Exploring the accuracy of analytic methods in predicting the evolution of large-scale structure

Alberto Acuto

A thesis submitted in partial fulfilment of the requirements of
Liverpool John Moores University
for the degree of
Doctor of Philosophy.

November 2021

Declaration

The work presented in this thesis was carried out at the Astrophysics Research Institute, Liverpool John Moores University. Unless otherwise stated, it is the original work of the author.

While registered as a candidate for the degree of Doctor of Philosophy, for which submission is now made, the author has not been registered as a candidate for any other award. This thesis has not been submitted in whole, or in part, for any other degree.

Alberto Acuto
Astrophysics Research Institute
Liverpool John Moores University
IC2, Liverpool Science Park
146 Brownlow Hill
Liverpool
L3 5RF
UK

February 2022

Abstract

Cosmology is at a crossroads. Experiments are providing an unprecedented amount of data that, in theory, should lead to clear solutions to the many open questions in cosmology. However, with new data comes new questions and recently uncovered tensions between the predictions of the standard model of cosmology and observations are leading some to question the very foundations on which the standard model is built. To explore the vast cosmological landscape, numerical simulations are often employed, but given the broad parameter space that needs to be explored other faster (but more approximate) methods need to be adopted to maximise the coverage and the possible extensions surveyed. In this panorama one of the options is the halo model, a simple and elegant way to study the clustering of matter in the Universe. However, this method is not free from assumptions and associated uncertainties.

In this thesis I explore the uncertainties associated with the halo model making use of cosmological numerical simulations. I use the BAHAMAS simulations to obtain data products such as the mass density profiles of the haloes and the number density of haloes over a wide range of masses and I use these quantities in the halo model formalism in order to make a self-consistent comparisons against the simulations results. Aside from this application, I calibrate a fitting function on the Einasto function, which has been shown to be a good representation of the matter distribution inside haloes, and I use a standard form for the halo mass function. Comparing against the simulation matter power spectrum at different redshift, I show the accuracy of the halo model predictions is strongly dependent on the mass definitions used with differences over 50%. In particular, the transition region between the 1-halo and the 2-halo terms and in the smallest scales sampled ($k \approx 10\,h\,\mathrm{Mpc}^{-1}$). This picture applies to both collisionless and hydrodynamical simulations, where galaxy formation processes are taken into account. In contrast to the poor ability in reproducing the matter clustering, the halo model

can reproduce the relative impact of baryons on the matter clustering to a competitive accuracy ($< 5\%$) in line of next-generation observations predictions.

In the second part of this work, I analyse the halo model applications of large-scale structure observables as gravitational weak lensing and thermal Sunyaev-Zel'Dovich effect. To explore these observables, I have built the halo model using the electron pressure inside haloes (relevant for the tSZ effect), and I have made several realisations of the matter power spectrum up to z=3 for the lensing observables, in both the collisionless and hydrodynamical cases. In this analysis, I have compared against observational data (e.g., KiDS-450 survey and *Planck*) and results obtained from light-cones from the BAHAMAS simulations. I examined the dependence of the results on the different mass definitions and the baryonic effects, in particular the baryonic suppression that can be inferred from this set of observables.

Publications

In the course of completing the work presented in this thesis, the following paper have been submitted and accepted for publication in a refereed journal:

Alberto Acuto, Ian G. McCarthy, Juliana Kwan, Salcido Jaime, Stafford G. Sam and Andreea S. Font. *The BAHAMAS project: evaluating the accuracy of the halo model in predicting the non-linear matter power spectrum.* 2021, MNRAS, Volume 508, Issue 3, pp. 3519–3534.

February 2022

Acknowledgements

It is not possible to start without thanking my supervisor, Ian McCarthy, for the countless helpful comments, for all the suggestions, insightful discussions, and lessons during these years. Your mentoring guided me during this PhD, you taught me to question all the results and go deeper in every aspect of the project, always providing a throughout viewpoint on the results.

I would like to thank Andreea Font and Rob Crain for their scientific discussions and support. And, again, Rob for the many tips on carrying on using IDL. Another mention is for all the BAHAMAS postdocs: Juliana Kwan, Jaime Salcido, Violeta Gonzalez-Perez and Victoria Yankelevich for their continuous support and their feedback in every stage of my project.

I would like to express my sincere gratitude to my viva examiners Prof. Ivan Baldry and Prof. Scott Kay. I have appreciated the time you spent reading and critically evaluating the work presented in the Thesis and for the many insightful questions and genuine interest in the results presented.

The ARI has been a special place to do my PhD, full of great scientists and kind people. I wish to thank all the PhD fellows with whom I had the honour to share this journey. Thanks for the lovely chats, scientific discussions, and shared moments in the ARI. I also thank everyone involved in the LiV.DAT CDT for the learning opportunities and all the support provided.

From the hot and uncomfortable rooms in 'Medicina legale' to the conference halls in the middle of the Dolomites. From the first Montecarlo codes, passing through many exams and lab experiments to the long discussions trying to make sense of all of these

questions. We have gone far, (now Dr.) Alessandro Ignesti. You know you have a considerable share in this achievement.

Eleonora, the smile of my days and the light in the night, thank you for being by my side.

Without a supportive and caring family this would have not been possible. You are always in my heart and mind.

"E quindi uscimmo a riveder le stelle"

Dante Alighieri, Inferno XXXIV, 139

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The main purpose of this work is to explore the halo model approach in reproducing large-scale structure predictions and to test its accuracy by making detailed comparisons with cosmological hydrodynamical simulations. This work highlights the power of using the relative impact of baryons in the halo model approach to explore, not only the matter clustering signal but also large-scale observables with an accuracy comparable to next-generation observational requirements.

In the introduction I aim to provide a general picture of the standard cosmological framework (1.1) with an emphasis on large-scale structure theory (1.1.1) and large-scale structure observables (1.1.2) considered in this work. I will, also, present the methods used to explore the cosmological parameter space such as cosmological numerical simulations (1.3) and the halo model (1.4), highlighting the strengths and weaknesses of each of one of those methods.

1.1 The ΛCDM cosmology

ΛCDM is a simple, predictive theory that describes the geometry and the evolution of the Universe from its origin up to its final days. Its name comes from the two major components of the Universe: the cosmological constant denoted by Λ and the cold dark matter (CDM). The main components of the Universe are radiation in form of photons and neutrinos, ordinary matter (baryons and leptons), (non-relativistic) dark matter and a cosmological constant, Λ. The first component, radiation, had a significant role in the first instants of the life of the Universe, and its contribution can be measured in

the Cosmic Microwave Background (hereafter CMB), which provides a picture of the Universe a few hundred thousand years after the origin, often called 'Big Bang'[1].

Dark matter is the fundamental building block of galaxies and, therefore, of the Universe itself. The current assumption is that dark matter is 'cold'[2]. This means that the particles are relatively heavy and were not-relativistic when they decoupled from the radiation field, allowing the start of the formation of potential wells large enough to trigger the later baryonic collapse that led to the development of the first objects. The debate about the origin and the nature of dark matter (which particle is it, how massive is the particle, does it interact in other ways other than via gravitational force?) are fundamental questions that remain unsolved.

The baryons are the main visible components of stars and galaxies that grant them the status of the 'real' observable tracers of the Universe. Without the baryons, we would not have had the first stars that have reionized the 'opaque' Universe, which became cold, after having lost all the energy through the emission of the photons that compose the CMB (reionizing phase, Barkana & Loeb 2001; Koopmans et al. 2015). Without the baryons we would not have, as well, light to trace the underlying dark matter haloes.

Finally, there is the cosmological constant, Λ. The Λ constant, called 'dark energy', is the last component of the concordance model for a spatially flat Universe. The standard model assumes a cosmological constant which implies the energy density is time invariant (Linder, 2003, 2005; Hu, 2005; Riess et al., 1998, 2004). However, the physical nature of this component (e.g., whether it represents a scalar field, zero-point energy, vacuum energy) and whether it is truly time-independent is currently unknown and represents one of the most pressing questions in physics.

It is useful to introduce a variable that will help us describe the time evolution of the Universe. This variable, *redshift*, is related to the expansion parameter, a, and it is noted as z. I define, starting from an observation point of view, the redshift as :

[1]The redshift of the CMB emission is $z_{\rm CMB} \approx 1089$ and it is assumed to be a flat plane distribution. Therefore using the Friedmann's equations we can obtain the age of the Universe at that event following:

$$t_e = t_H \int_{z_{\rm CMB}}^{\infty} \frac{dz'}{(1+z')E(z')} \ ; \tag{1.1}$$

where $E(z)$ is defined in eqn. 1.12 and $t_{\rm H} = H_0^{-1} \approx 9.78\,h^{-1}$Gyrs ($10^9$ years) also called Hubble time (Hogg, 1999). Doing this calculation using *WMAP*9 cosmology the age of the Universe when the CMB was emitted was $t_e \approx 400000$ years.

[2]The dark matter temperature is a measure of when this component is not in equilibrium with the other components and starts its collapsing evolution. A 'hot' dark matter would have been relativistic and, therefore, it would have decoupled at a similar time as radiation, leading to smaller potential wells. A 'cold' model is privileged to justify the observed size and clustering of haloes and the hierarchical growth of structures.

$$z = \frac{\lambda_o - \lambda_e}{\lambda_e} \ , \tag{1.2}$$

with λ_o wavelength of radiation from the observed source at the origin (which is assumed, for simplicity, the origin of the coordinate system) at a certain time t_0 and, consequently, emitted in a previous time t_e, with the corresponding emission wavelength λ_e. The source is moving with the expansion of the Universe at the comoving coordinate r_e so:

$$\int_{t_e}^{t_o} \frac{cdt}{a(t)} = d \ , \tag{1.3}$$

with distance d. The light emitted is moving at a constant speed therefore d is constant for both couples (t_o,t_e) and ($t_o + \delta t_o$,$t_e + \delta t_e$) and that allows us to write:

$$\int_{t_e}^{t_o} \frac{cdt}{a(t)} = \int_{t'_e=t_e+\delta t_e}^{t'_o=t_o+\delta t_o} \frac{cdt'}{a(t')} \rightarrow \frac{\delta t_o}{a_o} = \frac{\delta t}{a} \ . \tag{1.4}$$

In eqn. 1.4 I have $a_o = a(t_o)$ and I can define the frequency (e.g., ν_e for the emission) as δt_e^{-1} and obtain, with few logical passages using the associated wavelengths, to:

$$\nu_e a = \nu_o a_o \rightarrow \frac{a}{\lambda_e} = \frac{a_o}{\lambda_o} \rightarrow 1 + z = \frac{a_o}{a} \ , \tag{1.5}$$

making use of the z definition (eqn. 1.2) I have obtained the physical formulation of redshift in terms of the scale factor, a. It is customary to define the current scale factor, a_0, as being equal to one. In this way, the scale factor is relative to the present size of the Universe.

After presenting the general components of the Universe, it is worth putting a frame on these components and presenting the formal cosmological setup guided by general relativity. General relativity (GR) is the theory that describes gravitational interactions on several physical and cosmological scales. The 'cosmological principle' states that the Universe is, under a statistical point of view, homogeneous and isotropic in both matter and space on large scales (tens of megaparsec scale, inhomogeneity in galaxy cluster or cosmic web are ignored). The Friedmann-Lemaître-Robertson-Walker (FLRW) metric is the mathematical formalism that describes the cosmic distances as:

$$ds^2 = c^2 dt^2 - a(t)^2 \left(\frac{dr^2}{1 - Kr^2} + r^2 d\theta^2 + r^2 \sin^2 \theta d\phi^2 \right) \ , \tag{1.6}$$

that emerges naturally from the cosmological principle. In eqn. 1.6 I have used spherical polar coordinates to determine the distance s: r, θ, ϕ those are in comoving units (r

by convention is dimensionless), t is the proper time and $a(t)$ is the scale or expansion factor. K is the normalised curvature parameter which can assume the values -1, 0 and 1 corresponding to three possible metric solutions (open, flat and closed). By combining the 4-dimensions (t, r, θ, ϕ) of the space-time FLRW metric and the Einstein's field equations we obtain the Friedmann cosmological equations:

$$\dot{a}^2 + Kc^2 = \frac{8}{3}\pi G\rho a^2 + \frac{\Lambda c^2}{3} \ ; \tag{1.7}$$

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}\left(\rho + \frac{3p}{c^2}\right) + \frac{\Lambda c^2}{3} \ ; \tag{1.8}$$

defined in terms of expansion parameter a (eqn. 1.5) and normalised such $z = 0 \rightarrow a(t_0) = 1$. The overdot notation defines the temporal derivative. In eqns. (1.7)(1.8) G is the gravitational constant, ρ is the density, p is the pressure and c speed of light. With eqns. (1.7)(1.8) I can define the Universe critical density:

$$\rho_c \equiv \frac{3}{8\pi G}\left(\frac{\dot{a}}{a}\right)^2 \ , \tag{1.9}$$

that allows the introduction of the density parameter Ω defined as follows:

$$\Omega(t) = \frac{\rho}{\rho_c} \ . \tag{1.10}$$

With the introduction of the density parameter it is possible to rewrite, accordingly, the FLRW equations with these parameters explicitly:

$$\begin{aligned} &\underbrace{\frac{8\pi G}{3H^2}}_{\Omega_m}\rho \underbrace{- \frac{K}{a^2H^2}}_{\Omega_K} + \underbrace{\frac{\Lambda}{3H^2}}_{\Omega_\Lambda} = 1 \ , \\ &\Omega_m + \Omega_\Lambda = 1 - \Omega_K \ , \\ &H^2 = \frac{\dot{a}^2}{a^2} = \frac{8\pi G\rho + \Lambda c^2}{3} \ . \end{aligned} \tag{1.11}$$

As shown in eqn. 1.11, the Hubble parameter H is introduced and expressed in terms of the expansion parameter. The equations are expressed in term of the dimensionless density parameter, Ω, which states that, for a spatially flat universe ($\Omega_K = 0$), the sum of the components (matter and Λ) must be equal to one. Note that Ω_m here includes the contributions of both baryons (often identified as Ω_b) and dark matter and the contribution of radiation is neglected, which as I presented above, is relevant only in the early phases of the Universe. Another component is introduced into the standard CDM paradigm is massive neutrinos, that are believed to have a significant role before

recombination. The neutrino contribution is calculated as $\Omega_\nu = M_\nu/(93.12\,h^2)$ with the assumption that all neutrinos are non-relativistic (which is valid by $z \lesssim 100$ for 90% of those Ali-Haïmoud & Bird (2013)).

The cosmological evolution of the parameters is usually expressed by the, so-called, Peebles function (Peebles, 1993):

$$E(z)^2 = \Omega_m(1+z)^3 + \Omega_K(1+z)^2 + \Omega_\Lambda \ , \tag{1.12}$$

that can also be expressed in terms of scale factor a. Equation 1.12 shows the time (redshift or scale factor) evolution in terms of the matter density, Λ and curvature parameters and it is bounded to the geometry of the Universe.

The geometry of the Universe is relevant not only for what happened in the past, in terms of physical distances, but it is also important to understand the fate of the Universe, as different geometries lead to different outcomes. For a spatially flat universe (which is consistent with current measurements, Planck Collaboration (2018); Efstathiou & Gratton (2020)[3]), we should expect it to keep expanding at a constant pace, however, the presence of dark energy causes an accelerated expansion. In the hypothetical case of a matter-dominated Universe ($\Omega_\Lambda = 0$), the geometry can describe, as well, the scenarios of the 'big crunch' ($K > 0$, the matter density is above the critical threshold, thus the gravity can pull the universe back to a singularity) or the 'open' cases ($K < 0$, a continuous expansion).

Since I have already introduced and discussed distances, it is worth defining the various distances in cosmology that will be relevant in the later discussions and Chapters. I start with the Hubble distance, which is the distance travelled by the light in a Hubble-time:

$$D_H = \frac{c}{H_0} \ , \tag{1.13}$$

since H_0^{-1} is the Hubble-time (t_h). With this, I can define the comoving distance, which is worth highlighting can be differentiated between 'line-of-sight' (LOS) and 'transverse' (T) (Hogg, 1999):

$$D_C^{\rm LOS}(z) = \chi(z) = D_H \int_0^z \frac{dz'}{E(z')} \ . \tag{1.14}$$

[3]In recent years, some authors have analysed in detail the measurements of the curvature of the Universe using *Planck* data trying to explain some of the tensions present. For instance, Di Valentino et al. (2020) found that using the same *Planck* data without any priors imposed the best-fitting model prefers a curvature parameter significantly different from 0.

Eqn. 1.14 presents the distance between two events, in the notation I use I define the comoving distance with the Greek letter χ. For the transverse case, I define it in a more general way making the cosmological curvature dependence explicit:

$$f_K^T(z) = \begin{cases} K^{-1/2}\sin(K^{1/2}\chi(z)), & K > 0; \\ \chi(z), & K = 0; \\ -K^{-1/2}\sinh(-K^{1/2}\chi(z)), & K < 0. \end{cases} \tag{1.15}$$

Having defined the comoving distance, I now introduce the angular distance as the product between the comoving distance and the scale factor:

$$D_A(z) = \frac{f_K^T}{1+z} = a f_K^T \ . \tag{1.16}$$

The angular distance is the separation between two objects on the sky plane and it is often used to convert object positions on the telescope field onto real locations in the sky. Finally, I define the luminosity distance as the ratio between the integrated flux S and the integrated luminosity L:

$$D_L = \sqrt{\frac{L}{4\pi S}} \ , \tag{1.17}$$

which is related to the previous quantities as $D_L = a^{-1} f_K^T = a^{-2} D_A$. Hereafter, to keep the notation more clear, I avoid using the (T) notation when I write the transverse comoving distance, f_K.

Currently, the best constraints on cosmological models comes from detailed analysis of fluctuations in the CMB, starting in the early 90s with the COBE satellite[4] (Smoot et al., 1992; Fixsen et al., 1996). Since then, our understanding of the Universe has improved significantly. The measurements from the WMAP[5](Spergel et al., 2003) and *Planck* [6] satellites (Planck Collaboration et al., 2014) provided strong constraints on the cosmological parameters such as the amplitude of the fluctuations σ_8, the spectral index n_s and the Hubble constant (H, or its normalised version $h = H\,100^{-1}$ km/s/Mpc). Not only measurements from CMB observations confirm that the Universe is well represented by the ΛCDM paradigm, results from galaxy clustering (e.g., BOSS[7] survey, Ivanov et al. (2020)) and weak lensing provide (e.g., KiDS-1000[8], Heymans et al. (2021)) strong proofs on the flatness of the Universe and the hierarchical growth of structure.

[4] https://lambda.gsfc.nasa.gov/product/cobe/, COsmic Background Explorer.
[5] https://map.gsfc.nasa.gov/ , Wilkinson Microwave Anisotropy Probe.
[6] https://www.cosmos.esa.int/web/planck, named after the German 1918 Nobel winner Max Planck.
[7] BOSS = Baryon Oscillation Spectroscopic Survey.
[8] KiDS = Kilo Degrees Survey.

The last two parameters introduced, σ_8 and n_s, describe the distribution of fluctuations on the large-scale and are obtained by measuring the matter power spectrum:

$$\sigma_8^2 = \frac{1}{2\pi^2} \int T(k)P(k)W^2(kR)k^2 dk \ ; \tag{1.18}$$

where k is the comoving wavenumber, $P(k)$ is the primordial power spectrum, $T(k)$ is the transfer function which scales the primordial power spectrum to the present day assuming linear evolution. W is the Fourier transform of the top-hat window function of a real-space sphere of radius $R = 8\,h^{-1}$Mpc. The Gaussian fluctuations that happened in the inflationary period[9] left the seeds for the gravitational collapse that can be measured today with the positions and clustering of haloes. The primordial power spectrum can be expressed as:

$$P(k) = A_s k^{n_s} \ , \tag{1.19}$$

introducing the power spectrum amplitude, A_s, and the explicit dependence on the spectral index, n_s. Hypothetically, if the Universe had scale invariant fluctuations, we would have $n_s = 1$ (Harrison (1970); Peebles & Yu (1970); Zeldovich (1972), the Harrison-Zel'Dovich power spectrum) but the current, accepted, inflationary paradigm (as the slow-roll inflation, Linde (1982)) predicts that $n_s \approx 0.96$, which has been observationally verified (Planck Collaboration, 2018).

The ΛCDM model has successfully explained a wide range of observations over the past twenty years, but recent detailed comparisons have revealed some interesting tensions. Two tensions prevail in the current cosmological landscape: the value of Hubble's constant and LSS statistics of Ω_m and σ_8, often recast in the joint constraint parametrised as S_8[10]. The value of the Hubble's constant, H_0, presents a large discrepancy from early Universe measurements, as CMB and BAO (Planck Collaboration, 2016, 2018; Abbott et al., 2018), and late Universe probes as supernovae Ia (Riess et al., 2016) or strong lensing time delays (e.g., H0LiCOW, Wong et al. (2020); Verde et al. (2019)). The tension comparing the measurements of Ω_m and σ_8 arise because *Planck* data favour a higher value of one of those parameters (or both) whilst LSS data sets (e.g., Heymans et al. (2013); Hildebrandt et al. (2017); Leauthaud et al. (2017); McCarthy et al. (2018)) appear to prefer a relatively lower value. These tensions escalated recently due to the disagreement between the estimation of these from early Universe measurement of the CMB and large-scale structure tracers (hereafter LSS). This situation exposed some key trigger points in the solid theoretical background allowing variations and extensions of

[9] The inflationary epoch is a theorised phase in which the Universe underwent a phase of exponential expansion. This theory finds many confirmations by the 'cosmological principle', which assures that, statistically, the Universe is homogeneous and isotropic and explains the emission features of the CMB.

[10] $S_8 \equiv \sigma_8 \sqrt{\Omega_m/0.3}$.

TABLE 1.1: Cosmological parameter values from *Planck*2018 survey and *WMAP* year 9 results. For the *Planck* case I present the parameters 68% interval using CMB power spectra in combination with CMB lensing reconstruction and BAO (TT,TE,EE+lowE+Lensing+BAO) for *WMAP*9, instead, I present the ΛCDM best-fitting parameters with BAO and H_0 priors. In this table I present the dark matter density parameter as Ω_c.

Parameter	*Planck* (Planck Collaboration, 2018) 68% limits	*WMAP*9 (Hinshaw et al., 2013) 68% limits
H_0 [km s^{-1}Mpc^{-1}]	67.66 ± 0.42	69.33 ± 0.88
Ω_m	0.311 ± 0.0056	0.288 ± 0.01
Ω_b	0.049 ± 0.0003	0.0472 ± 0.0010
Ω_c	0.262 ± 0.0001	0.2408 ± 0.009
Ω_Λ	0.6889 ± 0.0056	0.712 ± 0.010
σ_8	0.8102 ± 0.006	0.830 ± 0.018
n_s	0.9665 ± 0.0038	0.971 ± 0.010
$10^9 A_s$	2.105 ± 0.030	2.427 ± 0.078

this theory. Different species of dark matter, as a warm component or self-interacting one, time-evolving dark energy, presence and feedback of massive neutrinos in the large-scales growth of structure are some of the many hypotheses surveyed. This plethora of options is heavily investigated in many different works, trying to dig out the information and find the right solution (Di Valentino et al., 2021).

In Table 1.1 I present a recap of the main ΛCDM parameters shown previously in the text from the *Planck* 2018 (Planck Collaboration, 2018) and *WMAP*9 surveys (Hinshaw et al., 2013). The parameters are expressed with the 68% confidence level and are obtained by combining the CMB power spectra (in the *Planck* case with also the EE, TE, TT polarisation spectra available), BAO measurements and CMB lensing in the first case and with priors on BAO and H_0 in case of latter. In this table I have used the parameter Ω_c to describe the dark matter density parameter (please note that $\Omega_c + \Omega_b = \Omega_m$).

1.1.1 Large-scale structure

Large-scale structure refers to the distribution of haloes and matter on the largest scales. With haloes, I talk about objects from galaxy to group-cluster of galaxies size, in a more general term. The distribution of haloes and matter in the Universe, in what is commonly known as the cosmic web (Bond et al., 1996), creates several patterns. The galaxies that occupy groups and clusters are the nodes developed from the intersection between long filaments and sheets of matter. Surrounding the elongated filaments and walls of matter there are near-empty void regions (van de Weygaert, 2016; Libeskind et al., 2018). The primordial density field presented Gaussian random fluctuations, that

through gravitational instabilities developed the patterns we observe in the cosmic web (Bartolo et al., 2004).

Haloes are not glued to a single position in the sky as they wander driven by gravity, collapsing and merging into larger objects in a process called hierarchical structure formation (White & Frenk, 1991; Kauffmann et al., 1999). In fact, the most massive galaxy clusters we can observe in the 'recent' Universe ($z \lesssim 0.5$) are the pinnacle of the growth of structure. The accretion phenomena granted them such strong potential wells that we can consider them as approximately closed cosmic laboratories.

While the growth of structure is well understood in the context of gravitational instability within the ΛCDM model, detailed comparisons to observations require that we must also carefully model the role of baryons and associated non-gravitational processes (e.g., feedback, radiative cooling, star formation). Furthermore, the details of the gravitational collapse are altered in detail when massive neutrinos are incorporated and/or the nature of dark energy is modified.

One of our most powerful tools for testing models of large-scale structure is the matter power spectrum and its evolution with redshift. The challenge for next-generation weak lensing surveys is to obtain this quantity to a percentage level accuracy, allowing us to distinguish between different cosmologies and galaxy formation feedback.

1.1.1.1 Matter power spectrum

A detailed description and characterisation of the matter power spectrum at different redshifts is the ultimate goal for modern cosmology. Our actual knowledge is definitely more refined than before, thanks to many results from CMB and all-sky surveys (Ross et al., 2020; Wang et al., 2020; Zhao et al., 2021).

From a theoretical perspective, the matter clustering can be described in a statistical formalism from the dark matter density field as an n-point correlation function of the density fluctuations $\delta(\mathbf{x})$. Any process that develops initial random fluctuation would, eventually, produce Gaussian distributed perturbations thanks to the central limit theorem. If we express the density perturbations in the Universe in terms of a relative background mean density, $\overline{\rho}$, we have:

$$\delta r = \frac{\rho(r)}{\overline{\rho}} - 1 \ ; \tag{1.20}$$

where r is a 3D position in space while $\mathbf{x}$ is the 3D normalised position in space. Assuming that the real coordinate space can be expressed by n-point correlation functions,

without any loss of information, it is possible to write:

$$\langle \delta(\mathbf{x_i}) ... \delta(\mathbf{x_n}) \rangle \equiv \xi_n(\mathbf{x_1} ... \mathbf{x_n}) \ , \tag{1.21}$$

with ξ_n denoted as the n-point correlation function. In Fourier space all the calculations of the n-point correlations can be easily re-written in a more concise form as:

$$A(\mathbf{x}) = \int \frac{d^3\mathbf{k}}{(2\pi)^3} A(\mathbf{k}) \exp(i\mathbf{kx}) \ , \tag{1.22}$$

that with the knowledge of the Fourier form of the Dirac delta function δ_D:

$$\delta_D(\mathbf{k}_{1..i..n}) = \int \frac{d^3\mathbf{x}}{(2\pi)^3} \delta(\mathbf{k}) \exp(i\mathbf{kx}) \ , \tag{1.23}$$

we can describe the real space density field as the sum over the Fourier modes:

$$\delta(\mathbf{x}) = \int \frac{d^3\mathbf{k}}{(2\pi)^3} \delta(\mathbf{k}) \exp(i\mathbf{kx}) \ . \tag{1.24}$$

From eqn. 1.21 we know that the density field is equivalent to the measure of n^{th} correlation function between those, meaning we can express the n^{th} correlation point using eqn. 1.24 as:

$$\begin{aligned} \langle \delta(\mathbf{k_1}) \delta(\mathbf{k_2}) \rangle &= (2\pi)^3 \delta_D(\mathbf{k_{12}}) P(k) \ , \\ \langle \delta(\mathbf{k_1}) \delta(\mathbf{k_2}) \delta(\mathbf{k_3}) \rangle &= (2\pi)^3 \delta_D(\mathbf{k_{123}}) B(\mathbf{k_1}, \mathbf{k_2}, \mathbf{k_3}) \ , \\ &.... \end{aligned} \tag{1.25}$$

where P is the 2-point correlation function in k-space, with $k = \|\mathbf{k}\|$. To clarify further, the Dirac delta application means that if $k_i = k_j \rightarrow \delta_D^{ij} = 1$ otherwise is $\delta_D^{ij} = 0$ and this can be generalised in the n-dimension version of this operator. The 2-point correlation function is known as power spectrum in Fourier space, B is the 3-point correlation function, often called bispectrum, and so on with the other n-point statistics. Each n-point correlation describes a particular feature of the Universe, for example, the power spectrum describes the clustering of haloes, non-Gaussianity[11] can be quantified with the bispectrum (Scoccimarro et al., 1998; Meerburg et al., 2019).

[11] The initial fluctuations that are measured by the power spectrum are assumed to be Gaussian, but primordial interactions and dynamics in the inflationary period are expected to influence this assumption (Meerburg et al., 2019).

1.1.2 Large-scale structure tracers

In the next short sub-sections, I go into detail about the various large-scale structure traces used in this work. I focus on the theoretical background, applications and cosmological dependencies exploring the work undertaken so far.

1.1.2.1 Gravitational weak lensing

To study the LSS we have several strong tools, one of these is the gravitational weak lensing signal. The presence of collapsed objects (identified as lenses) along the line of sights has a significant impact on disturbing the photon path-lines coming from objects (sources) that lie at redshifts further away. The weak lensing is the measure of the distortion that affects the background sources due to the effect of the lenses resulting in both shape and signal differences from the original distribution. The background sources can be photons either from galaxies or the CMB itself.

Gravitational weak lensing can be studied in statistical analysis using the coherence of the shear signal over the sky. This happens because observations cannot describe the true 3D distribution of matter of the galaxies, so it is not easy to differentiate between a 'real' distortion of the galaxy observed compared to one created by weak lensing. Lensing, in a broader sense, has two separate effects: the distortion and the magnification of the light of the background sources. The magnification, an enhancement of the flux, can be used as optical lenses to study objects further away with better resolution and details (Bonvin et al., 2017).

Gravitational weak lensing measurements have been used to identify the underlying dark matter haloes in spiral and elliptical galaxies (Brainerd et al., 1996; Griffiths et al., 1996; Bartelmann & Schneider, 2001) and map the mass distribution of galaxy clusters (e.g. Kaiser 1998; Hoekstra et al. 2013) providing strong constraints on the growth of structure in a CDM Universe.

Lensing is not sensitive only to the geometry of the Universe (as supernovae Ia or baryon acoustic oscillations that are pure geometric probes) but also a direct measurement of the growth of structure at different epochs (Hoekstra et al., 2013; Weinberg et al., 2013). The geometry is present because, similarly to linear optics, the distances between source-lens-observers can measure the cosmological distances as well as different growth status between those (like the increase of distances caused by the acceleration, for instance due to the effect of dark energy).

The weak lensing formalism starts from the observed surface brightness distribution, $f^{\rm obs}$, that must be mapped back to the original surface brightness distribution, f^{s},

under the assumption that the angle between the source and the lens is small (Hoekstra & Jain, 2008). This can be done applying a distortion matrix $\mathcal{A}$[12]:

$$
\begin{aligned}
f^{\rm obs}(\theta_i) &= f^s(\mathcal{A}_{ij}\theta_j) \ ; \\
\mathcal{A} = \frac{\partial(\delta\theta_i)}{\partial\theta_j} \equiv (\delta_{ij} - \Psi_{,ij}) &= \begin{pmatrix} 1-\tilde{k}-\gamma_1 & -\gamma_2 \\ -\gamma_2 & 1-\tilde{k}+\gamma_1 \end{pmatrix} \ ; \\
\Psi_{,ij} &\equiv \frac{\partial^2\Psi}{\partial\theta_i\partial\theta_j} \ .
\end{aligned}
\tag{1.26}
$$

In eqn. 1.26 I introduce the 2D gravitational potential Ψ and its second order spatial derivative $\Psi_{,ij}$ in terms of the observed lensed positions in the sky θ_{ij}. γ_1 and γ_2 are the components of the complex shear $\gamma \equiv \gamma_1 + i\gamma_2$, often expressed in terms of the shear orientation angle α as $\gamma = \gamma\exp(2i\alpha)$. These components are defined in terms of $\Psi_{,ij}$ as follows:

$$
\gamma_1 = \frac{1}{2}(\Psi_{,11} - \Psi_{,22}) \quad \text{and} \quad \gamma_2 = \Psi_{,12} \ . \tag{1.27}
$$

$\tilde{k}$ is the lensing convergence, a scalar quantity obtained by a weighting projection of the matter density fluctuation as:

$$
\tilde{k}(\theta) = \frac{1}{2}\nabla^2\Psi(\theta) = \int d\chi W(\chi)\delta[\chi, \chi\theta] \ ; \tag{1.28}
$$

with ∇^2 Laplace's operator defined using the flat sky approximation (so $\theta_i = \theta_j$[13]) as $\nabla^2 \equiv \partial^2/\partial\theta^2$. χ is the comoving distance, depending on the curvature and geometry of the Universe as shown in eqn. 1.15, that in a spatially flat Universe is equivalent as $K = 0 \rightarrow f_K = \chi$ (Seljak, 1998; Harnois-Déraps et al., 2015). In the treatment, here and throughout, I have assumed the Universe is spatially flat. The $W(\chi)$ is the lensing efficiency function (or 'kernel') and is obtained as:

$$
W(\chi) = \frac{3}{2}\Omega_m H_0^2 \chi a^{-1}(\eta(\chi)|z) \int d\chi_s n_s(\chi_s) \left(1 - \frac{\chi}{\chi_s}\right) \ , \tag{1.29}
$$

with H_0 Hubble parameter at $z = 0$, matter density parameter Ω_m and a scale factor. $\eta(\chi)$ is commonly used to define the coformal time of the distance and is often substituted in the formalism instead of redshift. The source-redshift distribution $n_s(\chi_s)$ is defined such $\int_0^\infty d\chi_s n_s(\chi_s) = 1$. Under the assumption of a single source redshift distribution,

[12] $\mathcal{A}$ is the Jacobian matrix of the transformation as can be seen from eqn. 1.26.
[13] In the case of $\theta_i \neq \theta_j$ the Lapalce's operator is defined as : $\nabla^2 \equiv \partial^2/\partial\theta_i\partial\theta_j$.

z_s, n_s function collapses to a Delta function as:

$$n_s(\chi) = \delta_D(\chi - \chi_s) \begin{cases} 1, & \text{if } \chi = \chi_s, \\ 0, & \text{otherwise.} \end{cases} \tag{1.30}$$

As I have already said, the weak lensing signal is measured by estimating the shear correlation functions from observations of galaxy shape catalogues in both optical/ near-infrared (e.g., CFHTLenS, Erben et al. (2013)) and radio images surveys (e.g., SKA[14], Brown et al. (2015)). The 2-point correlation function of the shear (that is the same for the convergence) for a catalogue is defined as:

$$\xi_{\gamma_i,\gamma_j}(\theta) = \langle \gamma_{\mathbf{i}}(\theta_1) \cdot \gamma_{\mathbf{j}}(\theta_2) \rangle \ ; \tag{1.31}$$

eqn. 1.31 is evaluated in general terms for two redshift bins (i^{th} and j^{th}) and with $\theta = \|\theta_1 - \theta_2\|$. Since it is more complex to deal with vectorial products, it is standard to define eqn. 1.31 in terms of sum and multiplication of the shear fields as:

$$\begin{aligned} \xi_+(\theta) &= \langle \gamma_{i+}(\theta_1)\gamma_{j+}(\theta_2) \rangle \ , \\ \xi_\times(\theta) &= \langle \gamma_{i\times}(\theta_1)\gamma_{j\times}(\theta_2) \rangle \ ; \end{aligned} \tag{1.32}$$

with the parallel term (+) and the perpendicular (×). With this simplification I can rewrite eqn. 1.31 as $\xi_{\gamma_i\gamma_j} = \xi_+ + \xi_\times$. The shear power spectrum at a particular angular wave number ℓ is the Fourier transform of $\xi_{\gamma_i\gamma_j}$ and can be expressed as the projection of the mass density power spectrum. I will go into detail about the calculations of the shear power spectrum in Section 2.21 with the different applications and the differences between the single source CMB weak lensing, the tomographic lensing analysis and the source redshift distributions used.

Weak lensing observations have been used to constrain cosmological parameters via the shear correlation functions defined above, but also indirectly via their use in estimating the halo masses of galaxies, groups and clusters. The mass estimation is extremely important since it provides a less biased measurement of the total halo mass, which is less subject to hydrodynamical distortion phenomena that can have a non-negligible effect on the baryonic (mainly on gas properties in X-ray/SZ analysis) counterpart (see the discussion on the hydrostatic mass bias in section 1.1.2.3). For what concerns cosmological parameter estimation via cosmic shear, the most sensitive parameters are the ones characterising the geometry of the Universe (being the source-lens distances heavily affected) such as the total amount of matter (Ω_m) and the dark energy component

[14] https://www.skatelescope.org/the-ska-project/, Square Kilometre Array.

(Ω_Λ). Furthermore, weak lensing studies can provide constraints on the large-scale clustering parameter σ_8 and next-generation observations aim to efficiently disentangle the constraints on Ω_m and σ_8, which so far have been strongly correlated (Heymans et al., 2021).

The source of uncertainties in weak lensing, shear imaging, analysis comes from various origins. Some of these uncertainties, like the PSF (point spread function of the telescope) and its calibration, come from technical difficulties that are taken into account in a generally straightforward way. Another weak spot of these analyses is accurately estimating the redshift distribution of the sources, due to the higher uncertainty associated with photometric redshift measurements. Perhaps even more challenging is distinguishing a lensed distorted galaxy image from a real galaxy shape. The intrinsic alignment (IA) in weak lensing analysis is a problem that has no easy solution. Numerical studies using hydrodynamical simulations have provided tools (e.g. identify which component is less susceptible to be influenced by the dark-matter halo shape or large-scale tidal force) to identify the most effective observable to understand the weak lensing signal (Velliscig et al., 2015, 2017; Hill et al., 2021). In addition, the presence of baryons and their impact in the non-linear power spectrum can introduce biases onto the cosmological analysis performed (Jing et al., 2006; van Daalen et al., 2011; Schneider & Teyssier, 2015; Chisari et al., 2019; Debackere et al., 2020, 2021; Schneider et al., 2019; van Daalen et al., 2020; Angulo et al., 2021; Aricò et al., 2021; Martinelli et al., 2021).

In the above treatment of weak lensing, I have adopted the so-called Born approximation. The reconstruction of the lensing potential and the distortion of the light can, in principle, be quite complex. However, in the case of weak lensing light rays will, by definition, only undergo small deflections. Assuming these deflections to be small and the angles between the lines of sight and original photon path lines are small, allows us to describe the lensing convergence signal (eqn. 1.28) as a function of the lowest order lensing potential. Consequently, the convergence signal $\tilde{k}$ is the integrated matter density on unperturbed line of sight weighted by the lensing kernel (eqn. 1.29, Schäfer et al. (2012); Petri et al. (2017)).

This approximation is enough for quick analytical predictions for 2 point correlation functions (as these halo model predictions, Schneider et al. (1998)) but it might not be enough accurate for precision measurements, where a post-Born approximation can solve the analytical issues (Cooray & Hu, 2002; Fabbian et al., 2018) or in simulations with a full ray-tracing procedures (Giocoli et al., 2016; Wei et al., 2018).

1.1.2.2 CMB weak lensing

The primary CMB is well understood and focus in recent years has shifted somewhat to new probes that exploit the CMB in different ways, such as secondary anisotropies arising from interactions with matter along the line of sight between the CMB and the observer. One of the most interesting is CMB weak lensing, which is the use of the CMB photons as source and their variations is associated by potential gradients along the line of sight (Hu, 2000; Lewis & Challinor, 2006).

The first authors that presented the possibility of having gravitational effects on the CMB anisotropies were Blanchard and Schneider in 1987 (Blanchard & Schneider, 1987). But the first detection of those effects were only hinted at following WMAP observations using cross-correlations with large-scale structures by galaxy surveys (Hirata et al., 2004) and within the CMB observation itself with WMAP data (Feng et al., 2012), ACT (Atacama Cosmology Telescope, Das et al. (2011)), SPT (South Pole Telescope, Keisler et al. (2011)) and, at last, with *Planck* (Planck Collaboration et al., 2018).

In the case of unlensed CMB power spectrum, the temperature measured is approximately uniform (of 2.725 K) with fluctuations in the order of 10^{-5} due to the rise of acoustic baryonic oscillations and signal from large-scale growth of structures associated with integrated Sachs-Wolfe effect (ISW, Sachs & Wolfe (1967)) due to evolving potentials along the line of sight. Small scale CMB temperature spectrum is influenced, also, by the reionization of neutral gas (mainly between $6 < z < 20$ due to stars and quasars feedback).

The deflection of light, as presented in the previous section, is influenced by both under- (e.g. weak lensing of cosmic voids, Davies et al. (2018, 2021)) and over-densities that can lead a deflection up to $\delta\beta \approx 10^{-4}$, where β is the deflection angle, and corresponds to 2° in angular scales for this CMB case (Lewis & Challinor, 2006).

This probe has an advantage compared to other weak lensing studies, since the source distribution is a well measured (almost) flat distribution of photons which have an associated lower measurement error compared to the redshifts of the galaxies in galaxy surveys (i.e., the associated higher uncertainty with photometric redshift measurement). Cross-correlating this observable with other lensing observations or probes, such as thermal Sunyaev-Zel'Dovich or X-ray can provide further details on the growth of structure (Hill & Spergel, 2014; Hand et al., 2015; Liu & Hill, 2015; Harnois-Déraps et al., 2016, 2017; Darwish et al., 2021).

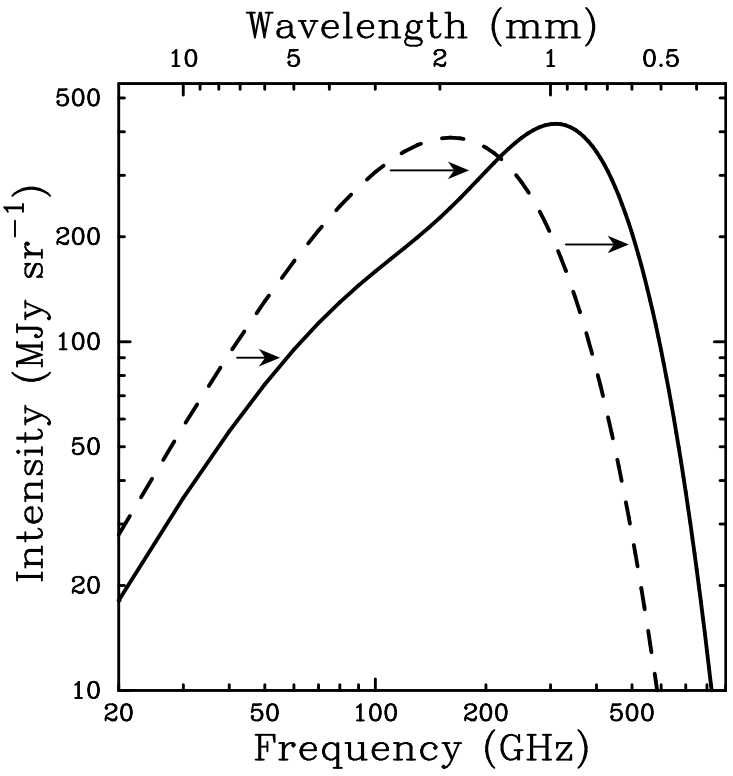

FIGURE 1.1: Thermal Sunayev-Zel'Dovich spectral shift feature. Figure adapted from Carlstrom et al. (2002).

1.1.2.3 Thermal Sunyaev Zel'Dovich effect

The thermal Sunyaev-Zel'Dovich effect (hereafter tSZ, Sunyaev & Zeldovich (1981)) is a frequency dependent distortion of the CMB spectrum due to the inverse Compton effect from the CMB photons and the high-energy free electrons in the intra-cluster medium of the galaxy clusters (Birkinshaw, 1999). This interaction creates a unique and characteristic signature on the CMB spectrum. This spectral feature shifts photons below $\nu \approx 220$ GHz into an excess above that energy range, resulting in a secondary source of anisotropies in the CMB spectral energy distribution.

Fig. 1.1 shows a simplistic representation of the CMB spectral distortion due to tSZ effect. The figure is adapted from Carlstrom et al. (2002).

Alongside the pure thermal contribution, if the collapsed object is moving with respect to the CMB rest frame this adds a velocity contribution, as a Doppler shift, on the spectral feature that is known as the kinetic Sunyaev Zel'Dovich (kSZ, Cooray & Chen (2002)). This secondary effect is smaller compared to the thermal contribution, but with upcoming surveys and detailed observations, the kSZ will become an important observable (Deutsch et al., 2018; Münchmeyer et al., 2019; Coulton et al., 2020).

The spectral distortion of the CMB of the tSZ effect can be expressed in terms of change of the temperature as $\Delta T_{\rm SZ}$ at a dimensionless frequency, $x \equiv \frac{h_p \nu}{k_B T_{\rm CMB}}$, as:

$$\frac{\Delta T_{\rm SZ}}{T_{\rm CMB}} = y f_{\rm SZ}(x) = f_{\rm SZ}(x) \int n_e \sigma_T \frac{k_B T_e}{m_e c^2} d\ell \ , \tag{1.33}$$

with y called Compton y-parameter, σ_T is Thomson cross-section, k_B is the Boltzmann constant, h_p Planck constant, T_e is the electron temperature and $m_e c^2$ is the electron rest mass energy (please note that in dimension $k_B T$ is an energy) integrated along the line of sight. $f_{\rm SZ}$ is the frequency dependence function and it has a general shape like:

$$f_{\rm SZ}(x) = \left(x \frac{e^x + 1}{e^x - 1} - 4 \right) (1 + \delta_{\rm SZ}(x, T_e)) = (x \coth(x/2) - 4)(1 + \delta_{\rm SZ}(x, T_e)) \ ; \tag{1.34}$$

with $\delta_{\rm SZ}$ as term of the relativistic correction that can be applied to the frequency dependence. In the non relativistic Rayleigh-Jeans limit[15] we have that $f_{\rm SZ}(x) \to -2$.

Eqn. 1.33 shows that in the spectral signature of the tSZ there is no dependency on the redshift providing a very strong tool for detecting distant objects (Hurier et al., 2017; Lin et al., 2021), studying galaxy cluster dynamics (Baldi et al., 2019), mapping the cosmic web (Tanimura et al., 2020a) and constraining cosmological parameters (Komatsu & Seljak, 2002; Komatsu et al., 2011; Battaglia et al., 2012b; McCarthy et al., 2014; Bolliet et al., 2020; Tanimura et al., 2021).

The cosmological dependence is primarily through parameters that control the growth of structure, such as σ_8 and Ω_m, therefore tSZ offers a constraining tool via both the thermal power spectrum as well as tSZ cluster number counts. These dependencies can help constrain the redshift evolution of the cluster number counts, test models of dark energy (Bolliet et al., 2018) but also provide an overview of cosmological thermal energy evolution (Chiang et al., 2020, 2021). In fact, the key role in this observable are clusters of galaxies, that are themselves important for cosmology. The evolution of galaxy clusters, from cluster abundances, clustering of clusters, and the thermodynamics of the ICM contain a wealth of information about our cosmology and important baryonic processes (such as feedback from supernovae and active galactic nuclei).

The energy distribution, in terms of thermal pressure, is where most uncertainties lie. An accurate description of this observable is needed, since it has been shown that this

[15] The Rayleigh-Jeans limit is an approximation, under classical physics assumptions, of the black body radiation emission (Rybicki & Lightman, 1986). This result predicts correctly the radiative frequencies below $\approx 10^5$ GHz and can be applied in low frequency regimes as submillimeters tSZ observations.

quantity is extremely sensitive to even the smallest changes in baryonic feedback. Starting from X-ray selected massive clusters, Arnaud et al. (2010) provided a widely used pressure profile characterisation, however, as other independent work have proven, a correct inclusion of AGN feedback is mandatory for a precise and realistic representation of this observable especially from simulations as Komatsu & Kitayama (1999); Battaglia et al. (2010, 2012a,b).

Another complex issue in tSZ modelling is the overall halo mass. Cluster mass measurements from weak lensing and X-ray observations tend to differ by $10 - 20\%$ (Hurier & Angulo (2018) and references in this paper). This problem, known as the hydrostatic mass bias, comes from the assumption that the gas inside haloes is in a hydrostatic configuration (so in perfect equilibrium), but this is a strong assumption that impacts significantly the overall mass estimation of the haloes involved (Salvati et al., 2018, 2019; Tanimura et al., 2021). The tSZ observations are, together with X-ray observations since both measure the same quantity (hot gas distribution), the most affected by this bias that can have a strong influence on the measurements (Bocquet et al., 2019; Ettori et al., 2019; Ruppin et al., 2019). This issue is less worrying in simulations, given the fact that in simulations there is already the knowledge of the 'true' mass of haloes and can help constrain and reconcile the differences between the observations (Barnes et al., 2020; Gianfagna et al., 2021).

In recent years, the tSZ effect has enabled us to measure and trace gas distribution far outside the virial radius of galaxies by stacking multiple images enhancing, significantly, the signal to noise ratio (Tanimura et al., 2020a). SZ stacking images, together with weak lensing techniques (Mead et al., 2010), X-ray stacking (Tanimura et al., 2020b) and study on the galaxy distribution (Malavasi et al., 2020), will enhance our understanding of the cosmic web as well.

1.2 Need for methods to explore parameters space

Stage IV cosmic shear surveys such as Euclid[16], the Rubin Observatory Legacy Survey of Space and Time (LSST)[17], and the Nancy Grace Roman Space Telescope (NGRST)[18] aim to measure the matter power spectrum to percent level accuracy, in principle allowing constraints to be placed on important cosmological parameters. Comparatively tight

[16] https://www.euclid-ec.org/
[17] https://www.lsst.org/
[18] https://roman.gsfc.nasa.gov/

constraints are also expected from forthcoming X-ray surveys as eRosita[19], Sunyaev-Zeldovich effect surveys with SPT-3G[20], Advanced ACTPol[21], and Simons Observatory[22] and optical surveys (e.g. galaxy clustering) such as LSST and Euclid and DESI[23]. In order to deliver on the aims of these surveys, a clear requirement is that we must be able to predict the observables (e.g., weak lensing power spectrum, galaxy clustering, SZ power spectrum) for a given cosmology, to an accuracy that is at least as precise as the statistical measurement errors. Otherwise, we risk biasing the derived cosmological parameters. In the specific case of weak lensing, this means predicting the matter power spectrum to percent level accuracy. The problem is particularly challenging, as most of the signal from current LSS tests comes from non-linear scales, thus requiring theoretical models to accurately follow matter as shells cross and collapse into 'haloes', with galaxies potentially forming at their centres.

The requirement for tools to explore extensively the cosmological parameter space accurately are one of the main challenges at present in cosmological research. To model the non-linear scales numerical simulations are, by far, the most accurate and reliable but are time and resource consuming, leading to a trade-off between simulation size, parameter space sampling and inclusion of baryons. Therefore analytical or semi-analytical methods are the most common choice since they are flexible, quick and easier to calibrate for different purposes.

A promising solution to tackle these problems is using emulation techniques (e.g., Kwan et al. 2015; Heitmann et al. 2016; DeRose et al. 2019; Nishimichi et al. 2019; Rogers et al. 2019; Pellejero-Ibañez et al. 2020; Bose et al. 2020, 2021; Donald-McCann et al. 2021; Spurio Mancini et al. 2021), whereby a grid of cosmological simulations, spanning some cosmological parameter landscape, is first run and then an emulator (e.g., based on Gaussian process interpolation or neural networks) is used to quickly and accurately interpolate the results (e.g., the matter power spectrum) for any choice of cosmological parameters found within the boundaries of the initial grid. Such emulators, which can typically be run in fractions of a second, can be implemented in cosmological Markov chains[24].

While emulation of cosmological simulations is clearly going to be an important tool going forward, it does have limitations. For example, predictions are confined to the

[19] https://www.mpe.mpg.de/eROSITA
[20] https://pole.uchicago.edu/
[21] https://act.princeton.edu/
[22] https://simonsobservatory.org/
[23] https://www.desi.lbl.gov/
[24] Markov chains Monte-Carlo (MCMC) is a series of methods built on the Bayes theorem used to solve highly dimensional parameter space. The basic and simplified idea is to 'walk' through the parameter values between the set boundaries (priors), finding the new posterior values and evaluating the likelihood of such prediction, please check Sharma (2017) for a more in-depth presentation of this method.

parameter space defined in the initial base grid of simulations, where there is a trade-off between accuracy of the emulator prediction, the volume of the cosmological landscape being surveyed (i.e., the range of parameter values included), and the number of simulations that can feasibly be run for the base grid. In addition, adding new extensions (e.g., beyond ΛCDM) or probing a larger (or different) cosmological parameter spaces generally requires one to construct a whole new grid or partially re-train the existing emulator, limiting the adaptability of this tool.

In addition, at the moment most emulators are based on simulations that do not incorporate the hydrodynamics of the galaxy formation processes.

For example, it has been shown in recent work based on cosmological hydrodynamical simulations that baryons can alter the matter power spectrum by up to a few tens of percent (e.g., Jing et al. 2006; van Daalen et al. 2011; Schneider & Teyssier 2015; Mummery et al. 2017; Chisari et al. 2019; van Daalen et al. 2020), which is significantly larger than the anticipated statistical error of future weak lensing measurements.

A widely used approach is the halo model, that given its intuitive and physically-motivated formalism and its quick and easy calibration can reproduce several LSS tracers and the non-linear correction in the matter power spectrum. I will discuss in detail the halo model approach in the next Sections after having presented the different types of simulations.

1.3 Role of cosmological simulations to do precision cosmology

As discussed above, cosmological simulations are the most accurate method for following the growth of structure. Here I refer specifically to simulations that follow the evolution of matter within large volumes (at least hundreds of Mpc on a side) so that they are cosmologically representative. Other astrophysical simulations are, instead, focused on smaller objects or systems, such as single galaxies or groups of galaxies aiming to characterise some particular aspects of galaxy evolution and small scale interactions (Vogelsberger et al., 2020). Cosmological simulations have multiple targets in their developments: in the first place, we want them to reproduce a realistic matter distribution of the Universe, secondly, identify clear dependence of the observables on cosmological parameters. In addition, we might want to explore the role of baryons and extensions of the standard CDM cosmology.

Cosmological simulations can be differentiated in two separate ways: N-body simulations (also called collisionless) that can describe very well the pure gravitational interplay

between particles and hydrodynamical simulations, which attempt to model the complex non-linear physics of galaxy formation processes. At fixed resolution and volume, N-body simulations tend to be considerably cheaper than hydrodynamical simulations and can describe carefully the dark matter distribution, they are easily scalable so they can reproduce huge volumes and explore a vast parameter space, such as that done in the Quijote simulation project[25] (Villaescusa-Navarro et al., 2020) as a latter example.

Hydrodynamical simulations are much more complex and expensive. Hydrodynamical simulations are this complex because they need to solve non-linear hydrodynamical equations, alongside the gravitational interactions between particles. Aside from these equations that describe the fluid motions, a number of other astrophysical processes have to be complemented to shape the galaxy population. Some of these processes include the production of heavy elements from stellar feedback, star formation on galaxy scales and cooling and heating (e.g., AGN feedback) phenomena taking place on cluster scales. The resolution and volume of the simulations are dictated by the final goal of the project. If the main interest of the project is galaxy formation, then the simulations must have higher resolution and smaller volume. Examples are the EAGLE[26] and IllustrisTNG[27] simulations (Schaye et al., 2015; Crain et al., 2015; Pillepich et al., 2018). However, if the final goal is to build simulations for cosmological purposes, like for Cosmo-OWLS and BAHAMAS[28] projects (Schaye et al., 2010; Le Brun et al., 2014; McCarthy et al., 2017), larger volumes are optimal, thus the resolution must be lower[29].

In the next section, we discuss briefly the different types of simulations.

1.3.1 N-Body simulations

The first numerical experiment can be back-traced to 1941 with a 'simple' two body interaction made by Holmberg (Holmberg, 1941) using thirty-seven light-bulbs as 'particle type' elements to measure the gravitational interaction using the bulb brightness. The computing power and the complexity of the problems have scaled up to today where we are able to create realistic realisations of the Universe with billions of particles (the MilleniumXXL simulation has 303 billion particles and 4.1 pGpc, physical Gigaparsecs, box Angulo et al. (2012)).

To follow the distribution of collisionless particles, we need to determine the gravitational force acting on each mass element. Gravity solvers, indeed, must determine these forces

[25] https://quijote-simulations.readthedocs.io/en/latest/
[26] http://eagle.strw.leidenuniv.nl/.
[27] https://www.tng-project.org/.
[28] https://www.astro.ljmu.ac.uk/ igm/BAHAMAS/.
[29] Note that all of these simulations have approximately the same number of particles, which is mainly limited by the computer memory.

on each mass element by solving the Poisson's equation. These solvers subdivide a significant portion of the universe into many particles and compute the forces between those particles and evolve, in discrete time-steps, the system (Somerville & Davé, 2015; Vogelsberger et al., 2020).

There are different approaches to solving the Poisson's equation: the direct summation (brute-force), tree methods, and particle-meshes. The discrete direct summation, the so-called particle-particle scheme, is the simplest and most accurate form to solve the particles gravitational interactions involving the calculation of the forces between every pair of particles. However, modern N-body simulations use a large number of particles making it impossible, even for current machines, to compute all the gravitational forces using direct summation methods (please note that this method scales as $\mathcal{O}(\mathcal{N}^2)$, with $\mathcal{N}$ number of particles). Approximate methods, such as the tree method, are employed to accelerate these calculations, where short-range interactions are evaluated using the direct summation method and long-range interactions are approximated by the lowest order term of a multipole expansion (this reduces the computational cost from $\mathcal{O}(\mathcal{N}^2)$ to $\mathcal{O}(\mathcal{N}\log\mathcal{N})$). Particle-mesh methods assign particles to cells in a grid calculating the gravitational forces using the Fourier transforms of the density field (the complexity is $\mathcal{O}(\mathcal{N})$, where $\mathcal{N}$ is the number of grid cells). In hybrid schemes, different methods are combined, the direct summation evaluates short-range interactions in the approximated tree-like method, while Fast Fourier Transform (FFT) techniques are used as a force solver for large scales.

An advantage of tree codes lies in how accurately they can represent the forces on the particles down to a lower limit, identified as force softening length[30], while the particle-mesh codes are limited by the grid size.

In N-body simulations there is no need of employing relativistic gravity instead of using classical Newtonian gravity, which provides, anyway, a good approximation for both the linear and non-linear large-scale growth. Furthermore, the velocities induced by these interactions are far below the relativistic limit.

The general results obtained using this type of simulation are quite astonishing:

- dark matter distribution on large scale is extremely well recovered, analysing the clustering signal;
- halo demographics (halo mass functions);
- the internal halo structure (e.g. the Navarro-Frenk-White, NFW, and Einasto profiles).

[30] Often defined as ϵ. It is standard to use a softening length to smooth high density environments to avoid unphysical two-body scatterings.

All these (semi) independent results are comparable to the observational results, with respect to the changes and corrections such as galaxy bias to match dark matter halo clustering when comparing galaxy clustering.

This starting point provides solid foundations (a 'skeleton') for structure formation. To reproduce the *observable* universe, however, we need to add baryons and associated non-gravitational physics.

1.3.2 Hydrodynamical simulations

Simulating baryons allows us, in principle, to reproduce the visible component of the Universe providing a more direct comparison with the observations.

The baryon component is described as an ideal inviscid fluid, and in the beginning, composed of a mixture of hydrogen and helium. To solve the hydrodynamical equations (the Euler's equations, which are a form of Navier-Stokes equations assuming no viscosity or conduction, Springel (2010)) there are two ways: the Eulerian and Lagrangian approaches.

The Lagrangian approach assumes an observer that follows an individual fluid unit, with its own properties, like density, as it moves through space and time. The Eulerian approach focuses on specific locations in space through which the fluid flows as time passes (Vogelsberger et al., 2020). The Eulerian method discretizes the fluid onto grid cells and computes the advection of the fluid properties across the cell boundaries. Smoothed particle hydrodynamics (SPH) is the most widely used mesh-free Lagrangian technique for solving the continuum dynamics of fluids using sampling particles. In this case, the fluid is discretized in mass and the particles carry the information about the fluid, which is evaluated by a kernel-weighted sum over several neighbouring particles, inside a specified smoothing length.

Eulerian methods can accurately model the dynamic instabilities in the gas, like Kelvin-Helmholtz or Rayleigh-Taylor instabilities which are fundamental for fluid mixing (e.g., Zavala et al. (2012)). SPH-based codes, with the addition of a form of artificial viscosity and conduction, can reproduce, effectively, the mixing of different fluids providing, as well, an increased dynamic range (e.g., Price (2008)). An advantage of the particle-based methods is that, being able to follow how the masses move, it is straightforward to understand how mass assembles into galaxies, and to track ejected material from galaxies (Somerville & Davé, 2015).

At this stage, we have explored how the gravitational and hydrodynamical forces are characterised in the simulations, however, there are a host of galaxy formation processes

not described in the treatment above. These processes take place on scales below the simulation's resolution, the *'sub-grid'* (or 'sub-resolution') models. These models are introduced to approximate the astrophysical processes allowing us to produce more accurate theoretical predictions.

In the treatment so far, we have considered properties of dark matter and baryons, but we have not included any prescription for radiation. Radiative cooling is implemented in most cosmological simulations and describes the energy dissipation of the gas through processes, like collisional excitation and ionization, inverse Compton, recombination and free-free emission. These processes are coupled to the energy equation via cooling functions (e.g., Wiersma et al. (2009b)) because they describe phenomena happening on atomic level, that cannot be resolved.

Some models characterise the stellar life cycle, from birth to death. Star formation takes place in dense, cold molecular clouds and its modelling is crucial. However, to model accurately the different phases of the galaxy environments is rather difficult. In addition, the timescales on which these processes occur are shorter than the simulation time-step. To solve these issues, the dense gas is often described by a polytropic equation of state (i.e., $T \propto \rho^{\gamma(\rho)}$), which relates the temperature of the gas to the density (Springel & Hernquist, 2003; Dalla Vecchia & Schaye, 2008). Simulating the molecular phase of the interstellar medium is rather challenging because it requires the modelling of the interaction between different actors as gas, dust and radiation (Vogelsberger et al., 2020). Following these difficulties, to model star formation, it is standard to adopt a probabilistic approach based on the calculated star formation rate. The criteria that need to be met to transform a gas particle into a star particle are based on: density thresholds, Jeans length and virial parameters ensuring that the star formation takes place in dense regions, close to gravitational instabilities (Katz et al., 1996; Teyssier et al., 2013; Schaye et al., 2015; Hopkins et al., 2018).

The stars have interactions with their surrounding gas through the injection of both energy and momentum leading to a feedback loop regulating star formation (Springel & Hernquist, 2003; Vogelsberger et al., 2013; Pillepich et al., 2018). Supernovae explosions and stellar winds are the principal sources to inject energy and momentum into the interstellar medium. Supernovae explosion must be capable of igniting galactic outflows that eject particles, metals and large amounts of energy into the surrounding gas. This energy heats the gas and helps regulate star formation. However, given the variety of implementations of these phenomena, together with several schemes to treat the combined effects of heating and cooling in the interstellar medium, more work is required to fully understand these stellar feedback channels.

The presence of supermassive black holes in the centres of galaxies is well known, hence simulations must include models to encapsulate their feedback. The formation of those objects, however, is not well understood at all, so it is standard to artificially seed the dark matter haloes ($M > 10^{10}\,\mathrm{M}_{\odot}$) allowing them to give rise to active galactic nuclei (AGN) that will accrete matter onto them and expel, via jets and winds, large amounts of high energy radiation (Di Matteo et al., 2005; Springel, 2005; Sijacki et al., 2008; Booth & Schaye, 2009). These AGN feedback phenomena are much more powerful than the supernovae and bring about the heating of the intracluster medium and the regulation of star formation in massive galaxies [31].

The main results obtained using hydrodynamical simulations are:

- global galaxy properties such as the stellar mass function, galaxy clustering, scaling relations (e.g., the Tully-Fisher);
- properties of the late-type galaxies (formation of a multi-component stellar disk, presence of cooling phases perpendicularly to the disk);
- properties of the early-type galaxies (formation of spheroid galaxies, with matched early star formation history and matched size-mass-velocity dispersion);
- intracluster medium, circumgalactic medium and diffuse gas (Lyman-α emission and metal emission/absorption lines from hot gas).

1.4 The Halo model approach

The halo model (Ma & Fry, 2000; Peacock & Smith, 2000; Seljak, 1998, 2000; Cooray & Sheth, 2002; Smith et al., 2003)[32] is a powerful tool that can solve some of the issues that emulators or simulations have. The halo model approach provides a simple physically-motivated picture for describing the clustering of matter and haloes.

If we focus on its standard (and simplest) form, the halo model requires as input the distribution of matter within haloes (i.e., their density profiles), the demographics of haloes (i.e., the abundances of haloes as a function of mass and redshift, the halo mass function), and in the 'linear regime' a prescription for a halo bias which describes how the

[31] In this Section, we have not explored the inclusion of magnetic fields in the simulations and their sub-grid modelling. Those simulations have higher complexity because the hydrodynamics equations must fulfil also the magnetic component. Popular sub-grid models include the presence of magnetic fields in the interstellar medium (relevant criteria for triggering star formation), production and feedback of cosmic rays and turbulence.

[32] The modern halo model formalism is presented in work as Seljak (1998) and Peacock & Smith (2000) but the first theoretical approach on this method can be traced back to McClelland & Silk (1977) and Scherrer & Bertschinger (1991).

clustering of haloes is related to the underlying clustering of matter. These ingredients can be used to calculate the so-called 1-halo and 2-halo clustering terms. The 1-halo term describes the clustering of particles inside the same halo, whilst the 2-halo term describes the clustering between different haloes. The halo model often assumes the Universe is spatially flat (though this assumption can be relaxed) and, usually, employs the so-called Limber approximation (which includes clustering only in the plane of the sky but not along the line of sight) when computing projected clustering observables such as the angular correlation function of galaxies or the tSZ effect power spectrum. The Limber approximation may fail when considering extremely large scales (in particular, it is relevant for CMB spectrum fluctuations in the large scale regime). A beyond-Limber approach has gathered momentum in explaining odd multipoles in weak-lensing and CMB analysis (Lemos et al., 2017; Fabbian et al., 2018; Fang et al., 2020).

In this brief introductory Section, I present an overview of halo model applications in predicting the matter clustering and the several LSS observables like galaxy weak lensing and thermal Sunyaev-Zel'Dovich.

The halo model used to predict the clustering of matter has been approached in different ways in order to maximise its reliability and its accuracy. The halo model weak spot is the transition between the two regimes (the 1-halo and 2-halo terms) in which the arising non-linear physics is not captured with sufficient accuracy by both the descending 2-halo component and rising 1-halo term. Different ways of mitigating this apparent problem have been proposed. For example, Mead et al. (Mead et al., 2015, 2016, 2021) developed a modified halo model (HMcode[33]) which introduced ad hoc parameters to grant the halo model additional flexibility and then determined those parameters by fitting the model to a large suite of cosmological simulations (CosmicEMU[34], Heitmann et al. (2016)). To tackle the non-linearity involved, solutions as the use of modification of the terms using effective field theory (EFT[35], Philcox et al. (2020); Sullivan et al. (2021)) or a Zel'Dovich approach (Mohammed & Seljak, 2014) have shown interesting results. The assumption of the conservation of cosmological laws (i.e., mass conservation) may reduce, as well, the tension between the halo model predictions and the simulations results (Chen & Afshordi, 2020).

Extensions of the standard approach include considering the cosmic voids, as presented in Voivodic et al. (2020), deviations from the spherical symmetry (Smith & Watts, 2005), addition of secondary terms of substructure (Cooray & Hu, 2002; Sheth & Jain, 2003) or exclusion terms (van den Bosch et al., 2013) and the inclusion of massive neutrino as in Massara et al. (2014).

[33] https://github.com/alexander-mead/hmcode, for the updated 2020 version.
[34] https://github.com/lanl/CosmicEmu.
[35] https://effectivehalos.readthedocs.io/en/latest/.

It is relatively straightforward to adjust the halo model to incorporate the impact of baryons. The free parameters associated with the baryon physics can either be constrained by cosmological hydrodynamical simulations (e.g.,Rudd et al. (2008), Fedeli (2014); Fedeli et al. (2014), Semboloni et al. (2011, 2013) and Mead et al. (2020, 2021)), external observations (Debackere et al., 2020), or they can be marginalised over when fitting the cosmological data set (e.g., cosmic shear) to jointly fit cosmological and baryon parameters.

The knowledge and the chance to explore cosmological extensions with the matter power spectrum with great accuracy is the basis of its extensive use in weak lensing analysis. The halo model has been applied in cosmic shear analysis (Giocoli et al., 2010; Schneider et al., 2019; Giocoli et al., 2017, 2020), tomographic weak lensing (Köhlinger et al., 2017; Heymans et al., 2021) and cross-correlations between different probes to explore further the parameter space and constrain the growth of structure (Robertson et al., 2021)

The halo model has been extensively employed to model the tSZ effect, yielding constraints on both cosmology and baryon physics. This observable is sensitive to physics involved in the characterisation of the hot free gas in massive haloes which are influenced by both supernova and AGN feedback (McCarthy et al., 2014; Battaglia et al., 2012a,b). Different extensions of the standard ΛCDM are also tested using this probe (Horowitz & Seljak, 2017; Bolliet et al., 2018).

In recent years, this approach has been used to test and predict the kinetic Sunyaev-Zel'dovich effect features (Giri & Smith, 2020). The kinetic Sunyaev-Zel'Dovich is an interesting next-generation observable which could tell more about the internal energy budget of haloes and the internal dynamics, next-generation observations will have enough (spectral) resolution to observe and identify coherent and turbulent gas motion that will shape this feature. Already, using stacking methods, it is possible to observe and measure this quantity but it is, still, in its early days.

In conclusion, the halo model ability to predict different baryonic and cosmological influences on LSS observables has been valuable in a number of works considering the cross-correlations between the tSZ effect and gravitational probes such as CMB (Battaglia et al., 2015; Hill & Spergel, 2014; Robertson et al., 2021; Koukoufilippas et al., 2020) and galaxy weak lensing power spectrum (Hojjati et al., 2017; Osato et al., 2018; Shirasaki et al., 2020).

1.5 This Thesis

In this thesis, I present a critical assessment of the internal accuracy of the halo model. This is achieved by specifying the ingredients of the halo model using cosmological simulations (the BAHAMAS), such that there are no tunable parameters remaining, and then testing how well the halo model reproduce the clustering of matter in the simulations. I explore how the different choices of the halo mass definition affect the results and the impact of baryons. After having explored these topics, I use the halo model to examine large-scale structure observables via auto and cross-correlations (such as cosmic shear, tSZ and CMB weak lensing) and I test how accurately the halo model matches observational and simulation-based results and their sensitivity to baryon physics.

Chapter 2

Methods

In this second Chapter of the Thesis, I summarise and highlight the main components behind the halo model formalism for matter clustering and LSS probes. Most of the content presented here is gathered from seminal papers on the topic, including Peacock & Smith (2000),Cooray & Sheth (2002) and Mead et al. (2015, 2016).

2.1 Halo Model formalism

I have introduced the main halo model applications in Section 1.4, while here I focus onto the mathematical formalism behind this tool.

The halo model describes haloes as the main component of the Universe, with some assumptions:

- all particles belong to haloes (no matter outside the halo boundaries, Cooray & Sheth (2002));
- the haloes have a spherical symmetry;
- it is possible to describe the clustering in two independent terms;
- the matter (or the observable, as pressure or hot gas) can be described as a continuous function dependent on the radial distance from the centre (e.g. spherical radius), the mass of the halo and redshift (Cole & Kaiser, 1988; Seljak, 1998, 2000; Battaglia et al., 2012b; Mead et al., 2015).

The halo model formalism often makes use of the Limber approximation (Limber, 1954) on a flat-sky limit. The Limber approximation describes the clustering of haloes and it

has been proved to provide a good results for $\ell > 100$ (Loverde & Afshordi, 2008; Lemos et al., 2017). At smaller multipoles, so very large scales, this approximation may fail but in this analysis I am interested in multipoles in a range from $100 < \ell < 10000$ (or in k-space above $0.001\,h^{-1}$Mpc).

The general prescription for the standard halo model can be summarised in the contribution of the two independent components as described from the Fourier transform of the 2-point correlation function (Cole & Kaiser, 1988):

$$\begin{aligned} C(\ell)^{\rm tot} &= C(\ell)^{\rm 1h} + C(\ell)^{\rm 2h} \ , \\ P(k)^{\rm tot} &= P(k)^{\rm 1h} + P(k)^{\rm 2h} \ ; \end{aligned} \tag{2.1}$$

where the first term (with notation 1h) describes the clustering of particles within a single halo (also called intra-halo clustering and 1-halo term) while the second term describes the clustering of particles belonging different haloes (2-halo term, Smith et al. (2011)). The formulation presented in eqn. 2.1 differentiates from the angular clustering defined as $C(\ell)$ and power spectrum defined as $P(k)$. In this introductory section I will carry both $C(\ell)$ and $P(k)$ notations as an overview, then I will focus on each of those independently.

The two terms can be described generally as:

$$\begin{aligned} C(\ell)^{\rm XY}_{\rm 1h} &= \int dz \frac{d^2V}{dz d\Omega} \int dM \frac{dn(M,z)}{dM} |\tilde{X}_\ell(M,z)\,\tilde{Y}_\ell(M,z)| \ , \\ C(\ell)^{\rm XY}_{\rm 2h} &= \int dz \frac{d^2V}{dz d\Omega} P_{\rm lin}(k,z) \left[\int dM_{\rm X} \frac{dn(M_{\rm X},z)}{dM_{\rm X}} b(M_{\rm X},z) |\tilde{X}_\ell(M_{\rm X},z)| \right] \\ &\quad \times \left[\int dM_{\rm Y} \frac{dn(M_{\rm Y},z)}{dM_{\rm Y}} b(M_{\rm Y},z) |\,\tilde{Y}_\ell(M_{\rm Y},z)| \right] \ ; \end{aligned} \tag{2.2}$$

$$\begin{aligned} P(k)^{\rm XY}_{\rm 1h} &= \int dM \frac{dn(M,z)}{dM} |\tilde{X}_k(M,z)\,\tilde{Y}_k(M,z)| \ , \\ P(k)^{\rm XY}_{\rm 2h} &= P_{\rm lin}(k,z) \left[\int dM_{\rm X} \frac{dn(M_{\rm X},z)}{dM_{\rm X}} b(M_{\rm X},z) |\tilde{X}_k(M_{\rm X},z)| \right] \\ &\quad \times \left[\int dM_{\rm Y} \frac{dn(M_{\rm Y},z)}{dM_{\rm Y}} b(M_{\rm Y},z) |\,\tilde{Y}_k(M_{\rm Y},z)| \right] \ ; \end{aligned} \tag{2.3}$$

with the auto-correlation case defined as X=Y, otherwise is the cross-correlation case. In the 1-halo term, I do not have to differentiate the mass, which identifies the haloes, since we are doing the analysis on the same halo for both X and Y observable. In the 2-halo, instead, I need to differentiate because the signal comes from different haloes. In eqn. 2.2 there is an integral on the comoving volume dV and redshift, differently from eqn. 2.3 which, instead, is at fixed redshift. $d\Omega$ is the comoving volume of the sky, that

under Limber approximation is $d\Omega \approx 1$, meaning all sky coverage (Komatsu & Seljak, 2002). dn/dM (also called Φ) is the halo mass function (HMF), so the abundance of haloes of a certain mass at certain redshift. $b(M)$ is linear halo bias and, finally, $P_{\rm lin}(k)$ the linear matter power spectrum computed here using the CAMB software[1](Lewis & Challinor, 2006). $\tilde{X}$ is the Fourier transform of the 3D spherical profiles, u, convoluted with the Fourier transform of the top-hat window function expressed as:

$$\tilde{X}_{\ell|k}(M,z) = \frac{1}{\overline{u}} \int_0^{R_\Delta} 4\pi r^2 u(r,M,z) \frac{\sin(\hat{r})}{\hat{r}} dr \quad , \tag{2.4}$$

with $\hat{r} = \ell r \ell_s^{-1}$ or kr depending on which observable we are considering while $\overline{u}$ is the normalisation to keep $\tilde{X}$ dimensionless. ℓ_s is a scale multipole defined accordingly in each cases (see for instance in the tSZ case, Section 1.33). The Limber approximation lies in the $\sin(\hat{r})/\hat{r}$ term, that spreads the dimensionless profiles contributions, u, at several Fourier-space modes (usually to convert from k to ℓ this relation is used $k \equiv \ell \chi^{-1}$). With the formulation just presented the halo model can reproduce several different observables in both auto and cross-correlations, by replacing the spherical 3D profile provided and the correct physical normalisation accordingly (in eqn. 2.4).

In the next subsections I focus on the different elements composing the power spectrum such the halo profiles, the halo mass function and the linear bias.

2.1.1 Halo profiles

A key component in the halo model's prediction is the way that matter is distributed inside haloes; i.e., their total matter density profiles or the pressure density in the case of the tSZ. A common choice in this regard for the matter distribution, which is motivated on the basis of collisionless (N-body) cosmological simulations, is the Navarro-Frenk-White profile (NFW, Navarro et al. (1997)):

$$\rho = \frac{\rho_0}{\left(\frac{r}{r_s}\right)\left(1+\frac{r}{r_s}\right)^2} \quad , \tag{2.5}$$

where r_s is the scale radius and ρ_0 is the normalisation. The scale radius is a free parameter, whereas one can either leave the normalisation (ρ_0) free or specify it through the halo mass definition (e.g., chosen so that the mean density within $R_{200,\rm crit}$ from the simulations is 200 times the critical density). The scale radius is often recast in terms of the halo concentration, $c_\Delta \equiv R_\Delta / r_s$, where R_Δ is the radius used in the halo mass

[1] https://camb.info/. Other Boltzmann solver often used are CLASS (Blas et al., 2011), CMBEASY (Doran, 2005) and PYCOSMO (Refregier et al., 2018).

definition. The concentration is known to depend on halo mass, redshift, and the choice of cosmological parameters and various fitting functions for this behaviour have been proposed (Duffy et al., 2008; Ludlow et al., 2014, 2016; Diemer & Kravtsov, 2015). Using these fitting functions for the concentration, one completely specifies the distribution of mass within haloes given a total halo mass, redshift, and the cosmological parameters.

While the NFW profile provides a reasonably good description of the typical density profiles of collisionless simulations, it performs less well in describing the total matter density profiles in cosmological hydrodynamical simulations (Duffy et al., 2010; Dutton & Macciò, 2014; Schaller et al., 2015a,b; Sereno et al., 2016). One can generalise the NFW form to allow for additional freedom (Nagai et al., 2007):

$$\rho(r, M, z) = \rho_0 \left(\frac{r}{r_s}\right)^{\alpha} \left[1 + \left(\frac{r}{r_s}\right)^{\gamma}\right]^{-\beta} \quad , \tag{2.6}$$

where ρ_0, α,γ and β are free parameters. This parametric form is often used to model the pressure distribution of the hot gas around groups and clusters (e.g., Arnaud et al. 2010; Battaglia et al. 2012b) but would also be suitable for the mass density distribution. In principle the free parameters of the generalised NFW form are also functions of mass and redshift, which leads to an even larger number of free parameters which would be expected to have significant degeneracies.

For what matters the electron pressure profiles, that will be relevant in the tSZ modelling, the common choice is to use the generalised NFW (GNFW) as presented in eqn. 2.5 and firstly calibrated on a small subset of X-ray relaxed selected clusters (REXCESS, Arnaud et al. (2010), or more recently as Sayers et al. (2016); Ghirardini et al. (2017, 2018)). Others, as Battaglia et al. (2012b) and Le Brun et al. (2015), have calibrated this profile on numerical hydrodynamical simulations. The susceptibility of this profile on galactic-radiative feedback leads to associated uncertainties and biases, as the hydrostatic bias or selection bias, affecting the accuracy of the results and predictions (Ruppin et al., 2019; He et al., 2021).

Our approach is to allow for additional freedom relative to the original NFW form, but with fewer free parameters than in the generalised NFW case. In particular, I adopt the so-called Einasto profile (Einasto, 1965), which recent work has shown that reproduces better the matter distribution in haloes in collisionless simulations (Springel et al., 2008; Navarro et al., 2010; Dutton & Macciò, 2014; Brown et al., 2020). This is due to its additional flexibility relative to NFW (it has an additional free parameter) which ought to allow it to better describe hydrodynamical simulations as well. The Einasto profile

can be expressed as:

$$\rho(r, M, z) = f_0(M, z) \exp\left[-A(M, z) r^{\alpha(M,z)}\right] \quad , \tag{2.7}$$

with three main parameters f_0, A and α which need to be fit for. I am also using the Einasto profile for the electron pressure profiles used in the tSZ.

2.1.2 Halo mass function

The second most important component in this formulation is the halo mass function (HMF, Φ). This quantity can be derived from analytic/semi-analytic theoretical arguments, such as those put forward by Press & Schechter (1974) and Sheth et al. (2001). However, more accurate representations can be derived from fits to large suites of cosmological simulations (e.g., Jenkins et al. 2001; Tinker et al. 2008; Bocquet et al. 2016; Despali et al. 2016; Bocquet et al. 2020; Diemer 2020).

It is commonplace to parametrise the halo mass function from cosmological N-body simulations as:

$$\frac{dn(M, z)}{dM} = f(\sigma) \frac{\overline{\rho}}{M} \frac{\ln \sigma^{-1}}{dM} \quad , \tag{2.8}$$

where the $f(\sigma)$ is a function fit to the simulation where the universal changes in redshift and σ are encapsulated:

$$f(\sigma) = A \left[\left(\frac{\sigma}{b}\right)^{-\alpha} + 1\right] e^{-c/\sigma^2} \quad , \tag{2.9}$$

with $\sigma(R)$, the rms density fluctuation in a sphere of radius R, is defined as:

$$\sigma^2(R) = \frac{1}{2\pi^2} \int P(k) \hat{W}^2(kR) k^2 dk \quad , \tag{2.10}$$

where $P(k)$ is the linear matter power spectrum, $\hat{W}$ is the Fourier transform of the top-hat window function[2] (Tinker et al., 2008). This form has been shown to reproduce the halo mass function from simulations to $\approx 10\%$ accuracy (Tinker et al., 2008; Diemer, 2020). Note that the cosmology dependence of the HMF enters in through both the σ term (which depends on the cosmology-dependent linear power spectrum) and the mean matter density ($\overline{\rho}$). The accuracy of the analytic arguments such as Press & Schechter (1974) and Sheth et al. (2001) HMFs are typically 20% with a general overprediction of the abundance of the most massive objects (Mead et al., 2015; Del Popolo et al., 2017).

[2] $\hat{W}(k) = 3x^3(\sin(x) - x\cos(x))$, where $x = kR$.

Note that these halo mass functions are typically derived from collisionless (*DM-Only*) simulations and therefore they do not account for any baryonic processes (e.g., feedback) effecting the haloes. Given the important role of galaxy formation processes in setting the mass distributions of haloes, I will make use of the BAHAMAS HMFs to build a 'correction' function to allow us to study the impact of baryons on the matter power spectrum via the halo model (see Section 3.3.1).

2.1.3 Halo bias

With the matter distribution within haloes (density profiles) and the number density (mass function) specified, the remaining ingredient is to clarify how haloes cluster in space, in terms of their 2-point correlation function (or, in Fourier space, their power spectrum). I adopt the standard assumption that haloes are linearly-biased tracers of the overall matter distribution. The linear bias is evaluated as :

$$b(k) = \sqrt{\frac{P_{\rm h,mm}(k)}{P_{\rm lin,mm}(k)}} \ , \tag{2.11}$$

where $P_{\rm hh}(k)$ and $P_{\rm lin,mm}(k)$ are the halo-halo and linear matter power spectra, respectively. On large (linear) scales $b(k) \rightarrow$const., which is what is typically referred as just the linear bias, which we use here.

In this thesis we use the linear bias-peak height relation of Tinker et al. (2010) to calculate the linear bias as a function of halo mass and redshift, with relative cosmological parameters. Please note that the peak height is defined as $\nu = \delta_{\rm crit}/\sigma(M)$ where $\delta_{\rm crit}$ is the density threshold for collapse (usually assumed to be equal to 1.686) and $\sigma(M)$ is the linear matter variance measured within the Lagrangian scale, R, corresponding to halo mass M.

A physical condition that must be met is that, when integrated over all halo masses, the bias must be unity. That is, the total matter is unbiased, by definition; i.e.:

$$\begin{aligned} \int_0^\infty b(\nu) f(\nu) d\nu = 1 \ , \\ f(\nu)d\nu = \frac{M}{\overline{\rho}_m}\frac{dn(M)}{dM}dM \ , \end{aligned} \tag{2.12}$$

(Tinker et al., 2010; Desjacques et al., 2018). This equation can also be verified in physical space by fulfilling:

$$\int_0^\infty b(M) M \frac{dn(M)}{dM} dM = \overline{\rho} \ , \tag{2.13}$$

where $\overline{\rho}$ is the mean density of the Universe at $z = 0$ and $b(M)$ and $dn(M)/dM$ are function of the total mass of the haloes.

With the linear bias all the main components of the halo model formalism are presented. Now I focus further in detail of the calculations of the matter power spectrum (Section 2.2) and LSS tracers (Section 2.3).

2.2 Matter power spectrum

The 3D matter power spectrum is, by far, the most challenging and interesting halo model application given its importance in tracing the matter distribution of the Universe. To model this quantity I use the 3D density profiles to feed eqn. 2.4 and then build the 1 and 2 halo terms as shown in eqn. 2.3 at a certain redshift.

In this quantity there are some corrections and adjustments aimed to improve and enhance, physically, the reliability of the halo model predictions. Starting with the fulfilment of eqn. 2.12. The HMF and bias are calibrated[3] to fulfil eqn. 2.12 for an infinite range of masses. This is true when, ideally, you are integrating from $[0, +\infty]$ (Schmidt, 2016; Mead et al., 2020). This lack of inclusion of small, and non-resolved, haloes from the simulations affects the ability to recover the large-scale linear power spectrum using the halo model (van Daalen & Schaye, 2015; Mead et al., 2020).

A correction to solve this issue is to add artificially the small haloes contribution. Following previous studies (Schmidt, 2016; Mead et al., 2020; Philcox et al., 2020) I add the contribution into the 2-halo term obtaining the corrective term derived from eqn. 2.12 as:

$$A_{\rm low} = 1 - \frac{1}{\overline{\rho}} \int_{M_{\rm min}}^{M_\infty} b(M) \frac{dn(M)}{dM} dM \ ; \tag{2.14}$$

where $M_{\rm min}$ is the minimum halo mass resolved in the simulation, and in BAHAMAS case I assume $\approx 4 \times 10^{11}\,{\rm M}_\odot\,h^{-1}$ (in the next chapter I will discuss further about this choice). The upper limit of integration is dependent on the size of simulation, BAHAMAS volume is large enough to have massive haloes around $4 \times 10^{15}\,{\rm M}_\odot\,h^{-1}$ which can be approximated as ∞ given the lower statistical impact of those, unique, objects in the Universe. With the above equation I construct the additive component

[3] Tinker et al. (2008) provide two different sets of parameters for the HMFs one that is built to converge on low mass haloes, $M < 10^8\,{\rm M}_\odot\,h^{-1}$, while other is not convergent. The differences in the simulation resolved haloes ($10^{11} - 10^{16}\,{\rm M}_\odot\,h^{-1}$) are below 1%.

for the 2-halo term as:

$$C_{\rm add} = \frac{A_{\rm low}\tilde{X}_k(M_{\rm min})}{M_{\rm min}} \ ; \tag{2.15}$$

where $\tilde{X}_k$ is the Fourier transform of the 3D density profiles of the lowest resolved halo mass of $M_{\rm min}$. The term in eqn. 2.15 is added to the 2-halo calculation before being multiplied by the linear matter power spectrum as:

$$P(k)^{\rm 2h} = \left[\int_{M_{\rm min}}^{M_\infty} \frac{dn(M,z)}{dM} b(M,z)|\tilde{X}_k(M,z)|dM + C_{\rm add}(M_{\rm min})\right]^2 \times P_{\rm lin}(k,z) \ . \tag{2.16}$$

As noted above, this approach guarantees that the constructed halo model reproduces the linear clustering of matter on large scales. Alternatively, a common choice is replace the 2-halo term with the linear matter power spectrum, as computed using CAMB software, and virtually obtain comparable results. The only significant difference occurs on small scales where the 1-halo is already dominant.

2.2.1 Additional (ad hoc) considerations

In this section, I will briefly explore some additional, ad hoc, adjustments of the standard halo model which have been implemented in previous works. Specifically, I follow some of the adjustments introduced in Mead et al. (2015) to avoid several non-physical artefacts in the standard halo model. These artefacts come from the Fourier expansions that spread the contributions far-off the expected range of the 1 and 2 halo terms, contributing power in scales which is not expected.

In this section I will treat the correction as suggested by Mead et al. (2015) using the dimensionless power spectra defined as:

$$\Delta^2(k) = 4\pi \left(\frac{k}{2\pi}\right)^3 P(k) \ . \tag{2.17}$$

Firstly, I apply a smooth cut-off of the 2-halo term on quasi-linear scales. As discussed in Mead et al. (2015, 2016), linear theory overpredicts the matter power spectrum on quasi-linear scales and does not accurately capture the damping of the baryonic acoustic oscillations (BAO) peaks in the particular range $k \approx 0.2 - 0.4\,h^{-1}$Mpc. Following Mead et al. (2015) (see their section 3.2.1) I, therefore, apply a tapering to the 2-halo term on quasi-linear scales using:

$$\Delta'^2_{\rm 2h}(k) = \left[1 - f\tanh^2(k\sigma_{\rm v,d}/\sqrt{f})\right]\Delta^2_{\rm 2h}(k) \ ; \tag{2.18}$$

with $\sigma_{\rm v,d}$ is the 1D linear-theory displacement variance defined as:

$$\sigma^2_{\rm v,d} = \frac{1}{3}\int_0^\infty \frac{\Delta^2_{\rm lin}(k)}{k^3} dk \ , \tag{2.19}$$

where f in eqn. 2.18 is the introduced damping factor, and $\Delta^2_{\rm lin}$ is the dimensionless linear power spectrum. Mead et al. (2016) find that f has a small dependence on $\sigma_{\rm v,d}$ as $f = 0.095\sigma^{1.37}_{\rm v,d}(z)$. Note that the application of eqn. 2.20 only affects $P(k)$ by about a percent on large scales and therefore has no significant impact on the results or conclusions of our study, but we include it for completeness.

As also discussed in the previous paper presented, but firstly presented in several papers as Cooray & Sheth (2002) (see also Smith et al. (2011); Valageas & Nishimichi (2011)), the standard (unmodified) 1-halo term displays unphysical behaviour at very large scales. In short, the 1-halo term becomes larger and larger than that predicted by linear theory on very large scales, because the halo model implicitly assumes that haloes are randomly distributed on large scales when, in reality, they tend to be more clustered and distributed more smoothly. Following Mead et al. (2015) (see section 3.2.2), I truncate the 1-halo term on large scales using:

$$\Delta'^2_{\rm 1h} = \left[1 - e^{-(k/k_*)^2}\right] \Delta^2_{\rm 1h} \ . \tag{2.20}$$

This, ad hoc, correction suppresses non-physical 1-halo power at scales $k \lesssim k_*$. Mead et al. find the value of k_* depends on the 1D linear-theory displacement variance as $k_* = 0.548\,\sigma^{-1}_{\rm v,d}(z)$.

In addition to the above modifications, Mead et al. (2015) (see also Mead et al. 2020, 2021) consider a number of other modifications of the halo model (e.g. changes in the linear collapse threshold, $\delta_{\rm crit}$, mass-concentration relation in haloes and smoothing between the 1-2 halo transition region) designed to provide a better fit to the non-linear matter power spectra of cosmological simulations. While allowing for extra degrees of freedom does allow the halo model to provide an improved fit to the simulations, one could argue that in doing so we are sacrificing the physical intuitiveness of the model for new parameters whose interpretation is ambiguous. Whether these parameters should depend on baryon physics or cosmology is also unclear. Therefore, I take a different approach and simply evaluate the accuracy (see Chapter 3) of the standard halo model, with the solely addition of eqns. (2.18)(2.20), and assess to what extent it can be reliably applied in this era of precision large-scale structure cosmology.

2.3 LSS tracers

After having presented the general description for the halo model approach on the case of the matter power spectrum I present other large-scale tracers under the halo model approach highlighting the differences whenever it is possible. I start presenting the weak lensing, since it is a direct application of the matter power spectrum, then I move on the tSZ and cross-correlations.

2.3.1 Weak lensing shear and tomography

In principle, one can again compute singularly the 1-2 halo terms using the halo density profiles, halo demographics and halo bias obtaining an equivalent weak lensing angular power spectrum using the kernel function and the source redshift distributions (Bartelmann & Schneider, 1999). However, the knowledge of the matter power spectrum, at different redshifts, allows to skip the first steps and integrate only in redshift using the appropriate weighting functions and source distributions (as discussed in Section 1.26).

As I have shown in the beginning of the weak lensing theory section, most of the information needed for weak lensing measurements lie in the lensing kernel, g, and in the non-linear matter power spectrum. In the matter power spectrum ($P(k)$) there is, already, the knowledge of the halo demographics, matter density profiles and biases while in the kernel there is the geometric knowledge of the lenses and sources.

So for a generic weak lensing shear power spectrum (defined as kk) I have:

$$C^{\rm kk}(\ell) = \int_0^{\chi_H} d\chi \frac{g^i(\chi) g^j(\chi)}{a(\chi)^2} P_{\rm mm}\left(k \equiv \frac{\ell + 1/2}{f_K(\chi)}, \eta(\chi)\right) \ ; \tag{2.21}$$

where the lensing kernel (or efficiency as seen in eqn. 1.29) is defined for a particular (i or j) source redshift distribution as:

$$g^i(\chi) = \frac{3H_0^2\Omega_m}{2c^2} \int_\chi^{\chi_H} d\chi^{'} n_\nu^i(\chi^{'}) \frac{f_K(\chi^{'} - \chi)}{f_K(\chi^{'})} \ . \tag{2.22}$$

a is the scale factor, n_ν^i is the normalised source redshift distribution as $\int d\chi n_\nu^i(\chi) = 1$, and $f_K(\chi)$ is the comoving distance of them evaluated as in eqn. 1.15 depending on the geometric curvature of the Universe. χ_H is the comoving distance of the source distribution horizon, in the special case of the CMB weak lensing that is equivalent to

the last scattering horizon (Kaiser, 1998). $\eta(\chi)$ is a coformal time at which the photon source is at a distance χ.

In the general cosmic shear weak lensing power spectrum, $g^i \equiv g^j$, that means assuming a unique source distribution. For what matters tomographic weak lensing analysis I have that $g^i \neq g^j$, meaning that different source redshift distributions are used for each bin, providing details by cross-correlating different populations of sources and distances between the sources and lenses (given the differences in the relative distances).

In weak lensing observations and analyses it is common use to adopt the angular correlation function instead of the standard Fourier version. To convert from angular power spectrum ($C(\ell)$) to angular correlation function ($\xi(\theta)$) one must do:

$$\xi^{ij}_{\pm}(\theta) = \frac{1}{2\pi}\int_0^\infty C^{ij}_{\rm kk}(\ell) J_{0/4}(\ell\theta)\ell d\ell \ , \tag{2.23}$$

with $C_{\rm kk}$ convergence shear power spectrum (eqn. 2.21), $J_n(x)$ is the nth order Bessel function of the first kind (Cole & Kaiser, 1988; Loverde & Afshordi, 2008). I have left explicit ij since this is valid for both shear and tomographic applications. The two ξ (monopole,+, and quadrupole,−) are computed using two different Bessel functions (the zeroth and fourth order respectively). To compute numerically the correlation function I have used the improved python package HANKEL[4] which allows to calculate efficiently the Hankel transform integral defined as:

$$F_\nu(k) = \int_0^\infty f(r) J_\nu(kr) r dr \ . \tag{2.24}$$

For the CMB power spectrum, one can use the convergence signal or the lensing potential ϕ_ℓ. The formulation for the individual terms of the latter can be expressed as:

$$\tilde{\phi}_\ell(M,z) = \frac{2W_{\rm CMB}(z)}{\ell(\ell+1)} \frac{4\pi R_\Delta}{\ell_s^2} \tilde{X}_\ell(M,z) \ ; \tag{2.25}$$

$W_{\rm CMB}$ is the CMB lensing kernel that can be written in terms of comoving distances as:

$$W_{\rm CMB}(z) = \frac{4\pi G \chi(z)(\chi_* - \chi(z))}{c^2 \chi_* (1+z)} \ ; \tag{2.26}$$

[4] https://github.com/steven-murray/hankel Murray & Poulin (2019).

where χ is the comoving distance, χ_* is the source comoving distance ($\chi_* = \chi(z_{\rm CMB})$) and G is the gravitational constant (Hu, 2000; Lewis & Challinor, 2006; Planck Collaboration et al., 2018). The CMB weak lensing can also be defined in terms of the lensing potential, $\phi(\hat{n})$, as:

$$\phi(\hat{n}) = -2\int_0^{\chi_*} d\chi \frac{f_K(\chi_* - \chi)}{f_K(\chi_*) f_K(\chi)} \Psi(\chi\hat{n}, \eta(\chi)) \ , \tag{2.27}$$

where $\hat{n}$ describes the direction and Ψ is the gravitational potential. This enters into the angular power spectra as:

$$C(\ell)^{\phi\phi} = \int_0^{\chi_*} \frac{d\chi}{f_K(\chi)^2} \left[-2\frac{f_K(\chi_* - \chi)}{f_K(\chi_*) f_K(\chi)}\right]^2 P_\Psi\left(k = \frac{\ell}{f_K(\chi)} \ , \eta(\chi)\right), \tag{2.28}$$

with $P_\Psi(k, \eta(\chi))$, power spectrum of the gravitational potential related with the matter density through the Poisson's equation (Hojjati & Linder, 2016; Osato et al., 2016) following:

$$P_\Psi = \frac{9\Omega_m^2(z) H^4(z)}{8\pi^2 k} P(k, z) \ . \tag{2.29}$$

Please note that to convert a lensing potential power spectrum to a convergence power spectrum you can:

$$\begin{gathered} \tilde{k} = -\frac{1}{2}\nabla \cdot \mathbf{d} = -\frac{1}{2}\nabla^2\phi \quad \text{where } \mathbf{d} = \nabla\phi \ , \\ C(\ell)^{kk} = [\ell(\ell+1)]^2 C(\ell)^{\phi\phi} \ . \end{gathered} \tag{2.30}$$

In eqn. 2.30, $\mathbf{d}$ is the deflection vector, which is related to the lensing potential ϕ and $H(z) = H_0 E(z)$ (see eqn. 1.12).

2.3.2 Thermal Sunyaev - Zel'Dovich angular power spectrum

In Section 1.1.2.3, I have explored the basic general description and use of the tSZ and its role as cosmological tracer, in this section I explore in deeper detail the halo model approach to produce the angular power spectrum.

The basic formalism is similar to what is presented for the matter clustering with the difference of the use of electron pressure profiles instead of matter density profiles. I start presenting the Compton-y parameter in terms of the electron pressure P_e:

$$y = \int n_e \sigma_T \frac{k_b T}{m_e c^2} dl = \frac{\sigma_T}{m_e c^2} \int P_e(l) dl \ , \tag{2.31}$$

in eqn. 2.31 I express the Compton-y in physical units as the integral of the electron pressure along the line of sight (l). This quantity then enters in the 3D Fourier transform as:

$$\tilde{y}(M,z) = \frac{4\pi r_s}{\ell_s^2}\frac{\sigma_T}{m_e c^2}\int_0^{R_\Delta} P_e(x_y r_{s,y}, M, z) x_y^2 \frac{\sin(\ell x_y/\ell_s)}{\ell x_y/\ell_s} dx_y \ , \tag{2.32}$$

where x_y is the normalised 3D radius defined as $x_y = r/R_\Delta$, ℓ_s is the scale multipole and it is defined as the ratio between the pressure scale length and the angular diameter distance (defined in eqn. 1.16) as follows $\ell_s = r_{s,y} D_A(z)^{-1}$.

Substituting eqn. 2.32 in the general halo model equations I can obtain the two tSZ angular power spectrum terms. Again, it is possible to use Limber approximation because I am considering a sensitive ℓ range ($\ell \gg 1$), in fact, below this limit the Limber assumption fails (Hill & Pajer, 2013).

The 2-halo term has less significant contribution, around $\approx 10\%$ of the total power, and uniquely in multipoles $\ell < 1000$ and it is related to clustering signal from different haloes (Komatsu & Kitayama, 1999; Cooray, 2001; Refregier & Teyssier, 2002; Horowitz & Seljak, 2017). The 1-halo term is, instead, dominant at all scales providing a further reason why this power spectrum is extremely interesting. Since most power comes from the correlation of the matter inside the same halo, thermal phenomena, such as AGN feedback and gas cooling, can heavily influence the electron energy distribution inside haloes and can be quantified using in this observable (Battaglia et al., 2012a,b; McCarthy et al., 2014; Ramos-Ceja et al., 2015; Khatri & Gaspari, 2016; Dolag et al., 2016; Battaglia et al., 2017).

The peak of the power spectrum distribution ($\ell \approx 3000$) is both influenced by astrophysical and cosmological parameters (Shaw et al., 2010; Trac et al., 2011; Dolag et al., 2016). In fact, following Bolliet et al. (2018), I can measure that the power spectrum is proportional to some cosmological parameters as:

$$C_\ell^{yy} \propto \sigma_8^{8.1} \Omega_m^{3.2} B^{-3.2} h^{-1.7} \ ; \tag{2.33}$$

with B, hydrostatic mass bias. The mass bias is related to the difference between the halo mass measured using weak lensing methods and SZ/X-ray as I have explored in Section 1.1.2.3, and it can have a non negligible effect onto the power spectra characterisation and accuracy. Therefore, some authors, use the hydrostatic bias as a free explicit parameter.

2.3.3 Weak lensing cross-correlations

Single probe can prove many interesting features of the collapsed objects and details of the astrophysics involved but, cross-correlating different probes can provide much more information both on the astrophysical and cosmological side. The halo model simplistic approach allows to do cross-correlations between several LSS probes, allowing to disentangle the key actors in each of the analyses, do forecast for next-generation observations and predictions on not yet studied probes (such as the upcoming use of kinetic Sunyaev-Zel'dovich).

In this thesis, I focus on cross-correlations between weak lensing measurements (galaxy lensing and CMB) with tSZ and between galaxy and CMB weak lensing.

The halo model simplistic approach allows to write the weak lensing cross correlations as follows:

$$C^{ij}(\ell) = \int_0^{\chi_H} d\chi W_i(\chi) W_j(\chi) P_{\rm mm}\left(k = \frac{\ell + 1/2}{\chi}, z\right) \quad , \tag{2.34}$$

with W_i and W_j kernel functions for each of the weak lensing probes. This is the simplified version, since the underlying matter distribution is the same and what changes is the source redshift distribution.

To do cross-correlations with tSZ observations, instead, I prefer using the general terms shown in eqn. 2.2 adapting accordingly the definition of $\tilde{y}$ (eqn. 2.32) for the pressure distribution, $\tilde{\phi}_\ell$ (eqn. 2.25) when using the CMB lensing (Hill & Spergel, 2014) or $\tilde{X}_k$ (eqn. 2.4) for convergence weak lensing (Hojjati & Linder, 2016; Osato et al., 2018, 2020; Yan et al., 2021).

2.4 BAHAMAS simulations

I employ the BAHAMAS (BAryons and HAloes of MAssive Systems) suite of cosmological hydrodynamical simulations to obtain haloes details to use in our halo model calibration and testing ground. The BAHAMAS simulations consist of $400\,h^{-1}$Mpc comoving on a side, periodic box containing 2×1024^3 particles McCarthy et al. (2017, 2018). The Boltzmann code CAMB (Lewis et al. (2000), version April 2014) was used to compute the transfer functions which were supplied to a modified version of V. Springel's code N-GenIC[5] to create the initial conditions at a starting redshift $z = 127$. The code was modified by S. Bird to include second-order Lagrangian Perturbation Theory (2LPT) and support for massive neutrinos.

[5] https://github.com/sbird/S-GenIC.

The BAHAMAS simulations were run with the Lagrangian TreePM-SPH code GADGET3 (Springel, 2005). The subgrid physics are adapted from OWLS simulations and Cosmo-OWLS (OWLS = OverWhelmingly Large Simulations, Schaye et al. 2010; Le Brun et al. 2014). These prescriptions are adapted from OWLS simulations and updated for latest results in order to match the galaxy stellar mass function at $z = 0$ and cluster baryon fraction (see McCarthy et al. (2017) for further details on the calibration). The prescriptions for the stellar components are related to star formation, evolution, metallicity evolution and supernova feedback (shocks, mass loss and chemical enrichment Dalla Vecchia & Schaye 2008; Schaye & Dalla Vecchia 2008; Wiersma et al. 2009a; Wiersma et al. 2009b). The AGN feedback and black holes mergers and accretion are adapted from Booth & Schaye (2009) and Ali-Haïmoud & Bird (2012).

While in OWLS (and cosmo-OWLS) no attempt was made to calibrate the feedback parameters to reproduce observations, the approach of BAHAMAS was to explicitly calibrate the efficiencies of the stellar and AGN feedback to reproduce the local ($z \approx 0$) galaxy stellar mass function and the gas fractions of galaxy groups and clusters. The objective in doing so was to ensure that the most massive haloes (massive galaxies up to clusters), which contribute the most to the matter power spectrum (van Daalen & Schaye, 2015; Mead et al., 2021), have the correct baryon fractions. van Daalen et al. (2020) have shown that the baryon fraction on the group scale ($M \approx 10^{14}\,\mathrm{M}_\odot\,h^{-1}$) is the key quantity in determining the impact of baryon physics on the matter power spectrum.

A standard Friend-Of-Friend (hereafter FOF) algorithm (SUBFIND, Dolag et al. (2009)) with a linking length of $b = 0.2$ times the mean interparticle separation is run to identify FOF haloes. Please note that because we first identify haloes with a FOF algorithm and then compute the spherical overdensity masses (and associated quantities), haloes cannot overlap spatially.

The BAHAMAS simulations and theirs calibrated baryonic feedback have been used to extend the standard ΛCDM parameter space to non-standard cosmologies like massive neutrinos (Mummery et al., 2017), running spectral index (Stafford et al., 2020a,b), dynamical dark energy (Pfeifer et al., 2020). In future work, there is the interest in explore further 'non-standard' cosmologies.

2.4.1 Light-cones and map-making

Simulations are a fantastic tool, however, they lack some observational features needed to make a fair and proper comparison with 'real' observations. A way to tackle these issues is to build 'observational-like' outputs from the simulations to narrow down the differences, which means computing 'light-cones'.

The way to build light-cones is to stack randomly rotated and translated simulations snapshots along the line of sight (da Silva et al., 2000) back to a certain redshift achieving the convergence of the various LSS diagnostics (for most BAHAMAS LSS results, such as tSZ, this means $z = 3$ but for CMB weak lensing we should have the knowledge of the matter distribution up to $z \approx 1100$). The way the BAHAMAS light-cones are built is presented in McCarthy et al. (2018) paper, and here I just summarise the main basic points.

The BAHAMAS simulations have 15 snapshots with different Δ in redshift as: $\Delta z = 0.125$ for $0 < z \leq 0.5$, and $\Delta z = 0.25$ from $z = 0.5$ to $z = 3$. That means that, under *WMAP*9 cosmology, $z = 3 \rightarrow \chi(z) \approx 4600\,h^{-1}$Mpc meaning that a minimum of 11 snapshots are needed to be stacked along the line of sight ($11 \times 400 = 4400\,h^{-1}$Mpc, where $400\,h^{-1}$Mpc is the simulation box size). For the maximum redshift I obtain a field that is slightly bigger than 5 degrees by $\theta_{\rm max} = L_{\rm box}/\chi(z = 3) \approx 400/4600 \times 180/\pi \approx 5.2°$. Therefore the light-cones (and consequent maps created from those) have resolution of 5×5 square degrees.

The maps which I will compare against are the tSZ maps produced computing the Compton-y parameter. This quantity is computed starting from:

$$\Upsilon \equiv \sigma_T \frac{k_b T}{m_e c^2} \frac{m}{\mu_e m_H} \ , \tag{2.35}$$

with T particle temperature, m particle mass, μ_e electron fraction and m_H hydrogen mass (Roncarelli et al., 2006, 2007). This quantity is associated to each particle in the light-cone and then a realistic map of this observable can be computed by summing all the contributions and dividing by the pixel angular size as $y \equiv \Upsilon / L^2_{\rm pix}$. For the maps created the angular pixel size is 10 arcseconds (arcsecs) which is a better resolution compared to actual tSZ telescopes (next-generation observatories like Simons Observatory will have an angular resolution of 0.5° (Ade et al., 2019) or CMB-S4 aims to get < 1.5 arcmins at 150 GHz (Abazajian et al., 2019; Carlstrom et al., 2019)). Then to create the 'observable' image they map the gas particles into a 2D grid using a simplistic nearest grid point algorithm summing all components from all particles on the line of sight.

To compare against weak lensing results I have used the convergence maps. For the 3D 'convergence' fields, commonly called $k(x)$, they compute the quantity using the matter overdensity $\delta(x)$ as :

$$2\tilde{k}(x) = \nabla^2 \psi(x) = \frac{3}{2} \Omega_m H_0^2 (1 + z) \delta(x) \ . \tag{2.36}$$

In eqn. 2.36 ψ is the local gravitational potential (in literature also often called Φ but for unambiguous definition I use ψ) that is related to the local matter overdensity. To

obtain a realistic results there is the need of considering the projection of this quantity and that converts eqn. 2.36 into:

$$k(\theta) = \frac{3\Omega_m H_0^2}{2c^2} \int_0^{\chi(z_{\rm max})} \frac{g(\chi)}{a(\eta(\chi))} \delta(\theta, \eta(\chi)) d\chi \ , \tag{2.37}$$

with $g(\chi)$ the lensing kernel defined as eqn. 1.29 and a is the scale factor (defined as eqn. 1.5).

The relation between $\eta(\chi)$ and z is linear therefore it is common use to describe this kind of equations in one of these variables. The lensing source redshift distribution $n_\nu(z|\chi)$ enters in the lensing kernel and has different shapes in case of a realistic source redshift distribution or if we compare against a single source distribution, where n_ν collapses into a Dirac delta function (as shown in Section 1.26). To compute these quantities in light-cones the standard procedure is to break these quantities into segments with fixed width of $\Delta z = 0.05$ (that is similar to imaging surveys as KiDS and DES, Harnois-Déraps et al. (2012)) as follows:

$$k(\theta) = \frac{3\Omega_m H_0^2}{2c^2} \sum_{i=1}^{N} \frac{g(\chi(z_i))}{a(z_i)} \delta_i(\theta) \Delta\chi_i \ ; \tag{2.38}$$

with the kernel described as:

$$g(\chi(z_i)) = \chi(z_i) \sum_{j=i}^{N} n_\nu(z_j) \left[1 - \frac{\chi(z_i)}{\chi(z_j)}\right] \Delta z \ , \tag{2.39}$$

where the sums are done over the i^{th} and j^{th} segments. Eqns. (2.38)(2.39) are valid only under the Born approximation, assuming small deflection angles (for further discussion see Subsection 1.1.2.1) which is enough for the purposes of these tests.

For weak lensing cosmic shear, there is the need to compute the γ_1 and γ_2 shear moments from the convergence maps. The method used is the one presented in Clowe et al. (2004) and Bahé et al. (2012) where they evaluate the Fourier transform of the complex shear as $\gamma = \gamma_1 + i\gamma_2$ more explicitly:

$$\tilde{\gamma} \equiv (\tilde{\gamma_1}, \tilde{\gamma_2}) = \left(\frac{\hat{k}_1^2 - \hat{k}_2^2}{\hat{k}_1^2 + \hat{k}_2^2} \tilde{k}, \frac{2\hat{k}_1 \hat{k}_2}{\hat{k}_1^2 + \hat{k}_2^2} \tilde{k} \right) \ , \tag{2.40}$$

where $\tilde{\gamma}$ and $\tilde{k}$ are the Fourier transform of γ and k respectively and $\hat{k}$ are the appropriate wave vectors. In real observations, there is a higher level of complexity compared to this simplistic approach. For instance, the role of intrinsic alignment of galaxies and their

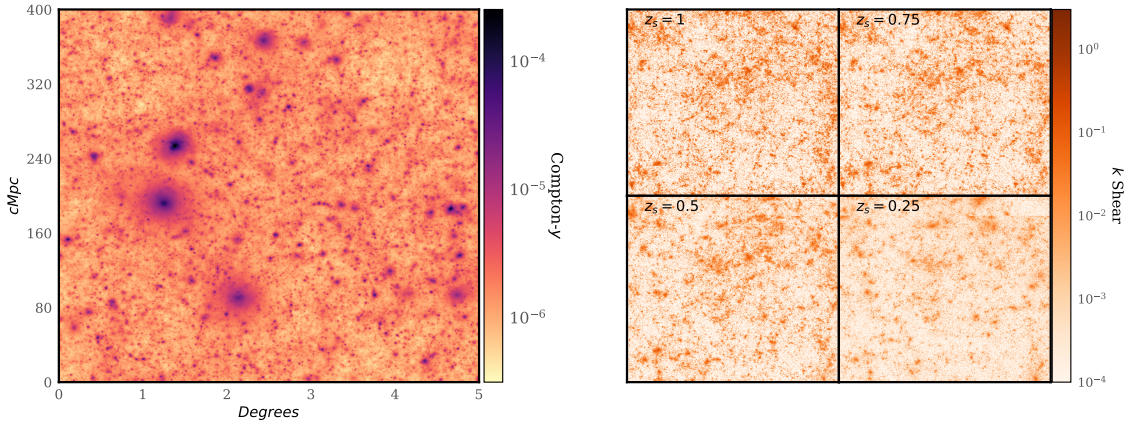

FIGURE 2.1: tSZ map from light-cone (left) and composite convergence images with four different source planes ($z_s = [0.25, 0.5, 0.75, 1.0]$) from light-cones (right). The light-cones used are the ones built from the AGN TUNED WMAP9 BAHAMAS simulations. In the left panel I show also the different scales of image in comoving Mpc on the y-axis and on the x-axis the degrees scale, in the right, for solely plotting constraint I have not shown any axis labelling.

components (as dark matter, gas and stars as shown in Velliscig et al. (2015) and Hill et al. (2021)) can alter the predictions but those effects seem to have a smaller influence on LSS statistics (Hildebrandt et al., 2017).

Before closing this Chapter, I present in Table 2.1 the cosmological parameters used in the BAHAMAS simulations for the *WMAP* and *Planck* runs, with explicit definition of the mass of the DM and GAS particles and the neutrino masses as stated in McCarthy et al. (2017, 2018) papers.

TABLE 2.1: Cosmological parameter values for the simulations used here. The columns are: (1) The summed mass of the 3 active neutrino species (we adopt a normal hierarchy for the individual masses); (2) Hubble's constant; (3) present-day baryon density; (4) present-day dark matter density; (5) present-day neutrino density, computed as $\Omega_\nu = M_\nu/(93.14 \text{ eV } h^2)$; (6) spectral index of the initial power spectrum; (7) amplitude of the initial matter power spectrum at a CAMB pivot k of $2 \times 10^{-3}\,h$ Mpc^{-1}; (8) present-day (linearly-evolved) amplitude of the matter power spectrum on a scale of 8 h^{-1}Mpc (note that we use A_s rather than σ_8 to compute the power spectrum used for the initial conditions, thus the ICs are 'CMB normalised'). In addition to the cosmological parameters, we also list the following simulation parameters: (9) dark matter particle mass; (10) initial baryon particle mass.

(1) M_ν (eV)	(2) H_0 (kms^{-1}Mpc^{-1})	(3) Ω_b	(4) $\Omega_{\rm cdm}$	(5) Ω_ν	(6) $n_{\rm s}$	(7) A_s (10^{-9})	(8) σ_8	(9) $M_{\rm DM}$ [$10^9\,h^{-1}$M$_\odot$]	(10) $M_{\rm bar,init}$ [$10^8\,h^{-1}$M$_\odot$]
***WMAP*9-based**									
0.0	70.00	0.0463	0.2330	0.0	0.9720	2.392	0.8211	3.85	7.66
0.06	70.00	0.0463	0.2317	0.0013	0.9720	2.392	0.8069	3.83	7.66
0.12	70.00	0.0463	0.2304	0.0026	0.9720	2.392	0.7924	3.81	7.66
0.24	70.00	0.0463	0.2277	0.0053	0.9720	2.392	0.7600	3.77	7.66
0.48	70.00	0.0463	0.2225	0.0105	0.9720	2.392	0.7001	3.68	7.66
***Planck*2015/$A_{\rm Lens}$-based**									
0.06	67.87	0.0482	0.2571	0.0014	0.9701	2.309	0.8085	4.25	7.97
0.12	67.68	0.0488	0.2574	0.0029	0.9693	2.326	0.7943	4.26	8.07
0.24	67.23	0.0496	0.2576	0.0057	0.9733	2.315	0.7664	4.26	8.21
0.48	66.43	0.0513	0.2567	0.0117	0.9811	2.253	0.7030	4.25	8.49
***Planck*2013-based**									
0.0	67.11	0.0490	0.2685	0.0	0.9624	2.405	0.8341	4.44	8.11
0.24	67.11	0.0490	0.2628	0.0057	0.9624	2.405	0.7759	4.35	8.11

Chapter 3

Evaluating the accuracy of the halo model in predicting the non-linear matter power spectrum

3.1 Introduction

In this Chapter, I go into detail in the accuracy tests I have applied to the 'BAHAMAS informed' halo model. The idea of testing the internal accuracy of the halo model starts from the need of assessing the reliability of this tool, improve the method and deepening the knowledge of where this model may reveal some issues that would corrupt its competitiveness.

Weak lensing studies of the large-scale structure of the Universe will, in the near future, constrain the distribution of collapsed objects with an unprecedented accuracy guiding tighter boundaries on the cosmological parameters possibly consolidating either the ΛCDM or ruling out proposed extensions. To validate and explore the parameter spaces, as I have presented in Section 1.2, there are several, equally interesting, ways with their own strengths and weaknesses. The halo model is often chosen for its simplistic approach, easy calibration and speed in obtaining results.

I explain in Section 3.2.1, how I select particles in haloes to build matter density profiles from collisionless and hydrodynamical simulations. Then, I use the, aforementioned, profiles to fit a generalised Einasto profile and I test the accuracy of these profiles in

recovering the original halo masses for the four different mass definitions and three redshifts here considered.

After presenting the profiles, I focus on the baryonic correction applied to the halo mass functions using the BAHAMAS simulations (Sections (3.3)(3.3.1)). Finally, I present the results of these applications in reproducing the matter power spectrum at different redshifts using the tabulated density profiles, the fit density profiles for the four different mass thresholds and the differences between the two cases (*DM-Only* and *AGN*). In the last section, I discuss the matter power spectrum suppression due to baryons.

Most of the content present in this Chapter is part of the submitted and accepted paper.

3.2 Informing the halo model with BAHAMAS

I start presenting the construction of the matter density profiles used in the analysis.

3.2.1 Matter density profiles

To calibrate the halo model I use the matter profiles obtained from BAHAMAS suite of simulations, described in detail in Section 2.4. I select FOF-identified haloes from $10^{11} < M_{\Delta}[\mathrm{M}_{\odot}\, h^{-1}] \lesssim 5 \times 10^{15}$ and from $0 \leq z \leq 3$, where the halo mass, M_{Δ}, is defined according to one of the four mass definitions that I present below (see Table 3.1). I extract the simulation particles attached to the FOF SUBFIND catalogues from the FOF-identified halo centres to R_{Δ}, then I select the FOF haloes in mass bins of $\mathrm{dlog}_{10}\, M = 0.125$ to create mass stacked density profiles. To create these profiles I sum all the particles inside ≈ 150 spherical shells scaled by R_{Δ} logarithmically spaced from $10^{-3}R_{\Delta}$ to R_{Δ}, where Δ is generalised to cover the four different mass definitions (see, again, Table 3.1). For a given radial bin, I compute the mass-weighted mean radius (r_w):

$$r_w = \frac{\Sigma_i m_i r_i}{\Sigma_i m_i}. \tag{3.1}$$

Note that I am able to reach such small inner radii ($10^{-3}R_{\Delta}$) because I am considering all the particles in many haloes stacked together. Note, also, that for the lowest mass haloes I consider that $10^{-3}R_{\Delta}$ can actually probe scales below the softening length of the simulations (4 pkpc h^{-1}), but this does not effect our ability to evaluate the consistency of the halo model, since the softening will also effect the power spectrum, $P(k)$, in the same way.

As already noted, I will use both the tabulated density profiles directly and parametric fits to those profiles, using an Einasto form. To allow for potential halo mass and redshift dependencies of the three main parameters in the Einasto profile (see eqn. 2.7), I model them with a simple power law dependence on both quantities, as follows:

$$f_0 = f_{0,\mathrm{int}} \left(\frac{M}{M_{\mathrm{ref}}} \right)^{f_M} (1+z)^{f_z} \ , \tag{3.2}$$

$$A = A_{0,\mathrm{int}} \left(\frac{M}{M_{\mathrm{ref}}} \right)^{A_M} (1+z)^{A_z} \ , \tag{3.3}$$

$$\alpha = \alpha_{0,\mathrm{int}} \left(\frac{M}{M_{\mathrm{ref}}} \right)^{\alpha_M} (1+z)^{\alpha_z} \ , \tag{3.4}$$

where M_{ref} is a reference mass (or pivot point) used for normalisation of the function, which I adopt as 10^{13} $\mathrm{M}_{\odot}\, h^{-1}$.

I determine the best-fitting parameters using a nonlinear least-squares Levenberg-Marquardt approach (Markwardt, 2009) with the IDL routines CURVEFIT and MPCURVEFIT using the partial derivatives with respect to each parameter to help the convergence of the fit (the partial derivatives for each parameters are explicitly presented in Appendix B.1). I simultaneously fit to the stacked density profiles over the full range of radial bins, halo mass bins, and redshifts described above.

In Table 3.1 I present the best-fitting parameters for the *DM-Only* and *AGN* total matter density profiles for the four different mass definitions. I note that there are likely to be large degeneracies between the derived parameters, but this is generally unimportant for the purposes, since I only require that the function provides a good fit to the simulated profiles for the range of halo masses, radii, and redshifts that I consider. Because of the degeneracies between the parameters, the best-fitting values themselves do not necessarily have important physical significance.

In Table 3.1 I have introduced the four different halo mass definitions that I consider, corresponding to spherical overdensities of either 200 or 500 times either the critical or mean density of the universe at a given redshift, where the critical density, $\rho_{\mathrm{crit}}(z)$, is defined as $3H(z)^2/8\pi G$ and the mean density is just $\Omega_m(z)\rho_{\mathrm{crit}}(z)$. I fit to the mass density profiles normalised by either ρ_{mean} or ρ_{crit} (depending on the halo mass definition), thus the density normalisation parameter f_0 is dimensionless, and the radial bins are normalised by the corresponding overdensity radius. For a given halo mass definition, M_{Δ}, I fit the profiles out to R_{Δ}.

TABLE 3.1: Best-fitting Einasto parameter values (see eqns. 2.7 and 3.4) describing the density profiles of the BAHAMAS *DM-Only* and *AGN* cases for four different halo mass definitions.

	Type	$M_{\rm ref}[{\rm M}_\odot\, h^{-1}]$	$F_0(\times 10^9)$	f_m	f_z	A_0	a_m	a_z	α_0	α_m	α_z
$\Delta = 200m$	*DM-Only*	10^{13}	0.3474	−0.0056	−1.804	14.01	−0.00475	−0.1891	0.298	−0.0182	−0.053
$\Delta = 200c$	*DM-Only*	10^{13}	0.2295	0.2633	−1.196	12.04	0.01594	−0.0062	0.255	−0.0522	−0.161
$\Delta = 500m$	*DM-Only*	10^{13}	0.3395	0.0025	−1.624	12.77	−0.00371	−0.1861	0.242	−0.0218	−0.096
$\Delta = 500c$	*DM-Only*	10^{13}	0.1354	0.3520	−0.869	10.42	0.0259	0.0213	0.279	−0.0711	−0.239
$\Delta = 200m$	*AGN*	10^{13}	0.5083	−0.099	0.642	16.63	−0.0115	0.006	0.189	−0.019	−0.025
$\Delta = 200c$	*AGN*	10^{13}	14504.6	−0.263	−1.76	25.29	−0.0147	−0.023	0.105	−0.032	−0.104
$\Delta = 500m$	*AGN*	10^{13}	12491.88	−0.316	−2.342	25.52	−0.0173	−0.118	0.104	−0.027	−0.115
$\Delta = 500c$	*AGN*	10^{13}	35538.27	−0.287	−1.793	25.09	−0.0152	−0.025	0.096	−0.033	−0.126

3.2.2 Tests and density profiles comparisons

To verify and quantify the ability of the profiles to reproduce accurately the total matter inside haloes, I run a test where I integrate the total density profiles over the halo spherical volume (depending on R_Δ):

$$M_\Delta^{\rm int} = \int_0^{R_\Delta} 4\pi r^2 \rho(r, M, z) \Delta\overline{\rho}\, dr \ . \tag{3.5}$$

Eqn. 3.5 shows the integral used to evaluate the total mass obtained using the profiles (stacked and fit) for the four different overdensities. $\Delta\overline{\rho}$ is needed to transform the normalised density profiles, ρ, into physical units. In figs. (3.3)(3.6) I have used eqn. 3.5 to compute the total mass and compare against the mean $M_{\rm FOF,\Delta}$ of the mass bins from simulations[1].

In fig. 3.1 and fig. 3.2 I present the *DM-Only* results comparing the stacked density profiles and the Einasto fits. In fig. 3.1 I present the results at three different redshifts ($z = [0, 1, 2]$) for three, well, populated mass bins ($\log_{10}(M) = [12.8, 13.5, 13.8]\,{\rm M}_\odot\, h^{-1}$) for the $\Delta = 200m$ overdensity. The density has been normalised by $200\overline{\rho}$ and multiplied by $(r/R_{\Delta=200m})^2$ in order to reduce the dynamic range of the plots. In the bottom panels, I present the residuals, defined as $(\rho_{\rm fit} - \rho_{\rm sim})/\rho_{\rm sim}$ with a shaded area that represents the $\pm\,10\%$ (0.1) agreement.

In fig. 3.2, instead, I present the residual plots for the four mass definitions at $z = 0$ for three different mass bins ($\log_{10}(M) = [13.0, 14.0, 15.0]\,{\rm M}_\odot\, h^{-1}$) aiming to show a broader comparison in terms of mass bins and different mass definitions.

In both figures we verify that the generalised Einasto profiles can reproduce well (within 10%) the radial trend of the profiles from 0.02 to $\approx R_\Delta$ at all redshifts and for the four mass definitions.

I evaluate the total masses using the density profiles integral (eqn. 3.5) for the four overdensity definitions. This will be relevant while considering the large-scale clustering as explained in Section 2.2.

In fig. 3.3 I present the results of the tests with the integral of stacked density profiles and the fitting profiles for the four mass definitions at three different redshifts $z = [0, 1, 2]$. In distinct shades of blue I show the four mass definitions and in the shaded region the 1% for the stacked density and 5% for the Einasto profile results. It is possible to state that the stacked density profiles can recover better than 5% the starting halo masses. This offset might be due to the fact that using mass-weighted radial bins the

[1] Using the mean or the median values of the mass bins presented less than 1% difference.

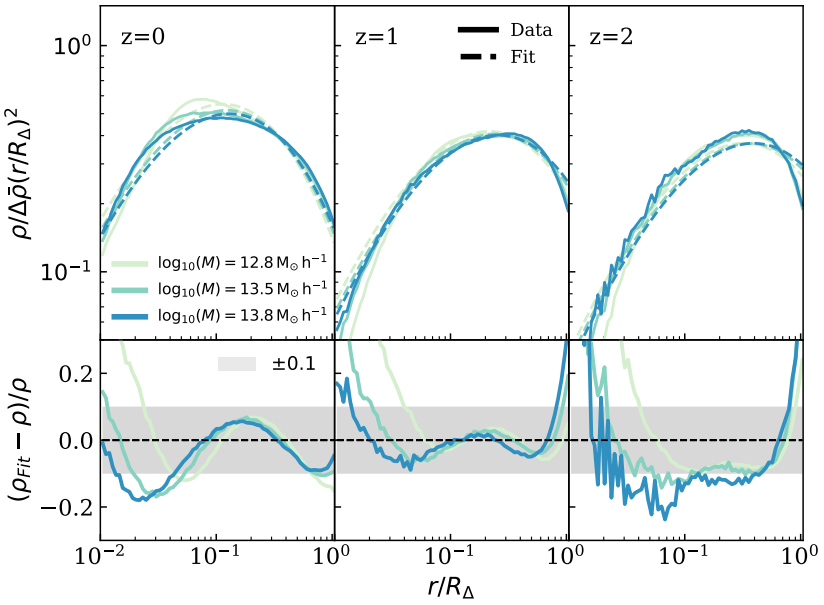

FIGURE 3.1: Stacked density (solid lines) and fit density profiles (dashed lines) for the *DM-Only* cases of three mass bins at three different redshifts ($z = [0, 1, 2]$) and the residuals for $\Delta = 200,\mathrm{Mean}$ mass definition. In the top panels I show the profiles normalised by $\Delta\bar{\rho}$ and multiplied by r/R_Δ squared to reduce the dynamic range. I have chosen the mass bins $\log_{10}(M) = [12.8, 13.5, 13.8]\,\mathrm{M}_\odot\,h^{-1}$ because they are well represented in all three redshifts. In the residual plots I see that the fitting function can reproduce the actual data at the three redshifts bins. The area grey-shaded identifies the ±10% accuracy. Overall the differences are below 10% in the radial range from 0.02 to R_Δ.

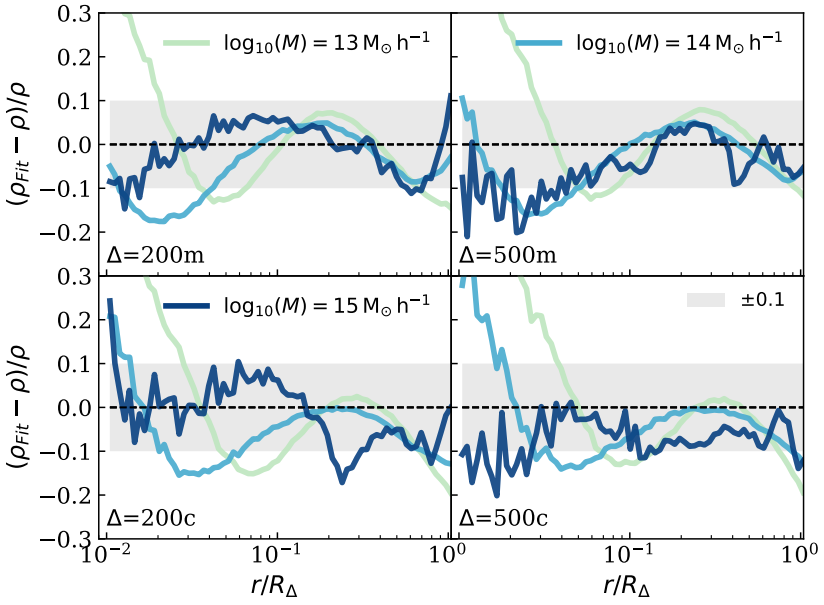

FIGURE 3.2: Extended comparisons with the *DM-Only* profiles for the four mass definitions at $z = 0$ for three mass bins ($\log_{10}(M) = [13.0, 14.0, 15.0]\,\mathrm{M}_\odot\,h^{-1}$). Here are presented in terms of residual plot defined as $(\rho_{\mathrm{Fit}} - \rho)/\rho$.

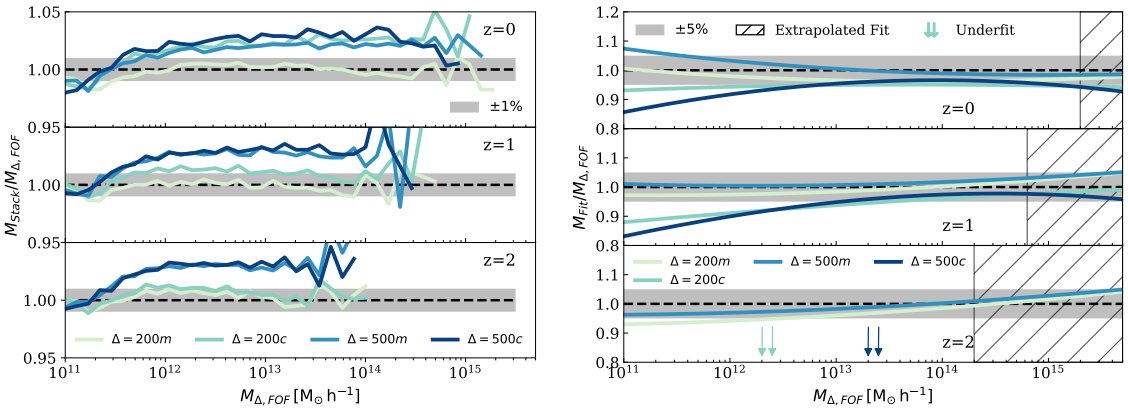

FIGURE 3.3: In left panel there are ratio plots between the integrated *DM-Only* stacked profiles compared to the original $M_{\Delta,\mathrm{FOF}}$ for the four mass definitions and at three redshifts bins ($z = 0$ top panel, $z = 1$ middle and $z = 2$ bottom). In the right panel, same ratio plot structure, but using the Einasto fitting profiles.

final integration bins are not perfectly centred at same FOF radius definitions, leading to these small mismatches. The stacked profiles using $\Delta = 200$ (lighter blue lines), both in the mean and critical case, can reproduce better than $\approx 2\%$ the total dark matter masses in a wide mass range ($10^{11} - 5 \times 10^{15}\,\mathrm{M}_\odot\,h^{-1}$) in all three redshifts. This result means that the spherical overdensities used to describe the dark matter profiles are enough detailed to reproduce the starting halo masses. The $\Delta = 500$ profiles (darker blue lines), instead, can reproduce the mass values in a slightly worse accuracy, these profiles tend to overpredict the total masses by $3 - 4\%$ evenly in redshift. Again, these differences might be related to the way the outer boundaries are defined and the way the centre of haloes are defined and scaled accordingly by R_Δ.

In case of the fitted density profiles (right plot) we can see that most of the densities are recovered better than 10% at all redshifts. The mass definitions are colour-coded as the stacked density case. The area with oblique lines is, instead, where the fit is extrapolated due to the lack of data points in the fitted data set. The Δ_{mean} mass definition tends to provide a better agreement with the expected starting halo mass at all redshifts (always below 5%) but for the case of Δ_{crit} it is possible to check that the accuracy at higher redshift ($z = 2$) is well above 20% (in the plot is highlighted with small coloured arrows showing the underfit).

A possible explanation is that the most influential mass bins ($10^{13} - 10^{14}\,\mathrm{M}_\odot\,h^{-1}$) in the stacked profiles were not enough detailed in the Δ_{crit} case at higher redshift, and this lack of definition has strongly impacted the overall agreement in all mass bins, even, in the lowest ones available. The shape of the ratio plot for the Δ_{crit} cases present a starting accuracy around 15% below $10^{12}\,\mathrm{M}_\odot\,h^{-1}$ rising up to $< 5\%$ at $\approx 10^{13}\,\mathrm{M}_\odot\,h^{-1}$

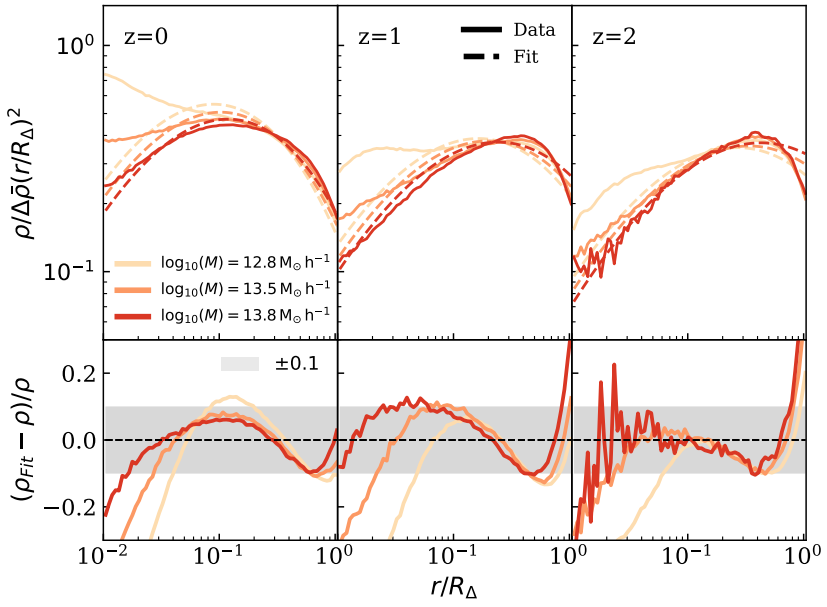

FIGURE 3.4: Same as fig. 3.1 but for the *AGN* case. I have used the same mass and redshift bins as above. To differentiate from the companion figure I have colour-coded the *AGN* case in shades of red.

and being almost constant until the last mass bins are reached. This, reinforces the fact that, if the central mass bins were not enough characterised, might have led to a worse fit in that regime.

Now I present the stacked density and the fit for the *AGN* case. For the total matter profiles I have computed each simulation particle types (DM, GAS and STAR) the stacked density profiles and I have sum them together to obtain the total matter density profiles. I found that the mass-weighted radii of each component agree better than 1% to the others so I am summing them without any changes or corrections[2].

In fig. 3.4 and fig. 3.5 I present the companion plots between the fit and the stacked density profiles for the $\Delta = 200m$ in the first figure and for the four overdensities at $z = 0$. In the first figure it is possible to see that the lowest mass bin involved ($\log_{10}(M) = 12.8\,M_\odot\,h^{-1}$ in light orange) presents a clear complex structure at all redshift in radii $r < 0.08R_\Delta$. This complex feature might be related to some stellar structure that is more relevant in this mass bin and less accurately reproduced in the Einasto fit. In the second figure, instead, I present the four different mass definitions comparison with three mass bins at $z = 0$. The fit here shows some difficulties in reproducing better the stacked density profiles, but the general agreement is always below 15%(0.13).

[2]As I have shown in Table 2.1 DM and GAS particles have different masses and while accounting for the weighted-mean radii, eqn. 3.1, that means that there was the risk of misplacement of a portion of particles. I have verified, that this is not the case, in fact, the weighted profiles match perfectly.

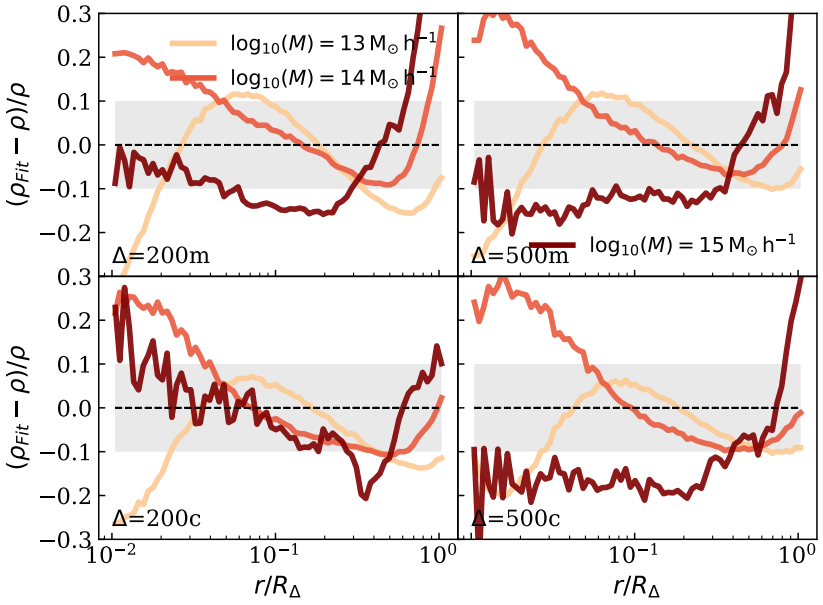

FIGURE 3.5: Companion figure of fig. 3.2 but for the *AGN* case. The same mass and redshift bins are presented here. As before, I am using a different colour-grading to show the three different mass bins involved.

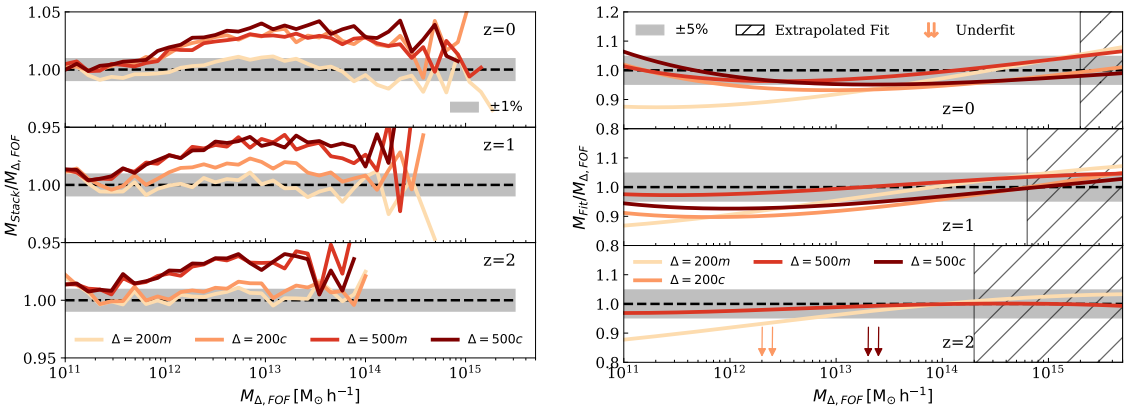

FIGURE 3.6: The figure is a companion of fig. 3.3 but for the *AGN* cases. In left panel there are the ratio plots between the integrated stacked profiles compared to the original $M_{\Delta,\mathrm{FOF}}$ for the four overdensities and at three redshifts bins ($z = 0$ top panel, $z = 1$ middle and $z = 2$ bottom). In the right panel same ratio plot structure but using the Einasto fitting profiles.

In fig. 3.6 I present the integral tests for the stacked density profiles and the fit density profiles for the *AGN* case. As was seen for the *DM-Only* case, the stacked density profiles can match the original $M_{\Delta,\mathrm{FOF}}$ with accuracy of $1-3\%$ for all redshifts and mass definitions. For the fitting profiles it is possible to see a similar behaviour as the *DM-Only* case with the mean overdensities that perform well for a wide mass ranges and redshifts while the Δ_{crit} suffers from the same issues presented in the previous case. Comparing the results from fig. 3.18 and fig. 3.6 look like that the inaccuracy of the fit, compared to the stacked profiles in the stellar components in the inner parts of the haloes have a quite small impact on the overall accuracy on the total halo mass.

In principle one could include an additional component to the parametric model to better fit the inner regions. Indeed, this would be recommended when modelling real data. However, since I also evaluate the halo model using the tabulated profiles directly from the simulations, I can still assess the accuracy of the halo model without including such a component. By comparing the tabulated and parametric versions, I can directly assess the impact of neglecting an additional component designed to better capture the central galaxy.

3.3 Halo mass function

I, now, consider the halo mass function (HMF) from BAHAMAS as input for the halo model. At a given redshift and for a given halo mass definition (spherical overdensity), I compute the HMF of FOF haloes using a bin width of $d\log_{10}(M) = 0.0625$, over a mass range $10^{11} - 5\times10^{15}\,\mathrm{M}_{\odot}\,h^{-1}$. To compute the mass function, dn/dM (called here also Φ), I simply count the number of FOF haloes in a given bin and divide by the linear bin width and simulation comoving volume.

I present in fig. 3.7 a comparison between the HMF from the BAHAMAS *DM-Only* run with the Tinker et al. (2008) prediction for the four different mass definitions at three redshifts bins. I have computed the Tinker HMF using the Colossus Toolkit (Diemer, 2018) [3].

I see that there is generally good agreement between the two independent mass functions. In the small panels below the main ones I present the residuals between the BAHAMAS and the Tinker HMFs for the three different redshifts (with the same lines). Small differences can be seen at high masses which are likely a result of cosmic variance and relatively poor statistics (Poisson errors) in the BAHAMAS volume.

[3] https://bdiemer.bitbucket.io/colossus/.

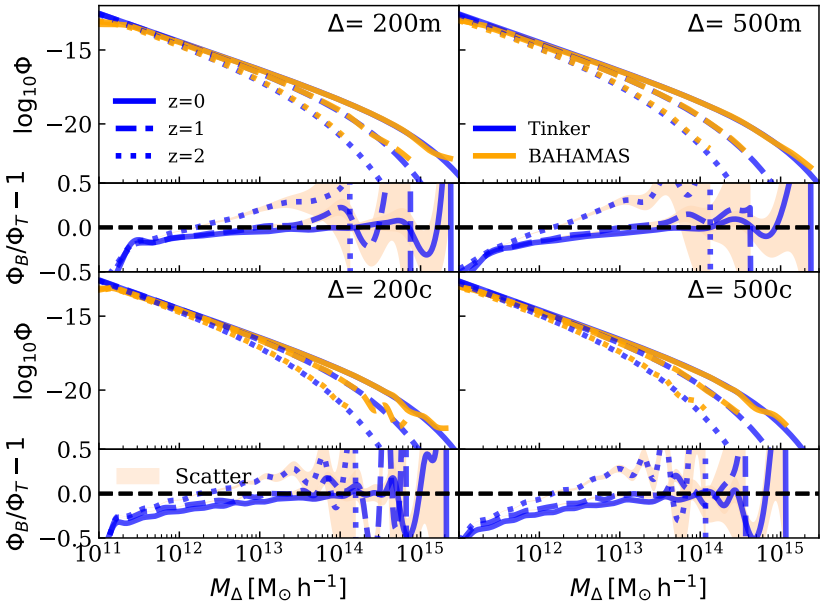

FIGURE 3.7: Halo mass function (HMF, $\Phi = dn/dM$ [$\mathrm{Mpc}^{-3}h^4\mathrm{M}_\odot^{-1}$]) comparison between the BAHAMAS *DM-Only* case (Φ_B, orange curves) and the Tinker et al. (2008) prediction (Φ_T, blue curves) at three redshifts bins (solid, dashed and dotted curves) for four halo mass definitions. In the smaller subpanels residuals between the BAHAMAS HMF and the Tinker HMF are shown. The shaded orange regions represent the Poisson errors for the BAHAMAS HMFs. Differences between the Tinker and BAHAMAS mass functions are likely due to cosmic variance and Poisson uncertainties at the high-mass end and finite resolution and differences in how haloes are identified (FOF for BAHAMAS and spherical overdensity for Tinker) at the low-mass end (see text). I examine how the differences in the mass functions affect the resulting non-linear power spectrum in fig. 3.13.

Regardless of the origin of the differences, they should be taken into account evaluating the internal accuracy of the halo model. For example, if through cosmic variance the BAHAMAS volume has somewhat more very massive clusters than expected on the basis of the Tinker HMF, this could also affect the overall non-linear $P(k)$ of the simulation. Therefore, by using the actual HMF from BAHAMAS I can more accurately test the halo model formalism.

At low masses ($\sim 10^{11}\,\mathrm{M}_\odot\,h^{-1}$), the BAHAMAS simulations predict a lower abundance of haloes compared to the Tinker expectation. This is likely due to two effects: finite resolution of the BAHAMAS simulations and differences in the way haloes are identified in BAHAMAS and Tinker et al. (2008). There is a clear resolution effect at masses below $\approx 3 \times 10^{11}\,\mathrm{M}_\odot\,h^{-1}$, where the BAHAMAS HMF stops increasing with decreasing mass. Here the simulations are approaching the 20 particle limit imposed on FOF groups. At somewhat higher masses, there is still a deficit with respect to the Tinker prediction of $\approx 10 - 20\%$. This is likely due to differences in the way haloes are identified.

For BAHAMAS, haloes are identified with a FOF algorithm after which spherical overdensity masses are computed, whereas Tinker et al. (2008) identify haloes using the spherical overdensity method and haloes are allowed to partially overlap (Tinker et al., 2008). Consequently, more intermediate/low mass haloes are identified in the vicinity of larger haloes using the spherical overdensity method, whereas a FOF algorithm will combine haloes into larger group in which they are sufficiently close to one another. These differences have been previously discussed in the literature (e.g., Bocquet et al. 2020) so I will not discuss them further here. However, such differences in the HMFs will propagate through the halo model and affect the predictions for $P(k)$. I will show that the differences in the HMFs will impact the $P(k)$ predictions only slightly at low redshift, but play a relatively larger role at higher redshift ($z \approx 2$).

3.3.1 Role of baryons on the HMFs

In the HMF comparison presented above, I examined the *DM-Only* run from BAHAMAS and compared it with the predictions of Tinker et al. (2008), who used a large suite of collisionless (dark matter-only) cosmological simulations to calibrate an approximately universal form (to $\sim 10\%$ accuracy) for the HMF (see eqn. 2.8). Thus, the comparison was a consistent one. However, as several authors have shown previously, the halo profiles and HMFs can be affected by baryonic processes such as feedback from supernovae and AGN (Cui et al., 2014; Velliscig et al., 2014; Bocquet et al., 2016; Mummery et al., 2017; Pfeifer et al., 2020; Stafford et al., 2020a), with effects as large as 20% in the HMF which is large enough to have a non-negligible impact on cosmological parameter inference (Cusworth et al., 2014; Castro et al., 2021; Debackere et al., 2021).

To evaluate the impact of baryons on the HMF and how these translate to predictions of the halo model, I extend the formalism presented in Velliscig et al. (2014) to correct the masses and HMFs. I explore two separate ways of accounting for baryons in the HMF. In the first case, I can exploit the fact that the *DM-Only* and *AGN* runs have the same phases in the initial conditions, making it possible to match haloes between the two runs (using the unique particle IDs) on a halo-by-halo basis, as done previously in Pfeifer et al. (2020) and Stafford et al. (2020a) when evaluating both the impact of baryons and cosmological extensions (dynamical dark energy and a running of scalar spectral index, respectively) on the HMF in BAHAMAS. With this approach, one can directly determine how the halo mass has changed as a result of baryonic processes. In appendix A.2 I briefly explain the matching technique by using particles IDs from two separate simulations. In the second approach, one can simply compare the HMFs of the *DM-Only* and *AGN* runs, effectively computing the ratio of abundances in a given halo mass bin (e.g., Velliscig et al. 2014). By default I use the halo matching scheme

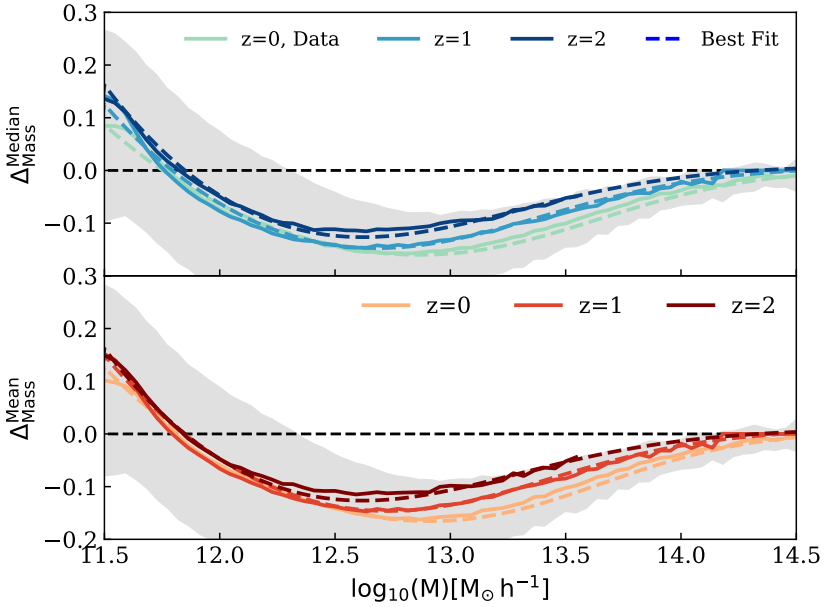

FIGURE 3.8: Fractional change in halo mass (eqn. 3.6) between the BAHAMAS *AGN* and *DM-Only* simulations as a function of *DM-Only* halo mass. Shown are the median (top panel) and mean (bottom panel) trends, along with the best-fit functions (eqn. 3.7), for the $\Delta = 200m$ case at three different redshifts. The scatter, shown in shaded grey area, represents the 68% confidence region at $z = 0$.

to derive a HMF correction factor, but I have also explored the results using the HMF ratio method (see appendix A.3). In short, while both approaches yield similar results, I find the halo matching scheme to be more accurate (less noisy).

In fig. 3.8 I present a comparison between the mean and median values of $\Delta_{\rm mass}$ at three redshifts bins for the $\Delta = 200m$ case, where $\Delta_{\rm mass}$ is the fractional difference in the halo mass between the *AGN* and *DM-Only* runs:

$$\Delta_{\rm mass} = \left(\frac{M_{AGN} - M_{DM\text{-}Only}}{M_{DM\text{-}Only}} \right) \quad . \tag{3.6}$$

Consistent with previous studies, I find that the halo masses are most strongly affected on the scale of galaxy groups ($10^{13} - 10^{14}\,{\rm M}_\odot\, h^{-1}$), where AGN feedback is able to expel a large fraction of the baryons. At higher masses (above a few $10^{14}\ {\rm M}_\odot\, h^{-1}$) the increased binding energy of the haloes prevents significant gas expulsion, while at lower masses ($\lesssim 10^{12}\,{\rm M}_\odot\, h^{-1}$) AGN feedback is generally not yet active and stellar (supernova) feedback is not sufficiently energetic to eject a significant amount of baryons.

Following Velliscig et al. (2014), I model the change in halo mass (mass shift) due to baryons with the following functional form:

TABLE 3.2: Best-fitting parameters for the baryonic mass correction (eqn. 3.7) for the median and mean fractional changes in halo mass, $\Delta_{\rm mass}$. In fig. 3.8 I show the comparison between our best-fit models and the median and mean trends.

	$\Delta_{\rm Mean}$		$\Delta_{\rm Median}$	
P	a_0	a_z	a_0	a_z
A	22291.3	0.260	19539.1	0.417
B	−0.327	0.289	−0.312	0.414
C	−13.574	−0.033	−13.552	−0.037
D	−0.395	0.218	−0.417	0.216

$$\Delta_{\rm mass}(M_{DM\text{-}Only}) = \frac{A}{\cosh[\log_{10}(M_{DM\text{-}Only})]} + \frac{B}{1+\exp\left[-\frac{\log_{10}(M_{DM\text{-}Only})-C}{D}\right]} \quad (3.7)$$

Note that in eqn. 3.7 I have added a hyperbolic cosine term that allows the function to better reproduce the increase in $\Delta_{\rm mass}$ towards low halo masses. In addition, to account for the redshift evolution of the halo mass shift, I allow the four parameters (A,B,C and D) to have power law redshift dependencies, e.g.,:

$$A(z) = a_0(1+z)^{a_z} \ . \quad (3.8)$$

To improve the fit I have computed the partial derivatives of each parameters of eqn. 3.7 and I have listed them in Appendix B.2.

In fig. 3.8 I see that the fitting functions can reproduce the halo mass shift for all mass bins ($> 3\times10^{11}$ $\rm M_\odot\, h^{-1}$) and at the three different redshifts shown. Note that the lower limit of $3\times10^{11}\,\rm M_\odot\, h^{-1}$ is dictated by the minimum number of matched most-bound particles (50) that I require to match haloes between two BAHAMAS runs.

Overall, the accuracy of the best-fitting functions to the mean and the median values of $\Delta_{\rm mass}$ is better than 10% in all mass bins and redshifts sampled. In Table 3.2 I present the best-fitting parameter values for the mean and median versions of the mass shift $\Delta_{\rm mass}$.

As an aside, I find that the effects of baryons on the HMF (and presumably density profiles as well) are slightly cosmology dependent. I have determined this by testing our model against a BAHAMAS *Planck*13 run that has a different universal baryon fraction ($f_b^{\rm Planck} = 0.15433$) and I have found that the first fitting parameter, A, should be corrected by a factor of 0.87 to account for the different value of f_b. We remind that the

value of f_b is 0.165 for our WMAP9 cosmology run. This cosmology correction works for both the mean and median $\Delta_{\rm mass}$ results. In Appendix A.3 I present a similar analysis for what concerns the cosmology dependence on the direct HMF baryonic correction.

While the halo mass correction procedure derived above could be applied on a halo-by-halo basis to the BAHAMAS *DM-Only* run to derive a baryon-corrected HMF, such a procedure would generally not be possible for published HMFs based on collisionless simulations, since the individual halo masses (halo catalogues) are generally not available. Thus, I wish to derive a simple correction factor that can be applied to existing collisionless HMFs in the literature.

To do this, I first shift the halo mass bins from a collisionless HMF (in this case the BAHAMAS *DM-Only* HMF) using the baryonic correction procedure above. I use the mean correction function in Table 3.2. This creates a new set of mass bins. I, next, rescale the abundances (Φ) by the relative ratio between the *DM-Only* and *AGN* mass bins as:

$$\frac{dn}{dM_{AGN}} = \frac{dn}{dM_{DM\text{-}Only}} \frac{(M_{DM\text{-}Only,i} - M_{DM\text{-}Only,i-1} \equiv dM_{DM\text{-}Only})}{(M_{AGN,i} - M_{AGN,i-1} \equiv dM_{AGN})} \quad , \tag{3.9}$$

where M_{AGN} is the corrected halo mass, derived using the (uncorrected) mass, $M_{DM\text{-}Only}$, and eqns. 3.6 and 3.7, and i refers to the i$^{\rm th}$ mass bin. Essentially, this procedure works because the number of haloes does not change as a result of feedback/baryons, only their masses change. This means that both the x-axis (halo mass) and the y-axis ($\phi \equiv dn/dM$, through the change in dM) change due to baryons[4]. With this procedure, I can correct existing HMFs derived from collisionless simulations for the presence of baryons.

In fig. 3.9 I present the result of the application of the method shown before on the BAHAMAS *DM-Only* HMFs with respect the *AGN* version. It is possible to see that the method provide a $\approx 5\%$ accuracy in most mass bins. Larger deviations are present at the very highest masses, which are due to poor sampling statistics. The differences using the mean or median values of $\Delta_{\rm mass}$ are small and have a slight effect on the largest masses regime.

In fig. 3.10 I present a comparison between the baryon-corrected Tinker HMFs and the BAHAMAS *AGN* HMFs, as I have done for the *DM-Only* case in fig. 3.7. I can see that the method applied also provides good results in this case.

Concluding this section on the halo mass function I remind that the formalism of the linear bias is the same both for the *DM-Only* and *AGN* cases, as supported by recent

[4] An underlying assumption of this procedure is that the *rank ordering* of haloes by mass does not change through the inclusion of baryons.

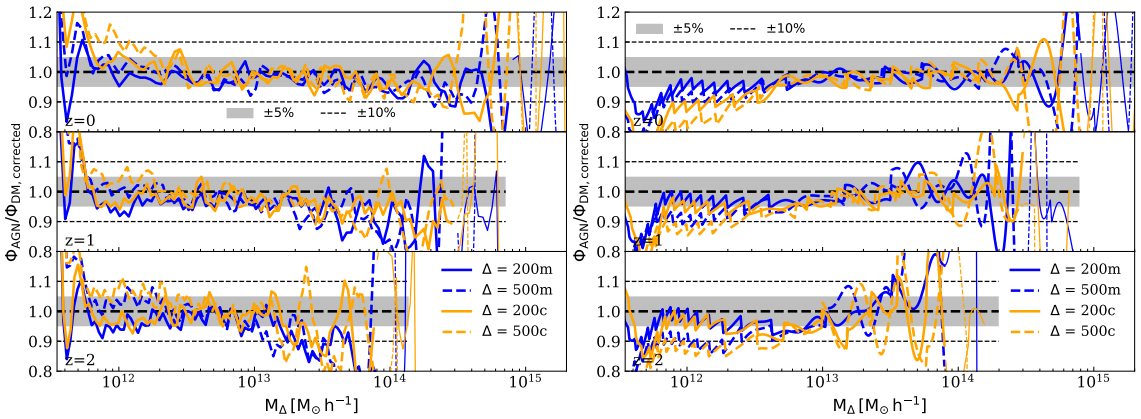

FIGURE 3.9: Ratio plots between the BAHAMAS HMFs for the *AGN* case over the corrected *DM-Only* using the baryonic correction using the mean (left plot) and median (right) values. I present the three redshifts bins and all four mass definitions (orange for the critical mass definitions and blue for the mean ones). I can see that the baryonic correction helps recover the HMFs better than 5% in most cases and all redshifts with a better agreement on the regime above $5 \times 10^{13}\,M_\odot\,h^{-1}$. The huge spikes visible in the high mass regime is due to the variance of the ratios and the small differences between the two cases.

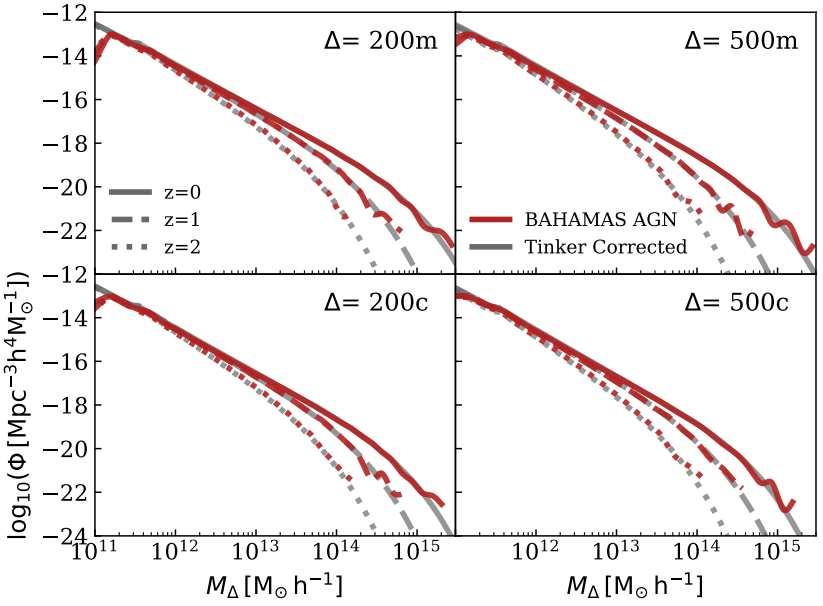

FIGURE 3.10: Same as fig. 3.7 but for the *AGN* case, with the Tinker HMF (grey curves) corrected using eqn. 3.7 and compared with the BAHAMAS *AGN* HMFs (red curves).

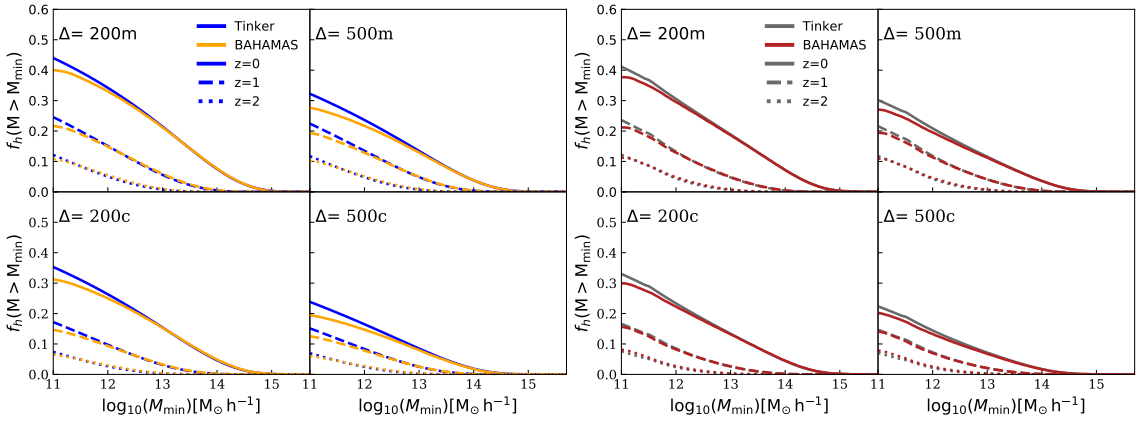

FIGURE 3.11: Halo fraction comparison between the Tinker and BAHAMAS HMFs as function of the M_{min} in terms of mass definitions and for the three redshifts. In the left plot, I present the *DM-Only* version with Tinker results in blue lines and BAHAMAS in orange with solid line for z=0, dashed for z=1 and dotted for the last z=2. In the right-hand side plot, I present the *AGN* version with Tinker, mass corrected, in grey and BAHAMAS in red.

work as Castro et al. (2021). Therefore I have not modified the linear bias also when working with the *AGN* matter power spectrum by the baryon effect. Also, I will use the standard linear bias when I will use the BAHAMAS HMF because the differences between the latter and the Tinker HMf is small and the impact should be small as well.

3.3.1.1 Halo fraction of the total mass

As pointed out in van Daalen & Schaye (2015), it is important to understand the share of mass that is inside haloes compared to the total amount of mass that there is in the Universe. That accounts for the overall 2-point correlation function that guides the clustering at large scales ($k < 0.1\, h\mathrm{Mpc}^{-1}$). A method to quantify the shared quantity is to integrate the HMFs between a M_{min} to a maximum mass (which I set to $10^{16}\, M_\odot\, h^{-1}$). I compare this quantity, called f_h halo fraction, between the different mass definitions and redshifts analysed in this thesis and for the *DM-Only* and *AGN* cases. f_h is defined, for clarity, as function of M_{min} as:

$$f_h(M_{min}, z) = \frac{1}{\bar{\rho}} \int_{M_{min}}^{\infty} M\phi(M, z) dM \quad . \tag{3.10}$$

In fig. 3.11 I present the four mass definitions estimates of f_h for both *DM-Only* (left-hand side plot) and *AGN* cases (right-hand side plot). I present the three redshifts bins as solid, dashed and dotted lines ($z = [0, 1, 2]$). In orange and red I have integrated the BAHAMAS HMFs while in blue and grey the Tinker HMFs (in the *AGN* cases I

have used the one baryonically corrected). It is possible to examine that for the limited range available ($10^{11} - 10^{16}\,\mathrm{M}_{\odot}\,h^{-1}$, I define the upper limit ∞ since the small statistical impact of this) the mass shared in haloes accounts only for 0.43 for Tinker at z=0 in the *DM-Only* case and 0.41 for the *AGN* case. Changing mass definition, also reduces that value to a mere 0.25 for the $\Delta = 500c$ for the *DM-Only* case and ≈ 0.22 for the *AGN* one.

At higher redshifts the halo fraction goes down below 0.3 (z=1) for $\Delta = 200m$ and 0.12 (z=2). This results show that there is huge portion of mass that is either not in haloes, as found out by van Daalen & Schaye (2015), which prevents the halo model approach and to be thoroughly descriptive to the reality of the Universe. Simulations, as well, contain a significant number of unresolved objects that contributes in the overall matter power spectrum.

3.4 Matter power spectrum

In the previous sections I have explained how I extract the ingredients necessary from the BAHAMAS simulations to be able to evaluate the halo model. I have shown the accuracy with which the Einasto fits reproduce the stacked density profiles computed from the simulations. I have also examined the differences in the HMFs, deriving a baryon correction factor that can be applied to HMFs from collisionless simulations. Below I apply these quantities to calculate the non-linear power spectrum, $P(k)$, using the halo model and I compare this with the actual power spectrum measured from the BAHAMAS simulations.

3.4.1 Collisionless matter power spectrum

In this section I present a comparison of the (BAHAMAS-informed) halo model predictions for the non-linear matter power spectrum alongside power spectrum predictions from the BAHAMAS simulations themselves. I also show linear theory prediction computed by the software CAMB (Lewis & Challinor, 2006) and the non-linear power spectrum from the (collisionless) HALOFIT package (Takahashi et al., 2012). Note that HALOFIT provides a non-linear correction factor for the linear power spectrum, which Takahashi et al. (2012) have derived by fitting to a large suite of collisionless simulations spanning a wide range of cosmologies.

I begin by presenting the results for the collisionless (*DM-Only*) case. As already discussed, I explore different versions of the halo model, where, for the density profiles I

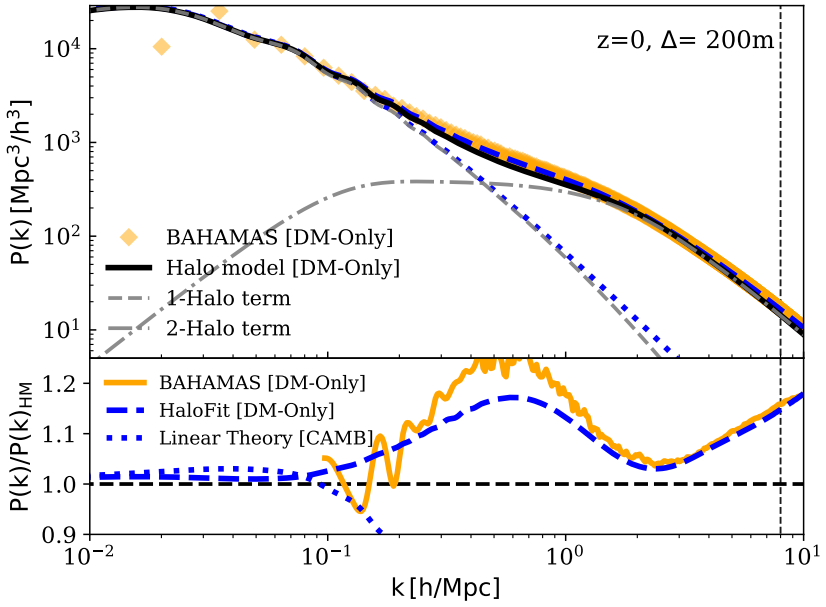

FIGURE 3.12: *T*op: Matter power spectra comparison for the collisionless case. The solid black curve represents the halo model computed with the stacked density profiles and the BAHAMAS HMF, while the dashed and dot-dashed grey curves represent the 1-halo and 2-halo terms separately. The power spectrum from the BAHAMAS *DM-Only* run is represented with orange symbols. The dashed and dotted blue curves represent the non-linear power spectrum predicted by HALOFIT and the linear matter power spectrum predicted by CAMB. The vertical dotted line represents 0.5 times the Nyquist frequency of the BAHAMAS simulation volume. *B*ottom: Ratios with respect to the halo model. On very large scales, the halo model reproduces linear theory to percent level accuracy, by construction. On small scales, the 1-halo term dominates and reproduces the simulated (BAHAMAS) power spectrum to typically 5% accuracy. In the 1-halo/2-halo transition region, the halo model predicts up to 20% less power than in the BAHAMAS simulations.

use either the tabulated profiles extracted directly from the simulations or a smooth parametric fit to them and for the halo mass function I use either the HMF directly from the simulations or forms from the literature (specifically Tinker et al. 2008). I also explore the impact of changing the halo mass definition, by varying the overdensity criteria used to define a halo's mass and its radial extent.

In fig. 3.12 I present the comparison between the halo model prediction (black solid line), its 1-halo and 2-halo terms (grey lines) and the BAHAMAS power spectrum (orange diamonds), as well as the predictions of linear theory and HALOFIT (dotted and dashed blue curves, respectively). Note that for this comparison, the halo model is computed using the tabulated mass density profiles (as opposed to Einasto fits to them) and the HMF directly from the BAHAMAS *DM-Only* run. In the bottom panel I present the ratio of the different cases with respect to the BAHAMAS-informed halo. The BAHAMAS simulation power spectrum is computed using the software NBODYKIT[5](Hand

[5]https://nbodykit.readthedocs.io/en/latest/.

et al., 2018).

Qualitatively speaking, the halo model does capture the general trends of the simulation non-linear power spectrum well, including the shape of the power spectrum from the simulations (top panel). For example, there is a strong increase in power with respect to linear theory on small scales, as expected. Focusing on the bottom panel for a quantitative comparison, I can see that the halo model predictions match those of linear theory at large scales to percent level accuracy, which is by construction, after accounting for haloes that lie below the mass resolution limit of the simulations (see eqn. 2.16). Note that the application of eqn. 2.18 (2-halo truncation) only affects $P(k)$ by about a percent on large scales and therefore has no significant impact on the results or conclusions of our study, but I include it for completeness.

The most challenging region is between $0.1 \lesssim k\,[h\,\mathrm{Mpc}^{-1}] \lesssim 2$ which corresponds to the transition region between the 1-halo and 2-halo terms. Here the halo model's prediction can deviate from the simulations by up to 15%. This is qualitatively consistent with previous findings (e.g., Giocoli et al. (2010), Massara et al. (2014), Mead et al. (2015),Chen & Afshordi (2020) and Voivodic et al. (2020)), although note that this test is more stringent due to the fact that I am using the same simulation to inform and then test the halo model.

At small scales I see that the level of agreement improves again ($< 5\%$ at $1 < k\,[h\,\mathrm{Mpc}^{-1}] < 4$) between the BAHAMAS-informed halo model predictions and the simulations and theoretical predictions. However, the error increases again at still smaller scales. While the error increases as the Nyquist frequency is approached[6] (see dotted vertical line), the fact that the HALOFIT prediction is very similar to that of BAHAMAS suggests that the error is not solely due to aliasing effects in the simulation $P(k)$. Further tests exploring the minimum radius and halo mass, as well as the radial and mass binning strategies, in the halo model show the results to be numerically robust. Plausible physical explanations for the deviation at very small scales include differences in the clustering of substructures compared to the smooth dark matter profile, asphericity of the mass distribution, and intrinsic scatter in the mass density profiles.

In fig. 3.13 I explore the effects of changing the halo mass definition, the profiles (tabulated vs. parametric fit), and the HMF (BAHAMAS vs. Tinker et al. 2008) at a number of different redshifts. There are four sets of plots, corresponding to the four halo mass definitions that I explore ($\Delta = 200m$, $500m$ and $\Delta = 200c$, $500c$). The top row of panels in each plot set correspond to the case where the tabulated profiles directly from the simulations are used in the halo model, whereas the bottom row of panels use the Einasto

[6]The Nyquist frequency is defined as $\nu_y = 2\pi N_{\mathrm{cell}}/L_{\mathrm{Box}}$ where N_{cell} is the number of cells used in the Fourier transform (to the one-third power) when evaluating $P(k)$ and L_{Box} is the box size.

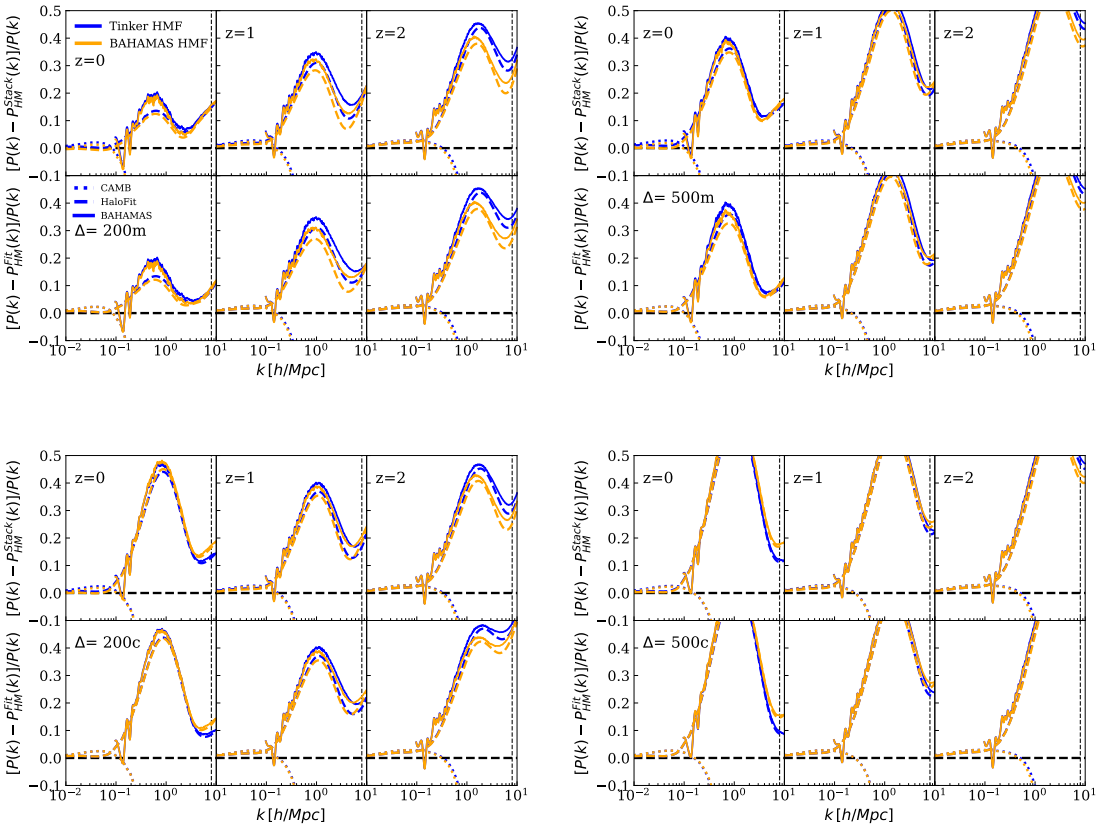

FIGURE 3.13: Residual plots between the matter power spectrum and the halo model predictions at three redshifts bins for the four different mass definition as follows: top-left $\Delta = 200m$, top-right $\Delta = 500m$, bottom-left $\Delta = 200c$ and bottom-right $\Delta = 500c$. The top (bottom) row of panels in each plot set correspond to the case where the tabulated profiles directly from the simulations (Einasto fits to) are used in the halo model. The orange (blue) curves correspond to the case where I use the BAHAMAS *DM-Only* simulation (Tinker) HMF. I compare against linear theory predictions using CAMB (dotted curves) and the non-linear HALOFIT prediction (dashed curves) and BAHAMAS *DM-Only* simulation (solid curves). The vertical dotted line represents 0.5 times the Nyquist frequency of the BAHAMAS simulation volume. All models recover the large-scale limit by construction while the 1-halo/2-halo transition region is (at best, corresponding to the $\Delta = 200m$ case) recovered to 10-15% accuracy. In general, the accuracy decreases with decreasing radial extent of the haloes (due to changing halo mass definition, see text) and increasing redshift.

fit to the density profiles. The blue curves correspond to the case where I use the Tinker HMF, whereas the orange curves used the BAHAMAS *DM-Only* simulation HMF. Note that here I present residuals, defined as $[P(k) - P^{\rm HM}(k)]/P(k)$ (where $P^{\rm HM}(k)$ is the halo model prediction), whereas in the bottom panel of fig. 3.12 I showed a simple ratio.

I focus first on the top left set of plots, corresponding to a spherical overdensity case of $\Delta = 200m$. Scanning from left to right, it is clear to see the halo model increasingly struggles to capture the 1-halo/2-halo transition region with increasing redshift. This is true regardless of which mass function I use (BAHAMAS or Tinker) or whether I use

tabulated or fitted density profiles (top vs. bottom rows). Interestingly, examining the other spherical overdensity cases (see the other three sets of plots in fig. 3.13), it appears that when the overdensity criteria are defined with respect to the mean background density, the precision of the model worsens with increasing redshift while the accuracy is mostly independent of redshift when the critical density is used. The fact that there is a relation between the accuracy of the halo model and the mass definition was also hinted at in Mead et al. (2021), where they identified differences between using $\Delta = 200m$ and $\Delta = 200c$.

Comparing the top and bottom rows of the top left set of plots, there are no significant differences in the ability of the halo model to recover the simulation $P(k)$. This implies that the Einasto form I have used reproduces the simulated matter density profiles sufficiently well for the purposes of predicting $P(k)$, since the result does not change when I use tabulated profiles directly (top row) vs. the Einasto fitting function (bottom row).

Comparing the solid orange (BAHAMAS HMF) and solid blue (Tinker HMF) curves, I see that using the actual BAHAMAS simulation HMF results in an improved agreement between the halo model and the simulation $P(k)$, particularly at higher redshifts. Thus, the halo model is more accurate than what might have been inferred using a generic halo mass function to test it.

Scanning between the four sets of plots, another trend that is clearly visible is that changing the halo mass definition has a significant impact on the accuracy of the halo model with respect to the simulations. The change in the halo mass itself is not what is driving this trend: since I essentially integrate over *all* haloes, how I label their masses should not matter. However, by changing the halo mass definition, I am also changing the radial extent (size) of a halo (given the spherical overdensity definition) and this clearly will impact where the 1-halo and 2-halo terms intersect, due to the change in the extent of the 1-halo term. These results are consistent with the findings of van Daalen & Schaye (2015), who showed the importance of the radial selection of particles on the resulting power spectrum of cosmological simulations (see figure 3 of that study).

I find that the larger the radial extent of the halo (noting that at $z = 0$, R_{200m} is the largest and R_{500c} is the smallest) the better the halo model is able to capture the 1-halo/2-halo transition region in the simulations. This suggests that one way to help further improve the halo model is to radially extend the 1-halo term (as recently presented in García et al. (2021)). For example, even if the halo mass function and bias are defined with respect to some standard choice of overdensity (e.g., $\Delta = 200c, m$), the profiles could, for example, be extended to several times the corresponding spherical overdensity radius, with the optimum extent determined by fitting to the simulation

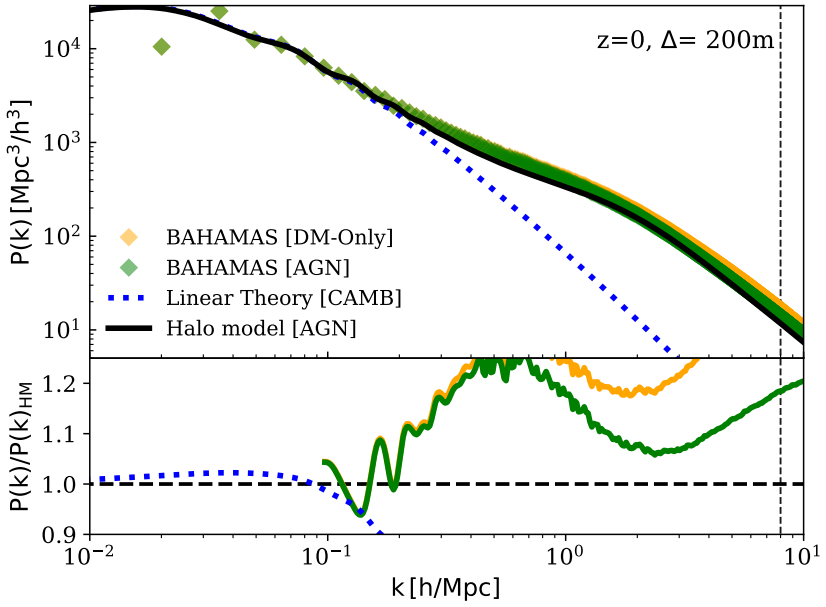

FIGURE 3.14: Same as fig. 3.12 but for the hydrodynamical (*AGN*) case. The BAHAMAS *AGN* matter power spectrum is represented by green symbols. The bottom panel I show the ratios with respect to the baryonic version of the halo model (i.e., using the density profiles and HMF from the BAHAMAS *AGN* run). The level of agreement between the baryonic halo model and the hydrodynamical simulations is similar to that seen in the comparison of the collisionless halo model and collisionless simulations in fig. 3.12.

$P(k)$. However, whether such an approach is strongly cosmology dependent is unclear. Alternatively, it may be possible to adopt a consistent mass and radius definition but simply lower the overdensity value (e.g., $\Delta = 100$) or adopt an alternative physical mass/radius scale such as the 'splashback' radius (e.g., Diemer & Kravtsov 2015; Diemer 2020; O'Neil et al. 2021).

3.4.2 Matter power spectrum including baryon physics

In fig. 3.14 I present an analogous plot as in fig. 3.12, where the green diamonds represent the power spectrum from the BAHAMAS *AGN* run and in the bottom panel the various ratios are now with respect to the baryon version of the halo model. Note that the baryon version of the halo model corresponds to either using tabulated profiles directly from the *AGN* run or an Einasto fit to them, as well as using either the BAHAMAS *AGN* HMF or a Tinker HMF with a baryon correction applied. For fig. 3.14 I use the tabulated profiles and HMF from the BAHAMAS *AGN* run.

As in the case of the *DM-Only* version, our baryon halo model prescription recovers the linear regime ($k < 0.1\,h\mathrm{Mpc}^{-1}$) to better than percent level accuracy, by construction. Consistent with the collisionless comparison, the agreement is worst at the 1-halo/2-halo

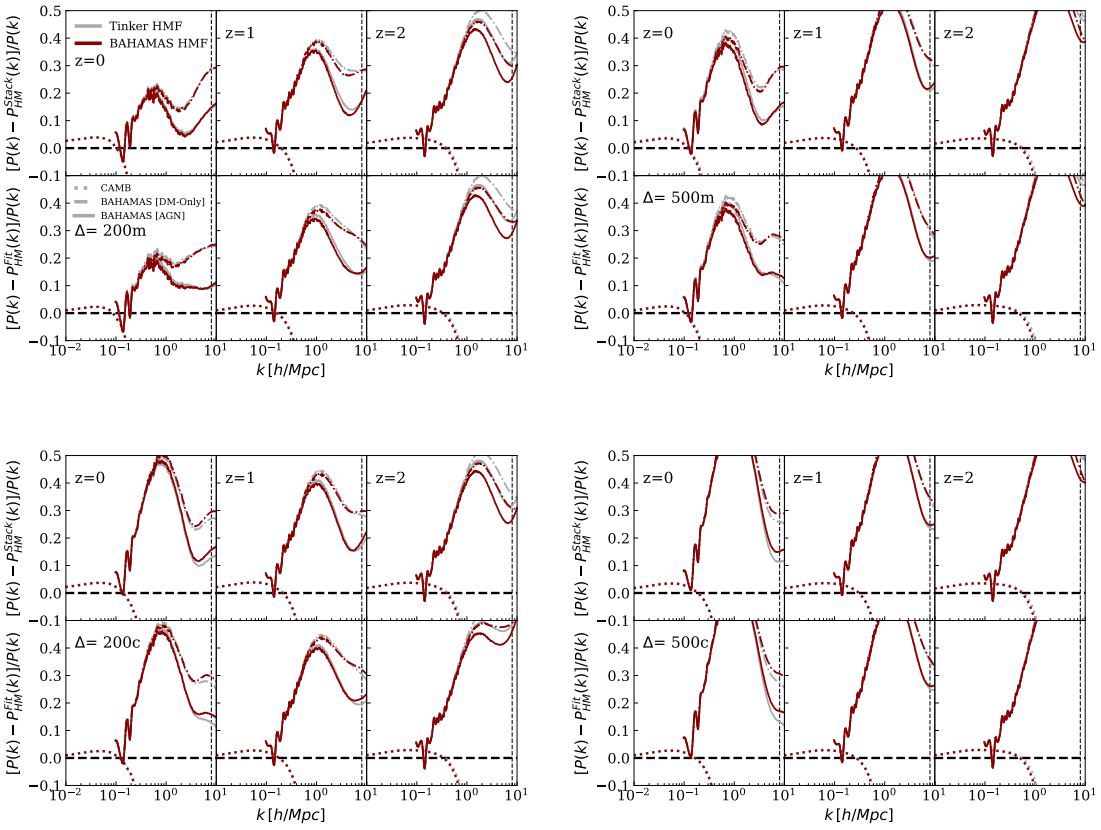

FIGURE 3.15: Residuals plots between the baryonic matter power spectrum and the halo model predictions at three different redshifts for the four different mass definition following the same format as the *DM-Only* companion plot in fig. 3.13. I also show, instead of the HALOFIT prediction, the difference between the baryonic halo model compared the BAHAMAS *DM-Only* $P(k)$ (dashed curve). The overall trends and level of agreement are very similar to those found for the *DM-Only* case, though there are differences in detail (see text).

transition region, deviating from the simulation prediction by up to 20%. The agreement improves again at smaller scales, though still deviates by ≈10%.

In fig. 3.15 I show the residuals plots in the same way I have presented for the *DM-Only* case, with the stacked and fitted density profiles and two different forms for the HMFs, at different redshifts ($z = [0, 1, 2]$), and for the four halo mass definitions. The grey curves correspond to the cases using the baryon-corrected Tinker HMF and the red curves correspond to the cases using the BAHAMAS *AGN* HMF. Overall, I find very similar trends to those presented in fig. 3.13 for the *DM-Only* case. Specifically, the mass definition that works best is again $\Delta = 200m$, which can recover the 1-halo/2-halo transition region to 20% at $z = 0$ and ≈ 35% at $z = 1$. The 1-halo region ($k > 2\,h\mathrm{Mpc}^{-1}$) is generally recovered to 10% at $z = 0$ independent of the choice of halo mass definition, HMF, and non-parametric vs. parametric profiles. At higher redshifts, the discrepancy

with respect to the simulations increases for both the 1-halo/2-halo transition region and in the deep non-linear (1-halo) region.

Upon closer inspection, it is apparent in some cases that there is a difference at very small scales (high k values) between the accuracy of the halo model when using either the parametric (Einasto) or tabulated mass density profiles. For example, at $z = 0$ in either the $\Delta = 200m$ or $\Delta = 200c$ cases, the residuals increase towards smaller scales when using the tabulated profiles, whereas for the parametric case they are approximately independent of k scale. I attribute this difference in behaviour to the inability of the Einasto form to fully capture the behaviour of the density profiles at small scales, due to the increasing importance of the central galaxy (see fig. 3.4). Thus, in this case, using the more accurate tabulated density profiles demonstrates that the halo model is actually less accurate in reproducing the non-linear power spectrum on small scales.

Overall, therefore, the trends in the accuracy of the baryon version of the halo model are very similar to those for the collisionless version, when the models are compared to the hydrodynamical and collisionless BAHAMAS simulations respectively. In particular, I find that the absolute accuracy is worse at the 1-halo/2-halo transition and typically worsens at higher redshifts when the halo mass definition is defined with respect to the mean background density. The choice of halo mass definition is also important. Given that the trends are very similar between the baryon and collisionless cases, it raises the interesting question of whether the halo model would actually be better suited at predicting the *ratio* (or suppression) of the matter power spectrum due to baryons, as opposed to predicting the absolute $P(k)$. I explore this possibility below.

3.4.3 Matter power spectrum suppression

I now explore the halo model predictions for the matter power spectrum suppression, sometimes also referred to as the 'suppression factor'. There are many recent studies of the suppression factor using cosmological hydrodynamical simulations in the literature (e.g., van Daalen et al. 2011, 2020; Schneider & Teyssier 2015; Chisari et al. 2019; Schneider et al. 2019; Debackere et al. 2020). Here I explore the accuracy with which the halo model can recover the suppression of the matter power spectrum in the BAHAMAS simulations.

In fig. 3.16 I show the suppression effect of the baryons with respect to the *DM-Only* simulations, where the suppression is defined simply as $S(k) \equiv P_{AGN}(k)/P_{DM}(k)$. The structure of the plots is similar to the previous ones that I have shown for the power spectra comparison (figs. 3.13 and 3.15) but in this case I show the suppression power spectra (top panels) and the ratio between the BAHAMAS results and the halo model

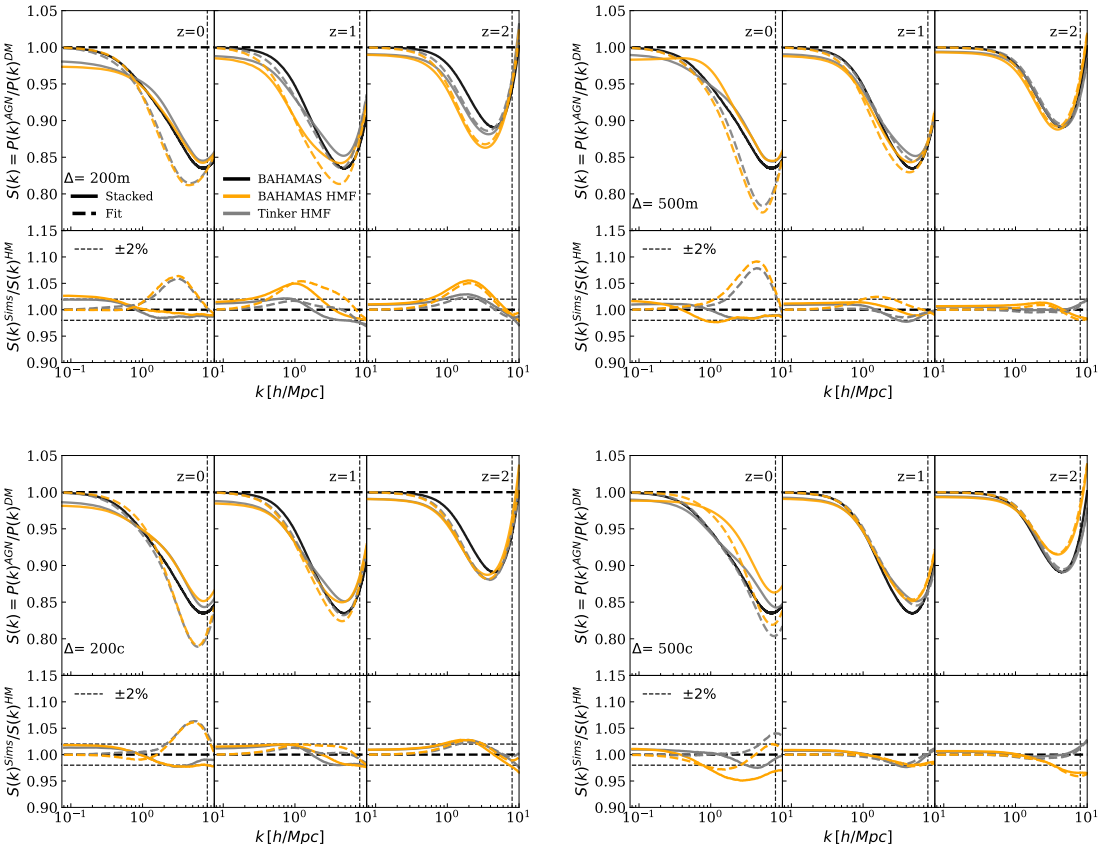

FIGURE 3.16: Matter power spectrum suppression plots between the total matter power spectrum and the *DM-Only* power spectra predictions using the halo model results at three different redshifts for the four different mass definition following the same structure as the *DM-Only* and *AGN* plots fig. 3.13 and fig. 3.15. In the bottom panel I show the accuracy of our prediction with the ratio between the simulation expected results and the halo model prediction highlighting the 2% difference using dashed black curves. The BAHAMAS predictions are shown in solid black curve. I show the prediction with BAHAMAS HMFs in solid orange curves while in grey I show the Tinker prediction, in dashed curves I show the case using the fit density profiles while in solid curves the stacked density profiles. The halo model reproduces the simulated suppression factor to typically a few percent accuracy, independent of details such as the halo mass definition.

results (bottom panels). I show in solid (dashed) curves the predictions using the tabulated (fitted) density profiles. In orange I show the predictions using the BAHAMAS HMFs and in grey using the Tinker HMFs (with the baryonic correction applied to the *AGN* cases).

On a qualitative level, I can see that the ratio of the baryon to collisionless halo models (top row of panels in each plot set) has a 'spoon'-like form that closely mimics that found by taking the ratio of power spectra from hydrodynamical and collisionless simulations. Examining the ratio of power spectrum suppression of the simulations with respect to that from the halo model (i.e., a ratio of ratios, in the bottom row of panels of each plot

set in fig. 3.16), I can also see that there is no evidence of an issue of near the 1-halo/2-halo transition region, nor of any particular systematic issues as a function of halo mass definition or redshift. Slight differences exist depending on which set of density profiles I use (tabulated vs. parametric), but it is nevertheless abundantly clear that the halo model formalism is considerably more accurate in predicting the matter power spectrum suppression factor, as opposed the absolute $P(k)$. Typically, I find that the ratio of halo models is accurate at the ≈2-3 percent level. The suppression comes principally from the baryonic correction applied to the HMFs then, in a minor part, from the density profiles.

3.4.4 Halo model comparison with other models

In Section 3.4, I have compared the calibrated halo model predictions against standard linear theory and non-linear results such as CAMB and HALOFIT in order to validate our results. In this short Section, I present the comparison against other, up-to-date, non-linear results such as CosmicEMU (Heitmann et al., 2016), HaloMOD (Murray et al., 2020) and HMCode (Mead et al., 2015, 2021) in order to verify our results in the light of the latest development on the topic. Please note that the first one presented, CosmicEMU, is a matter power spectrum emulator while the other two are halo model based software.

In fig. 3.17 I present in detail the comparison of our halo model prediction against other non-linear code, similarly to what I have done previously in fig. 3.12 in Section 3.4. I show the non-linear codes in dotted and dashed blue lines (see figure caption for further details) and the simulation output in orange square dots in the top panel, while in the bottom panel I show the ratio of the models over our prediction.

It is possible to see in the large scale limit ($k < 0.2\,h\,\mathrm{Mpc}^{-1}$) all predictions agree better than 5%, I can also see a feature in the HMCode in the region $0.07 < k\,[h\,\mathrm{Mpc}^{-1}] < 0.2$ where there is a different treatment of the baryonic acoustic peaks. In the transitional region, $0.1 < k\,[h\,\mathrm{Mpc}^{-1}] \lesssim 2$, all model are in reasonable agreement, therefore our prediction, that is not able to accurately reproduce this region, have an agreement between 10 and 20%, that grows back to 5% in the 1-halo region ($k \gtrsim 2\,h\,\mathrm{Mpc}^{-1}$) as was already shown in previous analysis.

In the small scale regime, I see a similar behaviour compared to the previous analysis. Please note that the different non-linear models have different largest k values (e.g., CosmicEMU has a maximum k scale at $\approx 7\,h\mathrm{Mpc}^{-1}$).

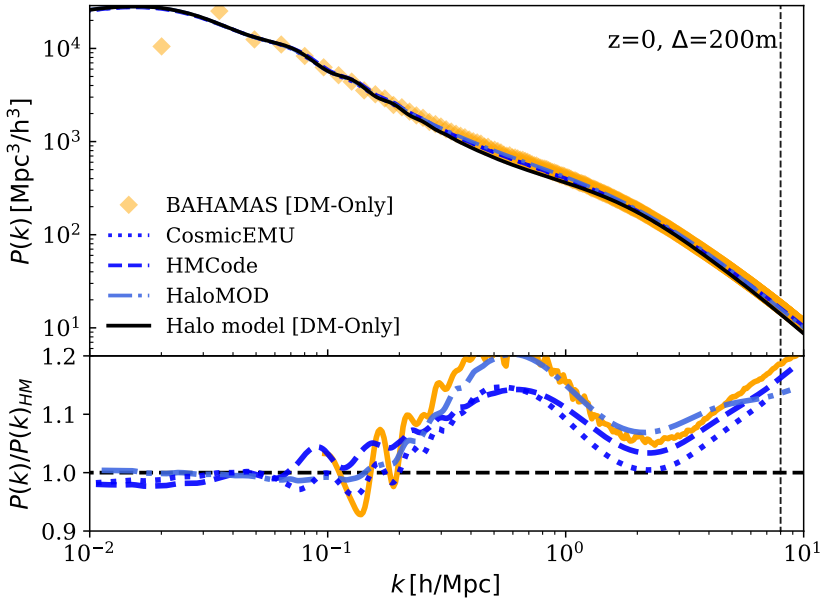

FIGURE 3.17: In the top panel I present the *DM-Only* BAHAMAS matter power spectrum as orange squared symbols, in dotted blue line the COSMICEMU results, in dashed blue the HMCODE and in dot-dashed line the HALOMOD results, finally with a solid black line our prediction using the stacked density profiles of $\Delta = 200m$. In the bottom panel I present the ratio of the previous results as presented in fig. 3.12. We see a good agreement in the large scale limit ($< 4\%$) with all the models, in the 1-2 halo transition region I see an agreement between $10 - 15\%$. Finally in the 1-halo region ($k \gtrsim 2\,h\,\mathrm{Mpc}^{-1}$) I see that our prediction has a better than 5% agreement with all models which is the same result that I have found comparing against HALOFIT and BAHAMAS power spectrum.

In fig. 3.18 I present a comparison between our *AGN* calibrated halo model against BAHAMAS power spectrum and HMCODE (v2016 and v2020) baryon matter power spectrum. As before, our prediction are in good agreement with these updated codes.

On the large scale regime there is a concordance better than 5% and $< 15\%$ agreement in the 1-2 halo transition region. It is interesting to see that the model denoted as v2016 in the HMCODE is in great agreement with the simulation output but the v2020 is a bit lower, under-predicting by more than 5% the simulation output being strongly comparable to our halo model predictions with an agreement in the order of 3%. The differences between the two version of the HMCODE lie in the different baryon feedback parameters used ($BAR = 2.75$ for the v2016 version) and the new calibrated one on a AGN heating temperature of $T_{\rm Heat} = 10^8\,\mathrm{K}$ for v2020. Further and more detailed discussion can be found in Mead et al. (2016, 2021).

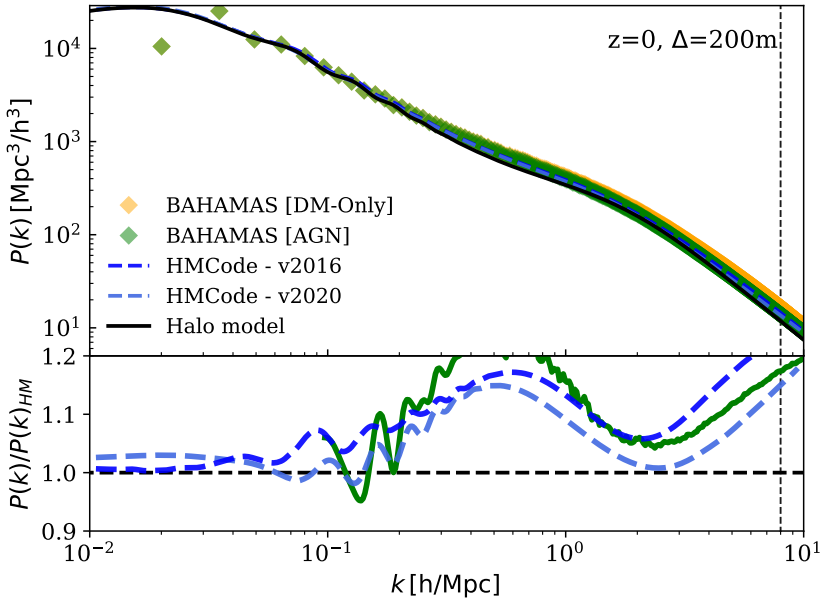

FIGURE 3.18: Same plot as fig. 3.17 but for the *AGN* case (companion plot with fig. 3.14 in Section 3.4). In the top panel I present the *DM-Only* and *AGN* BAHAMAS matter power spectrum as orange and green squared symbols, in dashed blue lines the results from HMCODE from two separate versions (v2016 with baryon feedback parameters $BAR = 2.75$ darker line, and v2020 with calibrated $T_{\rm Heat} = 10^8\,K$ in lighter blue), finally with a solid black line our prediction. In the bottom panel I present the ratio of the previous results as presented in fig. 3.17. As was already visible in the *DM-Only* case there is a good agreement in the large scale limit ($< 4\%$) with the HMCODE and an agreement between $10 - 15\%$ in the 1-2 halo transition region. Finally in the 1-halo region ($k \gtrsim 2\,h\,{\rm Mpc}^{-1}$) it is possible to see that our prediction has a better than 5% agreement with all models. A little surprising is the fact that the newest calibrated baryon feedback from HMCODE has a similar behaviour compared to our halo model prescription compared to BAHAMAS results.

3.5 Discussion and Summary

In this Chapter, I have assessed the accuracy of the halo model to predict the non-linear matter power spectrum, which is the basis of many large-scale structure cosmological probes. The advantages of the halo model are its speed, flexibility, and its intuitive physical nature. However, its accuracy in predicting the non-linear power spectrum needs to be carefully assessed and here I have posed a simple question: how well does the halo model predict the non-linear power spectrum, $P(k)$, from a cosmological simulation when the ingredients of the halo model (namely the halo mass function and mass density profiles) are extracted from the same simulation? Although the question is simple, the test is in fact a demanding one, since once the mass function and density profiles (and cosmology) are specified, there are no free parameters in the standard halo model.

I briefly summarise the main results below:

- I have computed the stacked (mean) total mass density profiles in bins of halo mass and redshift for the BAHAMAS *DM-Only* and *AGN* simulations (see figs. 3.1 and 3.4, respectively) and provided Einasto profile fits to these profiles (see eqn. 2.7 and Table 3.1).

- Using the BAHAMAS simulations, I have derived a correction to the halo mass function that encapsulates the presence and impact of baryons on haloes. This correction works for every overdensity and up to $z = 2$ with an accuracy better than 5% (see fig. 3.8 and eqn. 3.7).

- Using density profiles and halo mass functions extracted from the BAHAMAS simulations, I have calculated the non-linear power spectrum, $P(k)$, using the standard halo model. Qualitatively speaking, the standard halo model reproduces the power spectrum in both the collisionless and baryon cases (see fig. 3.12 and fig. 3.14, respectively) I have considered, correctly capturing both the large-scale, linear limit and the deep non-linear regime.

- In detail, I find that the halo model struggles to quantitatively reproduce the simulation power spectrum on intermediate scales ($0.1 \lesssim k\,[h\,\mathrm{Mpc}^{-1}] \lesssim 5$) that mark the transition from the so-called 2-halo term (the clustering of nearby, correlated haloes) to the 1-halo term (the mass density distribution inside a single halo). For example, at $z = 0$ and adopting a halo mass defined with respect to 200 times the mean background density, the halo model predicts a $P(k)$ that is systematically lower than predicted by the cosmological simulations by up to 15-20% (see figs. 3.13 and 3.15). This result follows previous works (e.g. Massara et al. 2014; Voivodic et al. 2020) but narrows down the source of uncertainties by using the halo mass function and density profiles directly from the simulations.

- I have shown that the choice of halo mass definition (defined with respect to the critical or mean background density and the choice of overdensity) has a significant impact on the 1-halo/2-halo transition region offset. This effect is due to the change in the radial extent of the haloes depending on the mass definition, with larger radial extents (lower overdensities) generally resulting in an improved match between the halo model and the simulations.

- The 1-halo dominated region is recovered to 5% at $z = 0$ and better than 10% for all mass definitions, although the accuracy decreases at higher redshifts.

- While the standard (unmodified) halo model cannot predict the absolute power spectrum to better than 15% accuracy on intermediate scales (at best), I have shown that these systematic errors largely cancel when considering the ratio of the baryon to collisionless cases. Typically, the halo model can reproduce the

suppression seen in the simulations to a few percent accuracy, independent of the details such as the halo mass definition (fig. 3.16).

One of the key findings of this study is that the accuracy of the halo model in reproducing the simulations is strongly affected by the halo mass definition, through its impact on the radial extent of haloes. In essence, adopting higher overdensities implies smaller radial extents (for a given mass) which effectively confines the 1-halo contribution to smaller scales, resulting in lower power at the 1-halo/2-halo transition region and poorer agreement with the simulations. One possibility is to simply radially extend the profiles associated with a given mass definition (García & Rozo, 2019; García et al., 2021). Alternatively, one can retain the link between the halo mass and radius and simply adopt a lower overdensity, or perhaps another physical scale (at typically low overdensities) such as the splashback radius. In addition, Mead & Verde (2021) have shown that accounting for non-linear bias in the 2-halo term also helps to mitigate the error in the transition region. Note that the standard halo model assumes a linear bias which is independent of scale, but in principle I expect the clustering to be scale-dependent on quasi-linear scales (Smith et al., 2007; Baldauf et al., 2012).

The other major finding of this study is that the ratio of power spectra (baryon case to collisionless case) can be much more robustly predicted with the standard halo model than can the absolute power spectra. Interestingly, previous studies have similarly concluded that the effects of including massive neutrinos or of altering the nature of dark energy or gravity on the matter power spectrum are also most reliably captured with the halo model in terms of ratios (e.g., Schmidt et al. 2010; Mead 2017; Cataneo et al. 2019, 2020; Bose et al. 2020, 2021). In these studies, the ratio is sometimes referred to as the 'response' or the 'reaction' to a cosmological change. Our results regarding the ratio of the baryon and collisionless halo models could therefore be termed as a 'baryon response' or 'baryon reaction'. One possibility, is to use the halo model to predict the baryon response and combine this with other methods for computing the absolute power spectrum in the collisionless limit.

Chapter 4

Halo model predictions for large-scale structure observables

In this Chapter of the Thesis, I explore the applications of the halo model in reproducing large scale structure auto- and cross-correlations used to constrain cosmological and baryonic feedback parameters. I will focus on how inaccuracies in the modelling of the non-linear power spectrum propagate through to predictions for these observables and quantify the impact of baryon physics. Being able to distinguish and quantify the baryonic effects on each separate observables could, in principle, improve the quality of the predictions made by the halo model and avoid the introduction of biases in the parameters inferred with this method.

4.1 Introduction

In the previous Chapter and Sections, I have presented several tests to validate the internal accuracy of the halo model prediction for the non-linear matter power spectrum by making use of the BAHAMAS simulations. The previous results showed that the standard halo model approach reproduces the large and small scale limits fairly well but struggles to accurately recover the 1-halo to 2-halo transition region better than 15% accuracy at z=0.

In this Chapter, I explore the accuracy of the standard approach in reproducing LSS observables. Those LSS probes (weak lensing and tSZ) are the perfect testing ground for exploring the calibrated halo model results and investigating the separability between the baryonic and accuracy effects. It is worth mentioning that not all observables considered are impacted in the same way by the halo model overall accuracy, for instance, the tSZ

is strongly influenced by the 1-halo term so the lack of accuracy in transition region will not affect this observable. As I have presented earlier, the use of the 'boost factor' or baryonic suppression can remove some of the uncertainties present in the standard halo model approach over a wide range of redshifts and k scales. Interestingly, I have shown that the mass definitions have a significant impact on the accuracy of the matter clustering with a relative discrepancy up to 50% at z=0. However, the dependence of the accuracy of the halo model on the choice of halo mass definition in predicting $P(k)$ appears to be independent of whether the collisionless or hydrodynamical simulations are used to constrain the halo model. I will explore the extent to which this issue is true for other observables.

To validate and explore the LSS probes, I will use the fit to the density profiles of the four mass definitions, using the Tinker HMF for both the *DM-Only* and *AGN* cases and I will apply the same parameters presented in Mead et al. (2016) as explained in the previous Chapter. The largest mass definition, Δ =200 Mean, showed the highest accuracy in reproducing the matter clustering and will be considered our 'fiducial' setup below.

To make these tests with LSS probes, I have computed the matter power spectrum up to z=3 (three hundred realisations between $0 \leq z \leq 3$) in a mass range from $4 \times 10^{11} < M\,[\mathrm{M}_{\odot}\,h^{-1}] < 5 \times 10^{15}$, using the Tinker HMF and linear bias, corrected when working with the hydrodynamical case accordingly. The redshift range is enough for many of the LSS probes presented, excluding CMB weak lensing which, as a general rule, requires the knowledge of the matter distribution at least up to $z \approx 10$ (Hill & Spergel, 2014) due to the width of the associated lensing efficiency kernel. For the latter probe, I will only focus on the baryonic suppression to compare against BAHAMAS simulations, as presented in Chung et al. (2020).

In the following Sections, I will present the weak lensing source redshift distributions used in the analysis, the tSZ data I am comparing against and how each probe is evaluated, as well as presenting the baryonic effect on those probes.

4.2 LSS predictions

The basic theory of how these observables are computed and their cosmological parameters dependency is presented thoroughly in Chapter 2 and for sake of brevity and clarity, they will not be repeated here. In the next subsection I present the weak lensing data I use to build these observables and compare against. I use separate data sets in order to

maximise the number of tests of the halo model and match some of the results presented in McCarthy et al. (2018).

4.2.1 Weak lensing data

To validate the halo model accuracy, I compare the predictions for angular power spectrum results for fixed source distributions using the BAHAMAS light-cones maps, as presented in Section 2.4.1, and in real space correlations with data sets of some of the latest weak lensing survey results. Throughout this Chapter, we present observational results as obtained by the original authors and we refer to their papers accordingly.

CFHTLenS

The Canada-France-Hawaii Telescope Lensing Survey (CFHTLenS)[1] is a weak lensing survey that obtained deep multi-colour data on 154 square degrees in five different photometric observational bands with an apparent magnitude limit of ≈ 25.5 in the filter $i^{'}$ (Hildebrandt et al., 2012; Heymans et al., 2012, 2013; Benjamin et al., 2013)[2]. The images were taken with the MegaPrime multi-chip instrument based on the Hawaii islands (Erben et al., 2013).

RCSLenS

The Red Cluster Sequence 2 Survey (RCSLenS)[3] is an application of the techniques developed for CFHTLenS to the ≈ 800 deg^2 multi-band imaging data survey (Hildebrandt et al., 2016). The observational bands are the $griz$ bands over an area of 785 deg^2 to a depth of ≈ 24.4 magnitude in the r-band for a 7σ point source. This survey uses the same instrument setup as CFHTLenS. This survey was designed to target an optically selected sample of galaxy clusters over a wide range of redshift (Choi et al., 2016; van Uitert et al., 2016).

This survey overlaps with Sloan Digital Sky Survey (SDSS) mostly in the northern sky with $\approx$ 400 deg^2 and with Baryon Oscillation Spectroscopic Survey (BOSS), a spectroscopic follow-up of SDSS-III (Eisenstein et al., 2011). RCSLenS overlaps SDSS by roughly 184 deg^2 thanks to the BOSS large sample of Luminous Red Galaxies (LRG), ≈ 50000.

Another overlap is with WiggleZ Dark Energy Survey (Drinkwater et al., 2010), a redshift survey of emission line galaxies in the southern sky with a total of 80000 shared objects

[1] https://www.cfhtlens.org/.

[2] The five photometric observational bands used are $u^*g^{'}r^{'}i^{'}z^{'}$ those correspond to 5 different photometric filters with effective wavelengths midpoint of $365, 464, 658, 806, 900$ nm. The first is in the UV band, $g^{'}r^{'}$ are in the optical bands and $i^{'}z^{'}$ are in the near-infrared wavelengths.

[3] https://www.rcslens.org/.

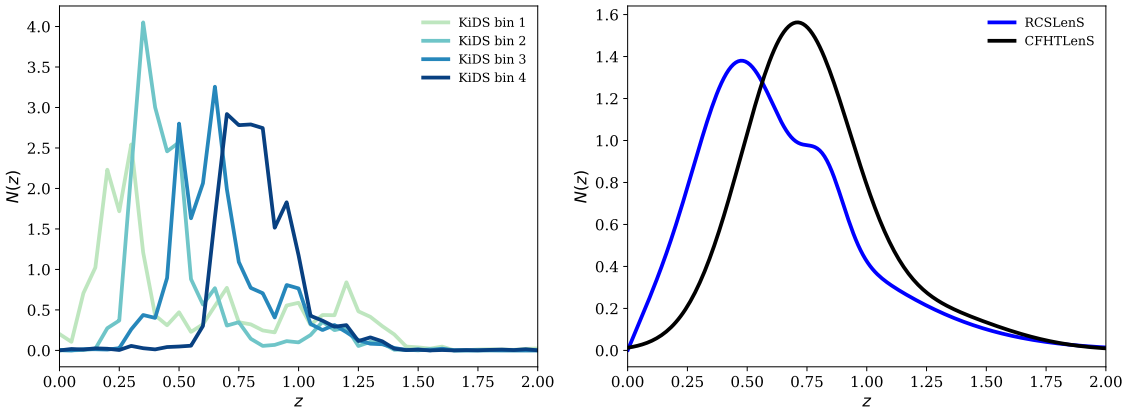

FIGURE 4.1: In the left plot, I present the KiDS tomographic source redshift distributions (Hildebrandt et al., 2017). I show in different colours the four tomographic bins used to compute the tomographic weak lensing measurements. In the right panel, instead, I present the continuum redshift source distributions for RCSLenS and CFHTLenS from Hildebrandt et al. (2016) and Harnois-Déraps et al. (2015) using eqns. (4.2)(4.1). I apply a cut at $z = 2$ due to the small number or sources above that threshold, $n_s(\chi|z) \approx 0$.

from $0 < z < 1$ (unmasked 181 deg^2). Finally, with the DEEP2 galaxy redshift survey (Newman et al., 2013) with roughly 6000 sources in an area of 1 deg^2.

KiDS

KiDS, the Kilo Degrees Survey[4], is a four-band imaging survey made using the Omega-CAM CCD mosaic camera mounted on the European southern observatory VLT Survey Telescope (VLT) on the Chilean mountains[5]. Observations are carried out similarly to the SDSS configurations with u, g, r, i bands with magnitudes at the optimal setup of 24.3, 25.1, 24.9, 23.8 (5σ in a 2 arcsec aperture). The observations targets are two strips of $\approx 10 \times 75$ deg on the celestial equator and around the South Galactic Pole (respectively called KiDS-N and KiDS-S, Hildebrandt et al. (2017)). The KiDS source redshift distributions were acquired from the survey database cited above.

In fig. 4.1 I present the source redshift distribution for the KiDS tomographic cases (left-hand side plot) and the continuum distributions (i.e., a single wide redshift bin) of RCSLenS and CFHTLenS (right-hand side plot). Sources are plotted up to $z = 2$ because there are very few detected above that threshold. In fig. 4.2 I present the lensing kernels of the source distributions presented in fig. 4.1. The values are presented in normalised units by dividing with the maximum value of the distribution for the cases of RCSLenS and CFHTLenS, while for KiDS, I have divided by the maximum of the last tomographic bin (fourth bin). The equation used to compute the kernel is presented

[4] http://kids.strw.leidenuniv.nl/.
[5] https://www.eso.org/public/italy/teles-instr/paranal-observatory/vlt/.

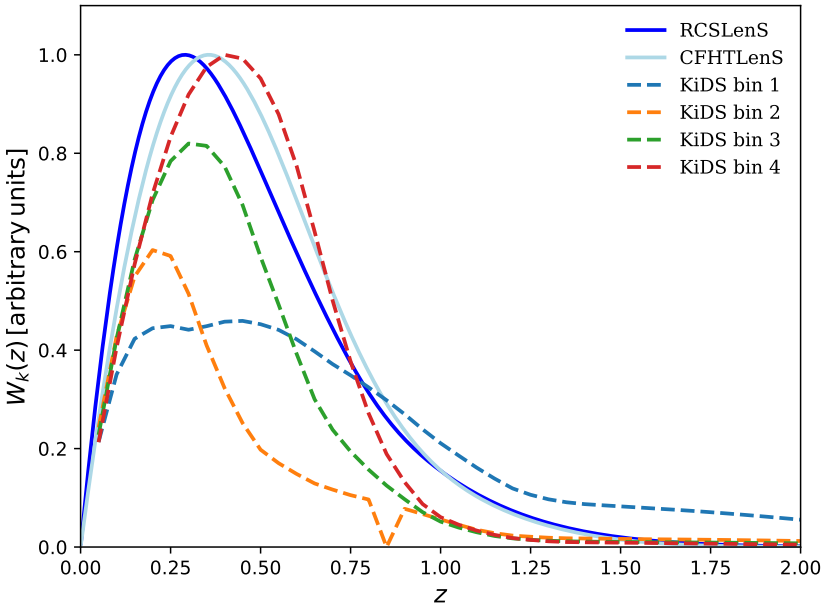

FIGURE 4.2: Weak lensing kernels for CFHTLenS, RCSLenS and KiDS source redshift distributions. The values are normalised, for plotting purposes, to the maximum of the W_k for the first two cases, while it is normalised to the maximum of the fourth bin for KiDS case. The kernels are computed using eqn. 2.22.

in eqn. 2.22. It is possible to see that the long tail of sources, present in the first bin of KiDS survey, creates a broader lensing kernel compared to the other cases.

The RCSLenS and CFHTLenS have source redshift distributions that can be approximated by the following functions:

$$N_{\rm CFHTLenS}(z) = N_0 \exp\left(\frac{-(z-z_0)^4}{S_0^2}\right) + N_1 \exp\left(\frac{-(z-z_1)^4}{S_1^2}\right) + N_2 \exp\left(\frac{-(z-z_2)^4}{S_2^2}\right) , \tag{4.1}$$

$$N_{\rm RCSLenS}(z) = zN_0^{'} \exp\left(\frac{-(z-z_0^{'})^2}{C_0^2}\right) + zN_1^{'} \exp\left(\frac{-(z-z_1^{'})^2}{C_1^2}\right) + zN_2^{'} \exp\left(\frac{-(z-z_2^{'})^2}{C_2^2}\right) . \tag{4.2}$$

In Table 4.1 I present the free parameters used in eqn. 4.1 for the CFHTLenS results (Kilbinger et al., 2013; Van Waerbeke et al., 2013; Harnois-Déraps et al., 2015) and eqn. 4.2 for the RCSLenS ones (Hildebrandt et al., 2016).

TABLE 4.1: CFHTLenS and RCSLenS free parameters of the analytical functions shown in eqns. (4.1)(4.2) from Van Waerbeke et al. (2013); Harnois-Déraps et al. (2015) and Kilbinger et al. (2013) for CFHTLenS and Hildebrandt et al. (2016) and Hojjati et al. (2017) for RCSLenS.

CFHTLenS (Harnois-Déraps et al., 2015)								
N_0	z_0	s_0	N_1	z_1	s_1	N_2	z_2	s_2
0.5482	0.699	0.074	0.596	0.815	0.216	0.207	1.133	0.308
RCSLenS (Hildebrandt et al., 2016)								
$N_0^{'}$	$z_0^{'}$	C_0	$N_1^{'}$	$z_1^{'}$	C_1	$N_2^{'}$	$z_2^{'}$	C_2
3.126	0.419	0.979	1.678	0.404	0.25	0.40	0.813	0.121

The knowledge of the distribution of the sources enters in the calculations of the weak lensing kernels, using eqn. 2.22, that is needed to compute the convergence power spectrum as shown in eqn. 2.21.

4.2.2 tSZ pressure profiles

To compute the thermal Sunyaev-Zel'Dovich power spectrum we need a way to characterise the energy distributed inside the haloes. To do so, following the general description of this observable provided in Chapter 1 of this Thesis, I compute the electron pressure density inside haloes.

I recall the description of the electron pressure density in terms of the gas density, ρ, and depending on the mean molecular weight per free electron, μ_e, and the proton mass, m_p, in eqn. 4.3 :

$$P_e = k_B T \frac{\rho}{\mu_e m_p} \quad . \tag{4.3}$$

Following the methodology applied to the mass density profiles, I build a catalogue of stacked pressure profiles spanning a wide range of masses and redshifts. Since the tSZ is mostly influenced by the group-cluster regime, I cut the haloes catalogue below $10^{12}\,\mathrm{M}_\odot\,h^{-1}$. In analogy to what was done with the density profiles, I compute energy weighted radial bins. The energy weighted radial bins, r_w^{SZ}, are built as follows:

$$r_w^{\mathrm{SZ}} = \frac{\sum_i k_B T_i m_i r_i}{\sum_i k_B T_i m_i} \quad ; \tag{4.4}$$

where k_B is the Boltzmann's constant, m_i and T_i are the mass and temperature of each particle inside the radial shells considered. In the analysis, I have excluded particles from dense, low temperature, clouds ($T < 10^{5.2}$K and/or $n_e > 0.1$) and belonging to sub-haloes (in contrast to what is done with the mass density profiles). In this analysis, I ignore small relativistic corrections (e.g. Lee et al. (2020)).

In contrast to previous works, which used a gNFW profile to model the pressure profiles (Nagai et al., 2007; Arnaud et al., 2010; Battaglia et al., 2012b), I have fitted those profiles with a generalised Einasto profile, which has fewer free parameters.

TABLE 4.2: Best-fitting Einasto parameter values (see eqns. 2.7 and 3.4) describing the electron pressure profiles of the BAHAMAS for four halo mass definitions.

	$M_{\rm ref}[\mathrm{M}_\odot\, h^{-1}]$	F_0	f_m	f_z	A_0	a_m	a_z	α_0	α_m	α_z
$\Delta = 200m$	7×10^{13}	24.445	0.812	0.13	5.82	0.13	-0.066	0.77	-0.097	-0.235
$\Delta = 200c$	7×10^{13}	10.743	0.781	0.521	4.494	0.14	0.10	0.75	-0.10	-0.095
$\Delta = 500m$	7×10^{13}	16.379	0.708	0.0315	4.904	0.135	-0.10	0.68	-0.062	-0.30
$\Delta = 500c$	7×10^{13}	7.543	0.718	0.30	3.82	0.151	0.03	0.70	-0.091	-0.109

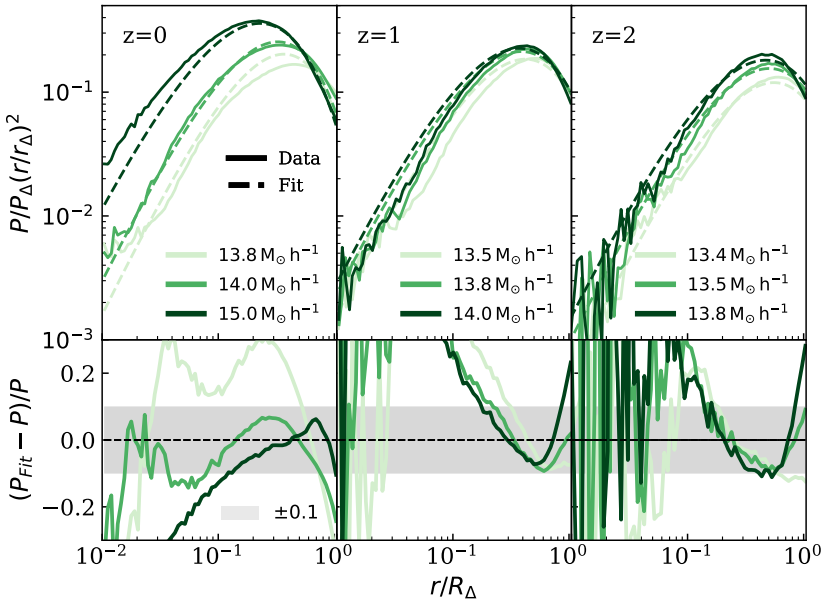

FIGURE 4.3: Same as fig. 3.1 but in the case of electron pressure profiles. The format of the plot is similar to the previous ones used to show the agreement between the fit and the stacked pressure profiles. In the top panels I show the comparison at three redshift bins for three separate mass bins. In the bottom, instead, I show the residual between the two models.

In Table 4.2 I present the best-fitting parameters for the Einasto fit of the pressure profile used in fig. 4.3 and fig. 4.4.

In fig. 4.3 I show the agreement between the stacked pressure profiles and the Einasto fitting functions using $\Delta = 200m$ mass definition. In the top panel I compare three different, well populated, mass bins for each redshift: $\log_{10}(M) = [13.8, 14, 15]\,\mathrm{M}_\odot\,h^{-1}$ for z=0, $\log_{10}(M) = [13.5, 13.8, 14]\,\mathrm{M}_\odot\,h^{-1}$ for z=0 and $\log_{10}(M) = [13.4, 13.5, 13.8]\,\mathrm{M}_\odot\,h^{-1}$ for z=2. The profiles are normalised by P_Δ and by (r/R_Δ) squared to reduce the dynamic range. P_Δ, also called 'virial pressure', is the self-similar pressure amplitude at R_Δ with Δ that identifies the mass definition, defined as:

$$\begin{aligned} P_\Delta &= n_{e,\Delta} k_B T_\Delta \ , \\ k_B T_\Delta &= \mu m_p \frac{G M_\Delta}{2 R_\Delta} \ , \\ n_{e,\Delta} &= \Delta \frac{\overline{\rho}(z) f_b}{\mu_e m_H} \ ; \end{aligned} \tag{4.5}$$

with f_b baryon fraction ($f_b = 0.165$) and $n_{e,\Delta}$ is the mean electron density inside R_Δ knowing the baryon fraction (Voit, 2005; Battaglia et al., 2012b; Le Brun et al., 2014; McCarthy et al., 2014). In the bottom panels I present the residuals as done in the companion plots (see figs. (3.1)(3.4)).

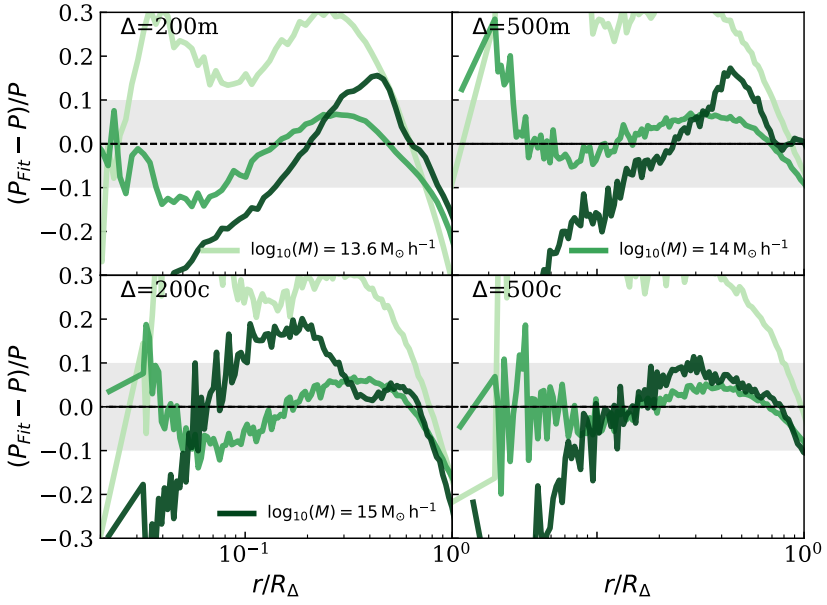

FIGURE 4.4: Residuals plot of the comparison between the stacked profiles and the fitting functions for the four mass definitions at z=0 in three mass bins (shown in different scales of green).

We see that the accuracy of the fit depends on the mass bins involved. In fact, larger haloes show less disturbance and a better fit compared to smaller ones (e.g. see the haloes of $7 \times 10^{13}\,\mathrm{M}_\odot\,h^{-1}$ at z=0). Smaller mass ranges might be influenced by ongoing AGN feedback and recent dynamical interactions. Those phenomena could have perturbed significantly the internal energy distribution. Overall, the Einasto function fits the simulated pressure profiles to typically 10-15% accuracy.

In fig. 4.4 I present, as done for the *DM-Only* and *AGN* profiles, a general comparison at z=0 for the four mass definitions of three mass bins. In this case I present the mass bins $\log_{10}(M) = [13.6, 14.0, 15.0]\,\mathrm{M}_\odot\,h^{-1}$ that are well reproduced in the simulation. In fact, the accuracy of the fitting is well below 10% for the largest mass bins while it is possible to see some variations (around 25%) for the lowest mass samples. Please note that for the $\Delta = 200m$ I have used a different, lower, mass bins compared to fig. 4.3.

In fig. 4.5 I present a test using the stacked electron profiles and the Einasto fit of the pressure for the four mass definitions at three redshifts bins. While discussing the density profiles, I checked that the density profile integrals were able to recover the total masses of the haloes ($\mathrm{M}(< R_\Delta)$). It is possible to do a similar test using the electron pressure profiles by comparing them with the total energy inside haloes.

However, the total energy budget in haloes is not fully modelled by uniquely the electron pressure. In fact, using M_Δ and the virial theorem, we would have a higher value for the total energy inside the haloes. So, to avoid these differences I use the integral of

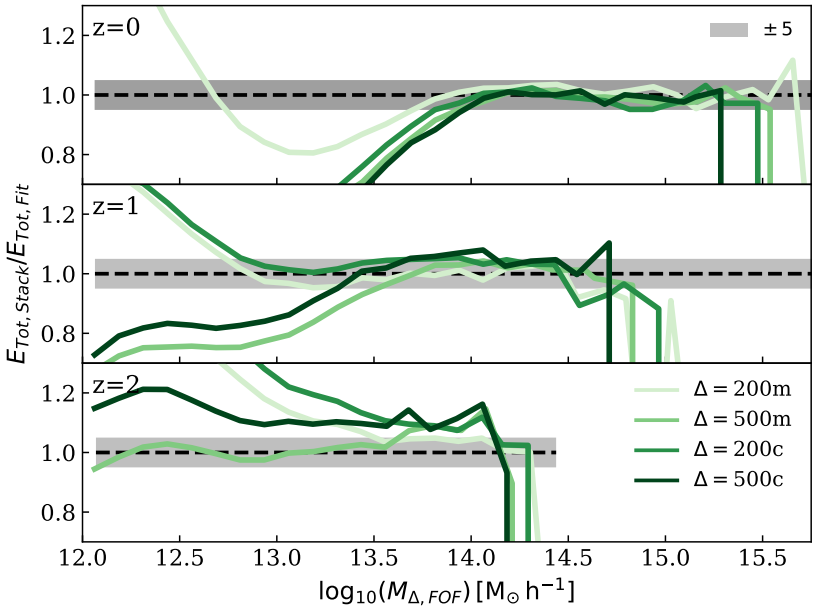

FIGURE 4.5: Integral comparison of the total energy, obtained by the electron pressure, between the stacked profiles and the Einasto fit for the four mass definitions and three different redshifts ($z = [0, 1, 2]$). In different scales of greens I present the four mass definitions.

the stacked electron pressure profiles, noted as $E_{\rm Tot,Stack}$ and I compare against the ones obtained using the fit profiles. I use different version of eqn. 3.5 to compute the value of $E_{\rm Tot}$ in this test:

$$E_{\rm Tot} = \int_0^{R_\Delta} 4\pi r^2 P(r, M, z) P_\Delta dr \ . \tag{4.6}$$

I remind the reader that I have a reduced integration limit of $10^{12} < M[\,{\rm M}_\odot\,h^{-1}] < 3\times10^{15}$ as discussed while presenting the tSZ power spectrum. It is possible to see that all models can reproduce well ($< 5\%$) at all redshifts in the mass range $\gtrsim 3\times10^{13}\,{\rm M}_\odot\,h^{-1}$ where the AGN influence is less strong to perturb the energy distribution inside haloes. Since the fit was done on stacked profiles, the injection of energy coming from dynamical effects such as accretion or minor mergers are most likely undetectable owing to the large number of particles present in each spherical shell. These effects can be seen also in fig. 4.4 with the larger discrepancy, in all mass definitions, at masses below $10^{14}\,{\rm M}_\odot\,h^{-1}$ (see also the fig. 4.3 for the case $\Delta = 200m$).

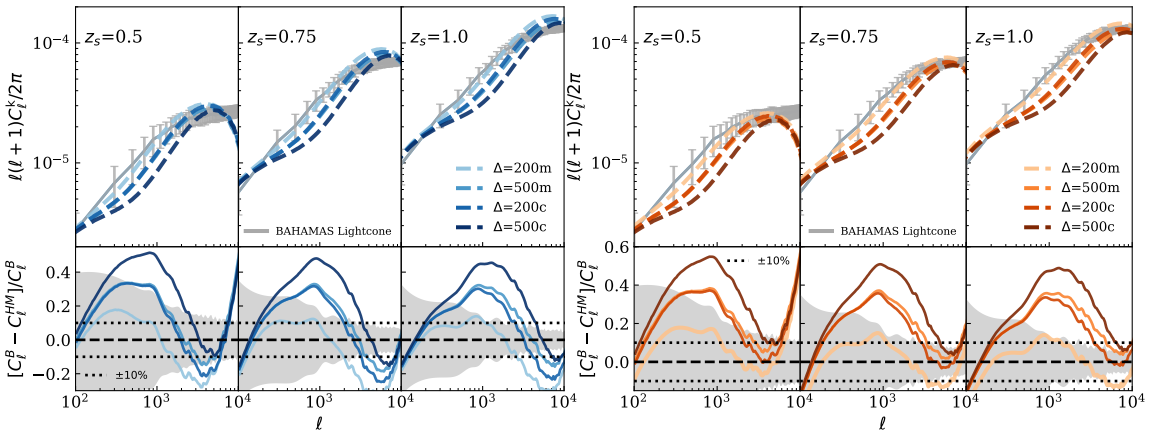

FIGURE 4.6: Single source plane distribution comparison between the *DM-Only* halo model predictions (left plot) and *AGN* (right plot) and the BAHAMAS maps for three different source plane locations ($z_s = [0.5, 0.75, 1.0]$). In different colours (blue for *DM-Only* case and orange for the *AGN*) are present the four different mass definitions used and in grey lines are shown the BAHAMAS maps prediction with the error bars associated with the scatter obtained from the 25 maps available. In the bottom panels I present the residual plot between the mean of the BAHAMAS maps and the halo model prediction using the same colours as before.

4.3 Weak lensing single source distribution test

In order to verify the halo model predictions presented above, I present some weak lensing tests carried out by making comparisons with the weak lensing maps built from BAHAMAS light-cones using the AGN TUNED WMAP9 runs.

The weak lensing light-cones, built using BAHAMAS simulations boxes, are presented in further detail in Section 2.4.1. Here, I present the convergence weak lensing maps made by a 'single source distribution'. I test the accuracy of the angular power spectrum built using both the *DM-Only* and *AGN* cases for three source planes $z_s = [0.5, 0.75, 1.0]$ using the four different mass definitions ($200m$, $500m$, $200c$, $500c$). The details of how to compute this observable are presented in Section 2.21.

The way I have built the matter-matter power spectrum for each of the four mass definitions is presented in Section 4.1.

In fig. 4.6 I present the comparison between three weak lensing maps computed for three separate single plane source distributions built using the BAHAMAS *AGN* simulations. The BAHAMAS (grey lines) results are presented with error-bars built with the scatter from the twenty-five light-cones available. The maps are built using the *AGN* BAHAMAS simulations. In the four different dashed lines in scales of blue and orange (*DM-Only* and *AGN* cases respectively) I compare the halo model predictions for the

four mass definitions to see how much the accuracy on the power spectrum reflects on the weak lensing predictions.

In the bottom panels, I present the residuals between the mean values of the BAHAMAS simulations against our predictions using the same colours scheme presented earlier. The 10% differences are highlighted using dotted black lines, while the intrinsic scatter from the twenty-five light-cones is shown with the grey area. Please note that residuals, presented as $(C(\ell)^B - C(\ell)^{HM})/C(\ell)^B$, are on different scales from the *AGN* to the *DM-Only* versions to allow a better visualisation.

To compute the weak lensing power spectrum I put a hard cut at $k > 10\,h\,\mathrm{Mpc}^{-1}$, which is the limit imposed to our halo model predictions (as discussed in the previous Chapter). This choice might be one of the reasons why there is a greater difference in $\ell \gtrsim 8000$, other explanation might be the accuracy in the prediction on the halo model itself, that at $k > 8\,h\mathrm{Mpc}^{-1}$ differs more than 10% from the simulation results.

As expected, the accuracy in reproducing the matter distribution has a strong influence on the overall accuracy in reproducing the weak lensing signal, especially in the range $200 < \ell < 3000$ corresponding to the 1-2 halo transition region. We see that, as the sources move to higher redshift, there is a more general agreement ($\approx 10\%$) on all scales while at lower redshift there is a significant difference at low multipoles and in the small-scale regime.

The different mass definitions have a similar and coherent behaviour: all models agree on the largest scales due to our efforts in matching the linear theory on the largest k-modes, and show a 15% agreement on the scales $2000 < \ell < 6000$, area dominated by the 1-halo signal. The sharp drop at smaller scales is mainly due to two factors: the halo model accuracy and the integration cut imposed. The differences in the *DM-Only* and *AGN* cases are more evident in the small scale regime, where the *DM-Only* have up to 15% more power compared to the simulation maps.

I have checked whether the number of matter power spectrum redshift realisations (three hundred, as stated earlier) used have an impact on the accuracy of the prediction but this seems not to be the case, achieving comparable results within a 2% difference on all scales.

The differences are still inside the scatter presented by the other maps, excluding only the small scale regime in the case of $z = 0.5$ which is not recovered well enough. I speculate that this is strongly affected by the k-cut applied at the halo model matter power spectrum and, partially, by the general poorer agreement between the halo model predictions and BAHAMAS results (as could be seen in Section 3.4).

4.4 tSZ effect

In this section, I present the tSZ halo model predictions using the fit of the electron pressure profiles described in Section 4.2.2 comparing against BAHAMAS tSZ maps, presented in Section 2.4.1 and observational results.

To compute the angular tSZ power spectrum, I integrate haloes from $0 < z < 3$ and in a mass range $10^{12} < M\,[\mathrm{M}_\odot\,h^{-1}] < 3 \times 10^{15}$, which is a smaller range compared to the matter-matter power spectrum presented earlier in Section 4.1. Massive ($\approx 3 \times 10^{15}\,\mathrm{M}_\odot\,h^{-1}$) clusters at low redshift have a strong influence on the characterisation of the largest modes of this observable, providing a non-negligible amount of power, indeed those are often excluded from tSZ power spectrum analysis. To avoid this issue, I have cut out integration up to haloes masses of that mass values[6].

In this analysis, I present two different approaches: one when keeping the halo mass fixed and a second where I integrate the halo mass using the density profiles computed in this work, for the four different mass definitions and for the *DM-Only* and *AGN* cases accordingly. These approaches differ due to the different link between the halo masses and the pressure profiles associated: in the second instance we use a fixed index (the halo mass) to identify both the density and pressure profiles, the density profiles then provide a more realistic halo masses that better fit the pressure profiles identified. With this approach, I can enhance further the baryonic effects.

In fig. 4.7 I present the halo model results in comparison with BAHAMAS light-cone maps (McCarthy et al., 2018), *Planck*13 results (grey diamonds, Planck Collaboration (2014)), *Planck*15 (black squares, Planck Collaboration et al. (2016)) and the more recent South Pole Telescope (SPT) measurements (George et al., 2015) and Atacama Cosmology Telescope (ACT, Sievers et al. (2013), upward arrow and downward arrow). The observational data from Planck survey, ACT and SPT are elaborated by the original authors. The latter measurements presented are an independent constrain at $\ell \approx 3000$ and are consistent one with the other. Instead, the two Planck power spectra show clearer differences at multipoles below ≈ 1000. The error bars presented are dominated by the systematic foreground subtraction uncertainty (due to the presence of infrared contamination such intergalactic dust, point sources and cosmic infrared background - CIB, see the relevant discussion in the papers cited above).

In the bottom panel, I present the residuals between the mean values obtained from BAHAMAS maps and the halo model predictions. I show in the grey shaded area the

[6] e.g., Virgo and Coma clusters are masked in the Planck analysis when computing the tSZ power spectrum.

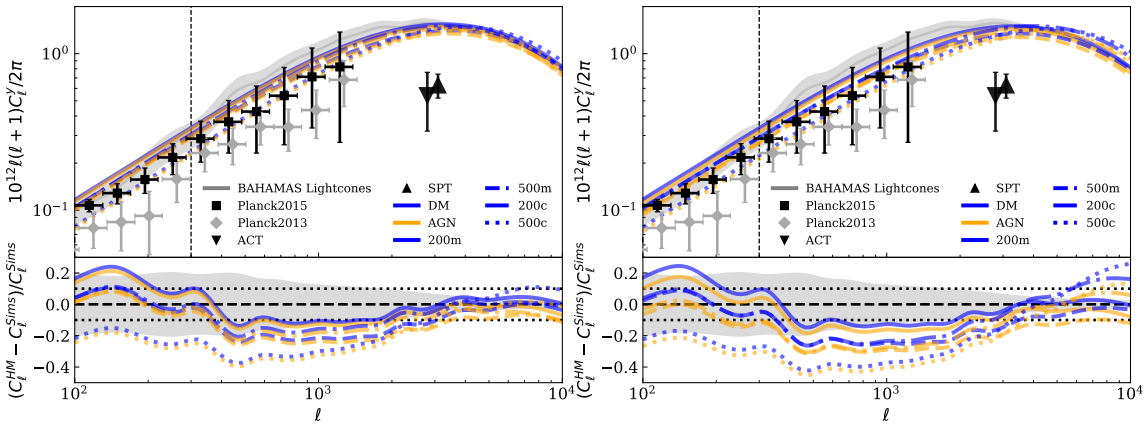

FIGURE 4.7: In the left plot, I have computed the tSZ using the fix halo mass, as discussed in the text, and in the right plot, I have used the masses integrated using the matter density profiles. The structure of each plot is the same and follows:
Top: tSZ auto-power spectrum comparison between the calibrated halo model predictions for the four mass definitions (blue for the *DM-Only* predictions and orange for the *AGN*, and four different line styles for each overdensities), the BAHAMAS light-cone results (grey lines, the line is the mean value of the twenty-five realisations and the shaded area is the error bars associated), *Planck*13 (grey diamonds) and *Planck*15 (black squares) observational results and ACT and SPT measurements (black arrows). The vertical dashed line highlights the scales of $\ell = 300$, below which the simulation finite box size influences the results.
Bottom: residuals, computed as $(C_\ell^{HM} - C_\ell^{\rm Sims})/C_\ell^{\rm Sims}$, between the BAHAMAS mean values and the halo model predictions (with the same colours and line styles as before) with shaded region computed using the scatter from the light-cones. In dotted black lines I show the 10% (± 0.1) agreement.

scatter obtained using the error bars measured on the twenty-five maps available. The dotted lines highlight the 10% differences.

I find that, as shown previously by other works (Horowitz & Seljak, 2017; Bolliet et al., 2018), the 2-halo term is sub-dominant in this observable, accounting for at most 10% of the total spectrum in the large scale limit ($\ell < 400$, here for clarity I have plotted only the total power spectrum without highlighting the two separate components). This large scale limit is a difficult region to quantify properly because it is both at the limit of light-cones size and it is strongly influenced by cosmic variance, which can contribute up $\approx 20\%$. Additionally, in observations, these scales are hugely influenced by massive low-redshift clusters (Planck Collaboration et al., 2016).

The halo model predictions for the four mass definitions are interesting. All predictions are in line with the expected results constraining the peak of the distribution better than 10% when using the fixed halo masses, for both the *AGN* and *DM-Only* cases, while in the integrated case (right plot) there is a slightly larger scatter. The increased difference in the small-scale regime we see using $\Delta = 500c$ mass definition, can be

explained, partially, by the fact that the density profiles overpredict (see fig. 3.3) the halo masses at high-redshift. The larger mass definitions (e.g. $200m$) provide a better representation of the power spectrum leading to a maximum 10% difference in the scales $700 < \ell < 2000$. These differences, which are present in both fix and integrated mass approaches and for each mass definition, can be associated with the lack of inclusion of outer boundaries of haloes which are a source of some clustering signal.

The peak at $\ell \approx 3000$ is influenced both by the choice of the cosmological parameters (Bolliet et al., 2018) as well as modelling of astrophysical processes (e.g., stellar and AGN feedback) and the minimum mass integration limits. The choice of integrating from $10^{12}\,\mathrm{M}_{\odot}\,h^{-1}$ is a conservative choice because smaller haloes do not contribute significantly to the tSZ power spectrum (see for reference Battaglia et al. (2012b), where the influence of different mass bins is clearly highlighted in the shaping of the power spectrum, as well in McCarthy et al. (2014) in a simulation context). Also the ability to fully reconstruct the AGN feedback is a key factor in this spectrum region (Battaglia et al., 2012a; McCarthy et al., 2014, 2017, 2018).

The two approaches on the use of fixed or integrated halo masses lead to comparable results, but with some relevant differences. As explained above, the different use of the halo mass to identify the profiles causes not only a shift in the prediction but also larger differences between the *AGN* and *DM-Only* cases. Indeed, the different mass values do reflect also on the HMFs leading to an increased difference between the two sets of predictions.

Using a *DM-Only* or a baryonically corrected HMF reduces the power spectrum values in a nearly scale-independent way. Therefore, the combined use of an accurate halo mass and the correction on the HMFs has an effect that is not a mere shift of the prediction but results in an enhancement of the baryonic feedback on this observable. A more detailed discussion about the feedback will be done in the next Sections.

The agreement with observations is more complex. The largest mass definitions predictions are within the uncertainties of the Planck measurements in the range $400 < \ell < 1200$, while smaller halo definitions seem to be in more agreement with *Planck*13 and/or *Planck*15 results. The differences between the two Planck data releases are tightly bound to the different treatment of cosmic dust and infrared background contamination. Indeed, Bolliet et al. (2018) examined the noise subtraction closely in the latest Planck result and found that it was closer to the previous (2013 data release) power spectrum result. Regarding the ACT and SPT measurements, these are not recovered at all, coherently with other BAHAMAS results.

One cannot choose the overdensity to match the data. What we should be doing, instead, is using all gas that contributes to the total power spectrum, independently from the mass definition. Matching, in this way, what is observed. Using larger mass definitions, e.g. the $\Delta = 200m$, could be the starting point for adding up even larger scales to obtain a fully convergent result. As said, the tSZ signal is not only coming from the central parts of the haloes but also from the outskirts, that in the light-cones measurements are considered and it results in an enhancing of the signal measured. It is, also, known that the calibration of halo masses is quite complex and the introduction of mass biases are quite common while trying to match the Planck modelling of the measurements (see the discussion reported while presenting the tSZ power spectrum in Section 1.33).

With all this information aligned, it is difficult to draw a conclusive solution: the halo model predictions can help understand the influence of the right baryon modelling and explore how different feedback approach can influence the ability to fit the data, but there is the need for novel and high-quality data from observations.

4.5 Weak lensing results

After having presented the weak lensing tests using single source redshift distribution in Section 4.3, those were done to investigate the influence of the overall accuracy of the halo model matter distribution in weak lensing analysis. Below I explore the weak lensing measurements in auto and cross-correlations comparing against real observations as well as BAHAMAS results.

In this Section, I present weak lensing predictions for different observational surveys (namely CFHTLenS and tomographic analysis using KiDS-450). The interest in these analyses is two-fold: on one side I am interested in seeing how much the internal matter power spectrum accuracy impacts the overall accuracy in weak lensing analysis and on the other side, making use of different tomographic redshift bins, I can highlight how the accuracy varies as a function of redshift. Furthermore, using the *DM-Only* and *AGN* versions of the halo model, I can constrain the baryonic effects and understand at what scales those take place and influence the predictions.

In fig. 4.8 I present the angular weak lensing predictions using the CFHTLenS redshift source distribution. I compare against BAHAMAS results presented in McCarthy et al. (2018) paper (solid grey line) and CFHTLenS observational data (black points) from Kilbinger et al. (2013). In blue and orange lines I present the halo model predictions using the *DM-Only* and *AGN* matter distributions presented in this work. The $C(\ell)$, weak lensing convergence power spectrum, is computed using eqn. 2.21 and the source redshift

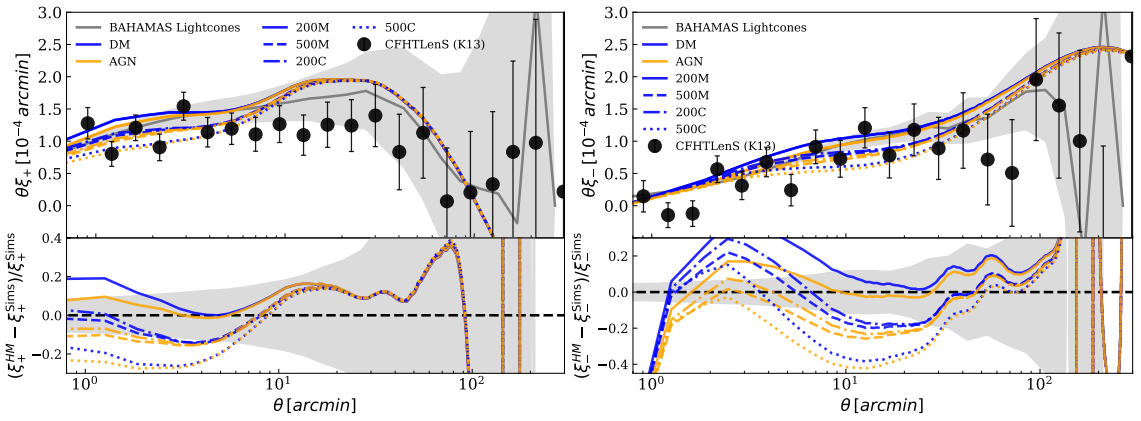

FIGURE 4.8: Weak lensing ξ_+ (left) and ξ_- (right plot) predictions using the CFHTLenS redshift source distribution, comparing against observational data (black dots, Kilbinger et al. (2013)) and BAHAMAS predictions (solid grey lines). In different colours (blue and orange) I present the halo model predictions for the *DM-Only* and *AGN* cases, and in different line styles the four mass definitions. In the bottom panels I show the residuals between the BAHAMAS predictions and the halo model, colour coded and with the same line style as the panel above. The shaded grey area is the scatter associated with the BAHAMAS maps used.

distribution is obtained using eqn. 4.1. The CFHTLenS source redshift distribution is shown in fig. 4.1 (and the relative lensing kernel in fig. 4.2). In order to convert from convergence power spectrum to angular correlation ($\xi_\pm$) I have used eqn. 2.23. Multiplying $\xi_\pm$ by the angular distance, θ, I reduce the dynamic range allowing an easier model-to-data comparison.

The agreement between the data and the models is good for both statistics, but there are some interesting differences. Those arise from the separate effects obtained from the tangential or cross values of the weak lensing effect. The ξ_+ case shows a good agreement (15%) in the range $1 < \theta\,[\text{arcmins}] < 100$ with both the data and the simulations, please note that both the BAHAMAS and the halo model overpredict the observational values in the scales $8 < \theta\,[\text{arcmins}] < 40$. At larger angular distances, we see a sharp drop in both simulations and halo model predictions and an increase of uncertainty, highlighted by the shaded grey area. The largest scales are more difficult to model due to the finite size of the light-cones and the cut at large scale of $k \approx 0.01\,h\,\text{Mpc}^{-1}$ of the halo model.

Differently, we do not see any of this larger discrepancy on the large scale in the ξ_- case. However, we see a wider spread in the range of $1 < \theta\,[\text{arcmins}] < 25$ between the different mass definitions. This observables is more influenced by small scale effects as can be visible in the scale of the baryonic suppression. In the large scale regime, instead, all predictions from the different mass definitions collapse to a similar result. The sharp drop at $\theta < 1$ arcmin in the residuals is mainly due to the few points sampled in the

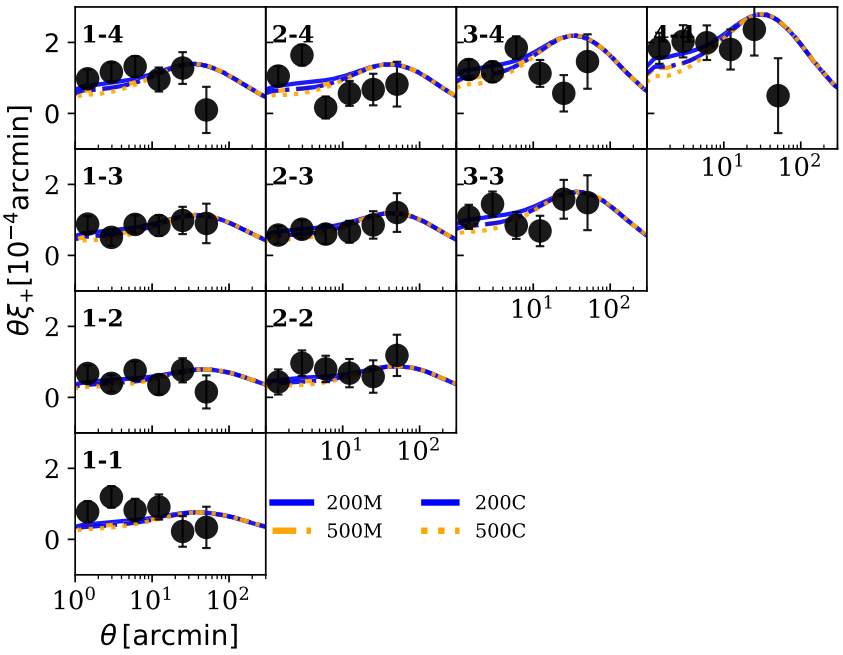

FIGURE 4.9: KiDS-450 tomographic weak lensing ξ_+ measurements (Köhlinger et al., 2017) with *DM-Only* halo model predictions with the four mass definitions. The different tomographic bins are highlighted in top right of each panel. I do not show the *AGN* version of the predictions for plotting clarity.

halo model and simulations, in fact, in the main panel it is possible to see that they share a similar trend.

On the large scales, we see that the mass definitions behave similarly in both cases: the fiducial setup is in accord with the predictions of the simulations only for a small portion of the spectrum surveyed, with the *DM-Only* case which overpredicts the largest modes (as expected). The $\Delta = 200c$ and $\Delta = 500m$ share a similar trend with differences of 20% in these scales. We note, as well, that in both statistics the behaviour is coherent with the expectations.

4.5.1 Tomographic weak lensing with KiDS-450 data

To explore the growth of structure and understand how the Universe evolves weak lensing tomography techniques provide a uniquely powerful tool. This method is based on the calculation of the convergence power spectrum using different source redshift distributions for each ij bin determining the lensing kernels g_{ij}. To explore this possibility I have used the four tomographic bins from the KiDS-450 survey shown in fig. 4.1 (as well as the lensing kernels in fig. 4.2) in Section 4.2.1.

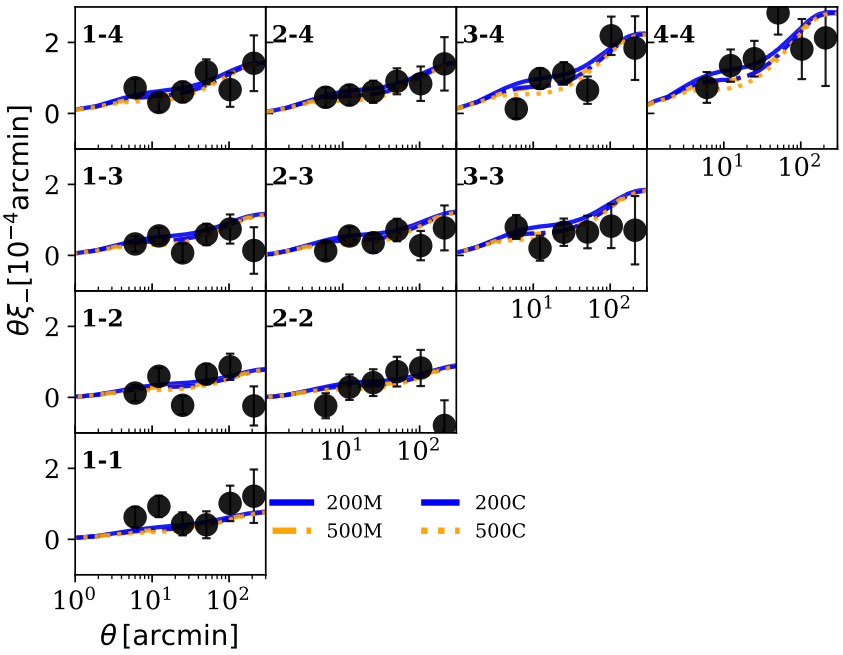

FIGURE 4.10: Same as fig. 4.9 but for ξ_- cases.

In figs. (4.9)(4.10) I present the tomographic weak lensing analysis using the four source redshift bins of KiDS-450 survey (Köhlinger et al., 2017) data and I compare the halo model predictions using the *DM-Only* case with the four mass definitions. I do not show the *AGN* predictions only for plotting clarity, as can be seen in the previous weak lensing analyses, they show very little differences compared to the *DM-Only* counterparts. The KiDS-450 data are shown in black dots with error bars (Hildebrandt et al., 2017). The triangular plot shows the cross-correlations between the four different redshift bins identified from 1 to 4. In the case where $i = j$ (e.g. 1-1 or 2-2) those are auto-correlations of each redshift bin.

We see that the auto-correlations at high redshift (4-4, with distribution peaked at $z \approx 0.9$) shows larger differences between the four overdensity definitions analysed. One explanation can be traced to the peak at high redshift of the distribution, where the different levels of accuracy in reproducing the matter clustering in the models influence these results. In general terms, it is possible to see that at all bins the different mass definitions trace quite well the data with major differences only at scales $\theta < 3$ arcmins, where the models differentiate the most.

In the other bins, both in auto and cross-correlation cases there is good agreement on all scales, with some larger mass definitions that are significantly higher compared to the actual data. This result is in agreement with what was found in McCarthy et al. (2018).

I find that the large scale limit ($\theta > 100$ arcmins) is slightly higher compared to trend of the data for the ξ_+ case, similar to what was evident in the CFHTLenS analysis[7].

Taking into consideration the last two analyses using weak lensing statistics, it is possible to explore further the impact of the mass definitions, meaning the influence of the halo model accuracy in the matter clustering and its effect on observational comparisons. In general terms, the ability of the halo model to reproduce to a competitive accuracy the matter distribution is visible and verifiable only in the ξ_- cases, which appears to be less sensitive to the broader inaccuracies and redshift dependencies. ξ_+ cases, on the other hand, highlight clearly how the discrepancies in the matter clustering are forecast in this observable. The weak lensing data used in this Thesis have themselves broad uncertainties that do not allow to draw clear conclusions on the ability to distinguish between accuracy or baryonic effects from the halo model. More recent data (e.g., DES or KiDS-1000) and upcoming results will help ease this task.

4.6 Weak lensing cross-correlations

With tomographic weak lensing analysis, I have introduced the role of cross-correlating different source distributions or different observables. The power of using this method is in the possibility of exploring further the composition of the Universe and seeing more in detail how much different probes explore the matter clustering in the Universe. In this Section I will explore the main weak lensing cross-correlations presented in theoretical details in Section 2.3.3 and then move on to the case of tSZ-weak lensing.

Weak lensing cross-correlations between CMB source and photons coming from galaxies at different redshifts is one of the main tools that can probe different epochs of the Universe, highlighting the growth of structure and constraining the geometry of the Universe. Thanks to the number of weak lensing surveys available and the detailed Planck results, this procedure has been applied multiple times, and it is expected to be a significant tracer with next-generation observations.

In fig. 4.11 I present the halo model predictions for the CMB - galaxy weak lensing cross-correlations (called also $k_\phi \times k$) using a wide unique bin made from the four tomographic bins of KiDS-450, in orange lines, I present the halo model *AGN* predictions while *DM-Only* predictions are depicted using blue lines. The observational data points are computed using KiDS-2D (from KiDS-450 data release, Harnois-Déraps et al. (2017)) and *Planck*15 data. The grey line represents the mean of the maps built from the BAHAMAS light-cones and in grey shaded area the scatter associated with those maps.

[7]Please note that the halo model predictions in the large-scale limit are also slightly higher compared to the BAHAMAS results, see McCarthy et al. (2018)

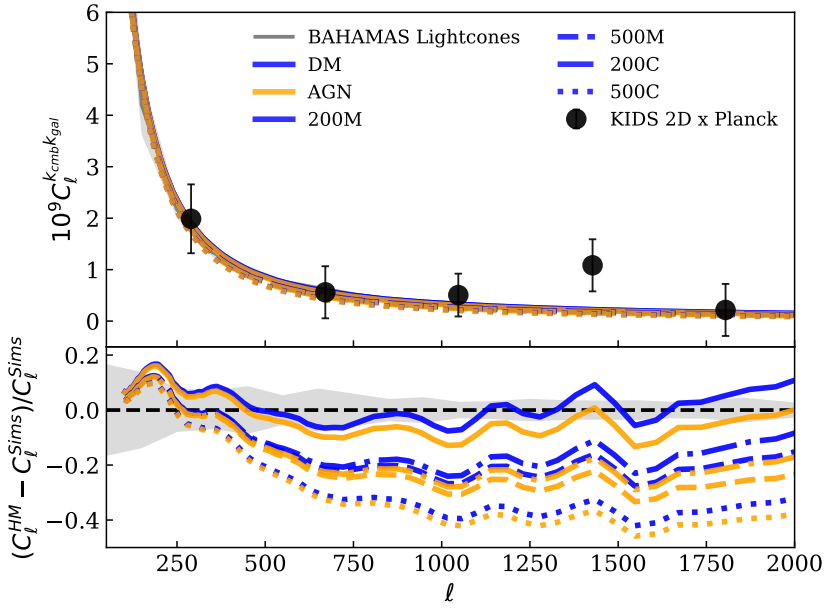

FIGURE 4.11: CMB-KiDS weak lensing cross-correlations ($k_\phi \times k$) comparison between the observations (black dots, Harnois-Déraps et al. (2017)), the BAHAMAS light-cone predictions (grey lines) and halo model *DM-Only* (blue) and *AGN* (orange lines) for the four different mass definitions (as presented earlier). In the bottom panel I present the residual between the halo model prediction and the BAHAMAS light-cones, with the grey shaded area highlighting the scatter of the maps used.

With the halo model cut imposed ($\ell < 100$), it is not possible to see the peak of the distribution that is around $\ell \approx 70$, where the Limber approximation is not solid enough (see in Chapter 2 the discussion about the limits of the Limber approximation, predominantly in CMB analysis). All mass definitions trace well the data point distribution, so it is quite complicated to distinguish between the different baryonic effects or levels of accuracy due to the separate mass definitions. In the residual plot with the calibrated BAHAMAS maps, it is possible to see that the fiducial model $\Delta = 200m$ seems to predict well on all scales, by $\approx 10\%$, the simulation results. The collisionless case is in slightly better agreement with the data and BAHAMAS maps. The $200c$ and $500m$ cases present a lower level of agreement (25%) and share similar trend, even lower, is the $500c$ case, which has shown a poorer ability in reproducing the matter clustering effectively.

All halo models predictions (and simulations as well) are below the data point at $\ell \approx 1400$ which is slightly (1.5σ) over the other data points.

4.6.1 tSZ cross-correlations

After having presented the weak lensing cross-correlations, I analyse here the cross-correlations with the tSZ power spectrum ($y \times k_\phi$), which had shown, already, some interesting features. As previously done for the tSZ power spectrum, I analyse two separate cases: one with the fixed halo mass and the other with the halo masses obtained by integrating the density profiles. In principle, the latter approach should be a fairer comparison with weak lensing measurements, since the masses are obtained using the, aforementioned, density profiles.

To continue making the comparison with McCarthy et al. (2018) paper, I should take into consideration cross-correlations between the tSZ and both CMB and galaxy lensing (in this case, we use RCSLenS because this survey has more overlap with the Planck maps and targets more clusters of galaxies, compared to KiDS or CFHTLenS surveys). However, I will focus only on the first presented since it has less mask-modelling required to match the results from observations. Galaxy lensing - tSZ cross-correlations, instead, are required to undergo several matching and masking iterations due to Planck beam size effect, that alters the signal above $\ell \approx 1000$, see, for instance, the discussion about this observable in Harnois-Déraps et al. 2015; Hojjati et al. 2017 and more recently with the first of the papers on the results using DES year 3[8] and Planck tSZ measurements (Gatti et al., 2021). The CMB weak lensing- tSZ cross-correlation, instead, does not require much modelling since the data comes from the 'same' instrument allowing to ignore these mask-modelling related issues.

In any case, I will present the baryonic suppression as well for the galaxy lensing - tSZ case ($y \times k$) since, mask modelling aside, the signal from the baryons should be distinguishable and measurable and, as we will see, different from the comparison with the CMB weak lensing and tSZ.

In fig. 4.12 I present the cross-correlation predictions using the halo model by comparing with the BAHAMAS simulations (grey lines) and observational results from Planck observations (Hill & Spergel, 2014). In the left plot, I present the result using a fixed halo mass when computing the tSZ power spectrum and, on the right-hand side, the ones obtained by computing the total halo mass by integrating the density profiles.

It is possible to see that the BAHAMAS results are a narrow band that is well reproduced by the largest halo mass definitions. In particular, there is a good agreement ($< 10\%$) when *AGN* HMF is adopted (please remember that the light-cones were built on the calibrated AGN BAHAMAS runs). It is interesting to see that using different mass definitions the power spectrum is shifted to lower values, which is something similar

[8] Dark Energy Survey, https://www.darkenergysurvey.org/.

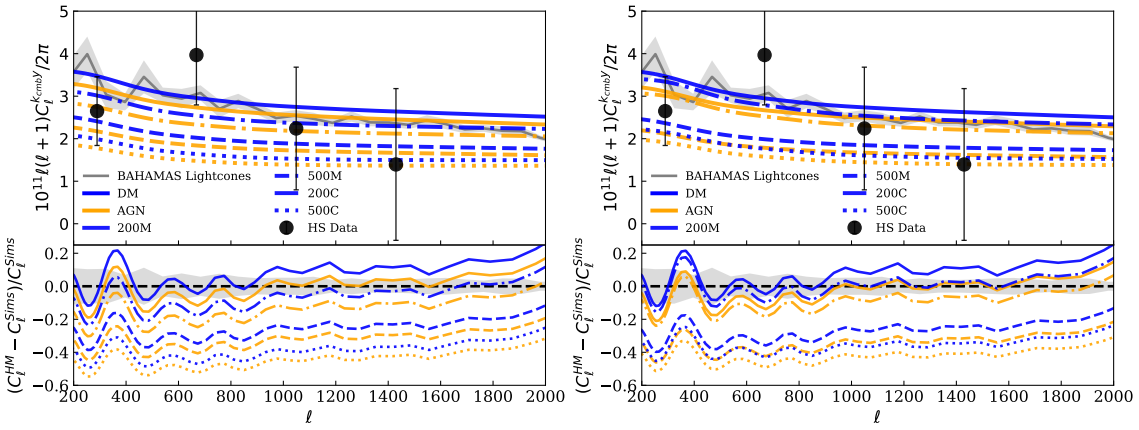

FIGURE 4.12: Cross correlation between the thermal Sunyaev-Zel'dovich and CMB weak lensing. As done previously in fig. 4.7 I explore on the left plot the predictions using the tSZ with fixed mass and on the right the one using the halo masses inferred from the integral of the density profiles for both the *DM-Only* and *AGN* cases. The grey lines (and shaded area) represents the BAHAMAS light-cone results (with scatter), the black dots the data from Hill & Spergel (2014) and the halo model predictions are shown in blue for the *DM-Only* cases and orange for the *AGN*. The different line styles represents the different mass definitions. In the bottom panel I present the residuals between the halo model predictions and the BAHAMAS results. The main difference between using the fix mass (left plot) and the integrated mass (right plot) is that the largest mass definitions ($200m$ and $200c$) tend to have a similar accuracy in the latter case. Using a fix mass, instead, each prediction shifts to lower and lower accuracy (all inside the data measurements) instead of being more clustered.

I have found in the tSZ modelling in section 4.4. This can be explained by the way the halo model, particularly the tSZ 1-halo component, traces efficiently the electron pressure distribution inside haloes and the broad impact on the size and outskirts of haloes.

It is, also, interesting to see that the use of fixed masses or the integrated ones provide some differences in the way the predictions match the expected results. With the integrated masses the largest mass definitions are more clustered, same as for the smaller mass definitions ($500m$ and $500c$). While, for the fixed masses, there is a broader spread of the predictions making it, on insight, more difficult to distinguish from other physical effects (similar suppression can be associated with neutrino feedback, see again McCarthy et al. (2018)).

Concluding the analyses on the cross-correlations explored, there are two interesting points to discuss: the pressure modelling in the tSZ shows broader and more complex effects on the probes sampled (together with baryonic effects). Weak lensing probes present, instead, a more clear baryonic signature and independent from other effects surveyed here (mainly the accuracy of the matter clustering). Those probes together,

with a more sophisticated and thorough approach, can, in principle, provide better constraining power to some modelling and baryonic uncertainties.

4.7 LSS baryonic suppression

In the previous sections, I have explored extensively the auto and cross-correlations of different LSS tracers using the four mass definitions and using collisionless and hydrodynamical setups. To quantify further the relative impact of baryons, as I have shown in Section 3.4.3, I analyse this quantity in the LSS probes surveyed highlighting where and how much the baryons play a role.

As stated earlier, the flaws in the modelling can bias, quite significantly, the LSS predictions, nevertheless, the use of the ratio between the *DM-Only* and *AGN* predictions could avoid the introduction of further uncertainties. These predictions can be used to quantify the impact of baryons and validate how much galaxy feedback can influence the relative cosmological and astrophysical parameters estimation.

In fig. 4.13 I present the baryon effects in the previously presented LSS probes. In the top left panel, I present the tSZ (noted also as $y \times y$) power spectrum, in the top right the CMB weak lensing ($k_\Phi \times k_\Phi$). In the centre-left the tSZ-CMB weak lensing cross-correlation ($y \times k_\Phi$), in the centre-right the weak lensing power spectra from both auto-correlation ($k \times k$) cases as the CFHTLenS cases (blue lines) and cross-correlations using the KiDS-450 bins (examples in red and orange lines). Bottom left, tSZ-Galaxy lensing power spectrum ($y \times k$, using RCSLenS source redshift distribution presented in eqn. 4.2 and fig. 4.2) and in bottom right the galaxy - CMB weak lensing ($k \times k_\Phi$). Please note that the top panels span from $10^2 < \ell < 10^4$, while the cross-correlation cases with tSZ and CMB weak lensing are presented up to a maximum of $\ell \approx 2500$ (for observational reasons due to Planck beam size) and the cosmic shear values are in angular scales $0.7 < \theta < 300$ arcmins.

In the tSZ panels, I present, again, the two separate approaches on the masses: one with just the correction on the halo mass function (blue line) with the fixed halo masses and the other with also a mass correction (orange lines). In the latter, the mass correction is computed by using the density profiles and integrating those up to R_Δ obtaining the final M_Δ mass using the *DM-Only* and *AGN* density profiles. While the first case presented, I have kept fixed the halo masses and changed only the halo mass function accordingly, as presented and explored in Section 4.4.

In the weak lensing panel, I present the results for the auto-correlations using CFHTLenS source redshift distribution (blue lines) and auto and cross-correlation examples using

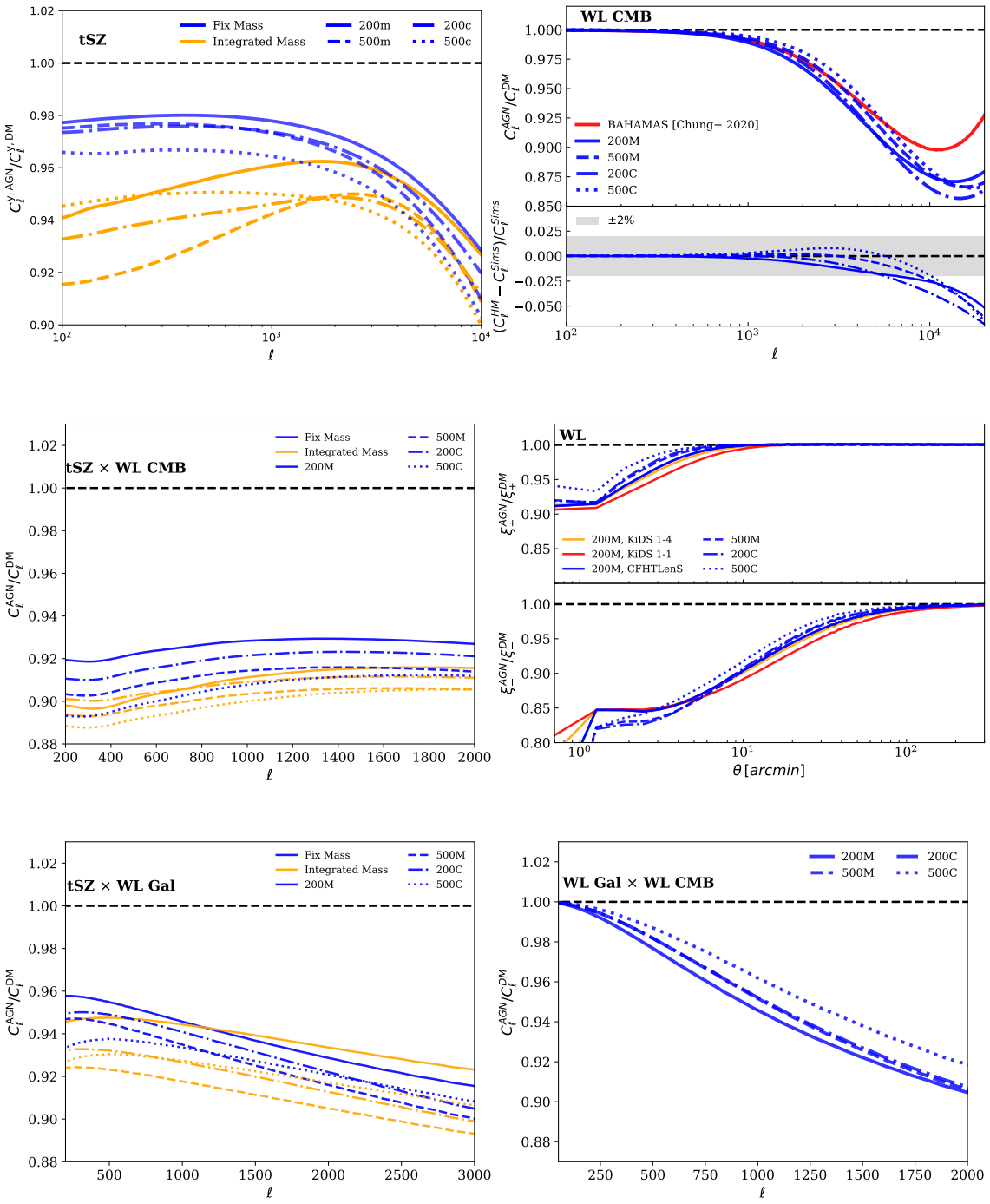

FIGURE 4.13: Baryonic effect in several LSS tracers. In the top left panel I present the tSZ case with the two approaches used. In the top right panel I present the CMB weak lensing power spectrum suppression comparing against BAHAMAS predictions from Chung et al. (2020) and in the bottom panel the ratio between the halo model predictions and the simulations. Central left plot I present the cross-correlation between the tSZ and CMB weak lensing. In the central right plot the weak lensing suppression ($\xi_{\pm}$). In the bottom left plot I present cross-correlation between tSZ and galaxy weak lensing (using RCSLenS source redshift distribution). In the bottom right plot I present the cross correlation between the CMB weak lensing and galaxy lensing using KiDS-2D source redshift distribution.

different KiDS-450 tomographic bins (red lines for 1-1 and orange lines for 1-4). The full KiDS-450 suppression analysis is presented in fig. 4.14 for the ξ_+ case and fig. 4.15 for the other. I present these two examples just as a comparison of the different scales and effects that can be measured and are not fully appreciable in the triangular plots below.

In the CMB weak lensing suppression (top right plot), I compare against the BAHAMAS predictions computed by Chung et al. (2020) that have made predictions for upcoming next-generation Stage IV CMB experiments. In this case the baryons have an effect in the small scale regime ($\ell > 5000$) leading to a suppression that is approximately comparable to the mean value of the boost factors obtained comparing the *DM-Only* and *AGN* at different redshifts as presented in Section 3.4.3. The agreement between the halo model predictions, for the four different mass definitions, and the simulations expectation is in line with the agreement shown while analysing $P(k)$ suppression presented at different epochs (see again Section 3.4.3 and fig. 3.16), reaching a 3% agreement on all scales below $\ell \approx 9000$. The bottom panel shows the residuals where we appreciate an increase of difference above $\ell \approx 9000$. These increasing differences could be due to multiple factors: the low mass cut applied ($M > 4 \times 10^{11}\,\mathrm{M}_\odot h^{-1}$) and the limit of the Nyquist frequency, as well as errors in the suppression that are carried on in the halo model predictions. Effectively quantifying the effect of baryons physics can help identify cosmological imprints on this probe, such as neutrino feedback (Osato et al., 2016; Green et al., 2021).

In the tSZ plots (left side plots, both considering auto and cross-correlations), it is interesting to note the different results for the use of the fixed mass and the integrated mass. The fixed mass case shows an almost uniform suppression on all scales, excluding the sharp drop at $\ell > 5000$ in the $y \times y$ case where the peak of the distribution is already reached. In the $y \times y$ case, the suppression is around 3% for all mass definitions for a large part of the spectrum, indicating that the impact of the HMF accounted for the baryonic effect is important but does not change the predictions significantly. The cross-correlations with weak lensing observations, however, highlight some interesting features. The case with CMB weak lensing ($y \times k_\phi$), shows a constant suppression of around 9% for all mass definitions for the ℓ scales taken into consideration (similar to the flat suppression found in tSZ case on the same scales) while in the case of galaxy lensing ($y \times k$) the suppression seems to rise at smaller scales, from a starting $\approx 5\%$ to a final $\approx 9\%$.

Using the integrated masses the picture is much more varied. For the $y \times y$ cases, it is possible to see that the large scale limit is more suppressed, with a mean value of $\approx 7\%$ for the four solutions provided, and grows back to a 5% at $\ell \approx 5000$ where, as

shown before, there is a sharp decrease. In the cross-correlation cases the suppression is similar to the other approach but higher, with values between 2 to 3% higher, enhancing a double effect on the baryonic suppression.

The $k \times k$ analysis (central right plot) shows the weak lensing statistics ($\xi_{\pm}$) using the CFHTLenS source redshift distribution and the tomographic bins of KiDS-450. As it was already evident in the comparisons against the observational data, the two statistics are sensitive to different parts of the spectrum showing a different level of suppression. The ξ_+ cases present a suppression only at angular scales below ≈ 7 arcmins, while for ξ_- the suppression evident on all the sampled range showing also an increase at the smallest angular scales.

Even using the tomographic bins, the effect of baryons does not vary significantly with redshift. I present two auto-correlations one at low redshift (1-1, in red line) and one higher (1-4 in orange line) to highlight that the baryon suppression seems not affected by the source distribution applied. In fact, these results, shown for clarity only for $\Delta = 200m$, are 1-2% different to the results from CFHTLenS, highlighting that the baryonic effects are smoothed in different redshift bins.

It is possible to verify that the differences between the *DM-Only* and *AGN* cases are very small in $\theta > 10$ arcmins for the ξ_+ and $\theta > 30$ arcmins for the ξ_- case. The suppression takes place at different θ scales for each of those probes providing, thus a powerful tool to understand where it will be possible to distinguish baryonic effects from cosmological inference. The suppression happens in the small scale regimes that is difficult to distinguish effectively from uncertainties (Joudaki et al., 2017). A recent analysis of DES Year 1 results (Huang et al., 2021) and an extension of KiDS-450 (Yoon & Jee, 2021) have also found similar and comparable results as ours using different baryonic feedback and modelling, opening up to further applications in future observations with more constraining power.

In the last panel, I analyse the galaxy-CMB weak lensing cross-correlations ($k \times k_\phi$). It is interesting to see that this probe is quite different from the others since the major differences between the *DM-Only* and *AGN* cases start at relatively small scales ($\ell < 200$) and grow back to 10% already at $\ell \approx 2000$ where there are hints of a flattening of the curve. This effect might be due to the combined effect from the suppression that takes place in the CMB signal, which is more sensitive to the distribution of the matter in the Universe and the cross-correlation with galaxies distributed at different epochs. This result is particularly interesting since it might open to a variety of applications in the near future with new observations available. This suppression is also higher at the same scales when compared to other weak lensing statistics surveyed. This result reinforces the narrative behind the great significance and power of cross-correlations.

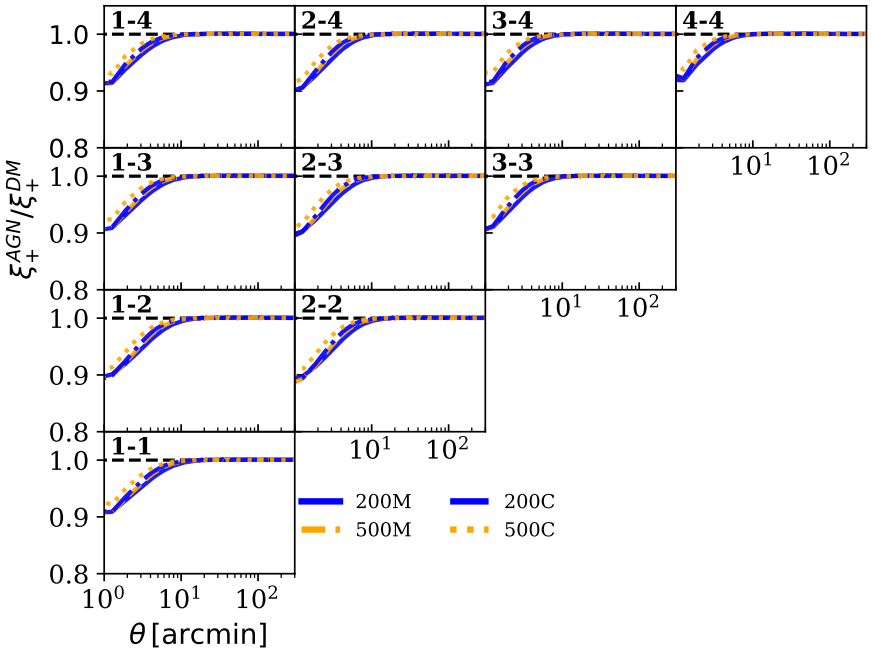

FIGURE 4.14: KiDS-450 tomographic weak lensing ξ_+ baryonic suppression measurements with the four mass definitions presented in different line styles and colours (blue for $\Delta = 200m, 200c$ and orange for $\Delta = 500m, 500c$). The different redshift bins used are highlighted in the top corner of each panel.

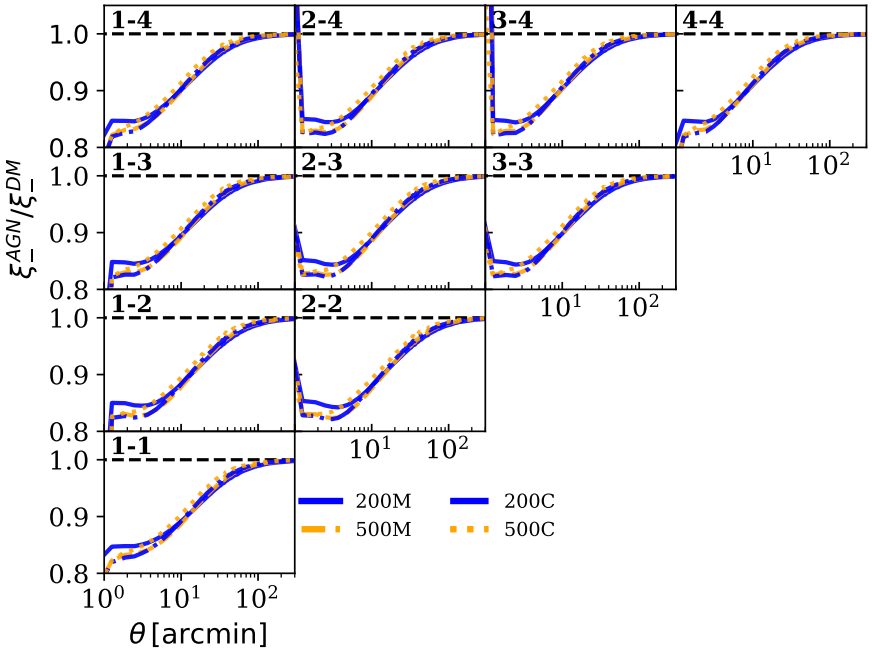

FIGURE 4.15: Same as fig. 4.9 but for ξ_- case.

In figs. (4.14)(4.15) I present the baryonic suppression for the $\xi_{\pm}$ cases using the KiDS tomographic redshift bins for the four mass definitions highlighting the small effect present in the different redshift bins considered. In fact, as it was evident in the panel of fig. 4.13 the cross or the auto-correlations of different source redshift distributions present only a small difference in the overall baryonic suppression. As presented above, the ξ_+ case shows almost a constant suppression (roughly 0.9 at $\theta \approx 1$ arcmin) in each of the redshift bins surveyed. As well, the ξ_- case presents a similar constant suppression in all bins with some differences that rise at the smallest scale ($\theta < 1$ arcmin) with some spikes (in both directions) that are due to the mismatch in the ratio between the collisionless and hydrodynamical cases (similar effect that can be visible in fig. 4.8).

In drawing to a close, I have explored in this section the effects of baryonic feedback on several LSS probes. It is possible to see that the baryonic feedback appears in different shapes and areas in each of the LSS surveyed allowing us to explore, efficiently, the whole spectrum. The most sensitive observations are the ones from tSZ auto and cross-correlations. Where the ability of the model to trace efficiently the pressure distribution can significantly be distinguished and it is possible to say the same for the baryonic effects. For weak lensing probes, I have verified that auto-correlations (like $k_\phi \times k_\phi$) show a smoothed suppression similar to the matter power spectrum suppression, occurring at the same scales as well, whilst in the cross-correlation cases (e.g., $k_\phi \times k$), present a suppression on different scales and generally broader in comparison to the first presented. For the shear measurements, the suppression happens in amplitude and scales, differently between the ξ_+ and ξ_- cases, with small differences ($< 3\%$) between the mass definitions as well as using a unique or a multiple source redshift distributions. More stringent and detailed observations (e.g. Dark Energy Survey or KiDS-1000) can constrain better the different mass definitions in addition to baryonic effects.

4.8 Summary and Conclusions

In this Chapter, I have presented an exploration of several LSS observables using the BAHAMAS calibrated halo model approach focusing on both the effects of different mass definitions and the role of baryons on those predictions.

I have compared the halo model predictions against observational results and observational-like maps built from the BAHAMAS light-cones for a fairer and more complete test of these analyses.

The weak lensing probes have shown some interesting features. The first set of tests provided, the comparison against single source distribution built from BAHAMAS maps,

showed a good agreement on the large-scale limit and some differences in the small scales, highlighting the limits of the calibrated halo model approach and its overall accuracy. Larger differences can be seen when using the lowest source redshift ($z_s = 0.5$) due to a major influence of the accuracy of the starting $P(k)$. It is possible to see a similar picture for all mass definitions involved. The sharp drop that happens at small scales ($\ell > 8000$) can be associated with the limit of the halo model in reproducing accurately the matter distribution at k scales around 8 $h\,\mathrm{Mpc}^{-1}$ (also where the simulation Nyquist frequency starts to influence the results). The results are coherent for both the *DM-Only* and *AGN* realisations.

Using the wide source redshift distribution, Section 4.5, I have studied the accuracy of the halo model against real observations and against the BAHAMAS predictions. This observable shows a similar level of agreement on both cases, ξ_+ and ξ_-, with differences that arise only in the large-scale regime (still below 20% at all scales). Both the halo model predictions and the simulations are above the data points in a significant portion of the spectrum, suggesting that the cosmology adopted could be a non-negligible factor. Very small differences can be found while comparing the collisionless and hydrodynamical cases on similar scales.

The tomographic weak lensing analysis using the KiDS-450 source redshift distributions instead highlighted that the poor halo model accuracy at high redshift reflects on a non-negligible impact in the auto-correlation cases (for instance in the 4-4 case, where it is possible to see the larger differences between the mass definitions). These differences are verifiable in both ξ_- and ξ_+ cases. In general terms, there is an overall good agreement between the observational data and the halo model predictions.

The tSZ power spectrum presented in Section 4.4 shows a good agreement between the halo model predictions, observational data and BAHAMAS results. I have explored a different approach for what regards the halo masses making clear when using a fixed halo mass or the one obtained by integrating the density profiles. Those two separate ways of computing the halo masses, and consequently the shift of the associated pressure profiles and halo mass functions, have a double effect on the power spectrum but, the baryonic feedback is still the most predominant. It is possible to verify those larger mass definitions ($200m$ or $200c$) trace better the power spectrum matching the simulation expectations but overestimating, as seen in the simulations, the Planck observational data. In the scales $400 < \ell < 2000$ we find larger differences, where I measure a minimum of 20% in comparison to the BAHAMAS predictions. This could be driven by the exclusion of outskirts of haloes that are still relevant for the tSZ signal. Smaller mass definitions have much more agreement with observational data (for observational data I always exclude ACT and SPT since those are in tension with BAHAMAS results as

well, see the discussion earlier). The peak of the distribution is fairly well reproduced by all mass definitions ($< 15\%$). The tSZ power spectrum is dominated by the 1-halo term contribution, when we use smaller mass definitions we, effectively, cut out gas particles reducing, artificially, the actual signal. One reason why the largest mass definitions and the BAHAMAS simulations predictions are higher in comparison to Planck results can be traced in the different cosmology adopted. It is still worth mentioning, again, that the tSZ predictions are dependent on many assumptions and adjustments such as the hydrostatic mass bias or the modelling of the pressure distribution inside haloes and outside the halo boundary chosen, those might alter, significantly, the final results.

The cross-correlations have opened up a world of possibilities in combining several observables and obtaining interesting constraints on both cosmological and astrophysical parameters (Shirasaki et al., 2020; Osato et al., 2016; Lu et al., 2021; Tröster et al., 2021; Schneider et al., 2021). In this thesis, I have explored the cross-correlations between galaxy and CMB weak lensing ($k \times k_\phi$) and tSZ with CMB weak lensing ($y \times k_\phi$). The first analysis provided an interesting picture, where the different mass definitions trace well the data however all tend to overestimate the simulation predictions in both collisionless and hydrodynamical cases. For the latter case, I see, instead, that the tSZ accuracy is different and the predictions span evenly the data points available. Again, only the largest mass definitions are comparable with the simulations results (similarly to what happened in the tSZ power spectrum). The last result hints, significantly, that the tSZ accuracy is the main driver on those probes.

The overall impact of baryons is extremely more intricate and complex in the analysis on such a variety of probes. I have shown to what extent the right treatment of baryons is needed for upcoming observational results. With the observational results considered, it is almost impossible to effectively distinguish between the *DM-Only* or *AGN* modelling. However, the knowledge of the overall baryonic suppression is fundamental for upcoming observational results. The weak lensing probes, as CMB or galaxy weak lensing in Fourier space, present a smoothed version of the matter-matter power spectrum suppression, allowing to constrain, efficiently, the influence of galaxy formation feedback present. In the case of angular correlations (ξ_+ and ξ_- cases), the differences are more evident. Thus, those allow distinguishing between accuracy and baryonic effects, even though the portion of the spectrum is subject to other small-scale effects and uncertainties. The tSZ showed a great variety of interesting features that have demonstrated that the baryonic effects can be separated by about 5% from a *DM-Only* treatment. More complex, and more evident, is instead the tSZ impact in the cross-correlations with weak lensing probes (both from galaxy lensing or CMB) with a suppression, on scales $100 < \ell < 2000$, around 9% and 7% independently from the mass definition adopted.

New deeper and detailed observations will reduce, significantly, the uncertainties associated with the measurements. At the same time, improved treatments of the mass definitions in the halo model approach (to a comparable accuracy level as the one needed from $P(k)$) will allow putting tighter constraints on the baryonic feedback parameters. These efforts in improving the modelling and the treatment of baryons are one of the key factors aimed to obtain the most unbiased constraints on the cosmological parameters in the near future.

Chapter 5

Summary and future work

5.1 Summary

In this Thesis, I have walked the reader on a journey trying to understand a bit more the halo model formalism and how, in its simplistic formulation, may fail and be inaccurate. Making use of detailed numerical simulations, I have explored how the halo model fails in reproducing the non-linear matter clustering accurately and how it is, strongly, dependent on the mass definitions, meaning on the size of the haloes considered. This picture here presented is not drawn from any analytic profiles but from the density profiles, spherically averaged, obtained directly from the simulations, both in the collisionless and hydrodynamical setups. Furthermore, the use of the number density of haloes from simulations (namely the halo mass function) shows that the difference is small (few %) compared to other analytical arguments (as the Press & Schechter (1974) or Tinker et al. (2008)).

The main and most important result of this work is that both using the tabulated density profiles or a calibrated fitting profile (using the Einasto fitting function), the halo model fails to reproduce the matter power spectrum accurately, but the relative impact of baryons, evaluated using the ratio between the collisionless and hydrodynamical cases, is recovered to % level accuracy, that is a competitive measure in the era of precision cosmology.

In fact, in the analysis presented the standard halo model shows a lack of power in certain parts of the spectrum, mainly the 1 to 2 halo transition region and the smallest scales, where smaller haloes are not fully incorporated (see the discussion about the role of the Nyquist frequency and the discrepancy at k$\approx 8\,h\,\mathrm{Mpc}^{-1}$ in Section 3.4.1). These issues are present in both setups (*DM-Only* and *AGN*) and they cancel each other out while evaluating the ratio, leaving the pure clustering signal.

In a broader way, this result follows what other authors have found in exploring non-standard cosmologies using the halo model approach. Recently, Cataneo et al. (2019, 2020) and Bose et al. (2020, 2021), with their ReACT[1] software, have shown that the standard halo model formalism is not accurate enough, if not after adding multiple components or exploring further the density fields (Mead et al., 2016; Philcox et al., 2020; García et al., 2021), in light of the requirements for next-generation weak lensing and CMB experiments. In spite of everything, the relative cosmological impact, whether it be the signature of massive neutrinos or the presence of dynamical dark energy, can be effectively and accurately constrained using the ratio respect to a standard cosmology.

The assessment of the power spectrum suppression due to baryonic feedback, theoretically, is the strongest constrain available that can be built from a variety of observations. In fact, the distribution of matter in the Universe can be explored under multiple point of view using different independent tracers.

Upcoming all sky surveys of several observables (tSZ, CMB and cosmic shear) will provide strong constraints on the cosmological parameters showing, again, the need of methods to understand better the Universe we observe.

To make fairer and more detailed comparisons with the upcoming surveys, I have explored how the halo model can reproduce LSS tracers and tested how much the different mass definitions impact the overall accuracy of these predictions. Additionally, I have explored the relative impact of baryons on those probes, understanding in which parts and how much the signature of galaxy formation processes can be detected.

Regarding the latest development, some recent weak lensing analyses (see Yoon & Jee (2021); Huang et al. (2021)) have explored the role of feedback parameters on KiDS-450 and DES year 1 results finding similar and comparable results as the one I have presented, reinforcing the idea that this type of approach and analysis will become more and more common in the years to come.

To wrap up the most relevant analyses and findings of this Thesis, I summarise here the most interesting points I have explored in the previous Chapters. In Chapter 3, I have presented an in-depth analysis of how accurately the halo model can reproduce the matter distribution making comparisons with the BAHAMAS suite of simulations and other linear and non-linear predictions. I have used products directly from the simulations (stacked density profiles and HMFs) to enhance the reliability of the tests. For comparison, I have calibrated an analytic fit, using Einasto profile, and used a standard HMF (Tinker et al., 2008), calibrating the baryonic correction accordingly when discussing the *AGN* cases. The results of these analyses found that the halo model

[1] https://github.com/nebblu/ReACT.

is able to partially recover the overall matter power spectrum, both from the simulations and other results, and the accuracy is strongly dependent on the mass definitions. I find that using both the stacked density or the fitting profiles, the differences are relatively small compared to the overall difference with the expected results. In analogy with the two different HMFs. The differences grow at higher redshift, which means that the halo boundaries are not capable of fully reproducing the matter distributed in the simulations accurately.

However, we have found the same differences in the $P(k)$ modelling, both the *DM-Only* and *AGN* setups. Instead, while considering the baryonic suppression, $S(k) = P_{AGN}(k)/P_{DM}(k)$, those discrepancies cancel out, which leaves the signal almost independent from mass definition and redshift and within 5% compared to simulation results.

In Chapter 4, I have explored beyond the matter clustering, using the halo model predictions in reproducing LSS tracers as galaxy lensing, CMB weak lensing, tSZ and cross-correlations between those probes. The interest in doing such analyses using LSS probes comes from the need to assess the possibility to distinguish between the halo model accuracy and baryonic effects.

I have presented the tSZ predictions that show that the largest mass definitions (namely the $\Delta = 200m$ and $\Delta = 200c$) are able to trace better the pressure distribution inside haloes and reproduce better the simulations results. Smaller haloes definitions, instead, are not able to fully recover the largest scales but can correctly model the smallest scales and the peak of the distribution. The galaxy lensing analysis showed that both in the case of a single wide source redshift distribution, as for CFHTLenS, or different tomographic bins, see KiDS-450 instead, there is a good general agreement with observations and simulations. This is valid for both ξ_+ and ξ_- statistics.

The weak lensing cross-correlations between the CMB and galaxy sources (in this case using a wide redshift bin of KiDS-450) show a generally good agreement between the data and simulations but, in fairness, it is not easy to distinguish accurately between the baryonic effects or the overall accuracy of the halo model. The tSZ-CMB weak lensing cross-correlation analysis shows interestingly a further dependence on the mass definitions, driven most likely by the tSZ probe that separates larger to smaller mass definitions by more than 20% difference.

The baryonic effects on the LSS probes show different and interesting results as well as possible applications in the near future. Starting with the CMB weak lensing auto-correlation which shows a suppression, similar to the matter power spectrum at fixed

redshift but more smoothed over a wider range of modes comparable with other simulations results (Chung et al., 2020). A similar result can be obtained using the $C(\ell)$ of the galaxy lensing cosmic shear. While using the angular correlations, $\xi_{\pm}$, the suppression takes place at different scales and, in a significantly larger way in the ξ_{-} cases and smaller in ξ_{+}. Very small differences (below 3%) can be found using a wide source redshift bin or different redshift bins (i.e., different tomographic bins using KiDS-450 data).

The cross-correlation case between the galaxy and CMB lensing shows a strong suppression ($\approx 10\%$) already in the largest scale regime ($200 < \ell < 2000$), stronger and more diffused compared to the previous ones analysed. We obtained a similar result considering the tSZ auto and cross-correlations, as explained in the text, I have used two separate approaches while dealing with halo masses: one is fixing a halo mass and using that for the profiles and HMFs and the other, instead, is using the halo masses obtained from the integration of the density profiles (accordingly for both the *DM-Only* and *AGN* cases). The suppression in those cases is slightly different, for instance, in the tSZ predictions with the fixed mass show less suppression (around 2.5%) that decreases only on the smaller scales ($\ell > 7000$), while the case with integrated masses shows higher suppression, roughly 6.5% and a smoother fall on the same regime as before. The cross-correlations show, for the smaller range sampled, an almost flat suppression, larger for the CMB cases (around 9% for the fixed mass and 10.5% for the integrated) and slightly lower for the galaxy lensing case (using RCSLenS source redshift distribution, around 5% for the fixed and 7% for the other case).

5.2 Future work

The halo model, in its 'vanilla' formulation, might be overrun in the next years in favour of better emulations techniques and extensive implementation of machine learning tools. Nevertheless, it is possible that it will still be used for quicker broader comparisons.

As I have pointed out, the transitional region between the 1-halo and the 2-halo terms is the most intriguing open question to answer. A non-linear bias treatment seems to provide a better description of that region (Mead & Verde, 2021) but there is also some work on a better characterisation of the halo extensions and halo boundaries (García & Rozo, 2019; García et al., 2021). Effective field theory applications are, as well, viable options that might result in a significant impact on the overall accuracy of the method (Philcox et al., 2020; Sullivan et al., 2021).

One of the most intriguing approaches that could be introduced in this formalism is the dropping of static definitions for the halo boundaries, as $\Delta = 200m$ or $\Delta = 500c$ can be considered, and start to define a more realistic, and dynamically motivated, boundary such as the 'splashback' radius. This new definition will enhance and expand the halo dimensions allowing to identify more efficiently the haloes, in particular at higher redshift, where those are expected to be undergoing more dynamical interactions. A dynamical identification of the size of haloes, with the right association of particles, could, in theory, fill the gap between the two separate terms in the halo model accurately.

Allowing to extend to lower density thresholds, e.g. $\Delta = 100$, could do a similar reduction in the tension between the halo model and the simulation results thanks to the inclusion of external particles, therefore adding clustering signal. But it might open up to more complex adjustments such as a greater chance of overlaps between haloes (mainly larger objects with multiple systems nearby) as well as further deviation from the spherical symmetry due to the inclusion of broader cosmic web structures. More complex treatment of the halo dimensions and shapes (e.g., allowing triaxiality of haloes, Smith & Watts (2005)) could be an interesting challenge for upgrading the halo model to a more flexible tool.

In the analysis of the LSS probes, I have shown that the overall accuracy of the halo model has a non-negligible impact, larger or smaller depending on each tracer surveyed, but the effects of baryons can be effectively distinguished. With the data used in this thesis, it is not possible to fully distinguish between the accuracy of the predictions from simple baryonic effects, therefore, better observations should identify more clearly the most accurate model.

Future all-sky surveys of different probes will open up a world of opportunities in LSS cosmology challenging the 1% accuracy, target for those observations. These predictions will lead to extremely tight constraints on both the growth of structures and galaxy formation parameters.

CMB weak lensing and tSZ observations from Simons Observatories and other stage IV submillimeter observations, in union with weak lensing from Vera Rubin observatory and Euclid, will provide data with unprecedented quality. This new data available will clear the table from theories and hypotheses no longer supported by observations.

To answer the needs of those high-quality data, the need for new hydrodynamical simulations with greater resolution on smaller haloes could push the modelling to smaller and smaller scales ($k \gg 10\,h\,\mathrm{Mpc}^{-1}$) allowing to put further constraints on many cosmological parameters (such as the spectral index or different flavours of dark matter). With high-resolution hydrodynamical simulations, as well as N-Body simulations, we

will have the chance to characterise the halo model to similar smaller scales, adding more physically motivated density profiles of haloes as well as more realistic HMFs in a wider range of masses, allowing us to use this method in both large and small-scale cosmology.

Appendix A

Appendix A

A.1 $\Delta = 2500$ case

In Chapter 3, I have presented an in-depth analysis of how the mass definitions (200 and 500 times the mean or the critical density of the Universe) have an impact on the overall accuracy of the halo model predictions. In this appendix, I present a short exploration of a smaller mass definition, $\Delta = 2500$, both in the mean and critical case.

This exploration aims to strengthen, further, the narrative presented earlier of how selecting particles associated and the size of haloes can track, poorly, the matter distribution in the simulations.

The $\Delta = 2500$ is a very small overdensity that identifies only the innermost and most bounded particles of haloes. As previously done, I present the density profiles, the halo mass function and then I show the $P(k)$ predictions for both the *DM-Only* and *AGN* cases. Finally, I comment on the suppression of the matter power spectrum in these two mass definitions.

In fig. A.1 I present a test that shows the result of integrating the density profiles (using eqn. 3.5) compared against the starting $M_{\Delta,FOF}$ for the *DM-Only* and *AGN* cases and the HMFs. The *DM-Only* cases are shown in blue lines while the *AGN* ones are in orange, the solid lines represent the $\Delta = 2500m$ case while the dashed the other one.

It is possible to see that the test on the integration of the profiles has an overall accuracy of 1% in the mass range considered, with larger differences on the smaller haloes for the *DM-Only* cases. Comparing the HMFs highlights how many and massive haloes are be used to trace the matter power spectrum.

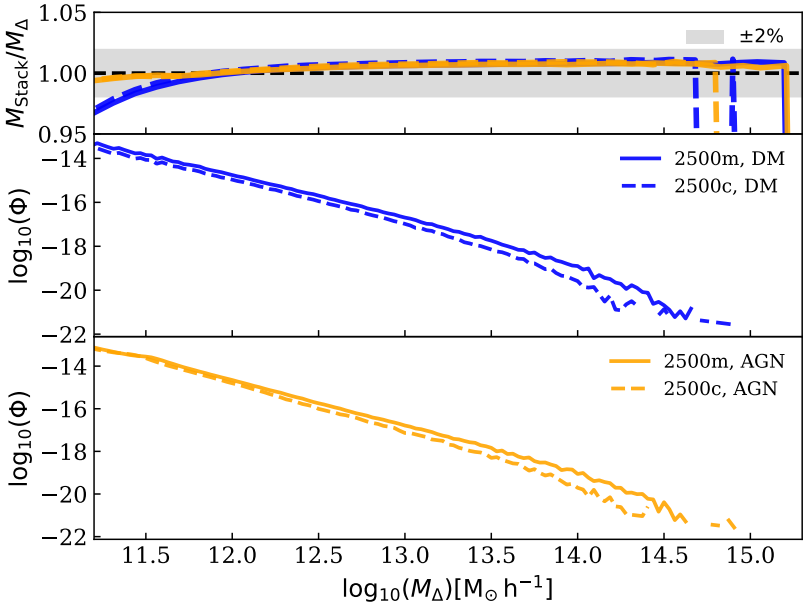

FIGURE A.1: Profile accuracy in the top panel, *DM-Only* HMF in the middle panel and *AGN* HMF in the bottom panel in the $\Delta = 2500$ cases for both the mean and critical cases. In blue I present the *DM-Only* cases while in orange the *AGN* counterparts. The $\Delta = 2500m$ case is shown in solid line and the $\Delta = 2500c$ in dashed lines.

To obtain the matter power spectrum I use the density profiles and the HMFs computed from the simulations, for the linear bias I use the Tinker et al. 2010 formalism for both the *DM-Only* and *AGN* runs as discussed in Chapter 3. Considering the halo fraction, as discussed in Section 3.3.1.1, I verify that at the minimum halo mass, $M_{\rm min} = 4 \times 10^{11}\,{\rm M}_{\odot}\,h^{-1}$, the $f_{\rm h}(M > M_{\rm min})$ for the mean cases is 0.14 and for the critical case is 0.07. That means that the haloes taken into consideration in this analysis are able to reproduce only the 14% and the 7% of the overall total mass in the simulations. The values do not differentiate much between the collisionless and hydrodynamical cases.

In fig. A.2 I present the matter power spectrum cases $\Delta = 2500m$ and $\Delta = 2500c$ using the stacked density profiles and the HMF directly from the simulations in both *DM-Only* and *AGN*. I compare against HALOFIT, CAMB and BAHAMAS (*DM-Only* and *AGN*) matter power spectra at z=0 and I evaluate how accurately the mass definition is able to reproduce the expected signal. I note the residual plot using the notation $R(k)$ defined as $[P(k) - P^{HM}(k)]/P(k)$ and $S(k)$ for the suppression defined as $P_{AGN}(k)/P_{DM}(k)$. The dashed vertical line identifies half the Nyquist frequency of the simulations.

It is possible to see that, as expected, these small mass definitions are able to capture only partially, and not better than 25%, the matter distribution even in the deep 1-halo regime. The lack of massive haloes, as can be seen in the HMFs plots, influences strongly the accuracy as well as the small volume of haloes explored. The picture is coherent

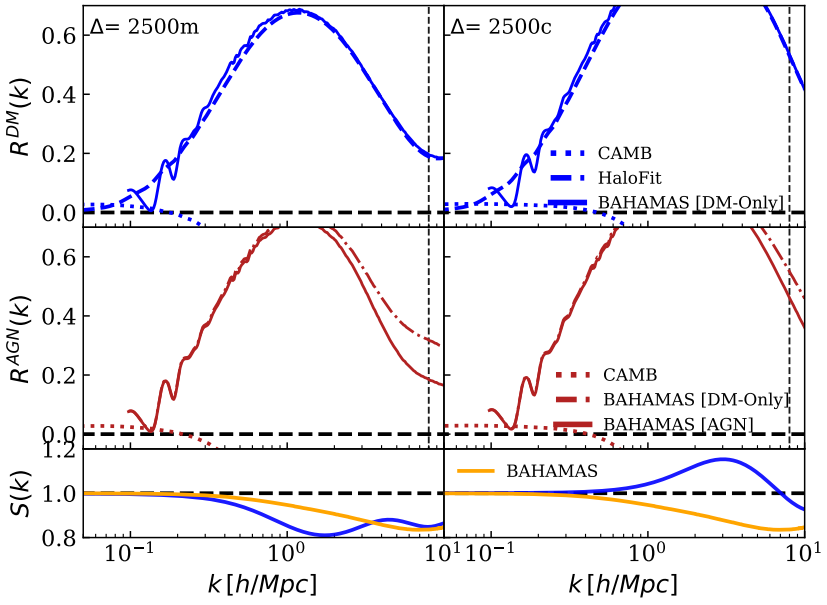

FIGURE A.2: Matter power spectrum and baryonic suppression for the $\Delta = 2500m$ (left column) and $\Delta = 2500c$ (right column) mass definitions in both *DM-Only* (top panels) and *AGN* (middle panels). In the bottom panels there is the matter power spectrum suppression compared against the BAHAMAS simulations results. The top and middle panels are presented, as before, the residuals as follow: $R(k) = [P(k) - P^{HM}(k)]/P(k)$.

from the previous findings, also, for the *AGN* cases. The suppression, shown in the bottom panels of the fig. A.2, shows quite interesting features. The suppression is not well recovered ($\approx 15\% - 20\%$) in the 1-halo region ($k \approx 0.8 - 3\,h\,\mathrm{Mpc}^{-1}$) in both the mean and critical cases, but in the mean case, the suppression actually happens, while in the other case there is an enhancement ($\approx 15\%$ at $k = 2.5\,h\,\mathrm{Mpc}^{-1}$). The small volume of haloes sampled with these mass definitions[1] does not allow to accurately explore the baryonic effects that impact these objects. A possible explanation is that the critical case (which is smaller compared to the mean ones) is capturing only the innermost part of haloes and effects such as gravitational softening overcome the baryonic feedback present. Another possible explanation is the small population of the largest mass bins in the critical case, which might influence the overall accuracy in the predictions and suppression.

A.2 Matching technique using Particles IDs

In this short appendix, I summarise the method behind the matching procedure between two sets of simulations by using particles IDs. The method was firstly presented in

[1]For reference the $R_{2500c} \approx 1/3R_{200c}$ and $R_{2500m} \approx 1/2R_{200m}$ for the cosmology adopted.

Velliscig et al. (2014) and have been used in many different works (Mummery et al., 2017; Pfeifer et al., 2020; Stafford et al., 2020a).

Each particle in the simulation has associated an unique *IDs*, particles belonging to the same haloes are associated a *groupnumber* by the FOF algorithm used to identify the haloes. This can, also, be applied to *group-subhaloes* groupnumbers.

Given this overview, I now explain, briefly, how the matching works:

- Select the *DM* particles data from a *DM-Only* and *AGN* simulations and assign an unique haloID;
- Match the *DM* particles as $ID_{AGN} \rightarrow ID_{DM\text{-}Only}$, with a minimum of N most-bound particles to identify the same halo between the two simulations;
- Check the goodness of the match by doing the inverse match $ID_{DM\text{-}Only} \rightarrow ID_{AGN}$,
- Obtain a catalogue of matched haloes between the two simulations.

With BAHAMAS particles mass I have imposed a cut at N=50 particles. This limits us in going below $\approx 3 \times 10^{11}\,\mathrm{M}_{\odot}\,h^{-1}$ given the mass resolution available in the BAHAMAS simulations. With the matched catalogue of haloes we can compare the different effects on the masses due to the baryon feedback. This matching, since it is based on the particles IDs, is independent on the mass definition chosen.

A.3 Direct baryonic HMF correction

In Section 3.3.1, I have presented how the baryons influence the masses of the haloes in hydrodynamical simulations and how that is relevant for the halo mass function baryonic characterisation. In this short appendix, I present an equivalent, but less accurate, method to correct directly the HMF for baryonic effects. The overall accuracy is lower compared to changing the masses beforehand but it might be useful in some applications.

In fig. A.3 I show the logarithmic suppression of the HMFs (Φ), as $\log_{10}(\Phi_{AGN}/\Phi_{DM})$, by comparing the simulations outputs (solid lines) and the best fit results (dashed lines) at the three redshifts bins. I have built the data to fit on, using the HMFs for the four mass definitions and I have computed the mean of those predictions and computed the scatter (shaded area). The scatter of the data is mainly due to the intrinsic differences between the HMFs and the ratio between the *AGN* and *DM-Only* cases. I have applied a cut at $2 \times 10^{14}\,\mathrm{M}_{\odot}\,h^{-1}$ for the $z = 1$ case and at $8 \times 10^{13}\,\mathrm{M}_{\odot}\,h^{-1}$ for $z = 2$ due to

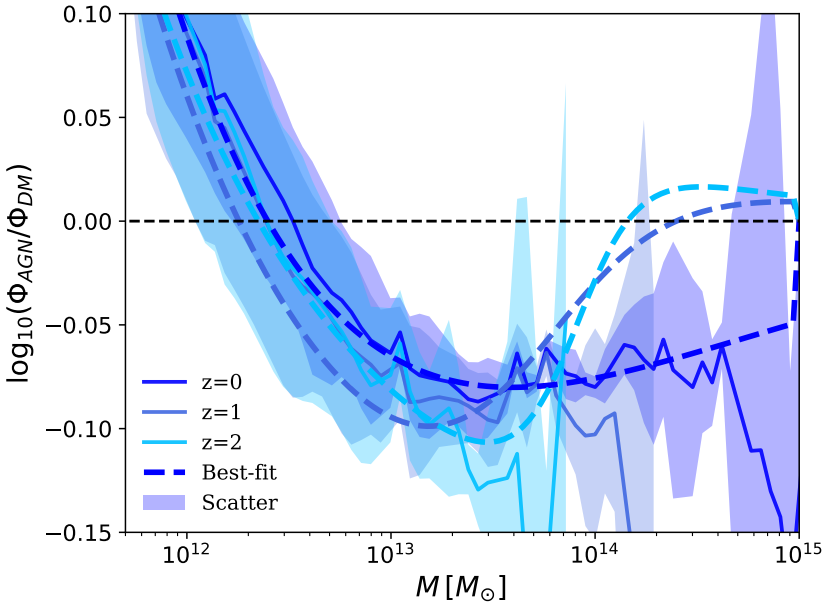

FIGURE A.3: Comparison plot between the best fit results and the data from the simulations for the HMF suppression. I present the three redshift bins $z = [0, 1, 2]$ using solid lines and different shades of blue. The 1σ scatter is obtained by computing the scatter from the mean of the four different mass definitions. In dashed lines I show the best-fitting analytic predictions. It is possible to see that there is a good agreement in the deep in the range $10^{13} < M\,[\mathrm{M}_\odot\,h^{-1}] < 3 \times 10^{14}$. At larger masses the scatter is larger due to cosmic variance.

TABLE A.1: Best-fitting parameters of the baryon correction for the HMF. In fig. A.3 I present the comparison between the BAHAMAS HMFs and the analytic results using this set of parameters.

P	A	B	C	D
a_0	27140.1	−0.312	−13.474	−1.148
a_z	−0.288	−0.546	0.029	−1.843

fact that there are very few haloes above those mass thresholds. Applying these cuts I have imposed that the two HMFs should have the same values above. We see that our fitting function reproduce well the data at $z = 0$ over a wide mass range and, for the other redshifts, it is possible to verify that the most suppressed part is well recovered alongside the rise at small halo masses.

The fitting procedure was done using the same IDL tools used previously for the mass correction. I provide here the best-fitting values for the eight parameters in Table A.1 for the eqn. 3.7 specified for $\log_{10}(\Phi_{AGN}/\Phi_{DM})$:

$$\log_{10}\left(\frac{\Phi_{AGN}}{\Phi_{DM}}\right) = \frac{A}{\cosh(\log_{10}(M_{DM}))} + \frac{B}{1+\exp\left(-\frac{\log_{10}(M_{DM})-C}{D}\right)} \quad . \tag{A.1}$$

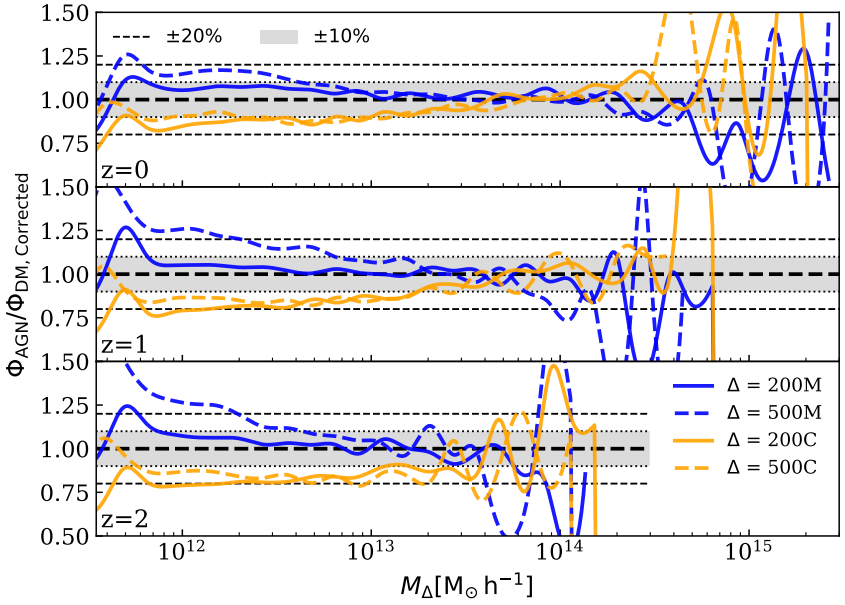

FIGURE A.4: Ratio plot between the BAHAMAS HMFs for the *AGN* case over the corrected *DM-Only* using the baryonic HMF correction. I present the three redshifts bins and all four mass definitions (orange for the critical mass definitions and blue for the mean ones). It is possible to see that the baryonic correction helps recover the HMFs better than 15% over a wide mass range and all redshifts. A better agreement is achieved on the regime above $5 \times 10^{13}\,\mathrm{M}_\odot\,h^{-1}$. Below $3 \times 10^{11}\,\mathrm{M}_\odot\,h^{-1}$ I have not applied the correction because of the small reliability of our predictions.

I find that two parameters (B and D) have a slight cosmology dependence built in terms of baryon fraction, similarly to eqn. 3.7, f_b as follows : $F(f_b) = \left(1 + \frac{f_b}{f_b^{WMAP9}}\right)^{-0.95}$, where $f_b^{WMAP9} = 0.165$, as was already shown in the baryonic mass correction.

In fig. A.4 I present the result of the application of the baryonic correction (eqn. 3.7) on the *DM-Only* HMF from BAHAMAS and I compare against the *AGN* version at the three redshifts considered. It is possible to see that for the four different mass definition, the correction helps recover the expected *AGN* HMF better than $\approx 15\%$ in most mass bins and for all mass definitions at $z = 0$.

The ratio plot is noisier at larger masses due to cosmic variance, as I have explained regarding fig. A.3.

In the first application presented, the agreement was better than 5% at all redshifts and over a wide range of masses. With this application, instead, we see at least a 15% discrepancy between the corrected halo mass function and the expected one. This poorer result is driven by the loss of details that comes from computing a general function as the HMFs where most of the effects on haloes are smoothed out by the statistics and it is harder to encapsulate the baryonic feedback accurately.

Appendix B

Appendix B

B.1 Einasto profile partial derivatives

In Section 3.2, I have presented the fitting functions of the *DM-Only* and *AGN* density profiles using the Einasto profile. To improve the convergence of the fit I have calculated the partial derivatives for each of the nine free parameters used to maximise the reliability of the fit and quicken the parameter space exploration. I remind that the Einasto profile has this form:

$$F(r, M, z) = f_0(M, z) \exp\left(-A(M, z) r^{\alpha(M,z)}\right) \quad , \tag{B.1}$$

with the three main parameters f, A and α have a double dependence on mass and redshift as shown in eqn. 3.4. I remind, also, that the fit we have performed in these analyses were meant to provide a description of how the matter is distributed inside the haloes and we have not put any physical constraint on the parameters, allowing for degeneracies between them. This applies as well for the pressure profiles, described in Section 4.2.2.

The partial derivatives for each free parameters are presented here:

$$\frac{\partial F}{\partial f_{0,\mathrm{int}}} = F/f_{0,\mathrm{int}};$$
$$\frac{\partial F}{\partial f_M} = F \times \ln\left(\frac{M}{M_{\mathrm{pivot}}}\right);$$
$$\frac{\partial F}{\partial f_z} = F \times \ln(1+z);$$
$$\frac{\partial F}{\partial A_{0,\mathrm{int}}} = F \times \left(\frac{A(M,z)}{A_{0,\mathrm{int}}}\right) \times r^{\alpha};$$
$$\frac{\partial F}{\partial A_M} = -F \times A r^{\alpha} \times \ln\left(\frac{M}{M_{\mathrm{pivot}}}\right);$$
$$\frac{\partial F}{\partial A_z} = -F \times A r^{\alpha} \times \ln\left(1+z\right);$$
$$\frac{\partial F}{\partial \alpha_{0,\mathrm{int}}} = -F \times A r^{\alpha} \ln(r) \left(\frac{\alpha}{\alpha_{0,\mathrm{int}}}\right);$$
$$\frac{\partial F}{\partial \alpha_M} = -F \times A \alpha r^{\alpha} \ln(r) \ln\left(\frac{M}{M_{\mathrm{pivot}}}\right);$$
$$\frac{\partial F}{\partial \alpha_z} = -F \times A \alpha r^{\alpha} \ln(r) \ln(1+z).$$

B.2 Updated Velliscig's fitting function partial derivatives

As I have done for the Einasto profile used for haloes profiles, I provide here the partial derivatives for the baryonic correction functions I have calibrated for masses and HMFs presented in Chapter 3.2. I remind the original fitting function expressed in terms of Δ_{mass} defined as:

$$\Delta_{\mathrm{mass}} = \frac{M_{AGN} - M_{DM-Only}}{M_{DM-Only}} \ ,$$
$$\Delta_{\mathrm{mass}}(M,z) = \frac{A(M,z)}{\cosh(\log_{10}(M))} + \frac{B(M,z)}{1+\exp\left(-\frac{\log_{10}(M)-C(M,z)}{D(M,z)}\right)} \ . \tag{B.2}$$

So the partial derivatives are:

$$\frac{\partial \Delta_{\text{mass}}}{\partial A_M} = \frac{(1+z)^{A_z}}{\cosh(\log_{10}(M))};$$
$$\frac{\partial \Delta_{\text{mass}}}{\partial A_z} = \frac{A\ln(1+z)}{\cosh(\log_{10}(M))};$$
$$\frac{\partial \Delta_{\text{mass}}}{\partial B_M} = \frac{(1+z)^{B_z}}{1+\exp\left(-\frac{\log_{10}(M)-C}{D}\right)};$$
$$\frac{\partial \Delta_{\text{mass}}}{\partial B_z} = \frac{B\ln(1+z)}{1+\exp\left(-\frac{\log_{10}(M)-C}{D}\right)};$$
$$\frac{\partial \Delta_{\text{mass}}}{\partial C_M} = \frac{B(1+z)^{C_z}\exp\left(-\frac{\log_{10}(M)+C}{D}\right)}{D\left[1+\exp\left(-\frac{\log_{10}(M)+C}{D}\right)^2\right]};$$
$$\frac{\partial \Delta_{\text{mass}}}{\partial C_z} = \frac{BC\ln(1+z)\exp\left(-\frac{\log_{10}(M)+C}{D}\right)}{D\left[1+\exp\left(-\frac{\log_{10}(M)+C}{D}\right)^2\right]};$$
$$\frac{\partial \Delta_{\text{mass}}}{\partial D_M} = \frac{-B(1+z)^{-D_z}(\log_{10}(M)+C)\exp\left((1+z)^{-D_z}-\frac{\log_{10}(M)+C}{D_M}\right)}{D_M^2\left[\exp\left((1+z)^{-D_z}-\frac{\log_{10}(M)+C}{D_M}\right)+1\right]^2};$$
$$\frac{\partial \Delta_{\text{mass}}}{\partial D_z} = \frac{-B(\log_{10}(M)+C)(1+z)^{-D_z}\ln(1+z)\exp\left(\frac{-(\log_{10}(M)+C)(1+z)^{-D_z}}{D_M}\right)}{D_M\left[\exp\left(\frac{-(\log_{10}(M)+C)(1+z)^{-D_z}}{D_M}\right)+1\right]^2}.$$

Bibliography

Abazajian K., et al., 2019, arXiv e-prints, p. arXiv:1907.04473

Abbott T. M. C., et al., 2018, Monthly Notices of the Royal Astronomical Society, 480, 3879

Ade P., et al., 2019, J. Cosmology Astropart. Phys., 2019, 056

Ali-Haïmoud Y., Bird S., 2013, Monthly Notices of the Royal Astronomical Society, 428, 3375

Ali-Haïmoud Y., Bird S., 2012, Monthly Notices of the Royal Astronomical Society, 428, 3375

Angulo R. E., Springel V., White S. D. M., Jenkins A., Baugh C. M., Frenk C. S., 2012, Monthly Notices of the Royal Astronomical Society, 426, 2046

Angulo R. E., Zennaro M., Contreras S., Aricò G., Pellejero-Ibañez M., Stücker J., 2021, Monthly Notices of the Royal Astronomical Society, 507, 5869

Aricò G., Angulo R. E., Contreras S., Ondaro-Mallea L., Pellejero-Ibañez M., Zennaro M., 2021, Monthly Notices of the Royal Astronomical Society, 506, 4070

Arnaud M., Pratt G. W., Piffaretti R., Böhringer H., Croston J. H., Pointecouteau E., 2010, Astronomy & Astrophysics, 517, A92

Bahé Y. M., McCarthy I. G., King L. J., 2012, Monthly Notices of the Royal Astronomical Society, 421, 1073

Baldauf T., Seljak U., Desjacques V., McDonald P., 2012, Phys. Rev. D, 86, 083540

Baldi A. S., Bourdin H., Mazzotta P., Eckert D., Ettori S., Gaspari M., Roncarelli M., 2019, Astronomy & Astrophysics, 630, A121

Barkana R., Loeb A., 2001, Phys. Rep., 349, 125

Barnes D. J., Vogelsberger M., Pearce F. A., Pop A.-R., Kannan R., Cao K., Kay S. T., Hernquist L., 2020, arXiv e-prints, p. arXiv:2001.11508

Bartelmann M., Schneider P., 1999, Astronomy & Astrophysics, 345, 17

Bartelmann M., Schneider P., 2001, Phys. Rep., 340, 291

Bartolo N., Komatsu E., Matarrese S., Riotto A., 2004, Phys. Rep., 402, 103

Battaglia N., Bond J. R., Pfrommer C., Sievers J. L., Sijacki D., 2010, The Astrophysical Journal, 725, 91

Battaglia N., Bond J. R., Pfrommer C., Sievers J. L., 2012a, The Astrophysical Journal, 758, 74

Battaglia N., Bond J. R., Pfrommer C., Sievers J. L., 2012b, The Astrophysical Journal, 758, 75

Battaglia N., Hill J. C., Murray N., 2015, The Astrophysical Journal, 812, 154

Battaglia N., Ferraro S., Schaan E., Spergel D. N., 2017, J. Cosmology Astropart. Phys., 2017, 040

Benjamin J., et al., 2013, Monthly Notices of the Royal Astronomical Society, 431, 1547

Birkinshaw M., 1999, Phys. Rep., 310, 97

Blanchard A., Schneider J., 1987, Astronomy & Astrophysics, 184, 1

Blas D., Lesgourgues J., Tram T., 2011, J. Cosmology Astropart. Phys., 2011, 034

Bocquet S., Saro A., Dolag K., Mohr J. J., 2016, Monthly Notices of the Royal Astronomical Society, 456, 2361

Bocquet S., et al., 2019, The Astrophysical Journal, 878, 55

Bocquet S., Heitmann K., Habib S., Lawrence E., Uram T., Frontiere N., Pope A., Finkel H., 2020, The Astrophysical Journal, 901, 5

Bolliet B., Comis B., Komatsu E., Macías-Pérez J. F., 2018, Monthly Notices of the Royal Astronomical Society, 477, 4957

Bolliet B., Brinckmann T., Chluba J., Lesgourgues J., 2020, Monthly Notices of the Royal Astronomical Society, 497, 1332

Bond J. R., Kofman L., Pogosyan D., 1996, Nature, 380, 603

Bonvin V., et al., 2017, Monthly Notices of the Royal Astronomical Society, 465, 4914

Booth C. M., Schaye J., 2009, Monthly Notices of the Royal Astronomical Society, 398, 53

Bose B., Cataneo M., Tröster T., Xia Q., Heymans C., Lombriser L., 2020, Monthly Notices of the Royal Astronomical Society, 498, 4650

Bose B., et al., 2021, Monthly Notices of the Royal Astronomical Society, 508, 2479

Brainerd T. G., Blandford R. D., Smail I., 1996, The Astrophysical Journal, 466, 623

Brown M., et al., 2015, in Advancing Astrophysics with the Square Kilometre Array (AASKA14). p. 23 (arXiv:1501.03828)

Brown S. T., McCarthy I. G., Diemer B., Font A. S., Stafford S. G., Pfeifer S., 2020, Monthly Notices of the Royal Astronomical Society, 495, 4994

Carlstrom J. E., Holder G. P., Reese E. D., 2002, Annual Review of Astronomy and Astrophysics, 40, 643

Carlstrom J., et al., 2019, in Bulletin of the American Astronomical Society. p. 209 (arXiv:1908.01062)

Castro T., Borgani S., Dolag K., Marra V., Quartin M., Saro A., Sefusatti E., 2021, Monthly Notices of the Royal Astronomical Society, 500, 2316

Cataneo M., Lombriser L., Heymans C., Mead A. J., Barreira A., Bose S., Li B., 2019, Monthly Notices of the Royal Astronomical Society, 488, 2121

Cataneo M., Emberson J. D., Inman D., Harnois-Déraps J., Heymans C., 2020, Monthly Notices of the Royal Astronomical Society, 491, 3101

Chen A. Y., Afshordi N., 2020, Phys. Rev. D, 101, 103522

Chiang Y.-K., Makiya R., Ménard B., Komatsu E., 2020, The Astrophysical Journal, 902, 56

Chiang Y.-K., Makiya R., Komatsu E., Ménard B., 2021, The Astrophysical Journal, 910, 32

Chisari N. E., et al., 2019, The Open Journal of Astrophysics, 2

Choi A., et al., 2016, Monthly Notices of the Royal Astronomical Society, 463, 3737

Chung E., Foreman S., van Engelen A., 2020, Phys. Rev. D, 101, 063534

Clowe D., De Lucia G., King L., 2004, Monthly Notices of the Royal Astronomical Society, 350, 1038

Cole S., Kaiser N., 1988, Monthly Notices of the Royal Astronomical Society, 233, 637

Cooray A., 2001, Phys. Rev. D, 64, 063514

Cooray A., Chen X., 2002, The Astrophysical Journal, 573, 43

Cooray A., Hu W., 2002, The Astrophysical Journal, 574, 19

Cooray A., Sheth R., 2002, Phys. Rep., 372, 1

Coulton W., Ota A., van Engelen A., 2020, Phys. Rev. Lett., 125, 111301

Crain R. A., et al., 2015, Monthly Notices of the Royal Astronomical Society, 450, 1937

Cui W., Borgani S., Murante G., 2014, Monthly Notices of the Royal Astronomical Society, 441, 1769

Cusworth S. J., Kay S. T., Battye R. A., Thomas P. A., 2014, Monthly Notices of the Royal Astronomical Society, 439, 2485

Dalla Vecchia C., Schaye J., 2008, Monthly Notices of the Royal Astronomical Society, 387, 1431

Darwish O., et al., 2021, Monthly Notices of the Royal Astronomical Society, 500, 2250

Das S., et al., 2011, Phys. Rev. Lett., 107, 021301

Davies C. T., Cautun M., Li B., 2018, Monthly Notices of the Royal Astronomical Society, 480, L101

Davies C. T., Cautun M., Giblin B., Li B., Harnois-Déraps J., Cai Y.-C., 2021, Monthly Notices of the Royal Astronomical Society, 507, 2267

DeRose J., et al., 2019, The Astrophysical Journal, 875, 69

Debackere S. N. B., Schaye J., Hoekstra H., 2020, Monthly Notices of the Royal Astronomical Society, 492, 2285

Debackere S. N. B., Schaye J., Hoekstra H., 2021, Monthly Notices of the Royal Astronomical Society, 505, 593

Del Popolo A., Pace F., Le Delliou M., 2017, J. Cosmology Astropart. Phys., 2017, 032

Desjacques V., Jeong D., Schmidt F., 2018, Phys. Rep., 733, 1

Despali G., Giocoli C., Angulo R. E., Tormen G., Sheth R. K., Baso G., Moscardini L., 2016, Monthly Notices of the Royal Astronomical Society, 456, 2486

Deutsch A.-S., Dimastrogiovanni E., Johnson M. C., Münchmeyer M., Terrana A., 2018, Phys. Rev. D, 98, 123501

Di Matteo T., Springel V., Hernquist L., 2005, Nature, 433, 604

Di Valentino E., Melchiorri A., Silk J., 2020, Nature Astronomy, 4, 196

Di Valentino E., et al., 2021, Classical and Quantum Gravity, 38, 153001

Diemer B., 2018, ApJS, 239, 35

Diemer B., 2020, ApJS, 251, 17

Diemer B., Kravtsov A. V., 2015, The Astrophysical Journal, 799, 108

Dolag K., Borgani S., Murante G., Springel V., 2009, Monthly Notices of the Royal Astronomical Society, 399, 497

Dolag K., Komatsu E., Sunyaev R., 2016, Monthly Notices of the Royal Astronomical Society, 463, 1797

Donald-McCann J., Beutler F., Koyama K., Karamanis M., 2021, arXiv e-prints, p. arXiv:2109.15236

Doran M., 2005, J. Cosmology Astropart. Phys., 2005, 011

Drinkwater M. J., et al., 2010, Monthly Notices of the Royal Astronomical Society, 401, 1429

Duffy A. R., Schaye J., Kay S. T., Dalla Vecchia C., 2008, Monthly Notices of the Royal Astronomical Society, 390, L64

Duffy A. R., Schaye J., Kay S. T., Dalla Vecchia C., Battye R. A., Booth C. M., 2010, Monthly Notices of the Royal Astronomical Society, 405, 2161

Dutton A. A., Macciò A. V., 2014, Monthly Notices of the Royal Astronomical Society, 441, 3359

Efstathiou G., Gratton S., 2020, Monthly Notices of the Royal Astronomical Society, 496, L91

Einasto J., 1965, Trudy Astrofizicheskogo Instituta Alma-Ata, 5, 87

Eisenstein D. J., et al., 2011, AJ, 142, 72

Erben T., et al., 2013, Monthly Notices of the Royal Astronomical Society, 433, 2545

Ettori S., et al., 2019, Astronomy & Astrophysics, 621, A39

Fabbian G., Calabrese M., Carbone C., 2018, J. Cosmology Astropart. Phys., 2018, 050

Fang X., Krause E., Eifler T., MacCrann N., 2020, J. Cosmology Astropart. Phys., 2020, 010

Fedeli C., 2014, J. Cosmology Astropart. Phys., 4, 028

Fedeli C., Semboloni E., Velliscig M., Van Daalen M., Schaye J., Hoekstra H., 2014, J. Cosmology Astropart. Phys., 8, 028

Feng C., Keating B., Paar H. P., Zahn O., 2012, Phys. Rev. D, 85, 043513

Fixsen D. J., Cheng E. S., Gales J. M., Mather J. C., Shafer R. A., Wright E. L., 1996, The Astrophysical Journal, 473, 576

García R., Rozo E., 2019, Monthly Notices of the Royal Astronomical Society, 489, 4170

García R., Rozo E., Becker M. R., More S., 2021, Monthly Notices of the Royal Astronomical Society, 505, 1195

Gatti M., et al., 2021, arXiv e-prints, p. arXiv:2108.01600

George E. M., et al., 2015, The Astrophysical Journal, 799, 177

Ghirardini V., Ettori S., Amodeo S., Capasso R., Sereno M., 2017, Astronomy & Astrophysics, 604, A100

Ghirardini V., Ettori S., Eckert D., Molendi S., Gastaldello F., Pointecouteau E., Hurier G., Bourdin H., 2018, Astronomy & Astrophysics, 614, A7

Gianfagna G., et al., 2021, Monthly Notices of the Royal Astronomical Society, 502, 5115

Giocoli C., Bartelmann M., Sheth R. K., Cacciato M., 2010, Monthly Notices of the Royal Astronomical Society, 408, 300

Giocoli C., et al., 2016, Monthly Notices of the Royal Astronomical Society, 461, 209

Giocoli C., et al., 2017, Monthly Notices of the Royal Astronomical Society, 470, 3574

Giocoli C., Monaco P., Moscardini L., Castro T., Meneghetti M., Metcalf R. B., Baldi M., 2020, Monthly Notices of the Royal Astronomical Society, 496, 1307

Giri U., Smith K. M., 2020, arXiv e-prints, p. arXiv:2010.07193

Green D., Kaplan D. E., Rajendran S., 2021, arXiv e-prints, p. arXiv:2108.06928

Griffiths R. E., Casertano S., Im M., Ratnatunga K. U., 1996, Monthly Notices of the Royal Astronomical Society, 282, 1159

Hand N., et al., 2015, Phys. Rev. D, 91, 062001

Hand N., Feng Y., Beutler F., Li Y., Modi C., Seljak U., Slepian Z., 2018, AJ, 156, 160

Harnois-Déraps J., Vafaei S., Van Waerbeke L., 2012, Monthly Notices of the Royal Astronomical Society, 426, 1262

Harnois-Déraps J., van Waerbeke L., Viola M., Heymans C., 2015, Monthly Notices of the Royal Astronomical Society, 450, 1212

Harnois-Déraps J., et al., 2016, Monthly Notices of the Royal Astronomical Society, 460, 434

Harnois-Déraps J., et al., 2017, Monthly Notices of the Royal Astronomical Society, 471, 1619

Harrison E. R., 1970, Phys. Rev. D, 1, 2726

He Y., Mansfield P., Rau M. M., Trac H., Battaglia N., 2021, The Astrophysical Journal, 908, 91

Heitmann K., et al., 2016, The Astrophysical Journal, 820, 108

Heymans C., et al., 2012, Monthly Notices of the Royal Astronomical Society, 427, 146

Heymans C., et al., 2013, Monthly Notices of the Royal Astronomical Society, 432, 2433

Heymans C., et al., 2021, Astronomy & Astrophysics, 646, A140

Hildebrandt H., et al., 2012, Monthly Notices of the Royal Astronomical Society, 421, 2355

Hildebrandt H., et al., 2016, Monthly Notices of the Royal Astronomical Society, 463, 635

Hildebrandt H., et al., 2017, Monthly Notices of the Royal Astronomical Society, 465, 1454

Hill J. C., Pajer E., 2013, Phys. Rev. D, 88, 063526

Hill J. C., Spergel D. N., 2014, J. Cosmology Astropart. Phys., 2, 030

Hill A. D., Crain R. A., Kwan J., McCarthy I. G., 2021, Monthly Notices of the Royal Astronomical Society, 505, 65

Hinshaw G., et al., 2013, The Astrophysical Journal Supplement Series, 208, 19

Hirata C. M., Padmanabhan N., Seljak U., Schlegel D., Brinkmann J., 2004, Phys. Rev. D, 70, 103501

Hoekstra H., Jain B., 2008, Annual Review of Nuclear and Particle Science, 58, 99

Hoekstra H., Bartelmann M., Dahle H., Israel H., Limousin M., Meneghetti M., 2013, Space Sci. Rev., 177, 75

Hogg D. W., 1999, preprint, pp astro–ph/9905116 (arXiv:astro-ph/9905116)

Hojjati A., Linder E. V., 2016, Phys. Rev. D, 93, 023528

Hojjati A., et al., 2017, Monthly Notices of the Royal Astronomical Society, 471, 1565

Holmberg E., 1941, The Astrophysical Journal, 94, 385

Hopkins P. F., et al., 2018, Monthly Notices of the Royal Astronomical Society, 480, 800

Horowitz B., Seljak U., 2017, Monthly Notices of the Royal Astronomical Society, 469, 394

Hu W., 2000, Phys. Rev. D, 62, 043007

Hu W., 2005, in Wolff S. C., Lauer T. R., eds, Astronomical Society of the Pacific Conference Series Vol. 339, Observing Dark Energy. p. 215 (arXiv:astro-ph/0407158)

Huang H.-J., et al., 2021, Monthly Notices of the Royal Astronomical Society, 502, 6010

Hurier G., Angulo R. E., 2018, Astronomy & Astrophysics, 610, L4

Hurier G., Aghanim N., Douspis M., 2017, arXiv e-prints, p. arXiv:1702.00075

Ivanov M. M., Simonović M., Zaldarriaga M., 2020, J. Cosmology Astropart. Phys., 2020, 042

Jenkins A., Frenk C. S., White S. D. M., Colberg J. M., Cole S., Evrard A. E., Couchman H. M. P., Yoshida N., 2001, Monthly Notices of the Royal Astronomical Society, 321, 372

Jing Y. P., Zhang P., Lin W. P., Gao L., Springel V., 2006, ApJ, 640, L119

Joudaki S., et al., 2017, Monthly Notices of the Royal Astronomical Society, 471, 1259

Kaiser N., 1998, The Astrophysical Journal, 498, 26

Katz N., Weinberg D. H., Hernquist L., 1996, ApJS, 105, 19

Kauffmann G., Colberg J. M., Diaferio A., White S. D. M., 1999, Monthly Notices of the Royal Astronomical Society, 303, 188

Keisler R., et al., 2011, The Astrophysical Journal, 743, 28

Khatri R., Gaspari M., 2016, Monthly Notices of the Royal Astronomical Society, 463, 655

Kilbinger M., et al., 2013, Monthly Notices of the Royal Astronomical Society, 430, 2200

Köhlinger F., et al., 2017, Monthly Notices of the Royal Astronomical Society, 471, 4412

Komatsu E., Kitayama T., 1999, ApJ, 526, L1

Komatsu E., Seljak U., 2002, Monthly Notices of the Royal Astronomical Society, 336, 1256

Komatsu E., et al., 2011, The Astrophysical Journal Supplement Series, 192, 18

Koopmans L., et al., 2015, in Advancing Astrophysics with the Square Kilometre Array (AASKA14). p. 1 (arXiv:1505.07568)

Koukoufilippas N., Alonso D., Bilicki M., Peacock J. A., 2020, Monthly Notices of the Royal Astronomical Society, 491, 5464

Kwan J., Heitmann K., Habib S., Padmanabhan N., Lawrence E., Finkel H., Frontiere N., Pope A., 2015, The Astrophysical Journal, 810, 35

Le Brun A. M. C., McCarthy I. G., Schaye J., Ponman T. J., 2014, Monthly Notices of the Royal Astronomical Society, 441, 1270

Le Brun A. M. C., McCarthy I. G., Melin J.-B., 2015, Monthly Notices of the Royal Astronomical Society, 451, 3868

Leauthaud A., et al., 2017, Monthly Notices of the Royal Astronomical Society, 467, 3024

Lee E., Chluba J., Kay S. T., Barnes D. J., 2020, Monthly Notices of the Royal Astronomical Society, 493, 3274

Lemos P., Challinor A., Efstathiou G., 2017, J. Cosmology Astropart. Phys., 5, 014

Lewis A., Challinor A., 2006, Phys. Rep., 429, 1

Lewis A., Challinor A., Lasenby A., 2000, The Astrophysical Journal, 538, 473

Libeskind N. I., et al., 2018, Monthly Notices of the Royal Astronomical Society, 473, 1195

Limber D. N., 1954, The Astrophysical Journal, 119, 655

Lin Z., Huang N., Avestruz C., Wu W. L. K., Trivedi S., Caldeira J., Nord B., 2021, Monthly Notices of the Royal Astronomical Society, 507, 4149

Linde A. D., 1982, Physics Letters B, 108, 389

Linder E. V., 2003, Phys. Rev. Lett., 90, 091301

Linder E. V., 2005, Phys. Rev. D, 72, 043529

Liu J., Hill J. C., 2015, Phys. Rev. D, 92, 063517

Loverde M., Afshordi N., 2008, Phys. Rev. D, 78, 123506

Lu T., Haiman Z., Zorrilla Matilla J. M., 2021, arXiv e-prints, p. arXiv:2109.11060

Ludlow A. D., Navarro J. F., Angulo R. E., Boylan-Kolchin M., Springel V., Frenk C., White S. D. M., 2014, Monthly Notices of the Royal Astronomical Society, 441, 378

Ludlow A. D., Bose S., Angulo R. E., Wang L., Hellwing W. A., Navarro J. F., Cole S., Frenk C. S., 2016, Monthly Notices of the Royal Astronomical Society, 460, 1214

Ma C.-P., Fry J. N., 2000, ApJ, 531, L87

Malavasi N., Aghanim N., Douspis M., Tanimura H., Bonjean V., 2020, Astronomy & Astrophysics, 642, A19

Markwardt C. B., 2009, in Bohlender D. A., Durand D., Dowler P., eds, Astronomical Society of the Pacific Conference Series Vol. 411, Astronomical Data Analysis Software and Systems XVIII. p. 251 (arXiv:0902.2850)

Martinelli M., et al., 2021, Astronomy & Astrophysics, 649, A100

Massara E., Villaescusa-Navarro F., Viel M., 2014, J. Cosmology Astropart. Phys., 2014, 053

McCarthy I. G., Le Brun A. M. C., Schaye J., Holder G. P., 2014, Monthly Notices of the Royal Astronomical Society, 440, 3645

McCarthy I. G., Schaye J., Bird S., Le Brun A. M. C., 2017, Monthly Notices of the Royal Astronomical Society, 465, 2936

McCarthy I. G., Bird S., Schaye J., Harnois- Deraps J., Font A. S., van Waerbeke L., 2018, Monthly Notices of the Royal Astronomical Society, 476, 2999

McClelland J., Silk J., 1977, The Astrophysical Journal, 217, 331

Mead A. J., 2017, Monthly Notices of the Royal Astronomical Society, 464, 1282

Mead A. J., Verde L., 2021, Monthly Notices of the Royal Astronomical Society, 503, 3095

Mead J. M. G., King L. J., McCarthy I. G., 2010, Monthly Notices of the Royal Astronomical Society, 401, 2257

Mead A. J., Peacock J. A., Heymans C., Joudaki S., Heavens A. F., 2015, Monthly Notices of the Royal Astronomical Society, 454, 1958

Mead A. J., Heymans C., Lombriser L., Peacock J. A., Steele O. I., Winther H. A., 2016, Monthly Notices of the Royal Astronomical Society, 459, 1468

Mead A. J., Tröster T., Heymans C., Van Waerbeke L., McCarthy I. G., 2020, Astronomy & Astrophysics, 641, A130

Mead A. J., Brieden S., Tröster T., Heymans C., 2021, Monthly Notices of the Royal Astronomical Society,

Meerburg P. D., et al., 2019, BAAS, 51, 107

Mohammed I., Seljak U., 2014, Monthly Notices of the Royal Astronomical Society, 445, 3382

Mummery B. O., McCarthy I. G., Bird S., Schaye J., 2017, Monthly Notices of the Royal Astronomical Society, 471, 227

Münchmeyer M., Madhavacheril M. S., Ferraro S., Johnson M. C., Smith K. M., 2019, Phys. Rev. D, 100, 083508

Murray S., Poulin F., 2019, The Journal of Open Source Software, 4, 1397

Murray S. G., Diemer B., Chen Z., 2020, arXiv e-prints, p. arXiv:2009.14066

Nagai D., Kravtsov A. V., Vikhlinin A., 2007, The Astrophysical Journal, 668, 1

Navarro J. F., Frenk C. S., White S. D. M., 1997, The Astrophysical Journal, 490, 493

Navarro J. F., et al., 2010, Monthly Notices of the Royal Astronomical Society, 402, 21

Newman J. A., et al., 2013, ApJS, 208, 5

Nishimichi T., et al., 2019, The Astrophysical Journal, 884, 29

O'Neil S., Barnes D. J., Vogelsberger M., Diemer B., 2021, Monthly Notices of the Royal Astronomical Society, 504, 4649

Osato K., Sekiguchi T., Shirasaki M., Kamada A., Yoshida N., 2016, J. Cosmology Astropart. Phys., 2016, 004

Osato K., Flender S., Nagai D., Shirasaki M., Yoshida N., 2018, Monthly Notices of the Royal Astronomical Society, 475, 532

Osato K., Shirasaki M., Miyatake H., Nagai D., Yoshida N., Oguri M., Takahashi R., 2020, Monthly Notices of the Royal Astronomical Society, 492, 4780

Peacock J. A., Smith R. E., 2000, Monthly Notices of the Royal Astronomical Society, 318, 1144

Peebles P. J. E., 1993, Principles of Physical Cosmology

Peebles P. J. E., Yu J. T., 1970, The Astrophysical Journal, 162, 815

Pellejero-Ibañez M., Angulo R. E., Aricó G., Zennaro M., Contreras S., Stücker J., 2020, Monthly Notices of the Royal Astronomical Society, 499, 5257

Petri A., Haiman Z., May M., 2017, Phys. Rev. D, 95, 123503

Pfeifer S., McCarthy I. G., Stafford S. G., Brown S. T., Font A. S., Kwan J., Salcido J., Schaye J., 2020, Monthly Notices of the Royal Astronomical Society, 498, 1576

Philcox O. H. E., Spergel D. N., Villaescusa-Navarro F., 2020, Phys. Rev. D, 101, 123520

Pillepich A., et al., 2018, Monthly Notices of the Royal Astronomical Society, 473, 4077

Planck Collaboration 2014, Astronomy & Astrophysics, 571, A20

Planck Collaboration 2016, Astronomy & Astrophysics, 594, A13

Planck Collaboration 2018, preprint, p. arXiv:1807.06209 (arXiv:1807.06209)

Planck Collaboration et al., 2014, Astronomy & Astrophysics, 571, A1

Planck Collaboration et al., 2016, Astronomy & Astrophysics, 594, A22

Planck Collaboration et al., 2018, arXiv e-prints, p. arXiv:1807.06210

Press W. H., Schechter P., 1974, The Astrophysical Journal, 187, 425

Price D. J., 2008, Journal of Computational Physics, 227, 10040

Ramos-Ceja M. E., Basu K., Pacaud F., Bertoldi F., 2015, Astronomy & Astrophysics, 583, A111

Refregier A., Teyssier R., 2002, Phys. Rev. D, 66, 043002

Refregier A., Gamper L., Amara A., Heisenberg L., 2018, Astronomy and Computing, 25, 38

Riess A. G., et al., 1998, AJ, 116, 1009

Riess A. G., et al., 2004, The Astrophysical Journal, 607, 665

Riess A. G., et al., 2016, The Astrophysical Journal, 826, 56

Robertson N. C., et al., 2021, Astronomy & Astrophysics, 649, A146

Rogers K. K., Peiris H. V., Pontzen A., Bird S., Verde L., Font-Ribera A., 2019, J. Cosmology Astropart. Phys., 2019, 031

Roncarelli M., Moscardini L., Tozzi P., Borgani S., Cheng L. M., Diaferio A., Dolag K., Murante G., 2006, Monthly Notices of the Royal Astronomical Society, 368, 74

Roncarelli M., Moscardini L., Borgani S., Dolag K., 2007, Monthly Notices of the Royal Astronomical Society, 378, 1259

Ross A. J., et al., 2020, Monthly Notices of the Royal Astronomical Society, 498, 2354

Rudd D. H., Zentner A. R., Kravtsov A. V., 2008, The Astrophysical Journal, 672, 19

Ruppin F., et al., 2019, Astronomy & Astrophysics, 631, A21

Rybicki G. B., Lightman A. P., 1986, Radiative Processes in Astrophysics

Sachs R. K., Wolfe A. M., 1967, The Astrophysical Journal, 147, 73

Salvati L., Douspis M., Aghanim N., 2018, Astronomy & Astrophysics, 614, A13

Salvati L., Douspis M., Ritz A., Aghanim N., Babul A., 2019, Astronomy & Astrophysics, 626, A27

Sayers J., et al., 2016, The Astrophysical Journal, 832, 26

Schäfer B. M., Heisenberg L., Kalovidouris A. F., Bacon D. J., 2012, Monthly Notices of the Royal Astronomical Society, 420, 455

Schaller M., et al., 2015a, Monthly Notices of the Royal Astronomical Society, 451, 1247

Schaller M., et al., 2015b, Monthly Notices of the Royal Astronomical Society, 452, 343

Schaye J., Dalla Vecchia C., 2008, Monthly Notices of the Royal Astronomical Society, 383, 1210

Schaye J., et al., 2010, Monthly Notices of the Royal Astronomical Society, 402, 1536

Schaye J., et al., 2015, Monthly Notices of the Royal Astronomical Society, 446, 521

Scherrer R. J., Bertschinger E., 1991, The Astrophysical Journal, 381, 349

Schmidt F., 2016, Phys. Rev. D, 93, 063512

Schmidt F., Hu W., Lima M., 2010, Phys. Rev. D, 81, 063005

Schneider A., Teyssier R., 2015, J. Cosmology Astropart. Phys., 2015, 049

Schneider P., van Waerbeke L., Jain B., Kruse G., 1998, Monthly Notices of the Royal Astronomical Society, 296, 873

Schneider A., Teyssier R., Stadel J., Chisari N. E., Le Brun A. M. C., Amara A., Refregier A., 2019, J. Cosmology Astropart. Phys., 2019, 020

Schneider A., Giri S. K., Amodeo S., Refregier A., 2021, arXiv e-prints, p. arXiv:2110.02228

Scoccimarro R., Colombi S., Fry J. N., Frieman J. A., Hivon E., Melott A., 1998, The Astrophysical Journal, 496, 586

Seljak U., 1998, The Astrophysical Journal, 506, 64

Seljak U., 2000, Monthly Notices of the Royal Astronomical Society, 318, 203

Semboloni E., Hoekstra H., Schaye J., van Daalen M. P., McCarthy I. G., 2011, Monthly Notices of the Royal Astronomical Society, 417, 2020

Semboloni E., Hoekstra H., Schaye J., 2013, Monthly Notices of the Royal Astronomical Society, 434, 148

Sereno M., Fedeli C., Moscardini L., 2016, J. Cosmology Astropart. Phys., 2016, 042

Sharma S., 2017, ARA&A, 55, 213

Shaw L. D., Nagai D., Bhattacharya S., Lau E. T., 2010, The Astrophysical Journal, 725, 1452

Sheth R. K., Jain B., 2003, Monthly Notices of the Royal Astronomical Society, 345, 529

Sheth R. K., Mo H. J., Tormen G., 2001, Monthly Notices of the Royal Astronomical Society, 323, 1

Shirasaki M., Lau E. T., Nagai D., 2020, Monthly Notices of the Royal Astronomical Society, 491, 235

Sievers J. L., et al., 2013, J. Cosmology Astropart. Phys., 2013, 060

Sijacki D., Pfrommer C., Springel V., Enßlin T. A., 2008, Monthly Notices of the Royal Astronomical Society, 387, 1403

Smith R. E., Watts P. I. R., 2005, Monthly Notices of the Royal Astronomical Society, 360, 203

Smith R. E., et al., 2003, Monthly Notices of the Royal Astronomical Society, 341, 1311

Smith R. E., Scoccimarro R., Sheth R. K., 2007, Phys. Rev. D, 75, 063512

Smith R. E., Desjacques V., Marian L., 2011, Phys. Rev. D, 83, 043526

Smoot G. F., et al., 1992, ApJ, 396, L1

Somerville R. S., Davé R., 2015, ARA&A, 53, 51

Spergel D. N., et al., 2003, ApJS, 148, 175

Springel V., 2005, Monthly Notices of the Royal Astronomical Society, 364, 1105

Springel V., 2010, ARA&A, 48, 391

Springel V., Hernquist L., 2003, Monthly Notices of the Royal Astronomical Society, 339, 289

Springel V., et al., 2008, Monthly Notices of the Royal Astronomical Society, 391, 1685

Spurio Mancini A., Piras D., Alsing J., Joachimi B., Hobson M. P., 2021, arXiv e-prints, p. arXiv:2106.03846

Stafford S. G., McCarthy I. G., Crain R. A., Salcido J., Schaye J., Font A. S., Kwan J., Pfeifer S., 2020a, Monthly Notices of the Royal Astronomical Society, 493, 676

Stafford S. G., Brown S. T., McCarthy I. G., Font A. S., Robertson A., Poole-McKenzie R., 2020b, Monthly Notices of the Royal Astronomical Society, 497, 3809

Sullivan J. M., Seljak U., Singh S., 2021, arXiv e-prints, p. arXiv:2104.10676

Sunyaev R. A., Zeldovich I. B., 1981, Astrophysics and Space Physics Reviews, 1, 1

Takahashi R., Sato M., Nishimichi T., Taruya A., Oguri M., 2012, The Astrophysical Journal, 761, 152

Tanimura H., et al., 2020a, Monthly Notices of the Royal Astronomical Society, 491, 2318

Tanimura H., Aghanim N., Kolodzig A., Douspis M., Malavasi N., 2020b, Astronomy & Astrophysics, 643, L2

Tanimura H., Douspis M., Aghanim N., Salvati L., 2021, arXiv e-prints, p. arXiv:2110.08880

Teyssier R., Pontzen A., Dubois Y., Read J. I., 2013, Monthly Notices of the Royal Astronomical Society, 429, 3068

Tinker J., Kravtsov A. V., Klypin A., Abazajian K., Warren M., Yepes G., Gottlöber S., Holz D. E., 2008, The Astrophysical Journal, 688, 709

Tinker J. L., Robertson B. E., Kravtsov A. V., Klypin A., Warren M. S., Yepes G., Gottlöber S., 2010, The Astrophysical Journal, 724, 878

Trac H., Bode P., Ostriker J. P., 2011, The Astrophysical Journal, 727, 94

Tröster T., et al., 2021, arXiv e-prints, p. arXiv:2109.04458

Valageas P., Nishimichi T., 2011, Astronomy & Astrophysics, 527, A87

Van Waerbeke L., et al., 2013, Monthly Notices of the Royal Astronomical Society, 433, 3373

Velliscig M., van Daalen M. P., Schaye J., McCarthy I. G., Cacciato M., Le Brun A. M. C., Dalla Vecchia C., 2014, Monthly Notices of the Royal Astronomical Society, 442, 2641

Velliscig M., et al., 2015, Monthly Notices of the Royal Astronomical Society, 454, 3328

Velliscig M., et al., 2017, Monthly Notices of the Royal Astronomical Society, 471, 2856

Verde L., Treu T., Riess A. G., 2019, Nature Astronomy, 3, 891

Villaescusa-Navarro F., et al., 2020, ApJS, 250, 2

Vogelsberger M., Genel S., Sijacki D., Torrey P., Springel V., Hernquist L., 2013, Monthly Notices of the Royal Astronomical Society, 436, 3031

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

Voit G. M., 2005, Reviews of Modern Physics, 77, 207

Voivodic R., Rubira H., Lima M., 2020, J. Cosmology Astropart. Phys., 2020, 033

Wang Y., et al., 2020, Monthly Notices of the Royal Astronomical Society, 498, 3470

Wei C., et al., 2018, The Astrophysical Journal, 853, 25

Weinberg D. H., Mortonson M. J., Eisenstein D. J., Hirata C., Riess A. G., Rozo E., 2013, Phys. Rep., 530, 87

White S. D. M., Frenk C. S., 1991, The Astrophysical Journal, 379, 52

Wiersma R. P. C., Schaye J., Smith B. D., 2009a, Monthly Notices of the Royal Astronomical Society, 393, 99

Wiersma R. P. C., Schaye J., Theuns T., Dalla Vecchia C., Tornatore L., 2009b, Monthly Notices of the Royal Astronomical Society, 399, 574

Wong K. C., et al., 2020, Monthly Notices of the Royal Astronomical Society, 498, 1420

Yan Z., et al., 2021, arXiv e-prints, p. arXiv:2102.07701

Yoon M., Jee M. J., 2021, The Astrophysical Journal, 908, 13

Zavala J., Balogh M. L., Afshordi N., Ro S., 2012, Monthly Notices of the Royal Astronomical Society, 426, 3464

Zeldovich Y. B., 1972, Monthly Notices of the Royal Astronomical Society, 160, 1P

Zhao G.-B., et al., 2021, Monthly Notices of the Royal Astronomical Society, 504, 33

da Silva A. C., Barbosa D., Liddle A. R., Thomas P. A., 2000, Monthly Notices of the Royal Astronomical Society, 317, 37

van Daalen M. P., Schaye J., 2015, Monthly Notices of the Royal Astronomical Society, 452, 2247

van Daalen M. P., Schaye J., Booth C. M., Dalla Vecchia C., 2011, Monthly Notices of the Royal Astronomical Society, 415, 3649

van Daalen M. P., McCarthy I. G., Schaye J., 2020, Monthly Notices of the Royal Astronomical Society, 491, 2424

van Uitert E., Gilbank D. G., Hoekstra H., Semboloni E., Gladders M. D., Yee H. K. C., 2016, Astronomy & Astrophysics, 586, A43

van de Weygaert R., 2016, in van de Weygaert R., Shandarin S., Saar E., Einasto J., eds, Vol. 308, The Zeldovich Universe: Genesis and Growth of the Cosmic Web. p. 493 (arXiv:1611.01222), doi:10.1017/S1743921316010504

van den Bosch F. C., More S., Cacciato M., Mo H., Yang X., 2013, Monthly Notices of the Royal Astronomical Society, 430, 725