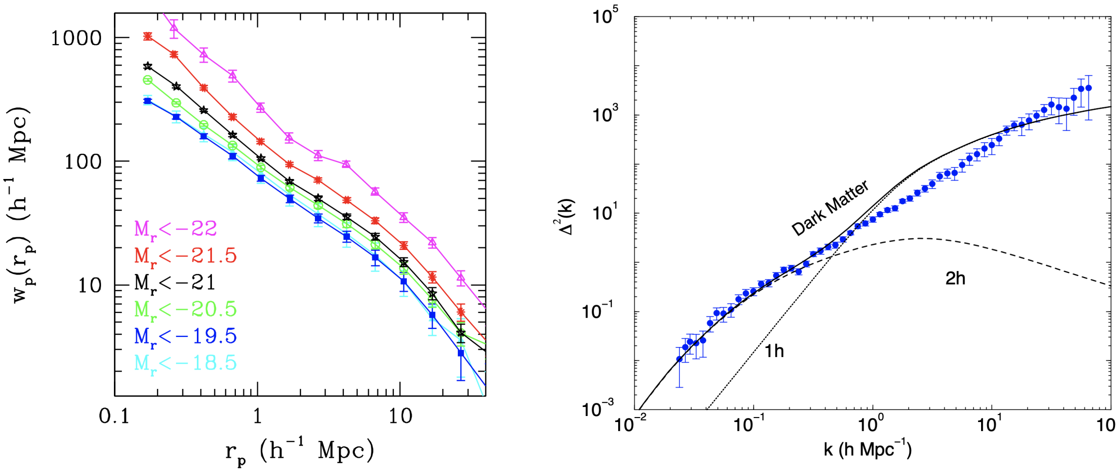


Figure 2.2: Measurements of the two-point correlation function (left) and power spectrum (right) from the literature. **Left**: The projected correlation function as a function of comoving separation, measured in SDSS by Zehavi et al. (2011). Measurements are shown for galaxies selected on different absolute magnitude thresholds, showing that more luminous galaxies are more strongly clustered together – evidence for the bias that depends on galaxy properties. **Right**: Galaxy power spectrum measurements (in blue circles) compared to a model of the dark matter power spectrum (black lines). On large scales (linear regime) the two are matched, but on smaller scales (non-linear regime) an anti bias is necessary to match the two – evidence for scale dependent bias. Figure taken from (Cooray and Sheth, 2002).

over an advantage in detecting weaker signals (Peebles, 2001) The power spectrum, and the correlation function, can also be defined for two different tracers to give the *cross power spectrum/correlation function*: $\xi_{ij}(r)$, $P_{ij}(k)$. With these one can describe the correlation between galaxies ξ_g, P_g, matter ξ_m, P_m, as well as the cross-correlation between matter and galaxies ξ_{gm}, P_{gm}.

From the power spectrum in three dimensions $P(k)$, one can derive its projection in two dimensions, the *angular power spectrum*. This is typically done using the Limber (1953) approximation in Fourier space (Kaiser, 1984), and can be written as an integral over the redshift as (Joachimi and Bridle, 2010)

$$C_{ij}(\ell) = \int \mathrm{d}z \frac{H(z)}{c} \frac{p_j(\chi)\, p_i(\chi)}{\chi^2} P\left(z, \frac{\ell + 1/2}{\chi(z)}\right). \tag{2.22}$$

Here, p_i and p_j are the weight functions of the two populations of tracers, which, for example for galaxies, is their redshift distribution.

The right panel of Fig. Fig. 2.2 presents the power spectrum as measured from a galaxy survey (Hamilton and Tegmark, 2002), along with a model of the dark matter one. The power spectrum here is defined as $\Delta(k) = k^3 P(k)/(2\pi^2)$. The galaxy one closely follows the dark matter one on large scales. On small, non-linear, scales there are deviations with more power in the dark matter spectrum. This is an evidence that the galaxy bias is also scale dependent – there is an anti-bias on small scales due to the particularities of the galaxy formation physics. Again, the powerful halo occupation distribution model has to be invoked to correctly model the galaxy power spectrum (Cooray and Sheth, 2002).

Halo mass function

An important statistical description of the collapsed matter in the Universe is the *mass function* of collapsed objects, or the *halo mass function* – this describes the number density of halos at a given mass. The first formalism, based on the spherical collapse model, that models the halo mass function (HMF) was developed by Press and Schechter (1974). The number density of halos depends on the cosmology, most notably on the amplitude of the density fluctuations, which is set by the normalization of the power spectrum, generally denoted σ_8. The HMF obtained from the Press and Schechter formalism is shown in Fig. 2.3 from $z = 0$ to $z = 12$. At all redshifts, it decreases with halo mass while the amplitude increases with cosmic time – in accordance with the hierarchical structure formation scenario. There are numerous HMF models in the literature, from purely analytical, to ones calibrated on N-body and hydrodynamical simulations.

The HMF is an important prediction of the standard cosmological model that can be compared with abundance measurements of clusters and galaxies. Such comparisons can provide constraints on the cosmology, as well as on the connection between galaxies and dark matter halos. The ΛCDM predictions of the HMF and the matter power spectra make the central ingredients in this Thesis used to connect galaxies to the underlying dark matter.

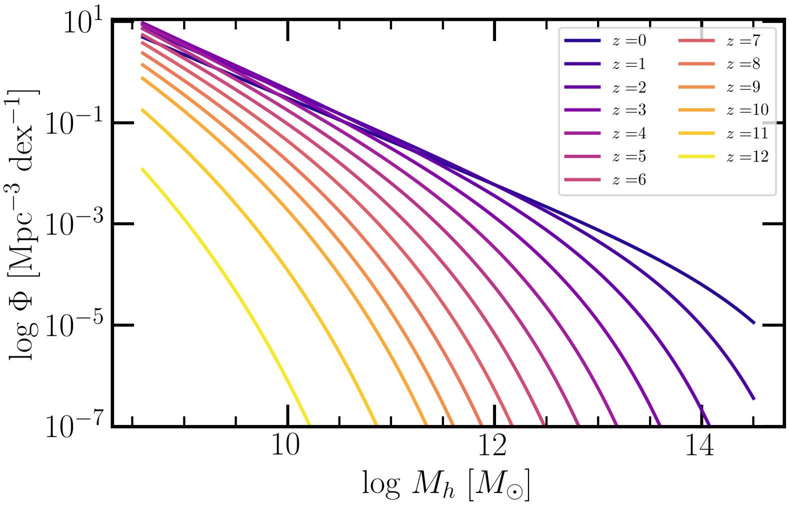


Figure 2.3: Example of the halo mass function by Press and Schechter (1974) from $z = 0$ to $z = 12$.

2.3 FORMATION AND EVOLUTION OF GALAXIES

Gas collapse in halos

Within the current paradigm, galaxies form and evolve within of dark matter halos governed by bayronic physics. Halos provide the potential well in which the baryonic gas can cool down, condense and fragment to form stars and shape the galaxy (White and Rees, 1978). The main physical process responsible for the formation of galaxies is gas cooling. As mentioned earlier, the halo shock heats the gas to the virial temperature, which would settle the gas in hydrostatic equilibrium at low density and large volume. However, baryonic gas is able to collapse by dissipating energy through several physical mechanisms depending on the temperature, density, and composition of the gas: 1) brehmsstrahlung emission of the free electrons in a completely ionized gas at the highest temperatures; 2) electron transition line emission from ions, atoms, and molecules that can be collisionally excited and then radiate to the ground state; 3) inverse Compton scattering of CMB photons by electrons in hot halo gas in the young Universe ($z > 6$). To characterize how fast the gas can cool down in halos, one defines a characteristic *cooling time* that can be written

$$t_{\rm cool} = \frac{3/2\,k\,T}{n_{\rm H}^2\,\Lambda(T,Z)}. \tag{2.23}$$

Here, the numerator represents the energy of the gas, defined by its temperature, and the denominator represents the rate at which the energy is radiated, depending on the number density hydrogen atoms and the cooling function $\Lambda(T,Z)$, which itself is a function of the gas temperature and metallicity. This relation indicates that gas cooling is more efficient with increasing density, thus contributing to a runaway process of gas collapse.

Since the gas is orbiting the center of the potential well of a system with a given mass M and radius R, this provides an internal pressure support counteracting the collapse. The time required for a particle to orbit this system is defined by the *dynamical time* given by

$$t_{\rm dyn} = \sqrt{\frac{1\pi^2\,R^3}{4G\,M}} \tag{2.24}$$

One of the key conditions for gas to be able to beat the hydrostatic equilibrium and actually collapse at the center of the potential well is that the cooling time is shorter than the dynamical time $t_{\rm cool} < t_{\rm dyn}$.

The dark matter halos also provide the initial angular momentum in which the collapsing gas can form a disk shaped by centrifugal forces. Furthermore, if the *star formation time*, i.e., the time that the galaxy would take to run out of its gas by forming stars is longer than the dynamical time and the cooling time, then this will give birth to *disk galaxies*. Inversely, if the star formation time is comparable to the dynamical time, then stars will be formed from the gas during the initial collapse and the galaxy will resemble an *elliptical*. However, the formation of the elliptical galaxies observed throughout the Universe requires other processes, namely *merging*.

Formation of the first stars and galaxies

After recombination, the Universe was completely neutral, dominated by hydrogen and helium, along with the primordial abundances of some lighter elements (i.e., D, Li, Be). Therefore, the very first stars that formed from the fragmentation of the collapsing gas in a halo and had chemical composition lacking elements heavier than H and He (i.e., zero

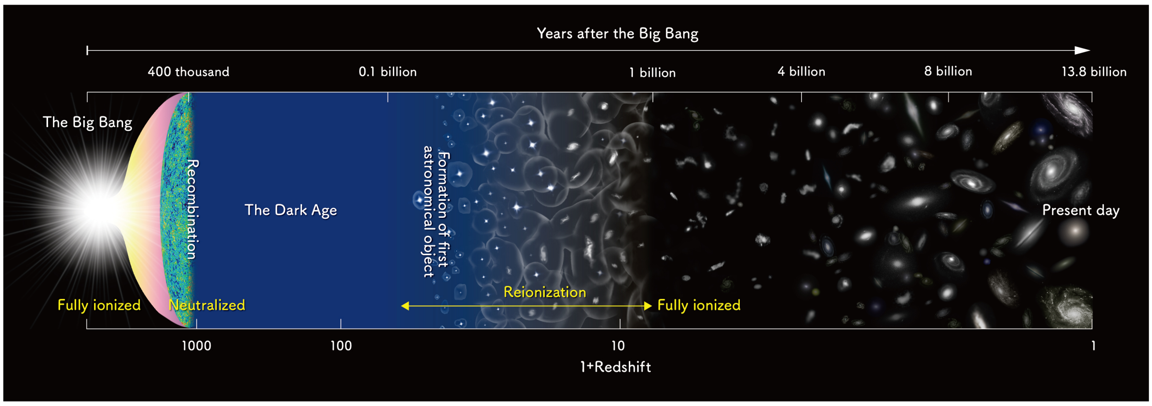
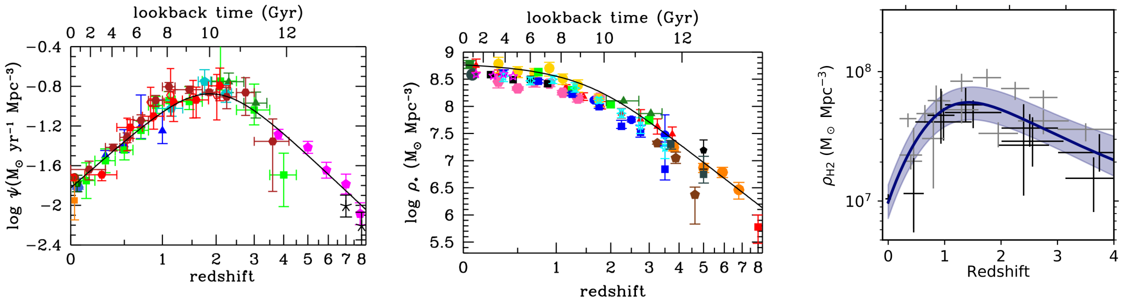


Figure 2.4: Illustration of the timeline of cosmic history. Credit: NAOJ

metallicity $Z = 0$). These are termed Population III (PopIII) stars, are estimated to have formed at $z \sim 20 - 30$ and were much more massive and short-lived than what is possible at present day. These newborn stars were the first sources of large numbers of photons in the UV part of the spectrum, energetic enough to ionize the surrounding neutral hydrogen. The ionizing UV photons create a so-called Strömgen sphere of ionized hydrogen (HII) around the young star. As such, the first galaxies, being a collection of massive, young and hot stars, created ionized bubbles around them that kept growing due to the increasing star-formation activity within them. The expansion of these ionized bubbles around a number of the first galaxies made them percolate and grow to eventually reionize the whole Universe. This process is referred to as *cosmic reionization*, and it is estimated to have been completed at $z \approx 6$. This epoch of phase transition of the baryonic Universe is referred to as the *Epoch of Reionization* and is at the heart of the scientific exploitation of all the major next-generation observatories on the ground and in space, most notably JWST (Robertson, 2021), Euclid and ELT.

PopIII stars, being so massive and short-lived, terminated their life as a supernovae (SNe), thus enriching the surrounding pristine gas with heavier elements and releasing considerable amounts of energy. This feedback from the luminous young stars as stellar winds and SNe heated the gas and prevented further collapse. Therefore, the first galaxies could not have been formed from this first generation of stars in relatively small halos $< 10^6\,M_\odot$ – this is how the physics of reionization sets the minimum mass of a halo to host a galaxy.

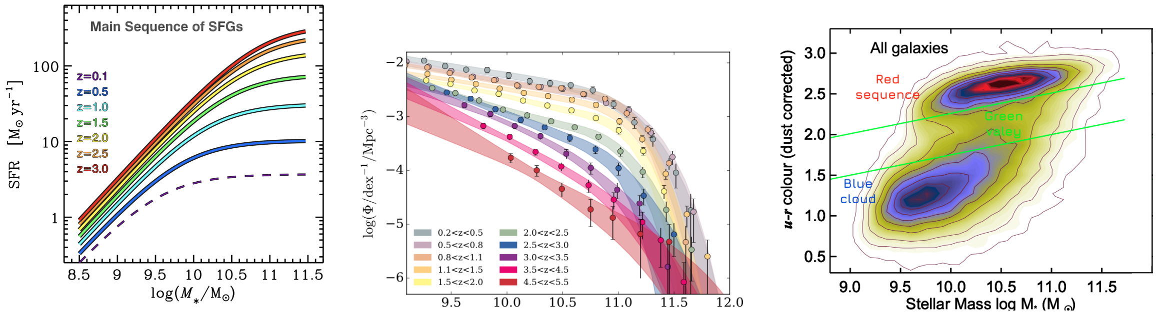
The first galaxies were born some time later in more massive halos whose gravitational well was deep enough to counteract the expulsive energy of stellar winds and SNe from the earliest stars. Such conditions are estimated to have been met in halos of $\sim 10^8\,M_\odot$ at $z \sim 10$. The most distant galaxy that has been spectroscopically confirmed is found at $z = 11$ (Oesch et al., 2016), which corresponds to just ~ 400 Myr after the Big Bang. Analyses of star formation histories of galaxies at $z > 9$ indicate a galaxy formation taking place at $z \sim 15$ (see e.g., Laporte et al., 2021, and references therein) The formation of the first galaxies is still poorly constrained by observations, most notably by the lack of observing capabilities in the rest frame optical at $z > 10$. However, with the successful



(a) Cosmic star formation rate density as a function of redshift from Madau and Dickinson (2014)

(b) Cosmic stellar mass density as a function of redshift from Madau and Dickinson (2014)

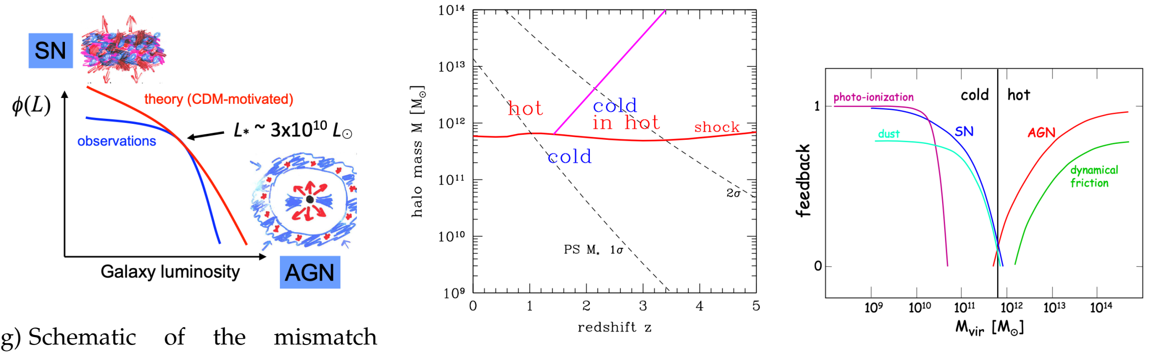
(c) Cosmic density of molecular hydrogen density as a function of redshift from Walter et al. (2020)



(d) Main sequence of star formation in six redshift bins from Förster Schreiber and Wuyts (2020)

(e) Galaxy stellar mass function in ten redshit bins from Davidzon et al. (2017)

(f) Bimodality in the color-stellar mass distribution of galaxies in the local Universe from Schawinski et al. (2014)



(g) Schematic of the mismatch between the galaxy luminosity function and the halo mass function (Silk and Mamon, 2012)

(h) Dominant modes of accretion as a function of halo mass and redshift from Dekel et al. (2009)

(i) Relative strength of various feedback processes at $z = 0$ from Dekel et al. (2009)

Figure 2.5

launch of JWST we now have an observatory that can hunt for the earliest galaxies (see Robertson, 2021, for an overview of the expected discoveries with JWST). Early results from JWST have already provided numerous good candidates for highest redshift galaxies, such as two potential $z \sim 16$ found by Atek, Shuntov et al., 2022

Cosmic star formation

Stars have been forming in the Universe at different rates in its history. A significant observational endeavor has been to measure the *cosmic star formation history*, i.e., the rate at which stars have been forming in the Universe as a function of redshift, expressed in units of $M_\odot\, yr^{-1}\,\mathrm{Mpc}^{-3}$. Various measurements of the star formation rate densities from surveys in the UV and IR – capturing all light emitted by star formation tracers from young stars in the UV to dust emission in the IR – converge towards a picture where the star formation rate density (SFRD) increased with cosmic time by a factor of > 10 from $z \sim 8$ to $z \sim 2$ where it peaks and then decreases again by a factor of 10 to today (see e.g., Madau and Dickinson, 2014; Bouwens et al., 2020). This evolution also corresponds to the cosmic evolution of the density of molecular hydrogen (see e.g., Walter et al., 2020, and references therein), a molecule that is a key ingredient in galaxy formation and its evolution, since stars form in the dense molecular phase of the gas (see e.g, Kennicutt and Evans, 2012, for a review). Fig. 2.5 shows the evolution of the SFRD (left) and molecular hydrogen density (right) from a compilation of measurements. Even though various observations are converging towards the same evolution of the SFRD, the picture at high redshifts remains less certain, mainly due to the difficulty of observing unbiased galaxy samples and correctly accounting for the effects of dust obscuration, as well as determining the relative contribution of various galaxy populations that might be optically invisible (Wang et al., 2019; Bouwens et al., 2020).

The stellar mass content in the Universe at a given moment is an accumulation of stars from earlier episodes of star formation. Therefore, one can obtain the redshift evolution of the stellar mass density (SMD) by integrating the SFRD while accounting for stellar loss due to stellar winds, SNe etc. One of the most remarkable successes of the theory and observations is that the SMD from the integrated SFRD is consistent with independent measurements of the SMD by measuring galaxy abundances as a function of stellar mass. This is shown in Fig. 2.5 (middle) showing that most stellar mass build-up happened at $z > 1$, with a fast increase by a factor of 100 from $z \sim 4$ to $z \sim 1$, and only marginal increase since.

Main sequence of star formation

The star formation rate (SFR) in individual galaxies (in units of $M_\odot\, yr^{-1}$) shows a tight linear correlation with the stellar mass – the *main sequence* of star formation (Daddi et al., 2007; Noeske et al., 2007; Elbaz et al., 2007). The main sequence hosts galaxies that are actively forming stars and are found to have disk-like morphology (see e.g., Wuyts et al., 2011). Most of the stars in the Universe have formed in the main sequence, with up to 80% of present day stars since $z \sim 2.5$. Galaxies lying above the main sequence are termed *starbursting* and have disturbed morphologies – an indication that their burst of star formation is due to a recent merging event. Below the main sequence, galaxies typically have large stellar masses, low star-formation rates and elliptical morphologies and make the *passive* population.

The main sequence of the SFR-M_* increases with redshift (Fig. 2.5d), meaning that earlier galaxies were forming stars more vigorously. This increase is accompanied by an increase in the available gas fraction for star formation (Daddi et al., 2010). This fact, along with the tightness of the main sequence, indicates that most of the star formation comes from gradual gas accretion and its transformation into stars, and only a limited importance of merger-driven starbursts (Elbaz et al., 2011; Rodighiero et al., 2011). However, pinpointing the exact contributions of different mechanism to fuel star formation, and how it relates to the dark matter environment of galaxies remains an active research area.

The linear increase of the SFR with stellar mass is interrupted at a certain mass scale of about $10^{10}\,M_{\odot}$ in the local Universe, after which it remains constant. At earlier times, this bending moves towards higher masses of about $10^{11}\,M_{\odot}$ at $z \sim 4$. This fact that the SFR starts to 'drop-off' from the main sequence as the mass increases is a signature of *quenching* of star-formation. The physical processes that contribute to this flattening and the onset of quenching are actively researched, and most recent investigations show that this is related to how fresh cold gas is funneled into massive and hot halos (Daddi et al., 2022a).

Bimodality of galaxy properties

Galaxies show a robust bimodal distribution in a number of their properties, generally being divided into two classes: "red" and "blue". Characteristic stellar and halo masses are typically associated to these two classes at about $M_{*} \sim 3 \times 10^{10}\,M_{\odot}$ and $M_{\mathrm{h}} \sim 10^{12}\,M_{\odot}$ (see e.g., Dekel and Birnboim, 2006, for a comprehensive summary).

In terms of colors, the bimodality is typically observed in $u - r$ rest-frame colors, where the *red sequence* and the *blue cloud* are separated by a *green valley* at about $u - r \sim 2$ (Fig. 2.5f). In terms of star-formation activity, red sequence galaxies are typically passive, populated by old stellar populations and gas-poor reservoirs, whereas blue galaxies are star-forming and gas-rich. In terms of morphology, red galaxies tend to be spheroidal while blue galaxies are disks. This dichotomy is also exhibited in the environmental dependence, with red sequence galaxies residing in more dense environments, or nodes of the cosmic web; blue galaxies tend to live in less dense environments, or in voids and filaments (Malavasi et al., 2017; Laigle et al., 2018; Kraljic et al., 2018). The environment density being strongly correlated with the host halo mass, this translates to halo masses of $\sim 10^{12}\,M_{\odot}$ below/above which these two populations (blue/red) are typically hosted.

The presence of this bimodality, including the low-density green valley, places a strong constraint on galaxy evolution. It indicates that blue sequence galaxies quench their star formation, transition through the green valley rather rapidly while burning out their remaining gas supply and settle into the red sequence. Once quenched and poor in cold gas, galaxies can further increase their mass via dry merging, leading to massive galaxies with old stellar populations formed at earlier times (Faber et al., 2007; Arnouts et al., 2007). The exact mechanisms that govern the green valley transition depends on the morphology of the galaxy, being rapid for early-type galaxies and long for late-type galaxies (see e.g., Schawinski et al., 2014). Moreover, galaxies undergo morphological changes during this transition, such as bulge growth and disk reddening (see e.g., Quilley and de Lapparent, 2022). Therefore, studying the transformation of galaxy properties as they transition from the red and blue sequences is a crucial aspect in understanding how galaxies assemble their stellar mass and quench.

Galaxy stellar mass function

One of the fundamental statistical measures of the stellar mass build-up in the Universe is the galaxy stellar mass function (GSMF). The stellar mass function $\Phi(M_*)$ measures the number of galaxies per logarithm of stellar mass per comoving volume and is expressed in units of $\mathrm{dex}^{-1}\,\mathrm{Mpc}^{-3}$. The way it evolves with redshift informs us about how galaxies have assembled their stellar mass. The integral of the GSMF results in the SMD, and is a measure of the integrated SFRD, as mentioned earlier. Precise measurements of the GSMF down to the earliest times are crucial to constrain galaxy evolution models, especially the effects of different physical processes that fuel and quench star formation. Fig. 2.5e presents the evolution of the GSMF from $z = 0$ to $z = 5.5$ in ten redshift bins. The number density of galaxies decreases with stellar mass as a power law down to the characteristic 'knee' at $M_* \sim 10^{11}\,M_\odot$, after which it drops exponentially. At lower masses, the slope of the GSMF steepens with redshift and the knee flattens. The evolution of the low mass end is more rapid than at high masses – this is an evidence of physical processes in massive galaxies and halos that suppress the growth of the massive galaxies. Additionally, the increase in number density for massive galaxies is rapid from $z \sim 4$ to $z \sim 2$, followed by only a marginal increase to $z \sim 0$, indicating that massive galaxies assembled their stars relatively early. The fact that low mass galaxies evolve more rapidly at $z < 2$ indicates that low mass galaxies formed their stars later and over a longer timescale – this is referred to as *downsizing* (De Lucia et al., 2006; Thomas et al., 2010). The evolution of the shape of the GSMF, especially when compared to that of the HMF, places strong constraints on how stellar mass can assemble in galaxies. There is a significant mismatch at lower redshifts at masses below and above the characteristic knee – there is deficiency in the number of both low and high mass galaxies compared to the number density of halos in which they are hosted. This is illustrated in Fig. 2.5g. To explain this mismatch, quenching mechanisms need to be invoked that act in low- and high-mass halos. At earlier times, the GSMF resembles a steep power law and is a better match to the HMF at intermediate and high masses – a clue that these quenching mechanisms, at least those in massive halos, were significantly less efficient in the early Universe (Davidzon et al., 2017).

Physical processes that regulate star formation and their relation to dark matter halos

The way galaxies assemble their stellar mass is regulated by processes that are closely related to their host halos. These processes include the hierarchical merging of halos, the accretion and cooling of baryonic gas and formation of stars that make galaxies grow. These are in competition with processes that quench growth by heating the gas and/or blowing it away, such as stellar, supernovae and active galactic nuclei feedback, as well as environmental quenching effects like harassment, stripping etc.

According to the current picture of galaxy formation, most of the stellar mass in galaxies has been assembled through star-formation, mainly driven by *baryon accretion*. Mergers are estimated to play only a minor role, especially major mergers, which are rare. Minor mergers are several times more frequent and are the second most important contributor; they are also estimated to play a more significant role at $z > 2.5$ (Genel et al., 2009; Bouché et al., 2010; Tacconi, Genzel and Sternberg, 2020). Increased gas accretion rates are also responsible for the peak of the cosmic SFRD discussed earlier.

The baryonic accretion can happen through two modes. *Hot mode* accretion refers to the scenario in which the infalling gas is shock heated at the virial radius, then dissipates energy and condenses at the center of the halo creating a disk. However, this mode alone can not explain the observed galaxy growth, scaling relations and bimodality in their properties. *Cold mode* accretion very efficiently funnels cold gas (as well as angular momentum) directly in the galaxy without getting shock heated, and is the dominant gas accretion mode (Kereš et al., 2005; Dekel et al., 2009). Cold flows happen through narrow and collimated streams along the cosmic web. Most importantly, the cold flows allow early and efficient star formation, even in massive and hot halos. Fig. 2.5h shows that the boundary that separates the regimes of cold flows in hot halos and purely hot accretion increases to more massive halos at earlier redshifts. The cold flows paradigm is crucial in explaining numerous observations of galaxy properties (Dekel et al., 2009). The role of the cold flows in shaping the relation between the assembled stellar mass in halos and the halo mass is one of the subjects of investigation in this Thesis.

However, the galaxy formation efficiency, which refers to the efficiency of the overall process of transforming gas into stars, is rather low, with only up to $\sim 20\%$ of all the available baryons converted into stars. This fraction peaks at halo masses of about $10^{12}\,M_{\odot}$ and steeply decreases at lower and higher masses – a trend present to high redshifts (see e.g., Behroozi, Wechsler and Conroy, 2013). Such a low efficiency is explained by the presence of powerful feedback mechanisms that prevent gas cooling and regulate star formation. Current studies are inconclusive regarding the redshift evolution of the galaxy formation efficiency, mainly due to the difficulty of observing galaxy properties (such as GSMF and clustering) to high redshifts in homogeneous surveys with consistent selection functions and methodology of estimating galaxy physical parameters (Behroozi, Conroy and Wechsler, 2010). Studying the galaxy formation efficiency across the majority of cosmic history from observations is the central scientific motivation of this Thesis.

Quenching and feedback

Various aspects of the observed distributions of galaxy properties – from the bimodality of blue, SF galaxies quickly transitioning to the red passive sequence, to the Schechter function form of the GSMF and its mismatch compared to the HMF – require some physical mechanisms to *quench* star-formation. The strength and the type of the physical process responsible for quenching is strongly dependent on both stellar and halo mass. Quenching, that happens as a result of an outpouring of energy into the ISM thus heating it or expelling it, comes from processes related to star-formation and active galactic nuclei activity and are called *feedback* processes. To explain quenching, typically two broad classes of feedback processes are invoked that act in different halo mass regimes.

Stellar feedback is the dominant mechanism in halos $< 10^{12}\,M_\odot$, and refers to injection of energy from processes related to star formation such as stellar winds, supernovae (SNe), radiation pressure and photoheating (Hopkins, Quataert and Murray, 2012; Somerville and Davé, 2015). Recent studies based on hydrodynamical studies reveal that magnetic fields and cosmic rays on galactic scales are a key ingredient in the feedback budget in order to account for the observed stellar mass in low mass galaxies and halos (Naab and Ostriker, 2017).

Active galactic nucleii (AGN) feedback in halos more massive than $10^{12}\,M_\odot$ is considered the dominant mode, and is crucial to explain the observed mass function at the high mass end (Croton et al., 2006; Bower et al., 2006; Dubois et al., 2014). Typically, large halos above this mass scale form a hot gas corona by shock heating the gas to the virial temperature. This prevents the gas from cooling, and shock heats any inflowing cold gas from the intergalactic medium (IGM) (referred to as *halo preventive feedback* or *hot halo* mode). However, an additional source of energy is required to prevent the gas from cooling and to quench the star-formation to the rate that is observed in surveys (Gabor and Davé, 2015). This additional source of energy comes from AGN, which are powered by an accreting black hole that can come in two main modes.

- *Radiative mode*, also referred to as *bright* or *quasar* mode, happens when the AGN is accreting near the Eddington rate[2] and releases high amounts of energy that can photoionize and heat the gas. Additionally, quenching through this mode is also called *outflow quenching* because the AGN can drive powerful winds that can remove cold gas, thus quenching star-formation. According to popular scenarios, this can happen after a gas-rich major merger that ignites the quasar-mode AGN, after which the remnant is an elliptical galaxy (Hopkins et al., 2006).
- *Mechanical mode*, also referred to as *radio* and *jet* mode, happens when the AGN releases the energy in collimated jets. The jets release enough energy to heat up the gas and keep it hot for a very long time, resulting into a so-called *maintenance mode feedback*. The radio mode is thought to be dominant in massive early type galaxies, in groups and clusters (Somerville and Davé, 2015). Additionally, recent hydrodynamical simulations show that the radio mode is also predominantly responsible for quenching following mergers (Rodríguez Montero et al., 2019).

Observationally, AGN activity, especially the radio mode, has been detected in nearby clusters by the presence of X-ray cavities coincident with radio lobes (e.g., Churazov et al., 2001). In normal galaxies, these are typically found based on radio observations out to high redshifts (e.g., out to $z \sim 6$ Delvecchio et al., 2017). Most recently, AGN activity has been detected in massive quiescent galaxies out to $z \sim 5$, supporting the paradigm of AGN induced quenching (Ito et al., 2022).

Within large halos, considerable stellar mass is deposited in satellite galaxies that have been accreted via minor mergers. Additional to the feedbacks presented above, satellites

---

[2] The Eddington rate refers to the maximum luminosity that the AGN reaches when there is a balance between the gravitational and radiation force.

are also subject to various environmental quenching mechanisms that act to suppress further growth (Peng et al., 2010; Gabor and Davé, 2015). One of the main responsible mechanisms is the hot halo mode, where the infalling gas is heated by virial shock heating (Birnboim and Dekel, 2003). However, additional environmental mechanisms act efficiently on satellites as strangulation (Larson, Tinsley and Caldwell, 1980; Balogh, Navarro and Morris, 2000), ram-pressure stripping (Gunn and Gott, 1972), and harassment (Moore et al., 1996).

However, many details regarding the role of various feedback mechanisms still remain poorly understood. The main method to study them is through hydrodynamical simulations that, although hugely insightful, remain limited by important caveats such as numerical resolutions and subgrid physics (see Chapter 5). Finally, it remains unclear whether quenching is mainly driven by halo imposed conditions – i.e., via the interplay of AGN feedback suppressing gas cooling and cold flows fueling star-formation – or galaxy-internal processes. This Thesis aims at shedding light at the halo imposed conditions using measurements of galaxy properties from observations.

Evolution of clusters

Clusters are the most massive collapsed structures in the Universe, lying at the intersections of filaments of the cosmic web. They represent a collection of galaxies with numbers that can range from tens to thousands of various types and morphologies. Galaxies inside clusters, however, have distinct properties and assembly histories shaped by these dense environments. For example, cluster galaxies tend to form earlier compared to those of same mass in less dense environments at same redshift, are more massive, passive and elliptical. Clusters have undergone their rapid formation phase at $z > 2$, when the bulk of the stellar mass was assembled. During this phase at $z > 2$ cluster galaxies were rich in gas and dust and have undergone a phase of vigorous star-formation activity, followed by a transformation into massive, quenched and elliptical galaxies observed in local clusters (e.g., Thomas et al., 2005) As such, high-z clusters are expected to be the first place hosting the first generation of quiescent galaxies (e.g., Kalita et al., 2021).

In the early universe, the most massive and overdense structures are often called galaxy protoclusters (see Overzier, 2016, for a review). Simulations suggest that the star formation in protoclusters contribute significantly to the cosmic SFRD, e.g., $20 - 50\%$ at $z > 3$ (Chiang et al., 2017). Observational constraints are still rare, but several protoclusters detected at $z \sim 4$ reveal a collection of dusty star-forming galaxy (DSFG) with an enormous, collective, dust-obscured SFR of $\sim 6000\ M_{\odot}\ \mathrm{yr}^{-1}$ (Oteo et al., 2018; Miller et al., 2018) This is because the most starbursting cluster members are often severely attenuated by dust, making it difficult to detect and measure SFR using UV/optical observations (Wang et al., 2019).

Another importance of $z > 3$ clusters is to test the paradigm of cold streams penetrating the shocks at the virial radii of massive, hot halos. (Proto)clusters are uniquely suited because their halos encompass large enough mass ranges of $M_h > 10^{13}\ M_{\odot}$ at $z \sim 3$ (Overzier, 2016). Only very recently, observational evidence of cold streams have been obtained by observations of collisionally excited Lyα emission around groups and clusters at $z \sim 3$

(Daddi et al., 2022a). However, this recent work suffers from small and inhomogeneous samples. One of the subjects in this Thesis is the search for (proto)clusters at $z > 3$ in wide and multiband surveys.

3

THEORETICAL FRAMEWORK TO STUDY GALAXIES IN RELATION TO THE LARGE-SCALE STRUCTURE

3.1 HALO MODEL OF THE LARGE-SCALE STRUCTURE

According to the current paradigm of the formation of the large-scale structure and of galaxies, presented in Chapter 2, overdensities with $\rho/\bar{\rho} \sim 200$ collapsed and virialized into dark matter halos. The *halo model* of the LSS postulates that all matter in the Universe is contained in these halos. As such, the halo model provides a very powerful formalism to model some essential observables – correlation functions and power spectra (see e.g., Mo and White, 1996). This is made possible by the fact that the properties of halos, such as their masses, profiles, spatial distribution (clustering) etc., depend only on dark matter[1]. These properties of the dark matter can be studied in detail using N-body simulations within the context of the ΛCDM model. The halo model, thus, offers a way to study the matter distribution in two steps: on large scales, the distribution of dark matter halos, and on small scales the distribution of matter within the halos. On the other hand, the spatial distribution of galaxies closely follows that of the dark matter halos. However, their properties strongly depend on the complex baryonic physics such as cooling, star-formation and feedback that is expensive to model. Using the halo model formalism, combined with a model that describes statistically how galaxies populate halos, one can model various statistics of the galaxy distribution (see e.g., Seljak, 2000; Peacock and Smith, 2000; Scoccimarro et al., 2001). This provides a very powerful framework to study both the LSS and galaxy evolution by analyzing the statistics of the spatial distribution of galaxies.

In this model, where all matter in the Universe is partitioned in halos, density at a given point in space $\vec{x}$ can be written as a sum of the contributions from each i-th halo

$$\rho(\vec{x}) = \sum_i \rho_{\mathrm{h}}(\vec{x} - \vec{x}_i) = \sum_i m_i u(\vec{x} - \vec{x}_i). \tag{3.1}$$

Here, $\rho_{\mathrm{h}}(\vec{x})$ is the density profile of a halo, while $u(\vec{x})$ is a normalized halo profile – essentially ρ_{g} divided by the total halo mass m. One of the most widely adopted functional

---

[1] at least down to the very small scales.

forms for the halo profile is the so-called Navarro-Frenk-White (NFW) profile (Navarro, Frenk and White, 1997)

$$\rho_{\mathrm{h}}(r) = \frac{\rho_s}{\frac{r}{r_s}\left(1+\frac{r}{r_s}\right)^2}. \tag{3.2}$$

The halo density profile here is written in terms of the radial distance from the center r. This relation is parametrized by two scale parameters ρ_s – the amplitude of the density and r_s – a characteristic scale radius. A halo has to be defined to a certain radius R_Δ, which is typically the radius within which the halo has a specified overdensity compared to a reference density. This typically is, for example, 200 times the critical density, 200 times the matter density or the virial radius. The scale factor is related to the halo radius definition as $R_\Delta = c(M)\, r_s$, where $c(M)$ is known as a concentration parameter, and is dependent on mass, so it is also called a concentration-mass relation. The mass of the halo, the normalization of the density ρ_s, the scale radius and the concentration parameter are then related as

$$m = 4\pi\, r_s^2\, \rho_s \left[\log(1+c) - \frac{c}{1+c}\right]. \tag{3.3}$$

Correlation function and power spectrum

Using this description of the matter density in the Universe in terms of halos, and the definition of a 2PCF Eqn. 2.18 one can derive the two-point correlation function. The correlation function can be decomposed into contributions from densities within the same halo (*one-halo* term) on small scales, and from contributions from two different halos (*two-halo* term) on large scales

$$\xi(r) = \xi^{1\mathrm{h}}(r) + \xi^{2\mathrm{h}}(r). \tag{3.4}$$

The two terms can be written as (for a more complete derivation, see Cooray and Sheth, 2002, and references therein)

$$\begin{aligned}
\xi^{1\mathrm{h}}(\vec{x}-\vec{x}') &= \int \mathrm{d}m \frac{m^2 n(m)}{\bar{\rho}^2} \int \mathrm{d}^3\vec{y}\, u(\vec{y}|m)\, u(\vec{y}+\vec{x}-\vec{x}'|m) \\
\xi^{2\mathrm{h}}(\vec{x}-\vec{x}') &= \int \mathrm{d}m_1 \frac{m_1 n(m_1)}{\bar{\rho}} \int \mathrm{d}m_2 \frac{m_2 n(m_2)}{\bar{\rho}} \int \mathrm{d}^3\vec{x}_1\, u(\vec{x}-\vec{x}_1|m_1) \\
&\quad \times \int \mathrm{d}^3\vec{x}_2\, u(\vec{x}'-\vec{x}_2|m_2)\, \xi_{\mathrm{hh}}(\vec{x}_1-\vec{x}_2|m_1,m_2).
\end{aligned} \tag{3.5}$$

In these relations, n(m) is the number density of halos, or the halo mass function, and $\bar{\rho} = \int \mathrm{d}m\, m\, n(m)$ is the mean density. The two-halo term contains the 2PCF of halos of mass m_1 and m_2. Halos can also be considered as biased tracers of the underlying matter, so their correlation function can also be related to the matter one by

$$\xi_{\mathrm{hh}}(r|m_1,m_2) \approx b_h(m_1)\, b_h(m_2) \xi_{\mathrm{m}}(r), \tag{3.6}$$

where $b(m_1)$ and $b(m_2)$ are halo bias functions. On large scales, the halo 2PCF approaches the linear 2PCF, $\xi^{2\mathrm{h}} \approx \xi_{\mathrm{lin}}$, which provides a crude but convenient modeling.

The integrals in Eqn. 3.5 involve convolutions of two similar halo profiles $u(r|m)$, weighted by the total number density of halo pairs. These convolutions are easier to work with in Fourier space, where they become multiplications of the Fourier transforms of the halo profiles. As for the 2PCF, the power spectrum can be written as a sum of the one-halo and two-halo terms

$$P(k) = P^{1\mathrm{h}}(k) + P^{2\mathrm{h}}(k), \tag{3.7}$$

where the two components are written as

$$\begin{aligned} P^{1\mathrm{h}}(k) &= \int \mathrm{d}m\, n(m) \left(\frac{m}{\bar{\rho}}\right)^2 |u(k|m)|^2, \\ P^{2\mathrm{h}}(k) &= \int \mathrm{d}m_1\, n(m_1) \left(\frac{m_1}{\bar{\rho}}\right) |u(k|m_1)| \\ &\quad \times \int \mathrm{d}m_2\, n(m_2) \left(\frac{m_2}{\bar{\rho}}\right) |u(k|m_2)|\, P_{\mathrm{hh}}(k|m_1, m_2). \end{aligned} \tag{3.8}$$

In these equations, $u(k|m)$ is the Fourier transform of the dark matter distribution in a halo of mass m, and $P_{\mathrm{hh}}(k|m_1, m_2)$ is the power spectrum of halo pairs of masses m_1 and m_2. As similarly as for the 2PCF, the halo-halo power spectrum can also be written in terms of the halo bias and the linear power spectrum

$$P_{\mathrm{hh}}(|m_1, m_2) \approx b_h(m_1)\, b_h(m_2) P_{\mathrm{lin}}(k). \tag{3.9}$$

The linear power spectrum, governing the very large scales, is a good approximation that can be modeled from perturbation theory (see Chapter 2).

Having presented how the clustering properties of dark matter halos can be determined, we now turn our attention to the more complex inhabitants of these halos – galaxies.

3.1.1 *Halo occupation distribution*

The halo model formalism provides a powerful description of the clustering properties of dark matter halos. Most importantly, it provides a framework in which to study the statistical properties of galaxies, which is our main interest. This is possible within the paradigm that the halo properties, such as mass and angular momentum, determine the efficiency of the galaxy formation processes, such as gas accretion and cooling rate, star formation rate etc. All these processes shape how galaxies are distributed inside halos (Peacock and Smith, 2000).

The halo occupation distribution (HOD) framework has been developed to describe, statistically, how galaxies populate dark matter halos. Within the HOD the number distribution of a given class of galaxies in halos (or biasing of galaxies) is fully defined by a probability distribution of the number of galaxies residing in halos conditioned on some criteria, usually on the mass $P(N|M_{\mathrm{h}})$ (Berlind and Weinberg, 2002). To determine this probability distribution, Kravtsov et al. (2004) and Zheng et al. (2005), have shown, using N-body and smoothed particle hydrodynamics simulations, that one can separate the contributions from the central and satellite galaxies inside the same halo. Typically, centrals are assumed to follow a Bernoulli distribution, while the number of satellites follows a Poisson distribution. Under these assumptions, the HOD is specified by the average number of galaxies in halos of a given mass $\langle N(M_h)\rangle$, that has a central and satellite component written as

$$\langle N_{\rm tot}(M_h)\rangle = \langle N_{\rm cent}(M_h)\rangle + \langle N_{\rm sat}(M_h)\rangle. \tag{3.10}$$

The central and satellite mean occupation distributions are given by

$$\begin{aligned} \langle N_{\rm cent}\,(M_h)\rangle &= \frac{1}{2}\left[1+{\rm erf}\left(\frac{\log M - \log M_{\rm min}}{\sqrt{2}\,\sigma_{\log M}}\right)\right], \\ \langle N_{\rm sat}\,(M_h)\rangle &= \langle N_{\rm cent}\,(M_h)\rangle \left(\frac{M-M_0}{M_1'}\right)^{\alpha}. \end{aligned} \tag{3.11}$$

Defined as such, the central mean occupation distribution is an error function that reaches unity – there can be at most one central in a massive halo, while the satellites follow a power law that asymptotes to zero at a certain halo mass scale. This parametrization of the mean occupation numbers has 5 free parameters:

– $M_{\rm min}$, which sets the characteristic minimum mass in which a halo can host a central galaxy, it is defined such that $N(M_{\rm min})\rangle = 1/2$

– $\sigma_{\log M}$ is a characteristic width of the cutoff profile when approaching lower masses. This parameter describes the fact that in a halo of fixed mass, there can be a scatter in the galaxy stellar mass due to different physical processes that govern galaxy formation.

– α is the slope of the power law that governs the mean number of satellites in a halo

– M_0 is a characteristic mass scale below which no satellites can be found in halos

– M_1' sets the normalization.

Fig. 3.1 shows an example of the HOD implementation by Coupon et al., 2012 to model the angular correlation function in Canada-France-Hawaii Telescope Legacy Survey (CFHTLS)-Wide. The left panel shows the measurements and the best fit model, decomposed into the one-halo and two-halo terms. The one- and two-halo terms combine in such a way that their sum results in a power law (Watson, Berlind and Zentner, 2011). The right panel shows the mean occupation distribution as a function of halo mass, decomposed into central and satellites.

One of the advantages of the HOD modeling is that it distinguishes galaxies by central and satellites. This allows to study satellite galaxy properties, that can be shaped by different physical processes that act on satellites. For example, the fraction of satellite galaxies is a valuable constraint on galaxy formation models (Zheng et al., 2005). This parametrization has had great success in modeling the observed galaxy correlation function in SDSS (see Zheng, Coil and Zehavi, 2007; Zehavi et al., 2011, e.g.,), as well as numerous galaxy surveys since (see e.g., Coupon et al., 2012; McCracken et al., 2015; Harikane et al., 2016; Ishikawa et al., 2020). However, other parametrizations also exist, and for the main work of this thesis we will use a different one that will be described in more detail in Chapter 4.

Having a relation of how galaxies populate halos in terms of mean occupations, and how halos are distributed in the LSS, the models for the power spectrum of galaxies can

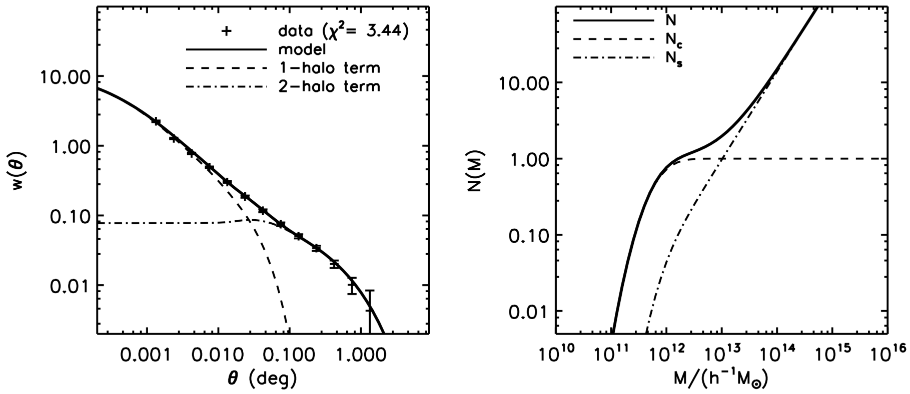


Figure 3.1: Example of an HOD model fit on measurement of the angular 2PCF from CFHTLS by Coupon et al., 2012. The left panel shows the measurements and the best fit model, decomposed into the one-halo and two-halo terms. The right panel shows the mean occupation distribution as a function of halo mass, decomposed in central and satellites.

be written similarly to Eqn. 3.5. Again, it can be separated in contributions from clustering of galaxies within the same halo (1-halo term) and between different halos (2-halo term)

$$P_{\rm gg}(k) = P^{\rm 1h}_{\rm gg}(k) + P^{\rm 2h}_{\rm gg}(k). \tag{3.12}$$

The 1-halo and 2-halo terms are given by

$$\begin{aligned} P^{\rm 1h}_{\rm gg}(k) &= \frac{1}{\bar{n}_g^2}\int {\rm d}M_h \frac{{\rm d}n}{{\rm d}M_h}\left[\langle N_{\rm sat}\rangle^2 u_s^2(k) + 2\langle N_{\rm sat}\rangle\, u_s(k)\right], \\ P^{\rm 2h}_{\rm gg}(k) &= \frac{1}{\bar{n}_g^2}\left[\int {\rm d}M_h \frac{{\rm d}n}{{\rm d}M_h} b_h(M_h)\left[\langle N_{\rm cent}\rangle + \langle N_{\rm sat}\rangle\, u_s(k)\right]\right]^2 P_{\rm lin}(k). \end{aligned} \tag{3.13}$$

In these equations,

$$\bar{n}_g = \int {\rm d}M_h \frac{{\rm d}n}{{\rm d}M_h}(\langle N_c\rangle + \langle N_s\rangle) \tag{3.14}$$

is the mean number density of galaxies and $b_h(M_h)$ is the large-scale halo bias. $u_s(k)$ is the Fourier transform of the density profile of galaxies in a halo. Typically, the density profile of galaxies in a halo can be assumed to follow the NFW profile Navarro, Frenk and White, 1997 with a specified mass-concentration relation. This is an approximation and the density profile of galaxies might slightly differ from the overall mass-concentration relation due to e.g., subhalo segregation, tidal stripping, dynamical friction etc (Watson et al., 2012).

Finally, one can compute the 2PCF from the power spectrum by Fourier transformation

$$\xi_{\rm gg}(r) = \frac{1}{2\pi^2}\int {\rm d}k\, k^2\, P_{\rm gg}(k)\frac{\sin kr}{kr}. \tag{3.15}$$

In this thesis we will mostly be working with angular 2PCF that can be computed by projecting the galaxy power spectrum along the redshift distribution of galaxies $p(z)$ for which we measure it. This is done using the Limber, 1953 approximation, following Bartelmann and Schneider, 2001

$$w(\theta) = \int \mathrm{d}\, z p^2(z) \left(\frac{\mathrm{d}\chi}{\mathrm{d}z}\right) \int \mathrm{d}k \frac{k}{2\pi} P_{\mathrm{gg}}(k) J_0(\chi\theta k), \tag{3.16}$$

where J_0 is the zeroth-order Bessel function of first kind, and $\chi(z)$ is the comoving distance

The important thing to note in the expressions is that the galaxy power spectrum is defined in terms of the occupation distributions of centrals and satellites $\langle N_c \rangle$ and $\langle N_s \rangle$ specified by Eq. 3.11. It is also independent of the exact parametrization of the HOD, as we will see in Chapter 4 where we will employ a different one. Therefore, this HOD formulation of the bias (or how galaxies occupy dark matter halos) is complete and sufficient to describe the clustering of galaxies and their relation to that of dark matter. In the words of Berlind and Weinberg: *"The most important strength of the HOD formulation of bias is its completeness... the HOD tells us everything a theory of galaxy formation has to say about the statistics of galaxy clustering, in real space and redshift space, and on small, intermediate and large scales"*. As an example, the authors state *"It would be more illuminating to learn, for example, that a given theory predicts too many red galaxies in halos of mass* $10^{13} - 10^{14}\, M_\odot$ *than to learn that it predicts the wrong 3-point correlation function of such galaxies"*. The HOD is also one of the central ingredients in semi-analytical models where halos from N-body simulations are populated in galaxies according to an HOD – a methodology that is key in studying galaxy formation and preparing future surveys. Furthermore, the HOD also provides a framework to consistently model other observables of the LSS and galaxies such as weak lensing (§ 3.2), thus allowing for joint analysis along with clustering. However, this analytical modeling using the HOD also has its limitations. For example, it has been shown to underpredict the clustering of galaxies on intermediate-to-large scales ($\sim 1 - 20$ Mpc/h (Mead et al., 2015; Mead et al., 2015; Hadzhiyska et al., 2020), but corrections do exist (Mead et al., 2021). The topic of this thesis being the study of galaxy properties in relation to their dark matter environment by combining several observables, the HOD formulation, then, makes a natural choice for methodology in this thesis.

3.2 WEAK GRAVITATIONAL LENSING

Photons coming from distant objects are deflected by the gravitational potential of intervening matter structures. This effect is called *gravitational lensing* since it draws similarities with the deflection of light in media with different optical refractive indices. Light emitted from objects with extended apparent size (e.g., from a galaxy) travel in bundles. Light bundles are subjects to tidal gravitational field due to the non-vanishing gradient[2] of the gravitational potential around the inhomogeneous matter distribution. Such light bundles are therefore differentially deflected, creating distortions of the images of resolved extended objects (e.g., galaxies) that we can observe. The effects of lensing are geometrical, and photons can be neither created nor destroyed in the process, thus conserving the surface

---

[2] Strictly speaking, the distortions on the light bundles come from the 2nd derivatives of the potential.

brightness of the lensed object. But, having the shape and size changed means that the flux of the object that we measure is also changed. This can push the flux of distant galaxies over the detection limit, thus making faint and distant galaxies observable (see e.g., Lotz et al., 2017). Since this is purely gravitational effect, it is the total matter content both *dark* and *baryonic* that causes the deflection. As we will see in a moment, depending on the spatial configuration of the source-lens system, the light deflection can be so severe that it creates multiple images of the background source. In other cases, the light deflection is so subtle that the gravitational effects are only measurable statistically over an ensemble of galaxies. Be it *strong* or *weak*, this gravitational effect has opened a new observational window of the largest scales and has revolutionized our understanding of the Universe, especially its dark sector.

Of particular interest in studying the LSS of the Universe is the *weak gravitational lensing*. In weak lensing, the shapes and sizes of distant galaxies are coherently distorted by the inhomogeneous (dark) matter distribution. Galaxies whose light bundles have crossed the same matter structures will have correlated orientations of their shapes. Therefore, by studying the statistical correlation of the shapes and sizes of background galaxies, one is able to probe the properties of the LSS. Performing these measurements in tomographic redshift bins allows probing the growth rate of cosmic structures. This effect is referred to as *cosmic shear* or *weak cosmological lensing* and is one of the most promising probes of the LSS, its growth with cosmological time, as well as the nature of dark energy. Weak cosmological lensing has provided some of the most stringent constraints on cosmological parameters (e.g., Hildebrandt et al., 2020; Abbott et al., 2022). Additionally, it is at the heart of the scientific objectives of some of the most advanced observatories on Earth and in space, such as *Euclid* (Laureijs et al., 2011), Vera C. Rubin observatory, (Ivezić et al., 2019), DES (Dark Energy Survey Collaboration et al., 2016) and Roman (Spergel et al., 2015).

Weak lensing has also provided a direct probe of the properties of the dark matter structures around galaxies, such as their mass profiles and biasing. This opened a new window to study the inextricable connection between galaxies and dark matter. A very successful probe in this respect is *galaxy-galaxy lensing*. In principle, every matter structure, i.e., galaxy in the foreground, distorts the shape of a background galaxy in a direction tangential to the line connecting the two, albeit by a tiny and unmeasurable amount. However, a statistical measurement using numerous foreground-background galaxy pairs can stack the lensing signal. In this way, one is able to measure the preferred tangential orientation of the ellipticity of background galaxies around a population of foreground galaxies. The amplitude of this tangential alignment is a measure of the mass properties of the foreground galaxy population and also depends on the redshift distributions of the foreground and background galaxies. The galaxy-galaxy lensing signal on small scale probes the mass profiles of galaxies (including the dark matter (DM)) and the mass properties of the environment (e.g., the group or cluster if the galaxies happen to be embedded in) on intermediate scales. On large scales, $\gtrsim 1$ Mpc, background galaxies are too far apart to have their shapes distorted by individual foreground galaxies. Instead, it is the matter distribution of the LSS in which the foreground galaxies are embedded that cause the lensing. This correlation between the galaxy positions, matter distribution and the shape distortion of the background galaxies allows us to measure the galaxy biasing with respect to the overall matter distribution (see e.g., Pandey et al., 2021, for recent results).

Lensing studies require extremely precise and unbiased shape measurements because of its very low amplitude and statistical nature. What we observe are galaxies who have had their intrinsic shape sheared by the intervening matter, blurred by the atmosphere and/or the optics of the telescope, and then pixelated on the detector. Therefore, measuring accurately the shapes of galaxies poses one of the major challenges in lensing studies. Several methods have been developed and extensively studied that can be broadly separated in two classes. *Parametric*, or model-fitting methods measure morphological properties by fitting a galaxy light profile model convolved with the instrument point spread function (PSF). *Non-parametric* methods are based on the moments of the light distribution that can provide a measurement of the ellipticity. In recent years, new approaches are developed based on machine learning (ML) and Bayesian inference. However, all of these methods suffer from biases in the shape measurement that can be separated into a *multiplicative* and *additive bias*. Extensive effort has been put by the lensing community to study and quantify these biases, such as the GREAT3 challenge (Mandelbaum et al., 2014) based on blind simulations

These difficulties in measuring galaxy shapes limit the employment of cosmic shear or galaxy-galaxy lensing to the sample of background galaxies that have measurable shapes. This greatly limits the redshift range that can be studied. However, there exist other weak lensing effects that don't rely on shape measurements. *Magnification bias* is a Weak Lensing (WL) effect in which the number density of background galaxies is modified by the foreground lenses. This is a combination of two effects: 1) decrease in number density due to the geometrical stretching and 2) increase in number density due to flux amplification – as the flux of background sources is increased the more numerous, fainter ones, are pushed within the flux limit of the survey. Magnification bias therefore offers an additional lensing observable that does not rely on shape measurements and is free of some of the most plaguing systematic errors affecting the latter. This means it can be employed to higher redshifts to probe matter properties and/or be included in joint analyses to provide additional constraints. However, its typical low S/N ratio makes it difficult to compete with more traditional methods (e.g., galaxy-galaxy lensing) in the regime where shape measurements are available. The difficulties that magnification measurements typically encounter are related to the difficulty of achieving a perfect uniform flux selection in a wide survey (due to variations in the background, PSF, effects from bright stars etc). In this thesis, magnification bias is one of the central techniques that was developed and extensively studied in order to be employed as a probe of dark matter properties in state-of-the-art surveys, and in particular, in the upcoming *Euclid* space mission.

In this chapter, we will review the most relevant theoretical concepts that allow us to study the structure of the Universe at large scales through the gravitational lens.

3.2.1 *Light propagation in an inhomogeneous universe*

Within the framework of GR, light in the universe travels through null geodesics, which gives the minimal travel time in a space-time curved by mass. This means that it has a vanishing line element ds, so setting d$s = 0$ in metric **??** we get the travel time of photons as an integral over the light path in proper coordinates:

$$t = \frac{1}{c}\int\left(1 - \frac{2\Phi}{c^2}\right)\mathrm{d}r. \tag{3.17}$$

A quantity of interest is the *deflection angle* $\hat{\boldsymbol{\alpha}}$ that is defined by the difference between the vectors of the emitted and received light ray. The deflection angle can be obtained by using the variational principle along the line of sight, which results in:

$$\hat{\boldsymbol{\alpha}} = -\frac{2}{c^2}\int\nabla_\perp\Phi\,\mathrm{d}r. \tag{3.18}$$

Performing the integration along the line of sight it is the gradient of the potential perpendicular to the LoS that matters. This is a very general equation that calculates the deflection from any mass distribution defined by the potential. Observationally, the deflection for galaxies is not measurable, since there is no knowledge of its exact location in the absence of lensing.

Propagation of light bundles

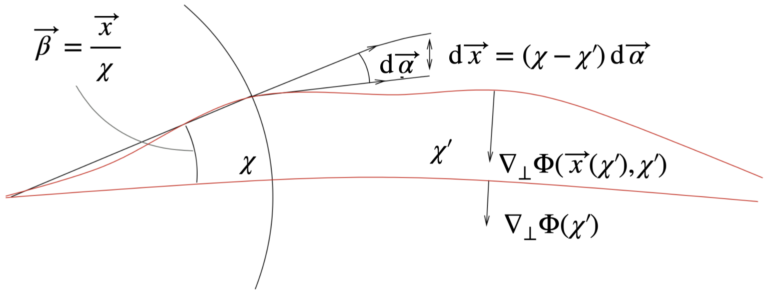


Figure 3.2: Illustration of the deflection of light bundles. Adapted from Schneider, 2005

Since we are interested in the distortion of the images of background galaxies, we will consider the deflection of light bundles. To derive the equations that describe how light bundles propagate in a clumpy universe, we will consider two neighboring geodesics coming from a distant galaxy with small intrinsic (non lensed) angular size $\vec{\beta}$. Differential deflection of the light bundles from the intervening matter will project the source in an image as seen by the observer, under an angle $\vec{\theta}$. The two are related by the deflection angle and give the *lens equation*

$$\vec{\beta} = \vec{\theta} - \vec{\alpha}. \tag{3.19}$$

To derive the relation for the deflection angle, we can study the change in the separation vector. The separation vector between the two close light rays is simply the transverse comoving separation[3] $\vec{x}_0(\chi) = \chi\,\vec{\theta}$. The change, defined by the deflection $\mathrm{d}\vec{\alpha}$, can be written as

$$\mathrm{d}\vec{x} = (\chi - \chi')\,\mathrm{d}\vec{\alpha}. \tag{3.20}$$

Here, χ is the comoving coordinate and χ' is the comoving distance from the observer to the point in space in which the potential Φ is considered. The differential deflection at a lens position χ' and transverse separation $\vec{x}$ can be rewritten in the comoving frame

$$\mathrm{d}\vec{\alpha} = -\frac{2}{c}\,\nabla_\perp \Phi(\vec{x}(\chi'),\chi')\,\mathrm{d}\chi'. \tag{3.21}$$

To obtain the total separation we can integrate along χ'. For simplification, it is useful to consider only the deflection of one light ray, the other can be kept fiducial (whose potential we can denote with $\Phi^{[0]}$) and can be subtracted; we thus get

$$\vec{x}(\chi) = \chi\,\vec{\theta} - \frac{2}{c^2}\int_0^{\chi} \mathrm{d}\chi'(\chi-\chi')\left[\nabla_\perp \Phi(\vec{x}(\chi'),\chi') - \nabla_\perp \Phi^{[0]}(\chi')\right]. \tag{3.22}$$

The fact that the potential is a function of χ and the integration is carried along the LoS, and the figuring potential also as a function of the LoS, means that this relation is generalized to any matter distribution. Dividing it by the comoving distance will give us the angle under which the source would be seen without lensing $\vec{\beta} = \vec{x}(\chi)/\chi$. Furthermore, the relation can be simplified by using the *Born approximation* where for small deflections the integration path is approximated by a straight line; this allows to set $\vec{x} \equiv \vec{x}_0 = \chi\,\vec{\theta}$. We will then arrive at the lens equation Eqn. 3.19 is by defining $\vec{\alpha}$ as

$$\vec{\alpha}(\vec{\theta}) = \frac{2}{c^2}\int_0^{\chi} \mathrm{d}\chi' \frac{\chi-\chi'}{\chi}\left[\nabla_\perp \Phi(\chi'\,\vec{\theta},\chi') - \nabla_\perp \Phi^{[0]}(\chi')\right]. \tag{3.23}$$

Note here that $\vec{\alpha}$ is a function of $\vec{\theta}$, which captures the differential deflection.

It is convenient to define a *lensing potential*, that contains all the terms that describe the mass configuration of the lens. To do this, we can also rewrite the derivative in terms of $\vec{\theta}$ so that $\nabla_\perp$ becomes $\nabla_{\vec{\theta}}$; this change will introduce another $1/\chi'$. This will allow us to rewrite the deflection angle as the gradient of the lensing potential,

$$\vec{\alpha}(\vec{\theta}) = \nabla_{\vec{\theta}}\psi(\vec{\theta},\chi), \tag{3.24}$$

where the lensing potential ψ is defined as:

$$\psi(\vec{\theta},\chi) = \frac{2}{c^2}\int_0^{\chi} \mathrm{d}\chi' \frac{\chi-\chi'}{\chi\chi'}\Phi(\chi'\,\vec{\theta},\chi'). \tag{3.25}$$

Finally, the lens equation in terms of the lensing potential writes as

$$\vec{\beta} = \vec{\theta} - \nabla_{\vec{\theta}}\psi(\vec{\theta},\chi). \tag{3.26}$$

---

[3] This relation is valid in a homogeneous and flat Universe.

The lens equation (Eqn. 3.19) represents a mapping $\vec{\theta} \mapsto \vec{\beta}$ between the lensed (image) and unlensed (source) planes. A linear mapping from the image coordinates to the source coordinates can be described by the Jacobian matrix $\mathcal{A}$

$$\mathcal{A}(\vec{\theta}) = \frac{\partial \vec{\beta}}{\partial \vec{\theta}} = \frac{\partial}{\partial \vec{\theta}} \left(\vec{\theta} - \vec{\alpha}(\vec{\theta}) \right) = \frac{\partial}{\partial \vec{\theta}} \left(\vec{\theta} - \nabla \psi(\vec{\theta}) \right) \tag{3.27}$$

The partial derivation with respect to $\vec{\theta}$ and the gradient can be written in terms of its components θ_i

$$\mathcal{A}(\vec{\theta}) = \frac{\partial}{\partial \theta_i} \left(\theta_j - \frac{\partial \psi}{\partial \theta_j} \right) = \left(\delta_{ij} - \frac{\partial^2 \psi}{\partial \theta_i \partial \theta_j} \right), \tag{3.28}$$

which can finally be written as a matrix

$$\mathcal{A}(\vec{\theta}) = \begin{pmatrix} 1 - \kappa - \gamma_1 & -\gamma_2 \\ -\gamma_2 & 1 - \kappa + \gamma_1 \end{pmatrix}. \tag{3.29}$$

The amplification matrix contains two new terms: the *convergence* κ and the *shear* γ, that has two components $\gamma = \gamma_1 + \mathrm{i}\gamma_2 = |\gamma| e^{2\mathrm{i}\varphi}$. All these terms are defined in terms of the ψ derivatives, as

$$\kappa = \frac{1}{2} \left(\frac{\partial^2 \psi}{\partial \theta_1^2} + \frac{\partial^2 \psi}{\partial \theta_2^2} \right) = \frac{1}{2} \nabla^2 \psi, \quad \gamma_1 = \frac{1}{2} \left(\frac{\partial^2 \psi}{\partial \theta_1^2} - \frac{\partial^2 \psi}{\partial \theta_2^2} \right), \quad \gamma_2 = \frac{\partial^2 \psi}{\partial \theta_1 \partial \theta_2}. \tag{3.30}$$

This transformation of the source image from its source plane to the lens plane is described by the inverse Jacobian matrix $\mathcal{A}^{-1}$. This transformation has two effects on the image: 1) isotropic magnification quantified by the convergence κ and 2) anisotropic stretching due to tidal gravitational field quantified by the shear γ. An illustration of these effects is given in Fig. 3.3. The convergence and shear encompass information about the projected mass distribution (since they are related to the lensing potential). They quantify how lensed images are distorted from the foreground mass distribution and are the main observables in weak lensing studies.

Projected overdensity

In the linearized mapping that was introduced above, the convergence κ was defined in terms of the lensing potential ψ via the two-dimensional Poisson equation (Eqn. 3.30). As such, the convergence can be interpreted as a projected surface mass density. Consequently, lensing by three-dimensional matter distribution can be considered as lensing from an equivalent lens located in a two-dimensional plane. To write explicitly these relations, we can start by relating κ to the three-dimensional density contrast δ. Since $\kappa = \nabla_\theta^2 \psi / 2$, we can write explicitly

$$\kappa(\vec{\theta}, \chi) = \frac{1}{c^2} \int_0^\chi \mathrm{d}\chi' \frac{\chi - \chi'}{\chi \chi'} \nabla_\theta^2 \Phi(\chi' \vec{\theta}, \chi'). \tag{3.31}$$

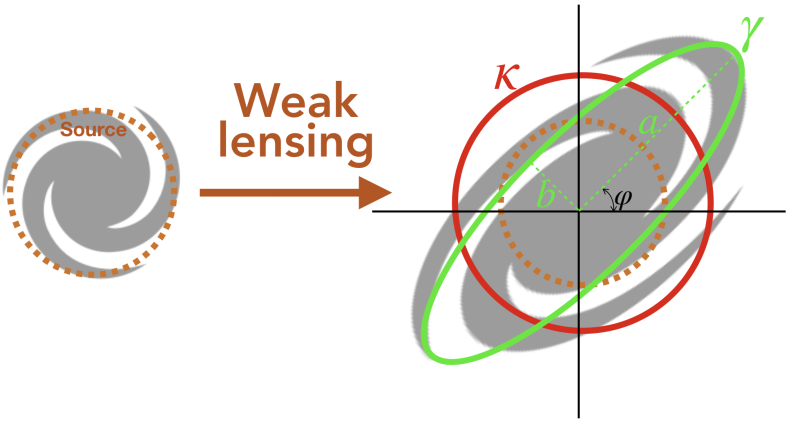


Figure 3.3: Scheme of the effect of convergence (κ) and shear (γ) on the image of a source. a and b are the major and semi-major axis of the ellipse of the stretched source image and φ the angle defining the rotation in the transformation.

Since we are interested in the density contrast δ, this can be made apparent using the three-dimensional Poisson equation of the gravitational potential

$$\nabla^2\Phi = 4\pi\, G\, a^2\, \delta\bar{\rho} = \frac{3H_0^2\,\Omega_\mathrm{m}\,\delta}{2a}, \tag{3.32}$$

where $\bar{\rho} = \bar{\rho}_0\, a^{-3}$ and $\Omega_\mathrm{m} = 8\pi\, G\, \bar{\rho}_0 / 3H_0^2$ were used. The Laplacian can again be rewritten in terms of θ which introduces a χ'^2 term, and after putting everything together the convergence can be written as:

$$\kappa(\vec{\theta},\chi) = \frac{3H_0^2\,\Omega_\mathrm{m}}{2c^2}\int_0^{\chi}\frac{\mathrm{d}\chi'}{a(\chi')}\,\frac{\chi'(\chi-\chi')}{\chi}\,\delta(\chi'\vec{\theta},\chi'). \tag{3.33}$$

This relation integrates the density contrast over the comoving coordinates and as such it expresses the projected density along the LoS and is proportional to the matter density Ω_m. Additionally, it is weighted by the distances that describe the lens-source configuration.

Eqn. 3.33 expresses the projected density from one single source plane located at χ. In principle, sources also follow some distribution along the LoS given by $p(\chi)\mathrm{d}\chi = p(z)\mathrm{d}z$. To obtain the convergence for a realistic distribution of source galaxies, we integrate over it

$$\kappa(\vec{\theta}) = \int_0^{\chi_\mathrm{lim}} \mathrm{d}\chi\, p(\chi)\,\kappa(\vec{\theta},\chi), \tag{3.34}$$

where the integration is done until the limiting distance χ_lim. Inserting the full expression for $\kappa(\vec{\theta},\chi)$ we can finally write the *mean convergence* for a population of source galaxies

$$\kappa(\vec{\theta}) = \int_0^{\chi_\mathrm{lim}} \mathrm{d}\chi G(\chi)\,\delta(\chi\vec{\theta},\chi), \tag{3.35}$$

where we define $G(\chi)$ as the *lensing efficiency*

$$G(\chi) = \frac{3H_0^2\,\Omega_\mathrm{m}}{2c^2}\,\frac{\chi}{a(\chi)}\int_{\chi}^{\chi_\mathrm{lim}} \mathrm{d}\chi'\,\frac{(\chi'-\chi)}{\chi'}\,p(\chi'). \tag{3.36}$$

The lensing efficiency contains the terms that describe the distances of the lens-source configuration and the source distribution. The lensing efficiency has a maximum at a distance that is approximately half-way to the mean of the source distribution. Convergence is an especially useful quantity since it is a measure of the total matter density projected along the LoS.

3.2.2 *Effects of lensing on images of galaxies*

To illustrate the effects of lensing on the observed image, we can consider a background source with a surface brightness I_ν which subtends on the sky an unlensed solid angle of $d\omega_0$; it will have a monochromatic flux of $f_{\nu,0} = I_\nu d\omega_0$. The differential deflection due to lensing will change the solid angle of the image to $d\omega$. In gravitational lensing no photons are emitted or absorbed, and therefore the surface brightness of the source is conserved[4] $I_\nu = const$; the flux of the lensed image will be $f_\nu = I_\nu d\omega$. Thus, the light deflection will induce a change in the flux of the observed lensed image by a factor

$$|\mu| = \frac{f_\nu}{f_{\nu,0}} = \frac{d\omega}{d\omega_0}. \tag{3.37}$$

$|\mu|$ is called *magnification*, and it is the ratio between the observed flux of the image after lensing and the intrinsic unlensed flux of the source (or consequently the ratio between the lensed and unlensed solid angles subtended by the source). Since the magnitude is related to the flux as $m = -2.5\log(f) + C$, the change of magnitude due to magnification is expressed as

$$m = m_0 - 2.5\log\mu. \tag{3.38}$$

The magnification μ can also be written in terms of the determinant of the Jacobian matrix $\mathcal{A}$

$$\mu = \frac{1}{\det\mathcal{A}} = \frac{1}{(1-\kappa)^2 - |\gamma|^2}. \tag{3.39}$$

The second effect of lensing on images of galaxies is shearing, governed by γ in Eqn. 3.30. The complex shear is expressed as $\gamma = |\gamma| e^{2i\varphi}$, which means that its amplitude describes the degree of distortion, whereas its phase φ gives the direction of distortion (see Fig. 3.3 for an illustration). The effect of shear on the intrinsic ellipticity of the galaxy ϵ_0, in the weak lensing regime, approximates to

$$\epsilon \approx \epsilon_0 + \gamma, \tag{3.40}$$

where ϵ would be the observed ellipticity of the galaxy. Whereas it is impossible to know the intrinsic ellipticity of background galaxies, they have no preferred orientation, there-

---

[4] Also a consequence of Liouville's theorem

fore the expectation value of an ensemble of sources is zero $\langle \epsilon_0 \rangle = 0$. This means that observed mean ellipticity is an unbiased estimator of the shear

$$\langle \epsilon \rangle = \gamma. \tag{3.41}$$

This fact allows the measurement of the shear locally, on a patch of the sky, from a sample of background galaxies whose ellipticity can be measured. This, obviously, requires very accurate and unbiased ellipticity measurements, where the observational effects have been carefully taken care of.

Magnification and shear provide lensing observables that can be used to probe the properties of the foreground matter distribution. For example, one of the most widely used methods to reconstruct mass maps of foreground matter structures (like galaxy clusters or even the LSS) uses measured ellipticities of galaxies (Kaiser and Squires, 1993). An alternative approach, uses magnification for the same purpose (see e.g., Dye et al., 2002).

3.2.3 *Galaxy-galaxy lensing*

Any matter distribution, even that of individual galaxies in the foreground, distorts the images of background galaxies. It is possible to measure the mean tangential shear $\langle \gamma_{\rm t} \rangle$ of background galaxies with respect to the lens. This is referred to as galaxy-galaxy lensing (GGL).

The mean tangential shear at a given radial distance θ from a lens (or, deflector) can be written in terms of the mean total projected density within this radius, and the azimuthally averaged surface density at the radius θ as:

$$\langle \gamma_{\rm t} \rangle(\theta) = \bar{\kappa}(\leq \theta) - \langle \kappa \rangle(\theta). \tag{3.42}$$

Here, $\bar{\kappa}(\leq \theta)$ is the mean convergence within the circular aperture θ, and $\langle \kappa \rangle(\theta)$ is the convergence azimuthally averaged around at θ. The convergence, which as we discussed earlier, measures the projected mass density. Since we are interested in lensing from individual lenses (galaxies), Eqn. 3.33 can be written for a single lens located at an angular diameter distance $D_{\rm L}$

$$\kappa(\theta) = \frac{4\pi G}{c^2} \frac{D_{\rm L} D_{\rm LS}}{D_{\rm S}} \int_{\Delta D_{\rm L}} {\rm d}D \Delta\rho(D\theta, D), \tag{3.43}$$

where the integration is carried over the extent of the mass distribution of the lens $\Delta D_{\rm L}$, with a density contrast $\rho(D\vec{\theta}, D)$, where D is the proper longitudinal coordinate. $D_{\rm S}$ is the distance to the source and $D_{\rm LS}$ the distance between the lens and the source. A *critical surface mass density* can be defined

$$\Sigma_{\rm crit} = \frac{c^2}{4\pi G} \frac{D_{\rm S}}{D_{\rm LS} D_{\rm L}}, \tag{3.44}$$

where $\Sigma_{\rm crit}$ is a characteristic (critical) surface mass density that defines strong ($\Sigma > \Sigma_{\rm crit}$) and weak ($\Sigma < \Sigma_{\rm crit}$) lens. It is completely determined by the distances between the lens and source. In terms of the critical mass density and the surface mass density of the source (defined by the integral), the convergence can be written as:[5]

$$\kappa(\theta) = \frac{\Sigma(D_L\theta)}{\Sigma_{\rm crit}}, \tag{3.45}$$

With these definitions, Eqn. 3.42 can be rewritten as:

$$\begin{aligned} \langle\gamma_{\rm t}\rangle(\theta) \times \Sigma_{\rm crit} = \Delta\Sigma(\theta) = \\ = \bar{\Sigma}(\leq \theta) - \langle\Sigma\rangle(\theta) \end{aligned} \tag{3.46}$$

This relation provides a way to measure GGL observationally. This is usually done by computing the critical mass density and source tangential shear for every i-th lens and j-th source pairs. For the critical mass density, this involves computing the distances between the two. The estimator of the excess projected mass density is

$$\Delta\Sigma_{ij}(\theta_{ij}) = \langle\gamma_{ij,\rm t}\rangle(\theta) \times \Sigma_{ij,\rm crit} \tag{3.47}$$

The signal of a single lens-source pair is very weak, and is usually stacked over many lenses, usually all lenses within your surveys. Selecting different lens populations, e.g., based on stellar mass, type etc., allows measuring matter profiles for these different galaxy populations.

Galaxy-galaxy lensing is a probe of the matter distribution and the LSS because of the correlation between galaxies and matter, defined by the two-point cross correlation function $\xi_{\rm gm}$. This cross-correlation between galaxies and matter is related to the projected surface mass density (or convergence) with the two-point correlation function (Bartelmann and Schneider, 2001)

$$\begin{aligned} \langle\kappa\rangle(\theta) = \langle\kappa(\vec{\theta})\delta_{\rm g}(\vec{\theta}+\vec{\phi})\rangle_{\vec{\theta}} = \\ = \frac{\bar{\rho}}{\Sigma_{\rm crit}}\int {\rm d}D\,\xi_{\rm gm}\left(\sqrt{(D\theta)^2 + (D-D_{\rm L})^2}\right) \end{aligned} \tag{3.48}$$

Given the fact that the $\xi_{\rm gm}$ and $\xi_{\rm m}$ are related with the galaxy biasing parameter, and the latter can be modeled theoretically, galaxy-galaxy lensing can be used to study the biasing of galaxies with respect to the dark matter distribution. Additionally, these correlation functions can be modelled in such a way (e.g., using the halo model formalism) that they parametrize the properties of the dark matter halos in which the galaxies reside. Therefore, constraining them against GGL measurements provides insight about halo masses and their density profiles for galaxies as a function of their stellar mass, luminosity, type etc. (see e.g., Mandelbaum et al., 2006; van Uitert et al., 2011; Leauthaud et al., 2012; Coupon et al., 2015; Uitert et al., 2018).

---

[5] Notice that this definition shows the convergence as a dimensionless surface mass density

3.2.4 *Magnification bias*

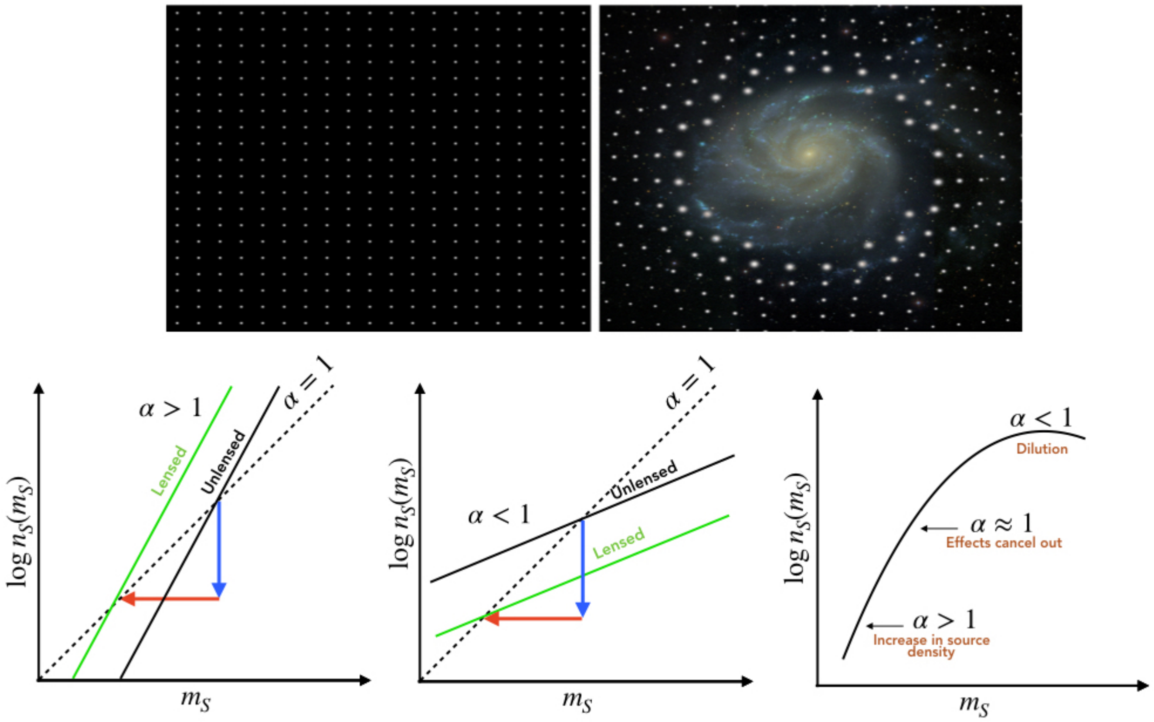


Figure 3.4: Magnification effects on the background source population. Top: an unlensed background source distribution (left) is modified by the presence of a lens in the foreground (right). Bottom: magnification effects with respect to the slope α. For $\alpha > 1$ ($\alpha < 1$) there is an increase (decrease) in source density due to magnification. The two effects cancel out at $\alpha \approx 1$ and the magnification can not be observed.

As presented previously, weak lensing by foreground matter structures affects the shape and size of the images of background sources. Due to the magnification of sizes and the conservation of surface brightness, the flux of the source is modified according to Eq. 3.37. Furthermore, since magnification is a geometrical effect, it also stretches the solid angle of the observed sky by the same relation Eq. 3.37. Let's consider an unlensed, intrinsic, number density of galaxies with a flux above $> f_0$ given by $n_0(> f_0)$. Magnification has two effects on the unlensed number density of the background galaxies in a flux limited survey: 1) dilution due to solid angle stretching $n = n_0/\mu$ and 2) increase in number density due to flux amplification $f = \mu f_0$ (or effectively decrease in the flux detection threshold) Therefore, the observed number density can be written[6]

$$n(> f) = \frac{1}{\mu} n_0 \left(> \frac{f}{\mu} \right) . \tag{3.49}$$

This concept is illustrated in Fig. 3.4 (top). By assuming a power law distribution for the unlensed source number counts, at least locally at $\sim f$,

$$n_0(> f) = A f^{-\alpha}, \tag{3.50}$$

---

[6] $n(> f)$ refers to the cumulative whereas $n(f)$ refers to the differential number count

and replacing it into the right side of Eq. 3.49 we obtain a relation between the observed and intrinsic number counts

$$\frac{n(>f)}{n_0(>f)} = \mu^{\alpha-1}. \tag{3.51}$$

This relation shows that whether there is an increase or decrease in the source number density depends on the slope α of the unlensed source number counts $n_0(>f)$. This is illustrated in Fig. 3.4 (bottom). For $\alpha > 1$ ($\alpha < 1$) there is an increase (decrease) in source density due to magnification. The two effects cancel out at $\alpha \approx 1$ and the magnification bias can not be observed. In terms of magnitudes, which are related to the flux as $m = -2.5\log(f) + const$, α can be expressed with respect to the magnitude number counts $n(<m)$ as[7]

$$\alpha = 2.5\frac{d\log n(<m)}{dm}. \tag{3.52}$$

Hence, α can be estimated from a magnitude number count (that is readily available in any survey) provided that it is representative of a wide enough field in which the overall effect of magnification will be vanishing $\mu \approx 1$. It should be noted that sample incompleteness (present in any flux limited survey) significantly changes the number counts at the faint end. Erroneous estimation of the slope will lead to significant errors when interpreting the magnification signal (Hildebrandt, 2016).

Weak lensing approximations

In typical observational scenarios, the effects of lensing are weak, such that $\kappa \ll 1$ and $\gamma \approx 0$. This allows to simplify the magnification relation Eq. 3.39 by making a Taylor expansion

$$\mu \approx (1-\kappa)^{-2} \approx 1 + 2\kappa. \tag{3.53}$$

By using this approximation in Eq. 3.51, and Taylor expanding around $\kappa = 0$ we can express the ratio between the observed and unlensed source counts (in terms of magnitude) as

$$\frac{n(<m)}{n_0(<m)} = 1 + 2(\alpha - 1)\kappa. \tag{3.54}$$

In the weak lensing regime the magnification is close to unity $\mu \approx 1$, so it can be written in terms of a magnification fluctuation $\delta\mu$ as $\mu = 1 + \delta\mu$, where $\delta\mu \ll 1$. Note that in these terms $2\kappa = \delta\mu$, meaning that the magnification fluctuation is twice the convergence.

3.2.4.1 *Correlation between foreground and background galaxies due to magnification*

The fact that magnification modifies the number density of background sources means that there is a correlation between the number densities of foreground lenses and the background sources. Such correlations are measurable with the angular 2PCF. This can be formally shown by considering the fluctuations in the number density of foreground and background galaxies.

---

[7] by inverting $m = -2.5\log(f) + const$, replacing for f in Eq. 3.50, and taking the derivative

The number density contrast of a galaxy population i, selected around a redshift bin z_i, projected on a position in the sky $\vec{\theta}$ will be a sum of two terms (Joachimi and Bridle, 2010)[8][9]: density fluctuations due to physical clustering of galaxies n_g and density fluctuations due to magnification n_μ

$$n^i(\vec{\theta}) = n^i_g(\vec{\theta}) + n^i_\mu(\vec{\theta}). \tag{3.55}$$

The angular number density fluctuations due to physical clustering can be written as a projection of the three-dimensional number density contrast of galaxies $\delta_g(\chi\,\vec{\theta},\chi)$

$$n^i_g(\vec{\theta}) = \int_0^{\chi_{\rm lim}} {\rm d}\chi\, p^i(\chi)\, \delta_g(\chi\vec{\theta},\chi), \tag{3.56}$$

where $p^i(\chi)$ is the redshift distribution of the i-th population. Since galaxies are a biased tracer of the underlying matter distribution, the galaxy density contrast can be written in the linear bias approximation in terms of the intrinsic matter density contrast as $\delta_g = b\delta_{\rm m}$, where b is the galaxy bias and $\delta_{\rm m}$ is the intrinsic matter density contrast.

In order to write the density fluctuations due to magnification, we can use Eqn. 3.51, written in terms of three-dimensions

$$\begin{aligned}\delta_\mu(\chi\vec{\theta},\chi) = \frac{n(<m,\,\chi\vec{\theta},\chi)}{n_0(<m,\,\chi\vec{\theta},\chi)} &= \mu^{\alpha(\chi)-1}(\vec{\theta},\chi) - 1\\ &\approx 2[\alpha(\chi)-1]\,\kappa(\vec{\theta},\chi).\end{aligned} \tag{3.57}$$

Note here that α is a function of the comoving distance χ, meaning that it concerns the number counts slope of the population at χ, (or consequently at z). To obtain the density contrast projected on the sky for a population of galaxies i, we carry out an integration over its distribution $p^i(\chi)$

$$\begin{aligned}n^i_\mu(\vec{\theta}) &= \int_0^{\chi_{\rm lim}} {\rm d}\chi\, p^i(\chi)\delta_\mu(\chi\vec{\theta},\chi)\\ &= \int_0^{\chi_{\rm lim}} {\rm d}\chi\, p^i(\chi)\, 2[\alpha(\chi)-1]\,\kappa(\vec{\theta},\chi)\\ &= 2(\alpha-1)\int_0^{\chi_{\rm lim}} {\rm d}\chi\, G(\chi)\,\delta(\chi\vec{\theta},\chi)\\ &= 2(\alpha-1)\,\kappa(\vec{\theta})\end{aligned} \tag{3.58}$$

Note that the number density contrast due to magnification in this form is written in terms of the convergence, which is a measure of the projected surface mass density, and is given by Eqn. 3.33. The last two lines simplify the relation by using the lensing efficiency (Eqn. 3.36) and the mean convergence (Eqn. 3.35). Note also the dependence on the three-dimensional *matter*[10] density contrast. In the last two lines, we adopt the assumption that the value of the slope (usually at the faint end of the number counts) does not vary appreciably over the integration range.

---

[8] This addition of the different contributing terms is valid in the case where magnification is small

[9] For brevity, we will omit the uncorrelated shot noise term

[10] since it is the total matter distribution that is responsible for the lensing

Any correlation between the positions of two galaxy populations (i, j) would be measurable with the angular 2PCF, which is defined in terms of the number overdensities as

$$w^{ij}(\theta) = \left\langle n^i(\vec{\theta}) n^j(\vec{\theta}') \right\rangle . \tag{3.59}$$

By plugging in Eqn. 3.55, we can express all the contributing terms to the correlations between i and j sample as:

$$\begin{aligned} w^{ij}(\theta) = \left\langle \left[n^i_g(\vec{\theta}) + n^i_\mu(\vec{\theta}) \right] \left[n^j_g(\vec{\theta}') + n^j_\mu(\vec{\theta}') \right] \right\rangle = \\ = \left\langle n^i_g(\vec{\theta}) n^j_g(\vec{\theta}') \right\rangle + \left\langle n^i_\mu(\vec{\theta}) n^j_\mu(\vec{\theta}') \right\rangle + \\ + \left\langle n^i_g(\vec{\theta}) n^j_\mu(\vec{\theta}') \right\rangle + \left\langle n^i_\mu(\vec{\theta}) n^j_g(\vec{\theta}') \right\rangle . \end{aligned} \tag{3.60}$$

The meaning of each of the 4 terms is the following. The first term describes the correlations due to physical clustering of matter; this term should be vanishing for a cleanly selected sample that do not overlap in redshift. The second term describes the correlations between the magnification density contrasts in two different redshift bins and can be neglected. The fourth erm is vanishing because a matter distribution at $z_j > z_i$ can not cause magnification correlations. $\left\langle n^i_g(\vec{\theta}) n^j_\mu(\vec{\theta}') \right\rangle$ is the magnification correlation between matter at $p^i(z)$ and magnification density contrast at $p^j(z)$ that we are interested in. Finally, since we are interested in the correlation between two populations separated in redshifts (lenses L, and sources S), we can model the angular 2PCF as

$$w^{\mathrm{LS}}(\theta) = w^{\mathrm{LS}}_{gg}(\theta) + w^{\mathrm{LS}}_{g\mu}(\theta) + w^{\mathrm{LS}}_{\mathrm{err}}(\theta). \tag{3.61}$$

w^{LS}_{gg} is a correlation term due to intrinsic clustering, which is equal to zero if the lens and source redshift bins are well separated. $w^{LS}_{g\mu}$ is the term describing the correlations due to the magnification bias and for cleanly separated samples is the only contribution. Due to uncertain redshifts, as is often the case in photometric surveys, a selection based on photo-z can include interlopers – objects selected in the lens bin that have their true redshift in the source bin and vice versa. Therefore, the cross-correlation of such samples will include clustering correlations, described by the term w^{LS}_{err}. This term can be a source of significant systematic error in magnification measurements, because any residual correlation coming from interlopers clustering will bias high the correlation amplitude.

3.2.4.2 *Magnification correlations as a probe of the matter distribution*

These magnification-induced correlations between foreground and background, as measured by the angular 2PCF, characterize the matter distribution of the foreground sample. $w^{\mathrm{LS}}_{g\mu}$ can be written in terms of the *galaxy-matter* power spectrum P_{gm} in the following way (Cooray and Sheth, 2002)

$$w^{\mathrm{LS}}_{g\mu}(\theta) = 2(\alpha - 1) \int_0^{\chi_{\mathrm{lim}}} \mathrm{d}\chi\, p^{\mathrm{L}}(\chi) G^{\mathrm{S}}(\chi) \int_0^\infty \frac{k\mathrm{d}k}{2\pi} P_{\mathrm{gm}}(k) J_0(k\chi\theta). \tag{3.62}$$

Here, G^{S} is the lensing efficiency where the integral is carried over the source redshift distribution and J_0 is the 0-th order Bessel function. Therefore, magnification measurements can be used to constrain parameters that describe the galaxy-matter power spectrum (Moessner and Jain, 1998).

This modeling of the magnification correlations is a convenient way to be incorporated within the HOD framework. Using the HOD formalism as described in § 3.1.1, the one and two-halo terms of the galaxy-matter power spectrum can be written as

$$\begin{aligned} P^{1\mathrm{h}}_{\mathrm{gm}}(k,z) &= \frac{1}{\bar{n}_g \,\bar{\rho}_m} \int \mathrm{d}M_h \frac{\mathrm{d}n}{\mathrm{d}M_h} M_h \left[\langle N_c\rangle + \langle N_s\rangle\, u_s^2(k)\right] u_m(k), \\ P^{2\mathrm{h}}_{\mathrm{gm}}(k,z) &= \frac{1}{\bar{n}_g \,\bar{\rho}_m} \int \mathrm{d}M_h \frac{\mathrm{d}n}{\mathrm{d}M_h} M_h \, b_h(M_h,z)\, u_m(k) \times \\ &\quad \times \int \mathrm{d}M_h \frac{\mathrm{d}n}{\mathrm{d}M_h} b_h(M_h,z) \left[\langle N_c\rangle + \langle N_s\rangle\, u_s(k)\right] P_{\mathrm{lin}}(k,z). \end{aligned} \tag{3.63}$$

Here, since we're correlating dark matter halos with galaxies, we have the quantity $u_m(k)$ which is the Fourier transform of the halo profile, that we take to be the NFW profile. This modeling allows for a consistent framework, based on the HOD, to simultaneously model and constrain the HOD parameters from different sets of measurements – in this case, galaxy clustering and weak lensing. Incorporating such a joint modeling is one of the subjects of this thesis and will be presented in Chapter 7 and Chapter 8.

Finally, because the correlation function, essentially measures the background overdensity stacked over all lenses fixed at the origin, it can be written as

$$\begin{aligned} w^{\mathrm{LS}}(\theta) &= \mu^{\alpha-1}(\theta) - 1 \\ &\approx 2(\alpha-1)\kappa(\theta). \end{aligned} \tag{3.64}$$

Therefore, the magnification correlations, being related to the projected mass density profile of the lens populations, can be used to probe lens mass profiles (see e.g., Van Waerbeke et al., 2010).

It is important to note how the magnification correlations scale with α, the magnitude number counts slope. Depending on the magnitude selection of the background sample, one can observe positive correlations for $\alpha > 1$, zero for $\alpha = 1$ and negative correlations for $\alpha < 1$. Fig. 3.5 shows an example of a measured cross-correlation by Hildebrandt, Waerbeke and Erben, 2009. The study uses the Canada-France-Hawaii-Telescope Legacy Survey (CFHTLS) Deep, that contains four independent fields of 1 square degree. The lenses are selected in the redshift bin of $0.1 < z_{\mathrm{L}} < 1.0$ in the i-band magnitude range of $17 < i < 24$, and the sources are Lyman-break galaxies at $z_{\mathrm{S}} \sim 3.2$. Magnification-induced cross-correlations are shown for background samples selected in four different magnitude bins hence with different values for the number count slope.

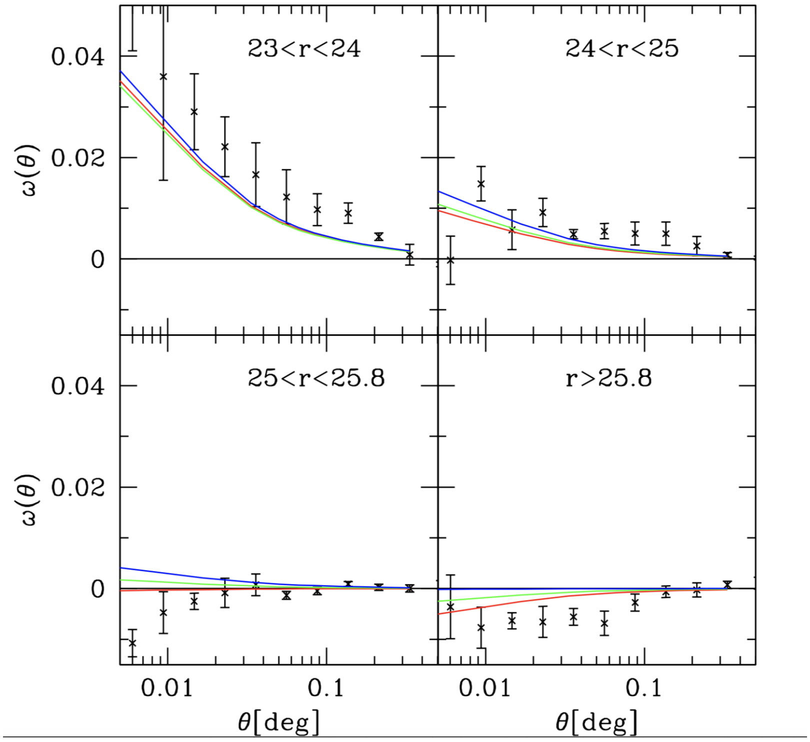


Figure 3.5: Magnification induced cross-correlation between foreground and background galaxies, measured by Hildebrandt, Waerbeke and Erben, 2009. The foreground lenses are selected in the redshift range of $0.1 < z < 1.0$. The four panels show the cross-correlations with four different background samples selected in different r-band magnitude bins. The background sample are Lyman-break galaxies at $z \sim 3.2$. The solid lines show the model predictions, where α is estimated from luminosity functions. *Figure adopted from Hildebrandt, Waerbeke and Erben, 2009*

3.2.5 *Comparison of the signal-to-noise of magnification and shear measurements*

Shear measurements rely on the fact that galaxies, have random orientations of their ellipticities with no preferential orientation and dispersion $\sigma_\epsilon \approx 0.4$. However, they are only possible with a relatively low number of resolved galaxies with accurate shape measurements N_γ. Magnification bias, on the other hand, can be measured with a significantly higher number of sources N_μ, but requires a measurement of the slope of the number counts and a clean separation between foreground and background sources. The signal-to-nose ratios for the shear and magnification method measurements can be written (Schneider, 2005)

$$\begin{aligned} (S/N)_\gamma &= \frac{|\gamma|}{\sigma_\epsilon}\sqrt{N_\gamma} \\ (S/N)_\mu &= 2\kappa|\alpha-1|\sqrt{N_\mu} \end{aligned} \tag{3.65}$$

The two methods can be compared by taking the ratio

$$\frac{(S/N)_\gamma}{(S/N)_\mu} = \frac{|\gamma|}{\kappa}\frac{1}{2\sigma_\epsilon|\alpha-1|}\sqrt{\frac{N_\gamma}{N_\mu}}. \tag{3.66}$$

This ratio tells us that in order to have magnification beat the S/N of shear, magnification measurements need to use more galaxies. This has, historically, been one of the main limitations of magnification and the reason why it was not so widely used in the literature. However, recent advances in wide and deep surveys with accurate photometry of even unresolved sources make magnification a promising probe, because they can increase N_μ, while limiting N_γ to well resolved galaxies.

4

GALAXY-HALO CONNECTION

4.1 INTRODUCTION

Within the current paradigm of galaxy formation, galaxies form and evolve within dark matter halos (described in more detail in Chapter 2). Therefore, the properties of the two are inextricably connected. The multivariate distribution of properties of halos and galaxies is referred to as the *galaxy-halo connection* (Wechsler and Tinker, 2018, for a review). Characterizing its many facets is key in understanding the physics of galaxy formation, probing the properties and distribution of dark matter in an unbiased way, as well as inferring cosmological parameters.

Galaxy formation is an inefficient process: the ratio of stellar mass to halo mass, M_*/M_h, is quite low (Shankar et al., 2006; Mandelbaum et al., 2006; Zheng, Coil and Zehavi, 2007; Conroy and Wechsler, 2009; Behroozi, Conroy and Wechsler, 2010). This quantity is a strong function of halo mass and rises to a peak at a characteristic peak halo mass, suggesting that at most only 20% of all the available baryons in the halo have turned into stars. At lower and higher halo masses, the M_*/M_h ratio decreases rapidly, which is seen as a signature of different feedback processes that suppress star-formation and act at different halo mass-scales: stellar feedback in low-mass halos and active galactic nuclei (AGN) in high-mass halos (e.g. Silk and Mamon, 2012, for a review).

Central and satellite galaxies contribute to the total stellar mass content of halos. In lower mass halos, the central galaxy makes up most of the stellar mass content, but its growth is regulated by stellar feedback mechanisms such as supernovae (SNe), stellar winds, radiation pressure and photoheating. All of these mechanisms are important to explain the observed amount of stellar mass in lower mass galaxies; otherwise, galaxy masses are overpredicted (Hopkins, Quataert and Murray, 2012). On the other hand, in cluster-scale halos the satellite galaxies dominate the stellar mass budget, mostly due to their high number (Leauthaud et al., 2012; Coupon et al., 2015). Within large halos, the stellar mass assembly in satellites is described by interplay between the hierarchical merger tree, where smaller halos accrete into larger ones and in-situ star formation facilitated by cold gas inflows and the various environmental quenching mechanisms that act to suppress further growth (Peng et al., 2010; Gabor and Davé, 2015). These include the so called 'hot-halo', where the infalling gas is heated by virial shock heating (Birnboim and Dekel, 2003), as well as mechanisms such as strangulation (Larson, Tinsley and Caldwell, 1980; Balogh, Navarro and Morris, 2000), ram-pressure stripping (Gunn and Gott, 1972) and harassment (Moore et al., 1996). In such massive halos, AGN feedback keeps the gas hot through the so-called 'radio' mode and is necessary to explain the break in the local galaxy luminosity/stellar mass function at the bright/massive end (Croton et al., 2006; Bower et al., 2006). All of these factors shape the total (central + satellite) SHMR, and its evolution with redshift can indicate the relative importance of these processes in shaping the star-formation efficiency as a function of the halo mass.

This Thesis aims to shed light on the relation between the mass of the halo and the stellar mass of the galaxy. Unveiling this relation across a vast redshift span is crucial in order to understand galaxy stellar mass assembly, their baryon conversion efficiency, and the physical mechanisms that regulate their growth. In this chapter, we will first give a brief overview of the galaxy-halo connection and the methods typically used to study it. In subsequent sections, we will present the stellar-to-halo mass relation, along with the theoretical formulation of the methods we use to study it. We will then discuss what can this relation teach us about the star formation efficiency in galaxies and how different physical processes can contribute in its regulation.

4.2 STELLAR-TO-HALO MASS RELATION

The relation between the stellar mass of a galaxy and the mass of its host dark matter halo (stellar to halo mass relation SHMR) gives an indication of how efficiently the baryons available in a dark matter halo have converted into stars. Assuming that the content of baryon matter in dark matter halos is given by $f_\mathrm{b} = \Omega_\mathrm{b}/\Omega_\mathrm{m} \approx 16\%$, the relation between the stellar and the halo mass can be written as

$$M_\star = \epsilon_\mathrm{SF}\, f_\mathrm{b}\, M_\mathrm{h}, \tag{4.1}$$

as well as between their number densities

$$\Phi_\star(M_\star, z) = \Phi_\mathrm{h}(M_\star\, f_b^{-1}\, \epsilon_\mathrm{SF}^{-1}, z) \tag{4.2}$$

where ϵ_SF is the efficiency of converting baryons to stars, i.e., the star-formation efficiency (SFE). $\Phi_\star(M_\star)$ and $\Phi_\mathrm{h}(M_\mathrm{h})$ are the stellar mass and halo mass function respectively. Eqn. 4.2 imposes an upper limit on the maximum stellar mass and stellar mass function that can be formed in a DM halo if the efficiency is 100%, i.e. all the available baryons are turned into stars (Behroozi and Silk, 2018). In this case, the SMF and HMF would have the same shapes, only offset by this f_b. However, observationally, this is shown not to be the case – the SMF and HMF differ by several orders of magnitude especially at the low and high mass regimes indicating the galaxy formation process is highly inefficient (Silk and Mamon, 2012; Behroozi, Wechsler and Conroy, 2013).

Studies on the SHMR show that there exists a specific mass $M_\mathrm{h}^\mathrm{peak}$ for the halo at which the M_*/M_h peaks. This peak occurs at $M_\mathrm{h}^\mathrm{peak} \approx 10^{12} M_\odot$, and at M_*/M_h of just a few percent, meaning that only $\sim 20-30\%$ of the baryons turned into stars (see Fig. 4.1). At this $M_\mathrm{h}^\mathrm{peak}$, the conversion of baryons to stars has been most efficient. At lower and higher halo masses $M_*/M_{\rm h}$ decreases sharply (more so at the low-mass end) – galaxy formation is even less efficient. This decrease in efficiency is explained by several feedback mechanisms that reheat the gas and thus suppress star formation. At low halo masses, the dominant feedback mechanisms are supernovae explosions and stellar winds that can reheat or eject the gas. Massive halos can host galaxies massive enough to form supermassive black holes in their nuclei that are highly active. These active galactic nuclei (AGN) heat the gas and suppress star formation. The way in which the strength of the different feedback mechanisms affect the SHMR is also illustrated in Fig. 2.5i. At the peak halo mass, both the stellar and AGN feedback are the least efficient. Fig. 4.1 shows a compilation of measurements of the SHMR at $z = 0$.

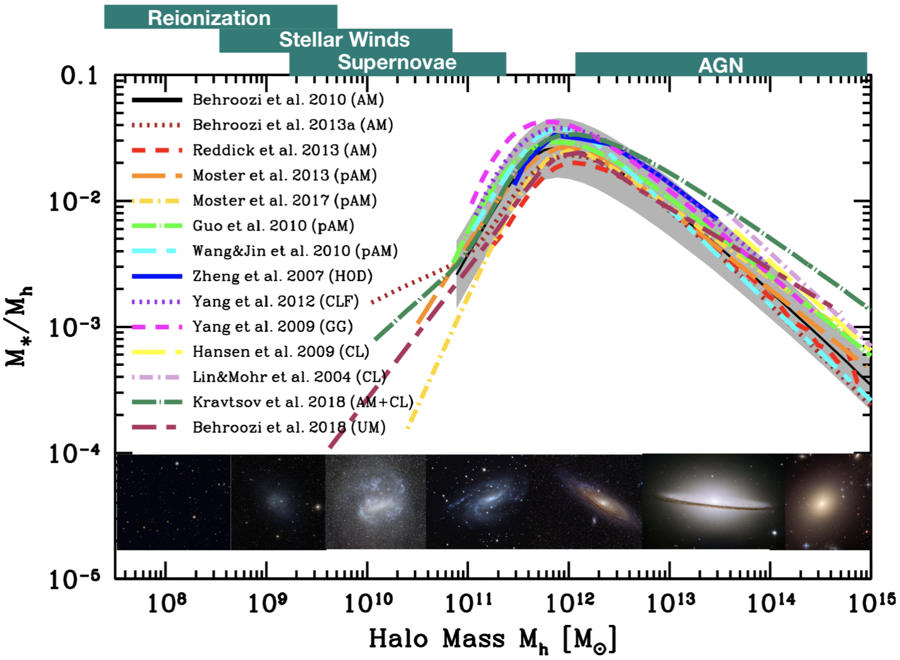


Figure 4.1: Compilation of measurements of the stellar-to-halo mass ratio of central galaxies as a function of the halo mass at z = 0; measurements are obtained from various methods. The bottom panel shows typical galaxies that are expected to live in a halo of a given mass. On top, the different feedback processes are indicated accordingly to the halo mass range on which they act. Figure adapted from Wechsler and Tinker, 2018; Behroozi et al., 2019

Functional form for the SHMR

The SHMR has been parametrized in a functional form by several authors in the literature. Some of the most widely used include power laws (Moster et al., 2010; Yang, Mo and van den Bosch, 2009) and a combination of a power law and a sub-exponential law (Behroozi, Conroy and Wechsler, 2010). These parametrizations capture the mismatch between the GSMF and HMF caused by feedback effects.

In this thesis, we study the SHMR in a large redshift range by constraining it using measurements of GSMF and clustering. For this purpose, we adopt the approach of Leauthaud et al., 2011, hereafter L11, who laid the theoretical framework to model galaxy clustering, weak lensing and GSMFs based on the HOD formalism with as a starting point a functional form of the SHMR. This function captures the different growth rates of galaxies as a function of the halo mass that is shaped by various feedback processes that operate at different mass scales. Behroozi, Conroy and Wechsler, 2010 presented a functional form of the SHMR that we adopt for our analysis

$$\log\left(f_{\mathrm{SHMR}}^{-1}(M_*)\right) = \log(M_h) = \\ \log(M_1) + \beta \log\left(\frac{M_*}{M_{*,0}}\right) + \frac{\left(\dfrac{M_*}{M_{*,0}}\right)^{\delta}}{1+\left(\dfrac{M_*}{M_{*,0}}\right)^{-\gamma}} - \frac{1}{2}. \tag{4.3}$$

This relation is parameterized by a characteristic halo mass, and stellar mass scales given by the parameters M_1 and $M_{*,0}$ respectively. M_1 controls the normalization of M_h as a function of M_* and $M_{*,0}$ controls the normalization along the M_*-axis. Together, two parameters govern the transition mass scale between the low-mass and high-mass regime of the SHMR. The low-mass regime ($M_* \lesssim 10^{10.5}\,M_\odot$) is described by a power-law regulated by the parameter β. The high-mass regime follows a sub-exponential law (increasing more rapidly than a power-law but less rapidly than an exponential) regulated by the parameter δ (Behroozi, Conroy and Wechsler, 2010). Intermediate mass galaxies in the transition regime between the power-law and the sub-exponential can have their growth affected by different feedback mechanisms. The width of the transition from the low-mass to the high-mass regime is allowed to vary by introducing the parameter γ. In total, this makes 5 parameters that describe the SHMR and that we seek to fit. The way the SHMR responds to the change of the values of each parameter is shown in Fig. 4.2 and Fig. 4.3.

The ratio between the stellar mass and the halo mass obtained with Eq. 4.3, (i.e. M_*/M_h) can be considered as the efficiency of the galaxy formation process that encapsulates all the processes that lead to the conversion of baryons to stars (from gas cooling and star formation to stellar and AGN feedback); this can be referred to as the baryon conversion efficiency or star formation efficiency (SFE). Since we can consider that baryonic matter content of halos is equal to the universal baryonic fraction $f_b = \Omega_\mathrm{b}/\Omega_\mathrm{m} \approx 0.16$, the M_*/M_h ratio will inform us of the fraction of the baryons available in the dark matter (DM) halo that have converted into stars. At a given redshift, the M_*/M_h ratio gives the baryon conversion efficiency integrated over the lifetime of the halo and therefore includes the combination of all the different physical processes that regulate star formation throughout the halo life (e.g., gas accretion, mergers, feedback). The shape of the M_*/M_h ratio is a strong function of halo mass, which indicates that various feedback mechanisms operate on different halo mass scales to regulate star-formation. This is the principal quantity of interest in this work that we extensively discuss in Chapter 7.

As L11, we assume that this SHMR concerns only central galaxies occupying the central regions of the dark matter halos. These halos can contain smaller sub-halos that orbit the potential well and can host satellite galaxies. Satellites can undergo different stellar growth than centrals, since various distinct processes affect satellites in the dark matter halos that can regulate their growth (such as stripping and harassment). This will reflect in the SHMR, therefore a different relation for satellites is a more accurate assumption. This is one of the advantages of this methodology to infer the SHMR, as compared to the commonly implemented abundance matching (AM) technique. In AM, one treats satellites as centrals in their own sub-halo which then assumes that centrals and satellites follow the same SHMR. The formalism that we employ is able to constrain the contribution of both centrals and satellites in the total stellar mass content in halos of given mass (Section 4.3.3).

Scatter between stellar mass and halo mass

Equation. 4.3 gives the mean logarithmic stellar mass as a function of halo mass $f_{\mathrm{SHMR}}(M_h) = \langle [M_*(M_h)] \rangle$. Due to the effects of the various galaxy formation processes, merger histories etc., there exists a scatter of galaxy stellar mass (as well as other galaxy properties) at a fixed halo mass. For example, studies based on numerical simulations have pointed out the effects of assembly bias: halos and galaxies can experience different assembly histories at fixed masses, which translates into dependence of clustering on secondary halo properties (see e.g. Wechsler and Tinker, 2018). Furthermore, recent studies based on hydrodynamical simulations have unveiled the dependence of galaxy properties on the Cosmic Web (CW) environment (for example on distance to CW filaments) at fixed halo masses (Martizzi et al. 2020; Song et al. 2021).

This stochastic nature of the SHMR can be modelled with a conditional function that describes the probability of observing a central galaxy with M_* at a given M_h, which can be chosen to be a log-normal distribution in the form

$$\begin{aligned}\Phi_c(M_*|M_h) =& \frac{1}{\ln 10 \sqrt{2\pi}\, \sigma_{\mathrm{Log}M_*}} \times \\ & \times \exp\left(- \frac{[(M_*) - (f_{\mathrm{SHMR}}(M_h))]^2}{2\, \sigma^2_{\mathrm{Log}M_*}} \right),\end{aligned} \tag{4.4}$$

described by the log-normal scatter $\sigma_{\mathrm{Log}M_*}$, and centered at the stellar mass calculated from the SHMR; the function is normalized such that the integral is equal to unity. $\sigma_{\mathrm{Log}M_*}$ then describes the scatter in stellar mass at fixed halo mass, and can include contributions due to intrinsic scatter (from the processes mentioned earlier) and due to observational uncertainties. Assuming a Gaussian error distribution, the total scatter would simply be a quadrature sum of the intrinsic scatter and the one due to measurement uncertainties. In the fitting procedure, we are sensitive to the quadrature sum of the two. However, their contribution can be distinguished provided an understanding of the measurement uncertainties $\sigma^{\mathrm{meas}}_{\mathrm{Log}M_*}$. In this work, we constrain $\sigma_{\mathrm{Log}M_*}$ without disentangling all the different contributors. An interesting avenue is investigating the various different contributions to the scatter: from measurement uncertainties to the effect of the anisotropic environment (the CW), which we leave for future work in the light of the upcoming deep and wide surveys.

Following previous work (e.g. Moster et al., 2010; Leauthaud et al., 2012; Tinker et al., 2013), in this work we consider $\sigma_{\mathrm{Log}M_*}$ to be independent of halo mass, and we leave it as a free parameter to be fitted in each z-bin. Hydrodynamical simulations, however, show that at $z = 0$, $\sigma_{\mathrm{Log}M_*}$ generally decreases with M_h going from ~ 0.32 at $M_h \sim 10^{11}$ to ~ 0.15 at $M_h \sim 10^{12}$ and staying constant to higher masses (Pillepich et al., 2018). We checked that varying $\sigma_{\mathrm{Log}M_*}$ over this range changes the clustering correlation function insignificantly (and within the measurements' error bars), and given the fact that this parameter is mostly constrained by the GSMF, taking $\sigma_{\mathrm{Log}M_*}$ independent of the halo mass is a safe assumption for our purposes.

Scatter in halo mass at fixed stellar mass

Apart from the scatter in stellar mass at fixed halo mass, there is also a scatter in halo mass at fixed stellar mass $\sigma_{\mathrm{Log}M_\mathrm{h}}$. In other words, a galaxy of a fixed mass can be a central in a group-sized halo, or a satellite galaxy in a massive cluster. The relation between $\sigma_{\mathrm{Log}M_*}$ and $\sigma_{\mathrm{Log}M_\mathrm{h}}$ can be written in terms of the derivative of the SHMR as

$$\sigma_{\mathrm{Log}M_\mathrm{h}} = \sigma_{\mathrm{Log}M_*} \frac{\mathrm{d}\log M_h}{\mathrm{d}\log M_*}. \tag{4.5}$$

In the simple case of a power law dependence between M_* and M_h, then there would be a linear relation between the two scatters. This is true for the low mass regime below the peak, which is indeed modeled as a power law. In the massive regime, the relationship between the two masses is sub-exponential. This means that $\sigma_{\mathrm{Log}M_\mathrm{h}}$ also increases with halo mass, at a rate governed by γ and δ parameters in Eqn. 4.3.

4.3 HOD APPROACH TO STUDY THE SHMR

To study the SHMR and its redshift evolution using measurements of GSMF, clustering and lensing, we adopt an HOD framework to model these observables. The HOD framework to model GSMF, and two-point correlation function was described in more detail in Chapter 3. Here we lay down the HOD based model we use in this thesis.

4.3.1 *Central occupation distribution*

In summary from Chapter 3, the HOD describes the statistical occupation of galaxies in dark matter halos. It assumes a probability distribution of the number of galaxies residing in halos conditioned on some criteria, usually on the mass $P(N|M_h)$. Typically, centrals are assumed to follow a Bernoulli distribution, while the number of satellites follows a Poisson distribution (see e.g., Zheng et al., 2005, and references therein). Under these assumptions, the HOD is described by the average number of galaxies with stellar masses higher than some threshold in halos of a given mass $\langle N(M_h| > M_*^{\rm th})\rangle$, This can be computed by integrating the conditional stellar mass function Eq. 4.4 from $M_*^{\rm th}$ to infinity that results in

$$\begin{aligned} &\left\langle N_{\rm cent}\left(M_h| > M_*^{\rm th}\right)\right\rangle = \\ &\frac{1}{2}\left[1 - {\rm erf}\left(\frac{\left[\log(M_*^{\rm th}) - \log(f_{\rm SHMR}(M_h))\right]}{\sqrt{2}\,\sigma_{{\rm Log}M_*}}\right)\right]. \end{aligned} \tag{4.6}$$

$\langle N_{\rm cent}\left(M_h| > M_*^{\rm th}\right)\rangle$ is a monotonic function increasing from 0 to 1. $f_{\rm SHMR}(M_h)$, whose inverse function is defined by Eq. 4.3 and gives the stellar mass at the halo mass, while $\sigma_{{\rm Log}M_*}$ is the scatter in stellar mass at fixed halo mass. We note that all the other parameters regulating the central HOD parametrize the functional form of the SHMR. An alternative approach employed by many studies (e.g., Zheng, Coil and Zehavi, 2007; Zehavi et al., 2011; Coupon et al., 2012; McCracken et al., 2015; Ishikawa et al., 2020) specifies the central HOD assuming the SHMR to be a simple power law, with a parameter quantifying the minimum halo mass to host a galaxy $M_{\rm min}$. The downside of this model is the difficulty in the interpretation of $M_{\rm min}$ as the halo mass at the stellar mass threshold, that is, $M_{\rm min} = f_{\rm SHMR}^{-1}(M_*^{\rm th})$, especially at high masses where the deviation from a power-law of the SHMR is clear (see L11 for more details).

4.3.2 *Satellite occupation distribution*

The occupation of halos by satellites can be modeled by a power-law at high halo masses with an exponential cut-off at low masses, given by:

$$\begin{aligned} &\left\langle N_{\rm sat}\left(M_h| > M_*^{\rm th}\right)\right\rangle = \\ &\left\langle N_{\rm cent}\left(M_h| > M_*^{\rm th}\right)\right\rangle \left(\frac{M_h}{M_{\rm sat}}\right)^{\alpha_{\rm sat}} \exp\left(\frac{-M_{\rm cut}}{M_h}\right), \end{aligned} \tag{4.7}$$

where $\alpha^{\rm sat}$ is the power-law slope, $M_{\rm sat}$ is the halo mass scale for the satellites defining the amplitude of the power law and $M_{\rm cut}$ is the cutoff scale. HOD studies have shown that the satellite mass scale is proportional to $f_{\rm SHMR}^{-1}$ at the threshold stellar mass (e.g. Zheng, Coil and Zehavi, 2007; Zehavi et al., 2011). This allows us to parametrize $M_{\rm sat}$ and $M_{\rm cut}$ as power laws by introducing four more parameters:

$$\begin{aligned} \frac{M_{\rm sat}}{10^{12}M_\odot} &= B_{\rm sat}\left(\frac{f_{\rm SHMR}^{-1}(M_*^{\rm th})}{10^{12}M_\odot}\right)^{\beta_{\rm sat}}, \\ \frac{M_{\rm cut}}{10^{12}M_\odot} &= B_{\rm cut}\left(\frac{f_{\rm SHMR}^{-1}(M_*^{\rm th})}{10^{12}M_\odot}\right)^{\beta_{\rm cut}}. \end{aligned} \tag{4.8}$$

The HOD is fully specified by the average occupation number of galaxies in halos, as given by Eq. 4.6 & 4.7. Finally, the total number of galaxies including centrals and satellites is simply $\langle N_{\rm tot}\rangle = \langle N_{\rm cent}\rangle + \langle N_{\rm sat}\rangle$. We can also compute the average number of galaxies in a mass bin of $M_*^{\rm th1} < M_* < M_*^{\rm th2}$ by simply taking the difference

$$\begin{aligned}&\left\langle N_{\rm cent/sat}\left(M_h|M_*^{\rm th1}, M_*^{\rm th2}\right)\right\rangle = \\ &\left\langle N_{\rm cent/sat}\left(M_h| > M_*^{\rm th1}\right)\right\rangle - \left\langle N_{\rm cent/sat}\left(M_h| > M_*^{\rm th2}\right)\right\rangle .\end{aligned} \tag{4.9}$$

The model has a total of 11 parameters. The SHMR for the centrals has five parameters (M_1, $M_{*,0}$, β, δ, γ) with one additional parameter that describes the scatter in stellar mass at a fixed halo mass of $\sigma_{{\rm Log}M_*}$. The occupation distribution for satellites is modeled with five parameters ($\alpha_{\rm sat}$, $B_{\rm sat}$, $\beta_{\rm sat}$, $B_{\rm cut}$, $\beta_{\rm cut}$).

4.3.3 *Total stellar content in halos*

From the model of the conditional mass function, it is possible to compute the total stellar mass contained in halos of a given mass by performing an integration over the stellar mass. Since we do not have a model of $\Phi_s(M_*|M_h)$, we can use the occupation distributions of centrals and satellites because they are also integrals of the conditional mass function. Therefore, the contribution of centrals and satellites to the total stellar mass content in halos can be computed as:

$$\begin{aligned}&M_*^{\rm tot}\left(M_h|M_*^{\rm th1}, M_*^{\rm th2}\right) = \\ &M_*^{\rm tot,cent}\left(M_h|M_*^{\rm th1}, M_*^{\rm th2}\right) + M_*^{\rm tot,sat}\left(M_h|M_*^{\rm th1}, M_*^{\rm th2}\right) = \\ &\int_{M_*^{\rm th1}}^{M_*^{\rm th2}} \langle N_{\rm cent}(M_h| > M_*)\rangle\ M_* - [\langle N_{\rm cent}(M_h| > M_*)\rangle\ M_*]_{M_*^{\rm th1}}^{M_*^{\rm th2}} + \\ &\int_{M_*^{\rm th1}}^{M_*^{\rm th2}} \langle N_{\rm sat}(M_h| > M_*)\rangle\ M_* - [\langle N_{\rm sat}(M_h| > M_*)\rangle\ M_*]_{M_*^{\rm th1}}^{M_*^{\rm th2}} .\end{aligned} \tag{4.10}$$

This equation computes the contribution of galaxies (centrals and satellites) in a stellar mass bin $M_*^{\rm th1} < M_* < M_*^{\rm th2}$ to the total stellar mass content in halos of M_h. The total SHMR then shows the overall efficiency of the galaxy formation process in halos, integrated over the halo's history, that is a combination of the in situ conversion of gas to stars and ex-situ from merging with satellites.

4.4 OVERVIEW OF THE METHODS TO STUDY THE GALAXY-HALO CONNECTION

There is an abundance of methods in the literature to study the galaxy-halo relation. They can range from purely physical models to empirical ones (Wechsler and Tinker, 2018). In physical models one starts from first principles, i.e., the known physical models, and infers the relation between galaxies and halos, usually in a forward modelling approach such as hydrodynamical simulations. Empirical models typically use data and statistical functions to relate galaxy properties to halo properties computed from theory or a simulation. These models can be constrained with numerous observations such as galaxy abundance, clustering, group and cluster catalogs, weak lensing etc. Here we give a brief overview of the methods typically employed in the literature.

Some of the most predictive models that give direct insight into the physical processes that shape the galaxy-halo connection come from hydrodynamical simulations (e.g. Vogelsberger et al., 2014; Dubois et al., 2014; Schaye et al., 2015; Pillepich et al., 2018). However, the known physics implemented in the simulations imposes a strong prior on the galaxy evolution and cannot provide information about physical processes previously unexpected to contribute to the galaxy-halo connection (Behroozi and Silk, 2018). Additionally, an important caveat of simulations is that they cannot simulate the full physics at all scales, so they rely on various parametrizations below the resolution scale – 'subgrid physics' (Chapter 5). The subgrid physics varies from one simulation to another, and so do their conclusions on the galaxy-halo connection. Even though hydrodynamical simulations are the most predictive, they are computationally expensive, which makes it difficult for them to be constrained against observations using techniques such as Markov Chains Monte Carlo (MCMC).

Empirical models offer the most flexibility (Wechsler and Tinker, 2018). Many methods have been described in the literature to compute the SHMR from observational data sets or numerical simulations. Starting from the basic assumption that most massive galaxies inhabit the most massive dark matter halos is the fertile approach of abundance matching (AM), that has provided numerous constraints on the SHMR. In the simplest application of AM, the abundance (i.e. number density in a comoving volume) of galaxies of a given mass is matched to the abundance of dark matter (sub)halos which then gives the halo mass – in practice one matches the galaxy stellar mass function (GSMF) and halo mass function (HMF) (Marinoni and Hudson, 2002; Kravtsov et al., 2004; Vale and Ostriker, 2004; Tasitsiomi et al., 2004). This match of the number densities as a function of mass is also presented in Eqn. 4.2. It is also possible to match other properties, such as the galaxy luminosity. For halos, one can also use the maximum circular velocity of halos (Kravtsov et al., 2004), which is less dependent to effects like tidal stripping that strongly affects the mass; for this purpose one can also use the halo mass at the time when the subhalo accretes onto the main halo.

Another method based on a statistical description of the way galaxies populate halos is the *halo occupation distribution* (HOD) model. The HOD can model one- and two-point statistical observables such as the GSMF and galaxy clustering as measured by the 2PCF) (e.g. Peacock and Smith, 2000; Seljak, 2000; Scoccimarro et al., 2001; Berlind and Weinberg, 2002; Zehavi et al., 2002). The advantage of the HOD-based models is that they can constrain satellite-related quantities (e.g. Zheng et al., 2005; Zheng, Coil and Zehavi, 2007; Zehavi et al., 2011; Coupon et al., 2012; McCracken et al., 2015) (also Chapter 3). Furthermore, since they describe how galaxy occupy halos, HOD models can be used in semi-analytic models to populate halos in N-body simulations and build mock galaxy catalogs.

AM and HOD are phenomenological approaches that are based on statistical measures (GSMF and 2PCF), they indirectly probe the dark matter halos and are agnostic about the physical processes that shape their relation with the galaxies. An additional drawback of these phenomenological models is that they rely on accurate knowledge of the halo mass function, that has to be calibrated using numerical simulations, and are sensitive to various definitions of the halo profile, radius, mass, concentration and bias. However, this modelling is not restrained by strong priors from the physics of galaxy evolution – it is almost entirely constrained by observations and as such can reveal signatures of new physical processes that shape galaxy properties.

Empirical models of galaxy formation histories bridge the gap between AM and HOD models and semi-analytic models of dark matter simulations. These models use N-body simulations to extract halo merger trees, populate and trace galaxies within their dark matter halo over time. These models are fitted to various observables at different epochs such as the galaxy stellar mass and luminosity functions, specific and cosmic star formation rates, correlation functions due to clustering and lensing amongst others. Once constrained against data, these models can predict the SHMR as well as galaxies' star-formation histories as a function of halo mass (e.g. Conroy and Wechsler, 2009; Moster, Naab and White, 2018; Behroozi et al., 2019).

There exists a class of methods that directly probe the dark matter halos, such as gravitational lensing (e.g., Leauthaud et al., 2011; van Uitert et al., 2011) and kinematics of satellite galaxies (van den Bosch et al., 2004). A popular weak lensing probe is galaxy-galaxy lensing, in which the gravitational potential (caused by both dark and baryonic matter) around foreground galaxies (lenses) causes coherent distortions of the shapes of background galaxies. The observable is statistical cross-correlation between background galaxy shapes and foreground matter (including dark) as measured by a two-point correlation function. Lensing observables, such as galaxy-galaxy lensing, can also be modelled within the HOD formalism and as such can be incorporated in a joint analysis along with abundance and clustering to offer complementary constrains on the HOD and SHMR (Leauthaud et al., 2011; van Uitert et al., 2011; Leauthaud et al., 2012; Tinker et al., 2013; Coupon et al., 2015). Galaxy-galaxy lensing relies on accurate measurements of galaxy ellipticities, which becomes increasingly difficult at $z > 1$ and poses problems in its implementation to probe a redshift evolution over a vast interval.

4.5 LIMITATIONS AND OPEN QUESTIONS

The advent of large photometric and spectroscopic galaxy surveys has offered statistically significant samples to study the SHMR including its dependence with galaxy properties and redshift. Simulations, both semi-analytical and hydrodynamical, have worked in synergy and provided a more detailed look into the various physical mechanisms that act on galaxies in halos. However, there remain a number of limitations in the study approaches which leaves us with open questions regarding the relation between stellar and halo mass, and especially its redshift evolution. In this section, we discuss some limitations and open questions that remain, non-extensively, focusing on those relevant to what this thesis aims to shed light upon.

Most of the investigations in the literature employ (sub)halo abundance matching to study the SHMR. (SH)AM techniques are based on a simple but powerful assumption of how galaxies reside in dark matter halos – they are matched in their abundance. AM techniques are a powerful tool, but suffer from some shortcomings. One of the more relevant is that the satellites are treated as centrals in their own sub-halo so that they only predict the SHMR of central galaxies. As we saw in the previous sections, the HOD approach coupled with clustering (and lensing) measurements are sensitive to the satellite galaxies and can allow one to estimate the stellar mass budged of satellites in dark matter halos. This is the approach adopted in this thesis.

An important aspect in understanding the cosmic stellar mass assembly in halos and the effect of the physical mechanisms that govern it is measuring the redshift evolution of the SHMR across a vast redshift range. The redshift evolution of the SHMR has been measured in a number of works in the literature (e.g., Conroy and Wechsler, 2009; Moster et al., 2010; Behroozi, Wechsler and Conroy, 2013; Moster, Naab and White, 2018; Behroozi et al., 2019). However, at redshifts of $z > 3$ the constraints on the SHMR in the literature are less certain and diverge. One of the main limitation has been the compilation of observables to constrain the models with. A number of works, for example, constrain the SHMR at high redshifts using GSMF derived from UV-to-M_* conversions to estimate stellar masses, which comes with caveats and uncertainties. One of the main limitations of previous works is that they have relied on observables from heterogeneous data sets to constrain their models. Different data sets can have different selection functions and methods of estimating galaxies' physical parameters that can propagate various systematic biases that can muddle the interpretations (Behroozi, Conroy and Wechsler, 2010). The aim of the work in this thesis is to use a homogeneous dataset to study the redshift evolution SHMR, and be free from such 'inter-observations' systematic errors. Finally, measurements of the contribution of both central and satellite to the SHMR exist in the literature but are limited only to $z < 1$ (e.g. Leauthaud et al., 2012; Coupon et al., 2015), or a single z-bin measurement at $2 < z < 3$ as in Cowley et al., 2019.

One of the main questions regarding the SHMR evolution is how does the SFE evolve with redshift. Currently, there is no consensus that the SFE evolves with redshift at all. On one side, some studies suggest that the SFE does not evolve considerably with redshift. These include studies from HOD modeling of clustering (Harikane et al., 2018), abundance matching (Stefanon et al., 2021), as well as empirical models of galaxy formation histories (Tacchella et al., 2018). On the other side, other studies find that the SFE does evolve with redshift. These include abundance matching from Legrand et al. (2019), empirical models of Moster, Naab and White (2018), Behroozi et al. (2019) and Girelli et al. (2020).

The debate of whether the SFE evolves or not, especially as we enter the redshifts of the epoch of reionization is, at the moment of writing of this thesis, a crucial question. This is because results from analyzing the first batch of *JWST* data have shown an excess of bright and massive galaxies at $z > 10$ compared to theoretical predictions based on non-evolving

SFE. One way to reconcile this observational challenge of the theoretical models is to allow a significantly higher (up to a factor of 10) SFE, as discussed in Furtak, Shuntov et al., 2022. Determining the SHMR and consequently the SFE is a key open question that needs to be answered in light of *JWST*.

In measuring the redshift evolution of the SHMR, it is important to have accurate stellar mass measurements that are based on optical rest frame observations red wards of the Balmer break. This ensures to sample the light from older stars that dominate the stellar mass budged of a galaxy. At $z > 3-5$, which is a regime that becomes increasingly interesting, this is made possible with observations in the infrared from *Spitzer* and now *JWST*. Therefore, large-area surveys with deep multi-band imaging, especially in the IR, such as the COSMOS survey (Scoville et al., 2007) and the upcoming COSMOS-Web (Kartaltepe et al., 2021), are crucial in constraining the redshift evolution of the SHMR

This thesis aims to address the questions laid here, that is to measure the evolution of the SHMR across a large redshift range (out to $z \sim 5$) using clustering and GSMF measurements from a single dataset (the COSMOS2020 catalog) by employing an HOD modeling approach that constrains both the central and satellite galaxy contribution to the stellar mass budget in a dark matter halo. Additionally, it incorporates and investigates the prospects an additional probe – weak lensing magnification

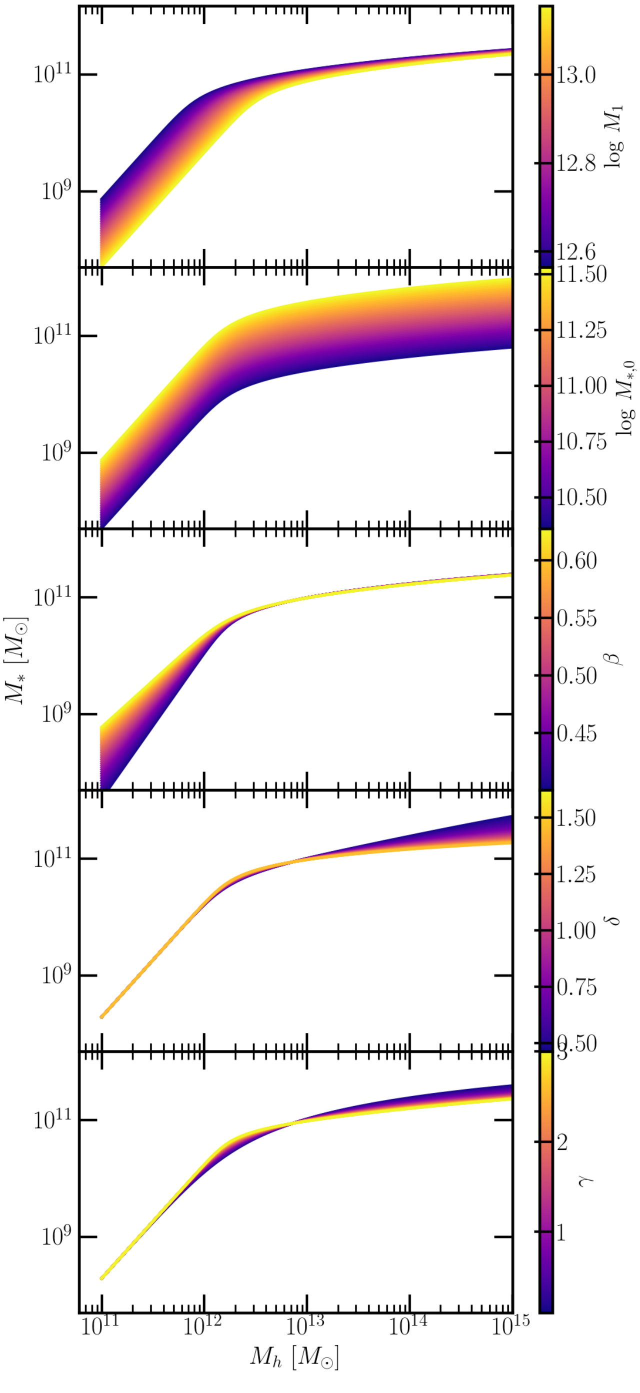


Figure 4.2: Stellar mass as a function of halo mass as parametrized by Eq. 4.3. In the different panels, we show how the SHMR changes by varying one parameter at a time, while fixing the others to their fiducial values: $\log M_1 = 12.80$, $\log M_{*,0} = 10.92$, $\beta = 0.51$, $\delta = 0.92$, $\gamma = 2.85$.

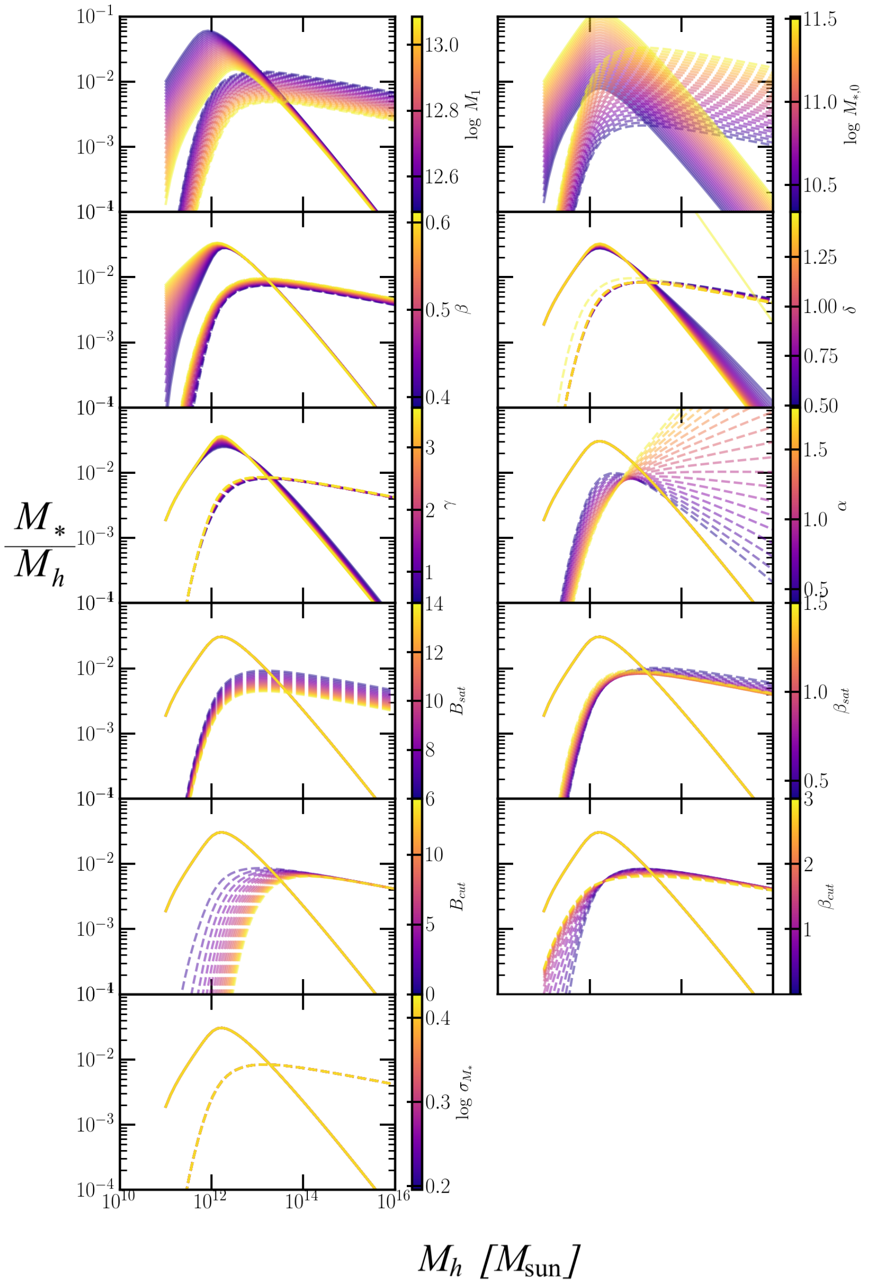


Figure 4.3: Total stellar-to-halo mass relation as parametrized by Eq. 4.10. In the different panels, we show how the total SHMR changes by varying one parameter at a time, while fixing the others to their fiducial values: $\log M_1 = 12.80$, $\log M_{*,0} = 10.92$, $\beta = 0.51$, $\delta = 0.92$, $\gamma = 2.85$, $\alpha = 0.87$. $B_{\rm sat} = 8.46$, $\beta_{\rm sat} = 0.89$, $B_{\rm cut} = 0.41$, $\beta_{\rm cut} = 0.13$, $\sigma_{{\rm Log}M_*} = 0.31$. The solid/dashed lines show the change in the central/satellite contribution to the total SHMR.

5

NUMERICAL SIMULATIONS

In establishing our current understanding of the formation of the large-scale structure, galaxies and their subsequent evolution, numerical simulations play a key role. The advent of ever more powerful computers allowed to first simulate the gravitational interaction of dark matter particles with an increasing precision and volume, and simulate the hydrodynamics of baryonic matter. Today, numerical simulations are an indispensable tool to conduct research in astrophysics. First, they provide a framework in which we can implement all of our known physical laws and models of the evolution of the dark and baryonic components of the Universe, create mock observations that when compared to real observations can help us improve our models. Second, the effects of various physical processes on shaping the properties of the LSS and galaxies can be disentangled and studied in greater detail by comparing and calibrating them with real observations. Such an approach has led, for example, to establishing the current paradigm of cold flows, stellar and AGN feedback regulating galaxy growth in DM halos (see § 2.3). Third, they are essential in the preparation of future observations, surveys and scientific projects. Mock catalogs can realistically simulate a survey by applying the instrument characteristics and predict the performance to ensure that the project requirements are met. Given their importance, and their use in this thesis, we give a brief overview of numerical simulations.

Numerical simulations are typically separated in two classes, depending on the matter component (dark vs. baryonic) and physical interactions that they simulate (gravity vs. gas hydrodynamics) (see e.g., Somerville and Davé, 2015; Vogelsberger et al., 2020, for reviews).

Semi-analytical models rely on dark-matter simulations that involve only the laws of gravity in a cosmological context to simulate structure formation and evolution. Since galaxies form and evolve within DM halos, they provide the backbone for galaxy formation models. In these simulations, the dark matter in a representative volume of the Universe is divided into $N \sim 10^{12}$ particles whose forces of gravity are computed by numerically solving the Poisson's equation and the system is evolved forward in discrete time-steps. Such methods are called N-body and can be particle-based, mesh-based or hybrid. Virialized halos are identified in the simulations with one of several methods such as friends-of-friends, spherical overdensity or phase-space based methods. The evolution and merging history of dark matter halos is recorded in *merger trees* that trace a halo in the current epoch to its progenitors in previous epochs. In semi-analytical models, galaxies are added in DM halos by following simplified physical prescriptions of gas cooling, star formation, supernova and AGN feedback, and galaxy merging. The resulting galaxy population can then be compared against observations and different model variations can be tested. Semi-analytical models are very efficient with reasonable computational resources and have played a key role in unveiling how gas collapses, forms stars and galaxies within the dark matter environment and the role supernovae and AGN feedback has played. However, they have limited realism given the fact that they rely on simplified recipes and detailed studies of gas properties and physical mechanisms that form galaxies is not directly possible. Because of this they fail to reproduce, for example, the dispersion and diversity of observed galaxy properties as a function of mass.

Hydrodynamical simulations simulate the co-evolution of dark matter and baryons consistently. For the baryonic part, this is done by solving the Euler equations that govern the conservation of mass, momentum and energy of ideal gas assuming a non-relativistic equation of state. Furthermore, they can also simulate the magnetic fields, cosmic rays and radiation fields, thus increasing the realism. The hydrodynamical equations are typically solved with three classes of methods: Lagrangian, Eulerian or arbitrary Lagrange-Eulerian techniques, all having their advantages and caveats and employed depending on the purpose of the simulation. The physical processes that govern galaxy formation such as gas cooling, star-formation, stellar and AGN feedback are implemented on a finite resolution scale of $\sim$ kpc for cosmological simulations – termed *subgrid physics*. The free parameters of the subgrid models usually are calibrated so that the simulation reproduces observations of a given distribution of galaxy properties. Typically, these observations are the galaxy stellar mass function, stellar-to-halo mass relation, black hole-galaxy relation and star-formation rate density, most often at $z = 0$. Hydrodynamical simulations offer us the most detailed insight into the physical processes that govern galaxy formation and evolution. However, they are computationally expensive and thus are limited in resolution and volume. For example, one of the largest hydrodynamical simulation BAHAMAS (McCarthy et al., 2017) has a volume of $571^3\,\mathrm{Mpc}^3$, but low dark/gas mass and spatial resolution of 5.5×10^9 / $1.1 \times 10^9\,M_\odot$ and 4 kpc respectively. Furthermore, different implementations of the subgrid recipes still leads to some discrepancies between the simulated and observed properties of galaxies, more so in the regimes in which the simulations have not been calibrated.

The computational efficiency of hydrodynamical simulations continues to increase in parallel with refining the solving algorithms, improving resolution and increasing complexity and realism of the implemented physical processes. As such, they are set to play an ever-increasing role in improving our detailed understanding of the physical processes that shape the formation and evolution of galaxies. Given their importance and insight, in this Thesis we use three state-of-art cosmological hydrodynamical simulations of galaxy formation to compare our principal observational results: Horizon-AGN (Dubois et al., 2014; Kaviraj et al., 2017), EAGLE (Crain et al., 2015; Schaye et al., 2015) and Illustris TNG-100 (Springel et al., 2018; Nelson et al., 2018; Pillepich et al., 2018; Marinacci et al., 2018; Naiman et al., 2018).

Numerical simulations continue to increase in realism and resolution, and coupled with a steady increase in computing power, they are set to continue playing a key role in understanding galaxy formation and evolution. In this Thesis, we use hydrodynamical simulations to compare our results with observations. The goal of such comparisons is to obtain a more detailed insight into which physical processes might govern the observed behavior of measured quantities, such as GSMF, correlation functions and the relation between stellar and halo mass.

Part II

Unveiling the galaxy-halo connection with photometric surveys

6

PHOTOMETRIC SURVEYS

6.1 HISTORICAL DEVELOPMENT OF PHOTOMETRIC SURVEYS

Modern photometric surveys are observations of wide-field observations of the sky in a given wavelength range of the electromagnetic spectrum. The ability to capture and quantify light coming from objects near and extremely far in space has had a transformative effect on our understanding of the Universe, and such surveys represent an important component of modern astrophysics. The first photometric surveys of the sky date back to Bigourdan (1888), and used photographic plates to quantify the light emitted by celestial objects. Such surveys have led to a detailed map of the large scale distribution of matter and have revealed the true size and scale of our Universe (Hubble, 1934; Shane and Wirtanen, 1954). More recently, charge-coupled device (CCD) detectors have allowed a real breakthrough in rapidly and deeply imaging the extragalactic sky, while providing precise photometric measurements. Instruments such as the SDSS' large-format CCD camera (Gunn et al., 1998), CFHT's 8K mosaic camera (UH8K) (Lilly et al., 1995) and MegaCam (Boulade et al., 2003) allowed the first degree-scale deep photometric surveys that opened a window into earlier epochs of the history of our Universe. Multi-object spectroscopic instruments like the Visible Multi-Object Spectrograph (VIMOS) (Le Fèvre et al., 2003) complemented the photometric surveys by providing spectroscopic redshifts for many objects, crucial for photometric redshift calibrations and precise measurements of physical parameters such as stellar mass and star formation rate.

6.2 OBSERVING THE DISTANT UNIVERSE FROM SPACE

A breakthrough in studying the faintest and more distant sources was made with the launch of the Hubble Space Telescope (HST). A space-based telescope, free from atmospheric effects, meant that we could survey the faintest sources with exquisite resolution. This resulted in the first ultra-deep fields in multiple optical bands, such as the Hubble Deep Field (HDF) using the Advanced Camera for Surveys (ACS) (Williams et al., 1996), Great Observatories Origins Deep Survey (GOODS) (Giavalisco et al., 2004), and Hubble Ultra Deep Field (HUDF) using both ACS and Wide-Field Camera 3 (WFC3) (Beckwith et al., 2006). These surveys, in combination with other ground based data, allowed the measurement of the cosmic star-formation and stellar mass assembly throughout the majority of the history of the Universe (e.g., Marchesini et al., 2009). However, the relatively small field-of-view of HST's instruments (e.g., 202 × 202 arcsec for ACS and 135 × 127 arcsec for WFC3) meant that it was only capable of small pencil-beam surveys (e.g., two fields of 16 × 10 arcmin fields for GOODS) that are inefficient in mapping the large-scale structure as well as detecting populations of rare high-redshift galaxies. A step towards surveying larger ares was made with Cosmic Assembly Near-infrared Deep Extragalactic Legacy Survey (CANDELS) (Grogin et al., 2011; Koekemoer et al., 2011, a total area of ~ 800 arcmin^2 over 5 separate fields), but still suffered from cosmic variance limitations due to the small area of the individual fields. This remains one of the main limitations of ultra sensitive space telescopes like HST – it provides a very deep view of the faintest and most distant objects in a pencil beam volume, but can not provide census at lower redshifts $z < 1$. In this regime, ground based observatories in the optical and UV continue to excel.

Measuring the majority of light emitted by stars of all populations in galaxies, especially going to higher redshifts, requires a space-based telescope in the near and mid-infrared. A unique view into the infrared universe was made possible by the launch of the *Spitzer* Space Telescope (Werner et al., 2004). Its instruments IRAC, IRS and MIPS spanned the wavelength range from $\sim 3.6 - 160\,\mu$m and quantified the light emitted from more evolved stellar populations and dust in galaxies. The *Spitzer*/IRAC camera (Fazio et al., 2004) was a crucial instrument in all deep extragalactic surveys seeking to measure photometric redshifts and stellar masses of objects at $z > 3$ and find the most distant galaxies. Even after its decommissioning in 2021, the legacy data from *Spitzer* remains an important component of deep and wide photometric surveys such as the Euclid Deep Survey (Euclid Collaboration et al., 2022a; Moneti et al., 2021), that promise to provide a census on the cosmic evolution of the brightest, most massive and rarest galaxies.

The next breakthrough in studying the visible and near-infrared Universe from space is set to come from the upcoming space mission *Euclid*, set for launch in 2023. The *Euclid* mission is designed with a core scientific goal to understand the accelerated expansion of the Universe and investigate the nature of dark energy and dark matter via weak lensing and galaxy clustering (Laureijs et al., 2011). To this end, the *Euclid* survey is designed to image about 15 000 deg^2 of the extragalactic sky over six years (Euclid Collaboration et al., 2022a, for a detailed description of the *Euclid* survey). The telescope is equipped with a Visible Imager (VIS) and Near Infrared Spectrometer and Photometer (NISP). The imager will obtain broad-band photometry in one very broad visible filter covering the r, i, z bands ($I_{\rm E}$) and three near-infrared filters ($Y_{\rm E}$, $J_{\rm E}$, $H_{\rm E}$). The large field of view (0.5 deg^2) and a design optimized to result in a pristine PSF, the VIS imager will provide exceptional quality for the images of galaxies, whereas NISP will measure broad band photometry in the observer NIR. These characteristics of Euclid, including its planned Deep Survey of ~ 50deg^2 will enable a panoply of galaxy formation and evolution science.

In the longer term, the currently under development Nancy Grace Roman Space Telescope will provide an unprecedented imaging and spectroscopy with its large field of view (0.28 deg^2) excellent resolution in optical to near-infrared wavelengths ($0.5 - 2.3\,\mu$m) that can reach 5σ depths of AB $\sim 26 - 28$ in one hour of imaging. It will allow accurate studies of the nature of dark energy and dark matter and a rich dataset for a panoply of galaxy evolution studies.

6.3 INTRODUCING THE COSMOS FIELD

A major step towards deep and wide multi-band surveys was made with the Cosmic Evolution Survey (COSMOS, Scoville et al., 2007) that began in 2003. HST's ACS imaged a 1.7 deg^2 mosaic, reaching a 5σ depth of ~ 27 AB at $\sim 0.8\,\mu$m. Since then, the COSMOS field has been extensively covered by most ground and space-based observatories. These observations span virtually the entire electromagnetic spectrum[1]: FUV and NUV bands from GALEX (Zamojski et al., 2007), broad, intermediate and narrow optical bands from Subaru Suprime-Cam and Hyper Suprime-Cam (HSC) (Capak et al., 2007; Taniguchi et al., 2007; Taniguchi et al., 2015), near-infrared from the VISTA telescope (McCracken et al., 2012), near and mid-infrared from the Spitzer Space Telescope (Sanders et al., 2007), far-infrared from Herschel (Oliver et al., 2012), sub-millimeter and radio from the SCUBA-2 camera on East Asian Observatory's James Clerk Maxwell Telescope and the Atacama Large Millimeter Array (ALMA) (Le Fèvre et al., 2020). A next chapter in the COSMOS survey will be opened with the recent launch of the James Webb Space Telescope (JWST) and the COSMOS-Web survey (Kartaltepe et al., 2021) that will provide 0.6 deg^2 wide observations at $\sim 1-4.5\,\mu$m down to ~ 28 AB. This survey promises to revolutionize our understanding of the spatial distribution of the reionization and characterize the masses and star formation histories of the earliest quiescent galaxies. Additionally, one of the main scientific goals of COSMOS-Web will be to measure the evolution of the relation between stellar and halo mass, which is also the subject of this Thesis. To this end, part of the work during this Thesis (especially in the last 8 months) has been devoted to the preparation of the photometric catalog extraction in COSMOS-Web.

The rich deep and multi-band data over ~ 2 deg^2 in COSMOS allowed the estimation of accurate photometric redshifts for millions of galaxies down to the epoch of reionization. Coupled with spectroscopic redshift measurements for many galaxies, necessary for the calibration of photo-z techniques, has enabled the construction of large statistical samples of galaxies with accurate photo-z and stellar mass measurements. Over the years, as new data was incrementally arriving, several COSMOS photometric catalogs have been publically released that have been widely used by the extragalactic community (Capak et al., 2007; Leauthaud et al., 2007; Ilbert et al., 2009; Ilbert et al., 2013; Muzzin et al., 2013; Laigle et al., 2016; Weaver et al., 2021).

6.4 DEPTH VS. AREA

Photometric surveys, when being designed, are typically faced with a choice of how to spend the awarded telescope time: cover a large but shallow area, or a small but very deep

---

[1] At least the range of interest in extragalactic studies, from the ultraviolet to radio

area. This is determined by the scientific goal of the survey. Fig. 6.1 presents the depth versus area diagram for several surveys. It shows *Spitzer*/IRAC surveys and their channel 1 (3.6 μm) depth for several surveys in the open circles and squares and for the Euclid Deep fields in lines (Moneti et al., 2021). The relation between the logarithms of the area and the flux depth is roughly linear – deep surveys have smaller area, while wide surveys have shallower depth. The next generation photometric surveys such as LSST, Roman Space Telescope and Euclid Deep will cover unprecedented depths and area $\sim$ 28 AB at 5σ and $\sim 18,000$ deg^2, $\sim 2,000$ deg^2 and ~ 50 deg^2 respectively.

Large-area surveys, sufficiently deep to detect galaxies at redshifts in the epoch of reionization ($z > 6$), will be able to find statistically representative samples of the rarest, brightest and most massive galaxies in the early Universe. This will be made possible by the huge cosmological volumes that they will probe thanks to their large areas. Assembling representative samples of bright and massive galaxies will be crucial to settle the debate on the nature of the bright/massive ends of the luminosity/stellar mass functions and consequently of the assembly and evolution of this population when the universe was less than 1 Gyr old (e.g., Stefanon et al., 2019; Bowler et al., 2020; Kauffmann et al., 2022, and references therein). The statistical power of these wide surveys will also allow us to study in better detail the evolution of the relation between stellar and halo mass at high redshifts. The difficulty of assembling large samples of massive galaxies with increasing redshift means that the massive end of the SHMR remains largely unconstrained in the literature. Another open question is that of the unknown evolution of the SHMR (or consequently the star-formation efficiency) at redshifts beyond the cosmic noon, as discussed in § 4.5. The current and upcoming wide, deep and multi-band surveys such as COSMOS and Euclid Deep provide the ideal grounds to study these questions.

6.5 THE ADVANTAGE OF PHOTOMETRIC SURVEYS TO STUDY THE GALAXY-HALO CONNECTION

In Chapter 4 we described how the HOD formalism can be used to model the relation between stellar mass and halo mass. The convenience of this modeling is in the fact that the HOD can also be used to model key observables such as the two-point correlation function and GSMF. In order to measure them for complete galaxy populations down to high redshifts, one needs deep and multi-band photometric surveys covering the near-infrared. These allow detecting, measuring photometry and deriving physical parameters for well-defined sample of galaxies down to a known mass completeness limit. The detection in near-infrared bands allows achieving high completeness in stellar mass down to high redshifts (i.e., down to $z \sim 6$ and $\log(M_\star/M_\odot) > 10$ in COSMOS2020, Weaver et al., 2021). This means that key observables can be accurately measured for various galaxy populations selected based on their photometric redshift, stellar mass, type, etc. For this reason, to tackle the main goal of this Thesis, which is the study of the cosmic evolution of the galaxy-halo connection, photometric surveys provide the main dataset.

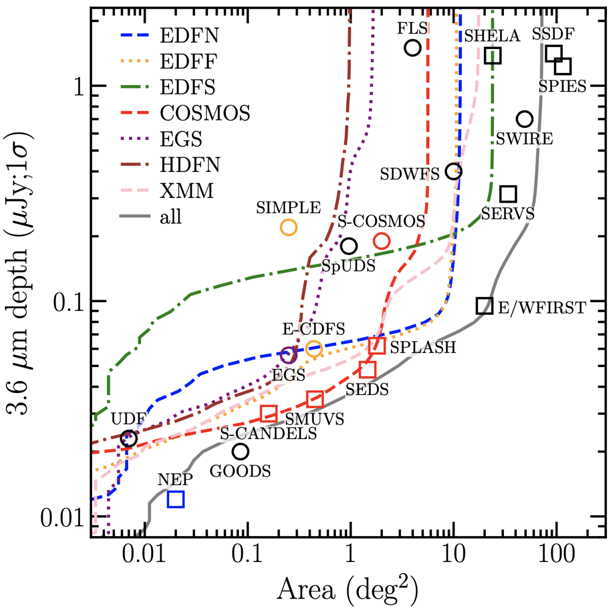


Figure 6.1: IRAC channel 1 depth as a function of survey area for different surveys. The colored lines correspond to the 1σ depths measured in 2.5 arcsecond diameter apertures in the Euclid Deep fields, while the gray line corresponds to the total survey area (Moneti et al., 2021).

6.6 PREPARATION OF PHOTOMETRIC CATALOGS

One of the main studies carried out in this Thesis is based on the COSMOS2020 photometric catalog (Weaver et al., 2021 and described in detail in Chapter 7). To this end, a considerable amount of time in the first year of the Thesis was devoted to the preparation of the COSMOS2020 catalog. This involved the extraction of photometry in the *Spitzer*/IRAC images, construction of survey geometry masks, validation of the photometry and source detection strategy by measuring the two-point angular correlation function as a function of apparent magnitude. This work is briefly described below.

IRAC photometry for COSMOS2020

We extracted photometry on the *Spitzer*/IRAC channels 1 (at 3.6 μm) and 2 (at 4.5 μm) images using the `IRACLEAN` software (Hsieh et al., 2012) for COSMOS2020. These include the latest and deepest IRAC observations of COSMOS, which is an improvement compared to the previous iteration of the catalog. The IRAC images have a larger PSF (with FWHM between 1.6 and 2.0 arcseconds) compared to the optical data and are significantly affected by source confusion which prevents reliable photometric extraction. To tackle this issue, `IRACLEAN` uses a high-resolution image (and its segmentation map) as a prior to identify the centroid and the boundaries of the source, and iteratively subtract a fraction of its flux ('cleaning') until it reaches some convergence criteria specified by the user. The high-resolution image and segmentation map is derived from chi-squared detection image constructed from the $izJYHK_s$ bands. `IRACLEAN` works in the approximation that an IRAC source can be modeled as a scaled Dirac delta function convolved with the PSF. The fi-

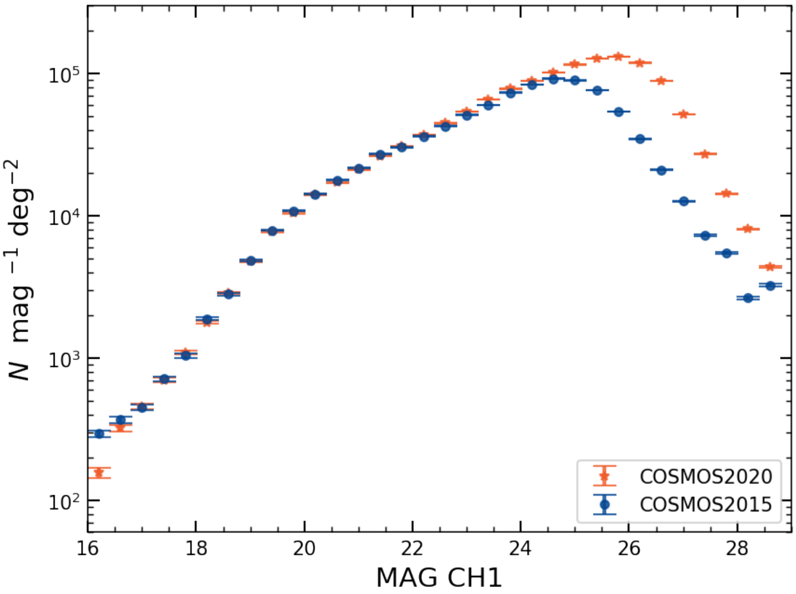


Figure 6.2: IRAC photometry in COSMOS2020. IRAC channel 1 (at 3.6 μm) number counts from the COSMOS2020 compared to the previous iteration of the catalog, COSMOS2015.

nal flux of each object is the sum of the fluxes subtracted at each step. Since the centroid position is allowed to change at every iteration, the source is eventually modeled by a combination of Dirac delta functions that are not necessarily centered at the same point. The flux error is computed using the residual map by measuring the fluctuations in a local area around the object.

Fig. 6.2 shows the resulting magnitude number counts compared to the previous iteration of the catalog, COSMOS2015 (Laigle et al., 2016). The improved depth by about a magnitude comes from deeper IRAC observations and from the deeper high resolution prior made from the $izJYHK_\mathrm{s}$ images. This allows improving the stellar mass completeness limit and improve the accuracy of photo-z and stellar mass estimates for the fainter and more distant galaxies.

Construction of masks for COSMOS2020

Photometric extraction of sources can be significantly affected by the spurious flux of nearby bright stars, galaxies, and various other artifacts in the images. Thus, it is required to mask these sources. For this purpose, the COSMOS2020 catalogs provide flags for objects in the vicinity of bright stars, and for objects affected by various artifacts.

Additionally, clustering measurements using the 2pt correlation function require an accurate knowledge of the survey geometry, which is provided by these masks. As such, they are an important ingredient in the main work of this Thesis that involves the computation of correlation functions.

The bright-star masks from the HSC-SSP PDR2 (Coupon et al., 2018) are used to flag problematic sources near stars. In particular, masks are taken from the Incremental Data Release 1 revised bright-star masks that uses Gaia DR2 as a reference star catalog, where stars brighter than $G = 18$ mag are masked. About 18 % of sources in the catalog are

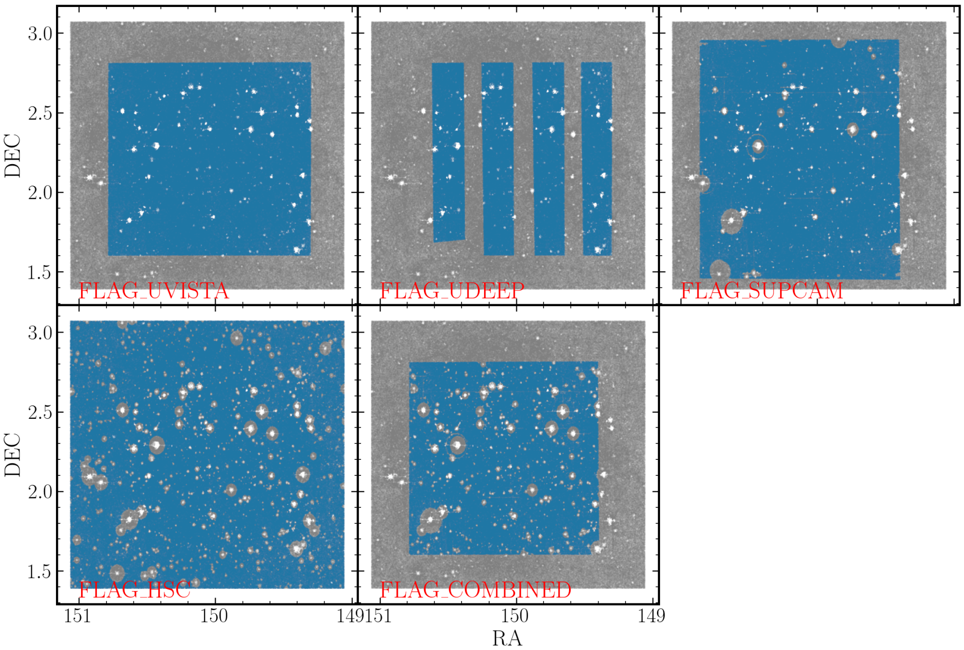


Figure 6.3: Masks in the COSMOS2020 catalog. Each panel shows the mask that corresponds to the catalog keyword written in the panel.

found within the masked regions in the vicinity of bright stars. Furthermore, artifacts in the Suprime-Cam images are masked using the same masks as in COSMOS2015.

The masks are shown in Fig. 6.3. A conservative combined mask is prepared for sources within 1.27 deg^2 which have coverage from HSC, UltraVISTA, and IRAC but which are not close to bright stars or large artifacts.

Catalog validation with the angular correlation function

Finally, we performed a validation of the catalog and its detection strategy by measuring the two-point angular correlation function $w(\theta)$ and its scaling with apparent magnitude. The principal strategy in COSMOS2020 is to detect objects in a chi-squared combination of $izJYHK_s$ bands (described in more detail in Chapter 7). For comparison, we also measure $w(\theta)$ for a K_s selected catalog. The angular correlation function for sources selected on apparent magnitude follows a power law with a decreasing amplitude for fainter samples. Its power law behavior (especially at smaller scales) and its scaling with magnitude is a standard test if the catalog is a fair sample and the survey completeness is well understood (e.g., Peebles, 2001).

The measurements of $w(\theta)$ in an $izJYHK_s$ and a K_s detected catalog is shown in Fig. 6.4 in orange and blue points correspondingly. The measurements are done for all sources in the catalog without removing stars. This does not pose a problem since stars are expected to be uncorrelated and would only introduce noise, but no bias. The first four panels correspond to essentially upper magnitude thresholds. The fifth to ninth panels show $w(\theta)$

in bins of K_s magnitude. The parameters of the power law fit are shown in each panel on the top right for the $izJYHK_s$-selected and bottom left for the K_s-selected catalog. The fit is performed on small scales $\theta < 0.01$ deg. The amplitude of the correlation decreases with apparent magnitude, as expected. Additionally, the slope of $w(\theta)$ becomes flatter towards fainter magnitudes, in agreement with previous measurements in COSMOS (McCracken et al., 2007).

The agreement between the two, the power law behavior and the scaling with apparent magnitude serve as a qualitative validation of the catalog. We note that there is a deviation from the power law with an extra power at $\sim 0.1\text{deg}^2$. This can be explained by the known deviation of the correlation function from a power law due to the two-halo contribution, as explained in Chapter 3. Another reason for this enhancement can be from the known overabundance of rich structures in the COSMOS field at $z \sim 1$ (McCracken et al., 2007; Meneux et al., 2009). At even larger scales, the sudden drop in power is due to the limited size of the survey $\sim 1.3\,\text{deg}^2$ and the effects of the integral constraint (described in more detail in Chapter 7). These tests served as a validation that the survey geometry masks and completeness are valid and understood, and we did not proceed to a more quantitative analysis of the resulting $w(\theta)$.

IRAC photometry extraction and validation in the Cosmic Dawn Survey

The Cosmic Dawn Survey is a multi-wavelength, deep, and wide survey in the optical and NIR. It comprises imaging from the Subaru telescope and Hyper SuprimeCam, complemented in the future by the Vera C. Rubin Observatory in the optical bands, and *Spitzer*/IRAC in the NIR bands. To this end, new observations and all relevant archival data from *Spitzer*/IRAC in the Euclid Deep and Auxiliary Fields have been processed to provide photometry at $3 - 5\,\mu$m down to AB ~ 25 (Moneti et al., 2021).

During this Thesis, a considerable effort was put in the photometric validation and photometry extraction in the *Spitzer*/IRAC observations of the Euclid Deep and Auxiliary fields. Here we briefly present this work.

We extracted source catalogs from channels 1 (at $3.6\,\mu$m) and 2 (at $4.5\,\mu$m) stacks of all fields using `SExtractor`. We adopted the usual approach, where we detect objects that contain a specified minimum number of connected pixels above a noise threshold (in this case 2σ). Photometry is then measured for the detected objects in fixed apertures. Due to the large PSF, IRAC data are known to contain many blended sources, and in the case of our moderately deep data, this approach can miss faint sources. However, this shallower catalog is sufficient for our quality assessment purposes. The measured fluxes were converted from MJy/sr to AB magnitude using a zero-point of 21.58, and the latter were converted to total magnitude using the aperture corrections given in the IRAC Instrument Handbook for the warm mission (-0.1164 and -0.1158 for channel 1 and channel 2, respectively)

We computed the differential number counts in channel 1 in each field using the corrected $7''$ aperture magnitudes. These are shown in Fig. 6.5, where the red circles with uncertainties present our measurements, and the lines show the number counts from the

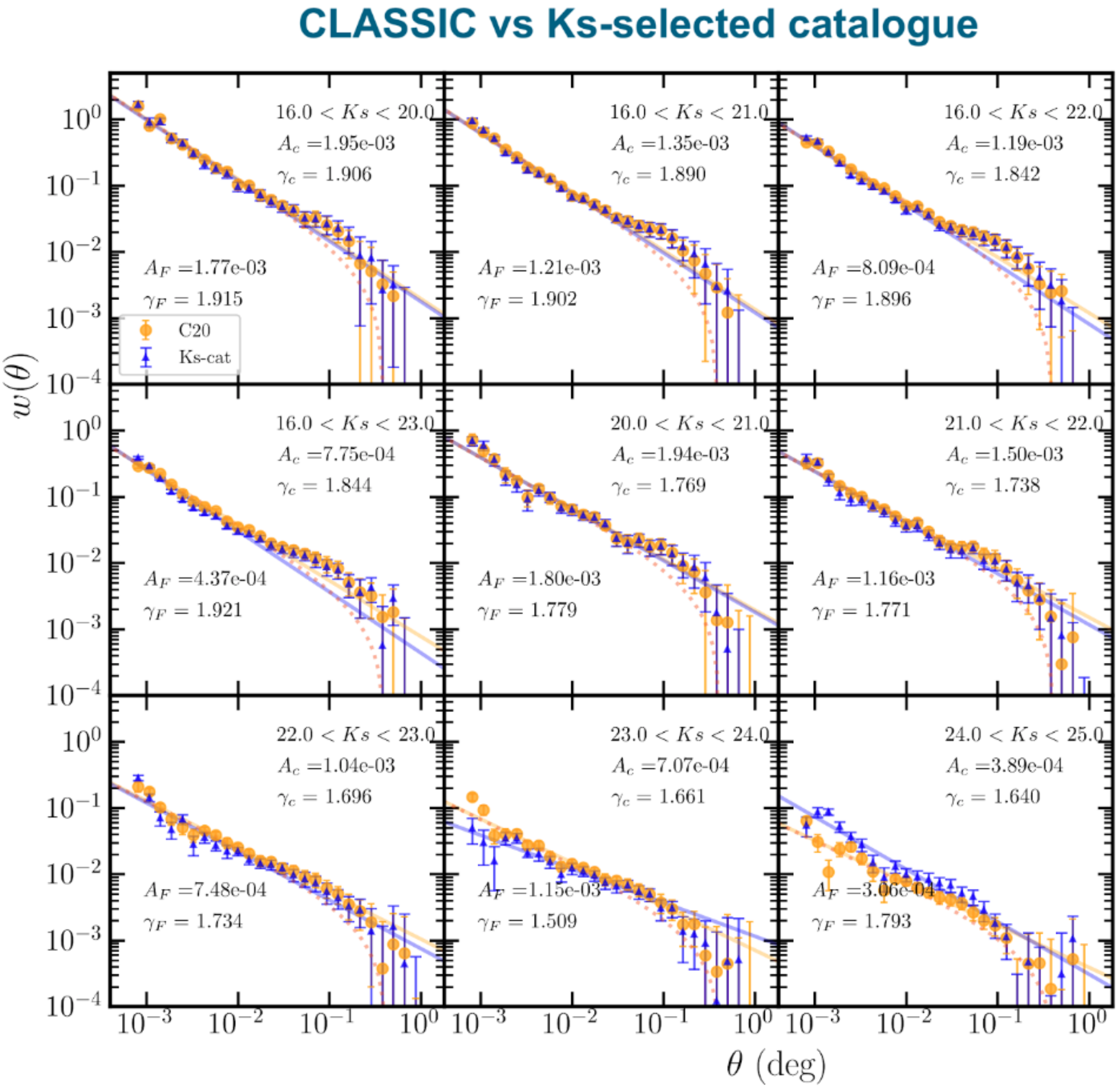


Figure 6.4: Scaling of the two point angular correlation function with apparent magnitude in COSMOS2020. Each panel shows the 2PCF for a different magnitude bin in catalogs extracted from a K_s and chi-mean $izJYHK_s$ (CLASSIC) detection image in blue and orange points correspondingly. The parameters of the power law fit are shown in each panel on the top right for the $izJYHK_s$ catalog and bottom left for the K_s catalog. The agreement between the two, the power law behavior and the scaling with apparent magnitude serve as a qualitative validation of the catalog

literature; the bottom-right panel shows the mean of all fields. We compared our number counts with those presented in Ashby et al. (2013), which also surveyed many of our fields and with those computed using the new COSMOS2020 photometric catalog (COSMOS2020) that we used as the reference.

There is a general agreement in the number counts in all the fields with Ashby et al. (2013) and COSMOS2020 for $16 < [3.6] < 22$. At brighter magnitudes, the COSMOS2020 counts drop off as bright sources were not included. At fainter magnitudes, our aperture-based catalogs are confusion-limited and thus incomplete. Conversely, the COSMOS2020 catalog, which uses a high-resolution prior for the detection and a profile-fitting method for the measurement, is complete up to significantly fainter magnitudes. Counts for EDF-N are slightly higher than the other fields at bright magnitudes. We investigated that this is due to the higher density of stellar sources in EDF-N.

The IRAC photometry in the Cosmic Dawn Survey was used in this Thesis to search for galaxy clusters at $z > 3$ by searching for overdensities of red IRAC sources. This is presented in Chapter 10.

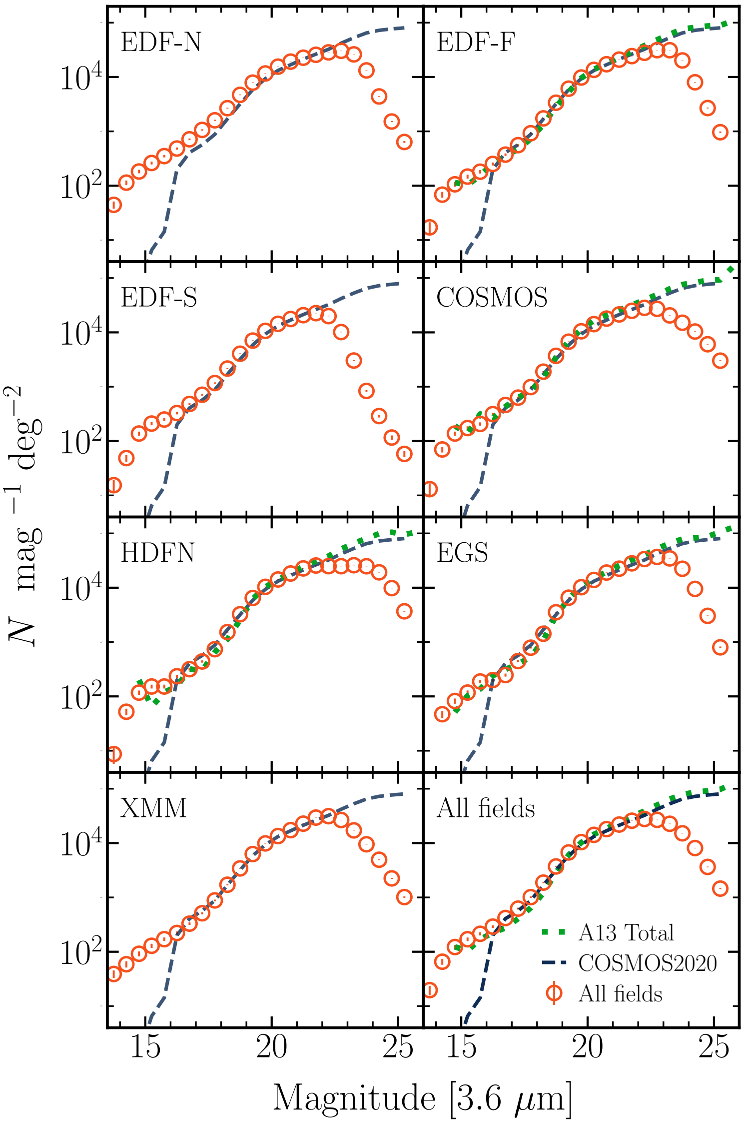


Figure 6.5: Magnitude number counts in channel 1 (red circles) together with COSMOS2020 (long dashed lines) and Ashby et al. (2013, hereafter A13, short dotted green lines). The bottom right panel shows the mean of all fields, and the legend there applies to all panels. Figure adopted from Moneti et al. (2021)

7

STELLAR-TO-HALO MASS RELATION IN COSMOS2020

7.1 INTRODUCTION

The aim of the work in this Chapter is to constrain the redshift evolution of the SHMR up to $z \sim 5$ by applying an HOD-based analysis consistently on a homogeneous data set: the COSMOS2020 photometric catalog. COSMOS2020 (Weaver et al., 2021) is the latest iteration of the photometric catalog of the COSMOS (Scoville et al., 2007) survey that includes the latest data-releases of deep imaging, covering wavelengths from the ultraviolet to the near-infrared. The deep multi-band photometry allows for the estimation of accurate photometric redshifts and stellar masses, and selection of complete samples up to high redshift. By adopting an HOD-based model to jointly fit for galaxy abundance and clustering, our analysis aims at constraining the satellite contribution to the total stellar mass budget in halos across a vast redshift range. This allows us to infer a coherent picture of how the stellar mass assembles as a function of the halo mass throughout cosmic history.

The novelty of our work is the use of a single dataset to perform all the measurements and probe the SHMR to $z \sim 5$ for both central and satellite galaxies. Most of the investigations in the literature have relied on observables from heterogeneous data sets to constrain their models (e.g. Behroozi, Wechsler and Conroy, 2013; Moster, Naab and White, 2018; Behroozi et al., 2019). Different data sets can have different selection functions and methods of estimating galaxies' physical parameters that can propagate various systematic biases that can muddle the interpretations (Behroozi, Conroy and Wechsler, 2010). Therefore, our work is free from such 'inter-observations' systematic errors.

This work builds up on the literature in several ways. Legrand et al. (2019) is the only work that measures the SHMR using a single data set (COSMOS2015 of Laigle et al., 2016), up to $z \sim 5$ using sub-halo abundance matching. The shortcoming of this approach is the satellites are treated as centrals in their own sub-halo so that they only predict the SHMR of central galaxies. Our approach allows to decouple centrals and satellites to compute the contribution of both to the total mass content of halos. Previous work that have measured both central and satellite SHMR are limited only to $z < 1$ (e.g. Leauthaud et al., 2012; Coupon et al., 2015), or a single z-bin measurement at $2 < z < 3$ as in Cowley et al. (2019). Therefore, this work presents the only measurement of the SHMR for both centrals and satellites up to $z \sim 5$ using a homogeneous dataset: COSMOS2020.

The organization of this chapter is as follows. In Section 7.2 we describe the COSMOS dataset we use and the mass-selected samples in the ten redshift bins of our analysis. In Section 7.3 we present the methods we employ to perform our measurements of galaxy abundance and clustering. In Section 7.4 we lay out the HOD-based modelling of our observables with a parametrization of the SHMR as a starting point. In Section 7.5 we present our measurements of the observables and the results on the redshift evolution of the HOD and SHMR. In Section 7.6 we discuss the physical mechanisms that may regulate the growth of central and satellite galaxies in dark matter halos. We also compare our results with hydrodynamical simulations and discuss the possible origins of the discrepancies that we find. Finally, we summarize our finding in Section 7.7

Throughout this work we adopt a standard ΛCDM cosmology with $H_0 = 70\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$, $\Omega_{\mathrm{m},0} = 0.3$, where $\Omega_{\mathrm{b},0} = 0.04$, $\Omega_{\Lambda,0} = 0.7$, $\sigma_8 = 0.82$ and $n_s = 0.97$. Galaxy stellar masses, when derived from spectral energy distribution (SED) fitting, scale as the square of the luminosity distance (i.e., D_L^2), therefore as h^{-2}; dark matter halo masses, usually derived from dynamics in numerical simulations, scale as h^{-1}. The h scaling factors are retained implicitly for all relevant measurements, unless explicitly noted otherwise (see Croton, 2013, for an overview of h and best practices). When comparing to the literature, we rescale all the measurements to the cosmology adopted for this work. All magnitudes are expressed in the AB system (Oke and Gunn, 1983). Stellar masses are obtained assuming Chabrier (2003) initial mass functions (IMF) and when comparing to the literature, stellar masses are rescaled to match the IMF adopted in this work.

7.2 DATA

7.2.1 *COSMOS2020 galaxy catalog*

This thesis relies heavily on the COSMOS2020 catalog (Weaver et al., 2021). This deep multi-wavelength near-infrared selected catalog uses deep observations over the $2\,\mathrm{deg}^2$ COSMOS field in 35 photometric bands from ultraviolet (UV) to near-infrared (NIR). This unique combination of depth, area, and wavelength coverage allows an accurate estimation of photometry, photometric redshifts and stellar masses for around a million sources up to $z \sim 10$.

Broad-band optical data in the *grizy* broad bands comes from the Hyper Suprime-Cam (HSC) Subaru strategic program (HSC-SSP Aihara et al., 2018) DR2 (Aihara et al., 2019) that covers about $2.2\,\mathrm{deg}^2$. Data in the u-band were taken from Canada-France-Hawaii Telescope's (CFHT) Large-Area u-band Deep Survey (CLAUDS Sawicki et al., 2019) and comes in two images in the u and u^* bands.

Near-infrared $YJHK_s$ data comes from the Data Release 4 (DR4) of the UltraVISTA survey (McCracken et al., 2012) covering $1.5\,\mathrm{deg}^2$ of COSMOS. The UltraVISTA depth is not homogenous: four 'ultra-deep' stripes covering $0.6\,\mathrm{deg}^2$ have deeper exposure time. Here we analyze the full UltraVISTA area and make a magnitude cut on this depth where appropriate.

Mid-infrared data, that is essential for the estimation of stellar masses and the selection of stellar mass complete samples up to high redshifts, comes from *Spitzer*/IRAC in the [3.6] and [4.5] bands that have been processed by the Cosmic Dawn Survey (Moneti et al., 2021). As part of the Cosmic Dawn Survey all existing IRAC data in COSMOS field were re-processed and the final images reach a 3σ depth of 26.4 and 26.3 in [3.6] and [4.5] respectively. The rest of the 35 bands comprising the catalog come from data in 12 medium and two narrowband filters in optical from Subaru (Taniguchi et al., 2007; Taniguchi et al., 2015), and UV photometry from GALEX (Zamojski et al., 2007).

The unique combination of number of observing bands, depth and area allows to accurately extract photometry and estimate redshifts and stellar masses of about a million sources up to $z \sim 10$. For this purpose, the object detection is carried out on a 'chi-squared' combination (Szalay, Connolly and Szokoly, 1999) of deep near-IR images in $izYJHK_s$ (AB mag 3σ depth in 2 apertures of 27.6, 27.2, 25.3, 25.2, 24.9, 25.3). COSMOS2020 comprises two photometric catalogs. One catalog uses the traditional approach of measuring fluxes in fixed apertures using `SExtractor` (Bertin and Arnouts, 1996), whilst the second uses a profile-fitting technique using `The Farmer` (Weaver et al. in prep) built around `The Tractor`. Photometric redshifts (photo-*z*s), stellar masses and other physical parameters are estimated for these two photometric catalogs using two spectral energy distribution (SED) fitting codes `LePhare` (Arnouts et al., 2002; Ilbert et al., 2006) and `EAZY` (Brammer, Dokkum and Coppi, 2008).

In this analysis, we use the Classic catalog with photometric redshifts estimated using `LePhare`. This is principally because photometric measurements with `The Farmer`, as it is based on a model fitting technique, can fail to converge in certain cases, especially in crowded regions or near bright sources. This spatially variable completeness is problematic for measurements of the angular correlation function; we have measured relative correlation function differences of up to about 30% on scales of 1 between the two catalogs. The Classic catalog, on the other hand, contains photometric measurements for almost all sources within the survey masks. There is a caveat, however, since aperture photometry is not very reliable in crowder regions and around bright sources neither. So whereas Farmer is a pure catalog, since it photometers all the reliable sources, Classic is a more complete catalog, photometrying almost all sources.

Accurate measurements of photometric redshifts for all the galaxies in our sample is crucial in order to study the cosmic evolution of the SHMR. The photo-*z*s are obtained from `LePhare` in the configuration described in Weaver et al., which is the same as Laigle et al. (2016) and Ilbert et al. (2013). Briefly described, photo-*z*s are estimated using a template fitting technique, where a library of 33 galaxy SED templates covering various galaxy types are fitted to the measured fluxes of each galaxy. Dust attenuation is included as a free parameter, where $E(B-V)$ is allowed to vary from 0 to 0.5; two different attenuation curves (Calzetti et al., 2000; Prevot et al., 1984) are considered. Emission lines are added to the SEDs using an empirical relation between the intrinsic UV luminosity and [OII] line emission flux, as well as fixed ratios between dust-corrected emission lines Ilbert et al. (2009).

Compared to COSMOS2015, COSMOS2020 reaches similar photometric redshift precision almost one magnitude fainter – this is shown in Fig. 17 of Weaver et al. (2021), where the 1σ uncertainty of the photo-zs is plotted as a function of redshift and magnitude bin. The normalized median absolute deviation (σ_{NMAD}[1]) at $i < 22.5$ is below $0.01\,(1+z)$ and stays below $0.05\,(1+z)$ to $25 < i < 27$. The outlier fraction[2] is below 1% and 20% for the corresponding magnitude bins. The bias[3] ranges from -0.003 to -0.014 in the bright and faint magnitude bins respectively.

Accurate and complete stellar mass estimates over a wide redshift range are necessary to accurately measure the SHMR. In COSMOS2020, this is enabled by the inclusion of the deep near-IR data from the UltraVISTA survey (McCracken et al., 2012) DR4 in $YJHK_s$ and mid-IR data from the Cosmic Dawn Survey (Moneti et al., 2021) *Spitzer*/IRAC observations in channels 1-2 ($3.6\,\mu$m, $4.5\,\mu$ m). Stellar masses are estimated with `LePhare` using SED templates produced from stellar population synthesis models by Bruzual and Charlot (2003) and initial mass functions by Chabrier (2003). The SEDs are fixed at $z = z_{\mathrm{phot}}$ then and fitted to the multi-wavelength photometry (for more details, see Weaver et al. 2021 and Laigle et al. 2016). The point estimate of the stellar mass is the median of the resulting PDF marginalized over all other parameters, with the 16th and 84th percentiles of the PDF giving the 1σ confidence interval. The improved depth (e.g. $K_s = 25.3$, $[3.6] = 26.4$ at 3σ) translates to higher stellar mass completeness compared to the previous versions of the catalog. This enables a selection of samples based on stellar mass that are complete down to $\log M_*/M_\odot \sim 8.2$ at $z \sim 0.3$ and $\log M_*/M_\odot \sim 9.3$ at $z \sim 4$.

Throughout this work, we use 2 aperture magnitudes and apply aperture-to-total and Milky Way extinction corrections using the Schlafly and Finkbeiner (2011) dust map. We apply masks to remove sources near bright stars and in regions near artifacts. This leaves an effective area of $1.27\,\mathrm{deg}^2$ which corresponds to the footprint of the UltraVISTA survey[4].

Finally, we use two star-galaxy classifications to remove uncorrelated stars from the catalog. One uses morphological information from *HST*/ACS and Subaru/HSC images where half-light radii and magnitudes classify as stars all point-like sources at $i < 23$ and $i < 21.5$ in ACS and HSC images respectively. This criterion is also satisfied by point-like AGN sources. The second, SED-based criterion, classifies as stars sources with the χ^2 of the best-fit stellar template lower than the χ^2 of the best-fit galaxy template. We performed tests by measuring the correlation function of sources classified as stars, while further removing point-like AGN sources based on their χ^2. The correlation function of our stellar sample is zero, indicating a clean separation.

7.2.2 *Sample selection*

Measuring galaxy clustering and abundance requires complete stellar mass-selected and volume-limited samples. To select stellar mass-complete samples we use the stellar mass

---

[1] $\sigma_{\mathrm{NMAD}} = 1.48 \times \ \mathrm{median}[(|\Delta z - \mathrm{median}(\Delta z)|)/(1+z_{\mathrm{spec}})]$; $\Delta z = z_{\mathrm{phot}}$ - z_{spec}

[2] defined as $|\Delta z| > 0.15(1+z_{\mathrm{spec}})$

[3] defined as $\mathrm{median}(\Delta z)$

[4] In the catalog this is selected using the keyword `FLAG_COMBINED`

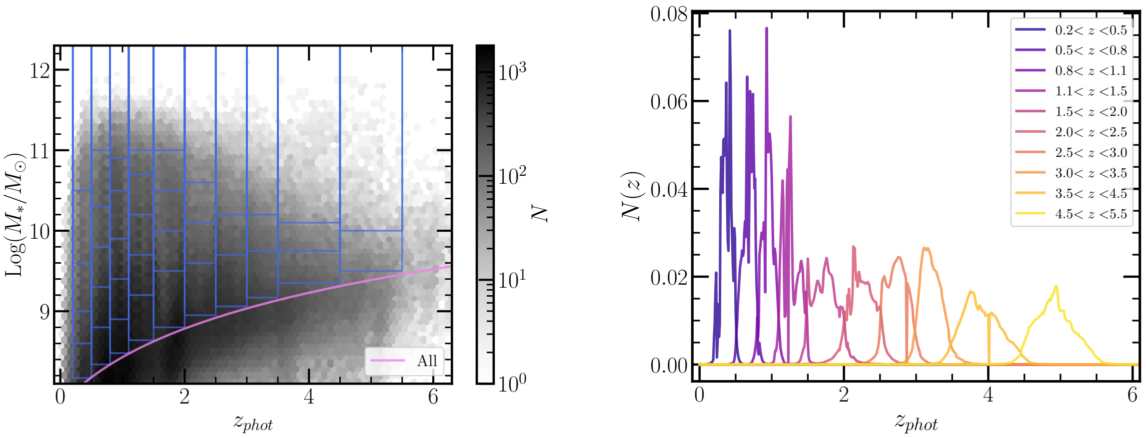


Figure 7.1: **Left:** Sample selection in the stellar mass-redshift plane. The solid grid lines show the mass threshold for each sample. The solid violet curve indicates the stellar mass completeness limit. We note that the histogram does not correspond to the final selection since further selection criteria (e.g. based on PDF(z) width) are applied. **Right:** Redshift distribution of the ten galaxy samples used for the clustering and abundance measurements. The redshift distribution is obtained by stacking the posterior photo-z distributions for all the sources in a given bin

completeness limit which is computed following the method prescribed by Pozzetti et al. (2010) and is described in catalog paper of Weaver et al. (2021). Briefly, a rescaled mass $M_{\rm resc}$ is computed as the mass that the galaxy would have if its apparent magnitude were equal to the limiting magnitude of the survey $m_{\rm lim}$; this is done using the rescaling relation $\log M_{\rm resc} = \log M + 0.4\,(m - m_{\rm lim})$. For each z-bin we take 90% of the galaxies with the lowest χ^2 of the fitted template, and we rescale them to their $\log M_{\rm resc}$. The 95th percentile of this distribution of rescaled masses is then the $M_{\rm lim}$ in the z-bin. This is repeated for several z-bins and fitted with a polynomial in order to interpolate the $M_{\rm lim}$ for any redshift. In COSMOS2020, limiting mass is computed using IRAC channel 1 magnitudes following the justification in Davidzon et al. (2017) and Weaver et al. This choice is motivated by the fact that channel 1 probes the bulk of the stellar emission in the galaxy SED at $z > 2.5$, where the Balmer break is redshifted to wavelengths larger than 2 μm. Therefore, a selection based on the deeper [3.6] magnitude is unbiased down to lower masses (Davidzon et al., 2017). To ensure complete samples, the channel 1 limiting magnitude $[3.6]_{\rm lim} = 26$ is computed with the help of the deeper CANDELS-COSMOS catalog (Nayyeri et al., 2017) which is used for completeness check. All samples throughout this work are selected to be brighter than $[3.6]_{\rm lim} = 26$ across the full redshift range, despite the fact that a K_s based selection is also suitable at low redshifts.

To probe the cosmic evolution of the observables we bin the samples in 10 redshift bins from $z = 0$ to $z = 5.5$ with varying width. The widths were chosen to ensure roughly the same number of galaxies in each bin. The redshift bins are listed in Table 7.1. Additionally, we require that each galaxy has its lower and upper 1σ values ($z_{\rm low}$, $z_{\rm up}$) within ±0.5 of the z-bin limits. This criterion removes any galaxies with highly uncertain redshifts that can introduce errors. To ensure that the samples remain as complete as possible, we don't

Table 7.1: Sample selection in redshift and stellar mass thresholds. The columns indicate the redshift bin, mass threshold, median mass and number of galaxies in for each sample used to measure clustering.

z-bin	$\log(M_*^{\text{threshold}}/M_\odot)$	Median $\log(M_*/M_\odot)$	N
$0.2 < z < 0.5$	8.17	8.86	23 346
	8.60	9.25	15 234
	9.00	9.64	10 000
	9.60	10.16	5 229
	10.50	10.76	1 495
$0.5 < z < 0.8$	8.34	8.99	43 752
	8.80	9.40	27 104
	9.30	9.90	15 325
	9.80	10.32	8 588
	10.30	10.65	4 454
	11.00	11.15	736
$0.8 < z < 1.1$	8.48	9.18	50 964
	8.90	9.55	34 306
	9.40	10.02	20 211
	9.90	10.43	11 683
	10.50	10.80	5 167
	10.90	11.07	1 864
$1.1 < z < 1.5$	8.64	9.33	53 285
	9.20	9.78	30 882
	9.70	10.22	17 095
	10.20	10.58	8 823
	10.70	10.91	3 239
	11.00	11.15	1 139
$1.5 < z < 2.0$	8.80	9.41	47 100
	9.50	10.03	20 941
	10.00	10.43	10 910
	10.50	10.77	4 745
	11.00	11.14	936
$2.0 < z < 2.5$	8.95	9.45	31 247
	9.60	10.01	12 205
	10.10	10.47	5 209
	10.60	10.82	1 869
$2.5 < z < 3.0$	9.06	9.54	25 660
	9.70	10.04	9 625
	10.20	10.48	3 475
$3.0 < z < 3.5$	9.17	9.62	15 018
	9.75	10.06	5 739
	10.20	10.43	2 059
$3.5 < z < 4.5$	9.35	9.70	12 249
	9.75	10.02	5 569
	10.10	10.35	2 247
$4.5 < z < 5.5$	9.50	9.79	3 374
	10.00	10.24	932

impose any other selection criteria, for example, based on S/N, number of bands in which a source was photometered, χ^2 of the SED fit etc.

One of the ingredients of the model of the galaxy correlation function is the redshift distribution $N(z)$. We use the z-likelihood from `LePhare` to build $N(z)$. Formally, for each galaxy, there is a likelihood of the observed photometry (denoted by the vector of fluxes o) given the redshift $\mathcal{L}(o|z)$. $N(z)$ is then constructed by simply stacking the individual $\mathcal{L}(o|z)$ in each z-bin

$$N(z) = \sum_{i}^{N_{\text{sample}}} \mathcal{L}_i(o|z), \tag{7.1}$$

where the sum is over the number of objects of the redshift and stellar mass threshold-selected sample N_{sample}.

Constructing the $N(z)$ in this way directly accounts for the uncertainty in the photometric redshift when selecting galaxies in a bin – the fact that galaxies can still have their true z outside the bin limits. However, it has been shown by Ilbert et al. (2021) that such a procedure can lead to biases in the mean redshift that can be inferred from the $N(z)$. They have quantified that these biases can reach up to $\sim 0.01\,(1+z)$. By using the model of $w(\theta)$ (described in Section 7.4.2) we have tested the effects of bias in the mean redshift, as well as different estimates of $N(z)$. Our conclusion is that biases of order of magnitude $\sim 0.01\,(1+z)$ result in relative difference in $w(\theta)$ of less than about 3%, which is considerably smaller than the typical relative error of the measurement (about 10%). On the other hand, the shape of $N(z)$ (notably, the width of its wings) can lead to significant difference in $w(\theta)$ that can bias the inferred SHMR parameters. The reason for this is the mix of physical scales when considering larger volumes: the angular correlation of galaxies selected in a wider radial interval is inevitably lower, since they can be far apart in the radial direction but close in angular separation. For example, considering $N(z)$ to be a Gaussian distribution centered at the z-bin mean and with width half of the z-bin width, can lead to a relative difference of about 20% at $z \sim 0.4$ and more than 50% at $z \sim 2.5$ (see Appendix A.1). Ilbert et al. (2021) have shown that the $\mathcal{L}_i(o|z)$ as output from `LePhare` can lead to biased and miss-calibrated $N(z)$ as evidenced by comparison with the true redshift histogram and the Probability Integral Transform (PIT) statistic (see Fig. 4 in Ilbert et al., 2021). The authors show that an $N(z)$ that is better representative of the true distribution can be obtained by using a posterior distribution, such as

$$N(z) = \sum_{i}^{N_{\text{sample}}} \mathcal{P}_i(o|z) = \sum_{i}^{N_{\text{sample}}} \mathcal{L}_i(o|z)\,\Pr(z|m_0), \tag{7.2}$$

where $\Pr(z|m_0)$ is the so called 'photo-z prior' (Brodwin et al., 2006). This prior can be constructed for every z-bin by summing the likelihoods per magnitude bins such as

$$\Pr(z|m_0) = \sum_{i}^{N_{z-\text{bin}}} \mathcal{L}_i(o|z)\,\Theta(m_{0,i}|m_0), \tag{7.3}$$

where $\Theta(m_{0,i}|m_0)$ is equal to 1 if the object's magnitude $m_{0,i}$ is within the magnitude bin centred at m_0 and zero otherwise. The outcome of this procedure and the effects on $w(\theta)$ are presented in more detail in Appendix A.1. We adopt the $N(z)$ obtained using Eq. 7.2 for our analyses. The $N(z)$ is constructed for each considered sample, including the mass threshold-selected samples. The resulting distributions for all z-bins for galaxies above the mass-completeness limit are shown in Fig. 7.1. We note that there are some narrow (of width ~ 0.01) dips at several z-values (e.g., $z \sim 1.3,\ 2.9,\ 4.0$). These come from the individual likelihoods being close to zero at these exact values. The template fitting outputs minimal likelihoods at exactly these 3 z-points, and being ~ 0.01 narrow they do not affect the selection nor the modeling.

To probe the clustering strength as a function of galaxy mass, we further select samples in stellar mass thresholds. In each z-bin we define samples selected in several stellar mass thresholds starting from the mass completeness limit; these are indicated by the horizontal solid lines in Fig. 7.1 and listed in Table 7.1. We chose the thresholds rather arbitrarily to ensure a good signal-to-noise (S/N) for the clustering measurement of each mass threshold-selected sample within a z-bin.

7.3 MEASUREMENTS OF GALAXY CLUSTERING AND ABUNDANCE

For the purpose of inferring the stellar-to-halo mass relation up to $z \sim 5$ we use measurements of galaxy clustering and abundance. As we described in detail in the previous chapters, these two are directly related to the properties of the dark matter halo through the halo occupation distribution (HOD): the halo model formalism provides a way to relate measurements of galaxy clustering and abundance to the properties of the dark matter halos. In this section we present the methods used to perform the measurements, who themselves are presented in Section 7.5.2 along with the best-fit models.

7.3.1 *Galaxy clustering*

We measure the two-point angular correlation function $w(\theta)$ using the Landy and Szalay (1993) estimator[5]

$$w(\theta) = \frac{DD - 2\,DR + RR}{RR}, \tag{7.4}$$

where DD is the number of data-data pairs in a given angular separation bin $[\theta, \theta + \delta\theta]$, RR is the number of pairs in the random catalog in the same bin, and DR is the number of pairs between the data and the random catalog. The data and random pairs are normalized by the total number of galaxies and random objects. We construct the random catalog with the same survey geometry mask as the data catalog and $N_{\mathrm{random}} \sim 3 \times 10^6$ which is more than 50 times the number of galaxies in each considered bin.

---

[5] The correlation functions are computed using the `TreeCorr` code (Jarvis, Bernstein and Jain, 2004)

The covariance matrix is computed using the jackknife method by subdividing the full area in $N_{\mathrm{patch}} = 22$ patches and recomputing the correlation function, removing one patch at a time. The covariance matrix is then estimated as

$$C_{lk} = \frac{N_{\mathrm{patch}} - 1}{N_{\mathrm{patch}}} \sum_{i=1}^{N_{\mathrm{patch}}} \left(w_i(\theta_l) - \bar{w}\right)^T \left(w_i(\theta_k) - \bar{w}\right), \tag{7.5}$$

where $\bar{w}$ is the mean correlation function and w_i the correlation function with the i-th patch removed. The final covariance matrix thus includes uncertainties due to sample (cosmic) variance, dominating at larger scales, and due to Poisson statistics from counting objects in bins, dominating mostly at small scales. However, due to the limited survey size, this method still cannot accurately estimate the uncertainties due to cosmic variance. To capture the cosmic variance effects in the covariance matrix, one can use computationally expensive simulations, which is out of the scope of this work. This, for example, is done in Leauthaud et al. (2011), where they show that cosmic variance has an impact on the covariance matrix on large scales (i.e. in the two-halo regime) – cosmic variance increases the correlation in the data at large scales.

Due to the limited size of the survey ($1.27\,\mathrm{deg}^2$), the clustering measurement suffer from the effects of the integral constraint (IC, Groth and Peebles, 1977). This leads to an underestimation of $w(\theta)$ at large scales comparable to angular size of the survey by a constant factor w_{IC}, such that the true correlation function is

$$w(\theta) = w_{\mathrm{meas}}(\theta) + w_{\mathrm{IC}}. \tag{7.6}$$

We incorporate this correction into our model, described later.

7.3.2 *Galaxy abundance*

We measure stellar mass functions in ten redshift bins using the $1/V_{\mathrm{max}}$ technique (Schmidt, 1968). This estimator weighs each galaxy by the maximum volume in which it would be observed given the redshift range of the sample and magnitude limit of the survey. For the i-th galaxy the $V_{\mathrm{max},i}$ is computed as

$$V_{\mathrm{max},i} = \frac{4\,\pi}{3} \frac{\Omega_{\mathrm{survey}}}{\Omega_{\mathrm{sky}}} \left(d_c(z_{\mathrm{max},i})^3 - d_c(z_{\mathrm{min},i})^3\right), \tag{7.7}$$

where $\Omega_{\mathrm{survey}} = 1.27\,\mathrm{deg}^2$ and $\Omega_{\mathrm{sky}} = 41\,253\,\mathrm{deg}^2$ are the surface area of the survey and the full sky respectively, $d_c(z)$ is the comoving distance at redshift z, and The comoving volume V_{max} is computed between z_{min} and z_{max}, where z_{min} is the lower redshift limit of the z-bin and $z_{\mathrm{max}} = \min(z_{\mathrm{bin,up}}, z_{\mathrm{lim}})$, where $z_{\mathrm{bin,up}}$ is the upper redshift limit of z-bin and z_{lim} is the maximum redshift up to which a galaxy of a given magnitude can be observed given the magnitude limit of the survey. For this purpose, we use IRAC channel-1 magnitudes and the limit of $[3.6] = 26$. Finally, the SMF is computed as

$$\Phi(M)\,\Delta\log M = \sum_{i}^{N(M)} \frac{1}{V_{\max,i}}. \tag{7.8}$$

We compute the SMFs in the mass range starting from the mass completeness limit of each z-bin, with a bin size of $\Delta\log M = 0.25$.

Uncertainties in the SMF include contributions from Poisson noise (σ_{Pois}), cosmic variance (σ_{cv}), and SED fitting uncertainties (σ_{fit}). We compute the uncertainties due to cosmic variance following Steinhardt, Jespersen and Linzer (2021) The starting point is the cosmic variance "cookbook" code of Moster et al. (2011), that computes the stellar-mass dependent cosmic variance and performs well up to intermediate redshifts and masses but becomes increasingly underestimated at high redshift and mass. Steinhardt, Jespersen and Linzer (2021) extend the recipes to the early universe ($z > 3$) by using linear perturbation theory.

Additionally, due to photometric errors and degeneracies in the SED fits, there are uncertainties in the M_* measurements that propagate to the SMF. To estimate SED fitting uncertainties on the SMF we use the PDF(M_*). We assign to each galaxy a weight that corresponds to its probability to be found in the given mass bin

$$w_i = \frac{1}{\int_{\mathrm{bin}} \mathrm{d}M_*\ \mathrm{PDF}_i(M_*)}. \tag{7.9}$$

We then use this weight to compute the combined Poisson and SED fitting uncertainties in the following way

$$\sigma_{\mathrm{Pois+fit}} = \sqrt{\sum_{i}^{N_{\mathrm{g}}} \frac{w_i^2}{V_{\max,i}^2}}. \tag{7.10}$$

The final uncertainty is then $\sigma_\Phi^2 = \sigma_{\mathrm{Pois+fit}}^2 + \sigma_{\mathrm{cv}}^2$.

The measurements for the clustering and SMF are presented in Fig. 7.4 and discussed in Section 7.5.2

7.4 MODELING

7.4.1 *Model of the galaxy stellar mass function*

From the defined occupation distribution of halos (Eq. 4.6 & Eq. 4.7) one can obtain the number density of galaxies in a given mass bin by integrating over the halo mass function (HMF) $\mathrm{d}n/\mathrm{d}M_h$

$$\Phi\left(M_*^{\mathrm{th1}}, M_*^{\mathrm{th2}}\right) = \int_0^\infty \mathrm{d}M_h \left\langle N_{\mathrm{tot}}\left(M_h | M_*^{\mathrm{th1}}, M_*^{\mathrm{th2}}\right)\right\rangle \frac{\mathrm{d}n}{\mathrm{d}\log M_h}. \tag{7.11}$$

Table 7.2: Adopted ingredients in the halo model

Ingredient	Assumption
HMF	Despali et al. (2016)
Halo bias $b_h(M_h)$	Tinker et al. (2010)
Halo mass-concentration relation $c(M_h)$	Duffy et al. (2008)
Halo and satellite over-density profiles	NFW profile, Navarro, Frenk and White (1997)
Halo mass definition	Virial

This allows also compute the GSMF of centrals and satellites by using their respective occupation distributions in Eq. 7.11. The literature abounds with prescriptions of the HMF obtained under various assumptions and methods, and for our work we use the HMF of Despali et al. (2016). The adoption of different HMFs has an effect on the modelled GSMF and inevitably on the inferred model parameters. The HMF also depends on the choice of halo mass definition, and since the models are computed using the HMF, the final results will also depend on these definitions. For example, a different halo mass definition would result in a systematic shift in halo masses. To compute the HMF we use the code `COLOSSUS` (Diemer, 2018), and we use the virial overdensity (Bryan and Norman, 1998) halo mass definition for the results we show in this chapter.

The model for the SMF shows a high sensitivity to the parameters describing the central SHMR, and coupled with the high signal-to-noise of the measurements has the most constraining power.

7.4.2 *Model of the 2-pt angular correlation function*

The model of the 2-pt angular correlation function as describe in Chapter 3, Eqn. 3.16. One of the main ingredients is the galaxy power spectrum that is written in terms of the HOD (Eqn. 3.13). We use the HOD a specified in Eqn. 4.6 and Eqn. 4.6 to compute the galaxy power spectrum and consequently the angular correlation function. For the computation of the 2-pt angular correlation function, we rely on LSST Dark Energy Science Collaboration's Core Cosmology Library (`CCL`)[6]. CCL is a library of routines to calculate a range of cosmological observables and is still under active development. The validation of the software along with a range of benchmark tests are presented in Chisari et al. (2019). The main ingredients that enter the modeling along with the prescriptions and assumptions that we adopt are given in Table 7.2. Finally, in Fig. 7.2 we give a schematic representation of the modeling of the observables.

Due to the relatively small volume probed by the COSMOS survey, the integral constraint affects $w(\theta)$ at large scales. We adjust the model to take this into account. The

---

[6] `https://github.com/LSSTDESC/CCL`

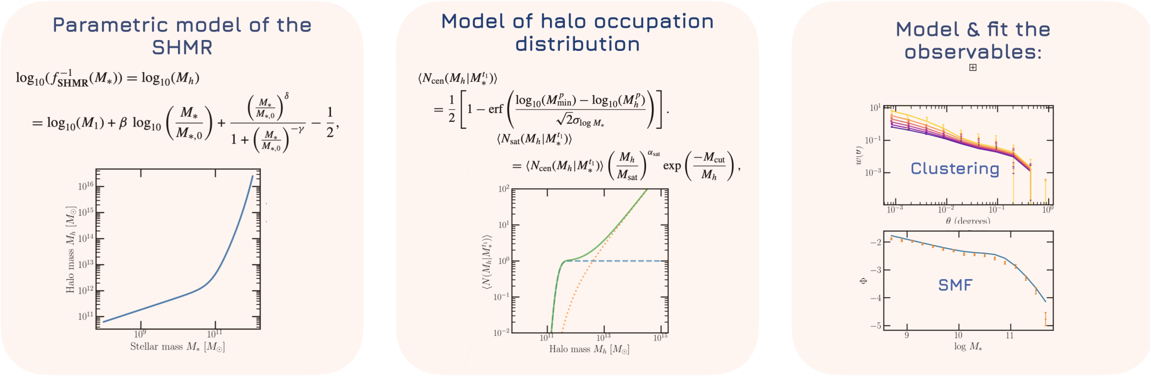


Figure 7.2: Schema of the HOD based modeling of the observables clustering and GSMF, with the parametric model of the SHMR as a starting point.

correction factor due to the IC can be estimated from the double integration of the true correlation function over the survey area,

$$w_{\rm IC} = \frac{1}{\Omega^2}\int w_{\rm true}\,{\rm d}\Omega_1{\rm d}\Omega_2. \tag{7.12}$$

This integration can be carried out using the random-random pairs from the random catalog following Roche and Eales (1999)

$$w_{\rm IC} = \frac{\sum w_{\rm true}(\theta)\,RR(\theta)}{\sum RR(\theta)}, \tag{7.13}$$

where $w_{\rm true}(\theta)$ is HOD-predicted model. Finally, the model that we fit against the data is simply $w(\theta) = w_{\rm true}(\theta) - w_{\rm IC}$.

7.5 RESULTS AND ANALYSIS

7.5.1 *Fitting procedure*

We fit the models of the $w(\theta)$ and the SMF to our measurements using a Markov Chain Monte Carlo (MCMC) approach, minimising χ^2:

$$\begin{aligned}\chi^2 = &\sum_i^{N_{M-\rm thres}} (w_i - \tilde{w}_i)^T C^{-1} (w_i - \tilde{w}_i) + \\ &\sum_i^{N_{M-\rm bins}} \left(\frac{\Phi(M_{*,i}) - \tilde{\Phi}(M_{*,i})}{\sigma_\Phi}\right)^2,\end{aligned} \tag{7.14}$$

where w are the measurements vector containing w at θ for all mass thresholds and $\tilde{w}$ and $\tilde{\Phi}$ are the models for a given set of parameter values. The first line of Eq. 7.14 corresponds to the clustering likelihood and the second line to GSMF likelihood. We use the affine-invariant ensemble sampler implemented in the `emcee` code (Foreman-Mackey et al., 2013). We use 200 walkers for our 11 parameters and rely on the auto-correlation time τ to assess the convergence of the chain. To consider the chains converged, we require that the auto-correlation time is at least 60 times the length of the chain and that the change in τ is less than 5%. We discard the first $2 \times \max(\tau)$ points of the chain as the burn-in phase and thin the resulting chain by $0.5 \times \min(\tau)$. We impose flat priors on all parameters; for the mass parameters, the flat priors are on the log quantities.

For best-fit parameters values, we take the medians of the resulting posterior distribution, with the 16-th and 84-th percentiles giving the lower and upper uncertainty estimates. The best-fit parameters and their uncertainties for all the 10 z-bin are listed in the Appendix A.3 The posterior distributions for the 11 parameters in the redshift bins are shown in Appendix A.2.

7.5.2 *Measurements and best-fit models*

The measurements of the $w(\theta)$ and GSMFs are shown in Fig. 7.3 – where we isolate the measurement in $0.5 < z < 0.8$ and compare with clustering measurements from the literature; and Fig. 7.4 – where we show the measurements in all the other z-bins. In the upper panel we show the clustering measurements (open circles with error bars) for different mass-threshold samples and in the bottom panel the SMF measurement. The solid lines show the best-fit models with the color code corresponding to the mass-threshold measurement (in the top panel). Table A.3 also shows the reduced chi-squared value χ^2_{reduced} for the-best fit parameters. Their values range from 2.5-6 for most bins, except for $0.8 < z < 1.1$ where $\chi^2_{\text{reduced}} = 10.2$. Given the number of data points, that ranges from 75 at $0.8 < z < 1.1$ to 30 at $4.5 < z < 5.5$, these values of χ^2_{reduced} indicate a reasonably well fit. A possible explanation can be the larger complexity of the data that is not completely captured by the fits. For example, we are simultaneously fitting for several mass-selected clustering measurements. Due to uncertainties in the stellar masses, there can exist some super-covariance between all the measurements, whose modeling is out of the scope of this work.

Description of the clustering measurements

The clustering measurements exhibit the usual behavior with $w(\theta)$ following a power-law at small scales that breaks at intermediate scales (~ 1). The origin of this break comes from the fact that the power at small scales is dominated by galaxies residing in the same halo (1-halo term) that drops off quickly at intermediate scales; at larger scales the power mainly comes from large-scale clustering of halos (2-halo term) and the transition between these two regime creates the characteristic shape (Zehavi et al., 2004). The relative contribution from these two terms becomes more apparent at high masses and high redshifts, with the 1-halo term dominating the power at small scales with a steep slope and the 2-halo term dominating the large scales with a shallower slope at the transition. The clustering amplitude increases with increasing mass threshold – a familiar behavior as massive galaxies trace high-density and more clustered regions.

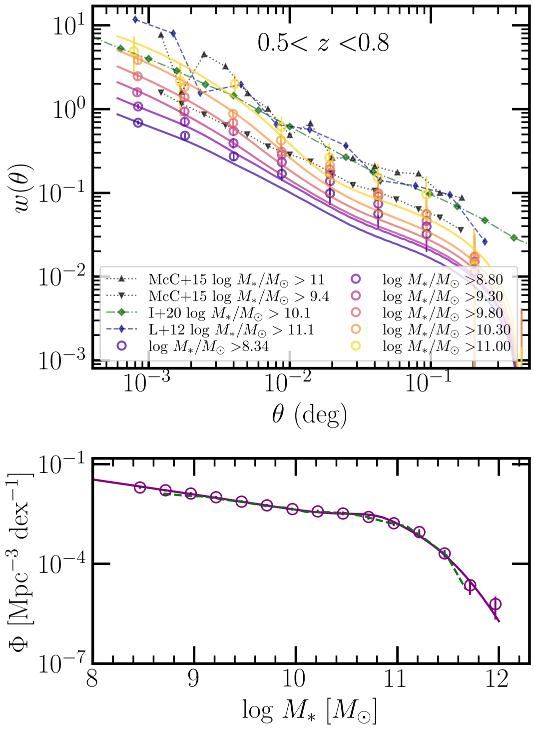


Figure 7.3: Best-fit models for clustering and abundance compared with measurements at $0.5 < z < 0.8$. *Top*: Clustering measurements for the 6 mass threshold-selected samples (empty circles with error bars) along with the best-fit models in solid lines in corresponding colors. *Bottom*: Measurements of the stellar mass function together with the best fit model. The dashed lines show the SMF in the same redshift bin obtained by Davidzon et al. (2017) for comparison.

At scales larger than 0.1 deg, there is a sharp drop in power due to the effects of the integral constraint. The best-fit models shown in solid lines generally agree well with the measurements. It should be noted that the measurements show an excess of power at scales of $\gtrsim 0.02$ deg (~ 0.5 Mpc at $z \sim 7$), and the fits are systematically below the data points. This can be due to the effects of the non-linear halo bias effect (scale dependent halo bias). While in this work we use the scale-independent halo bias of Tinker et al. (2010), some studies suggest that the halo bias is scale dependent in the quasi-linear regime at scales of about 1 Mpc (Jose et al., 2017). This will add power in the correlation at scales of ~ 0.04 deg. Furthermore, at $0.5 < z < 1.5$, a contribution can also come from the known overabundance of rich structures in the COSMOS field at these redshifts, as discussed by McCracken et al. (2007), (Meneux et al., 2009), and McCracken et al. (2015). This excess of power will decrease the SHMR, indicating an even lower efficiency of converting baryons to stars. Clustering is particularly sensitive to the satellite content within the halo, therefore the parameters regulating the satellite HOD will be constrained by clustering.

Literature comparison

In Fig. 7.3 we show COSMOS clustering measurements at $0.5 < z < 0.8$ from McCracken et al. (2015) and Leauthaud et al. (2012), and in HSC-SSP Wide survey from Ishikawa

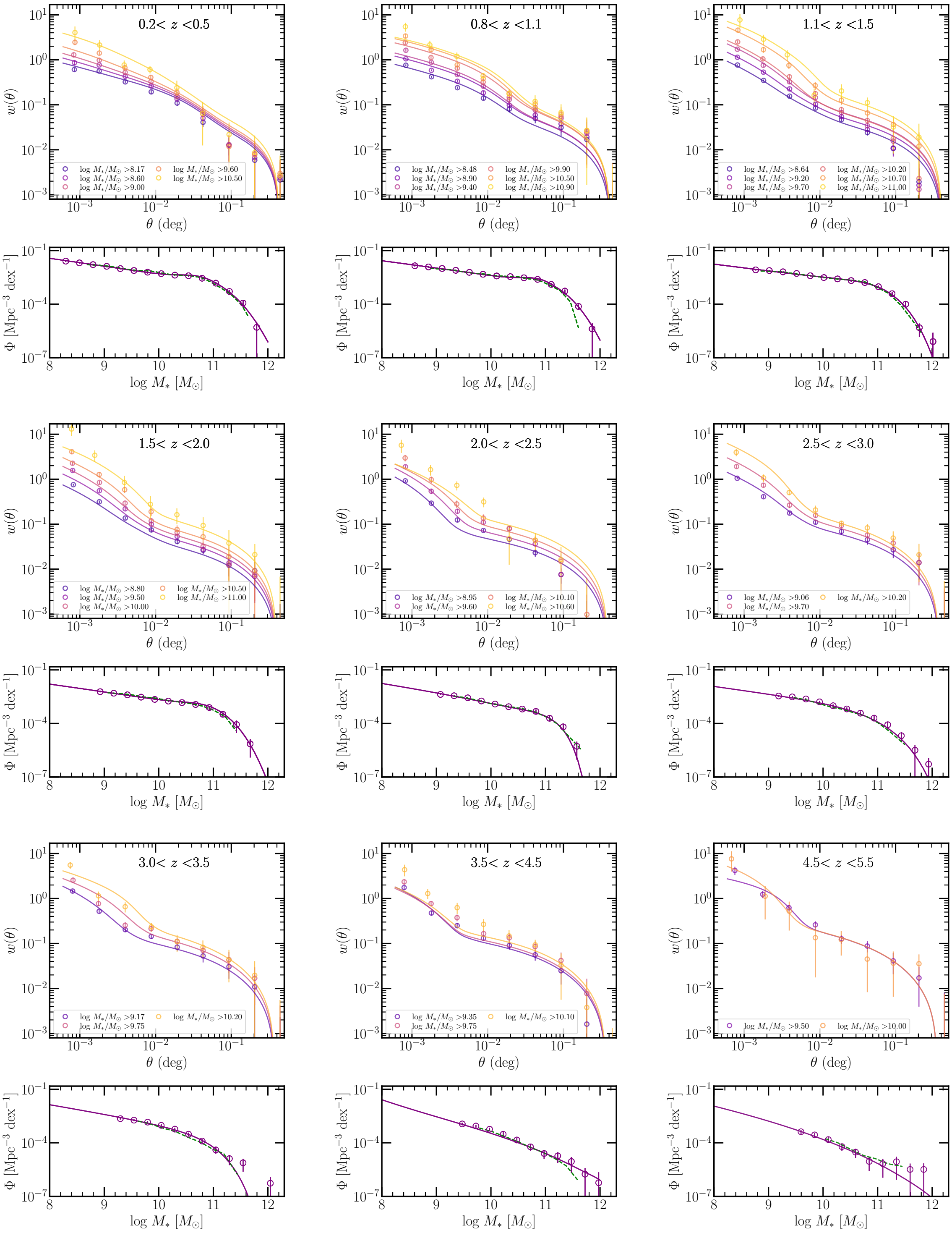


Figure 7.4: Best-fit models of clustering and abundance plotted over the measurements for all z-bins apart from $0.5 < z < 0.8$, which is shown in Fig. 7.3.

et al. (2020). McCracken et al. (2015) uses the 1.5 deg^2 COSMOS footprint of UltraVISTA DR1 (McCracken et al., 2012) to measure clustering for mass-threshold selected samples. We also show the correlation function for galaxies with $\log M_*/M_\odot > 9.4$ and $\log M_*/M_\odot > 11.0$ in black wedges and triangles. Qualitatively, measurements from the literature are in agreement with our work, although the 2PCF for the $\log M_*/M_\odot > 11.0$ sample has a slightly higher amplitude, especially at small scales. Leauthaud et al. (2012) uses the 1.64 deg^2 of COSMOS as imaged by HST/ACS in F814W (Koekemoer et al., 2007) to measure the 2PCF for mass-threshold samples at $0.48 < z < 0.74$. The measurement for $\log M_*/M_\odot > 11.1$ is shown in grey hexagons in Fig. 7.3, which are consistent with our measurements and those of McCracken et al. (2015). Finally, in green diamonds, we show the measurements from Ishikawa et al. (2020) in 145 deg^2 in the HSC-SSP Wide for a sample of $\log M_*/M_\odot > 10.1$ galaxies at $0.55 < z < 0.80$. The amplitude of Ishikawa et al. (2020) 2PCF corresponds to what we measure in this work between $\log M_*/M_\odot > 10.3$ and $\log M_*/M_\odot > 11.0$. This could come from incompleteness in their samples, or/and uncertain stellar masses that were estimated with optical (*grizy*) bands only.

Redshift evolution of the clustering

With respect to redshift, to show a possible evolution of the clustering amplitude, in Fig. 7.5 we recompute the correlation function for galaxies selected above the same mass threshold in all z-bins: $M_* > 10^{10}\,M_\odot$. Although the clustering amplitude of dark matter decreases with increasing redshift, the evolution of the clustering amplitude for galaxy samples selected at the same mass-threshold depends on the galaxy formation model. The clustering of galaxies depends on how galaxies occupy DM halos, which can change with redshift. N-body simulations combined with semi-analytical models of galaxy formation indicate that the clustering amplitude of similarly selected galaxies decreases from $z = 0$ to $z = 1.5$, remains constant up to $z = 2.5$, and then increases again at higher redshifts (Kauffmann et al., 1999). Qualitatively, this behavior can be observed in our measurements in Fig. 7.5: the correlation amplitude is the highest in the lowest redshift bins, reaches the lowest amplitude for intermediate z-bins of about $z \sim 1.5$ and then increases again at $z > 2.0$.

Redshift evolution of the SMF

The SMF measurements in Fig. 7.5 also show the usual evolution with redshift (see e.g. Ilbert et al., 2013; Davidzon et al., 2017): the normalization decreases and the knee at $M_* \sim 10^{11}\,M_\odot$ becomes less and less prominent with increasing redshift; the slope of the low-mass end remains constant up to $z = 2$ but steepens at higher redshifts where the SMFs resemble more a power-law (e.g. in the $z > 4.5$ bin); the redshift evolution is strongly dependent on mass: the low-mass end evolves more rapidly than the high-mass end. The SMFs, having the most constraining power over the model parameters (due to the small measurement errors and sensitivity of the model), show an excellent fit of the models to the measurements. The dashed lines in Fig. 7.4 show the SMFs measured by Davidzon et al. (2017) using the previous version of the catalog, COSMOS2015. On overall, they are in agreement with our measurements over the whole redshift range.

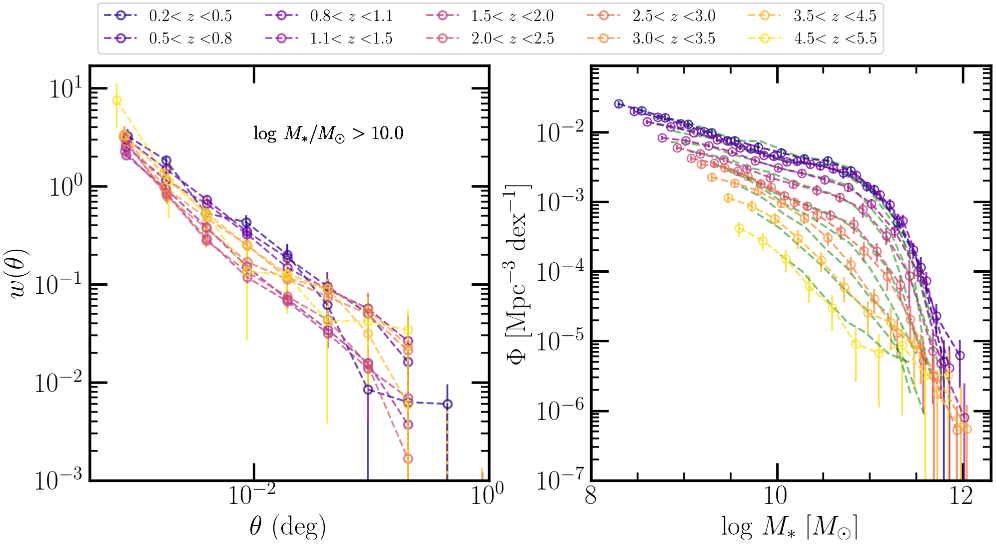


Figure 7.5: Correlation of galaxies with $M_* > 10^{10}\,M_\odot$ (left panel) and GSMF (right panel) for all ten redshift bins. The green dashed lines in the right panel correspond to the GSMF of Davidzon et al. (2017).

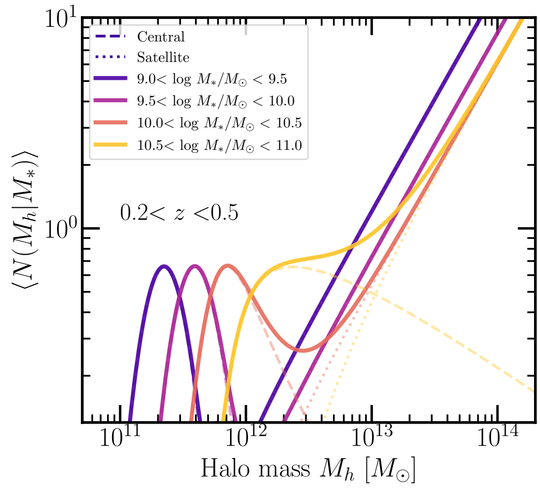


Figure 7.6: Mean number of galaxies with stellar masses in a given mass bin as a function of the mass of the halo that they occupy. We show the mean halo occupation function for galaxies in 4 stellar mass bins (color coded accordingly) at $0.2 < z < 0.5$. The thick solid lines show the total $\langle N_{\rm tot}\rangle$ and the dashed and dotted lines show the centrals and satellites.

7.5.3 *Evolution of the mean halo occupation with redshift*

The mean halo occupations, as defined by Eq. 4.6 and Eq. 4.7, are shown in Fig. 7.6 for $0.2 < z < 0.5$. We show the mean number of galaxies in four mass bins: $\log M_*/M_\odot = \{[9.0, 9.5], [9.5, 10.0], [10.0, 10.5], [10.5, 11.0]\}$, as a function of halo mass for all galaxies in the thick solid lines, and for satellites and centrals in dotted and dashed lines, respectively. It is immediately evident that the mean halo occupation shifts toward high halo masses for more massive galaxies, as it requires more massive halos to host more massive galaxies. Furthermore, the central occupation peaks at some characteristic mass. Halos that have

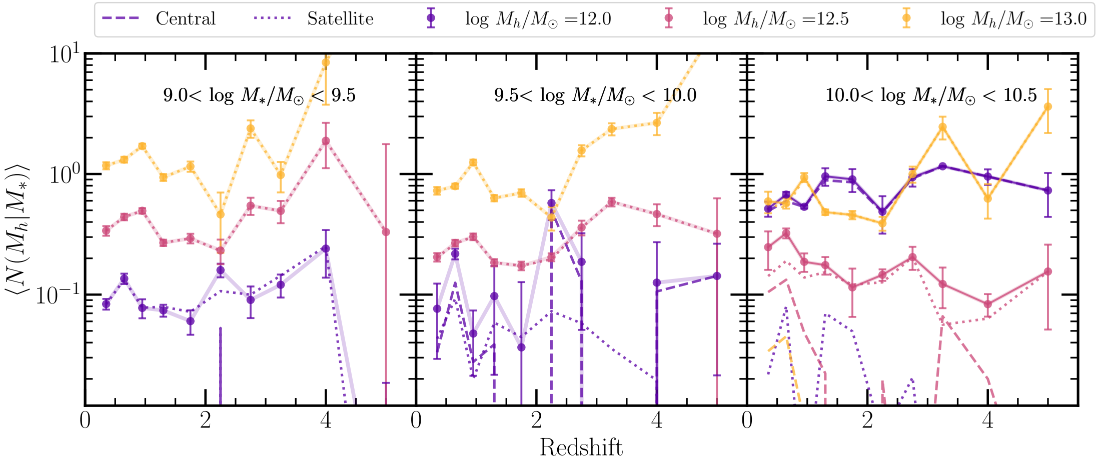


Figure 7.7: Mean halo occupation in halos of a given mass as a function of redshift. In each panel, we show the $\langle N_{\rm tot}(M_h)\rangle$ at $\log M_h/M_\odot = [12.0, 12.5. 13.0]$. The three different panels show the mean occupations by galaxies in three different stellar mass bins $\log M_*/M_\odot = \{[9.0, 9.5], [9.5, 10.0], [10.0, 10.5]\}$. The dashed and dotted lines show the central and satellite occupations, while the points connected with a transparent solid line show the total.

this characteristic mass can be considered as most likely to host a central galaxy in a given stellar mass bin.

As the halo mass increases, the number of satellites starts to increase sharply. The mean occupation for low-mass galaxies shows that there can be halos of intermediate mass that do not host any low-mass galaxies. For example, halos of $M_h \sim 10^{12}\,M_\odot$ have a very low probability of hosting of $10^{9.5} < M_*/M_\odot < 10^{10}$ galaxies. The central and satellite decomposition (dashed and dotted lines) show that this is because galaxies in this mass bin cannot be centrals in $M_h \sim 10^{12}\,M_\odot$ halos and can only be satellites in even more massive halos. We also note that as their stellar mass increases, central galaxies are more likely to occupy halos with a larger variety of masses (looking at the dashed line, for higher mass bins there is shallowing of the slope at which the central occupation decreases with halo mass). This behavior can come from a quenching of massive galaxies – as their stellar mass growth stops, the halo they inhabit continues to grow in mass. Finally, in clusters ($M_h > 10^{13}\,M_\odot$), low-mass satellites dominate the number of galaxies in the halo. This can also be seen as a consequence of quenching: satellites stop their growth because of quenching in the halo and remain less massive, while the halo can grow by merging with other halos containing more satellites of low masses.

To investigate the redshift evolution, in Fig. 7.7 we show the mean occupation distribution for galaxies with $9.0 < \log M_*/M_\odot < 9.5$ (left panel), $9.5 < \log M_*/M_\odot < 10$ (middle panel) and $10.0 < \log M_*/M_\odot < 10.5$ (right panel) as a function of redshift at three different halo masses $\log M_h/M_\odot = [12.0, 12.5, 13.0]$. Dashed and dotted lines show the central and satellite mean halo occupations, while the points connected with a transparent solid line show the total. The panels show that the total $\langle N(M_h/M_*)\rangle$ of $M_h \leq 10^{12}\,M_\odot$ halos is dominated by centrals at all redshifts, whereas satellite dominate at higher halo masses at all redshifts. An exception are galaxies with $9.0 < \log M_*/M_\odot < 9.5$ (left panel) which are found as satellites in halos of $M_h \geq 10^{12}, M_\odot$ and at all redshifts. In each panel and for

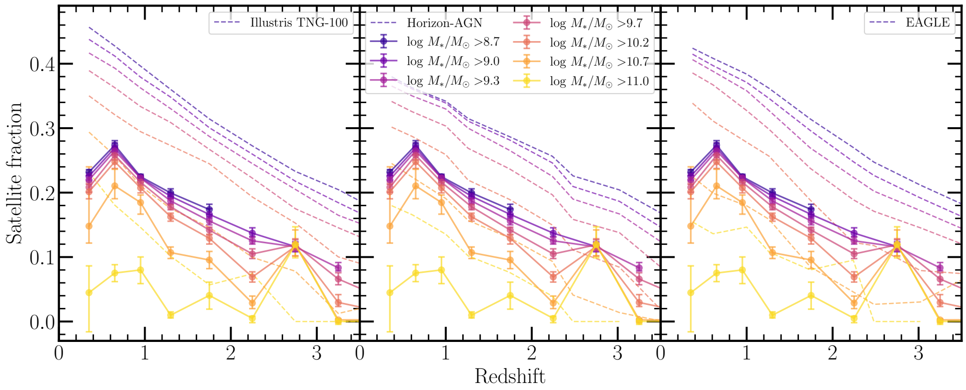


Figure 7.8: Fraction of satellite galaxies with masses above a given threshold as a function of redshift (solid lines). In each of the three panels, we compare with the satellite fractions measured in the hydrodynamical simulations (dashed lines) TNG100 (left), HORIZON-AGN (center), and EAGLE (right). The redshift evolution of the satellite fraction is shown for galaxies with masses above $\log M_* / M_\odot > [8.7, 9.0, 9.3, 9.7, 10.2, 10.7, 11.0]$.

every halo mass, we detect little-to-no evolution of the mean occupation number, in accordance with previous findings based on N-body simulations (e.g., Kravtsov et al., 2004). At $z > 2.5$ there are variations toward higher mean halo occupation number but with overly large uncertainties to be significant. These results indicate that (statistically) in terms of mean occupation numbers, galaxies populate DM halos in the same way throughout cosmic time.

7.5.4 *Satellite fraction and its evolution with redshift*

Dark matter halos are usually inhabited by a massive central galaxy and a number of smaller satellite galaxies orbiting the potential well of the halo. At a fixed stellar mass, a galaxy can be either a central in a relatively low-mass halo or a satellite in a massive one. The number of satellite galaxies in a halo and its evolution with redshift reflects the halo's evolutionary history in terms of its hierarchical merger tree, but it also reflects the physical processes and environmental effects that can affect the assembly of satellites. Using our constraints on the HOD in the broad redshift span up to $z \sim 5$, we can study the evolution of the satellite fraction and get insights into the halos' evolutionary history.

Within the HOD framework, we can compute the fraction of satellite galaxies, summed over all halos and with masses above a given stellar mass threshold; then, using our best-fit parameters in the ten z-bins reconstruct its evolution with redshift. To compute the satellite fraction, we perform the following integration:

$$f_{\rm sat}(z| > M_*^{\rm th}) = 1 - \frac{1}{\bar{n}_g} \int {\rm d}M_h \frac{{\rm d}n}{{\rm d}M_h} \Big\langle N_{\rm cent}(M_h| > M_*^{\rm th}) \Big\rangle, \tag{7.15}$$

where, as before, n/M_h is the halo mass function, $\bar{n}_g$ is the mean number density of galaxies with $M_* > M_*^{\rm th}$ and $\langle N_{\rm cent}(M_h| > M_*^{\rm th}\rangle$ is the mean occupation function for centrals with the best-fit parameters

Our results on satellite fraction of galaxies with masses above $\log M_*/M_\odot > [8.7, 9.0, 9.3, 9.7, 10.2, 10.7, 11.0]$ as a function of redshift are shown in Fig. 7.8. The general trend at all mass thresholds is an increase of the satellite fraction as cosmic time flows. For example, galaxies with masses $\log M_*/M_\odot > 9.7$ see an increase from about 10% at $z \sim 3$ to $\sim$ 18% at $z \sim 1.5$ all the way up to $\sim$ 25% at $z \sim 0.9$. In the lowest bin $0.2 < z < 0.5$, $f_{\rm sat}$ appears to systematically drop by about 3-4% for all stellar mass thresholds. This is likely results from a feature in the data, since the survey is not optimized for low redshifts. The fraction of satellites depends on the stellar mass threshold – at all redshifts, there are more low-mass satellites than high-mass ones. Furthermore, the increase with redshift is different with respect to the stellar mass threshold – the fraction of high mass satellites increases more slowly, only to reach $\sim$ 8% at $z \sim 0.6$. We note that the $f_{\rm sat}$ in $2.5 < z < 3.0$ have all very similar values, which is an artifact arising from systematic errors in the HOD parameters. We investigated that this is mainly driven by $\beta_{\rm cut}$ parameter which is poorly constrained.

The satellite fraction $f_{\rm sat}$ as a function of stellar mass threshold rises sharply from very massive to intermediate-mass satellites, but then reaches a plateau for intermediate to low-mass thresholds, especially at low z. This can be explained by the fact that low-mass galaxies are preferentially central galaxies in smaller halos rather than being satellites in more massive halos. This can be understood considering the halo mass function and the halo occupation function: even though the number of satellites increases as a power law with halo mass, there are simply more low-mass halos that can host a lower mass central; furthermore, the exponential high-mass cut-off of the halo mass function means that high mass halos that can host many low-mass satellites are very rare. Therefore, at a fixed low redshift, the satellite fraction increases with decreasing stellar mass threshold and reaches a plateau at about 30%.

7.5.5 *Inferred SHMR for centrals*

The SHMR and M_*/M_h ratio for centrals are shown in Fig. 7.9 at all z in the top and bottom panels, respectively. The shaded region envelops the 16th and 84th percentiles of the distribution of M_* at a given M_h that is obtained by plugging in the parameters of the MCMC chain in Eq. 4.3. The solid line corresponds to the 50th percentile of this distribution. In the remainder of the chapter, the 1σ confidence intervals are always computed in this way, unless stated otherwise. On the right-hand side of the $y-$axis, we show the corresponding halo star-formation efficiency (SFE) in percentages.

The SHMR increases monotonically with halo mass, changing slope at $M_h \approx 10^{12}\,M_\odot$ and $M_* \approx 5 \times 10^{10}\,M_\odot$. Below this pivot mass, the SHMR increases steeply with a slop that remains constant with redshift. Above the pivot mass, the slope suddenly decreases

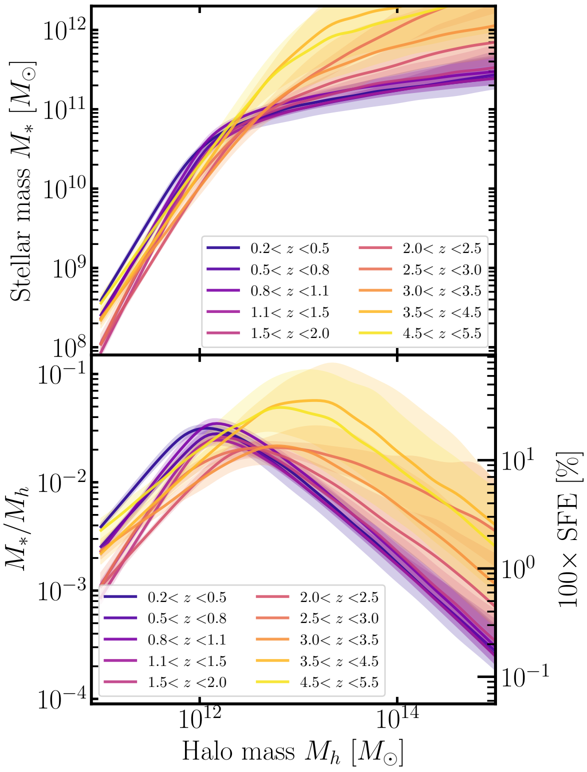


Figure 7.9: Stellar-to-halo mass relation (top) and M_*/M_h ratio (bottom) in the ten redshift bins. The solid lines and shaded regions show our inferred SHMR and 1σ confidence interval, color-coded according to the redshift bin.

and the stellar mass increases more slowly with halo mass. The SHMR is higher at low-z for masses below the pivot, and lower at low-z for masses above the pivot.

The M_*/M_h ratio, which can be considered as the star-formation efficiency integrated over the halo's lifetime, strongly depends on halo mass. The SFE can be defined as $\epsilon = f_b^{-1} M_*/M_h$ to quantify how efficiently baryons are converted into stars in galaxies residing in halos of a given mass – it is essentially a ratio between the star-formation rate and halo growth rate multiplied by the universal baryonic fraction. Our results, in line with previous findings, show that at all halo masses and at redshifts at least up to $z \sim 3$ the SFE is lower than 20%, indicating a globally inefficient galaxy formation process. In the last three z-bins above $z > 3$, our results become very uncertain – the large error bars on the fitted parameters propagate into large uncertainties on the SHMR that make the interpretation difficult. This can be due to increasingly smaller sample, especially of high-mass galaxies, as well as uncertainties in the physical parameters and possible cosmic variance effects.

The SFE peaks at 17% occurs at halo masses of about $M_h = 2 \times 10^{12} M_\odot$. It then decreases rapidly at lower and higher halo masses – about a 15% decrease in SFE for a decrease of 1 dex in halo mass, and a $\sim$ 10% decrease for an increase of 1 dex in halo mass. This behavior indicates that the majority (around two-thirds) of star-formation occurs in a relatively narrow range of halo masses around this peak (see e.g., Behroozi, Wechsler and Conroy, 2013; Behroozi et al., 2019). This peak corresponds to stellar mass of about, $M_* = 5 \times 10^{10}\,M_\odot$, which is the typical M_* mass scale of Milky Way-like galaxies ($\sim 6 \times 10^{10}\,M_\odot$, Licquia and

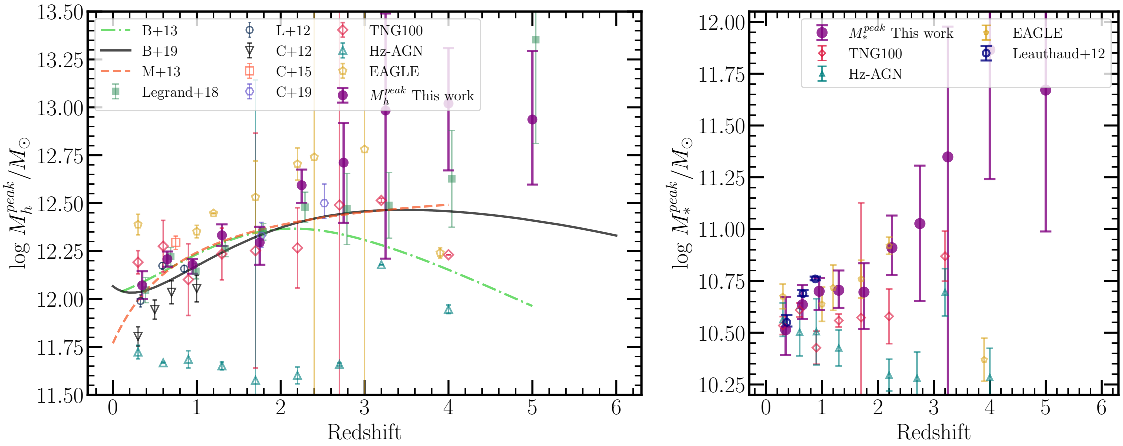


Figure 7.10: Evolution of the peak halo (left panel) and peak stellar mass (right panel) with redshift. The results from our analysis are shown in purple points. For comparison, we show measurements from the literature rescaled to match our chosen value for $H_0 = 70\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$. The literature measurements include Legrand et al. (2019), Leauthaud et al. (2012, pp. L+12), Coupon et al. (2012, pp. C+12), Coupon et al. (2015, pp. C+15), Cowley et al. (2018, pp. C+18), Moster, Naab and White (2013, pp. M+13) , Behroozi, Wechsler and Conroy (2013, B+13), Behroozi et al. (2019, B+19), and from the hydrodynamic simulations Horizon-AGN, TNG100, and EAGLE (references in the main text).

Newman, 2015). With respect to redshift, the peak SFE shows only a mild evolution, generally toward lower values with increasing redshift. This is further discussed in the next subsections.

7.5.6 *Redshift evolution of the peak mass quantities*

The peak halo mass (M_h^{peak}) in the M_*/M_h ratio represents the mass at which the galaxy formation process, integrated over the entire history of the halo, has been most efficient. Since the feedback mechanisms also depend on halo mass, the redshift evolution of M_h^{peak} informs us about the halo mass scales at which different feedback mechanisms become more important throughout cosmic time. We compute the peak SFE from the central M_*/M_h.

Redshift evolution of the peak halo mass.

Figure 7.10 shows the redshift evolution of M_h^{peak} inferred from our analysis, compared to a compilation of measurements from the literature. To obtain M_h^{peak} and its error bars we compute the peak mass for each parameter set of the MCMC samples; then from this distribution we compute the median, 16th and 84th percentile. Our results show that the peak halo mass increases with redshift from $M_h^{\mathrm{peak}} = 1.43 \times 10^{12}\, M_\odot$ at $z = 0.35$ to $M_h^{\mathrm{peak}} = 4.89 \times 10^{12}\, M_\odot$ at $z = 2.75$. The peak halo mass continues to increase up to $M_h^{\mathrm{peak}} = 7.25 \times 10^{12}\, M_\odot$ in our highest bin at $z = 5$. At $z > 3$, the uncertainty of the peak position increases due to the large uncertainties in the SHMR. While at low redshifts the peak value and evolution is in agreement with the literature, at $z > 3$ there is a large scatter in the literature with M_h^{peak} values ranging from $M_h \sim 10^{12}\,M_\odot$, as found by Behroozi, Wechsler and Conroy (2013), to $M_h \sim 2.5 \times 10^{13}\,M_\odot$, as found by Legrand et al. (2019).

Redshift evolution of the peak stellar *mass*

The right panel of Fig. 7.10 shows the evolution of the peak stellar mass M_*^{peak}. At the peak stellar mass, galaxies can be considered to have been the most efficient in converting baryons to stars. We find an increase of M_*^{peak} from $3.1 \times 10^{10}\,M_\odot$ at $z = 0.35$ to $8.7 \times 10^{10}\,M_\odot$ at $z = 2.75$. This increase means that at earlier times, more massive galaxies have been more efficient in the star-formation process; as time elapses, this efficiency moves toward lower mass galaxies. However, the co-evolution of both the peak halo and stellar mass leaves the M_*/M_h ratio nearly constant with time (further discussed in Section 7.5.8).

The trends of increase with increasing redshift of both the peak halo and peak stellar mass is a signature of the downsizing scenario. Downsizing, in its most general sense, refers to the decrease with time of some mass scale parameter that is related to stellar growth or star-formation (Cowie et al., 1996). In our case, these mass scale parameters are the peak halo and stellar mass, which are related to the efficiency of the star-formation process. Their increase with redshift means that higher mass halos and galaxies were more efficient in converting the baryon reservoir to stars at higher redshifts. Consequently, the feedback mechanisms, especially the ones active at the massive end, were less efficient in the past.

Literature comparison

From the literature compilation, we remark on comparisons with the following works. Legrand et al. (2019) used the previous iteration of the photometric catalog in the COSMOS field – COSMOS2015 – to infer the SHMR by fitting the same functional form using parametric sub-halo abundance matching. Their results (shown in green squares) are in close agreement with our results. Next, in three z-bins up to $z < 1$, we include the results from Leauthaud et al. (2012), which serve as our main reference for the theoretical modeling. Their analysis is based on a joint abundance, clustering, and galaxy-galaxy lensing fit on measurements done in the COSMOS field; their results are shown in dark blue circles. Unsurprisingly, our results are in agreement with the trend. The higher value in the $z \sim 0.65$ bin found by our work, Leauthaud et al. (2012) and Legrand et al. (2019), is likely a feature of the COSMOS field. Indeed, using 10 000 spectroscopic redshifts, Kovač et al. (2010) reported a very large overdense structure in COSMOS at these redshifts. Behroozi, Wechsler and Conroy (2013) and Behroozi et al. (2019) used empirical modeling where galaxies are populated in dark matter halos and are traced within their halos over time; the models are constrained to match a set of observables such as the SMF, luminosity function, and cosmic star formation rate, among others. Their results are in general agreement at low redshifts, but show that the peak halo mass turns over and starts to decrease at $z \sim 2$ or $z \sim 3$. Our results, as well as Legrand et al. (2019), on the other hand, show a peak halo mass increasing up to $z \sim 4$. It is possible that this effect is driven by effectively the same

COSMOS dataset used in ours and Legrand et al. (2019) analyses. Behroozi et al. (2019), for example, constrained the SHMR at high redshifts using GSMF derived from Song et al. (2016) data. It is important to note that they used UV-to-$M*$ conversions to estimate stellar masses, which comes with some caveats. The SHMRs at $z \sim 3-4$ of Behroozi et al. method are sensitive to the GSMF at $z > 4$, and clearly sensitive to the choice of observational constraints (see e.g., Behroozi and Silk, 2015).

Finally, in the left panel of Fig. 7.10, we compare the peak stellar mass with the literature. Our results are in good agreement with Leauthaud et al. (2012). Interestingly, no hydrodynamical simulations show a clear downsizing trend in the peak stellar mass.

In summary, our results show that the peak halo mass increases monotonically with redshift in agreement with the literature up to $z < 3$, including findings in hydrodynamic simulations. At higher redshift, the literature suggests a turnover and decrease of the peak halo mass, which is not captured by our analysis. The increase of both peak halo and peak stellar mass with redshifts is in accordance with the downsizing scenario, where the more massive halos and galaxies were more efficient in star-forming earlier in the universe and the peak efficiency shifts toward lower mass halos.

7.5.7 *Total SHMR*

As the mass of the halo increases, the number of satellite galaxies that occupy it also increases and, naturally, their contribution to the total stellar mass budget of the halo becomes important. In massive halos, stellar mass is assembled from in-situ star-formation and from mass accretion via merging of halos, while the growth is regulated by various quenching mechanisms. The ratio between the total stellar mass and the halo mass (total SHMR) can then inform us about the efficiency of the combination of both effects. The model adopted in this study allows us to compute the total stellar mass contained in a halo of a given mass using Eq. 4.10. To obtain the total stellar mass, an integration is carried out over the stellar masses with lower and upper mass limits. Ideally, we would integrate over the whole range of possible stellar masses, but in our case that would mean an extrapolation of the models beyond the stellar masses probed in our analysis. This can introduce inaccuracies, especially for computing the satellite contribution. However, we checked that most of the contribution to the total stellar mass content in $M_h > 10^{12}\,M_\odot$ halos comes from satellites in the mass range of $10^{10}\,M_\odot < M_* < 10^{11}\,M_\odot$, well within the mass scales probed by our analysis; this is also stated in Leauthaud et al. (2012). For our purposes, to compute $M_*^{\rm tot,sat}$, we set the lower integration limit to $M_* = 10^{8.5}\,M_\odot$. This lower stellar mass limit is below our completeness limit at $z > 2$, but we expect that the extrapolation at lower masses is not inaccurate enough to bias the results.

Figure 7.11 shows our results on the total stellar content as a function of the host halo mass in the top panel and $M_*^{\rm tot}/M_h$ ratio in the bottom panel in $0.2 < z < 0.5$ and for the other nine z-bins in Fig. 7.12. We show the central SHMR in dashed purple, while the dotted purple line shows the satellite contribution. The shaded purple area envelops the 1σ uncertainty in the total SHMR. In Fig. 7.12, both the central and satellite are displayed in

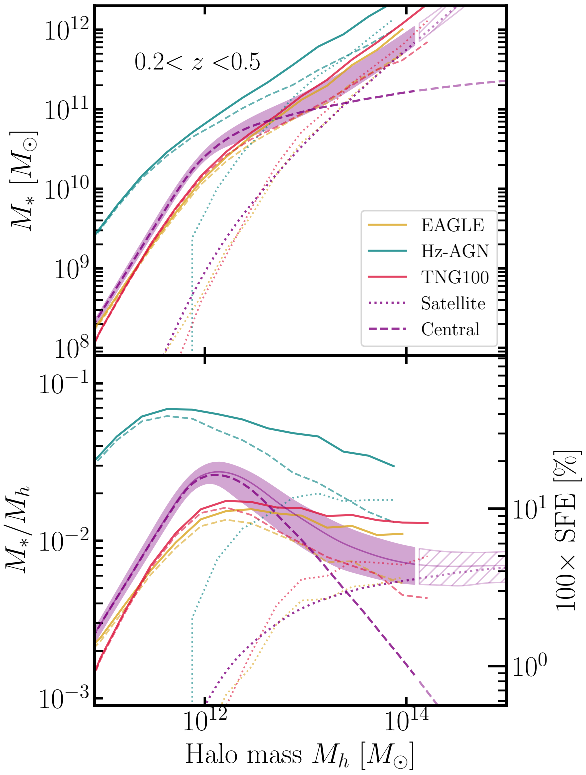


Figure 7.11: Total stellar-to-halo mass relation (top panel) and total M_* / M_h (bottom panel) at $0.2 < z < 0.5$ compared to hydrodynamical simulations. The purple dashed and dotted lines show our central and satellite contribution to the total SHMR respectively, with the shaded region showing 1σ confidence interval of the sum of the two. The break in solid purple lines and shaded regions indicate the highest stellar mass probed in our analysis, which we take to be the highest mass bin in the SMF. The transparent purple lines and hatched region is an extrapolation at higher masses. We overplot the SHMRs measured in the hydrodynamical simulations HORIZON-AGN in teal, TNG100 in red, and EAGLE in dark yellow, where the dashed dotted and solid lines show the central, satellite, and total SHMR.

solid purple with the 1σ envelope of the total $M_*^{\rm tot} / M_h$ ratio. The break in the lines and in the shaded region indicates the upper stellar mass limit probed by our analysis, as well as the extrapolation to higher masses is shown in transparent purple lines and hatched region. On the right-hand side y-axis, we show the integrated SFE. We recall here that by definition, the integrated SFE describes how efficiently stellar mass has been assembled in halos integrated over the halo's lifetime. This inevitably mixes various assembly paths, such as stellar mass formed in low-mass halos (where different mechanisms regulate growth) that are later accreted as satellites in massive halos.

At masses below $M_h \lesssim 10^{13} \, M_\odot$, the total stellar mass content is completely dominated by central galaxies. It rises sharply up to the peak halo mass at around $M_h \sim 10^{12} \, M_\odot$, and then falls more gradually with increasing halo mass. The peak of the total $M_*^{\rm tot} / M_h$ ratio is completely set by the centrals, meaning that the physical processes shaping the peak have to be related to the quenching of the central galaxy. Since the satellite contribution only becomes important at halo masses almost one order of magnitude higher, the accumulation

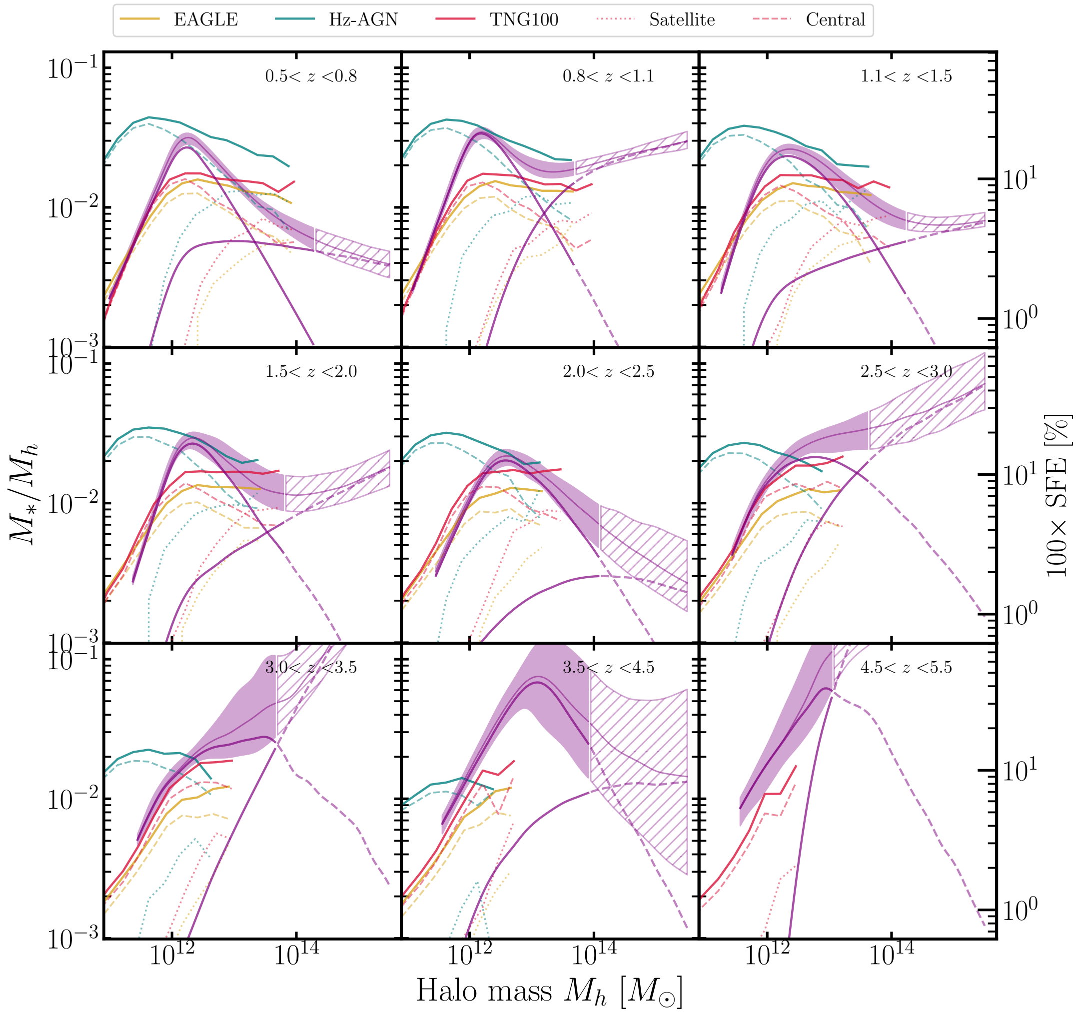


Figure 7.12: Total M_*/M_h in the different redshift bins compared to hydrodynamical simulations. The purple lines show our inferred central and satellite contribution to the total SHMR with the shaded region showing a 1σ confidence interval of the sum of the two. The break in solid purple lines and shaded regions indicate the highest stellar mass probed in our analysis, which we take to be the highest mass bin in the SMF. The dashed purple lines and hatched region is an extrapolation at higher masses. The comparison includes total M_*/M_h found in the hydrodynamical simulations HORIZON-AGN in teal, TNG100 in red and EAGLE in dark yellow. The dashed and dotted lines for the simulations indicate the central and satellite contributions, respectively.

of stellar mass in satellites (instead of the central growth or mergers), cannot be responsible for setting the peak efficiency.

For masses higher than $M_h \sim 10^{12}\,M_\odot$, the satellite contribution to $M_*^{\rm tot}$ rise as the central contribution drops, and a transition occurs at about $M_h = 2 \times 10^{13}\,M_\odot$ where satellites start to dominate. Going to higher masses, the satellite $M_*^{\rm tot}/M_h$ starts to flatten out, meaning that the stellar mass keeps up with the growth rate of halos in group- and cluster-scale halos. Excluding the $2.5 < z < 3.5$ and $4.5 < z < 5.5$ bins, the total $M_*^{\rm tot}/M_h$ ratio always remains below the peak set by the centrals at an SFE below 5%. This suggests that even if all the satellites were to merge into the central, the SFE would still be lower than the peak, indicating strong environmental quenching mechanisms in massive halos. In other words, the accumulation of mass in satellites is not responsible for the low M_*/M_h.

7.5.8 *Redshift dependence of M_*/M_h at fixed halo mass*

The M_*/M_h ratio (i.e., the integrated star-formation efficiency) might evolve with redshift differently depending on the mass of the halo. This dependence, especially when considering the contributions from both centrals and satellites, could shed light on the importance of the feedback mechanisms that regulate star-formation at different halo mass-scales. Figure 7.13 shows the M_*/M_h ratio as a function of redshift at different halo masses $\log M_h/M_\odot = [11.50,\ 12.00,\ 13.00,\ 13.60]$, as well as for M_h^{peak}. For the most massive halos at $M_h = 10^{13.6}\,M_\odot$, we also decomposed it in the contributions from satellites and centrals. We restricted our analysis to $z < 2.5$ since the large uncertainties at higher z prohibit a meaningful quantitative analysis.

Low-mass halos, below the peak halo mass ($M_h < 10^{12}\,M_\odot$), show their SFE steadily increasing from $z \sim 2.3$ up to present day. For $M_h = 10^{11.5}\,M_\odot$ the SFE goes from $\sim 2.5\%$ at $z = 2.3$ to $\sim 7\%$ at $z = 0.3$. For slightly more massive halos of $M_h = 10^{12}\,M_\odot$, the SFE increases even faster from $\sim 8\%$ at $z = 2.3$ to $\sim 17\%$ at $z = 0.3$. In contrast, for high-mass halos above the peak halo mass ($M_h > 10^{13}\,M_\odot$), the SFE remains almost constant, with a slight decrease with decreasing redshift. Furthermore, more massive halos show an even lower SFE by several percents.

These trends are a signature of the downsizing scenario, which was already mentioned in Section 7.5.6. Downsizing refers to the observation that, contrary to the hierarchical formation scenario in which small halos are formed first and subsequently grow by merging and accretion, more massive and early-type galaxies have stellar populations that are formed earlier (De Lucia et al., 2006; Thomas et al., 2010). This downsizing is observed in our results from the increase of the SFE of low-mass halos and the slight decrease in the efficiency of high-mass halos with decreasing redshift. Low-mass halos having been more efficient in forming stars at later times means that the stellar populations of lower mass galaxies inhabiting them would also be younger.

The SFE at the peak halo mass shows only a weak evolution with redshift, remaining constant to $z = 0.95$ then dropping from 20% to 16% at $z = 2.5$. Previous findings also point to a peak efficiency constant with redshift (Behroozi et al., 2019; Moster, Naab and White, 2018). This behavior suggests that the M_h^{peak} can be considered as the halo mass scale at which the integrated SFE history remains constant with redshift.

Satellites dominate the mass budget in high-mass halos, as can be seen in Fig. 7.13 for $M_h = 10^{13.6}\,M_\odot$ where we also show M_*/M_h for centrals and satellites separately. The satellite dominance is stronger at lower redshift, whereas from $z > 1.2$, the central contribution catches up and both remain comparable. This is unsurprising given the fact that the satellite fraction increases with decreasing redshift.

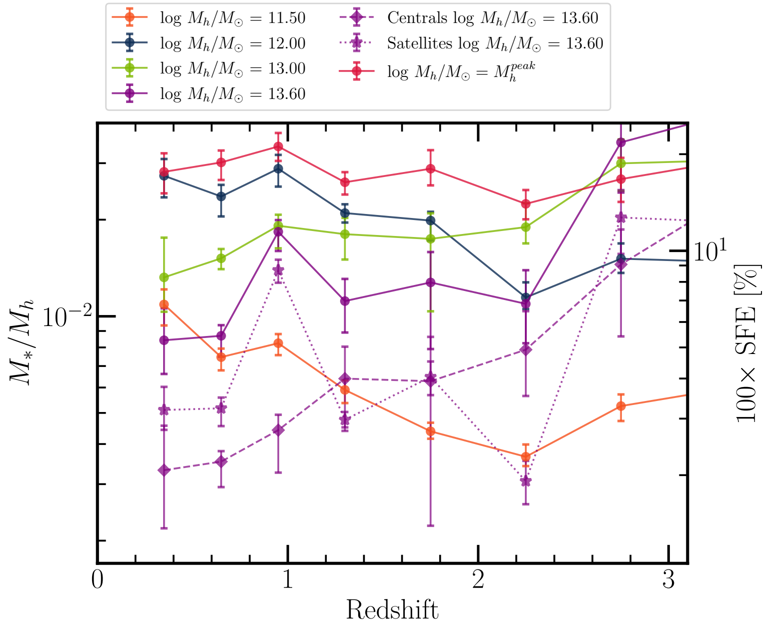


Figure 7.13: Redshift dependence of M_*/M_h at fixed halo mass. We show the total (centrals and satellites) M_*/M_h fixed at $\log M_h/M_\odot = [11.50,\ 12.00,\ 13.00,\ 13.60,\ M_h^{\text{peak}}]$, while the M_*/M_h for log $M_h/M_\odot = 13.60$ is also shown for centrals and satellites separately with star and diamond symbols, respectively.

7.6 DISCUSSION

7.6.1 *Physical mechanisms that regulate the stellar mass assembly of centrals*

The shape of the SHMR as a function of halo mass can provide us with qualitative information on the stellar mass growth mechanisms. For example, at $M_h < 10^{12}\,M_\odot$, the steep increase in the SHMR with halo mass tells us that the stellar mass growth in galaxies is mainly driven by in situ star formation, as opposed to mergers (Conroy and Wechsler, 2009; Leauthaud et al., 2012). The explanation comes from the fact that halos grow by merging with lower mass halos, in which the stellar mass drops as a power law (very low mass halos can even be devoid of stars). Since merging will increase the halo mass more significantly compared to the stellar mass, the stellar mass growth in low-mass halos has to be driven by star formation in order to obtain the steep rise of the SHMR.

The shape of the SFE with its characteristic peak is mostly driven by different feedback processes, whose efficiencies are dependent on halo mass. At low masses, stellar feedback from supernovae (SNe), stellar winds, radiation pressure, and photoheating can heat up and prevent the baryons from collapsing into stars (e.g., Dekel and Silk, 1986; Dubois and Teyssier, 2008; Hopkins et al., 2014; Kimm et al., 2015). As the halo mass increases, its potential well deepens and cold flows feed the halo with gas that fuels star-formation. The SNe are not powerful enough to prevent these cold flows, so the SFE increases with halo mass. As the halo mass increases even further, virial shock heating and AGN activity heat up the gas and quench star formation, thus driving the SFE back down (see e.g., Silk, Di Cintio and Dvorkin, 2014, for a review). Vogelsberger et al. (2013) show in their hydrodynamical simulations that the mechanical, radio-mode AGN feedback is the most responsible for the suppression of star formation in these massive halos (see also e.g., Dubois et al., 2010). Similar conclusions have been reached from AGN observations in the COSMOS field, where Vardoulaki et al. (2021) show that the radio-mode AGN feedback plays a significant role in shaping the SHMR at lower redshifts. Additionally, Gabor and Davé (2015) in their hydrodynamical simulations, these authors show that galaxies above the peak mass indeed tend to be quenched. These quenched centrals can still continue to grow in stellar mass via dry merging, but halo mass growth being faster – since halos keep accreting from their large-scale environment, while the gas accreted by galaxies is not efficiently turned into stars any longer due to feedback – leads the M_*/M_h ratio to decrease.

In Fig. 7.10, we showed that more massive halos are more efficient in forming stars at higher redshifts – the peak halo mass increases with z. This peak in the central SHMR is shaped by the interplay between cold flows being more efficient as the halo mass increases and shock heating in the hot halo. The increase of the peak halo mass is a signature that cold flows become more important in driving star formation at $z > 1.5$ in $M_h > 10^{12}\,M_\odot$ (Dekel et al., 2009; Oser et al., 2010; Dubois et al., 2013).

With respect to redshift, we observe in Fig. 7.12 that the peak generally flattens out at higher redshift. This is directly connected to the fact that the SMF at higher redshifts increases its low mass slope and the knee is smoothed (i.e., the SMF and HMF become more similar in shape). This indicates that the high mass halos decrease in their star-forming efficiency with time and that the AGN feedback has a larger impact at later times.

7.6.2 *Physical mechanisms that regulate the stellar mass assembly in satellites*

In Section 7.5.8 and Fig. 7.13, we presented how the total SHMR ratio evolves with redshift at fixed halo mass. Comparing the contributions of centrals and satellites to the total stellar mass content at different M_h can shed some light on the relative importance of the quenching mechanisms acting on the central and satellite galaxies in a halo. Figure 7.13 shows that in high-mass halos of about $M_h = 10^{13.6}\,M_\odot$, the M_*/M_h is about 0.2 dex lower than M_*/M_h at the peak halo mass. At least up to $z < 2.5$, the total M_*/M_h ratio for high-mass halos is always lower than M_*/M_h at the peak halo mass. This suggests that the low M_*/M_h ratio for centrals is not due to the stellar mass being accumulated in satellites instead of the central because, even if all satellites were to merge with the central, the total stellar mass content would still be lower than at the peak. This is a clear indication of powerful feedback present in massive halos that prevents gas from cooling and forming stars, and these feedback processes are acting on both the central and the satellites.

In halos more massive than $\sim 10^{12}\,M_\odot$, which includes group- and cluster-scale halos, the virial shock heats the halo gas (Birnboim and Dekel, 2003). Naturally, then, satellite galaxies that reside in these massive halos have their SF impeded by the hot gas. This hot-halo mode is likely to be responsible for the satellite quenching. This mechanism is also called environment quenching (Peng et al., 2010; Gabor and Davé, 2015) and is found to be independent of stellar and halo mass – galaxies of all stellar masses reside in the densest environments in halos of $M_h > 10^{12}$. Additional environmental effects such as strangulation (Larson, Tinsley and Caldwell, 1980; Balogh, Navarro and Morris, 2000), ram-pressure stripping (Gunn and Gott, 1972), and harassment (Moore et al., 1996) further prevent the stellar mass assembly in satellites. Altogether, these processes could explain the low SFE at high halo masses in our results. Our results at high masses remain too uncertain to unveil a consistent picture of the redshift evolution of the efficiency of the satellite quenching mechanisms. Incorporating lensing measurements to better constrain the satellite contribution and separating active versus passive populations could shed more light on this aspect. This is a natural extension to this work that we plan to carry out in the future.

7.6.3 *Comparison with hydrodynamical simulations*

We compare our results with several state-of-art cosmological hydrodynamical simulations of galaxy formation: HORIZON-AGN[7] (Dubois et al., 2014; Kaviraj et al., 2017), EAGLE[8] (Crain et al., 2015; Schaye et al., 2015), and TNG100 of the ILLUSTRISTNG project [9](Nelson et al., 2019; Springel et al., 2018; Nelson et al., 2018; Pillepich et al., 2018; Marinacci et al., 2018; Naiman et al., 2018). A brief recap of the main features of these simulations is given in Appendix B.1. Once central and satellite galaxies are matched with DM halos as described in Appendix B.1, the SHMR is directly measured in halo mass bins at different redshifts, as the mean of the ratio between the stellar and the halo mass. Baryonic physics has a small but real impact on the halo mass function. Beltz-Mohrmann and Berlind (2021) showed that the halo mass function in DM-only simulations is overestimated (with respect to hydrodynamical simulations) by a factor of about ∼1.1-1.2 depending on the mass range at $z = 0$, though this effect disappears at high redshift and high mass (see also e.g., Bocquet et al., 2016; Desmond et al., 2017; Castro et al., 2020). This aspect should be kept in mind, knowing that the modeling presented in Sec. **??** relies on DM-only halo mass function, while for the hydrodynamical simulations, the mass of the DM host halos is naturally impacted by baryonic physics. This could lead to a small systematic underestimation of the SHMR in the observational data with respect to the simulated ones.

---

[7] `www.horizon-simulation.org`

[8] `http://eagle.strw.leidenuniv.nl/`

[9] `https://www.tng-project.org/`

Redshift evolution of the peak halo mass

In Fig. 7.10, we show the evolution of the peak halo mass in the three hydrodynamic simulations Horizon-AGN, TNG100, and EAGLE. We directly compute the peak halo mass using the measured SHMR in the simulations. The error bars are obtained by computing the peak of the lower and upper 1σ values of the M_*/M_h relation as measured in the simulations. Here, TNG100 shows the best agreement with both our results and the literature, whereas EAGLE shows slightly higher peak halo masses up to $z \sim 3$. We also note that in both simulations, the peak halo mass decreases from $z \sim 3$ to $z \sim 4$, which is a priori in disagreement with our results. That being said, the ratio M_*/M_h at $z > 3$ is also highly uncertain (as seen in Fig. 7.12) because the simulations start to miss massive halos with masses comparable to the peak due to their fixed small volume and the rarity of these halos. Horizon-AGN shows the biggest discrepancy with our results and the literature. In this simulation, the peak efficiency is reached in much lower mass halos: $M_h^{\mathrm{peak}} = 10^{11.75}\,M_\odot$ at $z \sim 0.35$ that decreases up $10^{11.55}\,M_\odot$ to $z \sim 2$ and then increase at higher redshifts. This discrepancy can be explained by the inefficiency of the SN feedback in low mass galaxies as implemented in the simulation (see e.g., Hatfield et al. (2019) or Kaviraj et al. (2017) for further discussions).

The simulated SHMR in the low-mass regime

We present the M_*^{tot}/M_h for Horizon-AGN (in teal), TNG100 (red), and EAGLE (dark yellow color) in Figs. 7.11 and 7.12. At the low-mass end, where M_*^{tot}/M_h increases with halo mass, TNG100 and EAGLE show a reasonable agreement with our results with a similar normalization, slope and position of the peak, albeit the peak SFE is lower by a few percent. Horizon-AGN, on the other hand, has a higher normalization of the M_*^{tot}/M_h and a peak efficiency at lower halo masses (as already noted above) probably driven by insufficient SN feedback especially at high redshift. This issue is also responsible for the overall overestimation of the stellar mass in low-mass galaxies. The good agreement of TNG100 and EAGLE with our results in this regime suggests that the implementation of feedback from SNe and stellar winds in these simulations is well calibrated to reproduce observational data. We note that we cannot conclude from this measurement that the star-formation model at the subgrid scale is realistic nor that stellar feedback in these simulations is physically meaningful (some feedback mechanisms could be missing, such as radiation from young stars that suppresses star formation in low-mass galaxies, and others could be overestimated). What we can conclude is simply that the cumulative strength of the implemented feedback processes leads to realistic SHMR relations (see the discussion on subgrid models in Appendix B.1.2).

Simulated SHMR in the high-mass regime

The most significant discrepancies between the observational data and all simulations appear for masses above the peak. At least up to $z = 3.5$ the central M_*^{cent}/M_h for all the simulations shows a peak, meaning that AGN feedback are powerful enough to quench the central galaxy growth compared to the halo growth. However, at all redshifts in TNG100 and EAGLE, the central M_*/M_h decreases with increasing halo mass more gradually (shallower slope) than in observations, while HORIZON-AGN shows this tendency only at $z < 1$. Furthermore, the contribution of satellites relatively to central at all redshifts is higher in the simulations. These two facts contribute to a flattening of the peak and a higher SFE with increasing mass in the simulations, especially in TNG100 and EAGLE.

Redshift evolution of the satellite fraction

In Fig. 7.8 we also compare the satellite fractions against the three hydrodynamical simulations. The dashed lines in the left, middle, and right panels show the $f_{\rm sat}$ in TNG100, HORIZON-AGN, and EAGLE, where the different color-coded lines correspond to the mass thresholds indicated in the legend. The satellite fraction in all simulations was computed with the same criterion as in Eq. 7.15 – the ratio between the number of $M_* > M^{\rm th}$ galaxies that are satellites and the total number of $M_* > M^{\rm th}$ galaxies for all the halos in the simulation. The satellite fraction in all the simulations increases linearly with decreasing redshift, a trend which is in agreement with our results. However, at fixed stellar mass thresholds, the simulations exhibit higher numbers of satellites at all redshifts compared to our analysis. In TNG100 and EAGLE, $f_{\rm sat}$ goes from $\sim 20\%$ at $z \sim 3$ to $\sim 42\%$ at $z \sim 0.3$ for satellites with $\log M_*/M_\odot > 9.0$. Compared to our results, the satellite fraction is higher by about a factor of 2 for the lowest mass thresholds and a factor of 1.2 for the highest mass thresholds. In HORIZON-AGN, the satellite fraction is lower by about 0.5 compared to the other two simulations, but still a factor of 1.4 higher than our results for the lowest mass thresholds.

The difference in the satellite fractions can be due to a degeneracy between the satellite fraction and the spatial distribution of satellites within halos (Xu et al., 2016) – steeper satellite profiles at small scales increase the correlation, thus decreasing the need for high satellite fraction. This raises the possibility that the higher satellite fraction in the simulations is due to flatter satellite profiles compared to our models, which assume an NFW. Bose et al. (2019) show that the mean radial number density of luminous satellites in TNG100 matches the NFW profile. Furthermore, we compared the radial satellite density distribution in all three simulation and find that they follow the NFW profiles reasonably well, with TNG100 showing the best match, while the EAGLE and HORIZON-AGN show flatter profiles that are consistent with NFW but for lower concentration parameters. The agreement in TNG100 suggests that satellite profiles cannot serve as the explanation behind the different satellite fractions. In EAGLE and HORIZON-AGN, on the other hand, the flattening of the profiles can be interpreted as a consequence of AGN feedback (see e.g. Peirani et al., 2017, for comparison of DM halo profiles with and without AGN).

Excess of satellites in halos "above fixed mass thresholds" would translate into a higher level of small-scale clustering, which is measurable by the two-point correlation function of low-mass and red galaxies. Artale et al. (2017) compares the small-scale correlation function of galaxies in EAGLE and the GAMA survey at $z = 0.1$. They confirm that low-mass red galaxies have a considerably higher correlation at small scales, which is consistent with high fractions of satellite galaxies. Although they did not investigate this feature specifically, Springel et al. (2018) mentioned that the present-day clustering of red galaxies less massive than $10^{10}\ M_\odot$ was overestimated in TNG100 with respect to the SDSS. This difference in the satellite fractions corroborates our measurement of the M_*/M_h in Fig. 7.12, where the satellite contribution in the simulations is usually higher than our measurements. A higher fraction of satellites above a mass threshold could be the consequence of satellite galaxies having grown more massive due to lack of quenching, at least during a certain period of cosmic time. Due to the power-law increase of the SMF at the low-mass end, inefficient satellite quenching would shift the satellite SMF toward higher masses. Then, above a stellar mass threshold, there would be more satellites. This indicates that satellite quenching in the hydrodynamical simulations is inefficient, or that it happens at later times in the halo lifetime, or both.

It is worth noting that Donnari et al. (2021) made a thorough comparison of the quenching fraction between TNG100 simulation and observations up to $z = 0.65$. They concluded that TNG quenched fractions of centrals and satellites are qualitatively in agreement with observations. The scenario behind our findings might therefore lie at higher redshift: if strong satellite quenching occurs too late in the simulations, these galaxies would have had time to grow more massive than in the observations, leading to the overestimation of the satellite SHMR at high halo mass. We note that, at least for TNG100 and EAGLE, this quenching inefficiency must be specific to galaxies that have already been accreted as satellites: indeed the SHMR of low-mass central galaxies at high redshift is in good agreement with our observational measurements[10]. It is possible to argue that this issue is related to the resolution limit of the simulations. The coarse resolution reached in the circum-galactic medium might be insufficient to correctly model high gas temperature inherited from virial shock heating, preventing an overly strong quenching of satellite galaxies in massive halos (Gabor and Davé, 2015). However, this interpretation does not hold to explain the too large contribution of satellites at high redshift in halos less massive than $M_h > 10^{12}\, M_\odot$ (which are less prone to shock heating). For those very small satellites, we could suspect that other environmental mechanisms (e.g., ram-pressure and tidal stripping, harassment) are not correctly modeled, also due to the lack of resolution in the simulations. Interestingly, Costa, Rosdahl and Kimm (2019) also shows that the efficiency of satellite tidal stripping depends on the degree of pre-processing of low-mass galaxies by radiative stellar feedback (which are not implemented in the simulations studied here. See also Katz et al., 2020 for a discussion of stellar radiation during reionization). Therefore, the lack of satellite quenching at high redshift in the cosmological simulations studied here could also be a consequence of the absence of pre-processing by radiative stellar feedback.

---

[10] The issue is different for HORIZON-AGN, as previously noted. In this simulation, all low-mass galaxies (central and satellites) are concerned by inefficient quenching, which suggests a different type of failure or mismodeling in the simulation.

7.6.4 *Sources of uncertainties and the effects of model assumptions on the inferred SHMR*

The measurements presented here can suffer from a number of systematic errors. Firstly, due to the relatively small volume probed by COSMOS, the effects of sample variance can lead to biased measurements. This might be the case at $z \sim 1$, where several works indicate an overabundance of rich structures (McCracken et al., 2015, and references therein). Such an overdensity increases the normalization of the SMF and adds extra power on intermediate and large scales in $w(\theta)$. The effect of these overabundances on the inferred SHMR would be to decrease this ratio – indicating an even lower efficiency of converting baryons to stars – and a shift of M_h^{peak} toward higher masses.

A second source of systematic error are uncertainties in the estimation of physical parameters, for instance, stellar masses and redshifts [11]. Stellar mass uncertainties propagate into the Eddington bias at the high-mass end of the SMF. The effect of the Eddington bias on the inferred parameters is an increase in the value of $\sigma_{\text{Log}M_*}$ which sets the scatter in stellar mass at fixed halo mass.

In Fig. 7.14, we show the uncertainties in the SMF as a function of stellar mass and redshift. We show the fractional error (σ/Φ) from the cosmic variance (CV) and SED fitting. We do not show the Poisson errors because they are subdominant. Cosmic variance dominates the error budget by about 0.8 dex at $M_* < 10^{11}\,M_\odot$ compared to the SED fitting. Both increase with mass, but the SED fitting uncertainties increase exponentially at $M_* > 10^{11}\,M_\odot$ and become dominant over CV at the most massive end ($M_* > 6 \times 10^{11}\,M_\odot$). The most massive galaxies are rare and reside in the densest regions, therefore, a small survey is more likely to get a biased view of these objects. However, due to photometric errors and degeneracies between different SEDs, their stellar masses come with uncertainties. Due to the exponential shape of the SMF at the high mass end, these uncertainties in the M_* translate into large uncertainties in the SMF (Eddington bias) and become dominant. The SED fitting uncertainties in the M_* uncertainties being amplified by the Eddington bias into large σ_{SED} means that it will be very difficult, even for future surveys, to improve on these uncertainties.

Additional systematic errors arise from the number of assumptions in the model. The ingredients that go into the HOD-based model of the observables are the halo abundance (HMF), their clustering properties (halo bias, b_h), the radial distribution of dark matter and galaxies within halos, and halo mass definitions. The literature abounds with prescriptions for all of these ingredients, which can all lead to different results for the SHMR. This often makes the comparison with the literature difficult. Coupon et al. (2015) has investigated the effect on various model assumptions such as the σ_8 value, $b_h(M_h)$ relation, assembly bias, mass concentration relation, and halo profiles, on the error budget of several HOD parameters. Their conclusion is that the model systematics can lead to errors comparable to the statistical errors.

---

[11] for a detailed analysis of the uncertainties affecting the SHMR see Behroozi, Conroy and Wechsler (2010)

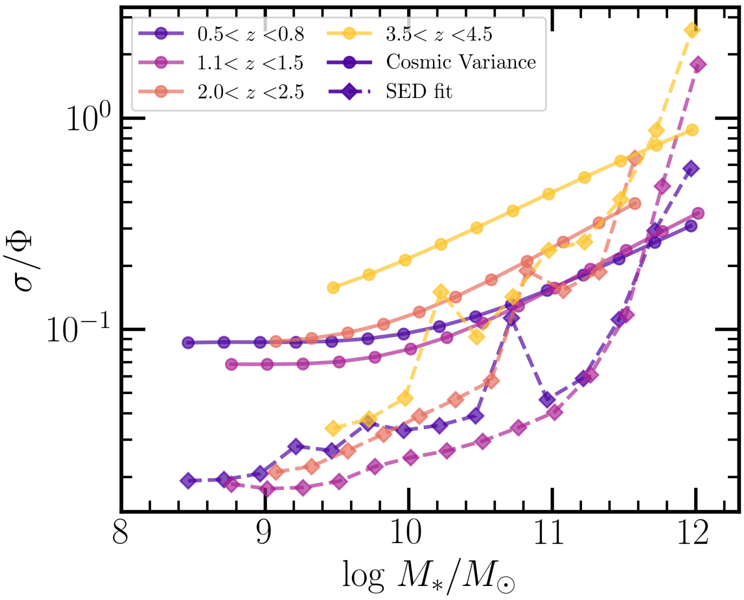


Figure 7.14: Uncertainties in the SMF expressed as a fractional error σ/Φ. We show the uncertainties from cosmic variance with solid line-connected points and the uncertainties from SED fitting in dashed line-connected diamonds. We show the uncertainties for four redshift bins.

7.7 SUMMARY AND CONCLUSIONS

This chapter presents the redshift evolution of the stellar-to-halo mass ratio to $z \sim 5$, derived from measurements of the galaxy stellar mass function and angular correlation function in the COSMOS2020 catalog fitted to a phenomenological model. The advantages of our work is the use of a single, homogenous dataset to perform all the measurements. Additionally, the HOD-based modeling allows us to consistently probe the contribution to the total stellar mass budget in halos of both central and satellite galaxies over this large redshift range, for the first time in the literature.

Our principal results are as follows:

- The mean halo occupation shows little-to-no evolution with redshift at fixed stellar mass, suggesting that galaxies occupy dark matter halos similarly throughout cosmic history.
- The M_*/M_h ratio for central galaxies – which may be interpreted as the integrated star formation efficiency (SFE) – strongly depends on halo mass, increasing up to a peak at halo masses of around 2×10^{12} and then decreasing again as the halo mass increases. The SFE shows little-to-no evolution with redshift and remains lower than 20% at least up to $z \sim 3$, indicating a globally inefficient galaxy formation process. The peak levels off with increasing redshift, consistent with a scenario in which AGN feedback in higher mass halos is less important at earlier times.

- The halo mass and stellar mass scale at which the SFE peaks, $M_h^{\rm peak}$ and $M_*^{\rm peak}$, increase continuously with redshift at least to $z \sim 4$. This stands in contrast to other works (e.g., Behroozi et al., 2019) where the peak halo mass decreases beyond $z \gtrsim 3$. However, given our errors, the peak halo mass evolution at these high redshifts remains uncertain.
- The total stellar mass content of halos, $M_*^{\rm tot}/M_h$, shows that at $M_h \lesssim 10^{13}\,M_\odot$ central galaxies completely dominate the stellar mass budget of the halo at all redshifts. The peak of the $M_*^{\rm tot}/M_h$ ratio is set by the central galaxy, indicating that the physical processes that shape the peak efficiency are related to the quenching of central galaxies.
- Satellite galaxies start to dominate the total stellar mass budget at $M_h > 2 \times 10^{13}\,M_\odot$ and the $M_*^{\rm tot}/M_h$ ratio flattens out as the halo mass increases. The fact that in the satellite-dominated regime, the $M_*^{\rm tot}/M_h$ ratio is lower than the peak means that strong quenching mechanisms must be present in massive halos that quench the mass assembly of satellites.
- For all samples, the satellite fraction decreases at higher redshifts. Moreover, there are always more low-mass satellites than high-mass ones. The satellite fraction increases more steeply for lower-mass satellites and reaches $\sim 25\%$ at low redshifts, whereas the most massive galaxies reach up to $\sim 15\%$.

We compared our SHMR measurements with three state-of-the-art hydrodynamical simulations HORIZON-AGN, TNG100, and EAGLE. For low halo masses, our results are in general agreement with TNG100 and EAGLE in terms of slope and peak of the $M_*^{\rm tot}/M_h$ ratio. However, there is a significant discrepancy with HORIZON-AGN that can be explained by insufficient stellar feedback. The most significant discrepancies with the simulations are for masses above the peak. The $M_*^{\rm tot}/M_h$ ratio in the simulations flattens out at the peak and has a larger value at higher masses compared to our results, which is mainly driven by a higher satellite contribution at all redshifts in the simulations – this excess in the satellite contribution relative to the central at the high-end is higher in TNG100 and EAGLE than in HORIZON-AGN. Furthermore, the simulations show higher fractions of satellite galaxies at all redshifts and all masses by about a factor of two. Both these findings at the high-mass end – excess of M_*^{tot}/M_h and high satellite fractions – suggest that the environmental quenching of satellites is less efficient in the simulations or that quenching occurs much later. This can be due to either an inefficient hot halo quenching mode, or from other environmental effects, such as ram-pressure or tidal stripping or harassment from other satellites, which are not well captured in the simulations – probably due to resolution effects. Lack of pre-processing by stellar radiative feedback could also have an impact.

To date, the COSMOS photometric redshift catalogs have provided the only homogeneous dataset to consistently study the evolution of the SHMR over a large redshift range (since $z \sim 5$) for a statistically representative area of the sky. However, the $2\,\deg^2$ of the survey does not eliminate the effects of cosmic variance; at $z > 2.5$, the uncertainty in our results makes it difficult to provide a convincing interpretation and comparison with simulations.

Our analysis is based on a phenomenological model that cannot provide any further insight into the physical processes governing the shape of the SHMR, especially the relative contribution of different feedback modes. But comparisons with hydrodynamical simulations, where the effects of feedback on the SHMR can be traced back, could provide insights into the relative importance of the different feedback mechanisms acting at different mass scales and environments and, ultimately, can help bridge the gap between observations and simulations.

One of the main issues raised by our comparisons with hydrodynamical simulations concerns satellite quenching. To investigate how satellites are quenched with respect to their DM halo throughout cosmic time, we could further separate the sample into star-forming and quenched galaxies. Upcoming works and surveys will make this possible. In the imminent future, COSMOS-Web will provide JWST observations in four NIR bands down to AB ~ 27 over 0.6 deg^2 in COSMOS. This unprecedented resolution and depth in NIR will enable lensing measurements (such as galaxy-galaxy lensing) to $z \sim 2.5$ and make it possible to measure the SHMR dependence on star-formation activity and even color gradients. At a slightly longer timescale, the Cosmic Dawn Survey will carry out deep NUV to MIR observations in ~ 50 deg^2 in the Euclid Deep Fields. Accurate photometric redshift and stellar mass measurements from this survey will probe the most massive end of the SHMR while greatly reducing cosmic variance.

8

STUDYING THE GALAXY-HALO CONNECTION WITH MAGNIFICATION

8.1 THE UNIQUE OPPORTUNITY FOR MAGNIFICATION

As we have seen in § 3.2.4, magnification offers an additional probe of the matter distribution, whose main observables are galaxy positions on the sky and magnitudes. On the other hand, other probes based on lensing effects, such as the galaxy-galaxy lensing, have the tangential shear as a crucial observable. This leads to its main limitation: accurate shape measurements are only available for a limited number of background galaxies, typically bright, at relatively low redshift. For example, state-of-the-art ground-based surveys reach $z \sim 1$ for DES (Pandey et al., 2021) and KIDS+VIKING-450 (Hildebrandt et al., 2020), whereas next generation space missions like *Euclid* will reach $z \sim 2$.

One of the most promising avenues for magnification is to provide lensing based measurements of dark matter halo properties at high redshifts beyond the regime of applicability of shear based probes. Additionally, since it is usually based on measurements of galaxy number densities (but also sizes in some approaches), it is subject to different systematic errors and can therefore provide complementary constrains in joint analysis with shear (van Waerbeke, 2010; Joachimi and Bridle, 2010). There is a body of work in the literature that explores the prospects of magnification for various purposes, such as estimating cosmological parameters, probing halo properties of clusters, groups and stacked individual galaxies.

One of the first measurements of magnification has been carried out in the SDSS survey by Scranton et al., 2005. They measured the 2PCF cross correlations between $\sim 13 \times 10^6$ galaxies at $z \sim 0.25$ and 2×10^5 quasars at $z \sim 1.75$, and successfully measured the expected behavior of the magnification 2PCF with magnitude. Hildebrandt, Waerbeke and Erben, 2009 measured magnification in the CFHTLS survey using samples of normal high-z galaxies in the background, selected as Lyman break dropouts. One of the earliest works, showcasing how magnification can be used to probe dark matter halos, is from Van Waerbeke et al., 2010, who laid a framework to estimate halo masses. Perhaps the most successful implementation of magnification in the literature has been to calibrate the mass-richness relation for clusters and estimate their masses by measuring cross-correlations with Lyman-break galaxies (e.g., Ford et al., 2014; Tudorica et al., 2017; Chiu et al., 2020). Other works have used sub-millimetre galaxies as background sources to probe cluster mass density profiles (Fernandez et al., 2021), constrain HOD parameters (González-Nuevo et al., 2017), as well as constrain the HMF (Cueli et al., 2022) and cosmological parameters (González-Nuevo et al., 2021). Additionally, Hildebrandt et al., 2013 used magnification to measure the masses of the dark matter halos of sub-millimetre galaxies by cross-correlating them with background Lyman break galaxies.

This Thesis aims to exploit the avenue opened by magnification to probe dark matter halo masses at $z > 1$. In contrast to what has already been done in the literature, our focus is to probe halo masses for normal foreground galaxies selected based on their photometric redshift and stellar masses. For this purpose, we rely on photometric redshift catalogs to select samples of normal galaxies in the foreground and background. The upcoming surveys will reveal the full potential of magnification, which in our context is to probe the dark matter mass for galaxies selected based on different properties (e.g., stellar mass, type, etc). This will be made possible by accurate photometric redshifts and physical properties measurements that will allow a clean selection of foreground and background sources. Additionally, foreground galaxies can further be selected in sub-samples based on their physical properties and therefore measure halo properties for galaxies of different stellar mass, type etc. As we also presented the S/N considerations of magnification measurements in § 3.2.4, magnification probes need large areas and large numbers of galaxies to achieve significant S/N. The *Euclid* Deep survey (§ 9.2.1), with its 50 deg^2 of deep multi-band imagery, will provide the ideal dataset for exploiting the constraining power of magnification.

In preparation for the Euclid survey, this Thesis explores the prospects of magnification in probing the relation between galaxy stellar mass and halo mass in real and existing data – the COSMOS2020 catalog. The COSMOS field is one of the *Euclid* calibration fields, and the excellent photometric redshifts and physical parameters provide a convenient testing ground for magnification. Its drawback is the relatively small area for lensing standards $\sim 2\mathrm{deg}^2$. In this Chapter, we first give a brief overview of how magnification can be used to probe the SHMR (section § 8.2), while referring to the theoretical framework developed in earlier chapters. As a showcase of the prospects of magnification, in § 8.3 we present magnification measurements in the COSMOS2020 catalog, while in § 8.4 we use these measurements to fit for the SHMR parameters using the HOD model. We summarize our results in § 8.5.

8.2 MAGNIFICATION TO PROBE THE SHMR

Magnification bias, as measured by the cross-correlation between a foreground and a background sample of galaxies, can be modeled with the HOD formalism, as presented in § 3.2.4. Therefore, by selecting the same foreground sample to measure magnification, clustering and abundance, one can combine them in a joint fit in order to better constrain the HOD parameters in the foreground z-bin. Magnification, being a lensing effect, is a direct probe of the dark matter halos and can help break degeneracies in the HOD parameter space. In particular, lensing probes are sensitive to parameters governing the satellite population (see Leauthaud et al., 2011, for a more detailed discussion on the constraining power of galaxy-galaxy lensing).

As presented in § 3.2.4, the correlation function due to magnification $w_{\mu}^{\mathrm{LS}}(\theta)$, can be modeled with the HOD with Eqn. 3.62. One of the main ingredients is the galaxy-matter power spectrum (Eqn. 3.63), which is specified in terms of the HOD. We can write this schematically as

$$w_{\mu}^{\mathrm{LS}}(\theta) \propto 2(\alpha-1)\, f(P_{\mathrm{gm}}(\Pi_{\mathrm{HOD}})), \tag{8.1}$$

where Π_{HOD} are the HOD parameters. In this work, we are interested in measuring the SHMR with magnification, therefore we use the same approach of HOD modeling of the SHMR as for clustering, described in § 4.3 and § 7.4.2. Therefore, since the HOD is written in terms of the SHMR parameters, Eqn. 8.1 can be used to constrain the SHMR. Additionally, with this formalism, the magnification correlation function is modeled consistently with the clustering correlation function and GSMF. This allows to leverage their complementarity and use them jointly to measure the SHMR.

Another important ingredient is the slope of the magnitude number counts α. The slope needs to be measured in the same band used to select the background sample. An accurate estimation of the slope is crucial, especially when selecting fainter samples that are close to the completeness limit of the survey. Errors in the slope estimations can lead to severe biases in the interpretations of the magnification measurements (Hildebrandt, 2016).

Finally, it is worth noting that magnification measurements can be used to assess the quality of photometric redshifts by looking for residual cross-correlations, as captured by the term w_{LS}^{err} in Eq. 3.61. For example, once the model parameters are calibrated using clustering and abundance, one can compute $w_{LS}^{g\mu}$ and compare it against the measurements. Any excess of amplitude in the measurements compared to the model would be an indication of clustering systematic due to redshift interlopers.

8.3 MAGNIFICATION MEASUREMENTS IN COSMOS2020

We explore the prospects of weak lensing magnification for probing the SHMR in the COSMOS2020 photometric catalog (see § 7.2.1). We measured the magnification bias using the angular cross-correlation function. For the foreground lens samples, we use the same samples as in the clustering and abundance measurement up to $z = 1.5$, as presented in § 7.2.2. These include the four redshift bins and the stellar mass threshold selected samples. We select the background sample in the redshift bin of $2.7 < z < 4.0$. The redshift distributions of the foreground and background samples is shown in Fig. 8.1.

The magnitude number counts of the background source sample is shown in Fig. 8.2 (Left). We use the cumulative number counts to impose a magnitude cut for the source sample that maximizes the S/N ratio (see § 3.2.5). To determine this, we compute the slope as a function of magnitude along with an estimate of the signal-to-noise $S/N \propto |2.5\,\alpha - 1|\sqrt{N_S}$, where N_S is the number of background sources brighter than m_{K_S}. To compute $\alpha(m)$, we fit $\log N(<m)$ with a polynomial function $0.5\,am_j^2 + bm_j + c$. This function is

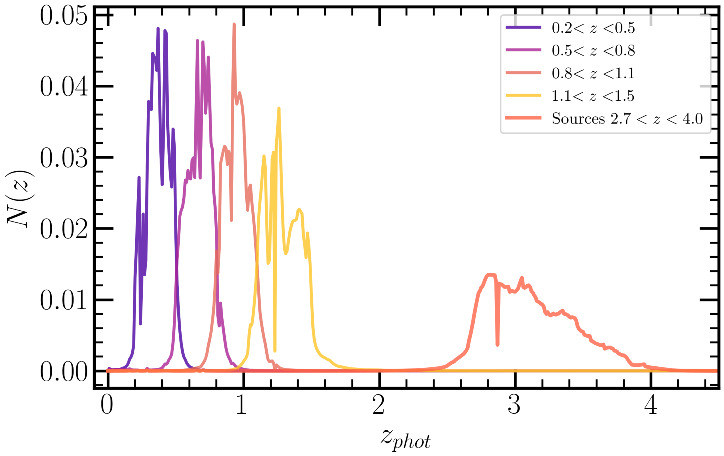


Figure 8.1: Redshift distribution of the samples used to measure magnification bias

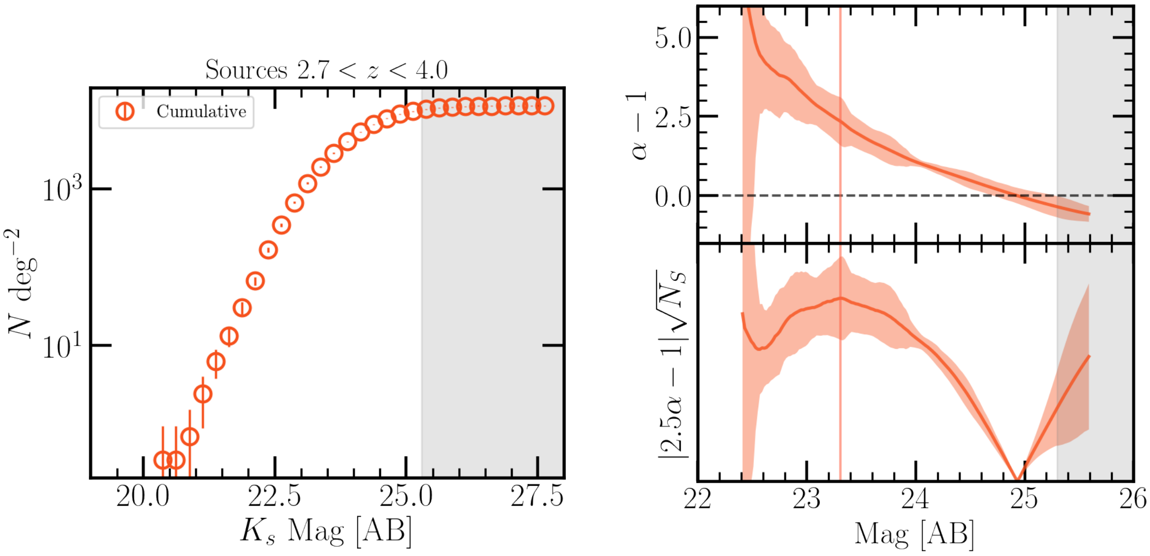


Figure 8.2: **Left:** Magnitude number counts for the sources sample in K_S band. **Right:** The slope of the number counts as a function of K_S magnitude (top panel) and the S/N estimate in the bottom panel. The vertical line marks the magnitude cut that we impose to maximise S/N, while the gray area marks the completeness limit.

fit in a magnitude bin j of width $m_j \pm 0.7$. From the fitted polynomial in each j-bin we compute the slope by taking the derivative as $\alpha_j = 2.5\,(am_j + b)$. Finally, we interpolate the slope at any m within the fitted range. This is shown in Fig. 8.2 (Right). We impose a cut in K_s-band magnitude of 23.3 for the background sources, which maximizes the S/N, as shown in Fig. 8.2 (Right, bottom panel). This results in $N_{\rm src} = 3713$ background sources.

The measurements of the angular cross-correlation function due to magnification for each of the four z-bins are shown in Fig. 8.4. For each z-bin we measure the magnification bias for samples selected in the same stellar mass thresholds as for the clustering case. The magnification measurements show a positive correlation, and more massive lenses have a higher amplitude. This is well expected, since the lensing potential is higher around more massive galaxies living in more massive halos. This showcases the potential of magnification to probe the halos around galaxies of different type.

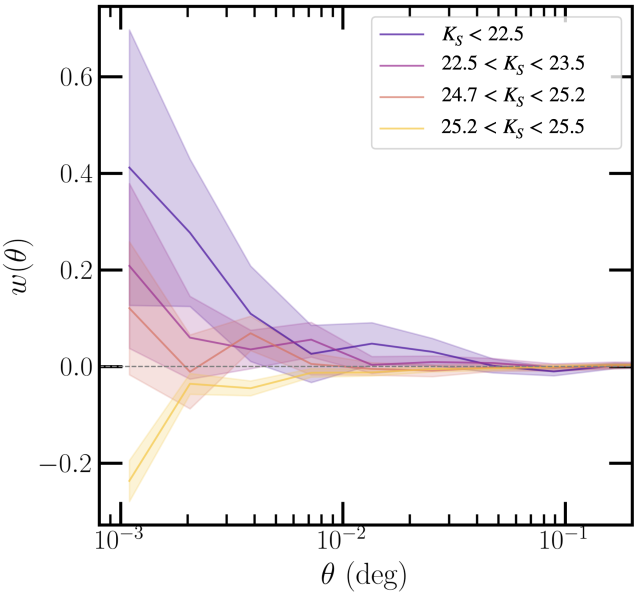


Figure 8.3: Dependence of the magnification bias measurements on the magnitude selection of the background source sample. We show the cross-correlation function between lenses at $0.5 < z < 0.8$ and sources at $2.7 < z < 4.0$ selected in four different magnitude bins. The magnitude bins correspond to slopes $\alpha > 3.5$ (violet), ~ 2.5 (purple) $\alpha = 1$ (red) and $\alpha < 1$ (yellow)

Check for systematic errors

The source of most severe systematic errors that can bias the magnification measurement comes from leaking clustering signal. In our sample, this can happen from interlopers between the lens and source sample due to catastrophic z_{phot} (§ 3.2.4). We expect this to be negligible in COSMOS2020 due to the accurate z_{phot} . The fraction of z_{phot} outliers in the full sample, as mentioned in § 7.2.1, is $\sim 10\%$. Out of this, the fraction of interlopers, i.e., the fraction of objects selected in the source z_{phot} bin but have their true redshift in the lens sample and vice versa, in our case is $< 0.2\%$. Such a fraction of interlopers is not expected to contribute to significant biases in the magnification signal. We do a quantitative analysis to demonstrate this in Chapter 9.

A convincing check of whether the magnification measurement is clean from the leaking clustering term is to investigate the w_{μ}^{LS} dependence on the magnitude of the background sample. As shown in Eqn. 8.1, the amplitude depends on the slope. Therefore, if the measured magnification amplitude is zero for a background sample having $\alpha = 1$, and a negative amplitude if $\alpha < 1$, then that is a proof that the measurements do not suffer from the interlopers clustering. If clustering is present, then the amplitude tends strongly towards higher values. In Fig. 8.3 we show magnification measurements for background samples selected in four different magnitude bins that correspond to slopes of $\alpha > 3.5$ (violet), ~ 2.5 (purple) $\alpha \sim 1$ (red) and $\alpha < 1$ (yellow). The $\alpha = 1$ measurements are consistent with zeros, whereas for $\alpha < 1$ the correlations are clearly negative. This proves that our lens/source selection has minimal interlopers that do not catastrophically bias the magnification measurements.

8.4 JOINT FITS OF MAGNIFICATION WITH CLUSTERING AND GSMF TO CONSTRAIN THE SHMR

To explore the magnification constraining power, we perform a joint fit of magnification, clustering and abundance in the four redshift bins up to $z = 1.5$. We do joint fitting of Magnification+GSMF and Magnification+Clustering+GSMF. We show in Fig. 8.5 the resulting posterior distributions for the SHMR parameters from the magnification joint fits along with the Clustering+GSMF fit from § 7.5 in $0.5 < z < 0.8$. Due to the relatively low S/N of magnification, the Magnification+GSMF posteriors (in orange) are broader and the parameter are rather poorly constrained. The results are consistent when using all three probes, with the results of Clustering+SMF. The much higher S/N measurements of clustering and GSMF dominate the constraining power. These results show that our magnification measurements don't bring a significant contribution in constraining the SHMR. This is unsurprising because the signal-to-noise (S/N) of the magnification measurements is considerably lower than e.g. the clustering measurements or what one would measure with shear-based probes, as discussed in § 3.2.5. The low S/N limits the constraining power of magnification for a relatively small survey area (according to lensing standards) and consequently low number of sources ~ 3700.

The resulting best-fit models from the joint Magnification+Clustering+SMF fit are shown in the solid lines in Fig. 8.4. The magnification models are in a reasonable agreement with the measurements. In the lowest two z-bins there is an extra power in the measurements at small scales, and an excess of power at the largest scales in all bins. The small-scale discrepancy can come from the modeling of the halo profile. In the case of magnification, it is the dark matter halo profile that enters the equations (see Eq. 3.63) whereas for clustering is the over-density profile of satellites (see Eq. 3.13). In our analysis, we are assuming that both follow the same NFW profile, which is not necessarily true. With the magnification model, one can relax this assumption and set the parameters defining the halo profile free. Fitting for the dark matter halo profile parameters using magnification, one can allow halos to have different density profiles compared to the density profile of satellites. A change of the halo profile will impact more strongly the one-halo term governing the correlation amplitude at small scales. Performing such a fit would allow reconciling the models and the measurements at small scales and constrain additional parameters concerning the dark matter halos. Even though we explored the effects of releasing this degree of freedom on the magnification models, we did not perform a formal fit for the halo profile parameters, and we leave this avenue open for future works.

8.5 SUMMARY AND CONCLUSIONS

In this Section, we presented measurements of the magnification bias through the two-point cross-correlation function. We successfully measured the magnification bias in four tomographic bins down to $z \sim 1.3$, and for stellar-mass selected samples. By investigat-

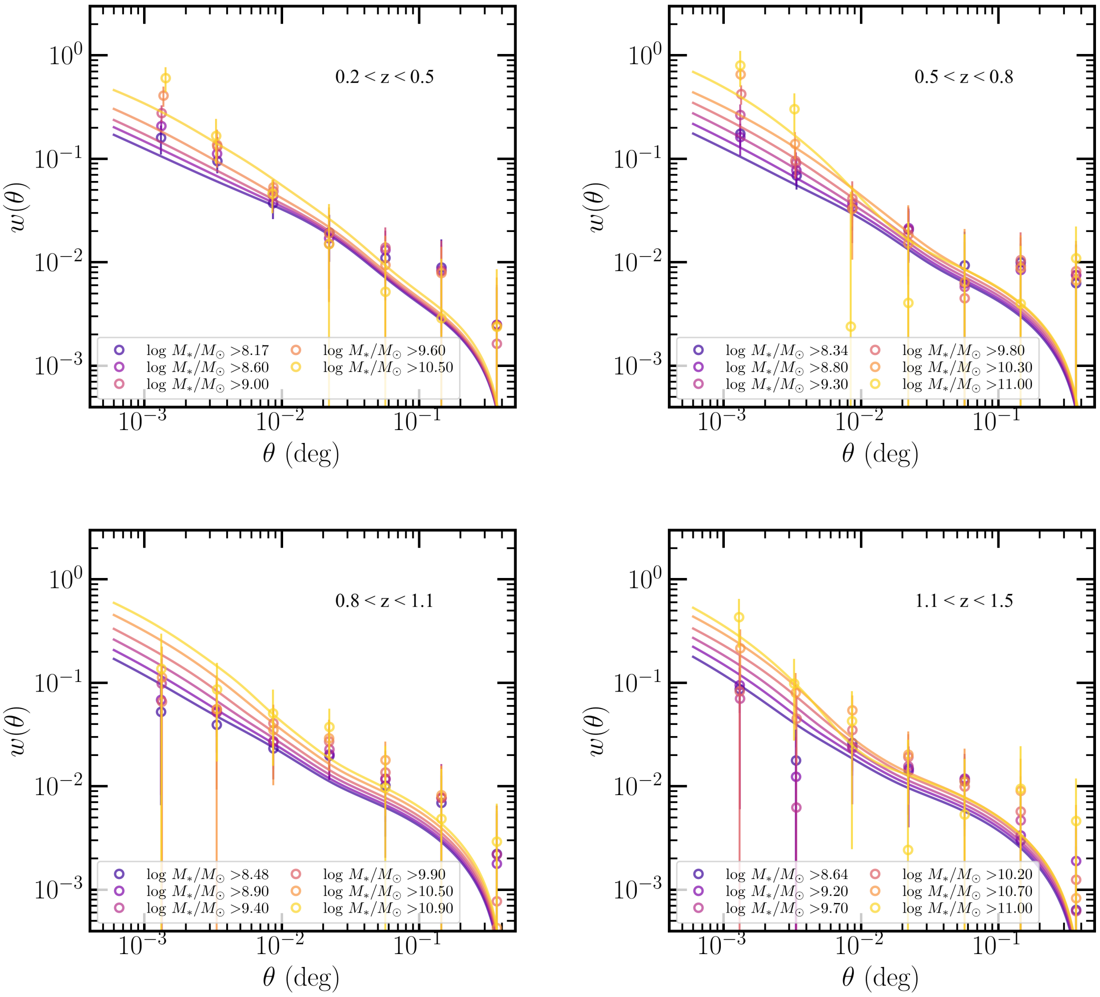


Figure 8.4: Magnification correlation measurements with best-fit models. Each panel shows the 2-point angular cross correlation due to magnification for foreground samples in each of the four redshift bins. Measurements are shown for each stellar mass threshold and are color coded accordingly. The solid lines show the best-fit models of the joint Magnification+Clustering+SMF fit.

ing the magnification correlation amplitude dependence on the number counts slope of the background sample, we showed that the magnification signal is not dominated by systematic errors coming from redshift interlopers. We performed a joint fit of Magnification+GSMF and Magnification+Clustering+GSMF using the HOD modeling with the parametrization of the SHMR as a starting point that we presented in Chapter 3 and Chapter 8. We showed that the best-fit models when including magnification in the fit is in reasonable agreement with the measurements. However, the small survey size of COSMOS does not allow for high S/N measurements and, consequently, magnification does not bring a significant constraining power on the SHMR model parameters. Nonetheless, this demonstrates that magnification can be successfully measured for normal galaxies in a photometric redshift catalog. Being a lensing probe and thus a direct probe of halos, it can be used to open other degrees of freedom in the modeling. For example, it can probe the galaxy distribution within halos. Here we measured magnification in a real data set

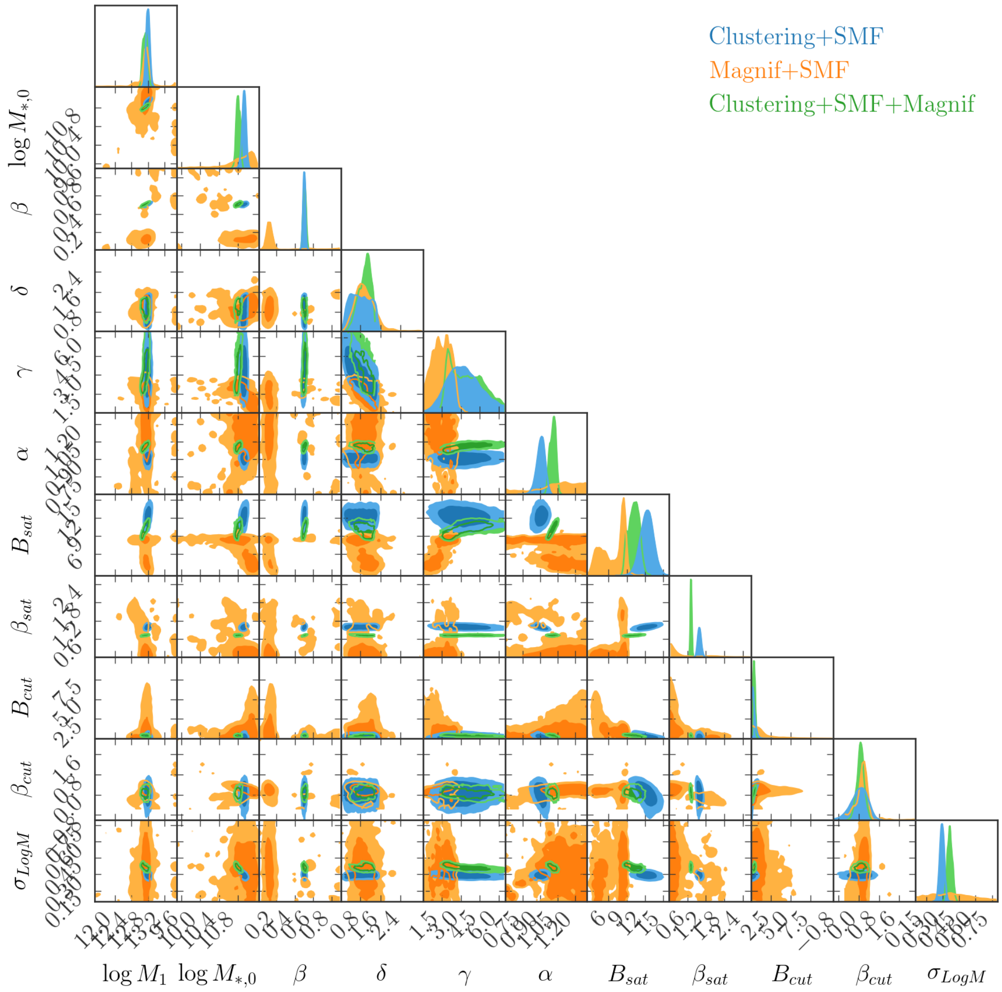


Figure 8.5: Corner plot of the posterior distribution of the parameters from joints fits in $0.5 < z < 0.8$. For comparison, we show joint fits from Magnification+GSMF, Clustering+GSMF and Magnification+Clustering+GSMF.

which sets the scene for its use in the upcoming deep and wide photometric surveys, such as Euclid Deep, which we explore in detail in Chapter 9.

9

PROSPECTS OF MAGNIFICATION IN THE EUCLID DEEP SURVEY

9.1 INTRODUCTION

In the search for a lensing probe of the dark matter halo properties, magnification offers a promising opportunity. The previous chapters described how the magnification bias can be used to probe halo properties and how does its signal-to-noise ratio compare to shear based probes. These considerations, coupled with the first measurements of magnification bias of normal galaxies in COSMOS2020, showed that to fully exploit its potential one needs large and deep photometric surveys.

One of the crucial ingredients necessary to successfully measure magnification for normal galaxies is quality photometric redshifts. Additionally, our interest is using magnification to probe halo properties as a function of redshift, especially in the redshift regime where shear-based probes become increasingly difficult, i.e., $z > 1$. Achieving this typically requires deep observations in as many bands as possible from the observer optical to IR part of the spectrum. This is time-consuming, and poses a significant challenge for wide surveys, so much so, that so far this has only been done for scale degree surveys (e.g., COSMOS). However, upcoming next generation surveys are designed to tackle exactly this issue.

The *Euclid* mission will comprise a Deep survey, in support for its cosmological core science Wide survey. The Euclid Deep Survey is designed to obtain deep NIR photometry in 3 bands and one VIS band over ~ 50 deg^2. Additionally, it will be complemented by deep ground based observations in the optical bands by some of the biggest facilities (current and upcoming) on ground such as the Vera C. Rubin Observatory and Subaru/Hyper SuprimeCam. This massive undertaking will provide a wealth of legacy science opportunities in galaxy formation and evolution, and will provide the perfect grounds for the exploitation of magnification.

In this chapter of the thesis, we analyze in detail the prospects of magnification in the Euclid Deep Survey. We rely on the Euclid FLAGSHIP2.0 simulation to construct realistic samples of galaxies for magnification measurements. Along with simulations of the photo-zs in a Euclid Deep survey strategy, we investigate how well magnification can be measured and what are the contributions from the different sources of systematic errors. This chapter is organized as follows. In § 9.2, we present the Euclid Deep Survey, the FLAGSHIP2.0 simulation and the preparation of its mock catalog for the purposes of this work.

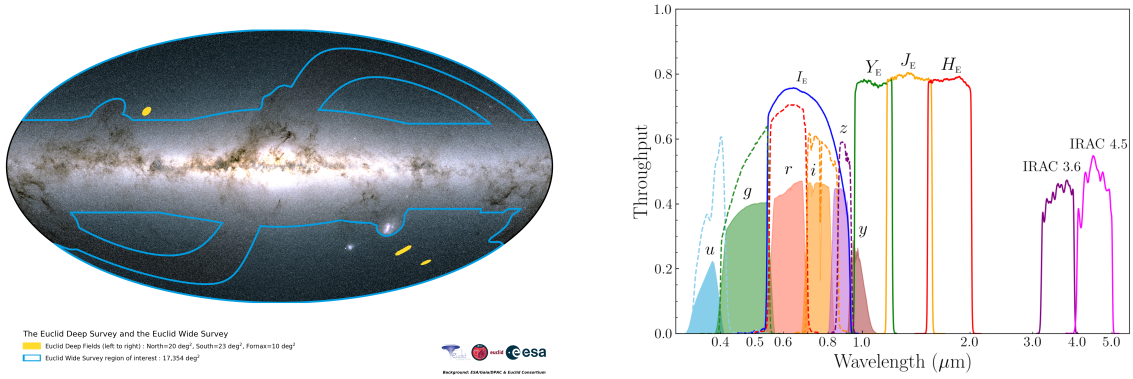


Figure 9.1: **Left:** *Euclid* survey: Wide, marked in blue and Deep, yellow regions. **Right:** Photometric broad band filters from Euclid, along with those from Rubin (filled, solid curves), Subaru/HSC (open, dashed curves) and *Spitzer*/IRAC (adopted from van Mierlo et al. submitted).

In § 9.3 we present the main results for the measurement of magnification for different sample selections, and we investigate the performance of using a photo-z based selection. Finally, in § 9.4 we discuss and summarize our results.

9.2 DATA AND METHODOLOGY

9.2.1 *Euclid Deep Survey*

The *Euclid* mission is designed with a core scientific goal to understand the accelerated expansion of the Universe and investigate the nature of dark energy and dark matter via weak lensing and galaxy clustering (Laureijs et al., 2011). To this end, the *Euclid* survey is designed to image about 15 000 deg^2 of the extragalactic sky over six years (Euclid Collaboration et al., 2022a, for a detailed description of the *Euclid* survey). The telescope is equipped with a VIS and NISP. The imager will obtain broad-band photometry in one very broad visible filter (I_{E}) and three near-infrared filters (Y_{E}, J_{E}, H_{E}). For its core science goal, *Euclid* will carry out a wide imaging and spectroscopy survey. It will achieve an image quality of about 0.18 arcsecond for the VIS band (0.1"/pixel) and 0.45 arcsecond for the near infrared (0.3"/pixel). Coupled with the large field of view (0.5 deg^2) and a design optimized to result in a pristine PSF, the VIS imager will provide exceptional quality for the images of galaxies, necessary to achieve its weak lensing requirements. Spectroscopic redshifts will be measured with in a slitless low-resolution near-infrared spectroscopy mode, that will provide the main sample for accurate clustering measurements. To enable the primary goal via characterization of the target galaxy populations, quantification of biases that arise in lensing and clustering measurements and calibration of photometric redshifts, *Euclid* will also survey about 50 deg^2 of deep and auxiliary fields. To do this, *Euclid* will have to achieve more than 40 times longer exposure per pixel. This will result in the Euclid Deep Survey (EDS), that will transcend the core mission goals and enable unprecedented galaxy formation and evolution legacy science.

The Euclid Deep Survey will consist of three deep fields: 1) Euclid Deep Field North (EDF-N), 20 deg^2 region centered on the north Ecliptic pole; 2) Euclid Deep Field Fornax (EDF-F), of 10 deg^2, centered on the Chandra Deep Field South; and 3) the Euclid Deep Field South (EDF-S), a completely new field of 20 deg^2 with no previous observations. EDS will go two magnitudes deeper than the Wide survey with 5σ depths of 28.2, 26.3, 26.5 and 26.4 AB in I_{E}, Y_{E}, J_{E} and H_{E} respectively. To put this into perspective, the EDS will be about 30 times larger and one magnitude deeper in NIR than COSMOS. Fig. 9.1 shows the Euclid Wide and Deep Survey footprints (right), and the photometric filters (left) accompanied by the Rubin, Subaru/HSC and Spitzer filters.

To maximize the scientific return of the mission and to enable science goals beyond the capability of Euclid alone, data products from other surveys will be combined. Here we give a brief overview of the more relevant for our work.

- Vera C. Rubin Observatory Legacy Survey of Space and Time in the optical. The Rubin observatory will observe the southern Euclid Deep Fields in u, g, r, i, z, y bands down to 5σ depth of AB 26.1, 27.4, 27.5, 26.8, 26.1, 24.9 respectively, and down to AB 28 in the deep drilling fields, after 10 years of observation. This synergy between Rubin and Euclid will enable a dense and deep sampling of the galaxy SED in the range of $0.3 - 2\,\mu$m that will be crucial in accurately measuring photo-zs and stellar masses (Guy et al., 2022).

- Cosmic Dawn Survey: Hawaii-Two-0 (H20) Subaru telescope/Hyper SuprimeCam (HSC) survey in the optical. The H20 survey will observe 10 deg^2 of EDF-N and EDF-F each in the g, r, i, z filters down to 5σ depth of AB 27.5, 27.5, 27.0, 26.0 respectively (McPartland et al., in prep). In addition, u band data down to a 5σ depth of AB 26.2 will be obtained by CHFTT. These bands will be key in sampling the UV-optical portion of the galaxy SED beyond cosmic noon ($z > 2$) and determining Lyman break photo-z. The H20 survey, even though overlapping with LSST, will have the data timely available for Euclid, long before LSST can achieve its depth, and is the only survey to provide data in the EDF-N.

- Cosmic Dawn Survey: *Spitzer* observations in the infrared. New observations and all relevant archival data from *Spitzer*/IRAC in the Euclid Deep and Auxiliary Fields have been reprocessed to provide photometry at $3 - 5\,\mu$m down to AB ~ 25 (Moneti et al., 2021). Crucially, IRAC photometry samples the rest frame optical portion of the SED beyond the Balmer break for $z > 3.5$ galaxies that will be essential to obtain accurate stellar masses and minimize the Lyman/Balmer break degeneracy in the photo-z estimation.

These joint data products will enable the spanning of a large wavelength range to sample the galaxy SED from the UV to IR and maximize the accuracy in measuring photometric redshifts and physical properties of galaxies. As a result, numerous galaxy formation and evolution studies are currently being prepared that consider the contribution of all of these data. In this chapter, for the purposes of investigating the prospects of magnification in the EDS we will also consider the inclusion of this complementary data.

9.2.2 *Euclid Flagship2.0 simulation*

To prepare the panoply of science that will be enabled by Euclid, the consortium has produced the FLAGSHIP simulation – the largest N-body dark matter simulation (Potter, Stadel and Teyssier, 2017). In this work we rely on the latest run, the FLAGSHIP2.0, which we briefly describe. The simulation is run with a flat ΛCDM cosmology with $\Omega_{\mathrm{m},0} = 0.319$, $\Omega_{\mathrm{b},0} = 0.049$, $\Omega_{\Lambda,0} = 0.681$, $\sigma_8 = 0.83$, $n_\mathrm{s} = 0.96$, $h = 0.67$. It is run with 4.1×10^{12} DM particles in a box of 3600 Mpc/h a side, with DM particle mass of $10^9\, M_\odot/h$. There are Wide and Deep runs designed for the wide and deep surveys. In this work we use the Wide run which provides an all-sky 3D light cone up to $z = 3$. We note that this poses a limitation, since we are interested in pushing magnification to the highest redshifts, but the Wide run was the only one available at the time, and still provides a useful redshift range to test the methods. FLAGSHIP2.0 has been downloaded from CosmoHub (Carretero et al., 2017; Tallada et al., 2020)

Dark matter halos are identified with the `ROCKSTAR` algorithm (Behroozi, Wechsler and Wu, 2013), while galaxies are assigned to halos using HOD and SubHalo abundance matching techniques. These are calibrated against a compilation of observations of the luminosity function, galaxy clustering as a function of luminosity and color and color-magnitude diagrams at all redshifts. SEDs are assigned to the galaxies that are interpolated from those measured in the COSMOS survey, giving rise to a range of physical parameters and simulated photometry for many bands from different observatories. The lensing is implemented in the Born approximation with 0.4 arcmin resolution. The resulting mock galaxy catalog of one full octant contains about 3.65×10^9 galaxies.

9.2.3 *Preparation of the* FLAGSHIP2.0 *mock galaxy catalog*

The goal of this work is to investigate how well the magnification bias can be measured with the two-point angular cross-correlation by implementing a realistic Euclid survey scenario. The ultimate goal of such measurements is to use them to estimate halo properties for the foreground sample. As such, we will focus on implementing a sample selection based on photometric redshifts and physical parameters, such as stellar mass. Photometric redshifts in the Euclid Deep survey can be obtained with several survey strategies, such as combining the Euclid NIR bands with the optical from LSST or the Cosmic Dawn Survey/H20. In this work, we choose to adopt the Euclid Deep + H20 survey strategy because the H20 survey will be complete even before the Euclid Deep is, whereas LSST will take at least 10 years of observing.

To emulate the galaxy density of EDS, we select a contiguous area with $152^\circ < \mathrm{RA} < 156^\circ$ and $20^\circ < \mathrm{Dec} < 25^\circ$, resulting in $\sim 20\ \mathrm{deg}^2$ to work with. This selection is contiguous, which will not be the case with the real Euclid Deep + H20 that will contain two 10 deg^2 fields. This can have a potential effect on the correlation function measurements on very large scales, but those regimes are not the subject of this work. A caveat of this selection is that in the FLAGSHIP2.0, H_E band magnitudes are sharply cut at AB $\sim$ 26. We account for this in the analysis by always applying a cut at $H_\mathrm{E} = 26$.

At the time of this work, the FLAGSHIP2.0 mock galaxy catalog does not provide added value data such as photometric redshifts derived from SED fitting. To address this, we simulate our own redshifts to reproduce the errors in the photometric redshifts that are relevant for magnification measurements. These mostly include interlopers – sources that based on their measured redshift are selected in the foreground (background) sample but have their true redshift in the background (foreground) sample. To investigate the effect of photometric redshift errors on the magnification signal, we proceed in two ways. 1) We introduce a given fraction of interlopers by perturbing the true redshifts from the source to the lens population (and vice versa). 2) We simulate flux errors and run a SED-fitting code to obtain photometric redshifts and estimate the expected number of interlopers.

Simulation of interlopers

To simulate the interlopers, we proceed in the following way. We first select the lens and source sample based on their true redshifts, resulting in N^0_lens, N^0_src initial number of objects in each sample. Then we randomly draw $N^\mathrm{interlop}_\mathrm{lens}$, $N^\mathrm{interlop}_\mathrm{src}$ of objects in both the lens and source sample which we denote $\mathcal{L}_\mathrm{interlop}$ and $\mathcal{S}_\mathrm{interlop}$ respectively. Then for the lens→source interlopers, $\mathcal{L}_\mathrm{interlop}$, we assign redshifts drawn randomly from a uniform distribution in the redshift range of the source sample. Similarly, for the source→lens interlopers, $\mathcal{S}_\mathrm{interlop}$, we assign random redshifts from a uniform distribution in the redshift range of the lens sample. The final definition of the fraction of interlopers is

$$f_\mathrm{interlop} = \frac{N^\mathrm{interlop}_\mathrm{lens} + N^\mathrm{interlop}_\mathrm{src}}{N^0_\mathrm{lens} + N^0_\mathrm{src}}. \tag{9.1}$$

Fig. 9.2 shows an example of a lens and source sample redshift distribution. The lenses

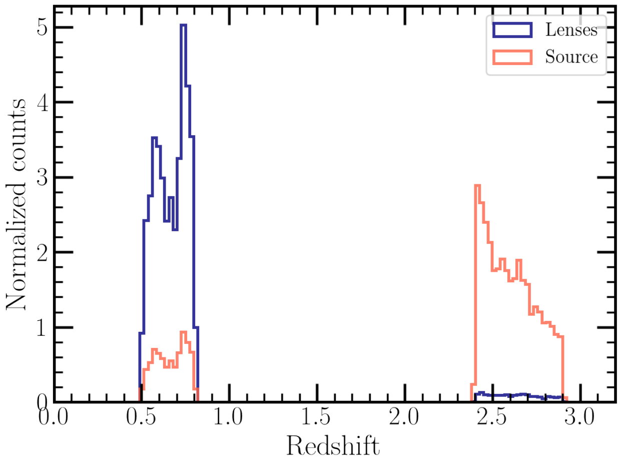


Figure 9.2: Redshift distribution of lens and source samples, including a total of 10% of interlopers

are selected based on their true redshifts in $0.5 < z < 0.8$ and the sources in $2.4 < z < 2.9$. After introducing a fraction of interlopers $f_\mathrm{interlop} = 10\%$, there is a mix in the redshift distribution between the two populations. As described in § 3.2.4, this will introduce clustering signal in the cross-correlation measurements. However, it has to be noted that in a real survey, where the selection is typically made on photo-z, the real redshift distributions of the two populations will not have such clear cuts in the z-bins, instead, they will have smoother distributions. In this case, the interloper fraction would have to be computed from the overlap of the two distributions (as we will see in § 9.3.5). Nonetheless, the approach outlined here is still useful in order to investigate the effect of f_{interlop} on the measured magnification signal.

Simulation of photometric errors

In its current version, the Flagship2.0 mock catalog comes only with true values for the fluxes (including internal and Milky Way extinction), and no errors in the photometry. In order to estimate photometric redshifts with SED-fitting, we simulate photometric errors in the following way. We rely on the simulated Euclid and H20 photometry from the Horizon-AGN virtual observatory (Laigle et al., 2019) to reconstruct the magnitude error vs. magnitude distribution, from which we randomly draw the errors for the Flagship2.0 mock.

The Horizon-AGN virtual observatory is a realistic photometric catalog based on the Horizon-AGN hydrodynamical simulation (Dubois et al., 2014; Kaviraj et al., 2017). The photometry is consistently processed in the hydrodynamical lightcone from the realistic SEDs of a large diversity of galaxy star-formation histories, metallicity enrichment, dust and IGM absorption. From these SEDs, the photometry is computed and errors are added such that they reproduce the S/N distribution and sensitivity limit of an Euclid Deep + H20 survey. The errors are derived assuming a Gaussian noise, where the standard deviation is given by the quoted depths of the survey in each band. The errors do not include the Poisson noise and potential systematic effects like blending, object fragmentation, imperfect background sky subtraction and offsets due to the rescaling of fixed aperture to total fluxes. Finally, the fluxes are perturbed according to their 1σ error. The resulting magnitude number counts of Horizon-AGN in the Euclid + H20 survey bands are shown in Fig. 9.3.

To simulate photometric errors for the Flagship2.0 mock catalog, we proceed in the following way. For each band, we split the Horizon-AGN dataset in small magnitude bins, and for each bin we fit the magnitude error distribution with a Gaussian mixture model. Then in Flagship2.0, for each magnitude, we draw its error from the fitted Gaussian mixture model in the corresponding magnitude bin. Finally, we perturb the magnitudes by drawing from a Gaussian with a standard deviation equal to their 1σ error. The resulting magnitude error as a function of magnitude for all 8 bands of a Euclid Deep + H20 survey is shown in Fig. 9.4. We consider the following bands: u from CFHT, g, r, i, z from HSC and $Y_{\mathrm{E}}, J_{\mathrm{E}}, H_{\mathrm{E}}$ from Euclid, that we are going to use in the photo-z estimation. Fig. 9.5 and Fig. 9.6 show the comparison between the perturbed and the true magnitude as a function of the true magnitude. The reason the scatter plot looks stripy is because of the finite size of the magnitude bins used in the error simulation. The sharp decrease in point density

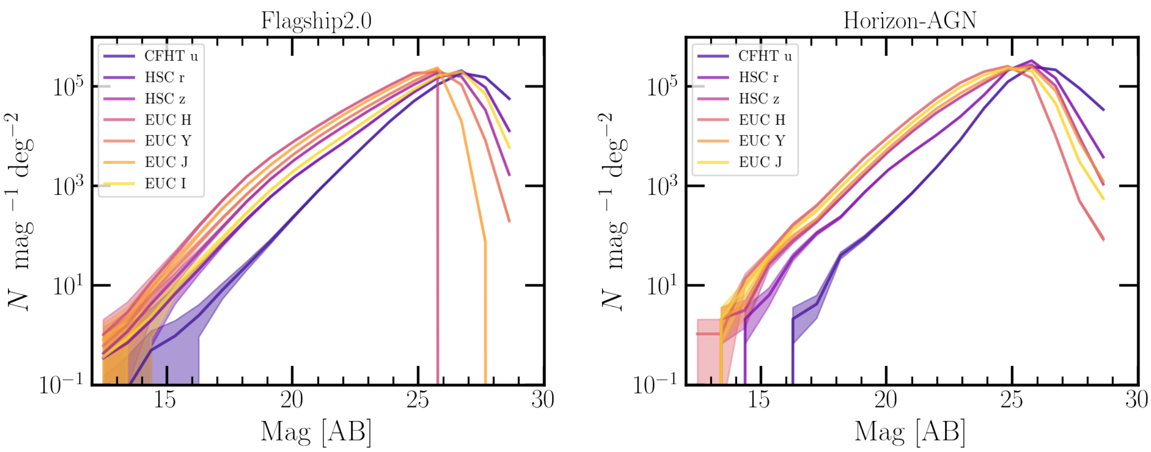


Figure 9.3: Magnitude number counts in Flagship2.0 (left) and Horizon-AGN (right) virtual observatory in the bands corresponding to an Euclid + Cosmic Dawn/H20 survey. For Flagship2.0 this is made using the true magnitudes, whereas for Horizon-AGN using the perturbed ones.

in the plots displaying the Euclid bands, especially the H_{E} band, is because of the sharp drop in the magnitude number counts in Flagship2.0, as shown in Fig. 9.2. We use these perturbed magnitudes along with their errors in the estimation of the photo-zs.

Photometric redshifts

We estimate photometric redshift using the perturbed magnitudes described in the previous section with the SED-fitting code `EAZY` (Brammer, van Dokkum and Coppi, 2008). `EAZY` fits a non-negative linear combination of a set of basis templates to the observed flux densities for each galaxy. The fit is performed on the perturbed magnitudes and their errors, converted to flux units internally in the Euclid Deep ($Y_{\mathrm{E}}, J_{\mathrm{E}}, H_{\mathrm{E}}$) + H20 ($u$ from CFHT and g, r, i, z from HSC) survey bands. We use the template set `"tweak_fsps_QSF_1_v3.param"` that is derived from the Flexible Stellar Population Synthesis models (FSPS) (Conroy, Gunn and White, 2009; Conroy and Gunn, 2010). The allowed redshift range is set to $0.01 < z < 4$, with a flat prior on the r band magnitude. The comparison of the resulting photometric, and the true redshifts is shown in Fig. 9.7. The left panel shows the true vs. photometric redshifts for all galaxies having $I_{\mathrm{E}} < 26$. This cut corresponds to the completeness of the mock catalog. The right panel shows the redshift distributions of both. The quality of the photometric redshifts is typically quantified with the normalized median absolute deviation, defined as

$$\sigma_{\mathrm{NMAD}} = 1.48 \times \text{ median} \left[\frac{|\Delta z - \text{median}(\Delta z)|}{1 + z_{\mathrm{true}}} \right], \tag{9.2}$$

where $\Delta z = z_{\mathrm{phot}} - z_{\mathrm{true}}$. For the sample selected at $I_{\mathrm{E}} < 26$ we obtain $\sigma_{\mathrm{NMAD}} = 0.051$. This corresponds to the σ_{NMAD} obtained for `EAZY` in the study of Desprez et al. (2020), who tested the photometric redshift performance of the Euclid survey with different codes. The redshift distributions, shown in the right panel, are in a reasonable agreement, with the most notable difference in the two spikes at $z < 2$ where the code fixes the photo-

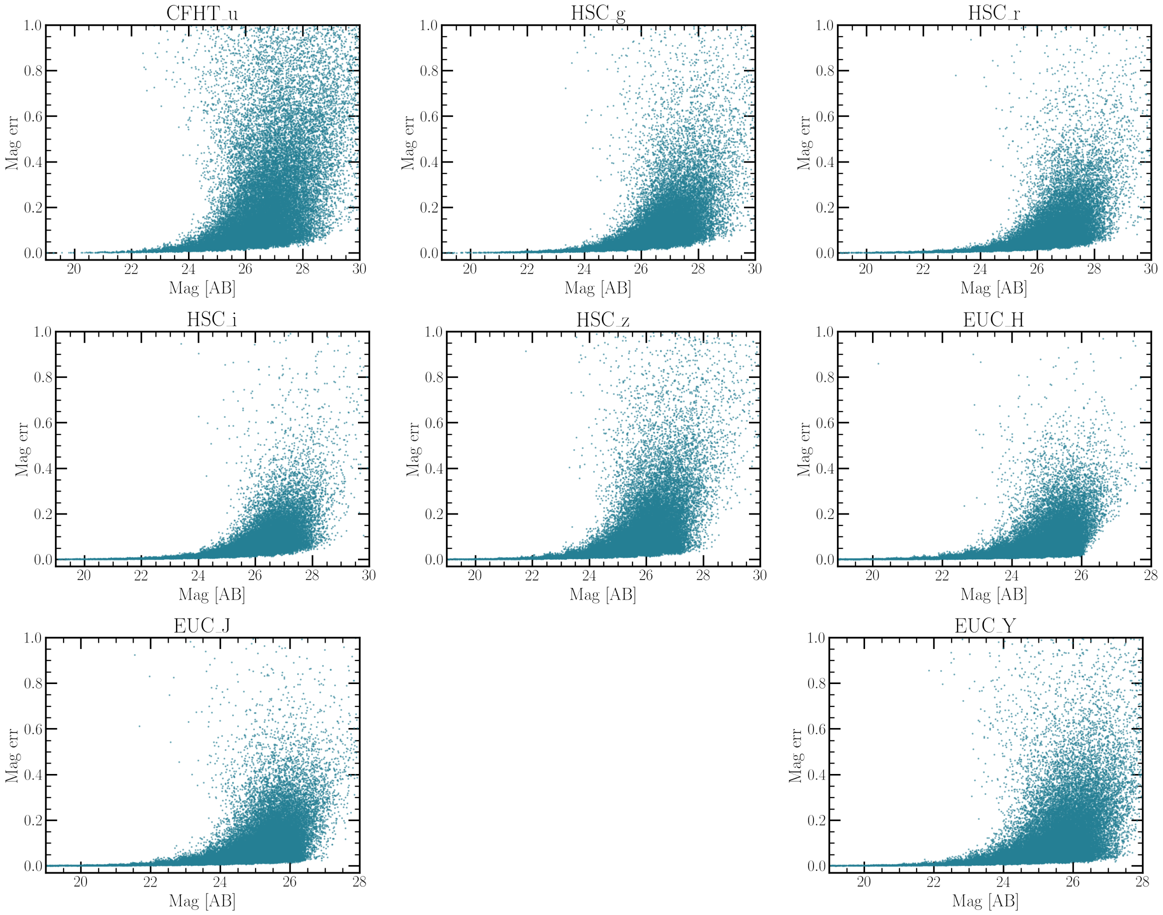


Figure 9.4: Simulated magnitude error vs. magnitude in the Euclid Deep ($Y_{\rm E}$, $J_{\rm E}$, $H_{\rm E}$) + H20 (u from CFHT and g, r, i, z from HSC) survey bands.

z solution at these particular values for certain galaxies. We use these photo-z to do a realistic selection of the foreground and background samples and investigate its effect on the measured magnification correlations.

9.2.4 *Optimal weighting of the correlation function*

As showed in the previous sections, in order to maximize the S/N one needs to apply a cut at relatively bright magnitude, which excludes a significant number of sources from the analysis. Ménard and Bartelmann (2002) showed that the S/N can further be increased by *optimally weighting* the data when computing the correlation function with $\alpha(m) - 1$, resulting in a second-order dependence of the lensing signal on magnitude, i.e., $[\alpha(m) - 1]^2$ (Ménard and Bartelmann, 2002; Scranton et al., 2005). In this way, the magnification signal is weighted proportionally to the expectation from its magnitude dependence – bright sources that exhibit positive correlations are positively weighted, while faint sources

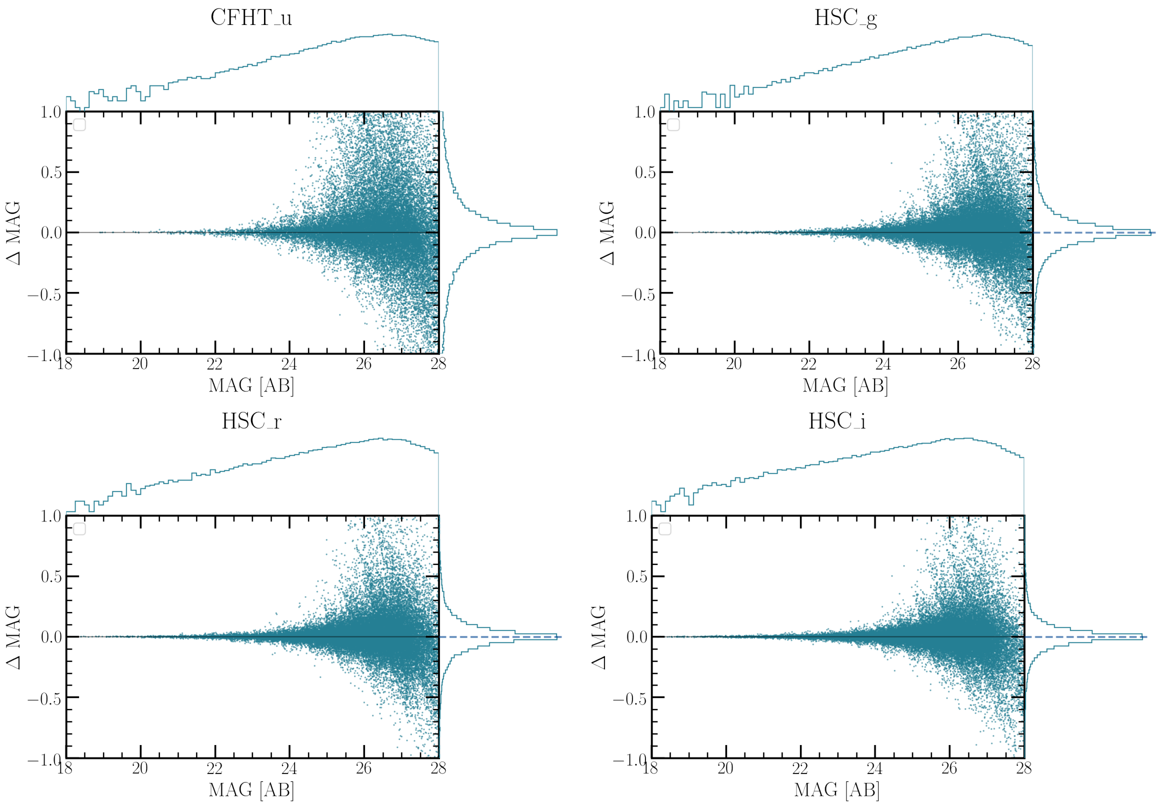


Figure 9.5: Comparison of the true and perturbed magnitudes using the simulated photometric errors. This figure presents the magnitude comparison for u band from CFHT and g, r, i bands from HSC. In each panel, the histogram on the top corresponds to the magnitude number counts and the histogram on the right corresponds to the ΔMAG = MAG$_{\text{pert}}$-MAG$_{\text{true}}$ distribution.

for which the correlations are negative have a negative weight. This corresponds to the following relation (Scranton et al., 2005)

$$\begin{aligned} w_\mu^{\text{optimal}}(\theta) &= \left\langle [\alpha(m) - 1]^2 \right\rangle w_\mu^0(\theta) \\ &= \langle \alpha - 1 \rangle_{\text{eff}}\, w_\mu^0(\theta), \end{aligned} \tag{9.3}$$

where we note $w_\mu^0(\theta)$ the underlying lensing signal. The second-order term can be expressed as an effective slope that can be computed as

$$\langle \alpha - 1 \rangle_{\text{eff}} = \frac{\int \mathrm{d}m\, N(m) [\alpha(m) - 1]^2}{\int \mathrm{d}m\, N(m)}. \tag{9.4}$$

This optimal weighting techniques will allow us to exponentially increase the number of galaxies by using the complete sample down to the completeness limit. It is important to note that beyond the completeness limit, it becomes virtually impossible to accurately compute the value of the slope, which will lead to severe systematic errors (Hildebrandt,

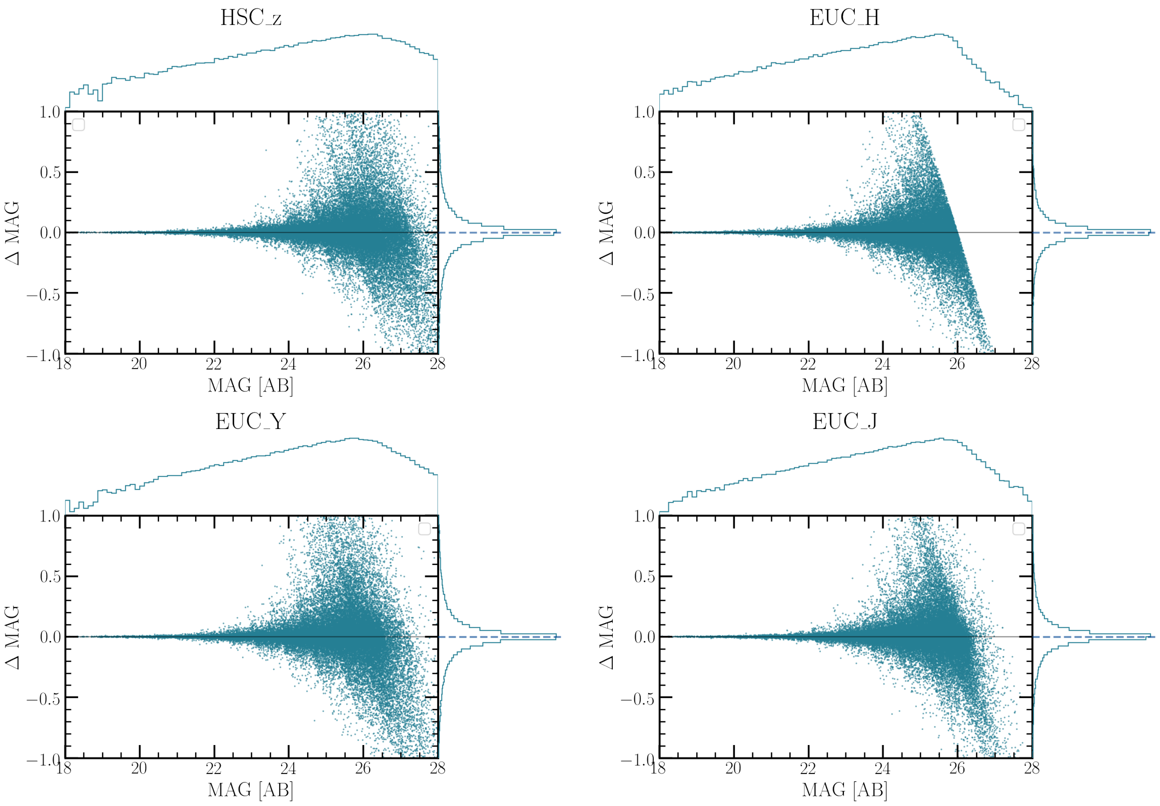


Figure 9.6: Same as Fig. 9.5 but for z band from HSC and Y_{E}, J_{E}, H_{E} bands from Euclid

2016). To apply the optimal weighting, the correlation function estimator is modified in the following way (Hildebrandt, Waerbeke and Erben, 2009)

$$w_{\mu}^{\text{optimal}}(\theta) = \frac{D_1^{\mathrm{w}} D_2 - D_1^{\mathrm{w}} R - \langle \mathrm{w} \rangle D_2 R + \langle \mathrm{w} \rangle RR}{RR}, \tag{9.5}$$

where $D_1^{w} D_2$ and $D_1^{w} R$ are the weighted pair counts with $\mathrm{w} = \alpha(m) - 1$, and $\langle \mathrm{w} \rangle$ is the average weight of the whole source sample.

9.2.5 *Selection of background (source) samples*

The amplitude of the magnification cross-correlation function depends linearly on the slope of the magnitude number counts of the background (source) sample. In the analysis that follows, the source samples are selected at $2.5 < z < 2.9$ and based on the I_{E} magnitude number counts, which are shown in Fig. 9.8 (left).

We note that the redshift selection of both the lens and source samples can be done in a more optimal way where they would be selected such that the lensing efficiency is maximized, i.e, lenses placed midway between the observer and the sources. This would further increase the S/N of the measurement, but we do not implement for simplicity.

We use the cumulative number counts to impose a magnitude cut for the source sample based on the expected S/N (see § 3.2.5. To determine this, we compute the slope as a function of magnitude along with an estimate of the signal-to-noise $S/N \propto |2.5\,\alpha - 1|\sqrt{N_S}$,

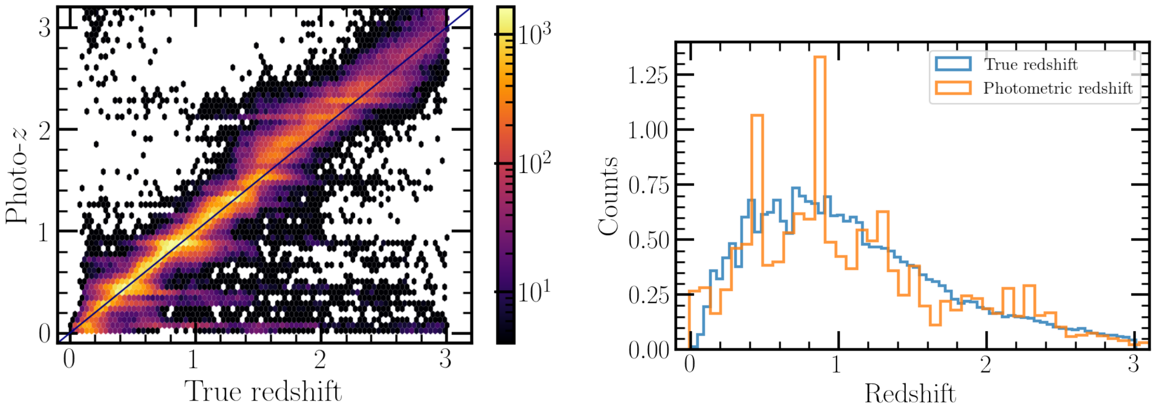


Figure 9.7: Comparison between the true and the photometric redshifts derived from SED-fitting with EAZY. **Left**: True vs. photometric redshifts for all galaxies having $I_E < 26$. **Right**: Comparison of the distributions of the true and photometric redshifts of galaxies with having $I_E < 26$

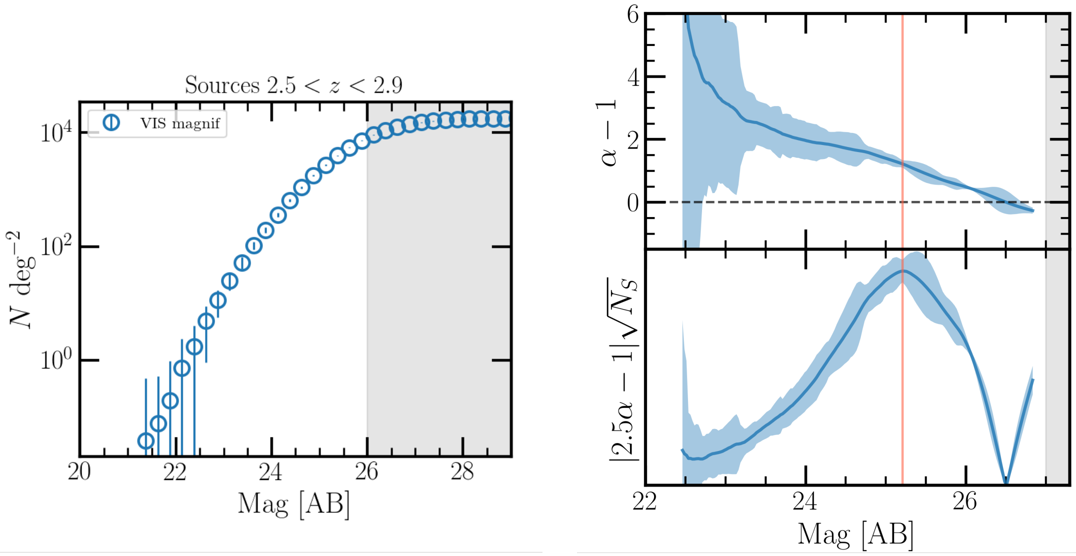


Figure 9.8: **Left:** Magnitude number counts for the source sample in I_E. **Right:** The slope of the number counts as a function of I_E magnitude (top panel) and the S/N estimate in the bottom panel. The vertical line marks the magnitude cut that we impose to maximize S/N, while the gray area marks the completeness limit.

where N_S is the number of background sources brighter than I_E. To compute $\alpha(m)$, we fit $\log N(< m)$ with a polynomial function $0.5\, am_j^2 + bm_j + c$. This function is fit in a magnitude bin j of width $m_j \pm 0.7$. From the fitted polynomial in each j-bin we compute the slope by taking the derivative as $\alpha_j = 2.5\,(am_j + b)$. Finally, we interpolate the slope at any m within the fitted range. The $\alpha - 1$ and S/N as a function of magnitude is shown in Fig. 9.8 (Right). Based on this, we select background subsamples by imposing a cut in I_E at a value that depends on its purpose. The maximum S/N is reached at $I_E = 25.3$ at a slope of $\alpha - 1 = 1.4$.

9.3 RESULTS

9.3.1 *Magnification bias as a function of lens stellar mass threshold*

We first measure the cross-correlation between the lens samples and a source sample selected at a $I_{\rm E}^{\rm cut} = 25.3$ that maximizes the S/N. This results in $N_{\rm src} = 74,376$ background sources. The foreground lens sample is selected in 5 tomographic z-bins: [0.2, 0.5], [0.5, 0.8], [0.8, 1.1], [1.1, 1.5] and [1.5, 2.0]. To measure the magnification bias dependence on lens stellar mass, we select the lenses in subsamples of three stellar mass thresholds: $\log(M_*/M_\odot) > [9.0, 9.7, 10.5]$.

Both samples are selected based on their true redshifts and magnitudes. Additionally, to compare the magnification cross-correlation to clustering, we also measure the auto-correlation function of the lenses. The clustering and magnification measurements in the five z-bins are shown in Fig. 9.9. For each z-bin and for both clustering and magnification measurements, the lenses are selected in the 3 stellar mass thresholds.

As expected, the amplitude of the angular 2PCF for clustering is higher by about two orders of magnitude compared to the magnification bias, with a considerably higher signal-to-noise of $S/N_{\rm clus} > 100$, compared to $S/N_{\rm magnif} = 5 - 18$. The amplitude for both depends on the stellar mass threshold – more massive galaxies are both more clustered and induce a higher magnification bias. As a function of redshift, the magnification bias amplitude loosely agrees with the redshift dependence of the lensing efficiency – the amplitude is highest for the lens redshift bins midway between the source sample and the observer i.e., $0.5 < z < 0.8$ and $1.1 < z < 1.5$, although this also depends on the width of the z-bin.

9.3.2 *Magnification bias as a function of source sample magnitude cut*

To investigate the dependence of the magnification bias with the magnitude cut of the background, mainly as a sanity check, we select samples on four different $I_{\rm E}$ cuts. These are chosen to capture the high slope regime $I_{\rm E} < 23.5$ ($\alpha - 1 = 2.43$), near maximum S/N $I_{\rm E} < 25.0$ ($\alpha - 1 = 1.26$), vanishing magnification bias with $\alpha = 1$ at $I_{\rm E} < 26.4$ and negative magnification bias at $I_{\rm E} < 28.0$ where $\alpha - 1 = -0.8$. We note that the latter selection is beyond the depth of the catalog and enters the incompleteness regime, so its results only serve as a qualitative check by measuring the anti-correlations.

Fig. 9.10 and Fig. 9.11 show the magnification measurements as a function of the magnitude cut for the three lens stellar mass thresholds and in each z-bin. These measurements are done using the samples selected on their true redshifts and magnitudes. For each z-bin different panels shows the magnification correlation function for lenses selected in a different stellar mass threshold. In each panel we show the correlation function for the steep slope sample, max S/N, $\alpha = 1$ and $\alpha < 1$. For all z-bins and lens stellar masses we observe the expected behavior with source magnitude cut – the amplitude decreasing as the slope decreases, with decreasing error bars corresponding to what one expects given the increase

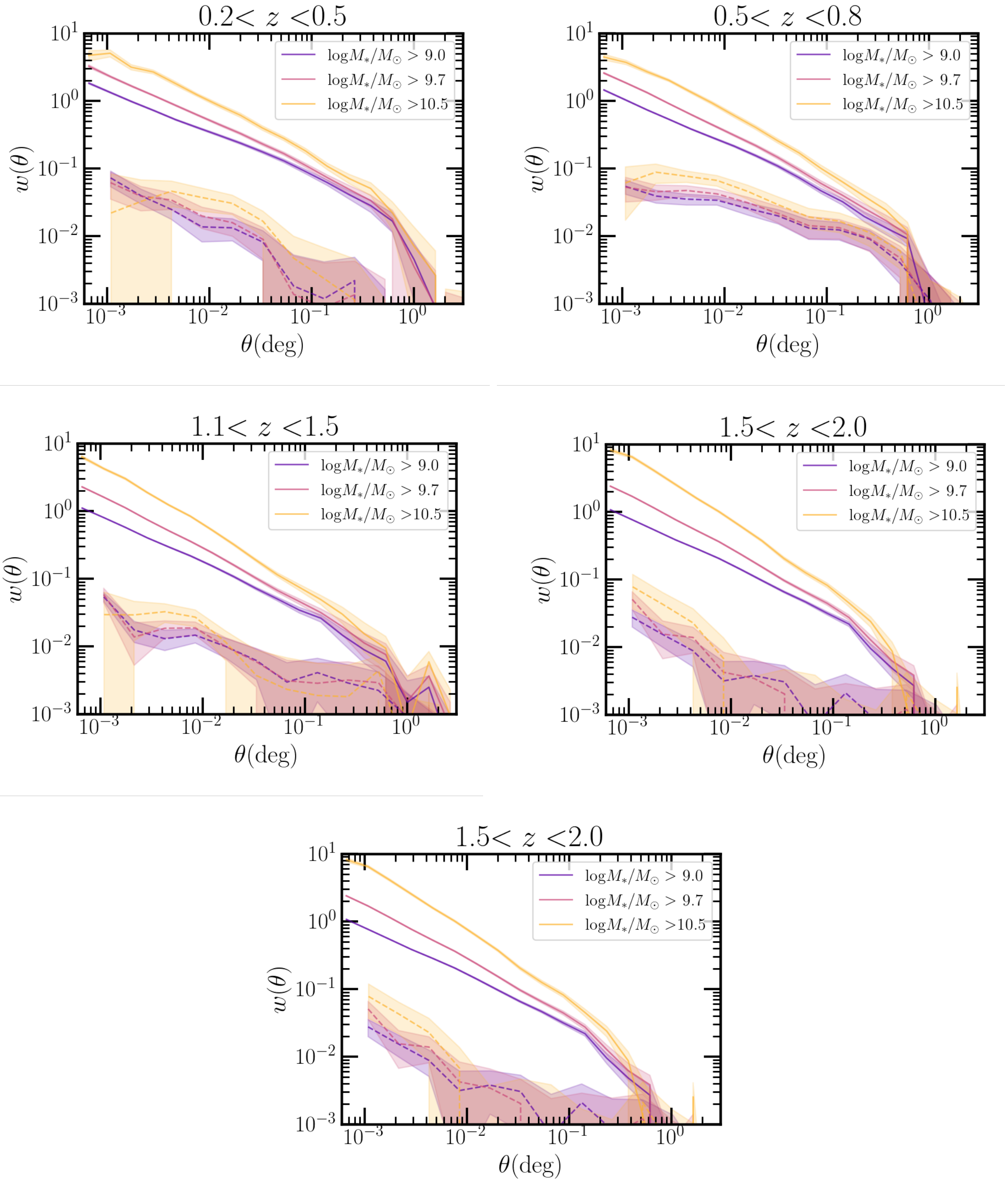


Figure 9.9: Magnification and clustering correlation measurements as a function of stellar mass for five z-bins. Each panel shows the 2PCF due to clustering (solid curves) and magnification (dashed curves) in each of the five redshift bins. Different color coded curves show the measurements for the three different stellar mass threshold selections. Magnification measurements show the expected behavior with redshift and stellar mass.

in number of sources as the magnitude cut is moved towards the faint end. As a validation of the magnification bias measurement, the correlation functions for the samples selected to have $\alpha = 1$ are consistent with zero for all lens stellar masses. Furthermore, the samples with $\alpha < 1$ show anti-correlations whose amplitude increases (in the negative direction) with increasing lens stellar mass.

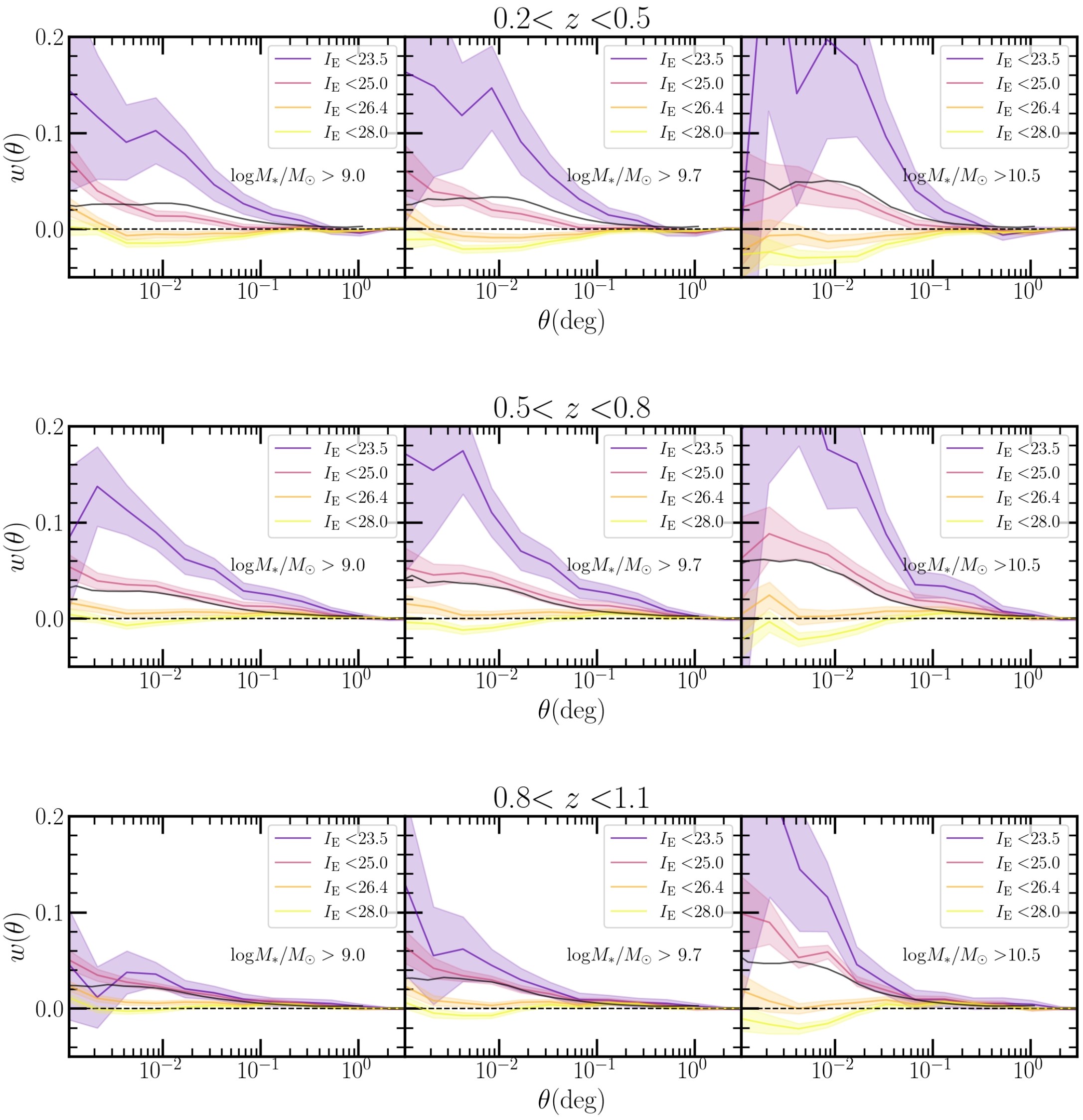


Figure 9.10: Magnification bias measurements as a function of source sample magnitude cut and stellar mass for the first three z-bins. Each panel shows the 2PCF due to magnification (dashed curves) in each of the five redshift bins. Different color coded curves show the measurements for the four different magnitude cuts of the background sample, while each panel corresponds to lenses selected at different stellar mass threshold Magnification measurements show the expected behavior with the source sample magnitude cut.

9.3.3 *Validation with the true lensing signal*

To validate the true lensing nature of the magnification measurements, we compare with the ground truth convergence profile. These are taken from the mock catalog described in § 9.2.2. To do this, we measure the convergence profile $\kappa(\theta)$ around the same lens sub-samples (redshift and stellar mass selected) that we use to measure the magnification cross-correlation. For every lens in the sub-sample l, where $l = 1, ..., N_{\text{lens}}$, we compute the mean $\kappa^l(\theta_i)$ in an annulus of $\theta_i \pm \mathrm{d}\theta$ for $i = 20$ radial bins from 0.0005 to 1.3 deg. For each

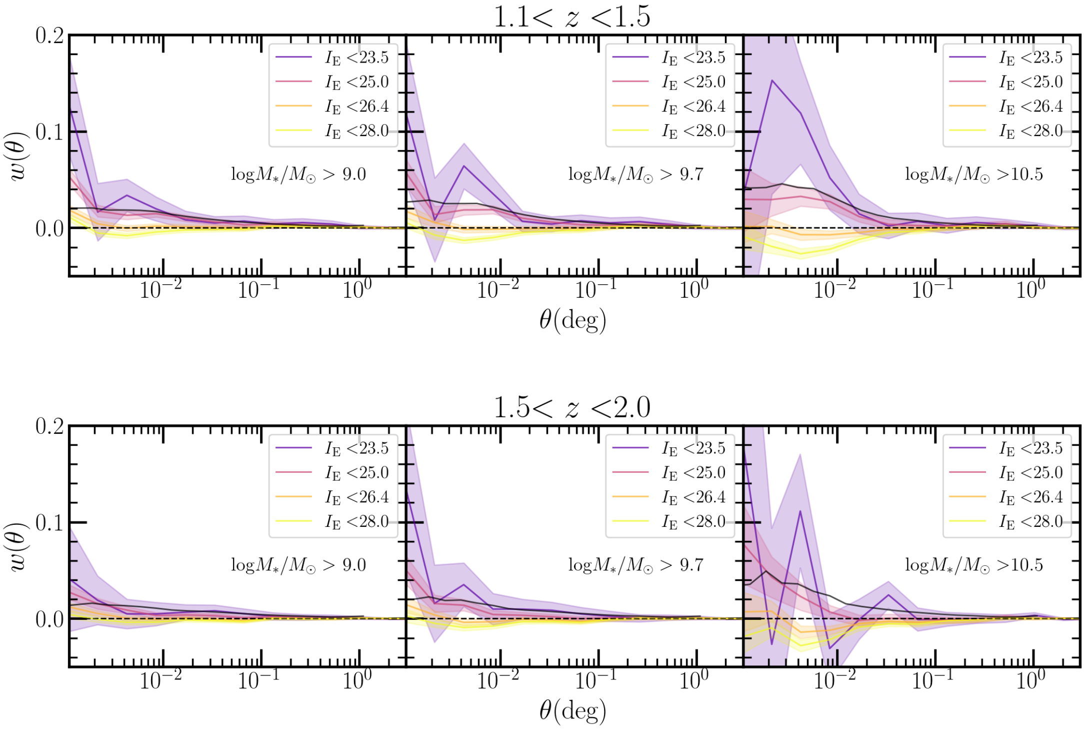


Figure 9.11: Same as Fig. 9.10 but for the last two z-bins

lens l, we take the values of κ_j^l that are associated to the $N_{\mathrm{src}}^{\text{r-bin}}$ sources lying in the radial bin, where $j = 1, ..., N_{\mathrm{src}}^{\text{r-bin}}$. To obtain $\kappa^l(\theta_i)$, we take the average of $\kappa_j^l(\theta_i)$ of all sources. Finally, we stack $\kappa^l(\theta_i)$ for all lenses to obtain the convergence profile $\kappa(\theta_i)$.

The convergence profile and the magnification angular correlation function are related from the fact that the 2PCF essentially measures the background overdensity stacked over all lenses; this, as we saw in Eqn. 3.64, can be written as

$$w_{\mathrm{true}}^{\mathrm{LS}}(\theta) \approx 2(\alpha - 1)\kappa(\theta). \tag{9.6}$$

We use this relation to compare the ground truth to our measurements.

Fig. 9.10 and Fig. 9.11 show the $w_{\mathrm{true}}^{\mathrm{LS}}(\theta)$ (in black solid curve) overplotted on the measurements. $w_{\mathrm{true}}^{\mathrm{LS}}(\theta)$ in this case, has been calculated for the background sources having $I_{\mathrm{E}} < 25.0$. This corresponds to $\alpha - 1 = 1.26$, which we use in Eqn. 9.6 to scale the convergence profile. The resulting $w_{\mathrm{true}}^{\mathrm{LS}}(\theta)$ is in good agreement with the measurements corresponding to the same source sample, shown in red. The good agreement persists for all lens stellar masses and redshifts bins. This agreement is a robust validation of the magnification measurements. Therefore, these measurements, obtained using the true redshifts and magnitudes, will serve as true magnification signal in the following analysis when we will introduce photometric errors.

9.3.4 *Systematic errors due to photometric redshift interlopers*

The source of the most important systematic error in the magnification measurement is a clustering signal coming from interlopers. We investigate the effect of interlopers on the measured magnification 2PCF in the following way. We select the lens sample based on their true redshifts in $0.8 < z < 1.1$ and in three different stellar mass thresholds $\log(M_\star/M_\odot) > [9.0, 9.7, 10.5]$ and the source sample at $2.5 < z < 3.0$ with a cut at $I_\mathrm{E} = 25.3$ corresponding to the maximum S/N (as described in § 9.2.5). We then introduce interlopers, in the way described in § 9.2.3, with four different fractions: $f_\mathrm{interlop} = 0.2\%$, 1%, 2% and 10%. We then measure the cross-correlation for four cases:

- 1) *main* case: which contains the interlopers and whose measurements will contain both magnification and clustering due to interlopers signals;
- 2) *only magnification* case, that does not contain interlopers and represents the clean magnification signal;
- 3) *no magnification* case, which contains interlopers but lensing is turned off, i.e., we are using the non-deflected positions and non-magnified magnitudes. These measurements would contain only the clustering signal due to the interlopers;
- 4) *pure* case, that has no interlopers and no magnification. The cross-correlations in this case should be consistent with zero, modulo some non-zero correlations due to the imperfect construction of the mock, whose investigation is beyond the scope of this work. These serve as sanity checks.

The angular cross-correlation measurements for the four different interloper fractions are presented in Fig. 9.12 and Fig. 9.13. The three panels show the measurements for the three lens stellar mass thresholds $\log(M_\star/M_\odot) > [9.0, 9.7, 10.5]$. The top panels show $w(\theta)$ for the main, magnification only, no magnification and pure cases. The bottom panels show the relative difference between the main and magnification only case. In the ideal case, where no interlopers are present, the relative difference between the main and magnification only should be consistent with zero, and the no magnification case should be consistent with the pure case. For a $f_\mathrm{interlop} \sim 0.2\%$ the main and the magnification only cases are in a good agreement, with only marginal difference well within the errorbars of the measurements. This is also shown in the only marginal difference between the pure and no magnification cases. This means that for $f_\mathrm{interlop} \sim 0.2\%$ the measured cross-correlation function can be considered as a just measurement of the magnification, not catastrophically affected by the leaking clustering signal. However, the situation changes for $f_\mathrm{interlop} \gtrsim 1\%$ where the clustering signal starts to significantly bias the cross-correlation measurement. At $f_\mathrm{interlop} \gtrsim 2\%$, the clustering from the interlopers starts to dominate the measured cross-correlations, and the magnification measurement is catastrophically biased.

This analysis can be used to place requirements from the photometric redshifts of a survey necessary to obtain an unbiased magnification measurement fit for scientific analyses. The analysis presented here shows that an interloper fraction of the order of $f_\mathrm{interlop} \sim 0.2\%$

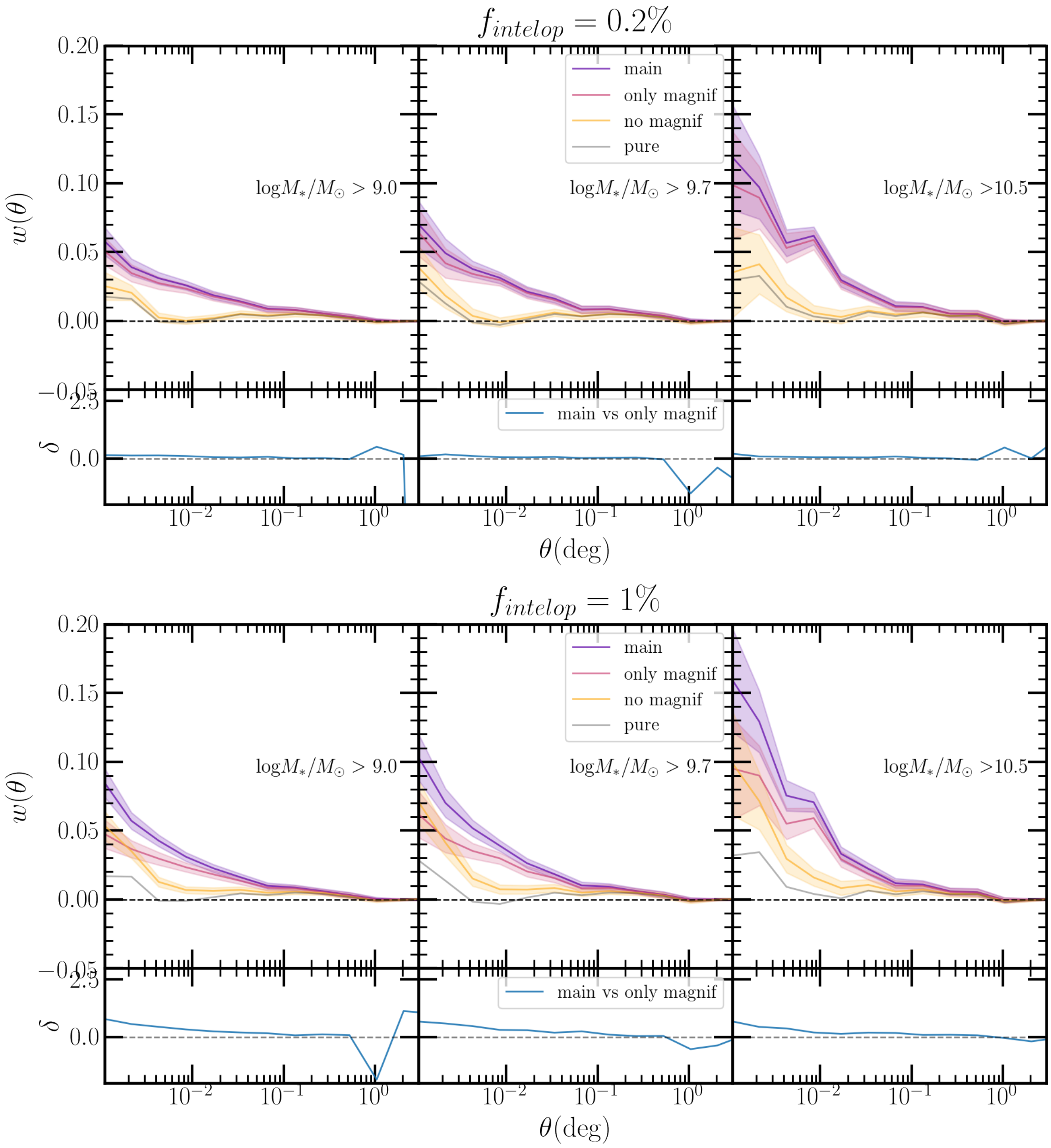


Figure 9.12: Cross-correlation measurements for interloper fractions of $f_{\text{interlop}} = 0.2\%$ (top figure) and $f_{\text{interlop}} = 1\%$ (bottom figure). The three panels show the measurements for the three lens stellar mass thresholds $\log(M_\star/M_\odot) > [9.0, 9.7, 10.5]$. The top panels show $w(\theta)$ for the main, magnification only, no magnification and pure cases. The bottom panels show the relative difference between the main and magnification only case

would provide a just measurement of the magnification. We use these results in the next subsection to verify the interloper fraction from the selection based on the photometric redshifts, and how justly the magnification can be measured.

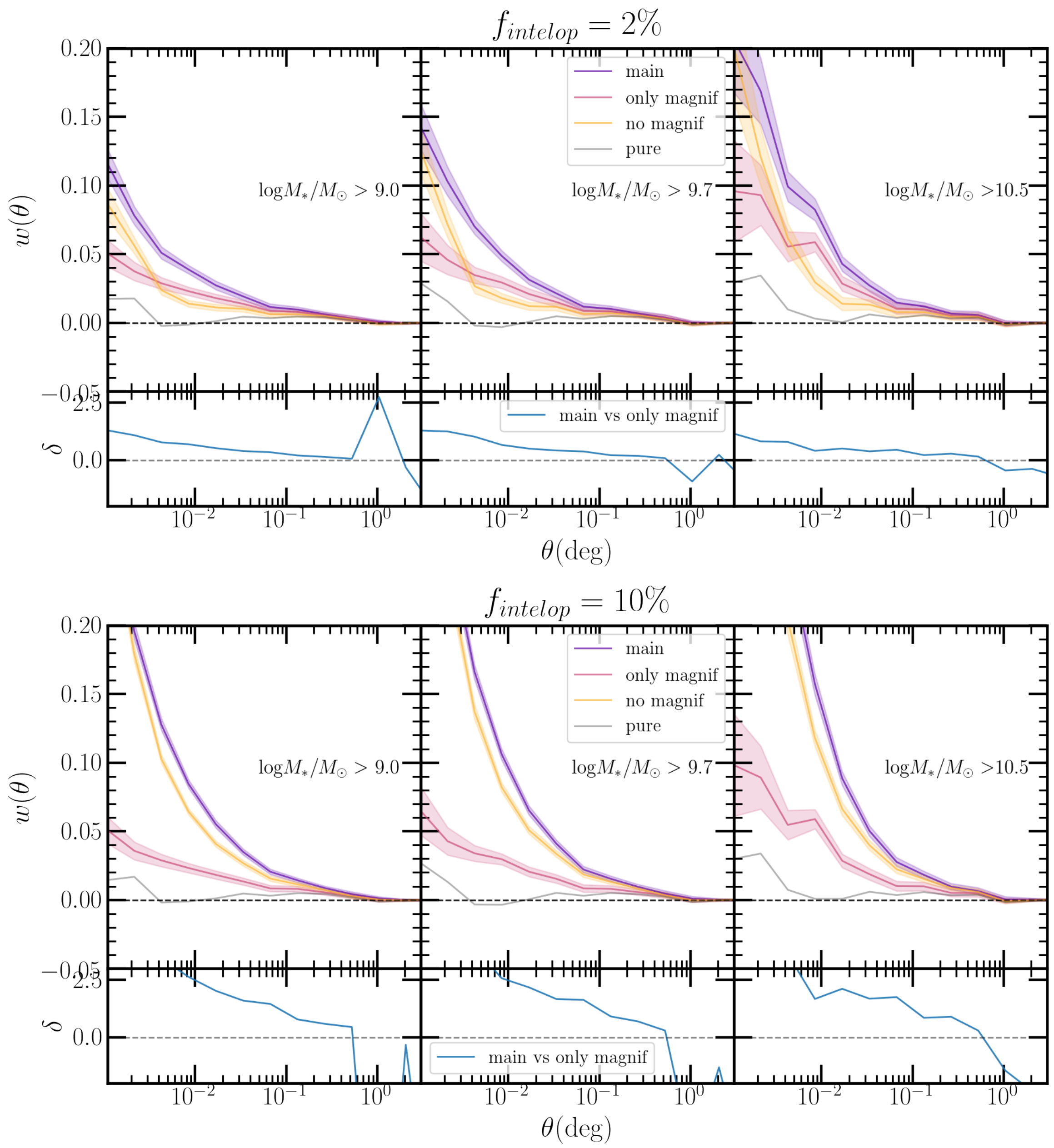


Figure 9.13: Same as Fig. 9.12, but for interloper fractions of $f_{\text{interlop}} = 2\%$ (top figure) and $f_{\text{interlop}} = 10\%$

9.3.5 *Measurements with photometric redshift selection*

Finally, we use the photometric redshifts we obtained in Fig. 9.2.3 to simulate a realistic selection of the lens and source sample in an Euclid Deep + H20 survey. We select the lenses at $0.5 < z_{\text{phot}} < 0.8$ with $\log(M_\star/M_\odot) = 9.7$ and I_E, J_E, Y_E, $H_E < 26$, and $\chi^2 < 20$ for the goodness of the SED fit. Sources are selected at $2.4 < z_{\text{phot}} < 2.9$ and at the magnitude cut that maximizes the S/N, $I_E < 25.3$. The redshift distribution of the two samples and the photo-z vs. true-z comparison is shown in Fig. 9.14, right and left panels correspondingly. The redshift distribution is plotted in log scale to accentuate the otherwise small overlap between the two distributions. The comparison with the true redshifts results in an interloper fraction of $f_{\text{interlop}} = 0.3\%$. In this case, the interloper fraction is computed from the overlap of the two distributions in Fig. 9.14 (right).

Fig. 9.15 shows the 2pt cross-correlation function for the photo-z selected sample (in blue), as well as for the case where magnification is turned off (orange) and for samples selected in true-z (green). The figure shows that the magnification measurement made with the photo-z selection, for which the interloper fraction of $f_{\text{interlop}} = 0.3\%$, is in a very close agreement with the selection on true redshift (and no interlopers). The difference between the two come from the different redshift distributions of the true and photo-z selection – the smooth distributions of the latter, cause the decrease in amplitude. The measurement where the magnification is turned off, which should only contain signal due to clustering of interlopers, is consistent with the null test. The null test is a measurement of the cross-correlation of the two samples selected in true-z and has magnification turned off, therefore it shows only residual correlations that might be present for various reasons (which is out of the scope of this work).

Here, we only presented measurements for the lens sample selected in $0.5 < z < 0.8$ which showed the lower interloper fraction. But this can change depending on the selection. For example, the $0.8 < z < 1.1$ lens sample selection results in an interloper fraction of $\sim 1\%$. As we've seen from the analysis above, this is the regime where the magnification measurements start to be significantly affected by clustering systematics.

This analysis shows that the photometric redshift selection of lens and source samples in an Euclid Deep + H20 survey can yield sufficiently clean selection in order to measure the magnification signal, without severe contamination from clustering. This, of course, will have to be done carefully. The redshift selection of both lens and source sample will have to be done in a way such that the contamination is minimized. For example, regions in the true-z vs. photo-z space that are populated with catastrophic photo-zs from the Balmer vs. Lyman break confusion will need to be avoided.

In scientific analyses, where one would use magnification to infer, for example, the HOD parameters, the systematic due to clustering of interlopers can be dealt with in the modeling. One can fit the measured cross-correlation with a model that takes into account the clustering contribution. This would require a good estimation of the interloper fraction that can be achieved from a well representative and unbiased spectroscopic redshift calibration sample, which the Euclid survey will in any case need to provide to achieve its main scientific goals.

9.3.6 *Optimally weighted magnification correlation function*

As presented in § 9.2.4, the S/N ratio of the magnification measurement can be increased by considering all sources down to the completeness limit, where the slope can be accurately computed. We apply the optimal weighting technique to compute the magnification cross correlation between lenses at $0.9 < z < 1.1$, $\log(M_\star/M_\odot) = 9.7$, $I_{\text{E}} < 26$, and sources at $2.4 < z < 2.9$ and $I_{\text{E}} < 26$. In this case, we select sources down to the completeness limit.

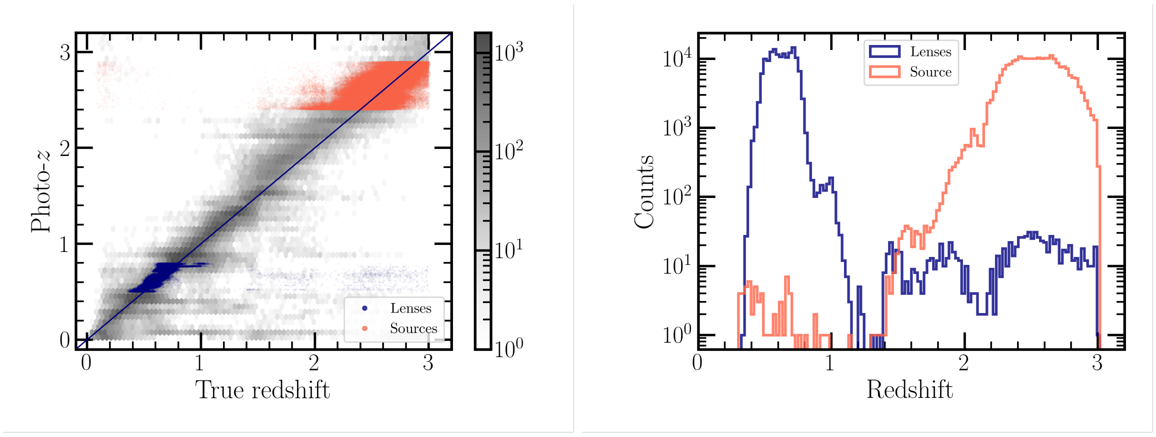


Figure 9.14: Comparison between the photometric and true redshifts (left panel) and the distributions (right panel) of the photo-z selected samples. The redshift distributions are shown in log scale to accentuate the otherwise small overlap between the two distributions

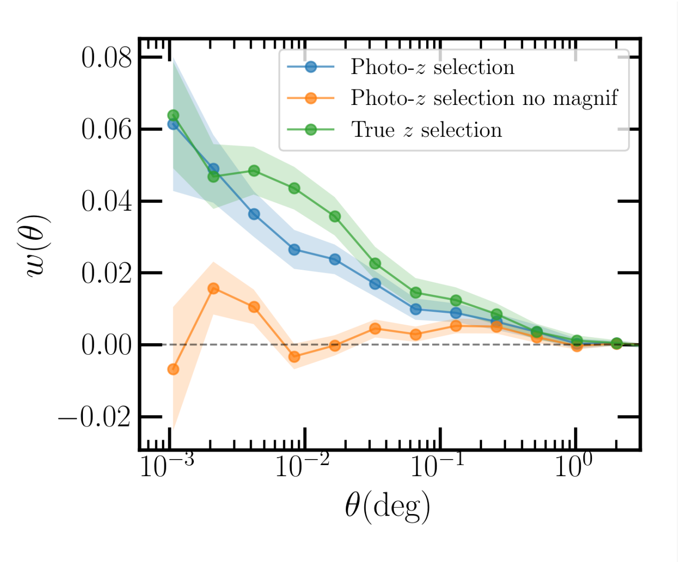


Figure 9.15: 2pt cross-correlation function for the photo-z selected sample (in blue), as well as for the case where magnification is turned off (orange) and for samples selected in true-z (green). The fact that the blue measurements are in close agreement with the green, and the fact that the orange is consistent with zero, means that the photo-z selection can provide a robust measurement of the magnification

Fig. 9.16 shows the correlation measurement using the optimal weighted technique (orange) and without it (blue). This results in a higher amplitude of the optimally weighted correlation function, due to the larger effective slope, compared to non-weighted one which has a $\alpha - 1 \sim 0.5$. This results in a $S/N_{\rm weigh} = 16.7$, compared to $S/N_{\rm weigh} = 8.9$, more than a factor of two difference.

The depth of the Flagship2.0 mock prohibits us to go deeper in magnitude and down to the real depths of Euclid Deep (see caveats in § 9.3.7). Nonetheless, this provides an example of the power of the optimally weighted technique to significantly increase the S/N of the magnification measurement and leverage the statistical power of the exponential increase of the number of galaxies at fainter magnitudes. Given the fact that shear based measurements will be limited to the number of background galaxies that can have a re-

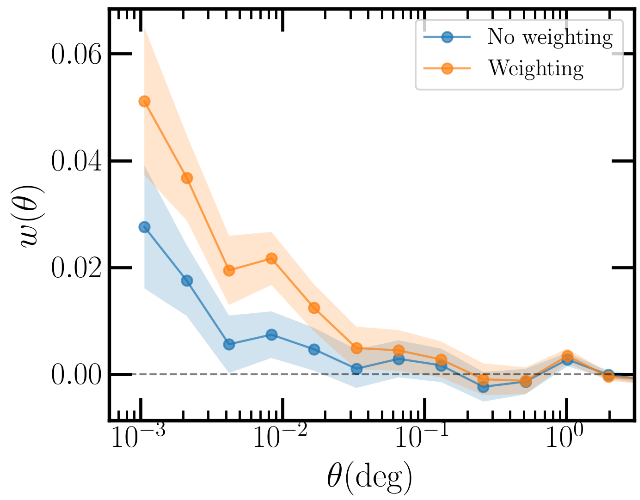


Figure 9.16: Optimally weighted correlation function (orange) due to magnification between lenses at $0.9 < z < 1.1$, $\log(M_\star/M_\odot) = 9.7$, $I_{\rm E} < 26$, and sources at $2.4 < z < 2.9$ and $I_{\rm E} < 26$. In blue, we show the correlation function for the same samples without applying the optimal weighting. The result is more than a factor of 2 higher S/N of the weighted correlation function.

solved shape, the optimal weighting of the magnification provides a competitive advantage over shear.

9.3.7 *Caveats and future extensions of the analysis*

In this work, we relied on the Flagship2.0 mock galaxy catalog, which only reaches the depths of the Euclid Wide survey. This is, of course, a major caveat and limitation for the analysis we are carrying out in this work, which focuses on the Deep survey. However, our approach is still useful to draw conclusions, at least in the regime of validity of Euclid Wide, down to AB ~ 26. This disables us to tap into the real advantage of Euclid Deep, with its improved depth by 2 magnitudes and the increased redshift range. Therefore, this analysis can be considered to provide upper limits to as what will be possible with the Euclid Deep.

An additional drawback is the unavailability of IRAC photometry in the current version of Flagship2.0. IRAC photometry at 3.6 μm and 4.5 μm probes the rest frame optical part of the SED, which is crucial to obtain accurate redshift and stellar mass measurements, especially at $z > 3$. This means that, again, our analysis on the photo-z performance and its effects on the magnification measurements, can be considered as an upper limit.

Magnification measurements are a subject of additional systematic errors, such as an inaccurate slope estimation, photometric uncertainties, dust extinction etc (Hildebrandt, 2016). The latter is not a major concern for our purposes where we use normal galaxies, but will have to be taken into account if the lens sample is a specific galaxy population such as sub-millimeter galaxies, which are very dusty. Erroneous estimation of the slope can severely affect the interpretations of the magnification measurements, since the models depend linearly on the slope. In this work we mainly relied on the true magnitudes, but ideally the effect of photometric errors on the estimation of the slope should be carefully investigated. We leave this for future work, which we plan for, in order to complete this Chapter and submit it as a paper.

Finally, the goal is to use magnification measurements to probe halo properties. Similarly to the analysis in **??**, the next extension of this work is to fit for the HOD parameters and compare them to the constraints from clustering and GSMF. Additionally, magnification is sensitive to the inner halo profile and its role in constraining halo profile parameters remains to be investigated.

9.4 DISCUSSION AND CONCLUSIONS

In this chapter, we presented an analysis of magnification measurements in an Euclid Deep-like survey. Using measurements of the two-point cross-correlation function between foreground and background galaxies, we showed the dependence of the magnification bias on the lens population stellar mass and source population magnitude cuts.

The magnification measurements using sources selected on different magnitude cuts yielded the expected behavior – the brightest sample with the steepest slopes has the highest amplitude, the sample with slope ~ 1 is consistent with zero and the faintest sample with slope lower than unity showed negative correlation. This serves as a validation of the lensing nature of the measured signal. To confirm this, we also compared it with the true lensing signal that we computed from the known convergence values in the mock catalog. This comparison showed a good agreement between the measured and the true signal, confirming the lensing nature of the correlation.

The stellar mass dependent measurements served to investigate that the sufficiently large area of Euclid Deep will allow separating the lens sample in several stellar mass thresholds and measure magnification with good S/N. Fitting these with models that constrain halo properties such as the median halo masses of the lens population (e.g., as we did in § 8.4), will allow measuring the relation between stellar and halo mass. This analysis is left for near future work.

One of the most important ingredients in cleanly selecting lens and source samples of normal galaxies in a photometric survey are the photometric redshifts. Due to erroneous photo-z the two samples can be contaminated by interlopers which introduce a clustering signal in the measured 2PCF which is a strong systematic error. We investigated the effect of different interloper fractions on the measured correlation function and estimated that f_{interlop} of about 0.3% and less can yield a relatively unbiased measurement of magnification. We then simulated photometric errors and photometric redshifts to more realistically estimate how an Euclid Deep + H20 survey would perform in this respect. Our conclusion is that such a survey would result in photo-zs sufficiently accurate to achieve minimal interloper fractions. However, the selection will have to be done carefully by avoiding outliers in the true-z vs. photo-z space, or the clustering contribution from the interlopers should be properly taken into account in the modeling.

Finally, we perform the optimal-weighting technique to include all source galaxies down to the completeness limit ($I_{\rm E} < 26$) by weighting them by the value of $\alpha(m) - 1$. We showed that this would result in an increase in the S/N by more than a factor of two compared to the non-weighted measurement. This will allow using all galaxies down to the completeness limit and exponentially increase the available number of galaxies for the analysis. Given the fact that shear based measurements will be limited to the number of background galaxies that can have a resolved shape, the optimal weighting of the magnification provides a competitive advantage over shear.

The purpose of this chapter was to show that in Euclid Deep we can successfully measure magnification with sufficient S/N and controlled systematic errors fit for scientific use, e.g., to study dark matter halo properties. It is important to also note the complementarity and advantage of a weak lensing probe such as magnification will have over other probes such as galaxy clustering and abundance probes. Lensing is especially sensitive to the satellite population within halos, therefore magnification can be used to probe the halo density profiles. This can be made possible by fitting for the NFW profile parameters within the HOD formalism (Chapter 3). Additionally, lensing directly probes halo masses around galaxies. In clustering analysis, this is done in a more convoluted way where masses can be indirectly inferred given a relationship between halo mass and clustering as well as constraints on the halo abundance in a cosmological model. Furthermore, clustering measurements can be influenced by secondary halo properties other than mass (also called assembly bias). Lensing, being a direct halo mass probe, can thus be helpful in keeping in check the systematic errors from the assembly bias in constraining the HOD from clustering measurements. Therefore, implementing weak lensing magnification in the upcoming deep and wide photometric surveys, such as Euclid Deep, will open fruitful avenues in studying in finer detail the intricate relation between galaxy and halo properties.

Magnification can also be a source of significant systematic error in cosmological studies. These are typically based on 3×2pt analyses that include 2pt correlation function (or power spectra) measurements of cosmic shear, galaxy-galaxy lensing and clustering. Magnification can introduce additional correlation in these that if not taken into account can lead to $> 1\sigma$ biases in the estimation of some cosmological parameters such as σ_8 and w_0 (e.g., Duncan et al., 2022; Euclid Collaboration et al., 2022b, and references therein). Additionally, including magnification into the analysis can also reduce the error bars on cosmological parameters such as σ_8, $n_{\rm s}$ and $\Omega_{\rm m,0}$ (Euclid Collaboration et al., 2022b). For these reasons, accounting for the magnification biases in cosmological analyses is identified as key importance in the *Euclid* collaboration.

10

GALAXY CLUSTERS AT $z \gtrsim 3$

10.1 INTRODUCTION

In the local Universe, galaxy clusters represent the most overdense environments gravitationally bound by massive dark matter halos. They host passively evolving, elliptical galaxies which have a negligible contribution to the cosmic star-formation rate density (CSFRD). In the early Universe, the most massive and overdense structures are often called galaxy protoclusters (see Overzier, 2016, for a review), which are rich in gas and dust and undergoing vigorous star-formation and AGN activities in the core region (e.g., Thomas et al., 2005; Dannerbauer et al., 2014; Emonts et al., 2016; Jin et al., 2021). Simulations suggest that the star formation in protoclusters contribute significantly to the CSFRD, e.g., 20–50% at $z > 3$ (Chiang et al., 2017), while observational constraints are still rare. This is because the most starbursting cluster members are often severely attenuated by dust, making it difficult to detect and measure SFR using UV/optical observations. Thanks to the strong negative K-correction in the Rayleigh-Jeans tail of the dust spectral energy distributions (SEDs) and the strength of molecular lines, advanced (sub)millimetre facilities allow us to detect and spectroscopically confirm protoclusters at $z > 4$.

Recently, using LABOCA and ALMA observations, Miller et al., 2018 confirmed a SPT-selected candidate SPT2349-56 to be a protocluster at $z = 4.3$, hosting a core of massive galaxies and a collective SFR of $\sim 6000\ M_\odot\,\mathrm{yr}^{-1}$. Oteo et al., 2018 discovered a $z = 4$ protocluster DRC, which was selected from red Herschel colors and then confirmed by CO and CI detection with ALMA and ATCA. It revealed a collection of at least 10 DSFGs distributed over $260 \times 310\ \mathrm{kpc}^2$ with a total, dust-obscured SFR of $\sim 6500\ M_\odot\ \mathrm{yr}^{-1}$, suggesting a fraction of SFRD up to 25% at $z \sim 4$. However, the observational constraints on CSFRD from protoclusters are still poor at $z > 3$ due to small sample size and selection bias, which hampers the direct comparison to simulations (e.g., Lagos et al., 2019; Bassini et al., 2020). On the other hand, protoclusters host galaxies that are rapidly evolving and undergoing a transformation of their properties. Furthermore, they are expected to host the first generation of quiescent galaxies (e.g., Kalita et al., 2021). These transformations are closely linked to the gas accretion mode of massive structures (Daddi et al., 2022a), highlighting the importance of high-z protoclusters in shaping galaxy properties.

Another importance of $z > 3$ clusters is to test the paradigm of cold streams penetrating the shocks at the virial radii of massive, hot halos, that are key in explaining the evolution of massive galaxies at high redshifts (Kereš et al., 2005; Dekel et al., 2009). Only very recently, observational evidence of cold streams have been obtained by observations of collisionally excited Lyα emission around groups and clusters at $z \sim 3$ (Daddi et al., 2022a). However, this recent work suffers from small and inhomogeneous samples. Compiling larger samples of clusters at $z \sim 3$ is crucial for full characterization of cold streams in massive halos. Therefore, discovering more protoclusters out to higher redshifts is necessary to understand the role they played in the early Universe.

Using the final processing of all archival *Spitzer*/IRAC data in the Euclid Deep Field North (EDF-N) (Moneti et al., 2021), which was done as part of this thesis (see Chapter 6), we have selected a sample of 60 Red IRAC Overdensity (RIO) associated with red Herschel colors, which are promising candidates of protoclusters at $z > 3$. Using these candidates, we proposed NOrthern Extended Millimeter Array (NOEMA) observations to solidly confirm their redshifts. This proposal was accepted, and the observations were carried out in June 2022 under the NOEMA project S22CW (PI: M. Shuntov).

Given the importance of clusters in shaping galaxy properties, including their role in understanding cold streams at high redshifts, studying clusters is in close alignment with the topic of this Thesis. This the main motivations behind pursuing this project of searching for galaxy clusters at high redshifts with the Cosmic Dawn Survey IRAC data.

10.2 NOEMA OBSERVATIONS OF PROTOCLUSTERS IN THE COSMIC DAWN SURVEY

Using infrared data from *Spitzer*/IRAC and Herschel we carried out a search for protoclusters in the 10 deg^2 EDF-N field using the following criteria (also presented in Fig. 10.1)

- **(1) Red IRAC color [3.6]$-$[4.5]$>$ 0.** The IRAC channel-1 (3.6 μm) and channel-2 (4.5 μm) probe the Raleigh-Jeans tail of the stellar emission in the spectral energy distribution (SED) of galaxies that peaks at $\sim 1.6\,\mu$m. As the galaxy SED is redshifted at $z \gtrsim 0.7$, the peak shifts towards the channel-2, resulting in a characteristic colour-redshift relation that linearly rises to $z \sim 1.5$ (Sorba and Sawicki, 2010). This is also verified by photo-z and spec-z catalogs in the COSMOS field, this color criterion would select $z > 1.5$ galaxies, as presented in Fig. 10.1-left. Furthermore, this is consistent with the similar colors of most known (proto)cluster members at $z > 2$. We exploit this IRAC color-redshift relation to select $z > 1.5$ galaxies.

- **(2) Significant overdensity of red IRAC sources** $\Sigma_5 > 4.8\sigma$ or $\Sigma_{10} > 4.5\sigma$. For IRAC sources with magnitudes $[4.5] < 22$ and colors $[3.6] - [4.5] \geq 0$, we compute the surface density using a local galaxy density estimators ρ_5 and ρ_{10}. ρ_N is defined as $N/(\pi r_N^2)$, where r_N is the distance to the N-th nearest neighbor ($N = 5, 10$). We implement an iterative 2σ clipping fit of a Gaussian on ρ_5 and ρ_{10} in order to model the normal distribution on the low density end. Using the fitted Gaussian, we then compute the over-density significance Σ_5 and Σ_{10} of each source. To select the most significant over-densities, we define the selection criteria by performing the same procedure in the COSMOS field. This results in IRAC selected red sources that have

$\Sigma_5 > 5\sigma$ or $\Sigma_{10} > 4.5\sigma$. As shown in Fig. 10.1-middle, this criterion allows selecting the known clusters in COSMOS (Wang et al. 2016 at $z = 2.5$, and Daddi et al. 2021 at $z = 2.9$).

- **(3) Red Herschel colors** $S_{250} < S_{350} < S_{500}$. In order to target our selection of high-z clusters in their rapid formation phase of vigorous dust-obscured star formation, we require that these red IRAC over-densities correspond to red peaks in *Herschel*/Spire 250 μm, 350 μm and 500 μm, as well as peaks in SCUBA2 850 μm bands when available. This criterion is also justified by the fact that all the known protoclusters at $z \sim 4$ in the literature show extremely red colors with $S_{250} < S_{350} < S_{500}$ (aka. 500μ-riser) and detection at $\sim$ 850 μm. This color is redder than known clusters at $z = 2 - 3$, indicating higher-z. Similar 500 μm-risers have been confirmed at $z > 4$ (Miller et al., 2018; Oteo et al., 2018), as shown in Fig. 10.1-right. In this figure, we show the colors of our candidates obtained from measuring peak fluxes, compared to the colors of several known clusters. It is worth noting that the colors of high clusters by Oteo et al. and Miller et al. are obtained using total fluxes, which can explain their redder S_{350}/S_{250} colors.

This resulted in a large sample of $\sim$ 60 protocluster candidates in the 10 deg^2 EDF-N field, i.e., the most significant red IRAC overdensity, the bright and red Herschel fluxes. Such criteria make them candidates to be at $z \gtrsim 3 - 5$. We select from the initial sample the brightest candidates with $S_{500} > 60$ mJy, resulting in five targets for close examination. We show their multi-band cutouts in Fig. 10.2.

For these candidates, we expect that there is very little chance that such Herschel fluxes originate from a single source. If so, it would indicate that it is a super-starbursting galaxy at $z > 5 - 7$ with SFR$> 4000\ M_\odot$ yr^{-1} given the red Herschel colors and high FIR brightness, which has never been observed nor predicted by simulations. As such, these alternative scenarios would still represent exciting discoveries. Additionally, it is unlikely that these targets are magnified by strong gravitational lensing, as no massive galaxies are seen in their foreground.

Given the lack of deep NIR data in the EDF-N field, the photometric redshifts are unavailable for our targets. Alternatively, we estimate the redshifts using Herschel colors by comparing to high-z clusters in literature. As shown in Fig. 10.1-right, the S_{500}/S_{350} colors of our targets are significantly redder than Herschel-detected clusters ($z = 2.5$, 2.9) in the literature, indicating $z > 3$. Comparing to GN20 and known protoclusters at $z \sim 4$, our targets are similarly red in S_{500}/S_{350}, indicating similar redshift $z \sim 4$.

Based on these arguments, we proposed NOEMA observations in the 3 mm band to detect emission lines, measure redshifts and confirm the potential protoclusters. We estimated that collective SFRs of these targets would be 2000–6000 $M_\odot$ yr^{-1} if at $z = 3 - 5$, comparable with the vigorous star-formation activities found in known protoclusters (Daddi et al., 2021; Miller et al., 2018; Oteo et al., 2018). NOEMA 3 mm scans can detect at least two CO (or possible CI) lines thanks to the broad bandwidth of the receivers, thereby robustly confirming their redshifts.

These candidates are found in EDF-N which is still a relatively unexplored field. However, being an Euclid Deep field, it is becoming one of the premier extragalactic field with deep and multi-wavelength coverage from all the major observatories on Earth and in space; e.g. H20 in UV-NIR (McPartland et al., in prep) and Euclid in VIS-NIR. Once the cluster candidate redshift are identified, these complementary data will enable detailed studies of the individual cluster members' physical properties and morphologies, that will fully characterize them.

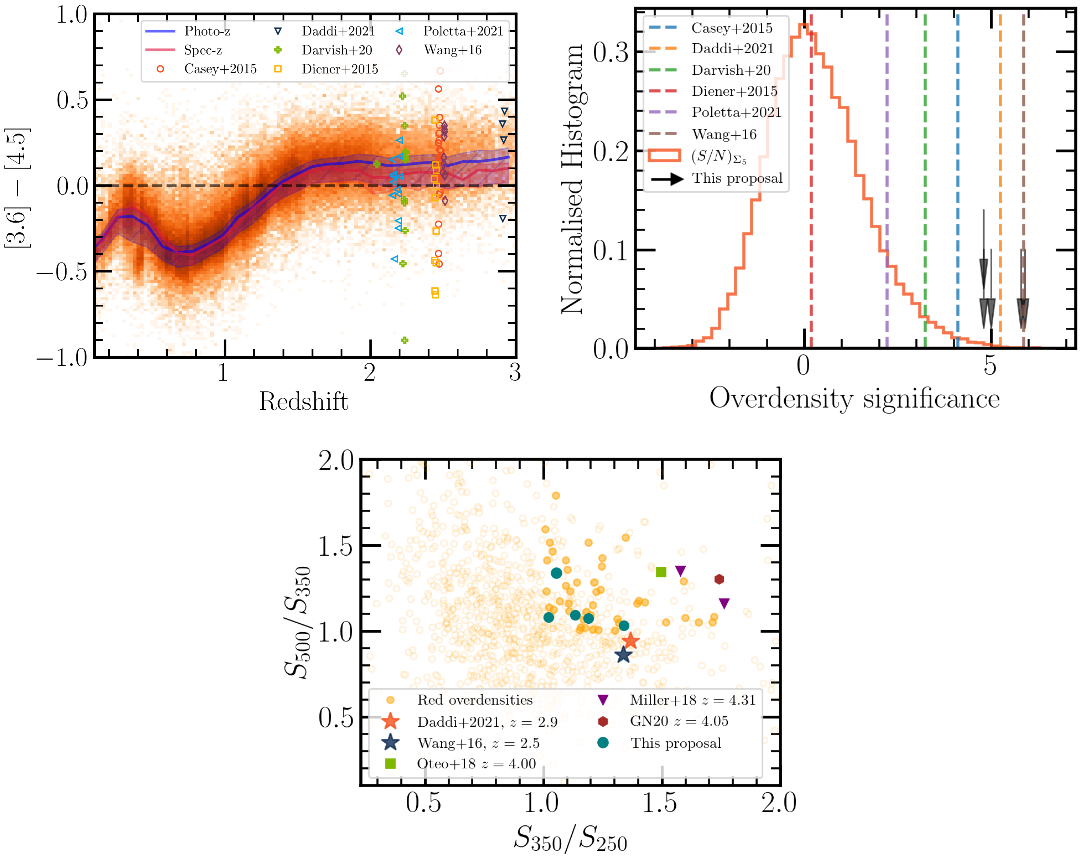


Figure 10.1: Selection of red IRAC overdensities coupled with Herschel 500μm-risers. **Top left:** IRAC color vs redshifts in the COSMOS field, overlaid by known (proto)cluster members at $z > 2$. The dashed line shows our color cut used for high-z cluster selection. **Top right:** Overdensity of red IRAC sources in with known (proto)clusters marked by vertical lines. **Bottom middle :** Herschel color-color plot for known clusters and our targets. Our targets are selected with the most significant overdensity of red IRAC sources and the reddest Herschel colors ($S_{250} < S_{350} < S_{500}$), indicating they are $z > 3$ protoclusters.

10.3 RIO-169: A PROTOCLUSTER CANDIDATE AT $z = 4.2$

The data of proposal S22CW have all been taken and reduced. All targets are detected with at least one emission line, and the redshifts are confirmed at $z = 2.6 - 4.2$. RIO-169 is the most extreme case among the five targets shown in Fig. 10.4, in which source RIO-169.a is confirmed to be at $z = 4.19$ with a strong CO(4-3) line and a 4.6σ but tentative

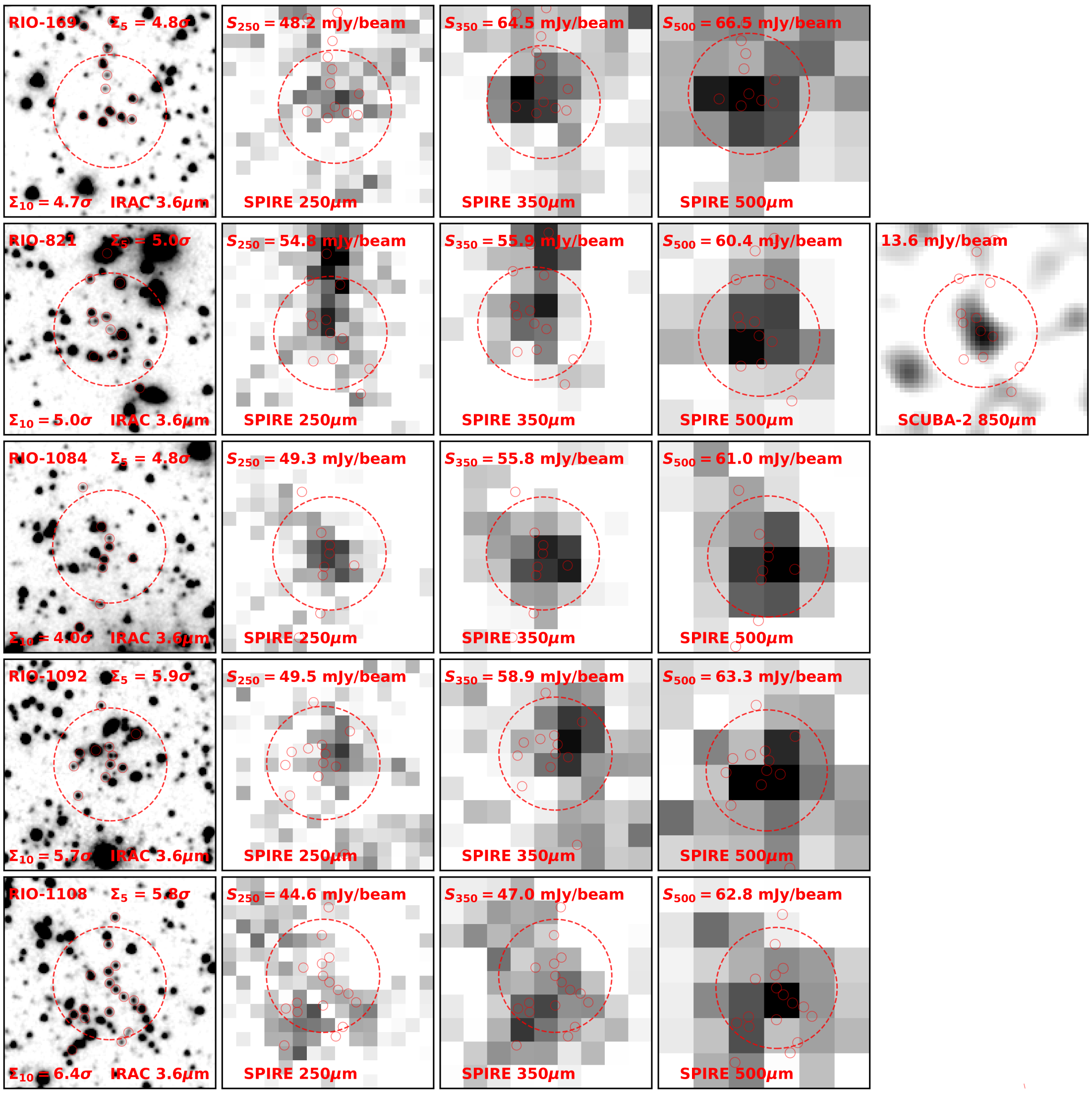


Figure 10.2: Multi-band cutouts of the cluster candidates in this proposal. We show the significance of Σ_5 and Σ_{10} overdensity in IRAC cutout, and the peak fluxes in Herschel/SPIRE and SCUBA2 cutouts. Large dashed circle is centering on the peak of IRAC overdensity with a size of $50''$ corresponding to NOEMA primary beam size at 100 GHz. Red IRAC sources ($S_{3.6} < S_{4.5}$) are marked with solid red circles of size $4''$.

$H_2O(4_{2,3}$–$3_{3,0})$ transition (rest frame 448 GHz). The CO(4-3) and the tentative $H_2O(4_{2,3}$–$3_{3,0})$ line are detected on the same position of its IRAC counterpart, thus confirming its redshift $z = 4.19$. We note that the [CI](1-0) transition is also covered by the 3mm observations, but not detected on the position of CO(4-3) emission. We further checked the signal by merging the channels of the [CI](1-0) range, only a 3σ peak is found $2''$ away from the CO emission (cyan contour in Fig. 2b), thus the [CI](1-0) is barely detected.

A second source RIO-169.b is detected with a 5σ line in the same 3mm pointing, as the source 'b' shown in Fig. 10.4. The line emission is well matched with the red IRAC counterpart, and the observed frequency is close to the CO(4-3) of RIO-169.a. Assuming the line is CO(4-3), the redshift of RIO-169.b would be $z = 4.25$, which is close to the redshift of RIO-169.a and thus supports the protocluster nature of RIO-169. The velocity offset between source a and b is 3400 km/s, which is larger than that in local clusters but

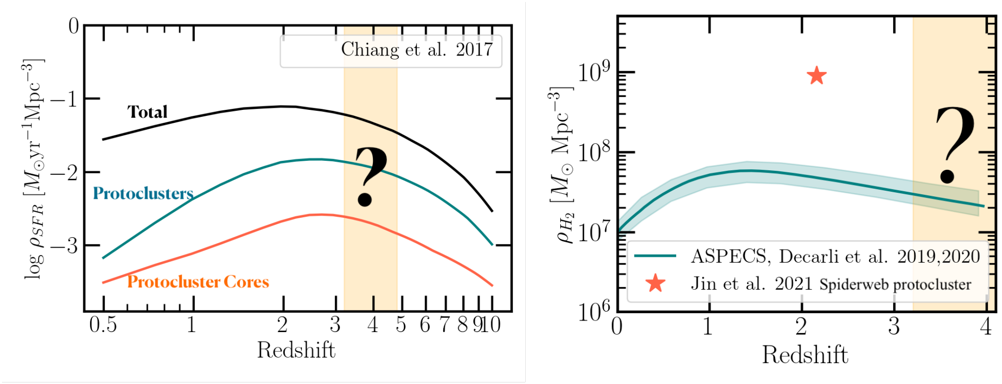


Figure 10.3: **Left**: SFR density of protoclusters predicted from semi-analytic models by Chiang et al., 2017, compared to the total SFRD of all galaxies. **Right**: Redshift evolution of molecular gas density from ASPECS (Decarli et al., 2019), and the measurement of the Spiderweb protocluster H_2 density by Jin et al., 2021. The objective is to place constraints on ρ_{SFR} and $\rho_{\mathrm{H_2}}$ from protoclusters at $z \sim 4$.

consistent with $z > 2$ protoclusters such as the Spiderweb (Jin et al. 2021) and Hyperion (Cucciati et al. 2018).

These together strongly suggest that RIO-169 is a protocluster at $z = 4.2$, one of the most distant starbursting structure to date, which the redshift is only lower the record-holder SPT2349-56 at $z = 4.3$ (Miller et al., 2018). In order to robustly confirm this structure, a second line detection for source 'b' and redshift confirmations of other potential cluster members are urgently needed. Therefore, we proposed 2 mm followup in the NOEMA 2022 winter call for proposals, to detect CO(7-6)+[CI](2-1), as well as dust continuum in RIO-169.

We note that for the moment optical data in this field are shallow, and all red IRAC sources are totally not detected even in the deepest Subaru/HSC I band image (see Fig. 10.4a,b). It is thus difficult to constrain redshifts using optical SEDs, and NOEMA is the only facility competent for this job. Given the faintness and non-detection of the secondary lines (Fig. 10.4c), CO(7-6)+[CI](2-1) at 2 mm is more efficient than deeper 3 mm integration. Meanwhile, as no continuum is detected in 3 mm observations, 2 mm dust continuum is expected to be 10× brighter and will be easily detected, which will provide us better constraints on SFRs and gas masses (Scoville et al. 2017).

10.3.1 *Revealing the physical properties of RIO-169*

In order to fully study in detail the physical properties of RIO-169 we proposed to carry out follow-up observations. These will allow detecting more sources belonging to the protocluster, and more lines that will allow measuring its physical properties. Observations with NOEMA 2 mm follow-up will cover the CO(7-6)+[CI](2-1) doublet and H2O transition. In addition to the data from the S22CW project, the new observations will allow the detection of CO(7-6)+[CI](2-1) (and potential H2O) transitions for the source 'a' and 'b' and other candidate members, solidly confirming one of the most distant starbursting protocluster.

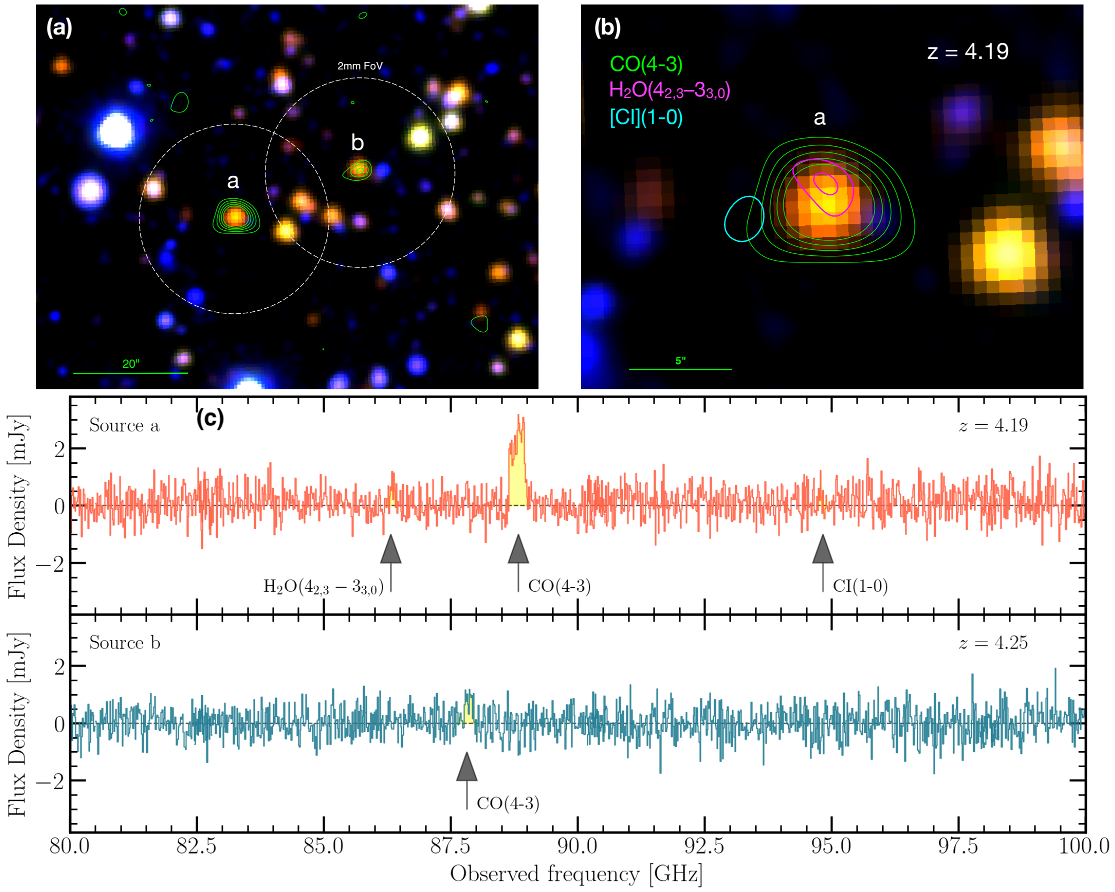


Figure 10.4: Line detections in RIO-169. **(a)** RGB color image of RIO-169 (blue: HSC I; green: IRAC 3.6μm; red: IRAC 4.5μm). White dashed circles mark the primary beams of the proposed 2mm observations. Green contours show CO(4-3) of source 'a' and the line of source #b, starting at 3σ in step of 1σ. Dashed circle marks the FoV of the proposed 2mm observation. **(b)** Line contours of source 'a'. We show the contours of H2O (blue) and the tentative [CI](1-0) detection (yellow), starting at 3σ in step of 1σ. **(c)** NOEMA 3mm spectra that are extracted by point source UVFIT on the positions of source 'a' and 'b'.

Once the redshifts are confirmed, we will measure the total SFR using Herschel together with the 2 mm dust continuum flux densities, and estimate gas mass using [CI] emissions. These will allow us to constrain the average SFE in overdense environment and explore the evolutionary trend by comparing to lower-z protocluster surveys (e.g., Jin et al., 2021; Daddi et al., 2021), which could shed light on the environmental effect on star-forming activities at very early cosmic time.

The total SFRs will provide a strong lower limit of the SFRD from protoclusters at $z \sim 4$ using Herschel data, which can be compared to models in, e.g., Chiang et al., 2017 (Fig. 10.3-left). The derived molecular gas masses will put constrains on the molecular gas density in dense environments at $z \sim 4$ (Fig. 10.3-right), and allow comparison to the gas density in field sample (Decarli et al., 2019, ASPECS) and simulations (SHARK, Lagos et al., 2019; DIANOGA, Bassini et al., 2020).

In addition, measurements of the stellar mass and the halo mass of this protocluster can be used to study the relation between the halo mass and the protocluster properties. In particular, it can be used to test the paradigm of cold streams in hot media, in which the transition mass below which cold streams can penetrate the hot media moves to higher halo masses with redshift. This protocluster would add a unique measurement at $z \sim 4.2$ (among the very few at these redshifts). This would require additional observations in NIR and MIR to detect more cluster members and estimate their stellar mass more accurately, which would be made possible by JWST. Halo masses can be estimated from several techniques such as velocity dispersion techniques or SHMR scaling relations. Furthermore, the presence of potential cold flows could be detected with potential observations of Lyα blobs around the cluster core (Daddi et al., 2022b, and references therein). This could be observed by facilities like VLT/MUSE (Oteo et al., 2018).

10.4 SUMMARY

In this chapter, we presented the discovery of one of the most distant protoclusters known to date at $z \sim 4.2$ by using IRAC, Herschel and NOEMA observations. The selection procedure of the cluster candidates involved searching for overdensities of galaxies with red IRAC colors (magnitudes [3.6] - [4.5] > 0) that correspond to red peaks in the Herschel bands (fluxes $S_{250\ \mu m} < S_{350\ \mu m} < S_{500\ \mu m}$). The most significant red overdensities were then observed with NOEMA at 3 mm and resulted in all protocluster candidates being detected with at least one emission line at $z = 2.6 - 4.2$. The most extreme case was the target RIO-169 that was detected with strong CO(4-3) and a tentative $H_20(4_{2,3} - 3_{3,0})$ line and at least two sources at $z = 4.19$. To better characterize this protocluster we are proposing additional observations with NOEMA at 2 mm to detect additional lines and other members and dust continuum. These will allow us to measure the total SFR, estimate the gas mass and the average SFE in the overdense environment. Estimates of the protocluster total stellar mass and its halo mass can help us study the cold streams in hot media paradigm. The S22CW program has demonstrated an efficient method in finding protocluster candidates at $z \gtrsim 3$ in the Euclid Deep Fields. This opens a unique opportunity of assembling larger samples of protoclusters of various properties and study their contribution to the CSFRD and how they transform galaxy properties in the epoch beyond cosmic noon. We plan to continue this program and extend it to other Euclid Deep Fields where IR data is currently available such as EDF-F.

11

CONCLUSIONS

11.1 SUMMARY

This Thesis aims at understanding the intricate relation between galaxies and dark matter halos, and to shed light on the various physical mechanisms that regulate the growth of galaxies in halos. To this end, this Thesis studied the cosmic evolution of the stellar-to-halo mass relation (SHMR) for central and satellite galaxies, and investigates weak lensing magnification as an additional probe. The novelty of this work is that this is done in a large redshift range $0.2 < z < 5.5$, using a single and homogeneous dataset – the multi-wavelength, near-infrared selected COSMOS2020 catalog. As a direct probe of halos out to high redshifts, this Thesis investigated an approach based on weak lensing magnification in the context of the Euclid Deep Survey. Finally, to probe how the most massive halos shape the properties of galaxies, this Thesis carried out a search of galaxy clusters at $z > 3$ using infrared data. The following sections summarize the main results.

Preparation of photometric catalogs

The first part of this Thesis focused on the preparation of photometric catalogs. This included photometric extraction in the infrared from *Spitzer*/IRAC images at 3.6 μm and 4.5 μm in the COSMOS and Cosmic Dawn Surveys. In COSMOS, this IRAC photometry was included in the COSMOS2020 catalog (Weaver et al., 2021), which was an important ingredient in obtaining the high quality photometric redshifts and complete samples down to low stellar masses and high redshifts (e.g., $\log M_\star/M_\odot \sim 8.2$ at $z \sim 0.3$ and $\log M_\star/M_\odot \sim 9.3$ at $z \sim 4$).

In addition, survey geometry masks were created that mask sources with potentially problematic photometry near bright stars and bad data. These masks are necessary in the measurement of the correlation function.

Finally, the resulting near-infrared selected catalog was validated using measurements of the two-point angular correlation function for apparent magnitude selected samples. The scaling of the correlation function with magnitude, in agreement with expectations and with previous measurements, served as a qualitative validation that the selection function of the catalog is well understood and fit for scientific analyses.

Stellar-to-halo mass relation since $z \sim 5$ in COSMOS2020

One of the principal studies of this Thesis was the measurement of the SHMR to $z \sim 5$. Using mass-complete samples in ten redshift bins in the COSMOS2020 catalog, the galaxy stellar mass function (GSMF) and angular correlation function were measured. These measurements were fitted with a model based on the halo occupation distribution (HOD) that allowed consistently probing the contribution to the total stellar mass budget in halos of both central and satellite galaxies. The advantages of this work is the use of a single, homogenous dataset to perform all the measurements and probe the SHMR divided by central and satellite galaxies in this large redshift range for the first time in the literature. These results were also compared with three state-of-the-art hydrodynamical simulations Horizon-AGN, EAGLE and TNG100 of the IllustrisTNG project.

Some of the principal results of this study suggest that there is little-to-no evolution in both the way galaxies occupy dark matter halos (in terms of mean numbers) and in the peak of the star-formation efficiency (SFE) at $z < 3$. The latter remains lower than 20% at least up to $z \sim 3$, indicating a globally inefficient galaxy formation process.

The halo mass and stellar mass scale at which the SFE peaks, $M_{\rm h}^{\rm peak}$ and $M_{\star}^{\rm peak}$, increase continuously with redshift, at least to $z \sim 4$. This result is in tension with previous works, where the peak halo mass decreases beyond $z \gtrsim 3$. However, given the error bars of the measurements in this work, the peak halo mass evolution at these high redshifts remains uncertain.

At $M_{\rm h} \lesssim 10^{13} M_{\odot}$, central galaxies completely dominate the stellar mass budget of the halo at all z, while satellite galaxies start to dominate at $M_{\rm h} > 2 \times 10^{13}\, M_{\odot}$. At this massive end, the $M_{*}^{\rm tot}/M_{\rm h}$ ratio flattens out and is always lower than the peak at $M_{\rm h} \sim \times 10^{12}\, M_{\odot}$, indicating that strong quenching mechanisms must be present in massive halos that quench the mass assembly of satellites.

Comparing to the hydrodynamical simulations, the SHMR is in general agreement at the low mass end below the peak. However, Horizon-AGN showed a higher SHMR in the low mass end, that can be explained by insufficient stellar feedback in the simulation. The most significant discrepancies are at the massive end, where the total stellar mass budged in halos is larger in the simulations. This suggests that the environmental quenching of satellites is less efficient in the simulations or that quenching occurs much later.

Magnification in COSMOS2020

For the purpose of revealing details of the galaxy-halo connection, this Thesis investigated the prospects of weak lensing magnification as a direct probe of halo properties. In COSMOS2020, magnification was measured using the two-point angular cross-correlation function between stellar-mass selected samples in four redshift bins at $0.2 < z < 1.5$ and $K_{\rm s}$ apparent magnitude selected background samples at $2.7 < z < 4.4$.

The magnification measurements showed the predicted dependence with lens stellar mass and source magnitude cut: the amplitude of the correlation increased with increasing lens stellar mass, while it decreased when selecting fainter samples of background galaxies.

The measurements were jointly fitted for the Magnification+GSMF and Magnification+Clustering+GSMF cases using the HOD modeling with the parametrization of the SHMR. We showed that the best-fit models including magnification is in reasonable agreement with the measurements. However, the small survey size of COSMOS does not allow for high S/N measurements and, consequently, magnification does not bring a significant constraining power on the SHMR model parameters. However, magnification allows opening other degrees of freedom in the model, such as the parameters governing galaxy distribution in halos and photo-z validation using cross-correlations. This remains to be investigated and implemented in future works.

Magnification in the Euclid Deep Survey

The prospects of magnification were also investigated within the context of the Euclid Deep Survey. Using the FLAGSHIP2.0 mock galaxy catalog, magnification was measured with the two-point angular cross-correlation function between stellar-mass selected samples in five redshift bins at $0.2 < z < 2.0$ and $I_{\rm E}$ apparent magnitude selected background samples at $2.5 < z < 2.9$. These were in a good agreement with the true convergence profile, meaning that magnification can indeed be accurately measured for various samples with high signal-to-noise (> 5).

The effects of the interloper clustering systematic errors were investigated by perturbing redshifts from the source to the lens sample (and vice versa). This led to the conclusion that to have negligible clustering systematic errors, the samples need to have $< 0.3\%$ of interlopers.

To investigate if such a clean selection can be achieved with photometric redshifts from the Euclid Deep Survey, photometric errors were simulated and photo-z were computed. This was done by combining the Euclid NIR and Cosmic Dawn Survey/H20 optical bands from Subaru Hyper SuprimeCam. This led to the conclusion that such a survey would result in photo-zs sufficiently accurate to achieve minimal interloper fractions. However, the selection will have to be done carefully by avoiding outliers in the true redshift vs. photo-z space, or the clustering contribution from the interlopers should be properly taken into account in the modeling.

Finally, by applying the optimal-weighting technique, the signal-to-noise of magnification can be significantly increased by including all galaxies down to the completeness limit, which would allow going deeper than analyses based on shear.

Search for galaxy clusters at $z > 3$

To probe the most massive objects in the Universe, this Thesis carried out a search of galaxy clusters at $z > 3$. This was done by looking for galaxy overdensities with red *Spitzer*/IRAC colors (magnitudes [$3.6\ \mu$m $- 4.5\ \mu$m > 0) that correspond to red peaks in the Herschel maps (fluxes $S_{250\ \mu m} < S_{350\ \mu m} < S_{500\ \mu m}$). This search resulted in five robust candidates that were observed with NOEMA in the 3 mm band. These observations resulted in all protocluster candidates being detected with at least one emission line at $z = 2.6 - 4.2$. Among these, is RIO-169 was revealed as one of the most distant star-bursting protoclusters at $z = 4.2$ by detecting CO(4-3) and a tentative $H_2O(4_{2,3} - 3_{3,0})$ emission line. This protocluster is a subject of follow-up proposal for NOEMA observations in the 2 mm band to detect more lines and cluster members and characterize in detail the protocluster. Along with additional observations of the other cluster candidates, this project aims to reveal the transformational effect of clusters on galaxy properties, their contribution to the total cosmic star-formation rate density and how cold streams fuel star-formation in the most massive halos.

11.2 THE RESULTS OF THIS THESIS IN CONTEXT AND FUTURE PROSPECTS

This Thesis focused on the Euclid Deep Survey, which is designed to provide a unique combination of area and depth in the near-infrared bands. Along with the spectroscopic coverage and ground-based observations, this will provide the necessary data to consistently study the cosmic evolution of the SHMR and provide measurements in currently poorly constrained regimes, such as at $z > 3$ and for different galaxy populations and in different environments. Additionally, most current studies of the galaxy-halo relation rely on shear-based methods to directly probe halos. The difficulty of accurately measuring shapes of background galaxies limits their use to $z \lesssim 1$. Therefore, this Thesis focuses on the development of a different approach using magnification to probe halos out to higher redshifts. As part of the Euclid Deep and Auxiliary fields, the COSMOS field currently offers the best dataset to do this. The study of the SHMR in COSMOS2020 in this Thesis highlighted some important discrepancies with the literature at $z > 3$. This includes the evolution of the peak halo mass, which this work measured to be continuously increasing with z, in contrast to the turn-over and decrease at $z \sim 3$ as found by empirical models by e.g., Behroozi et al. (2019).

Pinpointing the evolution of the peak halo mass is important in understanding the interplay between different feedback mechanisms that regulate the star-formation efficiency in halos. An interesting question is whether the halo mass scale at which cold streams start to penetrate halos corresponds to the peak halo mass. This is in part motivated by recent studies by Daddi et al. (2022a), who showed that the 'bending' of the star-formation main sequence traces this cold- to hot-accretion transition. Precise measurements of the peak halo mass would determine whether this transition also corresponds to the peak halo mass, which would shed light on the intricate interplay between different physical processes that shape galaxy growth. This question is, in part, the motivation behind the search for galaxy clusters at $z > 3$ (Chapter 10). In pinpointing the peak halo mass evolution at high redshifts, we are currently limited by the relatively small galaxy samples at high-z that are complete. Next-generation deep and wide surveys such as Euclid Deep are necessary to study this question.

Another important question regarding the SHMR evolution is how does the SFE evolve with redshift. Currently, there is no consensus that the SFE evolves with redshift at all. Some studies indicate that there is no evolution (Harikane et al., 2018; Stefanon et al., 2021;

Tacchella et al., 2018, e.g.,). Others, such as, Behroozi et al. (2019) find little change to $z \sim 5$ but significant evolution at higher redshifts. The results in this Thesis in COSMOS2020 show little-to-no evolution at least up to $z \sim 3$, but hint at evolution at higher redshifts. The large error bars of our work at $z > 3$ prevent us from concluding if there really is or not an evolution at these high redshifts. Therefore, this remains an open question that is starting to receive attention in the face of the results from early *JWST* data that indicate a significantly higher (up to a factor of 10) SFE (e.g., Furtak et al., 2022). Several upcoming surveys will allow measuring the SHMR down to the epoch of reionization. The Euclid Deep Survey will provide the necessary dataset to reach $z \sim 5-6$. Even deeper surveys from *JWST* will allow reaching deep in the reionization era. For example, COSMOS-Web (Kartaltepe et al., 2021) will reach a 5σ depth of AB $\sim$ 28 at $\sim 3\,\mu$m in 0.6 deg^2. This allows a considerable improvement in the mass-completeness, photo-zs and stellar masses to constrain the SHMR to $z \sim 6-7$ with clustering and GSMF.

The comparison of results in this Thesis to hydrodynamical simulations suggested that environmental quenching of satellites is either less efficient or quenching occurs later in the simulations. This result needs to be investigated from both sides. Constraining the high mass end of the SHMR requires statistically significant samples of high-mass galaxies. These are difficult to obtain and require large survey volumes in the near-infrared that can only be reached with the next-generation surveys such as Euclid Deep. On the other hand, these results can be driven by the limited realism of the simulations. For example, various other environmental quenching mechanisms are not well captured in the simulations, probably due to resolution effects. Therefore, to better understand the global contribution of satellites to the stellar mass budget in halos as well as their quenching mechanisms, improvements in both data and models based on hydrodynamical simulations are required.

A direct probe of halo properties would reveal more details about the coevolution of galaxies and halos. As discussed throughout this Thesis, weak lensing magnification offers a unique probe in a larger redshift regime than what is possible with shear-based methods. One of the most widely used assumptions in HOD based models is that the distribution of galaxies in a halo follows a similar NFW profile as the halo density. There is no strong evidence that this is true (e.g., Watson et al., 2012). Magnification, being a lensing probe, is sensitive to the inner halo profile and can therefore be used to place constrains. This is a direction in which the work of this Thesis should carry on in light of the upcoming surveys. To this end, extremely deep and relatively wide surveys such as the COSMOS-Web will provide sufficient background source densities to probe small scales with magnification.

Measurements of the SHMR are also important for calibration of hydrodynamical simulations and semi-analytical models that produce widely used mock catalogs. It is a standard benchmark to validate and even calibrate simulations with SHMR measurements, currently done at $z \sim 0$. Therefore, accurately measuring the SHMR, along with its dependence on redshift and galaxy properties such as colors and morphology, is of key importance. The upcoming *JWST* survey COSMOS-Web is designed to provide stringent constrains on the SHMR and its dependence on galaxy colors and morphology by incorporating galaxy-galaxy lensing at $z < 2.5$. This will be made possible in the 0.6 $\deg^2$ by the unprecedented resolution at $\sim 3\,\mu$m (Kartaltepe et al., 2021). Measuring the SHMR as a function of galaxy properties (e.g., for active and passive) is necessary to reveal more details on how quenching happens in halos. Finally, it would serve as a reference to compare and calibrate simulations, do stellar-to-halo mass conversions etc.

An area of interest which is still in the early stages is understanding the statistical relation between the cosmic web, dark matter halos and the surrounding gas that fuels galaxy growth. For example, hydrodynamical simulations have recently suggested a scenario where the vorticity of cosmic web filaments could play an important role in galaxy quenching at the edge of those filaments (Song et al., 2021). Such studies would require observations of HI in radio, Lyman-α absorption and X-ray to map the gas, as well as spectroscopic surveys to precisely map the cosmic web and measure gas properties. Such studies exist in the local Universe, but the challenge is to carry them out at high redshifts that correspond to the peak of the cosmic star-formation and the onset of quenching. This will be made possible by next-generation surveys such as Square Kilometre Array (SKA), WEAVE-QSO (Pieri et al., 2016), MOSAIC (Japelj et al., 2019), *Euclid* (Euclid Collaboration et al., 2022a) and Subaru Prime Focus Spectrograph (PFS) (Takada et al., 2014). In addition, the Nancy Grace Roman Space Telescope will provide an unprecedented imaging and spectroscopy with its large field of view (0.28 $\deg^2$) excellent resolution in optical to near-infrared wavelengths ($0.5 - 2.3\ \mu$m) that can reach 5σ depths of AB $\sim 26 - 28$ in one hour of imaging. All these observatories working in synergy will provide rich data sets to study various aspects of the galaxy-halo connection.

Finally, implications of understanding the galaxy-halo connection extend beyond the interest of galaxy formation and evolution and are important for precision cosmological studies. Incorporating galaxy clustering in cosmological studies is a standard procedure. However, if clustering measurements are pushed to small scales below the one-halo and non-linear regime, then models that consider the galaxy-halo connection become increasingly important. These small scales (~ 1 Mpc) are affected both by the cosmic expansion and baryons, and are thus difficult to model even with high-order perturbation theories. The galaxy-halo connection offers a non-linear model necessary for precision studies, provided the model is flexible enough to incorporate any effects from galaxy assembly bias, baryonic impact on clustering and abundance of galaxies in halos etc. This also highlights that theoretical models of the galaxy-halo connection need to continue developing in line with the data.

The galaxy-halo connection has many facets which are important in building a complete picture of our Universe. This Thesis focused on one particular facet – the stellar-to-halo mass relation. By studying its cosmic evolution for different galaxy populations, it aimed at providing new insights and a reference measurement of the SHMR to be various purposes such as stellar-to-halo mass conversions, comparison/calibration of simulations etc. This work took the approach that future studies will naturally take to investigate the galaxy-halo connection – namely, analyze a homogeneous multi-wavelength survey using flexible modeling techniques that consider small, non-linear scales affected by baryonic physics.

This Thesis naturally lays the foundations for analyzing next-generation surveys, which will gradually reveal the intricate connection between galaxies and halos.

A

APPENDIX 1

A.1 IMPACT OF $N(z)$ ON $w(\theta)$

The assumed redshift distribution of the sample can lead to differences in the modeled $w(\theta)$ that are considerably larger than the measurement uncertainties that may bias the inferred model parameters. We acknowledge, following the discussion in Ilbert et al., 2021, that the $N(z)$ constructed by stacking the posteriors $\mathcal{P}(z)$ for each individual galaxy is more representative of the true underlying distribution of the sample, compared to simply stacking the likelihoods $\mathcal{L}(z)$ as outputted from the template fitting code `LePhare`. In Fig. A.1, we show the $N(z)$ for the sample in the bin $0.8 < z_{\mathrm{phot}} < 1.1$ obtained using $\mathcal{L}(z)$ in red, $\mathcal{P}(z)$ in blue, and the histogram of the sources that have a spectroscopic redshift available in green. The figure shows that indeed stacking the individual $\mathcal{P}(z)$ agrees better with the spec-z histogram, especially in the tails which contribute the most in the $w(\theta)$ amplitude. Fig A.2 shows the model $w(\theta)$ of the two $N(z)$ (upper panel). The relative difference between the two is about 35% as shown in the bottom panel. Furthermore, we explore the effect of a bias in the mean redshift as inferred from the N(z) on the $w(\theta)$. For this we use a Gaussian distribution centered at the mean redshift of the bin and then move the mean by -0.02. This results in relative difference in less than 3% which is safely within the error bars of the measurements (which are about 10%).

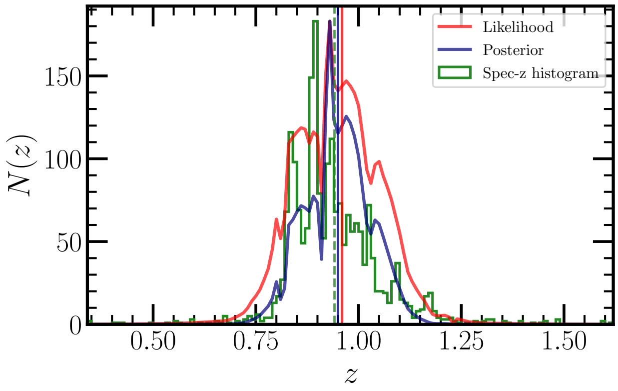


Figure A.1: Redshift distributions based on spec-z histogram and stacking of photo-z likelihood and posterior distributions. All distributions are normalized to the maximum of the spec-z histogram.

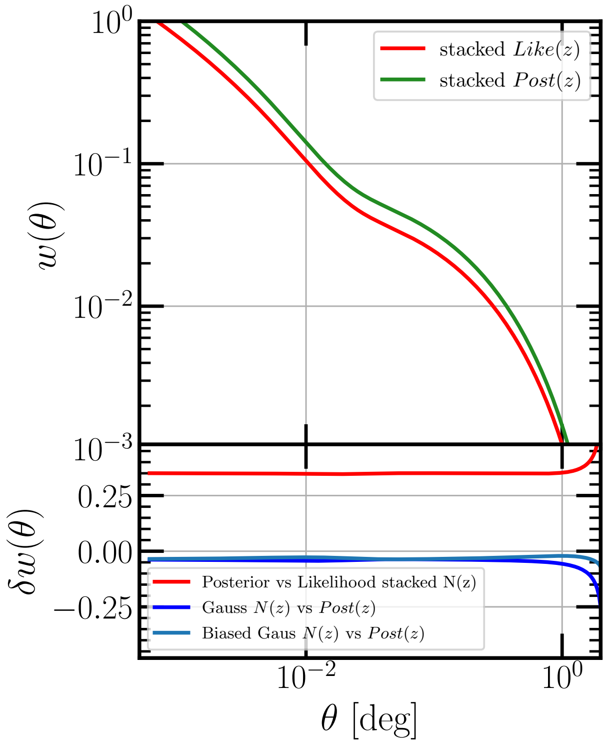


Figure A.2: Effect of the redshift distribution on the model correlation function

A.2 POSTERIOR PROBABILITIES OF THE PARAMETERS

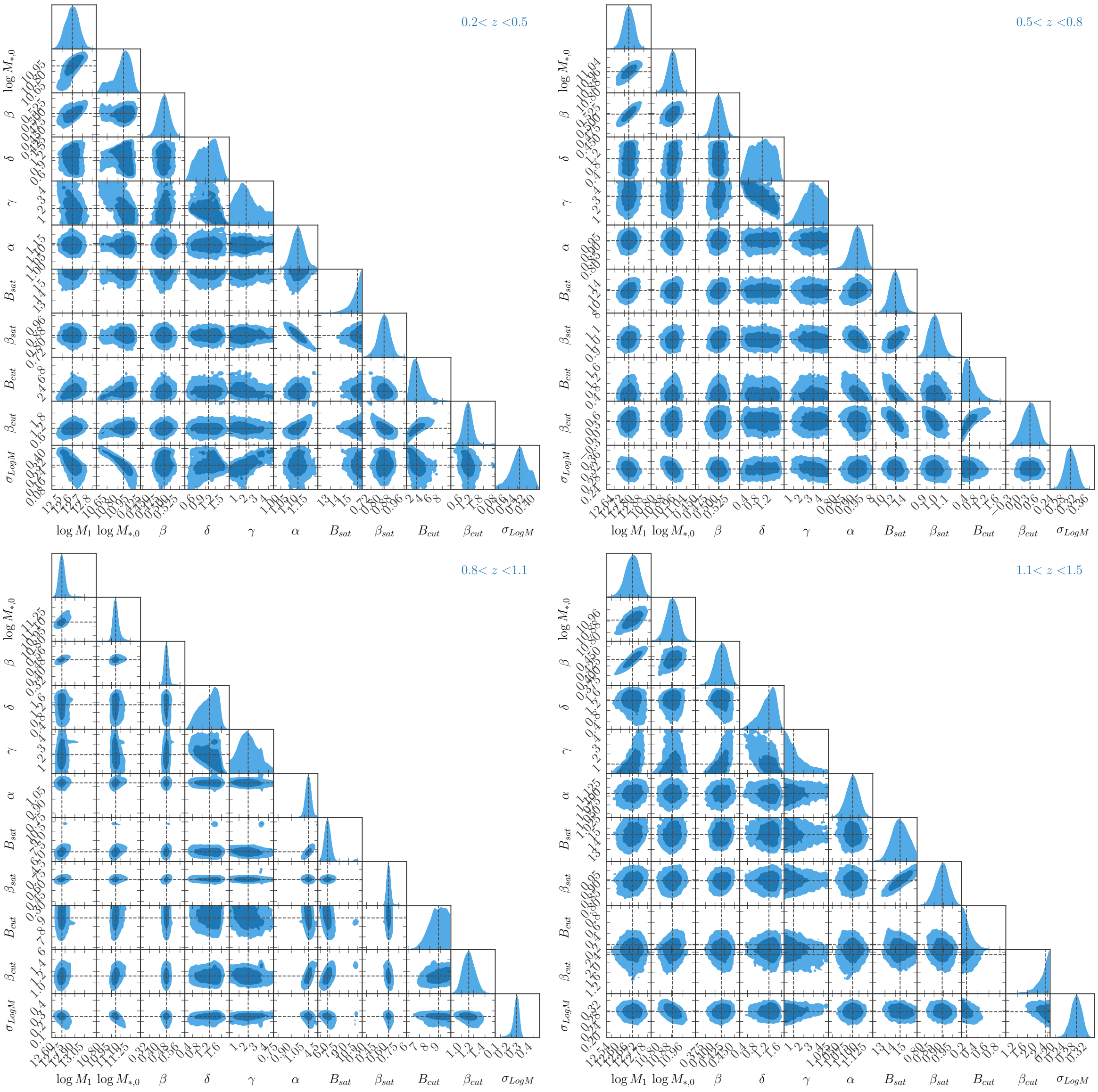


Figure A.3: Posterior distributions for each redshift fit

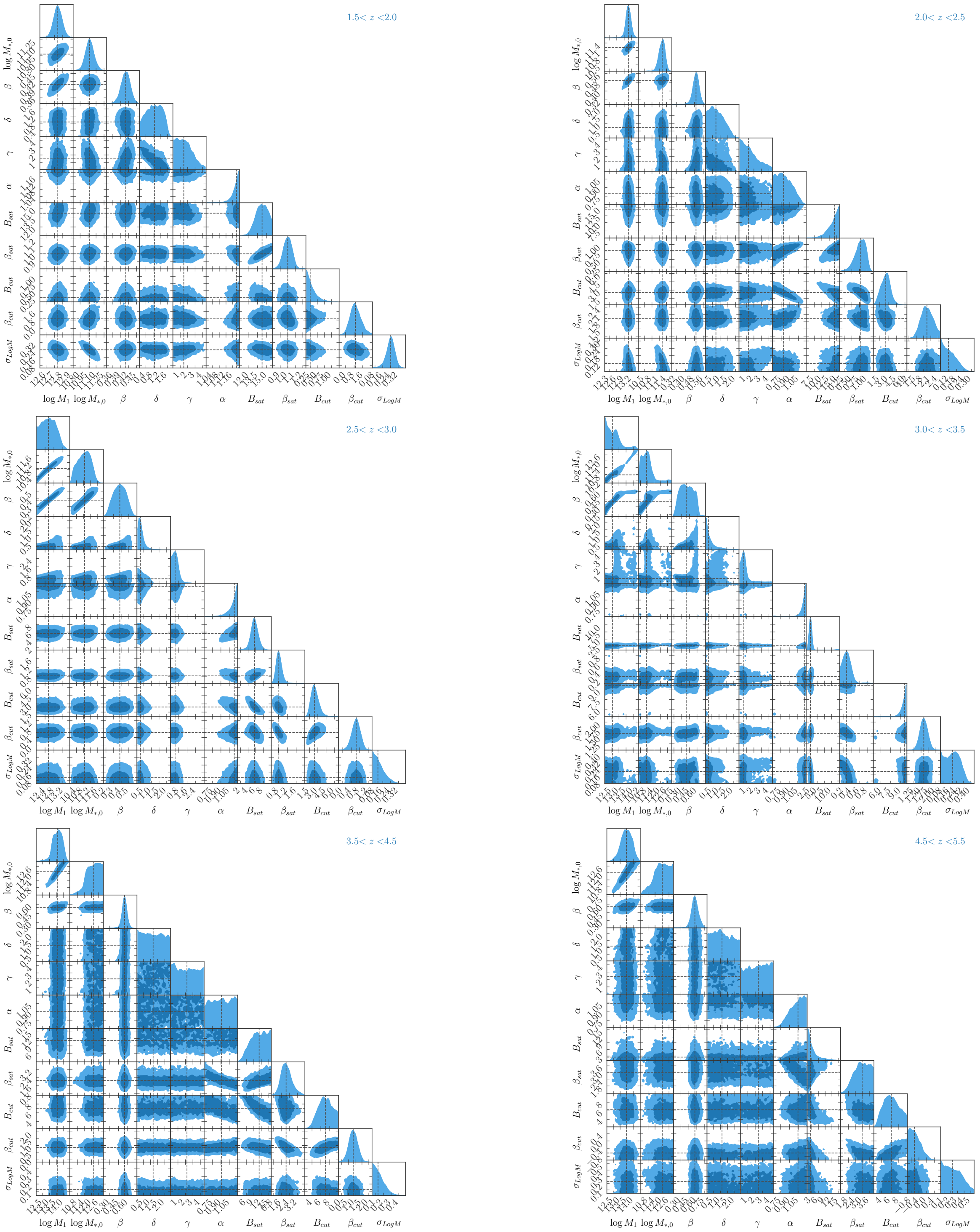


Figure A.4: Posterior distributions for each redshift fit

A.3 BEST-FIT VALUES OF THE MODEL PARAMETERS

z-bin	$\log M_1$	$\log M_{*,0}$	β	δ	γ	α	B_{sat}	β_{sat}	B_{cut}	β_{cut}	σ_{LogM}	$\chi^2_{\mathrm{reduced}}$
$0.2 < z < 0.5$	$12.629^{+0.069}_{-0.072}$	$10.855^{+0.099}_{-0.216}$	$0.487^{+0.016}_{-0.016}$	$0.935^{+0.295}_{-0.335}$	$1.939^{+2.515}_{-1.058}$	$1.056^{+0.029}_{-0.030}$	$15.467^{+0.394}_{-0.753}$	$0.845^{+0.040}_{-0.040}$	$2.045^{+1.288}_{-0.974}$	$0.745^{+0.231}_{-0.251}$	$0.268^{+0.009}_{-0.080}$	2.6
$0.5 < z < 0.8$	$12.793^{+0.044}_{-0.043}$	$10.927^{+0.038}_{-0.041}$	$0.502^{+0.013}_{-0.013}$	$0.802^{+0.446}_{-0.407}$	$3.132^{+1.488}_{-1.143}$	$0.905^{+0.033}_{-0.038}$	$12.221^{+1.082}_{-1.206}$	$0.985^{+0.057}_{-0.052}$	$0.277^{+0.290}_{-0.168}$	$0.076^{+0.303}_{-0.386}$	$0.293^{+0.002}_{-0.019}$	2.5
$0.8 < z < 1.1$	$12.730^{+0.044}_{-0.039}$	$11.013^{+0.057}_{-0.032}$	$0.454^{+0.014}_{-0.014}$	$1.109^{+0.329}_{-0.455}$	$1.925^{+1.441}_{-1.014}$	$1.065^{+0.030}_{-0.030}$	$5.416^{+0.552}_{-0.507}$	$0.612^{+0.019}_{-0.018}$	$8.845^{+1.261}_{-1.020}$	$1.098^{+0.113}_{-0.095}$	$0.250^{+0.002}_{-0.032}$	10.2
$1.1 < z < 1.5$	$12.673^{+0.065}_{-0.085}$	$10.967^{+0.071}_{-0.076}$	$0.393^{+0.022}_{-0.026}$	$0.746^{+0.292}_{-0.189}$	$0.335^{+0.443}_{-0.235}$	$1.078^{+0.017}_{-0.017}$	$15.015^{+0.639}_{-0.821}$	$0.906^{+0.031}_{-0.038}$	$0.101^{+0.065}_{-0.046}$	$4.197^{+0.222}_{-0.463}$	$0.167^{+0.008}_{-0.077}$	3.8
$1.5 < z < 2.0$	$12.787^{+0.067}_{-0.065}$	$11.040^{+0.163}_{-0.137}$	$0.410^{+0.015}_{-0.018}$	$0.716^{+0.480}_{-0.423}$	$1.312^{+1.431}_{-0.878}$	$1.213^{+0.027}_{-0.031}$	$14.168^{+0.981}_{-0.984}$	$0.951^{+0.063}_{-0.072}$	$0.099^{+0.168}_{-0.068}$	$1.848^{+1.667}_{-1.209}$	$0.211^{+0.012}_{-0.203}$	6.1
$2.0 < z < 2.5$	$13.097^{+0.087}_{-0.101}$	$11.254^{+0.101}_{-0.152}$	$0.495^{+0.022}_{-0.028}$	$0.668^{+0.562}_{-0.420}$	$1.077^{+1.325}_{-0.765}$	$0.793^{+0.158}_{-0.130}$	$14.156^{+1.358}_{-2.419}$	$0.751^{+0.164}_{-0.223}$	$2.501^{+0.776}_{-0.860}$	$1.968^{+0.290}_{-0.236}$	$0.050^{+0.010}_{-0.041}$	3.8
$2.5 < z < 3.0$	$12.627^{+0.325}_{-0.380}$	$10.920^{+0.333}_{-0.440}$	$0.393^{+0.077}_{-0.089}$	$0.274^{+0.218}_{-0.130}$	$0.446^{+0.342}_{-0.295}$	$1.246^{+0.041}_{-0.098}$	$6.539^{+0.921}_{-0.862}$	$0.772^{+0.107}_{-0.093}$	$1.763^{+0.679}_{-0.528}$	$0.686^{+0.237}_{-0.248}$	$0.083^{+0.007}_{-0.057}$	5.7
$3.0 < z < 3.5$	$12.820^{+0.972}_{-0.519}$	$11.067^{+1.124}_{-0.526}$	$0.465^{+0.138}_{-0.154}$	$0.354^{+0.988}_{-0.260}$	$0.741^{+2.066}_{-0.443}$	$1.251^{+0.037}_{-0.092}$	$2.592^{+0.425}_{-0.293}$	$0.067^{+0.109}_{-0.043}$	$13.122^{+1.195}_{-1.955}$	$2.075^{+0.181}_{-0.224}$	$0.013^{+0.001}_{-0.009}$	4.1
$3.5 < z < 4.5$	$13.638^{+0.303}_{-0.331}$	$12.222^{+0.533}_{-0.558}$	$0.551^{+0.038}_{-0.034}$	$1.557^{+1.187}_{-1.075}$	$3.149^{+2.279}_{-2.236}$	$0.930^{+0.253}_{-0.231}$	$9.930^{+4.068}_{-4.205}$	$1.838^{+0.694}_{-0.579}$	$7.588^{+2.985}_{-2.038}$	$1.017^{+0.415}_{-0.368}$	$0.085^{+0.008}_{-0.061}$	3.5
$4.5 < z < 5.5$	$13.547^{+0.378}_{-0.392}$	$12.105^{+0.605}_{-0.676}$	$0.567^{+0.066}_{-0.062}$	$1.427^{+1.262}_{-0.971}$	$3.225^{+2.248}_{-2.167}$	$1.031^{+0.194}_{-0.264}$	$2.630^{+4.383}_{-1.282}$	$3.108^{+1.816}_{-1.062}$	$7.020^{+3.104}_{-2.105}$	$-0.621^{+0.446}_{-0.274}$	$0.226^{+0.017}_{-0.151}$	1.9

Table A.1: Best-fit values of the model parameters in the ten redshift bins

B

APPENDIX 2

B.1 DESCRIPTION OF THE HYDRODYNAMICAL SIMULATIONS

B.1.1 *Main characteristics and catalog extraction*

HORIZON-AGN

has been produced with the adaptive mesh refinement code RAMSES (Teyssier, 2002), using WMAP7 cosmological parameters (Komatsu et al., 2011) in a box of size $100\,h^{-1}$ Mpc a side [1] and a DM mass resolution of $M_{\mathrm{DM,res}} = 8 \times 10^{7}\,\mathrm{M}_{\odot}$. The simulation includes gas heating, cooling, star-formation, feedback from stellar winds, type Ia and type II supernovae with mass, energy, and metal release in the interstellar-medium. The simulation also follows the formation of black holes and energy release from AGN in a quasar or radio mode depending on the accretion rate. Full details on the subgrid implementation are given in Dubois et al., 2014.

Galaxies and halos are identified using the ADAPTAHOP structure finder (Aubert, Pichon and Colombi, 2004; Tweed et al., 2009) applied to the distribution of star and DM particles respectively, as described in previous works (e.g., Dubois et al., 2014). Galaxy masses are obtained by summing the masses of all individual stellar particles within twice the effective radius of the galaxies, while halo masses are obtained by summing all DM particles within the halos. Galaxies and halos are positionally matched. The central galaxies of a halo is defined as the most massive galaxies within 0.1 times the halo virial radius.

We note that stellar mass losses in HORIZON-AGN were implemented assuming a Salpeter IMF (Salpeter, 1955). This can lead to ~ 0.1 dex more stellar masses at later time than when assuming a Chabrier IMF (Chabrier, 2003). The simulated masses were therefore rescaled accordingly for a consistent comparison with both the other simulations and the observational data (which assumes a Chabrier IMF when deriving the mass from the photometry).

The simulation is in relatively good agreement with observations up to $z = 4$. Known discrepancies include the overestimation of galaxy masses, especially at low mass (Kaviraj

---

[1] A light cone has also been extracted on-the-fly in order to build realistic mocks (Laigle et al., 2019; Gouin et al., 2019). In particular, based on these mocks, Hatfield et al., 2019 showed that the propagation of statistical and systematic uncertainties inherited from redshift and mass photometric estimates lead to an underestimation of the clustering amplitude by ~ 0.1 dex. In the present work, we use galaxy and halo catalogs extracted from the snapshot data, but we checked that using the light cone data does not significantly change the measured SHMR.

et al., 2017). These low-mass galaxies are on overall too quenched, an indication that star formation at high redshift was not regulated enough and the gas in the environment of these galaxies was too rapidly consumed. In addition, the bimodality is not as well-marked as in observations due to residual star-formation in massive galaxies, possibly because of a slightly insufficient strength for the AGN feedback.

TNG100-1

has been produced with the moving-mesh code AREPO (Springel, 2010) using cosmological parameters from Planck Collaboration et al., 2016 in a box of size 75 h^{-1} Mpc a side and a DM mass resolution of $M_{\rm DM,res} = 7.5 \times 10^6\,{\rm M}_\odot$. It follows magnetic fields in addition to the hydrodynamical processes above-mentioned for the Horizon-AGN simulation (with different subgrid implementations, see the discussion below). Full details on the subgrid implementation are given in Pillepich et al., 2018; Springel et al., 2018.

Halos, subhalos, and galaxies have been identified using the FoF (Davis et al., 1985) and SUBFIND (which, within a halo identified with the FoF technique, relies on all particle species to identify the galaxy, see Springel et al., 2001) algorithms, as described in, for instance, Pillepich et al., 2018. We downloaded the halo and galaxy catalogs from the public website[2]. In the following, galaxy masses are estimated by summing all stellar mass particles within twice the stellar effective radius, while halo masses are taken as being the sum of all individual dark matter particles in the identified halos, which matches the definition of halo and galaxy masses in Horizon-AGN. Galaxies that do not reside within R_{200} of a larger halo are identified as centrals. Galaxy clustering has been measured in Springel et al., 2018 and the SHMR has been presented in Pillepich et al., 2018 (see also Engler et al., 2020). In particular, these authors investigated how the chosen estimate for stellar mass modifies the measured relation. While we are aware of this discussion, we chose in this work to measure stellar mass within twice the effective radius for consistency between different simulated datasets. Such measurement is also closer to our observational estimate of stellar masses (which derive from SED-fitting on the total magnitudes – and not on aperture magnitudes).

EAGLE

has been produced using a modified version of the N-Body Tree-PM smoothed particle hydrodynamics (SPH) code GADGET-3 (Springel, 2005), adopting cosmological parameters from Planck (Collaboration et al., 2014). In this work, we use the reference run called Ref-L0100N1504,corresponding to a box of 100 comoving Mpc a side with a DM mass resolution of $M_{\rm DM,res} = 9.7 \times 10^6\,{\rm M}_\odot$. EAGLE follows gas heating and cooling, star-formation, feedback from stellar winds and AGB stars, along with type Ia and type II supernovae and from AGN. Full details on the subgrid implementation are given in Crain et al., 2015 and Schaye et al., 2015.

---

[2] https://www.tng-project.org/data/

As for TNG100-1, the halos, subhalos, and galaxies have been identified using the FoF (Davis et al., 1985) and SUBFIND (Springel et al., 2001) algorithms, as described in McAlpine et al., 2016. We downloaded halo and galaxy catalogs from the public website[3]. Galaxy and halo masses were obtained by summing the masses of all stellar and dark matter particles, respectively, which are part of the objects. The central galaxy is taken as the one which contains the particle with the lowest value of the gravitational potential. The redshift evolution of the SHMR for central galaxies in EAGLE simulation was briefly presented in Matthee et al., 2017.

B.1.2 *Implementation of subgrid recipes*

Our comparisons of the SHMR in observations and simulations has highlighted two main discrepancies: (1) in HORIZON-AGN, the SHMR is always overestimated with respect to observations and the two other simulations, especially at low masses; (2) in all simulations, the relative contribution of satellites with respect to centrals is too large compared to observations.

With regard to (1), we note first that in EAGLE, the free parameters of the subgrid (stellar and AGN) feedback model were calibrated so that the simulations reproduce the galaxy stellar mass function and SHMR, galaxy sizes, and the empirical relation between black hole mass and stellar mass all at $z = 0$ (Schaye et al., 2015). In TNG100-1, the calibration was performed against the star formation rate density as a function of cosmic time and the stellar mass function and SHMR both at $z = 0$ (Pillepich et al., 2018). In HORIZON-AGN however, the calibration was less constraining since only the efficiency of AGN feedback was tuned so that the black hole–galaxy scaling relation at $z = 0$ was reproduced (see Dubois et al., 2012, for details). It is therefore not surprising that both EAGLE and TNG100-1 better reproduce, overall, the stellar mass function (Schaye et al., 2015; Pillepich et al., 2018) and SHMR (Fig. 7.12) at low masses, because their calibration specifically constrain stellar feedback, which is not the case for HORIZON-AGN.

Without pretending to exhaustively discuss the differences in stellar feedback and star-formation implementation between HORIZON-AGN and the two other simulations, we instead opt to highlight a few aspects. In the EAGLE simulation as explained in Crain et al., 2015, the amount of injected energy from feedback depends on the local properties of the gas (it decreases with metallicity and increases with gas density), the calibration being adjusted in order to reproduce the local stellar mass function. This tuning contributes to increase the supernova feedback energy at high redshift beyond the energy available for their adopted Chabrier IMF (Crain et al., 2015). Furthermore, the stochasticity of energy deposition of supernovae is artificially increased to enhance the impact of their feedback in terms of wind mass loading and quenching of star formation[4] (Dalla Vecchia and Schaye, 2012). Finally, the EAGLE star-formation law does not follow the standard Kennicutt-Schmidt

---

[3] http://virgodb.dur.ac.uk:8080/Eagle/

[4] While this implementation of stochastic feedback enhance the impact of feedback in those (GADGET) SPH simulation, it makes no difference with standard thermal release when simulated with (RAMSES) AMR (Rosdahl et al., 2017).

prescription (Kennicutt, 1998, as adopted in Horizon-AGN); rather, it depends on pressure instead of density and includes a metallicity-dependent density threshold (against a simple density threshold in Horizon-AGN), which tends to reduce star-formation in metal-rich regions. Those features are likely to contribute to a more efficient quenching of star-formation in EAGLE compared to Horizon-AGN.

Star-formation in TNG100 follows the empirical Kennicutt-Schmidt relation, however, feedback implementation is sensibly different, as described in Pillepich et al., 2018. More specifically, the energy transfer from SNe to large-scale galactic winds is very efficient because of the hydrodynamical decoupling of the launched wind gas from the dense star-forming gas, until they recouple hydrodynamically with the circum-galactic gas (Springel and Hernquist, 2003; Vogelsberger et al., 2013). These authors used a wind velocity that is proportional to the local DM velocity dispersion and cosmic time so that the winds are faster in more massive halos and at lower redshift. Similarly to the EAGLE model, the given supernovae energy to the gas is higher for energy deposit in lower metallicity gas, with the same scaling with metallicity as in EAGLE. Taken together, these features contribute to make stellar feedback more effective at high redshift in TNG100 compared to Horizon-AGN.

Finally, we note that many missing mechanisms could be naturally added to the Horizon-AGN model in order to increase the strength of stellar feedback in low-mass galaxies at high redshift in a physically motivated way (without necessarily requiring empirical tuning of parameters). These processes include radiation from stars (see the discussion below), cosmic ray feedback from supernovae (e.g., Booth et al., 2013; Salem and Bryan, 2014; Dashyan and Dubois, 2020), a gravo-turbulent model for star formation (e.g., Nuñez-Castiñeyra et al., 2021; Dubois et al., 2021), or adopting an IMF varying with redshift, stellar density, or metallicity (e.g., Applebaum et al., 2020; Prgomet et al., 2021).

With regard to (2), however, the efficient stellar feedback implemented in TNG100 and EAGLE do not prevent the too large satellite fraction (Fig. 7.8) and excessive contribution of satellites to the total SHMR (Fig. 7.9). As previously noted in Sec. **??**, such a discrepancy must be due to a lack of satellite quenching at high redshift, making these satellites to grow too massive (with respect to observations) before being eventually quenched. Galaxy pre-processing by stellar radiation feedback (e.g., Rosdahl et al., 2013; Hopkins et al., 2020) resulting in less tightly-bound galaxies could be a prerequisite to efficiently quench satellites (once they have been accreted) through tidal stripping, as suggested by Costa, Rosdahl and Kimm, 2019.

Finally, some studies mention the possible role of AGN feedback from the central galaxy in quenching its satellites (e.g., Dashyan et al., 2019; Martín-Navarro, Burchett and Mezcua, 2019): the lack of quenching in high-redshift satellites could therefore also be the consequence of an imperfect AGN feedback implementation at high redshift, although hydrodynamical simulations show that the activity of SNe in low mass galaxies quench the growth of black holes and their associated AGN feedback (e.g., Dubois et al., 2015; Habouzit, Volonteri and Dubois, 2017; Anglés-Alcázar et al., 2017).

C

PUBLICATIONS

This Thesis includes work by the author that has been published in various journals. Here we list the first author publications and all the other papers published or submitted in collaborations.

C.1 FIRST AUTHOR PUBLICATIONS

C.1.1 *COSMOS2020: Cosmic evolution of the stellar-to-halo mass relation for central and satellite galaxies up to $z \sim 5$*

Shuntov et al., A&A 664, A61 (2022)

Abstract

We used the COSMOS2020 catalog to measure the stellar-to-halo mass relation (SHMR) divided by central and satellite galaxies from $z = 0.2$ to $z = 5.5$. Starting from accurate photometric redshifts, we measured the near-infrared selected two-point angular correlation and stellar mass functions in ten redshift bins. We used a phenomenological model that parametrizes the stellar-to-halo mass relation for central galaxies and the number of galaxies inside each halo to describe our observations. This model qualitatively reproduces our measurements and their dependence on the stellar mass threshold. Surprisingly, the mean halo occupation distribution only shows a mild evolution with redshift suggesting that galaxies occupy halos similarly throughout cosmic time. At each redshift, we measured the ratio of stellar mass to halo mass, M_*/M_h, which shows the characteristic strong dependence of halo mass with a peak at $M_h^{\rm peak} \sim 2 \times 10^{12}\, M_\odot$. For the first time, using a joint modeling of clustering and abundances, we measured the evolution of $M_h^{\rm peak}$ from $z = 0.2$ to $z = 5.5$. $M_h^{\rm peak}$ increases gradually with redshift from $\log M_h^{\rm peak}/M_\odot \sim 12.1$ at $z \sim 0.3$ to $\log M_h^{\rm peak}/M_\odot \sim 12.3$ at $z \sim 2$, and up to $\log M_h^{\rm peak}/M_\odot \sim 12.9$ at $z \sim 5$. Similarly, the stellar mass peak $M_*^{\rm peak}$ increases with redshift from $\log M_*^{\rm peak}/M_\odot \sim 10.5$ at $z \sim 0.3$ to $\log M_*^{\rm peak}/M_\odot \sim 10.9$ at $z \sim 3$. The SHMR ratio at the peak halo mass remains almost constant with redshift. These results are in accordance with the scenario in which the peak of star-formation efficiency moves toward more massive halos at higher redshifts. We also measured the fraction of satellites as a function of stellar mass and redshift. For all stellar mass thresholds, the satellite fraction decreases at higher redshifts. At a given red-

shift, there is a higher fraction of low-mass satellites and this fraction reaches a plateau at $\sim 25\%$ at $z \sim 1$. The satellite contribution to the total stellar mass budget in halos becomes more important than that of the central at halo masses of about $M_h > 10^{13}\,M_\odot$ and always stays below the peak, indicating that quenching mechanisms are present in massive halos that keep the star-formation efficiency low. Finally, we compared our results with three hydrodynamical simulations: HORIZON-AGN, TNG100 of the ILLUSTRISTNG project, and EAGLE. We find that the most significant discrepancy is at the high-mass end, where the simulations generally show that satellites have a higher contribution to the total stellar mass budget than the observations. This, together with the finding that the fraction of satellites is higher in the simulations, indicates that the feedback mechanisms acting in both group- and cluster-scale halos appear to be less efficient in quenching the mass assembly of satellites – and that quenching occurs much later in the simulations.

C.1.2 *PhotoWeb redshift: boosting photometric redshift accuracy with large spectroscopic surveys*

Shuntov et al., AA 636, A90 (2020)

Abstract

Improving distance measurements in large imaging surveys is a major challenge to better reveal the distribution of galaxies on a large scale and to link galaxy properties with their environments. As recently shown, photometric redshifts can be efficiently combined with the cosmic web extracted from overlapping spectroscopic surveys to improve their accuracy. In this paper we apply a similar method using a new generation of photometric redshifts based on a convolution neural network (CNN). The CNN is trained on the SDSS images with the main galaxy sample (SDSS-MGS, $r \leq 17.8$) and the GAMA spectroscopic redshifts up to $r \sim 19.8$. The mapping of the cosmic web is obtained with 680,000 spectroscopic redshifts from the MGS and BOSS surveys. The redshift probability distribution functions (PDF), which are well calibrated (unbiased and narrow, ≤ 120 Mpc), intercept a few cosmic web structures along the line of sight. Combining these PDFs with the density field distribution provides new photometric redshifts, , whose accuracy is improved by a factor of two (i.e., $\sigma \sim 0.004(1+z)$) for galaxies with $r \leq 17.8$. For half of them, the distance accuracy is better than 10 cMpc. The narrower the original PDF, the larger the boost in accuracy. No gain is observed for original PDFs wider than 0.03. The final PDFs also appear well calibrated. The method performs slightly better for passive galaxies than star-forming ones, and for galaxies in massive groups since these populations better trace the underlying large-scale structure. Reducing the spectroscopic sampling by a factor of 8 still improves the photometric redshift accuracy by 25%. Finally, extending the method to galaxies fainter than the MGS limit still improves the redshift estimates for 70% of the galaxies, with a gain in accuracy of 20% at low z where the resolution of the cosmic web is the highest. As two competing factors contribute to the performance of the method, the

photometric redshift accuracy and the resolution of the cosmic web, the benefit of combining cosmological imaging surveys with spectroscopic surveys at higher redshift remains to be evaluated.

C.2 OTHER PUBLICATIONS IN COLLABORATION

PUBLICATIONS BASED ON COSMOS2020

C.2.1 *COSMOS2020: A panchromatic view of the Universe to $z \sim 10$ from two complementary catalogs*

J. R. Weaver, O. B. Kauffmann, O. Ilbert, H. J. McCracken, A. Moneti, S. Toft, G. Brammer, M. Shuntov et al. ApJS 258 11 (2022)

Abstract

The Cosmic Evolution Survey (COSMOS) has become a cornerstone of extragalactic astronomy. Since the last public catalog in 2015, a wealth of new imaging and spectroscopic data has been collected in the COSMOS field. This paper describes the collection, processing, and analysis of this new imaging data to produce a new reference photometric redshift catalog. Source detection and multi-wavelength photometry is performed for 1.7 million sources across the 2 deg^2 of the COSMOS field, ∼966 000 of which are measured with all available broad-band data using both traditional aperture photometric methods and a new profile-fitting photometric extraction tool, , which we have developed. A detailed comparison of the two resulting photometric catalogs is presented. Photometric redshifts are computed for all sources in each catalog utilizing two independent photometric redshift codes. Finally, a comparison is made between the performance of the photometric methodologies and of the redshift codes to demonstrate an exceptional degree of self-consistency in the resulting photometric redshifts. The $i < 21$ sources have sub-percent photometric redshift accuracy and even the faintest sources at $25 < i < 27$ reach a precision of 5 %. Finally, these results are discussed in the context of previous, current, and future surveys in the COSMOS field. Compared to COSMOS2015, reaches the same photometric redshift precision at almost one magnitude deeper. Both photometric catalogs and their photometric redshift solutions and physical parameters will be made available through the usual astronomical archive systems (ESO Phase 3, IPAC IRSA, and CDS).

C.2.2 *COSMOS2020: UV selected galaxies at $z \geq 7.5$*

O. B. Kauffmann, O. Ilbert, J. R. Weaver, H. J. McCracken, B. Milvang-Jensen, G. Brammer, I. Davidzon, O. Le Fèvre, D. Liu, B. Mobasher, A. Moneti, M. Shuntov et al., eprint arXiv:2207.11740, accepted in A&A

Abstract

This paper presents a new search for $z \geq 7.5$ galaxies using the COSMOS2020 photometric catalogues. Finding galaxies at the reionisation epoch through deep imaging surveys remains observationally challenging. The larger area covered by ground-based surveys like COSMOS enables the discovery of the brightest galaxies at these high redshifts. Covering $1.4\,\mathrm{deg}^2$, our COSMOS catalogues were constructed from the latest UltraVISTA data release (DR4) combined with the final *Spitzer*/IRAC COSMOS images and the Hyper-Suprime-Cam Subaru Strategic Program DR2 release. We identify 17 new $7.5 < z < 10$ candidate sources, and confirm 15 previously published candidates. Using deblended photometry extracted by fitting surface brightness models on multi-band images, we select four candidates which would be rejected using fixed aperture photometry. We test the robustness of all our candidates by comparing six different photometric redshift estimates. Finally, we compute the galaxy UV luminosity function in three redshift bins centred at $z = 8, 9, 10$. We find no clear evolution of the number density of the brightest galaxies $M_{\mathrm{UV}} < -21.5$, in agreement with previous works. Rapid changes in the quenching efficiency or attenuation by dust could explain such lack of evolution between $z \sim 8$ and $z \sim 9$. A spectroscopic confirmation of the redshifts, already planned with JWST and the Keck telescopes, will be essential to confirm our results.

C.2.3 *COSMOS2020: Manifold learning to estimate physical parameters in large galaxy surveys*

Davidzon, I. ; Jegatheesan, K. ; Ilbert, O. ; de la Torre, S. ; Leslie, S. K. ; Laigle, C. ; Hemmati, S. ; Masters, D. C. ; Blanquez-Sese, D. ; Kauffmann, O. B. ; Magdis, G. E. ; Małek, K. ; McCracken, H. J. ; Mobasher, B. ; Moneti, A. ; Sanders, D. B. ; Shuntov, M et al., AA 665, A34 (2022)

Abstract

We present a novel method for estimating galaxy physical properties from spectral energy distributions (SEDs) as an alternative to template fitting techniques and based on self-organizing maps (SOMs) to learn the high-dimensional manifold of a photometric galaxy catalog. The method has previously been tested with hydrodynamical simulations in Davidzon et al. (2019, MNRAS, 489, 4817), however, here it is applied to real data for the first time. It is crucial for its implementation to build the SOM with a high-quality panchromatic data set, thus we selected "COSMOS2020" galaxy catalog for this purpose. After the training and calibration steps with COSMOS2020, other galaxies can be processed through SOMs to obtain an estimate of their stellar mass and star formation rate (SFR). Both quantities resulted in a good agreement with independent measurements derived from more extended photometric baseline and, in addition, their combination (i.e., the SFR vs. stellar mass diagram) shows a main sequence of star-forming galaxies that is consistent with the findings of previous studies. We discuss the advantages of this method compared to

traditional SED fitting, highlighting the impact of replacing the usual synthetic templates with a collection of empirical SEDs built by the SOM in a "data-driven" way. Such an approach also allows, even for extremely large data sets, for an efficient visual inspection to identify photometric errors or peculiar galaxy types. While also considering the computational speed of this new estimator, we argue that it will play a valuable role in the analysis of oncoming large-area surveys such as Euclid of the Legacy Survey of Space and Time at the Vera C. Rubin Telescope.

C.2.4 *COSMOS2020: Ubiquitous AGN Activity of Massive Quiescent Galaxies at 0 < z < 5 Revealed by X-Ray and Radio Stacking*

Ito, Kei ; Tanaka, Masayuki ; Miyaji, Takamitsu ; Ilbert, Olivier ; Kauffmann, Olivier B. ; Koekemoer, Anton M. ; Marchesi, Stefano ; Shuntov, Marko et al., ApJ 929 53 (2022)

Abstract

We report a galaxy group candidate HPC1001 at z3.7 in the COSMOS field. This structure was selected as a high galaxy overdensity at $z > 3$ in the COSMOS2020 catalog. It contains ten candidate members, of which eight are assembled in a 10×10 area with the highest sky density among known protoclusters and groups at $z > 3$. Four out of ten sources were also detected at 1.2 mm with Atacama Large Millimeter Array continuum observations. Photometric redshifts, measured by four independent methods, fall within a narrow range of $3.5 < z < 3.9$ and with a weighted average of $z = 3.65 \pm 0.07$. The integrated far-IR-to-radio spectral energy distribution yields a total UV and IR star formation rate SFR≈ $900\,M_\odot yr^1$. We also estimated a halo mass of $\sim 10^{13}\,M_\odot$ for the structure, which at this redshift is consistent with potential cold gas inflow. Remarkably, the most massive member has a specific star formation rate and dust to stellar mass ratio of $M_{\rm dust}/M$ that are both significantly lower than that of star-forming galaxies at this redshift, suggesting that HPC1001 could be a $z \approx 3.7$ galaxy group in maturing phase. If confirmed, this would be the earliest structure in maturing phase to date, and an ideal laboratory to study the formation of the earliest quiescent galaxies as well as cold gas accretion in dense environments.

C.2.5 *A galaxy group candidate at $z \sim$ 3.7 in the COSMOS field*

Bjerregaard Sillassen, Nikolaj ; Jin, Shuowen ; Magdis, Georgios E. ; Daddi, Emanuele ; Weaver, John R. ; Gobat, Raphael ; Kokorev, Vasily ; Valentino, Francesco ; Finoguenov, Alexis ; Shuntov, Marko et al., A&A Letter in press; doi:10.1051/0004-6361/202244661

Abstract

We report a galaxy group candidate HPC1001 at z3.7 in the COSMOS field. This structure was selected as a high galaxy overdensity at $z > 3$ in the COSMOS2020 catalog. It contains ten candidate members, of which eight are assembled in a 10×10 area with the highest sky density among known protoclusters and groups at $z > 3$. Four out of ten sources were also detected at 1.2 mm with Atacama Large Millimeter Array continuum observations. Photometric redshifts, measured by four independent methods, fall within a narrow range of $3.5 < z < 3.9$ and with a weighted average of $z = 3.65 \pm 0.07$. The integrated far-IR-to-radio spectral energy distribution yields a total UV and IR star formation rate SFR≈ $900\,M_\odot yr^1$. We also estimated a halo mass of $\sim 10^{13}\,M_\odot$ for the structure, which at this redshift is consistent with potential cold gas inflow. Remarkably, the most massive member has a specific star formation rate and dust to stellar mass ratio of $M_{\rm dust}/M$ that are both significantly lower than that of star-forming galaxies at this redshift, suggesting that HPC1001 could be a $z \approx 3.7$ galaxy group in maturing phase. If confirmed, this would be the earliest structure in maturing phase to date, and an ideal laboratory to study the formation of the earliest quiescent galaxies as well as cold gas accretion in dense environments.

PUBLICATIONS BASED ON THE COSMIC DAWN SURVEY

C.2.6 *Euclid preparation: XVII. Cosmic Dawn Survey. Spitzer observations of the deep fields and calibration fields*

Euclid Collaboration: Andrea Moneti, H.J. McCracken, M. Shuntov et al., A&A 658, A126 (2022)

Abstract

We present a new infrared survey covering the three deep fields and four other calibration fields using 's Infrared Array Camera (IRAC). We combined these new observations with all relevant IRAC archival data of these fields in order to produce the deepest possible mosaics of these regions. In total, these observations represent nearly 11 % of the total mission time. The resulting mosaics cover a total of approximately 71.5 deg^2 in the 3.6 and 4.5 μm bands, and approximately 21.8 deg^2 in the 5.8 and 8 μm bands. They reach at least 24 AB magnitude (measured to 5σ, in a 25 aperture) in the 3.6band and up to $\sim$ 5 mag deeper in the deepest regions. The astrometry is tied to the astrometric reference system, and the typical astrometric uncertainty for sources with $16 < [3.6] < 19$ is $\lesssim$ 015. The photometric calibration is in excellent agreement with previous WISE measurements. We extracted source number counts from the 3.6 μm band mosaics, and they are in excellent agreement with previous measurements. Given that the has now been decommissioned, these mosaics are likely to be the definitive reduction of these IRAC data. This survey

therefore represents an essential first step in assembling multi-wavelength data on the deep fields, which are set to become some of the premier fields for extragalactic astronomy in the 2020s.

PUBLICATIONS BASED ON JWST EARLY DATA

C.2.7 *Revealing Galaxy Candidates out to $z \sim 16$ with JWST Observations of the Lensing Cluster SMACS0723*

Hakim Atek, Marko Shuntov, et al., eprint arXiv:2207.12338, submitted to MNRAS

Abstract

One of the main goals of the JWST is to study the first galaxies in the Universe. We present a systematic photometric analysis of very distant galaxies in the first JWST deep field towards the massive lensing cluster SMACS0723. As a result, we report the discovery of two galaxy candidates at $z \sim 16$, only 250 million years after the Big Bang. We also identify two candidates at $z \sim 12$ and 11 candidates at $z \sim 10-11$. Our search extended out to $z \lesssim 21$ by combining color information across seven NIRCam and NIRISS filters. By modelling the Spectral Energy Distributions (SEDs) with `EAZY` and `BEAGLE`, we test the robustness of the photometric redshift estimates. While their intrinsic (un-lensed) luminosity is typical of the characteristic luminosity L^* at $z > 10$, our high-redshift galaxies typically show small sizes and their morphologies are consistent with disks in some cases. The highest-redshift candidates have extremely blue UV-continuum slopes $-3 < \beta < -2.5$, young ages $\sim 10-100\,\mathrm{Myr}$, and stellar masses $\log(M_\star/\mathrm{M}_\odot) = 8.4-8.8$ inferred from their SED modeling, which indicate a rapid build-up of their stellar mass. Our search clearly demonstrates the capabilities of JWST to uncover robust photometric candidates up to very high redshifts, and peer into the formation epoch of the first galaxies.

C.2.8 *Constraining the physical properties of the first lensed $z \sim 10-16$ galaxy candidates with JWST*

Lukas J. Furtak, Marko Shuntov, et al., eprint arXiv:2208.05473, submitted to MNRAS

Abstract

The first deep-field observations of the JWST have immediately yielded a surprisingly large number of very high redshift candidates, pushing the frontier of observability well beyond $z \gtrsim 10$. We here present a detailed SED-fitting analysis of the 15 gravitationally lensed $z \sim 10-16$ galaxy candidates detected behind the galaxy cluster SMACS J0723.3-7327 in Atek et al. (2022) using the `BEAGLE` tool. Our analysis makes use of dynamical considerations to place limits on the ages of these galaxies and of all three published SL

models of the cluster to account for lensing systematics. We find these galaxies to have relatively low stellar masses $M_\star \sim 10^7 - 10^8\,\mathrm{M}_\odot$ and young ages $t_\mathrm{age} \sim 10 - 100\,\mathrm{Myr}$. Due to their very blue UV-slopes, down to $\beta \sim -3$, all of the galaxies in our sample have extremely low dust attenuations $A_V \lesssim 0.02$. Placing the measured parameters into relation, we find a very shallow $M_\star - M_\mathrm{UV}$-slope and high sSFRs above the main sequence of star-formation with no significant redshift-evolution in either relation. This is in agreement with the bright UV luminosities measured for these objects and indicates that we are naturally selecting galaxies that are currently undergoing a star-bursting episode at the time they are observed. Finally, we discuss the robustness of our high-redshift galaxy sample regarding low-redshift interlopers and conclude that low-redshift solutions can safely be ruled out for roughly half of the sample, including the highest-redshift galaxies at $z \sim 12 - 16$. These objects represent compelling targets for spectroscopic follow-up observations with JWST and ALMA.

LIST OF FIGURES

LIST OF TABLES

ACRONYMS

2PCF	Two-point Correlation Function
ACS	Advanced Camera For Surveys
AGN	Active Galactic Nucleii
ALMA	Atacama Large Millimeter Array
AM	Abundance Matching
CANDELS	Cosmic Assembly Near-infrared Deep Extragalactic Legacy Survey
CFHTLS	Canada-France-Hawaii Telescope Legacy Survey
CMB	Cosmic Microwave Background
DES	Dark Energy Survey
DM	Dark Matter
DSFG	Dusty Star-forming Galaxy
EDF-F	Euclid Deep Field Fornax
EDF-N	Euclid Deep Field North
EDF-S	Euclid Deep Field South
EDS	Euclid Deep Survey
GGL	Galaxy-galaxy Lensing
GOODS	Great Observatories Origins Deep Survey
GR	General Relativity
GSMF	Galaxy Stellar Mass Function
H20	Hawaii-Two-0
HDF	Hubble Deep Field
HMF	Halo Mass Function
HOD	Halo Occupation Distribution
HSC	Hyper Suprime-Cam
HST	Hubble Space Telescope
HUDF	Hubble Ultra Deep Field
IGM	Intergalactic Medium
JWST	James Webb Space Telescope
LoS	Line Of Sight
LSS	Large Scale Structure

ML	Machine Learning
NFW	Navarro-Frenk-White
NISP	Near Infrared Spectrometer And Photometer
PFS	Subaru Prime Focus Spectrograph
PSF	Point Spread Function
SDSS	Sloan Digital Sky Survey
SED	Spectral Energy Distribution
SFR	Star Formation Rate
SFRD	Star Formation Rate Density
SHMR	Stellar-to-halo Mass Relation
SKA	Square Kilometre Array
SMD	Stellar Mass Density
VIMOS	Visible Multi-Object Spectrograph
VIS	Visible Imager
WFC3	Wide-Field Camera 3
WL	Weak Lensing

BIBLIOGRAPHY

Abbott, T. M. C. et al. (Jan. 2022). "Dark Energy Survey Year 3 results: Cosmological constraints from galaxy clustering and weak lensing". In: 105.2, 023520, p. 023520. doi: 10.1103/PhysRevD.105.023520. arXiv: 2105.13549 [astro-ph.CO].

Aihara, Hiroaki et al. (Jan. 2018). "The Hyper Suprime-Cam SSP Survey: Overview and survey design". In: 70, S4. issn: 0004-6264. doi: 10.1093/pasj/psx066. url: https://ui.adsabs.harvard.edu/abs/2018PASJ...70S...4A (visited on 21/07/2021).

Aihara, Hiroaki et al. (Dec. 2019). "Second data release of the Hyper Suprime-Cam Subaru Strategic Program". In: 71.6. issn: 0004-6264. doi: 10.1093/pasj/psz103. url: https://doi.org/10.1093/pasj/psz103 (visited on 21/07/2021).

Anglés-Alcázar, Daniel et al. (Nov. 2017). "Black holes on FIRE: stellar feedback limits early feeding of galactic nuclei". In: 472.1, pp. L109–L114. doi: 10.1093/mnrasl/slx161. arXiv: 1707.03832 [astro-ph.GA].

Angulo, Raul E. and Oliver Hahn (Dec. 2022). "Large-scale dark matter simulations". In: *Living Reviews in Computational Astrophysics* 8.1, 1, p. 1. doi: 10.1007/s41115-021-00013-z. arXiv: 2112.05165 [astro-ph.CO].

Applebaum, Elaad et al. (Feb. 2020). "A stochastically sampled IMF alters the stellar content of simulated dwarf galaxies". In: 492.1, pp. 8–21. doi: 10.1093/mnras/stz3331. arXiv: 1811.00022 [astro-ph.GA].

Arnouts, S. et al. (Jan. 2002). "Measuring the redshift evolution of clustering: the Hubble Deep Field South". In: 329.2. arXiv: astro-ph/0109453, pp. 355–366. doi: 10.1046/j.1365-8711.2002.04988.x.

Arnouts, S. et al. (Dec. 2007). "The SWIRE-VVDS-CFHTLS surveys: stellar mass assembly over the last 10 Gyr. Evidence for a major build up of the red sequence between z = 2 and z = 1". In: 476.1, pp. 137–150. doi: 10.1051/0004-6361:20077632. arXiv: 0705.2438 [astro-ph].

Artale, M. Celeste et al. (Sept. 2017). "Small-scale galaxy clustering in the EAGLE simulation". In: 470.2. arXiv: 1611.05064, pp. 1771–1787. issn: 0035-8711, 1365-2966. doi: 10.1093/mnras/stx1263. url: http://arxiv.org/abs/1611.05064 (visited on 15/12/2021).

Ashby, M. L. N. et al. (May 2013). "SEDS: THE SPITZER EXTENDED DEEP SURVEY. SURVEY DESIGN, PHOTOMETRY, AND DEEP IRAC SOURCE COUNTS". en. In: 769.1, p. 80. issn: 0004-637X, 1538-4357. doi: 10.1088/0004-637X/769/1/80. url: http://stacks.iop.org/0004-637X/769/i=1/a=80?key=crossref.ba6fb22cb53e08a607e3bee2fbd64e9c (visited on 11/03/2020).

Atek, Hakim et al. (July 2022). "Revealing Galaxy Candidates out to $z \sim 16$ with JWST Observations of the Lensing Cluster SMACS0723". In: *arXiv e-prints*, arXiv:2207.12338, arXiv:2207.12338. arXiv: 2207.12338 [astro-ph.GA].

Aubert, D., C. Pichon and S. Colombi (Aug. 2004). "The origin and implications of dark matter anisotropic cosmic infall on ~L* haloes". In: 352, pp. 376–398. issn: 0035-8711. doi: 10.1111/j.1365-2966.2004.07883.x. url: https://ui.adsabs.harvard.edu/abs/2004MNRAS.352..376A (visited on 26/09/2021).

Balogh, Michael L., Julio F. Navarro and Simon L. Morris (Sept. 2000). "The Origin of Star Formation Gradients in Rich Galaxy Clusters". In: 540, pp. 113–121. issn: 0004-637X. doi: 10.1086/309323. url: https://ui.adsabs.harvard.edu/abs/2000ApJ...540..113B (visited on 21/10/2021).

Bardeen, J. M. et al. (May 1986). "The Statistics of Peaks of Gaussian Random Fields". In: 304, p. 15. doi: 10.1086/164143.

Bartelmann, M. and P. Schneider (Jan. 2001). "Weak gravitational lensing". In: 340, pp. 291–472. doi: 10.1016/S0370-1573(00)00082-X. eprint: astro-ph/9912508.

Bassini, L. et al. (Oct. 2020). "The DIANOGA simulations of galaxy clusters: characterising star formation in protoclusters". In: 642, A37, A37. doi: 10.1051/0004-6361/202038396. arXiv: 2006.13951 [astro-ph.GA].

Beckwith, Steven V. W. et al. (Nov. 2006). "The Hubble Ultra Deep Field". In: 132.5, pp. 1729–1755. doi: 10.1086/507302. arXiv: astro-ph/0607632 [astro-ph].

Behroozi, Peter and Joseph Silk (July 2018). "The Most Massive Galaxies and Black Holes Allowed by ΛCDM". In: 477.4. arXiv: 1609.04402, pp. 5382–5387. issn: 0035-8711, 1365-2966. doi: 10.1093/mnras/sty945. url: http://arxiv.org/abs/1609.04402 (visited on 03/12/2021).

Behroozi, Peter et al. (Sept. 2019). "UniverseMachine: The Correlation between Galaxy Growth and Dark Matter Halo Assembly from z=0-10". In: 488.3. arXiv: 1806.07893, pp. 3143–3194. issn: 0035-8711, 1365-2966. doi: 10.1093/mnras/stz1182. url: http://arxiv.org/abs/1806.07893 (visited on 25/05/2021).

Behroozi, Peter S., Charlie Conroy and Risa H. Wechsler (July 2010). "A Comprehensive Analysis of Uncertainties Affecting the Stellar Mass - Halo Mass Relation for 0<z<4". In: 717.1. arXiv: 1001.0015, pp. 379–403. issn: 0004-637X, 1538-4357. doi: 10.1088/0004-637X/717/1/379. url: http://arxiv.org/abs/1001.0015 (visited on 03/07/2021).

Behroozi, Peter S. and Joseph Silk (Jan. 2015). "A Simple Technique for Predicting High-redshift Galaxy Evolution". In: 799.1, 32, p. 32. doi: 10.1088/0004-637X/799/1/32. arXiv: 1404.5299 [astro-ph.GA].

Behroozi, Peter S., Risa H. Wechsler and Charlie Conroy (June 2013). "The Average Star Formation Histories of Galaxies in Dark Matter Halos from z = 0-8". In: 770, p. 57. ISSN: 0004-637X. DOI: 10.1088/0004-637X/770/1/57. URL: https://ui.adsabs.harvard.edu/abs/2013ApJ...770...57B (visited on 22/07/2021).

Behroozi, Peter S., Risa H. Wechsler and Hao-Yi Wu (Jan. 2013). "The ROCKSTAR Phase-space Temporal Halo Finder and the Velocity Offsets of Cluster Cores". In: 762.2, 109, p. 109. DOI: 10.1088/0004-637X/762/2/109. arXiv: 1110.4372 [astro-ph.CO].

Beltz-Mohrmann, Gillian D. and Andreas A. Berlind (Nov. 2021). "The Impact of Baryonic Physics on the Abundance, Clustering, and Concentration of Halos". In: 921, p. 112. ISSN: 0004-637X. DOI: 10.3847/1538-4357/ac1e27. URL: https://ui.adsabs.harvard.edu/abs/2021ApJ...921..112B (visited on 14/12/2021).

Berlind, Andreas A. and David H. Weinberg (Aug. 2002). "The Halo Occupation Distribution: Towards an Empirical Determination of the Relation Between Galaxies and Mass". In: 575.2. arXiv: astro-ph/0109001, pp. 587–616. ISSN: 0004-637X, 1538-4357. DOI: 10.1086/341469. URL: http://arxiv.org/abs/astro-ph/0109001 (visited on 09/07/2021).

Bernardeau, F. et al. (Sept. 2002). "Large-scale structure of the Universe and cosmological perturbation theory". In: 367.1-3, pp. 1–248. DOI: 10.1016/S0370-1573(02)00135-7. arXiv: astro-ph/0112551 [astro-ph].

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

Bigourdan, G. (Jan. 1888). "Revue des publications astronomiques. Bulletin du comité international permanent pour l'exécution photographique de la Carte du Ciel. Premier fascicule". In: *Bulletin Astronomique, Serie I* 5, pp. 303–309.

Birnboim, Yuval and Avishai Dekel (Oct. 2003). "Virial shocks in galactic haloes?" In: 345, pp. 349–364. ISSN: 0035-8711. DOI: 10.1046/j.1365-8711.2003.06955.x. URL: https://ui.adsabs.harvard.edu/abs/2003MNRAS.345..349B (visited on 29/10/2021).

Bocquet, Sebastian et al. (Mar. 2016). "Halo mass function: baryon impact, fitting formulae, and implications for cluster cosmology". In: 456, pp. 2361–2373. ISSN: 0035-8711. DOI: 10.1093/mnras/stv2657. URL: https://ui.adsabs.harvard.edu/abs/2016MNRAS.456.2361B (visited on 14/12/2021).

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

Booth, C. M. et al. (Nov. 2013). "Simulations of Disk Galaxies with Cosmic Ray Driven Galactic Winds". In: 777.1, L16, p. L16. DOI: 10.1088/2041-8205/777/1/L16. arXiv: 1308.4974 [astro-ph.GA].

Bose, Sownak et al. (Dec. 2019). "Revealing the galaxy-halo connection in IllustrisTNG". In: 490.4, pp. 5693–5711. DOI: 10.1093/mnras/stz2546. arXiv: 1905.08799 [astro-ph.CO].

Bouché, N. et al. (Aug. 2010). "The Impact of Cold Gas Accretion Above a Mass Floor on Galaxy Scaling Relations". In: 718.2, pp. 1001–1018. DOI: 10.1088/0004-637X/718/2/1001. arXiv: 0912.1858 [astro-ph.CO].

Boulade, Olivier et al. (Mar. 2003). "MegaCam: the new Canada-France-Hawaii Telescope wide-field imaging camera". In: ed. by Masanori Iye and Alan F. M. Moorwood. Vol. 4841. Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, pp. 72–81. DOI: 10.1117/12.459890.

Bouwens, Rychard et al. (Oct. 2020). "The ALMA Spectroscopic Survey Large Program: The Infrared Excess of z = 1.5-10 UV-selected Galaxies and the Implied High-redshift Star Formation History". In: 902.2, 112, p. 112. DOI: 10.3847/1538-4357/abb830. arXiv: 2009.10727 [astro-ph.GA].

Bower, R. G. et al. (Aug. 2006). "Breaking the hierarchy of galaxy formation". In: 370.2, pp. 645–655. DOI: 10.1111/j.1365-2966.2006.10519.x. arXiv: astro-ph/0511338 [astro-ph].

Bowler, R. A. A. et al. (Apr. 2020). "A lack of evolution in the very bright end of the galaxy luminosity function from z 8 to 10". In: 493.2, pp. 2059–2084. DOI: 10.1093/mnras/staa313. arXiv: 1911.12832 [astro-ph.GA].

Brammer, Gabriel B., Pieter G. van Dokkum and Paolo Coppi (Oct. 2008). "EAZY: A Fast, Public Photometric Redshift Code". In: 686.2. arXiv: 0807.1533, pp. 1503–1513. ISSN: 0004-637X, 1538-4357. DOI: 10.1086/591786. URL: http://arxiv.org/abs/0807.1533 (visited on 20/06/2021).

Brammer, Gabriel B., Pieter G. van Dokkum and Paolo Coppi (Oct. 2008). "EAZY: A Fast, Public Photometric Redshift Code". In: 686.2, pp. 1503–1513. DOI: 10.1086/591786. arXiv: 0807.1533 [astro-ph].

Brodwin, M. et al. (Jan. 2006). "The Canada-France Deep Fields Survey. III. Photometric Redshift Distribution to IAB = 24". en. In: 162.1, p. 20. ISSN: 0067-0049. DOI: 10.1086/497990. URL: https://iopscience.iop.org/article/10.1086/497990/meta (visited on 05/07/2021).

Bruzual, G. and S. Charlot (Oct. 2003). "Stellar population synthesis at the resolution of 2003". In: 344, pp. 1000–1028. ISSN: 0035-8711. DOI: 10.1046/j.1365-8711.2003.06897.x. URL: https://ui.adsabs.harvard.edu/abs/2003MNRAS.344.1000B (visited on 23/06/2021).

Bryan, Greg L. and Michael L. Norman (Mar. 1998). "Statistical Properties of X-Ray Clusters: Analytic and Numerical Comparisons". In: 495.1, pp. 80–99. DOI: 10.1086/305262. arXiv: astro-ph/9710107 [astro-ph].

Calzetti, Daniela et al. (Apr. 2000). "The Dust Content and Opacity of Actively Star-forming Galaxies". In: 533, pp. 682–695. ISSN: 0004-637X. DOI: 10.1086/308692. URL: https://ui.adsabs.harvard.edu/abs/2000ApJ...533..682C (visited on 21/07/2021).

Capak, P. et al. (Sept. 2007). “The First Release COSMOS Optical and Near-IR Data and Catalog”. In: *The Astrophysical Journal Supplement Series* 172, pp. 99–116. ISSN: 0067-0049. DOI: `10.1086/519081`. URL: `http://adsabs.harvard.edu/abs/2007ApJS..172...99C` (visited on 08/05/2020).

Carretero, J. et al. (July 2017). “CosmoHub and SciPIC: Massive cosmological data analysis, distribution and generation using a Big Data platform”. In: *Proceedings of the European Physical Society Conference on High Energy Physics. 5-12 July*, 488, p. 488.

Castro, Tiago et al. (Nov. 2020). “On the impact of baryons on the halo mass function, bias, and cluster cosmology”. In: 500.2. arXiv: 2009.01775, pp. 2316–2335. ISSN: 0035-8711, 1365-2966. DOI: `10.1093/mnras/staa3473`. URL: `http://arxiv.org/abs/2009.01775` (visited on 14/12/2021).

Cautun, Marius et al. (July 2014). “Evolution of the cosmic web”. In: 441.4. arXiv: 1401.7866, pp. 2923–2973. ISSN: 1365-2966, 0035-8711. DOI: `10.1093/mnras/stu768`. URL: `http://arxiv.org/abs/1401.7866` (visited on 16/09/2021).

Chabrier, Gilles (July 2003). “Galactic Stellar and Substellar Initial Mass Function”. In: 115.809, pp. 763–795. DOI: `10.1086/376392`.

Chiang, Yi-Kuan et al. (Aug. 2017). “Galaxy Protoclusters as Drivers of Cosmic Star Formation History in the First 2 Gyr”. In: 844.2, L23, p. L23. DOI: `10.3847/2041-8213/aa7e7b`. arXiv: `1705.01634 [astro-ph.GA]`.

Chisari, N. E. et al. (Nov. 2018). “The impact of baryons on the matter power spectrum from the Horizon-AGN cosmological hydrodynamical simulation”. In: 480.3, pp. 3962–3977. DOI: `10.1093/mnras/sty2093`. arXiv: `1801.08559 [astro-ph.CO]`.

Chisari, Nora Elisa et al. (May 2019). “Core Cosmology Library: Precision Cosmological Predictions for LSST”. In: 242.1. arXiv: 1812.05995, p. 2. ISSN: 1538-4365. DOI: `10.3847/1538-4365/ab1658`. URL: `http://arxiv.org/abs/1812.05995` (visited on 09/07/2021).

Chiu, I. Non et al. (June 2020). “The richness-to-mass relation of CAMIRA galaxy clusters from weak-lensing magnification in the Subaru Hyper Suprime-Cam survey”. In: 495.1, pp. 428–450. DOI: `10.1093/mnras/staa1158`. arXiv: `1909.02042 [astro-ph.CO]`.

Churazov, E. et al. (June 2001). “Evolution of Buoyant Bubbles in M87”. In: 554.1, pp. 261–273. DOI: `10.1086/321357`. arXiv: `astro-ph/0008215 [astro-ph]`.

Collaboration, Planck et al. (Nov. 2014). “Planck 2013 results. XVI. Cosmological parameters”. In: 571. arXiv: 1303.5076, A16. ISSN: 0004-6361, 1432-0746. DOI: `10.1051/0004-6361/201321591`. URL: `http://arxiv.org/abs/1303.5076` (visited on 26/09/2021).

Conroy, Charlie and James E. Gunn (Apr. 2010). “The Propagation of Uncertainties in Stellar Population Synthesis Modeling. III. Model Calibration, Comparison, and Evaluation”. In: 712.2, pp. 833–857. DOI: `10.1088/0004-637X/712/2/833`. arXiv: `0911.3151 [astro-ph.CO]`.

Conroy, Charlie, James E. Gunn and Martin White (July 2009). “The Propagation of Uncertainties in Stellar Population Synthesis Modeling. I. The Relevance of Uncertain Aspects of Stellar Evolution and the Initial Mass Function to the Derived Physical Properties of Galaxies”. In: 699.1, pp. 486–506. DOI: `10.1088/0004-637X/699/1/486`. arXiv: `0809.4261 [astro-ph]`.

Conroy, Charlie and Risa H. Wechsler (May 2009). “Connecting Galaxies, Halos, and Star Formation Rates Across Cosmic Time”. In: 696.1. arXiv: 0805.3346, pp. 620–635. ISSN: 0004-637X, 1538-4357. DOI: `10.1088/0004-637X/696/1/620`. URL: `http://arxiv.org/abs/0805.3346` (visited on 30/09/2021).

Cooray, Asantha and Ravi Sheth (Dec. 2002). “Halo Models of Large Scale Structure”. In: *Physics Reports* 372.1. arXiv: astro-ph/0206508, pp. 1–129. ISSN: 03701573. DOI: `10.1016/S0370-1573(02)00276-4`. URL: `http://arxiv.org/abs/astro-ph/0206508` (visited on 30/06/2021).

Costa, Tiago, Joakim Rosdahl and Taysun Kimm (Nov. 2019). “The hidden satellites of massive galaxies and quasars at high redshift”. In: 489.4, pp. 5181–5186. DOI: `10.1093/mnras/stz2471`. arXiv: `1909.01360 [astro-ph.GA]`.

Coupon, J. et al. (June 2012). “Galaxy clustering in the CFHTLS-Wide: the changing relationship between galaxies and haloes since z ~ 1.2”. en. In: 542, A5. ISSN: 0004-6361. DOI: `10.1051/0004-6361/201117625`. URL: `https://ui.adsabs.harvard.edu/abs/2012A%26A...542A...5C/abstract` (visited on 09/07/2021).

Coupon, J. et al. (May 2015). “The galaxy-halo connection from a joint lensing, clustering and abundance analysis in the CFHTLenS/VIPERS field”. In: 449.2. arXiv: 1502.02867, pp. 1352–1379. ISSN: 1365-2966, 0035-8711. DOI: `10.1093/mnras/stv276`. URL: `http://arxiv.org/abs/1502.02867` (visited on 09/07/2021).

Coupon, Jean et al. (Jan. 2018). “The bright-star masks for the HSC-SSP survey”. In: 70, S7, S7. DOI: `10.1093/pasj/psx047`. arXiv: `1705.00622 [astro-ph.IM]`.

Cowie, Lennox L. et al. (Sept. 1996). “New Insight on Galaxy Formation and Evolution From Keck Spectroscopy of the Hawaii Deep Fields”. In: 112, p. 839. ISSN: 0004-6256. DOI: `10.1086/118058`. URL: `https://ui.adsabs.harvard.edu/abs/1996AJ....112..839C` (visited on 06/10/2021).

Cowley, William I. et al. (Jan. 2018). “The Galaxy–Halo Connection for $1.5z5$as Revealed by theSpitzerMatching Survey of the UltraVISTA Ultra-deep Stripes”. en. In: 853.1, p. 69. ISSN: 0004-637X. DOI: `10.3847/1538-4357/aaa41d`. URL: `https://doi.org/10.3847/1538-4357/aaa41d` (visited on 26/09/2021).

Cowley, William I. et al. (Mar. 2019). “The Stellar-to-Halo Mass Ratios of Passive and Star-Forming Galaxies at z~2-3 from the SMUVS survey”. In: 874.2. arXiv: 1901.10455, p. 114. ISSN: 1538-4357. DOI: `10.3847/1538-4357/ab089b`. URL: `http://arxiv.org/abs/1901.10455` (visited on 04/07/2021).

Crain, Robert A. et al. (June 2015). “The EAGLE simulations of galaxy formation: calibration of subgrid physics and model variations”. In: 450.2, pp. 1937–1961. DOI: 10.1093/mnras/stv725.

Croton, Darren (2013). “Damn You, Little h! (or, Real-World Applications Of The Hubble Constant Using Observed And Simulated Data)”. In: 30. arXiv: 1308.4150, e052. ISSN: 1323-3580, 1448-6083. DOI: 10.1017/pasa.2013.31. URL: http://arxiv.org/abs/1308.4150 (visited on 20/10/2021).

Croton, Darren J. et al. (Jan. 2006). “The many lives of active galactic nuclei: cooling flows, black holes and the luminosities and colours of galaxies”. en. In: 365.1. arXiv: astro-ph/0508046, pp. 11–28. ISSN: 00358711, 13652966. DOI: 10.1111/j.1365-2966.2005.09675.x. URL: http://arxiv.org/abs/astro-ph/0508046 (visited on 30/09/2021).

Cueli, M. M. et al. (June 2022). “Tomography-based observational measurements of the halo mass function via the submillimeter magnification bias”. In: 662, A44, A44. DOI: 10.1051/0004-6361/202142949. arXiv: 2203.14638 [astro-ph.CO].

Daddi, E. et al. (Nov. 2007). “Multiwavelength Study of Massive Galaxies at z∼2. I. Star Formation and Galaxy Growth”. In: 670.1, pp. 156–172. DOI: 10.1086/521818. arXiv: 0705.2831 [astro-ph].

Daddi, E. et al. (Apr. 2010). “Very High Gas Fractions and Extended Gas Reservoirs in z = 1.5 Disk Galaxies”. In: 713.1, pp. 686–707. DOI: 10.1088/0004-637X/713/1/686. arXiv: 0911.2776 [astro-ph.CO].

Daddi, E. et al. (May 2021). “Three Lyman-alpha emitting filaments converging to a massive galaxy group at z=2.91: discussing the case for cold gas infall”. In: 649. arXiv: 2006.11089, A78. ISSN: 0004-6361, 1432-0746. DOI: 10.1051/0004-6361/202038700. URL: http://arxiv.org/abs/2006.11089 (visited on 04/10/2021).

Daddi, E. et al. (Feb. 2022a). “Evidence for Cold-stream to Hot-accretion Transition as Traced by Lyα Emission from Groups and Clusters at $2 < z < 3.3$”. In: 926.2, L21, p. L21. DOI: 10.3847/2041-8213/ac531f. arXiv: 2202.03715 [astro-ph.CO].

— (Feb. 2022b). “Evidence for Cold-stream to Hot-accretion Transition as Traced by Lyα Emission from Groups and Clusters at $2 < z < 3.3$”. In: 926.2, L21, p. L21. DOI: 10.3847/2041-8213/ac531f. arXiv: 2202.03715 [astro-ph.CO].

Dalla Vecchia, Claudio and Joop Schaye (Oct. 2012). “Simulating galactic outflows with thermal supernova feedback”. In: 426.1, pp. 140–158. DOI: 10.1111/j.1365-2966.2012.21704.x. arXiv: 1203.5667 [astro-ph.GA].

Dannerbauer, H. et al. (Oct. 2014). “An excess of dusty starbursts related to the Spiderweb galaxy”. In: 570, A55, A55. DOI: 10.1051/0004-6361/201423771. arXiv: 1410.3730 [astro-ph.GA].

Dark Energy Survey Collaboration et al. (Aug. 2016). “The Dark Energy Survey: more than dark energy - an overview”. In: 460.2, pp. 1270–1299. DOI: 10.1093/mnras/stw641. arXiv: 1601.00329 [astro-ph.CO].

Dashyan, Gohar and Yohan Dubois (June 2020). “Cosmic ray feedback from supernovae in dwarf galaxies”. In: 638, A123, A123. DOI: 10.1051/0004-6361/201936339. arXiv: 2003.09900 [astro-ph.GA].

Dashyan, Gohar et al. (Aug. 2019). “AGN-driven quenching of satellite galaxies”. In: 487.4, pp. 5889–5901. DOI: 10.1093/mnras/stz1697. arXiv: 1906.07431 [astro-ph.GA].

Davidzon, I. et al. (Sept. 2017). “The COSMOS2015 galaxy stellar mass function: 13 billion years of stellar mass assembly in 10 snapshots”. In: 605. arXiv: 1701.02734, A70. ISSN: 0004-6361, 1432-0746. DOI: 10.1051/0004-6361/201730419. URL: http://arxiv.org/abs/1701.02734 (visited on 27/06/2021).

Davis, M. et al. (May 1985). “The evolution of large-scale structure in a universe dominated by cold dark matter”. In: 292, pp. 371–394. DOI: 10.1086/163168.

Davis, M. et al. (May 1985). “The evolution of large-scale structure in a universe dominated by cold dark matter”. In: 292, pp. 371–394. ISSN: 0004-637X. DOI: 10.1086/163168. URL: https://ui.adsabs.harvard.edu/abs/1985ApJ...292..371D (visited on 17/12/2021).

De Lucia, Gabriella et al. (Feb. 2006). “The formation history of elliptical galaxies”. In: 366.2, pp. 499–509. ISSN: 0035-8711. DOI: 10.1111/j.1365-2966.2005.09879.x. URL: https://doi.org/10.1111/j.1365-2966.2005.09879.x (visited on 27/09/2021).

Decarli, Roberto et al. (Sept. 2019). “The ALMA Spectroscopic Survey in the HUDF: CO Luminosity Functions and the Molecular Gas Content of Galaxies through Cosmic History”. In: 882.2, 138, p. 138. DOI: 10.3847/1538-4357/ab30fe. arXiv: 1903.09164 [astro-ph.GA].

Dekel, A. and Y. Birnboim (May 2006). “Galaxy bimodality due to cold flows and shock heating”. en. In: 368.1, pp. 2–20. ISSN: 0035-8711, 1365-2966. DOI: 10.1111/j.1365-2966.2006.10145.x. URL: https://academic.oup.com/mnras/article-lookup/doi/10.1111/j.1365-2966.2006.10145.x (visited on 29/10/2021).

Dekel, A. and J. Silk (Apr. 1986). “The Origin of Dwarf Galaxies, Cold Dark Matter, and Biased Galaxy Formation”. In: 303, p. 39. ISSN: 0004-637X. DOI: 10.1086/164050. URL: https://ui.adsabs.harvard.edu/abs/1986ApJ...303...39D (visited on 25/11/2021).

Dekel, A. et al. (Jan. 2009). “Cold streams in early massive hot haloes as the main mode of galaxy formation”. en. In: *Nature* 457.7228. arXiv: 0808.0553, pp. 451–454. ISSN: 0028-0836, 1476-4687. DOI: 10.1038/nature07648. URL: http://arxiv.org/abs/0808.0553 (visited on 29/10/2021).

Dekel, A. et al. (Jan. 2009). “Cold streams in early massive hot haloes as the main mode of galaxy formation”. In: 457.7228, pp. 451–454. DOI: 10.1038/nature07648. arXiv: 0808.0553 [astro-ph].

Delvecchio, I. et al. (June 2017). “The VLA-COSMOS 3 GHz Large Project: AGN and host-galaxy properties out to z $\lesssim$ 6”. In: 602, A3, A3. DOI: `10.1051/0004-6361/201629367`. arXiv: `1703.09720 [astro-ph.GA]`.

Desjacques, Vincent, Donghui Jeong and Fabian Schmidt (Feb. 2018). “Large-scale galaxy bias”. In: 733, pp. 1–193. DOI: `10.1016/j.physrep.2017.12.002`. arXiv: `1611.09787 [astro-ph.CO]`.

Desmond, Harry et al. (Oct. 2017). “On the galaxy–halo connection in the EAGLE simulation”. In: *: Letters* 471.1, pp. L11–L15. ISSN: 1745-3925. DOI: `10.1093/mnrasl/slx093`. URL: `https://doi.org/10.1093/mnrasl/slx093` (visited on 14/12/2021).

Despali, Giulia et al. (Mar. 2016). “The universality of the virial halo mass function and models for non-universality of other halo definitions”. In: 456.3. arXiv: 1507.05627, pp. 2486–2504. ISSN: 0035-8711, 1365-2966. DOI: `10.1093/mnras/stv2842`. URL: `http://arxiv.org/abs/1507.05627` (visited on 26/09/2021).

Desprez, G. et al. (Dec. 2020). “Euclid preparation. X. The Euclid photometric-redshift challenge”. In: 644, A31, A31. DOI: `10.1051/0004-6361/202039403`. arXiv: `2009.12112 [astro-ph.GA]`.

Diemer, Benedikt (Dec. 2018). “COLOSSUS: A Python Toolkit for Cosmology, Large-scale Structure, and Dark Matter Halos”. In: 239.2, 35, p. 35. DOI: `10.3847/1538-4365/aaee8c`. arXiv: `1712.04512 [astro-ph.CO]`.

Donnari, Martina et al. (Aug. 2021). “Quenched fractions in the IllustrisTNG simulations: comparison with observations and other theoretical models”. In: 506.4. arXiv: 2008.00004, pp. 4760–4780. ISSN: 0035-8711, 1365-2966. DOI: `10.1093/mnras/stab1950`. URL: `http://arxiv.org/abs/2008.00004` (visited on 15/12/2021).

Dubois, Y. and R. Teyssier (Jan. 2008). “On the onset of galactic winds in quiescent star forming galaxies”. In: 477.1, pp. 79–94. DOI: `10.1051/0004-6361:20078326`. arXiv: `0707.3376 [astro-ph]`.

Dubois, Yohan et al. (Dec. 2010). “Jet-regulated cooling catastrophe”. In: 409, pp. 985–1001. ISSN: 0035-8711. DOI: `10.1111/j.1365-2966.2010.17338.x`. URL: `https://ui.adsabs.harvard.edu/abs/2010MNRAS.409..985D` (visited on 25/11/2021).

Dubois, Yohan et al. (Mar. 2012). “Self-regulated growth of supermassive black holes by a dual jet-heating active galactic nucleus feedback mechanism: methods, tests and implications for cosmological simulations”. In: 420.3, pp. 2662–2683. DOI: `10.1111/j.1365-2966.2011.20236.x`. arXiv: `1108.0110 [astro-ph.CO]`.

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

Dubois, Yohan et al. (Oct. 2014). “Dancing in the dark: galactic properties trace spin swings along the cosmic web”. In: 444.2. arXiv: 1402.1165, pp. 1453–1468. ISSN: 0035-8711, 1365-2966. DOI: `10.1093/mnras/stu1227`. URL: `http://arxiv.org/abs/1402.1165` (visited on 17/09/2021).

Dubois, Yohan et al. (Sept. 2015). “Black hole evolution - I. Supernova-regulated black hole growth”. In: 452.2, pp. 1502–1518. DOI: `10.1093/mnras/stv1416`. arXiv: `1504.00018 [astro-ph.GA]`.

Dubois, Yohan et al. (July 2021). “Introducing the NEWHORIZON simulation: Galaxy properties with resolved internal dynamics across cosmic time”. In: 651, A109, A109. DOI: `10.1051/0004-6361/202039429`. arXiv: `2009.10578 [astro-ph.GA]`.

Duffy, Alan R. et al. (Oct. 2008). “Dark matter halo concentrations in the Wilkinson Microwave Anisotropy Probe year 5 cosmology”. In: *: Letters* 390.1. arXiv: 0804.2486, pp. L64–L68. ISSN: 17453925, 17453933. DOI: `10.1111/j.1745-3933.2008.00537.x`. URL: `http://arxiv.org/abs/0804.2486` (visited on 10/07/2021).

Duncan, Christopher A. J. et al. (Sept. 2022). “On cosmological bias due to the magnification of shear and position samples in modern weak lensing analyses”. In: 515.1, pp. 1130–1145. DOI: `10.1093/mnras/stac1809`. arXiv: `2111.09867 [astro-ph.CO]`.

Dye, S. et al. (Apr. 2002). “Lens magnification by CL0024+1654 in the U and R band”. In: 386, pp. 12–30. DOI: `10.1051/0004-6361:20020226`. arXiv: `astro-ph/0108399 [astro-ph]`.

Einasto, J. et al. (Nov. 1974). “Missing mass around galaxies - Morphological evidence”. In: 252, pp. 111–113. DOI: `10.1038/252111a0`.

Elbaz, D. et al. (June 2007). “The reversal of the star formation-density relation in the distant universe”. In: 468.1, pp. 33–48. DOI: `10.1051/0004-6361:20077525`. arXiv: `astro-ph/0703653 [astro-ph]`.

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

Emonts, B. H. C. et al. (Dec. 2016). “Molecular gas in the halo fuels the growth of a massive cluster galaxy at high redshift”. In: *Science* 354.6316, pp. 1128–1130. DOI: `10.1126/science.aag0512`. arXiv: `1612.00387 [astro-ph.GA]`.

Engler, Christoph et al. (Dec. 2020). “The distinct stellar-to-halo mass relations of satellite and central galaxies: insights from the IllustrisTNG simulations”. In: 500.3. arXiv: 2002.11119, pp. 3957–3975. ISSN: 0035-8711, 1365-2966. DOI: `10.1093/mnras/staa3505`. URL: `http://arxiv.org/abs/2002.11119` (visited on 10/06/2021).

Euclid Collaboration et al. (June 2022a). “Euclid preparation. I. The Euclid Wide Survey”. In: 662, A112, A112. DOI: `10.1051/0004-6361/202141938`. arXiv: `2108.01201 [astro-ph.CO]`.

Euclid Collaboration et al. (June 2022b). “Euclid preparation. XIX. Impact of magnification on photometric galaxy clustering”. In: 662, A93, A93. DOI: `10.1051/0004-6361/202142419`. arXiv: `2110.05435 [astro-ph.CO]`.

Faber, S. M. et al. (Aug. 2007). “Galaxy Luminosity Functions to z~1 from DEEP2 and COMBO-17: Implications for Red Galaxy Formation”. In: 665.1, pp. 265–294. DOI: 10.1086/519294. arXiv: astro-ph/0506044 [astro-ph].

Fazio, G. G. et al. (Sept. 2004). “Number Counts at 3 m < < 10 m from the Spitzer Space Telescope”. en. In: 154.1. Publisher: IOP Publishing, p. 39. ISSN: 0067-0049. DOI: 10.1086/422585. URL: https://iopscience.iop.org/article/10.1086/422585/meta (visited on 20/03/2020).

Fernandez, L. et al. (Nov. 2021). “Galaxy cluster mass density profile derived using the submillimetre galaxies magnification bias”. In: *arXiv:2111.05422 [astro-ph]*. arXiv: 2111.05422. URL: http://arxiv.org/abs/2111.05422 (visited on 16/11/2021).

Ford, J. et al. (Apr. 2014). “Cluster magnification and the mass-richness relation in CFHTLenS”. In: 439, pp. 3755–3764. DOI: 10.1093/mnras/stu225. arXiv: 1310.2295 [astro-ph.CO].

Foreman-Mackey, Daniel et al. (Mar. 2013). “emcee: The MCMC Hammer”. In: 125.925. arXiv: 1202.3665, pp. 306–312. ISSN: 00046280, 15383873. DOI: 10.1086/670067. URL: http://arxiv.org/abs/1202.3665 (visited on 03/07/2021).

Förster Schreiber, Natascha M. and Stijn Wuyts (Aug. 2020). “Star-Forming Galaxies at Cosmic Noon”. In: 58, pp. 661–725. DOI: 10.1146/annurev-astro-032620-021910. arXiv: 2010.10171 [astro-ph.GA].

Furtak, Lukas J. et al. (Aug. 2022). “Constraining the physical properties of the first lensed $z \sim 10 - 16$ galaxy candidates with JWST”. In: *arXiv e-prints*, arXiv:2208.05473, arXiv:2208.05473. arXiv: 2208.05473 [astro-ph.GA].

Gabor, J. M. and R. Davé (Feb. 2015). “Hot gas in massive haloes drives both mass quenching and environment quenching”. en. In: 447.1, pp. 374–391. ISSN: 1365-2966, 0035-8711. DOI: 10.1093/mnras/stu2399. URL: http://academic.oup.com/mnras/article/447/1/374/988785/Hot-gas-in-massive-haloes-drives-both-mass (visited on 27/09/2021).

Garnavich, Peter M. et al. (Dec. 1998). “Supernova Limits on the Cosmic Equation of State”. In: 509.1, pp. 74–79. DOI: 10.1086/306495. arXiv: astro-ph/9806396 [astro-ph].

Genel, Shy et al. (Aug. 2009). “The Halo Merger Rate in the Millennium Simulation and Implications for Observed Galaxy Merger Fractions”. In: 701.2, pp. 2002–2018. DOI: 10.1088/0004-637X/701/2/2002. arXiv: 0812.3154 [astro-ph].

Giavalisco, M. et al. (Jan. 2004). “The Great Observatories Origins Deep Survey: Initial Results from Optical and Near-Infrared Imaging”. In: 600.2, pp. L93–L98. DOI: 10.1086/379232. arXiv: astro-ph/0309105 [astro-ph].

Girelli, Giacomo et al. (Feb. 2020). “The stellar-to-halo mass relation over the past 12 Gyr”. In: 634. arXiv: 2001.02230, A135. ISSN: 0004-6361, 1432-0746. DOI: 10.1051/0004-6361/201936329. URL: http://arxiv.org/abs/2001.02230 (visited on 28/07/2021).

González-Nuevo, J. et al. (Oct. 2017). “H-ATLAS/GAMA: magnification bias tomography. Astrophysical constraints above ~1 arcmin”. In: 2017.10, 024, p. 024. DOI: 10.1088/1475-7516/2017/10/024. arXiv: 1707.03709 [astro-ph.GA].

González-Nuevo, J. et al. (Feb. 2021). “Cosmological constraints on the magnification bias on sub-millimetre galaxies after large-scale bias corrections”. In: 646, A152, A152. DOI: 10.1051/0004-6361/202039043.

Gouin, C. et al. (June 2019). “Weak lensing in the Horizon-AGN simulation lightcone. Small-scale baryonic effects”. In: 626, A72. DOI: 10.1051/0004-6361/201834199.

Grogin, Norman A. et al. (Dec. 2011). “CANDELS: The Cosmic Assembly Near-infrared Deep Extragalactic Legacy Survey”. In: 197.2, 35, p. 35. DOI: 10.1088/0067-0049/197/2/35. arXiv: 1105.3753 [astro-ph.CO].

Groth, E. J. and P. J. E. Peebles (Oct. 1977). “Statistical analysis of catalogs of extragalactic objects. VII. Two-and three-point correlation functions for the high-resolution Shane-Wirtanen catalog of galaxies.” In: 217, pp. 385–405. ISSN: 0004-637X. DOI: 10.1086/155588. URL: https://ui.adsabs.harvard.edu/abs/1977ApJ...217..385G (visited on 29/06/2021).

Gunn, J. E. et al. (Dec. 1998). “The Sloan Digital Sky Survey Photometric Camera”. In: 116.6, pp. 3040–3081. DOI: 10.1086/300645. arXiv: astro-ph/9809085 [astro-ph].

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

Gunn, James E. and J. Richard Gott III (Aug. 1972). “On the Infall of Matter Into Clusters of Galaxies and Some Effects on Their Evolution”. In: 176, p. 1. ISSN: 0004-637X. DOI: 10.1086/151605. URL: https://ui.adsabs.harvard.edu/abs/1972ApJ...176....1G (visited on 21/10/2021).

Guy, Leanne P. et al. (Jan. 2022). “Rubin-Euclid Derived Data Products: Initial Recommendations”. In: *Zenodo id. 5836022*. Vol. 58, 5836022, p. 5836022. DOI: 10.5281/zenodo.5836022. arXiv: 2201.03862 [astro-ph.IM].

Habouzit, Mélanie, Marta Volonteri and Yohan Dubois (July 2017). “Blossoms from black hole seeds: properties and early growth regulated by supernova feedback”. In: 468.4, pp. 3935–3948. DOI: 10.1093/mnras/stx666. arXiv: 1605.09394 [astro-ph.GA].

Hadzhiyska, Boryana et al. (Apr. 2020). “Limitations to the "basic” HOD model and beyond”. In: 493.4. arXiv: 1911.02610, pp. 5506–5519. ISSN: 0035-8711, 1365-2966. DOI: 10.1093/mnras/staa623. URL: http://arxiv.org/abs/1911.02610 (visited on 23/09/2021).

Hamilton, A. J. S. and Max Tegmark (Mar. 2002). "The real-space power spectrum of the PSCz survey from 0.01 to 300h Mpc^{-1}". In: 330.3, pp. 506–530. DOI: 10.1046/j.1365-8711.2002.05033.x. arXiv: astro-ph/0008392 [astro-ph].

Harikane, Yuichi et al. (Apr. 2016). "Evolution of Stellar-to-Halo Mass Ratio at z=0-7 Identified by Clustering Analysis with the Hubble Legacy Imaging and Early Subaru/Hyper Suprime-Cam Survey Data". In: 821.2. arXiv: 1511.07873, p. 123. ISSN: 1538-4357. DOI: 10.3847/0004-637X/821/2/123. URL: http://arxiv.org/abs/1511.07873 (visited on 14/09/2021).

Harikane, Yuichi et al. (Jan. 2018). "GOLDRUSH. II. Clustering of galaxies at z 4–6 revealed with the half-million dropouts over the 100 deg2 area corresponding to 1 Gpc3". In: 70.SP1. ISSN: 0004-6264. DOI: 10.1093/pasj/psx097. URL: https://doi.org/10.1093/pasj/psx097 (visited on 22/07/2021).

Hatfield, P. W. et al. (Dec. 2019). "Comparing galaxy clustering in Horizon-AGN simulated light-cone mocks and VIDEO observations". In: 490.4, pp. 5043–5056. DOI: 10.1093/mnras/stz2946.

Hildebrandt, H. (Feb. 2016). "Observational biases in flux magnification measurements". In: 455.4, pp. 3943–3951. ISSN: 0035-8711. DOI: 10.1093/mnras/stv2575. URL: https://doi.org/10.1093/mnras/stv2575 (visited on 14/10/2021).

Hildebrandt, H., L. van Waerbeke and T. Erben (Nov. 2009). "CARS: The CFHTLS-Archive-Research Survey III. First detection of cosmic magnification in samples of normal high-z galaxies". en. In: 507.2. arXiv: 0906.1580, pp. 683–691. ISSN: 0004-6361, 1432-0746. DOI: 10.1051/0004-6361/200912655. URL: http://arxiv.org/abs/0906.1580 (visited on 07/10/2021).

Hildebrandt, H. et al. (Mar. 2013). "Inferring the mass of submillimetre galaxies by exploiting their gravitational magnification of background galaxies". In: 429.4, pp. 3230–3237. DOI: 10.1093/mnras/sts585. arXiv: 1212.2650 [astro-ph.CO].

Hildebrandt, H. et al. (Jan. 2020). "KiDS+VIKING-450: Cosmic shear tomography with optical and infrared data". In: 633, A69, A69. DOI: 10.1051/0004-6361/201834878. arXiv: 1812.06076 [astro-ph.CO].

Hopkins, Philip F., Eliot Quataert and Norman Murray (Apr. 2012). "Stellar Feedback in Galaxies and the Origin of Galaxy-scale Winds". In: 421.4. arXiv: 1110.4638, pp. 3522–3537. ISSN: 00358711. DOI: 10.1111/j.1365-2966.2012.20593.x. URL: http://arxiv.org/abs/1110.4638 (visited on 04/10/2021).

Hopkins, Philip F. et al. (Mar. 2006). "A Unified, Merger-driven Model of the Origin of Starbursts, Quasars, the Cosmic X-Ray Background, Supermassive Black Holes, and Galaxy Spheroids". In: 163.1, pp. 1–49. DOI: 10.1086/499298. arXiv: astro-ph/0506398 [astro-ph].

Hopkins, Philip F. et al. (Nov. 2014). "Galaxies on FIRE (Feedback In Realistic Environments): stellar feedback explains cosmologically inefficient star formation". In: 445, pp. 581–603. ISSN: 0035-8711. DOI: 10.1093/mnras/stu1738. URL: https://ui.adsabs.harvard.edu/abs/2014MNRAS.445..581H (visited on 25/11/2021).

Hopkins, Philip F. et al. (Jan. 2020). "Radiative stellar feedback in galaxy formation: Methods and physics". In: 491.3, pp. 3702–3729. DOI: 10.1093/mnras/stz3129. arXiv: 1811.12462 [astro-ph.GA].

Hsieh, Bau-Ching et al. (2012). "THE TAIWAN ECDFS NEAR-INFRARED SURVEY: ULTRA-DEEP J AND K S IMAGING IN THE EXTENDED CHANDRA DEEP FIELD-SOUTH". In: *The Astrophysical Journal Supplement Series* 203.2, p. 23. DOI: 10.1088/0067-0049/203/2/23. URL: https://doi.org/10.1088%2F0067-0049%2F203%2F2%2F23.

Hubble, Edwin (Jan. 1934). "The Distribution of Extra-Galactic Nebulae". In: 79, p. 8. DOI: 10.1086/143517.

Ilbert, O. et al. (Oct. 2006). "Accurate photometric redshifts for the CFHT Legacy Survey calibrated using the VIMOS VLT Deep Survey". In: 457.3. arXiv: astro-ph/0603217, pp. 841–856. ISSN: 0004-6361, 1432-0746. DOI: 10.1051/0004-6361:20065138. URL: http://arxiv.org/abs/astro-ph/0603217 (visited on 09/07/2021).

Ilbert, O. et al. (Jan. 2009). "Cosmos Photometric Redshifts with 30-Bands for 2-deg^2". In: 690.2, pp. 1236–1249. DOI: 10.1088/0004-637X/690/2/1236. arXiv: 0809.2101 [astro-ph].

Ilbert, O. et al. (Aug. 2013). "Mass assembly in quiescent and star-forming galaxies since z 4 from UltraVISTA". en. In: 556, A55. ISSN: 0004-6361. DOI: 10.1051/0004-6361/201321100. URL: https://ui.adsabs.harvard.edu/abs/2013A&A...556A..55I/abstract (visited on 21/07/2021).

Ilbert, O. et al. (Mar. 2021). "Euclid preparation: XI. Mean redshift determination from galaxy redshift probabilities for cosmic shear tomography". In: 647. arXiv: 2101.02228, A117. ISSN: 0004-6361, 1432-0746. DOI: 10.1051/0004-6361/202040237. URL: http://arxiv.org/abs/2101.02228 (visited on 28/06/2021).

Ishikawa, Shogo et al. (Nov. 2020). "The Subaru HSC Galaxy Clustering with Photometric Redshift I: Dark Halo Masses Versus Baryonic Properties of Galaxies at 0.3<z<1.4". In: 904.2. arXiv: 1912.05668, p. 128. ISSN: 1538-4357. DOI: 10.3847/1538-4357/abbd95. URL: http://arxiv.org/abs/1912.05668 (visited on 09/07/2021).

Ito, Kei et al. (Apr. 2022). "COSMOS2020: Ubiquitous AGN Activity of Massive Quiescent Galaxies at $0 < z < 5$ Revealed by X-Ray and Radio Stacking". In: 929.1, 53, p. 53. DOI: 10.3847/1538-4357/ac5aaf. arXiv: 2203.04322 [astro-ph.GA].

Ivezić, Ž. et al. (Mar. 2019). "LSST: From Science Drivers to Reference Design and Anticipated Data Products". In: 873, 111, p. 111. DOI: 10.3847/1538-4357/ab042c. arXiv: 0805.2366.

Japelj, J. et al. (Dec. 2019). “Simulating MOS science on the ELT: Lyα forest tomography”. In: 632, A94, A94. doi: 10.1051/0004-6361/201936048. arXiv: 1911.00021 [astro-ph.GA].

Jarvis, M., G. Bernstein and B. Jain (July 2004). “The skewness of the aperture mass statistic”. In: 352, pp. 338–352. issn: 0035-8711. doi: 10.1111/j.1365-2966.2004.07926.x. url: https://ui.adsabs.harvard.edu/abs/2004MNRAS.352..338J (visited on 29/06/2021).

Jin, S. et al. (Aug. 2021). “COALAS. I. ATCA CO(1-0) survey and luminosity function in the Spiderweb protocluster at z = 2.16”. In: 652, A11, A11. doi: 10.1051/0004-6361/202040232. arXiv: 2103.08884 [astro-ph.GA].

Joachimi, B. and S. L. Bridle (Nov. 2010). “Simultaneous measurement of cosmology and intrinsic alignments using joint cosmic shear and galaxy number density correlations”. In: 523. arXiv: 0911.2454, A1. issn: 0004-6361, 1432-0746. doi: 10.1051/0004-6361/200913657. url: http://arxiv.org/abs/0911.2454 (visited on 10/07/2021).

Jose, Charles et al. (Aug. 2017). “Understanding the non-linear clustering of high-redshift galaxies”. In: 469.4, pp. 4428–4436. issn: 0035-8711. doi: 10.1093/mnras/stx1014. url: https://doi.org/10.1093/mnras/stx1014 (visited on 14/09/2021).

Kaiser, N. (Sept. 1984). “On the spatial correlations of Abell clusters.” In: 284, pp. L9–L12. doi: 10.1086/184341.

Kaiser, Nick and Gordon Squires (Feb. 1993). “Mapping the Dark Matter with Weak Gravitational Lensing”. In: 404, p. 441. doi: 10.1086/172297.

Kalita, Boris S. et al. (Aug. 2021). “An Ancient Massive Quiescent Galaxy Found in a Gas-rich z 3 Group”. In: 917.2, L17, p. L17. doi: 10.3847/2041-8213/ac16dc. arXiv: 2107.13241 [astro-ph.GA].

Kartaltepe, Jeyhan et al. (Mar. 2021). *COSMOS-Webb: The Webb Cosmic Origins Survey*. JWST Proposal. Cycle 1, ID. #1727.

Katz, Harley et al. (May 2020). “How to quench a dwarf galaxy: The impact of inhomogeneous reionization on dwarf galaxies and cosmic filaments”. In: 494.2, pp. 2200–2220. doi: 10.1093/mnras/staa639. arXiv: 1905.11414 [astro-ph.GA].

Kauffmann, Guinevere et al. (Aug. 1999). “Clustering of Galaxies in a Hierarchical Universe: II. Evolution to High Redshift”. In: 307.3. arXiv: astro-ph/9809168, pp. 529–536. issn: 0035-8711, 1365-2966. doi: 10.1046/j.1365-8711.1999.02711.x. url: http://arxiv.org/abs/astro-ph/9809168 (visited on 10/09/2021).

Kauffmann, O. B. et al. (July 2022). “COSMOS2020: UV selected galaxies at $z \geq 7.5$”. In: *arXiv e-prints*, arXiv:2207.11740, arXiv:2207.11740. arXiv: 2207.11740 [astro-ph.GA].

Kaviraj, S. et al. (June 2017). “The Horizon-AGN simulation: evolution of galaxy properties over cosmic time”. In: 467.4, pp. 4739–4752. doi: 10.1093/mnras/stx126.

Kennicutt Robert C., Jr. (May 1998). “The Global Schmidt Law in Star-forming Galaxies”. In: 498.2, pp. 541–552. doi: 10.1086/305588. arXiv: astro-ph/9712213 [astro-ph].

Kennicutt, Robert C. and Neal J. Evans (Sept. 2012). “Star Formation in the Milky Way and Nearby Galaxies”. In: 50, pp. 531–608. doi: 10.1146/annurev-astro-081811-125610. arXiv: 1204.3552 [astro-ph.GA].

Kereš, Dušan et al. (Oct. 2005). “How do galaxies get their gas?” In: 363.1, pp. 2–28. doi: 10.1111/j.1365-2966.2005.09451.x. arXiv: astro-ph/0407095 [astro-ph].

Kimm, Taysun et al. (Aug. 2015). “Towards simulating star formation in turbulent high-z galaxies with mechanical supernova feedback”. In: 451, pp. 2900–2921. issn: 0035-8711. doi: 10.1093/mnras/stv1211. url: https://ui.adsabs.harvard.edu/abs/2015MNRAS.451.2900K (visited on 25/11/2021).

Klypin, Anatoly (May 2000). “Numerical Simulations in Cosmology I: Methods”. In: *arXiv e-prints*, astro-ph/0005502, astro–ph/0005502. arXiv: astro-ph/0005502 [astro-ph].

Koekemoer, A. M. et al. (Sept. 2007). “The COSMOS Survey: Hubble Space Telescope Advanced Camera for Surveys Observations and Data Processing”. In: 172.1, pp. 196–202. doi: 10.1086/520086. arXiv: astro-ph/0703095 [astro-ph].

Koekemoer, Anton M. et al. (Dec. 2011). “CANDELS: The Cosmic Assembly Near-infrared Deep Extragalactic Legacy Survey—The Hubble Space Telescope Observations, Imaging Data Products, and Mosaics”. In: 197.2, 36, p. 36. doi: 10.1088/0067-0049/197/2/36. arXiv: 1105.3754 [astro-ph.CO].

Komatsu, E. et al. (Feb. 2011). “Seven-year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Cosmological Interpretation”. In: 192.2, p. 18. doi: 10.1088/0067-0049/192/2/18.

Kovač, K. et al. (Jan. 2010). “The Density Field of the 10k zCOSMOS Galaxies”. In: 708.1, pp. 505–533. doi: 10.1088/0004-637X/708/1/505. arXiv: 0903.3409 [astro-ph.CO].

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

Kravtsov, Andrey V. et al. (July 2004). “The Dark Side of the Halo Occupation Distribution”. en. In: 609.1, p. 35. issn: 0004-637X. doi: 10.1086/420959. url: https://iopscience.iop.org/article/10.1086/420959/meta (visited on 21/09/2021).

Lagos, Claudia del P. et al. (Nov. 2019). “From the far-ultraviolet to the far-infrared - galaxy emission at $0 \leq z \leq 10$ in the SHARK semi-analytic model”. In: 489.3, pp. 4196–4216. doi: 10.1093/mnras/stz2427. arXiv: 1908.03423 [astro-ph.GA].

Laigle, C. et al. (June 2016). “The COSMOS2015 Catalog: Exploring the 1<z<6 Universe with half a million galaxies”. In: 224.2. arXiv: 1604.02350, p. 24. ISSN: 1538-4365. DOI: 10.3847/0067-0049/224/2/24. URL: http://arxiv.org/abs/1604.02350 (visited on 23/06/2021).

Laigle, C. et al. (July 2019). “Horizon-AGN virtual observatory - 1. SED-fitting performance and forecasts for future imaging surveys”. In: 486.4, pp. 5104–5123. DOI: 10.1093/mnras/stz1054.

Laigle, Clotilde et al. (Mar. 2018). “COSMOS2015 photometric redshifts probe the impact of filaments on galaxy properties”. In: 474.4. arXiv: 1702.08810, pp. 5437–5458. ISSN: 0035-8711, 1365-2966. DOI: 10.1093/mnras/stx3055. URL: http://arxiv.org/abs/1702.08810 (visited on 07/06/2021).

Landy, Stephen D. and Alexander S. Szalay (July 1993). “Bias and Variance of Angular Correlation Functions”. In: 412, p. 64. ISSN: 0004-637X. DOI: 10.1086/172900. URL: https://ui.adsabs.harvard.edu/abs/1993ApJ...412...64L (visited on 29/06/2021).

Laporte, N. et al. (Aug. 2021). “Probing cosmic dawn: Ages and star formation histories of candidate $z \geq 9$ galaxies”. In: 505.3, pp. 3336–3346. DOI: 10.1093/mnras/stab1239. arXiv: 2104.08168 [astro-ph.GA].

Larson, R. B., B. M. Tinsley and C. N. Caldwell (May 1980). “The evolution of disk galaxies and the origin of S0 galaxies”. In: 237, pp. 692–707. ISSN: 0004-637X. DOI: 10.1086/157917. URL: https://ui.adsabs.harvard.edu/abs/1980ApJ...237..692L (visited on 21/10/2021).

Laureijs, R. et al. (2011). “Euclid Definition Study Report”. In: *arXiv e-prints*.

Le Fèvre, O. et al. (Nov. 2020). “The ALPINE-ALMA [CII] survey. Survey strategy, observations, and sample properties of 118 star-forming galaxies at $4 < z < 6$”. In: 643, A1, A1. DOI: 10.1051/0004-6361/201936965. arXiv: 1910.09517 [astro-ph.CO].

Le Fèvre, Oliver et al. (Mar. 2003). “Commissioning and performances of the VLT-VIMOS instrument”. In: ed. by Masanori Iye and Alan F. M. Moorwood. Vol. 4841. Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, pp. 1670–1681. DOI: 10.1117/12.460959.

Leauthaud, A. et al. (Sept. 2007). “Weak Gravitational Lensing with COSMOS: Galaxy Selection and Shape Measurements”. In: 172.1. arXiv: astro-ph/0702359, pp. 219–238. ISSN: 0067-0049, 1538-4365. DOI: 10.1086/516598. URL: http://arxiv.org/abs/astro-ph/0702359 (visited on 13/12/2021).

Leauthaud, A. et al. (Sept. 2011). “A theoretical framework for combining techniques that probe the link between galaxies and dark matter”. In: 738.1. arXiv: 1103.2077, p. 45. ISSN: 0004-637X, 1538-4357. DOI: 10.1088/0004-637X/738/1/45. URL: http://arxiv.org/abs/1103.2077 (visited on 30/06/2021).

Leauthaud, Alexie et al. (Jan. 2012). “New constraints on the evolution of the stellar-to-dark matter connection: a combined analysis of galaxy-galaxy lensing, clustering, and stellar mass functions from z=0.2 to z=1”. In: 744.2. arXiv: 1104.0928, p. 159. ISSN: 0004-637X, 1538-4357. DOI: 10.1088/0004-637X/744/2/159. URL: http://arxiv.org/abs/1104.0928 (visited on 22/04/2021).

Legrand, Louis et al. (July 2019). “The COSMOS-UltraVISTA stellar-to-halo mass relationship: new insights on galaxy formation efficiency out to \$z∼ 5\$”. In: 486.4. arXiv: 1810.10557, pp. 5468–5481. ISSN: 0035-8711, 1365-2966. DOI: 10.1093/mnras/stz1198. URL: http://arxiv.org/abs/1810.10557 (visited on 07/07/2021).

Licquia, Timothy C. and Jeffrey A. Newman (June 2015). “Improved Estimates of the Milky Way’s Stellar Mass and Star Formation Rate from Hierarchical Bayesian Meta-Analysis”. In: 806.1. arXiv: 1407.1078, p. 96. ISSN: 1538-4357. DOI: 10.1088/0004-637X/806/1/96. URL: http://arxiv.org/abs/1407.1078 (visited on 29/11/2021).

Lilly, S. J. et al. (Dec. 1995). “The Canada-France Redshift Survey. I. Introduction to the Survey, Photometric Catalogs, and Surface Brightness Selection Effects”. In: 455, p. 50. DOI: 10.1086/176555. arXiv: astro-ph/9507010 [astro-ph].

Limber, D. Nelson (Jan. 1953). “The Analysis of Counts of the Extragalactic Nebulae in Terms of a Fluctuating Density Field.” en. In: 117, p. 134. ISSN: 0004-637X. DOI: 10.1086/145672. URL: https://ui.adsabs.harvard.edu/abs/1953ApJ...117..134L/abstract (visited on 10/07/2021).

Lotz, J. M. et al. (Mar. 2017). “The Frontier Fields: Survey Design and Initial Results”. In: 837.1, 97, p. 97. DOI: 10.3847/1538-4357/837/1/97. arXiv: 1605.06567 [astro-ph.GA].

Madau, Piero and Mark Dickinson (Aug. 2014). “Cosmic Star-Formation History”. In: 52, pp. 415–486. DOI: 10.1146/annurev-astro-081811-125615. arXiv: 1403.0007 [astro-ph.CO].

Maddox, S. J. et al. (Jan. 1990). “Galaxy correlations on large scales.” In: 242, p. 43. DOI: 10.1093/mnras/242.1.43P.

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

Mandelbaum, Rachel et al. (May 2006). “Galaxy halo masses and satellite fractions from galaxy-galaxy lensing in the Sloan Digital Sky Survey: stellar mass, luminosity, morphology and environment dependencies”. In: 368.2, pp. 715–731. DOI: 10.1111/j.1365-2966.2006.10156.x. arXiv: astro-ph/0511164 [astro-ph].

Mandelbaum, Rachel et al. (May 2006). “Galaxy halo masses and satellite fractions from galaxy-galaxy lensing in the Sloan Digital Sky Survey: stellar mass, luminosity, morphology and environment dependencies”. In: 368, pp. 715–731. ISSN: 0035-8711. DOI: 10.1111/j.1365-2966.2006.10156.x. URL: https://ui.adsabs.harvard.edu/abs/2006MNRAS.368..715M (visited on 08/10/2021).

Mandelbaum, Rachel et al. (May 2014). “The Third Gravitational Lensing Accuracy Testing (GREAT3) Challenge Handbook”. In: 212.1, 5, p. 5. DOI: 10.1088/0067-0049/212/1/5. arXiv: 1308.4982 [astro-ph.CO].

Marchesini, Danilo et al. (Aug. 2009). “The Evolution of the Stellar Mass Function of Galaxies from z = 4.0 and the First Comprehensive Analysis of its Uncertainties: Evidence for Mass-Dependent Evolution”. In: 701.2, pp. 1765–1796. DOI: 10.1088/0004-637X/701/2/1765. arXiv: 0811.1773 [astro-ph].

Marinacci, Federico et al. (Nov. 2018). “First results from the IllustrisTNG simulations: radio haloes and magnetic fields”. In: 480.4, pp. 5113–5139. DOI: 10.1093/mnras/sty2206.

Marinoni, Christian and Michael J. Hudson (Apr. 2002). “The Mass-to-Light Function of Virialized Systems and the Relationship between Their Optical and X-Ray Properties”. In: 569, pp. 101–111. ISSN: 0004-637X. DOI: 10.1086/339319. URL: https://ui.adsabs.harvard.edu/abs/2002ApJ...569..101M (visited on 03/12/2021).

Martín-Navarro, I., Joseph N. Burchett and Mar Mezcua (Oct. 2019). “Quantifying the Effect of Black Hole Feedback from the Central Galaxy on the Satellite Populations of Groups and Clusters”. In: 884.2, L45, p. L45. DOI: 10.3847/2041-8213/ab4885. arXiv: 1909.12841 [astro-ph.GA].

Martizzi, Davide et al. (Feb. 2020). “Baryons in the CosmicWeb of IllustrisTNG – II: the Connection among Galaxies, Halos, their Formation Time and their Location in the Cosmic Web”. In: 491.4. arXiv: 1907.04333, pp. 5747–5758. ISSN: 0035-8711, 1365-2966. DOI: 10.1093/mnras/stz3418. URL: http://arxiv.org/abs/1907.04333 (visited on 10/06/2021).

Matthee, Jorryt et al. (Feb. 2017). “The origin of scatter in the stellar mass - halo mass relation of central galaxies in the EAGLE simulation”. In: 465.2. arXiv: 1608.08218, pp. 2381–2396. ISSN: 0035-8711, 1365-2966. DOI: 10.1093/mnras/stw2884. URL: http://arxiv.org/abs/1608.08218 (visited on 17/12/2021).

McAlpine, Stuart et al. (Apr. 2016). “The EAGLE simulations of galaxy formation: public release of halo and galaxy catalogues”. In: *Astronomy and Computing* 15. arXiv: 1510.01320, pp. 72–89. ISSN: 22131337. DOI: 10.1016/j.ascom.2016.02.004. URL: http://arxiv.org/abs/1510.01320 (visited on 17/09/2021).

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

McCracken, H. J. et al. (Sept. 2007). “The angular correlations of galaxies in the COSMOS field”. In: 172.1. arXiv: 0704.2545, pp. 314–319. ISSN: 0067-0049, 1538-4365. DOI: 10.1086/518693. URL: http://arxiv.org/abs/0704.2545 (visited on 14/07/2021).

McCracken, H. J. et al. (Aug. 2012). “UltraVISTA: a new ultra-deep near-infrared survey in COSMOS”. en. In: 544, A156. ISSN: 0004-6361, 1432-0746. DOI: 10.1051/0004-6361/201219507. URL: https://www.aanda.org/articles/aa/abs/2012/08/aa19507-12/aa19507-12.html (visited on 23/06/2021).

McCracken, H. J. et al. (May 2015). “Probing the galaxy-halo connection in UltraVISTA to z ~ 2”. In: 449, pp. 901–916. ISSN: 0035-8711. DOI: 10.1093/mnras/stv305. URL: https://ui.adsabs.harvard.edu/abs/2015MNRAS.449..901M (visited on 09/07/2021).

Mead, A. J. et al. (Dec. 2015). “An accurate halo model for fitting non-linear cosmological power spectra and baryonic feedback models”. In: 454.2, pp. 1958–1975. ISSN: 0035-8711. DOI: 10.1093/mnras/stv2036. URL: https://doi.org/10.1093/mnras/stv2036 (visited on 26/09/2021).

Mead, Alexander et al. (Feb. 2021). “HMcode-2020: Improved modelling of non-linear cosmological power spectra with baryonic feedback”. In: 502.1. arXiv: 2009.01858, pp. 1401–1422. ISSN: 0035-8711, 1365-2966. DOI: 10.1093/mnras/stab082. URL: http://arxiv.org/abs/2009.01858 (visited on 26/09/2021).

Ménard, B. and M. Bartelmann (May 2002). “Cosmological information from quasar-galaxy correlations induced by weak lensing”. In: 386, pp. 784–795. DOI: 10.1051/0004-6361:20020274. arXiv: astro-ph/0203163 [astro-ph].

Meneux, B. et al. (Oct. 2009). “The zCOSMOS survey. The dependence of clustering on luminosity and stellar mass at z=0.2-1”. en. In: 505.2, p. 463. ISSN: 0004-6361. DOI: 10.1051/0004-6361/200912314. URL: https://ui.adsabs.harvard.edu/abs/2009A%26A...505..463M/abstract (visited on 14/07/2021).

Miller, T. B. et al. (Apr. 2018). “A massive core for a cluster of galaxies at a redshift of 4.3”. In: 556.7702, pp. 469–472. DOI: 10.1038/s41586-018-0025-2. arXiv: 1804.09231 [astro-ph.GA].

Mo, H. J. and S. D. M. White (Sept. 1996). “An analytic model for the spatial clustering of dark matter haloes”. In: 282.2, pp. 347–361. DOI: 10.1093/mnras/282.2.347. arXiv: astro-ph/9512127 [astro-ph].

Moessner, R. and Bhuvnesh Jain (Feb. 1998). “Angular cross-correlation of galaxies: a probe of gravitational lensing by large-scale structure”. In: 294.1, pp. L18–L24. DOI: 10.1046/j.1365-8711.1998.01378.x. arXiv: astro-ph/9709159 [astro-ph].

Moneti, Andrea et al. (Oct. 2021). “Euclid preparation: XVIII. Cosmic Dawn Survey. Spitzer observations of the Euclid deep fields and calibration fields”. In: *arXiv:2110.13928 [astro-ph]*. arXiv: 2110.13928. URL: http://arxiv.org/abs/2110.13928 (visited on 28/10/2021).

Moore, Ben et al. (Feb. 1996). “Galaxy Harassment and the Evolution of Clusters of Galaxies”. In: *Nature* 379.6566. arXiv: astro-ph/9510034, pp. 613–616. ISSN: 0028-0836, 1476-4687. DOI: 10.1038/379613a0. URL: http://arxiv.org/abs/astro-ph/9510034 (visited on 21/10/2021).

Moster, Benjamin P., Thorsten Naab and Simon D. M. White (Feb. 2013). “Galactic star formation and accretion histories from matching galaxies to dark matter haloes”. In: 428, pp. 3121–3138. ISSN: 0035-8711. DOI:

10.1093/mnras/sts261. URL: https://ui.adsabs.harvard.edu/abs/2013MNRAS.428.3121M (visited on 22/07/2021).

— (June 2018). “EMERGE - an empirical model for the formation of galaxies since z 10”. In: 477, pp. 1822–1852. ISSN: 0035-8711. DOI: 10.1093/mnras/sty655. URL: https://ui.adsabs.harvard.edu/abs/2018MNRAS.477.1822M (visited on 22/07/2021).

Moster, Benjamin P. et al. (Feb. 2010). “Constraints on the relationship between stellar mass and halo mass at low and high redshift”. In: 710.2. arXiv: 0903.4682, pp. 903–923. ISSN: 0004-637X, 1538-4357. DOI: 10.1088/0004-637X/710/2/903. URL: http://arxiv.org/abs/0903.4682 (visited on 08/07/2021).

Moster, Benjamin P. et al. (Mar. 2011). “A COSMIC VARIANCE COOKBOOK”. en. In: 731.2, p. 113. ISSN: 0004-637X. DOI: 10.1088/0004-637X/731/2/113. URL: https://doi.org/10.1088/0004-637x/731/2/113 (visited on 17/12/2021).

Muzzin, Adam et al. (May 2013). “A Public Ks -selected Catalog in the COSMOS/ULTRAVISTA Field: Photometry, Photometric Redshifts, and Stellar Population Parameters”. In: *The Astrophysical Journal Supplement Series* 206, p. 8. ISSN: 0067-0049. DOI: 10.1088/0067-0049/206/1/8. URL: http://adsabs.harvard.edu/abs/2013ApJS..206....8M (visited on 08/05/2020).

Naab, Thorsten and Jeremiah P. Ostriker (Aug. 2017). “Theoretical Challenges in Galaxy Formation”. In: 55.1. arXiv: 1612.06891, pp. 59–109. ISSN: 0066-4146, 1545-4282. DOI: 10.1146/annurev-astro-081913-040019. URL: http://arxiv.org/abs/1612.06891 (visited on 03/12/2021).

Naiman, Jill P. et al. (June 2018). “First results from the IllustrisTNG simulations: a tale of two elements - chemical evolution of magnesium and europium”. In: 477.1, pp. 1206–1224. DOI: 10.1093/mnras/sty618.

Navarro, Julio F., Carlos S. Frenk and Simon D. M. White (Dec. 1997). “A Universal Density Profile from Hierarchical Clustering”. In: 490, pp. 493–508. ISSN: 0004-637X. DOI: 10.1086/304888. URL: https://ui.adsabs.harvard.edu/abs/1997ApJ...490..493N (visited on 09/07/2021).

Nayyeri, H. et al. (Jan. 2017). “CANDELS Multiwavelength Catalogs: Source Identification and Photometry in the CANDELS COSMOS Survey Field”. In: 228.1. arXiv: 1612.07364, p. 7. ISSN: 1538-4365. DOI: 10.3847/1538-4365/228/1/7. URL: http://arxiv.org/abs/1612.07364 (visited on 27/06/2021).

Nelson, Dylan et al. (Mar. 2018). “First results from the IllustrisTNG simulations: the galaxy colour bimodality”. en. In: 475.1, pp. 624–647. ISSN: 0035-8711, 1365-2966. DOI: 10.1093/mnras/stx3040. URL: https://academic.oup.com/mnras/article/475/1/624/4657187 (visited on 05/10/2021).

Nelson, Dylan et al. (May 2019). “The IllustrisTNG simulations: public data release”. In: *Computational Astrophysics and Cosmology* 6.1, 2, p. 2. DOI: 10.1186/s40668-019-0028-x. arXiv: 1812.05609 [astro-ph.GA].

Neyman, J. and E. L. Scott (July 1952). “A Theory of the Spatial Distribution of Galaxies.” In: 116, p. 144. DOI: 10.1086/145599.

Noeske, K. G. et al. (May 2007). “Star Formation in AEGIS Field Galaxies since z=1.1: The Dominance of Gradually Declining Star Formation, and the Main Sequence of Star-forming Galaxies”. In: 660.1, pp. L43–L46. DOI: 10.1086/517926. arXiv: astro-ph/0701924 [astro-ph].

Nuñez-Castiñeyra, A. et al. (Jan. 2021). “Cosmological simulations of the same spiral galaxy: the impact of baryonic physics”. In: 501.1, pp. 62–77. DOI: 10.1093/mnras/staa3233. arXiv: 2004.06008 [astro-ph.GA].

Oesch, P. A. et al. (Mar. 2016). “A Remarkably Luminous Galaxy at z=11.1 Measured with Hubble Space Telescope Grism Spectroscopy”. In: 819.2, 129, p. 129. DOI: 10.3847/0004-637X/819/2/129. arXiv: 1603.00461 [astro-ph.GA].

Oke, J. B. and J. E. Gunn (Mar. 1983). “Secondary standard stars for absolute spectrophotometry.” In: 266, pp. 713–717. ISSN: 0004-637X. DOI: 10.1086/160817. URL: https://ui.adsabs.harvard.edu/abs/1983ApJ...266..713O (visited on 20/10/2021).

Oliver, S. J. et al. (Aug. 2012). “The Herschel Multi-tiered Extragalactic Survey: HerMES”. In: 424.3, pp. 1614–1635. DOI: 10.1111/j.1365-2966.2012.20912.x. arXiv: 1203.2562 [astro-ph.CO].

Oser, Ludwig et al. (Dec. 2010). “The Two Phases of Galaxy Formation”. In: 725.2, pp. 2312–2323. DOI: 10.1088/0004-637X/725/2/2312. arXiv: 1010.1381 [astro-ph.CO].

Ostriker, J. P., P. J. E. Peebles and A. Yahil (Oct. 1974). “The Size and Mass of Galaxies, and the Mass of the Universe”. In: 193, p. L1. DOI: 10.1086/181617.

Oteo, I. et al. (Mar. 2018). “An Extreme Protocluster of Luminous Dusty Starbursts in the Early Universe”. In: 856.1, 72, p. 72. DOI: 10.3847/1538-4357/aaa1f1. arXiv: 1709.02809 [astro-ph.GA].

Overzier, Roderik A. (Nov. 2016). “The realm of the galaxy protoclusters. A review”. In: 24.1, 14, p. 14. DOI: 10.1007/s00159-016-0100-3. arXiv: 1610.05201 [astro-ph.GA].

Pandey, S. et al. (May 2021). “Dark Energy Survey Year 3 Results: Constraints on cosmological parameters and galaxy bias models from galaxy clustering and galaxy-galaxy lensing using the redMaGiC sample”. In: *arXiv e-prints*, arXiv:2105.13545, arXiv:2105.13545. arXiv: 2105.13545 [astro-ph.CO].

Peacock, J. A. and R. E. Smith (Nov. 2000). “Halo occupation numbers and galaxy bias”. In: 318.4, pp. 1144–1156. DOI: 10.1046/j.1365-8711.2000.03779.x. arXiv: astro-ph/0005010 [astro-ph].

Peacock, J. A. and R. E. Smith (Nov. 2000). “Halo occupation numbers and galaxy bias”. In: 318.4. arXiv: astro-ph/0005010, pp. 1144–1156. ISSN: 0035-8711, 1365-2966. DOI: 10.1046/j.1365-8711.2000.03779.x. URL: http://arxiv.org/abs/astro-ph/0005010 (visited on 09/07/2021).

Peebles, P. J. E. (Mar. 1975). "Statistical analysis of catalogs of extragalactic objects. VI. The galaxy distribution in the Jagellonian field." In: 196, pp. 647–651. DOI: 10.1086/153450.

— (Jan. 2001). "The Galaxy and Mass N-Point Correlation Functions: a Blast from the Past". In: *Historical Development of Modern Cosmology*. Ed. by Vicent J. Martínez, Virginia Trimble and María Jesús Pons-Bordería. Vol. 252. Astronomical Society of the Pacific Conference Series, p. 201. arXiv: astro-ph/0103040 [astro-ph].

Peirani, Sébastien et al. (Dec. 2017). "Density profile of dark matter haloes and galaxies in the HORIZON-AGN simulation: the impact of AGN feedback". In: 472.2, pp. 2153–2169. DOI: 10.1093/mnras/stx2099. arXiv: 1611.09922 [astro-ph.GA].

Peng, Y. et al. (Sept. 2010). "Mass and environment as drivers of galaxy evolution in SDSS and zCOSMOS and the origin of the Schechter function". In: 721.1. arXiv: 1003.4747, pp. 193–221. ISSN: 0004-637X, 1538-4357. DOI: 10.1088/0004-637X/721/1/193. URL: http://arxiv.org/abs/1003.4747 (visited on 26/09/2021).

Perlmutter, S. et al. (June 1999). "Measurements of Ω and Λ from 42 High-Redshift Supernovae". In: 517.2, pp. 565–586. DOI: 10.1086/307221. arXiv: astro-ph/9812133 [astro-ph].

Pieri, M. M. et al. (Dec. 2016). "WEAVE-QSO: A Massive Intergalactic Medium Survey for the William Herschel Telescope". In: *SF2A-2016: Proceedings of the Annual meeting of the French Society of Astronomy and Astrophysics*. Ed. by C. Reylé et al., pp. 259–266. arXiv: 1611.09388 [astro-ph.CO].

Pillepich, Annalisa et al. (Mar. 2018). "First results from the IllustrisTNG simulations: the stellar mass content of groups and clusters of galaxies". In: 475.1. arXiv: 1707.03406, pp. 648–675. ISSN: 0035-8711, 1365-2966. DOI: 10.1093/mnras/stx3112. URL: http://arxiv.org/abs/1707.03406 (visited on 23/09/2021).

Pillepich, Annalisa et al. (Jan. 2018). "Simulating galaxy formation with the IllustrisTNG model". In: 473.3, pp. 4077–4106. DOI: 10.1093/mnras/stx2656. arXiv: 1703.02970 [astro-ph.GA].

Planck Collaboration et al. (Sept. 2016). "Planck 2015 results. XIII. Cosmological parameters". In: 594, A13. DOI: 10.1051/0004-6361/201525830.

Planck Collaboration et al. (Sept. 2020). "Planck 2018 results. VI. Cosmological parameters". In: 641, A6, A6. DOI: 10.1051/0004-6361/201833910. arXiv: 1807.06209 [astro-ph.CO].

Potter, Douglas, Joachim Stadel and Romain Teyssier (May 2017). "PKDGRAV3: beyond trillion particle cosmological simulations for the next era of galaxy surveys". In: *Computational Astrophysics and Cosmology* 4.1, 2, p. 2. DOI: 10.1186/s40668-017-0021-1. arXiv: 1609.08621 [astro-ph.IM].

Pozzetti, L. et al. (Nov. 2010). "zCOSMOS - 10k-bright spectroscopic sample. The bimodality in the Galaxy Stellar Mass Function: exploring its evolution with redshift". In: 523. arXiv: 0907.5416, A13. ISSN: 0004-6361, 1432-0746. DOI: 10.1051/0004-6361/200913020. URL: http://arxiv.org/abs/0907.5416 (visited on 27/06/2021).

Press, William H. and Paul Schechter (Feb. 1974). "Formation of Galaxies and Clusters of Galaxies by Self-Similar Gravitational Condensation". In: 187, pp. 425–438. DOI: 10.1086/152650.

Prevot, M. L. et al. (Mar. 1984). "The typical interstellar extinction in the Small Magellanic Cloud." en. In: 132, pp. 389–392. ISSN: 0004-6361. URL: https://ui.adsabs.harvard.edu/abs/1984A&A...132..389P/abstract (visited on 21/07/2021).

Prgomet, Mateo et al. (July 2021). "EDGE: The sensitivity of ultra-faint dwarfs to a metallicity-dependent initial mass function". In: *arXiv e-prints*, arXiv:2107.00663, arXiv:2107.00663. arXiv: 2107.00663 [astro-ph.GA].

Quilley, Louis and Valérie de Lapparent (June 2022). "Aging of galaxies along the morphological sequence, marked by bulge growth and disk quenching". In: *arXiv e-prints*, arXiv:2206.04707, arXiv:2206.04707. arXiv: 2206.04707 [astro-ph.GA].

Roberts, M. S. and A. H. Rots (Aug. 1973). "Comparison of Rotation Curves of Different Galaxy Types". In: 26, pp. 483–485.

Robertson, Brant E. (Oct. 2021). "Galaxy Formation and Reionization: Key Unknowns and Expected Breakthroughs by the James Webb Space Telescope". In: *arXiv e-prints*, arXiv:2110.13160, arXiv:2110.13160. arXiv: 2110.13160 [astro-ph.CO].

Roche, Nathan and Stephen A. Eales (Aug. 1999). "The angular correlation function and hierarchical moments of ~70000 faint galaxies to R=23.5". In: 307, pp. 703–721. ISSN: 0035-8711. DOI: 10.1046/j.1365-8711.1999.02652.x. URL: https://ui.adsabs.harvard.edu/abs/1999MNRAS.307..703R (visited on 29/06/2021).

Rodighiero, G. et al. (Oct. 2011). "The Lesser Role of Starbursts in Star Formation at z = 2". In: 739.2, L40, p. L40. DOI: 10.1088/2041-8205/739/2/L40. arXiv: 1108.0933 [astro-ph.CO].

Rodríguez Montero, Francisco et al. (Dec. 2019). "Mergers, starbursts, and quenching in the SIMBA simulation". In: 490.2, pp. 2139–2154. DOI: 10.1093/mnras/stz2580. arXiv: 1907.12680 [astro-ph.GA].

Rodríguez-Monroy, M. et al. (May 2021). "Dark Energy Survey Year 3 Results: Galaxy clustering and systematics treatment for lens galaxy samples". In: *arXiv:2105.13540 [astro-ph]*. arXiv: 2105.13540. URL: http://arxiv.org/abs/2105.13540 (visited on 19/10/2021).

Rosdahl, J. et al. (Dec. 2013). "RAMSES-RT: radiation hydrodynamics in the cosmological context". In: 436.3, pp. 2188–2231. DOI: 10.1093/mnras/stt1722. arXiv: 1304.7126 [astro-ph.CO].

Rosdahl, Joakim et al. (Apr. 2017). "Snap, crackle, pop: sub-grid supernova feedback in AMR simulations of disc galaxies". In: 466.1, pp. 11–33. DOI: 10.1093/mnras/stw3034. arXiv: 1609.01296 [astro-ph.GA].

Rubin, V. C., Jr. Ford W. K. and N. Thonnard (Nov. 1978). “Extended rotation curves of high-luminosity spiral galaxies. IV. Systematic dynamical properties, Sa -> Sc.” In: 225, pp. L107–L111. DOI: `10.1086/182804`.

Salem, Munier and Greg L. Bryan (Feb. 2014). “Cosmic ray driven outflows in global galaxy disc models”. In: 437.4, pp. 3312–3330. DOI: `10.1093/mnras/stt2121`. arXiv: `1307.6215 [astro-ph.CO]`.

Salpeter, Edwin E. (Jan. 1955). “The Luminosity Function and Stellar Evolution.” In: 121, p. 161. DOI: `10.1086/145971`.

Salvato, Mara, Olivier Ilbert and Ben Hoyle (June 2019). “The many flavours of photometric redshifts”. In: *Nature Astronomy* 3, pp. 212–222. DOI: `10.1038/s41550-018-0478-0`. arXiv: `1805.12574 [astro-ph.GA]`.

Sanders, D. B. et al. (Sept. 2007). “S-COSMOS: The Spitzer Legacy Survey of the Hubble Space Telescope ACS 2 deg^2 COSMOS Field I: Survey Strategy and First Analysis”. In: 172.1, pp. 86–98. DOI: `10.1086/517885`. arXiv: `astro-ph/0701318 [astro-ph]`.

Sawicki, Marcin et al. (Nov. 2019). “The CFHT large area U-band deep survey (CLAUDS)”. In: 489, pp. 5202–5217. ISSN: 0035-8711. DOI: `10.1093/mnras/stz2522`. URL: `https://ui.adsabs.harvard.edu/abs/2019MNRAS.489.5202S` (visited on 21/07/2021).

Schawinski, Kevin et al. (May 2014). “The Green Valley is a Red Herring: Galaxy Zoo reveals two evolutionary pathways towards quenching of star formation in early- and late-type galaxies”. In: 440.1. arXiv: 1402.4814, pp. 889–907. ISSN: 1365-2966, 0035-8711. DOI: `10.1093/mnras/stu327`. URL: `http://arxiv.org/abs/1402.4814` (visited on 02/06/2021).

Schaye, Joop et al. (Jan. 2015). “The EAGLE project: simulating the evolution and assembly of galaxies and their environments”. In: 446.1, pp. 521–554. DOI: `10.1093/mnras/stu2058`.

Schlafly, Edward F. and Douglas P. Finkbeiner (Aug. 2011). “Measuring Reddening with Sloan Digital Sky Survey Stellar Spectra and Recalibrating SFD”. In: 737, p. 103. ISSN: 0004-637X. DOI: `10.1088/0004-637X/737/2/103`. URL: `https://ui.adsabs.harvard.edu/abs/2011ApJ...737..103S` (visited on 25/10/2021).

Schmidt, Maarten (Feb. 1968). “Space Distribution and Luminosity Functions of Quasi-Stellar Radio Sources”. In: 151, p. 393. ISSN: 0004-637X. DOI: `10.1086/149446`. URL: `https://ui.adsabs.harvard.edu/abs/1968ApJ...151..393S` (visited on 29/06/2021).

Schneider, Peter (Sept. 2005). “Weak Gravitational Lensing”. In: *arXiv e-prints*, astro-ph/0509252, astro–ph/0509252. arXiv: `astro-ph/0509252 [astro-ph]`.

Scoccimarro, Román et al. (Jan. 2001). “How Many Galaxies Fit in a Halo? Constraints on Galaxy Formation Efficiency from Spatial Clustering”. In: 546, pp. 20–34. ISSN: 0004-637X. DOI: `10.1086/318261`. URL: `https://ui.adsabs.harvard.edu/abs/2001ApJ...546...20S` (visited on 08/10/2021).

Scoville, N. et al. (Sept. 2007). “The Cosmic Evolution Survey (COSMOS): Overview”. In: 172.1, pp. 1–8. DOI: `10.1086/516585`. arXiv: `astro-ph/0612305 [astro-ph]`.

Scranton, Ryan et al. (Nov. 2005). “Detection of Cosmic Magnification with the Sloan Digital Sky Survey”. en. In: 633.2, pp. 589–602. ISSN: 0004-637X, 1538-4357. DOI: `10.1086/431358`. URL: `https://iopscience.iop.org/article/10.1086/431358` (visited on 14/10/2021).

Seljak, Uros (Oct. 2000). “Analytic model for galaxy and dark matter clustering”. In: 318.1. arXiv: astro-ph/0001493, pp. 203–213. ISSN: 0035-8711, 1365-2966. DOI: `10.1046/j.1365-8711.2000.03715.x`. URL: `http://arxiv.org/abs/astro-ph/0001493` (visited on 09/07/2021).

Shane, C. D. and C. A. Wirtanen (Sept. 1954). “The distribution of extragalactic nebulae”. In: 59, pp. 285–304. DOI: `10.1086/107014`.

Shankar, F. et al. (May 2006). “New Relationships between Galaxy Properties and Host Halo Mass, and the Role of Feedbacks in Galaxy Formation”. In: 643.1, pp. 14–25. DOI: `10.1086/502794`. arXiv: `astro-ph/0601577 [astro-ph]`.

Silk, Joe and Gary A. Mamon (Aug. 2012). “The Current Status of Galaxy Formation”. In: *Research in Astronomy and Astrophysics* 12.8. arXiv: 1207.3080, pp. 917–946. ISSN: 1674-4527. DOI: `10.1088/1674-4527/12/8/004`. URL: `http://arxiv.org/abs/1207.3080` (visited on 27/09/2021).

Silk, Joseph, Arianna Di Cintio and Irina Dvorkin (2014). “Galaxy formation”. In: *Proceedings of the International School of Physics Enrico Fermi* 186.New Horizons for Observational Cosmology. arXiv: 1312.0107, pp. 137–187. ISSN: 0074-784X. DOI: `10.3254/978-1-61499-476-3-137`. URL: `http://arxiv.org/abs/1312.0107` (visited on 27/09/2021).

Somerville, Rachel S. and Romeel Davé (Aug. 2015). “Physical Models of Galaxy Formation in a Cosmological Framework”. In: 53.1. arXiv: 1412.2712, pp. 51–113. ISSN: 0066-4146, 1545-4282. DOI: `10.1146/annurev-astro-082812-140951`. URL: `http://arxiv.org/abs/1412.2712` (visited on 08/10/2021).

Song, Hyunmi et al. (Jan. 2021). “Beyond halo mass: quenching galaxy mass assembly at the edge of filaments”. In: 501.3. arXiv: 2009.00013, pp. 4635–4656. ISSN: 0035-8711, 1365-2966. DOI: `10.1093/mnras/staa3981`. URL: `http://arxiv.org/abs/2009.00013` (visited on 08/07/2021).

Song, Mimi et al. (July 2016). “The Evolution of the Galaxy Stellar Mass Function at z = 4-8: A Steepening Low-mass-end Slope with Increasing Redshift”. In: 825.1, 5, p. 5. DOI: `10.3847/0004-637X/825/1/5`. arXiv: `1507.05636 [astro-ph.GA]`.

Sorba, Robert and Marcin Sawicki (Oct. 2010). “USING THE 1.6 m BUMP TO STUDY REST-FRAME NEAR-INFRARED-SELECTED GALAXIES AT REDSHIFT 2”. en. In: 721.2, pp. 1056–1078. ISSN: 0004-637X, 1538-

4357. DOI: 10.1088/0004-637X/721/2/1056. URL: https://iopscience.iop.org/article/10.1088/0004-637X/721/2/1056 (visited on 04/10/2021).

Spergel, D. et al. (Mar. 2015). "Wide-Field InfrarRed Survey Telescope-Astrophysics Focused Telescope Assets WFIRST-AFTA 2015 Report". In: *arXiv e-prints*, arXiv:1503.03757, arXiv:1503.03757. arXiv: 1503.03757 [astro-ph.IM].

Springel, Volker (Dec. 2005). "The cosmological simulation code GADGET-2". In: 364, pp. 1105–1134. ISSN: 0035-8711. DOI: 10.1111/j.1365-2966.2005.09655.x. URL: https://ui.adsabs.harvard.edu/abs/2005MNRAS.364.1105S (visited on 26/09/2021).

— (Sept. 2010). "Smoothed Particle Hydrodynamics in Astrophysics". In: 48, pp. 391–430. DOI: 10.1146/annurev-astro-081309-130914.

Springel, Volker and Lars Hernquist (Feb. 2003). "Cosmological SPH simulations: A hybrid multi-phase model for star formation". In: 339.2. arXiv: astro-ph/0206393, pp. 289–311. ISSN: 0035-8711, 1365-2966. DOI: 10.1046/j.1365-8711.2003.06206.x. URL: http://arxiv.org/abs/astro-ph/0206393 (visited on 04/10/2021).

Springel, Volker et al. (Dec. 2001). "Populating a cluster of galaxies - I. Results at [formmu2]z=0". In: 328, pp. 726–750. ISSN: 0035-8711. DOI: 10.1046/j.1365-8711.2001.04912.x. URL: https://ui.adsabs.harvard.edu/abs/2001MNRAS.328..726S (visited on 17/12/2021).

Springel, Volker et al. (Mar. 2018). "First results from the IllustrisTNG simulations: matter and galaxy clustering". In: 475.1, pp. 676–698. DOI: 10.1093/mnras/stx3304.

Stefanon, Mauro et al. (Sept. 2019). "The Brightest z $\gtrsim$ 8 Galaxies over the COSMOS UltraVISTA Field". In: 883.1, 99, p. 99. DOI: 10.3847/1538-4357/ab3792. arXiv: 1902.10713 [astro-ph.GA].

Stefanon, Mauro et al. (Nov. 2021). "Galaxy Stellar Mass Functions from z 10 to z 6 using the Deepest Spitzer-/Infrared Array Camera Data: No Significant Evolution in the Stellar-to-halo Mass Ratio of Galaxies in the First Gigayear of Cosmic Time". In: 922.1, 29, p. 29. DOI: 10.3847/1538-4357/ac1bb6. arXiv: 2103.16571 [astro-ph.GA].

Steinhardt, Charles L., Christian Kragh Jespersen and Nora B. Linzer (Dec. 2021). "Finding High-redshift Galaxies with JWST". In: 923.1, 8, p. 8. DOI: 10.3847/1538-4357/ac2a2f. arXiv: 2111.14865 [astro-ph.GA].

Szalay, Alexander S., Andrew J. Connolly and Gyula P. Szokoly (Jan. 1999). "Simultaneous Multicolor Detection of Faint Galaxies in the Hubble Deep Field". In: 117.1, pp. 68–74. DOI: 10.1086/300689.

Tacchella, Sandro et al. (Nov. 2018). "A Redshift-independent Efficiency Model: Star Formation and Stellar Masses in Dark Matter Halos at z $\gtrsim $4". en. In: 868.2, p. 92. ISSN: 0004-637X. DOI: 10.3847/1538-4357/aae8e0. URL: https://doi.org/10.3847/1538-4357/aae8e0 (visited on 22/07/2021).

Tacconi, Linda J., Reinhard Genzel and Amiel Sternberg (Aug. 2020). "The Evolution of the Star-Forming Interstellar Medium Across Cosmic Time". In: 58, pp. 157–203. DOI: 10.1146/annurev-astro-082812-141034. arXiv: 2003.06245 [astro-ph.GA].

Takada, Masahiro et al. (Feb. 2014). "Extragalactic science, cosmology, and Galactic archaeology with the Subaru Prime Focus Spectrograph". In: 66.1, R1, R1. DOI: 10.1093/pasj/pst019. arXiv: 1206.0737 [astro-ph.CO].

Tallada, P. et al. (2020). "CosmoHub: Interactive exploration and distribution of astronomical data on Hadoop". In: *Astronomy and Computing* 32, p. 100391. ISSN: 2213-1337. DOI: https://doi.org/10.1016/j.ascom.2020.100391. URL: http://www.sciencedirect.com/science/article/pii/S2213133720300457.

Taniguchi, Y. et al. (Sept. 2007). "The Cosmic Evolution Survey (COSMOS): Subaru Observations of the HST Cosmos Field". In: 172, pp. 9–28. ISSN: 0067-0049. DOI: 10.1086/516596. URL: https://ui.adsabs.harvard.edu/abs/2007ApJS..172....9T (visited on 21/07/2021).

Taniguchi, Yoshiaki et al. (Dec. 2015). "The Subaru COSMOS 20: Subaru optical imaging of the HST COSMOS field with 20 filters*". In: 67.6. ISSN: 0004-6264. DOI: 10.1093/pasj/psv106. URL: https://doi.org/10.1093/pasj/psv106 (visited on 21/07/2021).

Tasitsiomi, Argyro et al. (Oct. 2004). "Modeling Galaxy-Mass Correlations in Dissipationless Simulations". In: 614, pp. 533–546. ISSN: 0004-637X. DOI: 10.1086/423784. URL: https://ui.adsabs.harvard.edu/abs/2004ApJ...614..533T (visited on 08/10/2021).

Teyssier, R. (Apr. 2002). "Cosmological hydrodynamics with adaptive mesh refinement. A new high resolution code called RAMSES". In: 385, pp. 337–364. DOI: 10.1051/0004-6361:20011817.

Thomas, Daniel et al. (Mar. 2005). "The Epochs of Early-Type Galaxy Formation as a Function of Environment". In: 621.2, pp. 673–694. DOI: 10.1086/426932. arXiv: astro-ph/0410209 [astro-ph].

Thomas, Daniel et al. (Mar. 2010). "Environment and self-regulation in galaxy formation". In: arXiv: 0912.0259. ISSN: 00358711, 13652966. DOI: 10.1111/j.1365-2966.2010.16427.x. URL: http://arxiv.org/abs/0912.0259 (visited on 27/09/2021).

Tinker, Jeremy L. et al. (Dec. 2010). "The Large Scale Bias of Dark Matter Halos: Numerical Calibration and Model Tests". In: 724.2. arXiv: 1001.3162, pp. 878–886. ISSN: 0004-637X, 1538-4357. DOI: 10.1088/0004-637X/724/2/878. URL: http://arxiv.org/abs/1001.3162 (visited on 09/07/2021).

Tinker, Jeremy L. et al. (Nov. 2013). “Evolution of the Stellar-to-Dark Matter Relation: Separating Star-Forming and Passive Galaxies from z=1 to 0”. In: 778.2. arXiv: 1308.2974, p. 93. issn: 0004-637X, 1538-4357. doi: 10.1088/0004-637X/778/2/93. url: http://arxiv.org/abs/1308.2974 (visited on 31/05/2021).

Totsuji, H. and T. Kihara (Jan. 1969). “The Correlation Function for the Distribution of Galaxies”. In: 21, p. 221.

Tudorica, A. et al. (2017). “Weak lensing magnification of SpARCS galaxy clusters”. In: 608, A141, A141. doi: 10.1051/0004-6361/201731267. arXiv: 1710.06431 [astro-ph.CO].

Tweed, D. et al. (Nov. 2009). “Building Merger Trees from Cosmological N-body Simulations”. In: 506.2. arXiv: 0902.0679, pp. 647–660. issn: 0004-6361, 1432-0746. doi: 10.1051/0004-6361/200911787. url: http://arxiv.org/abs/0902.0679 (visited on 26/09/2021).

Uitert, Edo van et al. (June 2018). “KiDS+GAMA: Cosmology constraints from a joint analysis of cosmic shear, galaxy-galaxy lensing and angular clustering”. In: 476.4. arXiv: 1706.05004, pp. 4662–4689. issn: 0035-8711, 1365-2966. doi: 10.1093/mnras/sty551. url: http://arxiv.org/abs/1706.05004 (visited on 10/09/2021).

Vale, A. and J. P. Ostriker (Sept. 2004). “Linking halo mass to galaxy luminosity”. In: 353, pp. 189–200. issn: 0035-8711. doi: 10.1111/j.1365-2966.2004.08059.x. url: https://ui.adsabs.harvard.edu/abs/2004MNRAS.353..189V (visited on 08/10/2021).

van Daalen, Marcel P. et al. (Aug. 2011). “The effects of galaxy formation on the matter power spectrum: a challenge for precision cosmology”. In: 415.4, pp. 3649–3665. doi: 10.1111/j.1365-2966.2011.18981.x. arXiv: 1104.1174 [astro-ph.CO].

van den Bosch, Frank C. et al. (Aug. 2004). “Probing dark matter haloes with satellite kinematics”. In: 352.4, pp. 1302–1314. doi: 10.1111/j.1365-2966.2004.08021.x. arXiv: astro-ph/0404033 [astro-ph].

van Uitert, E. et al. (Oct. 2011). “Galaxy-galaxy lensing constraints on the relation between baryons and dark matter in galaxies in the Red Sequence Cluster Survey 2”. In: 534, A14, A14. doi: 10.1051/0004-6361/201117308. arXiv: 1107.4093 [astro-ph.CO].

van Waerbeke, L. (Jan. 2010). “Shear and magnification: cosmic complementarity”. In: 401.3, pp. 2093–2100. doi: 10.1111/j.1365-2966.2009.15809.x. arXiv: 0906.1583 [astro-ph.CO].

Van Waerbeke, L. et al. (Nov. 2010). “Magnification as a Probe of Dark Matter Halos at High Redshifts”. In: 723.1, pp. L13–L16. doi: 10.1088/2041-8205/723/1/L13. arXiv: 1004.3793 [astro-ph.CO].

Vardoulaki, Eleni et al. (Apr. 2021). “The M*-Mhalo relation at 0.08 < z < 1.53 in COSMOS: the role of AGN radio-mode feedback”. In: *arXiv:2104.04288 [astro-ph]*. arXiv: 2104.04288. url: http://arxiv.org/abs/2104.04288 (visited on 11/10/2021).

Vogelsberger, Mark et al. (Dec. 2013). “A model for cosmological simulations of galaxy formation physics”. In: 436.4. arXiv: 1305.2913, pp. 3031–3067. issn: 1365-2966, 0035-8711. doi: 10.1093/mnras/stt1789. url: http://arxiv.org/abs/1305.2913 (visited on 04/10/2021).

Vogelsberger, Mark et al. (Oct. 2014). “Introducing the Illustris Project: Simulating the coevolution of dark and visible matter in the Universe”. In: 444.2. arXiv: 1405.2921, pp. 1518–1547. issn: 0035-8711, 1365-2966. doi: 10.1093/mnras/stu1536. url: http://arxiv.org/abs/1405.2921 (visited on 20/12/2021).

Vogelsberger, Mark et al. (Jan. 2020). “Cosmological simulations of galaxy formation”. In: *Nature Reviews Physics* 2.1, pp. 42–66. doi: 10.1038/s42254-019-0127-2. arXiv: 1909.07976 [astro-ph.GA].

Walter, Fabian et al. (Oct. 2020). “The Evolution of the Baryons Associated with Galaxies Averaged over Cosmic Time and Space”. In: 902.2, 111, p. 111. doi: 10.3847/1538-4357/abb82e. arXiv: 2009.11126 [astro-ph.GA].

Wang, T. et al. (Aug. 2019). “A dominant population of optically invisible massive galaxies in the early Universe”. In: 572.7768, pp. 211–214. doi: 10.1038/s41586-019-1452-4. arXiv: 1908.02372 [astro-ph.GA].

Wang, Tao et al. (Sept. 2016). “Discovery of a Galaxy Cluster with a Violently Starbursting Core at z = 2.506”. In: 828.1, 56, p. 56. doi: 10.3847/0004-637X/828/1/56. arXiv: 1604.07404 [astro-ph.GA].

Watson, Douglas F., Andreas A. Berlind and Andrew R. Zentner (Sept. 2011). “A Cosmic Coincidence: The Power-law Galaxy Correlation Function”. In: 738.1, 22, p. 22. doi: 10.1088/0004-637X/738/1/22. arXiv: 1101.5155 [astro-ph.CO].

Watson, Douglas F. et al. (Apr. 2012). “The Extreme Small Scales: Do Satellite Galaxies Trace Dark Matter?” In: 749.1, 83, p. 83. doi: 10.1088/0004-637X/749/1/83. arXiv: 1108.1195 [astro-ph.CO].

Weaver, J. R. et al. (Oct. 2021). “COSMOS2020: A panchromatic view of the Universe to $z∼ 10$*fromtwocomplementarycatalogs*”. In: *arXiv:2110.13923 [astro-ph]*. arXiv: 2110.13923. url: http://arxiv.org/abs/2110.13923 (visited on 28/10/2021).

Wechsler, Risa H. and Jeremy L. Tinker (Sept. 2018). “The Connection between Galaxies and their Dark Matter Halos”. In: 56.1. arXiv: 1804.03097, pp. 435–487. issn: 0066-4146, 1545-4282. doi: 10.1146/annurev-astro-081817-051756. url: http://arxiv.org/abs/1804.03097 (visited on 08/07/2021).

Werner, M. W. et al. (Sept. 2004). “The Spitzer Space Telescope Mission”. In: 154, pp. 1–9. issn: 0067-0049. doi: 10.1086/422992. url: http://adsabs.harvard.edu/abs/2004\apjs..154....1W (visited on 19/03/2020).

White, S. D. M., M. Davis and C. S. Frenk (July 1984). “The size of clusters in a neutrino-dominated universe”. In: 209, 27P–31P. doi: 10.1093/mnras/209.1.27P.

White, S. D. M. and M. J. Rees (May 1978). “Core condensation in heavy halos: a two-stage theory for galaxy formation and clustering.” In: 183, pp. 341–358. doi: 10.1093/mnras/183.3.341.

Williams, Robert E. et al. (Oct. 1996). “The Hubble Deep Field: Observations, Data Reduction, and Galaxy Photometry”. In: 112, p. 1335. DOI: 10.1086/118105. arXiv: astro-ph/9607174 [astro-ph].

Wuyts, Stijn et al. (Dec. 2011). “Galaxy Structure and Mode of Star Formation in the SFR-Mass Plane from z ∼2.5 to z ∼0.1”. In: 742.2, 96, p. 96. DOI: 10.1088/0004-637X/742/2/96. arXiv: 1107.0317 [astro-ph.CO].

Xu, Haojie et al. (Aug. 2016). “On the clustering of faint red galaxies”. In: 460.4, pp. 3647–3659. DOI: 10.1093/mnras/stw1259. arXiv: 1510.00393 [astro-ph.GA].

Yang, Xiaohu, H. J. Mo and Frank C. van den Bosch (Apr. 2009). “Galaxy Groups in the SDSS DR4. III. The Luminosity and Stellar Mass Functions”. In: 695.2, pp. 900–916. DOI: 10.1088/0004-637X/695/2/900. arXiv: 0808.0539 [astro-ph].

Zamojski, M. A. et al. (Sept. 2007). “Deep GALEX Imaging of the COSMOS HST Field: A First Look at the Morphology of z ~ 0.7 Star-forming Galaxies”. In: 172, pp. 468–493. ISSN: 0067-0049. DOI: 10.1086/516593. URL: https://ui.adsabs.harvard.edu/abs/2007ApJS..172..468Z (visited on 21/07/2021).

Zehavi, Idit et al. (May 2002). “Galaxy Clustering in Early Sloan Digital Sky Survey Redshift Data”. In: 571.1, pp. 172–190. DOI: 10.1086/339893. arXiv: astro-ph/0106476 [astro-ph].

Zehavi, Idit et al. (May 2002). “Galaxy Clustering in Early Sloan Digital Sky Survey Redshift Data”. In: 571, pp. 172–190. ISSN: 0004-637X. DOI: 10.1086/339893. URL: https://ui.adsabs.harvard.edu/abs/2002ApJ...571..172Z (visited on 08/10/2021).

Zehavi, Idit et al. (June 2004). “On Departures from a Power Law in the Galaxy Correlation Function”. In: 608.1, pp. 16–24. DOI: 10.1086/386535. arXiv: astro-ph/0301280 [astro-ph].

Zehavi, Idit et al. (July 2011). “Galaxy Clustering in the Completed SDSS Redshift Survey: The Dependence on Color and Luminosity”. In: 736.1. arXiv: 1005.2413, p. 59. ISSN: 0004-637X, 1538-4357. DOI: 10.1088/0004-637X/736/1/59. URL: http://arxiv.org/abs/1005.2413 (visited on 21/04/2021).

Zel’dovich, Ya. B. (Mar. 1970). “Gravitational instability: An approximate theory for large density perturbations.” In: 5, pp. 84–89.

Zheng, Zheng, Alison L. Coil and Idit Zehavi (Oct. 2007). “Galaxy Evolution from Halo Occupation Distribution Modeling of DEEP2 and SDSS Galaxy Clustering”. In: 667.2. arXiv: astro-ph/0703457, pp. 760–779. ISSN: 0004-637X, 1538-4357. DOI: 10.1086/521074. URL: http://arxiv.org/abs/astro-ph/0703457 (visited on 21/04/2021).

Zheng, Zheng et al. (Nov. 2005). “Theoretical Models of the Halo Occupation Distribution: Separating Central and Satellite Galaxies”. In: 633.2. arXiv: astro-ph/0408564, pp. 791–809. ISSN: 0004-637X, 1538-4357. DOI: 10.1086/466510. URL: http://arxiv.org/abs/astro-ph/0408564 (visited on 21/04/2021).