Alma Mater Studiorum · Università di Bologna

DIPARTIMENTO DI FISICA E ASTRONOMIA
Corso di Laurea Magistrale in Astrofisica e Cosmologia

New cosmological constraints from a weak lensing analysis of the AMICO galaxy cluster catalogue in the Kilo-Degree Survey

Presentata da:
Lorenzo Gigante

Relatore:
Ch.mo Prof. Lauro Moscardini

Co-relatori:
Dott. Federico Marulli
Dott. Fabio Bellagamba

Appello IV
Anno Accademico 2017-2018

"Mò ca ddienti scienziatu tocca nchiani cu mmie sullu palcu de li comizi e tocca parli pé mie, ca ieu ormai sù bbecchiu e nu me fitu mutu cu bbausu la usce. Tocca cu li sputtani tutti sti politici sciacalli. De tutti, unu pe unu, amu mmusciare gli altarini ca tenenu scusi e te li dicu ieu, tutti. A tanti anni ca l'aggiu vvardati e schifati ne sacciu tante de cose ieu. Ma ni l'ha dire tie alli cristiani le cose comu stannu e comu l'amu cangiare, ca a mie nu me sentenu ca su bbecchiu e antipaticu. Ma tie sì ggiovane e scienziatu, e se parli tie cieddhri te pò dicere nienzi!"

Nonno Nino

Abstract

The spatial and dynamical properties of the large-scale structures of the Universe provide one of the most powerful probes to constrain the cosmological framework. In particular, the gravitational lensing produced by galaxy clusters can be efficiently exploited to constrain the main cosmological parameters. A crucial ingredient to use clusters as cosmological probes is the accurate calibration of the relation between the cluster observable properties and their masses. The *shear-cluster correlation* is independent of the nature and state of the mass distorting the background signal. On scales far beyond the cluster virial radius, the lensing signal is dominated by the correlated distribution of matter in haloes around clusters, described by the so-called *2-halo* term. The matter in correlated haloes provides sensible cosmological information, useful to test cosmological models.

In this thesis work, we present new cosmological constraints inferred from the analysis of weak lensing shear data, extracted from the KiDS-DR3 photometric survey. The data used for shear measurements are those in the r-band, being the ones with better seeing properties and with highest source density. The galaxy catalogue provides shear measurements for about 15 million galaxies over a total effective area of 360 deg^2. The galaxy photometric redshifts, used to weigh the lensing data, have been extracted from the 4 bands u, g, r, i. Reliable posterior probability distributions for the photometric redshifts are used in our analysis. We perform a conservative background source selection, applying both photometric redshift selections and colour selections. The galaxy cluster sample analysed in this work has been extracted using the *Adaptive Matched Identifier of Clustered Objects* (AMICO) algorithm (Bellagamba et al., 2018). This is a state-of-the-art algorithm for the detection and characterisation of galaxy clusters in photometric surveys, where data are affected by a noisy background. Indeed, this algorithm has been recently selected as the primary cluster detection algorithm for the European Space Agency (ESA) *Euclid* mission. The cluster catalogue has been constructed by selecting only the objects with signal-to-noise ratio (S/N) larger than 3.5, in the redshift range $0.1 < z < 0.6$. The final sample consists of 6972 galaxy clusters in total (Maturi et al., 2018). For the statistical lensing analysis, we divide the catalogue in 3 redshift bins, and in 14 bins of differential surface density profiles, stacking the clusters according to their signal amplitude A, that we use as a reliable mass proxy. A is a measure of the cluster galaxy abundance in units of the cluster model, which is built starting from the known data. The covariance matrix associated with the extraction of the shear signal and its relative stacking are computed with a bootstrap procedure with replacement: the source catalogue is re-sampled 10000 times.

A full Bayesian analysis is performed to model the shear profiles. The adopted likelihood function considers both the small-scale 1-halo term, used primarily to constrain the cluster masses, and the 2-halo term, that is used to constrain the matter density contrast Ω_M. Specifically, we model the cluster main halo and the associated baryonic component, that is the 1-halo term, through a *smoothly truncated* Navarro-Frenk-White profile model (Baltz et al., 2009). We also consider an additional component that parameterises the effect of haloes whose observed centre has a non-negligible scatter with respect to the centre assumed by the stacking procedure. This component is necessary not to introduce systematics that would lower the assessed cluster masses. The miscentring contribution is relevant mainly at the smallest scales ($\sim 0.5\,\mathrm{Mpc}/h$). To model the shear profiles up to large scales, we include the 2-halo term built upon the Eisenstein and Hu (1999) linear matter power spectrum and the Tinker et al. (2010) linear halo bias. The 2-halo term starts to dominate the signal on scales larger than about $10\,\mathrm{Mpc}/h$. The statistical analysis consists of two steps. Firstly, we collect the stacked shear signal produced by the 14 cluster bins of the sample, and estimate cluster masses by modelling the small-scale 1-halo signal in the radial range $0.1 < R < 3.16\,\mathrm{Mpc}/h$. Secondly, we re-extract the signal for the same cluster sample, but on a much larger scale range, $0.1 < R < 35\,\mathrm{Mpc}/h$, obtaining cosmological information from the 2-halo term.

We find that the adopted modelling is successful to assess both the cluster masses and the matter density parameter, when fitting shear profiles up to the largest scales probed. The derived masses of the cluster samples are in good agreement with those estimated by Bellagamba et al. (2019). Our work confirms the reliability of AMICO and of the KiDS-DR3 data. Moreover, our results provide a strong observational evidence of the 2-halo signal in the stacked gravitational lensing of galaxy clusters, and demonstrate the reliability of this probe for constraining cosmological parameters. The main result of this thesis work is a new, robust constraint on Ω_M, assuming a flat ΛCDM cosmology. Specifically, we find $\Omega_M = 0.285 \pm 0.023$, estimated from the full posterior

probability distribution. This value is consistent and competitive with that estimated by WMAP9, $\Omega_{M,WMAP9} = 0.273 \pm 0.049$, but slightly in tension with Planck18 constraints, $\Omega_{M,Planck18} = 0.317 \pm 0.008$.

Sommario

Le proprietà spaziali e dinamiche delle strutture su grande scala dell'Universo forniscono uno dei più potenti mezzi per testare i modelli cosmologici. In particolare, il segnale di lensing gravitazionale prodotto dagli ammassi di galassie può essere efficientemente sfruttato per vincolare i principali parametri cosmologici. Un ingrediente cruciale nell'utilizzo degli ammassi come sonde cosmologiche è un'accurata calibrazione della relazione fra le proprietà osservabili degli ammassi e le loro masse. La *correlazione shear-ammasso* è indipendente dalla natura e dallo stato della massa che distorce il segnale di background. Su scale molto più grandi del raggio viriale degli ammassi, il segnale di lensing è generato dalla distribuzione degli aloni di materia oscura ospitanti gli ammassi, descritto dal cosiddetto termine di *2-halo*, che fornisce importanti informazioni utili per testare i modelli cosmologici.

In questo lavoro di tesi, presentiamo nuovi vincoli cosmologici provenienti da un'analisi di dati di shear in regime di lensing debole, estratti dalla survey fotometrica KiDS-DR3. I dati usati per le misure di shear provengono dalla banda ottica nel rosso (r), essendo quella con migliori proprietà di *seeing* e con la più alta densità di sorgenti. Il catalogo di galassie contiene misure di shear per circa 15 milioni di galassie, in un'area effettiva totale di $360\,\mathrm{deg}^2$. I redshfits fotometrici delle galassie, usati per pesare i dati di lensing, sono stati estratti dalle 4 bande *u, g, r, i*. Nella nostra analisi abbiamo fatto uso di distribuzioni di probabilità a posteriori dei redshift fotometrici. Abbiamo eseguito una selezione conservativa delle galassie di background, applicando sia una selezione in redshift fotometrico che in colore. Il campione di ammassi di galassie analizzato in questo lavoro è stato estratto da KiDS utilizzando l'algoritmo *Adaptive Matched Identifier of Clustered Objects* (AMICO) (Bellagamba et al., 2018). Quest'algoritmo rappresenta lo stato dell'arte per la detezione e caratterizzazione di ammassi di galassie in surveys fotometriche i cui dati sono affetti da un background rumoroso. Infatti, è stato recentemente selezionato come principale algoritmo di detezione di ammassi per la missione *Euclid*, dell' European Space Agency (ESA). Il catalogo di ammassi è stato costruito selezionando solo oggetti il cui rapporto di segnale rumore è superiore a 3.5, e nel range di redshift $0.1 < z < 0.6$. Il nostro catalogo finale è composto da un totale di 6972 ammassi di galassie (Maturi et al., 2018). Per quanto riguarda l'analisi statistica del lensing, abbiamo diviso il catalogo in 3 intervalli di redshift e in 14 intervalli di profili di densità superficiale differenziale, sovrapponendo gli ammassi in accordo alla loro ampiezza di segnale A, che rappresenta una attendibile *proxy* della massa. A è una misura dell'abbondanza di galassie nell'ammasso in unità del modello di ammassi adottato. La matrice di covarianza, associata con l'estrazione del segnale di shear e relativa sovrapposizione, viene calcolata tramite una procedura di *bootstrap*: il catalogo di sorgenti è stato ri-campionato 10000 volte.

Per modellare i profili di shear è stata eseguita un'analisi Bayesiana. La funzione adottata per la *likelihood* tiene conto sia del termine di 1-halo, dominante a piccole scale e usato principalmente per stimare la massa degli ammassi, sia di quello di 2-halo, usato per vincolare il contrasto di densità di materia cosmologico Ω_M. Nello specifico, modelliamo il termine 1-halo del cluster, e la componente barionica associata, tramite un profilo NFW troncato (Baltz et al., 2009). Abbiamo anche considerato una componente aggiuntiva, che parametrizza l'effetto degli aloni i cui centri osservati hanno uno scatter non trascurabile rispetto al centro assunto dalla procedura di sovrapposizione. Tale componente è necessaria per non introdurre sistematici che sottostimerebbero le masse misurate dei cluster. Il contributo del decentramento diventa trascurabile per scale più grandi di circa ~ 0.5 Mpc/h. Per modellare i profili di shear fino alle grandi scale includiamo il temine di 2-halo, costruito assumendo uno spettro di potenza lineare della materia di Eisenstein and Hu (1999) e un bias lineare di Tinker et al. (2010). Il termine di 2-halo inizia a dominare il segnale di shear su scale maggiori di circa ~ 10 Mpc/h. L'analisi statistica è stata suddivisa in due parti. Per prima cosa abbiamo misurato il segnale di shear, prodotto dai 14 profili sovrapposti del campione, e abbiamo stimato le masse degli ammassi modellando il segnale prodotto dal 1-halo a piccole scale ($0.1 < R < 3.16$ Mpc/h). Successivamente, abbiamo misurato nuovamente il segnale prodotto dallo stesso campione di ammassi, ma su scale molto più grandi ($0.1 < R < 35$ Mpc/h), estraendo informazioni cosmologiche dal termine di 2-halo.

Il modello adottato è risultato efficace per derivare sia le masse degli ammassi che il parametro di densità di materia, quando modelliamo i profili di shear fino alle più grandi scale da noi sondate. Le masse derivate per gli ammassi del campione sono in ottimo accordo con quelle stimate da Bellagamba et al. (2019). Il nostro lavoro conferma l'attendibilità di AMICO e dei dati di KIDS-DR3.

Inoltre, i risultati che abbiamo ottenuto forniscono una forte evidenza osservativa del segnale del 2-halo nel lensing gravitazionale prodotto da ammassi di galassie, dimostrando anche l'attendibilità del nostro metodo statistico nel vincolare i parametri del modello cosmologico. Il principale risultato di questo lavoro è un nuovo e robusto vincolo su Ω_M, assumendo un modello di Universo ΛCDM piatto. Nello specifico, abbiamo ottenuto $\Omega_M = 0.285 \pm 0.023$, stimato dalla distribuzione di probabilità a posteriori di tutti i 14 profili di shear sovrapposti e marginalizzando sulle masse. Tale valore è consistente e competitivo con quello misurato da WMAP9, $\Omega_{M,WMAP9} = 0.273 \pm 0.049$, ma leggermente in tensione con i vincoli di Planck18, $\Omega_{M,Planck18} = 0.317 \pm 0.008$.

Contents

Introduction

Cosmology is the branch of Astrophysics that studies the cosmos in its entirety, aiming at understanding in which *type* of Universe we live, how the Universe was born and what its ultimate destiny will be. Our Universe is not static. Indeed, one of the first pillar of Cosmology was the observation of the *expansion* of the Universe by E. Hubble (Hubble, 1929). The subsequent discovery of *cosmic acceleration* (Riess et al., 1998), that can be attributed to an unknown form of smooth tension, the so-called dark energy, represented the major breakthrough in modern Cosmology. The *cosmological constant* Λ provides the simplest possible explanation for dark energy (Einstein, 1915; Carroll et al., 1992). The cosmological constant can be identified as the vacuum energy, that is a fixed amount of energy attached to every region of space, constant both in space and time. The standard Λ-Cold dark matter (ΛCDM) cosmological model provides the best and simplest parameterization of the so-called Concordance Model. Together with Λ, which contributes to about 68% of the total mass-energy in the Universe, the standard model assumes that a hypothetical, invisible form of matter constitutes about $\sim 27\%$ of the mass-energy density (Aghanim et al., 2018). The current consensus view is that this dark matter component is cold (CDM) for most part, being composed of particles moving at non-relativistic speeds. Only $\sim 4.9\%$ of the Universe is made up of ordinary matter (baryons), and even less by radiation ($10^{-5}\%$). From the Big Bang until the present day, the relative contributions of the four main components of the Universe have undergone drastic variations. Our Universe is currently dominated by the dark energy but, in the past, its influence was negligible. Note that the ΛCDM model says nothing about the nature and physical origin of CDM and dark energy. Nevertheless, it is currently the best cosmological model available to describe the large-scale structure of the Universe and its evolution. In fact, it is dubbed the Standard Cosmological Model (SCM). Moreover, the observations shows that the space of our Universe is approximately flat on large scales, and the SCM is often also-called the *flat ΛCDM model*.

In such a scenario, clusters of galaxies play an important role both for the study of the formation and evolution of the galaxies in their environment, and for the determination of cosmological parameters (Allen et al., 2011). Galaxy clusters are hosted by the most massive CDM haloes, and their evolution is dominated by the CDM. The shape and evolution of the cluster mass function (Press and Schechter, 1974) and the mass-to-light ratio (Schechter, 1976) allow to constrain the mass density parameter, Ω_m, and the amplitude of the matter power spectrum, σ_8 (Oukbir and Blanchard, 1992).

There are different methods to detect galaxy clusters, for instance through the X-ray emission of the intergalactic medium (Merloni et al., 2012; Clerc et al., 2014), from the optical emission of their galaxy population, or from the Sunyaev-Zel'dovich (SZ) effect (Sunyaev and Zeldovich, 1970, 1972). In fact, the cluster identification and characterisation are improved when using multi-band observations. By exploiting the optical and near infrared (NIR) bands, in particular, it is possible to identify the clusters by searching for the cluster-scale overdensities in the galaxy distribution. This technique enables us to cover large and deep areas of the sky through photometric surveys, that is required to collect a sufficiently large number of clusters to robustly infer their mass function. Ongoing and planned photometric surveys will allow us to increase quite a lot the present census of galaxy clusters, expanding the detection to lower masses and higher redshifts.

It is crucial to derive reliable cluster mass estimates to infer cosmological constraints. In particular, the measured number of galaxy cluster members, accessible in photometric or spectroscopic observations, has to be linked to the total mass of the cluster, made primarily by CDM. The difficulty is that this relation is hidden in complex, and not yet well understood, astrophysical processes. To estimate cluster masses we can rely on some cluster *scaling relations* (see, e.g. Giodini et al., 2013), directly derived from X-ray or SZ effect observations. However, such relations depend on empirical calibrations, that may undermine the reliability of the analysis. Other mass

proxies have been suggested in literature. The most used are the ones based on the number of (red) galaxies inside a given cluster radius (Andreon and Hurn, 2010; Rykoff et al., 2012), on photometric stellar mass estimates (Pereira et al., 2018), or on cluster luminosities (Mulroy et al., 2014). However, the most reliable method to estimate cluster masses is currently provided by the *gravitational lensing*. Gravitational lensing is a phenomenon that consists of the deflection of the radiation emitted by a source due to the presence of matter interposed between the source itself and the observer, as predicted by the General Relativity. The weak lensing effect, in particular, consists of slight distortions of the source shapes (called shear) and the intensification of their apparent brightness (magnification), generated by the convergence of light rays. In principle, we may exploit both effects to estimate the projected masses, because the amplitude of the distortion ensures a direct measure of the lens gravitational tidal field. The weak lensing comes into play when the source is at high angular distances from the geometric centre of the lens. Hence it is also very useful to probe the lensing cluster profiles up to large radii. Another helpful property of lensing is that the gravitational light deflection is independent of the nature and state of deflecting matter, hence the lensing signal depends only on the lens total mass. All of these properties made the weak lensing really powerful for characterising galaxies clusters (Bartelmann and Schneider, 2001).

A small number of background galaxies may limit the study of single clusters to virial masses of the order of a few times $10^{14} M_{\odot}$ or larger, for typical current optical surveys. Indeed, it is practice use to study lower mass systems by averaging, or *stacking* their signals: the resultant signal is that of the mass-weighted ensemble. Precisely, the *stacked gravitational lensing* represents the cross-correlation between foreground deflector positions and background galaxy shears. If we have a shallow but broad survey, to compensate for the low number density, stacking the signal around a large number of lenses is an efficient method. For example, this approach was applied with great success to the Sloan Digital Sky Survey (SDSS) data to measure the ensemble averaged lensing signal around groups and clusters (Mandelbaum et al., 2006b; Sheldon et al., 2009). By splitting the sample of clusters into subsets based on an observable property, such as amplitude or optical richness, it has been possible to determine the scaling relations between the observable and the mean mass of the sample (see, e.g., Bellagamba et al., 2019).

Furthermore, gravitational lensing is often one of the primary goals of current and upcoming photometric surveys, thus enabling the calibration of optical observables of photometrically-selected clusters within the same data-set. Therefore, altough large samples of clusters were also drawn from X-ray observations or through the detection of the SZ effect on the CMB, photometric observations remain the most promising source of discovery of new clusters of galaxies. This applies especially to the wide ongoing and future photometric surveys, such as the Dark Energy Survey (DES The Dark Energy Survey Collaboration, 2005) or the photometric Kilo Degree Survey (KiDS, Radovich et al., 2017; de Jong et al., 2017), that we exploit in this thesis work. Precisely, we use the Third Data Release of this survey (KiDS-DR3 de Jong et al., 2017), that contains reduced image data, source lists and a multi-band catalogue, and it is made by 440 survey tiles, with 48736591 sources in total. Its current total released data-set covers approximately $447 \mathrm{deg}^2$ (see Figure 5.1). This release also includes photometric redshifts with relative probability distribution functions (Kuijken et al., 2015; de Jong et al., 2017), a global improved photometric calibration, weak lensing shear catalogues (Kuijken et al., 2015; Hildebrandt et al., 2017) and lensing-optimised image data. Source detection, positions and shape parameters used for weak lensing measurements are all derived from the r-band images, since these typically have the best image quality and provide the most reliable measurements, while magnitudes are measured in all filters using forced photometry (Hildebrandt et al., 2017). The principal scientific goal of KiDS is to exploit the weak lensing and photometric redshift measurements to map the matter large-scale distribution in the Universe. This survey produces images with a high quality and depth and a very precise calibration.

Appropriate algorithms have to be used to identify galaxy clusters in photometric surveys. In this thesis work we exploit the *Adaptive Matched Identifier of Clustered Objects* (AMICO) algorithm, presented and tested in Bellagamba et al. (2018). It was used to extract the first cluster sample from KiDS-DR3 (Maturi et al., 2018). In this work we use the same cluster data-set for our analysis. AMICO is an algorithm for the detection of galaxy clusters in photometric surveys whose data-set is affected by a noisy background. It is build upon a *linear optimal matched filter method* (Maturi et al., 2005, 2007; Viola et al., 2010; Bellagamba et al., 2011), which maximises the S/N and takes advantage of the statistical properties of the field and cluster galaxies.

To build the shear profile of a given lens, through which we can reconstruct the stacked differ-

ential surface density profile, $\Delta\Sigma(R)$, we select the background galaxies following a conservative approach (Bellagamba et al., 2019). This approach aims at eliminating foreground and cluster galaxies from the KiDS catalogue by applying cuts based on photometric redshift distributions and colours. For every single lens we take into account all the background galaxies selected according to the method just mentioned, and then we calculate the shear signal along the tangential direction with respect to the centre of the considered cluster. We determine the weighted ellipticity of the background sources in radial bins, that is the tangential and cross components of the shear, though for the subsequent analysis we use only the tangential component. For most of the clusters in the sample, the S/N in the shear data is too low to constrain the density profile and, consequently, to measure a reliable mass. So we are obliged to measure the mean mass for ensembles of objects chosen according to their observables and redshift. In short words, we exploit the *shear-cluster correlation* (or stacked lensing), i.e the cross-correlation between the cluster (lens) density distribution and the shapes of the background source galaxies.

When collecting the stacked shear signal from KiDS-DR3, we split the cluster sample in three consecutive redshift bins: $0.1 \leq z < 0.3$, $0.3 \leq z < 0.45$ and $0.45 \leq z < 0.6$. The considered sample is composed of 6972 AMICO-KiDS clusters. We further split each redshift bin in a set of increasing amplitude bins[1], for a total of 14 stacked clustered objects, that covers a range of masses going from $\sim 10^{13}$ to 4.5×10^{14} $M_{\odot}/h$, in each of the three redshift bins. This cluster binning is the same one adopted in Bellagamba et al. (2019). The goal is to reconstruct the stacked differential (projected) surface density profiles of our cluster sample in two different radial ranges (from the assumed cluster centre), with the same redshift-amplitude binning. Precisely, we firstly collect the shear stacked signal from 0.1 to 3.16 $M_{\odot}/h$, in 14 radial annuli equally distanced in logarithmic space, and secondly from 0.1 to 35 $M_{\odot}/h$, in 30 radial annuli equally distanced in logarithmic space. The first spatial range covers the region occupied by the main halo, that is roughly up to the cluster virial radius, and with these data we perform an analysis aiming at inferring robust cluster mass estimates (see Table 6.1). When analysing shear data extracted up to 35 $M_{\odot}/h$ from the cluster centre, we are on scales where the signal is produced by the *2-halo term*. Precisely, its contribute begins to be dominant on scales larger than ~ 10 $M_{\odot}/h$ (see e.g Fig. 6.8). At such scales the shear signal is produced by correlated matter distribution around the location of the galaxy cluster and it manifests itself as a 2-halo contribute in the mass density profiles. This 2-halo analysis is performed to infer estimates of Ω_M.

We fit the $\Delta\Sigma(R)$ profiles with a model set up as the sum of the 1-halo and the 2-halo terms, $\Delta\Sigma(R) = \Delta\Sigma(R)_{1h} + \Delta\Sigma(R)_{2h}$ (Fig. 5.6 shows an example of this model). The 1-halo contribute refers to the virialized main cluster halo, described by a *smoothly-truncated NFW density profile* (Baltz et al., 2009) which is able to model cluster profiles far beyond the virial radius profile and produces less biased mass estimates, with respect to the original NFW. Inside the $\Delta\Sigma_{1h}$ model, we considered also a $\Delta\Sigma(R)_{off}$ term that takes into account the contribute of individual clusters whose centre does not coincide with the cluster centre assumed by the stacking procedure. Without this term, we would get biased low mass estimates. Furthermore, when assuming the truncated NFW with a given cosmology, we can infer the cluster mass as M_{200}, that represents the profile normalisation (see e.g Figure 6.3). Finally, the 2-halo term is characterised by the model given by (Oguri and Takada, 2011), that describes the cumulative effects of the large-scale structures in which galaxy clusters are located.

The analyses are performed following a Bayesian approach, through the use of Monte Carlo Markov Chains (MCMC) that allow us to infer the model parameters with a good accuracy (see Figures 6.11, 6.12, 6.13 and 6.18, that show the posterior distributions we obtained through the MCMC routine for the analysis results on the second shear profiles collection). The statistical numerical code (shown in appendix D) has been built entirely with the CosmoBolognaLib C++ libraries (Marulli et al., 2016), that provide all the functions used for the estimation of the various cosmological components of our fitting model.

This thesis work is organised as follows:

- Chapter 1 provides an overview of the cosmological context on which we build this thesis work, introducing the theoretical basis of the Standard Cosmological Model.
- Chapter 2 deals with the theory of structure formation and evolution, discussing the perturbative approach, both in the linear and non-linear regimes.

[1]The amplitude A is a measure of the cluster galaxy abundance, through an optimal filter, in units of the cluster model. We used it as a reliable mass proxy, as demonstrated by our analysis.

- Chapter 3 illustrates the main cluster properties, together with the methods to identify them in the observations. We focus on the cluster detection method in optical photometric surveys, discussing its applications and advantages. In particular, in Section 3.6 we describe the AMICO algorithm that allowed us to collect the clusters we use for the analysis.
- Chapter 4 deals with the basics of gravitational lensing, focusing on the properties of the weak lensing by clusters and how it can be exploited to derive cluster properties.
- Chapter 5 introduces the survey dataset on which this thesis work is based, namely KiDS-DR3, and how it was handled, through an appropriate shear data extraction. In particular, Section 5.4 introduces the model used to describe the cluster shear profiles extracted from the survey.
- Chapter 6 illustrates in detail the method used to constrain model and cosmological parameters from the survey dataset, and the results of the data analysis. Furthermore, we discuss the reliability of the adopted method, and compare our results with the most recent cosmological estimations. We also briefly deal with the possible systematic uncertainties affecting the cosmological parameters inference.
- Chapter 7 summarises the main results of this project and discuss possible future perspectives.

In all the analyses presented in this work, we assume a flat ΛCDM model with baryon density $\Omega_b = 0.05$ and scaled Hubble parameter $h = 0.7$. The matter density parameter is left free except in the first analysis stage where, to infer the cluster masses modelling the 1-halo term, it is fixed to $\Omega_M = 0.3$. When inferring Ω_M from the 2-halo modelling, we assume anyway a flat ΛCDM, hence $\Omega_\Lambda = 1 - \Omega_M$ is a derived parameter.

Chapter 1

An overview of the cosmological context

In this Chapter we will present the cosmological framework on which this thesis work is set up. We will supply the main notions and equations that are the foundation of the theoretical models, aiming at describing the growth and the dynamical evolution of the Universe and its content; the so-called Standard Cosmological Model.

1.1 Cosmological principles

Cosmology studies the formation and evolution of the Universe in its entirety, from the Big Bang (BB) to the present. Observational data pushes towards a model of the Universe that is difficult to explain and with an estimated composition (see Aghanim et al., 2018) given by:

- 4.9% of baryons
- 26.8% of dark matter
- 68.3% of dark energy
- $\sim 10^{-5}$% of radiation.

Nowadays the BB Model is widely accepted and it is founded on three pillars:

- inflation,
- anisotropies of the Cosmic Microwave Background (CMB),
- empirical confirmation of the accelerated expansion of Universe.

On the other hand, the BB Model is affected by three fundamental issues actually unresolved, not even through inflation: the coincidence problem (Velten et al., 2014), the origin and nature of dark matter and dark energy and the inevitability of the assumption of an origin of the space-time (i.e the BB).

The history of Universe is a *story of cooling*: starting from the BB, passing trough inflation and coming to the gravitational collapse of matter (baryonic and dark) and to the consequent formation of the first structures. Another main ingredient of the model is the so-called *Cosmological Principle (CP)*, that states that the Universe is *homogeneous and isotropic* on sufficiently large scales [1]. A first confirmation of this is the homogeneous distribution, on large scale, of galaxies together with the isotropy of CMB. In other words, homogeneity implies that all observers, anywhere in the Universe, observe the same properties and matter distribution, while isotropy involves that there are not preferred directions.

[1] much larger than 10 Mpc, that is much larger than the size of the largest galaxy clusters known to us

1.2 Basics of General Relativity

On large scale, the main force acting on the matter across the Universe is gravity. It dominates the dynamics and the evolution of the large cosmic structures, that are expected to be formed starting from tiny quantum fluctuations. The best description of the spacetime and of this force that models it, is given by the theory of General Relativity, proposed by Albert Einstein (Einstein, 1915).

Firstly, we have to set a metric for our Universe. In this context, the spacetime geometry is described by the *metric tensor* $g_{\alpha\beta}$, that measures the invariant distance between two events in spacetime, (t, x_1, x_2, x_3) and $(t', x_1', x_2', x_3') = (t + \delta t, x_1 + \delta x_1, x_2 + \delta x_2, x_3 + \delta x_3)$. This interval in spacetime is defined by:

$$\mathrm{d}s^2 = g_{\alpha\beta}\mathrm{d}x^\alpha \mathrm{d}x^\beta, \tag{1.2.1}$$

which in its explicit form reads:

$$\mathrm{d}s^2 = g_{00}\mathrm{d}t^2 + 2g_{0i}\mathrm{d}x^i\mathrm{d}t + g_{ij}\mathrm{d}x^i\mathrm{d}x^j, \tag{1.2.2}$$

where $g_{00}\mathrm{d}t^2$ is the time component, $2g_{0i}\mathrm{d}x^i\mathrm{d}t$ is the mixed component and $g_{ij}\mathrm{d}x^i\mathrm{d}x^j$ the spatial one. While $g_{\alpha\beta}$ describes the geometry of spacetime, $\mathrm{d}s^2$ tells us if the type of interval is: time ($\mathrm{d}s^2 > 0$), space ($\mathrm{d}s^2 < 0$) or light ($\mathrm{d}s^2 = 0$).

Now, we define the *Riemann tensor* as follows:

$$R^\mu_{\alpha\beta\gamma} \equiv \frac{\mathrm{d}\Gamma^\mu_{\alpha\gamma}}{\mathrm{d}x^\beta} - \frac{\mathrm{d}\Gamma^\mu_{\alpha\beta}}{\mathrm{d}x^\gamma} + \Gamma^\mu_{\sigma\beta}\Gamma^\sigma_{\gamma\alpha} + \Gamma^\mu_{\sigma\gamma}\Gamma^\sigma_{\beta\alpha}, \tag{1.2.3}$$

where Γ are the *affine connection* and $R^\mu_{\alpha\beta\gamma}$ is a fourth-order tensor that we can contract to the *Ricci tensor* $R^{\mu\nu}$, or further to the *curvature scalar* R (the track of Ricci tensor), respectively:

$$R_{\alpha\beta} \equiv R^\mu_{\alpha\beta\mu}, \qquad R \equiv R^\mu_\mu = g^{\mu\nu}R_{\mu\nu}. \tag{1.2.4}$$

To describe the energy-matter content of the Universe, we have to introduce the *energy-momentum tensor* $T_{\mu\nu}$, which we can connect with the spacetime metric into the *Einstein's Field Equations*:

$$R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R = \frac{8\pi G}{c^4}T_{\mu\nu}, \tag{1.2.5}$$

where G and c that are the gravitational constant and the speed of light, respectively. Considering the left-hand side of this equation we can define the *Einstein Tensor* $G_{\mu\nu}$ as:

$$G_{\mu\nu} \equiv R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R, \tag{1.2.6}$$

that is a tensor with zero covariant divergence, according to the kind of $T_{\mu\nu}$ we want to obtain, by virtue of the conservation laws of the spacetime. Combining the equations (1.2.5) and (1.2.6) we get the final *Einstein gravitational field equations* :

$$G_{\mu\nu} = \frac{8\pi G}{c^4}T_{\mu\nu}. \tag{1.2.7}$$

Note that the assumption of the CP requires an energy-momentum tensor referred to a perfect fluid (the mean free path of the fluid particles has to be lower than the physical scales of interaction). Furthermore the empirical proofs so far known seem to give reason to the original choice of Einstein to couple matter to the gravitational field through $T_{\mu\nu}$. The difficulty in solving Eq. (1.2.7) is due to the fact that they are non-linear equations, and therefore the principle of overlapping effects is not valid.

1.2.1 The Friedmann-Lemaître-Robertson-Walker metric

To be able to consider a set of homogeneous and isotropic models of the Universe, we must assume the CP, and find solutions to the Einstein's Field Equations. In this assumption, we can define a *proper time* t such that the spatial metric $\mathrm{d}l^2 = g_{ij}\mathrm{d}x^i\mathrm{d}x^j$ is identical in every space and time. Moreover, the isotropy implies that the mixed term g_{0i} in Eq. (1.2.2) is null. Thus we obtain the following generic form of the metric:

$$\mathrm{d}s^2 = c^2\mathrm{d}t^2 - g_{ij}\mathrm{d}x^i\mathrm{d}x^j = (c\mathrm{d}t)^2 - \mathrm{d}l^2, \tag{1.2.8}$$

that describes the geometrical properties of the spacetime. Now, we consider the time coordinate (the term $(cdt)^2$) on the right-hand side of the equation (1.2.8) and assume spherical polar coordinates where r, θ and φ are dimensionless co-moving coordinates and t is the proper time. According to the assumption of isotropy, we can derive the most general form of the equation (1.2.8) that verifies the CP as follows:

$$\mathrm{d}s^2 = (c\mathrm{d}t)^2 - a^2(t)\Big[\frac{\mathrm{d}r^2}{1-kr^2} + r^2(\sin^2\theta\mathrm{d}\varphi + \mathrm{d}\theta^2)\Big], \tag{1.2.9}$$

that defines the *Friedmann-Lemaître-Robertson-Walker metric (or FLRW metric)*, which embodies the CP.

We can note now the appearance of two new elements. The function $a(t)$ is the *cosmic scale factor* (also known as *expansion parameter*), having the dimensions of a length, and the *curvature parameter* k. The value of k and $a(t)$ can be derived by the Einstein's Field Equations, for any given energy-momentum tensor.

Let us consider a free massive particle at rest located, at a moment in time, at the origin of the co-moving system: the absence of preferred directions ensures that no velocity variation can be induced on this particle by gravity so, thanks to the homogeneity, the world lines $x^i = cost.$ coincide with the so-called *geodetics*.

1.2.2 The spatial curvature

Let us consider the simplified case of an homogeneous and isotropic two-dimensional Universe; this may have three geometric shapes, because the curvature of the spatial hyper-surfaces of the Universe can be positive, zero or negative. Indeed, the curvature of the spacetime depends on the value of the parameter k, which can be scaled in such a way that the only three possible values are +1, 0 or -1.

Specifically, the three possibilities are the following:

- A spherical surface with radius R (a *closed* spherical Universe with Gaussian curvature $1/R^2 > 0$, because $k = +1$) where, always in spherical coordinates $R > 0$, $0 \leq \theta \leq 2\pi$ and $0 \leq \varphi < 2\pi$, and with spatial metric $\mathrm{d}l^2 = R^2(\sin^2\theta\mathrm{d}\varphi^2 + \mathrm{d}\theta^2)$, setting $R = a$ and $r = \sin\theta$ we obtain that $\mathrm{d}r = \cos\theta\mathrm{d}\theta$, so $\mathrm{d}\theta^2 = \frac{\mathrm{d}r^2}{\cos^\theta} = \frac{\mathrm{d}r^2}{1-\sin^2\theta} = \frac{\mathrm{d}r^2}{1-r^2}$ and, finally, we have the *two-dimensional closed spherical spatial metric*:

$$\mathrm{d}l^2 = a^2\Big(\frac{\mathrm{d}r^2}{1-r^2} + r^2\mathrm{d}\varphi\Big). \tag{1.2.10}$$

- An hyperbolyc surface (an *open* hyperbolyc Universe with Gaussian curvature $1/R^2 < 0$, because $k = -1$) where, considering the trigonometrical hyperbolyc function $r = \sinh\theta$ with $0 \leq \rho < \infty$, we have a spatial metric $\mathrm{d}l^2 = R^2(\sinh^2\theta\mathrm{d}\varphi^2 + \mathrm{d}\theta^2)$ that, setting $R = a$, becomes the *two-dimensional open hyperbolyc spatial metric*:

$$\mathrm{d}l^2 = a^2\Big(\frac{\mathrm{d}r^2}{1+r^2} + r^2\mathrm{d}\varphi\Big). \tag{1.2.11}$$

- A Cartesian flat surface (a *flat Euclid*ean Universe with $R_{curvature} \longrightarrow \infty$, because $k = 0$) which, in polar coordinates $0 \leq \rho < \infty$ and $0 \leq \theta \leq 2\pi$, gives a spatial flat metric $\mathrm{d}l^2 = \mathrm{d}x^2 + \mathrm{d}y^2 = \mathrm{d}\rho^2 + \rho^2\mathrm{d}\theta^2$; then, introducing the dimensionless parameter $r \equiv \rho/a$ we obtain the *two-dimensional flat spatial metric*:

$$\mathrm{d}l^2 = a^2(\mathrm{d}r^2 + r^2\mathrm{d}\theta^2). \tag{1.2.12}$$

Now, we are able to write the *general form of a two-dimensional spatial metric*:

$$\mathrm{d}l^2 = a^2\Big(\frac{\mathrm{d}r^2}{1-kr^2} + r^2\mathrm{d}\varphi^2\Big), \tag{1.2.13}$$

in polar coordinates and with the dimensionless variable $r = \rho/a$.

By analogy in three-dimension, $\mathrm{d}\varphi^2 \longrightarrow \mathrm{d}\Omega^2 = \mathrm{d}\theta^2 + \sin^2\theta\mathrm{d}\varphi^2$ and the rest does not change, so we get back the *FLRW metric* (1.2.9):

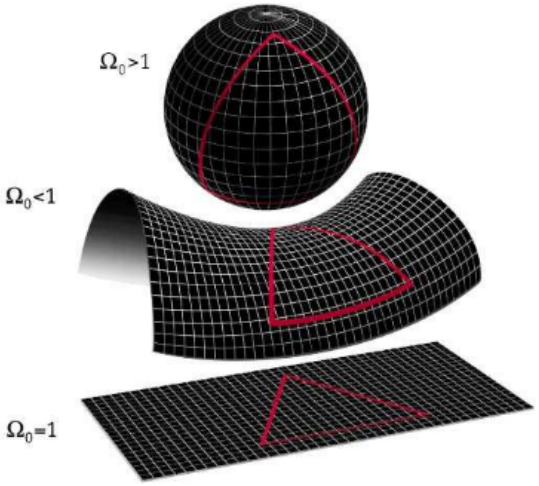

Figure 1.1: The different spatial geometries corresponding to the density parameter values $\Omega_0 > 1$, $\Omega_0 < 1$ and $\Omega_0 = 1$, respectively (see Eq. (1.6.5)). *Credits*: *wikipedia*

$ds^2 = (cdt)^2 - a^2(t)[\frac{dr^2}{1-kr^2} + r^2(\sin^2\theta d\varphi + d\theta^2)]$.

Resuming, the three values of the curvature parameter k lead to three different geometries of the Universe:

- if $k = 0$, the space has an infinite curvature and an *Euclid*ean (Cartesian) geometry that is, a *flat Universe*;
- if $k = +1$ the space is a hypersphere and has a finite volume but without boundaries, like a the two-dimensional case of a sphere; it is the case of a *closed Universe*;
- if $k = -1$ the space has an hyperbolic geometry; it is the case of an infinite *open Universe*.

The last two cases are two examples of pseudo-*Euclid*ean spatial geometries: for a closed Universe the sum of the internal angles of a triangle is $< \pi$, for a open Universe this sum is $> \pi$ (see Figure 1.1).

1.3 The Hubble-Lemaitre law

Let us integrate the FLRW metric (1.2.9) from the spatial point $P(0,0,0)$, system's origin, to a generic $P(r,0,0)$ and assume $d\theta = d\varphi = 0$ (namely θ and φ directed along the line of sight) and $dt = 0$ (ideal instantaneity of events). In this way we obtain the *proper distance*:

$$d_{pr} = \int_0^r \frac{a(t)dr'}{\sqrt{1-kr'^2}}\,, \tag{1.3.1}$$

but, being $a(t)$ not proportional to r and dependent only on time, it can be carried out from the integral (moreover, from here it can be seen that the non-static nature of the FLRW metric is due precisely to this dependence) and we can write:

$$d_{pr} = a(t)F(r) \qquad \text{with} \quad F(r) \equiv \int_0^r \frac{dr'}{\sqrt{1-kr'^2}}. \tag{1.3.2}$$

$F(r)$ is a function of radius that describes the spatial part of the metric and which, with variation of the curvature parameter k, gives a space:

$$\begin{aligned} \textit{Euclidean}\ \text{Flat if } k = 0 \quad &\Longrightarrow F(r) = r \\ \text{Closed Spherical if } k = +1 \quad &\Longrightarrow F(r) = \textstyle\int_0^r \frac{dr'}{\sqrt{1-r'^2}} \\ \text{Open Hyperbolic if } k = -1 \quad &\Longrightarrow F(r) = \textstyle\int_0^r \frac{dr'}{\sqrt{1+r'^2}}. \end{aligned}$$

However, because $a(t)$ depends on time, d_{pr} depends on time too; so we are able to define the proper distance calculated on the present time t_0, named *co-moving radial distance* d_c:

$$d_c \equiv d_{pr}(t_0) = a_0 F(r) = a_0 \int_0^r \frac{\mathrm{d}r'}{\sqrt{1-kr'^2}}, \tag{1.3.3}$$

where $a_0 \equiv a(t = t_0)$ is the present value of the expansion parameter, and that is called co-moving because it remains the same as the Universe expands.

It exists a direct link between the two kinds of distances just listed, called $d_c - d_{pr}$ *relation*:

$$d_{pr} = d_c \frac{a(t)}{a_0}, \tag{1.3.4}$$

where we can note that at present time t_0, $d_{pr} \equiv d_c$. Furthermore, because $a = a(t)$, this relation can change at different instants t of the look-back time of the Universe.

This time dependence implies the existence of a *radial velocity* v_r (between P and P_0), as derivative of d_{pr} with respect to t:

$$v_r = \frac{\mathrm{d}d_{pr}}{\mathrm{d}t} = \frac{\mathrm{d}[a(t)F(r)]}{\mathrm{d}t} = \frac{\mathrm{d}a(t)}{\mathrm{d}t} F(r) + a(t)\frac{\mathrm{d}F(r)}{\mathrm{d}t}, \tag{1.3.5}$$

but $\frac{\mathrm{d}F(r)}{\mathrm{d}t} = 0$ because of the time independence of $F(r)$, thus the relation (1.3.5) becomes:

$$v_r = \dot{a}(t)F(r), \tag{1.3.6}$$

but for the Eq. (1.3.2) we have: $F(r) = \frac{d_{pr}}{a(t)}$, hence we obtain finally the *Hubble-Lemaitre Law* (or simply the Hubble law):

$$v_r = \frac{\dot{a}(t)}{a(t)} d_{pr} = H(t) d_{pr} \quad \text{with } H(t) \equiv \frac{\dot{a}(t)}{a(t)}. \tag{1.3.7}$$

$H(t)$ is the *Hubble's parameter*.

The Hubble's law (1.3.7) is the linear relation existing between v_r and d_{pr}: v_r is directly proportional to d_{pr} thus, the more we look far into the Universe, the faster it expands; $H(t)$ gives us information about the isotropic *local* expansion of the Universe, at fixed cosmic time t. The motion of objects due solely to the expansion of the Universe is called the *Hubble flow*.

The parameter H is often, and erroneously, called *Hubble's constant*, because it has the same value across the Universe at a given cosmic time but it is time-dependent. At present time t_0, the most recent measures (Aghanim et al., 2018) give:

$$H_0 \equiv H(t = t_0) = 67.36 \pm 0.54 \quad \mathrm{kms}^{-1}\mathrm{Mpc}^{-1}, \tag{1.3.8}$$

but it is conventional to introduce a dimensionless parameter, h, redefining the Hubble parameter as:

$$H_0 = 100h \quad \mathrm{km\ s}^{-1}\mathrm{Mpc}^{-1}, \quad \text{with } h \approx 0.7. \tag{1.3.9}$$

Note that H_0 is expressed in terms of s^{-1}, thus H_0^{-1} provides us with a crude estimate of the current age of the Universe, that is approximately $13.8Gyr$. Furthermore the existence of the Hubble's law is a natural consequence of the CP for non-relativistic expansion velocities: in fact this law is identical for every place in Universe, according to CP.

1.3.1 The cosmological redshift

The *redshift* z is dimensionless and defined as the relative difference between the observed, λ_{oss}, and emitted, λ_{em}, wavelengths:

$$z = \frac{\lambda_{oss} - \lambda_{em}}{\lambda_{em}} = \frac{\Delta\lambda}{\lambda_{em}}, \tag{1.3.10}$$

or analogously for frequencies, by means of the relation $\lambda = c/\nu$. From Eq.(1.3.10), we see that $z > 0$ when the electromagnetic radiation from an object is increased in wavelength, and so shifted to the red region of the spectrum, that is the redshift; vice-versa we would have a blueshift for $z < 0$, but it is not the case for the Hubble flow. Substantially the redshift is a measure of the

shift of the electromagnetic spectrum emitted by a light source at $\lambda = \lambda_{em}$ and $t = t_{em}$, moving at a radial velocity v_r (1.3.7) with respect to observer, which sees the radiation at $\lambda = \lambda_{oss}$ and at $t_{oss} = t_0$.

Now, consider a source emitting a photon with wavelength λ_{em} at time t_{em} and an observer located at a distance d which will receive the signal with wavelength λ_{oss} at t_{oss}. Let us consider the FLRW metric (1.2.9) for this photon; it moves along null geodesics during the expansion of Universe thus, the term $\mathrm{d}s^2$ of the metric is null and we have:

$$(c\mathrm{d}t)^2 - a^2(t)\Big[\frac{\mathrm{d}r^2}{1-kr^2} + r^2(\sin^2\theta\mathrm{d}\varphi + \mathrm{d}\theta^2)\Big] = 0. \tag{1.3.11}$$

Then, let us assume θ and φ constants ($\mathrm{d}\theta = \mathrm{d}\varphi = 0$) for simplicity such that, integrating the metric along the path, the Eq. (1.3.11) becomes:

$$\int_{t_{em}}^{t_{oss}} \frac{c\mathrm{d}t}{a(t)} = \int_0^r \frac{\mathrm{d}r'}{\sqrt{1-kr'^2}} \equiv F(r). \tag{1.3.12}$$

Now let us suppose that another photon is emitted, by the same source, with a time delay δt_{em} with respect to the first one and reaches the observer at a time $t'_{oss} = t_{oss} + \delta t_{oss}$. Since the path of the two photons is the same, they differ only for the time employed to reach the observer because $F(r)$ does not change and we can write Eq. (1.3.12) as:

$$\int_{t_{em}}^{t_{oss}} \frac{\mathrm{d}t}{a(t)} = \int_{t_{em}+\delta t_{em}}^{t_{oss}+\delta t_{oss}} \frac{\mathrm{d}t}{a(t)} = F(r). \tag{1.3.13}$$

If we consider then the time interval small enough ($\delta t \to 0$), we can consider $a(t)$ almost constant and we obtain:

$$\frac{\delta t_0}{a(t_{oss})} = \frac{\delta t_{em}}{a(t_{em})}. \tag{1.3.14}$$

Since $\delta = 1/\nu$ and $\lambda = c/\nu$ we have then:

$$1 + z = \frac{a_{oss}}{a_{em}} \tag{1.3.15}$$

or, more generally for observers located at present time and for a generic instant t:

$$1 + z = \frac{a_0}{a(t)}. \tag{1.3.16}$$

The redshift is usually used to measure distances of extra-galactic sources; a_0 is a measurement of how much the scaling factor has changed over time from the photon emission to observation.

Note that it would be wrong to assert that the cosmological redshift is due to a pure Doppler effect, being the space that dilates itself and all structures contained inside it. For the same reason we have a velocity not because the source is moving but because the space is expanding itself ($z > 0$ always). Finally we can note that this Universe's expansion velocity grows up with the look-back time ($v_r \propto a^{-1}$).

For completeness we define the *deceleration parameter* $q(t)$:

$$q(t) \equiv -\frac{\ddot{a}(t)a(t)}{\dot{a}^2(t)} \quad \text{and that today: } q_0(t) = -\frac{\ddot{a_0}(t)a_0(t)}{\dot{a_0}^2(t)}. \tag{1.3.17}$$

It represents a way to indicate the second derivative of $a(t)$ but expressed to be dimensionless. $q(t)$ is defined with a minus sign because, with Friedmann equations (1.5.2) and (1.5.3) which we will see later, every possible Universe expands itself by decelerating, so that if $\ddot{a} < 0$ then $q > 0$.

1.4 Cosmological distances

In cosmology there is no absolute definition of distance. As we have already seen in Eq. (1.3.4), the system of co-moving coordinates is linked to the proper distance in spaces described by the FLRW metric, in the sense that d_{pr} is the distance measured on a hypersurface of constant proper

time. This kind of distance measure can not however be performed for astronomical objects of which we can only measure the light emitted, that takes a finite time to get to us.

The link that links the flow of energy l received from a source and its intrinsic luminosity L, dependent on the distance of the source d_L, is:

$$l = \frac{L}{4\pi d_L^2}, \tag{1.4.1}$$

where d_L is called *luminosity distance*. The concept of *standard candle* is based on this measure.

Indeed, we have to consider the presence of *redshift* and *time dilation*, with the latter coming into play because we are dealing with L, that is an energy per unit of time (of course, z influences the measure by reddening the radiation during its path from the source to us). Then the two factors come into play: $1 + z = \frac{a_0}{a}$ for z and $\delta t_0 = (\frac{a_0}{a})\delta t$ (because $L = L(\frac{a}{a_0})$) for the time's dilation. Finally we have:

$$l = \frac{L}{4\pi a_0^2 d_L^2}(1+z)^{-2} \tag{1.4.2}$$

and from this one we obtain the definition of *luminosity distance*:

$$d_L = a_0 r(1+z) = a(t)r(1+z)^2. \tag{1.4.3}$$

We have to use this definition when we measure luminosities to estimate cosmological distances.

Another classical measure is the *distance in angular diameter* or, simply, the *angular distance* d_A which is defined in such a way as to preserve the geometrical characteristics of a *Euclid*ean space. Substantially, the angle $\mathrm{d}\theta$ under which we see an object of known size D_{pr} (proper diameter of the object) gives us the distance of that object placed at coordinates r and t in the spacetime. If this angle $\mathrm{d}\theta$ represents the vertex of the isosceles triangle with base D_{pr} and height r and if we assume D_{pr} perpendicular to our line of sight, due to the assumption of a *Euclid*ean space, then this implies that every point of the proper diameter is aligned ($\mathrm{d}r = \mathrm{d}\varphi = 0$) and the light comes from every point at the same time ($\mathrm{d}t = 0$). Hence the FLRW metric for flat spaces becomes: $\mathrm{d}s^2 = a^2(t)r^2\mathrm{d}\theta^2 \equiv D_{pr}^2$, from which we get: $D_{pr} = a(t)r\mathrm{d}\theta$.

For small angular sizes (as it is the case with the typical dimensions of cosmological objects) we can define the *angular distance*:

$$d_A = \frac{D_{pr}}{\Delta\theta}. \tag{1.4.4}$$

This quantity is then connected to d_L through the *duality relation*:

$$d_A = a(t)r = \frac{d_L}{(1+z)^2} \tag{1.4.5}$$

which is valid only if the FLRW metric is founded because it implies the validity of the CP, hence this relation is a good test of the FLRW model.

1.5 The Friedmann equations

As already mentioned, at the base of each relativistic (General Relativity) cosmological model we have the *Einstein equations* (1.2.5): $R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R = \frac{8\pi G}{c^4}T_{\mu\nu}$, which assert that the spacetime geometry of the Universe, expressed by the *metric tensor* $g_{\mu\nu}$, is determined by its energetic content (or equally of matter, because of the famous relation $E = mc^2$), represented by the *energy-momentum tensor* $T_{\mu\nu}$. With only this assumptions we apply now the FLRW metric (1.2.9), that embodies the CP, to the Einstein equations (1.2.5), in order to obtain $R_{\mu\nu}$ and R (namely the spatial part of the equation (1.2.5)). As $T_{\mu\nu}$ we assume that of a *perfect fluid* (that is a fluid with density $\rho = const.$ and with negligible viscosity):

$$T_{\mu\nu} = -pg_{\mu\nu} + (p + \rho c^2)U_\mu U_\nu, \tag{1.5.1}$$

where p is the pressure, ρc^2 is energy density and U_k the four-velocity. The system provides sixteen equations ($\mu, \nu = 0, 1, 2, 3$), of which only two are independent, because of the assumption of homogeneity and isotropy. They provide the time evolution of $a(t)$ describing the growth of the Universe.

The first is the time-time component, or *first Friedmann equation*:

$$\ddot{a} = -\frac{4}{3}\pi G\Big(\rho + \frac{3p}{c^2}\Big)a, \tag{1.5.2}$$

and the second represents the space-space component, or *second Friedmann equation*:

$$\dot{a}^2 + kc^2 = \frac{8}{3}\pi G\rho a^2. \tag{1.5.3}$$

To create realistic cosmological models means solving these two equations.

Then, assuming the Universe as a *closed system* (reasonably, because there is no *outside* from the Universe), unable to lose energy, the two equations above can be linked to each other by the *adiabatic condition*, that is called *adiabatic expansion of Universe* in cosmology and which it can be expressed by the equations (there are three equivalent form of this):

$$\mathrm{d} = (\rho c^2 a^3) = -p\mathrm{d}a^3, \quad \dot{p}a^3 = \frac{\mathrm{d}}{\mathrm{d}t}[a^3(\rho c^2 + p)], \quad \dot{\rho} + 3(\rho + \frac{p}{c^2})\frac{\dot{a}}{a} = 0. \tag{1.5.4}$$

Each constituent of the Universe has its own equation of state, but all of them contribute to the density term ρ (in fact, the ρ in such equations is the sum of all the Universe components).

1.5.1 The cosmological constant

Looking at the *first Friedmann equation* (1.5.2) we can notice that to have a *static Universe* ($\ddot{a} = \dot{a} = 0$) we need that:

$$\rho = -\frac{3p}{c^2}, \tag{1.5.5}$$

i.e the energy density (or pressure) must be negative. Before the discovery of the Hubble flow, the scientific community supposed a static Universe and, in light of this, Einstein added to his field equations (1.2.5) a new term that modified the equations as follows:

$$R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R - \Lambda g_{\mu\nu} = \frac{8\pi G}{c^4}T_{\mu\nu}, \tag{1.5.6}$$

with Λ that is the so-called *cosmological constant* with dimension $[L^{-2}]$.

From what we said above we define:

$$\tilde{p} \equiv p - \frac{\Lambda c^4}{8\pi G} = p + p_\Lambda \quad \textit{and} \quad \tilde{\rho} \equiv \rho + \frac{\Lambda c^2}{8\pi G} = \rho + \rho_\Lambda, \tag{1.5.7}$$

called *effective pressure* and *density* with $p_\Lambda < 0$ and $\rho_\Lambda > 0$, respectively. With these quantities we ca re-write the Eq. (1.5.6) as:

$$R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R = \frac{8\pi G}{c^4}\tilde{T}_{\mu\nu}, \tag{1.5.8}$$

with $\tilde{T}_{\mu\nu}$ called *effective energy-momentum tensor*:

$$\tilde{T}_{\mu\nu} = T_{\mu\nu} + \frac{\Lambda c^4}{8\pi G}\rho_{\mu\nu} = \tilde{p}g_{\mu\nu} + (\tilde{p} + \tilde{\rho}c^2)U_\mu U_\nu. \tag{1.5.9}$$

From the equation (1.5.8) we re-obtain the *Friedmann equations* but with $\tilde{p}$ and $\tilde{\rho}$ instead of p and ρ. These equations admit *static solutions* for:

$$\tilde{\rho} = \frac{-3\tilde{p}}{c^2} = \frac{-3kc^2}{8\pi G a^2}. \tag{1.5.10}$$

When Hubble finally discovered the expansion of the Universe (Hubble, 1929), Einstein judged the introduction of Λ as ˝one of the biggest error of his life". Nevertheless, Λ-models have been reintroduced to describe the accelerated expansion of the Universe, for the first time discovered thanks to the observations of Supernovae (SNIa) (Riess et al., 1998). The Cosmological Constant has the effect of a source of repulsive gravitational force. It is a physical constant of the theory, and provides the simplest possible explanation for *dark energy (DE)*.

1.6 The Friedmann models

First of all, we assume a perfect fluid with a given ρ and p, without concern for now neither for the equation of state that links them, nor for the cosmological constant. We have to solve the Friedmann's equations, together with the adiabatic condition:

$$\begin{cases} \ddot{a} = -\frac{4}{3}\pi G(\rho + \frac{3p}{c^2})a & \text{first Friedmann equation} \\ \dot{a}^2 + kc^2 = \frac{8}{3}\pi G\rho a^2 & \text{second Friedmann equation} \\ \mathrm{d}(\rho a^3) = -3\frac{p}{c^2}a^2\mathrm{d}a & \text{adiabatic condition} \end{cases}$$

This system will allow us, in principle, to calculate the time evolution of $\rho(t)$, $p(t)$ and $a(t)$, once the state equation that links them is known (Friedmann, 1922).

Suppose we are nowadays, at t_0, then we know that $H_0 = \frac{\dot{a}_0}{a_0}$ and let us define the *critical density*:

$$\rho_{crit} \equiv \frac{3H^2}{8\pi G} \tag{1.6.1}$$

and the *today's critical density*:

$$\rho_{crit,0} \equiv \frac{3H_0^2}{8\pi G}. \tag{1.6.2}$$

If we re-write the second Friedmann's equation above into the form:

$$\Big(\frac{\dot{a}}{a_0}\Big)^2 = \frac{-kc^2}{a_0^2} + \rho_0\Big(\frac{\dot{a}_0}{a_0}\Big)^2\frac{H_0^2}{\rho_{crit,0}}, \tag{1.6.3}$$

then we can obtain the following equation:

$$H_0^2\Big(1 - \frac{\rho_0}{\rho_{crit,0}}\Big) = -\frac{kc^2}{a_0^2}, \tag{1.6.4}$$

that is the $H_0 - \rho_{crit,0}$ *relation*.

Let us introduce the *density parameter*:

$$\Omega \equiv \frac{\rho}{\rho_{crit}} \quad \text{and its today value} \quad \Omega_0 \equiv \frac{\rho_0}{\rho_{crit,0}}. \tag{1.6.5}$$

Equation (1.6.4) implies that for the three values of k we have:

- for $k = -1$: $(1 - \frac{\rho_0}{\rho_{crit,0}}) > 0 \Longrightarrow \rho_0 < \rho_{crit,0} \Longrightarrow \Omega_0 < 1$, hence a hyperbolic Universe;
- for $k = +1$: $(1 - \frac{\rho_0}{\rho_{crit,0}}) < 0 \Longrightarrow \rho_0 > \rho_{crit,0} \Longrightarrow \Omega_0 > 1$, hence a hyper-spherical Universe;
- for $k = 0$: $(1 - \frac{\rho_0}{\rho_{crit,0}}) = 0 \Longrightarrow \rho_0 = \rho_{crit,0} \Longrightarrow \Omega_0 = 1$, hence a flat Universe.

It is clear now that accurately estimating the value of $\rho_{crit,0}$ will help us discriminate the geometry of our Universe (via k). However, Eq. (1.6.4) is telling us that to estimate Ω_0 we must know $\rho_{crit,0}$, and then H_0, whose currently accepted value is that of Eq. (1.3.8), which is a fairly precise value. In other words, to know the geometry of Universe we have to know the current density ρ_0 of all the Universe components. This is done using the definition (1.6.2), from which it has been estimated a value, dependent on the Universe in which we live, through the value of the dimensionless parameter h (see (1.3.9)), and that it is estimated to be:

$$\rho_{crit,0} = 1.9 \times 10^{-29}h^2[gcm^{-3}] = 2.775 \times 10^{11}h^2[M_\odot Mpc^{-3}]. \tag{1.6.6}$$

Therefore, we have to accurately estimate the values of ρ_{DM} and ρ_{DE} to derive k.

Let us now see what happens by reintroducing the cosmological constant Λ, which has been seen to behave like a fluid with a *strange* equation of state. The Friedmann equations expressed by (1.5.2) and (1.5.3) will be changed, by replacing ρ and p with $\tilde{\rho}$ and $\tilde{p}$, and after some manipulations they can be written as follows:

$$\ddot{a} = -\frac{4}{3}\pi Ga\Big(\rho + \frac{3p}{c^2}\Big) + \frac{\Lambda c^2 a}{3} \quad \text{first Friedmann equation with } \Lambda, \tag{1.6.7}$$

$$\dot{a}^2 + kc^2 = \frac{8\pi}{3}G\rho a^2 + \frac{\Lambda c^2 a^2}{3} \quad \text{second Friedmann equation with } \Lambda. \tag{1.6.8}$$

Manipulating the second *modified* Friedmann equation (dividing both members by a_0^2 and developing it and then considering it at t_0) and keeping in mind the equation (1.5.8) and the definition of $H_0 \equiv \dot{a}_0/a_0$, we obtain:

$$H_0^2\Big(1 - \frac{\Lambda c^2}{3} - \frac{\rho_0}{\rho_{crit,0}}\Big) = -\frac{kc^2}{a_0^2}. \tag{1.6.9}$$

Then, remembering the *contribution of* Λ *to the total current density* $\rho_{tot,0}$:

$$\rho_{\Lambda,0} = \frac{\Lambda c^2}{8\pi G} \tag{1.6.10}$$

and defining the *contribute of* Λ *to the current density parameter*:

$$\Omega_{\Lambda,0} = \frac{\rho_{\Lambda,0}}{\rho_{crit,0}}, \tag{1.6.11}$$

we can re-write equation (1.6.9) as:

$$H_0^2(1 - \Omega_{\Lambda,0} - \Omega_0) = -\frac{kc^2}{a_0^2}. \tag{1.6.12}$$

Finally, defining a *total current density parameter*, as the sum of the contributions of the various components of the Universe:

$$\Omega_{tot,0} \equiv \sum_i \Omega_{i,0}, \tag{1.6.13}$$

the Eq. (1.6.12) becomes the *second Friedmann's equation expressed as a function of* H_0 *and* $\Omega_{tot,0}$:

$$H_0(1 - \Omega_{tot,0}) = -\frac{kc^2}{a_0^2}. \tag{1.6.14}$$

From Eq. (1.6.14) we understand that we can discriminate the values of the curvature parameter, k, based on the values of $\Omega_{tot,0}$, regardless of what the components of the Universe are.

1.6.1 Models with a perfect fluid

As we mentioned above in Section 1.2, the perfect fluid approximation for the Universe considered on a large scale is realistic. In the field equations, the perfect fluid enters through $T_{\mu\nu}$ because it succeeds in matching these equations with the CP because it does not admit anisotropic pressures.

The equation of state will take the generic form $p = p(\rho)$ which is assumed, as in many realistic physical situations in adiabatic conditions, in the form of the *general form of the equation of state*:

$$p = w\rho c^2 \qquad \text{with } w = const. \tag{1.6.15}$$

The w parameter lies in the so-called *Zel'dovich interval*:

$$w \in [0, 1[\tag{1.6.16}$$

the reason why $w = 1$ is not allowed is clear if we consider the definition of **sound velocity**:

$$c_S = \Big(\frac{\delta p}{\delta \rho}\Big)^{1/2}_{S=const.} = c\sqrt{w}. \tag{1.6.17}$$

The case $w = 1$ implies $c_S = c$ that is not acceptable. However if we had $w < 0$, then c_S would belong to the set of imaginary numbers but, as we will see, it will be an acceptable value (precisely for the Λ component). Anyway, w will assume different values depending on the pressure component we are dealing with.

The ordinary matter fluid can be in two different states: *non-relativistic* and *relativistic*. The first case is represented by baryons and CDM (called *dust* in this cosmological approach), having $w = 0$, i.e with $p = 0$. While a non-degenerate and ultra-relativistic fluid has an equation of state with $w = 1/3$ (hence $p = \frac{1}{3}\rho c^2$). This is the case for a *radiative fluid* or, more generally, for photons and relativistic particles like neutrinos.

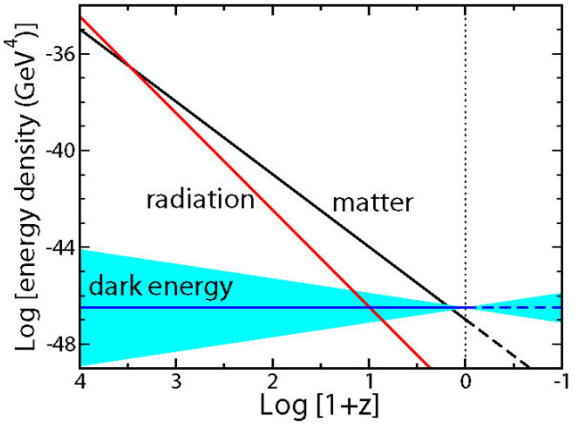

Figure 1.2: Redshift evolution of radiation, matter and dark energy densities. The cyan shaded band corresponds to $w = -1 \pm 0.2$. *Credits:* Frieman

Now, introducing the general form of the equation of state (1.6.15) into the *adiabatic condition* (1.5.4), and remembering that $1+z = \frac{a_0}{a(t)}$, after some simplifications we obtain the general relation between ρ and a:

$$\rho = \rho_0 \Big(\frac{a}{a_0}\Big)^{-3(1+w)} = \rho_0 (1+z)^{3(1+w)}. \tag{1.6.18}$$

For the matter ($w = 0$) and for the radiation ($w = 1/3$) we obtain, respectively:

$$\rho_M = \rho_{M,0}\Big(\frac{a}{a_0}\Big)^{-3} = \rho_{M,0}(1+z)^3 \quad \text{and} \quad \rho_R = \rho_{R,0}\Big(\frac{a}{a_0}\Big)^{-4} = \rho_{R,0}(1+z)^4. \tag{1.6.19}$$

The density evolution depends on the value of w. This can be used to determine which component dominates over the others at a given redshift. Indeed we can note that ρ_R decreases with time more quickly than ρ_M because only radiation is affected by the redshift caused by the Hubble flow (the matter is aggregated). With regard to the Λ (which we will assign to DE) we assume $w = -1$ so its contribution to ρ_{TOT} is:

$$\rho_\Lambda = \rho_{\Lambda,0} = const. \tag{1.6.20}$$

that means that ρ_Λ does not evolve with time (in fact, in the past matter and radiation dominated on the contribute on ρ_{TOT}).

The *current* density values of the various components are:

- $\rho_{R,0} = 10^{-34}[gcm^{-3}] \longrightarrow \Omega_{R,0} = 10^{-5}$,
- $\rho_{M,0} \approx 0.3\rho_{crit,0} \longrightarrow \Omega_{M,0} \approx 0.3$, where we consider the sum of the contributions of dark matter and baryons,
- $\rho_{\Lambda,0} \approx 0.7\rho_{crit,0} \longrightarrow \Omega_{\Lambda \equiv DE,0} \approx 0.7$,
- $\rho_{tot,0} \approx \rho_{crit,0} \longrightarrow \Omega_{tot,0} \approx 1$, i.e the data are consistent with a spatially flat Universe.

Plotting the evolution of the various density components as a function of redshift, we obtain the bheaviours shown in Figure 1.2. Currently, we are dominated by the Λ-component (the dark energy) and, in the future the energy density of matter and radiation will go further and further down going to the solution of de Sitter (a flat and empty Universe, hence with $k = p = \rho = 0$) while ρ_Λ, always remaining constant, from negligible that was at the beginning of the BB, it will dominate more and more in the future.

As a consequence of this, we could simplify the Friedmann equations considering only the component prevailing over the others, depending on the cosmic epoch we are studying. The transition from one regime to another is defined by some particular values of *equivalent redshift*:

- *matter-radiation equivalence*: $\rho_M = \rho_R \implies \rho_{M,0}(1+z_{eq})^3 = \rho_{R,0}(1+z_{eq})^4 \implies 1 + z_{eq} = \frac{\rho_{M,0}}{\rho_{R,0}} \approx 10^4 \longrightarrow z_{eq,M-R} \approx 10^4$

- *matter-Λ equivalence*: $\rho_M = \rho_\Lambda \implies \rho_{M,0}(1+z_{eq})^3 = \rho_{\Lambda,0} \implies 1+z_{eq} = \left(\frac{\rho_{\Lambda,0}}{\rho_{M,0}}\right)^{1/3} \approx 1.7 \longrightarrow z_{eq,\Lambda-M} \approx 0.3$

Let us rearrange now the second Friedmann equation(1.5.3) considering it for t_0, and dividing each member by a_0^2, while also considering the definition of H_0 (1.3.7), $\rho_{crit,0}$ (1.6.2) and the $\rho - w$ relation (1.6.18) to obtain:

$$\left(\frac{\dot{a}}{a_0}\right)^2 = H_0^2\left(1 - \frac{\rho_0}{\rho_{crit,0}}\right) + \frac{\rho_0\left(\frac{a_0}{a}\right)^3(1+w)}{\rho_{crit,0}}\frac{a^2}{a_0^2}H_0^2. \tag{1.6.21}$$

If we apply now the definition of Ω_0 of Eq. (1.6.5), we can rewrite the last equation (1.6.21) in the following two equivalent forms:

$$\left(\frac{\dot{a}}{a_0}\right)^2 = H_0^2\left[1 - \Omega_0 + \Omega_0\left(\frac{a_0}{a}\right)^{(1+3w)}\right] \quad \text{or} \quad H^2(t) = H_0^2\left(\frac{a_0}{a}\right)^2\left[(1-\Omega_0) + \Omega_0\left(\frac{a_0}{a}\right)^{(1+3w)}\right], \tag{1.6.22}$$

which provide us an analytic solution of the temporal evolution of $H(t)$ (through $a(t)$), in the case of *one-component Universe*. In the case of our real Universe with more components, we must resort to numerical solutions. However, to consider more components, we can just use the definition (1.6.13) instead of the simple Ω_0 for one component, and write it as function of z instead of t:

$$H^2(z) = H_0^2(1+z)^2\left[1 - \sum_i \Omega_{0,w_i} + \sum_i \Omega_{0,w_i}(1+z)^{(1+3w_i)}\right], \tag{1.6.23}$$

that is the *H-z relation* as a function of the w_i parameters or, in other words, the evolution of H with redshift z. Furthermore, it is usual to write it in the form:

$$H^2(z) = h_0^2E^2(z) \quad \text{with } E^2(z) = (1+z)^2\left[1 - \sum_i \Omega_{0,w_i} + \sum_i \Omega_{0,w_i}(1+z)^{(1+3w_i)}\right]. \tag{1.6.24}$$

Given the three components of the Universe currently known, to make it explicit it would be enough to replace the summation with the current values of the density parameters with the respective dependencies on w of $(1+z)$. Hence $(1-\Omega_{tot,0})$ is a measure of the curvature of the Universe.

1.7 Inevitability of the Big Bang in Friedmann Universes

If we consider the First Friedmann equation (1.5.2) together with the form (1.6.15) of the equation of state, we can write it as:

$$\ddot{a} = -\frac{4}{3}\pi G\rho(1+3w)a. \tag{1.7.1}$$

Then, from the definition of c_S with constant entropy (1.6.17), we have seen that to have $c_S < c$ we need that $(1+3w) > 0$ and then, ultimately, that $\ddot{a} < 0$, i.e a decelerating Universe (without considering Λ). But, all observations show that $H > 0$ (the Universe is expanding itself) thus, given its definition, this means that $\dot{a} > 0$.

In practice, the equation (1.7.1) is saying us that the function $a(t)$ must have negative concavity and no flexes: this means that the curve of $a(t)$, sooner or later, has to cross the t-axes when going back in time. In other words there must exist an *initial instant* in which the Universe began, i.e the *Big Bang*. The following growth of the Universe is not due to some kind of pressure, but rather it is intrinsically determined by the conditions of homogeneity and isotropy laid down by the CP. In the past there was a singularity (the BB), that is a point in which $a = 0$. It ensures us that the age of Universe $t_0 < H_0^{-1}$, too, with H_0^{-1} called *Hubble time*. We can understand better that, looking at Figure 1.3.

Furthermore, because $\rho \propto a^{-3(1+w)}$, when $a = 0$ (i.e at the BB where the Universe radius, given by a, was null) the density of Universe tends to infinite (and the temperature too). The *Big Bang model* cannot be avoided considering a Universe based on the Friedmann's equation, unless we change the definition of w (by adding Λ for example) not to give up the CP. When we are in proximity of the BB, we need necessarily a new gravitational theory which accounts for *quantum effects* on cosmological scale, to be able to study this singularity which seems to be only a limit of our classical theory of gravity (General Relativity).

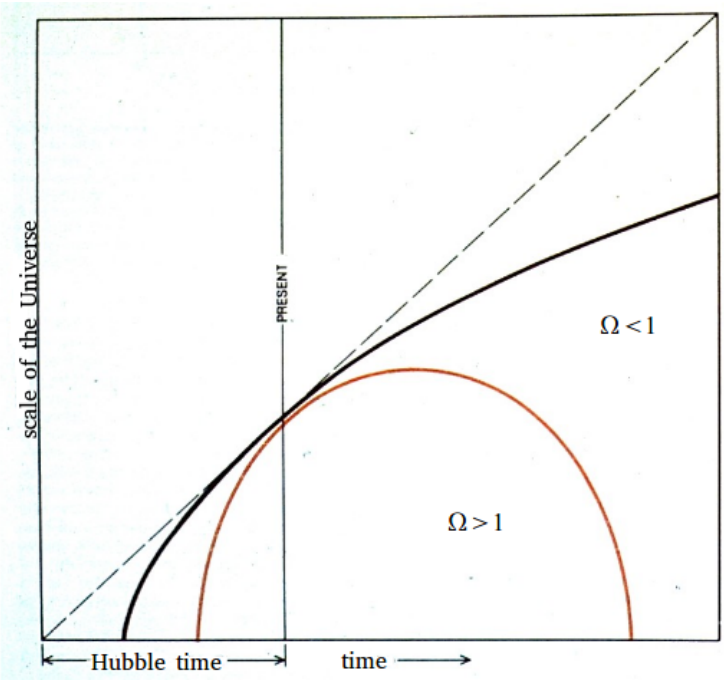

Figure 1.3: The time evolution of the Universe scale factor, $a(t)$, for open (black, $\Omega < 1$) and closed (red, $\Omega > 1$) geometries. The dashed line is the first derivative of $a(t)$, which intercepts the time axis at the 0-time (the BB). The vertical black line indicates t_0, i.e the present time. The *Hubble time* is indicated on the x-axis. *Credits*: lezione5.pdf

	general	pure matter ($w = 0$)	pure radiation ($w = \frac{1}{3}$)
a	$a(t) = a_0\left(\frac{t}{t_0}\right)^{\frac{2}{3(1+w)}}$	$a(t) \propto t^{2/3}$	$a(t) \propto t^{1/2}$
t	$t(z) = t_0(1+z)^{\frac{2}{3(1+w)}}$	$t(z) \propto (1+z)^{-3/2}$	$t(z) \propto (1+z)^{-2}$
H	$H(t) = \frac{2}{3(1+w)}t^{-1}$	$H(t) = \frac{2}{3t}$	$H(t) = \frac{1}{2t}$
t_0	$t_0 = \frac{2}{3(1+w)}\frac{1}{H_0}$	$t_0 = \frac{2}{3H_0}$	$t_0 = \frac{1}{2H_0}$
q	$q = \frac{1+3w}{2}$	$q = \frac{1}{2}$	$q = 1$
ρ	$\rho = \frac{1}{6\pi G(1+w)^2}\frac{1}{t^2}$	$\rho = \frac{1}{6\pi G}\frac{1}{t^2}$	$\rho = \frac{3}{32\pi G}\frac{1}{t^2}$

Table 1.1: This Table provides a list of some useful equations in the EdS Universe both for a generic fluid and for *matter-dominated* and *radiation-dominated* models, obtained keeping on mind the previous definition of the various quantities.

1.8 Flat and curved (open or closed) Universes: general results

Here we summarise the solutions of the Friedmann equation in the form (1.6.23) for the three models of possible Universes: flat, open and closed.

Let us start with the simplest model: a flat Universe with only one component (e.g. matter or radiation), $\Omega_0 = 1$, called *Einstein-de Sitter model (EdS)*. Equation (1.6.23), in this case, becomes:

$$H(z) = H_0(1+z)^{\frac{3(1+w)}{2}}. \tag{1.8.1}$$

It is interesting to note some properties of the EdS Universe for a generic fluid, arising from the table 1.1:

- the trend of $a(t)$ and $t(z)$, together with the assumption of the Zel'dovich interval (1.6.16) tells us that the expansion will never change its trend and it will continue to grow to infinite;
- $H(t)$ is a decreasing function of t, hence going into the past, the expansion is getting faster and faster and for $t \longrightarrow 0$ (i.e going towards the BB moment), $H \longrightarrow \infty$;

- from the definition of t_0 we can note that, for EdS models, the measure of H_0 gives directly the age of Universe;
- because $q = const.$, in the EdS Universes the expansion decelerates constantly;
- into EdS models, $\rho \propto t^{-2}$; indeed this is a reasonable fact because the expansion *dilutes* its components.

All these equations are valid at high-redshift, i.e very closely to the BB instant. Indeed, also curved Universes, approaching the BB, behave as flat. Furthermore, as we mentioned before, each model gives analytical solutions only for one-component Universes, so far from reality. For this reason, the EdS model is very useful in certain epochs of the history of the Universe, where only one component dominates over the others.

Let us now briefly analyse the case of *curved Universes*, considering models with a single component, characterised by $\Omega_w \neq 1$. Referring to equation (1.6.23), for small values of a or, equivalently, at high z we have that:

$$\Omega_0 \Big(\frac{a_0}{a}\Big)^{(1+3w)} \gg (1 - \Omega_0). \tag{1.8.2}$$

If we introduce a critical value $a*$ or, equivalently $z*$, such that:

$$\Omega_0 \Big(\frac{a_0}{a*}\Big)^{(1+3w)} = \Omega_0 (1 + z*)^{(1+3w)} = |1 - \Omega_0|, \tag{1.8.3}$$

for $z \gg z*$ (or $a \ll a*$), i.e closer and closer to the BB, the *curvature term* $|1 - \Omega_0|$ is negligible and the equation (1.6.23) simply becomes:

$$H(z) = H_0 \Omega_0^{1/2} (1 + z)^{\frac{3(1+w)}{2}}. \tag{1.8.4}$$

In terms of z we can say that the Universe starts to feel its curvature just for $z \ll z*$, that is, getting closer and closer to the present day. In other words, it is good practice to ignore the term of curvature until a young Universe ($z > 10$) is studied and to consider it with a good approximation an EdS-like Universe; vice versa to small redshifts. The impossibility to assert the curvature of the Universe, if we consider t too close to the Big Bang, is quite problematic.

Considering an open Universe with one component (where $\Omega_w < 1$ and $k = -1$) we are talking about a hyperbolyc Universe. Referring to equation (1.6.23) we can note that $\dot{a}$ will never be null and, reasonably supposing that $\dot{a} > 0$ at a certain instant t, then $\dot{a} > 0$ *forever*. In practice, a tends to grow indefinitely over time. For $a \gg a*$ ($z \ll z*$), the equation (1.6.23) becomes:

$$\dot{a} = a_0 H_0 (1 - \Omega_0)^{1/2} = const., \tag{1.8.5}$$

hence $a \propto t$, i.e a grows linearly with time in an open Universe. From the definition of the deceleration parameter (1.3.17) and from the equation (1.8.5) above we can deduce that:

$$q = -\frac{\ddot{a}a}{\dot{a}^2} \longrightarrow 0 \text{ asymptotically.} \tag{1.8.6}$$

Consequently, at a certain moment, the expansion continues indefinitely but with a constant velocity: it is a free and at constant velocity expansion that brings the Universe towards a sort of thermal death, where the spacetime is literally stretched heterogeneously. That is called *Big Rip*.

Considering a closed Universe with one component (where $\Omega_w > 1$ and $k = +1$) we are talking about hyperspherical Universe. In this case there is a moment t_{max} where $\dot{a}(t_{max}) = 0$. From Eq.(1.6.23) we can find the value of a which gives us $\dot{a} = 0$, i.e the maximum of the function $a(t)$:

$$a_{max} = a_0 \Big(\frac{\Omega_0}{\Omega_0 - 1}\Big)^{\frac{1}{(1+3w)}}, \tag{1.8.7}$$

that is the value of a for which the expansion reaches its maximum before to start to contract itself (in this maximum point the density has a minimum). In a closed Universe model there will be a second singularity with a Universe that closes on itself going towards the so-called *Big Crunch*.

Note that, with the same value of H_0, the age of closed Universes with positive curvature (i.e spherical) is smaller than that of flat Universes and, vice-versa for an open Universe (compared to a flat one): after all, having more matter ($\Omega > 1$), the expansion is more braked in a closed Universe.

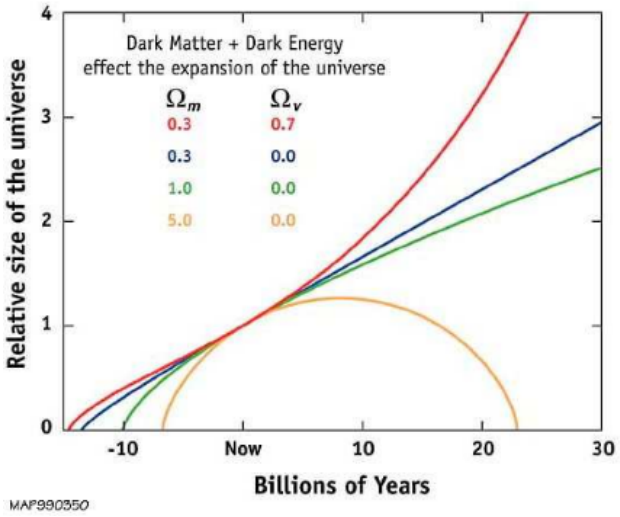

Figure 1.4: Four possible time evolution scenarios for the scale factor $a(t)$. The orange curve represents a one-component *ordinary* closed Universe ($\Omega_m = 5$). The blue one represents a one-component *ordinary* open Universe (hence hyperbolic). The green and red lines show a one-component and two-components flat Universes, respectively. The red one, in particular, refers to a Universe dominated by the dark energy component ($\Omega_\Lambda = 0.7$), that makes it accelerating at late times. This case is the closest one to our real Universe. *Credits*: $WMAP - Universe$

1.8.1 Time evolution of the density parameter

It is important to underline that the density parameter is a function of time, in each model. For a Universe with one component this dependence can be analytically expressed through z:

$$\Omega(z) = \frac{\Omega_0(1+z)^{(1+3w)}}{1-\Omega_0+\Omega_0(1+z)^{(1+3w)}}. \tag{1.8.8}$$

Another necessary observation: because for ordinary Universes ($0 \leq w < 1$) the exponent of $(1+z)^{(1+3w)}$ must be always positive, the geometry of Universe is not able to change from the BB until now.

1.9 Models with cosmological constant: the Lambda-CDM model

We have already said that observational data suggest our Universe expands itself accelerating, but this is incompatible with ordinary Universes, composed only by matter or radiation. Thus we suppose the existence of a *dark energy* which dominates over the other components and it causes an accelerated expansion. The simplest way to describe this effect is by including the cosmological constant, Λ, in the Einstein equations. We have already demonstrated that an ordinary Universe must expand decelerating, so a Universe of only matter is incompatible with the observed accelerated expansion.

Inserting the definition of $\tilde{p}$ and $\tilde{\rho}$ (1.5.7) into the Friedmann's equations (1.5.2) and (1.5.3), we obtain the *first Friedmann equation with the Λ contribution*:

$$\ddot{a} = -aH^2\frac{\Omega_m}{2} + aH^2\Omega_\Lambda \quad \text{with: } \Omega_\Lambda \equiv \frac{\Lambda c^2}{3H^2}. \tag{1.9.1}$$

From the definition of the deceleration parameter (1.3.17), the equation (1.9.1) brings to:

$$q = -\frac{\ddot{a}}{aH^2} = \frac{\Omega_m}{2} - \Omega_\Lambda, \tag{1.9.2}$$

hence, to have an accelerated expansion ($q < 0$), we must have $\ddot{a} > 0$. So we obtain the following condition for an accelerated expansion:

$$\Omega_\Lambda > \frac{\Omega_m}{2}. \tag{1.9.3}$$

At present days t_0, data suggest that $\Omega_{m,0} \approx 0.3$ and $\Omega_{\Lambda,0} \approx 0.7$, hence the condition is satisfied and it is estimated that $q_0 \approx -0.55$. With the same approach of Section 1.6.1, we can estimate the redshift at the moment at which Λ starts to accelerate the Universe is $z_\Lambda \approx 0.3$.

What we have described so far about the properties of a Λ which accelerates the expansion of the Universe, together with the percentages of the components of the Universe (see Chapter 1.1), that justify the estimated flat geometry, and together with the CP and the existence of the CMB, are the main properties of the Standard Cosmological Model. The model of the Universe exposed in this section is called *flat ΛCDM Model*, where CDM stands for *cold dark matter*. To understand the meaning of the CDM, we should confront it with the *Hot dark matter* (HDM) and introduce two scale-times (timed by the expansion factor):

- $a_{d.r.}$ is the *de-relativisation era*, i.e the instant in which the characteristic velocity dispersion of dark matter particles becomes non-relativistic (and it depends on the temperature of these particles): more massive particles become non-relativistic before the less massive ones;
- a_{dx} is the *era of decoupling from the cosmological fluid* (immediately after the Big Bang, there is only one *primordial broth* from which, with the passage of time, the various components of the present day Universe have been detached), i.e the instant after which the dark matter can interact (only gravitationally) with the other components of the Universe.

We do not yet have a clear idea of how dark matter is done but, through these two time-scales, we can distinguish two types:

- if $a_{dx} < a_{d.r.}$, i.e if the dark matter became non-relativistic *after* the decoupling, the typical mass of a dark matter particle must be little enough and in this case we use the definition *HDM*, whose best candidate turns out to be the neutrino;
- if $a_{dx} > a_{d.r.}$,i.e the dark matter became non-relativistic *before* the decoupling, the typical mass of a dark matter particle must be massive and in this case we speak about *CDM*, whose best candidate turns out to be the *WIMPs* (Weakly Interacting Massive Particles), that are particles able to interact only through gravity and weak force (for further information about the WIMPs see Jungman et al., 1996). A large list of dark matter particle candidates have been suggested so far. Other kind of particles involve both modifications of the theory of gravitay or exotic fields.

As we will see in Chapter 2, the dimensions of the primordial structures, progenitors of the current ones, depend on the type of dark matter assumed: the CDM model is the favorite one because it provides a hierarchical assembly of the matter where the first structures to virialize are the smaller ones. In fact, it is what we observe. This is the *Bottom-Up scenario*. The opposite case, is called *Top-Down scenario*.

In 1990, Efstathiou et al. (Efstathiou et al., 1990) noted that the CDM model was not able to explain the observed distribution of cosmological structures on very large scales. So they proposed that the Universe is now dominated by the Cosmological Constant. Nowadays, the ΛCDM model is accepted by almost the whole scientific community, though most of its theoretical bases are still unexplained. In particular, the physical origin of the accelerated expansion is still unknown. In equation (1.9.1), we have the contribution of Cosmological Constant in the FLRW metric. If we set $T_{\mu\nu}$ into the equation (1.5.8), only the Λ term remains. It represents the *vacuum energy*, constant in time, that provides a repulsive gravitational force, and it is described by an equation of state with $w = -1$. Until now the most plausible hypothesis is the association of the vacuum energy with the DE. However, the ΛCDM model says nothing about the physical origin of dark matter and dark energy.

Chapter 2

Theory of structures formation and evolution

In this Chapter, we will discuss the cosmological context in which the clusters of galaxies formed and evolved, through the Jeans theory. We will also discuss the aspects of the growth of perturbations in linear and non-linear regimes (spherical collapse and mass function) and the properties of the power spectrum, which is one of the main observables that characterise the perturbation growth.

2.1 The Jeans theory

The Jeans theory describes the formation and growth of cosmic structures as a consequence of *gravitational instabilities*. Density perturbations are believed to lead this process. They are originated as small fluctuations in the density of the primordial fluid, observationally detected in the CMB. Specifically, the density perturbation is a dimensionless quantity defined as:

$$\delta \equiv \frac{\delta\rho}{\bar{\rho}} = \frac{\rho - \bar{\rho}}{\bar{\rho}}, \tag{2.1.1}$$

where $\bar{\rho}$ is the mean matter density. We note that $-1 \leq \delta < +\infty$, since $\rho > 0$. Observationally, the primordial density perturbations are inferred from the CMB small temperature fluctuations, and are of the order of $\frac{\delta T}{T} \approx 10^{-5}$. In the local Universe, the density fluctuations are of the order of $\approx 10^2$ (nowadays we observe already virialized structures). So the question is how these primordial perturbations with $\delta \sim 10^{-5}$ at $z \approx 10^3$ evolved to current values, and how they drove the collapse and dynamical evolution of baryonic structures.

2.1.1 The Jeans scale

The Jeans scale determines the minimum size that an agglomerate of matter, with certain thermodynamical properties (ρ, T, etc.), must possess in order to collapse, and it depends on the cosmological model. The *not-collisional* dark matter particles decouple before the baryons. Therefore, as characteristic velocity of the dark matter component we will assume the velocity dispersion while, for the baryons, we will use the velocity of the sound in the medium c_S, as expressed in Eq. (1.6.17).

The Jeans theory is based on the idea that the collapse is dominated by two opposite processes: the gravitational collapse and the pressure inside the collapsing haloes, due to the thermal motions of the particles, which opposes it. Of course, the latter must be overcame by the gravity to allow the collapse. For a spherical collapse of a cloud of radius R, there are several approaches to determine the *Jeans scale* R_J, which we list here:

- An *energetic approach* where the gravitational potential energy $U = \frac{-GM^2}{R}$ is contrasted by the kinetic energy associated to the thermal agitation of particles $K = \frac{1}{2}Mv^2$. Equating these two forms of energy, we obtain the equilibrium point of the halo, i.e the Jeans radius:

$$R_J = v\sqrt{\frac{1}{2G\rho}}. \tag{2.1.2}$$

- A *balance of forces approach* where the gravitational force per unit mass $F_g = \frac{GM}{R^2}$ is contrasted by the thermal pressure force $F_p = \frac{pR^2}{M} = \frac{c_S^2}{\rho}$ and, equating these two forces we obtain a slightly different form of the Jeans radius:

$$R_J = \frac{c_S}{\sqrt{G\rho}}. \tag{2.1.3}$$

- The last approach consists of comparing the free-fall time $\tau_{ff} \propto \frac{1}{\sqrt{G\rho}}$ and the crossing time $\tau_{cross} = \frac{2R}{c_S}$ (the hydrodynamic time-scale necessary for a perturbation to propagate through the whole cloud under examination); so we obtain the last form of the Jeans radius:

$$R_J = \frac{c_S}{2\sqrt{G\rho}}. \tag{2.1.4}$$

The scales obtained with the three methods listed above are not identical for reasons of approximation, but all depend on the same thermodynamic quantity, i.e they are all expressed in a way that $R_J \propto \sqrt{\rho}$. In the following discussion, we will use the notation λ_J instead of R_J. It is also convenient to define a *Jeans mass*:

$$M_J = \frac{4}{3}\pi\rho R_J^3. \tag{2.1.5}$$

2.1.2 The Jeans theory in a static Universe

We start assuming a known solution to the problem, to which a small perturbation is added to study its temporal evolution. We will adopt a Newtonian treatment, considering a perfect fluid described by five quantities: the density ρ, the pressure p, the velocity $\vec{v}$, the cloud's gravitational potential Φ and the entropy S. The five *hydro-dynamic equations* that describe our system are the following:

$$\begin{cases} \frac{\partial\rho}{\partial t} + \vec{\nabla}\cdot(\rho\vec{v}) = 0 & \text{Mass conservation equation} \\ \frac{\partial\vec{v}}{\partial t} + (\vec{v}\cdot\vec{\nabla})\vec{v} = -\frac{1}{\rho}\vec{\nabla}p - \vec{\nabla}\Phi & \text{Eulero's equation (momentum conservation)} \\ \nabla^2\Phi = 4\pi G\rho & \text{Laplace's equation} \\ p = p(\rho, S) = p(\rho) & \text{Equation of state} \\ \frac{\mathrm{d}S}{\mathrm{d}t} = 0 & \text{Adiabatic condition} \end{cases} \tag{2.1.6}$$

Assuming an adiabatic collapse, the system reduces to four equations in four unknowns. Now suppose we know a *background solution* with constant values:

$$\begin{cases} \rho = \rho_b = const. \\ p = p_b = const. \\ \Phi = \Phi_b = const. \\ \vec{v} = 0 = const. \end{cases} \tag{2.1.7}$$

Then we apply *small perturbations* to the background, to study how the system evolves:

$$\begin{cases} \rho = \rho_b + \delta\rho = \rho_b(1+\delta) \\ p = p_b + \delta p \\ \Phi = \Phi_b + \delta\Phi \\ \vec{v} = \delta\vec{v}. \end{cases} \tag{2.1.8}$$

Since the perturbations are small, the theory of Jeans is *linear* and we can neglect all second-order terms.

Now let us see how the first three equations of the hydrodynamic system (2.1.6) are modified, once we insert the perturbations (2.1.8) (the equation of state will be used in the form given by (1.6.17), because of the constant entropy of the collapsing cloud):

$$\begin{cases} \frac{\partial\delta\rho}{\partial t} + \rho_b\vec{\nabla}\cdot\delta\vec{v} = 0 & \text{Perturbed mass conservation equation} \\ \frac{\partial\delta\vec{v}}{\partial t} = -\frac{c_S^2}{\rho_b}\vec{\nabla}\delta\rho + \vec{\nabla}\delta\Phi & \text{Perturbed Eulero's equation} \\ \nabla^2\delta\Phi = 4\pi G\delta\rho & \text{Perturbed Laplace's equation} \end{cases} \tag{2.1.9}$$

The idea is to look for solutions to the system (2.1.9) in the form of *plane waves*:

$$f(\vec{r},t) = f_k \exp[i\vec{k}\cdot\vec{r} + i\omega t], \tag{2.1.10}$$

with $\vec{r}$ the spatial vector, f_k the amplitude of the wave and $\vec{k} = \frac{2\pi}{\lambda}\hat{k}$ the wave vector. All the variables will be used in the phase space by means of Fourier transforms, and the new unknowns to be solved will be the amplitude of the waves. We can find the solution of the perturbed system in phase space by setting the determinant of the representative matrix of the system to zero. In this way we get the so-called *dispersion's relation*:

$$\omega^2 = k^2 c_S^2 - 4\pi G\rho_b, \tag{2.1.11}$$

which links the pulsation of plane waves to the speed of sound and to the wave vector. This equation has two possible solutions, depending on the sign of ω^2, that discriminate between real and imaginary values. Then, setting $\omega^2 = 0$ and resolving this one for $|\vec{k}|$, we find:

$$k_J = \frac{\sqrt{4\pi G\rho_b}}{c_S}. \tag{2.1.12}$$

This solution is the *Jeans wave number*, i.e the Jeans scale in the Fourier space. In the real space, this scale is:

$$\lambda_J = \frac{2\pi}{k_J} = c_S\sqrt{\frac{\pi}{G\rho_b}}, \tag{2.1.13}$$

that is the same R_J of Eq. (2.1.3), previously found, i.e the *Jeans scale*. With the equation (2.1.13) we can write the dispersion relation (2.1.12) as a function of the Jeans scale but always in the phase space:

$$\omega^2 = k^2 c_S^2\left[1 - \left(\frac{\lambda}{\lambda_J}\right)^2\right]. \tag{2.1.14}$$

Depending on the values of λ, we will have two kind of solutions:

- If $\lambda < \lambda_J \implies \omega^2 > 0$ and the pulsation is real. So, replacing the two real values of ω into the wave form of Eq. (2.1.10), adapted to each variable, we will have a solution whose temporal part is imaginary but with a real phase speed and a spatial part unchanged. Hence these two solutions become *progressive plane waves*, that propagate themselves in opposite directions.
- If $\lambda > \lambda_J \implies \omega^2 < 0$ and the pulsation is imaginary. Thus we obtain two imaginary solutions but, this time, with a real temporal part which modifies the wave amplitude; we obtain *stationary waves with amplitude increasing over time* whose solutions are:

$$\delta(\vec{r},t) = \delta_k \exp(i\vec{k}\cdot\vec{r}\exp(\mp|\omega|t). \tag{2.1.15}$$

These solutions incorporate exactly the phenomenon of the gravitational Jeans instability. The perturbation growth is exponential. However, all these solutions are valid only if $\delta \ll 1$.

2.2 The Jeans theory in expanding Universes

Let us see now how the Jeans theory is applicable in the case of our Universe, that is an expanding Universe (because of the Hubble flow), where the expansion tends to counteract the matter's collapse. In this framework, the background density evolves as a function of time, $\rho_b = \rho_b(t)$ and the velocities are not null but obey the Hubble law: $\vec{v_b} = H(t)\vec{x}$.

2.2.1 The cosmological horizon

Before proceeding we must define the so-called *cosmological horizon* R_H: this is the radius that defines the spherical region of the Universe in *causal connection* with the observers, that is the space traveled by all the photons emitted at a time less than or equal to Hubble time $t_H \approx H_0^{-1}$. Its formal definition is:

$$R_H(t) = a(t)\int_0^t \frac{c\,dt'}{a(t')}. \tag{2.2.1}$$

The cosmological horizon is finite and grows over time due to the expansion of the Universe, hence the causally connected region is finite and grows faster than the speed of light, due to the expansion caused by the Hubble flow. We can define another length, complementary to R_H, called *event horizon* R_E, which measures that region of the Universe that has not yet entered into causal connection with us:

$$R_E(t) = a(t) \int_t^{t_{max}} \frac{c\mathrm{d}t'}{a(t')}, \tag{2.2.2}$$

with t_{max} that is the maximum temporal extension of the Universe.

2.2.2 Applying the Jeans theory in expanding Universes

In the following discussion we must consider three fundamental scales:

- the *horizon scale*, λ_H, which is the scale below which there is causal connection, defined as R_H in Eq. (2.2.1). On scales larger than λ_H the only fundamental interaction is *gravity* (it is necessary to use the general relativity), while on smaller scales the other interactions of the matter also come into play in the collapse process.
- The *Jeans scale* λ_J: perturbations larger than this scale can increase or decrease in time, while perturbations below this scale propagate like plane waves.
- The *dissipation scale* λ_d which is the scale below which every perturbation dissipates because of the propagation of the waves that carry it (this dissipation occurs however separately for dark matter and baryons).

We must also consider two fundamental scale times:

- The *equivalence time* t_{eq}, i.e the moment in which the matter started to dominate on the radiative component and where, therefore, the radiation pressure was no longer able to contrast the gravitational collapse of matter (principally dark matter).
- The *decoupling time*, which is the instant after which the baryons decoupled from the radiation because the scattering rate between these two components became negligible and the free mean path of the photons is practically infinite.

As already said, on scales $\lambda > \lambda_H$ the gravity dominates, so it would be necessary to describe the system with general relativity. The same is also true on scales $\lambda_J \ll \lambda < \lambda_H$. Therefore, we will not make a complete treatment of this situation, but only a qualitative one. Let us consider the perturbations as small spherical ($\Omega > 1$) and over-dense Universes with respect to a larger, flat background Universe (EdS), in which they are immersed. With this approach we can use the Friedmann's equations to obtain perturbations of density dependent on time (as the Universe expands over time):

$$\delta(t) = \frac{3c^2}{8\pi G a^2 \rho_b} \propto a^{-2}(t)\rho_b^{-1}. \tag{2.2.3}$$

Thanks to the solutions to the Friedmann's equations for EdS Universes, shown in table 1.1, from the equation (2.2.3) we obtain that, on scales $\lambda_J \ll \lambda < \lambda_H$, for $t < t_{eq}$ (when the radiation dominated) $\delta_{rad} \propto a^2 \propto t$. Instead, for $t > t_{eq}$ the matter dominates and $\delta_{mat} \propto a \propto t^{2/3}$. Therefore, on scales smaller than the cosmic horizon, but much larger than the Jeans scale, all the components are decoupled and each one follows its own trend.

Now, we can analyse quantitatively the situation on scales $\lambda < \lambda_H$ and in a post-t_{eq} era, i.e our case of interest. In this treatment we shall use two different notations: the physical coordinates $\vec{r}$ and the co-moving ones $\vec{x}$, that are related by: $\vec{r} = a(t)\vec{x}$. The velocity in physical coordinates $\vec{u}$ is linked to that in co-moving coordinates by the relation:

$$\vec{u} = \dot{r} = H(t)\vec{r} + \vec{v}, \tag{2.2.4}$$

where the first term accounts for the Hubble flow and the second term $\vec{v}$ is called *peculiar velocity* of the fluid:

$$\vec{v} = a\dot{x}. \tag{2.2.5}$$

The peculiar velocity describes the perturbations of the velocity field with respect to the Hubble flow, so it is a measure of the velocity component of an object compared to the general component,

due to the expansion of the Universe. $\vec{v}$ is therefore an inhomogeneity of the overall velocity field, generated by a gravitational field acting on the matter. The two-velocity enters the hydrodynamic equations (2.1.6), seen in the statical approach, that now we re-write into the proper reference system (i.e in physical coordinates) as:

$$\begin{cases} \left(\frac{\partial \rho}{\partial t}\right)_{\vec{r}} + \vec{\nabla}_{\vec{r}} \cdot (\rho \vec{u}) = 0 \quad \text{Continuity equation} \\ \left(\frac{\partial \dot{r}}{\partial t}\right)_{\vec{r}} + (\vec{u} \cdot \nabla_{\vec{r}})\vec{u} = -\frac{1}{\rho}\nabla_{\vec{r}} p - \nabla_{\vec{r}}\Phi \quad \text{Eulero's equation} \\ \nabla^2_{\vec{r}}\Phi = 4\pi G\rho \quad \text{Laplace equation.} \end{cases} \tag{2.2.6}$$

We search for solutions to this system for Universes of matter and with the same approach of Section 2.1.2, i.e perturbing a known solution of the background:

$$\begin{cases} \rho = \rho_b + \delta\rho = \rho_b(1+\delta) \\ \vec{u} = H\vec{r} + \vec{v} \\ \Phi = \Phi_b + \delta\Phi \\ p = p_b + \delta p. \end{cases} \tag{2.2.7}$$

The final *perturbed system in physical coordinates* $\vec{r}$ is:

$$\begin{cases} \frac{\partial \delta\rho}{\partial t} + \rho_b \vec{\nabla} \cdot \vec{v} + H\vec{r}\vec{\nabla}\delta\rho + 3H\delta\rho = 0 \quad \text{Perturbed Continuity equation in physical coordinates} \\ \frac{\partial \vec{v}}{\partial t} + H\vec{v} + H\vec{r} \cdot \vec{\nabla}\vec{v} = -\frac{1}{\rho_b}\vec{\nabla}\delta p - \vec{\nabla}\delta\Phi \quad \text{Perturbed Eulero's equation in physical coordinates} \\ \nabla^2\delta\Phi = 4\pi G\delta\rho \quad \text{Perturbed Laplace equation in physical coordinates.} \end{cases} \tag{2.2.8}$$

These are the three Hydrodynamic equations (we have assumed the adiabatic condition) that link the perturbations to the physical space. Some terms are related exclusively to the expansion of the Universe and will therefore not be present in co-moving coordinates. Using the relation $\vec{r} = a\vec{x}$ and the transformation relations between co-moving and physical operators, we re-write the perturbed system in co-moving coordinates $\vec{x}$:

$$\begin{cases} \left(\frac{\partial \delta\rho}{\partial t}\right)_{\vec{x}} + \frac{\rho_b}{a}\vec{\nabla}_{\vec{x}}\vec{v} + 3H\delta\rho = 0 \quad \text{Perturbed Continuity equation in co-moving coordinates} \\ \left(\frac{\partial \vec{v}}{\partial t}\right)_{\vec{x}} + H\vec{v} = -\frac{c_S^2}{a}\vec{\nabla}\delta\rho - \frac{1}{a}\vec{\nabla}_{\vec{x}}\delta\Phi \quad \text{Perturbed Eulero's equation in co-moving coordinates} \\ \frac{1}{a^2}\nabla^2_{\vec{x}} = 4\pi G\delta\rho \quad \text{Perturbed Laplace equation in co-moving coordinates.} \end{cases} \tag{2.2.9}$$

With the equations system (2.2.9), we are able to develop the theory of Jeans for expanding Universes. As in the case of static Universes, we look for solutions in the form of plane waves but, this time, assuming the amplitude of fluctuations no longer constant over time: $f(\vec{x}, t) = f_k(t)\exp(i\vec{k} \cdot \vec{x})$. Then we substitute in the system (2.2.9) the wave forms of the hydrodynamic quantities involved in order to obtain:

$$\begin{cases} \dot{\delta}_k + \frac{ikv_{k,\parallel}}{a} = 0 \\ \dot{v}_{k,\parallel} + \frac{\dot{a}}{a}v_{k,\parallel} = -\frac{ik}{a}(c_S^2\delta_k + \delta\Phi_k) \\ k^2\delta\Phi_k = -4\pi G a^2\rho_b\delta_k. \end{cases} \tag{2.2.10}$$

The equation system just obtained contains the hydrodynamic equations perturbed in the Fourier space for expanding Universes on scales smaller than those of the horizon λ_H and after t_{eq}. We can note that we have used the modules of various vector quantities and, in particular, the parallel (with respect to $\vec{k}$) component of the velocity field that is non-rotational. This was done because the vertical non-divergent component of $\vec{v}$ decays rapidly during the expansion. After some manipulations of the system (2.2.10) we obtain a second order differential equation that describes the temporal evolution of the density field within an expanding Universe in physical coordinates, i.e the *dispersion relation in physical coordinates*:

$$\ddot{\delta}_k + 2\frac{\dot{a}}{a}\dot{\delta}_k + (k^2c_S^2 - 4\pi G\rho_b)\delta_k = 0 \tag{2.2.11}$$

where the second term takes into account the Hubble flow and depends on the cosmological model assumed through the Hubble parameter $\frac{\dot{a}}{a} = H(t)$, while the term $k^2c_S^2\delta_k$ takes into account the characteristic velocity field of the fluid under examination. Hence, the relation (2.2.11) is the general equation, valid for every cosmological model, to describe the growth of the perturbations of the density field.

2.3 Evolution of primordial density fluctuations

As described in Section 1.8, every kind of Universe approaching to the Big Bang behaves as flat, one-component EdS Universe. Therefore, to be able to analyse the evolution of the primordial density fluctuations, we have to solve the dispersion relation (2.2.11) for EdS Universes of dust. Moreover, the theory we are going to develop will be completely linear, as the initial perturbations start very small (we see it from the anisotropies of the CMB). In Section 2.4, we will discuss the non-linear evolution of the fluctuations at low redshits, where the linearity breaks down.

We are under the following conditions: on scales $\lambda < \lambda_H$, at an epoch $t > t_{eq}$ and in a flat Universe ($\Omega = \Omega_m = 1$). Referring to the solutions of the Friedmann's equations in table 1.1, we know that for a flat dusty Universe: $\rho_b = \frac{1}{6\pi G} t^{-2}$ and $H(t) = 2/3t^{-1}$ hence, replacing these values of ρ_b and $H(t)$ into the equation (2.2.11), we obtain the equation of the temporal evolution of δ_k:

$$\ddot{\delta_k} + \frac{4}{3}\frac{\dot{\delta_k}}{t} + \delta_k\left(k^2 c_S^2 - \frac{2}{3t^2}\right) = 0. \tag{2.3.1}$$

Furthermore, we will look for solutions in the form of a power law:

$$\delta_k \propto t^{\alpha}, \tag{2.3.2}$$

thus, replacing the solution (2.3.2) into equation (2.3.1), we obtain the *dispersion relation for EdS Universes*:

$$3\alpha^2 + \alpha + \frac{k^2 c_S^2}{2\pi G \rho_b} - 2 = 0, \tag{2.3.3}$$

which is a second degree equation in α that admits two solutions. The only variable able to discriminate between real and imaginary values is k_J, that is the Jeans scale in the Fourier space or, similarly, the respective scale in real space:

$$k_J = \frac{5}{c_S}\sqrt{\frac{\pi G \rho_b}{6}} \quad \text{from which: } \lambda_J \equiv \frac{2\pi}{k_J} = \frac{c_S}{5}\sqrt{\frac{24\pi}{G\rho_b}}. \tag{2.3.4}$$

The solutions to equation (2.3.3) are:

$$\alpha_{1,2} = \frac{-1 \pm 5\sqrt{1 - \left(\frac{\lambda_J}{\lambda}\right)^2}}{6}, \tag{2.3.5}$$

from which we see that the perturbations will not grow exactly as a power law, since the scale λ_J appears in their exponent. The solutions to the equation of temporal evolution of δ (2.3.1) will be found replacing $\alpha_{1,2}$ into the general power law form (2.3.2).

- For $\lambda < \lambda_J$, α_1 and α_2 are imaginary, hence $\delta_{k,\pm} \propto t^{i\alpha_{1,2}}$. Under the Jeans scale, the perturbations propagate while remaining sound waves.
- For $\lambda > \lambda_J$, α_1 and α_2 are real, hence $\delta_{k,\pm} \propto t^{\alpha_{1,2}}$ and the wave form solutions to the dispersion relation are: $\delta_{\pm}(\vec{x}, t) = \delta_{k,\pm} \exp(i\vec{k}\cdot\vec{x})$ with amplitude varying in time. If we have $\lambda \gg \lambda_J$, as in Universes EdS $t \propto a^{3/2}$, we will have the two solutions:

$$\begin{cases} \delta_-(\vec{x}, t) \propto t^{-1} \propto a^{-3/2} & \text{decreasing solution} \\ \delta_+(\vec{x}, t) \propto t^{2/3} \propto a & \text{growing solution.} \end{cases} \tag{2.3.6}$$

The first solution decays and does not give rise to the formation of structures, while the second one leads to gravitational instability. So the latter is the only cosmologically relevant solution (i.e. $\delta_+(\vec{x}, t)$).

In the case of EdS Universes, when perturbations can grow, they do it growing proportional to the expansion factor a. In comparison to the static Universes, where the perturbations grow with an exponential law, in expanding Universes the growth is slower because perturbations are affected by a negative feedback due to the Hubble flow. At a first approximation, the solution (2.3.6) is valid for every type of matter. However, the dark matter decouples from the radiation before the baryons, hence from the moment of decoupling, the various $\delta_{+,DM}$ will evolve following the trend

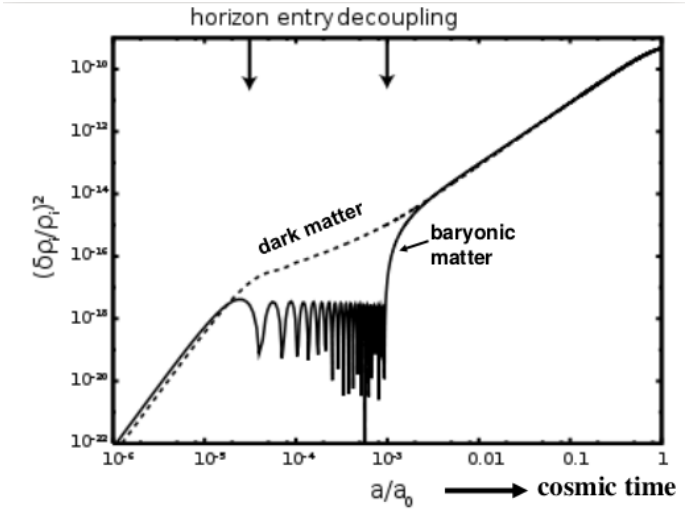

Figure 2.1: The time evolution of the density contrast of baryonic (solid line) and dark (dashed line) matter. The growth of baryon perturbations after decoupling tends asymptotically to that of dark matter. *Credits*: A. Cimatti, slides of "Formation and Evolution of Galaxies", A.A. 2016/2017, Alma Mater Studiorum - Università di Bologna

seen so far. When also the baryons are decoupled from the primordial fluid, the δ_{DM} will have already grown a bit and the result will be that the baryons perturbations δ_b will not evolve freely, but it will evolve following the gravitational potential wells of the already formed dark matter haloes. Consequently, the growth of the baryon's overdensities will be more rapid than $\delta_+ \propto a$. To be precise: $\delta_b \propto \delta_{DM}\left(1 - \frac{a_{dec}}{a}\right)$. This phenomenon is called *baryon catch up* (see Figure 2.1). The baryon fluctuations measured to the present day are more pronounced than what we would have predicted without the existence of dark matter: they should have been of the order of 10^{-2}, while in galaxy clusters they reach the order of 10^3. As the look-back time grows (going back in cosmic time) we expect, in general, a trend of the Universe with overdensities of minor amplitude, that is we must observe less galaxies and galaxy clusters.

For a flat Universe with a cosmological constant, the growing instability solution to the differential equation (2.2.11) can be written in integral form (as a function of redshift through $H(z)$):

$$\delta_+(z) = H(z) \int_z^\infty \frac{(1+z')\mathrm{d}z'}{a_0 H^3(z')}. \tag{2.3.7}$$

For other equations of state of DE, or for more components (curvature, radiation), no general integral expression exists, and the differential equation must be solved directly. We can provide an approximate formula to parametrize the previous findings, the so-called *growth factor* f, as a function of the density parameter evaluated at $z = 0$:

$$f(\Omega_0) \equiv \frac{\mathrm{d}\log\delta_+}{\mathrm{d}\log a} \approx \Omega_{m,0}^{0.55} + \frac{\Omega_\Lambda}{70}\left(1 + \frac{1}{2}\Omega_{m,0}\right). \tag{2.3.8}$$

This formula shows that Λ does not play a crucial role for the growth of the fluctuations. The value of the exponent γ depends on the gravity theory. In particular, $\gamma \approx 0.55$ in General Relativity. Constraining the value of γ is key to test the gravity theory on cosmological scales (for further information about this method, see: Coles and Lucchin, 2002). In this context, we did not analyse the solutions of the growth of the perturbations of non-flat Universes (for an analysis of these models we refer to Peebles, 1980).

2.4 The evolution theory in the linear regime

It is thought that the seeds of density perturbations are produced at the end of the inflationary period[1]. In the final oscillatory stages of the inflation, the density fluctuations δ are generated with

[1]The *inflation theory* describes the early Universe during its exponential expansion just after the Big Bang (as theorized by A. Guth in Guth, 1981)

a stochastic and quasi-Gaussian (it is a truncated Gaussian distribution because $-1 \ll \delta < \infty$) distribution. The density perturbations are defined as follows: $\delta(\vec{x}) = \frac{\rho(\vec{x})-\rho_b}{\rho_b}$. If we measured their value in different positions of the Universe, we would have that the probability to have a perturbation with δ value at a certain point, $P(\delta)$, is a Gaussian distribution. Note that this argument is currently valid on very large scales while, on galactic or cluster scales, we would not observe a Gaussian distribution. Furthermore, the $P(\delta)$ field grows spontaneously thanks to the homogeneity and isotropy of the Universe. Being $P(\delta)$ Gaussian, it is defined only by its average $< \delta(\vec{x}) >$ and variance $< \delta^2(\vec{x}) >$. However, by definition, the average of a stochastic fluctuation is null, hence the only value that defines this field is the variance.

To have a good statisticF, we would need a sufficiently large sample (i.e we would like to observe the phenomenon many times): it is not the case of our Universe, because the sample consists of only one element (our Universe itself) and we are not able to reproduce another Universe empirically, of course. To solve the problem we may invoke the *ergodic hypothesis*: the average between multiple volumes is equivalent to the spatial average between sufficiently large, and not overlapping, sub-volumes of a single volume. This hypothesis means that, thanks to the existence of the CP, we can consider our Universe as a group of many sub-Universes. It is based on the fact that sufficiently large regions of the Universe, and sufficiently separated, are statistically independent from each other[2].

Nevertheless, these sub-volumes will not stay independent long, because of the gravity action. So, it is convenient to describe a generic fluctuation as a superposition of plane waves in Fourier space, that have the advantage to be independent during the evolution. This approach requires a statistical treatment of the initial perturbations. The spatial Fourier transform of the $\delta(\vec{x})$ is:

$$\delta(\vec{k}) \equiv \frac{1}{(2\pi)^3} \int \delta(\vec{x}) \exp(-i\vec{k} \cdot \vec{x}) \mathrm{d}^3\vec{x}, \tag{2.4.1}$$

which has the dimensions of a volume (instead $\delta(\vec{x})$ is dimensionless).

2.4.1 The correlation function and the power spectrum

In statistical mechanics, the correlation function is a measure of the order level of a system that describes how the microscopic variables of a system are connected to each other. In cosmology, it measures how the fluctuations δ co-vary on average across the spacetime. Specifically, the *two-point correlation function* (2PCF) is defined as follows:

$$\xi(\vec{r}) =< \delta(\vec{x})\delta(\vec{x} + \vec{r}) >, \tag{2.4.2}$$

that is the average of the product between two values of the field of fluctuations, taken between two points at co-moving distance $|\vec{r}| = r$. This function quantifies the spatial clustering of the objects. If we define $\xi(\vec{r})$ as the Fourier anti-transform of the function:

$$< \delta(\vec{k})\delta^*(\vec{k}') >= (2\pi)^3 P(k)\delta_D^3(\vec{k} - \vec{k}'), \tag{2.4.3}$$

with $P(k)$ called the *power spectrum*, and $\delta_D(\vec{k} + \vec{k}')$ the delta of Dirac (we use $-k'$ because $\delta^*(k) = \delta(-k)$), we can rewrite it in the form:

$$\xi(\vec{r}) = \frac{1}{(2\pi)^3} \int P(k) \exp(i\vec{k} \cdot \vec{r}) \mathrm{d}^3\vec{k}. \tag{2.4.4}$$

Equation (2.4.4) expresses the *Wiener-Khintchine theorem*: the power spectrum $P(k)$ is the equivalent, in the Fourier space, of the correlation function $\xi(\vec{r})$ in the real space[3].

We note that $P(k) \propto < \delta(k)\delta(k') > \propto < \delta^2(\vec{k}) >$, hence the power spectrum tells us what is the mean quadratic measure (or variance), through a normalisation factor, of the amplitude of a perturbation with wave number $\vec{k}$, in the Fourier space.

[2] In this sense, the ergodic hypothesis, together with the CP, is called the Fair Simple hypothesis.

[3] The power spectrum has the dimension of the inverse of a volume, instead the correlation function is dimensionless. The field $\delta(\vec{x})$ is a real quantity, while its Fourier transform $\delta(\vec{k})$ is imaginary.

2.4.2 Punctual variance of the field and mass variance

The variance of a Gaussian field in three dimensions is defined as the mean quadratic value of the field itself and, in the case of the real density field $\delta(\vec{x})$, this variance is given by $\sigma^2 =< |\delta(\vec{x})|^2 >=< \delta^2(\vec{x}) >$. Applying the ergodic hypothesis, we can divide the Universe into a large number of independent V_∞ volumes[4] and compute σ^2 as a quadratic spatial mean of δ, within each volume and averaged across all volumes. So, the definition of power spectrum (2.4.3) helps us to define the *punctual variance of the field*:

$$\sigma^2 = \frac{1}{(2\pi)^3}\int_{-\infty}^{+\infty} P(k)\mathrm{d}^3\vec{k} \quad \text{or } \sigma^2 = \frac{1}{(2\pi)^3}\int_{-\infty}^{+\infty} k^2 P(k)\mathrm{d}\vec{k} \text{ (in spherical coordinates).} \tag{2.4.5}$$

It can be noted that σ^2 does not depend on spatial position but only on time. The variance is the integral of the power spectrum over the whole Fourier space (while it is called punctual because it is such in real space), save for a factor $(2\pi)^{-3}$. Thus, the variance is an integral information on the power spectrum.

However, measuring the density field of the Universe is not trivial because, from an observational point of view, we can only get information on over-dense regions that emit radiation. So the idea will be to count the galaxies (or clusters of galaxies) filling a given volume of the Universe. Anyway, this will introduce uncertainties mainly due to the fact that density is a continuous and punctually defined function while the distribution of objects in the Universe is a discrete quantity. Let us define the *fluctuations on the average number of galaxies within a given volume* V:

$$\delta_g = \frac{N_g(V) - < N_g(V) >}{< N_g(V) >}. \tag{2.4.6}$$

Since we want information about the mass distribution, and since nothing assures us that galaxies are the ideal tracers of matter distribution, we suppose for simplicity to have a distribution of identical galaxies, and correct δ_g for a *bias parameter* b:

$$\delta_g = \frac{N_g(V) - < N_g(V) >}{< N_g(V) >} = \frac{M_g(V) - < M_g(V) >}{< M_g(V) >} = b\delta_M. \tag{2.4.7}$$

The bias b contains all our ignorance about the formation and evolution of the galaxies used as tracers (that is the uncertainty about the percentage of dark matter associated with the different tracers). We will assume $b = const$ and the existence of a linear relationship between the number of galaxies and the total mass, contained in a volume (it is an approximation to the first order).

We can move from δ_M to the usual δ, carrying out a convolution that filters the field using a suitable *window function*. Usually, a *top-hat filter* is preferred, defined as a function constant inside a sphere with radius R_{TH} and null out of this radius. It is defined as:

$$W(\vec{x}, R_{TH}) = \frac{3}{4\pi R_{TH}^3}\theta(1 - \frac{|x - x'|}{R_{TH}}), \tag{2.4.8}$$

with $\theta(1 - \frac{|x-x'|}{R_{TH}})$ representing the Heavyside function.

The following function (2.4.9) links δ_M with δ:

$$\delta_M(\vec{x}) = \delta(\vec{x}) \otimes W(\vec{x}, R_{TH}). \tag{2.4.9}$$

We can note that the top-hat filter is a *low-pass filter* hence, with this approach, we lose every information on scales below R_{TH}. Also, δ_M will have a Gaussian distribution described only by variance, as it measures stochastic fluctuations, therefore with zero mean. To determine the Gaussian form of δ_M we define the mass variance:

$$\sigma_M^2 \propto < \delta_M^2 >=< \left(\frac{\delta M}{\bar{M}}\right)^2 >, \tag{2.4.10}$$

where we deliberately omit the bias. Applying the Parseval theorem along with that on the convolution, and remembering that $< |\delta^2(\vec{k})| >= P(k)V_\infty$, we obtain the complete form of the *mass variance*:

$$\sigma_M^2 = (2\pi)^{-3}\int_0^\infty < P(k) > \hat{W}^2(\vec{k}, R)\mathrm{d}^3\vec{k} \tag{2.4.11}$$

[4]From equation (2.4.3) we can note that $< \delta(\vec{k})\delta^*(\vec{k}') >= (2\pi)^3 P(k)\delta_D^3(\vec{k} - \vec{k}') = P(\vec{k})V_\infty$ hence the power spectrum is proportional to the mean quadratic value of the amplitudes of the fluctuations into the Fourier space.

that, in spherical coordinates, is:

$$\sigma_M^2 = \frac{1}{2\pi^2}\int_0^\infty k^2 < P(k) > \hat{W}^2(\vec{k}, R)\mathrm{d}\vec{k}, \tag{2.4.12}$$

where $\hat{W}$ is the Fourier-trasform of the window function, which is a function of R, and therefore of M. Note that $\sigma_M^2 \leq \sigma^2$. The main contribution to σ_M^2 is given by perturbations with $\lambda \approx k^{-1} > R$, because those with higher frequencies tend to be averaged out within the window volume. The Zel'dovich spectrum that we will introduce in Eq. (2.4.13) is a growing function of k, so waves with much larger λ contribute only a small amount. Furthermore, σ_M grows over time.

2.4.3 The primordial Zel'dovich spectrum

The shape of the initial spectrum is assumed to be imprinted in the Universe at some arbitrarily early time. Since the accepted inflation scenario states that the seeds of today's structures are generated by stochastic quantum fluctuations in a scalar field (i.e the *inflaton*), the initial $P(k)$ should produce a *power-law* form[5]:

$$P(k) = Ak^n, \tag{2.4.13}$$

with A is the amplitude, and the *spectral index* n is generally assumed to be close to unity, $n \approx 1$, if we assume that the inflation is the moment in which the fluctuations in the metric of the gravitational potential, which lead to the perturbations of the density field, were born (for a complete explanation, see Zeldovich, 1972). While the shape of the power spectrum may be fixed by these considerations, the amplitude is not and the gap should be plugged with observations. Nowadays the most reliable measure of A is obtained from the temperature fluctuations in the CMB.

Considering fluctuations in the gravitational potential ϕ on the scale R, we have $\phi_R \propto G\delta_M/R \propto \delta R^2 \propto \sigma_M R^2 \propto \sigma_M M^{2/3}$, because of the equation (2.4.10). Considering the *non-linear scaling relation* (in a $\delta \geq 1$ regime), observed and deduced by virialized and collapsed structures (i.e the galaxy clusters) assuming that gravity is the only responsible of collapse, one derives that if at some time t_i there is a density fluctuation $\delta(\vec{x}, t_i)$, its variation in linear regime does not depend on the scale and grows as a, or rather $\delta(\vec{x}, t) = \delta(\vec{x}, t_i)a(t)$ (For further information consult Giodini et al., 2013). So for the power spectrum we will have that:

$$P(k, t) = P(\vec{x}, t_i)a^2(t). \tag{2.4.14}$$

This means that whatever the power spectrum is, it grows in time in the same way, proportionally to $a^2(t)$. This is valid in an EdS Universe. Generalising, one has:

$$P(k, t) = P(\vec{x}, t_i)\delta_+^2(t). \tag{2.4.15}$$

So, regarding the mass variance, we have that: $\sigma_M^2 \propto \delta_+^2(t)k^{n+3}$. When $\delta \to 1$, $\sigma_M^2 \to 1$ whereby when the structures collapse, the mass variance tends to unity. Because $k \propto R^{-1} \propto M^{-1/3}$ we have:

$$\sigma_m^2 \propto \delta_+^2(t)M^{-(n+3)/3}. \tag{2.4.16}$$

An object of a given mass will form when $\delta_+^2(t)M^{-(n+3)/3} \propto 1$ that is, a mass collapses when reaches the value:

$$M_\star \propto \delta_+^{6/(n+3)}. \tag{2.4.17}$$

From the last relation (2.4.17) it emerges that the collapsing mass of the structures grows with time. The fact that small mass structures form before those of larger mass depends on the slope $-(n+3)/3$. For an EdS Universe we will have:

$$M_\star \propto a^{6/(n+3)} \propto (1+z)^{-6/(n+3)}. \tag{2.4.18}$$

Finally, from these relationships we obtain:

$$\phi_R \propto M^{(1-n)/6}, \tag{2.4.19}$$

[5] At the time of inflation there are no preferred scales (because the gravity, which is the only dominant force field at that time, does not hold preferred scales) and a functional form that respects this property is precisely the power law.

from which it can be deduced that, in order to have the fluctuations of the gravity potential with no scale dependence, it is necessary that $n = 1$ (Zel'dovich spectrum). As already mentioned, these relationships are valid in the case of the formation of structures is governed solely by gravity: this assumption is valid for very large structures such as clusters, but not for galaxies, for which, due to all non-gravitational processes connected to the physics of the gas, the assumption is no longer valid. In order to have a bottom-up hierarchical structure formation scenario it is necessary that the formation time is an increasing function of the mass and that the binding energy of the system (proportional to the average quadratic velocity) is an increasing function of the mass, because we want the thermal contribution to be due exclusively to the adiabatic response to the growth of the structure.

Another fundamental property of the Zeldovich spectrum is the fact that it corresponds to the spectrum for which the perturbations have the same height when they enter the event horizon a_H; in fact, the mass variance in this moment is:

$$\sigma_M(a_H) \propto M_H^{(1-n)/6}. \tag{2.4.20}$$

In other words, if the primordial spectrum is that of Zel'dovich, the density perturbations enter the event horizon at different times but all with the same amplitude, because $\sigma_M = const.$, as consequence of the stagnation phenomenon. The perturbations can not grow before the equivalence between radiation and matter, due to the so-called *Meszaros effect* or stagnation: the radiation, due to its high pressure, does not allow the growth of the matter perturbations, that are *frozen* (for a detailed description of this phenomeon, consult Coles and Lucchin, 2002). However, if $-3 < n < 1$ we will have hierarchical clustering (in agreement with the bottom-up scenario of the ΛCDM model).

2.5 The evolution theory in the non-linear regime

After the matter-radiation equivalence, the δ matter density fluctuations on scales larger than the Jeans mass in this era $M_J(a_{eq})$ can grow because the stagnation effects disappear. So far we have used linear approximation (valid as long as $\delta \ll 1$) but, in the present Universe, there are densities of matter, in particular the clusters, where $\delta \approx 10^{2-3}$. In order to study the evolution of structures we must develop an adequate non-linear theory, much more complex than the linear one (which provides analytical solutions) and that, excluding the analytical and ideal solution of spherical collapse, requires a numerical treatment.

2.5.1 The spherical collapse

The simplest approach to non-linear evolution is to study fluctuations with a simple geometric symmetry. There is an *analytical solution* (called spherical collapse Tolmann's solution) in the theory of non-linear perturbations that assumes a *perfectly spherical perturbation and with null initial co-moving speed* $v_i = 0$. For simplicity, we will assume an EdS background Universe (but the solutions can be obtained in any type of Universe). From equation (2.3.6) we have that the evolution of the perturbations in a linear regime follows the equation:

$$\delta = \delta_+(t_i)\left(\frac{t}{t_i}\right)^{2/3} + \delta_-(t_i)\left(\frac{t}{t_i}\right)^{-1}. \tag{2.5.1}$$

The co-moving collapse velocity, in the EdS Universes, is: $v_{com} = \frac{i\dot{\delta}}{ik_i\left(\frac{t}{t_i}\right)^{-2/3}}$, thus the relation between the proper velocity and $\dot{\delta}$, $v = \frac{i\dot{\delta}}{k}$, provides us the *evolution of the speed of collapse in linear regime*:

$$v = \frac{i}{k_i t_i}\left[\frac{2}{3}\delta_+(t_i)\left(\frac{t}{t_i}\right)^{1/3} - \delta_-(t_i)\left(\frac{t}{t_i}\right)^{-4/3}\right]. \tag{2.5.2}$$

We have imposed the condition that $v_i = v(t_i) = 0$, assuming the initial instant $t_i = 0$, thus, replacing those two values into equation (2.5.1), we obtain that: $\delta_-(t_i) = \frac{2}{3}\delta_+(t_i)$ and we can rewrite equation (2.5.1) as:

$$\delta(t_i) = \frac{5}{3}\delta_+(t_i). \tag{2.5.3}$$

In other words, of all the initial perturbation $\delta(t_i)$ only three fifths of it will grow and the other two fifths will decrease, so that the perturbation will grow over time according to the equation:

$$\delta(t) = \frac{3}{5}\delta_+(t_i)\Big(\frac{t}{t_i}\Big)^{2/3}. \tag{2.5.4}$$

The perturbation is assumed to be spherical, so that it can be treated as a Universe that is described by Friedmann's laws. Since that Friedmann's models predict that only closed Universes can collapse, we impose on the perturbation that $\Omega_p(t_i) = \Omega_{bkg}(t_i)(1+\delta) > 1$, thus obtaining the threshold condition for collapse:

$$\delta_+ > \frac{3}{5}\frac{1-\Omega_{bkg}(t_i)}{\Omega(t_i)(1+z_i)}. \tag{2.5.5}$$

From this report we find that, in closed and flat background Universes, any overdensity can collapse; instead, in the case of open Universes, the perturbation must exceed a certain threshold in order to collapse. The density of the perturbation will reach its maximum at the *turn-around time* t_{max} after which the collapse begins

$$t_{max} = \sqrt{\frac{3\pi}{32G}\rho_p^{-1}(t_{max})} \tag{2.5.6}$$

and with:

$$\rho_p(t_{max}) = \frac{3\pi}{32G}t_{max}^{-2}\,, \tag{2.5.7}$$

at this moment the perturbation is *detached* from the Hubble flow of the background Universe to reverse trend and collapse. It can be estimated that, at the turn-around time, the perturbation is already non linear: $\delta(t_{max}) \approx 4.6$. Note that, if we had used a linear approach we have obtained that $\delta_{lin}(t_{max}) \approx 1.07$, because the linear theory gives solutions poorly dependent on the assumed cosmology.

The simulations show that the perturbation ceases to collapse, reaching a physical dimension of equilibrium R_e, at the time $t_e \approx 2t_{max}$. After this last moment, the overdensity begins to stabilize, oscillating around its own R_e, until it is virialized to $t_{vir} \approx 3t_{max}$. Between t_e and t_{vir} the value of the perturbation δ remains unchanged and it is the background that instead expands in this period of time, decreasing its density as a^{-3}. For this reason, the density contrast between the overdensity and the background universe reaches up to about 400, to virialization. It should be noted that, despite the sphericity of the collapse being very far from reality, similar results are obtained assuming an ellipsoidal symmetry collapse. Moreover, further into the non-linear regime, bound structures form. In this stage, the baryonic content in these objects becomes important and the simple treatment presented in the previous Sections is no longer correct.

2.5.2 The mass and luminosity functions

In order to be able to test models of the evolution of cosmic structures, we must link these to observables: this can be done e.g. with the mass function of cosmic structures $n(M)\mathrm{d}M$, which gives us the number of objects with a mass between M and $M + \mathrm{d}M$ at a given time t per unit of volume. Let us see now how to get to define this observable. We have said that the Universe is born with a quasi-Gaussian distribution of density fluctuations but, while such fluctuations evolve, the peak of the original Gaussian distribution will be lowered as it increases the variance (over time) σ_M. Reasoning in terms of mass rather than density (we use the function $W(\vec{x}, R)$), for our Universe the Gaussian distribution is:

$$P(\delta_M)\mathrm{d}\delta_M = \frac{1}{\sqrt{2\pi\sigma_M^2}}\exp\Big(-\frac{\delta_M^2}{2\sigma_M^2}\Big)\mathrm{d}\delta_M, \tag{2.5.8}$$

which represents the probability of having a mass fluctuation between δ_M and $\delta_M + \mathrm{d}\delta_M$. Over time the $\delta_M(t)$ and $\sigma_M(t)$ values vary along with the increasing mode:

$$\begin{cases}\delta(t) = \delta_i\delta_+(t)\\ \sigma_M^2(t) = \sigma_M^2(t_i)\delta_+^2(t)\end{cases} \tag{2.5.9}$$

and their distribution remains Gaussian as long as we are in the linear regime. We expect structures to be formed where there will be overdensities. We therefore find the probability that a fluctuation

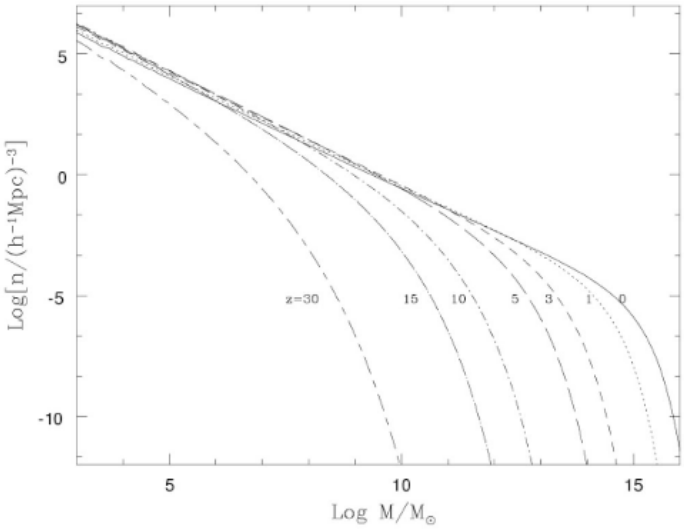

Figure 2.2: The dark matter halo mass function at different redshifts. *Credits:* caltech

δ_M exceeds a threshold value δ_c: $P_{>\delta_c}(M) = \int_{\delta_c}^{\infty} P(\delta_M)\mathrm{d}\delta_M$. This quantity is a function of mass, which in turn depends on the W filter used to change from δ to δ_M (see equation (2.4.9)). The idea of Press and Schechter (Press and Schechter, 1974) was to take as a threshold value the one found in the spherical collapse model at the equilibrium time of the structure t_e (expressed by Eq. (2.5.1)) but into a *linear approach* (found for an EdS background Universe but still valid for every cosmological model, depending poorly on the model adopted):

$$\delta_+(t_e) = \delta_c = 1.686 \tag{2.5.10}$$

We can suppose that all the fluctuations of the Gaussian distribution, now evolved quite a bit, at a certain moment have values $\delta_M > 1.686$ and can collapse and form the structures. The following is the *Press and Schecter mass function*:

$$n(M,z)\mathrm{d}M = \sqrt{\frac{2}{\pi}}\frac{\delta_c\bar{\rho}_M(z)}{\sigma_M(z)M^2}\left|\frac{\mathrm{d}ln\sigma_M}{\mathrm{d}lnM}\right|\exp\left(-\frac{\delta_c}{2\sigma_M^2(z)}\right) \tag{2.5.11}$$

which contains all the necessary cosmological information. The form of this distribution function will be that of a power-law, with an exponential cut from a certain value of the mass.

Assuming a power-law form for the power spectrum as in Eq. (2.4.13), where now the spectral index is a function of k (n=n(k)), and the variance is in the form $\sigma_M = \left(\frac{M}{M_0}\right)^{-\alpha}$, with $\alpha = \frac{n+3}{6}$ (as in (2.4.17)), we can simplify the term in the equation (2.5.11) as: $\frac{\mathrm{d}ln\sigma_M}{\mathrm{d}lnM} = \frac{M\mathrm{d}\sigma_m}{\sigma_M\mathrm{d}M} = \alpha$, so that the *Press and Schecter mass function* becomes:

$$n(M)\mathrm{d}M = \sqrt{\frac{2}{\pi}}\frac{\alpha\bar{\rho}_M}{M_*^2}\left(\frac{M}{M_*}\right)^{\alpha-2}\exp\left[-\left(\frac{M}{M_*}\right)^{2\alpha}\right]\mathrm{d}M, \tag{2.5.12}$$

where we use the *characteristic mass* (made entirely of constants but increasing over time together with σ_M because dependent on n, through α):

$$M_* = M_0\left(\frac{2}{\delta_c^2}\right)^{1/2\alpha}, \tag{2.5.13}$$

with M_0 the normalisation mass value. In Figure 2.2 we can note the cut-off of the function (2.5.12) for $M \approx M_*$ (also-called *knee*), where the primordial power-law power spectrum, with spectral index $n \approx 1$, evolves assuming an exponential form. It is simple to understand the reason of this passage looking at the two factors (the left power-law and the right exponential) of the mass function (2.5.12). At the present time, it is found that $M_* \approx 10^{15}\,, M_\odot/h$, which is the mass of a typical cluster of galaxies. Hence counting the number of clusters in the local Universe is equivalent

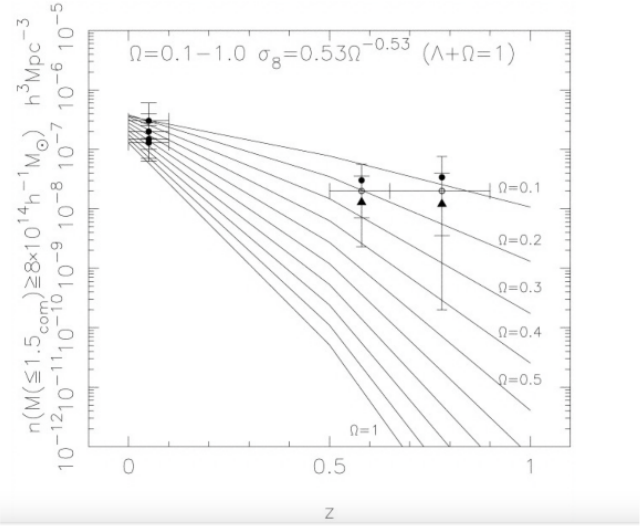

Figure 2.3: Measured number density of massive galaxy clusters, at three different redshifts, compared to Press-Schechter predictions for different values of the matter density contrast, as indicated by the labels. *Credits: A lightweight universe?*, Bahcall, Neta A.; Fan, Xiaohui, Proceedings of the National Academy of Sciences of the United States of America, Vol. 95, No. 11, 26.05.1998, p. 5956-5959.

to counting how much mass is contained in the exponential tail of the mass function and, finally, to placing constraints on two degenerate cosmological parameters: ρ_M (and then Ω_M) and the amplitude of perturbations A (see equation (2.4.13)). We focus on the objects of the exponential tail (i.e on the most massive objects) because they are objects with a higher and easily detectable emission and, above all, because they are extremely sensitive objects to cosmology, even if they are affected by very large errors due to the low statistics. Furthermore, this counting, made at different redshifts, allows to make speculations on the ΛCDM model assumed because as we grow up z we will have less and less objects of large mass ($n(M)$ drastically decreases after the cut off at M_*). For local clusters one has: $\sigma_8 \approx (A = 0.5, 0.6) * \Omega_{0,M}^{-0.5,-0.8}$ (Mandelbaum et al., 2013).

In Figure 2.3 the trend of the Press-Schechter mass function is shown as a function of redshift when the density parameter changes. Operationally, the position of the knee, and therefore of the exponential tail, will depend on the variance that we will write as:

$$\sigma_8^2 = (2\pi)^{-2} \int_0^\infty P(k) k^2 W(k, R_8) \mathrm{d}k = A(2\pi)^{-2} \int_0^\infty k^{n+2} T^2(k) W(k, R_8) \mathrm{d}k \tag{2.5.14}$$

with:

$$P(k) = A k^n T^2(k). \tag{2.5.15}$$

The term that heavily influences the variance is the normalisation A of the spectrum (the amplitude of the fluctuations) that assumes constant values in various periods of the Universe.

The mass function is analogous to the *Schechter luminosity function*:

$$\Phi(L) = A L^\beta \exp\Big(-\frac{L}{L_*}\Big), \tag{2.5.16}$$

with A the normalisation, $\beta = \alpha - 2$ the slope and L_* the *knee* luminosity (i.e the luminosity associated with objects with a characteristic mass M_*). For further information see the Subsection 3.2.1. The luminosity (Eq.(2.5.16)) and the mass (Eq.(2.5.12)) functions have exactly the same shape, so L_* marks the transition from the power-law to the exponential form. The Schechter luminosity function finds its theoretical justification, in fact, in the Press and Schechter mass function but is the one that actually is used in a first analysis, because L is a direct observable, contrary to the mass that must be inferred. Assuming a constant mass-luminosity ratio $\frac{M}{L} = const.$, we can link the mass function $n(M)$ to the luminosity one $\Phi(L)$ according to the relation:

$$\Phi(L) = n(M) \frac{\mathrm{d}M}{\mathrm{d}L}. \tag{2.5.17}$$

With reference to Figure 2.3, the galaxy clusters describe the part of the mass function corresponding to the high masses (the exponential tail). Smaller masses correspond to haloes hosting single galaxies, characterised by $\beta \approx 1 (\Longrightarrow \alpha = \beta + 2 \approx -1)$. Because $\alpha = (n+3)/6$, for galaxies we have $n \approx 3$, but to have hierarchical clustering it must be valid that $-3 < n < 1$. The fact that the matter has a Press-Schechter distribution which does not match the Schechter function of luminosity, means that there is an excess of dark matter haloes that are not bright. The basic issue is that the baryons are collisional and therefore it is not correct to assume $M/L = const.$ for galaxies while, for clusters, it is a valid assumption because their potential wells are so deep that they suffer much less from the energy of baryons.

Chapter 3

Clusters of galaxies

Galaxies are not evenly distributed in the Universe, but tend to cluster in groups or clusters of galaxies. In this thesis work we will use galaxy clusters as cosmological probes, in a way that will be clear later when we will analyse our data. In this Chapter and in the following one we will introduce the physical bases of our analysis. Specifically, this section will be focused on the main properties of galaxy clusters which, although not all of them necessary for our thesis work, will be quoted for completeness.

Galaxy clusters represent the largest gravitationally and virialised bound systems in the Universe: this makes them important in cosmology as they trace the highest density peaks of large-scale structures (see for instance Subsection 2.5.2). Clusters extend for about $1 - 4$ Mpc. Nevertheless, our final goal will be to study their extension far beyond the viral radius, thanks to the lensing distortion that they cause on the background galaxies, by means of the so-called 2-halo term. They have brightness in the optical ($400 - 700$ nm) and X-ray ($1 - 100$ keV) bands of $10^{46} erg/s$ and $10^{43-45} erg/s$, respectively. Clusters with masses of the order of $10^{13} M_{\odot}$ contain a small number of galaxies and are called *groups*. Instead, the biggest clusters reach masses of the order of $10^{15} M_{\odot}$ and contain up to about 1000 galaxies. Note that we are talking about the total mass: baryons plus dark matter. Although clusters are made up of a large number of galaxies, these ones contribute only to 5% to the total mass, which is composed of dark matter as dominant component (80%), and intracluster medium ($ICM, 15\%$), as evidenced by the gravitational lensing and by the X-ray emission of the ICM gas, respectively.

In this Chapter, we will present an overall description of the main cluster properties, including the various classification methods, the physical properties (luminosity function, spatial distribution and mass), the aspects related to X-ray observations and the main scaling relations that characterise the clusters. Finally, we will talk about various methods for identification of galaxy clusters from optical surveys. We will focus mainly on a powerful algorithm for cluster detection in optical surveys, called AMICO Bellagamba et al. (2018), explaining its main properties and detecting method, and how it was used by Maturi et al. (2018) to collect the cluster sample used in the analysis performed in this thesis work.

3.1 Classification of galaxy clusters

George Ogden Abell built a catalogue of galaxy clusters identifying areas of sky with overdensities of galaxies. He counted all the galaxies within a fixed projected radius and within a range of magnitudes $m_3 \leq m \leq m_3 - 2$, being m_3 the apparent magnitude of the third most bright galaxy of the cluster. These galaxies are placed within a circumference of angular radius $\theta_A \approx 1.7'/z$ ($0.02 \leq z \leq 0.2$). The redshift z is estimated from the assumption that the brightness of the tenth most brilliant galaxy in a cluster is the same for all the clusters (for a complete explanation see Abell, 1958; Abell and Seligman, 1965). Subsequently, an extension of this catalogue was made by Abell, Corwin and Olowin (ACO catalogue: Abell et al., 1989). However, cluster selection by galaxy overdensities presents problems related to projection effects since clusters are three-dimensional objects, while galaxy counts are based on the projection of galaxies in the sky. The random overdensities produced by projection along the line of sight can be erroneously considered as clusters of galaxies. On the contrary, due to the fluctuations in density of galaxies, a cluster at high redshift can be interpreted as a non-significant fluctuation. Despite these problems, the

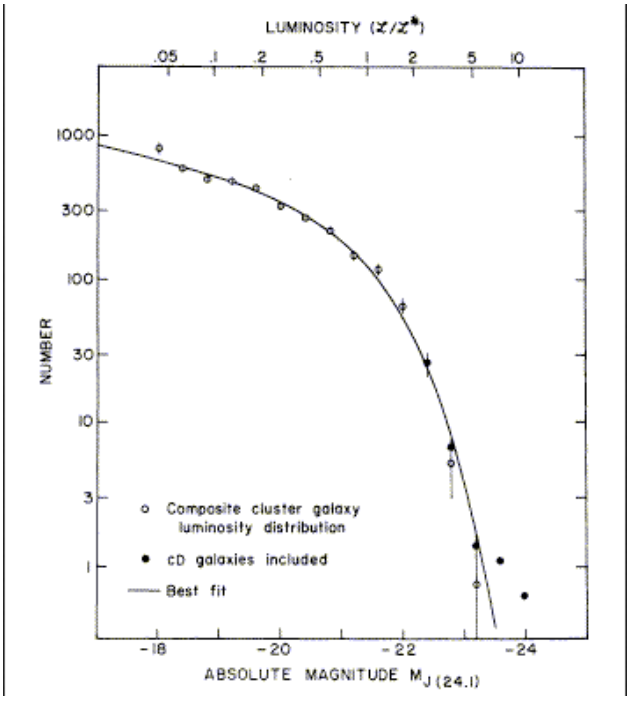

Figure 3.1: Measured composite cluster galaxy luminosity distribution compared to the Schechter (1976) model. Filled circles show the luminosity function including cD galaxies. *Credits*: Schechter

spectroscopic analysis performed for many clusters in the Abell catalogue did reveal that the redshift estimates were accurate, so the Abell and ACO catalogues are still widely used. Based on these catalogues, the clusters are divided into six classes of richness (i.e the excess of galaxies brighter than the limiting magnitude used to define the cluster), listed from 0 to 5 according to the number of galaxies belonging to the cluster. The class 5 corresponds to the clusters containing more than 300 galaxies brighter than the limiting magnitude, while the class 0 refers to groups containing from 30 to 49 galaxies.

Another type of cluster classification was made by Fritz Zwicky (Zwicky, 1933, 1937), who, taking into account the morphology, divided the clusters into regular (compact) and irregular (open). This classification gives an indication of the physical differences between the clusters, such as the correlation between the morphology and the properties of the clusters. For example, regular clusters are dominated by early-type galaxies, often present a central dominant (cD) galaxy in the centre, have a very high central density of galaxies and are characterised by a high richness. Instead, irregular clusters are less rich, contain a quantity of spirals similar to the distribution of field galaxies and are less dense at the centre.

3.2 Physical properties

3.2.1 The luminosity function

The Schechter luminosity function gives the numerical spatial density of galaxies having a certain luminosity. Looking at any galaxy cluster, we see a wide range of galaxy luminosities. The respective luminosity function specifies the relative number of galaxies at each luminosity. If we are interested in knowing the numerical density of galaxies with luminosity in the range between L and $L + \mathrm{d}L$, it is necessary to consider the quantity $N(L) = n(L)\mathrm{d}L$, with $N(L)$ the number of galaxies with luminosity L, and $n(L)$ the numerical density of galaxies with luminosity L. We did talk about its origin into the Subsection 2.5.2. The most used form is the one expressed by the Schechter luminosity function model (see Figure 3.1): (Schechter, 1976):

$$n(L)\mathrm{d}L = N_*\left(\frac{L}{L_*}\right)^{-\alpha} \exp^{-\frac{L}{L_*}} \mathrm{d}\left(\frac{L}{L_*}\right), \tag{3.2.1}$$

with $\alpha \approx -0.8, -1.3$ (from *flat* to *steep*), while N_* is the normalisation and L_* is a characteristic luminosity that discriminates between the power law regime and the exponential one. This is analogous to the mass function expressed by Eq. (2.5.16)[1]. The luminosity function contains information about the following phenomena:

- primordial density fluctuations,
- processes that form or destroy galaxies,
- processes that change one type of galaxy into another (e.g. mergers, stripping),
- processes that transform mass into light.

The luminosity function is thus a fundamental observational quantity, that any theory of galaxy formation and evolution must reproduce.

3.2.2 Spatial distribution of galaxies in clusters

The density of galaxies in most of the clusters shows a strong increase towards the centre. It can be safely assumed that the density of galaxies in clusters, and of ICM gas, follows a radially-spherical symmetry. Observing the distribution of the positions of the galaxies we go back to the projected two-dimensional density distribution $\Sigma(R)$, which is linked to the three-dimensional density $\rho(r)$ by the relation:

$$\Sigma(R) = \int_R^\infty \frac{r\rho(r)}{\sqrt{r^2 + R^2}} \mathrm{d}r, \tag{3.2.2}$$

where R is the projected radius and r the three-dimensional one. Inverting this relationship we yield the three-dimensional density. For example, the singular isothermal sphere (SIS) profile represents the simplest parameterization of the volumetric distribution of matter in a cluster and, furthermore, it assumes a Maxwellian distribution of the velocities of the galaxies within the cluster. The SIS profile is not properly physical because of the singularity that manifests itself at zero radius and of the fact that the total mass, estimated by integrating the function out to infinite, does not converge (Keeton, 2001). One of the most used isothermal models is the *King profile* (King, 1962, 1966) which is characterised by the following two equations, exploiting the further assumption that the speed of the galaxies does not depend on the distance from the centre:

$$\rho(r) = \rho_0 \left[1 + \left(\frac{r}{r_c}\right)^2\right]^{-3/2}, \tag{3.2.3}$$

$$\Sigma(R) = \Sigma_0 \left[1 + \left(\frac{r}{r_c}\right)^2\right]^{-1}, \tag{3.2.4}$$

with $\Sigma_0 \equiv 2\rho_0 r_c$, where r_c called core radius. This model describes accurately only the central regions of the clusters. Another model that can describe the total density distribution of clusters is the Navarro-Frenk-White (NFW) one (Navarro et al., 1997), that will be discussed in Chapter 4.

3.2.3 Estimation of the mass

There are various methods, more or less precise according to the observational conditions, to estimate the cluster masses. The most used exploits the following probes: the SZ effect, cluster scaling relations, ICM X-ray emission, gravitational lensing (that we will exploit later in this thesis work to derive the clusters density profiles) and the virial theorem. In this Subsection we will present the last one.

Clusters of galaxies typically contain galaxies with velocity dispersion along the line of sight of $\sigma_v \approx 10^3$ km/s, therefore the crossing time (which is the average time that a galaxy takes to cross once all the cluster) is of the order of:

$$t_{cross} \approx \left(\frac{r}{1Mpc}\right)\left(\frac{\sigma_v}{10^3 km/s}\right)^{-1} Gyr \approx 10^9 yr. \tag{3.2.5}$$

[1] In 1974 Press and Schechter calculated the mass distribution of clumps emerging from the young Universe, and in 1976 Paul Schechter applied this function to fit the luminosity distribution of galaxies in Abell clusters.

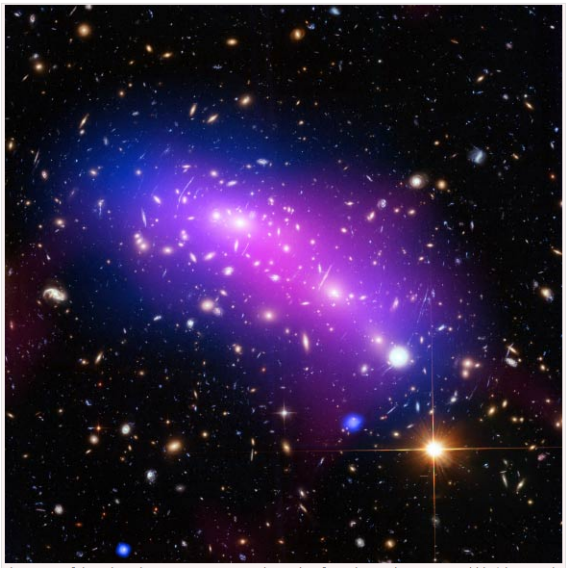

Figure 3.2: This image of the galaxy cluster MACS J0416 combines data from three telescopes – Hubble (showing the galaxies and stars), Chandra (diffuse X-ray emission, in blue), and the Jansky Very Large Array (diffuse radio emission, in pink). MACS J0416 acts as a powerful natural lens, bending and magnifying the light of distant background objects. This giant cluster has a total mass of about 160 trillion times the mass of the Sun. As all galaxy clusters, MACS J0416 contains a significant amount of dark matter, which leaves a detectable imprint in visible light by distorting the images of background galaxies. In this image, dark matter appears to align well with the blue-hued hot gas. The cluster contains many other features, such as a compact core of hot gas. *Credits*: HFF Team

The crossing time is an estimate of the time taken by the cluster to relax dynamically. Comparing it with the Hubble's time (which quantifies the age of the Universe) $t_H \approx 10^{10} h^{-1} yr$, with $h \approx 0.7$, we can infer that the clusters are dynamically stable. Therefore, the validity of the virial theorem can be assumed, from which an estimate of the mass of the clusters is obtained:

$$M_{vir} \approx \frac{R\sigma_v^2}{G} \approx 10^{15} \Big(\frac{\sigma_v}{10^3 km/s}\Big)^2 \Big(\frac{R_G}{Mpc}\Big) M_\odot, \tag{3.2.6}$$

with G the gravitational constant and $R_G = 2\frac{GM}{c^2}$ the gravitational radius. With this method we obtain a value of the order of $10^{15} M_\odot$ which, compared to the corresponding bolometric luminosity of the cluster $L_{bol} \approx 10^{13} L_\odot$, gives an M/L ratio:

$$\frac{M}{L_{tot}} \approx 300h \Big(\frac{M_\odot}{L_\odot}\Big) \tag{3.2.7}$$

As said in Subsection 2.5.2, this ratio is likely constant and reliable in the case of clusters, and it clearly indicates that clusters contain much more mass than that visible through galaxies and gas in emission, confirming the fact that they are dominated by dark matter.

3.3 X-ray emission from clusters

Clusters of galaxies are very bright X-ray sources: $L_x \sim 10^{43-45} erg/s$. This is an emission of an extended (i.e non-point-like) type, so it is not produced by single galaxies. A typical example of cluster diffuse emission, including that in X-ray, is shown in Figure 3.2. From the Spectral Energy Distribution (SED) it is deduced that the main process of thermal emission is the bremsstrahlung (or free-free emission) from collisionally ionized hot ICM gas (at a temperature of $10^{7-8} K$)[2]. This

[2]Indeed, in the centre of the almost all the relaxed clusters, where the Bright Central Galaxies (BCG) are present, there is also a strong synchrotron non-thermal X-ray emission produced by the central AGN (Active Galactic Nucleus).

hot gas is characterised by a bolometric emissivity:

$$\epsilon \propto T^{1/2} Z^2 n_e^2 \bar{g} \propto n_e^2 \Lambda_c(T), \tag{3.3.1}$$

where Z is the atomic number, $\bar{g}$ is the average Gaunt factor (which represents a correction term for the effects related to quantum mechanics and for distant collisions), n_e the number density of the electrons (that are the main actor in the free-free emission) and $\Lambda_c \propto T^{1/2}$ it is the cooling function for free-free emission of a ionized gas (see Cavaliere and Fusco-Femiano, 1976). The cooling time of this gas (the time-scale that quantifies the average time necessary to a gas, in hydrostatic equilibrium, to cool) is $t_{cool} = \frac{k_B T}{n\Lambda_c(T)} \propto T^{1/2} n^{-1}$. This last equation shows that in the inner regions, hotter and more dense, the cooling time decreases so as to pass part of the ICM in the cold phase, which does not emit more X-ray radiation. This is the so-called cool-core (Raymond and Smith, 1977). In the case of regular clusters, it is found that the X-ray brightness and the temperature increase with the mass of the cluster. Irregular clusters of a given mass, on the other hand, may be hotter or colder than regular ones because, being the result of a recent merger process, their temperature depends on the stage of this process.

The observation of the X-ray radiation allows to derive information on the characteristics of the ICM. Assuming that the gas in the clusters is in hydrostatic equilibrium is like saying that the condition:

$$\frac{\mathrm{d}p}{\mathrm{d}r} = \rho \frac{GM(r)}{r^2} \tag{3.3.2}$$

is valid. Then, taking advantage of the equation of state of perfect gases, we obtain that the mass within r is given by:

$$M(r) = \frac{k_B T r^2}{G\mu m_p} \Big(\frac{\mathrm{d}\ln \rho_g}{\mathrm{d}r} + \frac{\mathrm{d}\ln T}{\mathrm{d}r} \Big), \tag{3.3.3}$$

where p, ρ_g, T are, respectively, the pressure, the density and the temperature of gas, m_p is the proton mass and μ the average molecular mass. This relationship links the mass to the temperature and the density of the gas. However, it is not easy to measure the profiles. To derive the temperature and the density, one must consider the X-brightness, through the volumetric emissivity of bremsstrahlung: $\epsilon_{ff} \propto T^{-1/2} n_e^2$. This, however, can only be observed as a projection on the sky, i.e we observe its superficial brightness:

$$I_\nu(R) = 2 \int_R^\infty \frac{\epsilon_\nu(r) r}{\sqrt{r^2 - R^2}} \mathrm{d}r. \tag{3.3.4}$$

Therefore, in order to obtain the profiles it is necessary to carry out a deprojection. Usually the profiles are obtained by making data fit with the so-called *β-model*. This model exploits the assumption that the density of the total matter in the cluster follows an isothermal distribution (i.e the temperature does not depend on the radius), so that the density profile of the ICM becomes:

$$\frac{\mathrm{d}\ln \rho}{\mathrm{d}r} = -\frac{GM(r)}{\sigma_v^2 r^2}. \tag{3.3.5}$$

From Eq. (3.3.2) and from the perfect gas equation $p = nk_B T$, we obtain the *total radial density profile*:

$$\frac{\mathrm{d}\ln \rho_g}{\mathrm{d}r} = -\frac{\mu m_p}{k_B T} GM(R) r^2. \tag{3.3.6}$$

Comparing the equations (3.3.5) and (3.3.6) we obtain that:

$$\rho_g \propto [\rho(r)]^\beta \quad \text{with: } \beta = \frac{\mu m_p \sigma_v^2}{K_B T_g}. \tag{3.3.7}$$

Using the King model (Eq. (3.2.3) and (3.2.4)) for an isotherm mass distribution, it is found that the density and emission profiles of the ICM are respectively given by:

$$\rho_g(r) = \rho_{g,0} \Big[1 + \Big(\frac{r}{r_c} \Big)^2 \Big]^{-3\beta/2}, \tag{3.3.8}$$

$$I(R) \propto \Big[1 + \Big(\frac{R}{r_c} \Big)^2 \Big]^{-3\beta+1/2}. \tag{3.3.9}$$

Both relations describe well the central areas of galaxy clusters.

3.4 Scale relations of clusters

We have already mentioned and briefly introduced the scale relationships in Chapter 2, recalling the work of Giodini et al. (2013). Now, with referement to Mandelbaum et al. (2013), we summarily present the most used cluster scale relationships.

Mass-temperature relationship

One of the most solid scaling relation is that between mass and temperature. For a virialized mass, one has $T \propto \frac{M}{r_{vir}}$ (see Neumann and Arnaud, 1999). It is generally assumed that $r_{vir} = r_{200}$[3], for which the mass included within this radius is

$$M_{200} = \frac{4\pi}{3} 200 \rho_{crit} r_{200}^3. \tag{3.4.1}$$

Thus we obtain the *mass-temperature scale-relation*:

$$T \propto M_{200}^{2/3}. \tag{3.4.2}$$

The haloes are parameterised by the mass M_{200} which depends on the cosmological model and redshift z, because $\rho_{200} = \frac{\bar{\rho}(z)}{\Omega(z)} = \frac{3H^2}{8\pi G}$, where $\bar{\rho}(z)$ and $\Omega(z)$ are the average density of the Universe and the density parameter at a given redshift, respectively.

Mass-luminosity relation

Assuming that the ICM temperature is determined only by gravitational processes, it is obtained that:

$$L_x \propto \rho_g M_{200} T^{1/2}. \tag{3.4.3}$$

Combining this result with the scaling relation between mass and temperature (3.4.2), we obtain the *mass-luminosity scale-relation* for the bolometric luminosity (see Maughan, 2007):

$$L_x \propto M_{200}^{4/3}. \tag{3.4.4}$$

Y-mass relation

The Y Compton parameter is associated to the inverse Compton (IC) effect, which is related to the interaction between electrons and photons. In the IC effect the photons, being less energetic than the electrons with which they interact, gain energy in the collisions (with the photons of the CMB in this case) and their spectral distribution is distorted, allowing to identify clusters (for further information see Sunyaev and Zeldovich, 1980). Since this parameter is such that: $Y \propto TM$, we obtain (see Maughan, 2007) the *Y-mass scale-relation*:

$$Y \propto M^{5/3}. \tag{3.4.5}$$

3.5 Detection of clusters in optical surveys: methods, applications and advantages

Clusters of galaxies can not only be characterised but also identified using multi-band observations. By exploiting observational data in the Optical and Near Infrared bands, it is possible to identify clusters of galaxies by looking for overdensities of galaxies. Indeed, galaxy clusters were historically acknowledged as overdensities of galaxies in optical images (e.g Abell, 1958). This technique allows to cover large areas of the sky. However, more precise measurements are necessary on the individual galaxies. It is important to discriminate between proper physical clustered objects and clumps originated from random alignments and non-collapsed overdensities due to the matter large-scale distributions. Furthermore, during the observation of a specific area of the sky, we receive a signal from both galaxies that compose the cluster and from the galaxies of the field. In order to understand what are the cluster galaxies, one of the best ways is to derive the galaxy redshifts

[3] r_{200} is the radius within which there is a density contrast equal to 200 times the critical density of the Universe ρ_{crit} (see Eq. (1.6.1))

from their emission spectrum or, where there are no spectra, from photometric data, which are however less precise than spectroscopic ones.

Clusters of galaxies play an important role both in the study of the formation and evolution of the galaxies in their environment and in the determination of cosmological parameters. In the latter, the evolution of the mass spectrum of clusters and the mass-to-light ratio allow us to constrain the mass density contrast, Ω_m, and the amplitude of the perturbation spectrum, σ_8 (as done in Oukbir and Blanchard, 1992). So, the advantage offered by having a sample with a large redshift coverage is clear. Anyway, as noted by Yee and Gladders (2002): searching clusters using optical images is generally affected from one major deficiency. The projection of galaxies along the line of sight produces contaminating signals in the search for over-dense regions in 3-D space. This projection effect becomes increasingly more severe at higher and higher redshift, as the column of accumulated foreground galaxies renders the overdensity produced by a galaxy cluster almost impossible to detect, except for the richest clusters. The problem of projection contamination has caused optical searches to be considered as unreliable, and their selection functions difficult to quantify. This considerations were made in 2002. Instead, in 2019, Bellagamba et al. (2019) replied that ongoing and upcoming photometric surveys will increase the current census of clusters of galaxies by orders of magnitude, pushing the limits of detection towards lower masses and higher redshifts.

In order to make the most of this amount of photometric information useful for astrophysics and cosmology, it is essential to understand how to link the measured properties of a galaxies to the total mass of clusters, made almost entirely by dark matter. Unluckily, the relation between the total mass of a cluster and the properties of its galaxies is unseen in complex, and not yet well understood, astrophysical processes. The details of these processes are difficult to model both theoretically and through simulations. Because of this, there is not a clear mass proxy in the galaxy distribution and, furthermore, scaling relations between the mass and any galaxy observable can only be calibrated empirically. Various different mass proxies have been suggested in the literature, of which the most used the ones are based on the number of (red) galaxies inside a given radius (Andreon and Hurn, 2010; Rykoff et al., 2012), on photometric stellar mass estimates (Pereira et al., 2018) and on their luminosities (Mulroy et al., 2014).

One of the best ways to achieve well-founded cluster mass estimates, is to benefit from weak gravitational lensing, the mark on background galaxies observed shapes of the light deflection due to the intervening cluster gravitational potential. We will discuss about weak lensing in Chapter 4. Gravitational lensing is dependent only on the total matter distribution in the cluster and it supplies mass estimates which are not based on any assumption on the physical state of the clusters, contrarily to methods which depend on the gas properties (such as X-ray observations or the Sunyaev-Zel'dovich effect on the CMB). Furthermore, gravitational lensing is often one of the primary goals of current and future photometric surveys, thus enabling the calibration of optical observables of photometrically-selected clusters within the same data-set. Consequently, although large samples of clusters was also drawn from X-ray observations or through the detection of the Sunyaev-Zel'dovich (SZ) effect on the CMB, the photometric observations remain the most promising source of discovery and characterisation of new clusters of galaxies. This applies especially to the wide ongoing and planned photometric surveys such as the Dark Energy Survey (DES The Dark Energy Survey Collaboration, 2005), the Kilo Degree Survey (KiDS, Radovich et al., 2017; de Jong et al., 2017) and the European Space Agency Cosmic Vision mission *Euclid* (see Laureijs et al., 2011).

Galaxy clusters are detected, in photometric data, by running algorithms which search for not-negligible cluster-scale overdensities in the galaxy distribution. Usually one or more observables, related to the properties and abundance of galaxies, are estimated and employed in selecting the clusters, during this procedure. So the mass proxies, and the derived mass-scaling relations, are often estimated directly by the cluster finders (e.g. Parroni et al., 2017; Simet et al., 2017). In the next Section 3.6 we will present an algorithm which is able to efficiently perform this task.

3.6 The AMICO detection algorithm

AMICO, which stands for *Adaptive Matched Identifier of Clustered Objects*, is an algorithm for the detection of galaxy clusters in photometric surveys. AMICO is built on the *Optimal Filtering* technique, which maximizes the S/N of the clusters and takes advantage of the statistical properties of both field galaxies and galaxy clusters. The Optimal Filtering formalism enables to reliably

combine optical data with other observables such as weak lensing, SZ and X-ray observations. A more detailed description of AMICO can be found in Bellagamba et al. (2018), while for a further application of the algorithm to real data see Maturi et al. (2018). This algorithm adopts an updated iterative approach for the extraction of cluster candidates from the map produced by the filter. Moreover, AMICO can use an arbitrary number of physical quantities which are able to describe various galaxy cluster properties. In this section we will present its main properties and concepts, as well as the features recently implemented and the statistical assumptions on which it is based.

3.6.1 Linear optimal matched filtering

AMICO is built upon a *linear optimal matched filter method* (for information about the optimal filtering general technique, see e.g. Maturi et al., 2005, 2007; Viola et al., 2010; Bellagamba et al., 2011). A recent application of the method to photometric data is presented in Radovich et al. (2017). Within this context, the data, $d(\vec{x}) = s(\vec{x}) + n(\vec{x})$, are written as the superimposition of the signal we are interested in, i.e the *galaxy clusters signal* $s(\vec{X} - ray) = Ac(\vec{X} - ray)$, and a *noise component*, $n(\vec{X} - ray)$, characterising the contamination originated by the field galaxies. The filter itself is a kernel used to convolve the data, and it is obtained with a constrained minimisation procedure to ensure that the measured *signal amplitude* A is unbiased and with minimum variance. The amplitude A is a measure of cluster galaxy abundance in units of the cluster model. The Optimal Filtering allows to either consider or neglect an arbitrary number of galaxy properties in the model, such as magnitudes or colours, the presence of a BCG, or even photometric and morphological classifications. It is the algorithm itself that chooses the most suitable properties of the cluster detection in the available data-set, thus providing opportune weights to the galaxies in the catalogue. Despite this capability of AMICO, Maturi et al. (2018) limited their analysis to the case where the data points $\vec{x}_i = (\vec{\theta}_i, m_i, p_i(z))$ represent individual galaxies (marked with i) featured by sky coordinates, $\vec{\theta}_i$, an r-band magnitude, m_i, and a photometric redshift distribution, $p_i(z)$. As we will make use of their cluster catalogue in our analysis, in the following we will describe the application of AMICO to the same data-set and with the same settings. The previously mentioned convolution, returning an evaluation for A, is estimated on a three-dimensional grid, $(\vec{\theta}_c, z_c)$, with resolution of $0.3'$ through the sky and 0.01 in redshift, and is discretized as a sum over galaxies:

$$A(\vec{\theta}_c, z_c) = \alpha^{-1}(z_c) \sum_{i=1}^{N_{gal}} \frac{C(z_c; \vec{\theta}_i - \vec{\theta}_c, m_i) p(z_c)}{N(m_i, z_c)} - B(z_c). \tag{3.6.1}$$

In the previous equation, N and C describe the properties of the field and cluster galaxy distributions (at redshift z_c), respectively, as will be detailed in Subsection 3.6.4. The factor α deals with the filter normalisation, making sure that the amplitude A is a measure of the cluster signal in units of the cluster model M_c (see Eq. (3.6.9)). B is a background subtraction term quantifying the average contribution of the field galaxies to the total signal amplitude. The definition of B is given as $B(z_c) = \alpha^{-1}(z_c)\beta(z_c)$, where $\beta(z_c)$ is the total number of galaxies in the cluster model at redshift z_c. In fact, the subtraction of $B(z_c)$ in Eq. (3.6.1) ensure that $A = 0$ if the galaxy distribution around a given position corresponds to the field one. The expected r.m.s. (root mean square) of the amplitude is given by:

$$\sigma_A(\vec{\theta}_c, z_c) = \alpha^{-1}(z_c) + A(\vec{\theta}_c, z_c) \frac{\gamma(z_c)}{\alpha^2(z_c)}, \tag{3.6.2}$$

where the first term refers to the stochastic fluctuations of the background and the second one is related to the Poissonian fluctuations caused by the galaxies of a cluster with amplitude A. The factors B, α, and γ are properties of the filter, and only depend on the cluster[4] and field galaxy models. The definitions of α and γ are provided in Subsection 3.6.2.

Once the amplitude A has been computed for all angular and redshift positions, the first cluster candidate is then picked out as the location with positive amplitude and the *largest likelihood* (look at Appendix A for the definition of likelihood):

$$\mathcal{L}(\vec{\theta}_c, z_c) = \mathcal{L}_0 + A^2(\vec{\theta}_c, z_c)\alpha(z_c) \quad \Delta\mathcal{L}(\vec{\theta}_c, z_c) = A^2(\vec{\theta}_c, z_c)\alpha(z_c). \tag{3.6.3}$$

[4]The redshift distribution of the model is based on the average $p(z)$ of the input galaxies as detailed in Bellagamba et al. (2018).

In this context, $\mathcal{L}_0$ is a constant of no relevance which is no further discussed. Given a cluster detection marked by j, the probability of all galaxies in that region to be its members is estimated as:

$$P_i(j) = P_{f,i} \frac{A_j C_j(\vec{\theta}_i - \vec{\theta}_j, m_i) p_i(z_j)}{A_j C(\vec{\theta}_i - \vec{\theta}_j, m_i) p_i(z_j) + N(M_i, z_j)}, \tag{3.6.4}$$

with $P_{f,i} \equiv 1 - \sum_j P_i(j)$ that is the probability of the i-th galaxy to belong to the field. A galaxy could be associated to more than one detection because detections can partially overlap. To go ahead with the search of clusters, AMICO removes from the amplitude map the contribute of the last found detection and re-evaluates the likelihood and variance before identifying a new candidate. The elimination of a detection is done by exploiting the membership probabilistic association of the galaxies to a detection, $P_i(j)$, in this way:

$$A_{new}(\vec{\theta}_j, z_k) = A(\vec{\theta}_j, z_k) - \sum_{i=1}^{N_{gal}} \frac{C_j(\vec{\theta}_i - \vec{\theta}_c, m_i) p(z_k)}{N(m_i, z_k)}. \tag{3.6.5}$$

This signal subtraction enables a better identification of objects which might be confused with those with larger amplitudes. We refer to this process as *cleaning*. This iterative process proceeds until a chosen minimum S/N is reached, $S/N := A/\sigma_A$.

3.6.2 Features of the algorithm

To be able to exactly normalise the amplitude A and evaluate its uncertainty σ_A, AMICO calculates the quantities $\alpha(z_c)$ and $\gamma(z_c)$, which depend on the properties of the redshift probability distributions of the galaxy sample. In Bellagamba et al. (2018), this was done by using $q(z_c, z)$, the typical redshift probability distribution for a galaxy which lies at redshift z_c , computed as:

$$q(z_c, z) = \Big(\sum_{i=1}^{N_{gal}} p_i(z_c) \Big)^{-1} \sum_{i=1}^{N_{gal}} p_i(z - z_c + z_{peak,i}) p_i(z_c), \tag{3.6.6}$$

where $z_{peak.i}$ is the most probable redshift for the i-th galaxy. In Maturi et al. (2018), this treatment has been improved in two ways. Firstly, AMICO computes the photo-z properties as a function of the r-band magnitude, to catch the different precision of photo-zs that depends on the quality of the galaxy photometry. Secondly, it substitute q with two different statistics q_1 and q_2 defined, respectively, by:

$$q_1(m, z_p, z_c) = \Big(\sum_{z_{peak,i}=z_p} p_i(z_p) \Big)^{-1} \sum_{z_{peak,i}=z_p} p_i(z_p) p_i(z_c) \tag{3.6.7}$$

and

$$q_2(m, z_p, z_c) = \Big(\sum_{i=1}^{N_{gal}} p_i(z_c) \Big)^{-1} \sum_{z_{peak,i}=z_p} p_i(z_c) p_i(z_p), \tag{3.6.8}$$

where $z_{peak,i} = z_p$ means that the sum is done only on the galaxies whose photo-z probability peak corresponds to z_p. In other words, q_1 describes the typical photometric redshift (photo-z) distribution $p(z)$ that peaks at z_p, while q_2 describes the probability distribution for the peak, z_p, of a galaxy that is located at redshift z_c. They enable to measure the typical precision of the redshift probability distribution as a function of z, but also the small-scale features of the $p(z)$-peaks distribution, removing the smoothing that was implicit in equation (3.6.6). With these two new quantities, the constants $\alpha(z_c)$ and $\gamma(z_c)$ can be now defined as:

$$\alpha(z_c) = \int \frac{M_c^2(\vec{\theta} - \vec{\theta}_c, m) q_1(m, z_p, z_c) q_2(m, z_c, z_p)}{N(m, z_c)} \mathrm{d}^2\theta \mathrm{d}m \mathrm{d}z_p \tag{3.6.9}$$

and

$$\gamma(z_c) = \int \frac{M_c^3(\vec{\theta} - \vec{\theta}_c, m) q_1^2(m, z_p, z_c) q_2(m, z_c, z_p)}{N^2(m, z_c)} \mathrm{d}^2\theta \mathrm{d}m \mathrm{d}z_p. \tag{3.6.10}$$

3.6.3 Mass proxies and cluster richness

As introduced in Subsection 3.6.1, the main output of the linear optimal matched filter is the amplitude, A, expressed by the equation (3.6.5). In this subsection, we will show that AMICO gets two other mass proxies based on the probabilistic membership association of the galaxies (Eq. (3.6.4)). The first one is the *apparent richness* that is defined as the sum of the probabilities of all galaxies associated to the $j-$ th detection,

$$\lambda_j = \sum_{i=1}^{N_{gal}} P_i(j). \tag{3.6.11}$$

The apparent richness represents the number of visible galaxies belonging to a detection. Of course, this number depends upon the cosmic distance at which a cluster is situated so that λ depends on the redshift. Indeed, the farther is the cluster, the smaller the number of visible members. The convenience of this quantity with respect to the amplitude A is that it is connected to a direct observable, that is the number of visible galaxies.

Another proxy is the *intrinsic richness*, defined in an alike way but by summing over the galaxies brighter than $m_* + 1.5$ and within the virial radius, R_{200}:

$$\lambda_{*,j} = \sum_{i=1}^{N_{gal}} P_i(j) \quad \text{with} \quad \begin{cases} m_i < m_*(z_j) + 1.5 \\ r_i(j) < R_{200}(z_j) \end{cases} . \tag{3.6.12}$$

The radial truncation R_{200} and m_* are free parameters of the model used for the construction of the filter (see Subsection 3.6.4), that are adopted for internal consistency. Clearly each detection has its own R_{200}. We could scale the radial cut-off with the detection amplitude A or the apparent richness λ, but this would introduce an additional stochastic scatter given by the uncertainty of these quantities[5]. Nevertheless, λ_* is an almost redshift independent quantity which can better characterise the clusters mass. As in the work by Maturi et al. (2018), $m_* + 1.5$ should be well below the magnitude limit for the galaxy sample in the entire redshift range considered.

Summarising, AMICO delivers three mass proxy:

- The amplitude A, the natural output of the filter that is a measure of galaxy abundance through an optimal filter;
- The apparent richness λ, that indicates the number of visible galaxies;
- The intrinsic richness λ_*, the sum of membership probabilities in a consistent radial and magnitude range across redshift (as λ but for $r < R_{200}$ and $m < m_* + 1.5$).

3.6.4 Model description: cluster and field galaxies

In this section we will discuss specifically the model adopted in Maturi et al. (2018) since our thesis work is based on the galaxy cluster catalogue that they extracted.

The cluster model, $C(z_c; r, m)$, characterises the expected galaxy distribution as a function of the distance of the i-th galaxy from the cluster centre, $r = |\vec{\theta_i} - \vec{\theta_c}|$, and r-band magnitude m for a cluster at redshift z_c. The cluster model is built from a luminosity function $\Phi(m)$ and a radial profile $\Psi(r)$ as:

$$C(r, m) = \Phi(m)\Psi(r), \tag{3.6.13}$$

where we assume an implicit z_c redshift dependence. The parameters for these distributions are taken from Hennig et al. (2017), that analyse a cluster sample, detected through the SZ effect, from Dark Energy Survey (DES) data. These clusters cover a redshift range which is roughly comparable to the one of Maturi et al. (2018), and the detection via the SZ effect prevents any selection bias linked to the optical properties of the galaxies in clusters, which could introduce systematics in the detection process. Precisely, the luminosity function $\Phi(m)$ obeys a Schechter function (Schechter, 1976):

$$\Phi(m) = 10^{-0.4(m-m_*)(\beta+1)} \exp[-10^{-0.4(m-m_*)}]. \tag{3.6.14}$$

[5]Note that these definitions depend on the conversion from angular to physical distances and therefore imply the assumption of a specific cosmological model.

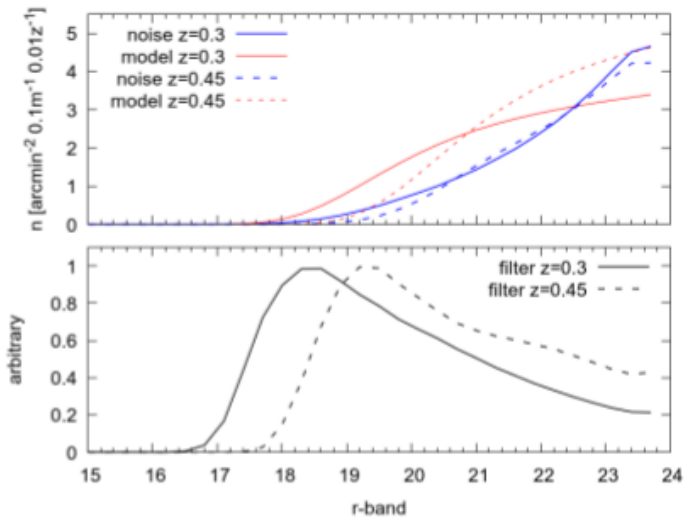

Figure 3.3: Top panel: the luminosity function of field galaxies, i.e the noise component (blue) and the one of the cluster model (red), at $z = 0.35$ (solid lines) and $z = 0.45$ (dashed lines) in the KiDS-DR3 cluster search with AMICO. Bottom panel: the magnitude distribution of the resulting filters. *Credits*: Maturi et al. (2018)

Note that in the previous equation we ignore the normalisation, as it will be fixed by the constants during the filter construction. We adopted a mean faint-end slope $\beta \approx -1$, in agreement with Hennig et al. (2017). A Navarro-Frenk-White radial profile (NFW Navarro et al., 1997) is exploited:

$$\Psi(r) = \frac{C_0}{\frac{r}{r_s}\left(1 + \frac{r}{r_s}\right)^2}, \tag{3.6.15}$$

where the scale radius r_s depends on the concentration parameter c via $r_s \equiv R_{200}/c_{200}$. For this cluster model it is considered a value of $c = 3.59$ and an R_{200} corresponding to a mass $M_{200} = 10^{14} M_\odot/h$, typical for the cluster sample at the redshift range we aim at. The normalisation parameter C_0 is such that the total number of galaxies N_{200} inside R_{200} and below $m_* + 2$ is coherent with what was found by Hennig et al. (2017). For a mass $M_{200} = 10^{14} M_\odot/h$, this corresponds to $N_{200} = 22.9$.

The field galaxies distribution $N(m, z_c)$ can be approximated by the total distribution in the galaxy sample, as the cluster component is small. For each redshift z_c , $N(m, z_c)$ is built weighing each galaxy with its redshift probability distribution $p(z_c)$. For illustration purposes, it is showed in the top panel of Figure 3.3 the luminosity function of field galaxies and of the cluster model at redshift $z = 0.35$ (solid lines) and $z = 0.45$ (dashed lines) in the KiDS-DR3 cluster search with AMICO (Maturi et al., 2018). The magnitude dependence of the algorithm filter, arising from the usage of the luminosity functions, is presented in the bottom panel of the same figure. Such filter turns out to be a band-pass filter which gives more weight to the galaxies with a certain luminosity on the bright end side: the higher the redshift, the larger the magnitude which the filter peaks at.

Chapter 4

Gravitational lensing

Gravitational lensing is a phenomenon that consists in the deflection of the radiation emitted by a source due to the presence of a very massive object interposed between the source itself and the observer, as predicted by General Relativity. Concretely, the curvature of spacetime is modified because of this mass, which behaves in a similar way to an optical lens, with refractive effects and an *apparent* time-delay. The resulting effect can consist in the production of multiple images (in case of strong lensing) or pure distortions of the source shape (almost always in presence of weak lensing) and in the intensification of the apparent brightness of this, generated by the convergence of light rays (as example, see Figure 4.1). In principle, we can measure and use both effects to estimate the projected masses, because the amplitude of the distortion ensures a direct measure of the lens gravitational tidal field. This is because the latter is independent of the nature of the dark matter or the dynamical state of the system of interest. In particular, this is really helpful for galaxy clusters, which are dynamically young and often in a state of merging.

When the source is at large angular distances from the geometric centre of the lens, its emitted radiation will be affected very weakly by the presence of the lens: in this case the effect is called *weak lensing*. In the opposite case of a source at an angular distance close to lens geometric centre, we talk about *strong lensing*. This Chapter will explain the fundamental notions of gravitational lensing, focusing mainly on weak lensing, as this is the main observational tool of the background source signal to probe the large-scale structures. In particular, the phenomenon of gravitational lensing produced by galaxy clusters will be described. Precisely, the effects of distortion (shear and amplification) that are induced on the light rays coming from the galaxies located behind the lenses will be highlighted. Weak lensing by clusters and also some of the techniques for obtaining the mass of the lens cluster through gravitational lensing will be discussed.

Most of the topics we will discuss up to the Section 4.4 are taken from the review that forms the: Weak Lensing part of the Saas-Fee Advanced Course on Gravitational Lensing, by Peter Schneider (Schneider, 2006) and, partly, from the lectures of Massimo Meneghetti (visit the the site: lectures to see the full document). The last part, which will deal with the specifics of weak lensing from clusters, draws the main notions from Bartelmann and Schneider (2001).

4.1 Basics of gravitational lensing

The *gravitational light deflection*:

- is independent of nature and state of deflecting matter causing the gravitational field, and depends only on the total mass of the lens,
- causes changes of apparent position of the light source or multiple imaging of this one,
- causes magnification and shape distortion of the source (differential light deflection).

For these reasons the gravitational lens effect can be used to estimate the mass distribution of the lens and the brightness distribution of the source. It can also be used to probe the geometry of the space between source, lens and observer and then, ultimately, the geometry of the Universe and its history of expansion through the so-called *cosmic shear*. The theory of cosmic shear was elaborated around the '90s, based on the concept that the Universe behaves like a huge gravitational lens: that is, because of massive structures along the line of sight, the shape of distant galaxies changes.

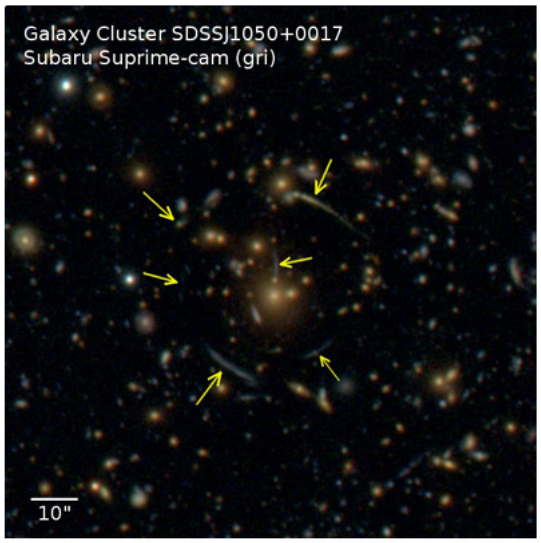

Figure 4.1: Subaru image of SDSS J1050+0017, one of the galaxy clusters used for the gravitational lensing analysis. The arrows indicate background galaxies whose appearance has been distorted by strong gravitational lensing, forming arcs and multiple images. *Credits*: subaru

Observing wide fields and averaging on a vast sample of galaxies, we can see how the cosmic shear (i.e the distortions of the shape of the galaxies) is correlated on different scales, obtaining information on the distribution of matter in the Universe, on the power spectrum and on the growth of structures.

Gravitational lensing describes phenomena of gravitational light deflection in the weak-field, small deflection limit. We define *weak lens* a massive object whose Newtonian gravitational potential Φ is such that: $\frac{\Phi}{c^2} \ll 1$. This is a virtually valid approximation for almost all cases of astrophysical interest (for a normal cluster $\frac{\Phi}{c^2} \approx 10^{-4}$). Strong-field light deflection (important for light propagation near black holes and neutron stars) usually is not covered by the gravitational lensing theory. The basic theory of gravity, and of light propagation in a gravitational field is General Relativity: photons travel along null geodesics of the space-time metric that are described by a second-order differential equation. In gravitational lensing, several simplifications apply, owing to restriction to weak fields, and thus *small deflections*.

4.2 Deflection angle and lens equation

Deflection by a point mass M Let us consider the deflection of a light ray by a *spherically symmetric mass M*. From the Schwarzschild metric (see Schwarzschild, 1916) one finds the *Einstein deflection angle*:

$$\hat{\alpha} = \frac{4GM}{c^2\xi} = \frac{2R_s}{\xi}, \tag{4.2.1}$$

with ξ the impact parameter of a light ray and $R_s = \frac{2GM}{c^2}$ the Schwarzschild radius that indicates the ideal event horizon radius for an object with mass M. Eq. (4.2.1) is valid for $\xi \gg R_s$ (or $\hat{\alpha} \ll 1$ i.e the small deflection angle limit) and in the weak-field regime $\frac{\Phi}{c^2} \ll 1$.

4.2.1 Deflection by a mass distribution

In the weak field regime, the General Relativity can be linearised as the deflection angle depends linearly on the mass. Thus, the deflection angle of a mass distribution can be considered as a superposition of deflections of mass elements: $\mathrm{d}m = \rho(\vec{r})dV$.

Let us indicate $(\xi_1(\lambda), \xi_2(\lambda), r_3(\lambda))$ as the light ray trajectory, with the incoming ray propagating along r_3. If the deflection is small then $\xi(\lambda) \approx const.$ near the mass concentration, and we can apply the Born approximation of a *geometrically thin lens*. Generally, the scale within which the deflection occurs (which corresponds to the size of the lens) is much smaller than the overall

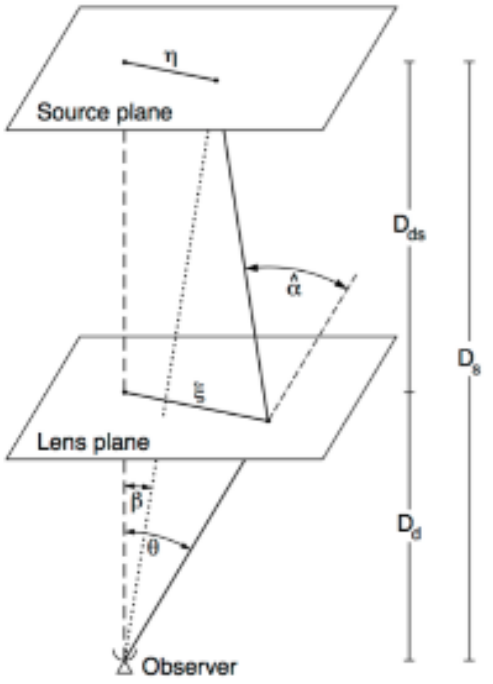

Figure 4.2: Scheme of the geometric system source-lens-observer in geometrically-thin lens approximation. *Credits*: Bartelmann and Schneider (2001)

dimensions of the observer, lens and source system, so the mass distribution of the lens can be projected along the line of sight and can be considered planar.

Considering more realistic lens models, three-dimensional distributions should be considered. However, in the case of clusters, this thin-screen approximation is justified by the fact that the lenses are much smaller than the overall distances of the system. Hence, with the assumption that the lens has a planar distribution, we get the *deflection angle for a planar lens*:

$$\hat{\vec{\alpha}}(\vec{\xi}) = \frac{4G}{c^2}\sum \mathrm{d}m(\xi_1', \xi_2', r_3')\frac{\vec{\xi}-\vec{\xi}'}{|\vec{\xi}-\vec{\xi}'|^2} = \frac{4G}{c^2}\int \mathrm{d}^2\xi' \frac{\vec{\xi}-\vec{\xi}'}{|\vec{\xi}-\vec{\xi}'|^2}\int \mathrm{d}r_3'\rho(\xi_1', \xi_2', r_3'). \tag{4.2.2}$$

Then, the r_3'-integration can be carried out, defining the *surface mass density of the lens*:

$$\Sigma(\vec{\xi}) \equiv \int \mathrm{d}r_3'\rho(\xi_1', \xi_2', r_3') \tag{4.2.3}$$

and, therefore, we can rewrite the Eq. (4.2.2) in a more compact form:

$$\hat{\vec{\alpha}}(\vec{\xi}) = \frac{4G}{c^2}\int \mathrm{d}^2\xi' \Sigma(\vec{\xi}')\frac{\vec{\xi}-\vec{\xi}'}{|\vec{\xi}-\vec{\xi}'|^2}. \tag{4.2.4}$$

Eq. (4.2.4) is valid if the impact parameter near the lens is much smaller than the characteristic scale of the mass distribution and if every kind of lens (galaxies, clusters, etc.) is localized[1] and all relevant angles are small (e.g. for galactic lenses $\hat{\alpha} \leq 1$" and for cluster lenses $\hat{\alpha} \leq 30$"). So, *the properties of a gravitational lens depend only on its surface-mass density* Σ.

The lens equation

With reference to Figure 4.2, we can write the equation which relates the true position of the source to its observed position[2]:

$$\vec{\eta} = \frac{D_s}{D_d}\vec{\xi} - D_{ds}\hat{\vec{\alpha}}(\vec{\xi}) \quad \text{with: } \vec{\eta} = D_s\vec{\beta} \text{ and } \vec{\xi} = D_d\theta. \tag{4.2.5}$$

[1] Eq. (4.2.4) is not valid for large-scale structures, i.e for the cosmic shear.

[2] In a first approach with a regular geometrical approximation, we can study this phenomenon with classical Newtonian gravity, without the necessity of the General Relativity.

Now, we can write the *lens equation* in its compact form:

$$\vec{\beta} = \vec{\theta} - \frac{D_{ds}}{D_s}\hat{\vec{\alpha}}(D_d\vec{\theta}) \equiv \vec{\theta} - \vec{\alpha}(\vec{\theta}), \tag{4.2.6}$$

with $\vec{\alpha}(\vec{\theta}) = \frac{D_{ds}}{D_d}\hat{\vec{\alpha}}$ called *scaled deflection angle*. Note that $D_d + D_{ds} \neq D_s$ because of the Hubble flow. However, this purely geometrical derivation is valid in a cosmological setting.

We can define the *dimensionless surface mass density* (or *convergence*):

$$\kappa(\vec{\theta}) := \frac{\Sigma(D_d\vec{\theta})}{\Sigma_{crit}}, \tag{4.2.7}$$

with Σ_{crit} called *critical surface mass density*:

$$\Sigma_{crit} \equiv \frac{c^2}{4\pi G}\frac{D_s}{D_d D_{ds}}, \tag{4.2.8}$$

that depends only on the distances. Then, we can rewrite the scaled deflection angle as a function of the convergence κ:

$$\vec{\alpha}(\vec{\theta}) = (\pi)^{-1}\int_{\mathbb{R}^2} \mathrm{d}^2\theta' \kappa(\vec{\theta'})\frac{\vec{\theta} - \vec{\theta'}}{|\vec{\theta} - \vec{\theta'}|^2}. \tag{4.2.9}$$

Now, using the convergence, we can redefine the method to distinguish the regime of a lens: lenses with $\kappa \sim 1$ are called *strong lenses* while, those with $\kappa \ll 1$ everywhere are *weak lenses*. The lens equation $\vec{\beta} = \vec{\theta} - \vec{\alpha}$ (4.2.6) maps $\vec{\theta} \to \vec{\beta}$ from the lens plane to the source plane but, in general, this mapping cannot be inverted: for a given source position $\vec{\beta}$, the lens equation can have multiple solutions $\vec{\theta}$, i.e *multiple images*.

Deflection and Fermat potentials

Since $\vec{\nabla}\ln|\vec{\theta}| = \vec{\theta}/|\vec{\theta}|^2$, the deflection angle can be written as:

$$\vec{\alpha} = \nabla\psi \quad \text{with: } \psi(\vec{\theta}) = (\pi)^{-1}\int_{\mathbb{R}^2} \mathrm{d}^2\theta' \kappa(\vec{\theta'})\ln|\vec{\theta} - \vec{\theta'}|, \tag{4.2.10}$$

with $\psi(\vec{\theta})$ called *deflection potential*. So mapping $\vec{\theta} \to \vec{\beta}$ is a gradient mapping.

From $\nabla^2\ln|\vec{\theta}| = 2\pi\delta_D(\vec{\theta})$, with δ_D the Dirac-delta, we finds the *two-dimensional Poisson equation*:

$$\nabla^2\psi = 2\kappa. \tag{4.2.11}$$

Then, defining the *Fermat potential* τ:

$$\tau(\vec{\theta}, \vec{\beta}) = \frac{1}{2}(\vec{\theta} - \vec{\beta})^2 - \psi(\vec{\theta}), \tag{4.2.12}$$

where $\vec{\beta}$ enters as a parameter, one sees that the lens equation can be written as:

$$\vec{\nabla}\tau(\vec{\theta}, \vec{\beta}) = \vec{0}. \tag{4.2.13}$$

4.3 Magnification and shape distortions

4.3.1 Magnification

The gravitational light deflection conserves surface brightness. There are no emission or absorption processes, thus the number of photons is conserved (it follows the Liouville's theorem[3]), so that:

$$I(\vec{\theta}) = I^{(s)}[\vec{\beta}(\vec{\theta})], \tag{4.3.1}$$

[3]Liouville's theorem is an important physical theorem of the Hamiltonian mechanics and classical statistics. It asserts that, along the trajectories of the system, the phase-space distribution function is constant. In other words, the density of system points near a given system point moving across the phase-space is constant with time. This density is independent of time and it is known as the classical a priori probability, in statistical mechanics.

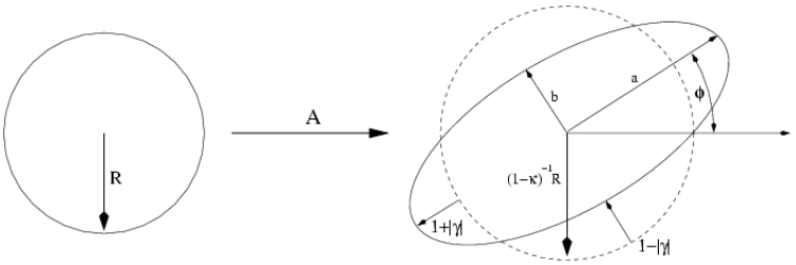

Figure 4.3: Representative scheme of the geometrical shape distortion of a circular light source into an ellipse, causing the lensing shear. *Credits*: Bartelmann and Schneider (2001)

with $I(\vec{\theta})$ and $I^{(s)}(\vec{\beta})$ the surface brightness in the image and source plane, respectively.

The differential light bending implies that the beams of light are distorted. For very small light bundles, the distortion is described by the *Jacobian matrix*:

$$\mathcal{A}(\vec{\theta}) = \frac{\partial \vec{\beta}}{\partial \vec{\theta}} = \left(\delta_{ij} - \frac{\partial^2 \psi(\vec{\theta})}{\partial \theta_i \partial \theta_j}\right) = \begin{pmatrix} 1-\kappa-\gamma_1 & -\gamma_2 \\ -\gamma_2 & 1-\kappa+\gamma_1 \end{pmatrix}, \tag{4.3.2}$$

where we introduced the *shear components*:

$$\gamma_1 = \frac{1}{2}(\psi_{11} - \psi_{22}) \quad \text{and} \quad \gamma_2 = \psi_{12}. \tag{4.3.3}$$

With this expression the meaning of convergence and shear clearly emerges. Eq. (4.3.2) shows that the convergence acts isotropically on the source, dilating it of the same factor in all directions, while the shear induces a deformation along a privileged direction.

For a small source centred on $\vec{\beta}_0 = \vec{\theta}_0 - \vec{\alpha}(\vec{\theta}_0)$, Eq. (4.3.1) becomes:

$$I(\vec{\theta} = I^{(s)}[\vec{\beta}_0 + \mathcal{A}(\vec{\theta}_0) \cdot (\vec{\theta} - \vec{\theta}_0)]. \tag{4.3.4}$$

Then, the image of a small circular source with radius r appears as an ellipse with semi-axes $\lambda_1 r$ and $\lambda_2 r$, where λ_1 and λ_2 are the *eigenvalues* of $\mathcal{A}$. The orientation of the ellipse is determined by the shear components γ_1 and γ_2 (see Figure 4.3).

If we define S and S_0 as the magnified flux and the intrinsic flux of the source, respectively, the area distortion produced by the differential deflection yields to define the scalar *magnification*:

$$\mu = \frac{S}{S_0} = \frac{1}{\det \mathcal{A}} = \frac{1}{(1-\kappa)^2 - |\gamma|^2} \quad \text{with: } |\gamma| = \sqrt{\gamma_1^2 + \gamma_2^2}. \tag{4.3.5}$$

Thus, the intensity of γ determines the magnitude of the distortion of the source image. Through this distortion the lens induces an amplification of the source, which is different depending on the position of the source with respect to the lens.

Then, since $\mathcal{A}$ is different for different multiple images, image fluxes are different. The observed flux ratios yield image magnification ratios. For this reason, the flux ratios can be used to constrain lens models but the micro-lensing made by stellar-mass compact object in lens galaxy (or cluster) or other small-scale mass structures (substructures) can affect flux ratios, rendering them less useful.

4.3.2 Shape distortions

A typical shape distortion of background galaxies, by a massive extended lens (a cluster in this case) is shown in Figure 4.1, where the distortion produces arcs. Of course, this effect is visible only if the lensed source is extended. Because of lensing, the shape of extended (resolved) sources changes, since the two eigenvalues of $\mathcal{A}$ are, in general, different from each other. So, we rewrite the Jacobian matrix (4.3.2) in the form:

$$\mathcal{A}(\vec{\theta}) = (1-\kappa)\begin{pmatrix} 1-g_1 & -g_2 \\ -g_2 & 1+g_1 \end{pmatrix}, \tag{4.3.6}$$

with

$$g_i \equiv \frac{\gamma_i}{(1-\kappa)} \tag{4.3.7}$$

the so-called *reduced shear*. Hence the shape distortion is determined by the reduced shear. Moreover, the reduced shear is the real direct observable that we estimate. Since the lens changes significantly on scales larger than those of the source, we can use $\mathcal{A}$ because the relationship between the position of the source and that of the image is linear. This represents the basis of the *weak lensing*, that works very well in a linear approach. Note that giant arcs cannot be described by linearised lens equations, as they are too big, that is they are observed in cases of quasi-strong lensing. In other words, the shear represents the projected tidal gravitational field of the deflector. However, the shear (and the reduced one) is a 2-component quantity, written as a complex number:

$$\gamma = \gamma_1 + i\gamma_2 = |\gamma| e^{2i\varphi} \quad ; \quad g = g_1 + ig_2 = |g| e^{2i\varphi}. \tag{4.3.8}$$

Its amplitude describes the degree of distortion, while its phase φ the direction. The reason for the factor 2 in phase factor is that an ellipse transforms into itself after a rotation of 180°. Note that the shear is not a vector but it is a 2×2 tensor. Vectors are defined trough their transformation behavior under rotations. The shear is the traceless part of a symmetric tensor, which determines its transformation behavior through $e^{2i\varphi}$ (the transformation in a different reference frame manifests itself in changing on the phase).

Referring to Figure 4.3, we consider a circular source with radius R. The lens distortion is mapped by the local Jacobi matrix $\mathcal{A}$ and the resulting source distorted image is an ellipse with axes:

$$\frac{R}{1-\kappa-|\gamma|} = \frac{R}{(1-\kappa)(1-|g|)} \quad \text{and} \quad \frac{R}{1-\kappa+|\gamma|} = \frac{R}{(1-\kappa)(1+|g|)}. \tag{4.3.9}$$

To first order, a circular background source is distorted into an elliptical image, with axis ratio:

$$r = \frac{1-|g|}{1+|g|}, \quad \text{valid for: } |g| < 1. \tag{4.3.10}$$

This shape distortion presents the primary weak lensing observable. Such shape distortion is very small (typically, $1\% < |g| < 10\%$), much smaller than the intrinsic ellipticity distribution of galaxies, so a statistical approach is required. When some galaxy images are distorted as heavily to form arcs (see Figure 4.1) there will be many more galaxy images around the cluster that are deformed much less. Here the distortion signal can be extracted only by statistical means. Hence, if circular sources could be distinguished, the measured image ellipticity would straightaway furnish the reduced shear, through the axis ratio, $|g| = \frac{1-b/a}{1+b/a} \Leftrightarrow \frac{b}{a} = \frac{1-|g|}{1+|g|}$, and the orientation φ of the major axis. Anyway, faint galaxies are not intrinsically round, thus the observed image ellipticity is a combination of intrinsic ellipticity and shear. To solve this problem, the strategy is to locally average over many galaxy images, assuming the intrinsic ellipticities as *randomly oriented*, so that we can calculate the local shear. Another problem is how to define the ellipticity for a source with arbitrary isophotes or, in addition, the effect of seeing by atmospheric turbulence that will blur, and thus circularize, the observed images. We will see how to solve this drawbacks in the next section.

4.4 Measurements of shapes and shear

The galaxies are not elliptical (their isophotes are in general not ellipses). Thus, to define an ellipticity for an arbitrary image shape, we will do it in terms of *brightness moments*. Let $I(\vec{\theta})$ be the brightness distribution of an image, assumed to be isolated on the sky. We define the centre of the image:

$$\vec{\bar{\theta}} \equiv \frac{\int \mathrm{d}^2\theta q_I[I(\vec{\theta})] I(\vec{\theta}) \vec{\theta}}{\int \mathrm{d}^2\theta q_I[I(\vec{\theta})] I(\vec{\theta})}, \tag{4.4.1}$$

where q_I is a suitably chosen weight function. We define now the tensor of second-order brightness moments:

$$Q_{ij} = \frac{\int \mathrm{d}^2\theta q_I[I(\vec{\theta})] I(\vec{\theta})(\theta_i - \bar{\theta}_i)(\theta_j - \bar{\theta}_j)}{\int \mathrm{d}^2\theta q_I[I(\vec{\theta})] I(\vec{\theta})}, \quad i,j \in \{1,2\}. \tag{4.4.2}$$

Note that, for an image with circular isophotes, $Q_{11} = Q_{22}$ and $Q_{12} = 0$. Thus, the trace of the tensor Q characterises the image size while, the traceless part of Q_{ij} includes the ellipticity information. From Eq. (4.4.2), we can define twocomplex ellipticities:

$$\chi \equiv \frac{Q_{11} - Q_{22} + 2iQ_{12}}{Q_{11} + Q_{22}} \quad \text{and} \quad \epsilon \equiv \frac{Q_{11} - Q_{22} + 2iQ_{12}}{Q_{11} + Q_{22} + 2(Q_{11}Q_{22} - Q_{12}^2)^{1/2}}, \tag{4.4.3}$$

and both have the same phase (since they have the same numerator), but different absolute value. Note that, in both cases, the denominator is given by the invariants of Q_{ij}, i.e invariant against rotations. Then, for an image with elliptical isophotes of axis ratio $r \leq 1$, we have:

$$|\chi| = \frac{1 - r^2}{1 + r^2} \quad \text{and} \quad |\epsilon| = \frac{1 - r}{1 + r}. \tag{4.4.4}$$

Which of these two definitions is more convenient depends on the context but, anyway, one can be transformed into the other:

$$\epsilon = \frac{\chi}{1 + (1 - |\chi|^2)^{1/2}} \quad \text{and} \quad \chi = \frac{2\epsilon}{1 + |\epsilon|^2}. \tag{4.4.5}$$

In total analogy, one can defines the second brightness tensor $Q_{ij}^{(s)}$, and the complex ellipticities $\chi^{(s)}$ and $\epsilon^{(s)}$ for the unlensed source.

From source to image ellipticities With referring to the the lens equation (4.2.6), from the equation:

$$Q_{ij}^{(s)} = \frac{\int \mathrm{d}^2\beta q_I[I^{(s)}(\vec{\beta})]I(\beta_i - \bar{\beta}_i)(\beta_j - \bar{\beta}_j)}{\int \mathrm{d}^2\beta q_I[I^{(s)}(\vec{\beta})]I}, \quad i, j \in \{1, 2\}, \tag{4.4.6}$$

one finds, with $\mathrm{d}^2\beta = \det \mathcal{A} \mathrm{d}^2\theta$ and $\vec{\beta} - \vec{\bar{\beta}} = \mathcal{A}(\vec{\theta} - \vec{\bar{\theta}})$, that:

$$Q^{(s)} = \mathcal{A}Q\mathcal{A}^T = \mathcal{A}Q\mathcal{A}, \tag{4.4.7}$$

where $\mathcal{A} \equiv \mathcal{A}(\vec{\theta})$ is the Jacobian matrix of the lens equation at the position $\vec{\theta}$. The brightness moments of second-order transformation between source and image are provided exclusively in terms of the Jacobian matrix $\mathcal{A}$, i. e. the locally linearised lens equation. Now, using the definitions of complex ellipticities (4.4.3), (4.4.4) and (4.4.5), and the reduced shear g, we find the transformations:

$$\chi^{(s)} = \frac{\chi - 2g + g^2\chi^*}{1 + |g|^2 - 2\mathfrak{Re}(g\chi^*)} \quad \text{and} \quad \epsilon^{(s)} = \begin{cases} \frac{\epsilon - g}{1 - g^*\epsilon}, & \text{if } |g| \leq 1 \\ \frac{1 - g\epsilon^*}{\epsilon^* - g^*}, & \text{if } |g| > 1. \end{cases} \tag{4.4.8}$$

The inverse transformations can be obtained interchanging source and image ellipticities, and $g \to -g$:

$$\chi = \frac{\chi^{(s)} + 2g + g^2\chi^{(s)*}}{1 + |g|^2 + 2\mathfrak{Re}(g\chi^{(s)*})} \quad \text{and} \quad \epsilon = \begin{cases} \frac{\epsilon^{(s)} + g}{1 + g^*\epsilon^{(s)}}, & \text{if } |g| \leq 1 \\ \frac{1 + g\epsilon^{(s)*}}{\epsilon^{(s)*} + g^*}, & \text{if } |g| > 1 \end{cases}. \tag{4.4.9}$$

4.4.1 Estimating the reduced shear

Keeping in mind the definition of *reduced shear* (4.3.7) and using the assumption that the intrinsic orientation of galaxies is random (i.e no direction in the Universe is singled out), $E(\chi^{(s)}) = E(\epsilon^{(s)}) = 0$ with E the expectation value, we can note that the expectation value of ϵ is:

$$E(\epsilon) = \begin{cases} g & \text{if } |g| \leq 1 \\ 1/g^* & \text{if } |g| > 1 \end{cases}. \tag{4.4.10}$$

So each image ellipticity provides an unbiased estimate of the local shear, though a very noisy one. The noise is determined by the *intrinsic ellipticity dispersion:* $\sigma_\epsilon = \sqrt{\langle \epsilon^{(s)}\epsilon^{(s)*} \rangle}$. The noise can be beaten down by averaging on many galaxy images: we live in a Universe where the sky is full of faint galaxies. The accuracy of the shear estimate depends then on the local number density of

galaxies for which the shape can be measured: it requires deep imaging observations. Note that in the weak lensing regime ($\kappa \ll 1$, $|\gamma| \ll 1$) we find that:

$$\gamma \approx g \approx \langle \epsilon \rangle \approx \frac{\langle \chi \rangle}{2}. \tag{4.4.11}$$

This regime is of primary interest in most weak lensing applications. Only close to strong lenses (like clusters) this weak lensing regime no longer olds. Note that the expectation value of χ depends on the intrinsic ellipticity probability distribution function.

Tangential and cross components of shear The shear components γ_1 and γ_2 are defined relative to a reference Cartesian coordinate frame but, often, it is useful to measure them with respect to a different direction (e.g. the arcs in clusters are tangentially aligned with them). One then wants to measure shear with respect to a given direction (here, the centre of cluster), i.e measuring the vector components in rotated reference frames. If Φ specifies a direction, we can define the *tangential* and *cross components* of the shear relative to this direction, respectively, as:

$$\gamma_t = -\mathfrak{Re}[\gamma e^{-2i\phi}] \quad \text{and} \quad \gamma_\times = -\mathfrak{Im}[\gamma e^{-2i\phi}]. \tag{4.4.12}$$

For example, in case of a circularly-symmetric matter distribution, the shear at any point will be oriented tangentially to the direction towards the centre of symmetry. Thus in this case, choosing ϕ as the polar angle of a point (the centre of the cluster here), then $\gamma_\times = 0$. The null value of $\gamma_\times$ is used as a test about the reliability of a weak lensing observation by clusters: if this value differs appreciably from 0, something went wrong during the interpretation of the lensed signal. This check is called *E-B mode*, in total analogy with electric and magnetic fields, where the E-mode field is curl-free and the B-mode field is divergence-free. Because gravitational lensing can only produce an E-mode field, the B-mode provides another test for systematic errors.

4.4.2 Magnification effects

Magnification changes apparent brightness of sources. There are two effects:

- the flux S from a source changes according to: $S = \mu S_0$ and, if $\mu > 1$, the sources appear brighter;
- a population of sources in the unlensed solid angle ω_0 is spread over the solid angle $\omega = \mu\omega_0$ due to magnification.

These two effects affect the number counts of sources differently. Which one wins depends on the slope of the number counts:

$$n(>S, z) = \frac{1}{\mu(\vec{\theta}, z)} n_0\Big(> \frac{S}{\mu(\vec{\theta}, z)}, z\Big), \tag{4.4.13}$$

with $n(>S, z)$ and $n_0(>S, z)$ the lensed and unlensed cumulative number counts of sources, respectively. We report now an illustrative example of this notion: if source counts follow a power-law, $n_0(>S) = aS^{-\alpha}$, we find, for the lensed counts in a region of the sky with magnification μ, that: $\frac{n(>s)}{n_0(>S)} = \mu^{\alpha-1}$. Hence, if $\alpha > 1 (< 1)$, the source counts are enhanced (depleted); the steeper the counts, the stronger the effect. E.g. the Quasi Stellar Object (QSO) number counts are steep at the bright end and flat at the faint end, thus in regions with $\mu > 1$, bright QSO should be over-dense and the faint ones underdense. This *magnification bias* is the reason why fraction of the lensed sources is much higher in bright QSO sample than in the fainter ones.

Magnification in the weak lensing regime

Provided that $\kappa \ll 1$ and $|\gamma| \ll 1$, then:

$$\mu \approx 1 + 2\kappa \quad \text{and} \quad \frac{n(>S)}{n_0(>S)} \approx 1 + 2(\alpha - 1)\kappa. \tag{4.4.14}$$

Thus, from measurement of the local number density $N(>S)$, the convergence κ can be inferred directly. For galaxies in B-band, $\alpha \sim 1$, but in redder bands $\alpha < 1$, so a depletion is expected for

these last ones. A further effect is that $\mu > 1$ causes galaxies with the same surface brightness to be larger than unlensed galaxies. In addition, the surface brightness of galaxies is a strong function of redshift (Tolman effect: $I \propto (1+z)^{-4}$). However, the magnification effect in weak lensing is difficult to measure, due to the required calibration accuracy.

4.5 Lensing by clusters

Hints of strong lensing. If the deflection of the light rays is huge enough, multiple images of the same source can be observed: these strong lensing events provide precise constraints on the mass on scales enclosed by these images (centred on the lens centre). The appearance of strong lensing effects by a cluster of galaxies is now seen very frequently, in particular in deep observations. Among the various applications of strong lensing one of the most effective is the determination of masses of galaxies or clusters. Unluckily, the effect is restricted to the inner regions, and a comparison with other mass proxies requires extrapolation to larger radii. Consequently, this phenomenon can be handled to obtain very precise constraints on the mass distribution but, only in cluster cores (for further information about strong lensing by clusters see Section 2.1 of Bartelmann and Schneider, 2001).

Weak lensing from clusters. Anyway, the lensing effect itself is not confined to small scales. At large radii the tidal field produces a tenuous change in the shapes of galaxies, resulting in a coherent alignment of the sources and that can be inferred statistically. The signal is proportional to the lens system mass so, for this reason, galaxy clusters were the first objects to be examined by means of the weak gravitational lensing technique. The alterations in the observable properties of galaxies, in presence of *weak gravitational lensing*, are smaller. So the measurement of the lensing signal asks for an ensemble of sources (e.g. the stacked lensing shear signal) to be appreciable. Respect to strong lensing studies, the statistical nature (because of the necessity of many different sources) of weak lensing restricts the precision with which the masses can be estimated. The resulting signal can then be measured, under the condition that the surface density of the sources is high enough.

Whenever κ and γ are small, the effect of gravitational lensing is to modify the unlensed ellipticity of a galaxy ϵ_{orig} (defined as the ratio $(a-b)/(a+b)$, with a and b the lengths of the major and minor semi-axis, respectively) to an observed value:

$$\epsilon_{obs} = \frac{\epsilon_{orig} + g}{1 + g^* \epsilon_{orig}} \approx \epsilon_{orig} + \gamma, \tag{4.5.1}$$

where g is the reduced shear (see Eq. (4.3.7)) and the asterisk denotes the complex conjugate. Hence only if $\gamma > \epsilon_{orig}$ the observed ellipticity of a single galaxy does supply a useful estimate of the shear. A population of intrinsically round sources would thence be ideal but, in practice, galaxies have an average intrinsic ellipticity of ~ 0.25 per component, where the exact value depends somehow on the galaxy type.In practice, clusters of galaxies generate systematic deformations on distant galaxies due to the high numerical density of the latter. Being on a much lower scale than the scale on which the gravitational potential of the entire cluster varies appreciably, the signals can be averaged using statistical techniques to determine the mass distribution of the cluster. Assuming that galaxies are randomly oriented, the average on their forms would tend to become circular. If the lensing is weak, the image of a circular source should appear elliptical. Under the assumption that the original angular positions of the galaxies are distributed randomly, the shear can be deduced by averaging over an ensemble of sources (see, e.g Mandelbaum et al., 2006a). The measurement of the shear component γ_i has a statistical uncertainty given by:

$$\sigma_{\gamma,i} = \sqrt{\frac{\langle \epsilon_i^2 \rangle}{N}} \approx \frac{0.25}{\sqrt{N}}, \tag{4.5.2}$$

where N is the number of galaxies that is averaged to obtain the lensing signal.

Considering an isolated (cluster) lens, the effect of the weak lensing is a systematic tangential alignment of the background galaxies images with respect to the lens. The *average tangential distortion* or, simply, *tangential shear*, defined as:

$$\gamma_T = -(\gamma_1 \cos 2\phi + \gamma_2 \sin 2\phi), \tag{4.5.3}$$

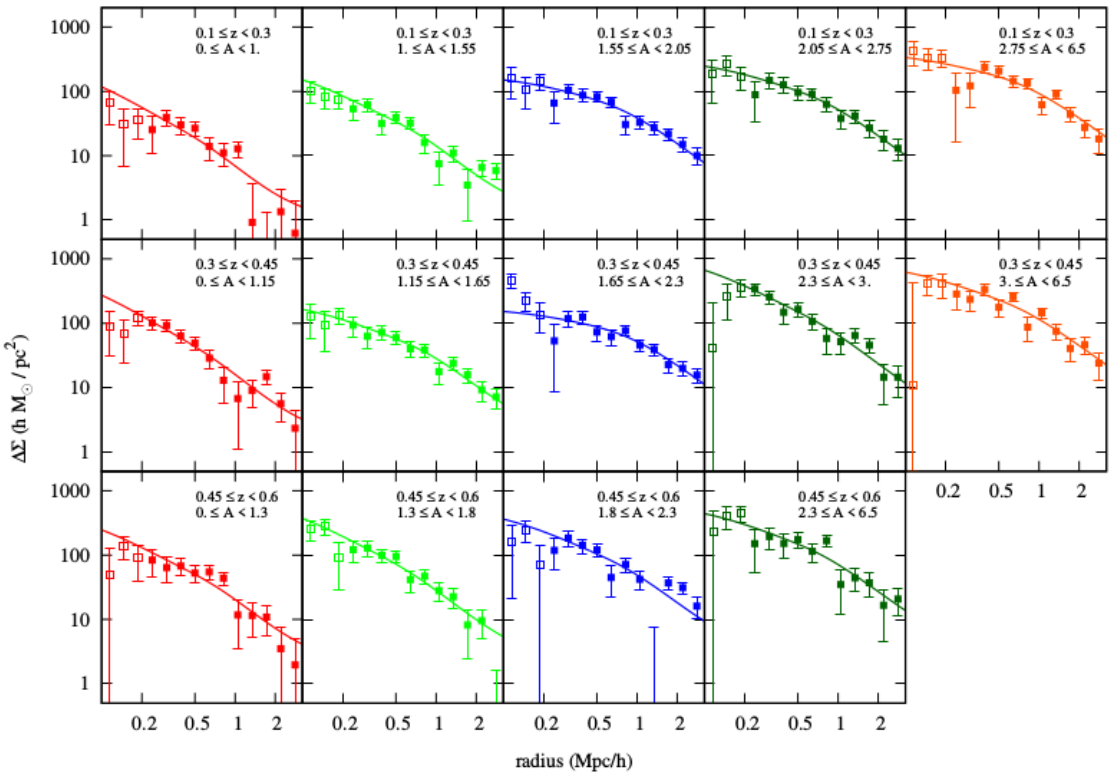

Figure 4.4: Example of stacked profiles of the differential surface density $\Delta\Sigma(R)$, in different bins of amplitude A and redshift (see Section 3.6.1 for the definition of A). The cluster halo density profiles are modelled with a truncated NFW profile (Eq. (5.4.1)). Redshift increases from top to bottom, amplitude increases from left to right. In each box, the curve represents the model with mean posterior parameters. Empty squares represent data at scales below $0.2\,\mathrm{Mpc}/h$, that have not been included in the fitting procedure. *Credits*: Bellagamba et al. (2019)

it can therefore be used to quantify the lensing signal, where ϕ is the azimuthal angle with respect to the lensing galaxy. This is a suitable quantity because, for *any* mass distribution, the azimuthally averaged tangential shear $\langle\gamma_T\rangle$, being function of the radius from the cluster centre, can be interpreted as a *mass contrast* (Miralda-Escude, 1991):

$$\langle\gamma_T\rangle(r) = \frac{\bar{\Sigma}(<r) - \bar{\Sigma}(r)}{\Sigma_{crit}} = \bar{\kappa}(<r) - \bar{\kappa}(r), \tag{4.5.4}$$

where $\bar{\Sigma}(<r)$ is the mean surface density within an aperture of radius r and $\bar{\Sigma}(r)$ is the azimuthally averaged surface density at distance r. It is also common to present the lensing signal as $\Delta\Sigma(r) = \Sigma_{crit}\gamma_T(r)$, which is a convenient quantity when the source and lense redshifts are measurable (see e.g Figure 4.4), as it enables a direct comparison between samples of a given lensing signal amplitude. Anyway, note that the observed signal coincides with the reduced shear $g = \gamma/1 - \kappa$. This is not a problem when considering fitting parametric models to the data, as the reduced shear can be computed.

4.5.1 Lensing cluster models

An important consideration for weak lensing studies is that not all clusters (if any) are well described by simple models (a complete list of cluster models is provided in Keeton, 2001). One of the most used lens models to characterise the phenomenon of gravitational lensing, is the axial symmetry lens. The latter have the advantage that the optical axis is chosen so that it intercepts the plane of the lens in the centre of the lens.

A simple and useful model to compare to the data is the Singular Isothermal Sphere (SIS), as it is obtained assuming that the lens is composed of a gas, in thermal and hydrostatic equilibrium, and

confined by a spherical symmetry gravitational potential, with density distribution $\rho(r) = \frac{\sigma_v^2}{(2\pi G r^2)}$ (σ_v is the velocity dispersion along the line of sight). This density radial profile can be compared to results from redshift surveys of cluster members, which provide a dynamical mass estimation. For this mass distribution, the tangential shear is expressed as $\gamma_T(r) = \kappa(r) = r_E/2r$. r_E is the Einstein radius, which can be expressed as $r_E = 28.9"(\sigma_v/1000km/s)^2\beta$, where $\beta = D_{ls}/D_s$ is the ratio of the angular diameter distance between the lens and the source and the angular diameter distance between the observer and the source, and which quantifies the dependence of the lensing signal amplitude on the source redshift. The slope $\partial\beta/\partial z$ quantifies the sensitivity to the source redshift distribution, which is also larger at high redshifts. This adds a noise in the mass measurements unless redshift information for the individual sources is obtained. However, note that this profile is mostly used to describe the distribution of matter in galaxies, as it reproduces to a flat rotation curve.

NFW profile in weak lensing

Another widely used profile is the Navarro-Frenk-White (NFW) profile, as it well characterise the radial density trend followed by the CDM haloes. Through numerical simulations it has been shown that haloes, over a wide range in mass, can be described by a simple fitting function (Navarro et al., 1997). The density profile is characterised by two parameters, usually the *halo mass* and the *concentration* c, which can be related to a characteristic scale (for its definition, see the Eq. (4.5.5) and its description). From strong or weak lensing alone, the concentration cannot be constrained well for individual clusters, on an observational level. The lensing is sensitive to the total density profile, therefore we should include the baryon contribute, though their effect should be most significant in the cluster cores, which are often excluded from the fit (e.g., as in the work of Bellagamba et al., 2019). How much the profiles are changed depends on the baryon physics. However the current knowledge about the baryons physics is weak. The uncertainty in the mass estimate is widened further by the presence structures along the line of sight. The *NFW density profile* itself is given by:

$$\rho_{NFW}(r) = \frac{M_{vir}}{4\pi f(c)} \frac{1}{r(r+r_s)^2}, \tag{4.5.5}$$

where M_{vir} is the mass enclosed within the virial radius r_{vir}. Note that for $r << r_s$, the density goes as r^{-1}; for $r >> r_s$ it goes as r^{-3}. The concentration parameter c is defined as the ratio between the r_{vir} and the scale radius r_s, $c = r_{vir}/r_s$ and it can be expressed as argument of the function $f(c)$, in the form: $f(c) = \ln(1+c) - c/(1+c)$. Analytic expressions for the projected cluster halo surface density and shear have been derived by Bartelmann (1996). The virial mass is defined with respect to the mean density of the Universe at the redshift of the cluster, where the virial overdensity is given by $\Delta_{vir} \approx 18\pi^2 + 82\xi - 39\xi^2$, with $\xi = \Omega(z) - 1$ (Bryan and Norman, 1998). For example for a standard ΛCDM cosmology, $\Delta_{vir}(z=0) = 337$. Other definitions are also used in the literature, with over densities defined with respect to the critical density ρ_c (defined in Eq. (1.6.1)) at the cluster redshift:

$$M_\Delta = \frac{4\pi}{3}\Delta\rho_c(z) r_\Delta^3, \tag{4.5.6}$$

where r_Δ refers to the radius where the mean density is $\Delta\rho_c(z)$, and M_Δ is the corresponding enclosed mass. M_{200} is often used as a proxy of the virial mass. Albeit similar in value, its definition is in effect different. As one must choose a definition of the mass, it is important that this is clearly defined when listing results.

Note that compared to the singular isothermal sphere, in the case of the NFW model, the radial profile of the lensing potential is flatter near the centre.

4.5.2 Choice of cluster centre

The choice of the cluster centre is very important to derive the lensing mass. When the adopted centre is offset from the right position, the contrast is lowered, with the result that one has a reduction in the observed signal and consequently in the lensing mass, as shown in Figure 4.5. The width of the bias depends on the range of scales used to deduce the lensing mass: if the lensing signal is measured at relatively large radii, the bias is decreased. Massive virialised (i.e relaxed) clusters typically have well defined central galaxies, whose locations coincide with the

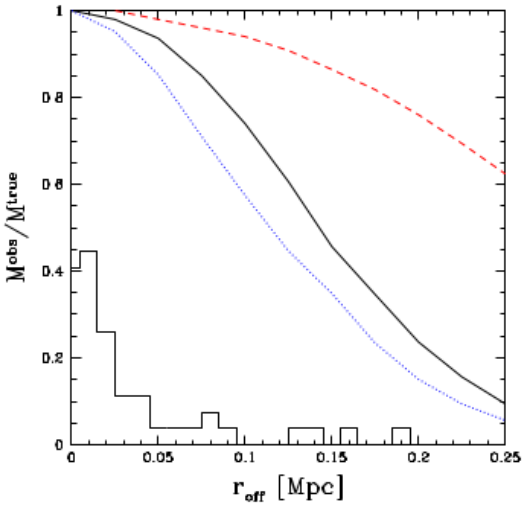

Figure 4.5: The ratio of the inferred lensing mass and true mass as a function of centroid offset r_{off}. The lensing mass is obtained by fitting an NFW model to the shear at $200 - 750\,h_{70}^{-1}$ Kpc (solid black curve) and $0.5 - 1.5\,h_{70}^{-1}$ Mpc (dashed red curve). The blue dotted line corresponds to the bias if all data within $750\,h_{70}^{-1}$ Kpc are used. The distribution of offsets observed for massive clusters is indicated by the histogram. *Credits*: Hoekstra et al. (2011)

X-ray emission peak.[4] The projected mass distribution reconstruction can, in principle, be used to choose the cluster centre. Anyway, even for the most massive clusters, the S/N is only of order 10, and this means that using the peak position itself, might introduce a bias that produces higher mass estimates. In case, it is possible to consider this when studying an ensemble of clusters, but it is more convenient to use external constraints. In fact, weak lensing cluster studies usually exploit the location of the brightest cluster galaxy, or the peak in the X-ray emission, as the centre. The situation is further complicated when considering huge samples of clusters, since high quality X-ray observations may not be available or cannot be reached because the masses of the targets are low. Analysing these large samples of clusters, it is noticed that the centre "handpicking" is therefore not functional, nor preferable. Indeed the effects of mis-centring can be considered statistically (Johnston et al., 2007). The existence of substructure in the central regions not only makes the choice of centre complicated, but it also leads to biases in the mass estimates when parametric models are fitted to the signal. In confront to a unimodal density distribution, this happens because of the substructure which also lowers the mean density contrast, as well as the shear.

Cosmic noise. Clusters are parts of the *cosmic web* and are located at the intersections of filaments. Hence some of this large-scale structure is physically associated with the clusters, which makes more difficult the interpretation of the lensing results, especially when one wants to compare with other mass indicators, or measure the density profiles (Hoekstra et al., 2013). This does introduce a bias which depends on how the mass is inferred from the lensing signal. For instance, if we restrict the model fit to R_{200}, we can reduce the bias, compared to fits to the signal, because the NFW model is not a good model for the description of the signal at larger radii, e.g. because substructures are present and due to the fact that clusters themselves are clustered. Gravitational lensing is affected by all structures along the line of sight. The large-scale structures bring excess correlations in the shapes of the galaxies, also known as *cosmic shear*. Furthermore, for weak lensing cluster studies, such excess variance behaves as a further source of noise. It is a consequence of the fact that we cannot differentiate the cluster signal from the large-scale structures contributions. However, the latter contribution disappears on average, and so it does not bias the mass measurements. In fact, the main contributes to the cosmic noise signal come from clusters and groups of galaxies that, in principle, we are able to identify.

[4]Note, however, that in merging clusters of galaxies the peak of the X-ray emission may not coincide with the main cluster haloes.

4.5.3 Cluster-mass cross-correlation function

A restricted number of background galaxies limits the study of single clusters to virial masses of the order of a few times $10^{14} M_{\odot}$ or larger. However, it is practice use to study lower mass systems by averaging, or *stacking* their signals: the resultant signal is that of the mass-weighted ensemble. Precisely, the stacked gravitational lensing represents the cross-correlation between foreground deflector positions and background galaxy shears. If we have a shallow but broad survey, to compensate for the low number density, stacking the signal around a large number of lenses is an efficacious method.

This approach was applied with great success to the Sloan Digital Sky Survey (SDSS) data to measure the ensemble averaged lensing signal around groups and clusters (Mandelbaum et al., 2006b; Sheldon et al., 2009). By splitting the sample of clusters into subsets based on an observable property, such as amplitude or optical richness (as in Figure 4.4, where a clear lensing signal is detected in all amplitude bins), we are able to determine scaling relations between the observable and the mean mass of the sample (see, e.g., Bellagamba et al., 2019). Anyway, note here that the interpretation of the result requires the knowledge of the intrinsic scatter between the mass and the quantity of interest. Only for high masses the latter can be observationally ascertained; for low masses we can rely on numerical simulations to supply a guide. The ensemble averaged signal corresponds to the *cluster-mass cross-correlation function*. This reflects the fact that only on small scales the signal coming from the main cluster halo dominates, while at large radii it is sensitive to the clustering of clusters, which in fact provides additional cosmological information. We are able to compute the combined signal through a (semi-)analytical approach (as suggested by Seljak, 2000): the halo-model. It uses the fact that the NFW profile well describes individual dark matter haloes and that we are able to estimate the halo mass function and the clustering of haloes for a given cosmology.

As already mentioned, *on small scales* the lensing signal is calculated from the NFW model, which is convolved with a probability of density distribution of offsets from the adopted centre. This can be deduced quite directly for massive clusters but, for large samples of (optically selected) clusters, a distribution of offsets needs to be adopted, e.g. via simulations. This first contribution to the signal gives information about the cluster dark matter halo (and implicitly assumes that the ICM comes behind a like distribution). The signal of the Brightest Central Galaxy (BCG) should also be added, and can be approximated by a point mass.

On large scales the fact that clusters themselves are part of the cosmic web, and therefore that are clustered, causes an additional signal. This can be seen clearly in Figure 4.4 for the lower mass clusters (top-left corner): the lensing signal shows a slight excess around the 2-3 Mpc (the solid line), compared to the NFW standard model. The signal from these adjacent haloes is described by the *2-halo term* (its usage can be seen, e.g., in Seljak, 2000), which requires the calculation of the correlation function in the linear perturbation theory and a prescription for the linear bias as a function of mass and redshift. A confrontation of the predicted amplitude with the observed value can be helpful to test models of structure formation or to constrain the normalisation of the matter power spectrum.

4.6 Cluster halo properties from weak lensing

To analyse the clusters density profiles at large radii, the weak lensing is well suited, especially when combined with strong lensing studies. As already mentioned, this is because the weak lensing signal can be measured up to large radii (from the cluster centre), thus investigating the gravitational potential on large scales. Even if the outskirts were measurable, the results from most dynamic techniques would be difficult to interpret because the most external areas of the clusters are not fully virialised.

4.6.1 Density profiles

One of the principal quantities that can be inferred from weak lensing is the *density profile of clusters*. Although this can be extrapolated for individual clusters, the allusion to a radial profile implicitly presumes a high degree of spherical symmetry, which may not be realistic. Moreover, the cosmic noise decreases the precision when considering only weak lensing measurements. Therefore a right question to ask is how the average density profile of clusters looks like. If the cluster

sample selection is not (too) sensitive to cluster orientation, the ensemble averaged cluster should be rather spherical. These results could be directly compared to the predictions of N-body dark matter simulations.

The studies of the density profile, through weak lensing, begins with the *tangential shear profile* $g_t(r)$, which then can be converted into a *projected density profile* $\Sigma(r)$. Instead, strong lensing and combined methods usually yield the profile via the lensing potential. Otherwise, one can fit a parametrized model to the tangential shear (as done, by means of stacking, by Bellagamba et al., 2019, ,see Figure 4.4).

Results from stacking

To average out the effects of the large-scale structures on the cluster identification, stacking is comprehensibly convenient. It also enables us to extend the range of analysis to much lower cluster masses and study their *statistical* properties. The stacking process must be executed in a sensible way that is, for example, taking into account the redshift scaling of the signals from clusters within a (redshift) bin. Johnston et al. (2007) performed the extraction of the 3D density and mass profiles from stacked clusters and developed numerical simulations to test the performance.

As it is easily understood, stacking methods are even more sensitive to mis-centring errors than weak lensing studies of individual clusters. Anyway, this can be modelled (see Subsection 4.5.2) and the drawback is exceeded by the cancellation of unrelated large-scale structures contaminants. Moreover, the analysis aims at diminishing additive errors in the shape analysis, because they do not contribute coherently, unless the clusters are targeted and always come into view on the same position in the field of view.

The SDSS did provide important contributes to the lenses statistical studies, ranging from galaxies to clusters of galaxies. For instance, Mandelbaum et al. (2006b) used the Luminous Red Galaxy (LRG) sample to define a sample of clusters of galaxies and galaxy groups and then fitted NFW models to the lensing signal $\Delta\Sigma(r)$. They found that the resulting concentrations agree well with ΛCDM model predictions and, furthermore, also review extensively possible systematic effects. Then, in a later work (Mandelbaum et al., 2008), the mass range was expanded to lower masses so as to provide constraints on the concentration parameter over more than two orders of magnitude in mass ($10^{12} - 5 \times 10^{14} M_{\odot}$). Another important contribute to these studies comes from the analysis presented in Sheldon et al. (2009), where it was studied the MaxBCG cluster catalogue and reported meaningful $\Delta\Sigma$ profiles in all tested richness and (total i-band) luminosity bins, spanning a scale between small galaxy groups and clusters. These measurements were analysed further in Johnston et al. (2007) who fitted a halo model (in a way similar to the one of Bellagamba et al. (2019) shown in Figure 4.4, but in luminosity bins). The optimal precision of the measurements enabled Johnston et al. (2007) to fix the concentration as a function of halo mass, finding good agreement with previous estimates.

About results and study cases and for individual clusters (that is not our case of interest) one can consult Section 5.1.2 of Bartelmann and Schneider (2001).

Shapes of dark matter haloes. Numerical simulations provide important predictions about the structure formation: dark matter haloes are tri-axial but this complicates the comparison of weak lensing masses to other methods. This prediction, however, can be investigated through gravitational lensing. Strong lensing studies are useful to analyse the inner regions, tough the presence of baryons may complicate the results interpretation. Instead, the dark matter halo shape at larger radii can be investigated through weak lensing. Although one could use mass reconstructions for this purpose, it is better to examine the effect of an elliptical mass distribution on the shear field.

4.6.2 Cluster masses from weak lensing

There have been and there are various successful effort to utilize the observed number density of clusters of galaxies as a precision probe of cosmological parameters. An accurate comprehension and interpretation of the observations relies upon the capability to correctly relate the observables to the mass. Such scaling relations (reviewed in detail by Giodini et al., 2013) need a calibration which is still an influential limiting factor for the precision of recent cluster surveys. Anyway, masses estimations from weak gravitational lensing are playing an increasingly important role, so

we highlight a few results in this subsection. One of the reasons why the weak lensing is so popular is that the method is becoming sufficiently mature, thanks to improved shear measurements. Furthermore, also important is the fact that the cluster lensing signal can be inferred out to large radii.

Last but not least, there is growing evidence on non-gravitational processes, like AGN feedback, which are crucial in cluster formation. For instance, the pressure support carried by turbulent motion of the ICM, introduces biases in the hydrostatic masses, which necessarily must be calibrated observationally. Being the weak lensing masses indifferent to the dynamical state of the cluster, the related results can be readily compared to cosmological numerical simulations. Note that this does not imply that weak lensing masses themselves are unbiased, only that the biases can be reliably quantified using numerical simulations. The main obstacle arising from comparing lensing masses with other tracers is the fact that clusters are complicated three-dimensional structures. Even if the lensing signal measurement itself does not require assumptions about the elaborated geometry of the cluster, comparison with other observations or predictions does rely on the assumed geometry. It is nowadays clear that the bias in the (de-projected) weak lensing mass relies on the methodology and the adopted radial range.

Results from stacking analyses

As already discussed, the signal-to-noise might be too low for clusters at high redshifts or for low mass systems. By stacking the lensing signal from an ensemble of objects we are able nevertheless to provide constraints on their (average) mass. For the first time, this process was performed for a sample of 50 galaxy groups from the CNOC2 field galaxy redshift survey by Hoekstra et al. (2011), finding that there was fairly good agreement between the lensing and dynamical mass. Although extended to a little larger sample of galaxies groups, by various researchers, these studies were limited by the relatively small number of systems.

Nowadays, the situation in the field of observations from stacked weak lensing is really improved, and photometric data are becoming more and more accurate, thanks to the improvement of observational tools and lensing signal extraction algorithms, as well as the quality and depth of current optical surveys. We can see these improvements in Chapter 5, when we will discuss about our data, on which our statistical analysis will be based.

Chapter 5

Cluster surface density profiles from shear stacked data

In this Chapter we will describe the survey dataset on which this thesis work is based and how it was handled for our purposes, together with the details of the models used to subsequently (in Chapter 6) infer cluster and cosmological parameters. With all the notions of the previous Sections in mind, we will present our data extraction of the weak lensing signal (the shear) of a galaxy sample of the KiDS survey, useful to reconstruct the surface density profiles of the galaxy clusters producing such signal. Clusters have been identified and extracted from the survey using the AMICO algorithm, described in Section 3.6, by Maturi et al. (2018). We have collected the background galaxy signal lensed by clustered structures, from the cluster centre up to their extreme outskirts, where the cluster halos tend to become filamentary and to connect with other clusters to form the so-called "cosmic web". The weak lensing S/N for each detected cluster is typically very low, so we followed a stacking approach for the signal measurement. This procedure allows to estimate various average profiles and masses of ensembles of clusters, but this last analysis step will be discussed in Chapter 6.

5.1 The KiDS survey

The Kilo-Degree Survey (KiDS, de Jong et al., 2013) is an optical wide-field imaging survey aiming at mapping 1550 square degrees of extra-galactic sky in two stripes (an equatorial one, KiDS-N, and one centred around the South Galactic Pole, KiDS-S) in four broad-band filters (u, g, r, i). The observations are performed with the 268 Megapixels OmegaCAM wide-field imager (Kuijken, 2011), owning a mosaic of 32 science CCDs and positioned on the VLT Survey Telescope (VST). This is an ESO telescope of 2.6 meters in diameter, located at the Paranal Observatory (for further technical information about VST see Capaccioli and Schipani, 2011). The principal scientific goal of KiDS is to exploit weak lensing and photometric redshift measurements to map the large-scale matter distribution in the Universe. To this end, it is essential to have unbiased and accurate measurements of weak lensing shear and photometric redshifts. This requires images with a high quality and depth and a very precise calibration. The VST-OmegaCAM is the best choice for such a survey, as it was expressly created to give excellent and uniform image quality over a large ($1\,\mathrm{deg}^2$) field of view with a resolution of 0.21 arcsec/pixel. Note that a field of view 1 deg wide corresponds to ~ 12.5Mpc/h at the median redshift of our lens sample, $z \sim 0.35$. Anyway, the huge data set provided by KiDS has many more possible applications (see for instance de Jong et al., 2013).

The current KiDS catalogue is composed of about 100,000 sources per square degree, for a total of almost 150 million sources over the full survey area, and the quantity of information of the whole survey final images reaches 15TB of memory. These data are attentively and uniformly calibrated and made usable in an easily accessible archive.

5.1.1 KiDS-DR3

DR3 stands for third Data Release. The KiDS-DR3 survey (de Jong et al., 2017) has been obtained with the same instrumental settings of the other two previous releases (de Jong et al., 2015) that

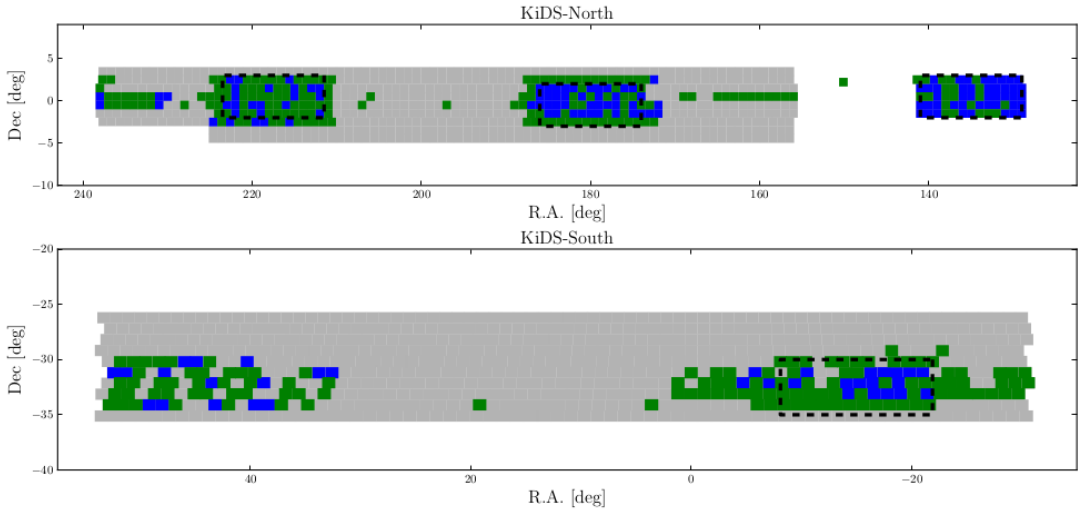

Figure 5.1: Sky distribution of the survey tiles released in KiDS-DR3 (green) and in the previous releases KiDS-DR1 and KiDS-DR2 (blue). Top panel: KiDS-North. Bottom panel: KiDS- South. The multi-band source catalogue covers the combined area (blue + green), while the full KiDS area is shown in grey. Black dashed lines delineate the locations of the GAMA fields; the single pointing at RA = 150 deg is centred at the COSMOS/CFHTLS D2 field. *Credits*: de Jong et al. (2017)

was made public in 2013 (DR1) and 2015 (DR2). The latter contains reduced image data, source lists and a multi-band catalogue, made by 148 survey tiles obtained with an automated masking observation process, for a total covering of about 160 $\deg^2$. KiDS-DR3 extends the total released data set coverage to approximately 447 $\deg^2$ with 440 survey tiles (48 736 591 sources), therefore, with respect to the other previous releases, it has produced a considerable survey area extension (see Figure 5.1). This release also includes photometric redshifts with the corresponding probability distribution functions, a global improved photometric calibration, weak lensing shear catalogues and lensing-optimized image data. Source detection, positions and shape parameters used for weak lensing measurements are all derived from the r-band images, since these typically have the best image quality and therefore they provide the most reliable measurements, while magnitudes are measured in all filters using forced photometry.

One of the main improvements of KiDS-DR3, compared to DR1 and DR2, is the enhanced photometric calibration and the inclusion of photometric redshift distribution probabilities (e.g., see Radovich et al., 2017; Maturi et al., 2018). Furthermore, the released DR1 and DR2 tiles formed a very patchy on-sky distribution, while the combined set of the 440 survey tiles of DR3 mostly comprises a small number of large contiguous areas. This enables a refinement of the photometric calibration that exploits both the overlap between observations within a filter as well as the stellar colours across filters (de Jong et al., 2017).

The data products that constitute the main DR3 release (stacked images, weight and flag maps, and single-band source lists for 292 survey tiles, as well as a multi-band catalogue for the combined DR1, DR2 and DR3 survey area of 440 survey tiles), are released via the ESO Science Archive, and they are also accessible via the Astro-WISE system (link) and the KiDS website http://kids.strw.leidenuniv.nl/DR3.

5.2 Our dataset

In this section, we will describe how the weak lensing measures of the KiDS-DR3 catalogue are collected, through an appropriate stacking algorithm, to fix the shear profiles of the used galaxy catalogue. We will also present the KiDS-AMICO clusters catalogue (Maturi et al., 2018) on which we build these shear profiles. The AMICO algorithm (presented and tested on simulations in Bellagamba et al., 2018) takes advantage of the known statistical properties of the field galaxies and of galaxy clusters. Moreover, we will discuss how this algorithm could ensure a good confidence level about the reliability of the clusters extraction. In fact, during the detection, AMICO has

estimated the probability for each KiDS-DR3 galaxy to be a member of the clustered structures detected by this extraction algorithm (see how in Section 3.6.1). The KiDS-DR3 shear data will be our tools, necessary to reconstruct excess surface density profiles for clusters detected by the algorithm, and to which we are interested in. Compared to the previous studies on the second data release (Radovich et al., 2017), the cluster catalogue collected through AMICO by Maturi et al. (2018), that we will exploit, takes advantage of a larger survey area, better data quality, and significant improvements in the cluster detection algorithm. Furthermore, the data processing and the galaxy catalogue extraction come from a KiDS-optimized pipeline running in the Astro-WISE environment (Verdoes Kleijn et al., 2012) .

We underline the fact that the KiDS-DR3-AMICO sample of galaxy clusters on which we have worked, has been entirely elaborated by Maturi et al. (2018). So, this step of the work is not our original work. We did simply exploit this sample for our purposes. Every information about the analysis coming from these data will be detailed into Chapter 6. For now, we will focus on the KiDS galaxies and on the clusters extracted by AMICO.

5.2.1 Galaxy catalogue

We have discussed the weak lensing distortions of the background sources signal by clusters in Section 4.5. The sources situated behind galaxy clusters look distorted by the gravitational potential of the intervening matter. We are interested in the distorted shapes of this background sources because they enclose the information about the matter lens distribution. We are also interested in the source redshifts which weigh the lensing information (the distortion). Here, we will simply review the main properties of the galaxies selected from the survey.

The KiDS galaxy catalogue provides the spatial coordinates, the 2 arcsec aperture photometry in four bands (u, g, r, i) and photometric redshifts for all galaxies down to the 5σ limiting magnitudes of 24.3, 25.1, 24.9 and 23.8 in the four bands, respectively, plus other quantities describing galaxies that we did not use. We chose to avoid the use of galaxy colors to reach a selection function that is as little as possible sensitive to the presence (or absence) of the red-sequence of clusters (see Maturi et al., 2018). For more precise information about the shear and redshift properties of the entire galaxy catalogue see de Jong et al. (2017) and Hildebrandt et al. (2017).

Galaxy shape measurements

The original KiDS data shear analysis was described in Kuijken et al. (2015) and Hildebrandt et al. (2017). The shape measurements was performed with an algorithm based on likelihood model-fitting called LENSFIT (Miller et al., 2007, 2013) and successfully used with KiDS in Fenech Conti et al. (2017), and in the analysis of other datasets such as CFHTLenS (Canada-France-Hawaii Telescope Lensing Survey Miller et al., 2013), for example. In this work, the data used for shape measurements are those in the r-band, being the ones with better seeing properties and with highest source density. The error on the multiplicative shear calibration, estimated from simulations with LENSFIT and benefiting from a self-calibration, is estimated on the order of 1% (Hildebrandt et al., 2017). The final catalogue gives shear measurements for ~ 15 million of galaxies, with an effective number density of galaxies of $n_{eff} = 8.53$ galaxies arcmin^{-2} (following the definition by Heymans et al., 2012), over a total effective area of $360\,\mathrm{deg}^2$.

Galaxies photometric redshifts

The original properties of photometric redshifts (photo-z) of KiDS galaxies are described in Kuijken et al. (2015) and de Jong et al. (2017). The photo-zs was extracted with BPZ (Benìtez, 2000; Hildebrandt et al., 2012), a bayesian photo-z estimator based upon template fitting, from the 4-band (u, g, r, i). Furthermore, BPZ returns a photo-z posterior probability distribution function which AMICO has fully exploited. When compared with spectroscopic redshifts from the GAMA (Galaxy And Mass Assembly, Liske et al., 2015) spectroscopic survey of low redshift galaxies, the resultant accuracy is $\sigma_z \sim 0.04(1+z)$, as shown in de Jong et al. (2017).

5.2.2 Cluster catalogue

As already mentioned, the galaxy cluster catalogue we use was extracted from KiDS-DR3 with AMICO (which is described in Section 3.6). In particular, we use the same identical cluster

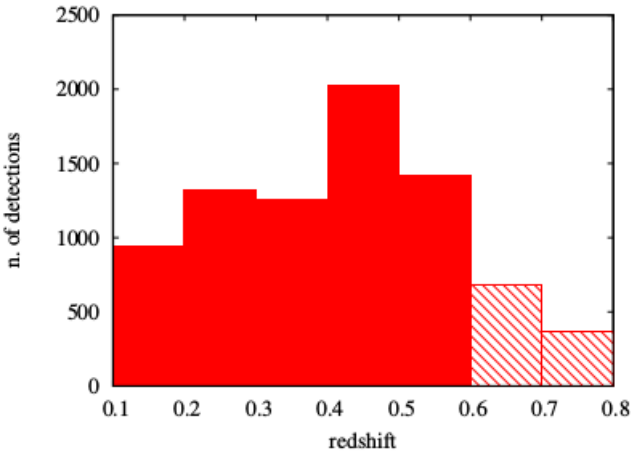

Figure 5.2: Redshift distribution of the cluster sample used in this work. The clusters have been detected by AMICO from KiDS-DR3 data. The striped bins referred to clusters not used in the analysis of this thesis. *Credits*: Bellagamba et al. (2019)

catalogue utilised by Bellagamba et al. (2019) and, consequently, its full description and validation can be found in Maturi et al. (2018). Here, we present the main characteristics and properties of the extracted cluster catalogue. Let us premise that the cluster detection through AMICO on KiDS-DR3 is an enhancement of the work by Radovich et al. (2017) on DR2, both in terms of total covered area (438 deg^2 against 114 deg^2) and of detection algorithm with respect to the previous matched filter method (Bellagamba et al., 2011). From the initial covered 438 deg^2, all clusters belonging to those regions heavily affected by satellite tracks, haloes created by bright stars, and image artefacts in general are rejected (Maturi et al., 2018). Furthermore, we select only signals with S/N > 3.5, for the final catalogue of clusters[1]. Finally, we narrowed the redshift range, where the objects fall, at $0.1 < z < 0.6$ for a final sample of 6972 galaxy clusters. Objects at $z < 0.1$ are discarded because of the reduced lensing power, while those at $z > 0.6$ because the background galaxies density in KiDS data is too small to consent a robust weak lensing analysis. Furthermore, collecting clusters with $z < 0.6$ we can robustly separate the clustered objects from the population of background galaxies.

AMICO searches for cluster candidates by convolving the 3D galaxy distribution with a redshift-dependent filter, produced as the ratio between a cluster model and a noise one. Such a convolution is able to create a 3D amplitude map, where every peak constitutes a detection. Then, for each identification, AMICO returns angular position, redshift, S/N and the *signal amplitude* A. We remind that A is a measure, through an optimal filter, of the cluster galaxy abundance in units of the cluster model, which is built starting from the known data. In Bellagamba et al. (2018) it has been demonstrated, using simulations, that the amplitude is a well-founded mass proxy, provided that the model calibration is correct. Its definition is given in Eq. 3.6.1. We rewrite such definition of the *signal amplitude* as:

$$A(\theta_c, z_c) \equiv \alpha^{-1}(z_c) \sum_{i=1}^{N_{gal}} \frac{M_c(\theta_i - \theta_c, m_i) p_i(z_c)}{N(m_i, z_c)} - B(z_c), \tag{5.2.1}$$

where M_c is the cluster model (i.e the expected density of galaxies per unit magnitude and solid angle) at the cluster redshift z_c, N is the noise distribution, $p_i(z)$, θ_i and m_i are the photometric redshift distribution, the sky coordinates and the magnitude of the i-th galaxy, respectively. Instead, α (Eq. (3.6.9)) and B (detailed in Subsection 3.6.1) are redshift dependent functions providing the exact normalisation and the background subtraction, respectively.

[1]This choice of minimum S/N is indispensable for realising well-founded mock catalogues and derive solid statistical properties for this kind of cluster samples. The method used to assess the quality of the detections is totally described in Section 6.1 of the Maturi et al. (2018) work, where is described the method used to produce realistic mock catalogues, constructed from the real data themselves. This mock catalogue production method is used to estimate the uncertainties on the quantities characterising the detections, as well as the purity and completeness of the entire sample.

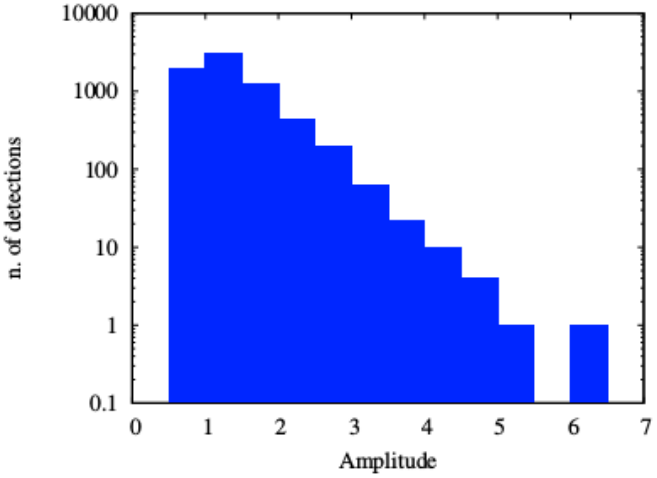

Figure 5.3: Amplitude distribution of the cluster sample, detected by AMICO from KiDS-DR3 data. The selection is carried out with a S/N threshold, hence the amplitude threshold is not constant in redshift. This produces the increase in cluster counts from the first to the second bin in amplitude.

Note that the cluster model M_c is described by a luminosity function and a radial profile, obeying to the formalism exposed in Bellagamba et al. (2018). The M_c parameter was derived from the observed galaxy population of clusters detected through the SZ-effect (Hennig et al., 2017), as described in detail in Maturi et al. (2018). Furthermore, AMICO attributes to each galaxy in a sky region a probability to be part of a given detection, defined as (see also Eq. 3.6.4):

$$P(i \in j) \equiv P_{f,i} \frac{A_j M_j(\theta_i - \theta_g, m_i) p_i(z_j)}{A_j M_j(\theta_i - \theta_j, m_i) p_i(z_j) + N(m_i, z_j)}, \tag{5.2.2}$$

where A_j, θ_j and z_j are the amplitude, the sky coordinates and the redshift of the j-th detection, respectively. $P_{f,i}$ is the field probability of the i-th galaxy before the j-th detection is defined. Indeed, in this application to KiDS-DR3 data, AMICO has considered only galaxy coordinates, r-band magnitude and the full photometric redshift distribution $p(z)$.

The distributions in redshift and amplitude of this sample of 6972 clusters are shown in Figure 5.2 and 5.3 , respectively. In the following analysis on shear data (see Chapter 6) we will divide this sample in three redshift bins:

- $0.1 \leq z \leq 0.3$,
- $0.3 \leq z \leq 0.45$,
- $0.45 \leq z \leq 0.6$.

5.3 Weak lensing profiles

Before proceeding, we refer the reader to Chapter 4 for a general overview of gravitational lensing and, in particular, of weak lensing by clusters (Section 4.5 and 4.6).

5.3.1 Background sources selection

To be able to collect the *shear profile of a given lens* (related to the density profile of the lens that, with its gravitational potential, perturbs the background signal distribution), first of all it is necessary to select the background galaxies, i.e sources positioned behind the lens and whose observed shape is distorted by the intervening mass distribution. If some galaxies belonging to the cluster, or positioned in the cluster foreground, are erroneously considered as background galaxies, the measured lensing signal can be crucially diluted (see, e.g Broadhurst et al., 2005;

Medezinski et al., 2007). A natural intuitive approach would be to consider all galaxies as accettable background sources, weigh them in accordance with their redshift probability distribution, and then deal with them individually to avoid the inclusion of some cluster members (Gruen et al., 2014; Melchior et al., 2017). However, following Sereno et al. (2017), we choose to apply a more conservative approach, aiming at eliminating foreground and cluster galaxies from the catalogue by executing cuts based on photometric redshift distributions and colours. To this purpose, a first selection is carried out to exclude the galaxies whose most likely redshift is not considerably higher than the lens one. Precisely, galaxies with photometric redshift z_s pass this cut if:

$$z_s > z_l + \Delta z, \tag{5.3.1}$$

where z_l is the lens redshift and Δz is conservatively set to 0.05, as this uncertainty is similar to the typical ones on photometric redshifts in the galaxy catalogue and, moreover, satisfactorily larger than the uncertainty on the cluster redshifts (Bellagamba et al., 2019).

A galaxy gains access to the background galaxy catalogue if it exceeds this preliminary selection and satisfies at least one of the two following criteria, based on photo-zs and on colours respectively.

Photo-z selection

This selection criterion is based on the photometric redshift distribution $p(z)$ and aims at selecting and maintaining only galaxies with a plausible $p(z)$ and which have a negligible probability of lying at redshift equal to or lower than that of the cluster. So the following selections are performed:

1. $z_{s,min} > z_l + \Delta z$
2. ODDS≥ 0.8
3. $0.2 \leq z_s \leq 1.0$.

The $z_{s,min}$ value is the lower boundary of the region that includes the 2σ (94.5%) of the probability density distribution. Requiring this value to be bigger than $z_l + \Delta z$, galaxies with nominal redshift higher than the lens one, but with a non-negligible probability of being at a redshift lower than or equal to the one of the lensing cluster, are automatically excluded. The ODDS parameter instead quantifies which fraction of the $p(z)$ is concentrated around the peak value, exploiting this threshold to exclude distributions with significant secondary solutions or wide tails (Coe et al., 2006; Bellagamba et al., 2012).[2] The third selection includes the range for which we expect the most reliable source photometric redshift estimations, given the available set of bands (u g r i), and which therefore excludes galaxies whose redshift estimation is less robust.

Colour selection

The main disadvantage of the photo-z selection just explained is the exclusion of a considerable fraction of galaxies situated behind the cluster, since they would be helpful to enhance the shear signal. This drawback is due to the stringent requests on the properties of the p(z) distribution. Some part of these excluded galaxies are recovered, while maintaining a very low contaminants fraction, by exploiting a selection based on galaxy colours (as done by Medezinski et al., 2010, 2018).

To be precise, the cut performed is:

$$(g - r < 0.3) \text{ OR } (r - i > 1.3) \text{ OR } (r - i > g - r),$$

[2]The measure of the integral of the $p(z)$ of a galaxy, in the close proximity of the most prominent peak, is able to provide an indication of the reliability of the redshift estimation. This can be done via the ODDS parameter, calculated by the BPZ estimator, which is defined as

$$ODDS = \int_{z'-2\Delta_z}^{z'+2\Delta_z} p(z),$$

where z' is the most probable redshift value, and Δ_z is a free parameter, set by default in BPZ to $0.067(1+z')$. This 0.067 factor represents the typical weight of the systematic uncertainties of this kind of measures. Thus an ODDS parameter is a measure of how much the distribution is concentrated around the principal peak, i.e it quantifies the relative importance of the most likely redshift (Hildebrandt et al., 2012). This is indeed a powerful way to assess the goodness of the estimate. Furthermore, the fraction of galaxies with a very bad redshift estimation decreases rapidly as ODDS increases.

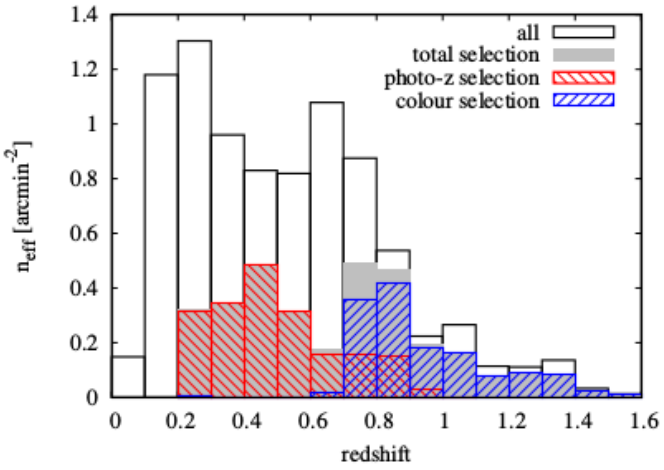

Figure 5.4: Effective number density of galaxies as a function of redshift of the total shear catalogue (black) and of the selected background sources (filled grey). The red and blue histograms referred to the samples selected according to the two considered criteria, based on photo-zs and colours, respectively. *Credits:* Bellagamba et al. (2019)

because these criteria efficiently select galaxies at $z_s > 0.7$. This cut was firstly proposed by Oguri et al. (2012) for the properties of the galaxies belonging to the COSMOS catalogue (Ilbert et al., 2009) and selecting galaxies at $z \geq 0.7$. It was afterward tested and utilized by Covone et al. (2014) and Sereno et al. (2017). In particular, the latter showed that the galaxies with high quality spectroscopic redshifts, detected with this cut in the CFHTLS-W1 and W4 fields with $z_{spec} > 0.63$, amount to 97%. This selection delivered an effective number density of ~ 3.15 arcmin^{-2} of useful background galaxies. The distribution of the selected sources as function of z_s is shown in Figure 5.4, where we can see how many background galaxies, excluded by the photo-z selection, are recovered by the colour selection.

5.3.2 Measuring the shear profile

In Subsection 4.2.1 we discussed how to compute, from the lensing distortion angle, the surface mass density of the lens (see Eq. (4.2.3)).

In the weak lensing regime, the shear tangential distortion of background galaxies (see sections 4.3 and 4.4 for a complete definition and derivation of the shear) is related to the lens *differential (projected) surface density profile* (see e.g. Sheldon et al., 2004) via:

$$\Delta\Sigma(R) = \bar{\Sigma}(< R) - \Sigma(R) = \Sigma_{crit}\gamma_T, \tag{5.3.2}$$

where γ_T denotes the *tangential component* of the *shear* (see Eq. 4.4.12), $\Sigma(R)$ is the mass surface density at radius R (Eq. 4.2.3), and $\bar{\Sigma}(< R)$ is its mean inside radius R. The tangential shear distortion γ_T is the main observable quantity for weak lensing studies. Σ_{crit} is the critical surface density for lensing (already seen in Eq. 4.2.8):

$$\Sigma_{crit} \equiv \frac{c^2}{4\pi G}\frac{D_s}{D_l D_{ls}}, \tag{5.3.3}$$

that encodes, for each single source redshift, the geometry of the lens-source system because it depends on the relative angular diameter distances between observer and source (D_s), observer and lens (D_l), and lens and source (D_{ls}).

For every single lens, we take account of all the background galaxies selected according to the method described in Subsection 5.3.1, and then we compute the shear signal along the tangential direction with respect to the centre of the considered cluster. In other words, we determine, in each radial bin, the weighted ellipticity of the background sources, i.e the tangential and cross component of the shear. We put in evidence the fact that the shear γ is actually inaccessible to observations. Concretely, in fact, we measure and estimate the *reduced shear* $g = \gamma/(1-\kappa)$, where

$\kappa = \Sigma/\Sigma_{crit}$ (see Subsections 4.3.2 and 4.4.1). However, in the *weak lensing regime*, we know that $\kappa \ll 1$ hence we can approximate $\gamma \approx g$. This allows us to build the observed $\Delta\Sigma$ profile by computing:

$$\Delta\Sigma(R_j) = \Big(\frac{\sum_{i\in j}(w_i\Sigma^{-2}_{crit,i})\gamma_{T,i}\Sigma_{crit,i}}{\sum_{i\in j}(w_i\Sigma^{-2}_{crit,i})}\Big)\frac{1}{K(R_j)}, \tag{5.3.4}$$

where j is the considered radial annulus with mean radius R_j and w_i is the weight assigned to the measurement of the source ellipticity (Sheldon et al., 2004). The function $K(R_j)$ is the average correction due to the multiplicative noise bias in the shear estimate and is computed as:

$$K(R_j) = \frac{\sum_{i\in j}(w_i\Sigma^{-2}_{crit,i})m_i}{\sum_{i\in j}(w_i\Sigma^{-2}_{crit,i})}, \tag{5.3.5}$$

where m_i represents the *multiplicative noise bias* for the i-th galaxy (for more detalis, see Fenech Conti et al., 2017).

We utilize the most probable redshift of the source, extracted with the BPZ, to calculate the critical density for the i-th galaxy $\Sigma_{crit,i}$. To evaluate and estimate the systematic uncertainty related to the source redshift, we use the full $p(z)$ (the systematic uncertainties of this method are listed and described in the Section 6.2). On the contrary, instead, we ignore the uncertainty in the lens redshift because is tipically of the order of $\sim 0.02(1+z)$ (Maturi et al., 2018) and, furthermore, its impact on Σ_{crit} is negligible with respect to the source one.

For our purposes, the excess surface density of clusters will be collected and estimated in two complementary and consecutive steps. Firstly, we estimate the stacked $\Delta\Sigma$ for 14 logarithmically spaced radial bins in physical units from 0.1 to 3.16 Mpc/h, for every cluster redshift-amplitude bin. This range corresponds to the main cluster halo profile, whose radial density profile is well described (but only up to its virial radius of ~ 1 Mpc/h) by a NFW profile (see Subsection 4.5.1). This step follows the analysis by Bellagamba et al. (2019) and it was executed to estimate the parameters of the model describing the excess surface density profiles of our clusters. Secondly, we extend our measurement up to 35 Mpc/h using 30 logarithmically spaced radial bins, to analyse the weak shear signal coming from the so-called *2-halo term* of clusters (depicted by the cosmic web filaments), to estimate cosmological parameters. In both steps, the cluster stacking is obtained by subdividing the first two redshift bin into five increasing amplitude bins and the third one into four. Note that every amplitude bin corresponds to a different cluster model on which the stacking was performed and, ultimately, to a different mass bin of the cluster. We will describe in detail the surface density modelling in Section 5.4.

5.3.3 Stacking the shear signal

For most of the clusters in the extracted sample, the S/N in the shear data is too low to constrain the density profile and, consequently, to measure a reliable mass or, however, the other model parameters. Consequently, we exploit the *shear-cluster correlation* (or stacked lensing), i.e the cross-correlation between the cluster (lens) density distribution and the shapes of the source galaxies. The results coming from the stacked shear profiles, and the shear profiles themselves, will be shown in Chapter 6 where, furthermore, we will see that exists an evident correlation between the excess surface density and the amplitude, in each redshift range. In practice, we obtain radial profiles as a function of the physical length scale by stacking the signal of clusters grouped according to their amplitude A. For this purpose, we have to define a weighted sum of the lensing signal for classes of objects as we are going to show.

The differential density profile for the K-th cluster bin is evaluated as:

$$\Delta\Sigma_K(R_j) = \frac{\sum_{k\in K} W_{k,j}\Delta\Sigma_k(R_j)}{\sum_{k\in K} W_{k,j}}, \tag{5.3.6}$$

where k runs over the clusters in the A bin, and $W_{k,j}$ is the total weight for the j-th radial bin of the k-th cluster,

$$W_{k,j} = \sum_i w_i\Sigma^{-2}_{crit,i} \tag{5.3.7}$$

and i runs over the background galaxies in the j-th radial bin.

The *covariance matrix* (that estimates variance and covariance, see appendix B) associated with the extraction of the shear signal and its relative stacking are computed with a *bootstrap procedure with replacement* (described in appendix C): the source catalogue was re-sampled 10000 times. Covariance also considers the residual contribution from large-scale projections but, however, it is negligible thanks to the large number of line of sight we stacked over. So every error associated with each stacked surface density profile point is a bootstrap error.

5.4 Modelling the shear profile

Our modelling of the lens shear profiles, through which we will constrain cluster and cosmological parameters, is splitted in two steps. The first one is performed on the central halo radial range, i.e up to $\sim 3\times$ the radius of the virialized matter halo, whose density profile is described by the NFW model seen in Subsection 4.5.5, modified to be a truncated NFW profile. The NFW profile, in lensing studies, gives analytical expression for the radial profile of the deflection angle, convergence and shear (Bartelmann, 1996), but has the drawback that the enclosed mass diverges logarithmically. Therefore we choose the truncated version. In this step, we will constrain values for four model parameters describing the main physical properties of the lens cluster.

The second step of this work, instead, is focused on the outskirts of the same clusters (far beyond the virial radius $r_{vir} \approx r_{200}$, up to $\sim \times 30$ this radius), to estimate the cosmological parameters. Here, the signal is produced by the matter in correlated cluster haloes which forms the so-called 2-halo term. Therefore our analytic model assembles the main halo model, described by a NFW truncated model, with that for the correlated matter around the main halo (2-halo term).

In the first step, we infer the parameters of the truncated NFW model describing the main halo which we keep fixed in the second step, as we will detail in Chapter 6, to infer cosmological parameters thanks to the 2-halo contribute. In both steps, the analysed lensing signal is in the weak lensing regime (described in Sections 4.5 and 4.6) and all the statistical analysis, described in Chapter 6, follows a Bayesian approach (consult appendix A). In this section we will describe the models we will use to analyse the signal.

5.4.1 Main halo lens model

The stacked radial density profile of all our clusters is modelled by a *smoothly-truncated NFW density profile* (an analytical hybrid model, tested in Baltz et al., 2009):

$$\rho(r) = \frac{\rho_s}{r/r_s(1+r/r_s)^2}\left(\frac{r_t^2}{r^2+r_t^2}\right)^2, \tag{5.4.1}$$

where ρ_s is the typical matter density (whose distribution, in a cluster halo, follows that generated by the potential well of the dark matter), r_s is the scale radius and r_t the truncation one, defined as $r_t = t \times r_s$, with t the truncation factor. The scale radius is commonly parametrized as $r_s = \frac{r_{200}}{c_{200}}$, with c_{200} that represents the concentration parameter (which is correlated with the halo mass and redshift). As already mentioned, the normalisation of this model can be expressed in terms of M_{200}. Note that the density parameter ρ_s is connected to the virial mass $M_{vir} \approx M_{200}$ (see Subsection 4.5.1) defined such that the average density within the virial radius becomes equal to the non-linear overdensity $\Delta_{vir} \approx \Delta_{200}$ (computed using the spherical collapse model described in Subsection 2.5.1), multiplied by the mean matter density of the Universe. In the strict sense, however, M_{vir} differs from a true virial mass defined from non-linear overdensities but, choosing it as normalisation, the "virial mass" of this model becomes rather close to the effective total mass (Oguri and Hamana, 2011).

This truncated version of the NFW model (see Eq. (4.5.5) for the "normal" NFW model) was well tested in simulations by Oguri and Hamana (2011), estabilishing that it better describes the cluster profiles very far beyond the virial radius (precisely $\sim \times 10$ this radius) with respect to the original NFW profile. In effect, at large radii $\rho(r) \propto r^{-7}$ and therefore the enclosed mass converges immediately. Indeed, one of the main advantages of this model is the truncation, as it removes the unphysical divergence of the total mass for large radii. Moreover, the truncation better describes the transition between the cluster main halo and the 2-halo contribute. Thus, the truncated NFW model produces less biased estimates of mass and concentration from shear profiles. Without the truncation, the mass would be underestimated and the concentration overestimated. Moreover,

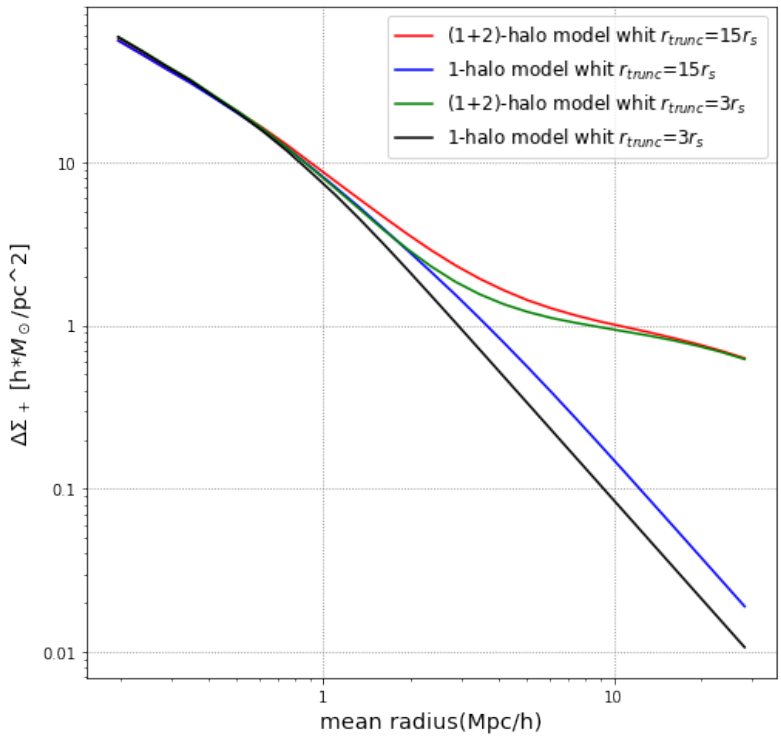

Figure 5.5: Differential projected density profiles for models with cluster parameters $M_{200} = 1.9 \cdot 10^{13} M_\odot/h$, $c = 6.8$, $f_{off} = 0.27$ and $\sigma_{off} = 0.23\,\mathrm{Mpc}/h$, at redshift $z = 0.19$, for our assumed cosmology. We can note the slight influence of a different truncation factor both on the 1-halo model and on the total (1+2)-halo model. The models with truncation radius $r_t = 3r_s$ (green and black lines) are representative of the ones used in our analysis. As an illustrative example, we also plot the models with unrealistic $r_t = 15r_s$ (red and blue lines) to illustrate the effect of this parameter on the 1-halo and (1+2)-halo models.

note that we are studying stacked shear profiles, hence the NFW model is a powerful fitting model as it is based on the assumption of spherical symmetry. When stacking various individual shear profiles the intrinsic ellipticity of each cluster tends to assume a spherical radial symmetry on the stacked profile, for pure statistical reasons. This is the reason why the weak lensing cross-correlation provides a *direct* estimate of the mean mass for clusters with some observable properties. We reasonably set $r_t = 3r_s$ for every amplitude bin as done by Bellagamba et al. (2019), following the results by Oguri and Hamana (2011). In Figure 5.5 we can see the effect of the truncation radius on the modelling.

As seen in Subsection 4.5.2, when analysing the weak lensing by clusters, an important source of bias is the wrong identification of the lens centre. In this analysis, we assume as the centre of the clusters those determined by AMICO in the detection procedure. However, the detection is performed through a grid which induces an intrinsic uncertainty due to the pixel size, that is <0.1 Mpc$/h$ (Bellagamba et al., 2018). Furthermore, it must be taken into account that the galaxy distribution centre may appreciably deviate from the mass centre of the system, e.g. because of ongoing merging. For example, in Johnston et al. (2007) it was found that the Brightest Central Galaxy (BCG), defining the cluster centre, might be misidentified and, in that case, the value of $\Delta\Sigma$ at small scales is underestimated, biasing low the measurement of the concentration c by $\sim 15\%$ and consequently undersestimating the mass by $\sim 10\%$. Moreover, George et al. (2012) found that the halo mass estimates from stacked weak lensing can be biased low by 5%-30% if inaccurate centers are considered and the issue of off-centring is not addressed.

We decide to parametrize this uncertainty by considering a second additional component to our model, which is the one produced by haloes whose observed centre has a non-negligible scatter with respect to the centre assumed by the stacking procedure, following the method adopted by Johnston et al. (2007) and Viola et al. (2015). We first define the *rms of the misplacement on the sky plane distribution of the haloes* σ_{off}, and we assume an azimuthally symmetric Gaussian

distribution. Thereby, the probability of a lens being at distance R_s from the assumed centre, or briefly the *probability distribution of the offsets*, can be defined as:

$$P(R_s) = \frac{R_s}{\sigma_{off}} \exp\Big[-\frac{1}{2}\Big(\frac{R_s}{\sigma_{off}}\Big)^2\Big], \tag{5.4.2}$$

where σ_{off} can be seen as a scale length, whose typical values are of the order of $\sigma_{off} \sim 0.4$ Mpc$/h$ (Johnston et al., 2007). During the modelling of the halo profile, we will assume that a *fraction* f_{off} of the lenses is miscentred. Now, we can introduce the *azimuthally averaged profile of a population misplaced by a distance* R_s in the lens plane (Yang et al., 2006):

$$\Sigma(R|R_s) = \frac{1}{2\pi}\int_0^{2\pi} \Sigma_{cen}(R)\Big(\sqrt{R^2 + R_s^2 + 2RR_s\cos\theta}\Big)\mathrm{d}\theta, \tag{5.4.3}$$

where $\Sigma_{cen}(R)$ refers to the centred surface brightness distribution (or briefly the centred profile) derived by integrating $\rho(r)$ along the line of sight. Finally, integrating Eq. (5.4.3) along R_s and weighing each offset distance according to Eq. (5.4.2), we can obtain the *mean surface density distribution of a miscentred haloes population*:

$$\Sigma_{off}(R) = \int P(R_s)\Sigma(R|R_s)\mathrm{d}R_s. \tag{5.4.4}$$

From now on, we refer ourselves to the cluster main halo term as *1-halo term*. Our final model for the small scales, where the shear profiles are dominated by the dark haloes contribute associated with the clusters, can be written as the sum of a centred population and an off-centred one:

$$\Sigma_{1h}(R) = (1 - f_{off})\Sigma_{cen}(R) + f_{off}\Sigma_{off}(R). \tag{5.4.5}$$

Generally, this 1-halo model depends on nine parameters but, when analysing the 1-halo radial range, we fix the cosmology to a flat Λ-CDM with $H_0 = 70km/s/Mpc$, $\Omega_m = 0.3$ and $\Omega_\Lambda = 0.7$, the truncation factor to 3 and the redshift to the effective redshift estimated by the stacking procedure for every cluster bin. Therefore our analysis on the 1-halo radial range uses a model that depends on four free parameters (M_{200}, c, σ_{off} and f_{off}).

Moreover, even if we did extract a shear signal starting from a radial distance of 0.1 Mpc/h from the cluster effective centre, in the following analysis we will neglect the first three radial bins, choosing to consider only the shear data in the range $0.2 \leq R \leq 3.16$ Mpc/h. There are three reasons for this choice. Firstly, from an observational point of view, the measure of photo-zs and shear close to the cluster centre is made very complex by the contamination of a high cluster galaxy concentration. Secondly, the shear signal analysis in close proximity to the cluster centre is very sensitive to the BCG contribute to the matter distribution, and to deviations from the weak lensing approximation used in the above derivation. Lastly, this choice mitigates the miscentring effects, too. Thus this choice principally reflects on the accuracy of the estimation of the concentration c, which would otherwise be overestimated.

5.4.2 2-halo lens model

Here, we introduce the 2-halo model we will use to describe the surface density profile beyond the cluster virial radius: we have collected the shear stacked signal from 0.1 to 35 Mpc/h, in 30 radial annuli equally distanced in logarithmic space. At such scales the shear signal is produced by the correlated matter distribution around the location of galaxy clusters and it manifests itself as a 2-halo term in the mass density profiles. In practice, the 2-halo term characterises the cumulative effects of the large-scale structures in which galaxy clusters are situated. The uncorrelated matter distribution along the line of sight does not give contibution to the stacked shear signal, consequently the 2-halo term is a function of the correlated matter distribution at the same cluster redshift (Oguri and Takada, 2011).

Anyway, its contribution on the 1-halo radial range is expected to be not very influential on the lensing profile because, at these scales, is really small, as it begins to manifest itself at scales $\gtrsim 9$Mpc$/h$. We add to the halo profile described by the Eq. (5.4.5) the contribution due to the matter in correlated haloes, i.e. the already mentioned *2-halo term*, that we model following Oguri and Takada (2011) which have elaborated it, as done by Oguri and Hamana (2011) and Sereno et al. (2017):

$$\Delta\Sigma_{2h}(\theta; M_{200}, z) = \int \frac{l\mathrm{d}l}{2\pi} J_2(l\theta)\frac{\bar{\rho}_m(z)b_h(M_{200}; z)}{(1+z)^3 D_l^2(z)} P_m(k_l; z), \tag{5.4.6}$$

where z is the cluster redshift (a photometric one for our measures). Every quantity is in physical units and was computed using the CosmoBolognaLib (CBL), a large set of Open Source C++ numerical libraries for cosmological calculations described in Marulli et al. (2016). Furthermore:

- θ is the *angular radius*, in radians. It is computed as the ratio between the projected radius, at which the shear signal is measured, and the corresponding *lens angular diameter distance* $D_l(z)$ [Mpc/h], evaluated at the halo redshift and assuming our flat ΛCDM cosmology.
- J_2 is the *Bessel function of second type* that gives solutions of the Bessel differential equation that have a singularity at the origin. It is function of $l\theta$, where l is the integration variable and the momentum of the wave vector k_l foe the linear power spectrum of matter fluctuations P_m.
- The *wave vector module* is evaluated as $k_l = l/((1+z)D_l(z))$.
- $\bar{\rho}_m(z)$ is the *mean cosmic background density*, at the redshift lens, in $M_\odot Mpc^{-3}h^2$.
- $P_m(k_l; z)$ is the *linear power spectrum of matter* (see setion 2.4) and it is computed according to the Eisenstein and Hu (1999) transfer function, normalised with $\sigma_8(z=0) = 0.8$, and assuming $\Omega_{baryon} = 0.046$. This model is constructed from approximated transfer functions and does not consider the Baryonic Acoustic Oscillations (BAO). Since the observational errors on the shear measurements are quite high, a more accurate calculation of P_m, with more sophisticated models, would have had a negligible impact on the final result. For instance, we compared it with the power spectrum form of a CAMB (Code for Anisotropies in the Microwave Background, consulted at the website) model, that considers the BAO, not finding remarkable differences. Furthermore, the dependence of this power spectrum model on redshift is negligible.
- $b_h(M; z)$ is the *halo bias*, for which we followed the prescription of Tinker et al. (2010), which successfully tested and calibrated numerically this model. It is expected to be constant at large scales and to evolve very weakly with redshift. It represents the bias of dark matter haloes, defined as the ratio of the halo power spectrum and the dark matter linear power spectrum. The bias amplitude decreases at small scales, i.e when the input parameter $\Delta(z)$ (the background cosmic overdensity at given redshift) increases or, in other words, it increases with the halo mass. However, we use the virial overdensity given a critical overdensity $\Delta_{vir}(z) = \Delta_{vir}^{crit}(z)/\Omega_M(z)$, as bias input parameter. For our redshift range, this overdensity ranges between ~ 300 and ~ 400, so that the details of the truncation near the virial radius do not affect the conversion of the masses so much (Oguri and Hamana, 2011). Of course, the halo bias function is computed using the same Eisenstein and Hu power spectrum P_m of Eq. (5.4.6).

Note that the 2-halo *shear* term around a single lens of mass M at redshift z, for a single source redshift, is the real signal that we measure, and it is modelled as:

$$\gamma_{T,2h}(\theta; M, z) = \int \frac{ldl}{2\pi} J_2(l\theta) \frac{\bar{\rho}_m(z) b_h(M; z)}{\Sigma_{crit}(1+z)^3 D_l^2(z)} P_m(k_l; z), \tag{5.4.7}$$

where we can see that our excess surface density profile of Eq. (5.4.6) is simply obtained by multiplying the shear in Eq. (5.4.7) by the relative Σ_{crit} , defined in Eq.(5.3.3).

Finally, to be able to correctly compare data and model, we will add the 2-halo term contribute of Eq. (5.4.6) to the model describing the total halo surface density profile of Eq. (5.4.5) obtaining:

$$\Delta\Sigma_{model}(R) = \Delta\Sigma_{1h}(R) + \Delta\Sigma_{2h}(R). \tag{5.4.8}$$

In Figure 5.6 an example of the different contributes to the total profile model is shown. Anyway when modelling exclusively the 1-halo, as we limit our analysis to the central 3.16 Mpc/h, the 2-halo term affects to a small degree only the few last radial bins in this radial range and principally in the smallest amplitude bins (which contain preferentially low-mass clusters). This makes our inference on the four 1-halo (NFW) model parameters much less dependent on the uncertain details of the halo bias model. In Figure 5.7 we can see how the chosen cosmological model affects the $\Delta\Sigma_{1h}$ profiles, when modelling them with the truncated NFW model. Anyway, we conservatively choose to consider the 2-halo contribute on this first modelling step but it did result negligible.

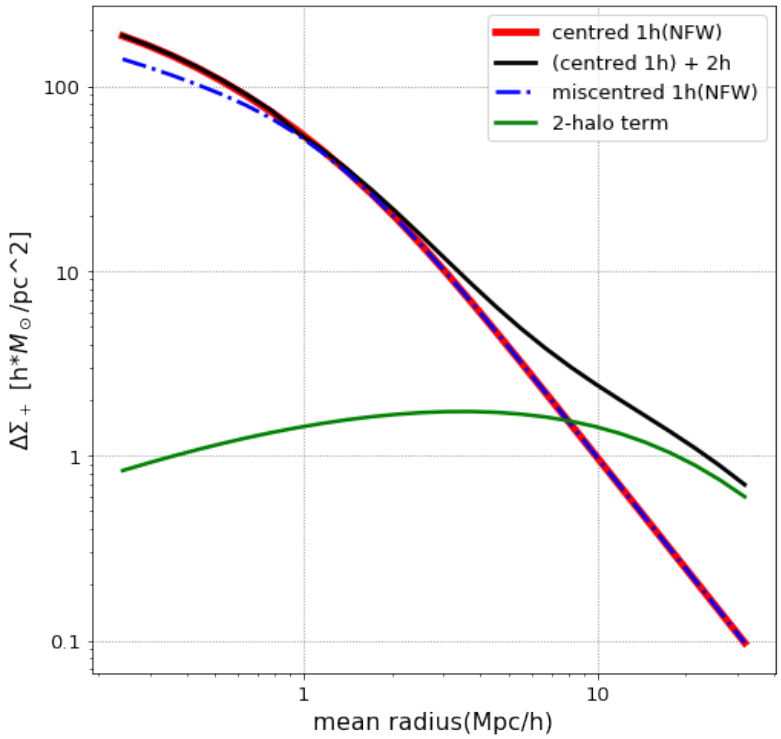

Figure 5.6: Differential projected density profile for a model with cluster parameters $M_{200} = 1.72 \cdot 10^{14} M_{\odot}/h$, $c = 3.25$, $f_{off} = 0.28$ and $\sigma_{off} = 0.29\,\mathrm{Mpc}/h$, $r_{trunc} = 3r_s$, at redshift $z = 0.23$, for our assumed cosmology. The 1-halo contribute dominates at $r \lesssim 10\,\mathrm{Mpc}/h$. The dash-dotted blue curve refers to a 1-halo model with $f_{off} = \sigma_{off}$=0: the miscentring causes a slight flattening of the profile at small scales ($r < 1\,\mathrm{Mpc}/h$).

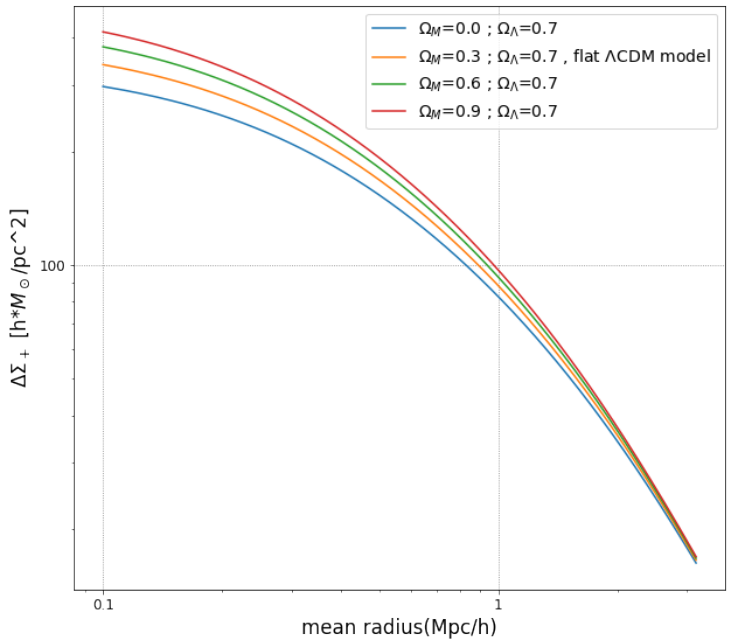

Figure 5.7: Truncated NFW profile models, for a cluster at $z = 0.5$, with $c = 2.7$, $M_{200} \approx 3 \cdot 10^{14} M_{\odot}/h$, $f_{mis} = 0.2$ and $\sigma_{off} = 0.2$, for different cosmological models, as indicated by the labels.

When we will pass to the second analysis step, we will extend the shear stacked data analysis performed on the 1-halo radial range to the larger radial range for the same amplitude and redshift bins. The first analysis step will be done to obtain a robust inference of the 1-halo four model parameters, to be able to fix them in the second analysis step, on the total halo surface density profile. There, we will infer the cosmological parameter Ω_M, as the 2-halo term is made principally by dark matter and contains much cosmological information. Among other things, cosmological interpretation of clusters relies on good estimations of their masses. We remind that gravitational lensing studies provide direct reliable measurements of the halo mass without assuming a scaling relation. So the estimate of M_{200} can be safely used to constrain cosmological parameters. Furthermore, to be able to compare the goodness of our inference of $\Delta\Sigma_{1h}$ model parameters, we have chosen the same analysis setting of Bellagamba et al. (2019), both in the splitting in amplitude A and in redshift, as we will see in Chapter 6, obtaining a perfect agreement. The results we are confronting with can be consulted in detail in Figure 8 (or, in this thesis work, in Figure 4.4) and table 2 of the work of Bellagamba et al. (2019). Also in the second step the analysis of the shear stacked data, extracted up to a 35 Mpc/h radius, will be executed on the model described by Eq. (5.4.8). Here, we will free the Ω_M parameter and study the correlation between Ω_M and M_{200} and c, to see how the halo terms of Eq. (5.4.8) affect each other and to test the reliability of the cosmological parameter inference. Having seen how the contribution of f_{off} becomes negligible already on scales >0.4-0.5 Mpc/h, in this second modelling we will fix the miscentring to the values obtained on the first analysis on the 1-halo.

Note that the highest S/N comes from small-scale lensing in this kind of measures, but it may be affected by systematic uncertainties, both in terms of theoretical interpretation and observational uncertainties (Mandelbaum et al., 2013). Anyway, we do not aim at connecting directly this small scale lensing to the cosmological background. We will just model it in terms of a physically motivated model, i.e the truncated NFW profile. So, in the context of the second step analysis, the four parameters inferred from the 1-halo can be considered as nuisance parameters, that we marginalize over to estimate the amplitude of the 2-halo term, as done by Sereno et al. (2015), for example.

Chapter 6

Estimates of the cosmological parameters

In this Chapter we will present the method used to infer the model parameters from the survey shear data and the resulting cosmological constraints. This is done through the model fitting of the KiDS-weak lensing data, as discussed in Section 5.4. Precisely, a Bayesian approach (see appendix A) is applied to obtain the statistics of interest. We will split this discussion following the two steps of the shear signal extraction and of the consequent modelling. Firstly we model the stacked shear profiles collected in the $0.1 < R < 3.16$ Mpc/h radial range, and then we model the profiles measured up to 35 Mpc/h (where the 2-halo term contribute dominates). Both steps of the analysis are based on a set of 14 stacked excess (differential) surface density profiles of clusters, obtained dividing the whole AMICO-KiDS cluster catalogue as follows:

- first redshift bin ($z = [0.1-0.3]$), split in 5 amplitude bins: $A = [0-1]$, $[1-1.55]$, $[1.55-2.05]$, $[2.05-2.75]$, $[2.75-6.5]$;
- second redshift bin ($z[0.3-0.45]$), split in 5 amplitude bins: $A = [0-1.15]$, $[1.15-1.65]$, $[1.65-2.3]$, $[2.3-3]$, $[3-6.5]$;
- third redshift bin ($z = [0.45-0.6]$), split in 4 amplitude bins: $A = [0-1.3]$, $[1.3-1.8]$, $[1.8-2.3]$, $[2.3-6.5]$.

This is the same division used in Bellagamba et al. (2019), and the cluster masses, estimated from the smoothly truncated NFW model, we obtain are in excellent agreement with the just mentioned work. Of course, every amplitude bin refers to a typical mass and/or richness of the cluster model. For increasing A, the typical mass and richness of the clusters grow, and as a consequence a smaller number of clusters is needed to obtain a significant shear signal. For this reason, the error on the shear profile data decreases with A and slightly grows with redshift (due to the decrease in the signal strength, as evident from Figures 5.2 and 5.4). For example, in the first redshift bin we stack 1066 clusters in the lower amplitude bin and only 41 clusters in the higher amplitude bin. Both measurements have been performed assuming a flat ΛCDM cosmology with $H_0 = 70 km/s/Mpc$ and $\Omega_\Lambda = 1 - \Omega_M$. Halo masses are given as M_{200}, i.e the halo mass, for a given cosmology, enclosed in a sphere of radius r_{200}, where the mean density is 200 times the critical density of Universe at the corresponding redshift (see Section 3.4). We will see that this M_{200} determines the normalisation of the density profile. Furthermore, during all the analyses we do not consider the error on the mean value of each radial point of the shear profiles.

This Section is organized as follows:

- In Section 6.1 we present and discuss the method used to infer the 1-halo model parameters, with a focus on the cluster mass, from the shear profiles stacked up to 3.16 Mpc/h from the cluster centres. This stage is necessary to fix the quantities describing the clusters, useful to extract cosmological information from the following study of the stacked shear signal collected at larger radii and modelled by the 2-halo term.
- In Section 6.2 we discuss the systematic uncertainties arising when inferring the mass-observable relation from weak lensing shear data, quantifying in detail each individual contribution affecting the mass estimate.

- In Section 6.3 we deal with the mass-amplitude relation, showing the narrow scaling relation between the two quantities which allows us to confirm the reliability of our mass measures.
- In Section 6.4 we present method and results of the analysis, performed on the stacked shear data collected up to 35 Mpc/h from the cluster centres, when the modelling is applied to infer exclusively the matter density parameter. This stage of the analysis aims at testing the reliability of our method, fixing the cluster parameters values to those measured in Section 6.1.
- In Section 6.5 we deal with the final results obtained with the same analysis performed as in Section 6.4, but in this case the modelling is used to constrain the matter density parameter together with mass and concentration of clusters. We show the final marginalized estimation of Ω_M, confronting it with the most recent cosmological measures and we discuss the reliability of our results.
- Finally, in Section 6.6 we discuss the possible systematic uncertainties arising when inferring Ω_M from weak lensing analysis on the scales corresponding to the 2-halo term.

6.1 Inference of model parameters from 1-halo radial range

We have a dataset D of stacked differential surface density profiles $\Sigma_{obs}(R)$ (from the collected shear measurements), derived as in Section 5.3, and a model $\Sigma_{mod}(R)$, described in Section 5.4, aiming at describing these data. Here, our goal is to infer the parameters of the model describing stacked shear data extracted in the 1-halo radial range, i.e $0.1 \leq R < 3.16$ Mpc/h in 14 equally spaced logarithmic bins.

In the context of our Bayesian approach, the posterior probability distribution of the set of our 1-halo model parameters $\theta = (M_{200}, c, f_{off}, \sigma_{off})$ comes from the Bayes theorem:

$$P(\theta|D) = \frac{\mathcal{L}(D|\theta)P(\theta)}{E(D)}, \tag{6.1.1}$$

where $\mathcal{L}(D|\theta)$ refers to the *likelihood* of the data given the model, $P(\theta)$ is the *prior* probability of the parameters and $E(D)$ is the *evidence*, which is independent of the parameters and that we therefore ignore in this framework. The likelihood in this context is expressed as (see Hartlap et al., 2007):

$$\mathcal{L} \propto \exp(-\chi^2/2), \tag{6.1.2}$$

with

$$\chi^2 = \sum_{i=1}^{N_{bin}} \sum_{j=1}^{N_{bin}} [\Delta\Sigma_{obs}(R_i) - \Delta\Sigma_{mod}(R_i)] C_{ij}^{-1} [\Delta\Sigma_{obs}(R_j) - \Delta\Sigma_{mod}(R_j)], \tag{6.1.3}$$

where C is the likelihood covariance matrix (see appendix B), while $N_{\rm bin}$ is the number of radial bins in the considered profile. Moreover, for the likelihood distribution we have chosen a Gaussian error, given the statistical nature of the shear-cluster cross correlation observations. The data used in the likelihood estimation of each cluster bin are 11, instead of 14, because we exclude from the analysis the first three radial points, i.e data at a distance less than 0.2 Mpc/h, for the reasons discussed in Subsection 5.4.1. We verified that the choice on the number of radial bins does not significantly affect the likelihood magnitude and the subsequent accuracy of our final posterior parameter estimation. We tried to perform our Bayesian analysis using different radial bins in the shear extraction noticing that there is not an appreciable difference on the posterior distribution of the model parameters, as shown in Figure 6.1.

Concretely, the main contributors to the off-diagonal terms in the covariance matrix are cosmic variance and correlated shape noise caused by the fact that same background galaxies can enter in more than one radial bin in the different shear profiles used in the stacking (Mandelbaum et al., 2013; Sereno et al., 2015; Viola et al., 2015). Being this analysis on the 1-halo radial range performed as in Bellagamba et al. (2019), we expect both term to be small enough in our case, too (see Section 6.2). The cosmic variance has been minimised as much as possible by sampling numerous different line of sight, so that its average would nullify itself. We can ignore correlated shape noise because the mean radial distance between cluster centres is significantly larger than the limiting radius (~ 3 Mpc/h) for the 1-halo shear profile (we can not say the same when we extend

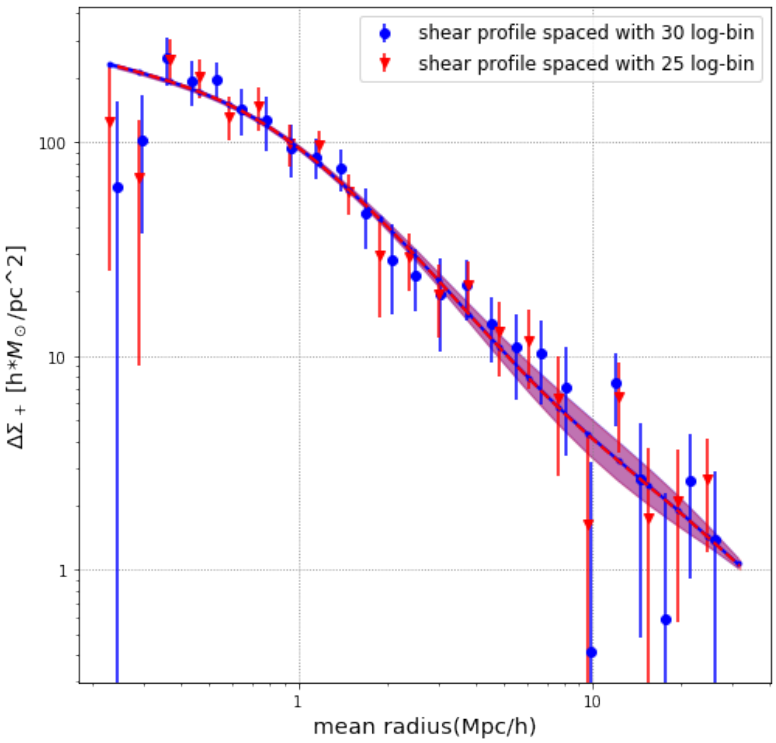

Figure 6.1: Stacked shear profile of the clusters in ranges $z = 0.1 - 0.3$ and $A = 2.75 - 6.5$, for two different radial binning extractions, as indicated by the labels. The measurements (points) are compared to best-fit posterior models (lines) with $1 - \sigma$ estimated uncertainties (shaded regions). We can see that the radial binning does not affect neither the best-fit model, nor its error.

the analysis up to very large radii). Indeed, the off-diagonal terms C_{ij} of the resulting covariance matrix are mostly compatible with a random noise over a null signal. Consequently, we decided to consider only the diagonal terms in this analysis, thus avoiding complications due to the inversion of a noisy estimate (we need C_{ij}^{-1} in Eq. (6.1.3)) of the covariance matrix (Hartlap et al., 2007). Furthermore, the values of the diagonal terms C_{ii} scale with radius as R^{-2} as expected for shape noise variance in logarithmically spaced bins.

To derive the posterior distribution of the parameters from each stacked profile, we use a Monte Carlo Markov Chain (MCMC) sampling method using the Metropolis-Hasting algorithm (see appendix A), whose computing code has been built entirely from the CBL (Marulli et al., 2016). Precisely, this algorithm exploits a code to perform a Bayesian fit to our set of data points with the model of Eq. (5.4.8). It is a code assembled starting from a simple CBL script called *fit.cpp* and built similarly to the MCMC Hammer algorithm (Foreman-Mackey et al., 2013), opportunely modified for our goals. The modified C++ code we are talking about is presented in appendix D. When the MCMC of each posterior sampling reaches the convergence we acquire our parameters distribution and build the model for the parameters.

Following Bellagamba et al. (2019), in our analysis we assume the following prior for the parameters:

- $\log(M_{200}/(M_{\odot}/h)$ uniform between 12.5 and 15.5
- c uniform between 1 and 20
- f_{off} uniform between 0 and 0.5
- σ_{off} uniform between 0 and 0.5 Mpc/h.

These priors values intervals aim to be conservative, assuming no a priori information about our cluster sample. The main results of our MCMC code is a sample of the 4-dimensional (one dimension for each parameter) posterior distributions, from which we can compute different estimators

and statistics. To be precise our code gives an output where, for every parameter posterior distribution, we obtain the mean with the relative standard error (variance), and the median with relative percentiles at 1σ (68.27%). Furthermore, this output contains the posterior surface density halo model as estimated from the inferred parameters, plus its error within 1σ. In the following, for each parameter, we quote the median with relative percentiles of the marginalized posterior distribution as the resulting typical value and its uncertainty. We choose to show the median, instead of the mean, because it is more stable in case of a skewed distribution of the posterior probability, even if in all the performed chains the mean and the median are in agreement with a relative difference ranging from less than 1% to a maximum of 6-7%.

In this phase of the Bayesian fitting on the 1-halo radial range, the modelling is performed on MCMC running with a set of 10 parallel chains constituted by 3000 chain steps each. These chains are sequences in the parameter space that represent a fair sampling of the full posterior probability distribution. Anyway, for exposing and calculation timing purposes, we choose a thin of 6 (the thin is the number with which we divide the final chain steps number to collect) and a burn-in of 20 (see the burn-in definition in Appendix A). We make this choice because we maximize the posterior starting chain value on that of the maximized likelihood, so as not to force the natural path of the chain (we do not have any a priori information about the cluster mass of every A bin). Therefore each posterior distribution, extracted through the MCMCs for our 14 bins in amplitude and redshift, is composed of 4980 statistical samples. In addition, we have performed various trials with different chains lengths, noticing that the convergence was reached even with shorter chains the one chosen. Hence about the convergence of the posterior distributions we have been conservative.

The results of this analysis are shown in Figure 6.3 and Table 6.1. From the figure we can clearly note that the signal grows with A, for each redshift bin. Consequently, M_{200}, which describes the normalisation of the profile, also grows with the amplitude, showing that A well correlates with the mass. To ascertain this, in addition to observing the $\Delta\Sigma$ values for each of the bins in A of a given bin in z, just note that as A increases, the quantity of background galaxies to stack (represented by cluster number to stack N_{cl}), necessary to have the desired S/N, decreases. Less background galaxies to stack means we have a stronger shear signal, and hence a deeper potential well that generates it. Furthermore, the reduced χ^2 values, reported in legends of Fig. 6.3, show that the model well fit the data. Anyway the percentiles represented by the filled contours are more satisfactory than a simple χ^2 because they are evaluated and showed for every radial point. Finally, from Table 6.1 we see that the inferred masses correlate with A and are affected by reasonable relative error ($\sim 15\%$).

There is an intrinsic degeneracy between the truncated NFW model parameters (M_{200} and c) and the miscentring represented by f_{off} and σ_{off}. So these two parameters remain unconstrained by the data (especially σ_{off}), as shown in Figure 6.2. Always in this figure we can note indeed the good constraints between mass and concentration. Anyway, in all the 14 bins the posterior distribution of c gives median values that range between ~ 1.5 and ~ 10.5, while between ~ 0.2 and ~ 0.3 for what concerns f_{off} and σ_{off}. Thence the only model parameter with a reliable posterior distribution is M_{200} (often a quasi-Gaussian distribution) and, accordingly, we concentrate on the results about it in the following (in table 6.1 we only show the results on the M_{200} posterior, in fact). The fine modelling discussed in Section 5.4 is unavoidable if we want to parameterize our ignorance about the other parameters and marginalize over their possible values. By doing so we can minimise the risk of biasing our results on M_{200}, and obtain realistic errors, i.e that take into account the uncertainties related to modeling and miscentring.

6.2 Systematic uncertainties

We can assert that the error on the parameter inference, as well as on the fitting model, comes directly from the error on the data, principally. Indeed, because of the stacking, shear observations at different radii are correlated and this effect is more significant at radii larger than the typical lens angular separation (Mandelbaum et al., 2013). The uncertainties deriving from the posterior distribution of M_{200} are the dominant error of our kind of measures but they are not the only ones. There are other error sources to consider when inferring the mass-observable relation from weak lensing data. In principle, by combining the signal from many lenses, the error on the mean shear profile and on the inferred median mass can be reduced to very low level. In that limit, systematic errors can start to overcome the statistical errors.

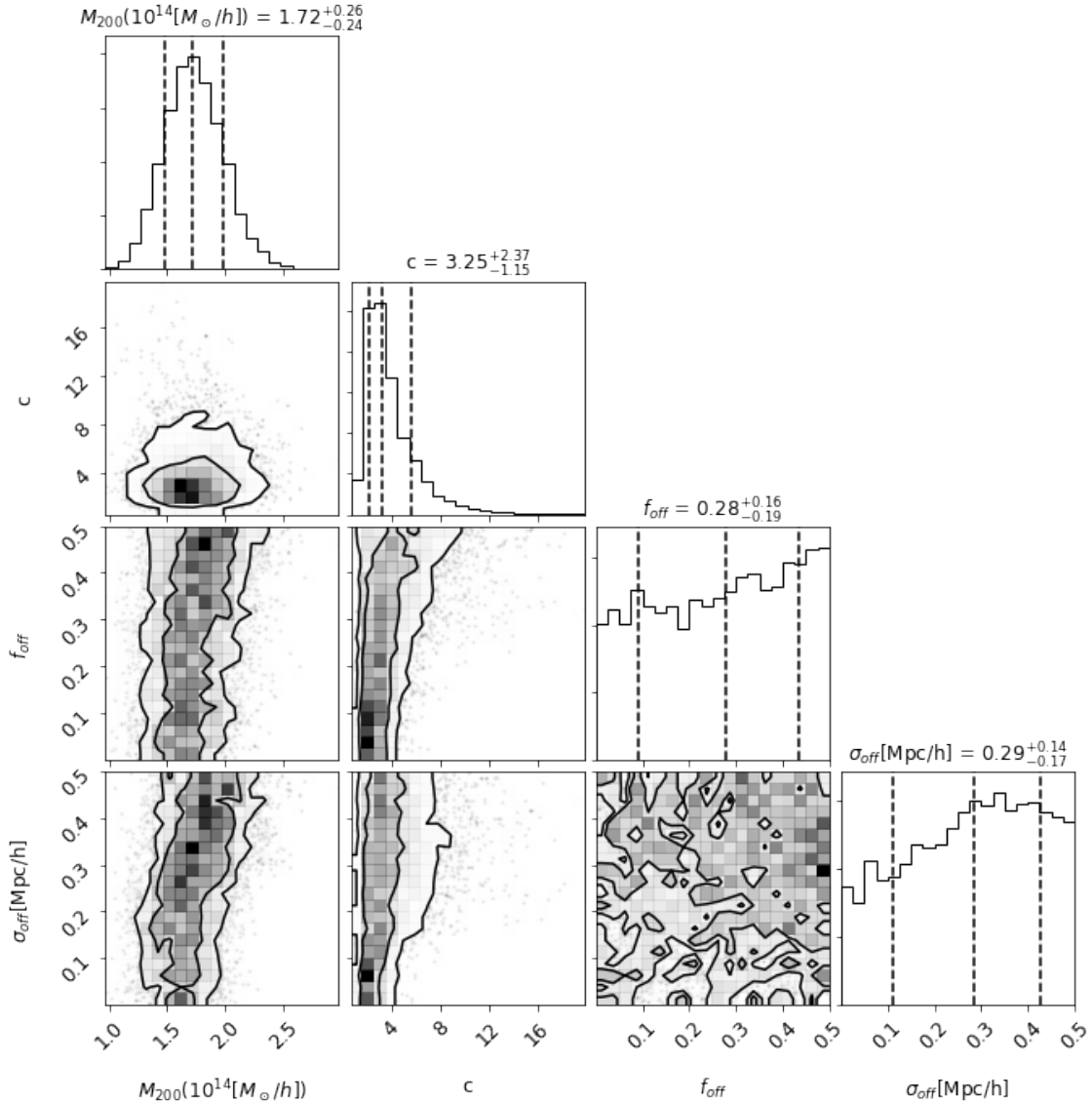

Figure 6.2: Posterior distributions of the four free parameters of the 1-halo lensing model (M_{200}, c, f_{off}, σ_{off}) for one of the 14 cluster samples of our analysis, in the smaller radial range, shown as an illustrative example of the reliability of the inferred constraints. Specifically, the contour plots shown refer to the fourth box of the first redshift bin of Figure 6.3, i.e the profiles built with $2.05 < A < 2.75$ in $0.1 < z < 0.3$. The dashed vertical lines represent the posterior medians, and the 18-th and 82-th percentiles. The values of these quantities are reported on top of each panel. The contour lines refer to $1-\sigma$ and $2-\sigma$.

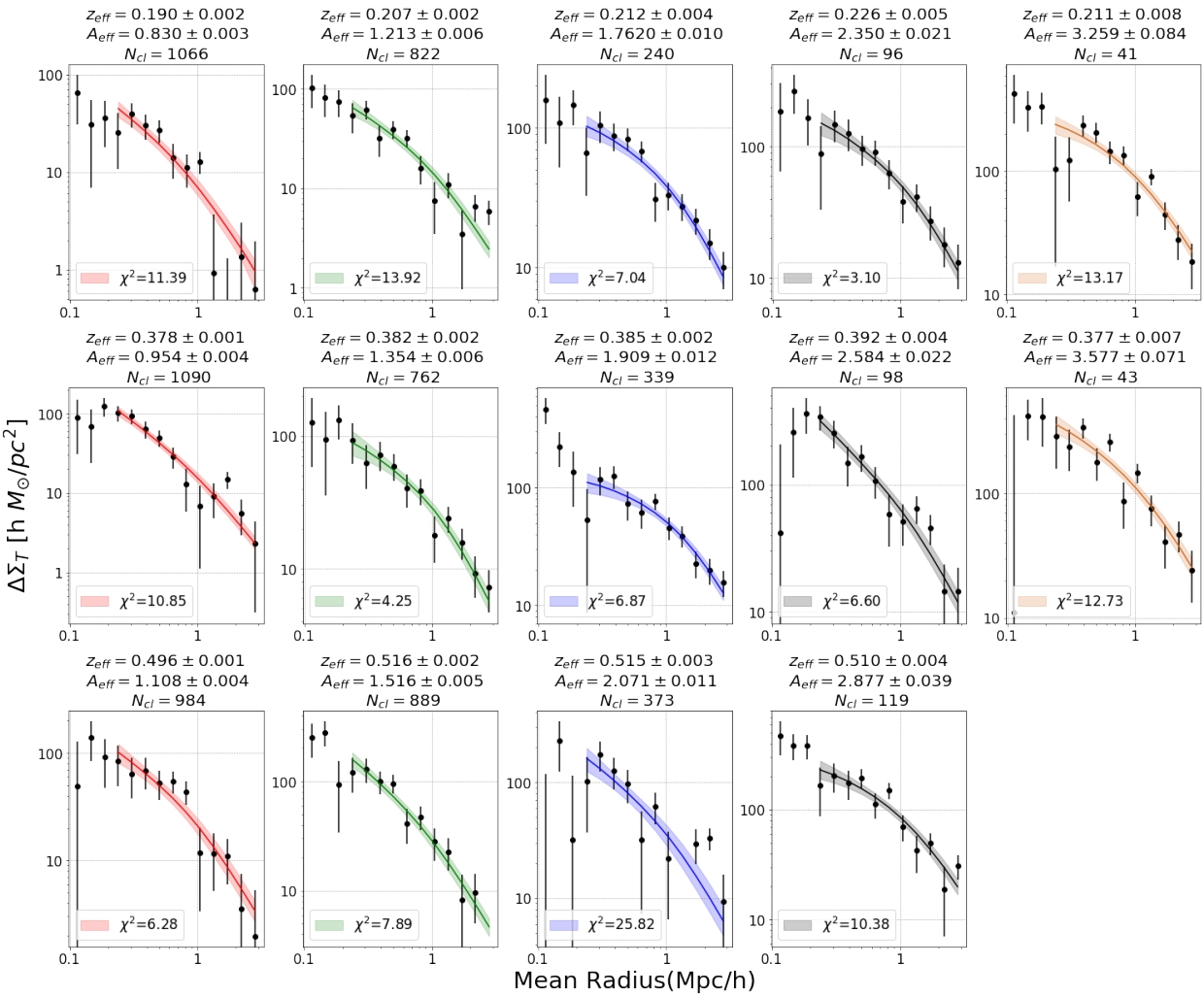

Figure 6.3: Stacked profiles for cluster samples in 14 bins of amplitude and redshift. The redshift increases from top to bottom, while the amplitude (hence M_{200}) from left to right. The error bars are the diagonal values of the bootstrap covariance matrix, estimated as described in Section 5.3.3. The effective redshifts $z_{\rm eff}$ and amplitudes $A_{\rm eff}$ (evaluated with Eq. (6.3.2)) and the numbers of lens $N_{\rm cl}$ are reported on top of each box. The coloured curves show the best-fit posterior 1-halo models derived from the MCMC statistical analysis, while the shaded regions indicate the estimated $1-\sigma$ uncertainties. The χ^2 values are also reported in the labels. The measurements at $r < 0.2{\rm Mpc}/h$ (the first three data points) have been excluded from the MCMC analysis (see the discussion in Section 5.4.1). Thus the number of degrees of freedom is 7 (11 data points - 4 free parameters). The median values of the M_{200} posterior distributions of each bin are reported in Table 6.1.

	$M_{200}(10^{14}M_{\odot}/h)$				
	A1	A2	A3	A4	A5
$0.1 \leq z < 0.3$	$0.16^{+0.03}_{-0.03}$	$0.35^{+0.05}_{-0.06}$	$1.17^{+0.14}_{-0.17}$	$1.72^{+0.22}_{-0.24}$	$3.69^{+0.40}_{-0.43}$
$0.3 \leq z < 0.45$	$0.41^{+0.05}_{-0.07}$	$0.74^{+0.10}_{-0.13}$	$1.60^{+0.20}_{-0.22}$	$2.19^{+0.38}_{-0.41}$	$4.44^{+0.57}_{-0.64}$
$0.45 \leq z < 0.6$	$0.53^{+0.09}_{-0.10}$	$0.78^{+0.11}_{-0.12}$	$1.01^{+0.20}_{-0.25}$	$3.03^{+0.36}_{-0.39}$	

Table 6.1: Table showing the posterior distribution values of the median of M_{200}, with relative 1σ-percentiles (18-th and 82-th percentiles of the posterior), of each of the 14 redshift-amplitude bins of Figure 6.3. The table structure is equal to the disposition of the panels in Figure 6.3. We call every bin in amplitude with An, with $n = 1, 2, 3, 4, 5$, to indicate the ordered A bin splitting we have set and that is listed at the beginning of Chapter 6. Note that every An varies from one redshift bin to another, and here it is only useful to enumerate the growing amplitude bins. We do not list the posterior values of the other three model parameters because they are degenerate and they are not of our concern.

Three aspects of the weak lensing analysis are particularly critical for the reliability of the mass estimation: the background galaxies selection, the photo-z estimates and the shear measurements. The uncertainties in the results of the just mentioned selection criteria directly impact on the derivation of $\Delta\Sigma$ from the data. Furthermore, consider that as the cross-correlation lensing method (i.e the stacking procedure) represents a statistical measurement of the lens-mass cross correlation function, the inferred median masses are not affected, on average, by *uncorrelated* matter to the source galaxies along the line of sight. Instead, on the cluster scales considered in this first analysis phase, the mean effects of the *correlated* mass along the line of sight, like in neighboring clusters or filaments, are negligible out to scales comparable with the cluster virial radius.

However, we can note that all the conservative performed data extraction, described in Sections 5.2 and 5.3, should let us predict that possible systematic uncertainties will not affect remarkably the mass-observable relation. Our selection of background galaxies implies that we should not expect significant contamination from foreground galaxies and that our selected galaxies have a robust photo-z estimate. Indeed, when estimating the shear profiles it is known that the biggest source of error is related to the source redshift estimate, that enters in $\Sigma_{\rm crit}$. Finally, the shear estimates of KiDS-DR3 are by their nature very reliable thanks to the accuracy of the survey (Hildebrandt et al., 2017).

Anyway we model their effects to take into account these possible sources of uncertainty. Thus, in order to understand how much they affect the mass estimation we should take into account the relation between the lens mass and its excess surface density profile. Obeying to the formalism in Melchior et al. (2017), one can define the logarithmic dependence of $\Delta\Sigma$ on M_{200} as:

$$\Gamma = \frac{\mathrm{d}\log\Delta\Sigma(M_{200})}{\mathrm{d}\log M_{200}}. \tag{6.2.1}$$

In Bellagamba et al. (2019) in the range of radii, redshifts and amplitude considered in the analysis, assuming a realistic concentration $c = 4$, they found that $\Gamma \sim 0.75$ is a good approximation. Thus, assuming this value, we can connect the uncertainties on the mass with those on $\Delta\Sigma$ as:

$$\frac{\delta M_{200}}{M_{200}} \sim \Gamma^{-1}\frac{\delta\Delta\Sigma}{\Delta\Sigma}. \tag{6.2.2}$$

A detailed study of the uncertainties deriving from background galaxies selection, photo-z estimates and shear measurements, as well as checks for residual systematics, can be found in Section 6 of Bellagamba et al. (2019). We report the main results of this systematic error analysis:

- the foreground contamination is of the order of $\sim$ 2% which, according to Eq. (6.2.2), becomes a 2.7% uncertainty on mass;
- the uncertainty related to the redshift distribution of background sources is estimated of 4.2% and it becomes of 5.6% on mass;
- the one relative to the shear measures instead is of the order of 1% (Hildebrandt et al., 2017), translated in 1.3% uncertainty on mass.

We take these results as also valid for our work because so far we follow the same methodology of Bellagamba et al. (2019), apart from having quoted the medians and not the averages. And, as expected, we have re-obtained similar results, as it can be confirmed comparing Table 6.1 and Fig. 6.3 with the table 2 and Fig.8 in Bellagamba et al. (2019). Even if our posterior values of M_{200} slightly differ, the difference is always well inside the errors, hence we could associate this minimum disagreement (our M_{200} posterior values are systematically slightly higher) to a different code used to perform the MCMC routines to extract the parameters. One can see Figure 6.4 to confirm as already said.

6.3 Mass-Amplitude relation

So far, we have not yet discussed the reliability of A estimates, evaluated by AMICO, for inferring mass estimates. Looking at Figure 6.3 and Table 6.1 we could already confirm the reliability of the method used. Moreover, the reduced-χ^2 values shown in Figure 6.3 confirm that the most likely model well describes the data in radial range 0.2<R<3.16 Mpc/h, for almost all the 14 redshift-amplitude bins. There is a clear correlation between the excess surface density (mass) and the amplitude, but we want to analyse more in detail the relationship between A and M_{200}.

The relation between the observable A and the weak lensing mass is modelled as:

$$\log \frac{M_{200}}{10^{14} M_{\odot}/h} = \alpha + \beta \log \frac{A}{A_{piv}} + \gamma \log \frac{E(z)}{E(z_{piv})}, \tag{6.3.1}$$

where $E(z) = H(z)/H_0$ takes into account the cosmology through the Hubble parameter evolution and A_{piv} and z_{piv} represent typical values of the observable amplitude and redshift of the total sample, respectively. The factor $\gamma \log E(z)$ instead considers a redshift evolution of the relation. We rely on this factor because the scaling of observable properties of galaxy clusters with mass evolves with time. Assessing the role of the evolution is crucial to study the formation and evolution of massive haloes and to avoid biases in the calibration (Sereno and Ettori, 2015). For each cluster bin, the typical value of an observable O_k is calculated by means of a *lensing-weighted average* (Umetsu et al., 2014):

$$O_k = \frac{\sum_{k \in K} W_k O_k}{\sum_{k \in K} W_k}, \tag{6.3.2}$$

where W_k represents the *total weight* for the k-th cluster of the bin, computed as:

$$W_k = \sum_i w_i \Sigma_{crit}^{-2}, \tag{6.3.3}$$

where i runs over all the background galaxies of the k-th cluster, regardless of the radial bin.

The covariance matrix for the mass estimates of the cluster bins is constructed as:

$$C_{M,ij} = \delta_{ij} E_i^2 + S^2 M_{200,i} M_{200,j}, \tag{6.3.4}$$

where E_i corresponds to the statistical error derived from the M_{200} posterior distribution and S represents the systematic uncertainty, estimated as described in Section 6.2. Furthermore, we underline that the systematic errors are considered as correlated here, because they affect almost homogeneously all cluster bins. Practically, we sum in quadrature the uncertainties for background selection, photo-redshifts and shear measurements and in this way we get $S = 0.064$ (Bellagamba et al., 2019). Anyway, this systematic error is much lower than the one derived from the M_{200} posterior for all the bins, so it can be assumed to be sub-dominant. The parameters α, β, γ of Eq. (6.3.1) can be computed with a Bayesian analysis of the same type as done in Section 6.1. Consequently, the likelihood is once more given by Eq. (6.1.2), where χ^2 becomes now:

$$\chi^2 = \sum_{i=1}^{N_{bin}} \sum_{j=1}^{N_{bin}} [M_{obs,i} - M_{mod,i}] C_{M,ij}^{-1} [M_{obs,j} - M_{mod,j}]. \tag{6.3.5}$$

In this equation, the sums run over the $N_{\rm bin}$ cluster bins, M_{obs} refers to the computed M_{200} while M_{mod} is the value of M_{200} derived from the scaling relation (6.3.1), and $C_{M,ij}$ is that derived in Eq. (6.3.4).

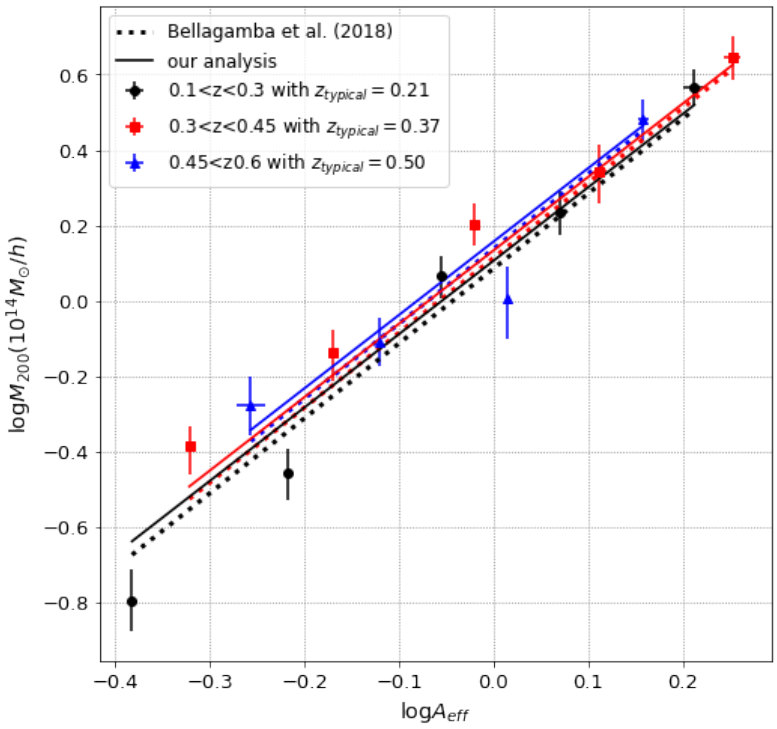

Figure 6.4: The mass-amplitude ($\log M_{200} - \log A_{eff}$) relation of the analysed 14 cluster bins, evaluated at three representative redshifts ($z = \{0.20, 0.37, 0.50\}$). The solid lines show the best-fit linear relations for each redshift. The error bars show the estimated $1-\sigma$ uncertainties on M_{200} (computed as the difference between the posterior 18-th and 82-th percentiles), plus the systematic uncertainty S. The mass-amplitude relations estimated by Bellagamba et al. (2019) (dotted lines) are reported for comparison.

Now, we can estimate the linear regression in Eq. (6.3.1) assuming the A_{eff} value of each cluster bin listed upon each box of Fig. 6.3. We use the median of the masses computed for each amplitude-redshift cluster bin (table 6.1) to constrain the $\log M_{200}$-$\log A$ relation, setting z_{piv}=0.35 and A_{piv}=2, because these are central values in the ranges covered by the whole sample. In Figure 6.4 we plot the 14 mass-amplitude data points, evaluated separately for the three typical redshifts of each z-bin ($z = 0.20$, 0.37, 0.50, computed through Eq. (6.3.2)), and their maximum-likelihood estimation curve, namely a simple linear regression optimized (i.e with optimal values for the parameters so that the sum of the squared residuals of Eq. (6.3.1) is minimised), to have an idea of the degree of the correlation.

From this linear regression we computed the common values of the parameters α (intercept), β (slope) and γ (redshift evolution) of the mass-amplitude relation. Precisely we get:

- α=0.13±0.04,
- β=1.95±0.11,
- γ=0.71±0.66.

The mass-amplitude relation is linear for the amplitude and redshift range covered by the sample and, moreover, this relation fits the data well for a mass range of more than one order of magnitude in all our redshift range. Since less massive objects are more abundant in the Universe, the cross-correlation lensing can be used over a huge range of lens masses. Therefore we can say that clusters in different amplitude regimes do not follow different relations, as there are no evidence of contamination by fake detections (even in the low-A samples where such contamination would be more influential). We can corroborate the close scaling relation between the amplitude of galaxy clusters and their mass.

6.4 Inference of cosmological parameters from the (1+2)-halo terms

In the second phase of our analysis, we re-extracted the stacked shear signal produced by the same clusters of Section 6.1, i.e with the same amplitude-redshift binning (listed into the introduction of this Chapter 6), but in this case the extraction was pushed up to very large radii. Precisely, the extraction was performed in the radial range $0.1 \leq R \leq 35$ Mpc/h on 30 radial annuli equally distanced in logarithmic space. At the mean redshift of the sample ($z \approx 0.35$) we make note that 35Mpc/h corresponds to about 1.9 deg on the sky. The total number of clustered objects is always 6972 and the same is true for the number of clusters stacked in each individual bin, as well as the total sample of stacked clusters that remains split in 14 amplitude bins. In each of these cluster bins is clearly evident the 2-halo term contribute (see Fig. 6.8), confirming the necessity of considering it when studying the shear signal on large scale.

The Bayesian approach used in the analysis of cluster shear profiles at large radii follows the same prescription used in the analysis described in Section 6.1 but, now, we aim at inferring posterior distributions of the cosmological value Ω_M. To fit the measured differential surface density profiles with our model $\Delta\Sigma_{1h+2h}(R)$ (Eq. (5.4.8)) we use again a Monte Carlo Markov Chain routine exploiting the Metropolis-Hasting algorithm. MCMCs are very useful for efficiently computing likelihoods in multi-dimensional parameter spaces but they work well even for inferring a single parameter, as in the case of this analysis where only Ω_M is free. When analysing the shear signal on large scales, we have set the length of the MCMC sampling with 10 chains of 5000 steps each running in parallel, and without assuming a burn-in period, thus every posterior distribution is constituted by 50000 points. Even here the convergence was reached with less chain steps but we choose to be conservative. Furthermore, here we did not assume a burn-in because every chain was initialized to the parameter values of maximized likelihood. We graphically check the convergence every time as shown in the example of Figure 6.5.

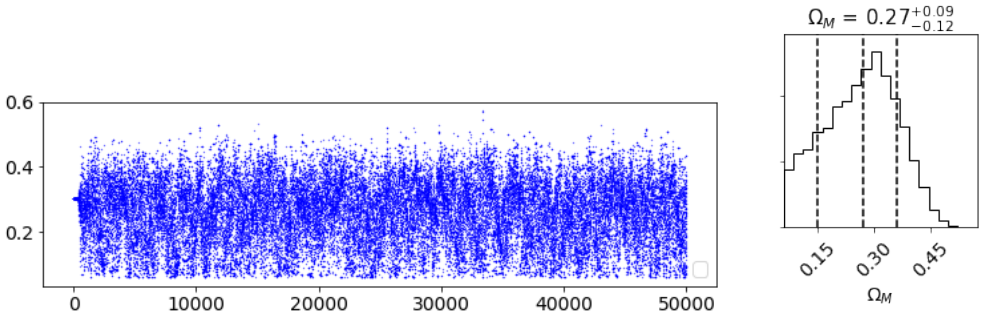

Figure 6.5: A typical MCMC chain output: on the x-axis are enumerated the consecutive chain steps, on the y-axes the Ω_M parameter values at each step. Specifically, the chain refers to the cluster sample with $0.1 \leq z < 0.3$ and $2.05 \leq A < 2.75$. We can appreciate the convergence of the chain. The posterior distribution is reported on the right panel, where the dashed vertical lines indicate the median value (central line) and the $1-\sigma$ percentiles.

When setting the prior on the MCMC routine we choose for Ω_M a uniform prior ranging between 0.06 and 0.6. The minimum prior aims at not excluding the almost total absence of dark matter (into the contribute to the total matter in Universe, the percentage of baryons is always present with a $\sim 5\%$ contribute). Instead, the maximum prior is far beyond typical values measured until now for Ω_M. Hence this prior setting is abundantly conservative. The remaining parameters were fixed to the posterior median values computed in the first phase of the analysis on the shorter radial range. In this way, we make sure that we are considering robust values of concentration and miscentring and, in particular, the cluster mass, to well fix a correct halo bias b_h value in the 2-halo term model of Eq. (5.4.6) that we are using to infer the cosmological parameter. The lens redshift was fixed to the one estimated by AMICO when extracting the shear profiles for the 1-halo region. In fact, the lens redshift associated to this extended extraction is more uncertain and overestimated because, into such large regions, in Eq. (6.3.2) the total weight of the various clusters changes, and this depends on the weights of the background galaxies.

Furthermore the signal coming from the cluster outskirts is very weak, sometimes it shows neg-

ative values too, because here the S/N is lower. Nevertheless, in each cluster bin, the contribution of the 2-halo term is clearly visible for radii larger than 10 Mpc/h (see Fig. 6.8). The truncation factor remains fixed to $3r_s$ and the cosmology to a ΛCDM flat model, therefore every posterior value of the matter parameter density is such that $\Omega_\Lambda = 1 - \Omega_M$. For the same reasons assumed when modelling the first extraction on the 1-halo range, we avoid to consider shear radial points at distances lower than 0.2 Mpc/h, i.e we exclude the first four radial points.

In Figures 6.8 and 6.7 the individual results of this analysis are shown. In Fig. 6.8 we can appreciate a good modelling coming from the posterior distribution of our Bayesian analysis as well as a small error on the model, considering the very large error on shear data in correlated matter region. Furthermore, from Fig. 6.7 we can note the purely statistical behavior of the 14 samples: there is no clear pattern in the posterior distributions as a function of redshift or amplitude, except a slight one in the first redshift bin. This is a confirmation of the "indipendence" of the data coming from the different cluster bins, and it allow us to combine all the 14 posterior probabilities as discussed in appendix A. In practice, to be able to obtain a single posterior distribution of Ω_M from all the cluster shear profiles, we simply multiply the 14 individual posterior distributions of Fig. 6.7. The result can be seen in Fig. 6.9: we are in excellent agreement with WMAP9 (black lines) but there is a tension with Planck18 (red lines). Furthermore the posterior product distribution tends to a Gaussian distribution, confirming the statistical behaviour of our measures. In Fig. 6.10 we show the posterior product evaluated for each redshift bin, where we see that there is not an appreciable trend with z and where, moreover, we can see that the lower redshift bin is the dominant one in determining the final product distribution.

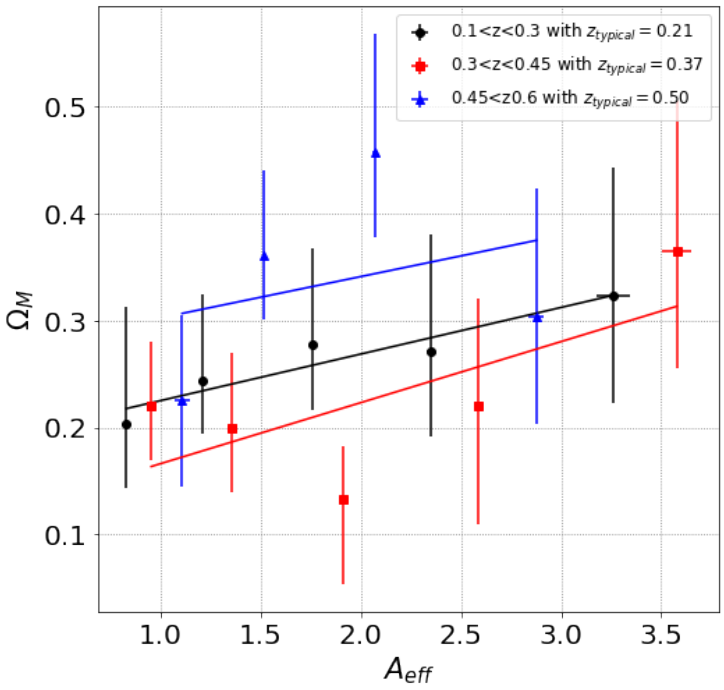

Figure 6.6: The inferred median Ω_M-Amplitude relations estimated from the modelling of the lensing profiles in three redshift ranges, as indicated by the labels. The amplitude, A_{eff}, has been estimated with Eq. (6.3.2). The solid lines are the best-fit linear relations in each redshift bin.

In Fig. 6.6 we plot and fit the relation between the median Ω_M and the amplitude for each redshift bin. The points are described by the following linear best-fit models:

- Ω_M=0.18A_{eff}+0.04, for 0.1<z<0.3 (black line)
- Ω_M=0.11A_{eff}+0.06, for 0.3<z<0.45 (red line)
- Ω_M=0.26A_{eff}+0.04, for 0.45<z<0.6 (blue line).

From the figure we can confirm that there is not a correlation with redshift for Ω_M but it seems to be present with amplitude. Anyway, we can not confirm this statement because of the large error

on the matter density parameter values, that produces a very large error on the inferred slopes. Therefore, the correlation is not significant.

Indeed, this analysis with free Ω_M aimed at being a test on the reliability of the method, i.e to see if we can infer reasonable cosmological parameter values from the 2-halo signal, but the strength of the method relies on a modelling that infer both M_{200} and Ω_M (see Sect. 6.5).

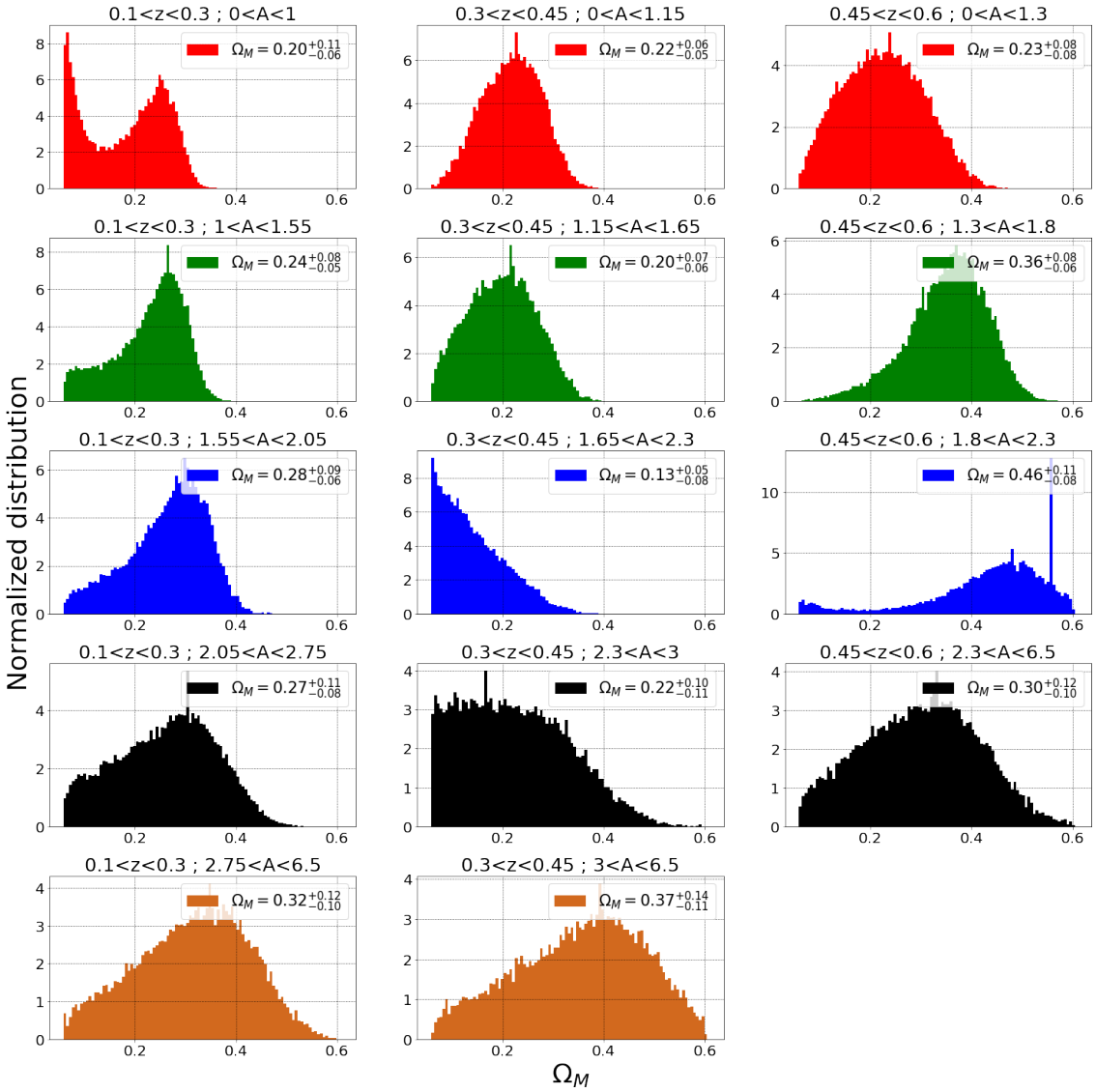

Figure 6.7: The MCMC marginalised posterior distributions of Ω_M, in 14 bins of amplitude and redshift, as indicated on top of each box (as in Fig. 6.8). The median values, with 1-σ uncertainties, are reported in the labels.

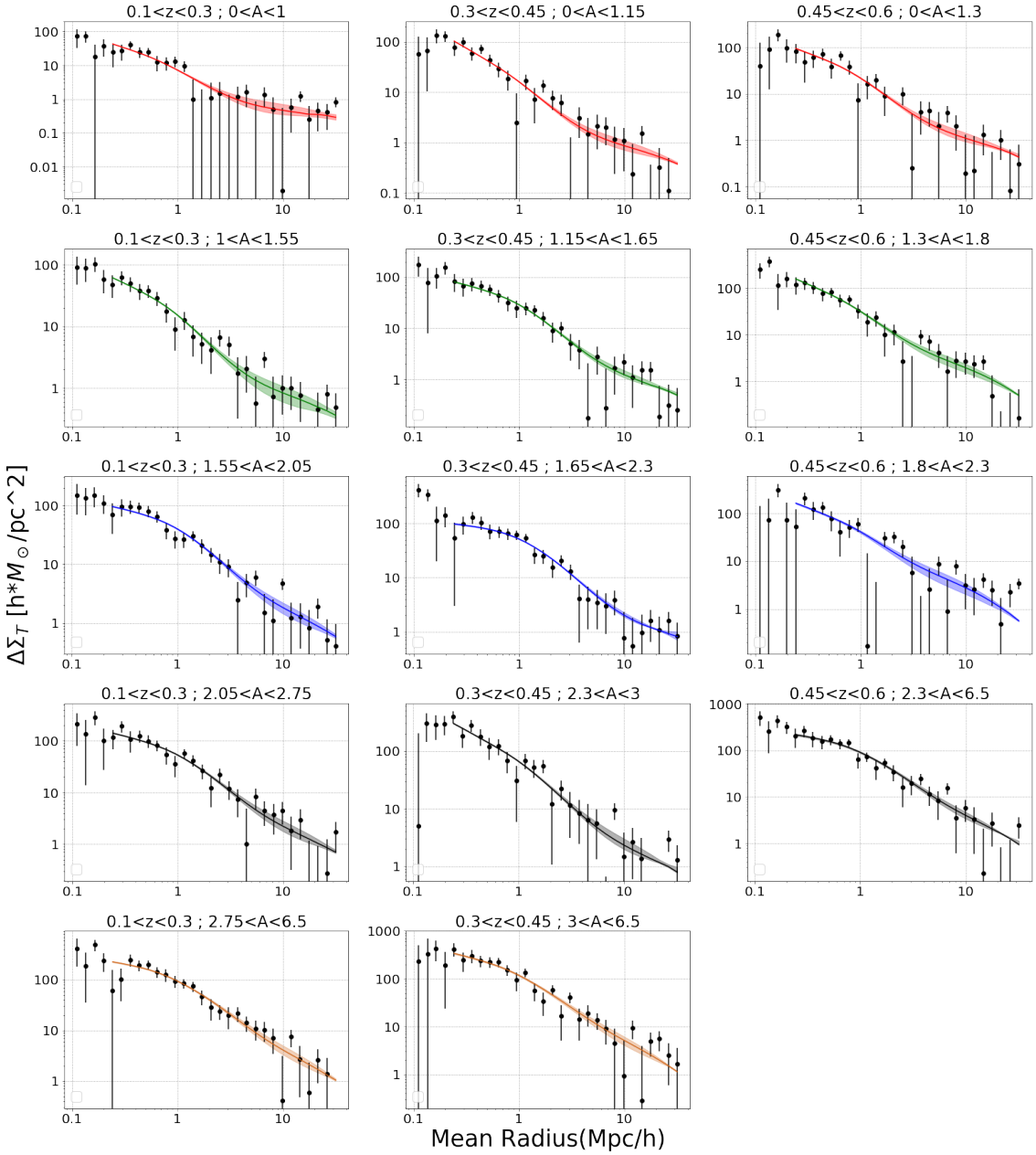

Figure 6.8: Stacked profiles for cluster samples in 14 bins of amplitude and redshift. The redshift increases from left to right, while the amplitude (hence M_{200}) from top to bottom. The error bars are the diagonal values of the bootstrap covariance matrix, estimated as described in Section 5.3.3. The coloured curves show the best-fit posterior (one+two)-halo models assessed with the MCMC statistical analysis, while the shaded regions indicate the estimated $1-\sigma$ uncertainties. The measurements at $r < 0.2\,\mathrm{Mpc}/h$ (the first four data points) have been excluded from the MCMC analysis.

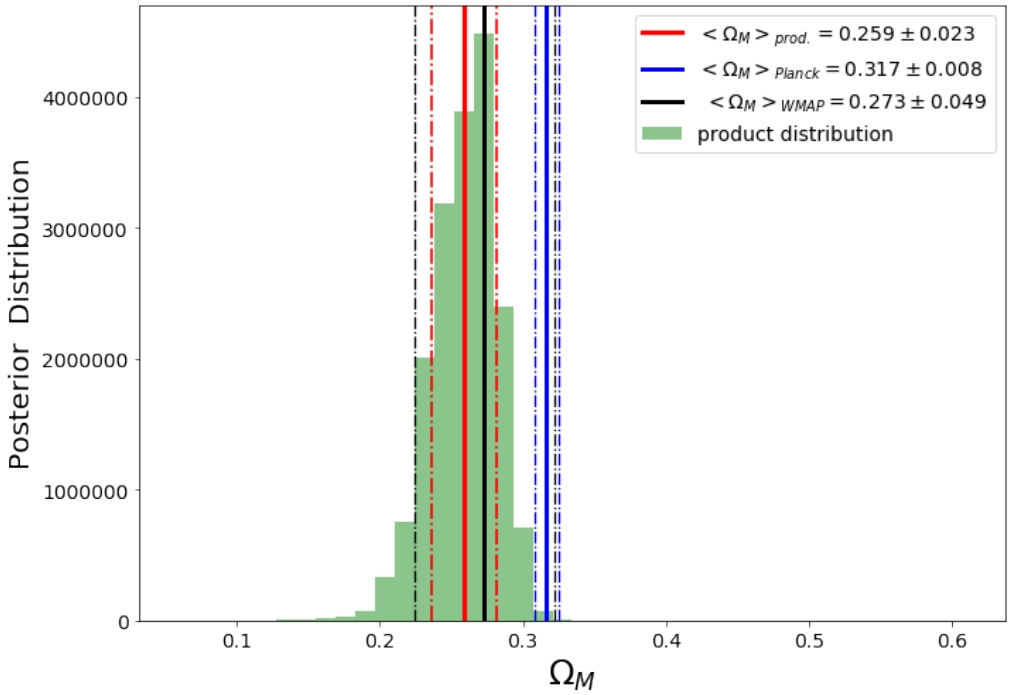

Figure 6.9: Combined Ω_M posterior distribution, estimated as the product of the 14 posterior distributions of each cluster sample shown in Fig. 6.7 (green histogram). The Planck18 (solid blue lines) and WMAP9 (solid black lines) constraints are shown for comparison, with the dot-dashed lines showing the respective errors (1σ), as indicated by the labels.

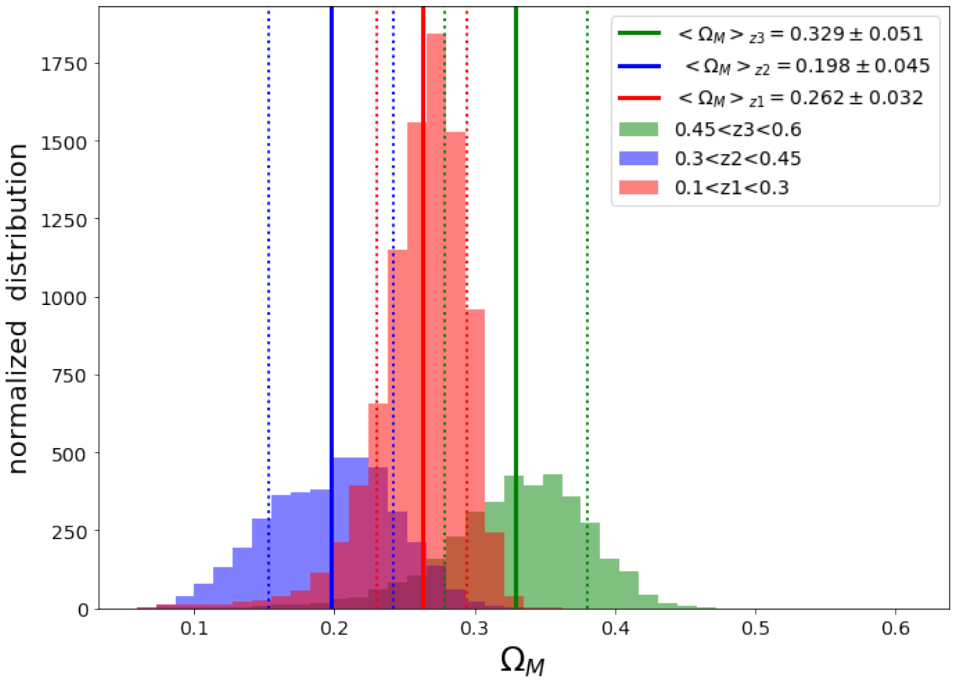

Figure 6.10: Combined Ω_M posterior distribution, estimated as the product of the posterior distributions in the three redshift range separately, that is $0.1 < z < 0.3$ (red), $0.3 < z < 0.45$ (blue) and $0.45 < z < 0.6$ (green). The mean and variance are indicated by the solid and dotted lines, respectively. The mean and variance of the product distribution of each redshift bin are shown in the legend.

6.5 Inferring Ω_M and M_{200} together

The spatial and dynamical properties of dark matter halos are a robust function of the cosmological parameters but to exploit clusters as cosmological probes we have to know the relation between their observable properties and their masses. From this, we understand the importance to have a close correlation between mass and amplitude (the amplitude can be seen as a parameterization of that observable that is the cluster richness), found in Section 6.3, to confirm the reliability of the cosmological parameter estimates from clusters. This is true especially in our situation, since our shear measurements come from weak lensing observations, that can map mass profiles up to large radii, and are therefore not influenced by the dynamics and the state of the baryons.

In this section we present the final task of our thesis work: a method to infer both cosmological parameters and cluster masses from a stacked shear data analysis, modelling together the 1-halo contribute and the 2-halo one. Here we simply introduce our analysis method with its theoretical bases and data modelling procedure and we also discuss the reliability of the measure of shear profiles from photometric signals on large scales in the KiDS-DR3 survey.

The shear data analysis was developed exactly as the one discussed in Section 6.4, both for the Bayesian analysis performed and for the used shear data. The fundamental difference now is that we are fitting the shear profiles, shown in Figure 6.8, leaving free mass and concentration together with Ω_M. M_{200} and c are parameters estimated exclusively from the 1-halo model. The priors range is the same assumed for Ω_M in Section 6.4, and for M_{200} and c in Section 6.1. In this way we allow to the MCMC runs to be free to model the total shear profiles extracted up to the extreme outskirts of the AMICO clusters.

We remind that M_{200} is computed as:

$$M_{200} \equiv M(r_{200}) = 200\rho_c(z)\frac{4}{3}\pi r_{200}^3, \tag{6.5.1}$$

where $\rho_c(z) = 3H^2(z)/(8\pi G)$ is the critical density at epoch z and the Hubble parameter satisfies $H^2(z) = H_0^2[\Omega_M(1+z)^3 + (1-\Omega_M)]$ for a flat ΛCDM Universe. Hence, when inferring the Ω_M value on the extended radial range, we expect some correlation between M_{200} and Ω_M. Ω_M will affect the estimation of halo mass, being this composed principally by dark matter.

The values obtained for M_{200} fitting the profiles up to large radii around the cluster itself are very similar to those inferred when modelling the 1-halo cluster profile only (Section 6.1), confirming both the reliability of the truncated NFW model and the negligibility of the 2-halo term contribute to the 1-halo profile. It is important to obtain a reliable estimation of the one halo mass because it determines the right value of the halo bias b_h, that grows exponentially with mass, and that is the term that more than all configures the surface density profile of the 2-halo term of Eq. (5.4.6). We do not care about the concentration because it is a degenerate parameter, but we underline that was necessary to leave it free in this analysis to make sure it did not bias the mass estimation. The miscentring parameters f_{off} and σ_{off} were fixed to the values inferred in the first step of the analysis on the 1-halo radial range because their contribute decreases rapidly beyond 0.5-0.6 Mpc/h and, consequently, they can not affect the 2-halo term. However, we could not ignore their contribute to the mass estimate.

In Figures 6.11, 6.12 and 6.13 the individual fitting models derived from the analysis through MCMC runs are shown for the first, the second and the third redshift bin, respectively. With the same splitting, the posterior distributions for model parameters are shown in Figures 6.14, 6.15 and 6.16. We can confirm the reliability of the used method by looking at the figures just mentioned. First of all, from Figures 6.11, 6.12 and 6.13, we can see the necessity of the 2-halo term to describe the stacked surface density profile on large scales: between 7 and 10 Mpc/h the 2-halo term starts to dominate over the 1-halo term, which is unable to fit the data far beyond the cluster virial radius. We separately put in these plots the 1 and 2-halo profile models (dot-dashed red curves) to show the scales where they fit the data and to check the good conjunction they reach on radii where they cross. We underline that they match with each other thanks to the flexibility introduced by the truncation in the NFW profile. Always from the profiles we can see how our posterior models are often in good agreement with the Ω_M values estimated by Planck18 (solid blue curves) and WMAP9 (black solid curves), representing the state of the art regarding the estimation of cosmological parameters. Furthermore, the error on the fitting models of Eq. 5.4.8 is much smaller with respect to the prior range of the free parameters: we plot the model for the minimum Ω_M-prior (violet dashed curve) and for the maximum one (dashed green) to show that our Bayesian analysis was able to narrow down the range of possible values that Ω_M

can take. Moreover, from the curves for the extreme values, we can see how the matter density parameter affects the shape of the 2-halo profile, since this value determines directly the power spectrum shape and the normalisation of the bias halo, in the $\Delta\Sigma_{2h}(R)$ profile (see Eq. 5.4.6). An interesting thing that emerges from the violet curves is that, when considering the minimum Ω_M prior, that assumes negligible dark matter contribute, the fitting model always shows an inflection point beyond which the model profile grows quickly, as if at those scales the related matter started to be more concentrated. If there is any physical reason behind this, we intend to further investigate after this thesis work.

In Fig. 6.17 we plot and fit the relation between the inferred median Ω_M and the M_{200}, for each redshift bin. The points are described by the following linear best-fit models:

- Ω_M=0.27M_{200}+0.03, for 0.1<z<0.3 (black line)
- Ω_M=0.17M_{200}+0.05, for 0.3<z<0.45 (red line)
- Ω_M=0.36M_{200}+0.04, for 0.45<z<0.6 (lue line)

We can see that there is some form of correlation between M_{200} and Ω_M. This would be somewhat worrying because the estimation of the cosmological parameter would depend on the derived values of the cluster mass. Anyway the errors on the different Ω_M values are so large that we can reach no definite conclusion about this possible dependence. In fact, the error on the slope of these relations is of the order of the fit parameters themselves. From the figure we can suppose there is not a correlation with redshift despite the magnitude of the errors does not allow a clear confirmation.

Finally, as done in Section 6.4, we show the product of the 14 individual cluster bin distributions to get our final posterior distribution of Ω_M. We did so considering the single Ω_M distributions of the three-dimensional posterior distributions of Figures 6.14, 6.15 and 6.16. In this case we obtain $< \Omega_M >= 0.285 \pm 0.023$ that is consistent and competitive with the one estimated by WMAP9 (95% confidence level). Instead, it is consistent at the limit of the agreement with Planck18 (68% level confidence), as shown in Figure 6.18. Our measures are in tension with the estimations inferred from the CMB by Planck18. Furthermore, the error on the posteriors is less than a half of that from WMAP9 and, unexpectedly, it has not grown compared to that resulting from the "test" discussed in Section 6.4 (we suggest to compare Figures 6.18 and 6.9). Indeed, the error deriving from the analysis with free M_{200}, c and Ω_M has grown as expected but on mass and concentration estimates, especially on the latter. Moreover, the estimates of the median mass have grown slightly and systematically with respect to those estimated from the extraction of the 1-halo profile only (Section 6.1). However, we believe that the latter are more reliable, since they suffer from negligible systematic uncertainties and are estimated mainly with the truncated NFW, which in literature has proved to be very reliable on the 1-halo radial ranges.

We find that the modelling approach through the sum of models of equations (5.4.1) and (5.4.6) was successful to compute both the lensing cluster mass and the matter density parameter. It is not trivial that the cosmological information contained into the 2-halo term (described by an analytical function taking into account the matter power spectrum and the halo bias) can help us to constrain the cluster halo mass together with a cosmological parameter, which enters in the 1-halo modelling simply to set the cosmological model adopted by the truncated NFW fitting function.

6.6 Possible systematics affecting the Ω_M estimate

There are still several tests to perform on this new method and on these data. For instance, we have to check the systematic uncertainties and how much they affect the measures. The shot noise is not the only source of uncertainties at large radii: if we consider clusters whose mean separation is smaller than the range over which we measure the shear profiles, we cannot simply rescale the lensing S/N as $S/N \propto \sqrt{N_{cl}}$ (Sereno et al., 2017). The uncertainty derived from cosmological parameters is reasonably underestimated in our analysis. This is because we have ignored the correlation between data at large scales and because we have fixed all the other cosmological parameters to those of Planck18, for example.

Now, we will discuss some advancements that might improve the results (see Figures 6.11, 6.12, 6.13 and 6.18) of our kind of lensing analysis.

A first source of uncertainty is based on the fact that the inference of cosmological densities requires that the measured signal is independent of the reference cosmological conditions. Instead,

when stacking in physical lengths and estimating the shear signal we needed to assume a cosmological model. We had to compute angular diameter distances to convert angular separations to physical lengths, to convert tangential shear to differential surface density contrasts and to weigh the lensing signal for sources at different redshifts. Also the masses were inferred assuming a cosmological model in this thesis work. A solution could be to assume a cosmology to pass from the ra, dec, z coordinates (observed) to the cartesians (co-moving) ones and model the possible geometrical distortion when we estimate the likelihood for the fitting model. Anyway, in literature these effects are considered small and negligible hence we did not measure and consider this bias in our work. In other word, we considered negligible the possible geometrical distortions induced by the assumption of a wrong cosmological model. However, stacked shear signal drawn in a fixed cosmology has been previously used to constrain the Ω_M parameter (Cacciato et al., 2013; Mandelbaum et al., 2013). A useful solution to get results without an unbiased procedure like this could be to stack the signal in angular radii for lenses in small redshift bins, adopting a unique population of background sources for the lensing analysis (Oguri and Takada, 2011). This approach will be available for wider and deeper surveys, like *Euclid*.

In this thesis work we did not carry out a detailed analysis of the possible systematic uncertainties arising when inferring Ω_M from shear profiles measured up to very large radii. In Section 6.2 we quantified the influence of systematic uncertainties on the mass estimate when fitting shear profiles on 1-halo radial scales. We concluded that our mass inference is robust and the main source of error that affects the mass estimates is principally of a statistical nature. This is due to our conservative approach in the background galaxies selection (see Section 5.2).

The 2-halo term signal is proportional to the halo bias times the normalisation of the amplitude of the matter power spectrum, $\propto b_h \sigma_8$. The halo bias is determined principally by the 1-halo mass. Hence we can suppose that the cosmological densities, estimated from the 2-halo term modelling, are reliable in our case. Moreover, we estimated the uncertainty covariance matrix with a bootstrap procedure with resampling (precisely, 10000 resampling of the clusters on mock catalogues), neglecting the non-diagonal terms that are at the level of a noise. While this method is valid for the masses inferred from shear profiles collected in the 1-halo radial range, it can be improved for what concerns the 2-halo scales, from which we infer the matter density parameter. Because of the stacking, shear observations at different radii are correlated and the effect increases at radii larger than the typical lens angular separation (Mandelbaum et al., 2013). Hence, the posterior error on Ω_M might be slightly underestimated, as we did not consider all possible systematic uncertainties that might affect our analysis. We defer this analysis to future work.

Anyway, a brief overview of how the error on Ω_M could affect these types of measures can already be thought. A first solution to reduce systematic uncertainties can be to restrict the number of radial bins when deriving the shear profiles. If the number of lenses to stack in each redshift-amplitude bin is considerably larger than the number of radial bins we can neglect a possible bias induced by the noisy estimator of the inverse covariance matrix (Hartlap et al., 2007). We know that having less radial bins would reduce the error on the shear profiles but it would enhance that on the modelling. Until we will have more numerous cluster samples, this road can not be undertaken. An alternative approach would require to create a noiseless covariance matrix (Mandelbaum et al., 2013). The shape noise is the first source of error at small radii but it scales as R^{-2} for logarithmically spaces bins. However, the total noise flattens at larger radii, mainly for two reasons. Primarily, when R is appreciably larger than the typical separation between lenses, the annular bins enclose many of the same sources lensed by neighboring clusters. So the shape noise is not able to decrease by stacking more lenses. Secondly, when R is of the same order of the field of view, an incomplete correction of the optical distortion can induce a coherent anisotropy of the point spread function ellipticities in the edge of the field of view, causing a residual systematic error. Like in our case, where we collect data on scales that are about twice the camera coverage. Considering these two systematic uncertainties sources, the diagonal terms of the uncertainty covariance matrix C can be modelled (Mandelbaum et al., 2013) as:

$$C(R_i, R_i) = AR^{-2}[1 - (R/R_t)^2],$$

where R_t is the turn-around radius above which the shape noise is subdominant (roughly $9 - 10$ Mpc/h) and A is the amplitude of the noise variance. The smoothing of the non-diagonal terms is difficult to model and probably the usage of a noisy version of the covariance matrix is a safer choice. Furthermore, the uncertainty covariance matrix can be established only when the number of clusters to stack is large enough. This is not the case for the less populated bins in amplitude,

i.e for more massive clusters.

A limited field of view in weak lensing observations would not allow to constrain cosmological parameters that are not $\propto b_h\sigma_8$ (Sereno et al., 2017). We strongly believe that the reasonable statistical uncertainties, obtained from KiDS data with our analysis on the Ω_M inference, counterbalanced the uncertainties due to the limited OmegaCAM field of view with respect to our observation scales. For these reasons, the increased spectroscopic sample associated with the *Euclid* survey will greatly improve the cosmological estimates of our kind of analysis.

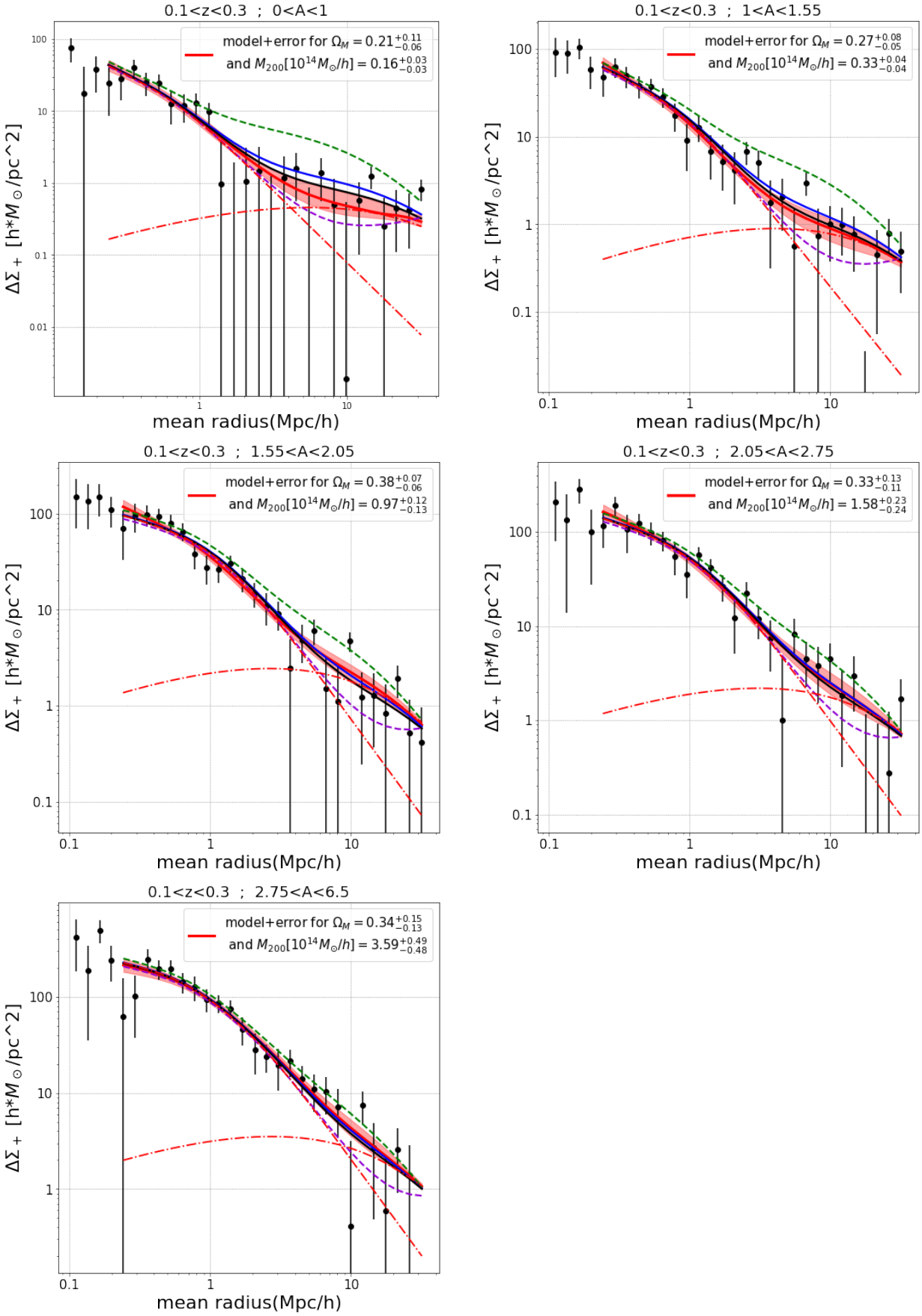
0.1<z<0.3 ; 0<A<1
model+error for $\Omega_M = 0.21^{+0.11}_{-0.06}$ and $M_{200}[10^{14}M_\odot/h] = 0.16^{+0.03}_{-0.03}$
$\Delta\Sigma_+$ [h*$M_\odot$/pc^2]
mean radius(Mpc/h)
0.1<z<0.3 ; 1<A<1.55
model+error for $\Omega_M = 0.27^{+0.08}_{-0.05}$ and $M_{200}[10^{14}M_\odot/h] = 0.33^{+0.04}_{-0.04}$
$\Delta\Sigma_+$ [h*$M_\odot$/pc^2]
mean radius(Mpc/h)
0.1<z<0.3 ; 1.55<A<2.05
model+error for $\Omega_M = 0.38^{+0.07}_{-0.06}$ and $M_{200}[10^{14}M_\odot/h] = 0.97^{+0.12}_{-0.13}$
$\Delta\Sigma_+$ [h*$M_\odot$/pc^2]
mean radius(Mpc/h)
0.1<z<0.3 ; 2.05<A<2.75
model+error for $\Omega_M = 0.33^{+0.13}_{-0.11}$ and $M_{200}[10^{14}M_\odot/h] = 1.58^{+0.23}_{-0.24}$
$\Delta\Sigma_+$ [h*$M_\odot$/pc^2]
mean radius(Mpc/h)
0.1<z<0.3 ; 2.75<A<6.5
model+error for $\Omega_M = 0.34^{+0.15}_{-0.13}$ and $M_{200}[10^{14}M_\odot/h] = 3.59^{+0.49}_{-0.48}$
$\Delta\Sigma_+$ [h*$M_\odot$/pc^2]
mean radius(Mpc/h)

Figure 6.11

Figure 6.11: The stacked profiles of galaxy clusters in all bins of amplitudes and redshifts, for the first redshift bin ($0.1 < z < 0.3$). The amplitude increases from left to right and from top to bottom. The error bars show the diagonal values of the bootstrap covariance matrix (continue to the next page). The measurements at $r < 0.2\,\mathrm{Mpc}/h$ (the first four data points) have been excluded from the MCMC analysis (see the discussion in Section 5.4.1). The red solid curves are the best-fit posterior models, while the shaded red regions show the estimated $1-\sigma$ uncertainties. The MCMC is performed with three free parameters, M_{200}, c and Ω_M. The blue and black curves show the models with $\Omega_{M,Planck} = 0.32$ and $\Omega_{M,WMAP} = 0.27$, respectively. The dashed violet and green curves show the models with $\Omega_M = 0.06$ and $\Omega_M = 0.6$), corresponding to the minimum and maximum limits of the assumed Ω_M prior, respectively. The dot-dashed red curves are the models for 1-halo and 2-halo, separately.

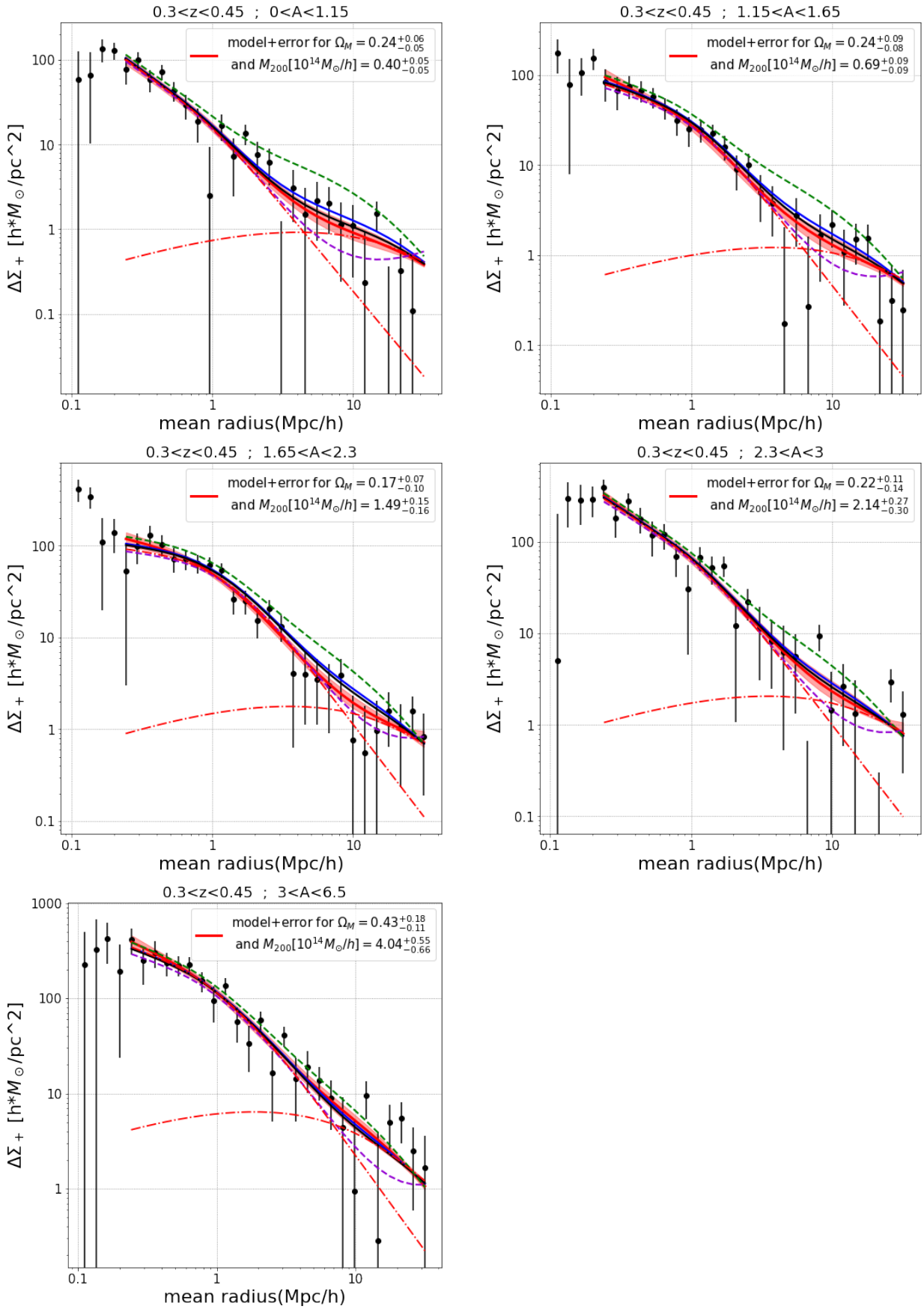

Figure 6.12: As Fig. 6.11, but for the second redshift bin analysed ($0.3 < z < 0.45$).

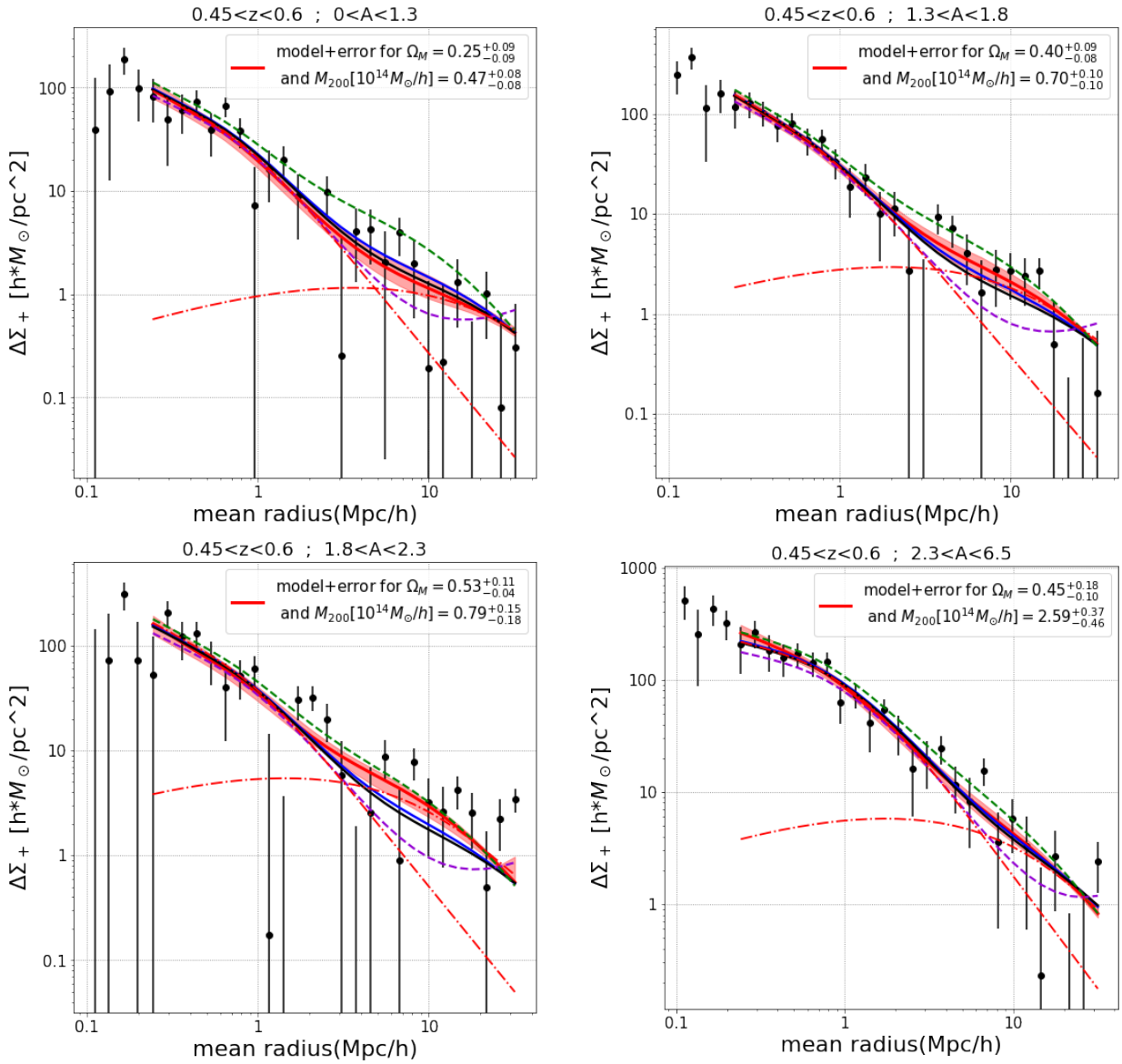

Figure 6.13: As Fig. 6.11, but for the third redshift bin analysed ($0.45 < z < 0.6$).

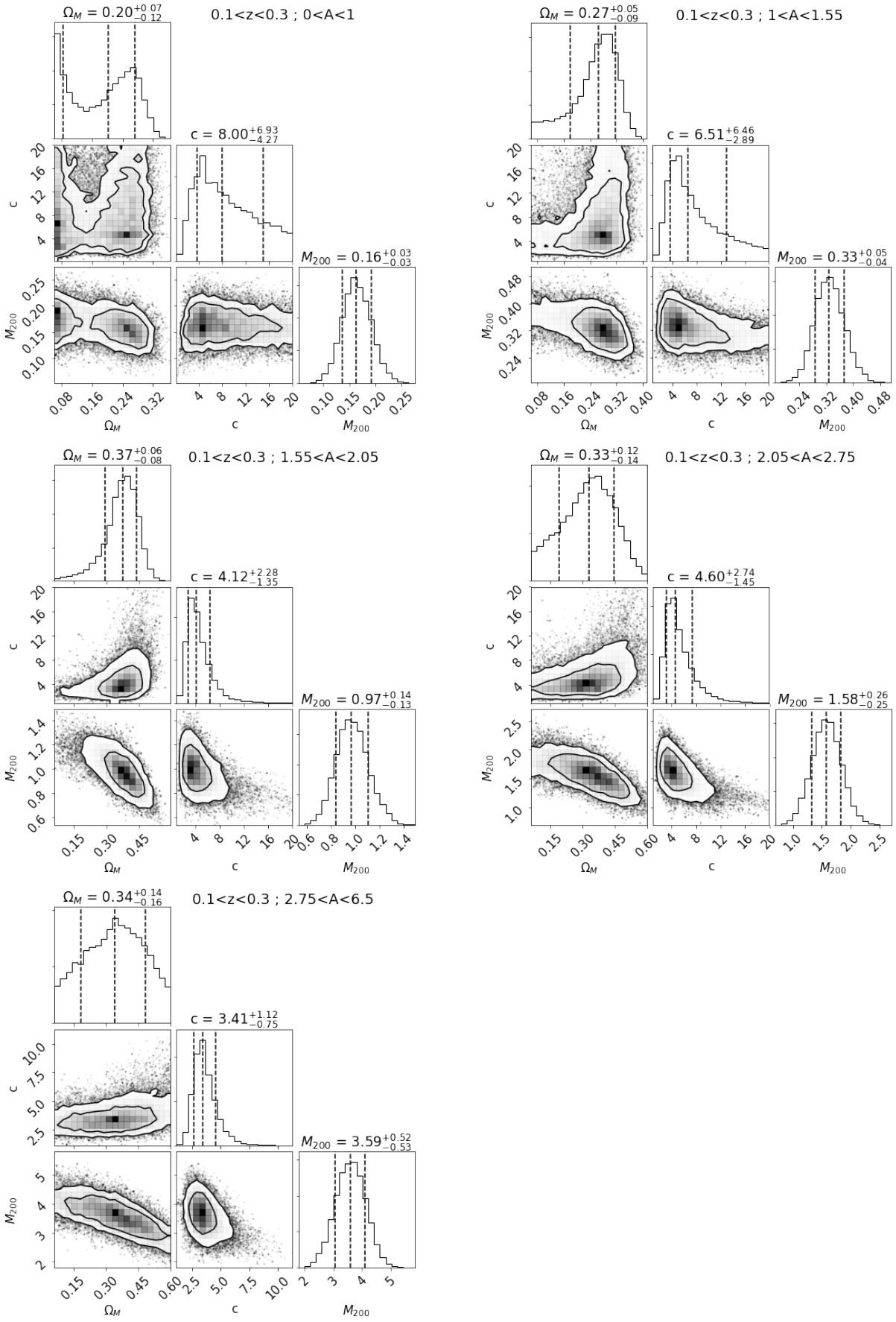

Figure 6.14: Posterior distributions of the three free parameters of the assumed (one+two)-halo lensing model (Ω_M, c, M_{200}) for the first redshift bin analysed ($0.1 < z < 0.3$). The dashed vertical lines represent the posterior medians, and the 18-th and 82-th percentiles. The values of these quantities are reported on top of each panel. The contour lines refer to $1-\sigma$ and $2-\sigma$.

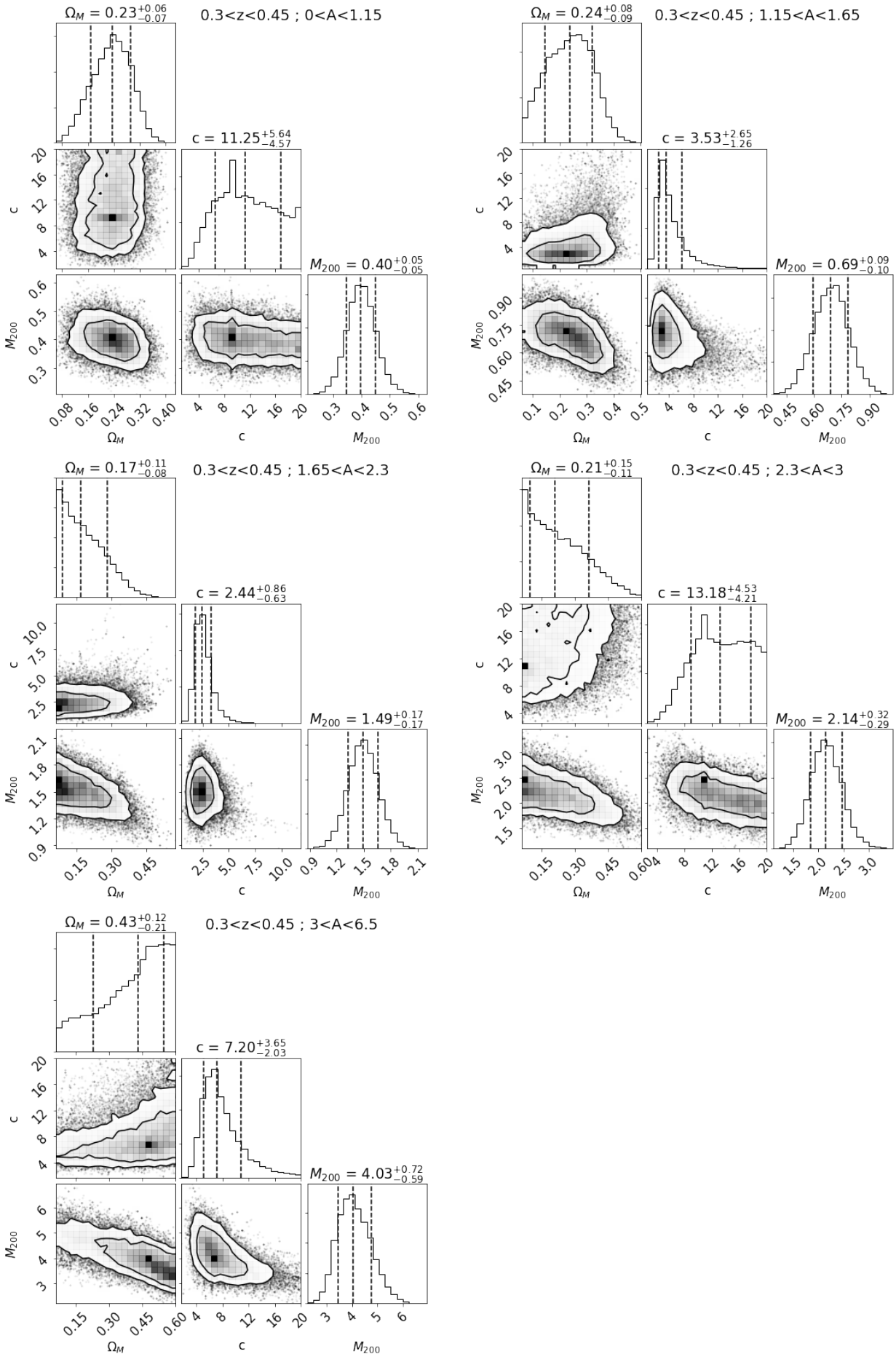

Figure 6.15: As Fig. 6.14, but for the second redshift bin analysed ($0.3 < z < 0.45$).

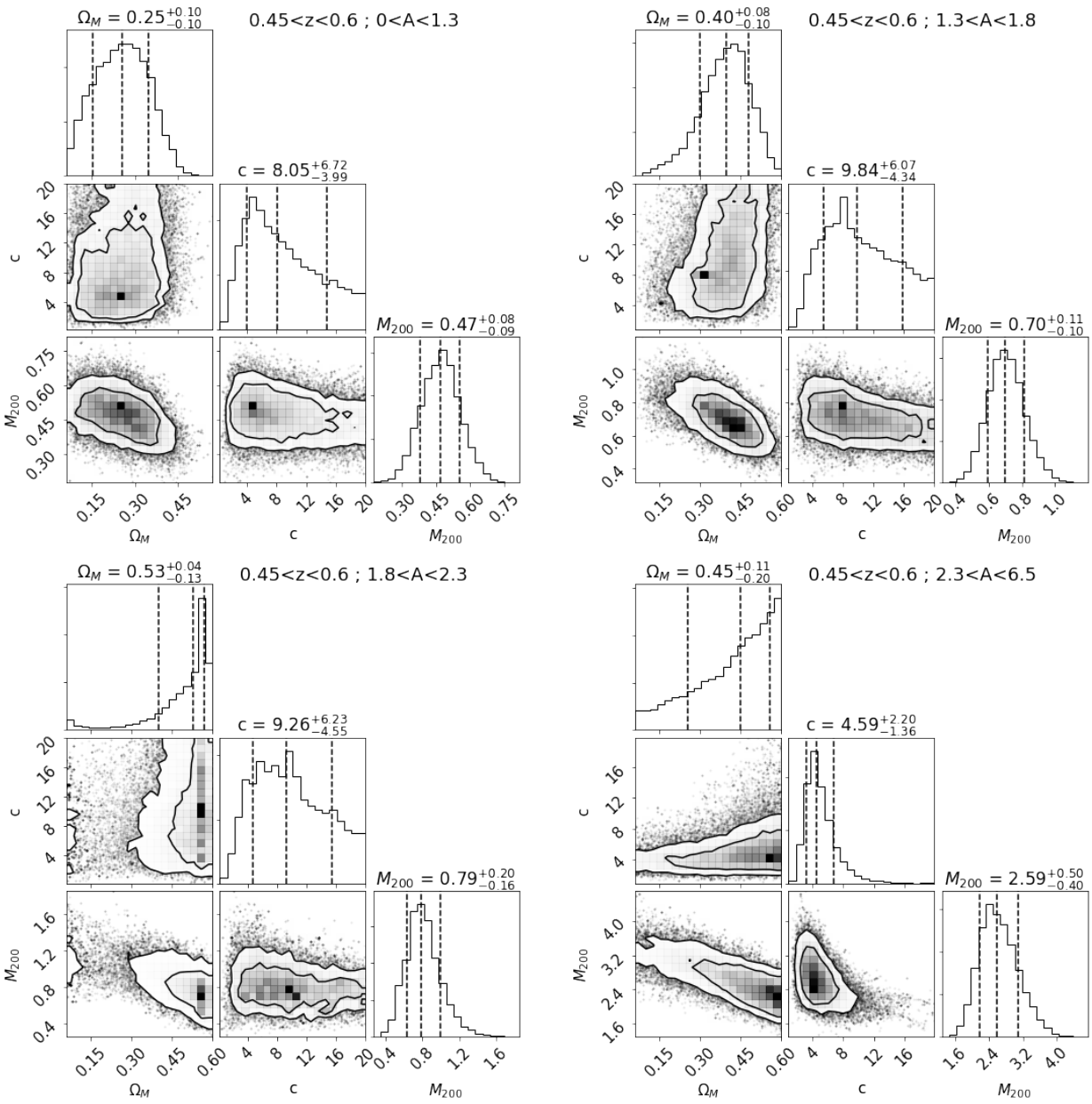

Figure 6.16: As Fig. 6.14, but for the third redshift bin analysed ($0.45 < z < 0.6$).

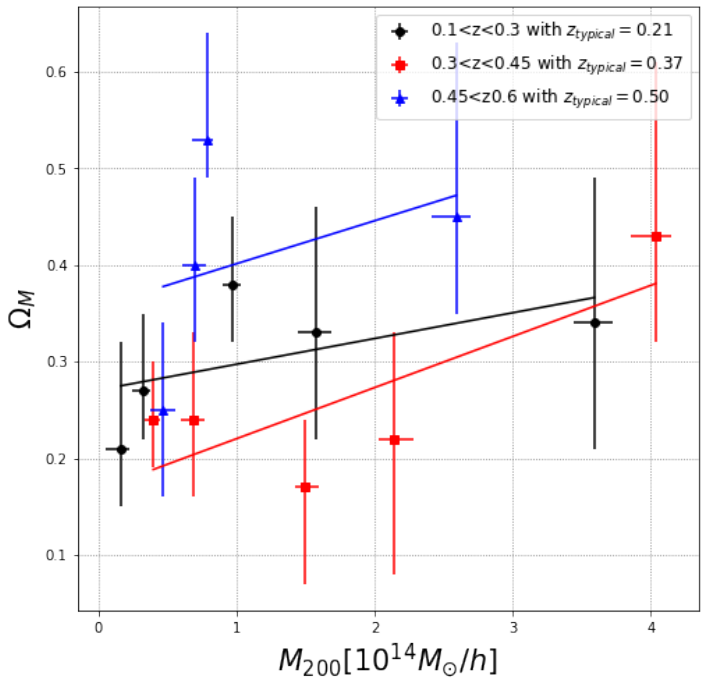

Figure 6.17: The Ω_M-M_{200} relation inferred from the modelling of the lensing profiles, marginalising over c. The error bars show the $1-\sigma$ uncertainties. The latter refer to the median values of the posteriors reported on top of the histograms in Figures 6.14, 6.15 and 6.16 for the first, the second and the third redshift bin, respectively. The solid lines are the best-fit linear relations in each redshift bin.

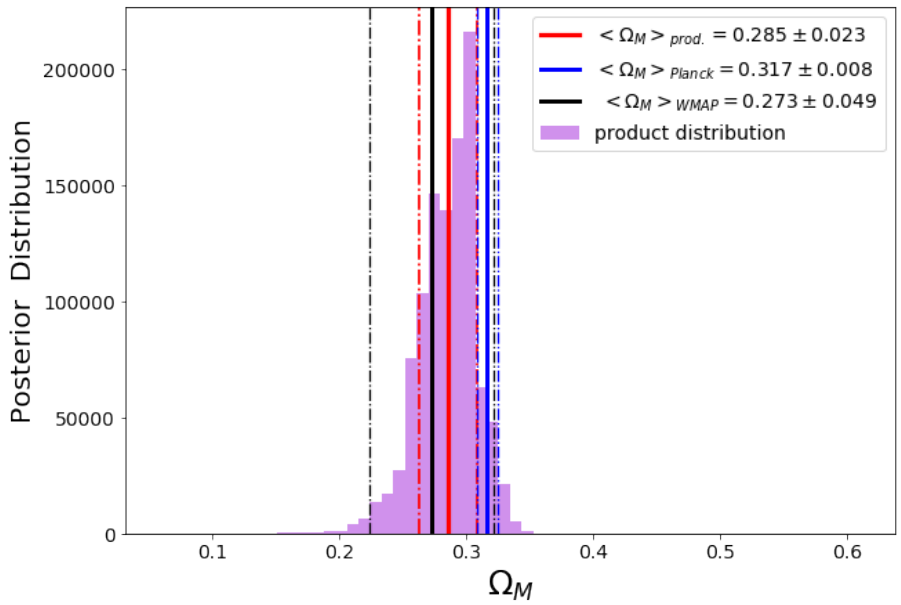

Figure 6.18: Combined Ω_M posterior distribution, estimated as the product of the 14 posterior distributions of each cluster sample shown in Fig. 6.14, 6.15 and 6.16. The symbols are the same as in Fig. 6.10. The difference is that here we marginalised over M_{200} and c.

Chapter 7

Conclusions and perspectives

In the era of precision cosmology, the progress in the establishment of independent methods to infer cosmological parameters is fundamental to test possible failures of the standard ΛCDM model, or hidden systematics. Galaxy clusters are biased tracers of the underlying mass distribution (Tinker et al., 2010). Gravitational lensing offers a unique probe into the total matter distributions in clusters. Indeed, the stacked lensing method has increasingly grown more and more successful in measuring the average masses of clusters, down to the less massive ones (Johnston et al., 2007; Mandelbaum et al., 2008; Covone et al., 2014; Bellagamba et al., 2019).

In this thesis work, we presented new cosmological results from the analysis of weak lensing signal produced by a galaxy cluster sample detected from the third data release of the KiDS survey (KiDS-DR3, de Jong et al., 2017; Hildebrandt et al., 2017). Specifically, we exploited photometric redshifts with relative probability distribution functions (Kuijken et al., 2015; de Jong et al., 2017), a global improved photometric calibration, weak lensing shear catalogues (Kuijken et al., 2015; Hildebrandt et al., 2017) and lensing-optimised image data. Our cluster sample, used to build the stacked shear profiles for the cosmological analysis, was obtained using the AMICO algorithm (Bellagamba et al., 2018) in Maturi et al. (2018). The redshift-amplitude binning of the cluster sample was set as in Bellagamba et al. (2019).

Using the CFHTLenS public shear catalogue, Covone et al. (2014) obtained stacked weak lensing signal of $N_{cl} = 1176$ galaxy clusters, in the redshift range $0.1 < z < 0.6$, with a number density of backgound galaxies $n_{gal} \sim 6$ per square arcminute. They built high S/N 1+2-halo shear profiles in six bins in optical richness, whith average masses spanning the range from 0.5 to $\sim 3 \times 10^{14} M_{\odot}/h$. In the same range of redshifts and, roughly, of masses, the cluster sample collected in Maturi et al. (2018) from KiDS-DR3, that we used in this work, is composed of 6972 objects. We exploited a survey area of ~ 360 deg^2, to be compared to the one analysed by Covone et al. (2014) of ~ 154 deg^2. This allowed us to collect shear profiles for 14 redshift-amplitude bins with $n_{gal} \sim 8.53$ per square arcminute. Moreover, the shear data and photometric redshift estimates of KiDS are better than those of its sister survey CFHTLenS.

Our studies support the use of the stacking method in weak lensing to calibrate galaxy cluster scaling relations (e.g Oguri and Takada, 2011). Furthermore, we corroborated the reliability of AMICO (Bellagamba et al., 2018) in detecting galaxy cluster samples. This algorithm was used to construct the first and only cluster sample from KiDS-DR3 data by Maturi et al. (2018). This cluster data-set successfully enabled us to infer cosmological parameters from shear measurements.

Modelling the shear profiles of these clusters up to large scales, we inferred robust cosmological constraints on Ω_M (see Figure 6.18). X-ray studies (Henry et al., 2009), together with CMB and cosmic shear measurements, highlighted the importance of understanding systematic effects associated with the use of directly observable quantities as proxies for mass. Our results show that AMICO detected structures down to few $10^{13} \mathrm{M}_{\odot}/h$ in the KiDS data, and provided a reliable mass proxy for each of them (see Table 6.1). Its application to planned KiDS Data Releases will provide one of the largest and deepest cluster samples with mass calibration available to date (Bellagamba et al., 2019).

We confirmed also the goodness of the modelling used in Bellagamba et al. (2019) to fit shear profiles (see Figures 6.11, 6.12 and 6.13). We considered a minimum radius of $0.2\,\mathrm{Mpc}/h$ in the modelling, properly taking into account the contribution of clusters miscentred with respect to the effective cluster centre. Anyway, since we exploited mainly the cosmological information in the

2-halo term, dominant at very large radii, our final constraints on Ω_M are negligibly affected by the choice of the lower fitting radius.

Note that the lensing signal, expected to arise in the 2-halo term, is generally too low to be carefully detected in the outskirts of individual clusters (but see Sereno et al., 2018). Stacking the lensing signal of a significant number of clusters enabled us to detect the 2-halo term at relatively high S/N. Hence we further confirmed the presence of the 2-halo contribute in the lensing signal produced by the large-scale structures of the Universe. Furthermore, we assessed the efficiency of the model given by Eq. (5.4.7) (Oguri and Hamana, 2011) to describe the weak lensing signal produced by the 2-halo term.

In our thesis work, we constrained simultaneously the cluster masses, M_{200}, and Ω_M (see Figures 6.14, 6.15, 6.16), assuming a robust scaling relation between the cluster masses and the observed amplitudes (see Figure 6.4), and using a 1+2-halo model. The masses were also inferred from the 1-halo model at small scales, assuming the truncated NFW model given by Eq. (5.4.1), finding consistent results. The results presented in this work are fully consistent with the theoretical predictions of the ΛCDM model. Marginalising over the cluster masses and concentrations, we get $\Omega_M = 0.285 \pm 0.023$, from the full MCMC posterior distribution, assuming a flat geometry. This value is competitive and consistent with that estimated by WMPA9 ($\Omega_{M,WMAP9} = 0.273 \pm 0.049$), but slightly in tension with the Planck18 constraints ($\Omega_{M,Planck18} = 0.317 \pm 0.008$). However, in this thesis work we did not carry out a detailed analysis on the possible systematic uncertainties arising when inferring Ω_M from shear profiles measured up to very large radii (we discussed it in Sect 6.6), so our assessed uncertainties might be slightly underestimated.

Surely, enhancing the number of clusters to stack, the uncertaintiy on the stacked shear data points decreases because of the nature of the cluster-shear cross-correlation. Hence we await the progress of KiDS to improve our stacked shear measurements, and subsequently the *Euclid* mission launch. In fact, upcoming deeper galaxy surveys will let us detect and stack lensing clusters in small redshift bins at $z \lesssim 1.5$, enabling us to break parameter degeneracies and limit systematics (Oguri and Takada, 2011). The accuracy in the estimate of Ω_M will significantly benefit from the advancement planned for KiDS when it will reach its final coverage of $1500\,\mathrm{deg}^2$ of sky, with a median redshift of ~ 0.8. The same is true for the upcoming optical galaxy surveys. As test bed, we can consider the large area survey conceived for the *Euclid* mission, whose launch is planned for 2022[1]. The lensing signal can be incremented by considering a larger number of clusters (which can be attained with either deeper or wider surveys), background sources (deeper surveys) and a wider survey area (in order to collect the 2-halo term lensing signal up to 50 Mpc/h). These three aspects will be reached by *Euclid* which will represent a significant improvement compared to the data-set used in this work. *Euclid* will cover an area of ~ 15000 deg^2 ($\times 10$ that planned for KiDS) and it will detect $n_g \sim 30$ galaxies per square arcminutes, with a median redshift larger than 0.9. Therefore, it will be conveniently exploited for weak lensing analyses (Laureijs et al., 2011). These primary properties are sufficient to predict the expected improvement in the lensing signal, as well as the corresponding accuracy expected in the determination of Ω_M from lensing analysis. Due to the improved photometry, a larger number of clusters will be detected to higher redshifts. In this context, AMICO, that has been recently selected as the primary cluster detection algorithm for the *Euclid* mission, is expected to provide excellent results. Anyway, most of the clusters are expected to be detected in low-mass haloes, that produce small lensing signal. Morever, the photometric redshift determination for small groups might be significantly uncertain, getting in the way of the weak lensing analysis. This drawback will be balanced by the significant redshift range extension of the background galaxies of the survey, because their lensing signal is maximised at high redshifts. Excluding measures of shear profiles at very large radii (2-halo region), the lensing S/N rescales with the number of clusters as $S/N \propto \sqrt{N_{cl}}$, hence we can consider a significant improvement of the lensing signal detected by *Euclid* with respect to the present analysis, due to the larger density of detected clusters. The same consideration is valid for the enhancing of the lensing signal due to a higher number of background galaxies in *Euclid* ($n_{gal} \sim 30$ galaxies per arcminute) with respect to KiDS ($n_{gal} \sim 22$). Finally, the broad coverage of the *Euclid* survey will allow us to detect the 2-halo lensing signal up to its full radial extent (50Mpc/h), decreasing the error in the lensing estimate of cosmological parameters (Sereno et al., 2017). Anyway, this effect would be counterbalanced by the increased overlap in the lensing area of near clusters. The Ω_M estimate might benefit from enhanced spectroscopic sample associated to the *Euclid* survey because spectroscopic redshifts are more precise than the photometric ones, drastically diminishing the main source of uncertainty

[1] https://www.*Euclid*-ec.org/

in our kind of analysis (see Section 6.2). The increased spectroscopic sample associated with the *Euclid* survey will greatly improve the cosmological estimates of our kind of analysis.

Another efficient approach to drastically improve the constraints on cosmological parameter estimates is to use the same cluster population used for the lensing analysis togheter with a joint halo clustering analysis, i.e inferring the correlation function of galaxy clusters. Furthermore, this method minimises the modelling by keeping only the information far beyond the cluster virial radius. The simultaneous analysis of stacked lensing and clustering proved to be particularly efficient to constrain the halo bias together with the matter linear power spectrum amplitude (Sereno et al., 2017). We aim at completing and improving our analysis for the inference on Ω_M with the same approach. The features of a joint analysis of clustering and weak lensing will determine the matter density parameter to a very good accuracy and, when considering the opportunities offered by *Euclid*, we expect even better results.

Appendix A

Bayesian statistics

This appendix is based on the review notes of the course: "Practical Statistics for Physics and Astrophysics" held by Prof. R. B. Metcalf at the Department of Physics and Astronomy of the University of Bologna. The notes can be downloaded at the following link: notes.pdf.

In the Bayesian approach, the probability is a measure of the degree of credibility of a proposition. So this is a *"subjective"* approach, in the sense that the probability of an event is estimated taking into account also the prior knowledge we might have on that event. This definition is applicable to any event. Through statistical Bayesian inference we can constrain models for physical processes, that describe the probabilistic distribution of the data. In particular, the Bayesian approach provides us the possibility to calculate the probability of a model, or of model parameters, given the data and any prior knowledge. The Bayesian probability is an inverse probability, that is, it consists in ascending from the data observed to a model probability. The probabilities are not interpreted as frequencies, but rather as levels of confidence in the occurrence of a given event. One of the advantages of the Bayesian interpretation of probability is that it enables us to assign the probability that a model is the best one in describing the data, with respect to the other models considered. On the other side, through the so-called frequentist approach, it is not possible to assign a probability to a model, as only the data are probabilistic. Nevertheless it can be shown that when a frequentist result exists, and in the limit of a very large sampling, the Bayesian and frequentist results coincide. The Bayesian statistics is based on the *Bayes theorem*:

$$P(A|B) = \frac{P(A)P(B|A)}{P(B)}\,. \tag{A.0.1}$$

This theorem comes from the logical requirement that $P(A, B)$ is the same as $P(B, A)$, and from the *product rule* $P(A, B) = P(A)P(B|A)$[1], where $P(A, B)$ is the *joint probability* of events A and B, while $P(B|A)$ is the *conditional probability* of B given that A is true.

The Bayes theorem allows to *convert* the a priori probability into the a posteriori probability: an estimate of the degree of credibility of a given hypothesis is used before data observation to associate a numerical value with the degree of credibility of that hypothesis, after the observation of the data.

Posterior, likelihood, prior and evidence

Let $\mathcal{D}$ be some amount of data and M_i a model that should describe it, being a member of the set $\{M_1, M_2, ...\}$. The latter might be totally different from each other, e.g. with different parameters, or they might differ by only the values of the model parameters. Let I represents everything else in the Universe, that we will take to be fixed or irrelevant to our experiment. Applying Eq. (A.0.1) we get:

$$P(M_i|\mathcal{D}, I) = \frac{P(\mathcal{D}|M_i, I)P(M_i|I)}{P(\mathcal{D}|I)} = \frac{P(\mathcal{D}|M_i, I)P(M_i|I)}{\sum_i P(\mathcal{D}|M_i, I)P(M_i|I)}, \tag{A.0.2}$$

where the second equality follows from $P(\mathcal{D}|I) = \sum_i P(\mathcal{D}, M_i|I) = \sum_i P(\mathcal{D}|M_i, I)P(M_i|I)$, which is the probability that the data will occur assuming the correct model is one of the M_i's. Here, I

[1] The *product rule* and the *sum rule* $(P(A) + P(\bar{A}) = 1$ (where $\bar{A}$ means "not A"), represent the two fundamental axioms of probability theory; from these ones we can derive all the necessary properties of probability.

is included only to emphasize the fact that every probability has some implicit assumptions which could be incorporated into the model but, if they have no effect on the outcome of the experiment or they were never changed when the experiments were conducted, they can be considered conditionals for all the probabilities. In the following, I will be considered implicit and not included.

Each factor in Eq.(A.0.2) has a special name:

- $P(M_i|\mathcal{D})$ is called the *posterior probability* for the model M_i, given the data. Estimating the full posterior probability distribution is the final *goal* of the Bayesian inference, though it is often sufficient to estimate the posterior median (or mode) and confidence regions.

- $P(\mathcal{D}|M_i)$ is called the *likelihood* and it is the probability of getting the observed data given the model M_i. It is often denoted as $\mathcal{L}(\mathcal{D}|M_i)$. Note that this is the probability used in frequentist methods. Often the likelihood function can be assumed to be Gaussian. It includes the model used to describe the data, and the covariance matrix which describes the noise. The one-dimensional Gaussian likelihood can be written as follows:

$$\mathcal{L}(\theta) = p(x|\theta) = \sqrt{\frac{1}{2\pi\sigma^2}} \exp\Big(-\frac{(x-\mu(\theta))^2}{2\sigma^2}\Big), \tag{A.0.3}$$

  where μ is the mean of the model parameter θ and σ^2 is the variance.

- $P(M_i)$ is called the *prior* and it is the probability of the model prior to the data, $\mathcal{D}$, being considered. This might take into account some previous experiment with data $\mathcal{D}'$, in which case it would be the posterior of that experiment, $P(M_i|\mathcal{D}')$. It might also take into account that some models, or range of parameters, is not possible, in which case $P(M_i) = 0$ for some i.

- $P(\mathcal{D}|I) = \sum_i P(\mathcal{D}|M_i, I)P(M_i|I)$ is called the *evidence* (or marginal likelihood) and it is not a function of M_i although it is implicitly dependent on the set of all models considered. Since the data do not change, the evidence will be a constant for a fixed set of models. For the estimation of the model parameters (which is done by fixing a model), the evidence is simply a constant that normalises the posterior probability, and it can be ignored. Instead, it becomes crucial when you want to compare different models. The evidence is sometimes denoted as $\epsilon(\mathcal{D})$.

$P(M_i|\mathcal{D}, I)$ can be considered as the prior conditional probability of a new experiment. This is referred as *Bayesian's rule*. The proceedings can be viewed as updating our knowledge of the model after we take into account new data or information and it can be thought of as a kind of chain where every bit of new information, data, updates our knowledge progressively.

Parameter estimation

The Bayesian inference can be effectively exploited to assess model parameters and their uncertainties. Let us assume to have a model written as a function of the parameters $\{\theta_1, \theta_2, ...\}$. We will assume that these parameters take on a continuous range of values, although this is not necessary. The sum in the evidence then becomes an integral and the posterior is:

$$P(\theta_1, \theta_2, ...|\mathcal{D}) = \frac{\mathcal{L}(\mathcal{D}|\theta_1, \theta_2, ...)p(\theta_1, \theta_2, ...)}{[\int \mathrm{d}^n\theta \mathcal{L}(\mathcal{D}|\theta_1, \theta_2, ...)p(\theta_1, \theta_2, ...)]}\,. \tag{A.0.4}$$

Eq.(A.0.4) provides the probability distribution for the set of parameter values, *given that the model is the correct one.*

Note that the *Bayesian analysis is always relative.* Almost always we compare one model to another or many others. A corollary to this is that we will always get an answer even when the model is completely wrong. The posterior for a model that fits the data very badly will often look just fine: there will usually be a set of parameters that fit the data best, but that does not meant they fit the data well. A Bayesian model selection is required to address this issue (see Section A).

Marginalization

Some model parameters might be not of interest for a particular statistical analysis. In the Bayesian context these parameters are called *nuisance parameters*. We can marginalize over them by integrating over the posterior distribution. Let us say parameters $\alpha_1, \alpha_2, ...$ are the parameters we are interested in, and parameters $\beta_1, \beta_2, ...$ are the nuisance parameters. We can write the likelihood as:

$$\begin{aligned} P(\alpha_1, \alpha_2, ...|\mathcal{D}) &= \int_{-\infty}^{\infty} \mathrm{d}\beta_1 \int_{-\infty}^{\infty} \mathrm{d}\beta_2 ... P(\alpha_1, \alpha_2, ..., \beta_1, \beta_2, ...|\mathcal{D}) \\ &= \frac{\int_{-\infty}^{\infty} \mathrm{d}\beta_1 \int_{-\infty}^{\infty} \mathrm{d}\beta_2 ... P(\mathcal{D}|\alpha_1, \alpha_2, ..., \beta_1, \beta_2, ...) P(\alpha_1, \alpha_2, ..., \beta_1, \beta_2, ...)}{\epsilon(\mathcal{D})}. \end{aligned} \tag{A.0.5}$$

Choice of priors

In the Bayesian approach, prior information is properly used to assign the probability to a given event. As its name hints, the prior represents the information we had about the parameters before using the current data to constrain them. This information might come from a previous experiment or observation, in which case the prior would be the posterior of that experiment. The prior can also express the theoretically permissible values range of a parameter. In fact there is often some boundaries we can set on parameter values, at least on a theoretical level. The parameter priors may be not important if the likelihood is small at the boundaries of parameter space that they will not affect the integral in the evidence, Therefore the posterior will be zero at these points

In the case of complete prior ignorance, the use of a uniform prior is the safest choice. This is the prior that is constant over a region of parameter space and zero outside of it. Of course, it is superfluous to specify the limits when likelihood is null at the boundaries, since the normalisation, that appears both in the prior and in the evidence, cancels out. The prior imposes a metric on the parameter space. The prior probability for a parameter α being in an infinitesimal region is $p(\alpha)\mathrm{d}\alpha$. What is a uniform prior for one set of parameters will not be a uniform prior for another set, even though they might describe the same model.

Note that the normalisation of the uniform prior is formally zero if the parameter range $-\infty < \alpha < +\infty$. These ranges are routinely used when the posterior (likelihood times prior) has a well defined integral. These are examples of improper prior distributions that are not valid distributions by themselves, but make sense in a posterior. Having a general way of including prior information is an important advantage to the Bayesian method; the frequentist approach does not have this advantage. In general, if the data strongly constrain the parameters beyond what was previously known, the choice of prior should not strongly affect the resulting posterior.

Model selection

Now we will address a more general problem than just parameter estimation. Let us assume there are different competing models that might describe the data, but that these models do not just differ from each other by having different parameter values. Specifically, the models might be described by different functions, with possibly different number of parameters. We can ask ourselves which model is more strongly supported by the data, or if it is justified to add the extra parameters: this is called *model selection*.

Let us consider a set of all possible models $\{M_1, M_2, ...\}$ that should describe the data. We can write down the posterior for model M_i using Bayes' theorem as in the parameter estimation case:

$$P(M_i|\mathcal{D}) = \frac{P(\mathcal{D}|M_i)P(M_i)}{P(\mathcal{D})} = \frac{P(\mathcal{D}|M_i)P(M_i)}{\sum_i P(\mathcal{D}|M_i)P(M_i)} \tag{A.0.6}$$

Model selection is often restricted to comparing the relative probability of two models, M_1 and M_2. This can be done by taking the ratio of their posteriors:

$$O_{1,2} = \frac{P(M_1|\mathcal{D})}{P(M_2|\mathcal{D})} = \frac{P(\mathcal{D}|M_1)P(M_1)}{P(\mathcal{D}|M_2)P(M_2)} = B_{1,2}\frac{P(M_1)}{P(M_2)}. \tag{A.0.7}$$

$O_{1,2}$ is called the *odds* of model 1 relative to model 2 and $B_{1,2}$, the ratio of the model likelihoods, is known as the *Bayes's factor*. If the prior probabilities are equal, as they often are, then the odds is

equal to Bayes' factor. Note that $P(\mathcal{D})$ cancels out so we can avoid needing to know the probability of the data over all possible models. If the odds is large then model 1 is favored, vice-versa if it is small.

Now, we can ask ourselves how we can calculate the likelihood $P(\mathcal{D}|M)$. In the parameter estimation problem we stayed within one model, and for this reason all the probabilities were conditional to this model being true, although that was not explicitly shown. Let us consider a generic parameter θ. We can write the Bayes' theorem again with the model conditionality explicitly shown:

$$P(\theta|\mathcal{D}, M) = \frac{P(\mathcal{D}|\theta, M)P(\theta|M)}{P(\mathcal{D}|M)}. \tag{A.0.8}$$

We can now see that $P(\mathcal{D}|M)$ is actually the evidence:

$$P(\mathcal{D}|M_i) = \int_{-\infty}^{+\infty} \mathrm{d}\theta P(\mathcal{D}|\theta, M_i)P(\theta|M_i) = \epsilon(\mathcal{D}|M_i), \tag{A.0.9}$$

where the integral is over all of parameter space within model M_i. In conclusion, Bayes' factor is the ratio of evidences for two models.

Calculating the evidence

It is often difficult, if not impossible, to achieve an analytic expression for the evidence, that is the normalisation of the posterior. In practice, it is usually calculated numerically by integrating the likelihood times the prior over the parameter space. This is usually a simple task if there are only one, two or three parameters because, in this case, we can simply grid the parameter space or use a standard integration routine.

Note that if we are estimating the parameters we only needs the posterior, because any factors in the likelihood and prior, that do not depend on the parameters, will cancel out. For this reason it is not necessary to normalise these probabilities individually, but just the product of them. This can save a lot of time and calculations, especially when the likelihood or prior are something strange of which we do not know the normalisation. On the other hand, the likelihood and prior need to be properly normalised to get the evidence. When the dimension of the parameter space is larger (≥ 3), numerical integration can be much more difficult.

Monte Carlo Markov Chain

In the ideal case, we should be able to construct an analytic expression for the likelihood, prior and posterior, and execute integrals over them, either analytically or by standard numerical integration methods, to find expectation statistical values or the integrated probability for a variable being in a certain region. However, sometimes these integrals are very difficult to carry out because the dimension of parameter(data)-space is high, and sometimes there is not an analytic expression for these probability distributions (for example when a simulation is used to go from model parameters to predictions).

To integrate over an analytic function, the best thing is having a large sample of deviates drawn from the distribution. With a sufficiently large sample drawn from a distribution, we can exploit the *law of large numbers* to estimate any expectation value:

$$E[g(x)] = \int_{-\infty}^{+\infty} \mathrm{d}^n x p(\mathbf{x}) g(\mathbf{x}) \approx \frac{1}{n} \sum_i g(\mathbf{x}_i)\,, \tag{A.0.10}$$

where $\mathbf{x}_i$'s are drawn from the distribution $p(\mathbf{x})$. In one dimension, it is often possible to efficiently sample from a standard distribution function. There are several methods used to find these deviates such as rejection and transformation methods, but we will not describe them in this appendix, because they are not used in our analysis.

The Monte Carlo integration is a way of exploiting the law of large numbers (A.0.10). It consists of extracting random numbers to estimate a multidimensional integral that cannot be computed analytically or by using a standard one-dimensional methods, such as for example the trapezoids. It is typically used when the number of dimensions is high and/or the boundaries to the region of integration are complicated.

The frequentist approach can be summarised as follows: determining a Maximum Likelihood Estimator MLE (which is a function of the data and therefore a random variable) and its distribution, that is a function of the data. In most cases this step is performed numerically by generating mock datasets and determining for each of them the MLE. The likelihood and the Monte-Carlo method are therefore the basis of the *frequentist* approach.

Markov chain

The Monte Carlo Markov Chain (MCMCs) method allows to numerically map the posterior, even in very complicated cases where, for example, the likelihood is obtained through numerical simulations, the parameter space has a large number of dimensions and the posterior has a complex structure with many peaks. Through MCMC we sample the posterior the most efficiently by concentrating in the regions where it is larger. In this way the number of points to be calculated increases approximately linearly rather than exponentially, with N. The term Monte Carlo refers to the fact that for the computation we resort to a repeated random sampling (through the generations of sequences of random numbers).

In statistics, a *chain* is an ordered series of random numbers $x_1, ..., x_n, ...$ where the conditional probability of each element, given the other elements, is specified: $p(x_n|x_1, ...)$. It is possible to think of the whole chain as being a single random object. The theory on chains is extensive. In fact, chains can be used to model almost everything. There are many different types of chains with different properties. Here we will concentrate only on the Markov chains, that is commonly used in scientific inference problems, and on the properties that are important to our application.

A MCMC is a chain where the conditional probability of any element x_n depends only on the previous element x_{n-1} and not from how we came to this state (from the history). The probability $p(\mathbf{x}_{n+1}|\mathbf{x}_n)$ is known as the Markov chain's *transition kernel*. If the transition kernel is independent of n it is said to be time-homogeneous. The chains we are interested in are *ergodic chains*. To be ergodic the chain must be:

- *irreducible*, i.e a chain starting at any state x_0 can reach any other state after a finite number of steps, not necessarily one step;
- *aperiodic*, i.e the chain will not return to the same state after some fixed number of steps;
- *positive recurrent*, i.e the expectation value for the number of steps between any two states is finite.

It is also true that a Markov chain is ergodic if there is a number N such that any state can be reached from any other state in N steps. The most important consequence of ergodicity is that the chain has a *unique stationary distribution* $f(\mathbf{x})$ such that

$$\int_{-\infty}^{+\infty} \mathrm{d}x_n f(\mathbf{x}_n) p(\mathbf{x}_{n+1}|\mathbf{x}_n) = f(\mathbf{x}_{n+1})\,, \tag{A.0.11}$$

which means that we are able to realize chains whose states are distributed according to $f(\mathbf{x})$, if we can find a transition kernel that satisfies this requirement, and the law of large numbers:

$$E[g(\mathbf{x})] = \lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N} g(\mathbf{x}_n) \tag{A.0.12}$$

will be applicable. Note that the transition kernel is not unique for a particular $f(\mathbf{x})$. We can also select one or two parameters and make a histogram which should be a representation of the stationary distribution. Also note that having a stationary distribution does not mean that each element in the chain is independent, i.e $f(\mathbf{x}_{n+1}, \mathbf{x}_n) \neq f(\mathbf{x}_{n+1}) f(\mathbf{x}_n)$. In fact, states that are very far separated in the chain are not always independent, but as the separation increases they will in case become independent. Normally, an MCMC algorithm generates Markov chains of samples, each of which is autocorrelated to those generated immediately before and after him. Consequently, successive samples are not independent but form a Markov chain with a certain degree of correlation.

Metropolis-Hastings algorithm

Like other MCMC methods, the aim of the Metropolis-Hastings (MH) algorithm (Hastings, 1970) is to generate a collection of states that follow a given distribution (in the Bayesian case, the posterior). The MH algorithm is based on finding a transition kernel that will have any desired stationary distribution. The kernel has to satisfy the *detailed balance equation*:

$$p(\mathbf{x}_{n+1}|\mathbf{x}_n)f(\mathbf{x}_n) = p(\mathbf{x}_n|\mathbf{x}_{n+1})f(\mathbf{x}_{n+1}) \tag{A.0.13}$$

for all n. It can be easily seen that if $p(\mathbf{x}_{n+1})$ satisfies the detailed balance it will also satisfy Eq. (A.0.11). We can think of this as if there were a flow of points out of state $\mathbf{x}_n$ into $\mathbf{x}_{n+1}$ and a counter flow out of $\mathbf{x}_{n+1}$ into $\mathbf{x}_n$. The flow is proportional to the probability of being in the first state times the probability of transitioning. Detailed balance requires that the flow and counter flow between every pair of states to be equal. The stationary state will then be the steady state and the time the chain spends in a given state will be proportional to $f(\mathbf{x})$. The MH algorithm runs, starting at state $\mathbf{x}_n$, as follows:

1. Choose a trial point $\mathbf{x}_t$ from a *proposal distribution* $q(\mathbf{x}_t|\mathbf{x}_n)$: this is the *initialization*.
2. Calculate the probability of acceptance of the chosen point:

$$\alpha(\mathbf{x}_t, \mathbf{x}_n) = \min\left\{1, \frac{q(\mathbf{x}_n|\mathbf{x}_t)f(\mathbf{x}_t)}{q(\mathbf{x}_t|\mathbf{x}_n)f(\mathbf{x}_n)}\right\} \tag{A.0.14}$$

3. At each iteration i, if $\alpha < 1$ we draw a uniform random number between 0 and 1. If α is larger than this number we accept the trial state and set $\mathbf{x}_{n+1} = \mathbf{x}_t$. Otherwise set $\mathbf{x}_{n+1} = \mathbf{x}_n$. In other words, we accept the trial state with probability $\alpha(\mathbf{x}_t, \mathbf{x}_n)$.
4. Go to step 1.

We can easily see that the detailed balance is satisfied by this algorithm.

Choosing a proposal distribution

Even if the MCMC is guaranteed to converge under the conditions mentioned above, it might take a very long time. The chain moves across the parameter space in a random walk and, if it does not reach every region of significant probability many times, it will not be a good approximation of an independent sampling from the stationary distribution. To reach a good mixing, the *rejection rate* of the proposed moves must not be too high or too low. If it is too high, the chain will have many duplicated points that will not fill the parameter space in an even way. If the rejection rate is too low the chain will move, but not fast enough to get around the space. There is a great deal of freedom in choosing a proposal function, and finding the right one for a particular problem requires much experience. There is nothing stopping one from making steps in one dimension at a time as long as all dimensions are possibly explored. One can for example cycle through the parameters or pick a parameter at random each step. Sometimes this can improve the convergence.

To initialize the chain we must guess a point in the parameter space. This will usually not ensure high probability unless you are a good guesser. The chain will be attracted by the high probability regions, but it might take a lot of time to get there. During this *burn-in* period, the chain is not close to the maximum of the distribution. For this reason one usually discards the first part of the chain. It is often useful to use a proposal distribution that jumps longer during this burn-in stage, while the chain is searching for the peak(s), and then reduce the jumps later on to get an acceptable rejection rate during the rest of the chain. On the other hand, if the chain starts near a maximum of the posterior distribution, it might not be necessary to discard a burn-in period.

The largest difficulties with MCMC arise in the following cases.

- When the *initial guess* is so far from any peaks and the probability is so flat out there that the chain never finds a peak. Sometimes this situation happens in low probability regions, where the calculation of $f(\mathbf{x})$ has numerical underflow errors, or it is dominated by numerical noise, in which case the $f(\mathbf{x})$ chain may wander around without getting anywhere.

- When the parameters are *degenerate*. Imagine a $f(\mathbf{x})$ that has a narrow ridge. If the proposal distribution is isotropic, it will be either too wide in one direction so that the rejection rate is too high, or it will be too small and creeps along the ridge very slowly. In the case of a linear degeneracy, you might be able to learn something about the distribution and then make your proposal distribution anisotropic in a way that improves convergence. A non-linear degeneracy may cause a much harder issue. In this case a proposal distribution that works well in one point of the parameter space might not work well in another. When possible, one should try to eliminate known non-linear degeneracies by changing the variables.

- When the distribution has *multiple modes*. This is probably the hardest problem to deal with. If there are multiple peaks in the distribution that are separated by regions of low probability, the chain can easily get caught in one peak where its probability of transitioning to the other is very small. We might adjust the proposal distribution to get a good rejection rate for one peak, but that might make the probability of jumping between peaks effectively zero. This problem is exacerbated in large dimensional space because peaks that might not seem to be far away by their *Euclid*ean distance d are in a volume that goes up like d^D, where D is the dimension of space. An effective strategy to get rid of this issue is to run multiple chains with different random initial states and see if they find different modes.

Convergence

It might be difficult in general to assess when the chain has converged. Unfortunately, there are no trivial proof ways to determine that. One thing we can do is to calculate the auto-correlation for each of the parameters as a function of the *lag* m, that is the separation in the chain. The cross-correlation can be defined as follows:

$$C_{\alpha,\beta}(m) = \frac{\sum_{i=1}^{N-m}(\alpha_i - \bar{\alpha})(\beta_{i+m} - \bar{\beta})}{\sqrt{\sum_{i=1}^{N-m}(\alpha_i - \bar{\alpha})^2 \sum_{i=1}^{N}(\beta_i - \bar{\beta})^2}}\,, \tag{A.0.15}$$

where α and β are the parameter values, and $C_{\alpha,\beta}(0) = 1$. In the case of the auto-correlation $\alpha = \beta$. Distant points along the chain should not be correlated: in this case this function should oscillate around zero for large lag. The first time this function drops to zero, or close enough to zero, is an estimate of the *correlation length*, that we call N_{corr}. Points separated by a length less than the correlation length will not be independent. We can define an *effective number of independent samples in the chain* as:

$$N_{eff} = \frac{N_{chain}}{N_{corr}}\,. \tag{A.0.16}$$

We want this number to be large. We also want any statistics we are interested in to depend on a number of points that is much larger than N_{corr}. For example, the difference between the 95% and 99% contour levels depends on only 4% of the points. This might be smaller than N_{corr}. To be conservative, $N_{corr} \sim 200 - 600$ but, in practice, this criterion can be fooling. For example, a chain that is caught in one mode of a multi-modal distribution might appear to be converging nicely.

Combining posterior probabilities of independent data

Here we will discuss how to combine the posterior probabilities estimated from an ensemble of *independent* data, when every data is used to estimate the same parameters. This discussion is taken from Bailer-Jones and Smith (2011).

As explained in Sec. A, Eq. (A.0.2) is the general equation to be used for updating probabilities. For the purposes of this discussion, we rewrite it in a different form, expressing the Bayes' theorem (Eq. (A.0.1)) adding the explicit dependence on a third parameter, E, that might be another measurement, or a theoretical assumption:

$$P(A|B,E) = \frac{P(B|A,E)P(A|E)}{P(B|E)}\,. \tag{A.0.17}$$

The meaning of Eq. (A.0.17) is the following: if we have an initial estimate of the probability of A given exclusively by the information E, i.e $P(A|E)$, and then we measure a new data B, Bayes' theorem gives us a way of updating our inference to assess the probability of A based on both B and E, $P(A|B,E)$. The posterior from the first inference, P(A|E), becomes the prior of the next one. In fact, it might be convenient to think in terms of just *independent pieces of information* expressed as conditional probabilities, rather than in terms of *priors* and *posteriors* (Bailer-Jones and Smith, 2011).

Let us consider now to have two sets of information, D_A and D_B, and that we have used each of them to independently infer some parameter C of a model describing these datasets. For instance, D_A and D_B might be two unrelated shear datasets coming from two different lenses. Suppose we have a model (in this example case, a unique model describing the surface density distribution of the lenses) with which we can estimate the $P(C|D_A)$ and $P(C|D_B)$ probabilities. We want to combine them to obtain the $P(C|D_A,D_B)$ probability, that is we can in case exploite them to assess our best estimation of the C parameter. To obtain this combination we resort to the Bayes theorem:

$$P(C|D_A,D_B) = \frac{P(D_A,D_B|C)P(C)}{P(D_A,D_B)}\,. \tag{A.0.18}$$

If we can assume that D_A and D_B are *independent measurements*, then this usually means that they are *conditionally independent* given a model[2], that is:

$$P(D_A,D_B|C) = P(D_A|C)\times P(D_B|C). \tag{A.0.19}$$

Consequently, Eq. (A.0.18) can be written as:

$$P(C|D_A,D_B) = \frac{P(D_A|C)P(D_B|C)P(C)}{P(D_A,D_B)}. \tag{A.0.20}$$

Now, using the Bayes theorem to rewrite $P(D_A|C)$ as $(P(C|D_A)P(D_A))/P(C)$ and the same for $P(D_B|C)$ in Eq. (A.0.20), we obtain:

$$P(C|D_A,D_B) = \frac{P(D_A)P(D_B)}{P(D_A,D_B)}\times\frac{P(C|D_A)P(C|D_B)}{P(C)} = a\frac{P(C|D_A)P(C|D_B)}{P(C)}, \tag{A.0.21}$$

which defines a, that is a probability class-independent term, as it depends only on the data. As we have already measured these data, its value does not concern us in this context (we are not doing model comparison), thus we consider it as a normalisation constant which only ensures that $\sum_k P(C_k|D_A,D_B)=1$. The equation (A.0.21) is our final achievement regarding the combination of results coming from two independent estimators (where *independent* here means in the sense of equation (A.0.18)).

We can safely extend this result to N independent classifiers (estimators), each using independent information (data) D_n, combining them as:

$$P(C|D_1,...,D_N) = a\frac{\prod_{n=1}^{N}P(C|D_n)}{P(C)^{N-1}}\,. \tag{A.0.22}$$

This is the equation to use to combine multiple *conditionally independent* posterior probability distributions. Note that all of the classifiers (the usual posterior probability), $P(C|D_n)$, as well as the prior $P(C)$, are implicitly assumed to be conditioned on the same background information. Conversely, if the D_n data are not independent (conditioned on C), we need to know their joint probability distribution.

[2]On the contrary, they are generally not *unconditionally independent*, i.e $P(D_A,D_B)\neq P(D_A)\times P(D_B)$.

Appendix B

Covariance matrix

The *covariance* is a measure of the joint variability of two random variables. Thus, it generalizes the notion of variance to multiple dimensions. Specifically, the covariance matrix represents the variation of each variable with respect to the others (including itself, in which case it is the variance). It is a symmetric matrix and is positive semi-definite. Let us consider a set of n elements on which k variables X_i are detected. Hence every X_i is a vector with n elements indicated with $x_{h,i}$, where $i = 1, ..., n$. The covariance matrix has dimensions $k \times k$, and each element is defined as:

$$\sigma_{ij} = \frac{1}{n} \sum_{h=1}^{n} (x_{h,i} - \mu_i)(x_{h,j} - \mu_j) \,,$$

where μ_i is the mean of the X_i variable. Each diagonal element, σ_{ii}, is the variance of the variable X_i and it is therefore always a non-negative value. Each element σ_{ij} (with $i \neq j$) is the covariance between the variables X_i and X_j. The sign of a covariance element indicates the trend of the relationship between the variables. Specifically, when σ_{ij} is positive, if the i variable grows then also the j variable grows. While if σ_{ij} is negative, the opposite happens. If the variables are statistically independent, the covariance is 0. On the other hand, while the inverse implication is not necessarily verified.

The units of the covariance elements are those of X_i times those of X_J, and the overall matrix is often indicated with $cov(X_i, X_j)$. The *correlation* matrix is defined as the normalised version of the covariance matrix.

Appendix C

Bootstrap procedure

We provide here a brief explanation of the bootstrap sampling method. A more complete and detailed description can be found in Efron and Tibshirani (1993).

The bootstrap method is a statistical technique that exploits a dataset re-sampling with replacement. It is used to approximate the sample distribution of a statistics, without requiring any a-priori assumption on the distribution. Its main goal is to estimate the standard errors with random sampling methods. In fact, the bootstrap technique is often used as an alternative method when parametric inference is difficult to perform, or if it requires complicated mathematical procedures for the computation of standard errors.

The need to re-sample is the following. If possible, we would like to extract large, non-repeated, samples from a population in order to generate a sampling distribution for some statistics. Unfortunately, as often happens, the available sample is only one. Nevertheless, this single sample can be exploited as a fraction of the entire population, from which to repeatedly extract small samples with replacement. As well as saving us a lot of time, bootstrapped samples can reliably approximate the population parameters.

A *bootstrap sample* is a smaller sample that is *bootstrapped* from a larger sample. In this kind of re-sampling, sizable numbers of smaller samples of the same size are recurrently extracted, with replacement, from a single original sample (see figure C.1). After a large number of iterations, the bootstrap statistics are drawn up into a bootstrap distribution. The large numbers law asserts that if we sample again and again, our dataset should approximate the true population dataset.

We have to distinguish two kinds of bootstrap samples: the *empirical* bootstrap sample that is derived from observations, and the *parametric* one that comes from a parameterised distribution (e.g. a normal distribution). In practice, the bootstrap procedure is generally performed numerically following these three steps:

1. the dataset is re-sampled n time;
2. then a summary statistic (the so-called bootstrap statistic) is measured for each of the n samples;
3. finally, the standard error for these bootstrap statistics is assessed through the standard deviation of the bootstrap distribution.

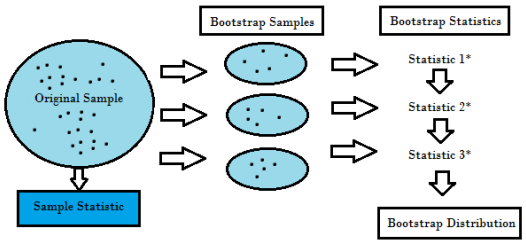

Figure C.1: Schematized procedure of the execution of a sampling useful to build a bootstrap sample.

In conclusion, starting from a sub-population of a large sample of data, the bootstrap method allows us to accurately estimate the errors associated to any statistics of sample.

Appendix D

The C++ code used for statistical inference

Here we present the specific C++ code used in this Thesis work to perform the Bayesian statistical analysis of our stacked shear profiles, exploiting the halo model (Eq. 5.4.8). The code imports several CBL classes to handle catalogues of extra-galactic sources, model all measured quantities and derive cosmological constraints. The C++ classes for the NFW truncated model used to describe the 1-halo term are the same ones used in Bellagamba et al. (2019), while the 2-halo model has been originally implemented for this Thesis work, making use of the CBL cosmological functions. We exploit principally the class *cbl::cosmology* that contains all the functions useful to the estimation of the various cosmological components of our fitting model. *cbl* is the global namespace of the cosmological libraries.

The CBL provide methods to perform the MCMC posterior sampling (see appendix A). The latter consists in sampling a target distribution using a correlated random walk: every step is extracted after a trial that depends only on the previous one (Markov process). The steps are collected in chains, that define the marginalised posterior probability of the model parameters (Marulli et al., 2016). Specifically, the CBL algorithms implements the Metropolis-Hastings algorithm (see appendix A) that consists in a single-particle sampling of the parameter space. At each step, t , the proposed parameter vector, θ', is extracted from the distribution $q(\vec{\theta'}|\vec{\theta}(t))$, centred on $\vec{\theta}(t)$ (Marulli et al., 2016). The C++ code, accompanied by explanatory comments, is reported below:

```
// import the CBL class used to perform all the cosmological calculations
#include "Cosmology.h"

// import the CBL classes to handle input data
#include "Data1D.h"

// import the CBL classes to perform Bayesian statistical inferences
#include "Posterior.h"

// import the class to construct the model for the 1-halo term
#include "nfwModel.h"

using namespace std;

string cbl::par::DirCosmo = DIRCOSMO, cbl::par::DirLoc = DIRL;

// function to evaluate the angular coordinates (radians) of lens density profile
double Theta (const double rad, const double z, const cbl::cosmology::Cosmology
    cosmology)
{
  return (rad/cosmology.D_A(z));
}

// vector containing the model
vector<double> model_function (const vector<double> x, const shared_ptr<void>
    modelInput, std::vector<double> &parameter)
{
  cbl::cosmology::Cosmology cosm = *static_pointer_cast<cbl::cosmology::Cosmology>(
```

```
    modelInput);
  cosm.set_OmegaM(parameter[0]);
  cosm.set_OmegaDE(parameter[1]);
  (void)modelInput;

  vector<double> model(x.size(), 0.);

  // define the cosmology adopted in the modelling
  cosmology my_cos(parameter[0], parameter[1], parameter[2]);
  nfwModel my_nfw(&my_cos, parameter[3], pow(10., parameter[4]), parameter[5], 3.,
    parameter[6], parameter[7], false);

  // define the density profile model NFW truncated

  // nfw parameters
  // 1) cosmology
  // 2) redshift
  // 3) M_200( log(M200)*10^14)
  // 4) concentration=r_200/r_s
  // 5) truncation factor (r_trunc = trunc_fact*r200)
  // 6) fraction of misaligned haloes
  // 7) rms of misplacement
  // 8) if true, reads 2-halo term from tables; if False use our 2-halo model

  // functions useful to build the 2-halo density profile
  double delta_bkg = cosm.Delta_vir(200, parameter[3]);
  double Bias = cosm.bias_halo(pow(10., parameter[4]), parameter[3], "Tinker", "
    EisensteinHu","test","Linear",delta_bkg);
  double Rho_m = cosm.rho_m(parameter[3]); // [Msun*Mpc^-3(*h^2)]
  double Dl = cosm.D_A(parameter[3]); // Mpc/h

  for (size_t i=0; i<x.size(); ++i) {
    double theta = Theta(x[i], parameter[3], cosm);

    // integrand function of the integral defining the 2-halo density profile: it's
     a Lambda-function
    std::function<double(double)> Func = [&] (const double l)
      {
      double Dl = cosm.D_A(parameter[3]); // Mpc/h
      double X-ray = theta*l;
      double kl = l/((1+ parameter[3])*Dl); // h/Mpc
      double P_k = cosm.Pk(kl, "EisensteinHu", false, parameter[3]); // Mpc^3 h^-3
      double J2 = cbl::j2(X);
      return (l*J2)*(P_k) ;
      };

      // integral of the 2-halo term
      const double integrand = cbl::gsl::GSL_integrate_qag(Func,1.0e-6,1000000.);
      const double twohalo= 1.e-12*((Bias*Rho_m)/(2.*cbl::par::pi))*(pow(Dl,-2)*pow
    (1+parameter[3],-3))*integrand;

      model[i] = my_nfw.get_deltasigma(x[i]) + twohalo; // total 1halo+2halo model
  }

  parameter[8] = pow(10.,parameter[4])*1.e-14; // parameter[8] is a derived
    parameter

  return model;
}


// ============================================================================



int main () {

  try {

    // --- set the input/output file/directories ---

    const string dir_output = cbl::par::DirLoc+"../output/CHAIN_2-HALO/";
```

```
92      const string file_output_start = "1+2h_model__z=0.1-0.3__A=2.05-2.75.dat";
93
94
95      // --- construct the dataset by reading the stacked_shear2HALO input file ---
96
97      const string file_input = cbl::par::DirLoc+"../input/risultatishear-2HALO/
        stacked_shear_AMPLITUDE_2-HALO__35Mpc--30bin.dat";
98      ifstream fin(file_input.c_str()); cbl::checkIO(fin, file_input);
99
100     string line;
101     int lines_to_skip = 24;
102     for (int i=0; i<lines_to_skip; ++i)
103       getline(fin, line);
104
105     vector<double> rad, DSigma, error, z_two;
106
107     while (getline(fin, line)) {
108       stringstream ss(line);
109       vector<double> num; double NUM;
110       while (ss>>NUM) num.emplace_back(NUM);
111
112       if (num[0]==0 && num[1]==3) {
113       if (num[6]>0.2) { // exclude the first 3 radial bins
114         z_two.emplace_back(num[2]);
115         rad.emplace_back(num[6]);
116         DSigma.emplace_back(num[8]);
117         error.emplace_back((num[14]>0) ? num[14] : num[8]);
118       }
119       }
120     }
121     fin.clear(); fin.close();
122
123
124     // --- read the input dataset used to set the priors ---
125
126     const string file_inputs = cbl::par::DirLoc+"../output/CHAIN/chains_testI_0
        .1-0.3_AMPL__2.05-2.75_parameters.dat";
127     ifstream FIN(file_inputs.c_str()); cbl::checkIO(FIN, file_inputs);
128
129     string Line;
130     int Lines_to_skip = 15;
131     for (int k=0; k<Lines_to_skip; ++k)
132       getline(FIN, Line);
133
134     vector <double> MEAN, MEDIAN;
135
136     while (getline(FIN, Line)) {
137       stringstream SS(Line);
138       vector<double> numb;    double NUMB;
139
140       while (SS>>NUMB) numb.emplace_back(NUMB);
141
142       if (numb.size()>0)
143       MEAN.emplace_back(numb[0]);
144       else cbl::Error("ERROR!");
145
146       if (numb.size()>2)
147       MEDIAN.emplace_back(numb[2]);
148       else cbl::Error("ERROR!");
149     }
150
151     FIN.clear(); FIN.close();
152
153
154     // --- construct the dataset ---
155
156     const cbl::data::Data1D data(rad, DSigma, error);
157     auto ptr_data = make_shared<cbl::data::Data1D>(data);
158
159
160     // --- set the model to construct the likelihood ---
161
162     // number of model parameters
```

```
163     const int nparameters = 9;
164
165     // names of the model parameters
166     const vector<string> parNames = {"OmegaM", "OmegaL", "h", "redshift", "LogM200"
        , "conc", "f_mis", "sigma_mis", "M200 [e+14 M_sun]"};
167
168     // starting values parameters
169     const double OmegaM = 0.3;
170     const double OmegaL = 1.-OmegaM;
171     const double h = 1.;
172
173     const double redshift = MEAN[0];
174     const double LogM200 = MEDIAN[1];
175     const double conc = MEDIAN[2];
176     const double f_mis = MEDIAN[3];
177     const double sigma_mis = MEDIAN[4];
178
179     // vector containing the 9 model parameters
180     vector<cbl::statistics::ParameterType> parType(nparameters-1, cbl::statistics::
        ParameterType::_Base_);
181     parType.emplace_back(cbl::statistics::ParameterType::_Derived_);
182
183     // set the stuff used to construct the model: here an object of class cosmology
184     const double Omegab = 0.046;
185     const double Omeganu = 0.;
186     const double massless_neutrinos = 3.04;
187     const int massive_neutrinos = 0;
188     const double Omega_radiation = 0.;
189     const double scalar_amp = 1.92e-9;
190     const double scalar_pivot = 0.05;
191     const double n_s = 1;
192     const double wa = 0.;
193     const double w0 = -1.;
194     cbl::cosmology::Cosmology cosmology {OmegaM, Omegab, Omeganu,
        massless_neutrinos, massive_neutrinos, OmegaL, Omega_radiation, h, scalar_amp,
        scalar_pivot, n_s, wa, w0};
195
196     auto ptr_modelInput = make_shared<cbl::cosmology::Cosmology>(cosmology);
197
198     // construct the model
199     const cbl::statistics::Model1D model(&model_function, nparameters, parType,
        parNames, ptr_modelInput);
200     auto ptr_model = make_shared<cbl::statistics::Model1D>(model);
201
202
203     // --- construct the likelihood ---
204
205     // define a Gaussian likelihood
206     cbl::statistics::Likelihood likelihood(ptr_data, ptr_model, cbl::statistics::
        LikelihoodType::_Gaussian_Error_);
207
208     // write the model used to construct the likelihood, at the starting parameter
        values
209     const vector<double> start = {OmegaM, OmegaL, h, redshift, LogM200, conc, f_mis
        , sigma_mis};
210
211     likelihood.write_model(dir_output, file_output_start, start);
212
213     //set the parameter space where priors can move
214     double minomegaM = 0.06, maxomegaM = 0.6;
215
216
217     // --- construct the priors ---
218
219     auto priorOmegaM = make_shared<cbl::statistics::PriorDistribution>(cbl::
        statistics::PriorDistribution(cbl::glob::DistributionType::_Uniform_, minomegaM
        , maxomegaM, 351));
220     auto priorOmegaL = make_shared<cbl::statistics::PriorDistribution>(cbl::
        statistics::PriorDistribution(cbl::glob::DistributionType::_Constant_, OmegaL))
        ;
221     auto priorH = make_shared<cbl::statistics::PriorDistribution>(cbl::statistics::
        PriorDistribution(cbl::glob::DistributionType::_Constant_, h));
222     auto prior_z = make_shared<cbl::statistics::PriorDistribution>(cbl::statistics
```

```
    ::PriorDistribution(cbl::glob::DistributionType::_Constant_, redshift));
223     auto priorLogM200 = make_shared<cbl::statistics::PriorDistribution>(cbl::
        statistics::PriorDistribution(cbl::glob::DistributionType::_Constant_, LogM200)
        );
224     auto priorConc = make_shared<cbl::statistics::PriorDistribution>(cbl::
        statistics::PriorDistribution(cbl::glob::DistributionType::_Constant_, conc));
225     auto priorFMis = make_shared<cbl::statistics::PriorDistribution>(cbl::
        statistics::PriorDistribution(cbl::glob::DistributionType::_Constant_, f_mis));
226     auto priorSigmaMis = make_shared<cbl::statistics::PriorDistribution>(cbl::
        statistics::PriorDistribution(cbl::glob::DistributionType::_Constant_,
        sigma_mis));
227
228     const vector<shared_ptr<cbl::statistics::PriorDistribution>>
        prior_distributions = {priorOmegaM, priorOmegaL, priorH, prior_z, priorLogM200,
         priorConc, priorFMis, priorSigmaMis };
229
230
231     // --- construct and sample the posterior ---
232
233     cbl::statistics::Posterior posterior(prior_distributions, likelihood, 796);
234
235     // sample the posterior
236     const int nwalkers = 2;
237     const int chain_size = 400;
238     const vector<double> POSTERIOR_start = { OmegaM };
239     posterior.initialize_chains(chain_size, nwalkers, 1.e-5, POSTERIOR_start);
240     posterior.sample_stretch_move(2, true);
241
242     // show the results on screen
243     const int burn_in = 0;
244     const int thin = 1;
245     posterior.show_results(burn_in, thin);
246
247     // write the chain ouput
248     posterior.write_results(cbl::par::DirLoc+"../output/CHAIN_2-HALO/", "
        chains_linear_relation__30bin_z=0.1-0.3__A=0-1", burn_in, thin);
249
250     // write the best-fit model
251     posterior.write_model_from_chain(cbl::par::DirLoc+"../output/FIT_2-HALO/", "
        model_from_chain_-30bin_z=0.3-0.45__A=1.65-2.3", {}, {}, burn_in, thin);
252
253   }
254
255   catch(cbl::glob::Exception &exc) { cerr << exc.what() << endl; exit(1); }
256   return 0;
257 }
```

Bibliography

Abell, G. O. (1958). The Distribution of Rich Clusters of Galaxies. *apjs*, 3:211.

Abell, G. O., Corwin, Jr., H. G., and Olowin, R. P. (1989). A catalog of rich clusters of galaxies. *apjs*, 70:1–138.

Abell, G. O. and Seligman, C. E. (1965). The Distribution of Clusters of Galaxies. *aj*, 70:317.

Aghanim, N. et al. (2018). Planck 2018 results. VI. Cosmological parameters.

Allen, S. W., Evrard, A. E., and Mantz, A. B. (2011). Cosmological Parameters from Observations of Galaxy Clusters. *Annual Review of Astronomy and Astrophysics*, 49:409–470.

Andreon, S. and Hurn, M. A. (2010). The scaling relation between richness and mass of galaxy clusters: a Bayesian approach. *mnras*, 404:1922–1937.

Bailer-Jones, C. and Smith, K. (2011). Combining probabilities. GAIA-C8-TN-MPIA-CBJ-053.

Baltz, E. A., Marshall, P., and Oguri, M. (2009). Analytic models of plausible gravitational lens potentials. *Journal of Cosmology and Astroparticle Physics*, 2009(01):015.

Bartelmann, M. (1996). Arcs from a universal dark-matter halo profile. *aap*, 313:697–702.

Bartelmann, M. and Schneider, P. (2001). Weak gravitational lensing. *physrep*, 340:291–472.

Bellagamba, F., Maturi, M., Hamana, T., Meneghetti, M., Miyazaki, S., and Moscardini, L. (2011). Optimal filtering of optical and weak lensing data to search for galaxy clusters: application to the COSMOS field. *mnras*, 413:1145–1157.

Bellagamba, F., Meneghetti, M., Moscardini, L., and Bolzonella, M. (2012). Accuracy of photometric redshifts for future weak lensing surveys from space. *Monthly Notices of the Royal Astronomical Society*, 422(1):553–562.

Bellagamba, F., Roncarelli, M., Maturi, M., and Moscardini, L. (2018). AMICO: optimized detection of galaxy clusters in photometric surveys. *Mon. Not. Roy. Astron. Soc.*, 473(4):5221–5236.

Bellagamba, F., Sereno, M., Roncarelli, M., Maturi, M., Radovich, M., Bardelli, S., Puddu, E., Moscardini, L., Getman, F., Hildebrandt, H., and Napolitano, N. (2019). AMICO galaxy clusters in KiDS-DR3: weak lensing mass calibration. *Monthly Notices of the Royal Astronomical Society*, 484(2):1598–1615.

Benìtez, N. (2000). Bayesian Photometric Redshift Estimation. *apj*, 536:571–583.

Broadhurst, T., Takada, M., Umetsu, K., Kong, X., Arimoto, N., Chiba, M., and Futamase, T. (2005). The surprisingly steep mass profile of a1689, from a lensing analysis of subaru images. *The Astrophysical Journal Letters*, 619(2):L143.

Bryan, G. L. and Norman, M. L. (1998). Statistical Properties of X-Ray Clusters: Analytic and Numerical Comparisons. *apj*, 495:80–99.

Cacciato, M., van den Bosch, F. C., More, S., Mo, H., and Yang, X. (2013). Cosmological constraints from a combination of galaxy clustering and lensing – III. Application to SDSS data. *Monthly Notices of the Royal Astronomical Society*, 430(2):767–786.

Capaccioli, M. and Schipani, P. (2011). The VLT Survey Telescope Opens to the Sky: History of a Commissioning. *The Messenger*, 146:2–7.

Carroll, S. M., Press, W. H., and Turner, E. L. (1992). The cosmological constant. , 30:499–542.

Cavaliere, A. and Fusco-Femiano, R. (1976). X-rays from hot plasma in clusters of galaxies. *aap*, 49:137–144.

Clerc, N., Adami, C., Lieu, M., Maughan, B., Pacaud, F., Pierre, M., Sadibekova, T., Smith, G. P., Valageas, P., Altieri, B., Benoist, C., Maurogordato, S., and Willis, J. P. (2014). The XMM-LSS survey: the Class 1 cluster sample over the extended 11 deg^2 and its spatial distribution. *mnras*, 444:2723–2753.

Coe, D., Benítez, N., Sánchez, S. F., Jee, M., Bouwens, R., and Ford, H. (2006). Galaxies in the Hubble Ultra Deep Field. I. Detection, Multiband Photometry, Photometric Redshifts, and Morphology. *aj*, 132:926–959.

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

Covone, G., Sereno, M., Kilbinger, M., and Cardone, V. F. (2014). Measurement of the Halo Bias from Stacked Shear Profiles of Galaxy Clusters. *apjl*, 784:L25.

de Jong, J. T. A., Verdoes Kleijn, G. A., Boxhoorn, D. R., Buddelmeijer, H., Capaccioli, M., Getman, F., Grado, A., Helmich, E., Huang, Z., Irisarri, N., Kuijken, K., La Barbera, F., McFarland, J. P., Napolitano, N. R., Radovich, M., Sikkema, G., Valentijn, E. A., Begeman, K. G., Brescia, M., Cavuoti, S., Choi, A., Cordes, O.-M., Covone, G., Dall'Ora, M., Hildebrandt, H., Longo, G., Nakajima, R., Paolillo, M., Puddu, E., Rifatto, A., Tortora, C., van Uitert, E., Buddendiek, A., Harnois-Déraps, J., Erben, T., Eriksen, M. B., Heymans, C., Hoekstra, H., Joachimi, B., Kitching, T. D., Klaes, D., Koopmans, L. V. E., Köhlinger, F., Roy, N., Sifón, C., Schneider, P., Sutherland, W. J., Viola, M., and Vriend, W.-J. (2015). The first and second data releases of the Kilo-Degree Survey. *aap*, 582:A62.

de Jong, J. T. A., Verdoes Kleijn, G. A., Erben, T., Hildebrandt, H., Kuijken, K., Sikkema, G., Brescia, M., Bilicki, M., Napolitano, N. R., Amaro, V., Begeman, K. G., Boxhoorn, D. R., Buddelmeijer, H., Cavuoti, S., Getman, F., Grado, A., Helmich, E., Huang, Z., Irisarri, N., La Barbera, F., Longo, G., McFarland, J. P., Nakajima, R., Paolillo, M., Puddu, E., Radovich, M., Rifatto, A., Tortora, C., Valentijn, E. A., Vellucci, C., Vriend, W.-J., Amon, A., Blake, C., Choi, A., Conti, I. F., Gwyn, S. D. J., Herbonnet, R., Heymans, C., Hoekstra, H., Klaes, D., Merten, J., Miller, L., Schneider, P., and Viola, M. (2017). The third data release of the Kilo-Degree Survey and associated data products. *aap*, 604:A134.

de Jong, J. T. A., Verdoes Kleijn, G. A., Kuijken, K. H., and Valentijn, E. A. (2013). The Kilo-Degree Survey. *Experimental Astronomy*, 35:25–44.

Efron, B. and Tibshirani, R. J. (1993). *An Introduction to the Bootstrap*. Number 57 in Monographs on Statistics and Applied Probability. Chapman Hall/CRC, Boca Raton, Florida, USA.

Efstathiou, G., Sutherland, W. J., and Maddox, S. J. (1990). The cosmological constant and cold dark matter. *nat*, 348:705–707.

Einstein, A. (1915). Zur allgemeinen Relativitätstheorie. *Sitzungsberichte der Königlich Preußischen Akademie der Wissenschaften (Berlin), Seite 778-786.*

Eisenstein, D. J. and Hu, W. (1999). Power spectra for cold dark matter and its variants. *The Astrophysical Journal*, 511(1):5–15.

Fenech Conti, I., Herbonnet, R., Hoekstra, H., Merten, J., Miller, L., and Viola, M. (2017). Calibration of weak-lensing shear in the Kilo-Degree Survey. *mnras*, 467:1627–1651.

Foreman-Mackey, D., Hogg, D. W., Lang, D., and Goodman, J. (2013). emcee: The MCMC Hammer. *Publications of the Astronomical Society of the Pacific*, 125:306.

Friedmann, A. (1922). Über die Krümmung des Raumes. *Zeitschrift fur Physik*, 10:377–386.

George, M. R., Leauthaud, A., Bundy, K., Finoguenov, A., Ma, C.-P., Rykoff, E. S., Tinker, J. L., Wechsler, R. H., Massey, R., and Mei, S. (2012). Galaxies in X-Ray Groups. II. A Weak Lensing Study of Halo Centering. *apj*, 757:2.

Giodini, S., Lovisari, L., Pointecouteau, E., Ettori, S., Reiprich, T. H., and Hoekstra, H. (2013). Scaling Relations for Galaxy Clusters: Properties and Evolution. *ssr*, 177:247–282.

Gruen, D., Seitz, S., Brimioulle, F., Kosyra, R., Koppenhoefer, J., Lee, C.-H., Bender, R., Riffeser, A., Eichner, T., Weidinger, T., and Bierschenk, M. (2014). Weak lensing analysis of SZ-selected clusters of galaxies from the SPT and Planck surveys. *mnras*, 442:1507–1544.

Guth, A. H. (1981). Inflationary universe: A possible solution to the horizon and flatness problems. *prd*, 23:347–356.

Hartlap, J., Simon, P., and Schneider, P. (2007). Why your model parameter confidences might be too optimistic. Unbiased estimation of the inverse covariance matrix. *aap*, 464:399–404.

Hastings, W. K. (1970). Monte carlo sampling methods using markov chains and their applications. *Biometrika*, 57(1):97–109.

Hennig, C., Mohr, J. J., Zenteno, A., Desai, S., Dietrich, J. P., Bocquet, S., Strazzullo, V., Saro, A., Abbott, T. M. C., Abdalla, F. B., Bayliss, M., Benoit-Lévy, A., Bernstein, R. A., Bertin, E., Brooks, D., Capasso, R., Capozzi, D., Carnero, A., Carrasco Kind, M., Carretero, J., Chiu, I., D'Andrea, C. B., daCosta, L. N., Diehl, H. T., Doel, P., Eifler, T. F., Evrard, A. E., Fausti-Neto, A., Fosalba, P., Frieman, J., Gangkofner, C., Gonzalez, A., Gruen, D., Gruendl, R. A., Gupta, N., Gutierrez, G., Honscheid, K., Hlavacek-Larrondo, J., James, D. J., Kuehn, K., Kuropatkin, N., Lahav, O., March, M., Marshall, J. L., Martini, P., McDonald, M., Melchior, P., Miller, C. J., Miquel, R., Neilsen, E., Nord, B., Ogando, R., Plazas, A. A., Reichardt, C., Romer, A. K., Rozo, E., Rykoff, E. S., Sanchez, E., Santiago, B., Schubnell, M., Sevilla-Noarbe, I., Smith, R. C., Soares-Santos, M., Sobreira, F., Stalder, B., Stanford, S. A., Suchyta, E., Swanson, M. E. C., Tarle, G., Thomas, D., Vikram, V., Walker, A. R., and Zhang, Y. (2017). Galaxy populations in massive galaxy clusters to z=1.1: colour distribution, concentration, halo occupation number and red sequence fraction. *Monthly Notices of the Royal Astronomical Society*, 467(4):4015–4035.

Henry, J. P., Evrard, A. E., Hoekstra, H., Babul, A., and Mahdavi, A. (2009). The X-Ray Cluster Normalization of the Matter Power Spectrum. , 691:1307–1321.

Heymans, C., Van Waerbeke, L., Miller, L., Erben, T., Hildebrandt, H., Hoekstra, H., Kitching, T. D., Mellier, Y., Simon, P., Bonnett, C., Coupon, J., Fu, L., Harnois Déraps, J., Hudson, M. J., Kilbinger, M., Kuijken, K., Rowe, B., Schrabback, T., Semboloni, E., van Uitert, E., Vafaei, S., and Velander, M. (2012). CFHTLenS: the Canada-France-Hawaii Telescope Lensing Survey. *mnras*, 427:146–166.

Hildebrandt, H., Erben, T., Kuijken, K., van Waerbeke, L., Heymans, C., Coupon, J., Benjamin, J., Bonnett, C., Fu, L., Hoekstra, H., Kitching, T. D., Mellier, Y., Miller, L., Velander, M., Hudson, M. J., Rowe, B. T. P., Schrabback, T., Semboloni, E., and Benítez, N. (2012). CFHTLenS: improving the quality of photometric redshifts with precision photometry. *mnras*, 421:2355–2367.

Hildebrandt, H., Viola, M., Heymans, C., Joudaki, S., Kuijken, K., Blake, C., Erben, T., Joachimi, B., Klaes, D., Miller, L., Morrison, C. B., Nakajima, R., Verdoes Kleijn, G., Amon, A., Choi, A., Covone, G., de Jong, J. T. A., Dvornik, A., Fenech Conti, I., Grado, A., Harnois-Déraps, J., Herbonnet, R., Hoekstra, H., Köhlinger, F., McFarland, J., Mead, A., Merten, J., Napolitano, N., Peacock, J. A., Radovich, M., Schneider, P., Simon, P., Valentijn, E. A., van den Busch, J. L., van Uitert, E., and Van Waerbeke, L. (2017). KiDS-450: cosmological parameter constraints from tomographic weak gravitational lensing. *mnras*, 465:1454–1498.

Hoekstra, H., Bartelmann, M., Dahle, H., Israel, H., Limousin, M., and Meneghetti, M. (2013). Masses of Galaxy Clusters from Gravitational Lensing. *ssr*, 177:75–118.

Hoekstra, H., Donahue, M., Conselice, C. J., McNamara, B. R., and Voit, G. M. (2011). The Mass-L_x Relation for Moderate Luminosity X-ray Clusters. *apj*, 726:48.

Hubble, E. (1929). A Relation between Distance and Radial Velocity among Extra-Galactic Nebulae. *Proceedings of the National Academy of Science*, 15:168–173.

Ilbert, O., Capak, P., Salvato, M., Aussel, H., McCracken, H. J., Sanders, D. B., Scoville, N., Kartaltepe, J., Arnouts, S., Le Floc'h, E., Mobasher, B., Taniguchi, Y., Lamareille, F., Leauthaud, A., Sasaki, S., Thompson, D., Zamojski, M., Zamorani, G., Bardelli, S., Bolzonella, M., Bongiorno, A., Brusa, M., Caputi, K. I., Carollo, C. M., Contini, T., Cook, R., Coppa, G., Cucciati, O., de la Torre, S., de Ravel, L., Franzetti, P., Garilli, B., Hasinger, G., Iovino, A., Kampczyk, P., Kneib, J.-P., Knobel, C., Kovac, K., Le Borgne, J. F., Le Brun, V., Le Fèvre, O., Lilly, S., Looper, D., Maier, C., Mainieri, V., Mellier, Y., Mignoli, M., Murayama, T., Pellò, R., Peng, Y., Pérez-Montero, E., Renzini, A., Ricciardelli, E., Schiminovich, D., Scodeggio, M., Shioya, Y., Silverman, J., Surace, J., Tanaka, M., Tasca, L., Tresse, L., Vergani, D., and Zucca, E. (2009). Cosmos Photometric Redshifts with 30-Bands for 2-deg^2. *apj*, 690:1236–1249.

Johnston, D. E., Sheldon, E. S., Tasitsiomi, A., Frieman, J. A., Wechsler, R. H., and McKay, T. A. (2007). Cross-Correlation Lensing: Determining Galaxy and Cluster Mass Profiles from Statistical Weak-Lensing Measurements. *apj*, 656:27–41.

Johnston, D. E., Sheldon, E. S., Wechsler, R. H., Rozo, E., Koester, B. P., Frieman, J. A., McKay, T. A., Evrard, A. E., Becker, M. R., and Annis, J. (2007). Cross-correlation Weak Lensing of SDSS galaxy Clusters II: Cluster Density Profiles and the Mass–Richness Relation.

Jungman, G., Kamionkowski, M., and Griest, K. (1996). Supersymmetric dark matter. *physrep*, 267:195–373.

Keeton, C. R. (2001). A catalog of mass models for gravitational lensing.

King, I. (1962). The structure of star clusters. I. an empirical density law. *aj*, 67:471.

King, I. R. (1966). The structure of star clusters. III. Some simple dynamical models. *aj*, 71:64.

Kuijken, K. (2011). OmegaCAM: ESO's Newest Imager. *The Messenger*, 146:8–11.

Kuijken, K., Heymans, C., Hildebrandt, H., Nakajima, R., Erben, T., de Jong, J. T. A., Viola, M., Choi, A., Hoekstra, H., Miller, L., van Uitert, E., Amon, A., Blake, C., Brouwer, M., Buddendiek, A., Conti, I. F., Eriksen, M., Grado, A., Harnois-Déraps, J., Helmich, E., Herbonnet, R., Irisarri, N., Kitching, T., Klaes, D., La Barbera, F., Napolitano, N., Radovich, M., Schneider, P., Sifón, C., Sikkema, G., Simon, P., Tudorica, A., Valentijn, E., Verdoes Kleijn, G., and van Waerbeke, L. (2015). Gravitational lensing analysis of the Kilo-Degree Survey. *mnras*, 454:3500–3532.

Laureijs, R., Amiaux, J., Arduini, S., Augueres, J. L., Brinchmann, J., Cole, R., Cropper, M., Dabin, C., Duvet, L., Ealet, A., Garilli, B., Gondoin, P., Guzzo, L., Hoar, J., Hoekstra, H., Holmes, R., Kitching, T., Maciaszek, T., Mellier, Y., Pasian, F., Percival, W., Rhodes, J., Saavedra Criado, G., Sauvage, M., Scaramella, R., Valenziano, L., Warren, S., Bender, R., Castander, F., Cimatti, A., Le Fèvre, O., Kurki-Suonio, H., Levi, M., Lilje, P., Meylan, G., Nichol, R., Pedersen, K., Popa, V., Rebolo Lopez, R., Rix, H. W., Rottgering, H., Zeilinger, W., Grupp, F., Hudelot, P., Massey, R., Meneghetti, M., Miller, L., Paltani, S., Paulin-Henriksson, S., Pires, S., Saxton, C., Schrabback, T., Seidel, G., Walsh, J., Aghanim, N., Amendola, L., Bartlett, J., Baccigalupi, C., Beaulieu, J. P., Benabed, K., Cuby, J. G., Elbaz, D., Fosalba, P., Gavazzi, G., Helmi, A., Hook, I., Irwin, M., Kneib, J. P., Kunz, M., Mannucci, F., Moscardini, L., Tao, C., Teyssier, R., Weller, J., Zamorani, G., Zapatero Osorio, M. R., Boulade, O., Foumond, J. J., Di Giorgio, A., Guttridge, P., James, A., Kemp, M., Martignac, J., Spencer, A., Walton, D., Blümchen, T., Bonoli, C., Bortoletto, F., Cerna, C., Corcione, L., Fabron, C., Jahnke, K., Ligori, S., Madrid, F., Martin, L., Morgante, G., Pamplona, T., Prieto, E., Riva, M., Toledo, R., Trifoglio, M., Zerbi, F., Abdalla, F., Douspis, M., Grenet, C., Borgani, S., Bouwens, R., Courbin, F., Delouis, J. M., Dubath, P., Fontana, A., Frailis, M., Grazian, A., Koppenhöfer, J., Mansutti, O., Melchior, M., Mignoli, M., Mohr, J., Neissner, C., Noddle, K., Poncet, M., Scodeggio, M., Serrano, S., Shane, N., Starck, J. L., Surace, C., Taylor, A., Verdoes-Kleijn, G., Vuerli, C., Williams, O. R., Zacchei, A., Altieri, B., Escudero Sanz, I., Kohley, R., Oosterbroek, T., Astier, P., Bacon, D., Bardelli, S., Baugh, C., Bellagamba, F., Benoist, C., Bianchi, D., Biviano, A., Branchini, E., Carbone, C., Cardone, V., Clements, D., Colombi, S., Conselice, C.,

Cresci, G., Deacon, N., Dunlop, J., Fedeli, C., Fontanot, F., Franzetti, P., Giocoli, C., Garcia-Bellido, J., Gow, J., Heavens, A., Hewett, P., Heymans, C., Holland, A., Huang, Z., Ilbert, O., Joachimi, B., Jennins, E., Kerins, E., Kiessling, A., Kirk, D., Kotak, R., Krause, O., Lahav, O., van Leeuwen, F., Lesgourgues, J., Lombardi, M., Magliocchetti, M., Maguire, K., Majerotto, E., Maoli, R., Marulli, F., Maurogordato, S., McCracken, H., McLure, R., Melchiorri, A., Merson, A., Moresco, M., Nonino, M., Norberg, P., Peacock, J., Pello, R., Penny, M., Pettorino, V., Di Porto, C., Pozzetti, L., Quercellini, C., Radovich, M., Rassat, A., Roche, N., Ronayette, S., Rossetti, E., Sartoris, B., Schneider, P., Semboloni, E., Serjeant, S., Simpson, F., Skordis, C., Smadja, G., Smartt, S., Spano, P., Spiro, S., Sullivan, M., Tilquin, A., Trotta, R., Verde, L., Wang, Y., Williger, G., Zhao, G., Zoubian, J., and Zucca, E. (2011). Euclid Definition Study Report. *arXiv e-prints*, page arXiv:1110.3193.

Liske, J., Baldry, I. K., Driver, S. P., Tuffs, R. J., Alpaslan, M., Andrae, E., Brough, S., Cluver, M. E., Grootes, M. W., Gunawardhana, M. L. P., Kelvin, L. S., Loveday, J., Robotham, A. S. G., Taylor, E. N., Bamford, S. P., Bland- Hawthorn, J., Brown, M. J. I., Drinkwater, M. J., Hopkins, A. M., Meyer, M. J., Norberg, P., Peacock, J. A., Agius, N. K., Andrews, S. K., Bauer, A. E., Ching, J. H. Y., Colless, M., Conselice, C. J., Croom, S. M., Davies, L. J. M., De Propris, R., Dunne, L., Eardley, E. M., Ellis, S., Foster, C., Frenk, C. S., Häußler, B., Holwerda, B. W., Howlett, C., Ibarra, H., Jarvis, M. J., Jones, D. H., Kafle, P. R., Lacey, C. G., Lange, R., Lara-López, M. A., López-Sánchez, Á. R., Maddox, S., Madore, B. F., McNaught-Roberts, T., Moffett, A. J., Nichol, R. C., Owers, M. S., Palamara, D., Penny, S. J., Phillipps, S., Pimbblet, K. A., Popescu, C. C., Prescott, M., Proctor, R., Sadler, E. M., Sansom, A. E., Seibert, M., Sharp, R., Sutherland, W., Vá zquez-Mata, J. A., van Kampen, E., Wilkins, S. M., Williams, R., and Wright, A. H. (2015). Galaxy And Mass Assembly (GAMA): end of survey report and data release 2. *mnras*, 452:2087–2126.

Mandelbaum, R., Hirata, C. M., Broderick, T., Seljak, U., and Brinkmann, J. (2006a). Ellipticity of dark matter halos with galaxy-galaxy weak lensing. *Mon. Not. Roy. Astron. Soc.*, 370:1008–1024.

Mandelbaum, R., Hirata, C. M., Ishak, M., Seljak, U., and Brinkmann, J. (2006b). Detection of large-scale intrinsic ellipticity—density correlation from the sloan digital sky survey and implications for weak lensing surveys. *Monthly Notices of the Royal Astronomical Society*, 367(2):611–626.

Mandelbaum, R., Reyes, R., Slosar, A., Baldauf, T., Seljak, U., Smith, R. E., Hirata, C. M., and Nakajima, R. (2013). Cosmological parameter constraints from galaxy–galaxy lensing and galaxy clustering with the SDSS DR7. *Monthly Notices of the Royal Astronomical Society*, 432(2):1544–1575.

Mandelbaum, R., Seljak, U., and Hirata, C. M. (2008). Halo mass - concentration relation from weak lensing. *JCAP*, 0808:006.

Mandelbaum, R., Slosar, A., Baldauf, T., Seljak, U., Hirata, C. M., Nakajima, R., Reyes, R., and Smith, R. E. (2013). Cosmological parameter constraints from galaxy-galaxy lensing and galaxy clustering with the SDSS DR7. *mnras*, 432:1544–1575.

Marulli, F., Veropalumbo, A., and Moresco, M. (2016). CosmoBolognaLib: C++ libraries for cosmological calculations. *Astronomy and Computing*, 14:35–42.

Maturi, M., Bellagamba, F., Radovich, M., Roncarelli, M., Sereno, M., Moscardini, L., Bardelli, S., and Puddu, E. (2018). AMICO galaxy clusters in KiDS-DR3: sample properties and selection function.

Maturi, M., Meneghetti, M., Bartelmann, M., Dolag, K., and Moscardini, L. (2005). An optimal filter for the detection of galaxy clusters through weak lensing. *aap*, 442:851–860.

Maturi, M., Schirmer, M., Meneghetti, M., Bartelmann, M., and Moscardini, L. (2007). Searching dark-matter halos in the GaBoDS survey. *aap*, 462:473–479.

Maughan, B. J. (2007). The Lx-Yx Relation: Using Galaxy Cluster X-Ray Luminosity as a Robust, Low Scatter Mass Proxy. *Astrophys. J.*, 668:772.

Medezinski, E., Battaglia, N., Umetsu, K., Oguri, M., Miyatake, H., Nishizawa, A. J., Sifón, C., Spergel, D. N., Chiu, I. N., Lin, Y.-T., Bahcall, N., and Komiyama, Y. (2018). Planck Sunyaev-Zel'dovich cluster mass calibration using Hyper Suprime-Cam weak lensing. *Publications of the Astronomical Society of Japan*, 70:S28.

Medezinski, E., Broadhurst, T., Umetsu, K., Coe, D., Ben'itez, N., Ford, H., Rephaeli, Y., Arimoto, N., and Kong, X. (2007). Using Weak-Lensing Dilution to Improve Measurements of the Luminous and Dark Matter in A1689. *apj*, 663:717–733.

Medezinski, E., Broadhurst, T., Umetsu, K., Oguri, M., Rephaeli, Y., and Ben'itez, N. (2010). Detailed cluster mass and light profiles of A1703, A370 and RXJ1347-11 from deep Subaru imaging. *mnras*, 405:257–273.

Melchior, P., Gruen, D., McClintock, T., Varga, T. N., Sheldon, E., Rozo, E., Amara, A., Becker, M. R., Benson, B. A., Bermeo, A., Bridle, S. L., Clampitt, J., Dietrich, J. P., Hartley, W. G., Hollowood, D., Jain, B., Jarvis, M., Jeltema, T., Kacprzak, T., MacCrann, N., Rykoff, E. S., Saro, A., Suchyta, E., Troxel, M. A., Zuntz, J., Bonnett, C., Plazas, A. A., Abbott, T. M. C., Abdalla, F. B., Annis, J., Benoit-Lévy, A., Bernstein, G. M., Bertin, E., Brooks, D., Buckley-Geer, E., Carnero Rosell, A., Carrasco Kind, M., Carretero, J., Cunha, C. E., D'Andrea, C. B., da Costa, L. N., Desai, S., Eifler, T. F., Flaugher, B., Fosalba, P., García-Bellido, J., Gaztanaga, E., Gerdes, D. W., Gruendl, R. A., Gschwend, J., Gutierrez, G., Honscheid, K., James, D. J., Kirk, D., Krause, E., Kuehn, K., Kuropatkin, N., Lahav, O., Lima, M., Maia, M. A. G., March, M., Martini, P., Menanteau, F., Miller, C. J., Miquel, R., Mohr, J. J., Nichol, R. C., Ogando, R., Romer, A. K., Sanchez, E., Scarpine, V., Sevilla-Noarbe, I., Smith, R. C., Soares-Santos, M., Sobreira, F., Swanson, M. E. C., Tarle, G., Thomas, D., Walker, A. R., Weller, J., and Zhang, Y. (2017). Weak-lensing mass calibration of redMaPPer galaxy clusters in Dark Energy Survey Science Verification data. *mnras*, 469:4899–4920.

Merloni, A., Predehl, P., Becker, W., Böhringer, H., Boller, T., Brunner, H., Brusa, M., Dennerl, K., Freyberg, M., Friedrich, P., Georgakakis, A., Haberl, F., Hasinger, G., Meidinger, N., Mohr, J., Nandra, K., Rau, A., Reiprich, T. H., Robrade, J., Salvato, M., Santangelo, A., Sasaki, M., Schwope, A., Wilms, J., and German eROSITA Consortium, t. (2012). eROSITA Science Book: Mapping the Structure of the Energetic Universe. *arXiv e-prints*, page arXiv:1209.3114.

Miller, L., Heymans, C., Kitching, T. D., van Waerbeke, L., Erben, T., Hildebrandt, H., Hoekstra, H., Mellier, Y., Rowe, B. T. P., Coupon, J., Dietrich, J. P., Fu, L., Harnois-Déraps, J., Hudson, M. J., Kilbinger, M., Kuijken, K., Schrabback, T., Semboloni, E., Vafaei, S., and Velander, M. (2013). Bayesian galaxy shape measurement for weak lensing surveys - III. Application to the Canada-France-Hawaii Telescope Lensing Survey. , 429:2858–2880.

Miller, L., Kitching, T. D., Heymans, C., Heavens, A. F., and van Waerbeke, L. (2007). Bayesian galaxy shape measurement for weak lensing surveys - I. Methodology and a fast-fitting algorithm. , 382:315–324.

Miralda-Escude, J. (1991). Gravitational lensing by clusters of galaxies - Constraining the mass distribution. *apj*, 370:1–14.

Mulroy, S., Smith, G., Haines, C., Marrone, D., Okabe, N., Pereira, M., Egami, E., Babul, A., Finoguenov, A., and Martino, R. (2014). Locuss: The near-infrared luminosity and weak-lensing mass scaling relation of galaxy clusters. *Monthly Notices of the Royal Astronomical Society*, 443(4):3309–3317.

Navarro, J. F., Frenk, C. S., and White, S. D. M. (1997). A Universal Density Profile from Hierarchical Clustering. *apj*, 490:493–508.

Neumann, D. M. and Arnaud, M. (1999). Regularity in the X-ray surface brightness profiles of galaxy clusters and the M-T relation. *aap*, 348:711–727.

Oguri, M., Bayliss, M. B., Dahle, H., Sharon, K., Gladders, M. D., Natarajan, P., Hennawi, J. F., and Koester, B. P. (2012). Combined strong and weak lensing analysis of 28 clusters from the sloan giant arcs survey*. *Monthly Notices of the Royal Astronomical Society*, 420(4):3213–3239.

Oguri, M. and Hamana, T. (2011). Detailed cluster lensing profiles at large radii and the impact on cluster weak lensing studies. *Monthly Notices of the Royal Astronomical Society*, 414(3):1851–1861.

Oguri, M. and Takada, M. (2011). Combining cluster observables and stacked weak lensing to probe dark energy: Self-calibration of systematic uncertainties. *Phys. Rev. D*, 83:023008.

Oukbir, J. and Blanchard, A. (1992). X-ray clusters in open universes. *aap*, 262:L21–L24.

Parroni, C., Mei, S., Erben, T., Waerbeke, L. V., Raichoor, A., Ford, J., Licitra, R., Meneghetti, M., Hildebrandt, H., Miller, L., Côté, P., Covone, G., Cuillandre, J.-C., Duc, P.-A., Ferrarese, L., Gwyn, S. D. J., and Puzia, T. H. (2017). Next generation virgo cluster survey. xxi. the weak lensing masses of the cfhtls and ngvs redgold galaxy clusters and calibration of the optical richness. *The Astrophysical Journal*, 848(2):114.

Peebles, P. J. E. (1980). *The large-scale structure of the universe*.

Pereira, M. E. S., Soares-Santos, M., Makler, M., Annis, J., Lin, H., Palmese, A., Vitorelli, A. Z., Welch, B., Caminha, G. B., Erben, T., Moraes, B., and Shan, H. (2018). Weak-lensing calibration of a stellar mass-based mass proxy for redMaPPer and Voronoi Tessellation clusters in SDSS Stripe 82. *mnras*, 474:1361–1372.

Press, W. H. and Schechter, P. (1974). Formation of Galaxies and Clusters of Galaxies by Self-Similar Gravitational Condensation. *apj*, 187:425–438.

Radovich, M., Puddu, E., Bellagamba, F., Roncarelli, M., Moscardini, L., Bardelli, S., Grado, A., Getman, F., Maturi, M., Huang, Z., Napolitano, N., McFarland, J., Valentijn, E., and Bilicki, M. (2017). Searching for galaxy clusters in the Kilo-Degree Survey. *aap*, 598:A107.

Raymond, J. C. and Smith, B. W. (1977). Soft X-ray spectrum of a hot plasma. *apjs*, 35:419–439.

Riess, A. G., Filippenko, A. V., Challis, P., Clocchiatti, A., Diercks, A., Garnavich, P. M., Gilliland, R. L., Hogan, C. J., Jha, S., Kirshner, R. P., Leibundgut, B., Phillips, M. M., Reiss, D., Schmidt, B. P., Schommer, R. A., Smith, R. C., Spyromilio, J., Stubbs, C., Suntzeff, N. B., and Tonry, J. (1998). Observational evidence from supernovae for an accelerating universe and a cosmological constant. *The Astronomical Journal*, 116(3):1009.

Rykoff, E. S., Koester, B. P., Rozo, E., Annis, J., Evrard, A. E., Hansen, S. M., Hao, J., Johnston, D. E., McKay, T. A., and Wechsler, R. H. (2012). Robust Optical Richness Estimation with Reduced Scatter. *apj*, 746:178.

Schechter, P. (1976). An analytic expression for the luminosity function for galaxies. *apj*, 203:297–306.

Schneider, P. (2006). Weak gravitational lensing. In *Proceedings, 33rd Advanced Saas Fee Course on Gravitational Lensing: Strong, Weak, and Micro: Les Diablerets, Switzerland, April 7-12, 2003*, pages 269–451.

Schwarzschild, K. (1916). "Uber das Gravitationsfeld eines Massenpunktes nach der Einsteinschen Theorie. *Sitzungsberichte der K"oniglich Preussischen Akademie der Wissenschaften (Berlin), 1916, Seite 189-196*.

Seljak, U. (2000). Analytic model for galaxy and dark matter clustering. *mnras*, 318:203–213.

Sereno, M., Covone, G., Izzo, L., Ettori, S., Coupon, J., and Lieu, M. (2017). PSZ2LenS. Weak lensing analysis of the Planck clusters in the CFHTLenS and in the RCSLenS. *Mon. Not. Roy. Astron. Soc.*, 472(2):1946–1971.

Sereno, M. and Ettori, S. (2015). CoMaLit - IV. Evolution and self-similarity of scaling relations with the galaxy cluster mass. *mnras*, 450:3675–3695.

Sereno, M., Giocoli, C., Izzo, L., Marulli, F., Veropalumbo, A., Ettori, S., Moscardini, L., Covone, G., Ferragamo, A., Barrena, R., and Streblyanska, A. (2018). Gravitational lensing detection of an extremely dense environment around a galaxy cluster. *Nature Astronomy*, 2:744–750.

Sereno, M., Veropalumbo, A., Marulli, F., Covone, G., Moscardini, L., and Cimatti, A. (2015). New constraints on 8 from a joint analysis of stacked gravitational lensing and clustering of galaxy clusters. *Monthly Notices of the Royal Astronomical Society*, 449(4):4147–4161.

Sheldon, E. S., Johnston, D. E., Frieman, J. A., Scranton, R., McKay, T. A., Connolly, A. J., Budavári, T., Zehavi, I., Bahcall, N. A., Brinkmann, J., and Fukugita, M. (2004). The Galaxy-Mass Correlation Function Measured from Weak Lensing in the Sloan Digital Sky Survey. *aj*, 127:2544–2564.

Sheldon, E. S., Johnston, D. E., Scranton, R., Koester, B. P., McKay, T. A., Oyaizu, H., Cunha, C., Lima, M., Lin, H., Frieman, J. A., Wechsler, R. H., Annis, J., Mandelbaum, R., Bahcall, N. A., and Fukugita, M. (2009). Cross-correlation Weak Lensing of SDSS Galaxy Clusters. I. Measurements. *apj*, 703:2217–2231.

Simet, M., McClintock, T., Mandelbaum, R., Rozo, E., Rykoff, E., Sheldon, E., and Wechsler, R. H. (2017). Weak lensing measurement of the mass-richness relation of SDSS redMaPPer clusters. *mnras*, 466:3103–3118.

Sunyaev, R. A. and Zeldovich, I. B. (1980). The velocity of clusters of galaxies relative to the microwave background - The possibility of its measurement. *mnras*, 190:413–420.

Sunyaev, R. A. and Zeldovich, Y. B. (1970). Small-Scale Fluctuations of Relic Radiation. , 7:3–19.

Sunyaev, R. A. and Zeldovich, Y. B. (1972). The Observations of Relic Radiation as a Test of the Nature of X-Ray Radiation from the Clusters of Galaxies. *Comments on Astrophysics and Space Physics*, 4:173.

The Dark Energy Survey Collaboration (2005). The Dark Energy Survey. *arXiv e-prints*, pages astro–ph/0510346.

Tinker, J. L., Robertson, B. E., Kravtsov, A. V., Klypin, A., Warren, M. S., Yepes, G., and Gottlöber, S. (2010). THE LARGE-SCALE BIAS OF DARK MATTER HALOS: NUMERICAL CALIBRATION AND MODEL TESTS. *The Astrophysical Journal*, 724(2):878–886.

Umetsu, K., Medezinski, E., Nonino, M., Merten, J., Postman, M., Meneghetti, M., Donahue, M., Czakon, N., Molino, A., Seitz, S., Gruen, D., Lemze, D., Balestra, I., Benítez, N., Biviano, A., Broadhurst, T., Ford, H., Grillo, C., Koekemoer, A., Melchior, P., Mercurio, A., Moustakas, J., Rosati, P., and Zitrin, A. (2014). CLASH: Weak-lensing Shear-and-magnification Analysis of 20 Galaxy Clusters. *apj*, 795:163.

Velten, H. E. S., vom Marttens, R. F., and Zimdahl, W. (2014). Aspects of the cosmological "coincidence problem". *European Physical Journal C*, 74:3160.

Verdoes Kleijn, G., de Jong, J. T. A., Valentijn, E., Kuijken, K., KiDS Consortium, and Astro-WISE Consortium (2012). Astro-WISE for KiDS Survey Production and Quality Control. In Ballester, P., Egret, D., and Lorente, N. P. F., editors, *Astronomical Data Analysis Software and Systems XXI*, volume 461 of *Astronomical Society of the Pacific Conference Series*, page 237.

Viola, M., Cacciato, M., Brouwer, M., Kuijken, K., Hoekstra, H., Norberg, P., Robotham, A. S. G., van Uitert, E., Alpaslan, M., Baldry, I. K., Choi, A., de Jong, J. T. A., Driver, S. P., Erben, T., Grado, A., Graham, A. W., Heymans, C., Hildebrandt, H., Hopkins, A. M., Irisarri, N., Joachimi, B., Loveday, J., Miller, L., Nakajima, R., Schneider, P., Sifón, C., and Verdoes Kleijn, G. (2015). Dark matter halo properties of gama galaxy groups from 100 square degrees of kids weak lensing data. *Monthly Notices of the Royal Astronomical Society*, 452(4):3529–3550.

Viola, M., Maturi, M., and Bartelmann, M. (2010). Constraints on the inner density profile of dark matter haloes from weak gravitational lensing. *mnras*, 403:859–869.

Yang, X., Mo, H. J., Van Den Bosch, F. C., Jing, Y. P., Weinmann, S. M., and Meneghetti, M. (2006). Weak lensing by galaxies in groups and clusters – i. theoretical expectations. *Monthly Notices of the Royal Astronomical Society*, 373(3):1159–1172.

Yee, H. K. C. and Gladders, M. D. (2002). Optical Surveys for Galaxy Clusters. In Chen, L.-W., Ma, C.-P., Ng, K.-W., and Pen, U.-L., editors, *AMiBA 2001: High-Z Clusters, Missing Baryons, and CMB Polarization*, volume 257 of *Astronomical Society of the Pacific Conference Series*, page 109.

Zeldovich, Y. B. (1972). A hypothesis, unifying the structure and the entropy of the Universe. *mnras*, 160:1P.

Zwicky, F. (1933). Die Rotverschiebung von extragalaktischen Nebeln. *Helvetica Physica Acta*, 6:110–127.

Zwicky, F. (1937). On the Masses of Nebulae and of Clusters of Nebulae. *apj*, 86:217.