



HAL
open science

From halos to cosmic web: three studies of the large-scale structures

Emma Ayçoberry

► **To cite this version:**

Emma Ayçoberry. From halos to cosmic web: three studies of the large-scale structures. Astrophysics [astro-ph]. Sorbonne Université, 2024. English. NNT : 2024SORUS339 . tel-04879063

HAL Id: tel-04879063

<https://theses.hal.science/tel-04879063v1>

Submitted on 10 Jan 2025

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

SORBONNE UNIVERSITÉ

ÉCOLE DOCTORALE 127: ASTRONOMIE ET ASTROPHYSIQUE D'ÎLE-DE-FRANCE

INSTITUT D'ASTROPHYSIQUE DE PARIS

From halos to cosmic web: three studies of the large-scale structures

DOCTORAL THESIS IN ASTROPHYSICS

Author:

Emma Ayçoberry

Supervisors:

Karim Benabed & Yohan Dubois

PRESENTED AND PUBLICLY DEFENDED ON SEPTEMBER, 20TH 2024
IN FRONT OF THE FOLLOWING JURY

Pierre Astier	Research Director, CNRS, LPNHE	President
Björn Malte Schäfer	Professor, Heidelberg University	Referee
Etienne Pointecouteau	Research Director, CNRS, IRAP	Referee
Elisabeth Krause	Associate Professor, University of Arizona	Examiner
Ludovic Van Waerbeke	Professor, University of British Columbia	Examiner
Karim Benabed	Astronomer, SU, IAP	Co-supervisor
Yohan Dubois	Researcher, CNRS, IAP	Co-supervisor

INSTITUT D'ASTROPHYSIQUE DE PARIS

Unité mixte de recherche 7095



CNRS - Sorbonne Université

Acknowledgements

I would like to express my heartfelt gratitude to all the people who supported me throughout this PhD journey. Thanks to all of you, this has been an amazing experience that has taught me so much, both scientifically and personally. This journey has for sure helped me grow.

First and foremost, my deepest thanks go to Karim Benabed and Yohan Dubois, my PhD advisors. It has been an absolute pleasure to work with both of you, and I feel very lucky to have you as advisors. It has been an enriching experience, and I have learned so much from your expertise. Your complementary skills have provided me with valuable insights and helped me develop as a scientist. I am grateful that your doors were always open, whether it was to answer questions - simple or complex - or to discuss various aspects of PhD life. Thank you for your guidance, feedback, and support throughout this journey.

I would also like to warmly thank my colleagues from the University of Arizona, with whom I had the chance to collaborate during my PhD. It was such a nice opportunity to broaden my expertise and learn from you.

A sincere thank you to the members of my PhD jury. I am deeply appreciative of the time you dedicated, as well as your thoughtful questions and remarks. There were very stimulating and left me plenty of ideas to reflect upon and explore. A special thanks to Björn Malte Schäfer and Etienne Pointecouteau for accepting to review my manuscript in detail. All your comments were very useful to me.

I am also grateful to the CNRS for funding and enabling my collaboration with the University of Arizona. My gratitude extends to the doctoral school and, in particular, Frédéric Daigne and Thierry Fouchet for their support through the different processes and their dedication to the well-being of PhD students.

My warmest thanks go to Florence Durret and Hakim Atek, members of my PhD committee. Thank you for your unconditional support and reassurance. Knowing that I could rely on you if needed was a source of great comfort throughout this journey.

I feel lucky to have completed my PhD at the Institut d'Astrophysique de Paris. The supportive environment created by the direction and administrative teams was a real pleasure. As a PhD representative for two years, I was able to interact with you even more closely, and I truly appreciated your openness.

More broadly, I want to extend my thanks to everyone at IAP. I will always remember the people I met, the science we discussed, and the fun moments we shared. I also want to thank all the scientists I met at various conferences—it was always a pleasure to connect and broaden my horizons.

There are a few other scientists I would like to thank individually: Sébastien Rodriguez, Antoine Lucas, Anaïs Abramian, Sandrine Codis, and Martin Kilbinger. Anaïs, thank you for helping clarify how I wanted to pursue my academic career. Sébastien and Antoine, thank you for giving me my first opportunity to work in research, even though it was far from my current field (dune morphodynamics on Titan). You were the first to spark my love for research. Sandrine, thank you for mentoring me during my internship and teaching me so much. Your advice at different stages of my career has been invaluable, and I've greatly enjoyed staying connected with you. Finally, Martin, thank you for the experience of working with you and the Cosmostat team. It has been such a positive environment that I'm thrilled to continue my postdoc with this group.

Of course, a special thank to all the PhD students I have had the pleasure of meeting. While I won't list everyone by name to avoid accidentally leaving someone out, you will recognize you. Thank you for all the shared lunches, breaks, and conversations that we had. It was a real pleasure to go through this journey with you.

A special thanks to Emilie, Simon, and Axel, my office mates for the past three years. Sharing this experience with you was amazing. It has helped a lot to move forward in any conditions, and I can confidently say that office 9 was the best.

I also want to thank my friends from different stages of my life for their curiosity and support. Thank you for all the laugh and needed breaks you provided. Meeting with you always helped me recharge and decompress.

Finally, I want to express my deepest gratitude to my family. Your enthusiasm and support for my work have been constant, even if some of you still struggle to explain exactly what I do! None of this would have been possible without your encouragement throughout my academic journey. Because you always believed in me, supported me, and pushed me to follow my passions, I've been able to achieve this journey. Thank you for everything!

Abstract

This thesis investigates the study of the complexities of the large-scale structures in the Universe, focusing on the thermal Sunyaev-Zel'dovich effect (tSZ), the cosmic web, and weak lensing peak counts. Understanding the evolution of these structures offers crucial insights into various properties of the Universe, including the amount of dark matter, dark energy, and baryonic matter. After providing an overview of cosmology and structure formation, I will introduce a few observables and the halo model framework. I will then describe the methods used to run simulations and present the specific simulations utilized in this research. Following this, I will present my results on the tSZ effect, particularly a comparison between tSZ properties measured in simulations and those predicted by a halo model. I will demonstrate that the halo model tSZ power spectrum shows discrepancies with simulation measurements. Additionally, I will discuss results on the tSZ effect within the context of the w CDM model, showing that it is indeed sensitive to the dark energy equation of state w . Next, I will present results on the theoretical prediction of the abundances of the different cosmic web environments – voids, walls, filaments, and knots – demonstrating that these theoretical predictions align well with measurements from simulations. Finally, I will discuss the impact of systematics on weak lensing peak count statistics. I will highlight the importance of careful consideration of systematics and uncertainties such as shear calibration, baryonic feedback, and the choice of signal-to-noise ratio in future analyses.

Keywords: large-scale structures of the Universe, cosmology, thermal Sunyaev-Zel'dovich effect, cosmic web, peak counts, simulations

Résumé

Cette thèse se concentre sur l'étude des complexités des grandes structures de l'Univers, en utilisant l'effet Sunyaev-Zel'dovich thermique (tSZ), la toile cosmique et la statistique de comptage de pics sur le lentillage gravitationnel. Comprendre l'évolution de ces structures fournit des informations sur diverses propriétés de l'Univers, telles que la quantité de matière noire, d'énergie noire et de matière baryonique. Après avoir rappelé le contexte cosmologique et la formation des structures, je présenterai quelques observables et le modèle de halo. Je décrirai ensuite les méthodes utilisées pour générer des simulations numériques, et présenterai celles que j'utilise dans cette thèse. Je continuerai en présentant mes résultats sur l'effet tSZ, en particulier une comparaison entre les mesures des propriétés tSZ dans les simulations et celles prédites par un modèle de halo. Je démontrerai que les spectres de puissance du tSZ prédits par ce modèle de halo montrent des différences avec ceux mesurés dans les simulations. De plus, je présenterai des résultats sur l'effet tSZ dans le contexte du modèle w CDM, montrant que celui-ci est en effet sensible à la valeur de l'équation d'état d'énergie noire w . Ensuite, je présenterai des résultats sur la prédiction théorique des abondances des différents environnements de la toile cosmique – vides, murs, filaments, et nœuds – démontrant que les prédictions théoriques sont en bon accord avec les mesures dans les simulations. Enfin, je discuterai de l'impact des systématiques sur les statistiques de comptage de pics de l'effet de lentille gravitationnelle, montrant que les systématiques et les incertitudes telles que la calibration du cisaillement, l'impact baryonique et le choix du rapport signal sur bruit doivent être soigneusement pris en compte dans les analyses futures.

Mots clés: grandes structures de l'Univers, cosmologie, effect Sunyaev-Zel'dovich thermique, toile cosmique, comptage de pics, simulations

Résumé long

Partie I - Contexte

Chapitre 1 - Cosmologie

Dans ce premier chapitre de ma thèse, je présente les bases théoriques de la cosmologie, en commençant par un aperçu du modèle standard de la cosmologie, le modèle Λ CDM. Ce modèle décrit l'Univers composé d'une constante cosmologique (Λ) responsable de l'expansion de l'Univers, de matière noire froide (CDM) et de matière baryonique. Je retrace également l'évolution de l'Univers depuis le Big Bang pour brièvement décrire ces différentes phases.

J'introduis ensuite la relativité générale, notamment le principe d'équivalence et la courbure de l'espace-temps. Je développe ensuite les équations qui permettent de décrire l'expansion et la structure d'un Univers homogène à travers les équations de Friedmann et le paramètre de Hubble. Ce cadre permet d'introduire les différents paramètres cosmologiques (quantité de matière, de radiation, d'énergie noire,...) ainsi que les différentes époques cosmologiques.

Ce chapitre poursuit avec plus de détail sur le modèle Λ CDM, notamment son accord avec les observations cosmologiques, mais aussi ses limitations, notamment sur des petites échelles. J'introduit brièvement des possibles extensions, telles que le modèle w CDM, qui remplace la constante cosmologique par une équation d'état d'énergie noire qui peut évoluer dans le temps.

Je présente ensuite deux périodes clés du début de l'Univers : l'inflation cosmique et le fond diffus cosmologique. L'inflation, une phase d'expansion très rapide de l'Univers, est une solution aux problèmes du modèle du Big Bang, notamment le problème de l'horizon et de la platitude. L'inflation permet également d'expliquer les perturbations primordiales qui ont conduit à la formation des structures à grande échelle que l'on peut observer aujourd'hui. Ces fluctuations sont observables dans le fond diffus cosmologique, qui est la première lumière de l'Univers. Je présente également l'impact des différents paramètres cosmologiques sur ce fond diffus cosmologique pour comprendre comment son étude peut nous aider à contraindre la cosmologie.

Enfin, je présente l'évolution des structures à grande échelle de l'Univers, en analysant les perturbations de densité et leur croissance sous l'effet de la gravité dans un cadre inhomogène. Les équation d'Euler et l'approximation de Zel'dovich sont présentées pour modéliser ces structures. On apprend comment les structures évoluent et peuvent être classées en différents environnements : vide, mur, filament et noeud. Je finis en présentant le modèle d'effondrement sphérique qui est une approximation permettant de décrire la formation des halos dans le régime linéaire.

En conclusion, ce chapitre fournit une base pour comprendre l'évolution des perturbations de densité et la formation des structures, ouvrant la voie aux analyses des grandes structures de l'Univers dans les chapitres suivants.

Chapitre 2 - Grandes structures de l'Univers

Ce deuxième chapitre est consacré à l'étude des structures à grande échelle de l'Univers en mettant en avant leur rôle fondamental dans la compréhension de l'évolution et de la composition de l'Univers.

Je commence par introduire des outils statistiques qui permettent d'étudier ces structures. Je présente notamment le spectre de puissance, qui permet de décrire la distribution d'une certaine observable (matière, pression,...) à différentes échelles spatiales. Cette statistique contient toute l'information Gaussienne et permet d'analyser les fluctuations de densité. Je présente ensuite des statistiques d'ordre supérieur, qui sont utilisées pour capturer les non-Gaussianités de la distribution de matière. On peut par exemple citer les fonctions à trois points ou les comptages de pics dans les cartes de lentillage gravitationnelle.

Je continue en détaillant différentes observables qui peuvent être utilisées pour étudier les grandes structures. La matière, majoritairement constituée de matière noire, est l'une des observables les plus fondamentales. Sa répartition est étudiée par des simulations numériques ou indirectement via des traceurs. On peut par exemple citer les galaxies, qui tracent la distribution de matière. Cette distribution est biaisé puisque majoritairement basée sur la matière visible. Le cisaillement gravitationnel faible est une autre observable, qui résulte de la déviation de la lumière par la matière. Il est donc sensible à la matière totale, y compris la matière noire, et est un outil complémentaire à d'autres observables. Finalement, j'introduis l'effet Sunyaev-Zel'dovich thermique (tSZ), qui trouve son origine dans le gaz chaud des amas de galaxies. Cet effet est donc sensible à la matière baryonique, et nous permet de contraindre à la fois les paramètres cosmologiques et les processus astrophysiques.

Dans la suite du chapitre, j'introduis le modèle de halo qui est un outil analytique permettant de modéliser le spectre de puissance de la distribution de matière et de ses traceurs. Le modèle de halo se base sur l'hypothèse que toute la matière se trouve dans des halos sphériques et symétriques. Les trois principales composantes de ce modèle sont : la fonction de masse des halos qui décrit la distribution des halos en fonction de la masse et du redshift, le biais de halo qui décrit comment les halos se regroupent par rapport à la matière et le profil de halo qui décrit comment le traceur est distribué autour d'un halo.

Le modèle de halo permet ensuite de prédire le spectre de puissance du traceur étudié comme la somme de deux termes : le terme un-halo et le terme deux-halo. Le terme un-halo correspond à la contribution de la matière au sein d'un seul halo, et est donc caractéristique des petites échelles. Le terme deux-halos décrit la corrélation entre différents halos, caractérisant donc les grandes échelles. Ce modèle permet de prédire rapidement les spectres de puissance de différentes observables, mais il présente des limitations. Par exemple, il suppose que toute la matière est contenue dans des halos sphériques, ce qui n'est pas toujours réaliste. Il est aussi dépendant des choix faits pour décrire les différentes composantes du modèle.

Ce chapitre illustre l'importance des structures à grande échelle pour contraindre les paramètres cosmologiques et comprendre les processus astrophysiques. La combinaison des outils théoriques, des observations et des simulations numériques constituent une approche essentielle pour raffiner notre connaissance de l'Univers et de son évolution.

Partie II - Simulations numériques

Dans ce chapitre (Chapitre 3), j'introduis les simulations cosmologiques utilisées pour étudier les structures à grande échelle de l'Univers.

Je présente tout d'abord les méthodes permettant de créer une simulation numérique. La première étape consiste à créer les conditions initiales d'une simulation. Je cite des méthodes comme l'approximation de Zel'dovich qui permet de reproduire les modes linéaires ou la théorie des perturbations lagrangiennes du second ordre (2LPT) qui est capable de capturer les modes non linéaires.

Je présente ensuite comment faire évoluer ces conditions initiales à travers l'exemple du code **RAMSES**, qui est un code de simulation basé sur un schéma de raffinement de maillage adaptatif (AMR), permettant une résolution élevée dans les régions denses. Pour faire évoluer une simulation ne contenant que de la matière noire, le solveur N-corps résout le système de Vlasov-Poisson, pour suivre l'évolution gravitationnelle des particules. Pour obtenir des simulations plus réalistes, il est nécessaire d'inclure les processus baryoniques tels que l'impact des noyaux de galaxies actifs (AGN) sur la distribution du gaz ou la formation stellaire. Dans ce cas, le solveur hydrodynamique résout les équations d'Euler sous forme conservative. **RAMSES** permet l'utilisation de critères pour gérer le raffinement et le choix des pas de temps de la simulation, permettant ainsi de mieux résoudre les régions dynamiques actives.

Je présente ensuite les différentes simulations utilisées dans cette thèse :

- **Horizon-AGN** : simulation hydrodynamique de $100 h^{-1}$ Mpc. La physique inclut un refroidissement du gaz, la formation stellaire, et les rétroactions des AGN et supernova.
- **Horizon-noAGN** : variante de la simulation précédente, sans AGN. Cette simulation permet d'évaluer l'impact de ce dernier.
- **Horizon-Large** : simulation de $896 h^{-1}$ Mpc avec une physique simplifiée (contient seulement le chauffage et refroidissement du gaz, et pas de physique galactique). Cette simulation permet d'étudier des plus grands volumes et donc des halos plus massifs à moindre coûts computationnels.
- **L896_wCDM** : simulations de $896 h^{-1}$ Mpc, avec la même physique que la simulation **Horizon-Large**, mais avec une cosmologie différente qui explore l'effet d'une énergie noire avec une équation d'état w variant $(-0,8, -1, -1,2)$.
- **Magneticum** : simulations de $896 h^{-1}$ Mpc utilisant le code **GADGET**, avec une physique similaire à la simulation **Horizon-AGN** mais utilisant une méthodologie différente.
- **BAHAMAS** : simulations hydrodynamiques de $400 h^{-1}$ Mpc. Ces simulations ne sont pas directement utilisées dans ma thèse, mais sont utilisées pour calibrer le modèle de halo **HMx** utilisé dans le chapitre 4.

Je finis ce chapitre en présentant la fonction de masse des halos des différentes simulations. Cette fonction traduit l'abondance des halos de différentes masses et dépend des processus physiques intégrés et du volume simulé. Les simulations **Horizon** et **Magneticum** montrent un bon accord global, mais les limites en résolution et volume influencent les halos accessibles. Pour les simulations **L896_wCDM**, on voit

que la fonction de masse est sensible à la valeur de w , encore plus marquée avec les halos plus massifs. Plus haut est w , plus massifs et plus nombreux sont les halos.

Dans ce chapitre, je souligne donc l'importance des simulations cosmologiques pour étudier les structures à grande échelle de l'Univers. Je décris leur utilisation et explicite celles étudiées dans cette thèse.

Partie III - Résultats et publications

Chapitre 4 - Propriétés de l'effet Sunyaev-Zel'dovich thermique

Dans ce chapitre, je présente avec plus de détails la modélisation, l'observation et la comparaison des propriétés de l'effet Sunyaev-Zel'dovich thermique (tSZ) avec des mesures issues de simulations numériques.

Je commence en présentant le modèle de halo HMx et les choix faits pour ces différentes composantes. Je détaille en particulier son utilisation pour prédire le spectre de l'effet tSZ. La composante majeure de ce spectre est le profil de pression électronique qui dépend de la densité et de la température du gaz, modélisé notamment par des paramètres ajustables comme l'indice polytropique. Les différents paramètres du modèle sont ajustés à partir des spectres de puissance des simulations BAHAMAS, et différents modèles existent en fonction des spectres utilisés. Dans notre cas, nous utilisons le modèle ajusté sur le spectre matière-pression.

Je présente ensuite quelques observations de l'effet tSZ, notamment la mesure sur des cartes couvrant de vastes champs, comme celles obtenues par Planck ou SPT et celles faites à l'échelle de halos individuels comme avec NIKA2.

Je finis cette partie en présentant une méthodologie d'analyse combinant plusieurs observables (tSZ, lentillage gravitationnel, lentillage gravitationnel du fond diffus cosmologique et relevé de galaxies). Cette étude montre une amélioration significative des contraintes sur les paramètres cosmologiques, qui peut s'améliorer en affinant la modélisation de l'effet tSZ qui incorpore beaucoup de paramètres libres au modèle.

La suite de ce chapitre est consacrée à la présentation des différents résultats obtenus. J'ai tout d'abord comparé les prédictions du modèle HMx avec les spectres de puissance mesurés dans plusieurs simulations (Horizon-AGN, Horizon-noAGN, Horizon-Large et Magneticum). Les résultats sur les spectres de l'effet tSZ montrent des divergences croissantes lorsque le redshift augmente. Pour mieux comprendre ces différences, je me suis intéressée à différentes propriétés du modèle de halo. La décomposition de HMx en termes un- et deux-halo montre que le terme deux-halo domine à grande distance, et est majoritairement à l'origine de l'excès de puissance prédit par le modèle. J'ai également étudié, dans les simulations, la contribution de la matière se trouvant à l'intérieur ou à l'extérieur d'un rayon du viriel des halos. Les résultats montrent que plus le redshift augmente, plus la contribution de la matière se trouvant en dehors des halos augmente, et devient même plus importante que la matière se trouvant dans un rayon du viriel à $z \sim 3$.

Comme le modèle est ajusté au niveau du spectre de puissance, des différences sont attendues et les propriétés telles que le profil doivent être vues comme des propriétés effectives, mais il est tout de même intéressant de comparer ces propriétés pour comprendre les dégénérescences. J'ai tout d'abord comparé les profils de pression électroniques qui montrent en effet des différences, en particulier à faible masse. J'ai également comparé d'autres propriétés telles que l'indice polytropique, le paramètre qui capture la déviation à une température du viriel, la fraction de gaz lié et la concentration. Les mesures de ces paramètres révèlent des différences avec ceux utilisés dans le modèle, mais il est surtout intéressant de noter que tout ces paramètres dépendent de la masse et du redshift, dépendances qui ne sont pas toujours prises en compte dans le modèle. Il serait intéressant d'étudier si un profil de pression plus en accord avec la mesure, ou des paramètres qui dépendent de la masse et du redshift aident à retrouver des spectres puis proche de ceux mesurés.

Je finis en présentant les résultats sur l'impact de la variation de l'équation d'état de l'énergie noire (w) à l'aide des simulations L896_wCDM. Les résultats montrent que le spectre de puissance tSZ, et celui de la matière, sont sensibles à la valeur de w , offrant un potentiel pour contraindre des modèles cosmologiques alternatifs. On remarque notamment que plus w est grand, plus l'expansion de l'Univers commence tôt et donc plus les structures croissent lentement. Ce comportement impose des structures plus développées plus tôt dans l'histoire de l'Univers, ce qui est observé dans ces simulations.

Pour conclure, ce chapitre se concentre sur l'étude du modèle de halo pour modéliser l'effet tSZ. La comparaison avec les simulations révèle des écarts importants qu'il est nécessaire de comprendre. L'effet tSZ reste une observable puissante pour explorer les propriétés baryoniques et cosmologiques, à condition d'améliorer la modélisation.

Chapitre 5 - La toile cosmique et la classification T-web

Dans ce chapitre, je m'intéresse à la description des différents environnements de la toile cosmique, en particulier à leurs classifications. J'introduis tout d'abord la théorie des perturbations pour décrire l'évolution non linéaire des structures, qui en particulier a lieu à bas redshift.

Je présente ensuite comment la toile cosmique est décrite comme un réseau de vides, murs, filaments et noeuds résultant de l'évolution gravitationnelle et je présente plusieurs méthodes de classification de ces structures. Je présente notamment la méthode du T-web, qui est la méthode que j'utilise et qui repose sur les valeurs propres des dérivées secondes du potentiel gravitationnel. Le nombre de valeurs propres supérieures à un seuil donné détermine le type d'environnement :

- vide : zéro valeur propre supérieure au seuil,
- mur : une valeur propre supérieure au seuil,
- filament : deux valeurs propres supérieures au seuil,
- noeud : trois valeurs propres supérieures au seuil.

La classification des structures avec cette méthode peut être faite dans des simulations numériques et je me suis intéressée à construire un modèle analytique dérivé de principes fondamentaux pour prédire l'abondance des environnements. Dans une première étape, je considère un champ de densité Gaussien et utilise la formule de Doroskevich pour décrire la distribution jointe des valeurs propres. Dans un second temps, des corrections non Gaussiennes sont introduites via une expansion de Gram-Charlier, qui prend en compte des termes d'ordre supérieur tel que la skewness. J'ai ensuite comparé les probabilités des environnements (vides, murs, filaments, noeuds) dans les simulations *Quijote* pour différents redshifts et échelles de lissage avec les prédictions analytiques. Les prédictions avec corrections non Gaussiennes montrent un très bon accord (bien meilleur que les prédictions Gaussiennes) avec les mesures des simulations, même à des échelles non Gaussiennes (petite échelle de lissage et faible redshift).

Ce chapitre présente un cadre théorique robuste pour prédire les abondances des environnements du réseau cosmique en utilisant le formalisme du T-web. Cette méthode offre une alternative analytique aux simulations coûteuses et ouvre des nouvelles perspectives.

Chapitre 6 - Statistiques de comptage de pics de l'effet de lentille gravitationnelle

Dans ce dernier chapitre, je présente une méthode statistique d'ordre supérieur, le comptage de pics, appliquée au cisaillement gravitationnel faible, qui est donc sensible aux non-gaussianités.

Dans ce travail, je me suis intéressée à l'impact de différentes incertitudes et systématiques sur les paramètres cosmologiques dérivés du comptage de pics. Les prédictions sont basées sur les simulations *MassiveNuS* et les données proviennent du relevé en cours *UNIONS*. Les comptages de pics sont obtenus en identifiant les maxima locaux de cartes de rapport signal/bruit (SNR). Ces maxima sont ensuite utilisés pour inférer les paramètres cosmologiques.

Un aspect important concerne la calibration du cisaillement, réalisée via la méthode de metacalibration. Cette méthode estime la matrice de réponse du cisaillement, qui relie le cisaillement réel au cisaillement mesuré tout en prenant en compte les biais multiplicatifs et additifs. Je me suis tout d'abord concentrée sur l'utilisation d'une méthode de calibration locale du cisaillement. Dans ce cas, la calibration est effectuée sur des sous-échantillons de différentes tailles (0.5 à 4 deg²) pour évaluer l'impact des effets locaux comme la fonction d'étalement des points (PSF) ou la taille des galaxies.

J'ai ensuite étudié l'impact de cette calibration locale sur les contraintes cosmologiques, ainsi que l'impact de différentes systématiques : le biais multiplicatif résiduel, l'incertitude sur le redshift, l'impact des effets baryoniques, l'alignement intrinsèque et facteur de boost.

Les analyses montrent que des calibrations effectuées sur des petits sous-échantillons (≥ 1 deg²) se rapprochent des calibrations globales avec des déviations standards faibles. Cela suggère qu'une calibration locale est utile mais doit être effectuée sur des échelles suffisamment grandes pour être robuste.

Les simulations *MassiveNuS* permettent de contraindre les paramètres M_ν , Ω_m , et A_s et je me suis intéressée à l'évolution des contraintes quand différentes systématiques sont prises en compte. Les résultats mettent en avant l'importance d'une calibration précise et d'une prise en compte rigoureuse de l'alignement intrinsèque et du facteur de boost pour obtenir des contraintes robustes.

Conclusions et perspectives

Mon travail s'est concentré sur l'étude des structures à grande échelle de l'Univers afin d'améliorer la capacité des différents observables à contraindre les paramètres cosmologiques et astrophysiques. Les résultats portent sur trois axes principaux : l'effet Sunyaev-Zel'dovich thermique (tSZ), la toile cosmique (cosmic web), et les statistiques de comptage de pics.

Concernant l'effet tSZ, mon travail a mis en évidence que le modèle de halo utilisé dans HMx ne parvient pas à reproduire précisément les propriétés de l'effet tSZ mesurées dans les simulations. Les écarts augmentent avec le redshift, et des différences de 20 à 50% sont observées dans le spectre de puissance angulaire du tSZ. Ces écarts pourraient être liés aux hypothèses simplifiées du modèle de halo, comme l'hypothèse que toute la matière se trouve dans des halos, ou à des paramètres ne capturant pas toute la complexité physique par exemple.

Par ailleurs, l'effet tSZ est sensible au modèle cosmologique, notamment à l'énergie noire qui influence la croissance des structures. Dans ce contexte, j'ai exploré les propriétés de l'effet tSZ dans des simulations w CDM, montrant que l'effet est bien sensible à la valeur de w . Ces résultats encouragent à poursuivre les efforts pour modéliser l'effet tSZ sous différentes cosmologies afin d'en faire un outil robuste pour sonder l'énergie noire et la matière baryonique.

Concernant la toile cosmique, j'ai développé un modèle théorique prédisant l'abondance des environnements en fonction du redshift et de l'échelle de lissage, en se basant sur la méthode du T-web. Les comparaisons avec les mesures des simulations *Quijote* montrent un excellent accord, y compris aux échelles non linéaires.

Finalement, concernant les comptages de pics du lentillage gravitationnel, j'ai étudié l'impact de différentes systématiques sur les contraintes cosmologiques des données du relevé UNIONS. Ces analyses montrent que les incertitudes et systématiques peuvent décaler les paramètres cosmologiques inférés, en particulier Ω_m . Ce travail souligne l'importance de considérer soigneusement ces effets pour obtenir des contraintes robustes dans les analyses futures.

Suite à ce travail, je conclus en mentionnant quelques perspectives. Tout d'abord, il serait important d'améliorer la modélisation de l'effet tSZ. Par exemple, développer un modèle permettant de reproduire précisément le spectre de puissance tSZ à partir d'un profil de pression donné pourrait être intéressant pour inférer les spectres de l'effet tSZ à partir de profils de pression mesurés avec NIKA2. Il serait également intéressant de prendre en compte des aspects plus réalistes, comme des halos non sphériques, ainsi que des paramètres variant avec la masse et le redshift pour améliorer la précision des prédictions.

Un autre aspect concerne l'étude de la contribution des différents environnements de la toile cosmique au spectre total de l'effet tSZ pour mieux comprendre les environnements à prendre en compte dans la modélisation.

Finalement, il est intéressant d'étendre les analyses à des modèles cosmologiques plus complexes, tels que des équations d'état d'énergie noire avec évolution temporelle ou des théories de gravité modifiée.

Sur les comptages de pics, il serait intéressant de continuer les efforts pour inclure plus précisément toutes les systématiques dans le modèle afin d'extraire des informations cosmologiques robustes.

Dans cette thèse, j'ai étudié différentes méthodes pour analyser les propriétés des structures à grande échelle de l'Univers. Chacune présente des avantages et des limites, rendant chaque approche pertinente pour des problématiques spécifiques. En exploitant les non-gaussianités et les statistiques d'ordre supérieur, ces travaux contribuent à améliorer notre compréhension des structures complexes de l'Univers, permettant ainsi de contraindre plus solidement les paramètres cosmologiques et astrophysiques.

Contents

Acknowledgements	iii
Abstract	v
Résumé	iv
Résumé long	v
Introduction	1
I Context	3
1 Cosmology	4
1.1 Concordant model and history of the Universe	5
1.2 Introduction to general relativity	6
1.2.1 Equivalence principle	6
1.2.2 Curvature of space-time	6
1.3 Homogeneous Universe	8
1.3.1 Friedmann-Lemaître-Robertson-Walker metric and Hubble parameter	8
1.3.2 Friedmann equations	9
1.3.3 Hubble radius and epoch of the Universe	10
1.4 Λ CDM model	12
1.4.1 Cosmological constant	12
1.4.2 Cold dark matter	12
1.4.3 Limitations of the Λ CDM model	12
1.4.4 Basic extensions to Λ CDM	12
1.5 Inflation and primordial fluctuations	13
1.5.1 Motivations	13
1.5.2 Model	13
1.5.3 Expectation	14
1.6 Cosmic microwave background	15
1.6.1 Observations	15
1.6.2 Foregrounds	16
1.6.3 Influence of the cosmological parameters	17
1.7 Inhomogeneous Universe	20
1.7.1 Eulerian view of the gravitational instabilities	21
1.7.2 Lagrangian view and Zel'dovich approximation	21
1.7.3 Structures in the non-linear regime	22
1.8 Conclusion	23
2 Large-scale structures	24
2.1 Statistics	25
2.1.1 Two-point correlation function and power spectrum	25
2.1.2 Angular power spectrum	26
2.1.3 Higher-order statistics	26
2.2 Observables	28
2.2.1 Matter	28
2.2.2 Galaxy clustering	28
2.2.3 Gravitational lensing	30

2.2.4	Thermal Sunyaev-Zel'dovich	33
2.2.5	Galaxy clusters	34
2.3	Halo model to model the power spectrum	36
2.3.1	Basic principle	36
2.3.2	Power spectrum computation	38
2.3.3	Application to matter	38
2.3.4	Halo properties	38
2.3.5	Angular power spectrum	40
2.4	Conclusions	41
II Simulations		42
3 Simulations		43
3.1	Initial conditions	44
3.2	Numerical overview of RAMSES	44
3.2.1	AMR scheme	44
3.2.2	N-body solver	45
3.2.3	Hydrodynamical solver	45
3.2.4	Time step control	46
3.2.5	Refinement strategy	46
3.2.6	Baryonic processes	47
3.3	Description of simulations	47
3.3.1	The Horizon suite of simulations	47
3.3.2	The Magneticum simulations	49
3.3.3	The BAHAMAS simulations	50
3.4	Halo mass function of the simulations	50
III Results and publications		53
4 Thermal Sunyaev-Zel'dovich properties		54
4.1	The HMx halo model	55
4.1.1	Theoretical modelisation	55
4.1.2	Angular power spectrum prediction	58
4.2	Thermal Sunyaev-Zel'dovich effect observations	59
4.2.1	Angular power spectrum on wide fields	59
4.2.2	Cross-correlation with other observables	59
4.2.3	Pressure profile	62
4.3	Comparison with measurement in simulations	63
4.3.1	Methodology	63
4.3.2	Mass cut	64
4.3.3	Power spectrum comparison	64
4.3.4	Angular power spectrum comparison	66
4.3.5	Halo model consequences	67
4.3.6	Halo pressure profiles	69
4.3.7	Difference of the simulations	70
4.4	Measurement of some properties	72
4.4.1	Properties in the simulations	72
4.4.2	Implication for the power spectrum and pressure profile	75
4.5	Beyond standard cosmology	78
4.5.1	Results on the power spectrum	78
4.5.2	Result on the pressure angular power spectrum	78
4.5.3	Discussion	79
4.6	Summary	82
5 The cosmic web and T-web classifier		85
5.1	Overview	86
5.1.1	Towards the non-linear regime	86
5.1.2	Cosmic web	87
5.1.3	Motivation of my work	91

5.2	"A theoretical view of the T-web statistical description of the cosmic web" (article)	91
6	Peak counts statistics in weak lensing	109
6.1	Overview	110
6.2	"UNIONS: The impact of systematic errors on weak-lensing peak counts" (article)	110
	Conclusions and perspectives	130
A	Physical constants & Conventions	133
B	Power spectrum in the Horizon-noAGN and Horizon-Large simulations	134
	Abbreviations	137
	Bibliography	138

List of Figure

Unsourced figures are from the author of the manuscript.

1.1	Evolution of the content of the Universe (Planck Collaboration et al., 2014)	7
1.2	Light cone	8
1.3	Comoving Hubble radius as a function of scale factor (Mishra et al., 2021)	11
1.4	Temperature anisotropies of the CMB observed by Planck	15
1.5	Planck 2018 TT angular power spectrum (Planck Collaboration et al., 2020b)	16
1.6	Contributions to the CMB temperature power spectrum (Baumann, 2018)	17
1.7	Foregrounds and CMB spectrum (Planck Collaboration et al., 2020a)	18
1.8	Influence of Ω_K on the CMB temperature angular power spectrum (Dodelson & Schmidt, 2020)	19
1.9	Influence of $\Omega_b h^2$ and $\Omega_c h^2$ on the CMB temperature angular power spectrum (Dodelson & Schmidt, 2020)	20
1.10	Galaxy distribution by SDSS	22
2.1	Matter power spectrum (Tegmark et al., 2004)	29
2.2	Illustration of the lensing effect (Bisnovatyi-Kogan & Tsupko, 2017)	31
2.3	Shear and convergence effect	31
2.4	From weak to strong lensing (Wambsganss, 1998)	32
2.5	Effect of CMB lensing (Dodelson & Schmidt, 2020)	33
2.6	Thermal Sunyaev-Zel'dovich effect	34
2.7	Thermal Sunyaev-Zel'dovich effect on the CMB spectrum (Carlstrom et al., 2002)	35
2.8	Schematic visualisation of the halo model process (Asgari et al., 2023)	36
2.9	Illustration of the one- and two-halo term	37
2.10	Matter power spectrum predicted by HMx (Mead et al., 2020)	37
2.11	Peak height ν as a function of mass	39
3.1	Pressure and matter projected maps from the Horizon-AGN simulation	49
3.2	Scale factor as a function of the lookback time in the L896_wCDM simulations	50
3.3	Halo mass function in the simulations	51
3.4	Halo mass function in the L896_wCDM simulations	52
4.1	Predicted pressure angular power spectrum as a function of mass and redshift	59
4.2	Planck Collaboration et al. (2016b) tSZ maps and angular power spectrum	60
4.3	3x2pt to 10x2pt constraints in the $S_8 - \Omega_m$ plane (Fang et al., 2024)	61
4.4	FoM in the $\Omega_m - S_8$ plane for 3x2pt to 10x2pt analysis Fang et al. (2024)	62
4.5	Matter power spectrum in the simulations	64
4.6	Pressure power spectrum in the simulations	65
4.7	Matter-pressure power spectrum in the simulations	66
4.8	Pressure angular power spectrum in the simulations	67
4.9	Pressure power spectrum: 1h and 2h contribution, and contribution of within and without $1R_v$ in Horizon-AGN	68
4.10	Contribution of the different mass bins to the total pressure power spectrum in Horizon-AGN	69
4.11	Pressure profiles in the simulations	70
4.12	Variance of the pressure power spectrum estimated from Horizon-Large	71
4.13	Pressure as a function of density in Horizon-AGN	73
4.14	Measurement of Γ in Horizon-AGN	73
4.15	Measurement of α in Horizon-AGN	74
4.16	Measurement of f_{bnd} in Horizon-AGN	75
4.17	Measurement of the concentration in Horizon-AGN	76

4.18	Impact of different parameters in the predicted pressure power spectrum	76
4.19	Impact of different parameters in the predicted pressure profiles	77
4.20	Matter and pressure auto-power spectrum in the L896_wCDM simulations	79
4.21	Pressure angular power spectrum in the L896_wCDM simulations	80
4.22	Density as a function of scale factor for different w	81
4.23	Growth factor as a function of the scale factor for different w	82
4.24	Ratio of the growth factor, pressure, and matter power spectrum in the L896_wCDM simulations	83
5.1	Probabilities of voids, walls, filaments, and knots	93
6.1	Response matrix use for the local shear calibration	111
6.2	Impact of the systematic on the cosmological parameters	112
B.1	Matter power spectrum in the simulations	135
B.2	Pressure power spectrum in the simulations	135
B.3	Matter-pressure power spectrum in the simulations	136

The large-scale structures of the Universe offer unique insights into its fundamental properties. Studying these structures allows us to understand the evolution and composition of the Universe, serving as a probe for both cosmology and astrophysics.

The Universe began as a Gaussian field, evolving into a non-Gaussian field with inhomogeneities that served as the seeds of the large-scale structures we observe today. The evolution of these structures is influenced by various properties of the Universe, such as the amount of dark matter, dark energy, and baryonic matter. Gravitational instabilities have led to the formation of structures such as galaxies, galaxy clusters, and voids that trace the cosmic web of the Universe.

Different observables can be used to study the large-scale structures of the Universe. The cosmic microwave background (CMB) is a snapshot of the Universe at the time of recombination when it became transparent. The CMB provides extensive information about the early Universe, including its geometry and composition. Other tracers, such as the distribution of galaxies, weak lensing, thermal Sunyaev-Zel'dovich (tSZ) effect, and other tracers of matter probe the late-time Universe. These tracers are sensitive to different components: for example, weak lensing is sensitive to total matter, while the tSZ effect is sensitive to baryonic matter. By studying and correlating these tracers, we can extract valuable cosmological information.

Additionally, the large-scale structures are well-modeled using numerical simulations. These simulations are valuable for studying cosmology, astrophysics, and the interplay between the two. They can be used to test various models and provide predictions for future observations, for instance. Modeling baryonic physics is more complex than dark matter physics and remains an active field of research.

Currently, the concordant model of cosmology is the Λ CDM model, which includes dark matter, dark energy, and baryonic matter. The Λ CDM model agrees with a wide range of observations, such as the CMB or large-scale structures. However, some tensions have been observed between different observations, such as discrepancies in the Hubble constant measured from the CMB and from supernovae. These tensions could result from systematic errors in the observations or indicate new physics beyond the Λ CDM model. Addressing these discrepancies is an active field of research. With the increasing precision of data, it is crucial to extract maximum information to test different scenarios. This can be achieved by extracting higher-order information from data, using non-Gaussianities of the field, or employing higher-order statistics. In this thesis, I focus on extracting higher-order information through different methodologies.

The first part of this thesis provides general context. In Chapter 1, I present a brief history of the Universe and introduce equations of evolution of a homogeneous and inhomogeneous Universe. The Λ CDM model, inflation, and CMB are also discussed. In Chapter 2, I introduce the large-scale structures of the Universe, explaining the relevant statistics and covering a few observables such as matter, galaxy, lensing, tSZ, and clusters. I also present the halo model framework.

The second part focuses solely on numerical simulations (Chapter 3). I describe the process of generating initial conditions for a simulation, the numerical methods used to evolve the simulation, and the halo properties. I conclude this part with a description of the different simulations employed in this thesis.

The third part details the results I have obtained. Chapter 4 focuses on the tSZ properties. I begin by describing the halo model that allows predictions of the power spectrum, followed by a discussion

of the state of the art on tSZ observations. I then present results comparing the tSZ properties (power spectrum, angular power spectrum, profile) measured in simulations with those predicted by the HM \times halo model. This comparison is extended to some intrinsic properties of the model. Additionally, I present results on the tSZ effect within the context of the w CDM model. In Chapter 5, I present a theoretical prediction on the abundances of the different environments within the cosmic web. Finally, in Chapter 6, I present the impact of systematics on weak lensing peak count statistics for cosmological constraints.

Through these different results, my work has focused on the complexities of large-scale structures. The tSZ effect is mainly sensitive to high-mass halos, thus probing the densest regions of the Universe. Its sensitivity to baryonic matter makes it a robust probe of baryonic physics. The cosmic web, by the identification of different environments, provides a way of extracting higher-order information. Finally, peak count statistic is a higher-order statistic that probes the non-Gaussianities of a field.

Part I
Context

1.1	Concordant model and history of the Universe	5
1.2	Introduction to general relativity	6
1.2.1	Equivalence principle	6
1.2.2	Curvature of space-time	6
1.3	Homogeneous Universe	8
1.3.1	Friedmann-Lemaître-Robertson-Walker metric and Hubble parameter	8
1.3.2	Friedmann equations	9
1.3.3	Hubble radius and epoch of the Universe	10
1.4	Λ CDM model	12
1.4.1	Cosmological constant	12
1.4.2	Cold dark matter	12
1.4.3	Limitations of the Λ CDM model	12
1.4.4	Basic extensions to Λ CDM	12
1.5	Inflation and primordial fluctuations	13
1.5.1	Motivations	13
1.5.2	Model	13
1.5.3	Expectation	14
1.6	Cosmic microwave background	15
1.6.1	Observations	15
1.6.2	Foregrounds	16
1.6.3	Influence of the cosmological parameters	17
1.7	Inhomogeneous Universe	20
1.7.1	Eulerian view of the gravitational instabilities	21
1.7.2	Lagrangian view and Zel'dovich approximation	21
1.7.3	Structures in the non-linear regime	22
1.8	Conclusion	23

To describe the Universe, its components, and its evolution, we need a comprehensive model. The standard model of cosmology is based on the theory of general relativity and the cosmological principle. The most widely adopted model is the Lambda Cold Dark Matter (Λ CDM) model. In this chapter, I will introduce the primary concepts of cosmology and general relativity that form the foundation of this model and describe the evolution of the Universe. I will begin in Sect. 1.1 with a brief history of the Universe, from the Big Bang to the present day. In Sect. 1.2, I will introduce the concepts of general relativity necessary for describing the Universe. I will then provide more detailed explanations of various epochs of the Universe: the homogeneous Universe in Sect. 1.3, the Λ CDM model in Sect. 1.4, inflation in Sect. 1.5, cosmic microwave background in Sect. 1.6, and the inhomogeneous Universe in Sect. 1.7. This chapter was written with the help of [Peter & Uzan \(2013\)](#); [Bernardeau \(2007\)](#); [Dodelson & Schmidt \(2020\)](#); [Codis-Decara \(2015\)](#).

1.1 Concordant model and history of the Universe

Concordant model The concordance model of cosmology is known as the Λ CDM model, which describes a Universe composed of a cosmological constant (Λ), cold dark matter (CDM), and baryonic matter. According to this model, the Universe began 13.8 billion years ago with the Big Bang. Since that moment, the Universe has been expanding. This model will be further discussed in Sect. 1.4.

The Planck era Up to 10^{-43} seconds after the Big Bang, the Universe existed in an era known as "the Planck era". This was the very first moment of our Universe when it was extremely hot and dense. The laws of physics as we know them cannot be applied to this state. During this era, the four fundamental forces – electromagnetic force, weak nuclear force, strong nuclear force, and gravity – had similar intensities, and it was impossible to describe them separately.

Grand Unified Theory From 10^{-43} seconds to 10^{-36} seconds, the Universe entered the "Grand Unified Theory" (GUT) era. During this period, the Universe continued to cool, and gravity "froze" out, becoming a force of distinct intensity and behaving independently of quantum effects. The other three fundamental forces remained unified.

Electroweak era & inflation The subsequent era, from 10^{-36} seconds to 10^{-12} seconds, is known as the "electroweak era". As the Universe continued to cool, it was now the weak nuclear force that "froze" out, becoming distinct. It was only at the end of this era that the electromagnetic and weak nuclear forces also separated. During this era, from 10^{-36} to 10^{-32} seconds, the Universe undergoes a period of exponential expansion known as "inflation". During inflation, the Universe was dominated by a scalar field called the inflaton. This era is responsible for the homogeneity and isotropy of the Universe and the origin of the primordial fluctuations that will give rise to the large-scale structures (LSS) we observe today. More details about inflation will be provided in Sect. 1.5.

Particle era & nucleosynthesis era The electroweak era is followed by the "particle era" which lasted until 10^{-3} seconds after the Big Bang and allowed the formation of the first particles. During this period, the Universe comprises quarks, electrons, neutrinos, photons, and their antiparticles. However, the temperature was still too high for the formation of atoms. Between 3 and 20 minutes after the Big Bang, the Universe entered the "nucleosynthesis era". By this time, the Universe had cooled enough to allow the formation of light elements. Protons and neutrons combine to create simple nuclei, such as hydrogen, helium, deuterium, and lithium. At this time, the baryonic Universe's composition was approximately 75% of hydrogen and 25% of helium, with traces of deuterium and lithium.

Nuclei era & atomic era The Universe then entered the "nuclei era", a period cool enough to allow the formation of atomic nuclei. Approximately 50,000 years after the Big Bang, atoms began to form, marking the "atomic era". During this epoch, matter begins to dominate the content of the Universe.

Recombination era & cosmic microwave background 380,000 years after the Big Bang, the Universe entered the "recombination era". Nuclei were able to capture free electrons to form neutral atoms. Consequently, the Universe is composed of neutral atoms, photons, and neutrinos. Before this period, electrons rapidly absorbed photons, but photons could now travel freely through the Universe. The Universe is now transparent, and the photons we observe today originate from this moment, producing the first observable relic light of the Big Bang known as the cosmic microwave background (CMB). The CMB manifests as a black-body radiation of 2.725 K in today's Universe due to the redshifting of its

wavelength by the expansion of the Universe. It is a powerful tool for studying the Universe and will be discussed further in Sect. 1.6.

First stars & galaxies A possible scenario is that the first stars began to form approximately 180 million years after the Big Bang. These early stars, composed of light elements, initiated the process of nuclear fusion, creating heavier elements than hydrogen or helium (so-called "metals") in their cores. Upon exhausting their nuclear fuel, these stars ejected these heavier elements into the interstellar medium, enriching it. This stage is followed by the creation of the first galaxies around 400 million years after the Big Bang. The initial metal-free stars (Population III stars) often ended their lives as black holes, and the first galaxies formed around their remnants. These initial galaxies were small and irregular. Over time, they merged and grew larger. Gravitational interactions will allow the formation of the complex structures we observe in the Universe today.

Dark energy era Finally, around 6 billion years after the Big Bang, the Universe entered the "dark energy era". During this period, an unknown fluid or the cosmological constant Λ began to dominate the dynamics of the Universe. This dark energy provides the repulsive force that accelerates the expansion of the Universe.

As discussed and illustrated in Fig. 1.1, the content of the Universe has evolved over time. In the early Universe, radiation and neutrinos dominated, whereas they are almost negligible today. Currently, the Universe is composed of $\sim 70\%$ of dark energy, $\sim 25\%$ of dark matter, and $\sim 5\%$ of baryonic matter. Studying the Universe's evolution gives us insights into its composition, structure, and history. As we will see in the following sections, the concordant model of cosmology, the Λ CDM model, provides a comprehensive framework for understanding the Universe but also has some limitations. For example, we will discuss the possibility that dark energy may not be a simple cosmological constant but could have more complex properties.

1.2 Introduction to general relativity

The theory of general relativity describes gravitation as the curvature of space-time, allowing us to describe astrophysical objects' motion. Gravitation, which is a long-range force, is the dominant attractive force at cosmological scales. It is essential to consider it to accurately describe both the expansion of the Universe and the growth of structures at these scales.

1.2.1 Equivalence principle

The equivalence principle is a fundamental law of physics stating that, locally, the effects of a gravitational field are indistinguishable from those of an accelerated reference frame. This principle, which is the foundation of general relativity, applies exclusively to gravity. It implies we can always find a reference frame where gravity is absent. The equivalence principle is based on three conditions:

- Weak equivalence principle: the trajectory of a body is independent of its internal structure and composition,
- Local position invariance: all non-gravitational experiments are independent of the position in space and time,
- Local Lorentz invariance: all non-gravitational experiments are independent of the motion of the laboratory as long as it is in free fall.

1.2.2 Curvature of space-time

Gravitation is described as the curvature of space-time under the influence of massive objects. The geometry of this space-time is characterized by a metric. The motion of a free-falling object follows the geodesic of this metric. Space-time is modeled as a four-dimensional manifold, where space locally resembles a three-dimensional Euclidean space but not necessarily globally. The general metric, representing the line element ds between two events, is given by:

$$ds^2 = g_{\mu\nu} dx^\mu dx^\nu, \quad (1.1)$$

also called the proper time, where $g_{\mu\nu}$ is the metric tensor, which is symmetric ($g_{\mu\nu} = g_{\nu\mu}$). The Greek indices can take the value 0, 1, 2, 3, where 0 represents the time coordinate, while 1, 2, 3 are the spatial

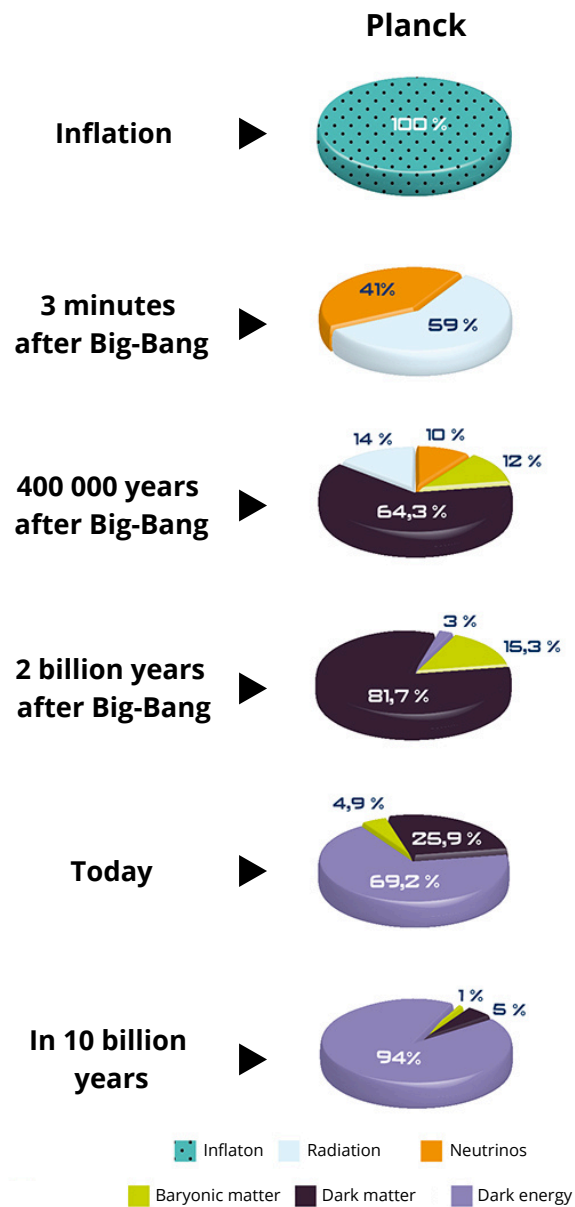


Figure 1.1: Figure adapted from public.planck.fr, which uses the results of [Planck Collaboration et al. \(2014\)](#). It represents the evolution of the content of the Universe (inflaton, radiation, neutrinos, baryonic matter, dark matter, and dark energy) at different periods.

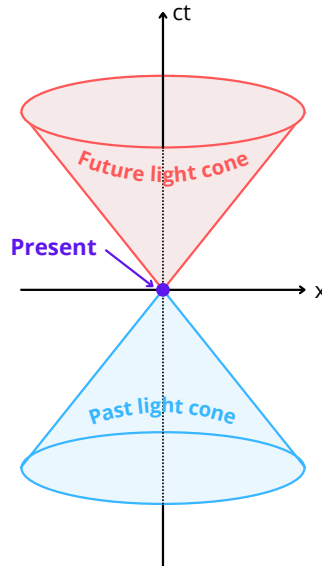


Figure 1.2: Representation of the light cone in the Minkowski space-time. The future light cone is in red, while the past light cone is in blue. Present time is represented by the violet dot.

coordinates. This metric is expressed using the Einstein summation convention, where repeated indices are summed over.

To describe special relativity, we use the Minkowski metric. In this framework, the proper-time can be written as: $ds^2 = -c^2dt^2 + dx^2 + dy^2 + dz^2$, as $g_{00} = -1$, $g_{11} = g_{22} = g_{33} = 1$ and $g_{\mu\nu} = 0$ for $\mu \neq \nu$.

In this metric, ds^2 can be negative, positive, or zero. A positive ds^2 indicates that the two events are spatially disconnected. The set of points $ds^2 = 0$ is called the light cone and is illustrated in Fig. 1.2. It represents all the points that can receive a light beam emitted today (referred to as the future light cone, shown by the red cone) or that received one from the past (past light cone, shown by the blue cone). Since no object in the Universe can travel faster than the speed of light, the light cone defines the boundary of the causally connected region of the Universe: events outside the light cone are not causally connected to the event at the origin of the cone.

1.3 Homogeneous Universe

In this section, I will describe and present the tools necessary for understanding the evolution of a homogeneous and isotropic Universe.

Cosmological principle The cosmological principle is the fundamental assumption that underlies our understanding of the Universe’s evolution. It postulates that the Universe is homogeneous and isotropic on sufficiently large scales (cosmological scales). This principle is supported by galaxy surveys, which show a uniform and isotropic distribution of galaxies on these scales, as well as measurements of the CMB. This principle implies that we do not occupy a special or unique location in the Universe.

1.3.1 Friedmann-Lemaître-Robertson-Walker metric and Hubble parameter

The Friedmann-Lemaître-Robertson-Walker (FLRW) metric describes the Universe according to the cosmological principle. It is given by:

$$ds^2 = -c^2dt^2 + a^2(t) \left(\frac{dr^2}{1 - Kr^2} + r^2d\Omega^2 \right), \quad (1.2)$$

where $a(t)$ is the scale factor, t is the cosmic time, c the speed of light, and K is a constant (equal to 0, -1 or $+1$) that depends on the local curvature of space. If $K = 0$, the spatial part of the metric is flat. This metric is expressed in spherical coordinates (r, θ, ϕ) , where $d\Omega^2 = d\theta^2 + \sin^2\theta d\phi^2$.

The Hubble parameter characterizes the cosmic expansion rate:

$$H(t) = \frac{\dot{a}(t)}{a(t)}, \quad (1.3)$$

where the dot denotes a derivative with respect to cosmic time t . The Hubble law describes the motion of galaxies, stating that the further a galaxy is, the faster it moves away from us. This law is a consequence of the expansion of the Universe and leads to observations that galaxies have redshifted spectra. We can thus define the redshift z as:

$$1 + z \equiv \frac{\lambda_{\text{obs}}}{\lambda_{\text{em}}} = \frac{a(t_{\text{obs}})}{a(t_{\text{em}})}, \quad (1.4)$$

where λ_{obs} and λ_{em} are the observed and emitted wavelengths of light, respectively, and t_{obs} and t_{em} are the time of observation and emission, respectively. At small distances, the Hubble law gives $z = H_0 d/c$, where d is the distance to the galaxy, and H_0 is the Hubble constant today. The scale factor is normalised to $a(t_{\text{obs}}) \equiv a_0 = 1$ today, thus $a = 1/(1+z)$.

1.3.2 Friedmann equations

The Friedmann equations are the Einstein equations applied to the FLRW metric, describing the evolution of the Universe. They are given by:

$$\frac{H^2}{c^2} = \frac{\kappa}{3}\rho - \frac{K}{a^2} + \frac{\Lambda}{3}, \quad (1.5)$$

$$\frac{\ddot{a}}{a} = -\frac{\kappa}{6}(\rho + 3P) + \frac{\Lambda}{3}, \quad (1.6)$$

where $\kappa = 8\pi G/c^4$, G is the gravitational constant, ρ is the energy density, P is the pressure, and Λ is the cosmological constant. The continuity equation is given by:

$$\dot{\rho} + 3H(\rho + P) = 0. \quad (1.7)$$

Equation of state

To solve these equations, we need to introduce an equation of state, $P = w\rho$. Pressureless matter is characterized by $w = 0$, while radiation is characterized by $w = 1/3$. For the cosmological constant, we have $w = -1$. The continuity equation defined in Eq. (1.7) can be integrated to give, if w is constant in time, $\rho \propto a^{-3(1+w)}$.

Cosmological parameters

To rewrite the first Friedmann equation and its components, it is convenient to define the critical density:

$$\rho_c = \frac{3H^2}{8\pi G}. \quad (1.8)$$

Today, the critical density is $1.88 \times 10^{-29} h^2 \text{ g} \cdot \text{cm}^{-3}$, using $H_0 = 100 h \text{ km} \cdot \text{s}^{-1}$. This corresponds to the density of a flat Universe ($K = 0$). We can then define the density parameters for matter, radiation, and cosmological constant as $\Omega_i = \frac{\rho_i}{\rho_c}$, in particular, we can obtain $\Omega_\Lambda = \frac{\Lambda}{3H^2}$, and the cosmological parameter for the curvature as $\Omega_K = -\frac{K}{H^2 a^2}$. The first Friedmann equation becomes:

$$\Omega_m + \Omega_r + \Omega_\Lambda + \Omega_K = 1. \quad (1.9)$$

With this notation, the continuity equation becomes:

$$E(z) \equiv \frac{H(z)}{H_0} = \sqrt{\Omega_{m0}(1+z)^3 + \Omega_{r0}(1+z)^4 + \Omega_{\Lambda0} + \Omega_{K0}(1+z)^2}, \quad (1.10)$$

where indices 0 denote quantities today.

The total content of the Universe, denoted Ω , is the sum of:

- Matter: Ω_m , itself the sum of baryonic matter Ω_b and cold dark matter Ω_c ,
- Radiation: Ω_r , itself the sum of photons Ω_γ and neutrinos Ω_ν ,
- Dark energy: Ω_Λ .

Some solutions

For simple cases, solving the first Friedmann equation to obtain the evolution of $a(t)$ is possible. Here are a few examples of solutions.

Example 1 When $K = 0$ and $w \neq -1$, the conservation equation reads $\rho \propto a^{-3(1+w)}$. The first Friedmann equation thus gives $a(t) \propto t^{\frac{2}{3(1+w)}}$.

Example 2 When $K = 0$ and $w = -1$, it implies that the density is dominated by a cosmological constant. The first Friedmann equation leads to $a(t) \propto e^{Ht}$. This case is called a de Sitter space and is characterized by an accelerated expansion. It is representative of the inflation era and the dark energy era if it is not a constant cosmological constant.

Example 3 When $K = 0$, $w = 0$, and $\Lambda \neq 0$, we are in an Euclidean space dominated by a pressureless fluid and cosmological constant. The first Friedmann equation leads to $a(t) \propto \left(\frac{1}{\Omega_{\Lambda 0}} - 1\right)^{1/3} \sinh^{2/3}\left(\frac{3\alpha t}{2}\right)$ with $\alpha = H_0\sqrt{\Omega_{\Lambda 0}}$. This solution is relevant for the late-time evolution of a flat Λ CDM Universe.

Example 4 When $K = 0$, $w = 0$, and $\Lambda = 0$, the first Friedmann equation leads to $a(t) \propto t^{2/3}$. This case is referred to as the Einstein-de Sitter Universe.

Expansion and contraction

Studying the signs of the expansion rate in different configurations is also possible. For a Universe with non-flat spatial sections, matter, radiation, and a cosmological constant, we have different possible scenarios:

- $K = 0$ or $K = -1$ and $\Lambda = 0$: the Universe will expand forever since the Big Bang,
- $K = 1$ and $\Lambda = 0$: the expansion will be followed by a contracting phase, leading to a Big Crunch,
- $K = 1$ and Universe has spherical spatial sections: the expansion lasts forever but was preceded by a contracting phase, the Universe thus undergoes a bounce,
- For other scenarios, it is possible to have a series of contraction and expansion phases, leading to an oscillating Universe.

Current constraints on the cosmological parameters are consistent with a flat Universe. For example, [Planck Collaboration et al. \(2020c\)](#) reports $\Omega_K = -0.0007 \pm 0.0037$ at the 95% confidence level, using Planck temperature, polarisation, lensing, and baryonic acoustic oscillation data.

1.3.3 Hubble radius and epoch of the Universe

Defining sub-Hubble and super-Hubble modes is crucial to describe the behaviour of perturbations in the Universe. A mode with a wavelength larger than the Hubble radius is a super-Hubble mode. These modes remain "frozen" until they re-enter the Hubble radius, at which point they evolve again. The sub-Hubble modes have wavelengths shorter than the Hubble radius and evolve through various physical processes. These are the modes that can collapse to form structures in the Universe. As the Hubble radius changes over time, modes can cross it, transitioning between sub-Hubble and super-Hubble states.

The evolution of the comoving Hubble radius as a function of the scale factor is presented in Fig. 1.3 in green. The blue dashed line illustrates a specific mode that exits and re-enters the Hubble radius at the associated scale factor a_k and a_p . Before a_k and after a_p , this mode is sub-Hubble and thus influenced by the physical process. We observe that the evolution of the scale factor varies across the different epochs. We can distinguish several epochs delimited by the brown dashed lines:

- Inflation era: ends at a_e , the Hubble radius varies as $(aH)^{-1} \propto a^{-1}$,
- Reheating: extends from a_e to a_{re} , for a duration of N_{re} ,
- Radiation dominated era: extends from a_{re} to a_{eq} , for a duration of N_{RD} , the Hubble radius varies as $(aH)^{-1} \propto a$. Photons were the dominant component of the Universe during this era,
- Matter dominated era: begins at a_{eq} , marking the scale factor at the equality between radiation and matter. Radiation and matter have the same energy density at this point, and matter becomes dominant after,
- Λ -dominated era: begins at a_{de} , not shown here. If the Universe is flat, the scale factor grows exponentially $a(t) \propto e^{Ht}$.

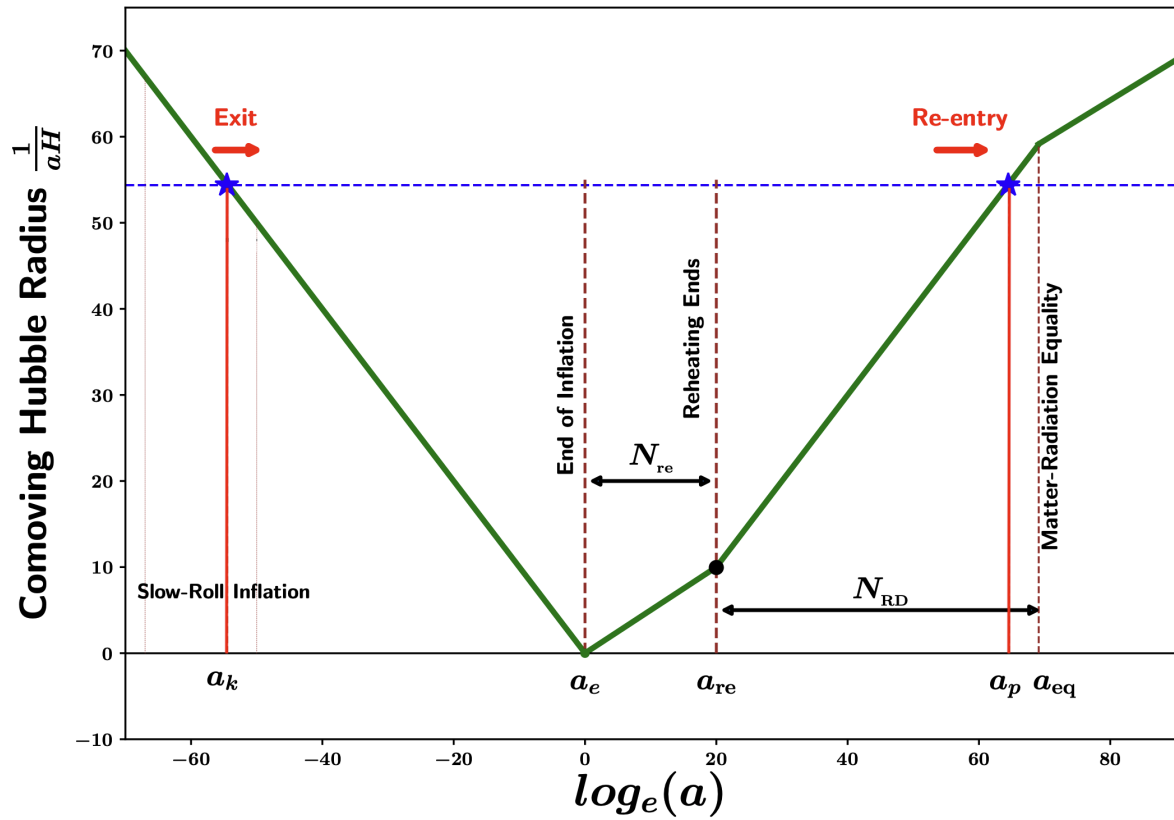


Figure 1.3: Figure adapted from [Mishra et al. \(2021\)](#). Evolution of the comoving Hubble radius as a function of the scale factor in green. The blue dashed line shows a given mode that exits and re-enters the Hubble radius, with the associated scale factor a_k and a_p . The brown dashed line shows the scale factor that delimits the different epochs: inflation, reheating, radiation domination, and matter domination. N_{re} and N_{RD} are the reheating and radiation domination epoch durations, respectively.

1.4 Λ CDM model

The Λ CDM model is the standard model of cosmology, describing the Universe as composed of a cosmological constant Λ , cold dark matter (CDM), and baryonic matter. This model assumes that the Universe is flat and is characterized by a set of cosmological parameters, including the baryon density Ω_b , the cold dark matter density Ω_c , and the Hubble constant H_0 discussed previously. Additionally, the model includes other parameters: τ , the optical depth of reionization; A_s , the amplitude of the primordial power spectrum; and n_s , the primordial spectral index. In this model, all the other parameters are fixed, and $\Omega_K \equiv 1 - \Omega_m - \Omega_\Lambda$.

1.4.1 Cosmological constant

The accelerated expansion of the Universe can be driven by any form of matter having an equation of state with $w < -1/3$. One still open question is understanding the origin of this accelerated expansion of the Universe we observe today. It can come from a cosmological constant, which is a particular form of dark energy having $w = -1$, or by any other form of dark energy. The Λ CDM model is the one where the accelerated expansion is described by a cosmological constant having $w = -1$, thus constant in time.

1.4.2 Cold dark matter

Evidence for dark matter arises from the study of galaxy rotation curves and the dynamics of galaxy clusters. In spiral galaxies, stars follow circular orbits, and their rotation velocity as a function of distance from the centre does not match the expected curve based on visible matter alone. This discrepancy can be explained by the presence of dark matter, which does not emit light and interacts only through gravity. A similar effect is observed in galaxy clusters, where the gravitational effect cannot be accounted for only by visible matter.

The most popular dark matter model is the Cold Dark Matter (CDM) model, which assumes that dark matter particles are massive and non-relativistic at the time of decoupling from the plasma. These particles are collisionless, cold, and interact with matter only through gravity. Because these particles were not coupled to radiation, they began to develop potential wells before the matter-radiation equality, allowing the formation of the first instabilities that attracted baryonic matter. The CDM model is the most successful model for describing the large-scale structures of the Universe.

1.4.3 Limitations of the Λ CDM model

The Λ CDM model aligns very well with observations on large scales. However, discrepancies emerge on smaller scales, where baryonic physics becomes significant. One such discrepancy is the "missing satellite problem", where the model predicts more satellite galaxies around the Milky Way than are observed. Another issue is the "cusp-core problem", where the model predicts a cusp in the density profile of dark matter halos while observations indicate a core (a flat density profile). Additionally, different observational methods yield to inconsistent values for cosmological parameters. One such example is the "Hubble tension", where the Hubble constant derived from the CMB does not match the value obtained from local measurements using supernovae. These discrepancies could come from astrophysical effects but can also suggest that the Λ CDM model might be incomplete and that extensions to the Λ CDM model may be needed.

1.4.4 Basic extensions to Λ CDM

Dark energy

Many models of dark energy exist, but here we focus on a common parametrization of a non-constant equation of state:

$$w(z) = w_0 + w_a(1 - a), \quad (1.11)$$

where w_0 and w_a are two constants, and a is the scale factor. This parametrization allows us to describe a large range of dark energy models and is used in the w CDM model, where the cosmological constant is replaced by a field having $w \neq -1$. It is possible to constrain these parameters using observations of the expansion rate of the Universe. For example, the **Dark Energy Survey**¹ (DES) is a ground-based telescope designed to probe the origin of the accelerating expansion of the Universe, by studying the nature of dark energy. DES gives constraints on the $w_0 - w_a$ parameters in [Abbott et al. \(2023\)](#) for example.

¹<https://www.darkenergysurvey.org/>

The cosmological constant mentioned previously is a particular case of this parametrization with $w_0 = -1$ and $w_a = 0$. Another particular case is the quintessence model, where a scalar field describes the dark energy. For quintessence dark energy, we also have $w_a = 0$, thus a constant equation of state in time.

Dark energy affects the growth of the matter density contrast through the modification of the evolution of the Universe and hence $H(a)$. These modifications will depend on the actual dark energy present in the Universe. It will modify the normalisation of the matter power spectrum and the time when modes enter the non-linear regime.

Other form of dark matter

Another extension to the Λ CDM model arises from the nature of dark matter. Some models propose alternative forms of dark matter. For example, Warm Dark Matter (WDM) is composed of relativistic particles at the time of decoupling. This model suppresses the formation of small-scale structures due to the initial thermal velocity of its particles. WDM can address some of the issues of the CDM model, such as the "missing satellite problem" and the "cusp-core problem". However, WDM does not explain the large-scale structures of the Universe as effectively as CDM.

Other alternatives to the Λ CDM model exist, such as modified gravity, but these are not discussed here. Despite its limitations, the Λ CDM model remains the most successful model for describing the Universe on large scales. The investigation of these limitations and the exploration of new models is an active field of research in cosmology.

1.5 Inflation and primordial fluctuations

At the very beginning of the Universe, it underwent a phase of accelerated expansion known as inflation. This epoch was crucial in establishing the homogeneity and isotropy observed in the Universe today. It seeded the primordial fluctuations that will evolve into the large-scale structures we observe. In this section, I will introduce the motivations, the model, and its consequences.

1.5.1 Motivations

Before inflation was introduced, the standard Big Bang model faced several unresolved issues. Introducing a period of accelerated expansion helped to address these challenges.

First, the absence of magnetic monopoles in the Universe posed a problem. While their existence was theoretically predicted, they are not observed. Inflation provided a mechanism for their dilution.

Secondly, the horizon problem arises from the observation that distant regions exhibit remarkable homogeneity and isotropy despite appearing causally disconnected. Inflation allows for the entire observable Universe to be in causal contact before inflation because two points in causal contact at the beginning of inflation can be separated by a distance larger than the Hubble radius at the end of inflation (because the comoving Hubble radius decreases in time), so they can seem causally disconnected today.

Lastly, the flatness problem concerned the flatness of the Universe, which required that the Universe was even more flat in the past. One way of explaining why the Universe is so flat today is to have a long enough inflation.

The duration of inflation is quantified by the number of e-folds, defined as:

$$N \equiv \log \left(\frac{a_f}{a_i} \right), \quad (1.12)$$

which measures the growth in the scale factor. a_f and a_i are the scale factors at the end (final) and at the beginning (initial) of inflation, respectively. To address the flatness and horizon problems, inflation typically requires $N \gtrsim 60$, ensuring a sufficiently prolonged period of accelerated expansion.

1.5.2 Model

The inflation model relies on one or several scalar fields known as inflaton. It relates the existence of a phase during which the Universe is close to a de Sitter space-time, and a slowly rolling scalar field rules the expansion.

Scalar field

The simplest solution is a single scalar field φ , which evolves in a potential $V(\varphi)$. The energy density and pressure of the inflation are defined as:

$$\rho_\varphi = \frac{\dot{\varphi}^2}{2} + V(\varphi), \quad P_\varphi = \frac{\dot{\varphi}^2}{2} - V(\varphi), \quad (1.13)$$

in conformal time. The Friedmann and Klein-Gordon equations during the inflation phase become:

$$H^2 = \frac{8\pi G}{3} \left(\frac{1}{2}\dot{\varphi}^2 + V \right) - \frac{K}{a^2}, \quad (1.14)$$

$$\frac{\ddot{a}}{a} = \frac{8\pi G}{3} (V - \dot{\varphi}^2), \quad (1.15)$$

$$\ddot{\varphi} + 3H\dot{\varphi} + V'(\varphi) = 0, \quad (1.16)$$

where the prime denotes a derivative with respect to φ . The Universe undergoes an accelerated expansion when $\dot{\varphi}^2 < V(\varphi)$. This expansion is quasi-exponential if the scalar field enters a regime of slow-roll. This concept is defined by certain conditions: we assume $K = 0$ due to its negligible influence during inflation and that the scalar field satisfies conditions of slow-roll: $\dot{\varphi}^2 \ll V$ and $\ddot{\varphi} \ll 3H\dot{\varphi}$. The slow-roll parameters are defined as:

$$\epsilon = -\dot{H}/H^2 \quad \delta = \epsilon - \frac{\dot{\epsilon}}{2H\epsilon} \quad \xi = \frac{\dot{\epsilon} - \dot{\delta}}{H}. \quad (1.17)$$

For inflation to occur, it is necessary to have $\ddot{a} > 0$, implying $w < -1/3$ and $\epsilon < 1$.

End of inflation

The end of inflation can be identified with the end of the slow-roll regime, which is characterized by $\max(\epsilon, \delta) \approx 1$. At this point, classical inhomogeneities have been exponentially diminished. If the inflaton potential has a minimum, the scalar field will oscillate around this minimum immediately after inflation. These oscillations can be considered as individual scalar particles. Due to Hubble expansion, these oscillations are gradually damped, leading to the decay of the scalar field into a multitude of particles. A significant portion of the matter constituting the Universe today traces its origins back to this decay process.

Quantum fluctuations

The emergence of LSS in the Universe is related to quantum fluctuations of the inflaton field. These fluctuations are amplified during inflation and redshifted to macroscopic scales. Their wavelengths grow exponentially until they exceed the Hubble radius, at which point they become frozen and converge towards an almost scale-invariant power spectrum. In reality, the Hubble parameter H varies over time during inflation, leading to a red-tilted power spectrum if the inflation occurs in a slow-roll regime. The spectral index n_s is defined as:

$$n_s - 1 \equiv \frac{d \log P(k)}{d \log k}, \quad (1.18)$$

where $P(k)$ is the power spectrum of the primordial fluctuations. A scale-invariant power spectrum corresponds to $n_s = 1$, while a red-tilted spectrum indicates $n_s < 1$, and a blue-tilted one indicates $n_s > 1$. Observational constraints on this parameter give $n_s \approx 0.96$ (Planck Collaboration et al., 2020b).

The perturbations of the inflation scalar field induce metric perturbations, which in turn lead to density perturbations. The inflation will also give rise to primordial gravitational waves (GW). The amplitude of the density perturbations and the GW are defined as:

$$\frac{\delta\rho}{\rho} \sim \frac{H^2}{2\pi\dot{\varphi}}, \quad h \sim \frac{H}{2\pi M_{\text{P}}}, \quad (1.19)$$

where M_{P} is the Planck mass.

1.5.3 Expectation

Inflation addresses various fundamental aspects of the Universe. First, it elucidates the observed homogeneity and isotropy by erasing classical inhomogeneities, rendering the observable Universe describable within a Friedmann-Lemaître model. Moreover, it accounts for the current flatness of the Universe by exponentially suppressing the curvature. It also elucidates the generation of scalar perturbations, thus

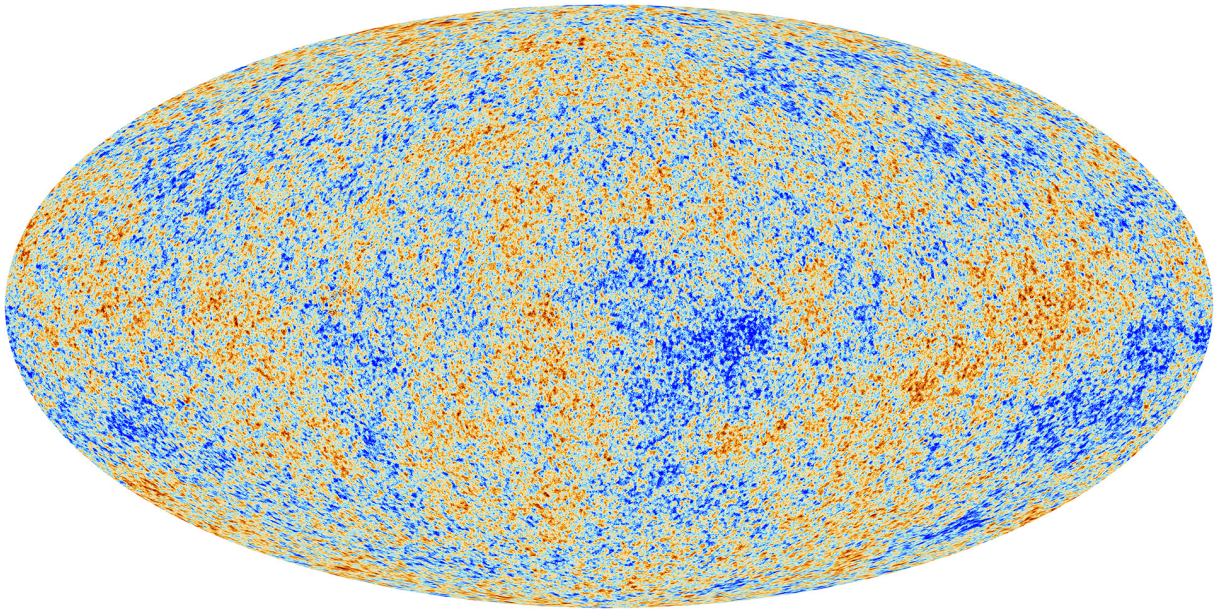


Figure 1.4: Temperature anisotropies of the CMB as observed by the Planck satellite.

explaining the origin of the density perturbations that manifest as LSS today. The inflation predicts an almost scale-invariant spectrum of primordial perturbations and presence of GW.

Observations of the CMB support these predictions, revealing a Gaussian distribution and observing adiabatic, nearly scale-invariant primordial fluctuations. Although primordial GW has yet to be detected, their confirmation would be a proof. Moreover, evidence of a non-flat Universe would challenge the inflation model, as it predicts a flat Universe.

1.6 Cosmic microwave background

The cosmic microwave background (CMB) is a relic radiation from the Big Bang, emitted when the Universe was 380,000 years old. It is a powerful tool for studying the Universe, providing insights into its composition, structure, and history. In this section, I will introduce CMB observations and discuss its different properties.

1.6.1 Observations

The CMB was first observed accidentally in 1965 by Penzias and Wilson ([Penzias & Wilson, 1965](#)), who discovered a background radiation at ~ 3.5 K, now estimated at 2.725 K. This radiation is isotropic, with temperature fluctuations of less than 10^{-5} K. Over the following years, several satellites have observed the CMB. The first satellite to observe the CMB was *Cosmic Background Explorer*² (COBE), launched in 1989. COBE was followed by different satellites, including the *Wilkinson Microwave Anisotropy Probe*³ (WMAP), launched in 2001, and *Planck*⁴, launched in 2009. The Planck satellite provided the most comprehensive measurements of the CMB to date. The CMB temperature map obtained by Planck is shown in Fig. 1.4. In recent years, ground-based experiments have also been conducted to observe the CMB, such as the *South Pole Telescope*⁵ (SPT) and the *Atacama Cosmology Telescope*⁶ (ACT). These experiments are typically conducted at high altitudes or very dry sites to avoid atmospheric fluctuations. Ground-based experiments are limited to observing half of the sky, while satellites allow for a full-sky survey.

CMB measurements are a powerful tool for studying cosmology due to the Gaussian nature of its statistics. From the observations, it is possible to compute the temperature, denoted as T , and the polarization, denoted as E , of the CMB. To constrain cosmological parameters with these observations, we use the angular power spectrum of the temperature, polarization, and their correlation. The angular power spectrum traces the correlation between the different points on the map, providing critical insights

²<https://lambda.gsfc.nasa.gov/product/cobe/>

³<https://map.gsfc.nasa.gov/>

⁴<https://www.cosmos.esa.int/web/planck>

⁵<https://pole.uchicago.edu/public/Home.html>

⁶<https://act.princeton.edu/>

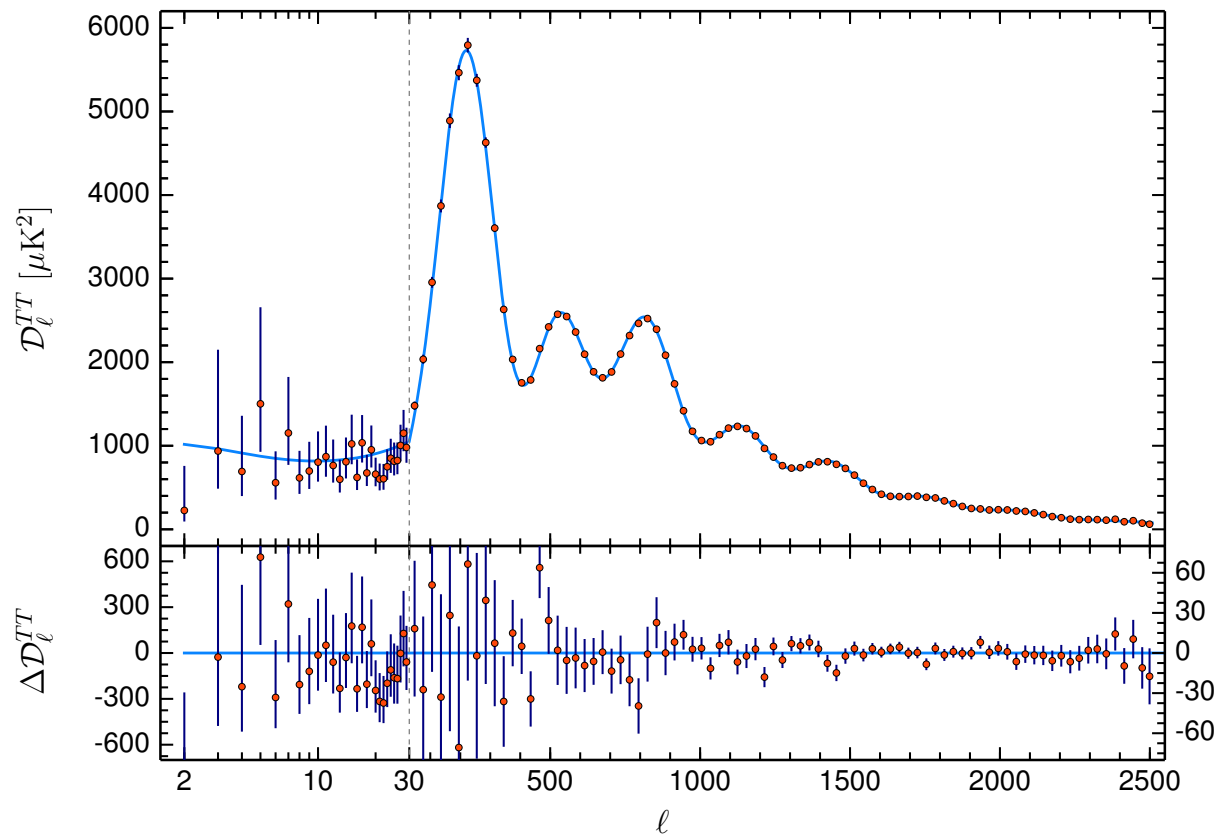


Figure 1.5: Planck 2018 temperature power spectrum from Planck Collaboration et al. (2020b). The red dots are the measurements with their 1σ errors. The blue line is the best fit of a Λ CDM model.

into the Universe’s properties. More details about the generic computation of an angular power spectrum will be given in Sect. 2.1.2. Fig. 1.5 shows the angular power spectrum of the temperature measured by Planck Collaboration et al. (2020b) with the red dots and the best fit in blue. The best fit is realized using a Λ CDM model and is based on the TT, TE, EE, and lensing measurements. The cosmological constraints obtained by Planck Collaboration et al. (2020b) remain the most comprehensive to date and often serve as a reference.

The shape and amplitude of the CMB power spectrum will depend on the cosmological parameters, and it contains different contributions that will be affected differently. The contribution of three different components are shown in Fig. 1.6 and are:

- Sachs-Wolfe (SW) (Sachs & Wolfe, 1967): due to the photons that oscillate in potential wells,
- Integrated Sachs-Wolfe (ISW): due to the change in the potential wells caused by dark energy,
- Doppler: due to the motion of the baryons with respect to the photons.

1.6.2 Foregrounds

The CMB is not the only source of radiation in the Universe. Several foregrounds can contaminate the CMB signal and must be removed to obtain a clean CMB map. The most important foregrounds are:

- Thermal Sunyaev-Zel’dovich effect (tSZ) (Sunyaev & Zeldovich, 1970): extragalactic foreground due to the scattering of CMB photons by the hot electrons in galaxy clusters, the tSZ effect will be further discussed in Sect. 2.2.4 and Chapter 4,
- Kinetic Sunyaev-Zel’dovich effect (kSZ) (Sunyaev & Zeldovich, 1980): extragalactic foreground due to the motion of the gas in galaxy clusters,
- Cosmic Infrared Background (CIB) (Partridge & Peebles, 1967; Puget et al., 1996): extragalactic foreground due to the emission of dust from faraway background galaxy,
- Radio: extragalactic foreground due to the emission of radio sources,

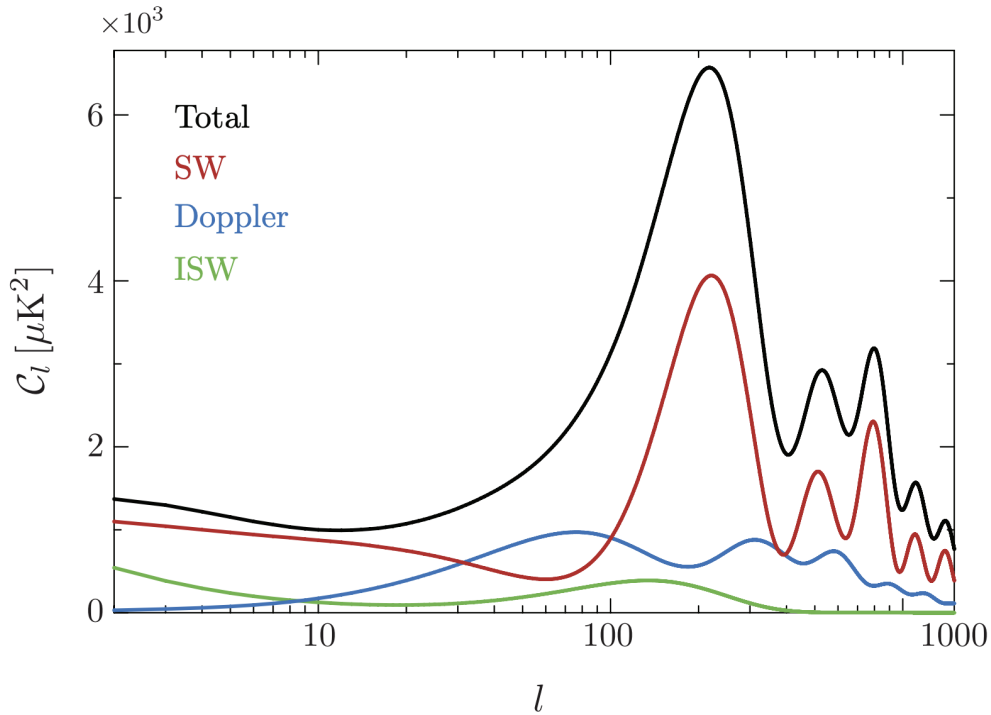


Figure 1.6: Different contributions to the rescaled power spectrum of CMB anisotropies [Baumann \(2018\)](#). The total CMB spectrum is shown in black, while the Sachs-Wolfe (SW), Integrated Sachs-Wolfe (ISW), and Doppler contributions are shown in red, green, and blue, respectively.

- Galactic synchrotron emission (e.g., [Page et al., 2007](#)): polarized galactic foreground due to the acceleration of electrons in the magnetic field of galaxy,
- Galactic dust emission: galactic foreground, that can be polarized or not, due to the thermal emission of dust grains,
- μ -distortion: galactic foreground for the temperature due to the Compton scattering of CMB photons by free electrons ([Sunyaev & Zeldovich, 1970](#)),
- free-free emission, also known as thermal bremsstrahlung: galactic foreground for the temperature due to the bremsstrahlung of electrons in the interstellar medium.

In Fig. 1.7, we see the CMB temperature angular power spectrum from [Planck Collaboration et al. \(2020a\)](#) at 143 GHz. The black dots are the measurement with the associated model in thick blue. The sum of the foregrounds is the thick orange line, and all the foregrounds are represented with different colours: CIB in orange, SZ in purple, point source (which includes radio and the point source of CIB) in green, and dust in red. The instrumental noise is represented by the grey dashed line. The shape and amplitude of the CMB power spectrum and the foregrounds will vary with frequency. Foregrounds, in particular the dust in red, dominate at large scales (small ℓ), while the CMB signal is dominant at large scales (small ℓ). Removing these foregrounds is a complex process, and it is often necessary to mask some regions of the sky to obtain a clean CMB map.

1.6.3 Influence of the cosmological parameters

The shape and amplitude of the angular power spectrum depend on the cosmological parameters and the initial conditions of the perturbations. The power spectrum can thus probe various cosmological parameters, even though some of them are correlated. Here, I will briefly present the influence of certain cosmological parameters on the CMB angular power spectrum. In the examples, the mass of the neutrinos is fixed at $\sum m_\nu = 0.06$ eV, the equation of state of dark energy is set to $w = -1$, and tensor modes are neglected. The non-varying parameters are fixed to the fiducial value from [Planck Collaboration et al. \(2020b\)](#) ($\Omega_c h^2 = 0.119$, $\Omega_b h^2 = 0.022$, $h = 0.677$, $n_s = 0.967$, $\tau = 0.056$, $\ln(10^{10} A_s) = 3.047$), and the fiducial case is represented by the black line in the figures.

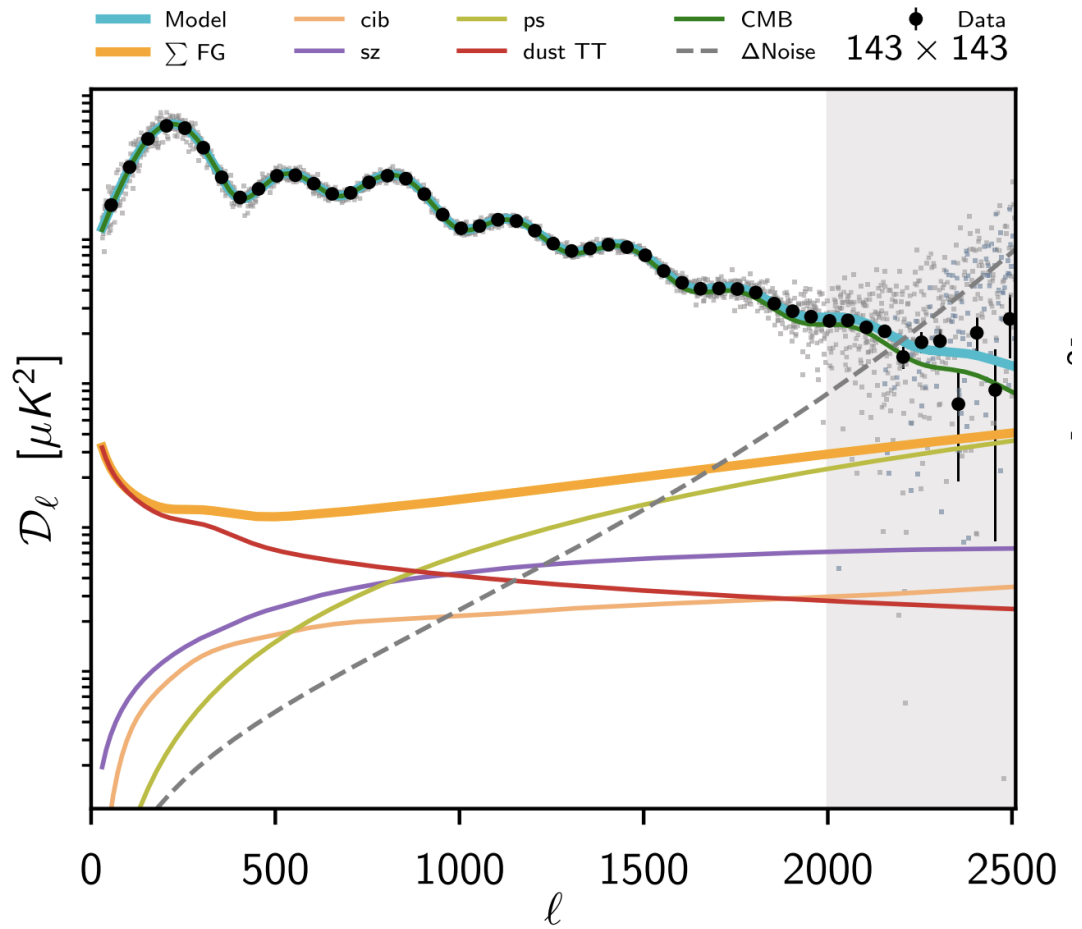


Figure 1.7: CMB temperature angular power spectrum (green line) and associated foregrounds (coloured lines) from [Planck Collaboration et al. \(2020a\)](#), at 143 GHz. The y-axis is $D_\ell = \ell(\ell + 1)C_\ell/(2\pi)$. The sum of all the foregrounds is in thick orange, the measurement in black dots, and the instrumental noise in dashed grey. The grey-shaded area is not used for cosmology.

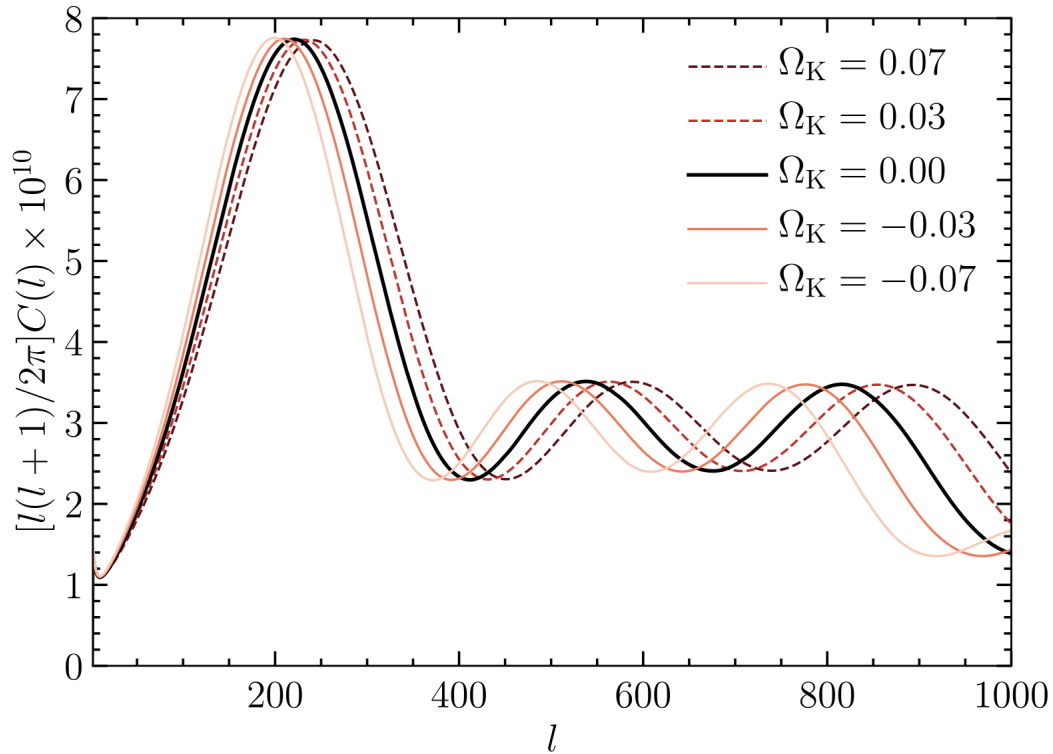


Figure 1.8: Influence of Ω_K on the CMB temperature angular power spectrum (Dodelson & Schmidt, 2020)⁷. The figure shows the power spectrum for $\Omega_K = 0$ (flat Universe) in the black line, for $\Omega_K > 0$ (open Universe) in dashed line, and for $\Omega_K < 0$ (close Universe) in solid line.

The effect of $\Omega_b h^2$ and $\Omega_c h^2$ (rather than Ω_b or Ω_c) will be studied, as the Hubble parameter can be derived from the relation $\Omega_m = 1 - \Omega_K - \Omega_\Lambda$. The reason for this focus is that CMB is more sensitive to the physical matter density (and physical baryon density) than to the density parameter Ω_m (or Ω_b).

Cosmological constant and curvature

The variation of the cosmological constant Ω_Λ will alter the matter-radiation equality and modify the relationship between angular distance and redshift. This variation will be evident in the position of the CMB angular power spectrum peaks. The variation of Ω_Λ will also influence the late ISW effect by increasing the power at large scales.

The variation in curvature $\Omega_K h^2$ will alter the angular diameter distance to the last scattering surface, resulting in similar effects to those induced by changes in the cosmological constant. The precise effect of having a close ($\Omega_K < 0$), flat ($\Omega_K = 0$), or open ($\Omega_K > 0$) Universe is shown in Fig. 1.8, represented by the solid lines, black line, and dashed lines, respectively. The figure shows that an open Universe shifts the peaks to smaller scales, while a closed Universe shifts the peaks to larger scales.

These two phenomena dominate in the present epoch, explaining why Ω_K and Ω_Λ constraints are degenerate if only CMB data are used in the analysis.

Amplitude, spectral index, and optical depth

The variation of the amplitude A_s will directly affect the CMB angular power spectrum. Specifically, changing A_s by a given factor will change the amplitude of the power spectrum by the same factor.

The spectral index n_s will alter the shape of the SW plateau observed at large scale and adjust the relative amplitudes of the acoustic peaks compared to the SW plateau.

Changing the optical depth τ will change the epoch of reionization, resulting in a modulation of the peak amplitudes.

As the effect of τ is degenerate with A_s , it is often the parameter $A_s \exp(-2\tau)$ which is considered. n_s is also correlated with A_s because n_s is normalised at a given ℓ and thus depends on the amplitude of

⁷Reprinted from Modern Cosmology, Scott Dodelson & Fabian Schmidt, Chapter 9 - The cosmic microwave background, Pages No. 231 to 269, Copyright 2021, with permission from Elsevier

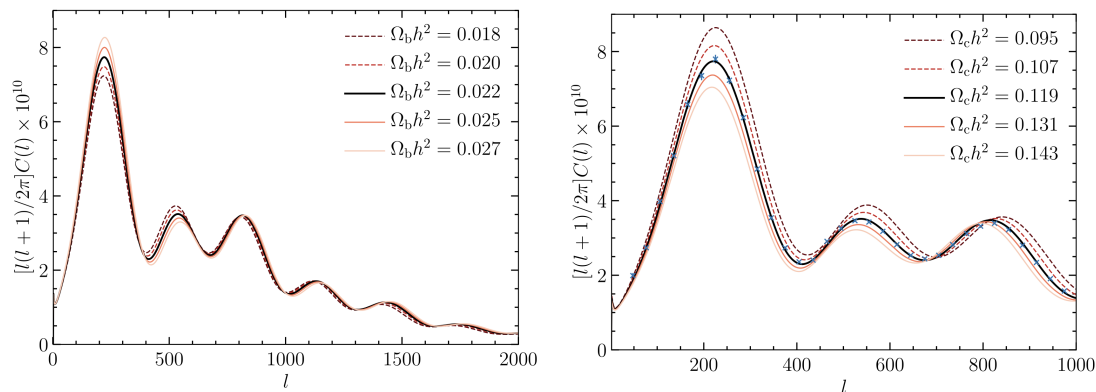


Figure 1.9: Influence of $\Omega_b h^2$ (left) and $\Omega_c h^2$ (right) on the CMB temperature angular power spectrum (Dodelson & Schmidt, 2020)⁸. Left: power spectrum for $\Omega_b h^2 = 0.022$ in the black line, for $\Omega_b h^2 < 0.022$ in dashed line, and for $\Omega_b h^2 > 0.022$ in solid line. Right: power spectrum for $\Omega_c h^2 = 0.119$ in the black line, for $\Omega_c h^2 < 0.119$ in dashed line, and for $\Omega_c h^2 > 0.119$ in solid line. Blue markers are the binned Planck measurements (Planck Collaboration et al., 2020b).

the power spectrum.

Baryon, CDM, and matter densities

In the following example, when changing $\Omega_b h^2$ and $\Omega_c h^2$, an Euclidian Universe is maintained by adjusting Ω_Λ .

The baryon density $\Omega_b h^2$ significantly impacts the CMB power spectrum in several ways. It affects the frequency of the photon-baryon plasma, leaving an imprint on the position and spacing the peaks in the CMB spectrum. The specific effects are shown in the left panel of Fig. 1.9. A higher Ω_b increases the relative amplitude of the odd peaks compared to the even peaks. In particular, as Ω_b increases, the second peak becomes smaller relative to the first. Additionally, it influences the Doppler effect, which affects the contrast between peaks. A higher Ω_b results in a greater contrast between the peaks. Finally, it will influence the Silk damping at small scales: a higher Ω_b increases the Silk damping effect, leading to more pronounced suppression of the small-scale fluctuations. Silk damping results from photon diffusion, which dissipates the small-scale fluctuations.

Modifying the cold dark matter density $\Omega_c h^2$ will modify the potential wells. Higher $\Omega_c h^2$ will lead to deeper potential wells, which prevents photons from escaping. The specific effects are shown in the right panel of Fig. 1.9. Increasing $\Omega_c h^2$ will decrease the amplitude of the peaks and will shift their position to smaller scales.

Primordial gravitational waves

If primordial gravitational waves were generated, they would dominate the power spectrum on large scales, decreasing the relative height of the acoustic peaks compared to the SW plateau.

1.7 Inhomogeneous Universe

The Friedmann-Lemaître solutions described in Sect. 1.3 provide a framework for understanding a homogeneous and isotropic Universe. However, the actual Universe exhibits inhomogeneities, as evidenced by galaxies, clusters, and superclusters. These structures emerged from the evolution of small perturbations in the matter density. The growth of primordial fluctuations, originating from the quantum fluctuations created by the inflation (see Sect. 1.5.2), is driven by gravitational instabilities that amplify the density fluctuations of matter. The observed structures in the Universe result from the interplay between gravitational forces, which amplify local density fluctuations, and cosmic expansion. To explain the structures we observe today, it is essential to understand the evolution of density perturbations under the influence of gravity in an expanding Universe. This requires solving the equations for an inhomogeneous Universe.

⁸Reprinted from Modern Cosmology, Scott Dodelson & Fabian Schmidt, Chapter 9 - The cosmic microwave background, Pages No. 231 to 269, Copyright 2021, with permission from Elsevier

1.7.1 Eulerian view of the gravitational instabilities

In this section, I will focus on solving the equations that describe the growth of inhomogeneities in a fluid of non-relativistic matter. In the linear regime, the density contrast and velocity remain small. The growth of inhomogeneities due to linear instabilities allows us to describe the emergence of structures at large scales. To understand the formation of local structures, such as galaxies, we must extend our analysis to the fully non-linear regime.

Using an Eulerian approach and under the approximation of a single flow (valid during the first phase of the gravitational instabilities, when there is no shell-crossing), the dynamics can be described by the Poisson, mass conservation, and Euler equations:

$$\Delta\Phi(\mathbf{x}) = 4\pi G\rho_0 a^2 \delta(\mathbf{x}), \quad (1.20)$$

$$\dot{\delta} + \frac{1}{a}\nabla_{\mathbf{x}} \cdot [(1 + \delta(\mathbf{x}))\mathbf{u}(\mathbf{x})] = 0, \quad (1.21)$$

$$\dot{\mathbf{u}}_i + \frac{\dot{a}}{a}\mathbf{u}_i + \frac{1}{a}\mathbf{u}_i\mathbf{u}_j = -\frac{1}{a}\nabla_i\Phi - \frac{1}{a\rho}\nabla_i P, \quad (1.22)$$

where Φ is the gravitational potential, ρ_0 is the mean density of the Universe, δ is the density contrast defined as $\delta(\mathbf{x}) = \frac{\rho(\mathbf{x}) - \rho_0}{\rho_0}$, \mathbf{u} is the peculiar velocity, P is the pressure, and a is the scale factor.

At the beginning of the gravitational instabilities, the density is only weakly perturbed compared to the homogeneous configuration. This allows us to linearize the equations, assuming that the density contrast and gradient are small: $\delta \ll 1$, and $(\frac{ut}{a})^2 \ll \delta$, where u is the characteristic fluid velocity and d the coherence length of density perturbations.

For the following equation, we introduce the dimensionless divergence of the peculiar velocity: $\theta(\mathbf{x}) = \frac{1}{aH}\nabla_{\mathbf{x}} \cdot \mathbf{u}(\mathbf{x})$. At linear order, the evolution of a fluid in the single-flow regime, the equations become:

$$\dot{\delta} + H\theta = 0, \quad (1.23)$$

$$\dot{\theta} + 2H\theta + \frac{\dot{H}}{H}\theta = -\frac{4\pi G\rho_0}{H}\delta, \quad (1.24)$$

where I have suppressed the \mathbf{x} dependence for clarity. The first equation is the mass conservation, and the second one corresponds to the Euler equation. Consequently, we can obtain the evolution for the density contrast as:

$$\ddot{\delta} + 2H\dot{\delta} = 4\pi G\rho_0\delta. \quad (1.25)$$

As the previous equation is a second-order differential equation, the evolution of the density perturbations admits two solutions. By evaluating the sign of the second derivative, we can determine that one solution is increasing while the other is decreasing. The general solution for the density contrast is:

$$\delta(t) = D_+(t)\delta_+(0) + D_-(t)\delta_-(0), \quad (1.26)$$

where D_+ is the growing mode, and D_- the decaying mode. Both modes depend on the initial conditions of the density. Here, δ_- and δ_+ denote the corresponding initial density field.

As the mass conservation can be written as $a\frac{\partial}{\partial a}\delta + \theta = 0$, the solution for $\theta(t)$ reads:

$$\theta(t) = \frac{\partial \log D_+}{\partial \log a} D_+(t)\delta_+(0) + \frac{\partial \log D_-}{\partial \log a} D_-(t)\delta_-(0). \quad (1.27)$$

Certain specific solutions can be derived. However, numerical integration is needed for solutions in the general case.

1.7.2 Lagrangian view and Zel'dovich approximation

In the previous section, I employed an Eulerian view, but an alternative approach is to adopt a Lagrangian view. The new coordinates are the initial position \mathbf{q} and the displacement field $\psi(\mathbf{q}, t)$. Under the Zel'dovich approximation (Zel'dovich, 1970), the linear solution for the displacement field is given by:

$$\nabla_{\mathbf{q}}\ddot{\psi} + 2H\nabla_{\mathbf{q}}\dot{\psi} = \frac{3}{2}\Omega H^2\nabla_{\mathbf{q}}\psi. \quad (1.28)$$

Similar to the Eulerian view, the solution is the sum of a growing and decaying mode. At linear order, we can neglect the decaying mode. The Zel'dovich approximation leads to the solution for the displacement field and density as:

$$\psi_{\text{ZA}}(\mathbf{q}, t) = D_+(t)\psi^+(\mathbf{q}), \quad (1.29)$$

$$\rho_{\text{ZA}}(\mathbf{q}, t) = \frac{\bar{\rho}}{|(1 - D_+(t)\lambda_1)(1 - D_+(t)\lambda_2)(1 - D_+(t)\lambda_3)|}, \quad (1.30)$$

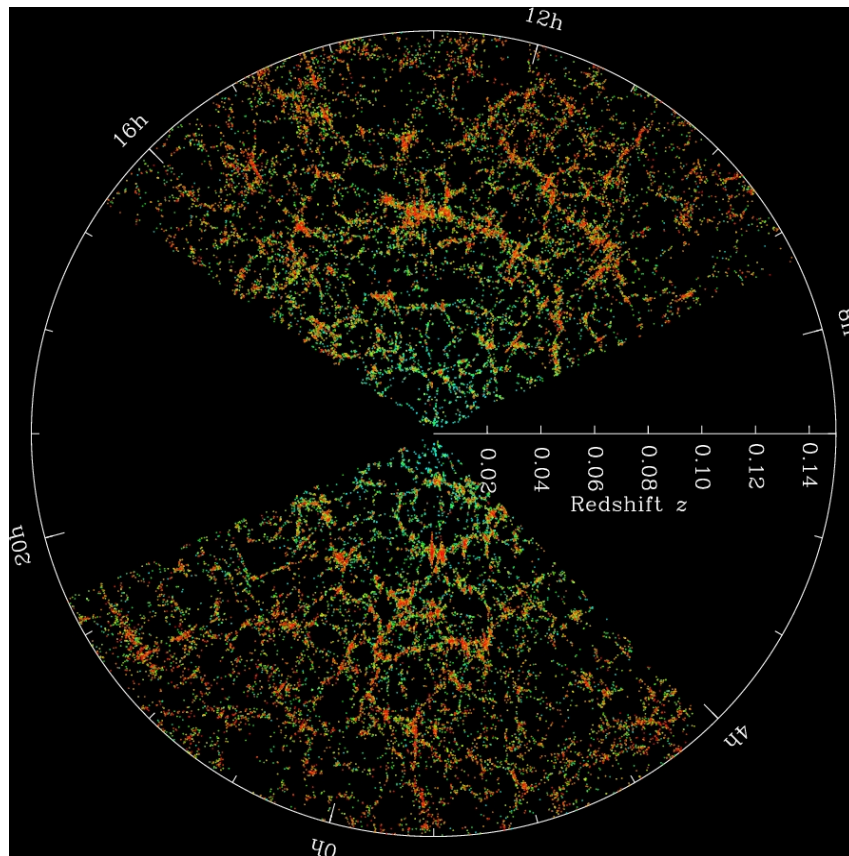


Figure 1.10: Galaxy distribution by the Sloan Digital Sky Survey. Each point is a galaxy, with higher density of galaxies represented in red. The cosmic web is clearly visible.

where $\bar{\rho}$ is the mean density in the Universe, λ_i are the eigenvalues of the deformation tensor $-\partial\psi_i^+/\partial q_j$: and are order as: $\lambda_1 < \lambda_2 < \lambda_3$. The density field is singular when the deformation tensor becomes singular, marking the occurrence of the first shell crossing. The Zel'dovich approximation is valid until this point and can be used to describe the evolution of the density field in the linear regime. Following the Zel'dovich approximation, we can describe the formation of the different structures observed in the Universe, known as the cosmic web. It is a network of complex structures with different types of environments that results from the evolution of the density field under gravitational instabilities. When $\lambda_3 > 0$, the density compression along one direction initiates the formation of the so-called pancakes, later called walls. When $\lambda_2 > 0$, the pancakes collapse along a secondary axis, giving rise to filaments. Finally, when $\lambda_1 > 0$, the filaments further collapse into knots. In regions where none of the eigenvalues are positive, the density is underdense, leading to the formation of voids. The galaxy distribution measured by the Sloan Digital Sky Survey⁹ (SDSS), shown in Fig. 1.10, illustrates these structural elements, with each point representing a galaxy. Denser regions are shown in red. From this picture, we can clearly see the presence of large voids, filaments, and knots. Over the years, various classifications of these four structures have emerged, a topic explored in more detail in Chapter 5, particularly in Sect. 5.1.

1.7.3 Structures in the non-linear regime

The linear approximation is valid for small density contrast or if the field is smoothed with a sufficiently large window function. The earlier a wavelength becomes sub-Hubble, the more it is amplified, causing small scales to be the first to enter the non-linear regime.

Scale where non-linearities become important

The normalisation of the initial power spectrum is typically performed such that the variance of the density field within a sphere of radius $R_8 = 8 h^{-1}\text{Mpc}$ is approximately unity:

$$\sigma_8^2 \equiv \left\langle \left(\frac{3}{4\pi R_8^3} \int_{|\mathbf{x}| \leq R_8} \delta(\mathbf{x}) d^3\mathbf{x} \right)^2 \right\rangle \sim 1, \quad (1.31)$$

⁹<https://www.sdss4.org/>

where δ is the density contrast and \boldsymbol{x} the spatial position. The scale where non-linearities become important must be greater than the scale of the non-linearity. The concordant model predicts that non-linearities are integrated for a length scale of $8h^{-1}\text{Mpc}$. The length scale R_8 is thus taken as the characteristic scale below which non-linearities cannot be neglected. It is the scale at which the density contrast becomes too large for the linear approximation to hold, necessitating the application of perturbation theory. Typically, the linear regime holds for $\sigma \lesssim 0.1$, and perturbation theory is applicable for $\sigma \lesssim 0.1 - 0.5$. Perturbation theory will be discussed in Sect. 5.1.1. Here, I present a simple model to solve the evolution of structures in the non-linear regime.

Spherical collapse model

The spherical collapse model is a simple model for describing the formation of structures in the non-linear regime, such as halos. This model assumes a spherically symmetric density field, where the density contrast is significant enough to enter the non-linear regime. It states that the evolution of the radius of the perturbation depends only on the mass enclosed within the sphere centred on the perturbation, not on the density profile itself. The equation of motion for the radius R is given by:

$$\ddot{R} = -\frac{GM(<R)}{R^2}. \quad (1.32)$$

In an Einstein-de Sitter Universe with a background density ρ_0 , after a given time, the overdensity of initial radius R_0 can be express as:

$$R = R_0(1 + \delta)^{-1/3} \sim R_0 \left(1 - \frac{\delta}{3}\right). \quad (1.33)$$

For an Einstein-de Sitter Universe, it is possible to determine the mean overdensity of a sphere and the moment when the perturbation collapses. The critical density is $\delta_c \approx 1.69$ and represents the threshold above which the perturbation will collapse. If the density contrast exceeds this value, the perturbation will undergo gravitational collapse.

The previous equations are only valid before shell-crossing. After spherical collapse, the sphere is expected to reach hydrostatic equilibrium through virialization. During this process, dissipative mechanisms convert kinetic energy into thermal motion, leading to an equipartition of energies. The sphere reaches a stationary regime and stabilizes its radius.

1.8 Conclusion

This chapter provides an understanding of the theoretical framework of cosmology. We have introduced the Friedmann-Lemaître models, the ΛCDM model, and the inflationary model. We have also discussed the CMB, its observations, and its properties. Finally, we presented the evolution of density perturbations in an expanding Universe from the linear to the non-linear regime using the simple gravitational collapse model. We have seen how initial perturbations can lead to the formation of large-scale structures, which will be the focus of the next chapter.

2.1	Statistics	25
2.1.1	Two-point correlation function and power spectrum	25
2.1.2	Angular power spectrum	26
2.1.3	Higher-order statistics	26
2.2	Observables	28
2.2.1	Matter	28
2.2.2	Galaxy clustering	28
2.2.3	Gravitational lensing	30
2.2.4	Thermal Sunyaev-Zel'dovich	33
2.2.5	Galaxy clusters	34
2.3	Halo model to model the power spectrum	36
2.3.1	Basic principle	36
2.3.2	Power spectrum computation	38
2.3.3	Application to matter	38
2.3.4	Halo properties	38
2.3.5	Angular power spectrum	40
2.4	Conclusions	41

The previous chapter presented the emergence and evolution of structures and the formation of the cosmic web. In this chapter, the focus shifts to the observables that enable the study of large-scale structures of the Universe. I will first present the statistics used to study the LSS, such as the two-point correlation function and power spectrum, and then discuss higher-order statistics in Sect. 2.1. In Sect. 2.2, I will discuss several observables used in LSS studies: matter distribution, galaxy clustering, gravitational lensing, CMB lensing, and the thermal Sunyaev-Zel'dovich effect, along with their inherent limitations. Finally, in Sect. 2.3, I will present the methodology of modeling observables using a halo model framework. The two first sections of this chapter (Sects. 2.1 and 2.2) were written with the help of [Bernardeau et al. \(2002\)](#); [Bernardeau \(2007\)](#); [Peter & Uzan \(2013\)](#); [Dodelson & Schmidt \(2020\)](#), while the third section (Sect. 2.3) was written with the help of [Hill & Pajer \(2013\)](#); [Mead et al. \(2020\)](#); [Asgari et al. \(2023\)](#).

2.1 Statistics

To study the statistics of the LSS of the Universe, we assume a fundamental principle known as the ergodicity principle. This principle states that the time average of a physical quantity over a long period for a single system is equal to the ensemble average over a large number of identical systems at a single instant in time. Applied to the LSS, which are, according to the cosmological principle, statistically homogeneous and isotropic, this principle allows us to relate ensemble averages with spatial averages over one realisation of a random field.

2.1.1 Two-point correlation function and power spectrum

Two-point statistics, such as the two-point correlation function and power spectrum, are the most common tools used to study the LSS of the Universe. They describe how points (such as galaxies) are distributed relative to each other at given distances. Two-point statistics capture all the Gaussian information, as the statistical properties of a Gaussian random field can be fully described by its two-point statistics, particularly by the mean and variance.

The two-point correlation function and the power spectrum can be applied to any scalar field. Here, I will use the density contrast $\delta(\mathbf{x})$, which is a random field.

The correlation function between two objects at positions \mathbf{x} and $\mathbf{x} + \mathbf{r}$ is defined as:

$$\xi(r) = \langle \delta(\mathbf{x})\delta(\mathbf{x} + \mathbf{r}) \rangle, \quad (2.1)$$

and only depends on the norm of \mathbf{r} because of statistic homogeneity and isotropy. If $\xi(r) > 0$, the two points are more likely to be found at a distance r than in a random distribution. Conversely, if $\xi(r) < 0$, the two points are less likely to be found at a distance r than if they were distributed randomly.

The Fourier transform and inverse Fourier transform are defined with the following normalisation:

$$\delta(\mathbf{x}) = \int \frac{d^3\mathbf{k}}{(2\pi)^{3/2}} \delta(\mathbf{k}) \exp(i\mathbf{k} \cdot \mathbf{x}), \quad \delta(\mathbf{k}) = \int \frac{d^3\mathbf{x}}{(2\pi)^{3/2}} \delta(\mathbf{x}) \exp(-i\mathbf{k} \cdot \mathbf{x}). \quad (2.2)$$

If $\delta(\mathbf{x})$ is real, then $\delta(\mathbf{k}) = \delta^*(-\mathbf{k})$. k is the wave number corresponding to a specific scale, with higher k values representing smaller scales.

It is then possible to apply Fourier transform to the two-point correlation, which becomes:

$$\langle \delta(\mathbf{k})\delta(\mathbf{k}') \rangle = \int \frac{d^3\mathbf{x}}{(2\pi)^{3/2}} \frac{d^3\mathbf{r}}{(2\pi)^{3/2}} \xi(r) \exp[-i(\mathbf{k} + \mathbf{k}') \cdot \mathbf{x} - i\mathbf{k}' \cdot \mathbf{r}] \quad (2.3)$$

$$= \delta_{\text{D}}(\mathbf{k} + \mathbf{k}') \int d^3\mathbf{r} \xi(r) \exp(i\mathbf{k} \cdot \mathbf{r}) \quad (2.4)$$

$$\equiv \delta_{\text{D}}(\mathbf{k} + \mathbf{k}') P(k), \quad (2.5)$$

where δ_{D} is the Dirac delta function. From these equations, we define $P(k)$ as the power spectrum, which is directly related to the correlation function. The power spectrum focuses on the distribution of power across different scales (frequencies) instead of distances. Because the Fourier modes are discrete, the power spectrum is only accessible at a finite number of modes k . A high value of $P(k)$ at a given k means that structures of this size are prominent in the distribution. It is also possible to obtain the inverse relation:

$$\xi(r) = \int \frac{d^3\mathbf{k}}{(2\pi)^3} P(k) \exp(i\mathbf{k} \cdot \mathbf{r}) \quad (2.6)$$

The two-point correlation function and the power spectrum are related, and we can obtain one from the other. Depending on the problem, one can be more valuable than the other. As observables are

isotropic, they are invariant under translation, leading to the decorrelation of the modes in Fourier space: $\langle P(k)P(k') \rangle = 0$, which is an advantage of working in Fourier space. For example, when studying galaxy clustering, the two-point correlation function is generally used as it is easier to measure and to account for masks in it (e.g., Zhang et al., 2024, which uses a two-point correlation function to study the clustering of Milky Way-like galaxies, using simulations). On the other hand, analyses of LSS, which include broader fields, are often done using the power spectrum to decorrelate the modes (e.g., Tegmark et al., 2004, which computes the power spectrum using SDSS observations).

Cross-correlation

In the previous derivation of the two-point correlation function and power spectrum, we looked at the correlation of density with itself. It is also possible to look at the correlation between two different fields. Cross-correlating different fields can improve the constraining power as we gain information. For example, the cross-correlation between CMB and one probe of LSS can help to break the Ω_K - Ω_Λ degeneracy mentioned in Sect. 1.6.3.

Cross-correlation is particularly useful when different observables are sensitive to different aspects of the underlying cosmological model. For instance, the CMB provides information about the early Universe and the initial conditions of structure formation, and LSS probes the growth of structures at later times. Additionally, different observables are more sensitive to different physics (baryonic matter or total matter, for instance). Combining datasets through cross-correlation leads to more precise constraints (e.g., Van Waerbeke et al., 2014, which perform the first detection of correlation between tSZ from Planck and weak lensing from Canada-France-Hawaii Telescope lensing survey (CFHTLenS)¹, or Fang et al., 2024, which evaluate the constraining power of combining weak lensing, CMB lensing, galaxy density and tSZ).

2.1.2 Angular power spectrum

The angular power spectrum C_ℓ characterizes the distribution of power of the fluctuations on the celestial sphere. It is measured in multipole ℓ , corresponding to different angular scales. Smaller ℓ corresponds to larger angular scales, while larger ℓ represents smaller scales, capturing finer details. Essentially, the angular power spectrum is a projection of the 3D power spectrum onto the sphere. One example of the angular power spectrum is the CMB temperature, as illustrated in Fig. 1.5.

The angular power spectrum is defined over the entire sky by integrating the power spectrum on the line of sight. To obtain its complete expression, the depth correlations must be accounted for. I will not provide the full equations here, but they can be found, for example, in Gao et al. (2024). A standard approximation known as the Limber approximation (Limber, 1954; Peebles, 1973) assumes that these depth correlations are negligible with respect to the spatial correlations because the radial window functions are broad compared to the scales of interest in the spatial fluctuations. This approximation is particularly accurate for small scales. To have robust estimation on all scales, different approaches, such as the one done in Gao et al. (2024), can be used.

In this thesis, I will use the angular power spectrum under the Limber approximation and the flat sky approximation, which neglects the sky curvature. A comparison between different approximations can be found in Gao et al. (2024). Under these two approximations, the angular power spectrum C_ℓ is expressed as:

$$C_\ell = \int \frac{d\hat{r}}{\hat{r}^2} W(\hat{r}) W'(\hat{r}) P(\ell/\hat{r}; \hat{r}), \quad (2.7)$$

where \hat{r} is the comoving distance, $W(\hat{r})$ is the window function tailored to the specific observational characteristics, and $P(\ell/\hat{r}; \hat{r})$ is the 3D power spectrum at the scale corresponding to the multipole ℓ at the comoving distance \hat{r} . A comoving distance is the distance between two points in the Universe at a given time. The physical distance r is given by $r = a(t) \times \hat{r}$, where $a(t)$ is the scale factor. The comoving distance is constant in time, while the physical distance evolves with time.

2.1.3 Higher-order statistics

Two-point statistics only capture the Gaussian information of a field. To probe non-Gaussianities, it is often beneficial to use higher-order statistics. This can be achieved through higher-order correlation functions and poly-spectra, which extend the concept of the second-order correlation function and power spectrum. Additionally, other statistics, such as Minkowski functionals, peak counts, etc., can be used. In the following, I will provide a few examples of higher-order statistics.

¹<https://www.cfht.hawaii.edu/en/>

Wick theorem

The Wick theorem allows to compute higher-order statistics of a Gaussian field, such as the four-point and higher-order correlation functions. It states that the expectation value of the product of a Gaussian random field is equal to the sum of all possible pairings of the field, with each pairing representing a two-point correlation function. This allows us to relate higher-order correlation functions to the two-point correlation function. The theorem is expressed as:

$$\langle \delta(\mathbf{k}_1) \dots \delta(\mathbf{k}_{2p+1}) \rangle = 0, \quad \langle \delta(\mathbf{k}_1) \dots \delta(\mathbf{k}_{2p}) \rangle = \sum \prod_{i,j} \langle \delta(\mathbf{k}_i) \delta(\mathbf{k}_j) \rangle. \quad (2.8)$$

The first equation indicates that for a Gaussian field, the expectation value of an odd number of fields is zero. This occurs because Gaussian distributions are symmetric about their mean. The second equation shows that for an even number of fields, the expectation value is the sum of all possible pairings of the fields.

Higher-order correlation functions and poly-spectra

The correlation function of order N , which quantifies the spatial correlations between N points, is defined as:

$$\xi_N(\mathbf{x}_1, \dots, \mathbf{x}_N) = \langle \delta(\mathbf{x}_1), \dots, \delta(\mathbf{x}_N) \rangle_c, \quad (2.9)$$

where $\langle \dots \rangle_c$ denotes the connected correlation function. The connected part represents the sum of all possible pairings of the fields, as specified by the Wick theorem.

From this higher-order correlation function, we can define the poly-spectra P_n :

$$\langle \delta(\mathbf{k}_1) \dots \delta(\mathbf{k}_n) \rangle_c = \delta_D(\mathbf{k}_1 + \dots + \mathbf{k}_n) P_n(\mathbf{k}_1, \dots, \mathbf{k}_n). \quad (2.10)$$

When $n = 3$, the poly-spectra is known as the bispectrum $P_3(k) \equiv B(k)$. The bispectrum is widely used to study non-Gaussianities in the Universe. It can, for example, be applied to galaxy clustering (Sefusatti et al., 2006), weak lensing (Takada & Jain, 2004), and other observable. It can also be used to study alternative models of gravity (e.g., Gil-Marín et al., 2011).

Apart from higher-order correlation functions and poly-spectra, other higher-order statistics can be used to study the LSS of the Universe. I mention a few examples of such statistics here.

Minkowski functionals

The Minkowski functionals are a set of geometric and topological measures used to assess the morphology and connectivity of structures within a field. In three-dimensional space, these functionals include the volume, surface area, mean curvature, and Euler characteristic (representing the number of independent connected regions minus the number of holes in each region). They are valuable tools for analysing the non-Gaussian characteristics and topological properties of the cosmic web (e.g., Einasto et al., 2014).

Peak counts

Peak count statistics involve the identification and enumeration of the local maxima (peaks) within a field. The histogram of these peaks can be used to constrain cosmological parameters. For example, peak counts can apply to various fields such as the distribution of galaxies or the weak lensing convergence map (e.g., Aycoberry et al., 2023, article discussed in Chapter 6).

Void statistics

Voids are large and underdense regions in the Universe. Studying their distribution and properties provides insights into the LSS and the mechanisms driving structure formation in the Universe. Void statistics include the abundance, size distribution, and shape within galaxy surveys. These statistics serve as a probe for testing models of dark energy and modified gravity, for example, (e.g., Contarini et al., 2022).

Density split statistics

Density split statistics involve dividing the density field into distinct regions based on different density thresholds, allowing for separate analyses of different regimes. This approach allows the study of the distinct properties associated with each type of region. For example, density split statistics is used to analyse the clustering properties of galaxies (e.g., Paillas et al., 2024).

2.2 Observables

The LSS of the Universe are probed through a range of observables, including matter distribution, galaxy clustering, gravitational lensing, CMB lensing, and the thermal Sunyaev-Zel'dovich effect. These observables provide valuable insights into the distribution of matter and the underlying cosmological model. In the following sections, I will present these observables and discuss their inherent limitations.

2.2.1 Matter

The matter distribution is one of the most fundamental observables in studying the LSS of the Universe. We expect that the contrast of the matter density is fully characterized by its power spectrum because of the Gaussian initial conditions in the Universe. However, on the underdense and overdense regions, the evolution of structures imprints non-Gaussian information that is not captured by the power spectrum. The power spectrum still provides essential information about how matter is distributed across different scales in the Universe. During the formation of structures, small-scale fluctuations enter the horizon before matter-radiation equality, suppressing the power spectrum. The turnover of the power spectrum at k_{eq} , which is the scale that enters the horizon at equality, is a key feature. Another critical scale is k_{NL} , marking where non-linearities become significant. At present epoch $k_{\text{NL}}(a=1) \sim 0.25 h\text{Mpc}^{-1}$. In earlier epochs when structures were less evolved, k_{NL} was larger, indicating that smaller structures could still be described within the linear theory.

There is no direct way to measure the matter power spectrum. Instead, it can be inferred through observations such as galaxy clustering and gravitational lensing at the cost of some potential bias. Fig. 2.1 illustrates the matter power spectrum measured from various observational techniques: CMB, SDSS galaxies, cluster abundance, weak lensing, and Lyman- α forest in green, black, blue, pink, and red, respectively. The agreement between these observations demonstrates their complementarity, enabling a robust characterisation of the matter power spectrum across different scales. For instance, CMB probes primarily large scales, while Lyman- α forest probes small scales.

2.2.2 Galaxy clustering

Galaxy clustering is another essential observable for studying the LSS of the Universe. Galaxies are tracers of the underlying matter distribution, and their clustering properties provide valuable insights into this distribution. The galaxy density field can be derived from galaxy redshift surveys, which provide the 3D positions of each galaxy (such as with the SDSS survey, see Fig. 1.10 and Tegmark et al., 2004). Two types of surveys are used: photometric and spectroscopic. Photometric surveys use different broadband filters to capture images of galaxies at different wavelengths, thus providing images of the integrated light. The redshift of galaxies is then estimated based on the colour of the galaxies using photometric redshift techniques. In contrast, spectroscopic surveys employ spectrographs to obtain the spectrum of each galaxy. By identifying spectral lines and measuring their shifts, a direct and precise measurement of the redshift can be obtained. In summary, photometric surveys allow for the rapid cover of large sky areas at the expense of less precise redshift estimations. On the other hand, spectroscopic surveys provide more accurate redshift measurements but cover smaller areas of the sky, as obtaining a spectrum for each galaxy is time-consuming. Typically, photometric surveys are used for wide-field analyses, while spectroscopic surveys are used for precise measurements and detailed analysis of galaxy properties.

The galaxy distribution differs from the matter distribution due to the effect of stellar formation and other physical processes, making it a biased tracer of LSS. To quantify the relationship between the galaxy and matter distributions, a bias factor b is introduced. The linear bias relation is given by:

$$\delta_{\text{g}}(\mathbf{x}) = b\delta_{\text{m}}(\mathbf{x}), \quad (2.11)$$

which holds only to linear order. Consequently, the galaxy power spectrum $P_{\text{g}}(k)$ is related to the matter power spectrum $P(k)$ through the bias factor as $P_{\text{g}}(k) = b^2 P(k)$. The bias factor depends on the galaxy sample and generally varies with redshift. It can be determined using the peak-background split formalism, which shows that the bias factor is larger for high-mass halos, as they cluster more strongly than the overall matter distribution. While it is possible to introduce a non-linear bias (see, for example, the comparison of using a linear or non-linear bias in Hoffmann et al., 2017), the following will focus on the linear bias only.

To obtain the true power spectrum of galaxies (and the underlying matter distribution), it is necessary to account for the redshift-space distortion (RSD) and cosmology. Galaxy clustering is affected by RSD (Kaiser, 1984), which arises due to the peculiar velocities of galaxies. These velocities cause galaxies to move away from us, leading to an increase in their observed redshift and thus affecting their inferred

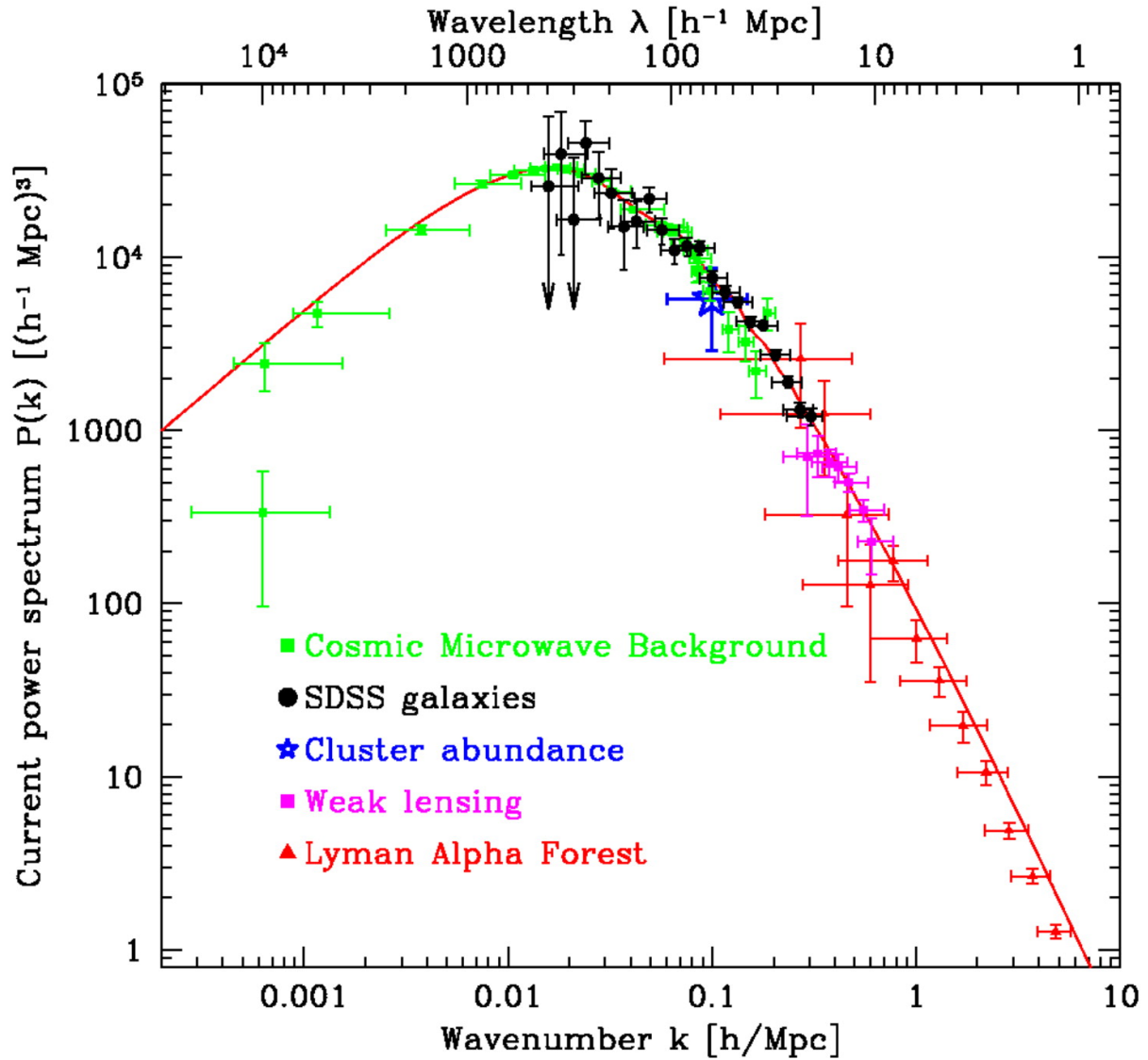


Figure 2.1: Matter power spectrum from Tegmark et al. (2004). The red line corresponds to the matter power spectrum for $\Omega_m = 0.28$, $h = 0.72$, $\Omega_b/\Omega_m = 0.16$, and $\tau = 0.17$. The different colours represent the power spectrum measured from different observable: CMB (green), SDSS galaxies (black), cluster abundance (blue), weak lensing (pink), and Lyman- α forest (red) with their associated error bars.

distances. The RSD effect provides a means to quantify the growth rate of structures, denoted by f (where $f \equiv \frac{d \log D(a)}{d \log a}$). This growth rate is related to the galaxy overdensity in redshift space, and accurately accounting for RSD is crucial for deriving the true galaxy power spectrum.

In the distant-observer approximation, only RSD is considered when analysing the observed overdensity. The redshift-space overdensity in Fourier space can be expressed as:

$$\delta_{g,\text{RSD}}(\mathbf{k}) = [b + f\mu_k^2] \delta_m(\mathbf{k}), \quad (2.12)$$

where μ_k is the cosine of the angle between the wavevector and the line of sight. When $f\mu_k^2 \geq 0$, the apparent overdensity in redshift space is larger than in real space. At the level of the power spectrum, the observed galaxy power spectrum in redshift space can be written as:

$$P_{g,\text{RSD}}(k, \mu_k, \bar{z}) = P_{\text{lin}}(k, \bar{z}) [b + f\mu_k^2]^2 + P_N, \quad (2.13)$$

where P_{lin} is the linear power spectrum, \bar{z} is the mean redshift, and P_N is the noise power spectrum. The noise is considered as white noise here, implying it is scale-independent. This approximation is valid if galaxies are Poisson-sampled from an underlying continuous field: $P_N = 1/\bar{n}_g$, where \bar{n}_g is the mean number of galaxies. This approximation is generally reasonable for low k but may not hold at higher k .

The other effect that must be accounted for is using an incorrect cosmology when assigning 3D positions to galaxies based on their redshift, as different cosmologies affect the distance-redshift relation. This is known as the Alcock-Paczynski effect (Alcock & Paczynski, 1979). In the density contrast and power spectrum, all the components of \mathbf{k} need to be accounted for. This effect is degenerate with RSD, but this degeneracy can be broken by examining the BAO peak. The BAO peak directly measures the comoving distance \hat{r} and the Hubble parameter H .

In observations, the power spectrum at low k (larger scales) exhibits large error bars due to cosmic variance, as only a finite volume of the Universe is observed. This is evident from the black dots in Fig. 2.1. On the other hand, at higher k (small scales), non-linearities become significant and must be accounted for.

2.2.3 Gravitational lensing

Gravitational lensing is a powerful tool for studying the LSS of the Universe. It arises due to the deflection of light from distant sources by inhomogeneous gravitational fields (lenses), providing a way to map the matter distribution in the Universe, including dark matter. These inhomogeneities induce distortions in the observed shapes of galaxies (images). The deflection of light and its impact on the formation of images are illustrated in Fig. 2.2. Gravitational lensing applies to different systems: a star can be lensed by another star, a galaxy can be lensed by another galaxy, a cluster or LSS, and the CMB can be lensed by the LSS. In this section, we will first focus on the lensing of galaxies by cluster and finish with CMB lensing. In the case of galaxies, since we do not know their intrinsic shape, we assume their shapes are randomly oriented, attributing any observed shape correlations to the effects of gravitational lensing.

The distortion caused by gravitational lensing arises because the deflection angle varies across the galaxy image. This distortion is quantified by the distortion matrix A , which describes how the gravitational lens alters the image of a source:

$$A \equiv \frac{d\theta_S}{d\theta} = \begin{pmatrix} 1 - \kappa - \gamma_1 & -\gamma_2 \\ -\gamma_2 & 1 - \kappa + \gamma_1 \end{pmatrix} \quad (2.14)$$

where θ_S is the position of the source and θ position of the image. κ is the convergence, which describes the isotropic magnification, and $\gamma = (\gamma_1, \gamma_2)$ is the shear, describing the anisotropic stretching of the image. These two effects are illustrated in Fig. 2.3, where the blue area is the source image, the red illustrates the impact of convergence, and the green the effect of shear. The magnification due to convergence affects the received flux of the galaxy. By measuring the ellipticities of galaxies, we can estimate the shear field, which depends on the gravitational potential of the lens structures.

Gravitational lensing can be categorized into two regimes: strong and weak lensing. Strong lensing results in giant arcs and multiple images of the source. Weak lensing occurs when the convergence and shear are small ($\kappa \ll 1$ and $\gamma \ll 1$), resulting in subtle distortions of the background images without creating multiple distinct images of the source. In the intermediate regime between weak and strong lensing lies the arclet regime, where images can be subject to strong distortions without necessarily producing multiple images. Fig. 2.4 is a simulation illustrating the effect of different lensing regimes. The central dot is the lens object, and the black lines represent the strong magnification zone. The coloured objects represent images of the source. From left to right: the simulation goes from weak to strong lensing with the apparition of an Einstein ring when the lens and the source alignment are nearly perfect.

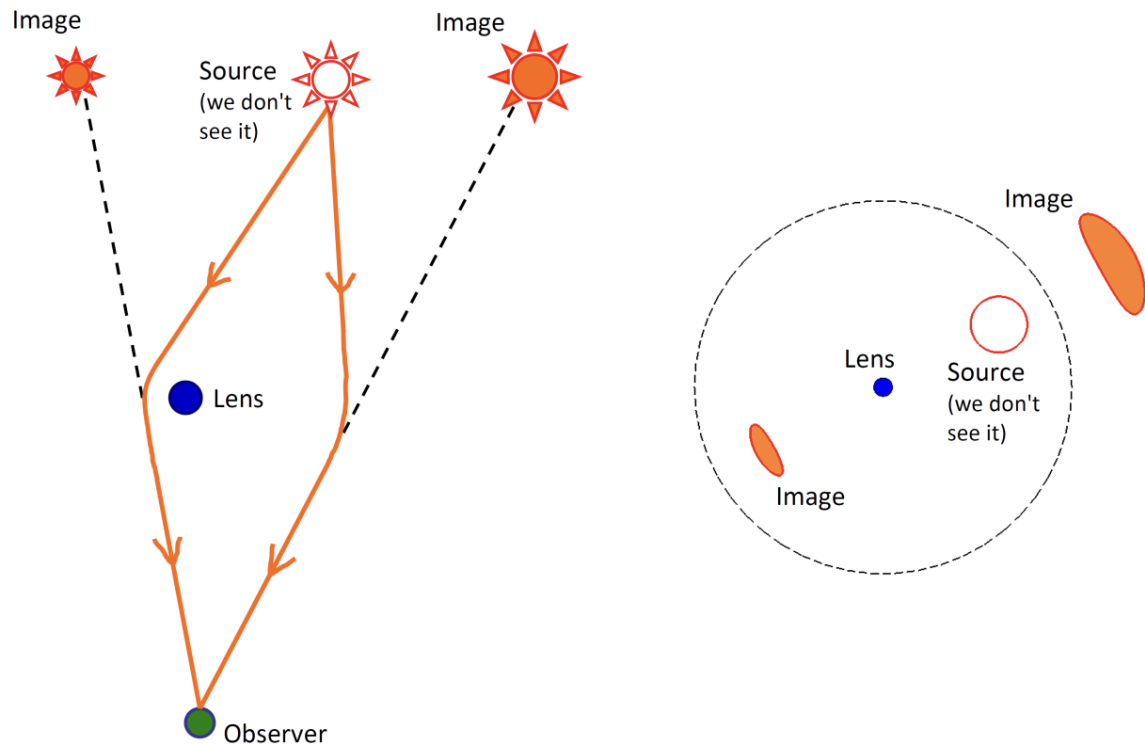


Figure 2.2: Illustration of the lensing effect (Bisnovatyi-Kogan & Tsupko, 2017). The left panel shows the geometry of the gravitational lensing and the trajectories of light. The right panel shows the position of the images created, with the position of the potential Einstein ring in dashed black. On both panels, the lens is represented by the blue object, and the images an observer will see are in orange.

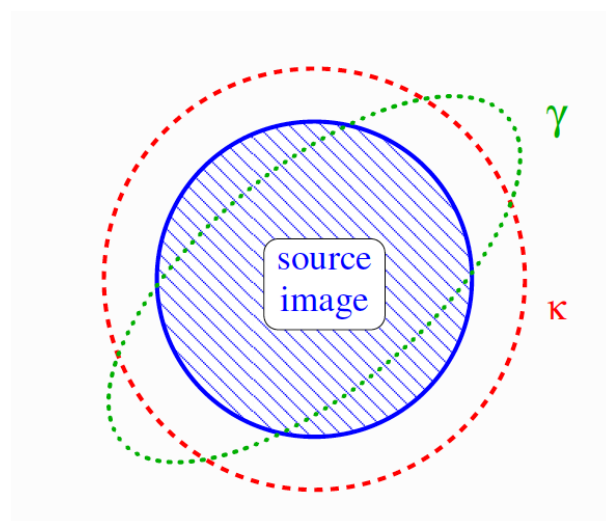


Figure 2.3: Illustration of the lensing effect. The non-lensed image is in blue. The convergence κ , represented in red, is responsible for the magnification, while the shear γ , represented in green, is responsible for the distortion of the image.

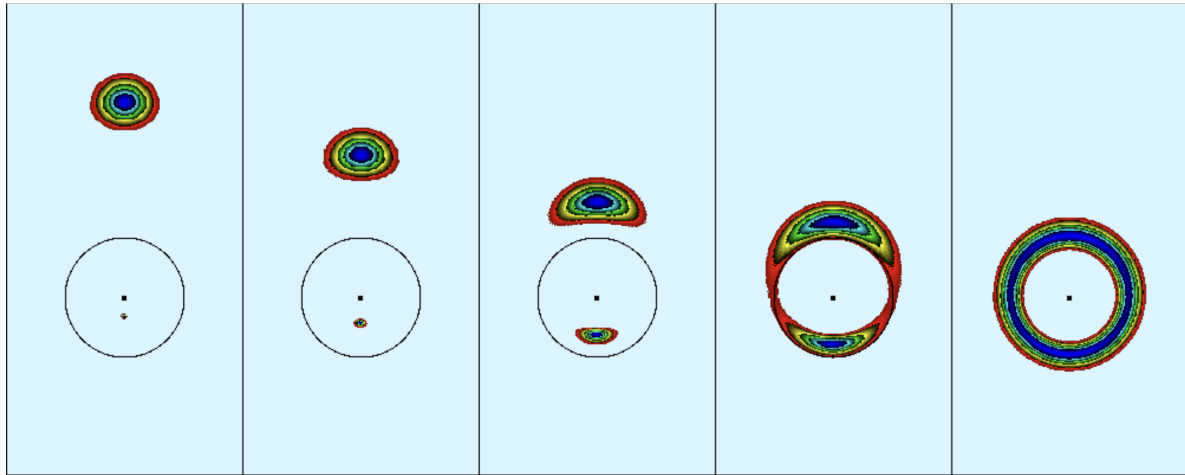


Figure 2.4: Five snapshots of a gravitational lensing simulation (Wambsganss, 1998). From left to right: the simulation goes from weak to strong lensing with an Einstein ring. This is due to the larger alignment between lens and source on the right.

For example, strong lensing can measure the mass distribution in clusters, such as in Oguri (2010). Weak lensing can be used to study the properties and evolution of LSS and geometry of the Universe (Kilbinger, 2015).

Measurement of the cosmic shear

The measurement of cosmic shear relies on analysing the shapes of background galaxies and their ellipticities. This measure is based on the assumption that the intrinsic ellipticities of source galaxies are random and uncorrelated. Any deviations from this random distribution and correlation observed in the ellipticities are attributed to gravitational lensing effects. However, this assumption is not always valid, as the environment can influence the shape of galaxies. Galaxies often reside in groups or clusters where external gravitational distortions can alter their shapes by a few percent. Additionally, alignments of galaxies shapes can occur when galaxies reside in filaments, for instance.

Limitations

When measuring the cosmic shear, several limitations or uncertainties can arise:

- Mass-sheet degeneracy: the addition of a uniform mass sheet to the lens can produce the same observed lensing effects while changing the inferred mass distribution,
- Redshift distribution of the sources: imperfect knowledge of the redshift distribution of the sources introduces uncertainties in the inferred mass distribution and resulting analyses,
- Intrinsic alignment: the intrinsic alignment of galaxies can influence the measured shear, potentially introducing biases into the analyses,
- Survey systematics: various systematics effects can impact measurements, including atmospheric turbulence, optical aberrations, or other instrumental artefacts.

CMB lensing

The CMB lensing occurs when photons from the CMB are deflected by the LSS of the Universe. While the mean temperature of the CMB remains unchanged, the directions of arrival of the photons are shifted due to gravitational lensing. This effect distorts the observed CMB temperature map. For instance, CMB lensing can be used to measure cluster mass (Huchet & Melin, 2024).

The CMB lensing is reconstructed by measuring the non-Gaussianities created by the lens on the CMB observables, which are supposed to be Gaussian. In Fig. 2.5, the angular power spectrum of the CMB temperature is shown for both the lensed (red) and unlensed (dotted black) cases. The figure illustrates how lensing smooths out the peaks and enhances power on very small scales.

²Reprinted from Modern Cosmology, Scott Dodelson & Fabian Schmidt, Chapter 13 - Probes of structure: lensing, Pages No. 373 to 399, Copyright 2021, with permission from Elsevier

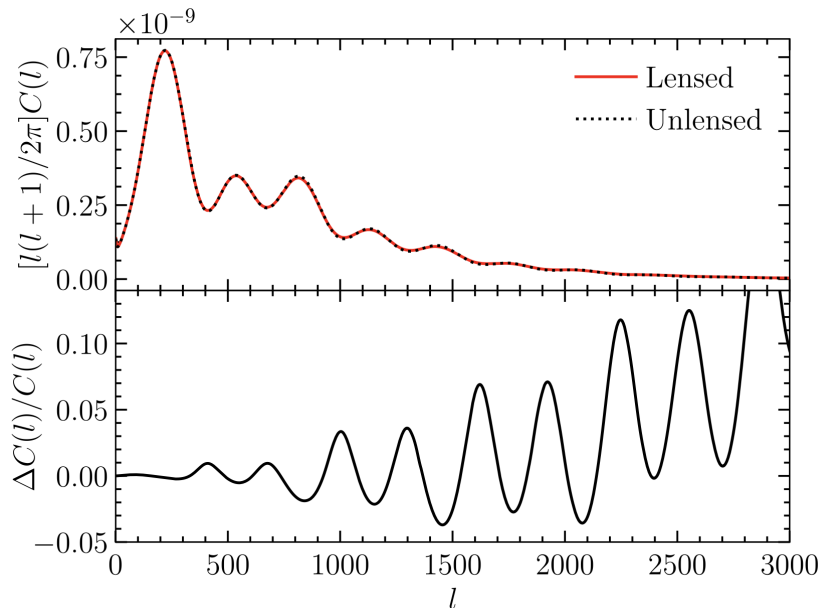


Figure 2.5: Effect of lensing on the CMB temperature angular power spectrum from [Dodelson & Schmidt \(2020\)](#)². Top: temperature angular power spectrum before (black dotted line) and after lensing (red line). Bottom: fractional difference between the two.

2.2.4 Thermal Sunyaev-Zel’dovich

The thermal Sunyaev-Zel’dovich (tSZ) effect [Sunyaev & Zeldovich \(1970\)](#) occurs when CMB photons interact with hot electrons in galaxy clusters through inverse Compton scattering. This interaction results in an energy shift of the CMB photons, as illustrated in Fig. 2.6. Since the tSZ effect depends on the electron pressure, it is a powerful probe of baryonic matter within these clusters.

The tSZ effect manifests as a distortion of the CMB black body spectrum, which varies with frequency as illustrated in Fig. 2.7. The undistorted CMB spectrum is shown in dashed line, while the distorted spectrum, resulting from a hypothetical cluster with a mass 1000 times greater than usual cluster, is shown in solid line. At frequencies below ~ 217 GHz, the tSZ effect decreases the CMB temperature, whereas, at higher frequencies, it causes an increment. The tSZ effect is primarily driven by contributions from galaxy clusters, where hot and free electrons reside. Additionally, since the effect only weakly depends on the distance to the cluster, it is particularly useful for cross-correlation with other observables.

By measuring this distortion, it is possible to infer the astrophysical properties of the hot gas in galaxy clusters, particularly their pressure, on which the tSZ signals scale, and also recover cosmological information. The tSZ effect is characterized by the Compton- y parameter, which quantifies the integrated pressure of the hot gas along the line of sight. The Compton- y parameter in a given direction \hat{n} is defined as:

$$y(\hat{n}) = \frac{\sigma_{\text{T}}}{m_e c^2} \int P_e ds, \quad (2.15)$$

where σ_{T} is the Thomson cross-section, m_e is the electron mass, c is the speed of light, P_e is the electron pressure, and ds the distance along the line of sight. The Compton- y parameter is related to the temperature of the electrons (T_e) and the gas density (n_e) through the relation $P_e = n_e T_e$. The number of free electrons in a halo scales with M , and the temperature scales as $M^{2/3}$. Consequently, the shape and amplitude of the spectra are influenced more by massive halos (as the Fourier transform of the pressure thus scales as $M^{5/3}$) than matter or galaxies, whose profiles scale as $\sim M$. The amplitude of the tSZ is thus extremely sensitive to σ_8 , as σ_8 influences the formation and evolution of structures, in particular the abundance and mass of clusters (e.g., [Komatsu & Kitayama, 1999](#); [Refregier et al., 2000](#); [Seljak et al., 2001](#); [McCarthy et al., 2014](#)).

The temperature shift of the CMB photons due to the tSZ depends on the frequency ν and is given by:

$$\frac{\Delta T}{T_{\text{CMB}}} = g(\nu)y, \quad (2.16)$$

where $g(\nu) = x \coth(x/2) - 4$ if relativistic corrections are neglected, $x = h\nu/(k_{\text{B}}T_{\text{CMB}})$, k_{B} is the Boltzmann constant, and h is the Planck constant. $g(\nu)$ is positive (negative) for frequencies above

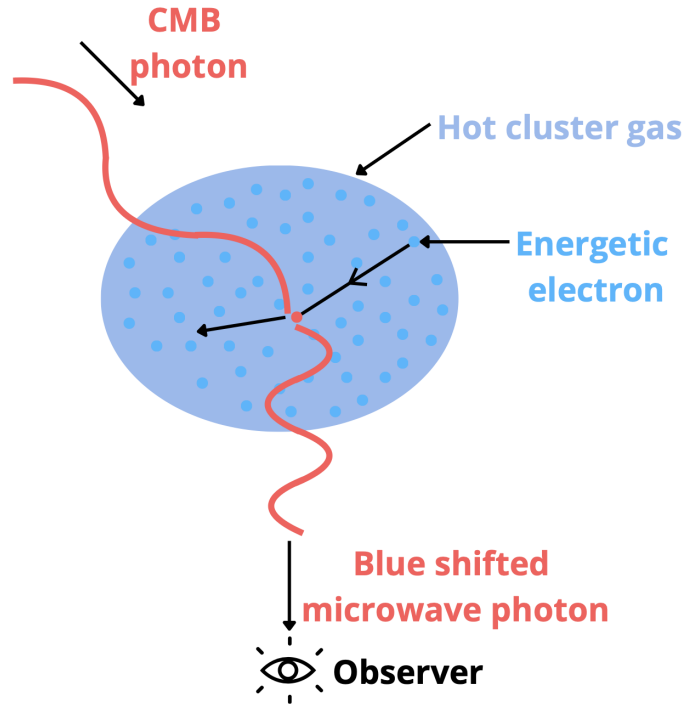


Figure 2.6: Illustration of the thermal Sunyaev-Zel'dovich effect. CMB photons (in red) interact with energetic electrons (in blue) through inverse Compton scattering, leading to a shift in the energy of the observed photon.

(below) $\nu \sim 217$ GHz. Examples of conversion factors between the CMB temperature and the Compton- y parameters can be found in [Planck Collaboration et al. \(2016b\)](#).

The tSZ is a foreground of the CMB and also a probe of the distribution of baryonic matter, which can help us obtain better astrophysical and cosmological constraints. These two aspects emphasize the need for having robust modeling of the tSZ effect. For instance, the tSZ effect can be used to study early dark energy models (e.g., [Sadeh et al., 2007](#); [Waizmann & Bartelmann, 2009](#)). Additionally, the tSZ effect is sensitive to astrophysics phenomena such as AGN feedback, which redistributes mass and modifies the pressure of the hot plasma in massive halos (e.g., [McCarthy et al., 2014, 2023](#); [Le Brun et al., 2015](#); [Spacek et al., 2018](#); [Lee et al., 2022](#); [Moser et al., 2022](#); [Pandey et al., 2023](#))

The tSZ effect will be further discussed in Chapter 4, where I will introduce observations both on the full sky and within individual clusters. I will also present a comparison between the predictions from an analytical model and measurements in simulations. This comparison includes (angular) power spectrum, pressure profile, and a few parameters of the model.

2.2.5 Galaxy clusters

The tSZ effect and photometric surveys mentioned previously can also be used to study the properties and evolution of galaxy clusters. In addition, X-ray observations, which we did not discuss here, are also a powerful tool for studying galaxy clusters. Each method of observation has its own advantages and limitations. For instance, optical and X-ray observations rely on direct detection of light from the cluster, making measurement challenging at large distances, whereas the tSZ effect can more easily probe higher redshifts. At low redshifts, however, X-ray identification has a lower mass threshold than the tSZ effect, with the opposite trend at high redshifts (e.g., [Bleem et al., 2015](#)), highlighting the complementarity of these methods.

The main challenge in studying galaxy clusters is estimating their mass. The mass can be inferred using different observables, including the Sunyaev-Zel'dovich effect, X-ray emission, and weak lensing, again demonstrating the complementarity of different observational techniques. The mass function of clusters can be used to constrain cosmological parameters, such as σ_8 .

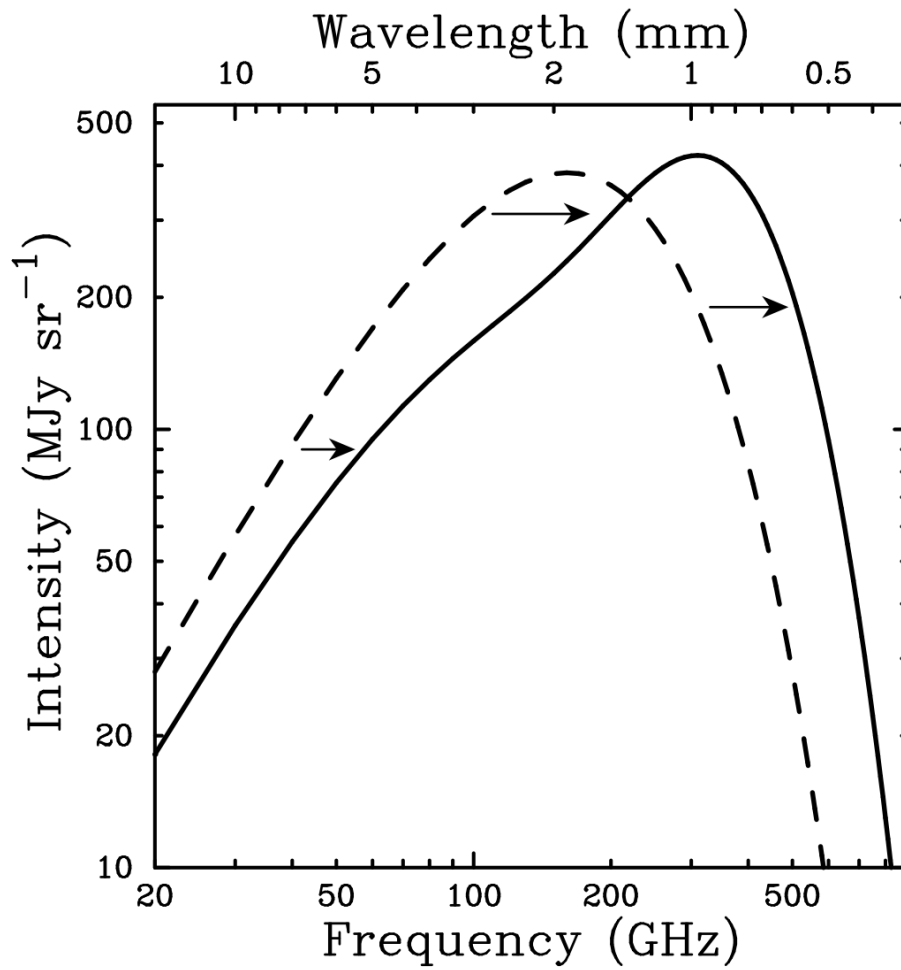


Figure 2.7: Illustration of the thermal Sunyaev-Zel'dovich effect on the CMB spectrum (Carlstrom et al., 2002). The dashed line represents the undistorted CMB spectrum, while the solid line represents the distorted spectrum.

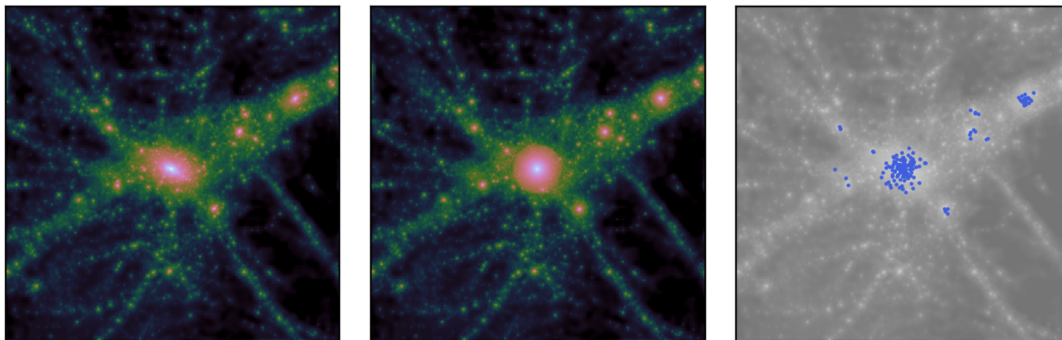


Figure 2.8: Schematic visualisation of the process of the halo model (Asgari et al., 2023). Left: matter density field in a $25^3 h^{-1} \text{Mpc}$ region of an N-body simulation, centred on the massive halo of $\sim 10^{14.5} h^{-1} M_{\odot}$ at $z = 0$. Centre: same region where all the identified halos are isolated and replaced with idealized spherical halos of the same mass. Right: result of populating these halos with galaxies according to a simple galaxy-occupation prescription.

2.3 Halo model to model the power spectrum

In the previous sections, I have described the (angular) power spectrum statistics and some of the different observables that can be used to probe the LSS of the Universe. In this section, I will present the halo model, a powerful tool for modeling the power spectrum of the matter distribution and its tracers. For instance, the matter and tSZ power spectrum predicted by a halo model will be discussed and compared with measurements from simulations in Chapter 4.

2.3.1 Basic principle

The halo model is a framework that proposes a simplified description of the distribution of matter in the Universe and allows for constructing a model for the non-linear matter density and its tracers. It assumes that the cosmic web can be a superposition of all its individual components: dark matter, gas, and galaxies. The halo model assumes that all matter resides in halos, which are spherically symmetric structures with a range of masses, and that each particle belongs to one single halo. Halos are distributed where the initial perturbations allow gravity to create overdensities, following the linear matter power spectrum. To understand how a halo model works, a schematic visualisation of how the halo model behaves in a simulation is shown in Fig. 2.8. Starting with a simulation and its matter density (left panel), the halos are identified, isolated, and replaced with idealized spherical halos of the same mass (centre panel). The halo model is then based on how these halos are populated with galaxies using a galaxy-halo occupation prescription (right panel). In Sect. 4.3.5, I will discuss the consequences of the halo model assumptions and the limitations of the halo model in the context of the tSZ effect.

The halo model is based on three main properties that characterize halos and encapsulate the complexity of the cosmic web: the halo mass function, the halo bias, and the halo profile. The halo mass function represents the number of halos of a given mass at a given redshift. The halo bias describes how halos cluster relative to matter. The halo profile, assumed to be spherically symmetric and only mass-dependent, details how tracers are distributed within a halo. These properties are typically obtained from numerical simulations. The halo bias is usually considered linear. To probe tracers other than matter, their connection with matter must be understood, and the signal from the tracers must originate from the halos. For instance, the halo model can be applied to the tSZ signal, which is sourced by the electron pressure, which is most intense within halos. The halo model framework is analytical. An alternative framework, known as the simulation-based approach, uses dark matter-only simulations, identifies halos, and populates them with tracers.

With the properties and distribution of halos, the halo model can estimate statistical properties such as the power spectrum. The power spectrum for matter and its tracers is the sum of the one-halo and two-halo terms, as illustrated in Fig. 2.9, where halos are represented in blue. The one-halo term corresponds to the correlation of two points within the same halo (shown in purple in the figure). The two-halo term corresponds to the correlation of two points in two different halos (shown in red in the figure). Fig. 2.10 shows the matter power spectrum in grey, with the contribution of the one-halo and two-halo terms in dotted purple and dashed red, respectively. The one-halo term dominates at small scales, probing the interiors of halos, while the two-halo term dominates at large scales, probing the large-scale clustering of halos and their spatial distribution.

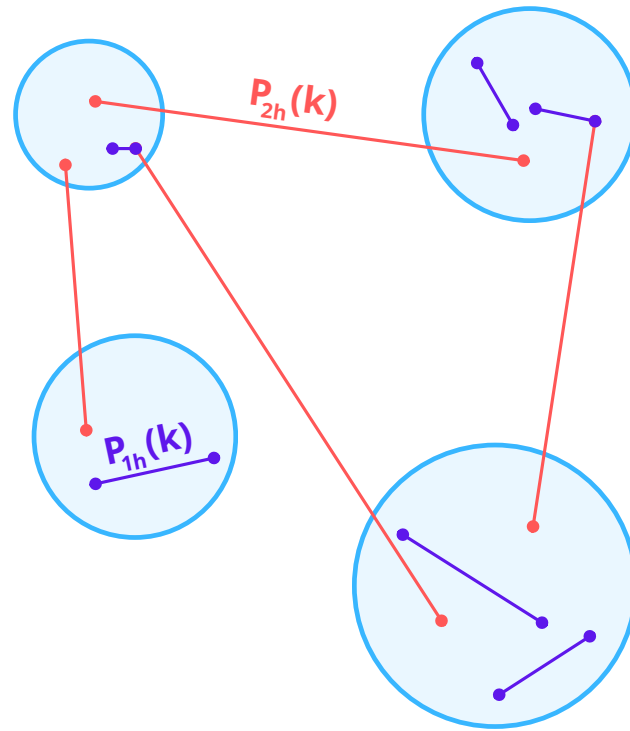


Figure 2.9: Illustration of the one- and two-halo term that appears in the power spectrum prediction done by a halo model. Blue circles are halos that are spherically symmetric. Examples of the one-halo term are represented by the purple lines and represent the correlation between two points that are in the same halo. Examples of the two-halo term are represented in red and correspond to the correlation between two points in two halos. The power spectrum is the sum of these two terms, with the one-halo term dominating at small scales and the two-halo term dominating at large scales.

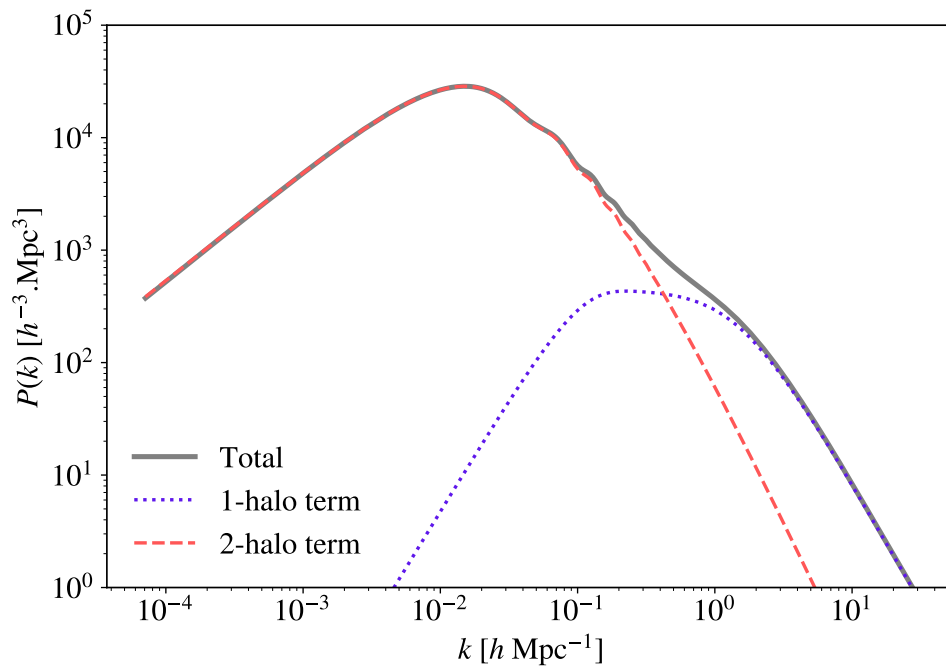


Figure 2.10: Matter power spectrum at $z = 0$ predicted by the HMx (Mead et al., 2020) halo model. The grey line represents the total power spectrum, the purple dotted line represents the one-halo term, and the red dashed line represents the two-halo term.

2.3.2 Power spectrum computation

Now that the principles of a halo model are understood, we can delve deeper into the computation and associated equations. The power spectrum between two fields u and v is the sum of the one-halo and two-halo term: $P_{uv}(k) = P_{1h,uv}(k) + P_{2h,uv}(k)$, defined as:

$$P_{1h,uv}(k) = \int_0^\infty W_u(M, k) W_v(M, k) n(M) dM, \quad (2.17)$$

$$P_{2h,uv}(k) = P_{\text{lin}}(k) \prod_{i=u}^v \left[\int_0^\infty b(M) W_i(M, k) n(M) dM \right], \quad (2.18)$$

where we have the Fourier transform of the field u , given by:

$$W_u(M, k) = \int_0^\infty 4\pi \hat{r}^2 \frac{\sin(k\hat{r})}{kr} \theta_u(M, \hat{r}) d\hat{r}, \quad (2.19)$$

with $\theta_u(M, \hat{r})$ the averaged radial profile for the field u , which is function of mass M and comoving radius \hat{r} . In this last equation, the integration over angular coordinates is possible due to the assumption of the halo model: spherical symmetry of halos and the dependence of each halo's properties only on its mass. The equations of the one-halo and two-halo terms assume linear halo bias. The two-halo term is proportional to the linear power spectrum times a normalisation that depends on the halo properties.

2.3.3 Application to matter

To apply this formalism to the matter density, the halo mass function $n(M)$ and linear bias $b(M)$ must satisfy two conditions:

$$\int_0^\infty M n(M) dM = \bar{\rho}, \quad (2.20)$$

$$\int_0^\infty M b(M) n(M) dM = \bar{\rho}, \quad (2.21)$$

where $\bar{\rho}$ is the mean density of the Universe. These conditions ensure that all matter is contained in halos, and on average, the matter is unbiased with respect to itself.

2.3.4 Halo properties

The halo model relies on selecting the appropriate halo mass function, halo bias, and halo profile. These components are typically calibrated using simulations or observational data. Different choices lead to distinct implementations and results. The choice of these properties can impact the accuracy of the model, and it is essential to calibrate them to improve the model's predictive capabilities. Examples of halo model are proposed by, for instance, Hill & Pajer (2013); Mead et al. (2020); Maniyar et al. (2021); Bolliet et al. (2023). In the halo model, halos are considered as discrete quantities, whereas their boundaries are not clearly defined in observations. A fundamental choice in simulations is how to identify halo masses, which can be done using different methods, including the friends-of-friends (FoF), spherical overdensity (SO), or Bryan & Norman (1998) definition, for example. The Bryan & Norman (1998) definition is the one I used in this thesis and will be further described in Sect. 3.4.

It is often standard practice to parametrize the halo mass function in terms of the peak height parameter ν or variance σ and to adopt a linear halo bias. The use of ν or σ instead of M comes from the fact that the halo mass function (and associated bias) demonstrates nearly universal behaviour across different cosmologies and redshifts when expressed in terms of these variables (e.g., Press & Schechter, 1974; Bond et al., 1991; Sheth & Tormen, 1999; Tinker et al., 2008).

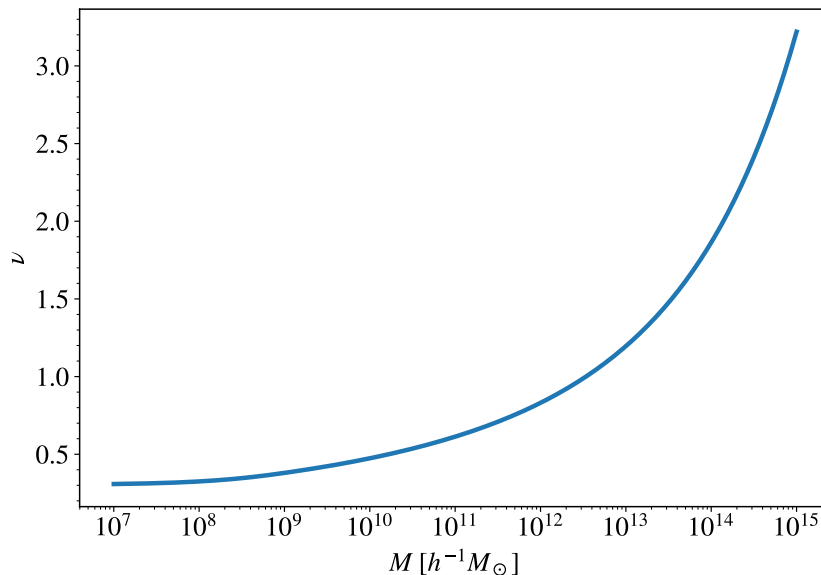
Peak height ν

To determine the peak height ν , we start with the variance of the linear matter overdensity smoothed on a comoving scale \hat{R} :

$$\sigma^2(\hat{R}) = \int_0^\infty 4\pi \left(\frac{k}{2\pi} \right)^3 P_{\text{lin}}(k) T^2(k\hat{R}) d \ln k, \quad (2.22)$$

where $T(k\hat{R})$ is the Fourier transform of the filter window function. Typically, in real space, $T(k\hat{R})$ is represented by a top hat function, with its Fourier transform given by:

$$T(k\hat{R}) = \frac{3}{(k\hat{R})^3} \left(\sin(k\hat{R}) - k\hat{R} \cos(k\hat{R}) \right). \quad (2.23)$$

Figure 2.11: Peak height ν as a function of mass.

The comoving scale \hat{R} corresponds to the comoving radius of a sphere in a homogeneous Universe which contains a given mass M :

$$M = \frac{4}{3}\pi\hat{R}^3\bar{\rho}. \quad (2.24)$$

The previous equations allow us to express $\sigma(\hat{R})$ in terms of the mass. We can now define the peak height ν as:

$$\nu(M) = \delta_c/\sigma(M), \quad (2.25)$$

where $\delta_c \approx 1.686/D(z)$ is the critical overdensity for collapse at redshift z (discussed in Sect. 1.7.3), and $D_+(z)$ is the linear growth factor normalised to 1 at $z = 0$. This change of coordinates is represented in Fig. 2.11 and can differ with the chosen mass definition. The peak height increases monotonically with M .

Halo mass function

The halo mass function $n(M)$ describes the number density of halos with mass M at a given redshift. It can be derived using different analytical approaches and fitting processes. In terms of peak height, the halo mass function is expressed as:

$$f(\nu)d\nu = \frac{M}{\bar{\rho}}n(M)dM. \quad (2.26)$$

The integral of $f(\nu)$ over all peak heights ν must satisfy the condition $\int_0^\infty f(\nu)d\nu = 1$, ensuring that all matter is accounted for within halos. A few standard halo mass functions are the one of Press & Schechter (1974); Sheth & Tormen (1999); Tinker et al. (2008), for example.

Linear halo bias

On sufficiently large scales, overdensities of halos can be approximated using the halo mass function derived from the peak-background split formalism (Cole & Kaiser, 1989; Mo & White, 1996; Sheth et al., 2001). According to this formalism, the density field consists of large-scale and small-scale components, with halos forming at the peaks of the large-scale density fluctuations. Using the peak-background split formalism, the linear halo bias can be approximated from any mass function $f(\nu)$ through:

$$b(\nu) = 1 - \frac{1}{\delta_c} \left[1 + \frac{d \ln f(\nu)}{d \ln \nu} \right]. \quad (2.27)$$

If $f(\nu)$ satisfies the mass normalisation condition in Eq. (2.20), then this automatically satisfies the bias normalisation condition in Eq. (2.21).

Other alternative approaches can be used to obtain the halo bias (Sheth et al., 2001; Tinker et al., 2010). It is also possible to use simulation-calibrated bias relations or non-linear halo bias models. Adopting a non-linear halo bias can enhance the accuracy of the model, particularly in the transition between the one-halo and two-halo terms (Mead & Verde, 2021).

Dark matter halo profile

To compute the matter power spectrum within the halo model framework, the dark matter halo profile is needed. For other tracers, this profile can also serve as a starting point. A commonly adopted profile for collisionless matter is the Navarro–Frenk–White (NFW) profile (Navarro et al., 1997a):

$$\rho(r) = \frac{\rho_s}{(r/r_s)(1+r/r_s)^2}, \quad (2.28)$$

where r_s is the scale radius and ρ_s the scale density. Both of these parameters depend on the halo mass. In a halo model context, the profile is usually truncated at the halo radius r_h . This radius is not necessarily the virial radius but is defined more generally such that:

$$M = \frac{4}{3}\pi r_h^3 \Delta_h \bar{\rho}, \quad (2.29)$$

where Δ_h is the halo overdensity, it can be taken as 200, as often used in simulations or based on other definitions.

The scale radius r_s is related to r_h through the concentration-mass relation $c_M = r_h/r_s$. This relation and the scale density are calibrated against simulations and depend on the specific halo and mass definitions employed.

Baryonic feedback

In the Universe, baryonic feedback processes can alter the halo properties. By comparing simulations with observations, it has been identified that baryonic processes inject energy into the system, leading to a regulation of the mass function. AGN feedback, for example, can expel gas from the centres of halo. The NFW profile is a reasonable choice to match to gravity-only simulated data. However, when accounting for baryonic feedback, the halo profile can be significantly altered.

Studies like those by Rudd et al. (2008) and Zentner et al. (2008) have shown that AGN activity can decrease small-scale power. The authors have demonstrated that this effect could be captured by changing the concentration-mass relation in the NFW profile used in the halo-model matter power spectrum calculation. The exact impact of baryonic feedback on the halo profile remains uncertain and will likely vary depending on the tracers of interest.

Tracers

This framework applies to any diffuse tracer of the LSS tracers as long as the halo profile can be defined. Examples of tracers include galaxies, intrinsic alignments of galaxies, X-rays, cosmic infrared background (CIB), neutral hydrogen, gamma rays, tSZ, kSZ, CIB,... For tracers that are too diffuse and not explicitly tied to halos, applying the halo model framework becomes more challenging.

Generally, the spectrum has a linear shape on large scales, with its amplitude determined by the statistics of how tracers populate halos. On smaller scales, the shape of the one-halo is governed by the tracer profiles.

2.3.5 Angular power spectrum

The power spectrum obtained from the halo model framework has a redshift dependence, and integrating them over comoving distance \hat{r} (or equivalently redshift) gives the angular power spectrum:

$$C_{uv}(\ell) = \int_0^{\hat{r}_H} \frac{X_u(\hat{r})X_v(\hat{r})}{f_k^2(\hat{r})} P_{uv}(k(\hat{r}), z(\hat{r})) dr, \quad (2.30)$$

where ℓ is the multipole moment, \hat{r}_H the Hubble radius, $z(\hat{r})$ is the redshift at comoving distance \hat{r} , the functions $X_u(\hat{r})$ and $X_v(\hat{r})$ are the window functions for the fields u and v , respectively, and $k(\hat{r}) = (\ell + 1/2)/f_k(\hat{r})$, with $f_k(\hat{r})$ the comoving angular-diameter distance:

$$f_k(\hat{r}) = \begin{cases} \sin(H_0\sqrt{|\Omega_K|}/\hat{r})/(H_0\sqrt{|\Omega_K|}) & \Omega_K < 0, \\ \hat{r} & \Omega_K = 0, \\ \sinh(H_0\sqrt{|\Omega_K|}/\hat{r})/(H_0\sqrt{|\Omega_K|}) & \Omega_K > 0. \end{cases} \quad (2.31)$$

The angular power spectrum can also be expressed in terms of the sum of the one-halo and two-halo terms: $C_{uv}(\ell) = C_{1h,uv}(\ell) + C_{2h,uv}(\ell)$.

In Chapter 4, I used the HMx (Mead et al., 2020) halo model to predict the pressure and matter auto- and cross-power spectrum. In this chapter, I will introduce the choice done by HMx for the different ingredients of the halo model and compare the predictions with measurements from simulations.

2.4 Conclusions

In this chapter, I have introduced the large-scale structures of the Universe and the observables that can be used to probe it. I discussed different statistical measures, including the angular power spectrum and higher-order statistics applicable to these observations. Finally, I presented the halo model as a robust framework for modeling the power spectrum of the matter distribution and its tracers.

While observations contain some inherent limitations and uncertainties, understanding how to model them effectively and selecting optimal statistics is crucial. Addressing these challenges can help to mitigate degeneracies in cosmological or astrophysical parameters, enhancing our understanding of the Universe.

Part II

Simulations

3.1	Initial conditions	44
3.2	Numerical overview of RAMSES	44
3.2.1	AMR scheme	44
3.2.2	N-body solver	45
3.2.3	Hydrodynamical solver	45
3.2.4	Time step control	46
3.2.5	Refinement strategy	46
3.2.6	Baryonic processes	47
3.3	Description of simulations	47
3.3.1	The Horizon suite of simulations	47
3.3.2	The Magneticum simulations	49
3.3.3	The BAHAMAS simulations	50
3.4	Halo mass function of the simulations	50

As discussed in Chapter 1, fully understanding the formation and evolution of structures in the Universe requires moving beyond linear approximations to account for the non-linear evolution. This can be achieved using perturbation theory, which will be described in Sect. 5.1.1 or through simulations, which is the focus of this chapter. Since perturbation theory is only valid before shell-crossing, simulations are a more robust method for studying the non-linear regime. However, they are more time-consuming and demand significant computational resources.

In this chapter, I will begin by describing how to obtain the initial conditions for running a simulation in Sect. 3.1. In Sect. 3.2, I will present the numerical methods used in N-body and hydrodynamical simulations, focusing on the RAMSES code. In Sect. 3.3, I will present the simulations used in this work: Horizon-AGN, Horizon-noAGN, Horizon-Large, L896_wCDM, Magneticum, and BAHAMAS. Finally, I will present the halo mass function in Sect. 3.4.

Most of my work focuses on the study and run of the Horizon suite of simulations, which are conducted using the RAMSES code. This is why I focus on the RAMSES description. While other codes exist, such as GADGET (used for Magneticum), I will not detail them.

The methods sections of this chapter (Sects. 3.1 and 3.2) were written with the help of Dodelson & Schmidt (2020) and Teyssier (2002) for RAMSES.

3.1 Initial conditions

The first step to run a cosmological simulation is to generate the initial conditions. Two main methods exist for this purpose: the Zel'dovich approximation, and the second-order perturbation theory (2LPT).

The Zel'dovich approximation (Zel'dovich, 1970) is a first-order solution of the Lagrangian perturbation theory. It allows for generating a Gaussian density field in Fourier space from the desired matter power spectrum. The Lagrangian potential and its gradient can be computed for every particle position, which leads to the displacement field. The particles are located on a grid and assigned with an initial velocity that follows the Zel'dovich approximation (see Sect. 1.7.2).

The Zel'dovich approximation solution reproduces accurately the linear growing mode of density and velocity perturbations. However, it is less accurate for non-linear growth. Therefore, simulations using this method must start at a higher redshift, where the non-linear growth is less significant.

The 2LPT method (Jenkins, 2010) solves the displacement field using a second-order Lagrangian perturbation theory. This approach is more accurate in reproducing the growing modes, including non-linearities, and allows to start the simulations at a lower redshift.

3.2 Numerical overview of RAMSES

Once the initial conditions are created, the next step is to evolve the system. Cosmological simulations can be N-body simulations (containing only dark matter) or hydrodynamical simulations. I will describe the methods used in RAMSES (Teyssier, 2002). If a simulation requires high masses, it can be achieved by increasing the number of particles or wavelengths in the computational volume. Additionally, a good spatial resolution will depend on the numerical methods employed.

The RAMSES code allows the production of both N-body and hydrodynamical simulations, enabling the study of structure formation with high spatial resolution, for example. The code can also produce zoom-in simulations of a single object, for example, but I am only using the production of large volumes to simulate the cosmic web. The code employs a tree-based Adaptive Mesh Refinement (AMR) hydrodynamical scheme using the "Fully Threaded Tree" data structure described by Khokhlov (1998). This structure allows for recursive grid refinement on a cell-by-cell basis. The N-body solver is similar to the one of the ART code (Kravtsov et al., 1997). The hydrodynamical solver uses a finite volume method with a second-order Godunov scheme. In the following sections, I will describe the AMR method in Sect. 3.2.1, the N-body solver in Sect. 3.2.2, and the hydrodynamical solver in Sect. 3.2.3. Additionally, I will discuss the time step control in Sect. 3.2.4 and the refinement strategy in Sect. 3.2.5. Finally, I will present the baryonic processes implemented in RAMSES in Sect. 3.2.6.

In RAMSES, all the solvers are implemented using conformal time as time variable. This approach enables the computation of comoving coordinates with only a few changes of variables (see Martel & Shapiro, 1998 and references therein).

3.2.1 AMR scheme

To follow the components of the simulations, different methods can be employed. Lagrangian or quasi-Lagrangian schemes (Gnedin, 1995; Pen, 1995) utilize a moving mesh that follows the flow geometry while

maintaining a constant number of grid points. Eulerian schemes, on the other hand, are based on a static mesh. Smooth Particles Hydrodynamical (SPH) is an intermediate solution that follows the Lagrangian evolution of the flow, with resolution elements defined as average.

The method used in RAMSES (Teyssier, 2002) is the Adaptive Mesh Refinement (AMR), initially described in Berger & Olinger (1984) and Berger & Colella (1989). AMR is an Eulerian hydrodynamic scheme, with a hierarchy of grids to locally increase spatial resolution. The resolution is proportional to the local cell size of the grid, making it space-dependent. The grid is continuously refined and coarsened during the simulation to follow the evolution of the system, maintaining a roughly constant mean number of particles per cell.

The data structure in RAMSES follows the "Fully Threaded Tree" of Khokhlov (1998). The basic elements are octs, group of 2^{dim} sibling cells. Each oct belongs to a level of refinement l . At the coarse level ($l = 1$, the least refined level), the grid is Cartesian and serves as the base of the tree structure.

Time integration is performed at each level independently, with two possible approaches. First, a single-step algorithm can be used, where the same time step is applied to all levels. Second, an adaptive time step can be used, where each level evolves with its own time step based on a stability condition, which will be discussed in more detail in Sect. 3.2.4.

To build the refinement map, cells are marked for refinement according to the user-defined criteria. To ensure a smooth transition in spatial resolution, an oct that will be refined must be surrounded by $3^{\text{dim}} - 1$ neighbouring parent cells. The modification of the tree structure involves splitting and destroying children cells according to the refinement map, a process that will be discussed further in Sect. 3.2.5.

3.2.2 N-body solver

N-body simulations are the simplest simulations, containing only dark matter particles. These simulations solve the Vlasov-Poisson system, which only includes gravitational interactions, to track the evolution of the system over time.

The N-body solver in RAMSES (Teyssier, 2002) is similar to that in the ART code (Kravtsov et al., 1997). It solves the Vlasov-Poisson system of equations for a collisionless N-body system in reduced units. For each particle p , the system is:

$$\Delta\Phi(x) = 4\pi G\rho, \quad \frac{d\mathbf{x}_p}{dt} = \mathbf{v}_p, \quad \frac{d\mathbf{v}_p}{dt} = -\nabla_x\Phi, \quad (3.1)$$

where \mathbf{x} is the position of the particle, and \mathbf{v} its velocity.

The grid-based scheme used to solve the equations follows these steps:

1. The mass density ρ is computed on the mesh, using a Cloud-In-Cell (CIC) interpolation (Hockney & Eastwood, 1981),
2. The Poisson equation is solved, and the potential Φ is obtained by a conjugate gradient,
3. The acceleration on the mesh is computed using a standard finite-difference approximation of the gradient,
4. The acceleration of each particle is computed using an inverse CIC interpolation,
5. The velocity of each particle is updated on a half-time step, according to its acceleration (often called the "kick" step),
6. The position of each particle is updated according to its velocity (often called the "drift" step),
7. The velocity of each particle is updated on the other half-time step, according to its acceleration at the new position ("kick" step). The succession of steps 5, 6, and 7 is called "leapfrog". This method allows for a symplectic integrator, which conserves volume in space-phase and thus energy, maintaining the stability of the orbit over time.

3.2.3 Hydrodynamical solver

To obtain more realistic simulations, it is necessary to include baryonic processes and solve the hydrodynamics. In RAMSES (Teyssier, 2002), hydrodynamics are solved using the Euler equation in their

conservative form:

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho u) = 0, \quad (3.2)$$

$$\frac{\partial}{\partial t}(\rho u) + \nabla \cdot (\rho u \otimes u) + \nabla P = -\rho \nabla \Phi, \quad (3.3)$$

$$\frac{\partial}{\partial t}(\rho e) + \nabla \cdot \left[\rho u \left(e + \frac{P}{\rho} \right) \right] = -\rho u \cdot \nabla \Phi, \quad (3.4)$$

where ρ is the mass density, u the fluid velocity, e the specific total energy, Φ the gravitational potential, and P the thermal pressure defined by:

$$P = (\gamma - 1) \left(e - \frac{1}{2} u^2 \right), \quad (3.5)$$

where γ is the adiabatic index. The energy equation defined in Eq. (3.4) conserves total fluid energy perfectly in the absence of gravity, ensuring no energy loss due to numerical errors. However, when gravity is included, the system is not strictly conservative, and the total energy conservation is maintained at the percent level. The equations are solved using a second-order Godunov scheme with a Riemann solver.

3.2.4 Time step control

As previously discussed, the time step for each level can be determined independently in RAMSES (Teyssier, 2002). Here, I will present the different stability criteria used to define the time step. In practice, the time step at each level is determined as the minimum of the following four constraints:

The first constraint comes from the gravitational evolution of the coupled N-body and hydrodynamical system. The time step difference Δt^l must be smaller than a fraction of the minimum free-fall time of the system.

The second constraint comes from the particle dynamics. Particles within the AMR grid should only move a fraction of the cell size during each time step.

The third constraint specifies that the expansion factor should not change by more than 10% over one-time step. This constraint is only applied during the early stages when gravitational clustering is linear.

The fourth constraint is imposed by the Courant Friedrichs Levy stability condition, which states that the time step should be less than:

$$\Delta t_{\text{CFL}}^l = cfl \times \frac{\Delta x^l}{\max(|u_x| + c_s, |u_y| + c_s, |u_z| + c_s)}, \quad (3.6)$$

where $cfl < 1$ is the Courant factor, Δx^l is the cell size at level l and $|u_x|$, $|u_y|$, $|u_z|$ are the velocity components in each direction, with c_s representing the speed of sound.

This procedure implies that the time step can be small in regions of high density or high velocity to capture the dynamics of the system accurately. However, this will also increase the computational cost of the simulation. In practice, we only save some time steps by specifying the scale factor we want to save.

3.2.5 Refinement strategy

The refinement strategy differs slightly between the N-body and hydrodynamic solvers in RAMSES (Teyssier, 2002), but they are functionally equivalent and based on a quasi-Lagrangian approach.

For the N-body solver, the refinement criterion ensures a constant number of particles per cell to control Poisson noise. Cells at refinement level l are refined if the dark matter density exceeds a level-dependent density threshold defined as:

$$\rho_l = M_c \times (\Delta x^l)^{\text{dim}}, \quad (3.7)$$

where M_c is the maximum mass (or number of particles) imposed per cell.

For the gas component, the density level-dependent threshold is:

$$\rho_l = M_b \times (\Delta x^l)^{\text{dim}}. \quad (3.8)$$

To ensure consistent Lagrangian evolution for both dark matter and gas, the typical baryonic mass per cell M_b is set as:

$$M_b = M_c \frac{\Omega_b}{\Omega_m - \Omega_b}. \quad (3.9)$$

Additionally, other refinement criteria based on gradients of the flow variables (such as pressure, density,...) or geometrical arguments are implemented in RAMSES. In practice, we define the maximum level of refinement we want to reach.

3.2.6 Baryonic processes

In the `RAMSES` code, different baryonic processes have been implemented and can be employed in the simulations. These processes include:

- Cooling (Sutherland & Dopita, 1993),
- Star formation (Rasera & Teyssier, 2006),
- Supernovae (Dubois & Teyssier, 2008b; Kimm & Cen, 2014),
- AGN (Dubois et al., 2012; Pfister et al., 2019; Massonneau et al., 2023),
- Magneto-hydrodynamics (MHD) (Teyssier et al., 2006; Fromang et al., 2006),
- Radiative transfer (Rosdahl et al., 2013),
- Cosmic rays (Dubois & Commerçon, 2016; Dubois et al., 2019),
- Thermal conduction (Dubois & Commerçon, 2016),
- Dust (Dubois et al., 2024),
- Tracer particles for baryon (Cadiou et al., 2019).

These processes can be used individually or in combination, depending on the specific requirements of the simulation. Additionally, users have the flexibility to modify or add modules to explore different physical scenarios.

To simulate the large-scale structures, the inclusion of cooling, star formation, AGN, and supernova feedback is the most important. These processes allow the formation of galaxies with properties that agree with observations. Supernovae contribute to the mass regulation of low-mass halos, whereas AGN feedback is crucial for regulating gas reservoirs within galaxies, thereby influencing the distribution of gas and affecting the properties of halos, especially at high masses. MHD is necessary for the study of the magnetic field. Radiative transfer affects the feedback of dwarf galaxies and reionization, for example. Cosmic rays also impact the feedback of galaxies. Thermal conduction is important for studying galactic disks. Dust is used to model spectral energy distribution. Finally, tracer particles are used to follow the baryonic component of the simulation.

3.3 Description of simulations

My work involves different simulations, which are described in the following. The simulations are the `Horizon` suite, which includes `Horizon-AGN`, `Horizon-noAGN`, `Horizon-Large`, and `L896_wCDM`, as well as the `Magneticum` simulations. Additionally, I describe briefly the `BAHAMAS` simulations, which are used to calibrate the `HMx` halo model.

The `Horizon-AGN` and `Horizon-noAGN` simulations were already completed, and I have analysed them. The `Horizon-Large` simulation was conducted by Yohan Dubois specifically for our work, and I have analysed its results. I slightly modified the `RAMSES` code to run the `L896_wCDM` simulations with the help of Yann Rasera and Yohan Dubois to modify the evolution of time, and I have analysed them. The `Magneticum` simulations were analysed by Pranjal Rajendra Singh, and I use the results to compare with the `Horizon` simulations. My thesis was a collaboration between the Institut d’Astrophysique de Paris and the University of Arizona. In this context, I collaborated with Pranjal Rajendra Singh, a PhD student at the University of Arizona for the collaboration.

3.3.1 The Horizon suite of simulations

In the following section, I will describe the `Horizon-AGN`, `Horizon-noAGN`, `Horizon-Large`, and `L896_wCDM` simulations. I will provide more detailed information on the `Horizon-AGN` simulation, as the other simulations are variants of this one.

The Horizon-AGN simulation

The `Horizon-AGN` simulation (Dubois et al., 2014) is a cosmological hydrodynamical simulation of $100 h^{-1} \text{Mpc}$ comoving volume, with 1024^3 dark matter particles, resulting in a resolution of $M_{\text{DM, res}} = 8.3 \times 10^7 M_{\odot}$. The simulation uses the AMR code `RAMSES` (Teyssier, 2002) (see Sect. 3.2 for a description of the code), which can refine up to a minimum cell size of $\Delta x \simeq 1 \text{kpc}$.

Initial conditions and refinement The initial conditions are generated with the `mpgrafic` software (Prunet et al., 2008), using the Zel’dovich approximation described in Sect. 3.1. The simulation contains 7 levels of refinement with a quasi-Lagrangian criterion based on the number of dark matter particles or density of baryonic mass. A new refinement level is created if there are more than 8 dark matter particles or if the total baryonic mass is 8 times the initial dark matter resolution.

Cosmology The cosmology is a standard Λ CDM cosmology compatible with WMAP-7 (Komatsu et al., 2011) with $\{\Omega_m, \Omega_\Lambda, \sigma_8, \Omega_b, n_s\} = \{0.272, 0.728, 0.81, 0.045, 0.967\}$ and $H_0 = 70.4 \text{ km s}^{-1} \text{ Mpc}^{-1}$. Multiple redshifts between $z = 0.018$ and $z = 38.3$ are available, allowing us to perform redshift space analysis. Below, I summarise the main aspects of the physics included in the simulation.

Cooling and heating The simulation incorporates gas cooling via hydrogen and helium cooling, with a contribution from metals (Sutherland & Dopita, 1993), and a uniform UV background (Haardt & Madau, 1996) with redshift of reionization $z_r = 10$. Metallicity is treated as a passive variable for the gas, and its amount is modified by the injection of gas ejecta during supernova explosions and stellar winds. The release of various chemical elements (O, Fe, C, N, Mg, Si) synthesized in stars and released by stellar winds, and supernova are also accounted for. The gas follows an equation of state for an ideal monoatomic gas with an adiabatic index $\gamma = 5/3$.

Star formation and stellar feedback Star formation follows a Schmidt law with 2% star formation efficiency, and the associated feedback from stellar winds, type II and type Ia supernovae (Dubois & Teyssier, 2008a). Star formation occurs in regions where the gas hydrogen number density exceeds $n_0 = 0.1 \text{ Hcm}^{-3}$, following a Poissonian random process (Rasera & Teyssier, 2006; Shandarin & Zel’dovich, 1989). To prevent excessive gas fragmentation and mimic the effect of stellar heating on the mean temperature of the interstellar medium, gas pressure is artificially enhanced above a threshold n_0 , assuming a polytropic equation of state $T = T_0 (n/n_0)^{\Gamma-1}$ with a polytropic index $\Gamma = 4/3$. (Springel & Hernquist, 2003).

AGN feedback The simulation includes feedback from active galactic nuclei (AGN) powered by Bondi-Hoyle-Lyttleton accretion limited at Eddington with jet/radio or heating/quasar mode depending on the accretion rate relative to Eddington (Dubois et al., 2012). Black holes are created where the gas mass density exceeds ρ_0 , with an initial seed mass of $10^5 M_\odot$. To prevent the formation of multiple black holes in the same galaxy, black holes cannot form at distances less than 50 kpc.

Projected maps The left panel of Fig. 3.1 shows the logarithm of the pressure, with the densest regions in red and less dense in blue. The right panel shows the logarithm of the matter, with the densest regions in yellow and less dense in blue. The maps represent one slice of $100 h^{-1} \text{ Mpc}$ in height and width and $10 h^{-1} \text{ Mpc}$ in thickness of the `Horizon-AGN` simulation. From these projection maps, we clearly see the cosmic web structure. We also see that the pressure is a tracer of matter, even if it is more diffuse than the matter.

The `Horizon-noAGN` simulation

The `Horizon-noAGN` simulation (Dubois et al., 2016) shares the same initial conditions, sub-grid modeling, and cosmology as the `Horizon-AGN` simulation, only the physics is different. This simulation contains no black hole formation and, therefore, no AGN feedback. This leads to a significant overshoot of the baryonic mass content in galaxy groups and clusters, and in particular of their gas fraction, at the high-mass end (Beckmann et al., 2017; Chisari et al., 2018). This simulation can be used to study the impact of AGN feedback on different observables.

The `Horizon-Large` simulation

The `Horizon-Large` simulation is a cosmological hydrodynamical simulations of $896 h^{-1} \text{ Mpc}$ comoving volume, with 1024^3 dark matter particles, leading to a resolution of $M_{\text{DM, res}} = 6 \times 10^{10} M_\odot$. The grid is allow to refine up to a spatial resolution of 10 kpc. The cosmology is the same as the `Horizon-AGN` and `Horizon-noAGN` simulations, but the initial conditions are different. The physics is simpler than the two other simulations, it only contains gas cooling and UV background heating below z_r and no galactic physics. We run this simulation to have a similar box size to the `Magneticum` simulation described below. Removing metals, supernova feedback, and star formation with respect to the `Horizon-noAGN` simulation implies less cooling.

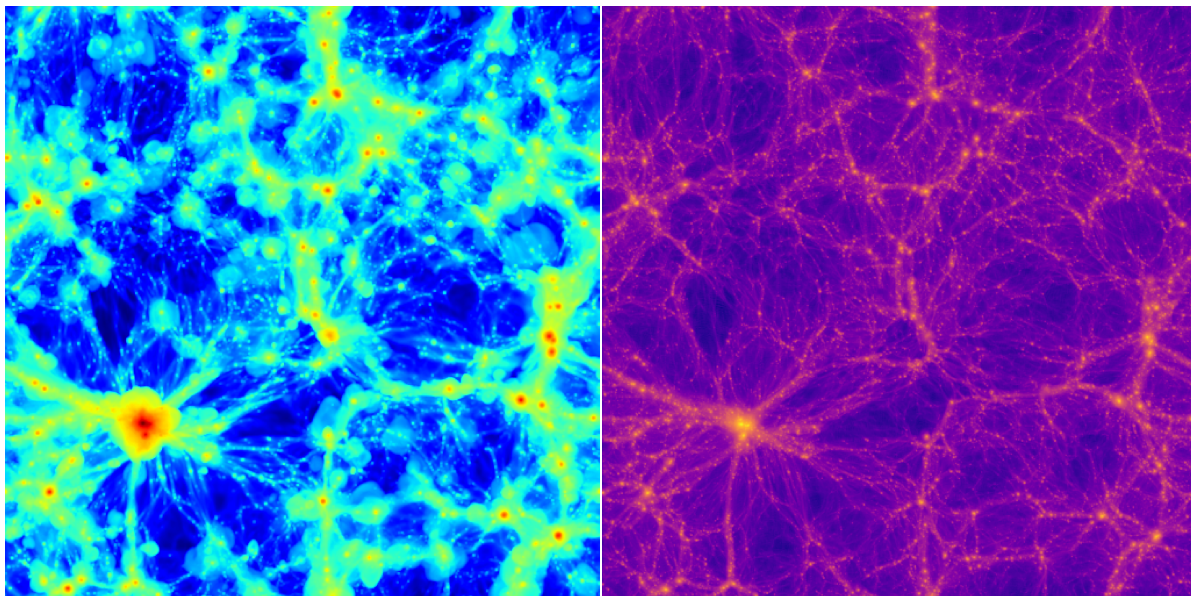


Figure 3.1: Left: logarithm of the pressure, with the densest regions in red and less dense in blue. Right: logarithm of the matter, with the densest regions in yellow and less dense in blue. Maps show one slice of $100 h^{-1}\text{Mpc}$ in height and width and $10 h^{-1}\text{Mpc}$ in thickness of the `Horizon-AGN` simulation.

The L896_wCDM simulations

The `L896_wCDM` simulations are a suite of three cosmological hydrodynamical simulations of $896 h^{-1}\text{Mpc}$ comoving volume, with 1024^3 dark matter particles, leading to a resolution of $M_{\text{DM, res}} = 6 \times 10^{10} M_{\odot}$. These simulations adopt the same sub-grid modeling as the `Horizon-Large` simulations, but differ in their initial conditions, which are generated using 2LPT, described in Sect. 3.1, instead of the Zel’dovich approximation. The simulations contain 7 levels of refinement with the same refinement condition as the one of the `Horizon-AGN` simulation.

Cosmology The cosmology of these simulations follow a w CDM cosmology with $\{\Omega_m, \Omega_{\Lambda}, \sigma_8, \Omega_b, n_s\} = \{0.272, 0.728, 0.81, 0.045, 0.967\}$ and $H_0 = 70.4 \text{ km s}^{-1} \text{ Mpc}^{-1}$. The simulations explore three different constant equations of state of dark energy: $w = -0.8$, $w = -1$ (compatible with Λ CDM), and $w = -1.2$. They are normalised such that σ_8 is the same for the three simulations.

Motivation These simulations were specifically run to start an investigation into the influence of dark energy on the tSZ effect. While cosmological constraints generally favour values of w closer to -1 (Planck Collaboration et al., 2020b), I chose these distinct w values to emphasize their impact. The constant equation of state of dark energy alters the evolution of the scale factor $a(t)$, which affects the passage of time in the simulations. The evolution of the scale factor as a function of the lookback time used in these simulations is shown in Fig. 3.2. The evolution for $w = -0.8$, $w = -1$, and $w = -1.2$ are shown in blue, black, and red, respectively. Using different values of w alters the age of the Universe and the slope of the scale factor, mainly in the past. For a given scale factor, the lookback time is higher for $w = -1.2$ than for $w = -1$ and lower for $w = -0.8$.

3.3.2 The Magneticum simulations

The `Magneticum` suite of simulations (Dolag et al., 2016) are cosmological hydrodynamical simulations with different box sizes and cosmologies. In this work we use the medium resolution `Box1a` simulation of $896 h^{-1}\text{Mpc}$ comoving volume, with 1512^3 dark matter and (initial) gas particle masses of $1.3 \times 10^{10} h^{-1} M_{\odot}$ and $2.6 \times 10^9 h^{-1} M_{\odot}$, respectively. For analysing the properties of lower mass halos (see Sect. 4.3.5), we use `Box2` which has a smaller volume of $352 h^{-1}\text{Mpc}$ comoving volume but a better mass resolution of $6.9 \times 10^8 h^{-1} M_{\odot}$ and $1.4 \times 10^8 h^{-1} M_{\odot}$ for dark matter and gas particles, respectively. The simulations were performed using the higher order SPH kernels (Dehnen & Aly, 2012) within the `P-GADGET3` code (Springel et al., 2005). The boxes that we are using also follow a WMAP-7 cosmology (Komatsu et al., 2011) with $\{\Omega_m, \Omega_b, \sigma_8, h, n_s\} = \{0.272, 0.0456, 0.809, 0.704, 0.963\}$. We also have access to the redshifts between 0 and 5. We summarise here the main physical aspects. The simulation includes

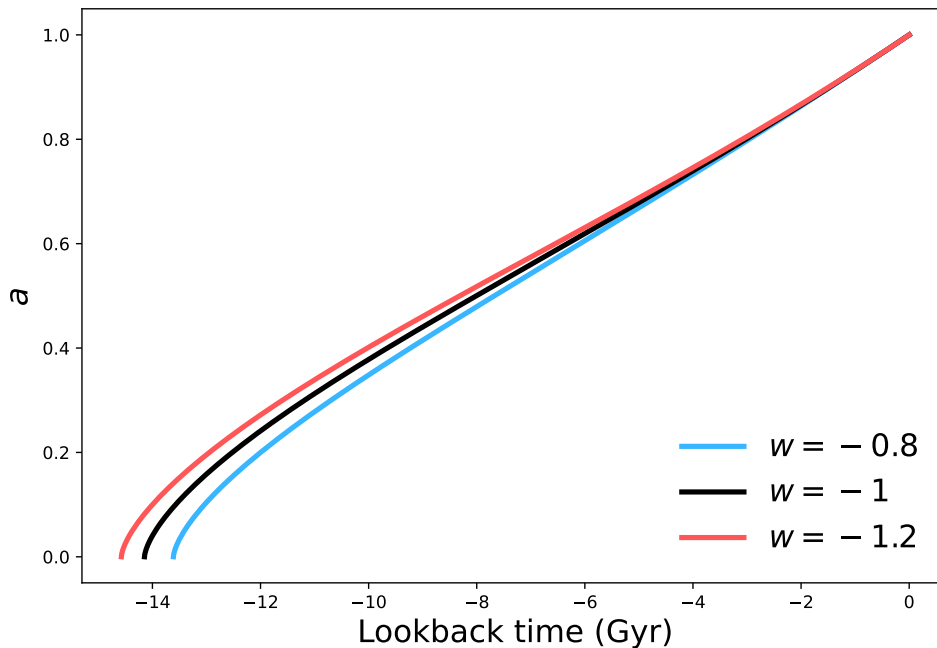


Figure 3.2: Evolution of the scale factor a as a function of the lookback time (in Gyr) used in the L896_ Λ CDM simulations. The evolution for $w = -0.8$, $w = -1$, and $w = -1.2$ are shown in blue, black, and red, respectively.

gas cooling, star formation, and winds (Springel & Hernquist, 2003). The metals, stellar population, and chemical enrichment, SN-Ia, SN-II, AGB follow Tornatore et al. (2003, 2006) with the new cooling tables of Wiersma et al. (2009). The thermal conduction follows Dolag et al. (2004), and the low viscosity scheme follows Dolag et al. (2005) and Beck et al. (2016). Magnetic fields are present as passive variables following Dolag & Stasyszyn (2009). There are also black holes and AGN feedback (Hirschmann et al., 2014). I use these simulations to compare with the Horizon simulations. As the implementation of the physics and the solver are different, we cannot predict the difference between the simulations.

3.3.3 The BAHAMAS simulations

The BAHAMAS simulations are a suite of hydrodynamical simulations of $400 h^{-1}$ Mpc with the WMAP-9 (Hinshaw et al., 2013) and Planck 2013 (Planck Collaboration et al., 2014) cosmology. The simulations contain 2×10^{24} particles leading to a resolution of $M_{\text{baryon, res}} = 3.85 \times 10^9 h^{-1} M_{\odot}$ ($M_{\text{baryon, res}} = 4.45 \times 10^9 h^{-1} M_{\odot}$) and $M_{\text{DM, res}} = 7.66 \times 10^8 h^{-1} M_{\odot}$ ($M_{\text{DM, res}} = 8.12 \times 10^8 h^{-1} M_{\odot}$), respectively for a WMAP-9 (Planck) cosmology. The hydrodynamic code and subgrid physics are the same as the ones in the OWLS (Schaye et al., 2010) and cosmo-OWLS (Le Brun et al., 2014; McCarthy et al., 2014) projects. The simulations include radiative cooling and heating, with a reionization that occurs at $z_r = 9$, a star formation rate, and a stellar evolution and chemical enrichment. It also contains a black hole and AGN feedback with three strengths of AGN feedback, from the smaller to the bigger: $10^{7.6}$ K, $10^{7.8}$ K, and $10^{8.0}$ K. More details about the physics are available in Schaye et al. (2010). I do not directly work with these simulations, but they are used to calibrate the HMx halo model, described in Sect. 4.1, that I am using. As the implementation of the physics and the solver are different, we cannot predict the difference between the simulations.

3.4 Halo mass function of the simulations

An important statistical property of the halos is their abundance, that is the number density of halos $dN/d\ln(M)$ within a logarithmic mass bin. The mass of the highest-mass halo will depend on the size of the simulations, while the lower-mass halos will be affected by the physics included in the simulations and by their merging history and local environment.

The abundance of massive clusters is highly sensitive to cosmology, particularly to the amplitude of the matter power spectrum, and can thus serve as a valuable cosmological probe.

In Fig. 3.3, I present the halo mass function of the different simulations I have worked with (Horizon-AGN, Horizon-noAGN, Horizon-Large, and Magneticum) in different colours (red, purple, green and blue) with

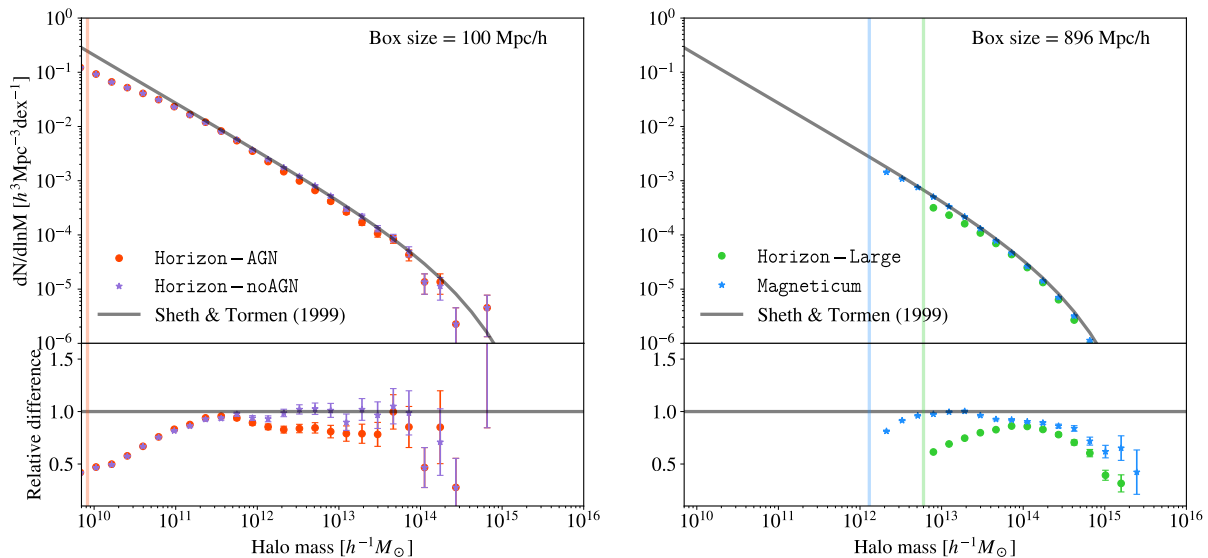


Figure 3.3: Halo mass function at $z = 0$ for the `Horizon-AGN`, `Horizon-noAGN`, `Horizon-Large`, and `Magneticum` simulations in red, purple, green, and blue, respectively, with their associated Poisson noise as error bar. Each mass function is computed up to the maximal halo mass in the corresponding simulation. The analytical mass function from [Sheth & Tormen \(1999\)](#) is shown in black for comparison. Vertical lines indicate the minimum halo mass accessible in each simulation, defined as 100 times the dark matter mass resolution. The lower limits for `Horizon-AGN` and `Horizon-noAGN` are identical.

their associated Poisson noise as error bar compared with the theoretical halo mass function from [Sheth & Tormen \(1999\)](#) (in black), which is the one used in the HMx halo model. For the `Horizon` simulations, the halos are identified with the `adaptaHOP` halo finder ([Aubert et al., 2004](#)), and we only keep the halos (and subhalos) with at least 100 particles. `adaptaHOP` find the structures using the peaks in the 3D density field and compute the virial mass and radius using only dark matter particles. For `Magneticum`, the halos are identified with a standard Friends-of-Friends algorithm and subhalos with the `subfind` module ([Springel et al., 2001](#); [Dolag & Stasyszyn, 2009](#)). The virial mass definition used is:

$$M_v = \frac{4\pi}{3} r_v^3 \Delta_v \rho_c(z), \quad (3.10)$$

where $\rho_c(z) = 3H(z)^2/(8\pi G)$ is the critical density of the Universe at redshift z , G is the gravitational constant, $H(z)$ is the Hubble expansion factor, and where Δ_v comes from the Λ CDM fitting function of [Bryan & Norman \(1998\)](#):

$$\Delta_v = 18\pi^2 + 82x - 39x^2, \quad (3.11)$$

where $x = \Omega(z) - 1$, with $\Omega(z) = \Omega_{m0}(1+z)^3/E(z)^2 = \Omega_{m0}(1+z)^3/(\Omega_{m0}(1+z)^3 + \Omega_{K0}(1+z)^2 + \Omega_\Lambda)$ if $\Omega_{r0} = 0$ (see Eq. (1.10) for the definition of $E(z)$). This is the same definition of virial mass as in [Mead et al. \(2020\)](#).

We see a relatively good agreement between the different simulations, nevertheless, they all have different maximal and minimal masses because of the difference in volume and mass resolution. We observe a lack of high-mass halos in the bigger simulations (`Horizon-Large` and `Magneticum`) and of low-mass halos in all the simulations.

Additionally, in Fig. 3.4, I present the halo mass function at $z = 0.02$ (left) and $z = 2.72$ (right) of the `L896_wCDM` simulations for $w = -0.8$, $w = -1$, and $w = -1.2$ in blue, black, and red, respectively. The fractional difference is computed with respect to the $w = -1$ simulation. The comparison between the three `L896_wCDM` simulations shows that the mass function is sensitive to the equation of state of dark energy. The $w = -0.8$ simulation has more halos of every mass, including massive halos, compared to the $w = -1$ simulation, which in turn has more halos than the $w = -1.2$ simulation. At $z = 0.02$, the difference increases with the mass of the halos. For halos below $\sim 10^{15} h^{-1} M_\odot$, there is less than 5% differences between the $w \neq -1$ simulations and $w = -1$ simulation, but for halos above $\sim 10^{15} h^{-1} M_\odot$, the difference can reach up to 10%. At $z = 2.72$, the differences are more important, with between $\sim 20\%$ and more than a 100% difference. It seems that the difference increases with the mass, but the trend is less clear as it starts to decrease at the highest masses. This can be due to the low number of massive halos. For both redshifts, there is an asymmetry in the difference: the fractional difference between the $w = -0.8$ and $w = -1$ simulations is larger than between the $w = -1.2$ and $w = -1$ simulations.

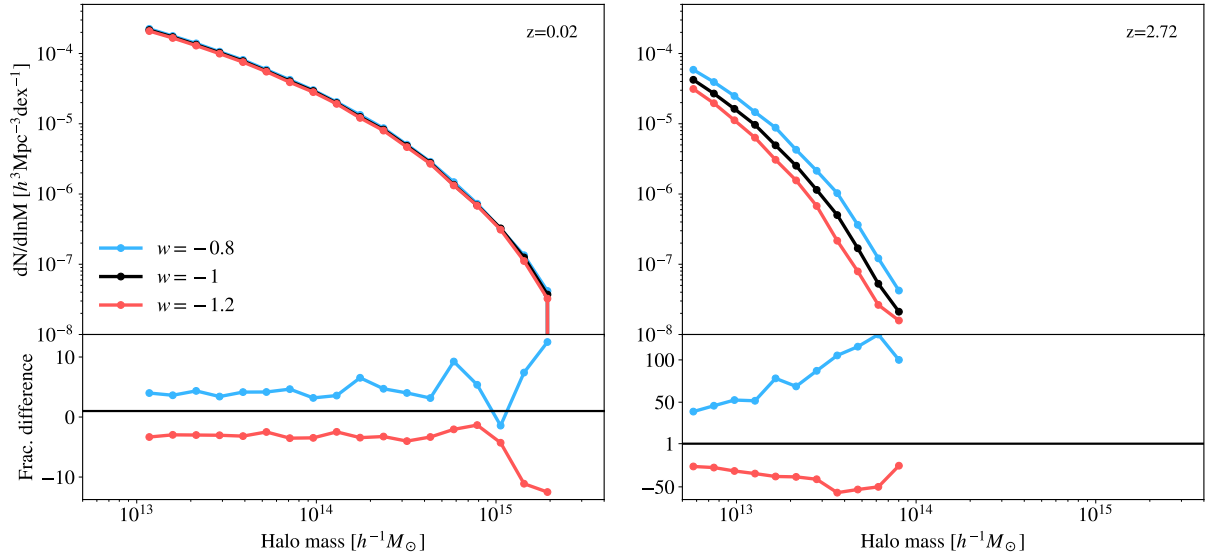


Figure 3.4: Halo mass function at $z = 0.02$ (left) and $z = 2.72$ (right) of the L896_wCDM simulations. The simulation with $w = -0.8$ is represented in blue, with $w = -1$ in black and with $w = -1.2$ in red. The fractional difference is computed with respect to the $w = -1$ simulation. Note: the y-axis for the fractional difference does not show the same range for the two redshifts to emphasize the differences present at each redshift. Note: these halo mass functions are computed with a version of the simulations which contain 4 levels of refinement.

The differences observed in the mass function can impact the predicted spectrum of the tSZ effect, as the tSZ effect is sensitive to the number of halos and their properties. The dependence of the spectrum on the cosmology will be studied in Sect. 4.5.

Part III

Results and publications

CHAPTER 4

THERMAL SUNYAEV-ZEL'DOVICH PROPERTIES

4.1	The H_{Mx} halo model	55
4.1.1	Theoretical modelisation	55
4.1.2	Angular power spectrum prediction	58
4.2	Thermal Sunyaev-Zel'dovich effect observations	59
4.2.1	Angular power spectrum on wide fields	59
4.2.2	Cross-correlation with other observables	59
4.2.3	Pressure profile	62
4.3	Comparison with measurement in simulations	63
4.3.1	Methodology	63
4.3.2	Mass cut	64
4.3.3	Power spectrum comparison	64
4.3.4	Angular power spectrum comparison	66
4.3.5	Halo model consequences	67
4.3.6	Halo pressure profiles	69
4.3.7	Difference of the simulations	70
4.4	Measurement of some properties	72
4.4.1	Properties in the simulations	72
4.4.2	Implication for the power spectrum and pressure profile	75
4.5	Beyond standard cosmology	78
4.5.1	Results on the power spectrum	78
4.5.2	Result on the pressure angular power spectrum	78
4.5.3	Discussion	79
4.6	Summary	82

As previously discussed, the thermal Sunyaev-Zel'dovich (tSZ) effect serves as a probe of baryonic physics and can play a crucial role in resolving degeneracies among cosmological parameters. Thus, achieving an accurate modelisation of the tSZ is needed for its effective utilisation. This will be explored with the discussion in Sect. 4.2.2 about cross-correlation with different probes, in particular, the 10x2pt analysis done in Fang et al. (2024), which is an article to which I contributed. In the following chapter, I will present a comparative analysis between the analytical modelisation of the tSZ effect and measurements obtained from various simulation suites. Sects. 4.1 and 4.3 are largely inspired by Aygoberry et al. (2024b), and some parts of the paper are reproduced here, with the addition of details and other computations. The analysis of the *Magneticum* simulations has been done by Pranjal R.S., while the analyses of the *Horizon-AGN*, *Horizon-noAGN*, *Horizon-Large*, and *L896_wCDM* simulations have been done by myself.

First, I will present the analytical modelisation of the tSZ effect as done by HMx in Sect. 4.1. Then, I will present the cosmological and astrophysical constraints that can be derived from tSZ observations in Sect. 4.2. I will continue in Sect. 4.3 by comparing the tSZ properties measured in simulations and those predicted by HMx, encompassing aspects such as the power spectrum and profile. In Sect. 4.4, I will compare different components of HMx with measurements from simulations. In Sect. 4.5, I will explore the impact of changing the equation of state of dark energy using the *L896_wCDM* simulations. Finally, in Sect. 4.6, I will present a summary and conclusion of this chapter.

4.1 The HMx halo model

In Sect. 2.3, I presented the general description of a halo model. In this section, I will focus on the HMx halo model (Mead et al., 2020), a specific halo model implementation that allows to predict the tSZ power spectrum. The first implementation of the tSZ power spectrum was developed using the Press-Schechter formalism (e.g., Komatsu & Kitayama, 1999; Refregier et al., 2000; Seljak et al., 2001), which only accounts for the one-halo term. The two-halo term was first introduced by Komatsu & Kitayama (1999). These early models evolved into halo model more aligned with the generic framework presented in Sect. 2.3. For instance, Refregier & Teysier (2002) conducted the first comparison between such a model and measurements from simulations.

Over the past years, different halo models for the tSZ power spectrum have been implemented to improve the agreement between models and simulations. One widely used halo model (e.g., Tröster et al., 2022) is the HMx halo model, which I will use for my analysis. The initial version of HMx, called *hmcode*, (Mead, 2015) was developed to predict the matter power spectrum. This model was improved, and I am using the version from Mead et al. (2020) that can predict power spectrum for different tracers, including matter, pressure, CIB,...

I will compare the prediction and components of the HMx halo model with measurements in simulations (*Horizon-AGN*, *Horizon-noAGN*, *Horizon-Large* and *Magneticum*). I recall here the principal components and behaviours inherent to the model. Furthermore, I will detail how to derive the angular power spectrum by integrating the power spectrum over redshift.

4.1.1 Theoretical modelisation

Halo mass function and linear bias

The mass definition used in HMx is the one described in Sect. 3.4, and I remind it here for clarity:

$$M_v = \frac{4\pi}{3} r_v^3 \Delta_v \rho_c(z), \quad (4.1)$$

where $\rho_c(z) = 3H(z)^2/(8\pi G)$ is the critical density of the Universe at redshift z , G is the gravitational constant, $H(z)$ is the Hubble expansion factor, and where Δ_v comes from the Λ CDM fitting function of Bryan & Norman (1998):

$$\Delta_v = 18\pi^2 + 82x - 39x^2, \quad (4.2)$$

where $x = \Omega(z) - 1$, with $\Omega(z) = \Omega_{m0}(1+z)^3/E(z)^2 = \Omega_{m0}(1+z)^3/(\Omega_{m0}(1+z)^3 + \Omega_{K0}(1+z)^2 + \Omega_\Lambda)$ if $\Omega_{r0} = 0$ (see Eq. (1.10) for the definition of $E(z)$).

The halo model then translates mass to peak height ν following the procedure described in Sect. 2.3.4. The mass function $g(\nu)d\nu$ adopted in HMx is the one from Sheth & Tormen (1999):

$$g(\nu)d\nu = A \left[1 + \frac{1}{(q\nu^2)^p} \right] \exp(-q\nu^2/2)d\nu, \quad (4.3)$$

with $p = 0.3$, $q = 0.707$, and $A \approx 0.216$.

The linear halo bias function $b(\nu)$ is derived from the mass function using the peak-background split formalism (Mo & White, 1996; Sheth et al., 2001):

$$b(\nu) = 1 + \frac{1}{\delta_c} \left[q\nu^2 - 1 + \frac{2p}{1 + (q\nu^2)^p} \right]. \quad (4.4)$$

Electron pressure profile

For the tSZ power spectrum, the field u appearing in Eq. (2.19) is the electron pressure. For tSZ in general, the electron pressure profiles are derived from theoretical arguments (e.g., Komatsu & Seljak, 2001) or fitting to observational data and simulations (e.g., Arnaud et al., 2010). In the HMx framework, the electron pressure is defined as

$$P_e(M, r) = \frac{\rho_{\text{bnd}}(M, r)}{m_p \mu_e} k_B T_g(M, r), \quad (4.5)$$

where ρ_{bnd} denotes the halo bound gas density, representing the gas contained within the radius r_v (the virial radius encapsulating a mean gas density Δ_v times the critical density), m_p is the proton mass, μ_e is the mean gas-particle mass per electron divided by the proton mass (HMx uses $\mu_e = 1.17$), and k_B is the Boltzmann constant. The gas initially associated with the initial density of the corresponding halo underwent some feedback processes that can redistribute the gas. Part of this gas stays within the virial radius and is the bound gas, and part of this gas is now located outside r_v . The more important is the feedback, the more gas will be ejected. This ejected gas is incorporated into the two-halo term only, assuming it does not contribute to the one-halo term, consistent with the approach taken by Fedeli (2014) and Debackere et al. (2020).

The halo bound gas density ρ_{bnd} is defined according to the density profile of Komatsu & Seljak (2001):

$$\rho_{\text{bnd}} \propto \left[\frac{\ln(1 + r/r_s)}{r/r_s} \right]^{1/(\Gamma-1)}, \quad (4.6)$$

where Γ is the polytropic index. Each components of the model are described by a spherical density profile $\rho(M, r)$ and is normalised. Applied to the bound gas, we have

$$f_{\text{bnd}}(M)M = \int_0^{r_v} 4\pi r^2 \rho_{\text{bnd}}(M, r) dr, \quad (4.7)$$

which can be inverted to obtain the normalisation factor of Eq. (4.6). The fraction of bound gas is required and can be obtained from Schneider & Teyssier (2015):

$$f_{\text{bnd}}(M) = \frac{\Omega_b}{\Omega_m} \frac{(M/M_0)^\beta}{1 + (M/M_0)^\beta}, \quad (4.8)$$

where M_0 allows the transition from the universal baryon fraction in high-mass halos to zero in lower-mass halos, while β governs the transition rate.

The model assumes that all the gas is ionized, with the gas temperature T_g determined by the Komatsu & Seljak (2001) profile:

$$T_g(M, r) = T_v(M) \frac{\ln(1 + r/r_s)}{1 + r/r_s}, \quad (4.9)$$

which assumes hydrostatic equilibrium, where T_v denotes the virial temperature:

$$\frac{3}{2} k_B T_v(M) = \alpha \frac{GM m_p \mu_p}{r_v}, \quad (4.10)$$

where μ_p is the mean gas-particle mass divided by the proton mass (HMx uses $\mu_p = 0.61$), α encapsulates deviations from a virial relation and thus serves as a hydrostatic bias and r_s is the halo scale radius parameter, determined via the concentration relation $c_M = r_v/r_s$.

The dark-matter concentration used in HMx is the Duffy et al. (2008) concentration:

$$c_D(M) = 7.85 \left(\frac{M}{2 \times 10^{12} h^{-1} M_\odot} \right)^{-0.081} (1+z)^{-0.71}. \quad (4.11)$$

Based on the discussion in Sect. 2.3.4 about the impact of baryons on the concentration, the HMx implementation modifies the Duffy concentration by introducing two parameters ϵ_1 and ϵ_2 to account for the influence of baryons on the concentration. The concentration-mass relation employed becomes:

$$c_M(M) = c_D(M) \left[1 + \epsilon_1 + (\epsilon_2 - \epsilon_1) \frac{f_{\text{bnd}}(M)}{\Omega_b/\Omega_m} \right], \quad (4.12)$$

where ϵ_1 and ϵ_2 allow to capture the influence of baryons, specifically caused by the ejection of gas. In this modified concentration, we recognize Ω_b/Ω_m , which is the quantity of baryons, which allows us to rescale the dark-matter concentration to a baryonic one. It is also rescaled by the fraction of bound gas to account for the fact that within one virial radius, only the bound gas needs to be accounted for. This normalisation $\frac{f_{\text{bnd}}(M)}{\Omega_b/\Omega_m}$ is greater than one, as we expect the baryon to be more concentrated than the dark matter. A halo that has lost all its gas will be multiplied by $(1 + \epsilon_1)$, while a halo that has kept all its gas will be multiplied by $(1 + \epsilon_2)$. Adding ϵ_1 and ϵ_2 allows the interpolation between dark matter and baryonic concentration and introduces an additional degree of freedom. As seen from Fig. 3.1, the pressure is more diffuse than the matter, and these parameters can allow us to capture that.

From the model, we discern that high-mass halos contribute the most, as low-mass halos are deficient in gas (as indicated by f_{bnd} in Eq. (4.8)). Additionally, the amplitudes of the one- and two-halo terms are more sensitive to high-mass halos. It can be explained by the fact that the electron pressure follows the gas density but originates primarily from the highest gas-density peaks, where elevated temperatures lead to enhanced electron pressure.

Power spectrum and angular power spectrum

Using the halo mass function, halo bias, and electron pressure profile defined previously, the equations (2.17) and (2.18) can be applied to obtain the pressure power spectrum.

To obtain the angular power spectrum, for the Compton- y parameter, the weight function appearing in Eq. (2.30) is:

$$X_y(\hat{r}) = \frac{\sigma_T}{m_e c^2} \frac{1}{a^2(\hat{r})}, \quad (4.13)$$

where a represents the expansion factor as a function of the comoving radius \hat{r} , σ_T is the Thompson scattering cross-section, m_e is the electron mass, and c is the speed of light. It is noteworthy that this weight function does not depend on redshift, unlike, for instance, weak lensing, which has a weight function that decreases quickly after $z = 1$. This characteristic of the tSZ effect is particularly advantageous as it enables to probe high redshifts.

Fitted parameters & different models

The model is characterized by different parameters, such as Γ , M_0 , β , α , ϵ_1 , and ϵ_2 encounter in Eqs. (4.6), (4.8), (4.10) and (4.12). With the addition of other parameters that appear in the model but not on the electron pressure implementation, these parameters are summarised in Table 4.1, with their default values on the second column. To improve the accuracy of the model, these parameters are fitted against the BAHAMAS simulations (McCarthy et al., 2017), briefly described in Sect. 3.3.

In detail, the fits are performed on the BAHAMAS simulation 3D power spectrum response. The parameters are fitted simultaneously for z between 0 and 1, with a linear weighting, and for k between 0.015 and $7 h^{-1}$ Mpc, with a logarithm weighting. Four different models are fitted:

- Stars: allow to fit M_* , A_* and η ,
- Matter: the parameters fitted for the star (M_* , A_* , η) are maintained fixed, and the parameters ϵ_1 , Γ , and M_0 are fitted,
- Matter & electron pressure: the parameters fitted for the stars (M_* , A_* , η) are maintained fixed, and the parameters ϵ_1 , Γ , M_0 , α , and T_w are fitted on matter and electron pressure jointly,
- Matter, CDM, gas & stars jointly: allow to fit M_* , A_* , η , ϵ_1 , Γ , and M_0 .

The fits have been done for the three strengths of feedback present in the BAHAMAS simulations. For the tSZ modelisation, I will use the value of the parameters fitted on the matter-electron pressure cross-power spectrum only, with the intermediate feedback strength. The value of the parameters is the third column of Table 4.1 and should not be taken as intrinsic properties of halos, as fits are done at the level of the response power spectrum. We see how the fitted values (including redshift dependencies in some cases) differ from their physics-inspired priors. In Mead et al. (2020), the authors argue that this approach provides the lowest error on the pressure auto-power spectrum, as it is difficult to obtain a robust fit on the pressure auto-power spectrum.

Using this model, the relative difference between the BAHAMAS simulation and prediction, averaged linearly over z between 0 and 1 and logarithmically over k between 0.015 and $7 h \text{Mpc}^{-1}$, is of 2% for the matter auto-power spectrum, 15% for the matter-pressure, 25% for the pressure auto-power spectrum. As can be seen, the prediction has a relatively low fidelity for the pressure auto-spectrum. Several reasons can explain this fact. The halo model makes several strong hypotheses, such as halos

Table 4.1: Halo model parameters adapted from Table 1 and Table 2 of Mead et al. (2020). The fitted values are the ones fitted with the matter-electron pressure model against the BAHAMAS simulations with an AGN heating temperature of $10^{7.8}$ K.

Parameter	Default value	Fitted value	Physical meaning
ϵ_1	0	$-0.1065 - 0.1073z$	Halo concentration modification for gas-poor halos
ϵ_2	0	N/A	Halo concentration modification for gas-rich halos
M_0	$10^{14} h^{-1} M_\odot$	$10^{13.5937} h^{-1} M_\odot$	Halo mass below which halos have lost more than half of their initial gas content
β	0.6	N/A	Low-mass power-law slope of halo bound gas fraction
Γ	1.17	1.1770	Polytropic index for the equation of state of gas that is bound in halos
A_\star	0.03	$0.0330 - 0.0088z$	Peak fraction of halo mass that is in stars
M_\star	$10^{12.5} h^{-1} M_\odot$	$10^{12.4479} \exp(-0.3521z)$	Halo mass of peak star-formation efficiency
σ_\star	1.2	N/A	Logarithmic width of star-formation efficiency distribution
η	-0.3	-0.3556	Power-law index for central-satellite galaxy split
α	1	0.8471	Ratio of halo temperature to that of virial equilibrium
T_w	$10^{6.5} \text{K}$	$10^{6.6545} \exp(-0.3652z)$	Temperature of the warm-hot intergalactic medium

trace the underlying linear matter distribution with a linear halo bias, halo profiles are perfectly spherical with no substructure and no scatter at fixed mass, and nothing prevents halos from overlapping. While these hypotheses provide a reasonable approximation of the physics at play in the case of the matter distribution, explaining the good precision in the case of the matter power spectrum and fair in the case of the cross-spectrum, they are not necessarily correct for the electron pressure distribution. It is still interesting to compare HMx with the simulation measurements as a first step to understanding and characterizing the key components of the model.

4.1.2 Angular power spectrum prediction

The angular power spectrum corresponds to an integration over redshift, and it is not useful to discuss differences occurring at redshifts that do not contribute significantly to the angular power spectra. For example, Komatsu & Seljak (2002) showed that the contribution of clusters at $z > 10$ is negligible to the pressure angular power spectrum. We can use the predictions from HMx to address this question. The pressure angular power spectrum obtained using different redshift ranges of integration is shown on the left panel of Fig. 4.1, where we look at the angular power spectrum when integrating up to different redshifts. Integration up to $z = 3$ or $z = 4$ captures more than 97% of the power for ℓ between 10 and 10^4 . Limiting ourselves to $z = 2$ will only lead to a dramatic loss of $\sim 17\%$ of validity after $\ell = 4 \times 10^3$. We have also looked at the contribution coming from the one-halo term when integrating up to a given z for different ℓ . Except when z is sufficiently small and ℓ large, corresponding to the interior of halos, more than 90% of the angular power spectrum comes from the one-halo term. This behaviour can be due to the change of slope in the pressure profiles of halos or to the fact that the contrast of pressure is smaller on the inner part of halos. These two conclusions imply that the modelisation of the electron pressure profiles up to $z = 4$ will impact our results and need to be well understood.

As discussed previously, the power spectra are dominated by high-mass clusters. However, depending on the simulation characteristics, the number of such high-mass objects in our simulations will be limited, which we need to take into account when comparing the model predictions with the simulations. As a first step in investigating the impact of high-mass objects, we show in the right panel of Fig. 4.1 the contribution of the different halo masses to the power spectra, as we vary the maximum mass considered in HMx (see also e.g., Refregier et al., 2000; Battaglia et al., 2012). Our reference here is the default maximum mass used in HMx: $M_{\max} = 10^{17} h^{-1} M_\odot$. We can note that the different ℓ are not affected in the same way; this choice of maximal mass impacts the most the values of $\ell = 50 - 60$. We can see that integrating up to $M_{\max} = 10^{16} h^{-1} M_\odot$ makes almost no differences. Using Gumbel statistics, Davis et al. (2011) found that it is very unlikely to have dark matter halos with $M > 10^{16} h^{-1} M_\odot$ within a volume

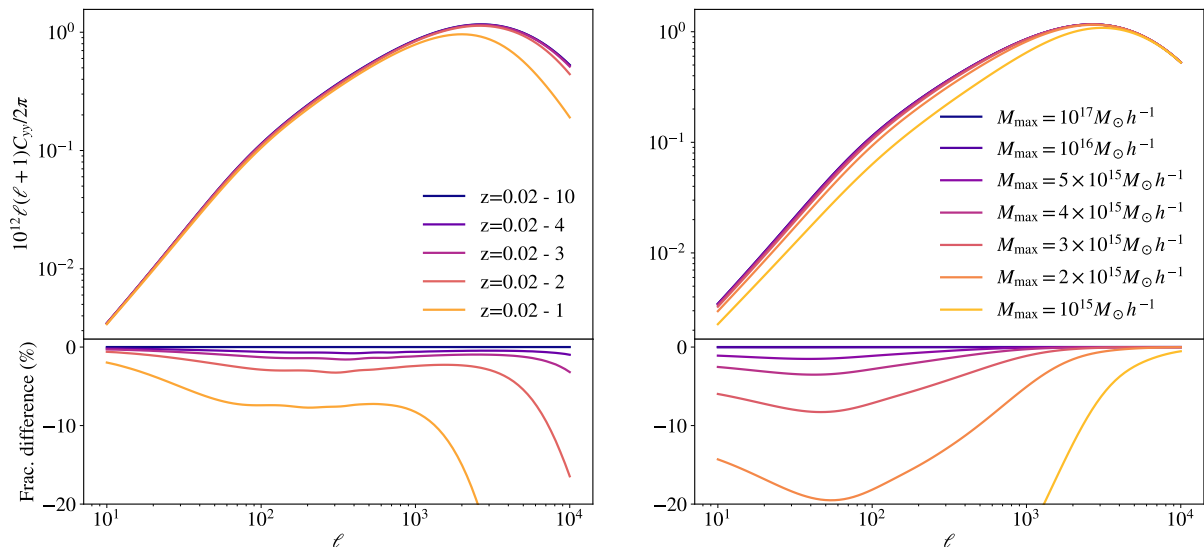


Figure 4.1: Predicted pressure angular power spectrum when integrating up to different redshifts in different colours (on the left) or between $z = 0.02$ and $z = 10$ for different maximal mass in different colours (on the right) as a function of the angular scale ℓ .

of $1 h^{-1}$ Gpc, we can extrapolate that it is also the case within the observable Universe. The prediction still changes a bit because the halo mass function predicts that it is possible to have such halos, but for analysis with real data, we will probably never use this maximal mass (Holz & Perlmutter, 2012). If we are integrating up to $M_{\max} = 2 \times 10^{15} h^{-1} M_{\odot}$ instead of $M_{\max} = 4 \times 10^{15} h^{-1} M_{\odot}$ we lose maximum 20% of the signal instead of a few percent. This implies that the masses of a few $10^{15} h^{-1} M_{\odot}$ must be well-modelled in our prescription.

4.2 Thermal Sunyaev-Zel'dovich effect observations

I describe the tSZ observable in Sect. 2.2.4. In this section, I will give more details about how the tSZ observations can be utilized to constrain cosmology or astrophysics.

4.2.1 Angular power spectrum on wide fields

The tSZ effect can be observed across the entire sky, allowing for the creation of tSZ maps using data from missions such as Planck (Planck Collaboration et al., 2016b; Chandran et al., 2023) or SPT (Bleem et al., 2022). Fig. 4.2 shows the tSZ map and angular power spectrum as measured by Planck Collaboration et al. (2016b). Notably, the amplitude of the Compton- y parameter is remarkably small, approximately $\sim 10^{-6}$, and the error bars on the angular power spectrum at low ℓ are rather large. With such an observable, it is then possible to explore higher-order statistics, such as bispectrum or one-dimension probability distribution function analysis, enabling constraints on cosmological parameters, as discussed in Planck Collaboration et al. (2016b). It is also possible to study the possible constraints obtained on cosmological and astrophysical parameters with the tSZ power spectrum from the full-sky CMB observations, such as those on primordial non-gaussianities f_{NL} and mass of the neutrinos M_{ν} (Hill & Pajer, 2013).

A robust modelisation of the tSZ is crucial for conducting component separation analyses such as the one conducted in Planck Collaboration et al. (2016a). Indeed, the tSZ is one of the foreground contaminants in the observation of the CMB. Other notable foregrounds include the cosmic infrared background (CIB), spinning dust, and others. Having robust modelisations for all the foreground components can allow the generation of cleaner CMB and foreground maps, which can then be employed for cosmological analyses. Using a halo model for the tSZ can be interesting, as it allows to employ similar framework to modelise other foregrounds such as the CIB or the kinetic SZ effect (kSZ). This unified approach allows cross-correlation analyses with a unique and coherent modelisation of the various effects.

4.2.2 Cross-correlation with other observables

To extract even more information, the tSZ effect is used in correlation with other probes. Numerous 3x2pt analyses have been conducted, which involve examining the auto- and cross-correlation between

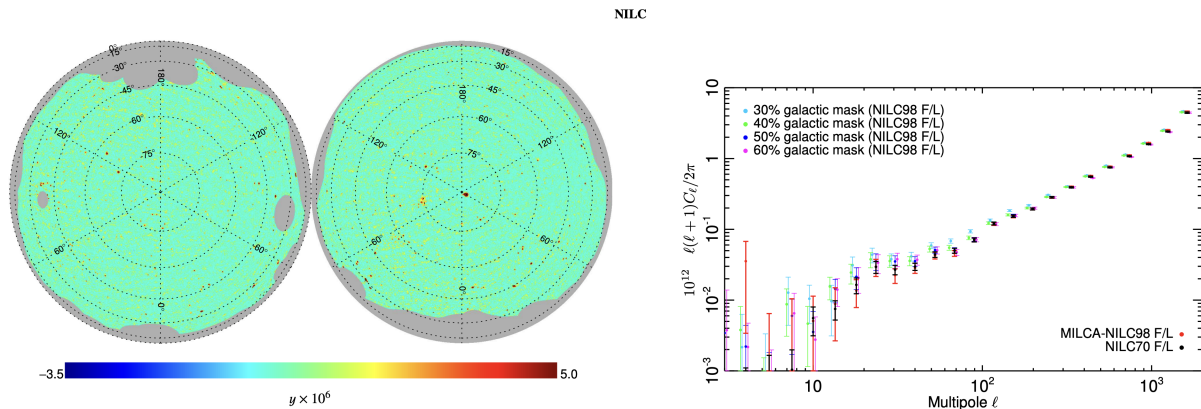


Figure 4.2: Figure adapted from Figs. 2 and 11 of [Planck Collaboration et al. \(2016b\)](#). Left: Reconstructed Planck all-sky Compton- y parameter maps for NILC in orthographic projections. Right: Angular cross-power spectra of the Planck NILC F/L reconstructed Compton- y parameter maps for different galactic masks.

the power spectra of two different fields. For example, the correlation between the tSZ signal and the gravitational weak lensing (WL) signal has been widely used to constrain both cosmological parameters and nuisance parameters. A few examples of such analyses include the one of [Van Waerbeke et al. \(2014\)](#) who were the first to report a cross-correlation between gravitational lensing and tSZ, highlighting the presence of a warm baryonic gas tracing the large-scale structure; [Ma et al. \(2015\)](#) who used the tSZ-WL correlation to constrain the diffuse baryonic distribution; [Osato et al. \(2020\)](#) who were able to constrain the hydrostatic mass bias of high redshift clusters and [Tröster et al. \(2022\)](#) who developed a pipeline using HMx to constrain cosmology and assess the impact of baryon for the first time with the KiDS-1000 and Planck and ACT surveys. Furthermore, the tSZ signal can also be correlated with other fields, such as in the study by [Maniyar et al. \(2021\)](#), which correlates the tSZ with the CIB. This correlation is significant as it is a foreground for kSZ analysis using the CMB power spectrum. Many other examples exist and are being studied, as the tSZ is a valuable probe of baryonic physics.

10x2pt analysis

In [Fang et al. \(2024\)](#), we conducted the first 10x2pt analysis to forecast outcomes for the joint [Simons Observatory](#)¹ and [Vera Rubin Observatory Legacy Survey of Space and Time](#)² using a coherent halo model. The 10x2pt analysis combines two-point functions (auto- and cross-power spectra) between fields of tSZ, weak lensing, CMB lensing, and galaxy clustering, resulting in an analysis with ten combinations. Our central objective was to evaluate the potential gains in cosmological constraints achievable by transitioning from 3x2pt (which includes only weak lensing and galaxy clustering) to 10x2pt analysis while employing a coherent and unified halo model.

In this analysis, we extend the existing 6x2pt modelling of `cosmolike` ([Eifler et al., 2014](#); [Krause & Eifler, 2017](#)) to include the tSZ auto- and cross-correlation. We employ a halo model framework for all observables to maintain a coherent model for all the 10x2pt statistics. For the tSZ, the halo model closely follows the HMx model of [Mead et al. \(2020\)](#) presented in Sect. 4.1.1. This model serves as a toy model to qualitatively explore information content. We performed a parameter inference on simulated data assuming a Gaussian likelihood and marginalized over the photo- z and shear calibration parameters.

Our objective was to assess the improvement in the constraining power when using different probes. In Fig. 4.3, we compare the constraints in the $S_8 - \Omega_m$ plane when using 3x2pt (containing only WL and clustering), 6x2pt (adding the CMB lensing) or 10x2pt statistics. By employing a 10x2pt analysis, the constraints improved by $\sim 70\%$ in the figure of merit (FoM) compared to 3x2pt and by $\sim 30\%$ (in the FoM) compared to 6x2pt. We also observed that the constraining power when using an 8x2pt (excluding the tSZ-clustering and tSZ-CMB lensing probes) is very similar to that of 10x2pt statistics. However, maintaining the 10x2pt pipeline is still beneficial as these probes can help to self-calibrate the systematic parameters. To further improve the constraints, small-scale modelling must be improved to reduce halo model uncertainties. We found that most of the uncertainty comes from the ϵ_1 , ϵ_2 (arising from the concentration defined in Eq. (4.12)), and Γ parameters (the polytropic index defined in Eq. (4.6)), which are added to the halo model due to the incorporation of the tSZ probe. On the other hand, the tSZ

¹<https://simonsobservatory.org/>

²<https://rubinobservatory.org/>

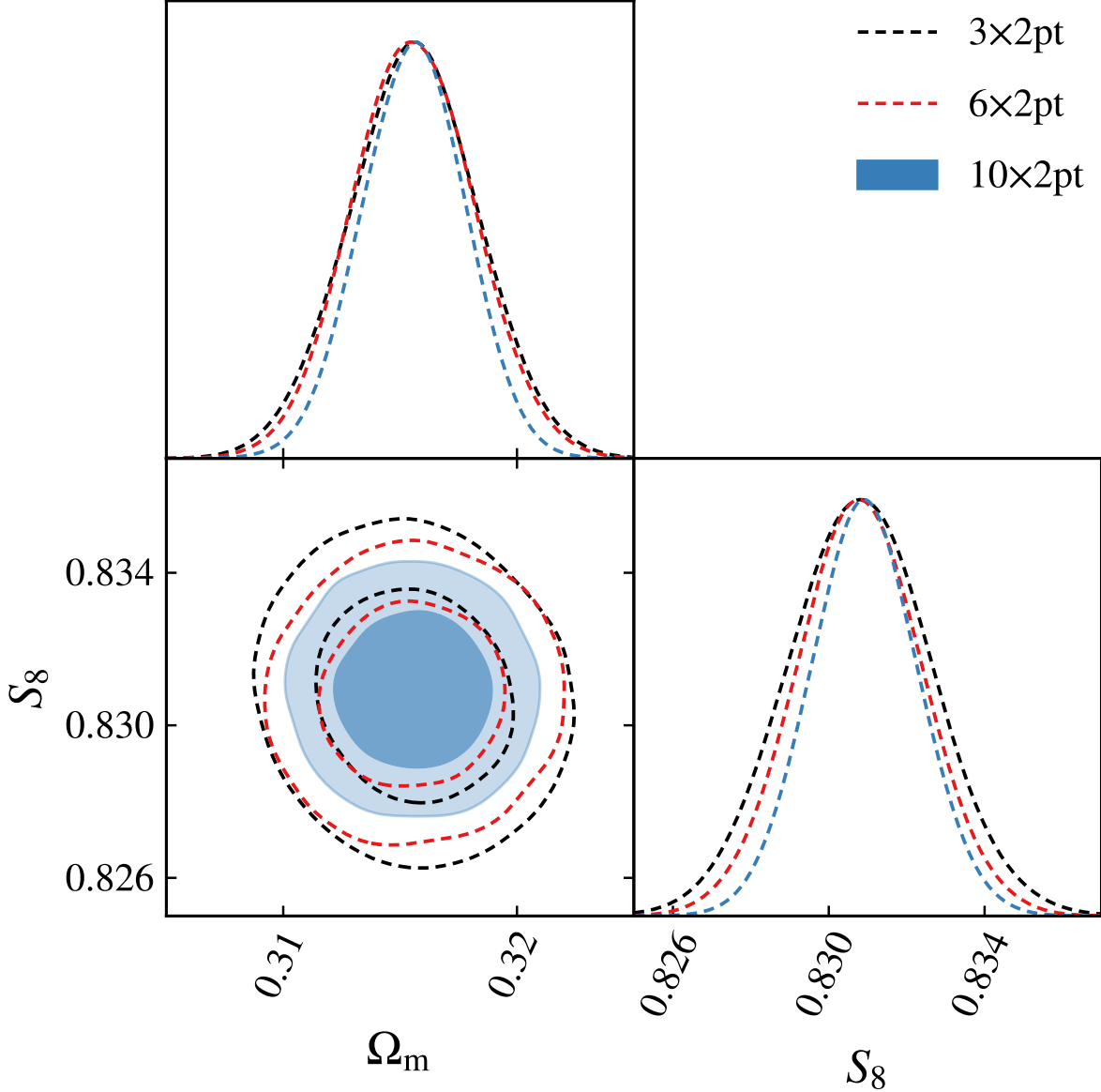


Figure 4.3: Increase in LSST Y6 + SO Y5 constraining power as the data vector increases from 3x2pt (black dashed) to 6x2pt (red dashed) to 10x2pt (blue shaded), for galaxy shear scale cuts $\ell_{\text{max}}^{\text{shear}} = 3000$ (Fang et al., 2024).

allows for a better self-calibration of the model, reducing uncertainties in the small-scale matter power spectrum.

More generally, the gain in the constraining power arises from the increased signal-to-noise ratio (SNR) and the self-calibration of the various nuisance parameters. The main conclusion of the paper is presented in Fig. 4.4 where we evaluate the improvement in the FoM in the Ω_m - S_8 plane and the required sky fraction needed to maintain a consistent SNR. Notably, we observed that when transitioning from a 3x2pt to a 10x2pt analysis (rescaling the sky fraction to maintain a similar SNR), we can gain a factor of two in the FoM. Additionally, adopting a 10x2pt analysis significantly reduces the required sky fraction by a factor of three, reaching 0.4. A sky fraction of 0.4 means that we need to observe $\approx 16,500 \text{ deg}^2$, or ≈ 1.15 times the area observed by the Euclid survey (the Euclid footprint covers $\approx 14,000 \text{ deg}^2$, which corresponds to a sky fraction of ≈ 0.35).

Digging deeper into the details, we found that most of the gain in constraining power comes from transitioning from 3x2pt to 6x2pt analysis, which involves adding the CMB lensing correlations (steps 1 and 2). Adding the auto tSZ power spectrum reduces the constraining power as additional nuisance parameters are required, and these parameters exhibit partial degeneracy with other cosmological parameters (step 3). Going from 6x2pt to 10x2pt analysis, we gain a bit of sky fraction at constant FoM because it enables self-calibration (step 4). Further improvement in the FoM, up to a factor of ~ 1.5 ,

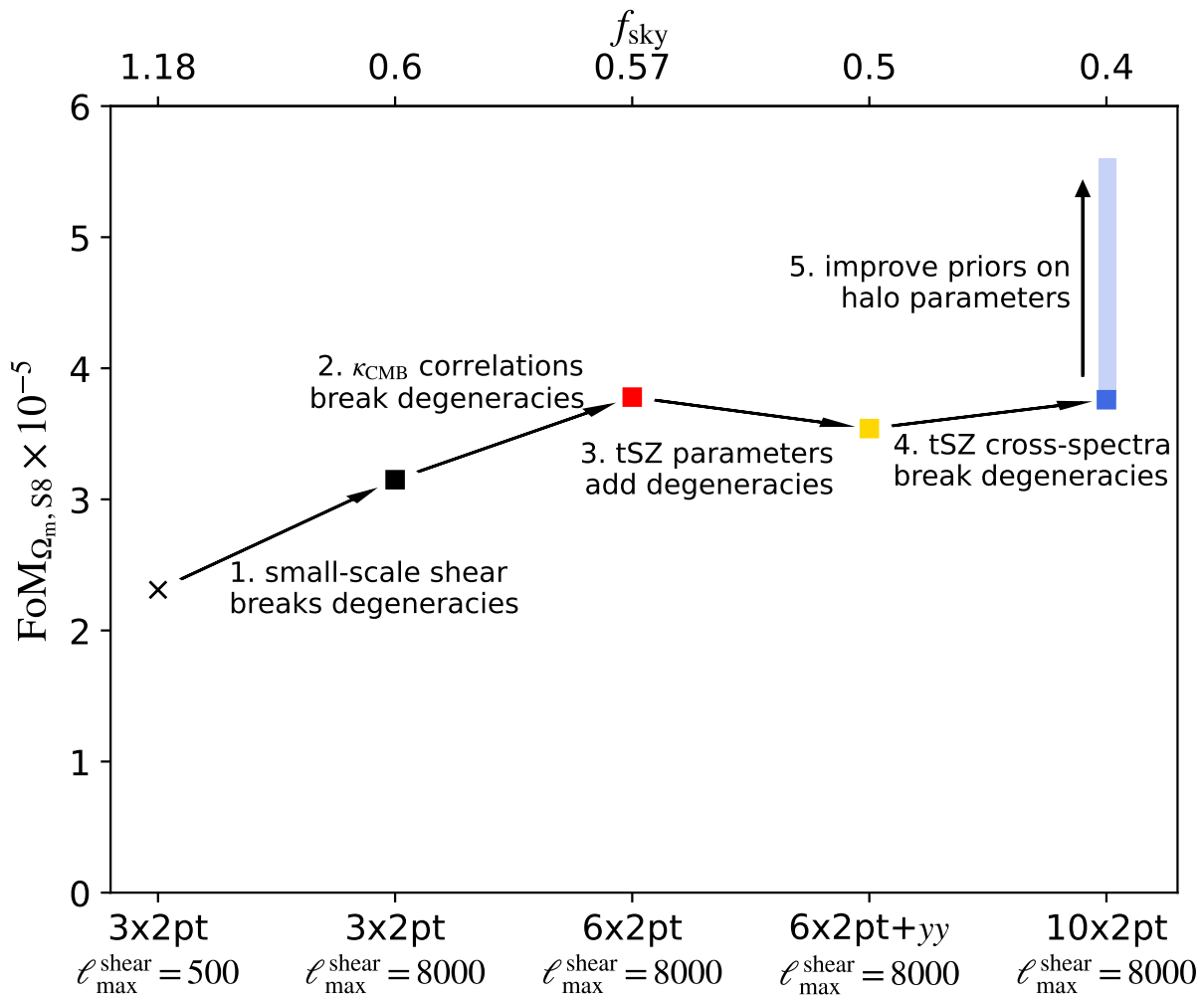


Figure 4.4: Evolution of the Figure-of-Merit in the Ω_m - S_8 plane and the required sky fraction needed to maintain a consistent signal-to-noise ratio when going from a 3x2pt to a 10x2pt analysis (Fang et al., 2024).

can be achieved by adding stronger priors on the halo model (step 5). The top of the blue vertical line represents the constraining power achievable if all the halo parameters are fixed.

In the context of this study, we found that refining priors and enhancing the modelling of the tSZ power spectrum is both important and timely. This effort holds the promise of tightening the constraints on cosmological parameters, potentially addressing issues such as the H_0 or S_8 tensions (e.g., McCarthy et al., 2023). Moreover, the implications of this investigation extend into the domain of baryonic feedback. By comparing the halo model with simulations, we aim to refine our understanding of the baryonic modelling within halos. This refinement is crucial for capturing the underlying physics of the model and improving its fidelity to observational data.

4.2.3 Pressure profile

I have discussed that tSZ observations can be conducted on wide fields, but it is also possible to observe individual halos. It enables the study of the evolution of clusters and their properties. For instance, Planck Collaboration et al. (2013) took benefit of the all-sky survey to study 62 nearby halos and reconstruct their pressure profiles. Analysis of these profiles enabled, for the first time, the study of the intracluster medium and provided constraints on the thermal pressure support in clusters. Such analyses can be utilized to better characterize the pressure profile of observed clusters to improve modelisation and can also help in studying astrophysical parameters such as gas mass fraction.

More recently, with the dedicated ground-based telescope New IRAM Kids Arrays³ (NIKA2), the tSZ surface brightness of halos is measured and compared with the X-ray surface brightness or the y -map from, for example, Planck (Ruppin et al., 2018). Ruppin et al. (2018) analysed, for the first time with

³<https://nika2.osug.fr/>

NIKA2, a galaxy cluster (PSZ2G144.82+25.11) and demonstrated that the impact of the ICM can be characterized through the study of the tSZ-mass scaling relation and the pressure profile. NIKA2 is currently measuring the pressure profiles of 38 halos to extract a mean pressure profile and the tSZ-mass scaling relation (Hanser et al., 2023) and to study the thermodynamics of these halos (Perotto et al., 2023). Having robust measurements of halos pressure profiles can also be an interesting and challenging avenue for predicting the angular power spectrum.

4.3 Comparison with measurement in simulations

In this section, I will present my results regarding the comparison of the tSZ properties predicted by HMx and those measured in the different simulations. Despite the expected differences, it is interesting to understand their origin and the underlying degeneracies. It can allow to think about improved tSZ modelisation so it can be used more in future cosmological analyses such as 10x2pt analysis.

4.3.1 Methodology

To obtain the different power spectra from the simulations, we follow different procedures for the `Horizon` and `Magneticum` suite of simulations.

Power spectrum computation in the `Horizon` suite of simulations

To compute the power spectra in the `Horizon` suite of simulations, we first need to project the component of interest onto a uniform three-dimensional grid. The matter component is the sum of the dark matter (DM), stars, and gas. DM and stars are projected with a cloud-in-cell interpolation on the grid. Gas quantities are already on the regular Cartesian grid, and we directly use the values of mass or pressure from the corresponding level of refinement. The simulations provide the total gas pressure, which we can be easily modified to obtain the electron pressure, assuming local thermodynamical equilibrium between ions and electrons for a fully ionized gas. Using the ideal gas law, the total gas pressure can be expressed as:

$$P = P_i + P_e = \frac{\rho k_B T_i}{\mu_i m_p} + \frac{\rho k_B T_e}{\mu_e m_p}, \quad (4.14)$$

where index i is for ions and index e for electrons. Because we assume local equilibrium, therefore temperature equilibrium, we have $T_i = T_e = T$, resulting in:

$$P = \frac{\rho k_B T}{m_p} \left(\frac{1}{\mu_i} + \frac{1}{\mu_e} \right) = \frac{\rho k_B T}{m_p} \left(\frac{1}{\mu} \right), \quad (4.15)$$

thus leading to

$$\frac{P_e}{P} = \frac{\mu}{\mu_e} \simeq 0.492, \quad (4.16)$$

where P is the total gas pressure, μ_i is the mean molecular weight for ions particles ($\mu_i = 1.136$), μ_e is the mean molecular weight for electron particles ($\mu_e = 1.219$), and μ is the mean molecular weight for gas particles ($\mu = 0.6$ for a fully ionized gas).

We project our quantity into a 512^3 grid, allowing us to reach a Nyquist frequency ($k_{\text{Ny}} = \pi \times N_{\text{mesh}}/L_{\text{box}}$) of $k_{\text{Ny}} \sim 16 h \text{ Mpc}^{-1}$ for `Horizon-AGN` and `Horizon-noAGN` and up to $k_{\text{Ny}} \sim 1.8 h \text{ Mpc}^{-1}$ for `Horizon-Large`. To obtain the angular power spectrum, it will be beneficial to project `Horizon-Large` into a 1024^3 grid to achieve $k_{\text{Ny}} \sim 3.6 h \text{ Mpc}^{-1}$.

Once the quantity is projected, we use the `Pylians` python package (Villaescusa-Navarro, 2018) to compute the 3D auto- and cross-power spectra deconvolved by the CIC mass-assignment scheme.

Power spectrum computation in `Magneticum`

For the `Magneticum` simulation, we assign each gas particle (labelled by i) an electron pressure $P_{e,i}$ according to the ideal gas law

$$P_{e,i} = \frac{N_{e,i} k_B T_i}{V_{\text{cell}}}, \quad (4.17)$$

$$N_{e,i} = \frac{m_i}{m_p \mu_{e,i}}, \quad (4.18)$$

where T_i is the particle temperature, V_{cell} is the cell volume, $N_{e,i}$ is the number of free electrons, m_i is the particle mass, $\mu_{e,i} (= 2/(2 - Y_i))$ is the mean mass per electron and Y_i is the Helium fraction.

We use `Pylians` to project the electron pressure onto a 1024^3 mesh based on the CIC assignment scheme and then measure the power spectrum. We thus achieve $k_{\text{Ny}} \sim 3.6 h \text{ Mpc}^{-1}$.

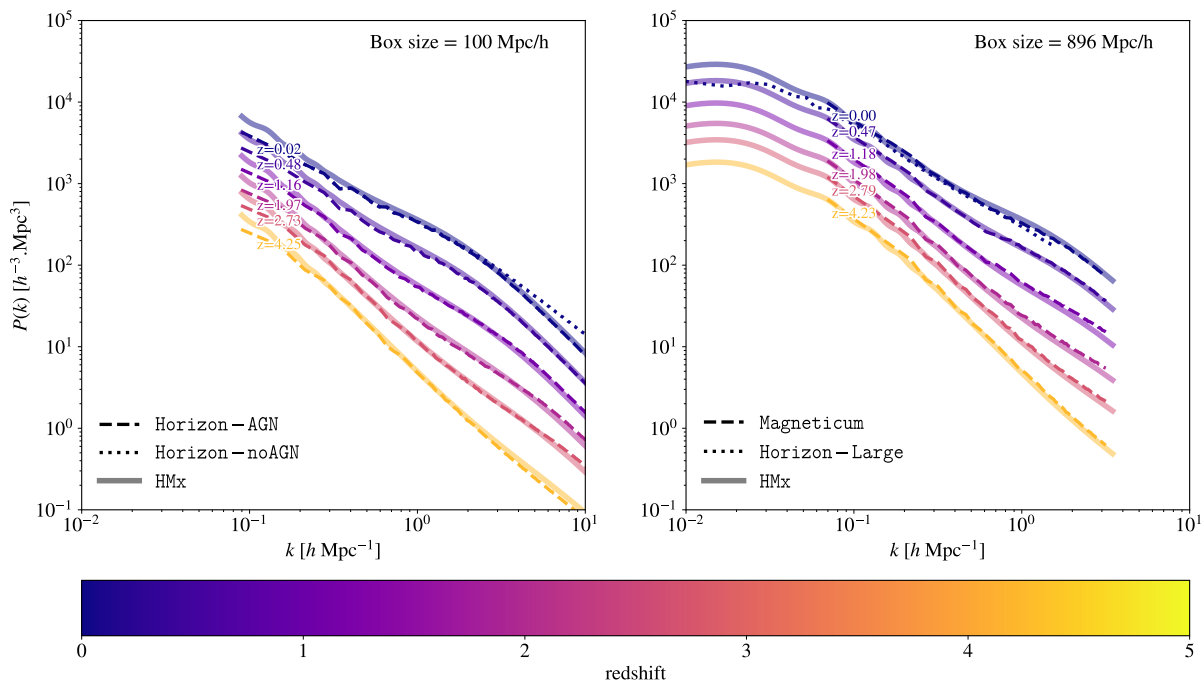


Figure 4.5: Matter auto-power spectrum as a function of redshift. The left panel shows the results for simulations with a box size of $100 h^{-1} \text{Mpc}$, thus **Horizon-AGN** in dashed line compared to **HMx** in solid line, and **Horizon-noAGN** at $z = 0$ in dotted line. The right panel shows the results for simulations with a box size of $896 h^{-1} \text{Mpc}$, thus **Magneticum** in dashed line compared to **HMx** in solid line, and **Horizon-Large** at $z = 0$ in dotted line. The power spectra go from $z = 0$ in dark blue to $z = 4.25$ in yellow.

4.3.2 Mass cut

The maximal mass chosen in **HMx** will significantly change the prediction on the pressure auto-power spectrum, as seen in Sect. 4.1.2. It is thus important to consider them when comparing the results from **HMx** and the one measured in the simulations. From the halo mass function of the simulations presented in Fig. 3.3, we see that the maximal masses of the $100 h^{-1} \text{Mpc}$ simulations (**Horizon-AGN** and **Horizon-noAGN**) is different than one in the $896 h^{-1} \text{Mpc}$ simulations (**Horizon-Large** and **Magneticum**). In the following of the analysis, I will thus use a different maximal mass to distinguish these two cases.

For the **Horizon-AGN** and **Horizon-noAGN** simulations, I will use $M_{\text{max}} = 6.4 \times 10^{14} h^{-1} M_{\odot}$ (maximum mass of **Horizon-AGN**). For the **Horizon-Large** and **Magneticum** simulations, I will use $M_{\text{max}} = 2.6 \times 10^{15} h^{-1} M_{\odot}$ (maximum mass of **Magneticum**). I have checked that the prediction using the maximal mass of **Horizon-noAGN** instead of the **Horizon-AGN** one (or **Horizon-Large** instead of the **Magneticum** one) makes almost no difference.

4.3.3 Power spectrum comparison

Matter auto-power spectrum

As an initial validation, we compared the matter auto-power spectrum of all simulations at different redshifts with the prediction by **HMx**, as illustrated in Fig. 4.5. The results obtained from **Horizon-AGN** and **Magneticum** are in dashed lines, while the predictions by **HMx** are in solid lines, varying from $z \sim 0$ in dark blue to $z \sim 4.25$ in yellow. The left panel shows the results for a box size of $100 h^{-1} \text{Mpc}$ and the right one for a box size of $896 h^{-1} \text{Mpc}$. Additionally, the results for **Horizon-noAGN** and **Horizon-Large** at $z \sim 0$ are included as dotted lines to emphasize the impact of different physics. However, we do not show their evolution with redshift since the trend is similar to the other simulations (as shown in Fig. B.1). For all the simulations and redshifts, we observe a good agreement. At low k , there is less power than predicted, which can be explained by the cosmic variance and the size of the simulated box, whereas at high k , the differences can be explained by the resolution. This test confirms the reliability of our pipeline before proceeding to analyse pressure auto- or cross-power spectra.

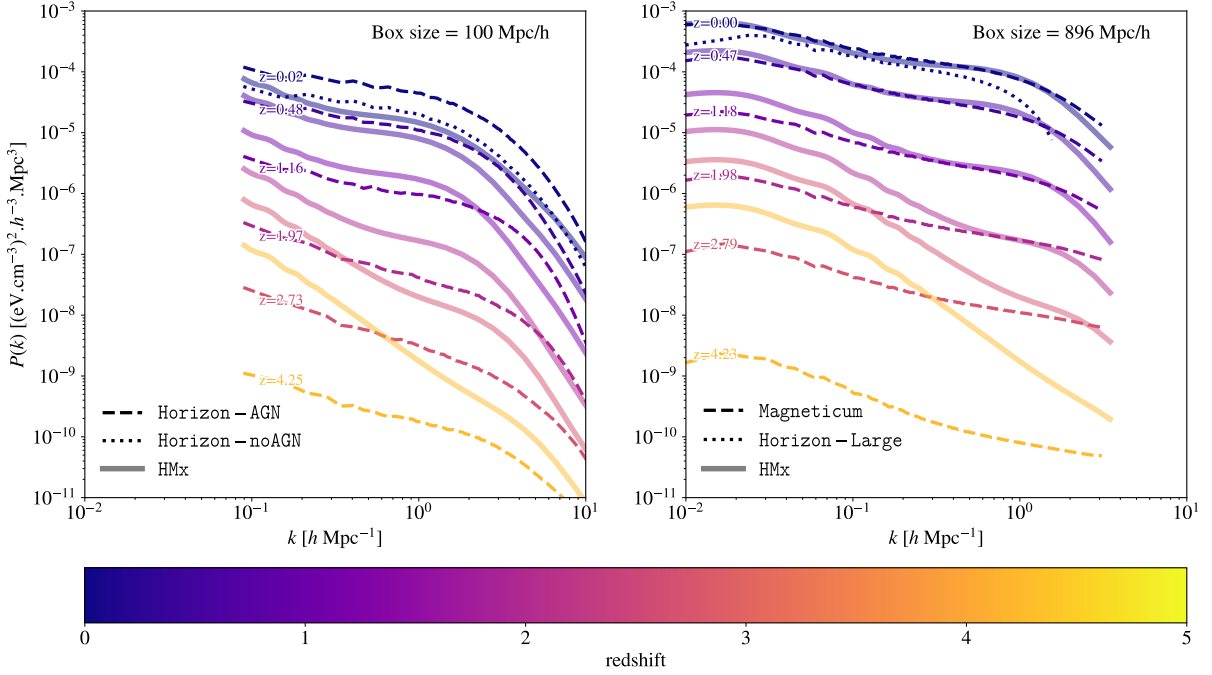


Figure 4.6: Pressure auto-power spectrum as a function of redshift. The left panel shows the results for simulations with a box size of $100 h^{-1}$ Mpc, thus **Horizon-AGN** in dashed line compared to **HMx** in solid line, and **Horizon-noAGN** at $z = 0$ in dotted line. The right panel shows the results for simulations with a box size of $896 h^{-1}$ Mpc, thus **Magneticum** in dashed line compared to **HMx** in solid line, and **Horizon-Large** at $z = 0$ in dotted line. The power spectra go from $z = 0$ in dark blue to $z = 4.25$ in yellow.

Pressure auto-power spectrum

We then compare the pressure auto-power spectrum in Fig. 4.6. The results obtained from the **Horizon-AGN** and **Magneticum** simulations are shown in dashed lines and the one predicted by **HMx** in solid lines, varying from $z \sim 0$ in dark blue to $z \sim 4.25$ in yellow. The left panel shows the results for a box size of $100 h^{-1}$ Mpc and the right one for a box size of $896 h^{-1}$ Mpc. Additionally, the result of **Horizon-noAGN** and **Horizon-Large** at $z \sim 0$ are included as dotted lines to emphasize the impact of different physics. However, we do not show their evolution with redshift since the trend is similar to the other simulations (as shown in Fig. B.2). At $z \sim 0$ (thus the darker blue lines), both **Horizon-AGN** and **Horizon-noAGN** predict an excess of power, even more important in **Horizon-AGN**. On the other hand, **Horizon-Large** shows a deficit in power. Finally, **Magneticum** is in relatively good agreement with the prediction at every scale. We also observe more power in the larger simulations, which is expected as they contain more massive halos (see also Sect. 4.1.2).

Then, we can examine the evolution with redshift. For all the simulations, we observe a relatively good agreement at low redshift with the predictions from **HMx** (up to $z \sim 1$ for **Horizon-AGN** and **Magneticum**). However, as the redshift increases, discrepancies become more evident. **HMx** predicts an excess of power at high redshift, indicating that the model’s physics fails to capture the nuances present in the simulations. We also notice that the measured power spectra are flatter than the predictions. Further investigations into these differences will be studied and discussed in Sect. 4.3.5.

Matter-pressure power spectrum

As we are using the matter-electron pressure model, we extend our analysis to examine the matter-pressure power spectrum to explore the agreement across different redshifts. Moreover, these spectra offer avenues for studying the correlation between lensing and pressure (that we are not doing here). In Fig. 4.7, we show the results obtained from the **Horizon-AGN** and **Magneticum** simulations in dashed lines and the one predicted by **HMx** in solid lines, varying from $z \sim 0$ in dark blue to $z \sim 4.25$ in yellow. The left panel shows the results for a box size of $100 h^{-1}$ Mpc and the right one for a box size of $896 h^{-1}$ Mpc. Additionally, the result of **Horizon-noAGN** and **Horizon-Large** at $z \sim 0$ are included as dotted lines to emphasize the impact of different physics. However, we do not show their evolution with redshift since the trend is similar to the other simulations (as shown in Fig. B.3). At $z \sim 0$ (thus the darker blue lines), as for the pressure auto-power spectrum, **Horizon-AGN** demonstrates an excess of power, **Horizon-Large** a lack of power, while **Magneticum** agrees well with the prediction. We now observe a good agreement

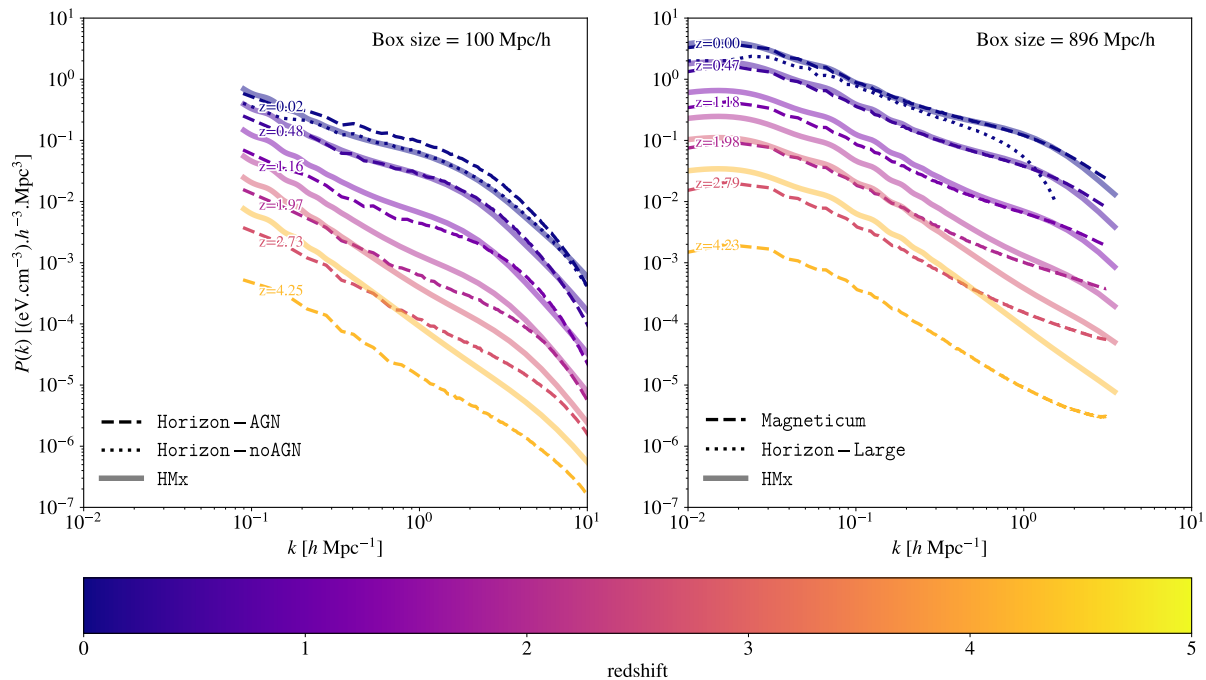


Figure 4.7: Matter-pressure power spectrum as a function of redshift. The left panel shows the results for simulations with a box size of $100 h^{-1}$ Mpc, thus **Horizon-AGN** in dashed line compared to **HMx** in solid line, and **Horizon-noAGN** at $z = 0$ in dotted line. The right panel shows the results for simulations with a box size of $896 h^{-1}$ Mpc, thus **Magneticum** in dashed line compared to **HMx** in solid line, and **Horizon-Large** at $z = 0$ in dotted line. The power spectra go from $z = 0$ in dark blue to $z = 4.25$ in yellow.

between **Horizon-noAGN** and **HMx**.

Then, we can examine the evolution with redshift. We observe a better agreement up to a comparable redshift ($z \sim 1.18$ for both simulations) than for the pressure auto-power spectrum. This outcome is expected as **HMx** has been calibrated on the matter-pressure power spectrum up to $z = 1$. Moreover, the matter auto-power spectrum agrees well across all redshifts, mitigating the discrepancies in the pressure auto-power spectrum. At higher redshifts, the discrepancies caused by the pressure auto-power spectra persist. Given the sensitivity of pressure to baryonic physics, such cross-correlations can serve as valuable tools for constraining astrophysical parameters (e.g., with the cosmic shear-tSZ cross-correlation with the **flamingo** simulations in [McCarthy et al., 2023](#) or with the lensing-tSZ cross-correlation from **KiDS-1000** (lensing), **Planck** and **ACT** (tSZ) in [Tröster et al., 2022](#)). Depending on the probes we are working with, it is crucial to adequately model different redshift ranges. With future surveys, we expect to be sensitive up to redshift two for probes such as the distribution of galaxies, tomographic studies, or weak lensing (e.g., with **Euclid**⁴, [Laureijs et al., 2011](#) or **Nancy Grace Roman Space Telescope**⁵, [Eifler et al., 2024](#)), our prediction thus needs to be trustable up to this redshift, which is qualitatively the case for the matter-pressure power spectrum.

4.3.4 Angular power spectrum comparison

The observable accessed through surveys is the angular power spectrum, and its accurate prediction is crucial. Using the power spectra computed on the simulations or predicted by **HMx**, we can use Eqs. (2.30) and (4.13) to obtain the pressure angular power spectrum. We integrate these spectra over the redshift range $z = 0.02$ to $z = 4$ and limit our analysis to the Nyquist frequency $k = k_{\text{Ny}}$. Thus, we are going up to $k \sim 16 h^{-1}$ Mpc for **Horizon-AGN** and **Horizon-noAGN**, while going up to $k \sim 3.6 h^{-1}$ Mpc for **Magneticum** and **Horizon-Large** (using a projection on 1024^3 to achieve a comparable k_{Ny} to that of **Magneticum**). At each z , the multipole ℓ range accessible with the simulations varies, depending on the available k range (which is influenced by the size of the simulations). Choosing an ℓ range accessible to all the simulations across all z between 0.02 and 4 would be quite narrow. To avoid this limitation and extract maximum information from the simulations, we opted for an interesting range of ℓ , filling the angular power spectrum with 0 for any inaccessible ℓ . We ensure to maintain the same behaviour in the

⁴<https://cnes.fr/projets/euclid>

⁵<https://science.nasa.gov/mission/roman-space-telescope/>

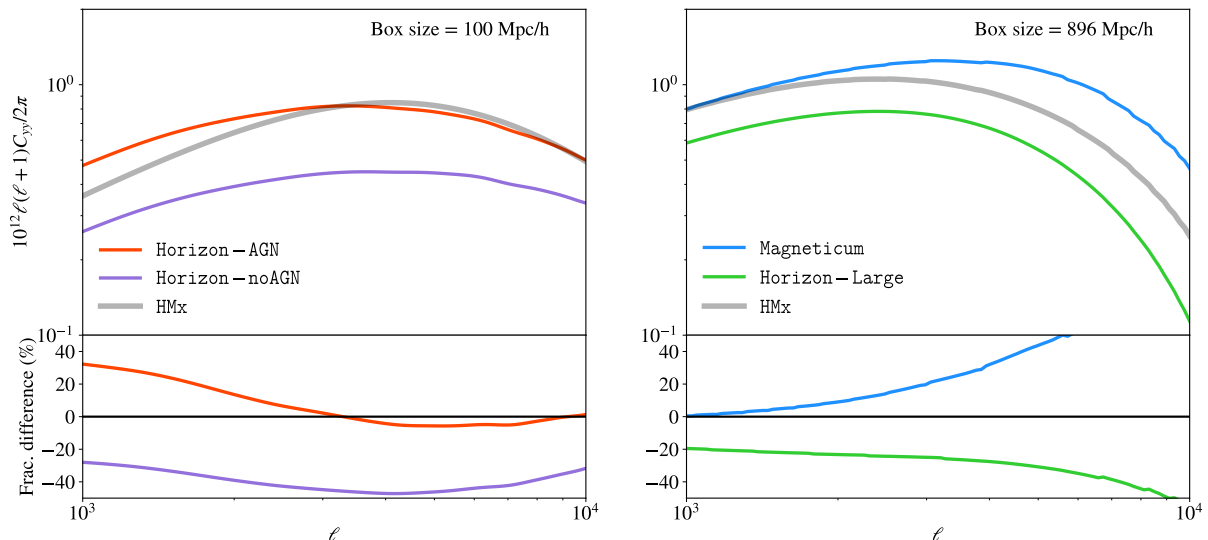


Figure 4.8: Pressure angular power spectrum integrated between $z = 0.02$ and $z = 4$ for the different simulations in different colours and HMx in dark grey. The left panel shows the results for simulation with a box size of $100 h^{-1}$ Mpc, thus **Horizon-AGN** in red and **Horizon-noAGN** in purple. The right panel shows the results for simulation with a box size of $896 h^{-1}$ Mpc, thus **Horizon-Large** in green and **Magneticum** in blue.

angular power spectrum computed with the power spectrum of HMx by cutting out at the same locations. We computed the angular power spectrum for ℓ between 10^3 and 10^4 to encompass a range where the tSZ becomes an important foreground but remains accessible with future surveys. This pipeline limits our predictive power and is not realistic but allows us to infer the properties and behaviour of the different simulations and HMx on the maximum modes possible.

The angular power spectra obtained are shown in Fig. 4.8. Results for the different simulations are illustrated with various colours: **Horizon-AGN** in red, **Horizon-noAGN** in purple, **Horizon-Large** in green, and **Magneticum** in blue, and for HMx in grey. The top left panel shows the result for a box size of $100 h^{-1}$ Mpc and the top right one for a box size of $896 h^{-1}$ Mpc. The bottom panels show, with the same colour scheme, the fractional difference between the simulation and the result from HMx. The agreement between **Horizon-AGN** and HMx improves at higher ℓ with differences reaching less than 10% between $\ell = 3 \times 10^3$ and $\ell = 10^4$. The power spectrum of **Horizon-AGN** (see Fig. 4.6) exhibits more power at low redshifts and less power at high redshifts compared to the prediction, and these discrepancies seem to compensate each other, resulting in a small difference in the angular power spectrum. In contrast, **Magneticum**, shows an opposite trend, with better agreement at low ℓ where differences are less than 10% difference between $\ell = 10^3$ and $\ell = 2 - 3 \times 10^3$. For **Horizon-noAGN** and **Horizon-Large**, the general behaviour is quite similar, the simulations always have between 20% and 50% less power than HMx.

4.3.5 Halo model consequences

As mentioned previously, a halo model framework contains some inherent limitations that can affect its predictive power. In this section, I am trying to understand these limitations and explore how they might bias or restrict our predictive power.

One- and two-halo terms decomposition

To understand better the differences observed in the predicted and measured pressure auto-power spectrum (see Fig. 4.6 and related text), we can explore the evolution of the one- and two-halo term contributions to the total power spectrum as a function of redshift. In Fig. 4.9, we can see this decomposition for different redshifts in the different panels: $z = 0.02$, $z = 1.16$, and $z = 3.01$ by focusing on the coloured lines and ignoring the grey lines, which will be addressed in the next subsection. For each redshift, we compare the **Horizon-AGN** power spectrum in coloured dashed lines, the one predicted by HMx in coloured solid lines, the one- and the two-halo term predicted by HMx in coloured thin and coloured dotted-dashed lines, respectively. As we increase the redshift, the contribution from the two-halo term becomes increasingly significant compared to the one-halo term at a given k . The increasing importance of the two-halo term is related to the scale at which the one- and two-halo terms intersect, which shifts towards higher

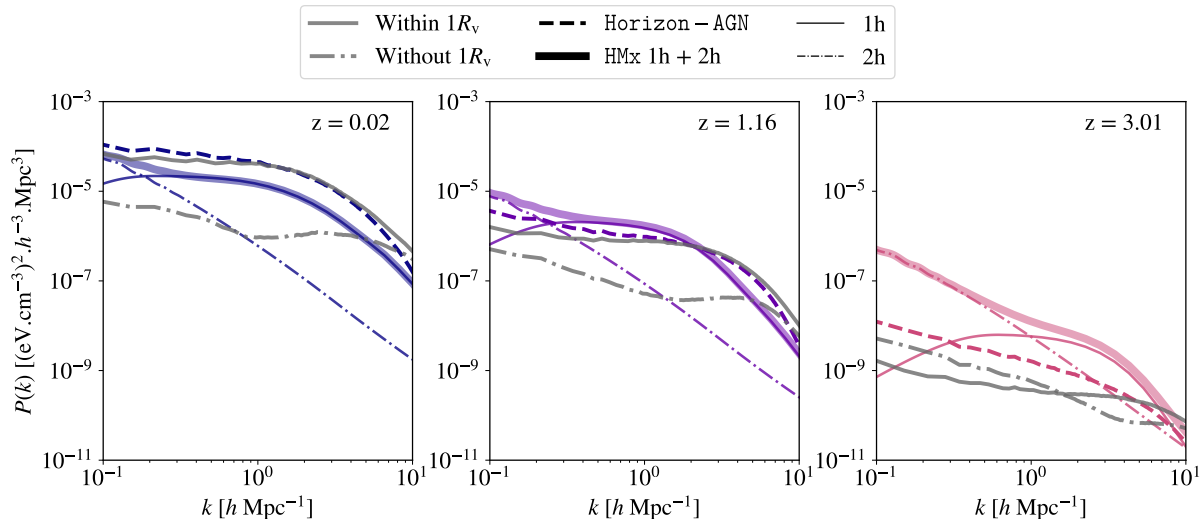


Figure 4.9: Pressure auto-power spectrum as a function of redshift in the different panels. For every redshift, we show the power spectrum of `Horizon-AGN` in a coloured dashed line, the `HMx` prediction in a coloured solid line, the predicted one-halo term in a coloured thin line, and the predicted two-halo term in a coloured dotted-dashed line. We superimpose the pressure auto-power spectrum within (without) one virial radius in a solid grey line (grey dotted-dashed line). On the left, we show the result for $z = 0.02$, on the middle for $z = 1.16$, and on the right for $z = 3.01$.

k values. We observe that the excess of power in the `HMx` power spectrum at higher redshifts is thus dominated by the two-halo term at low k and by the one-halo term at high k . The excess of power in `HMx` suggests potential discrepancies in the distribution between the one- and two-halo terms, including the amplitude of each term at a given k scale as a function of redshift. At higher redshifts, there are only a few halos, the observable thus resembles the matter distribution, whereas at lower redshifts, more, and more massive, halos have formed. Thus, at low redshifts, the hypothesis that all the matter is within the halos is more accurate and the halos contribute more to the total power. Also, the contribution from the halos increases more rapidly than that from the diffuse gas. Finally, differences can suggest an inaccurate representation of the intergalactic medium effects by the two-halo term, in particular its amplitude.

Validity of the halo model

The halo model contains some intrinsic limitations, as it assumes that all the matter is within spherical halos and that the one-halo term—which is the dominant contribution to the power spectra at low redshift (as shown by the coloured lines in Fig. 4.9 and discussed in the previous subsection)—only contains bound gas, which is the gas inside the virial radius. To better understand the validity of the halo model, we can investigate how the power spectrum differs when considering pressure inside or outside one virial radius of halos, compared to the prediction from `HMx`. These spectra are obtained by selectively masking gas pressure by masking either inside or outside one virial radius of halos. We have performed this study for the different simulations (`Horizon-AGN`, `Horizon-noAGN`, `Horizon-Large`, and `Magneticum`) and they all show similar behaviour, thus, we are showing the results for `Horizon-AGN` in Fig. 4.9 where we can now focus on the grey lines. We can compare the electron pressure auto-power spectra of `Horizon-AGN` (coloured dashed line) to the one coming from inside (outside) one virial radius in solid grey lines (grey dotted-dashed lines) for different redshifts: $z = 0.02$ on the left, $z = 1.16$ on the middle and $z = 3.01$ on the right. We observe that, at low redshift, most of the power comes from within one virial radius of halos. However, this assumption loses validity with increasing redshifts. Notably, at $z = 3.01$, there is more power coming from outside one virial radius of halos than inside. This implies the diminishing applicability of the halo model prescription at higher redshifts. Moreover, comparing the one-halo term (coloured thin lines) to the power spectrum coming from inside one virial radius (solid grey lines) reveals that at $z = 0.02$, the shape is consistent even if `HMx` lacks power. However, as redshift increases, discrepancies in shape emerge, contrary to the expectation that the one-halo term should represent the power spectrum within one virial radius of the halos. These observations can partially explain the discrepancies observed in Fig. 4.6 with increasing redshift. Given these limitations, it becomes imperative to consider them when evaluating the cross-correlation with other probes.

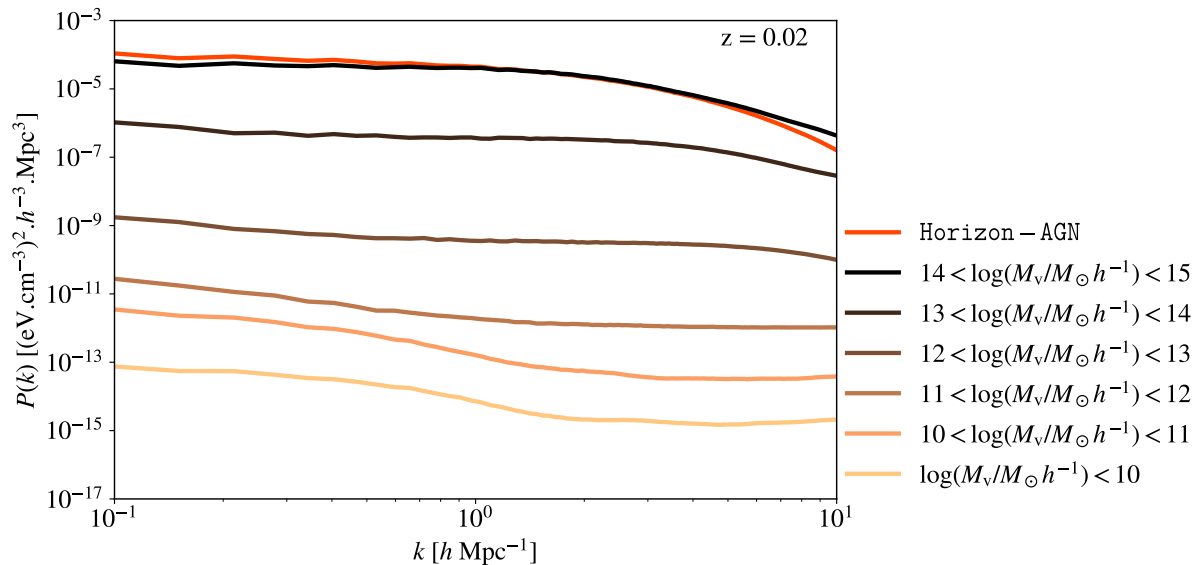


Figure 4.10: Pressure auto-power spectrum of the total **Horizon-AGN** simulation in red compared to the signal coming from within one virial radius of the different mass. The highest mass bin is in dark brown, and the lowest mass bin is in yellow.

Importance of the halo mass

To go one step further in our understanding of the halo model, we investigate the contribution of each mass bin to the overall power spectrum measured in a simulation. We have performed this study for the different simulations (**Horizon-AGN**, **Horizon-noAGN**, **Horizon-Large** and **Magneticum**), and they all show a similar trend, thus, we are only presenting the results for **Horizon-AGN** in Fig. 4.10. We are comparing the electron pressure auto-power spectra (red line) with the one coming from inside one virial radius of the halos in different mass bins, ranging from dark brown for the highest mass bin to yellow for the lowest mass bin at $z = 0.02$. We can clearly see that most of the power emanates from the highest mass bin, despite its relatively low population (for **Horizon-AGN**, the highest mass bin contains only 0.006% of the total number of halos). The lower the mass bin, the lesser its contribution to the total power spectrum. A reduction of approximately one order of magnitude is observed with each logarithmic decrease in mass bin. This shows the importance of ensuring that the halo model, particularly the electron pressure profile, accurately reflects the characteristics of the highest mass halos. However, given that the higher mass halos are less common occurrences, it may be worth masking them to mitigate the connected non-Gaussian covariance and tighten cosmological or astrophysical constraints, as explored in [Osato & Takada \(2021\)](#).

4.3.6 Halo pressure profiles

The main component of the model is the electron pressure profile, defined in Eq. (4.5). This profile not only represents a major aspect of the model but also constitutes the primary component of the one-halo term. As demonstrated in Fig. 4.9, the one-halo term is the predominant contribution to the power spectrum at low redshifts, further influencing the angular power spectrum. Consequently, in Fig. 4.11 we present the electron pressure profile measured in the simulations in different colours: **Horizon-AGN** in red, **Horizon-noAGN** in purple, **Horizon-Large** in green and **Magneticum** in blue, compared to the prediction from **HMx** in dark grey. We show these profiles at $z = 0$ (top) and $z = 1.18$ (bottom) across three mass bins: $12 < \log(M_v/M_\odot h^{-1}) < 12.5$, $13 < \log(M_v/M_\odot h^{-1}) < 13.5$ and $14 < \log(M_v/M_\odot h^{-1}) < 14.5$ from left to right. The consistency in trends between the two redshifts aligns with expectations, given that the measured profiles scale as $(1+z)^4$, as the predicted one. For all the **Horizon** simulations at both redshifts, as well as **Magneticum** at $z = 1.18$, discrepancies are noticeable, particularly in the low- and intermediate-mass bins, where the inner regions exhibit closer agreement, while deviations escalate towards outer regions. The **Magneticum** profiles at $z = 0$ have different behaviour, characterized by elevated inner pressures and convergence towards other simulations in the outer regions. For the high-mass bin, we see that the **Horizon** simulations have more power than the predictions at both redshifts, whereas the **Magneticum** profiles surpass the prediction at $z = 0$ and undershoot them at $z = 1.18$. Nevertheless, the overall shape remains reasonably consistent across all distances from the centre. As **HMx** should represent the mean behaviour of halos, we have added error bars that indicate the error on

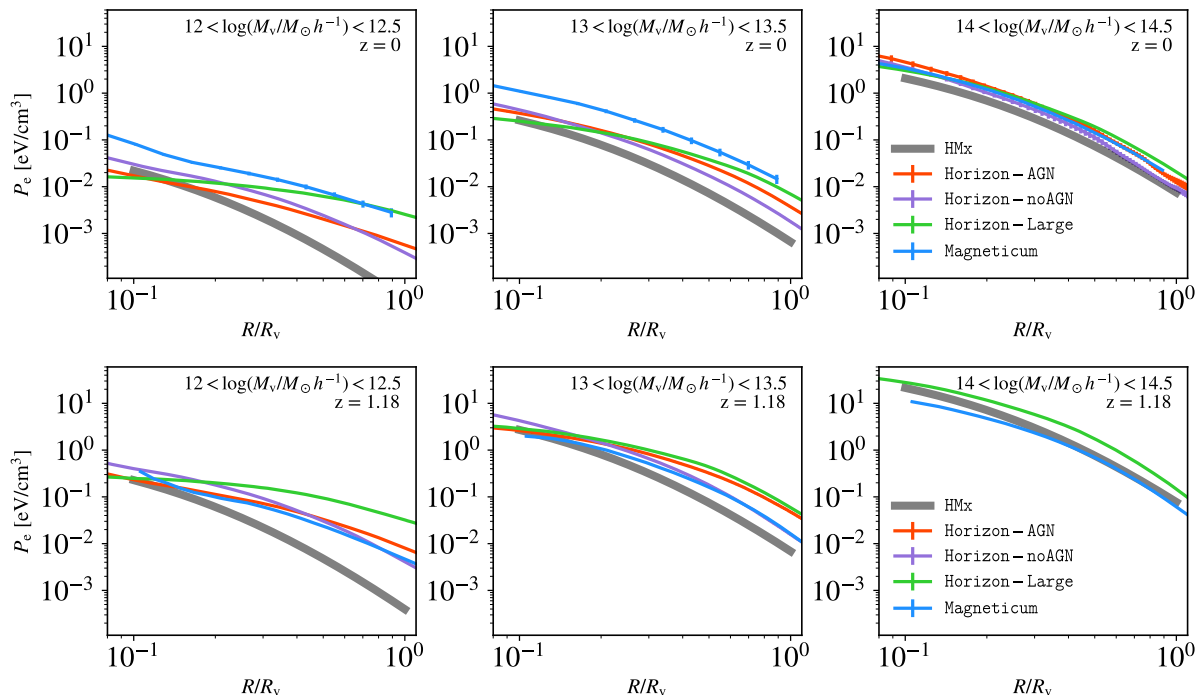


Figure 4.11: Pressure profile as a function of the distance to the centre of the halos compared with the one predicted by HMx in dark grey at $z = 0$ (first row) and $z = 1.18$ (second row). Each panel represents a different mass bin, and each colour is a different simulation: **Horizon-AGN** in red, **Horizon-noAGN** in purple, **Horizon-Large** in green, and **Magneticum** in blue. The error bars represent the error on the mean and are, most of the time, too small to be visible.

the mean and are often too small to be discernible. This observation suggests that while our mean is well estimated, it is not entirely compatible with the HMx prescription.

Since the model’s free parameters are tailored to fit the response power spectrum, achieving a perfect agreement on pressure profiles is not guaranteed. Moreover, the influence of high masses, which contribute the most, can alter even more the profiles of low-mass halos. It is still worth comparing the predicted and measured profiles to understand power spectrum disparities. Let us describe more the high-mass bin, which is the dominant part of the power spectrum. The difference observed in this bin could explain power spectrum differences. For example, at $z = 0$, all simulations (**Horizon-AGN**, **Horizon-noAGN**, **Horizon-Large** and **Magneticum**) have qualitatively similar pressure profiles (with more pronounced differences on the outer regions), containing more pressure than HMx. At $z = 1.18$, the **Horizon-Large** remains above the prediction, while **Magneticum** is under. On the left panel of Fig. 4.6, both **Horizon-AGN** and **Horizon-noAGN** power spectra also contain more power than HMx. However, on the right panel of Fig. 4.6, **Horizon-Large** power spectrum lies below HMx (with **Magneticum** showing relatively good agreement). These diverse behaviours suggest that profile differences alone cannot entirely account for observed power spectrum discrepancies. Additionally, we note that the lower-mass halos exhibit greater discrepancies, as anticipated.

To improve our analysis, it can be interesting to focus even more on high-mass halos. Future studies with larger volume simulations could provide a more comprehensive probe of these halos. Because of the current computational constraints, the resolution of such simulations cannot be as good as the one in **Horizon-AGN**, for example, which can potentially introduce additional biases. Building large simulations with zoom-in capabilities targeting big halos to assess the fidelity of baryonic physics can be an avenue. This approach could offer a new perspective for such analysis, which is currently limited by the noise on the number of these halos.

4.3.7 Difference of the simulations

The simulations that we are analysing are different: they are run with different computational codes, different physics models, different resolutions, and different box sizes. Different choices in terms of the included physics and their modelisation methodology are made and can influence the obtained results. For example, in Mead et al. (2020), models were fitted against three **BAHAMAS** simulations, each containing different strengths of AGN feedback, yielding to different values for the fitted parameters. In our

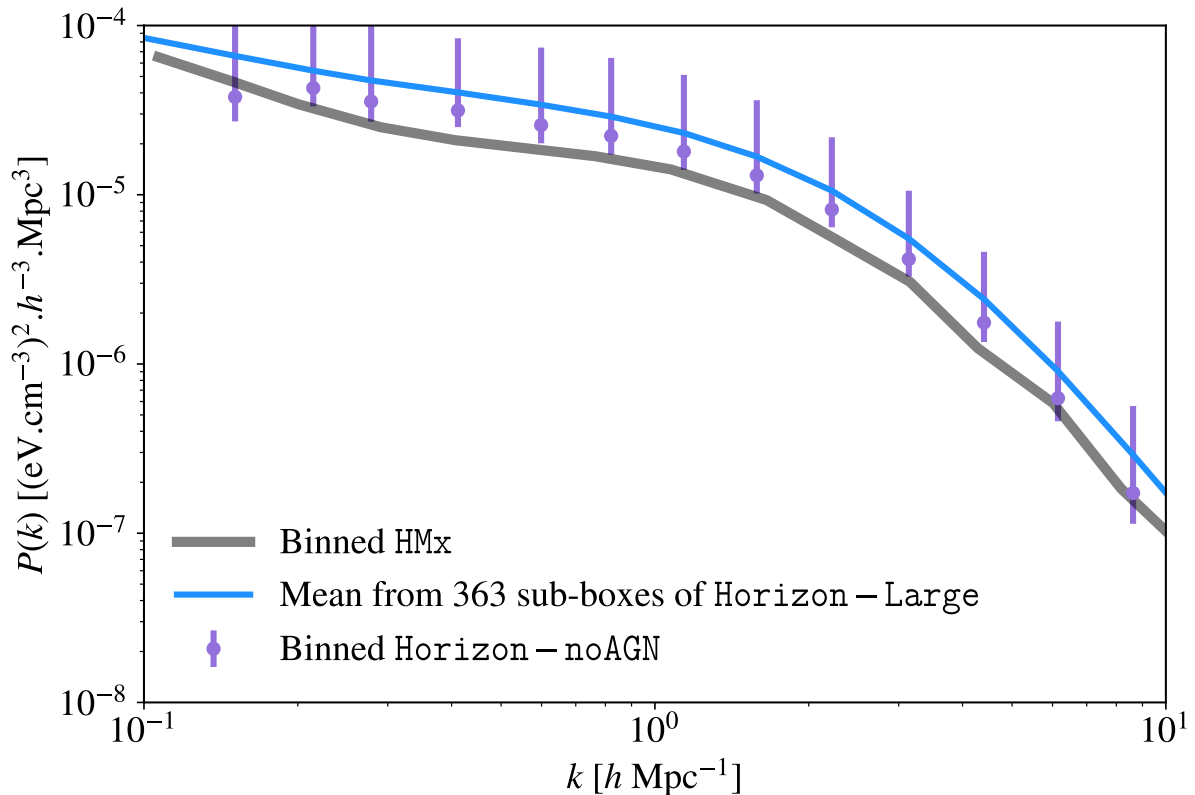


Figure 4.12: Pressure auto-power spectrum in Horizon-noAGN in purple compared to HMx in dark grey and the mean over 363 boxes of $100 h^{-1}$ Mpc from the Horizon-Large in blue. The purple error bars contain 68% around the mean of the 363 boxes of the Horizon-Large (see the text for more information). They are applied to the Horizon-noAGN simulations to represent its variance.

simulations, with the different choices, the strength of feedback is also different, but other choices can lead to many other differences that are challenging to precisely identify and define.

Another critical aspect that can be evaluated is the cosmic variance of the simulation. To quantify this variance, we took 500 non-overlapping boxes extracted from the Horizon-Large simulation, each with a dimension of $100 h^{-1}$ Mpc. As the power spectrum depends a lot on the high-mass halos, we applied selection criteria to retain only those boxes with maximal mass similar to the one in the Horizon-AGN and Horizon-noAGN simulations. This refinement yielded 363 non-overlapping boxes, from which we computed the pressure auto-power spectrum. We have noticed a strong correlation in k among these 363 pressure auto-power spectra, explained by both the multitude of bins in our power spectrum and the observed scatter in the power spectra themselves. Thus, direct computation of error bars would be artificially large because of this correlation and will not really emphasize the true cosmic variance. To mitigate this issue, we employed a constant binning in log space of k for our power spectrum, checking that the level of correlation between the bins is low. We then calculated a cumulative probability distribution function at each k to extract the values encompassing 68% of the signal. This approach allows us to derive the lower and upper bounds of our error bars.

In Fig. 4.12, we show the binned mean power spectrum of these boxes in blue, of the Horizon-noAGN simulation in purple, and of the HMx prediction in grey. The variance derived using the method described above is overlaid on the Horizon-noAGN simulation to represent the cosmic variance of such a simulation. We see that the error bars are non-Gaussian and of the same order of magnitude across different k ranges. They emphasize a tendency for a higher power spectrum than the one predicted by HMx.

Trying to build a halo model that more faithfully captures the physics embedded in the simulation (with the value of the parameters measured from the simulation, for example) can be an avenue for increasing the robustness of predictions. Such a procedure can pose inherent challenges, such as the parameter measurement methodologies. Another possibility could be to focus on refining pressure profiles to better match simulation data. However, it remains unclear whether such improvements would directly propagate into enhanced agreement in the power spectrum and angular power spectrum. We let the exploration of such analysis for future work.

4.4 Measurement of some properties

As discussed in the presentation of the HMx halo model in Sect. 4.1.1, the parameters of the model have been jointly fitted at the level of the response power spectrum. We thus expect the model to have differences when looking at other components, as explored in this section. As the parameters are linked to physical processes, learning them at the level of the power spectrum can lead to overfitting for this observable and not allow to reproduce the physics of each component. Even if the HMx framework assumes that these fitted parameters are not intrinsic properties of halos, it is interesting to compare the value used in the model with measurements from simulations to investigate the impact of varying these parameters on the predicted power spectrum. Moreover, the assumption of parameters constancy with mass and, to a large extent, with redshift can introduce biases. In this section, I undertake the comparison between the measured parameters in the Horizon-AGN simulation and those employed in HMx. Subsequently, I investigate the impact of varying these parameters in the predicted pressure auto-power spectrum and pressure profiles.

4.4.1 Properties in the simulations

Polytropic index Γ

The first studied parameter is the polytropic index Γ , which appears in the bound gas density profile, defined in Eq. (4.6). By default, $\Gamma = 1.17$, but after fitting, the model employs a value of 1.1770. It is assumed to remain constant across mass and redshift. To assess its agreement with the one measured in the Horizon-AGN simulation, I can derive it from Eqs. (4.5) and (4.6). This yields:

$$\frac{d \log P(M, r)}{d \log \rho(M, r)} = \Gamma(M, r). \quad (4.19)$$

To derive Γ , I examine the pressure-density relationship within the simulation for different mass bins. In Fig. 4.13, I show this pressure-density relation for two different mass bins in the two panels. The red line is the measurement from the Horizon-AGN simulation at $z = 0.02$, while the dark red line shows the relation at $z = 1.16$. The thick grey line is the relation used in HMx with a slope of 1.1770. Upon inspection, there is a change in slope within the curves at a certain density. This transition occurs between the inner of halos (low pressure and density, left of the vertical lines) and the intracluster medium (high pressure and density, right of the vertical lines). This transition occurs within a specific range of densities, typically around 10^{-1} H/cm^2 . However, there are variations in the density threshold for this transition across different mass bins and redshifts. As I am interested in the relation within halos, I apply a manual cutoff represented by the vertical lines in the plot, and I only fit the relation that is on the left of these vertical lines. With this restricted dataset, I fit the slope Γ .

The results of the Γ parameter estimation across different mass bins and redshifts are shown in Fig. 4.14. We compare the constant value used in the HMx ($\Gamma = 1.1770$) in thick grey line with the measurements obtained at $z = 0.02$ in red and $z = 1.16$ in dark red. The error bars are computed using the covariance provided by the fitting process. Their magnitude is determined by the square root of the associated coefficient of the covariance matrix. Assuming a constant over mass and redshift is not representative of the behaviour observed in the simulation. As the high-mass and low redshift are the dominant contributions to the angular power spectrum, I consider using a value of $\Gamma = 1.2847$, corresponding to the mean Γ from the four highest mass bins at $z = 0.02$, to refine our predictions.

Deviation from a virial relation in the temperature α

The parameter α is also taken as a constant with mass and redshift and should encapsulate the deviation from a virial relation in Eq. (4.10), it thus acts as a hydrostatic bias. By default, $\alpha = 1$, but after fitting, the model employs a value of 0.8471. To determine the value of α in the simulation, the measured gas temperature can be compared with the gas temperature predicted by Eq. (4.9), assuming $\alpha = 1$. I can then fit the slope α . Fig. 4.15 shows the α value used in HMx (0.8471) in thick grey line, the measurement in Horizon-AGN at $z = 0.02$ in red, and the measurement at $z = 1.16$ in dark red. The error bars are computed using the covariance provided by the fit. Their magnitude is determined by the square root of the associated coefficient of the covariance matrix. There is a clear trend where α increases with the mass of the halo, indicating higher temperature. At $z = 0.02$, around $M_v = 10^{13.3} M_\odot h^{-1}$, α becomes bigger than one, suggesting that halos above this mass are hotter than the virial temperature. As the concentration enters in the gas temperature equation, the discrepancies can indicate that the concentration is not well modelised or a non-virialized state of the halos. As the high-mass and low redshift are the dominant contributions to the angular power spectrum, I consider using a value of $\alpha = 0.9972$, corresponding to the mean of the four highest mass bins at $z = 0.02$, to refine our predictions.

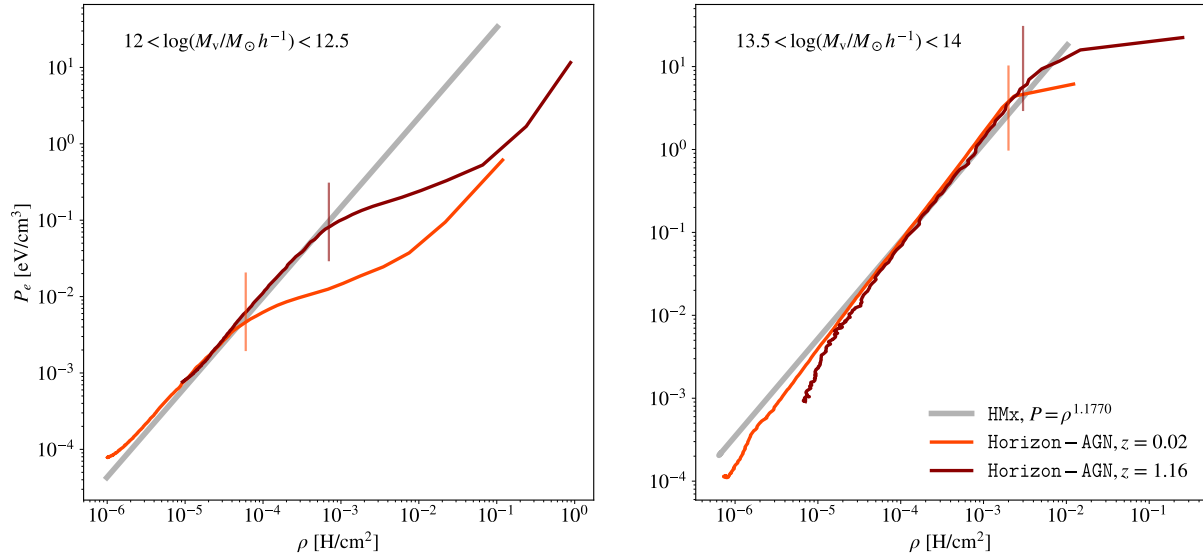


Figure 4.13: Electron pressure as a function of the density for two different mass bins in the different panels to study the slope Γ . Thick grey lines are the relation used in HMx with $\Gamma = 1.1770$. Red and dark red lines correspond to the measurement of the relation in Horizon-AGN at $z = 0.02$ and $z = 1.16$, respectively. The vertical lines correspond to the limit between what is considered as the inner of halos (left of the lines) and the ICM (right of the lines).

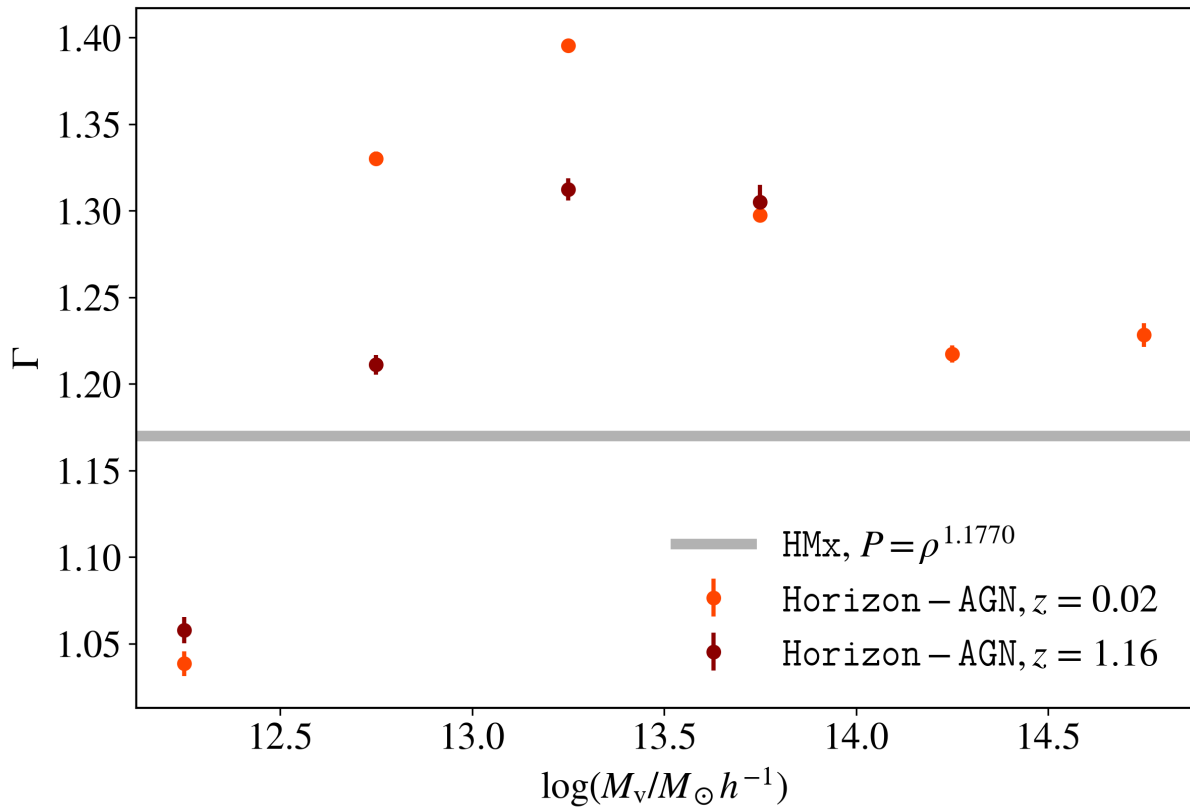


Figure 4.14: Slope Γ measured from the pressure-density relation in different mass bin in the Horizon-AGN simulation at $z = 0.02$ in red and $z = 1.16$ in dark red. Error bars are computed as the square root of the associated value in the covariance matrix given by the fit. The grey line represent the value used in HMx $\Gamma = 1.1770$.

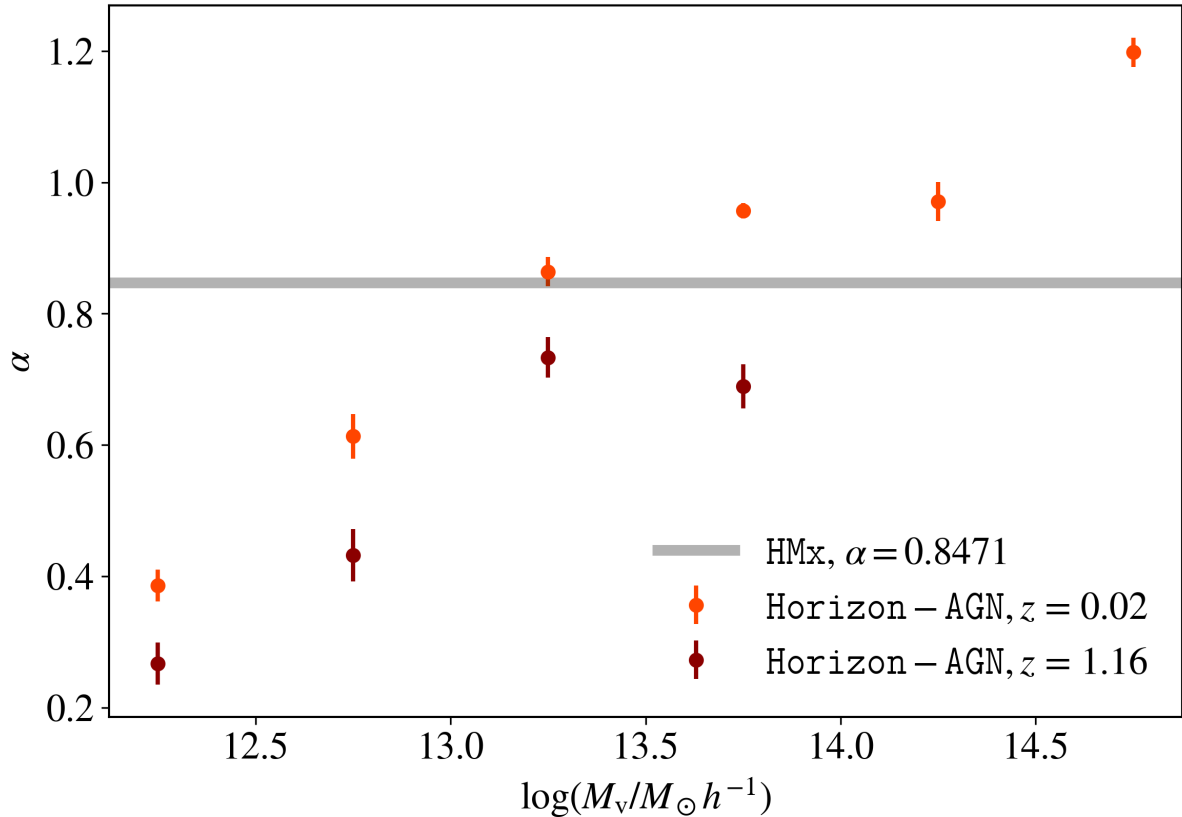


Figure 4.15: Measurement of α in the **Horizon-AGN** simulation at $z = 0.02$ in red and $z = 1.16$ in dark red compared to the value used in **HMx** in thick grey line. Error bars are computed as the square root of the associated value in the covariance matrix given by the fit.

β and M_0 in the fraction of bound gas

The parameters β and M_0 appear in the fraction of bound gas in Eq. (4.8) and govern its slope. The default value of M_0 is $10^{14} h^{-1} M_\odot$ but after fitting, the model employs a value of $10^{13.5937} h^{-1} M_\odot$ (which is equivalent to $3.92 \times 10^{13} h^{-1} M_\odot$). The default value of β is 0.6 and remains unchanged in the model as it was not subject to fitting. To measure the fraction of bound gas in the simulations, I evaluate the masses of the different components within one virial radius at $z = 0.02$ and $z = 1.16$:

$$f_{\text{bnd}} = \frac{M_{\text{gas}}}{M_{\text{gas}} + M_{\text{stars}} + M_{\text{DM}}} . \quad (4.20)$$

The results are shown in Fig. 4.16, where the fraction of bound gas of halos measured in **Horizon-AGN** is represented by blue dots, for $z = 0.02$ on the left panel and $z = 1.16$ on the right panel. The red dots (and dark red dots) and error bars are the mean and standard deviation of the individual halos in different mass bins. On the left panel, the red line represents the best fit obtained from the binned fraction of bound gas (red dots), while the **HMx** function is shown with the thick grey line. On the right panel, the dark red line represents the best fit obtained from the binned fraction of bound gas (dark red dots), while the **HMx** function is shown with the thick grey line. The dashed red line is the best fit obtained at $z = 0.02$ for comparison. The fraction of bound gas used in **HMx** is notably lower than that in the simulation, meaning that halos contain more gas than assumed by the model. This trend is accentuated at higher redshift. The best fit at $z = 0.02$ gives $M_0 = 1.07 \times 10^{13} \pm 1.67 \times 10^{12} h^{-1} M_\odot$ and $\beta = 0.6 \pm 0.06$. To obtain a good fit, β is similar to the value employed in **HMx**, while M_0 is approximately one order of magnitude smaller. The best fit at $z = 1.16$ gives $M_0 = 5.03 \times 10^{12} \pm 0.55 \times 10^{12} h^{-1} M_\odot$ and $\beta = 0.45 \pm 0.04$. To obtain a good fit, both M_0 and β are lower than the value employed in **HMx**. As the low redshifts are the dominant contribution to the angular power spectrum, I retain the $M_0 = 1.07 \times 10^{13} h^{-1} M_\odot$ to evaluate its impact on the prediction.

Concentration

The concentration assumed in **HMx** follows a modified [Duffy et al. \(2008\)](#) formulation to incorporate the impact of baryonic feedback, defined in Eq. (4.12). Although I do not intend to modify it, comparing

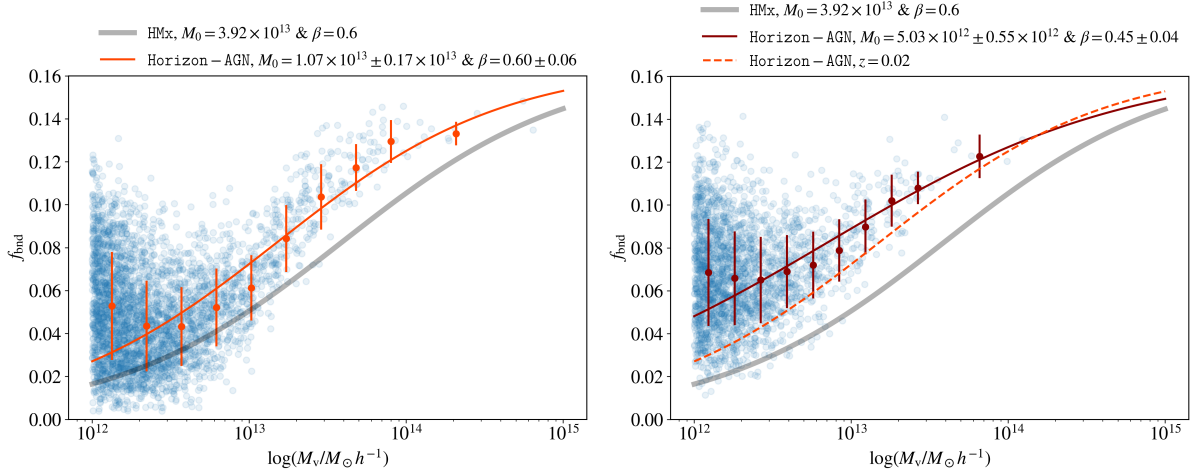


Figure 4.16: Measurement of the fraction of bound gas of individual halos of the **Horizon-AGN** simulation in blue, at $z = 0.02$ (left panel) and $z = 1.16$ (right panel). The red dots (left) and dark red dots (right) are the binning of the blue points, where the error bars are the standard deviation. The thick grey line shows the equation used in **HMx**. Left: the red line shows the best fit obtained from the red dots for our simulation for $z = 0.02$. Right: the dark red line shows the best fit obtained for the dark red dots for our simulation at $z = 1.16$ and the dashed red line is the best fit obtained at $z = 0.02$ for comparison. Note: the mass definition of this plot is different and is computed using the virial equilibrium. In practice, it shifts the mass by 0.1 dex.

measurements in the simulations with the prediction is still interesting. Fig. 4.17 shows a comparison between the concentration used in **HMx** at $z = 0.02$ and $z = 1.16$. First, at $z = 0.02$, the **HMx** concentration is in a thick grey line, and the measurements obtained from **Horizon-AGN** are the red dots. At $z = 1.16$, the **HMx** concentration is in a thick black line, and the measurements obtained from **Horizon-AGN** are the dark red dots. The error bars are computed by propagating the covariance of r_s given by the fit. The error bars on r_s (noted σ_{r_s}) are the square root of the associated coefficient of the covariance matrix, and then the error bars on c_M are computed as $\sigma_{c_M} = c_M \times \frac{\sigma_{r_s}}{r_s}$. To obtain the concentration in the simulations, I measure the dark matter profiles in different mass bins, then fit a NFW profile [Navarro et al. \(1997b\)](#):

$$\rho(r) = \frac{\rho_0}{\frac{r}{r_s} \left(1 + \frac{r}{r_s}\right)^2}, \quad (4.21)$$

to obtain ρ_0 and r_s in every mass bin. Finally, as $c_M = r_v/r_s$ and the virial radius is known, I derive the c_M value for each mass bin. At both redshifts, the concentration decreases as the mass increases, at a faster rate in the simulations compared to the prediction. When increasing the redshift, the concentration decreases in the simulations and in the **HMx** prediction, leaving the measured simulation higher than the one used in the model.

4.4.2 Implication for the power spectrum and pressure profile

Impact on the pressure power spectrum Now that we have measured and compared the value of various parameters in the **Horizon-AGN** simulation, it is interesting to examine their impact on the pressure auto-power spectrum. First, I assess the influence of each parameter by adjusting their values to the best fit, followed by analysing the collective impact of changing all parameters simultaneously. Fig. 4.18 shows the impact of changing different parameters to the predicted pressure auto-power spectrum. The **Horizon-AGN** power spectrum is in red line, and the default **HMx** prediction used in the above analysis is in thick grey line. The different black lines show the predictions when changing different parameters. The best fit for Γ is represented in dashed line, for α in dotted-dashed line, for M_0 , and β in dotted line and the prediction when changing all parameters simultaneously with the solid line. From the measurements and best fits, we see that changing α and M_0 primarily drives the differences. Modifying α and M_0 enables the prediction to have more power, thereby bringing it closer to the **Horizon-AGN** simulations, while Γ reduces the prediction even more at high k values. When combining the value of all the parameters, we observe an improvement in the prediction up to $k \sim 3 - 4 h \text{Mpc}^{-1}$, beyond which the performance deteriorates. Improving the prediction through this procedure is still insufficient for robust modelling of the tSZ for future, for instance, 10x2pt analyses. Nevertheless, it provides valuable insights into the role of different parameters and possible degeneracies.

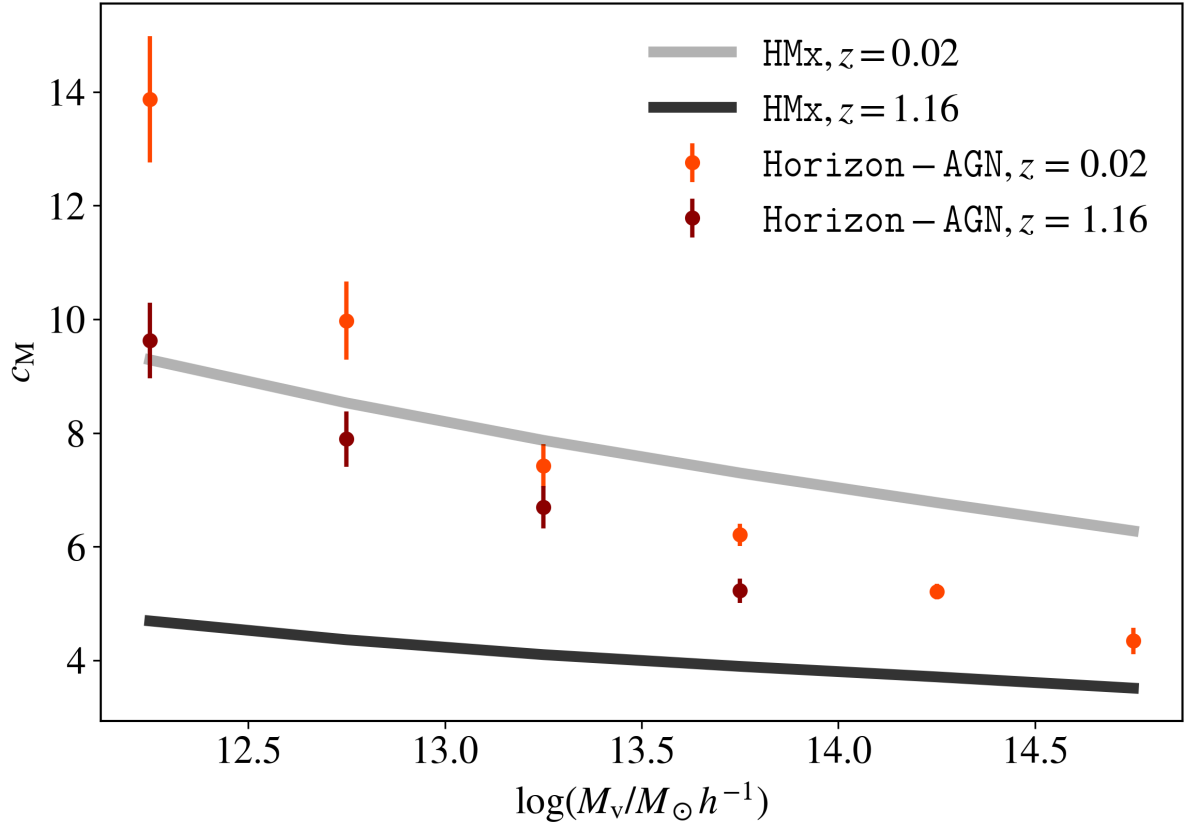


Figure 4.17: Measurement of the concentration in the **Horizon-AGN** simulation compared to the equation used in **HMx**. Measurements at $z = 0.02$ in different mass bins are in red and must be compared to the equation used in **HMx** in a thick grey line. Measurements at $z = 1.16$ in the different mass bins are in dark red and must be compared to the equation used in **HMx** in a thick black line. Error bars are propagated from the error bars on r_s given by the square root of the associated value in the covariance matrix given by the fit.

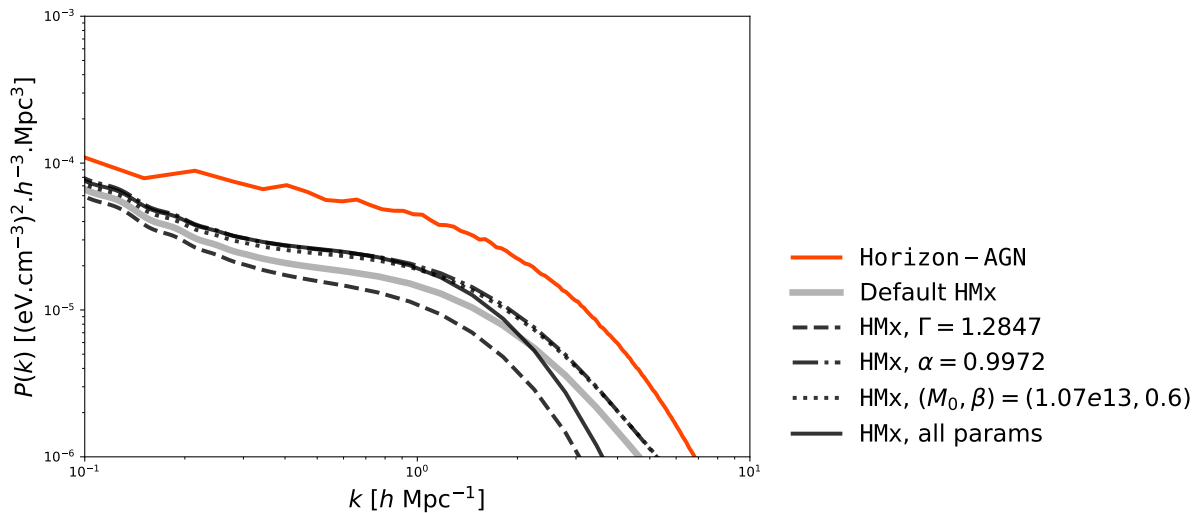


Figure 4.18: Pressure auto-power spectrum at $z = 0.02$ with different predictions. The red line shows the measured power spectrum in the **Horizon-AGN** simulation. The thick grey line shows the default **HMx** prediction used in the above analysis. The different black lines show the power spectrum predicted by **HMx** when changing the value of different parameters. We put our best fit for Γ in dashed line, for α in dotted-dashed line, for M_0 and β in dotted line and the prediction when changing all of them with the solid line.

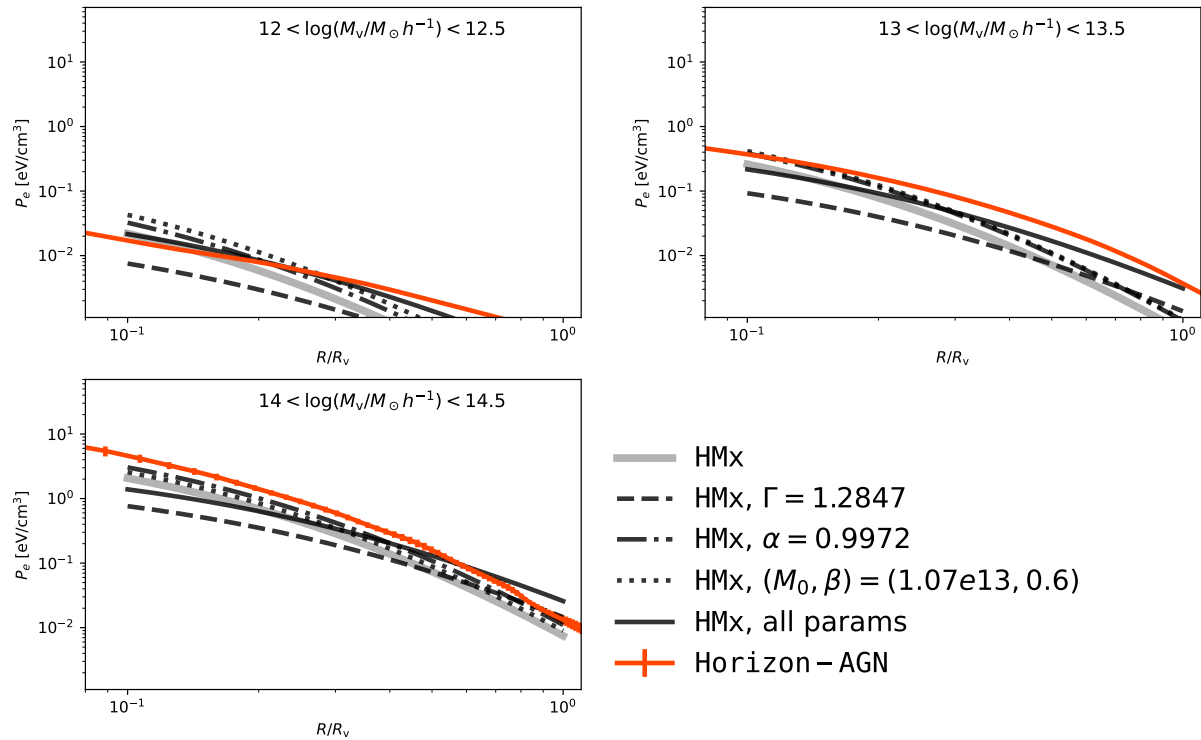


Figure 4.19: Pressure profile at $z = 0.02$ with different predictions. The red line shows the measured power spectrum in the `Horizon-AGN` simulation with its associated error bars. The thick grey line shows the default `HMx` prediction used in the above analysis. The different black lines show the power spectrum predicted by `HMx` when changing the value of different parameters. We put our best fit for Γ in dashed line, for α in dotted-dashed line, for M_0 and β in dotted line and the prediction when changing all of them with the solid line.

Impact on the pressure profiles As the profiles in `HMx` are not expected to match perfectly due to the parameters being fitted at the level of the power spectrum, it is interesting to examine how they change when using the value obtained from our best fit at $z = 0.02$ to see behaviour and potential degeneracies. Fig. 4.19 shows the impact of changing different parameters to the induced pressure profile in `HMx`. The `Horizon-AGN` pressure profile is in red line, and the default `HMx` profile used in the above analysis is in thick grey line. The different black lines show the profiles when changing different parameters. The best fit for Γ is represented in dashed line, for α in dotted-dashed line, for M_0 and β in dotted line, and the prediction when changing all parameters simultaneously with the solid line. As for the power spectrum analysis, modifying the value of α and M_0 tends to give more power, while Γ lowers the profiles across most of the radius range examined. The profiles obtained when changing all parameters demonstrate better agreement than the default one, but the agreement is still not compelling. These modifications are insufficient to fully capture all the physics present in the simulation, highlighting the importance of continued efforts to achieve a profile that completely matches the simulation and see if it leads to a better match in the power spectrum. Moreover, given the observed mass dependence on Γ and α , it can be beneficial to incorporate it. I did not correct the concentration used in `HMx` to align with the measured one, which could also help improve the agreement between the profiles. Nevertheless, it is noteworthy to observe the extent to which these parameters can influence the profile.

In the effort to build a halo model more closely aligned with the intrinsic physical properties of halos, I have worked on an initial investigation by measuring parameters in the `Horizon-AGN` simulation to update the values used in `HMx`. I found that using the best fit from the highest mass at $z = 0.02$ improves the predictions on some scales, but it is insufficient to fully reproduce the simulation's measurements. To pursue this analysis, it would be interesting to incorporate the mass and redshift dependence of the parameters to determine whether perfect alignment of these parameters leads to a better agreement in the power spectrum and pressure profiles. Additionally, exploring if a perfect match between predicted and measured pressure profiles can result in a sufficiently good agreement in the power spectrum would be of interest. If this is the case, obtaining a power spectrum from profile observations, for example, can be beneficial.

Finally, following the procedure described in `HMx` by fitting at the power spectrum level and comparing

the parameters' values with those measured directly, as done in this section, could provide insights. It could enhance our understanding of the correlation between the parameters and the impact of the different physics included in the simulations.

4.5 Beyond standard cosmology

The tSZ effect is expected to be sensitive to the dark energy due to its influence on the growth of structures. Before attempting to constrain w using the tSZ effect, it is crucial to understand how this effect, thus the electron pressure, is affected and to establish a robust model. To initiate this investigation, my focus is on a scenario where the dark energy equation of state is constant, parametrized solely by w . In Eq. (1.11), we adopt $w_a = 0$ and $w_0 = w$.

To explore the impact of dark energy, I ran and used the L896_ w CDM simulations, described in Sect. 3.3.1. For this initial exploration, I chose values of $w = -1.2$, $w = -1$, and $w = -0.8$. It is noteworthy that current constraints typically favour w values close to -1 Planck Collaboration et al. (2020b). In particular, values of $w < -1$ would necessitate more complex models to be explained, such as the phantom dark energy (e.g., Nojiri et al., 2005) model, which is not favoured by the current observations. Such a model is physically more complex as it requires negative mass.

4.5.1 Results on the power spectrum

I start the investigation by studying the matter and pressure power spectra within these three simulations. It is interesting to evaluate the impact in both tracers, discerning the direction and the magnitude of the effect.

Matter power spectrum

I have first examined the matter power spectrum in the L896_ w CDM simulations. The left part of Fig. 4.20 shows the matter power spectrum as a function of redshift. The top panel shows the power spectrum for the different values of w , while the bottom panel shows the fractional difference with respect to the $w = -1$ simulations. The power spectra go from $z = 0.02$ in dark blue to $z = 5.87$ in yellow. The $w = -1$ results are represented with the solid line, $w = -0.8$ by the dashed line, and $w = -1.2$ by the dotted line.

We observe that the power spectrum has more power when $w = -0.8$ and less power when $w = -1.2$ compared to the $w = -1$ simulation. The difference amplifies with increasing redshift, particularly at higher k values. At the highest redshift considered ($z = 5.87$), the $w = -0.8$ simulation shows between 15% to 30% more power than the $w = -1$ simulation, while the $w = -1.2$ simulation shows between 10% to 20% less power. There is an asymmetry between simulations with $w < -1$ and $w > -1$, with $w > -1$ simulations showing roughly twice the difference observed between the $w < -1$ simulation and the $w = -1$ baseline.

Pressure power spectrum

I pursue the analysis with the pressure power spectrum in the L896_ w CDM simulations. The right panel of Fig. 4.20 shows the pressure power spectrum as a function of redshift. The top panel shows the power spectrum for the different values of w , while the bottom panel shows the fractional difference with respect to the $w = -1$ simulations. The power spectra go from $z = 0.02$ in dark blue to $z = 5.87$ in yellow. The $w = -1$ results are represented with the solid line, for $w = -0.8$ by the dashed line, and for $w = -1.2$ by the dotted line.

Similarly to the matter power spectrum, the pressure power spectrum has more power when $w = -0.8$ and less power when $w = -1.2$ compared to the $w = -1$ simulation. This trend also increases with redshift but demonstrates a more consistent pattern across the k values. At the highest redshift considered ($z = 5.87$), the $w = -0.8$ simulation shows over 100% more power than the $w = -1$ simulation, while the $w = -1.2$ simulations show between 20% to 50% less power. The same asymmetry as in the matter spectrum is observed: the $w > -1$ simulations show approximately twice the difference observed between the $w < -1$ simulation and the $w = -1$ simulation.

4.5.2 Result on the pressure angular power spectrum

Following the computation of the power spectrum, I computed the pressure angular power spectrum to assess the impact of the w CDM simulations on the tSZ effect. The top panel of Fig. 4.21 shows the pressure angular power spectrum integrated between $z = 0.02$ and $z = 4$ as a function of multipole ℓ . The bottom panel shows the fractional difference relative to the $w = -1$ simulation. Results for $w = -1$ are

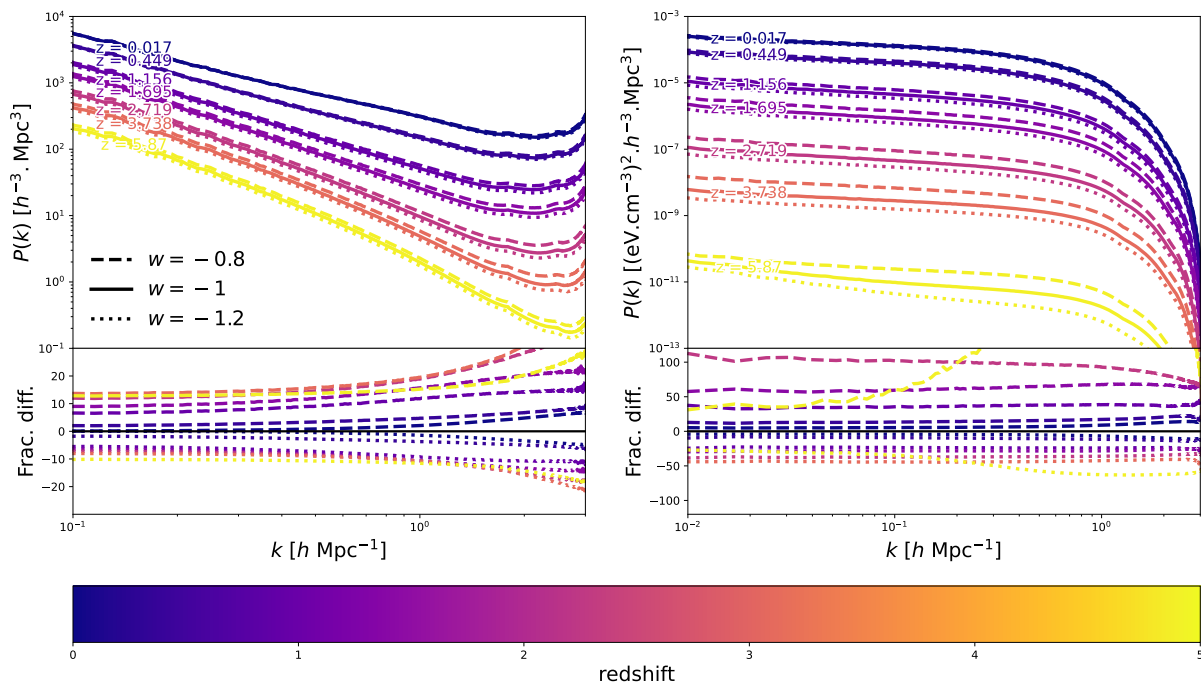


Figure 4.20: Top left (right): Matter (Pressure) auto-power spectrum in the L896_wCDM simulations as a function of redshift. We compare the result of simulations with different values of w . Bottom: fractional difference for the $w \neq -1$ with respect to the $w = -1$ simulations. Results for $w = -0.8$ are in dashed line, for $w = -1$ are in solid line, and for $w = -1.2$ in dotted line. The power spectra go from $z = 0.02$ in dark blue to $z = 5.87$ in yellow. Note: The range of the y-axis is different for the matter and pressure power spectrum and fractional difference, to better distinguish the trends and differences.

represented in black, $w = -0.8$ in blue, and $w = -1.2$ in red. The $w = -0.8$ simulation shows between 20% and 80% more power than the $w = -1$ simulation in the pressure angular spectrum. Conversely, the $w = -1.2$ simulation shows between 15% and 40% less power compared to the $w = -1$ simulation. In both cases, these differences increase with increasing ℓ , with the most significant differences at higher ℓ values.

4.5.3 Discussion

To understand the trends observed in the matter and pressure power spectra, we can investigate the contributions of the different components of the Universe across the various w scenarios. This is illustrated in Fig. 4.22, where the density as a function of the scale factor is qualitatively represented for different w values, excluding the radiation dominated era. The matter density is represented by the dashed line, the density of the cosmological constant $w = -1$ with the solid black line, and the density of the cosmological constant for $w = -0.8$ and $w = -1.2$ by the blue and red lines, respectively. The normalisation is set $z = 0$, corresponding to $a = 1$, as in the simulations. The matter dominated era and the cosmological constant dominated era are represented with respect to the case $w = -1$. To understand the impact of these different cosmologies on the spectra, we can consider the growth of structures in the different scenarios through the investigation of the dominant component of the Universe.

In the Λ CDM scenario ($w = -1$), during the matter dominated era, the expansion of the Universe goes at a rate comparable to the structure formation. However, for $w \neq -1$, this is no longer true. For $w > -1$ (e.g., $w = -0.8$), the expansion of the Universe starts earlier because the density of dark energy becomes more important than matter earlier. Consequently, the growth of structures is slower than the expansion of the Universe, causing structures to move away from each other.

Conversely, for $w < -1$ (e.g., $w = -1.2$), the dominance of expansion occurs later, and the growth of structures is faster compared to the expansion of the Universe.

If we impose the same σ_8 at present time, the $w = -0.8$ scenario implies slower structure growth. This is illustrated in Fig. 4.23, where the evolution of the growth factor as a function of the scale factor is represented for the three L896_wCDM simulations. The evolution for $w = -0.8$, $w = -1$, and $w = -1.2$ are shown in blue, black, and red, respectively. The larger the scale factor, the more differences appear in the growth factor. For larger w values ($w = -0.8$), the growth factor is smaller. However, to maintain the same structure density today, structures were more developed earlier to compensate, spending more time

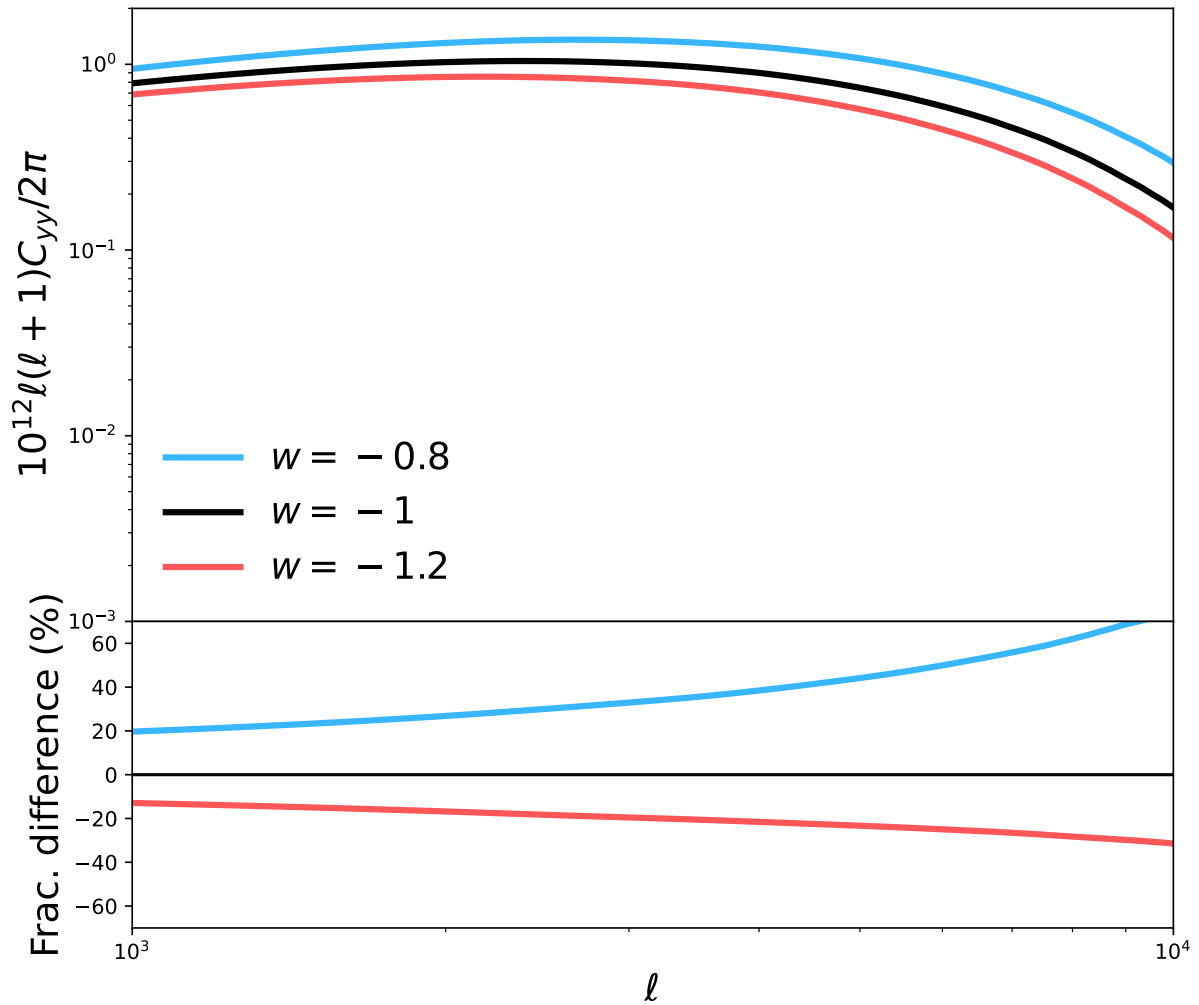


Figure 4.21: Top: Pressure angular power spectrum integrated between $z = 0.02$ and $z = 4$ in the L896_ w CDM simulations as a function of ℓ . We compare the result of simulations with different values of w . Bottom: fractional difference for the $w \neq -1$ with respect to the $w = -1$ simulations. Results for $w = -1$ are in black, for $w = -0.8$ in blue, and for $w = -1.2$ in red.

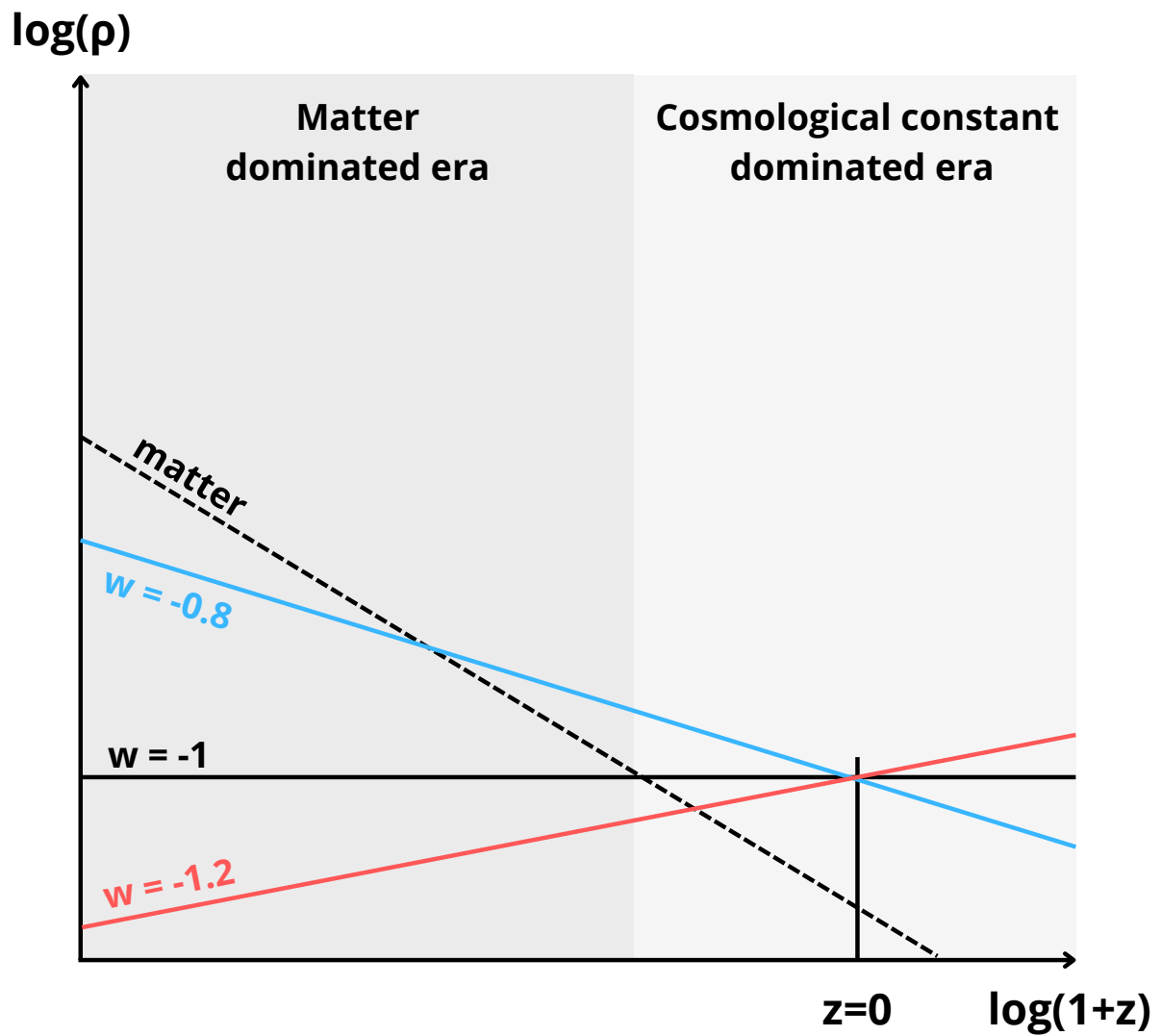


Figure 4.22: Schematic representation of the density as a function of scale factor for different w , assuming a normalisation at $z = 0$, equivalently $a = 1$. The matter density is represented in dashed line ($\rho \propto a^{-4}$), and the cosmological constant in solid black line ($\rho \sim \text{constant}$, and $w = -1$). The blue and red lines represent the density for a cosmological constant with $w = -0.8$ and $w = -1.2$, respectively. The matter dominated era and cosmological constant dominated era are represented with respect to the case $w = -1$.

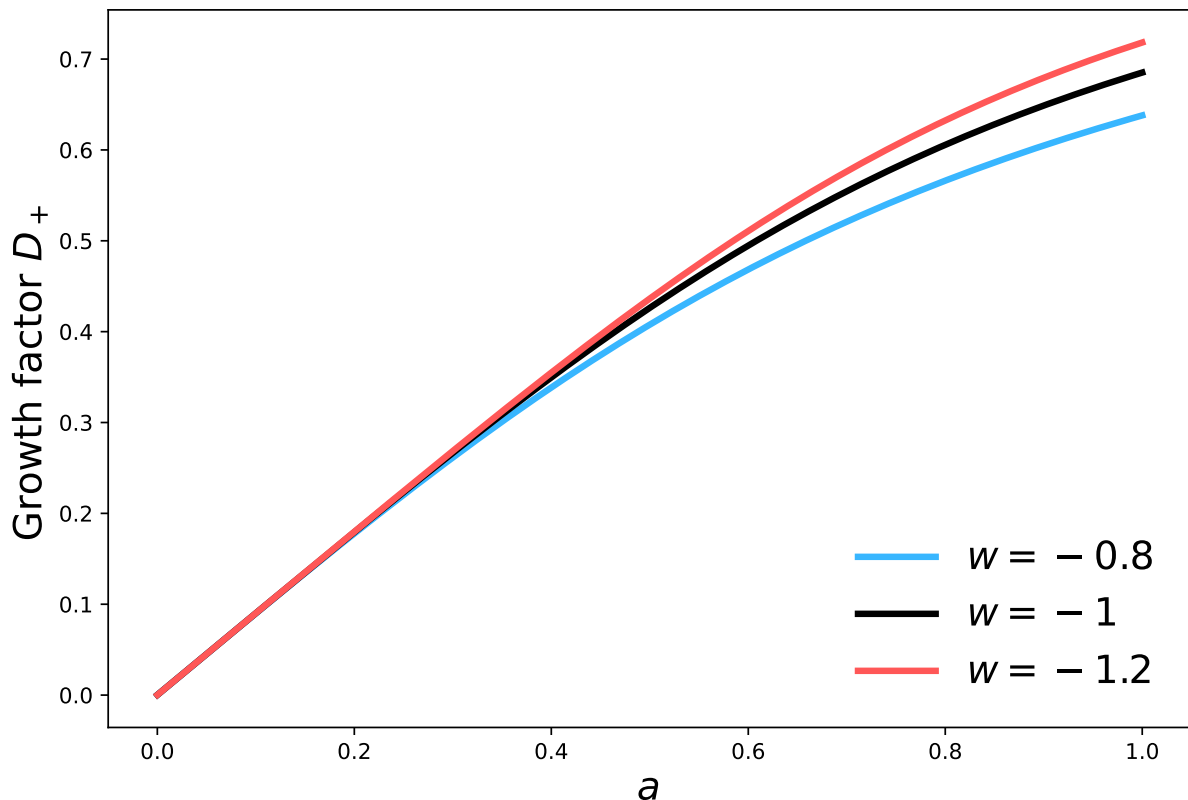


Figure 4.23: Evolution of the growth factor D_+ as a function of the scale factor a used in the L896_wCDM simulations. The evolution for $w = -0.8$, $w = -1$, and $w = -1.2$ are shown in blue, black, and red, respectively.

in the non-linear regime. This trend is reflected in the power spectrum, where the $w = -0.8$ simulation exhibits greater power, with differences increasing as redshift increases. The opposite trend is observed for $w = -1.2$: faster structure growth means structures were less developed at earlier times, resulting in a lower power spectrum.

In Fig. 4.24, I show the fractional difference at $z = 1.156$ of different quantities between the $w = -0.8$ and $w = -1$ simulations in blue and between the $w = -1.2$ and $w = -1$ simulations in red. The solid line is the ratio of the growth factor D_+ normalised at $z = 0$, the dashed line is the ratio of the matter power spectrum, and the dotted line is the ratio of the pressure power spectrum. As already seen in Fig. 4.20 and emphasized here with the dashed and dotted lines, the matter power spectrum reveals a trend with k , while the pressure power spectrum ratio is more flat. The comparison between the growth factor and the pressure power spectrum confirms that the effect of dark energy cannot be explained only by the growth factor; more complex effects are at play. Further investigations on the impact of the halo mass function, shown in Fig. 3.4, could provide additional insights into understanding and characterizing the impact of dark energy on the tSZ effect.

This study is the first investigation into the influence of dark energy on the tSZ effect. Results encourage further analysis with more complex dark energy models. As a next step, developing a robust model to predict these effects can be used to constrain dark energy models with observational data. It will also be interesting to explore if a combination of some simple quantities, such as the growth factor and halo mass function, can fully describe the impact of dark energy on the spectrum.

4.6 Summary

In this chapter, I explored how the tSZ effect can serve as a tool for constraining both cosmological and astrophysical parameters. To improve its constraining power, it is interesting to examine its cross-correlation with various observables. However, due to imperfect modelling of the tSZ effect and the introduction of numerous parameters into the pipeline, it is crucial to reduce the priors on these parameters to enhance the constraining power (Fang et al., 2024). Therefore, continued efforts to refine the tSZ modelisation are essential. With this objective, I conducted an analysis of the tSZ properties within the

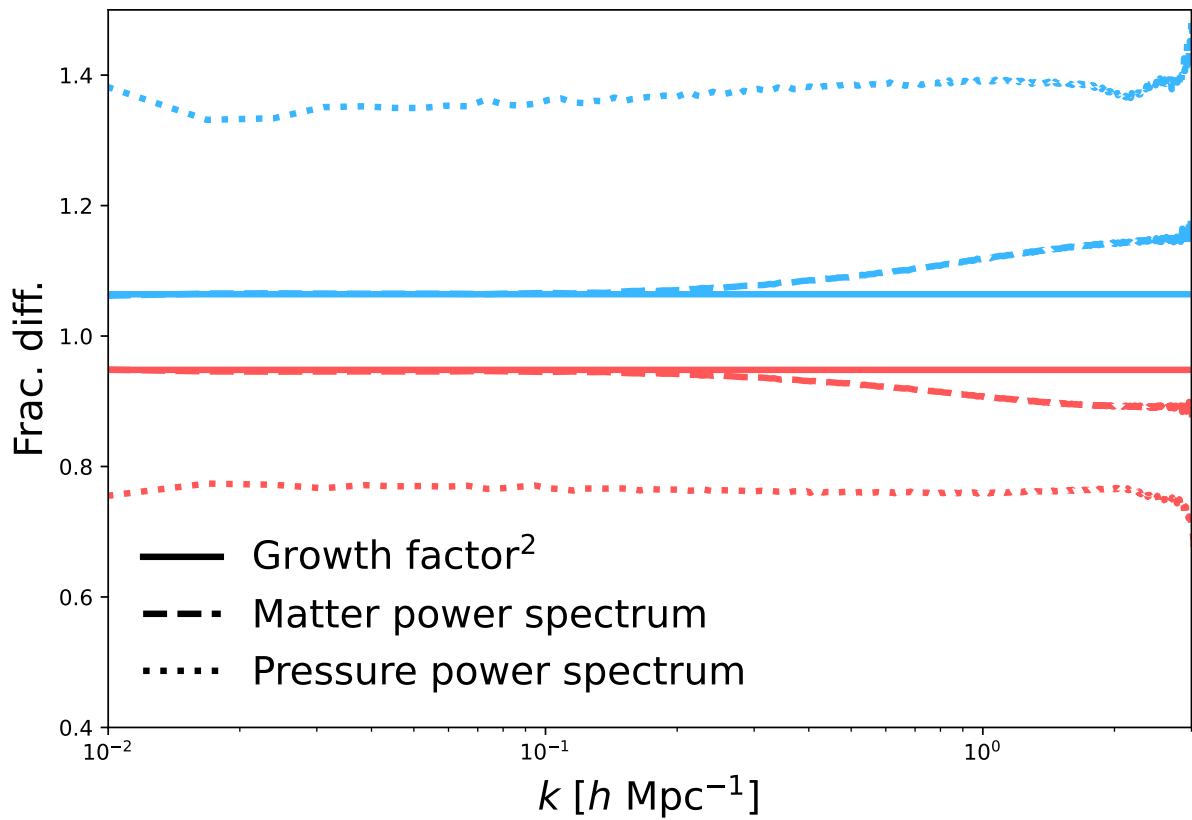


Figure 4.24: Fractional difference at $z = 1.156$ of different quantities, with respect to the $w = -1$ simulation. Ratio for $w = -0.8$ are blue for $w = -1.2$ in red. The solid line is the ratio of the growth factor D_+ normalised at $z = 0$, the dashed line is the ratio of the matter power spectrum, and the dotted line is the ratio of the pressure power spectrum.

Horizon and Magneticum suite of simulations and compared the results with the HMx predictions.

The first main result compares the power spectrum and angular power spectrum. The angular power spectrum measured in the simulations differs between 20 to 50% compared to the predictions provided by HMx. This discrepancy is attributed to variations in the pressure auto-power spectrum, which become more pronounced with increasing redshifts. Specifically, the predicted power spectrum shows higher power levels compared to the observed values in the simulations.

To understand the origin of these discrepancies, I explored different parameters and limitations inherent to the halo model. First, I have studied the one- and two-halo decomposition and observed an overestimation of the two-halo term at higher redshifts, which can be a problem in modelling the ICM. This contributes significantly to the excess power observed in the predictions. Second, I investigated the contribution of power from within or outside one virial radius of halos in the simulations at different redshifts. While at lower redshifts, the majority of power comes from within halos, indicating compatibility with the halo model assumption, at higher redshift (e.g., $z \sim 3$), a substantial portion of the power comes from regions outside one virial radius of halos, highlighting a failure of the halo model to capture this phenomenon. Then, I studied the importance of each mass bin to the total power spectrum in the simulations. Remarkably, almost all power comes from the highest mass bin ($14 < \log(M_v/M_\odot h^{-1}) < 15$) despite this bin containing only 0.006% of the total number of halos. Finally, I compared the pressure profiles for different mass bins and redshifts. While differences were anticipated due to the halo model being fitted at the level of the response power spectrum, it was still interesting to observe the behaviour and degeneracies. Across most mass and redshift ranges, the simulations have higher power levels and distinct profile shapes compared to the prediction. Future studies could focus on refining predicted profiles to better match those of the simulations, potentially addressing some of the observed discrepancies in the (angular) power spectrum.

Continuing the investigation into the parameters of the halo model, I compared the values of Γ , M_0 , β , and α measured in the Horizon-AGN simulation with those used in HMx. An important observation is that in HMx, these parameters do not depend on redshift and mass, whereas we clearly see a mass and redshift trend in the simulations. To assess the impact of these parameters on the predicted power spectrum and pressure profile, we use the best fit values for M_0 and β , and the average values from the four highest mass bins at redshift $z = 0$ for Γ and α . Updating the model with these values resulted in an improved agreement between the predictions and measurements. However, this adjustment alone is insufficient to remove all discrepancies. Understanding which parameters exert the greatest influence represents a crucial step towards developing a more robust and physically motivated tSZ model.

Finally, I investigated the influence of dark energy on the tSZ effect with the L896_wCDM simulations. Matter and pressure (angular) power spectra are affected by the value of w . Specifically, the $w = -0.8$ simulation shows enhanced power, while the $w = -1.2$ simulation shows reduced power compared to the $w = -1$ simulation. This is attributed to the impact of w on the growth of structure, considering our normalisation with respect to σ_8 today. This study represents an initial step in understanding the impact of dark energy on the tSZ effect and opens the door to future investigations with more complex dark energy models.

CHAPTER 5

THE COSMIC WEB AND T-WEB CLASSIFIER

5.1	Overview	86
5.1.1	Towards the non-linear regime	86
5.1.2	Cosmic web	87
5.1.3	Motivation of my work	91
5.2	"A theoretical view of the T-web statistical description of the cosmic web" (article)	91

To describe the Universe, it becomes crucial to consider non-linearities, especially at smaller scales and lower redshifts. One intriguing aspect of cosmic structures lies in their distribution throughout the Universe. The cosmic web, a network of structures arising from the gravitational collapse of the initial density fluctuations, is composed of four distinct environments: voids, walls, filaments, and knots. Identifying these environments can be useful to study various aspects, such as galaxy properties as a function of their environments. Learning more about galaxy properties can provide insights into their formation and evolution, which is a crucial aspect of cosmology. Galaxies are not uniformly distributed throughout the Universe, and studying the cosmic web can help to extract higher-order information compared to using two-point statistics such as the power spectrum. Additionally, the cosmic web can help us understand the impact of the environment on galaxy properties, such as morphology, star formation, spin, etc.

In this chapter, I will start in Sect. 5.1 with an overview of perturbation theory and a description of different methods applicable for classifying the cosmic web. In Sect. 5.2, I will describe and copy the article, of which I am the primary author, elucidating a theoretical prediction of environmental abundance using the T-web classifier.

5.1 Overview

In this section, I will provide an overview of the main equations of perturbation theory, serving as a foundational basis for the article presented in Sect. 5.2. I will then describe various approaches applicable for classifying the four cosmic web environments, with a particular focus on the T-web classifier, which is the one used in the article. I will finish the overview with a brief motivation for the article.

5.1.1 Towards the non-linear regime

In Sect. 1.3, I outlined the formation and evolution of structures within a homogeneous Universe, and I discuss the Zel'dovich approximation and spherical collapse model in Sect. 1.7 as an approximation to solve the evolution of structures in an inhomogeneous Universe. To go one step further and address the non-linearities, we can use perturbation theory. Non-linearities become significant as the redshift decreases because the matter distribution becomes increasingly asymmetric between regions of positive and negative density contrast. Non-Gaussianities in the density field grow in significance, rendering the linear approximation invalid. The scale where non-linearities become significant is presented in Sect. 1.7.3. Perturbation theory applies when the density contrast δ is small compared to unity ($\delta \ll 1$), thus $\sigma \lesssim \sigma_8$. Perturbation theory is generally used for $\sigma \lesssim 0.1 - 0.5$.

Perturbation theory

This section was written with the help of [Bernardeau et al. \(2002\)](#); [Bernardeau \(2007\)](#); [Peter & Uzan \(2013\)](#); [Codis-Decara \(2015\)](#); [Dodelson & Schmidt \(2020\)](#). All equations provided hereafter are derived for an Einstein de Sitter universe. Beginning with the Poisson equation, mass conservation, and Euler equation (Sect. 1.7.1) and assuming no shell-crossing, we can neglect the rotational part of the velocity field and compute the divergence of the Euler equation. In Fourier space, we thus obtain:

$$a \frac{\partial \delta(\mathbf{k}, a)}{\partial a} + \theta(\mathbf{k}, a) = \frac{-1}{(2\pi)^{3/2}} \int d^3 \mathbf{k}_1 d^3 \mathbf{k}_2 \delta_D(\mathbf{k} - \mathbf{k}_1 - \mathbf{k}_2) \alpha(\mathbf{k}_1, \mathbf{k}_2) \theta(\mathbf{k}_1, a) \delta(\mathbf{k}_2, a), \quad (5.1)$$

$$a \frac{\partial \theta(\mathbf{k}, a)}{\partial a} + \frac{1}{2} \theta(\mathbf{k}, a) + \frac{3}{2} \delta(\mathbf{k}, a) = \frac{-1}{(2\pi)^{3/2}} \int d^3 \mathbf{k}_1 d^3 \mathbf{k}_2 \delta_D(\mathbf{k} - \mathbf{k}_1 - \mathbf{k}_2) \beta(\mathbf{k}_1, \mathbf{k}_2) \theta(\mathbf{k}_1, a) \theta(\mathbf{k}_2, a), \quad (5.2)$$

where:

$$\alpha(\mathbf{k}_1, \mathbf{k}_2) = 1 + \frac{\mathbf{k}_1 \cdot \mathbf{k}_2}{k_1^2}, \quad (5.3)$$

$$\beta(\mathbf{k}_1, \mathbf{k}_2) = \frac{\mathbf{k}_1 \cdot \mathbf{k}_2}{2k_1^2} + \frac{\mathbf{k}_1 \cdot \mathbf{k}_2}{2k_2^2} + \frac{(\mathbf{k}_1 \cdot \mathbf{k}_2)^2}{k_1^2 k_2^2}, \quad (5.4)$$

and where we recall that δ is the density contrast and θ the dimensionless velocity divergence. These equations encapsulate all the non-linearities inherent in the equation of motion as long as there is no shell-crossing (implying as long as the rotational part of the velocity is zero). They illustrate that the evolution relies on the coupling between different modes k .

In perturbation theory, fields can be developed perturbatively at each order as a function of the initial field, both for density and velocity divergence:

$$\delta(\mathbf{x}, t) = \sum_n \delta^{(n)}(\mathbf{x}, t), \quad \theta(\mathbf{x}, t) = \sum_n \theta^{(n)}(\mathbf{x}, t), \quad (5.5)$$

where n is the order of the perturbation. The initial conditions are determined by the linear theory, and the evolution of the perturbation is governed by the equations derived earlier (Eqs. 5.1 to 5.4).

The time dependence of the growing mode of each term of order n follows a^n . Therefore, for each order n , perturbation theory yields:

$$\delta^{(n)}(\mathbf{x}) = \int \frac{d^3 \mathbf{k}_1}{(2\pi)^{3/2}} \delta(\mathbf{k}_1) \dots \frac{d^3 \mathbf{k}_n}{(2\pi)^{3/2}} \delta(\mathbf{k}_n) a^n F_n(\mathbf{k}_1, \dots, \mathbf{k}_n), \quad (5.6)$$

$$\theta^{(n)}(\mathbf{x}) = \int \frac{d^3 \mathbf{k}_1}{(2\pi)^{3/2}} \delta(\mathbf{k}_1) \dots \frac{d^3 \mathbf{k}_n}{(2\pi)^{3/2}} \delta(\mathbf{k}_n) a^n G_n(\mathbf{k}_1, \dots, \mathbf{k}_n) \quad (5.7)$$

where F_n and G_n are two kernels of order n . The kernels satisfy the property $F_1 = -G_1 = 1$, and for $n > 2$, can be expressed as:

$$F_n(\mathbf{k}_1, \dots, \mathbf{k}_n) = \sum_{i=1}^{n-1} \frac{G_i(\mathbf{k}_1, \dots, \mathbf{k}_i)}{(2n+3)(n-1)} [- (2n+1)\alpha(\mathbf{p}_1, \mathbf{p}_2) F_{n-i}(\mathbf{k}_{i+1}, \dots, \mathbf{k}_n) + 2\beta(\mathbf{p}_1, \mathbf{p}_2) G_{n-i}(\mathbf{k}_{i+1}, \dots, \mathbf{k}_n)], \quad (5.8)$$

$$G_n(\mathbf{k}_1, \dots, \mathbf{k}_n) = \sum_{i=1}^{n-1} \frac{G_i(\mathbf{k}_1, \dots, \mathbf{k}_i)}{(2n+3)(n-1)} [3\alpha(\mathbf{p}_1, \mathbf{p}_2) F_{n-i}(\mathbf{k}_{i+1}, \dots, \mathbf{k}_n) - 2n\beta(\mathbf{p}_1, \mathbf{p}_2) G_{n-i}(\mathbf{k}_{i+1}, \dots, \mathbf{k}_n)], \quad (5.9)$$

with $\mathbf{p}_1 = \mathbf{k}_1 + \dots + \mathbf{k}_i$ and $\mathbf{p}_2 = \mathbf{k}_{i+1} + \dots + \mathbf{k}_n$.

Skewness

One famous example of a non-linear effect is the skewness of the density field. It has been widely investigated in various works, including the one of Peebles (1980); Juszkiewicz et al. (1993); Lokas et al. (1995); Colombi et al. (2000), and I will summarise its main properties here.

The skewness arises from the third-order moment of the density field, which can be expressed as:

$$\langle \delta_1 \delta_2 \delta_3 \rangle = \left\langle \prod_{i=1}^3 (\delta_i^{(1)} + \delta_i^{(2)} + \dots) \right\rangle = \left\langle \prod_{i=1}^3 \delta_i^{(1)} \right\rangle + \sum_{j=1}^3 \left\langle \delta_j^{(2)} \prod_{i \neq j} \delta_i^{(1)} \right\rangle + \dots, \quad (5.10)$$

where the terms have been reorganised in increasing powers of the local linear density in the last expression, and the following terms are of higher order. For Gaussian initial conditions, the first term is equal to zero, and the second term is thus the leading order.

At order $n = 3$, under the assumption of Gaussian initial conditions and after applying a few simplifications, including the application of the Wick theorem mentioned in Eq. (2.8), the expression then becomes:

$$\langle \delta^3 \rangle \approx 3 \left\langle (\delta^{(1)})^2 \delta^{(2)} \right\rangle \quad (5.11)$$

$$= 6a^4 \int d^3 \mathbf{k}_1 \int d^3 \mathbf{k}_2 P(k_1) P(k_2) F_2(\mathbf{k}_1, \mathbf{k}_2), \quad (5.12)$$

where F_2 can be obtained from Eq. (5.8). Further details about the computation can be found in Appendix A of the article presented in the following section.

After integration, we obtain $\langle \delta^3 \rangle = 34/7 \langle \delta^2 \rangle^2$. The third moment can be rescaled to obtain the skewness, which is constant in time at tree-order (Peebles, 1980):

$$S_3 \equiv \frac{\langle \delta^3 \rangle}{\langle \delta^2 \rangle^2} = \frac{34}{7} + O(\sigma^2). \quad (5.13)$$

The skewness quantifies the asymmetry between underdense and overdense regions, which becomes evident in the high-density tail of the probability distribution function of the density. To compare the skewness with observations, it is necessary to incorporate a smoothing filter, which is what is done in the article.

5.1.2 Cosmic web

Initial fluctuations in the Universe evolve through gravitational collapse, sculpting a complex network of structures known as the cosmic web. Pioneering work by Zel'dovich (1970) laid the groundwork for its description (also see Sect. 1.7.2). The Zel'dovich approximation describes the first-order ballistic trajectories of particles with a Lagrangian framework. It predicts that matter streams along 'pancakes' that are elongated structures, often called walls, as the gravitational collapse takes effect. Over time, these walls

intersect and form filaments, which emerge as the largest cosmic structures. Matter continues to converge along these filaments, eventually coalescing into dense knots at their intersections. Voids represent the underdense regions of space and are surrounded by walls, filaments, and knots. Consequently, the cosmic web comprises four environments, ranging from the least dense to the densest: voids, walls, filaments, and knots (also referred to as clusters). The probability distribution function of the environments exhibits some overlap, as some regions can be classified as multiple environments. Extensive literature has explored the formation and evolution of these structures (see references in the introduction of the article presented in the next section). Studying the properties of galaxies or dark matter halos within these environments offers valuable cosmological and astrophysical insights and is still an active field. Various classification schemes have been developed to study the cosmic web and its impact on galaxies or dark matter halos. These classifiers can be categorized into different groups, and I will describe them, largely following the groups defined in Libeskind et al. (2017). I will first describe the Hessian methods including the T-web classification used in the article. Then, I will provide a brief overview, cite a few examples of other classifiers, and compare the various classifiers.

Hessian-based methods

Pioneering efforts to classify cosmic web structures using the Hessian of the gravitational potential were undertaken by Doroshkevich (1970), using the properties of the tidal field. This work laid the foundation for subsequent research, including the contribution by Hahn et al. (2007a), which opened the door to two classifications that are now widely used: the V-web and T-web.

In their work, Hahn et al. (2007a) implement a method based on the eigenvalues of a deformation tensor, which characterizes the geometrical nature of each point in space. The approach is based on the local stability criterion for the orbits of test particles and is closely connected to the Zel'dovich approximation. In this method, the deformation tensor is the tidal field, defined as:

$$T_{ij} \equiv \partial_i \partial_j \Phi, \quad (5.14)$$

where Φ is the gravitational potential and $i, j = 1, 2, 3$ are the coordinates.

The eigenvalues of the tidal field are denoted as $\lambda_{1,2,3}$ with $\lambda_1 < \lambda_2 < \lambda_3$. The classification is based on the number of positive eigenvalues: a positive (negative) eigenvalue represents gravitational contraction (expansion). This criterion is also equivalent to the dimension of the stable manifold at the point. Therefore, analogous to Zel'dovich, we have the following criteria:

- void: $\lambda_1 < \lambda_2 < \lambda_3 < 0$, equivalent to unstable orbits,
- wall: $\lambda_1 < \lambda_2 < 0 < \lambda_3$, equivalent to 1D stable manifold,
- filament: $\lambda_1 < 0 < \lambda_2 < \lambda_3$, equivalent to 2D stable manifold,
- knot: $0 < \lambda_1 < \lambda_2 < \lambda_3$, equivalent to an attractive fixed points..

In this method, the only free parameter is the smoothing scale R_s . Hahn et al. (2007a) applied this approach to dark matter simulations and confirmed that when the smoothing scale is sufficiently large ($R_s \gg 10 h^{-1} \text{Mpc}$), the density field approaches a Gaussian distribution, and the results converge towards the analytical predictions of Doroshkevich (1970). With increasing smoothing scale, several changes are observed: the number of halos in voids increases at the expense of the surrounding sheets, the size of massive clusters at the intersection of filaments increases at the expense of the ending points, and the clusters themselves become filamentary structures. This method can be used to investigate the formation history of galaxies and their properties as a function of their environments, as demonstrated by Hahn et al. (2007b), studying dark matter halos at $z = 1$.

Using these principles, Forero-Romero et al. (2009) developed a dynamical classification known as the T-web. This classification is based on the tidal shear tensor T , which is the tensor of the second derivatives of the gravitational potential:

$$T_{ij} = \frac{\partial^2 \Phi}{\partial r_i \partial r_j}, \quad (5.15)$$

where Φ is the gravitational potential and r_i with $i = 1, 2, 3$ are the spatial coordinates.

The previous equation assumes that the matter density field is known and smoothed using a finite kernel to enable finite derivatives. This approach enhances the method proposed by Hahn et al. (2007a), addressing concerns such as insufficient void volume and a lack of satisfactory visual agreement between the classified structures and the density field. The discrepancies are related to the choice of considering the number of positive eigenvalues instead of those above a given threshold. An infinitesimally positive

eigenvalue suggests collapse, but it may not occur immediately, thus, it does not provide the classification at present time. To rectify this, [Forero-Romero et al. \(2009\)](#) advocates for evaluating the number of eigenvalues above a threshold that needs to be chosen. Consequently, this model has two free parameters: the smoothing scale R_s (as in [Hahn et al., 2007a](#)) and the threshold λ_{th} . To date, no theoretical framework exists for selecting the threshold, which is typically chosen to achieve a good visual agreement.

In their study, [Forero-Romero et al. \(2009\)](#) apply this method to dark matter simulations and investigate the impact of different choices of threshold and smoothing scale. When $\lambda_{\text{th}} = 0$ and high smoothing scale are employed, once again, the analytical results from [Doroshkevich \(1970\)](#) are found.

In terms of volume and mass filling fraction (VFF and MFF), voids show significant sensitivity to the chosen threshold, unlike other environments that demonstrate consistent behaviour across varying values of λ_{th} and R_s . As λ_{th} increases, voids expand, at the expense of sheets and filaments, and become more connected. To fix a volume filling fraction for the different environments, the choice of a threshold becomes more critical than the choice of a smoothing scale. Through their analysis of VFF and MFF, [Forero-Romero et al. \(2009\)](#) suggests that $\lambda_{\text{th}} \lesssim 0.2$ appears to be a relevant threshold value.

In the same analysis, the size and percolation of voids were also studied. The percolation is quantified by the fraction of the volume of the largest void to the total volume occupied by all the voids. Once again, a significant dependency on the threshold parameter was observed. For threshold values λ_{th} ranging between 0 and 0.3, the number of voids decays roughly exponentially. Regarding percolation, a notable transition occurred within the range $0.1 \lesssim \lambda_{\text{th}} \lesssim 0.2$, wherein the ratio abruptly shifted from ≤ 0.1 to ≥ 0.9 .

Additionally, the study explored the fragmentation of filaments, which began to fragment as the threshold increased, creating additional void space. Consequently, there exists a threshold range ($0.2 \lesssim \lambda_{\text{th}} \lesssim 0.4$) where two distinct environments coexist: percolating voids with a network of interconnected filaments. This dual characteristics is expected to be discernable in the classification of the cosmic web. This study thus concludes that a threshold $\lambda_{\text{th}} = 0.1$ appeared to provide a satisfactory visual impression. For other simulations, the best choice of threshold may differ, and the choice of λ_{th} is still an open question.

Building upon the principles established by [Hahn et al. \(2007a\)](#) and [Forero-Romero et al. \(2009\)](#), [Hoffman et al. \(2012\)](#) introduced a method called the V-web. This method relies on the velocity shear tensor Σ :

$$\Sigma_{ij} = -\frac{1}{2} \left(\frac{\partial v_i}{\partial r_j} + \frac{\partial v_j}{\partial r_i} \right) / H_0, \quad (5.16)$$

where H_0 is the Hubble constant and v the velocity.

In their examination of this method using dark matter simulations, [Hoffman et al. \(2012\)](#) observed that, once again, the two free parameters are the smoothing scale R_s and the threshold λ_{th} . The V-web method demonstrates greater sensitivity to the choice of threshold compared to the T-web and is capable of discerning finer structures. Specifically, The choice of the threshold influences the width of the filaments and their visual appearance: monolithic or fragmented. While there is a correlation between the V-web classification and local density, there is no direct one-to-one correspondence between the type of cosmic web and the density level. This implies that each environment can be encountered at various density levels.

These methods offer a valuable approach to investigate the properties of galaxies or dark matter halos within the different cosmic web environments. For instance, [Codis et al. \(2012\)](#) explored the spin of dark matter halos relative to their environment, [Nuza et al. \(2014\)](#) studied galaxy morphology, and [Metuki et al. \(2015\)](#) investigated the halo mass function and stellar mass function across various environments. Additionally, these classifications can serve as a probe of cosmology. For example, [Biswas et al. \(2010\)](#) and [Lee & Park \(2009\)](#) used voids to probe dark energy, as the signature of dark energy can manifest in the shape of voids. Moreover, they can help to understand the missing baryon problem by working on filaments, as done by [Fard et al. \(2019\)](#). Cosmological insights can also be extracted from the power spectrum of the environments and their combinations, as shown by [Bonnaire et al. \(2022\)](#). Their findings suggest that more information is encoded in the environment compared to the matter power spectrum, offering a potential avenue to break degeneracies in cosmological parameters.

An improved set of techniques to classify the cosmic web utilizes a scale-space multiscale approach based on Hessian methods. For instance, [Aragón-Calvo et al. \(2007a\)](#) introduced the Multiscale Morphology Filter (MMF) for classifying the cosmic web. This method, based on visualisation and feature extraction techniques, involves smoothing the density field at various scales and applying a morphology response filter to measure the blobness, filamentariness, or wallness of the structures. The morphology having the maximum response is the one identified. This approach can be applied to study properties of dark matter halos, such as their spin and shape, as demonstrated by [Aragón-Calvo et al. \(2007b\)](#), and to

explore morphological properties, such as the density profile of filament or the correlation a cluster’s mass and the number of connected filaments, as demonstrated by [Aragón-Calvo et al. \(2010a\)](#). Subsequently, [Cautun et al. \(2013\)](#) extended this method with the NEXUS formalism, which can be applied not only to density field but also to tidal, velocity divergence, or velocity shear fields.

Graph and percolation methods

To quantify the filamentary structure of the cosmic web, researchers have developed a method based on the minimum spanning tree (MST) of galaxies, one of the oldest methods in the field. MST, a concept of graph theory, represents the tree connecting all points in a set with the minimum total edge length. It serves to characterize the filaments of the cosmic web and examine properties such as their length, width, orientation, or density profile. This approach has been employed by, for example, [Alpaslan et al. \(2014b,a\)](#) in their Adapted Minimal Spanning Tree method.

Stochastic methods

A method based on the statistical evaluation of stochastic geometric concepts has also been used for filament detection. For instance, algorithms employing Bayesian sampling of stochastic spatial point processes have been developed for this purpose. [Tempel et al. \(2014\)](#) introduced the Bisous model, which allows the detection of filaments in galaxy surveys. This approach can be directly applied to galaxy catalogue without the need to compute the density field. For example, this Bisous algorithm was employed by [Poudel et al. \(2017\)](#) on the SDSS DR10 to investigate the properties of groups and their central galaxies as a function of the luminosity density field and the cosmic web filaments.

Topological methods

Several methods based on the connectivity and topological properties of the fields have also been developed. Using criteria based on Morse theory, which describes the spatial connectivity of the density field from its singularities, [Novikov et al. \(2006\)](#); [Sousbie et al. \(2008\)](#); [Pogosyan et al. \(2009\)](#) introduced the local skeleton as a set of critical lines that connect the density field’s maxima through saddle points. Also using Morse theory, the SpineWeb was introduced by [Aragón-Calvo et al. \(2010b\)](#). This method relies on the watershed segmentation of the cosmic density field and examines the properties of the boundaries between the watershed regions.

An extension of such a method is the DisPerSE algorithm developed by [Sousbie \(2011\)](#), which enables a coherent multiscale identification of environments, particularly filaments. This approach is entirely scale-free and parameter-free and relies on the discrete Morse theory and persistence theory. It identifies environments from the discrete distribution of particles. The algorithm returns a set of critical points, lines, surfaces, and volumes corresponding to the knots, filaments, walls, and voids, respectively. This method has been used to study filaments, such as in [Hasan et al. \(2023\)](#), which investigate galaxy quenching within environments and the impact of the distance to the nearest filament and node. One advantage of this method is its applicability to discrete data. For example, it has been applied to the Galaxy And Mass Assembly¹ (GAMA) spectroscopic survey by [Kraljic et al. \(2018\)](#), which analyses galaxy properties (such as stellar mass or specific star formation rate) in the various environments and as a function of their distances to structure.

Lagrangian-based methods

While all the previous methods are Eulerian-based, another set of techniques leverages on the phase-space structure of the evolving mass distribution. These Lagrangian approaches monitor the trajectories of mass streams within the flow field, offering insights into the formation of non-linear structures. One such method is the ORIGAMI algorithm, introduced by [Falck et al. \(2012\)](#). This technique classifies structures in phase-space by counting the number of orthogonal axes along which a Lagrangian patch undergoes a shell-crossing event, occurring when Lagrangian cells collapse or invert.

Finally, another Lagrangian-based approach was introduced by [Lavaux & Wandelt \(2010\)](#) in their DIVA algorithm. This method relies on studying the eigenvalues of the shear of the displacement field. It is quite similar to the T-web, with a key distinction: T-web is an Eulerian method, whereas DIVA is a Lagrangian one. DIVA was subsequently refined by [Leclercq et al. \(2017\)](#) with the LICH algorithm, which incorporates both potential and vortical flows.

¹<https://www.gama-survey.org/>

Comparison of the classifiers

In the study conducted by Libeskind et al. (2017), the comparison of various classifiers revealed a nuanced landscape where the optimal choice depends on the targeted cosmological inquiry. Originally, each method was crafted to address specific questions, and divergent results may arise when employing alternative techniques.

We thus summarise here the main observations of Libeskind et al. (2017) on a comparison between different classifiers (some that we mentioned previously, and others one). In terms of visual representation of structures, Hessian methods tend to give voids that dominate the underdense volume. Some classifiers do not identify knots (SpineWeb) or focus solely on identifying filaments (MST, Bisous). When analysing the density probability distribution function, a relative consensus emerges for voids, while a wide variety for the walls, sheets, and knots are observed. Regarding volume and mass filling fraction, there exists notable diversity across all environments. Finally, examining the halo assignment and mass functions, it was found that while classifiers generate halo mass functions that agree, the level of agreement in halo assignment depends on the spatial distribution of the halos, with denser regions having greater agreement.

In their investigation, Leclercq et al. (2016) also compared different classifiers based on information theory. The authors study which classifiers should be used when doing parameter inference (such as the morphology of the cosmic web), model selection (such as studying dark energy), or prediction for new observations (such as galaxy colours). A comparison between the T-web (with a threshold $\lambda_{\text{th}} = 0$), ORIGAMI, and DIVA classifications of SDSS maps (Jasche et al., 2015) is performed. For cosmic web inference, the T-web surpasses DIVA, which, in turn, outperforms ORIGAMI. The superiority of the T-web comes from the high level of information contained in clusters. Conversely, DIVA can be more robust as it is less sensitive to artefacts. For model selection, Lagrangian methods like DIVA will outperform the T-web. This advantage comes from the ability of Lagrangian methods to retain memory of the initial conditions. Finally, for galaxy colours, all classifiers have a broad agreement in the trend. T-web and ORIGAMI notably outperformed DIVA, primarily attributed to DIVA's diminished sensitivity to the local density.

As highlighted by these two analyses, the selection of a classifier will depend on the specific inquiry and the nature of the available data. In the following section, I will describe the motivation of the article, which centres on the analysis of the abundance of the different environments.

5.1.3 Motivation of my work

The exploration of the cosmic web offers a window into the formation and evolution of structures in the Universe. For example, it is interesting to explore how the properties of galaxies or dark matter halos vary across diverse environments. Consequently, it becomes important to classify environment.

In their investigation, Cui et al. (2017) worked on various cosmological simulations to probe the influence of different baryonic processes on the cosmic web, in particular in the abundance of the four different environments. The authors used the V-web, with a threshold $\lambda_{\text{th}} = 0.1$ (following Hoffman et al., 2012; Carlesi et al., 2014) and the T-web, with a threshold $\lambda_{\text{th}} = 0.01$ (to ensure visual good agreement with the V-web) classifiers and compared the results at $z = 0$. Their findings revealed that the baryonic processes have almost no impact on the large-scale structure, with the gas component tracing the filamentary structure. The study is pursued to higher redshift (up to $z = 2$) solely using the V-web in Cui et al. (2019), which leads to analogous conclusions.

Given the absence of a theoretical framework derived from fundamental principles to forecast the abundance of the four environments, the objective of the presented article is to construct such a framework. It is an interesting avenue as it can allow the use of analytical prediction instead of costly simulations for cosmological analyses.

5.2 "A theoretical view of the T-web statistical description of the cosmic web" (article)

In this section, I will briefly describe the method and main result of the article (Ayçoberry et al., 2024a) and attach it at the end of the section.

Method In this article, we present a theoretical framework to predict the abundance of the four cosmic web environments: voids, walls, filaments, and knots. Our approach uses the T-web classification, which is based on the eigenvalues of the tidal shear tensor (see Eq. (5.15) and related text). We compare our theoretical predictions with measurements from the Quijote simulations (Villaescusa-Navarro et al., 2020). The four environments are identified by the number of eigenvalues above a threshold $\Lambda_{\text{th}} = 0.01$,

according to the value used in Cui et al. (2017). To obtain the theoretical predictions, we work with the density contrast and normalise the variables with the variance, resulting in a threshold of $\lambda_{\text{th}} = \Lambda_{\text{th}}/\sigma(z)$, where $\sigma(z)$ is the variance of the density field measured in the Quijote simulations.

We obtain the probability of the different environments by integrating the probability distribution function (PDF) under various criteria on the eigenvalues. For instance, the probability of the void is given by:

$$P_{\text{void}} = \int d\lambda_1 d\lambda_2 d\lambda_3 \mathcal{P}(\lambda_1, \lambda_2, \lambda_3) \text{Boole}(\lambda_1 < \lambda_2 < \lambda_3 < \lambda_{\text{th}}), \quad (5.17)$$

where $\mathcal{P}(\lambda_1, \lambda_2, \lambda_3)$ is the PDF of the eigenvalues of the tidal shear tensor. For the other environments, only the *Boole* function changes, to reflect the criteria on the number of eigenvalues above the threshold.

Our initial investigation focuses on determining abundances in the linear regime. In this case, the density field is Gaussian, and so its associated gravitational potential and its successive derivatives are also Gaussian-distributed. It allows us to use the PDF from the Doroskevich formula Doroshkevich (1970) or derive it by transforming variables in terms of rotational invariants to obtain minimally correlated variables (Pogosyan et al., 2009), that will be useful for the rest of the analysis. We will consider J_1 , J_2 , and J_3 , which are the rotational invariants of the tidal shear tensor. Following Pogosyan et al. (2009) and Gay et al. (2012), we derive the PDF of interest with these variables. An analytical integration over J_1 and J_3 is feasible, leaving a numerical integration over J_2 to be completed.

To extend our analysis, we consider mild non-Gaussian corrections to the abundances. The PDF is no longer Gaussian but includes non-linearities in a perturbative manner. To obtain it, we use a Gram-Charlier expansion, following Pogosyan et al. (2009); Gay et al. (2012); Codis et al. (2013). In terms of our variables, the Gram-Charlier expansion leads to:

$$\mathcal{P}(J_1, J_2, J_3) = \mathcal{P}_{\mathcal{G}}(J_1, J_2, J_3) \left[1 + \frac{1}{6} \langle J_1^3 \rangle (J_1^3 - 3J_1) + \frac{5}{2} \langle J_1 J_2 \rangle J_1 (J_2 - 1) + \frac{25}{21} \langle J_3 \rangle J_3 \right] + o(\sigma_0^2), \quad (5.18)$$

where $\mathcal{P}_{\mathcal{G}}(J_1, J_2, J_3)$ is the Gaussian PDF, and σ_0 is the linear variance of the density field. The additional terms represent the correction arising from the expansion. From this expression, we define three cumulants: $\langle J_1^3 \rangle / \sigma_0$, which is the well known skewness discussed in Sect. 5.1.1, $\langle J_1 J_2 \rangle / \sigma_0$, and $\langle J_3 \rangle / \sigma_0$. These three cumulants are constant at tree order.

Once the cumulants are determined, the expression of the PDF in Eq. (5.18) can be substituted into Eq. (5.17) (and its equivalent for the other environments). Finally, an analytical integration over J_1 and J_3 is possible, with a numerical integration over J_2 to complete the process.

Result The results of the PDF integration lead to the probabilities of the different cosmic web environments, constituting our main result presented in Fig. 2 of the article. This figure illustrates the probabilities of voids, walls, filaments, and knots as a function of redshift and smoothing scale. A portion of this figure is reproduced in Fig. 5.1. We compare the results obtained using a Gaussian field (dashed lines), non-Gaussian field utilizing a Gram-Charlier expansion (solid lines), and the measurements in the *Quijote* simulations (dots). The probabilities of void, wall, filaments, and knots are represented in blue, green, yellow, and red, respectively. At $z = 0$, with a smoothing scale of 5 Mpc/h, we examine the most non-linear scale in our study, whereas, at $z = 3$ and a smoothing scale of 45 Mpc/h, we explore the most linear scale in our study. Lower redshifts and smaller smoothing scales indicated a more non-linear regime, as structures are more evolved and smaller structures are probed.

We find that for scales as small as ~ 5 Mpc/h and down to redshifts ~ 0 (the most non-linear scale studied), our prediction using a Gram-Charlier expansion shows good agreement with the measurements from the simulations. Across all smoothing scales and redshifts, the non-Gaussian field prediction consistently outperforms the Gaussian field prediction, which quickly loses accuracy with decreasing redshift or smoothing scale.

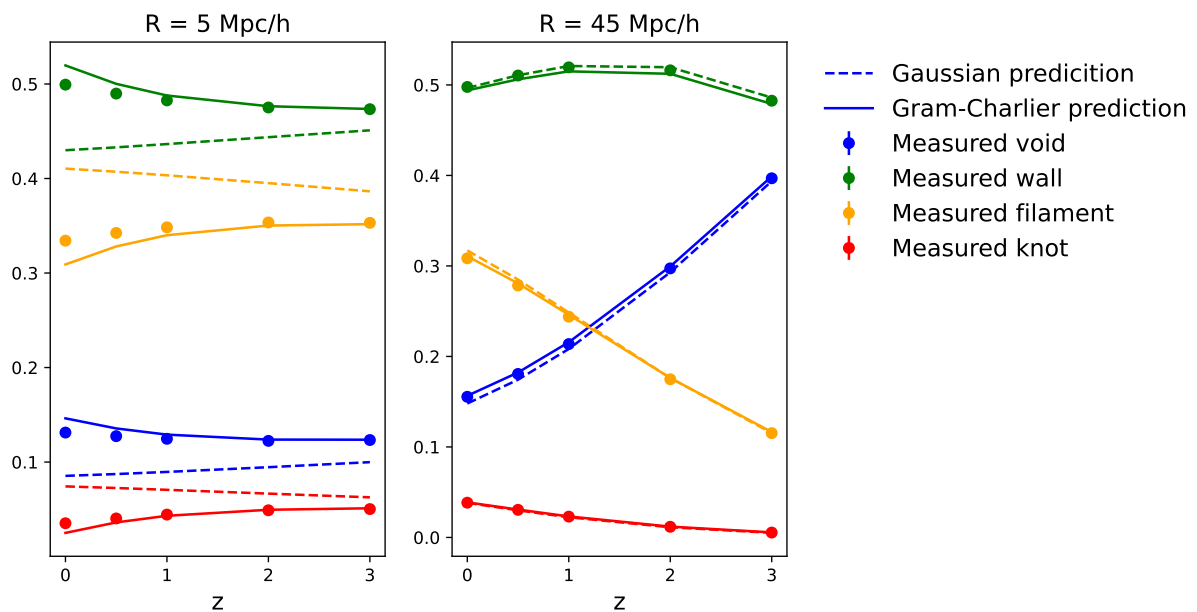


Figure 5.1: Probabilities of voids, walls, filaments, and knots as a function of the redshift. These probabilities are shown in blue, green, yellow, and red, respectively. Left panel is obtained for a smoothing scale of 5 Mpc/h, and the right one for a smoothing scale of 45 Mpc/h. Dots are measurements from the simulation, dashed lines are the Gaussian prediction, and solid lines are the prediction obtained with the Gram-Charlier formalism at next-to-leading order. The error bars are errors on the mean but are too small to be distinguished.

Our theoretical framework offers a new perspective on the statistical description of the cosmic web and opens up new avenues for future research. It provides a robust method for predicting the abundance of the four cosmic web environments, which can be used in cosmological analyses.

A theoretical view of the T-web statistical description of the cosmic web

Emma Aycoberry¹, Alexandre Barthelemy², and Sandrine Codis^{3,1}

¹ Sorbonne Université, CNRS, UMR7095, Institut d'Astrophysique de Paris, 98 bis Boulevard Arago, 75014 Paris, France
e-mail: emma.aycoberry@iap.fr

² Universitäts-Sternwarte, Fakultät für Physik, Ludwig-Maximilians Universität München, Scheinerstr. 1, 81679 München, Germany

³ Université Paris-Saclay, Université Paris Cité, CEA, CNRS, Astrophysique, Instrumentation et Modélisation Paris-Saclay, 91191 Gif-sur-Yvette, France

Received 5 October 2023 / Accepted 14 March 2024

ABSTRACT

Context. The objective classification of the cosmic web into different environments is an important aspect of large-scale structure studies, as it can be used as a tool to study the formation of structures (halos and galaxies) in more detail, and it forms a link between their properties and the large-scale environment; these different environments also offer another class of objects whose statistics contain cosmological information.

Aims. In this paper, we present an analytical framework to compute the probability of the different environments in the cosmic web based on the so-called T-web formalism, which classifies structures into four different classes (voids, walls, filaments, and knots) based on the eigenvalues of the Hessian of the gravitational potential, often called the tidal tensor.

Methods. Our classification method relies on studying whether the eigenvalues of this Hessian matrix are above or below a given threshold and thus requires knowledge of the joint probability distribution of those eigenvalues. We performed a change of variables in terms of rotational invariants, which are polynomials of the field variables and minimally correlated. We studied the distribution of those variables in the linear and quasi-linear regimes with the help of a so-called Gram-Charlier expansion, using tree-order Eulerian perturbation theory to compute the Gram-Charlier coefficients. This expansion then allowed us to predict the probability of the four different environments as a function of the chosen threshold and at a given smoothing scale and redshift for the density field. We checked the validity regime of our predictions by comparing those predictions to measurements made in the N -body Quijote simulations.

Results. Working with fields normalised by their linear variance, we find that scaling the threshold value with the non-linear amplitude of fluctuations allows us to capture almost the entire redshift evolution of the probabilities of the four environments, even if we assume that the density field is Gaussian (corresponding to the linear regime of structure formation). We also show that adding mild non-Gaussian corrections with the help of a Gram-Charlier expansion – hence introducing corrections that depend on third-order cumulants of the field – provides even greater accuracy, allowing us to obtain very precise predictions for cosmic web abundances up to scales of as small as $\sim 5 \text{ Mpc } h^{-1}$ and redshifts down to $z \sim 0$.

Key words. methods: analytical – methods: numerical – cosmology: theory – large-scale structure of Universe

1. Introduction

We have long since known that the large-scale distribution of matter throughout the Universe is well-approximated by a filamentary structure dubbed the cosmic web (de Lapparent et al. 1986). From a theoretical point of view, the first building blocks for a description of the cosmic web date back to the seminal work of Zel'dovich (1970) and many others that followed (e.g., Arnold et al. 1982; Klypin & Shandarin 1983). Indeed, the so-called Zeldovich approximation, which describes the first-order ballistic trajectories of particles in a Lagrangian description – predicts the existence of ‘pancakes’ (i.e., sheet-like tenuous walls), filaments, and clusters due to the collapse of anisotropic primordial fluctuations through gravitational instabilities in our expanding Universe. In the 90s, Bond et al. (1996) and Klypin & Shandarin (1983) showed that this ensemble of objects forms a connected network called the cosmic web based on the correlations imprinted in the primordial fluctuations. Initial peaks led to the formation of clusters at the nodes of the web while initial correlation bridges in between later form filaments that lie within walls, which themselves surround nearly empty void regions. Such a web classification can be achieved through the eigenvalues of the linear deformation tensor, which states that if all eigenvalues are negative (or below a threshold that is taken here to be zero) then the region is expanding in 3D, thus describing a void region; if all values are positive then the region is contracting, thus describing a knot, with the other two configurations of eigenvalues leading to either walls or filaments. It is possible to classify the cosmic web by means of its tidal field (Hessian of the gravitational potential). Pioneering work in this direction was presented by Doroshkevich (1970), and later by van de Weygaert & Bertschinger (1996), Rossi (2012), Desjacques et al. (2018), Castorina et al. (2016), Feldbrugge et al. (2018).

Another classification method was later introduced by Hahn et al. (2007a), Aragón-Calvo et al. (2007a) and Forero-Romero et al. (2009) using tidal fields, whereby either the Hessian of the gravitational potential is evaluated theoretically in the linear regime, or

the non-linear potential is estimated in dark-matter-only numerical simulations. In subsequent years, several authors developed different classification schemes, to improve the detection of these cosmic web structures in various types of data (continuous fields, simulated datasets, point-like galaxy surveys, etc). For instance, [Aragón-Calvo et al. \(2010a\)](#) used the SpineWeb topological framework to segment the density field, [Shandarin \(2011\)](#), [Falck et al. \(2012\)](#) and [Abel et al. \(2012\)](#) showed how to classify morphological structures using the Lagrangian phase space sheet to count for shell crossings, and various authors have applied techniques from (continuous and then discrete) Morse theory ([Colombi et al. 2000](#)) to identify topological structures in the cosmological density field ([Sousbie et al. 2008](#); [Aragón-Calvo et al. 2007a](#); [Sousbie et al. 2009](#); [Sousbie 2011](#)). Alternatively, [Hoffman et al. \(2012\)](#) introduced the V-web classification scheme, this time based on the shear of the velocity field, showing that it is able to better resolve smaller structures, which in turn allows for the study of finer dark-matter halo properties. An extension of this idea to Lagrangian settings was later proposed by [Fisher et al. \(2016\)](#). We note that the velocity divergence and density fields are closely related (e.g. through the mass-conservation equation in the Vlasov-Poisson system) and their statistics are virtually equivalent in the linear regime of structure formation. We refer readers to [Wang et al. \(2014\)](#) and [Wang & Szalay \(2014\)](#) for a thorough investigation of the differences between the dynamical and kinematical classifications.

Beyond the strict motivation of wanting to describe the cosmic web from a mathematical point of view, the classification schemes allow us to explore many different environmental effects on the properties of dark-matter halos ([Hahn et al. 2007a,b](#); [Aragón-Calvo et al. 2007b, 2010b](#); [Codis et al. 2012](#); [Libeskind et al. 2013](#); [Hellwing et al. 2021](#)) and galaxies within them ([Nuza et al. 2014](#); [Metuki et al. 2015](#); [Poudel et al. 2017](#); [Kraljic et al. 2018](#); [Codis et al. 2018a](#); [Hasan et al. 2023](#)). Cosmic web classification can also be used to discriminate between cosmological models, as shown for instance by [Lee & Park \(2009\)](#) or [Biswas et al. \(2010\)](#) using voids; [Codis et al. \(2013\)](#) with counts of cosmic web critical points; [Feldbrugge et al. \(2019\)](#) with Betti numbers; [Codis et al. \(2018b\)](#) using the connectivity of the filaments; [Bonnaire et al. \(2022\)](#) relying on the power spectrum of the various cosmic web environments; and [Dome et al. \(2023\)](#) with cosmic web abundances. Understanding how this cosmic web evolves with time and scale is therefore paramount for both cosmology and studies of the galaxy formation that takes place within this large-scale environment. This evolution has been studied theoretically in some contexts (e.g., the local skeleton; [Pogosyan et al. 2009b](#)) with semi-analytical approaches (e.g., [Fard et al. 2019](#)) and in various numerical works. One example of the latter is the recent work by [Cui et al. \(2017, 2019\)](#), who used simulations to investigate the abundances of the various environments and their time evolution, relying on the T- and V-web decompositions. However, to the best of our knowledge, no theoretical model in the quasi-linear regime, and based on first principles, has been explicitly derived so far for these cosmic web abundances. This objective is nevertheless within the reach of standard techniques used in large-scale structure theoretical studies.

In the present paper, we therefore propose to derive theoretical predictions for web classifications based on the T-web definition. We first focus on a Gaussian description of the cosmic density field before turning to mild non-Gaussian corrections.

To assess the validity regime of our theoretical formalism and predictions for the one-point statistics of all four cosmic environments, we compare our results with measurements made in a dark-matter-only N -body numerical simulation, the Quijote simulation ([Villaescusa-Navarro et al. 2020](#)). The simulated box used by these latter authors has a size of $1 \text{ Gpc } h^{-1}$, with 1024^3 dark matter particles. We are using their high-resolution fiducial cosmology snapshots at five different redshifts: $z = 0, 0.5, 1, 2, \text{ and } 3$. Their fiducial cosmology is $\{\Omega_m, \Omega_b, h, n_s, \sigma_8\} = \{0.3175, 0.049, 0.6711, 0.9624, 0.834\}$. Finally, in order to maximise the statistical relevance of our analysis, we typically average our measurements over 11 realisations of the Quijote suite and use those realisations to estimate error bars.

The outline of the paper is as follows. We first describe the T-web and V-web classifications in Sect. 2. In Sect. 3, we then present the theoretical formalism and the predictions obtained for the abundance of the different environments (as a function of threshold, redshift, and smoothing scale) in the linear regime of structure formation, that is assuming a Gaussian matter density field. Section 4 introduces a Gram-Charlier expansion to include non-Gaussian (i.e., non-linear) corrections to the joint probability distribution function (joint-PDF) of the elements of the Hessian of the gravitational potential, and we finally discuss our results and draw conclusions in Sect. 5. In the remainder of this paper, we will use ‘abundance’ or ‘probability’ equivalently.

2. T-web classification of the cosmic web

Among many possibilities ([Libeskind et al. 2017](#); [Leclercq et al. 2016](#)), one commonly used mathematical way to classify the different cosmological environments is based on the number of eigenvalues of a deformation tensor above a threshold. Several definitions can be used for this deformation tensor (strain tensor, mass tensor, etc.) but the classification can always follow the outline below:

- 0 eigenvalue above the threshold: void,
- 1 eigenvalue above the threshold: wall,
- 2 eigenvalues above the threshold: filament,
- 3 eigenvalues above the threshold: knot.

Notably, the T-web classification developed by [Forero-Romero et al. \(2009\)](#) is based on the tidal shear tensor T , which is the tensor of the second derivative of the gravitational potential

$$T_{ij} = \frac{\partial^2 \Phi}{\partial r_i \partial r_j}, \quad (1)$$

where Φ is the gravitational potential and r_i with $i = 1, 2, 3$ are the spatial coordinates. Alternatively, the V-web classification developed by [Hoffman et al. \(2012\)](#) is based on the velocity shear tensor Σ :

$$\Sigma_{ij} = -\frac{1}{2} \left(\frac{\partial v_i}{\partial r_j} + \frac{\partial v_j}{\partial r_i} \right) / H_0, \quad (2)$$

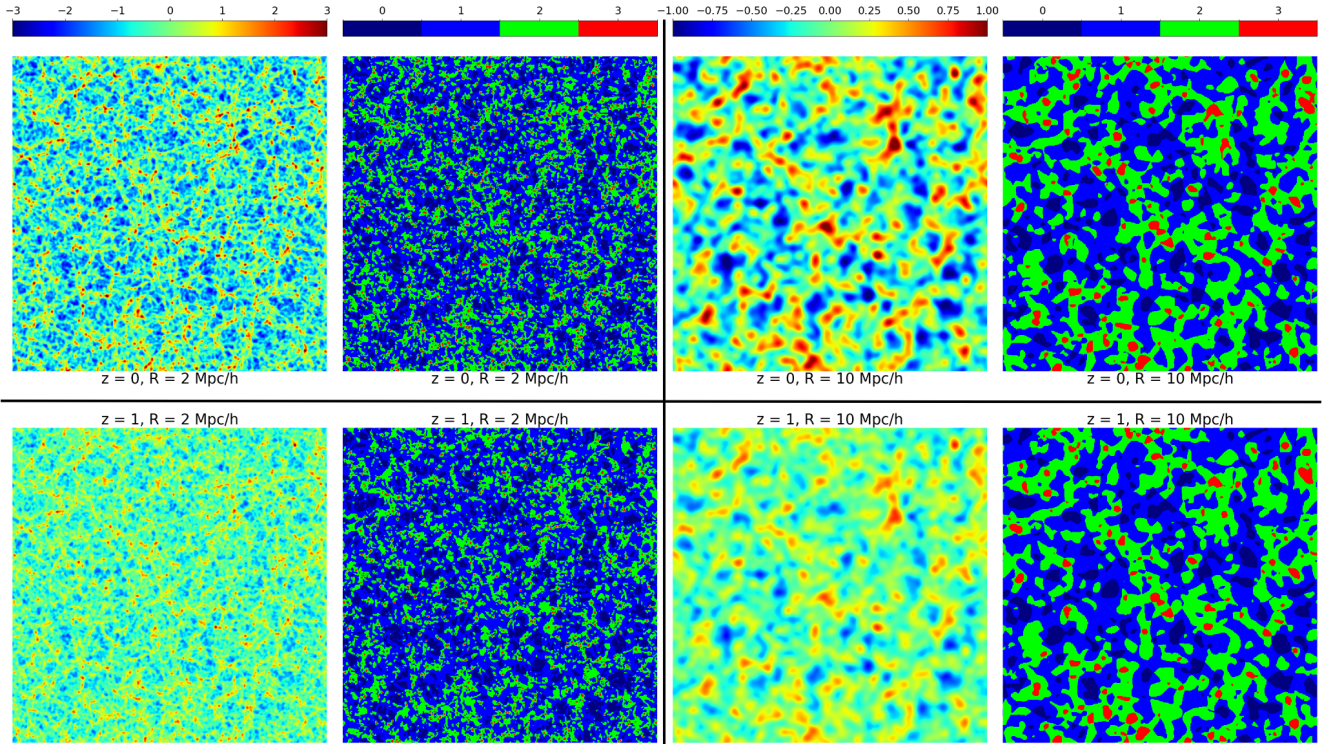


Fig. 1. Logarithm of the density contrast of one slice of $1 \text{ Gpc } h^{-1}$ in height and width and $2 \text{ Mpc } h^{-1}$ in thickness of the Quijote simulation (first and third column) compared with the classification obtained using the T-web (second and fourth column) at different redshifts and smoothing scales. For the T-web classification, the colour bar is the number of eigenvalues above our chosen threshold $\Lambda_{\text{th}} = 0.01$ chosen specifically for good visual agreement.

where H_0 is the Hubble constant and v the velocity. We note that both of the above classifications are similar to the one used by [Hahn et al. \(2007a\)](#). Given those definitions, in practice, one can start from field maps, compute the deformation tensor, diagonalise it, and obtain the number of eigenvalues above a given threshold Λ_{th} at each spatial position, thus dividing the cosmic web into four characteristic classes.

Usually, the value of this threshold is not fixed a priori in the literature and is often chosen in order to obtain satisfactory visual agreement between the log density field and the classification. Originally, [Hahn et al. \(2007a\)](#) took $\Lambda_{\text{th}} = 0$ to distinguish structures with inner or outer flows. Follow-up studies demonstrated that having a non-zero threshold leads to a better visual match (for example, [Forero-Romero et al. 2009](#); [Libeskind et al. 2017](#); [Suárez-Pérez et al. 2021](#); [Cui et al. 2017](#); [Hoffman et al. 2012](#); [Carlesi et al. 2014](#)). The threshold is generally taken as a positive constant of between 0 and 1, and usually does not depend on the redshift or smoothing scale, but we note that a positive threshold will tend to underline the most extreme knots and filaments (which is usually what is meant by good visual agreement) while a negative threshold will then underline the most empty voids.

In the remainder of this paper, we focus on the T-web formalism, although we emphasise that the V-web description would yield identical results at first perturbative order in a Lagrangian framework (the so-called 1LPT or Zeldovich approximation). This is indeed due to the usual expression for these ballistic trajectories where the velocity is given by the gradient of the gravitational potential up to a uniform time-dependent factor ([Zel'dovich 1970](#)). As an example of the accuracy of this classification scheme, [Fig. 1](#) presents a comparison of the density contrast (first and third columns, with a continuous colour bar) at different redshifts and smoothing scales, with the classification obtained using the T-web (second and fourth column, with a discrete color bar) at the same redshifts and smoothing scales. In the second and fourth columns, voids are shown in dark blue, walls in blue, filaments in green, and nodes in red. As expected from the T-web classification, we obtain the environments of the simulation by computing the second derivative of the potential and looking at the number of eigenvalues above our chosen threshold. Here, we are using a threshold of $\Lambda_{\text{th}} = 0.01$ for every redshift and smoothing scale based on the value used in [Cui et al. \(2017\)](#), which was chosen in order to have good visual agreement. For instance, the knots and voids (respectively in red and dark blue) can easily be identified by eye in both visualisations and are at the same position as rare maxima and minima in the density contrast maps.

To build a theoretical description of the cosmic web as defined by the T-web description, let us first define the variance of the contrast of the density field, $\delta = (\rho - \langle \rho \rangle) / \langle \rho \rangle$,

$$\sigma_0^2 = \langle \delta^2 \rangle. \quad (3)$$

Following [Pogosyan et al. \(2009a\)](#), we choose to normalise the derivatives of the gravitational potential by their variance, such that

$$v = \frac{1}{\sigma_0} \delta, \quad \phi_{ij} = \frac{1}{\sigma_0} \nabla_i \nabla_j \Phi. \quad (4)$$

Formally, given our classification method, the probability of each cosmic environment then depends on the joint probability distribution \mathcal{P} of the eigenvalues of the tidal tensor. Given our normalisation choice in [Eq. \(4\)](#), here we use the eigenvalues normalised

by their variance (or equivalently the eigenvalues of the normalised tidal tensor), which we denote $\lambda_1 \leq \lambda_2 \leq \lambda_3$. The probabilities of the four environments can then be written as

$$P_{\text{void}} = \int d\lambda_1 d\lambda_2 d\lambda_3 \mathcal{P}(\lambda_1, \lambda_2, \lambda_3) \text{Boole}(\lambda_1 < \lambda_2 < \lambda_3 < \lambda_{\text{th}}) \quad (5)$$

$$P_{\text{wall}} = \int d\lambda_1 d\lambda_2 d\lambda_3 \mathcal{P}(\lambda_1, \lambda_2, \lambda_3) \text{Boole}(\lambda_1 < \lambda_2 < \lambda_{\text{th}} < \lambda_3), \quad (6)$$

$$P_{\text{filament}} = \int d\lambda_1 d\lambda_2 d\lambda_3 \mathcal{P}(\lambda_1, \lambda_2, \lambda_3) \text{Boole}(\lambda_1 < \lambda_{\text{th}} < \lambda_2 < \lambda_3), \quad (7)$$

$$P_{\text{knot}} = \int d\lambda_1 d\lambda_2 d\lambda_3 \mathcal{P}(\lambda_1, \lambda_2, \lambda_3) \text{Boole}(\lambda_{\text{th}} < \lambda_1 < \lambda_2 < \lambda_3), \quad (8)$$

where $\{\lambda_i\}_{i=1,2,3}$ are the orderly eigenvalues of the $(\phi_{ij})_{1 \leq i, j \leq 3}$ matrix and *Boole* is a Boolean equal to 1 if the condition is satisfied, and 0 otherwise.

As we are now working with normalised eigenvalues, and keeping our earlier choice of a $\Lambda_{\text{th}} = 0.01$ value, our threshold will be given by $\lambda_{\text{th}} = (\Lambda_{\text{th}} = 0.01)/\sigma(z)$, with

$$\sigma^2(z) = 4\pi \int dk k^2 P(k, z) W(kR)^2, \quad (9)$$

where $W(kR)$ is the applied smoothing, which in this paper is Gaussian with a smoothing scale R such that

$$W_G(kR) = \exp\left(-\frac{1}{2}k^2 R^2\right), \quad (10)$$

and $P(k, z)$ is the matter density power spectrum. We denote the linear variance σ_0^2 when using the linear power spectrum, which is computed using the Boltzmann code CAMB (Lewis et al. 2000), and σ_{NL}^2 the non-linear variance (e.g., measured in the simulation or predicted with emulators for example).

3. Cosmic web abundances in the linear regime

To obtain some theoretical information about the abundance of the different cosmic environments in the density field and their redshift evolution, let us first consider the density field at linear order in Eulerian perturbation theory, this is, let us assume that it is Gaussian. For a Gaussian density field, the associated gravitational potential and its successive derivatives are also Gaussian-distributed, which means that we can use the Doroshkevich formula (Doroshkevich 1970), which gives the joint probability distribution of eigenvalues of a Gaussian symmetric matrix:

$$\mathcal{P}_D(\lambda_1, \lambda_2, \lambda_3) = \frac{675 \sqrt{5} e^{\frac{3}{4}(\lambda_1 + \lambda_2 + \lambda_3)^2 - \frac{15}{4}(\lambda_1^2 + \lambda_2^2 + \lambda_3^2)}}{8\pi} (\lambda_3 - \lambda_2)(\lambda_2 - \lambda_1)(\lambda_3 - \lambda_1). \quad (11)$$

We note that the general expression for the multi-dimensional PDF of the tidal tensor in Cartesian coordinates $X = \{\phi_{ij}\}$ in the Gaussian case reads

$$\mathcal{P}_G(X) = (2\pi)^{-N/2} |C|^{-1/2} \exp\left(-\frac{1}{2}XC^{-1}X\right), \quad (12)$$

where C is its covariance matrix.

From Eq. (11), we can then numerically integrate Eqs. (5)–(8) to obtain the probabilities of the various environments. In practice, those 3D integrals can be reduced to 1D as two degrees of freedom can be analytically integrated out. For that purpose, one needs to express the three degrees of freedom of the tidal tensor that are rotationally invariant as polynomials of its Cartesian coordinates ϕ_{ij} (in contrast to eigenvalues, which are not rotationally invariant). Following for instance Pogosyan et al. (2009a), a natural option can be the three coefficients of its characteristic polynomial:

$$I_1 = \text{Tr}(\phi_{ij}) = \phi_{11} + \phi_{22} + \phi_{33} = \lambda_1 + \lambda_2 + \lambda_3 = \nu, \quad (13)$$

$$I_2 = \phi_{11}\phi_{22} + \phi_{22}\phi_{33} + \phi_{11}\phi_{33} - \phi_{12}^2 - \phi_{23}^2 - \phi_{13}^2 = \lambda_1\lambda_2 + \lambda_2\lambda_3 + \lambda_1\lambda_3, \quad (14)$$

$$I_3 = \det[\phi_{ij}] = \phi_{11}\phi_{22}\phi_{33} + 2\phi_{12}\phi_{23}\phi_{13} - \phi_{11}\phi_{23}^2 - \phi_{22}\phi_{13}^2 - \phi_{33}\phi_{12}^2 = \lambda_1\lambda_2\lambda_3. \quad (15)$$

To get variables that are as uncorrelated as possible, a simplification proposed by Pogosyan et al. (2009a) is to use combinations of the $\{I_k\}_{1 \leq k \leq 3}$ – denoted $\{J_k\}_{1 \leq k \leq 3}$ – such that the first variable is again the trace of the tidal tensor but higher-order variables only depend on the traceless part of that tensor and are thus uncorrelated with the trace in the Gaussian case. After some algebra, one can easily find

$$J_1 = I_1, \quad J_2 = I_2^2 - 3I_2, \quad J_3 = I_1^3 - \frac{9}{2}I_1I_2 + \frac{27}{2}I_3. \quad (16)$$

For a Gaussian random field, one can easily show that the probability distribution function of the above-defined variables reads (Pogosyan et al. 2009a)

$$\mathcal{P}_G(J_1, J_2, J_3) = \frac{25\sqrt{5}}{12\pi} \exp\left[-\frac{1}{2}J_1^2 - \frac{5}{2}J_2\right]. \quad (17)$$

where $J_2 \geq 0$ and J_3 is uniformly distributed between its boundaries such that $-J_2^{3/2} \leq J_3 \leq J_2^{3/2}$. Let us note that the probability distribution of J_s can be easily mapped to the Doroskevitch formula for the distribution of the eigenvalues after taking into account the usual Vandermonde determinant.

To get the abundances for the different environments, we use a criterion on the number of eigenvalues superior to the threshold. As we are now working with variables (J_1, J_2, J_3) , we have to translate the criterion on eigenvalues into this new set of variables. Following Pogosyan et al. (2009a) and Gay et al. (2012), we can obtain the integration limits for the four different environments as follows:

$$P_{\text{void}} = \int_0^\infty dJ_2 \int_{-\infty}^{-2\sqrt{J_2+3\lambda_{\text{th}}}} dJ_1 \int_{-J_2^{3/2}}^{J_2^{3/2}} dJ_3 \mathcal{P}_G(J_1, J_2, J_3) + \int_0^\infty dJ_2 \int_{-2\sqrt{J_2+3\lambda_{\text{th}}}}^{-\sqrt{J_2+3\lambda_{\text{th}}}} dJ_1 \int_{-J_2^{3/2}}^{-\frac{1}{2}(J_1-3\lambda_{\text{th}})^3 + \frac{3}{2}(J_1-3\lambda_{\text{th}})J_2} dJ_3 \mathcal{P}_G(J_1, J_2, J_3), \quad (18)$$

$$P_{\text{wall}} = \int_0^\infty dJ_2 \int_{-2\sqrt{J_2+3\lambda_{\text{th}}}}^{\sqrt{J_2+3\lambda_{\text{th}}}} dJ_1 \int_{-\frac{1}{2}(J_1-3\lambda_{\text{th}})^3 + \frac{3}{2}(J_1-3\lambda_{\text{th}})J_2}^{J_2^{3/2}} dJ_3 \mathcal{P}_G(J_1, J_2, J_3), \quad (19)$$

$$P_{\text{filament}} = \int_0^\infty dJ_2 \int_{-\sqrt{J_2+3\lambda_{\text{th}}}}^{2\sqrt{J_2+3\lambda_{\text{th}}}} dJ_1 \int_{-J_2^{3/2}}^{-\frac{1}{2}(J_1-3\lambda_{\text{th}})^3 + \frac{3}{2}(J_1-3\lambda_{\text{th}})J_2} dJ_3 \mathcal{P}_G(J_1, J_2, J_3), \quad (20)$$

$$P_{\text{knot}} = \int_0^\infty dJ_2 \int_{\sqrt{J_2+3\lambda_{\text{th}}}}^{2\sqrt{J_2+3\lambda_{\text{th}}}} dJ_1 \int_{-\frac{1}{2}(J_1-3\lambda_{\text{th}})^3 + \frac{3}{2}(J_1-3\lambda_{\text{th}})J_2}^{J_2^{3/2}} dJ_3 \mathcal{P}_G(J_1, J_2, J_3) + \int_0^\infty dJ_2 \int_{2\sqrt{J_2+3\lambda_{\text{th}}}}^\infty dJ_1 \int_{-J_2^{3/2}}^{J_2^{3/2}} dJ_3 \mathcal{P}_G(J_1, J_2, J_3). \quad (21)$$

The analytical integration over J_1 and J_3 can now be performed and we thus obtain the following probabilities of the four environments (written in terms of 1D integrals over J_2):

$$P_{\text{void}} = \int_0^\infty \frac{25\sqrt{5}}{48\pi} e^{-\frac{5J_2}{2}} \left[-\sqrt{2\pi}(2J_2^{3/2} - 9J_2\lambda_{\text{th}} + 9(3\lambda_{\text{th}}^3 + \lambda_{\text{th}})) \operatorname{erf}\left(\frac{3\lambda_{\text{th}} - 2\sqrt{J_2}}{\sqrt{2}}\right) + \sqrt{2\pi}(2J_2^{3/2} - 9J_2\lambda_{\text{th}} + 9(3\lambda_{\text{th}}^3 + \lambda_{\text{th}})) \operatorname{erf}\left(\frac{3\lambda_{\text{th}} - \sqrt{J_2}}{\sqrt{2}}\right) \right. \\ \left. + 4\sqrt{2\pi}J_2^{3/2} \operatorname{erfc}\left(\frac{2\sqrt{J_2} - 3\lambda_{\text{th}}}{\sqrt{2}}\right) - 2e^{-\frac{1}{2}(2\sqrt{J_2}-3\lambda_{\text{th}})^2} (6\sqrt{J_2}\lambda_{\text{th}} + J_2 + 9\lambda_{\text{th}}^2 + 2) + e^{-\frac{1}{2}(\sqrt{J_2}-3\lambda_{\text{th}})^2} (6\sqrt{J_2}\lambda_{\text{th}} - 4J_2 + 18\lambda_{\text{th}}^2 + 4) \right] dJ_2, \quad (22)$$

$$P_{\text{wall}} = \int_0^\infty \frac{25\sqrt{5}}{48\pi} e^{-\frac{5J_2}{2}} \left[-\sqrt{2\pi}(2J_2^{3/2} + 9J_2\lambda_{\text{th}} - 9(3\lambda_{\text{th}}^3 + \lambda_{\text{th}})) \operatorname{erf}\left(\frac{3\lambda_{\text{th}} - 2\sqrt{J_2}}{\sqrt{2}}\right) + \sqrt{2\pi}(2J_2^{3/2} + 9J_2\lambda_{\text{th}} - 9(3\lambda_{\text{th}}^3 + \lambda_{\text{th}})) \operatorname{erf}\left(\frac{\sqrt{J_2} + 3\lambda_{\text{th}}}{\sqrt{2}}\right) \right. \\ \left. + 2e^{-\frac{1}{2}(2\sqrt{J_2}-3\lambda_{\text{th}})^2} (6\sqrt{J_2}\lambda_{\text{th}} + J_2 + 9\lambda_{\text{th}}^2 + 2) + 2e^{-\frac{1}{2}(\sqrt{J_2}+3\lambda_{\text{th}})^2} (3\sqrt{J_2}\lambda_{\text{th}} + 2J_2 - 9\lambda_{\text{th}}^2 - 2) \right] dJ_2, \quad (23)$$

$$P_{\text{filament}} = \int_0^\infty -\frac{25\sqrt{5}}{48\pi} e^{-\frac{5J_2}{2}} \left[-\sqrt{2\pi}(2J_2^{3/2} - 9J_2\lambda_{\text{th}} + 9(3\lambda_{\text{th}}^3 + \lambda_{\text{th}})) \operatorname{erf}\left(\frac{2\sqrt{J_2} + 3\lambda_{\text{th}}}{\sqrt{2}}\right) + \sqrt{2\pi}(2J_2^{3/2} - 9J_2\lambda_{\text{th}} + 9(3\lambda_{\text{th}}^3 + \lambda_{\text{th}})) \operatorname{erf}\left(\frac{3\lambda_{\text{th}} - \sqrt{J_2}}{\sqrt{2}}\right) \right. \\ \left. - 2e^{-\frac{1}{2}(2\sqrt{J_2}+3\lambda_{\text{th}})^2} (-6\sqrt{J_2}\lambda_{\text{th}} + J_2 + 9\lambda_{\text{th}}^2 + 2) + e^{-\frac{1}{2}(\sqrt{J_2}-3\lambda_{\text{th}})^2} (6\sqrt{J_2}\lambda_{\text{th}} - 4J_2 + 18\lambda_{\text{th}}^2 + 4) \right] dJ_2, \quad (24)$$

$$P_{\text{knot}} = \int_0^\infty \frac{25\sqrt{5}}{48\pi} e^{-\frac{5J_2}{2}} \left[-\sqrt{2\pi}(2J_2^{3/2} + 9J_2\lambda_{\text{th}} - 9(3\lambda_{\text{th}}^3 + \lambda_{\text{th}})) \operatorname{erf}\left(\frac{\sqrt{J_2} + 3\lambda_{\text{th}}}{\sqrt{2}}\right) + \sqrt{2\pi}(2J_2^{3/2} + 9J_2\lambda_{\text{th}} - 9(3\lambda_{\text{th}}^3 + \lambda_{\text{th}})) \operatorname{erf}\left(\frac{2\sqrt{J_2} + 3\lambda_{\text{th}}}{\sqrt{2}}\right) \right. \\ \left. + 4\sqrt{2\pi}J_2^{3/2} \operatorname{erfc}\left(\frac{2\sqrt{J_2} + 3\lambda_{\text{th}}}{\sqrt{2}}\right) - 2e^{-\frac{1}{2}(2\sqrt{J_2}+3\lambda_{\text{th}})^2} (-6\sqrt{J_2}\lambda_{\text{th}} + J_2 + 9\lambda_{\text{th}}^2 + 2) + e^{-\frac{1}{2}(\sqrt{J_2}+3\lambda_{\text{th}})^2} (-6\sqrt{J_2}\lambda_{\text{th}} - 4J_2 + 18\lambda_{\text{th}}^2 + 4) \right] dJ_2, \quad (25)$$

where $\operatorname{erf}(z)$ is the error function $\operatorname{erf}(z) = 2 \int_0^z e^{-t^2} dt / \sqrt{\pi}$ and $\operatorname{erfc}(z) = 1 - \operatorname{erf}(z)$.

It is then possible to perform numerical integrations and obtain the abundance of each environment. This is illustrated in Fig. 2, where the probabilities of the different environments are shown as a function of redshift. Each panel corresponds to a different Gaussian smoothing scale and displays the probability of voids, walls, filaments, and knots in blue, green, yellow, and red, respectively. The dots are the measurements obtained from the simulation, the dashed lines represent the Gaussian prediction obtained with the formalism described in this section, and the solid lines – which can be ignored for now – are the predictions at next-to-leading order obtained with a Gram-Charlier expansion, which is described in Sect. 4, below. As expected, we observe that the higher the redshift and/or the larger the smoothing scale – and thus the closer the simulation gets to the linear regime –, the closer the Gaussian prediction to the measurements. At lower redshift and smaller smoothing scales, non-Gaussianities are more important and departures from the Gaussian prediction thus appear. Here, all probabilities are computed using a threshold $\lambda_{\text{th}} = 0.01/\sigma_{\text{NL}}$ inspired from the literature (using simulations) for which the evolution of σ with redshift and smoothing scale is given in Table 1. We note that the variance used solely in the normalisation of the threshold is the non-linear variance measured from the simulation. This allows us to have a meaningful threshold – in terms of rarity – even though we describe the cosmic structures in linear theory. Given the good agreement with the simulation already, that is, simply with Gaussian theory for sufficiently large scale and redshift, this states that the probabilities of the different environments are roughly captured by the statistics

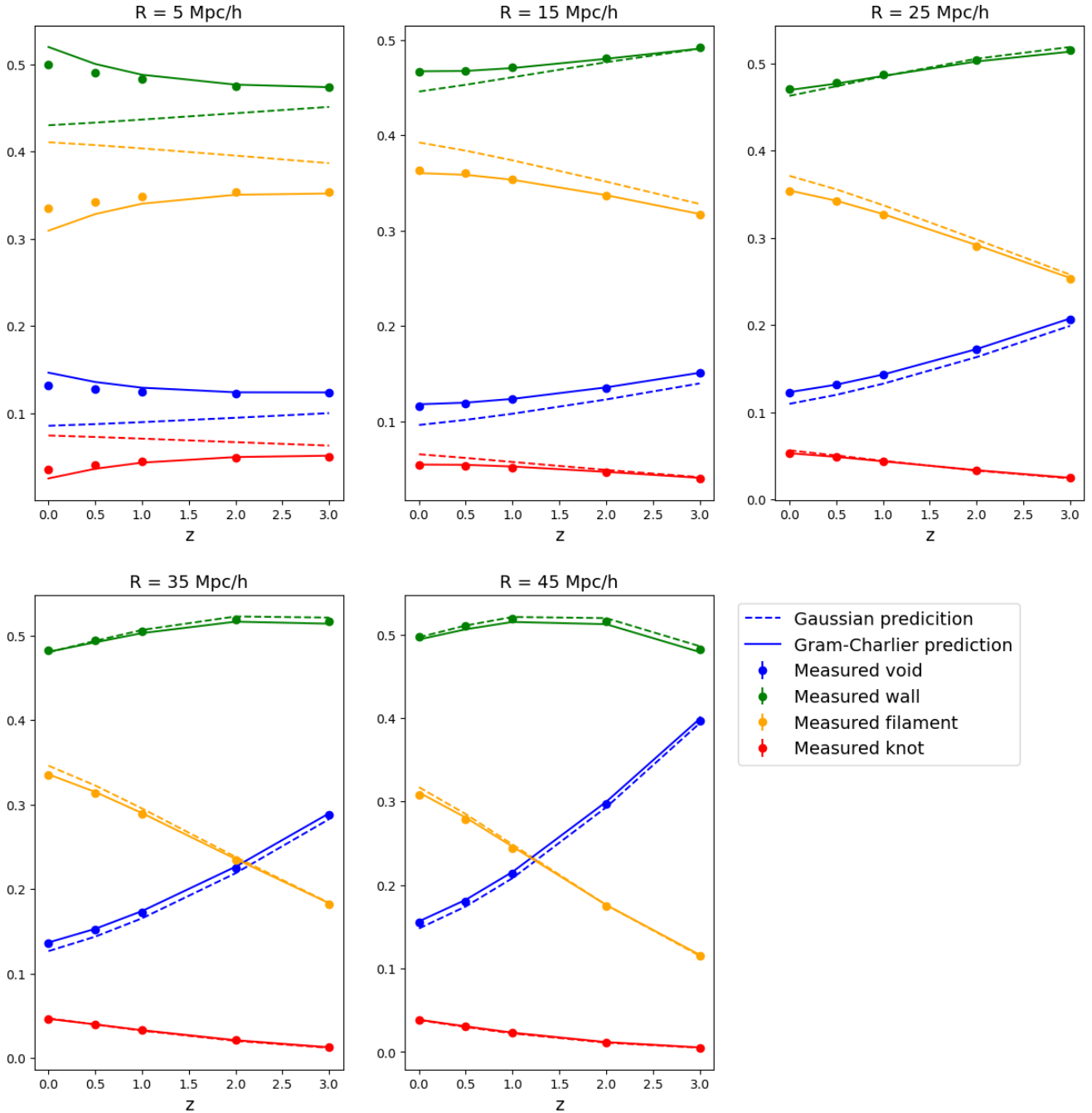


Fig. 2. Probabilities of voids, walls, filaments, and knots as a function of the redshift. These probabilities are shown in blue, green, orange, and red, respectively. Each panel is obtained at a given smoothing scale. Dots are measurements from the simulation, dashed lines are the Gaussian prediction, and solid lines are the prediction obtained with the Gram-Charlier formalism at next-to-leading order. The error bars are errors on the mean but are too small to be distinguished.

of a Gaussian field, at least for redshifts and scales that typically correspond to typical variances of $\sigma \lesssim 0.1$. Consequently, the redshift evolutions seen in previous works (Cui et al. 2017, 2019) with a fixed threshold can be roughly understood simply as the non-linear evolution of the amplitude of fluctuations for that threshold. This is because a fixed threshold in non-linear densities does not correspond to a fixed ‘rarity’ or ‘abundance’ threshold for which the cosmic evolution would be much less important. This interpretation is notably valid at sufficiently large scale and redshift. Beyond $\sigma \sim 0.1$, such a Gaussian field approximation starts to break down. In the following section, we show how the accuracy of the theoretical model with the help of Gram-Charlier corrections.

Before turning to the case of non-Gaussian corrections, let us emphasise that our Gaussian theoretical formalism allows us to obtain the probability of voids, walls, filaments, and knots as a function of the threshold itself, as illustrated in Fig. 3 for redshift $z = 0$ and Gaussian smoothing $R = 15 \text{ Mpc } h^{-1}$. The threshold used in the above analysis ($\Lambda_{\text{th}} = 0.01$) is the vertical black dashed line. For this choice of smoothing and redshift – which corresponds to a mildly non-linear regime – and for all the environments, we can

Table 1. Standard deviation at different redshifts and (Gaussian) smoothing scales.

z/R (Mpc h^{-1})	5	15	25	35	45	55	65
0	0.71 (0.68)	0.26 (0.26)	0.15 (0.15)	0.10 (0.10)	0.074 (0.072)	0.057 (0.055)	0.045 (0.043)
0.5	0.54 (0.52)	0.20 (0.20)	0.12 (0.12)	0.078 (0.077)	0.57 (0.056)	0.044 (0.042)	0.34 (0.033)
1	0.42 (0.41)	0.16 (0.16)	0.091 (0.091)	0.061 (0.061)	0.045 (0.044)	0.034 (0.033)	0.027 (0.026)
2	0.29 (0.28)	0.11 (0.11)	0.063 (0.062)	0.042 (0.042)	0.031 (0.030)	0.024 (0.023)	0.019 (0.018)
3	0.21 (0.22)	0.082 (0.082)	0.047 (0.047)	0.032 (0.031)	0.023 (0.023)	0.018 (0.017)	0.014 (0.014)

Notes. The first value is the measurement from the Quijote simulation σ_{NL} , while the second column in parenthesis is the linear prediction σ_0 .

draw the same conclusion as before: the qualitative picture is correctly captured but non-linear corrections are nonetheless necessary to improve the predictive power of our model. This figure could be helpful to guide the choice of threshold based on theoretical arguments. We note that an alternative approach could be to renounce defining a global threshold and choose it according to the studied environment(s). For example, one could determine the threshold that gives the 20% rarest knots and use the same threshold for the filaments, and obtain a threshold for the voids and walls in a similar fashion (which would be the opposite of the one used for knots and filaments in the simple Gaussian case). In this case, some spatial position may be in none of the environments and this will mean that this position is a transition between two environments. Another possibility would be to use the filling factor approach to fix a threshold in abundance (i.e., we fix the volume fraction occupied by the excursion above ν , [Gott et al. 1987](#); [Matsubara 2003](#)): this would lead to a remapping of our environments.

Hereafter, we adhere to the standard global-threshold strategy. The description and inclusion of the above-mentioned non-linear theoretical corrections are now the goals of the remainder of this paper.

4. Mild non-Gaussian corrections to cosmic web abundances

Relying on Gaussian random fields is valid only in the linear regime of structure formation. However, at low redshift and small scales, increasing numbers of non-Gaussianities appear in the density field and corrections to the Gaussian predictions need to be accounted for. To improve our previous Gaussian predictions, we now propose to work with a probability distribution function $\mathcal{P}(X)$ that is no more Gaussian but includes non-linearities in a perturbative manner. In practice, we use a Gram-Charlier expansion following previous works in the literature including [Pogosyan et al. \(2009a\)](#), [Gay et al. \(2012\)](#) and [Codis et al. \(2013\)](#).

4.1. Gram-Charlier expansion of the joint distribution

For a set X of random fields, the Gram-Charlier expansion of the joint PDF reads

$$\mathcal{P}(X) = \mathcal{P}_G(X) \left[1 + \sum_{n=3}^{\infty} \frac{1}{n!} \text{Tr}[\langle X^n \rangle_{\text{GC}} h_n(X)] \right], \quad (26)$$

where $\mathcal{P}_G(X)$ is a Gaussian kernel as defined in Eq. (12), $h_n(x)$ are the Hermite tensors $h_n(X) = (-1)^n \mathcal{P}_G^{-1}(X) \partial^n \mathcal{P}_G(X) / \partial X^n$ and $\langle X^n \rangle_{\text{GC}} = \langle h_n(X) \rangle$ are the Gram Charlier coefficients.

For our rotation-invariant variables J_1, J_2, J_3 , this translates into the following expression

$$\begin{aligned} \mathcal{P}(J_1, J_2, J_3) = \mathcal{P}_G(J_1, J_2, J_3) & \left[1 + \sum_{n=3}^{\infty} \sum_{k,l}^{k+2l=n} \frac{(-1)^l 5^l \times 3}{k!(3+2l)!} \langle J_1^k J_2^l \rangle_{\text{GC}} H_k(J_1) L_l^{(3/2)} \left(\frac{5}{2} J_2 \right) \right. \\ & \left. + \sum_{n=3}^{\infty} \sum_k^{k+3=n} \frac{25}{k! \times 21} \langle J_1^k J_3 \rangle_{\text{GC}} H_k(J_1) J_3 + \sum_{n=5}^{\infty} \sum_{k,l,m=1}^{k+2l+3m=n} \frac{c_{lm}}{k!} \langle J_1^k J_2^l J_3^m \rangle_{\text{GC}} H_k(J_1) F_{lm}(J_2, J_3) \right], \quad (27) \end{aligned}$$

where $\mathcal{P}_G(J_1, J_2, J_3)$ is the Gaussian case given by Eq. (17), H_n are the successive Hermite polynomials, $L_l^{(\alpha)}(x)$ are the generalised Laguerre polynomials, c_{lm} are the normalisation coefficients that we leave undetermined and the orthogonal polynomials F_{lm} associated with J_2 and J_3 are such that

$$\int_0^{\infty} dJ_2 \int_{-J_2^{3/2}}^{J_2^{3/2}} dJ_3 \mathcal{P}_G(J_1, J_2, J_3) F_{lm}(J_2, J_3) F_{l'm'}(J_2, J_3) = \delta_l^{l'} \delta_m^{m'}.$$

We have, for example, these special cases: $F_{l0} = \sqrt{3 \times 2^l \times l! / (3+2l)!} \times L_l^{3/2}(5J_2/2)$ and $F_{01} = 5J_3 / \sqrt{21}$.

Here, we focus on the first corrective term, $n = 3$, such that

$$\mathcal{P}(J_1, J_2, J_3) = \mathcal{P}_G(J_1, J_2, J_3) \left[1 + \frac{1}{6} \langle J_1^3 \rangle_{\text{GC}} H_3(J_1) - \langle J_1 J_2 \rangle_{\text{GC}} H_1(J_1) L_1^{(3/2)} \left(\frac{5}{2} J_2 \right) + \frac{25}{21} \langle J_3 \rangle_{\text{GC}} J_3 \right] + o(\sigma_0^2). \quad (28)$$

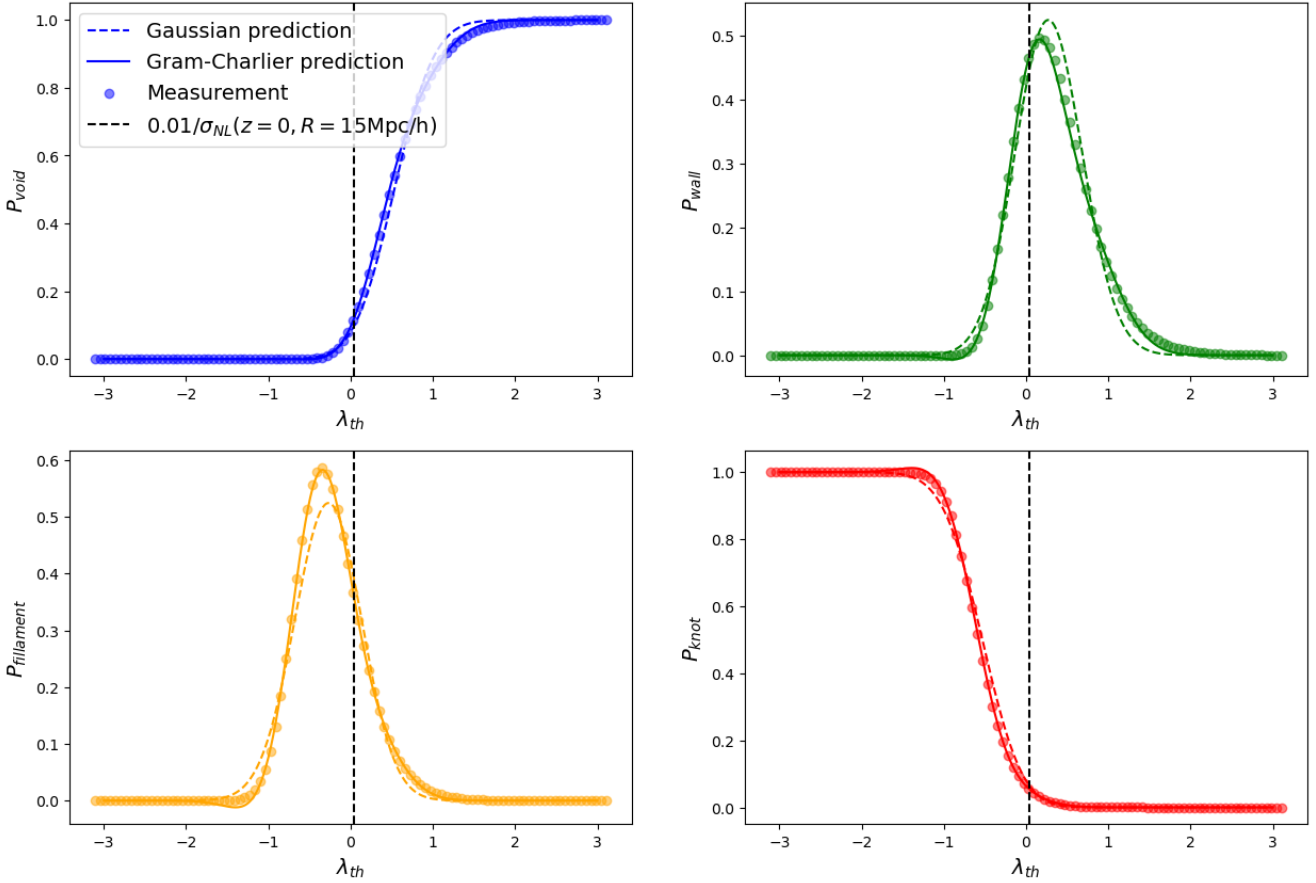


Fig. 3. Probability of the environment as a function of the threshold for a Gaussian random field (dashed lines), Gram-Charlier (solid lines), and measurement in the simulations (dots) at $R = 15 \text{ Mpc } h^{-1}$ and $z = 0$. The vertical black line is the threshold $\Lambda_{\text{th}} = 0.01$ (thus $\lambda_{\text{th}} = \Lambda_{\text{th}}/\sigma_{\text{NL}}$) used in this study.

The Gram-Charlier cumulants read

$$\langle J_1^3 \rangle_{\text{GC}} = \langle H_3(J_1) \rangle, \quad \langle J_1 J_2 \rangle_{\text{GC}} = -\frac{2}{5} \left\langle H_1(J_1) L_1^{(3/2)} \left(\frac{5}{2} J_2 \right) \right\rangle, \quad \langle J_3 \rangle_{\text{GC}} = \langle J_3 \rangle, \quad (29)$$

such that we finally obtain the Gram-Charlier expression at first non-linear order (NLO):

$$\mathcal{P}(J_1, J_2, J_3) = \mathcal{P}_{\text{G}}(J_1, J_2, J_3) \left[1 + \frac{1}{6} \langle J_1^3 \rangle (J_1^3 - 3J_1) + \frac{5}{2} \langle J_1 J_2 \rangle J_1 (J_2 - 1) + \frac{25}{21} \langle J_3 \rangle J_3 \right] + o(\sigma_0^2). \quad (30)$$

Three cumulants appear in the next-to-leading-order Gram-Charlier probability distribution function (Eq. (30)): $\langle J_1^3 \rangle$, $\langle J_1 J_2 \rangle$ and $\langle J_3 \rangle$. In Eulerian perturbation theory, those cumulants are linear in σ_0 at tree order¹ and we therefore introduce the reduced cumulants $S_3 = \langle J_1^3 \rangle / \sigma_0$, $U_3 = \langle J_1 J_2 \rangle / \sigma_0$ and $V_3 = \langle J_3 \rangle / \sigma_0$, which are constant in time at tree order. For example, S_3 is the usual cosmological skewness, whose analytical prediction at tree order is well known (for example, Peebles 1980; Juszkiewicz et al. 1993; Lokas et al. 1995; Colombi et al. 2000). The other two reduced cumulants can be computed in a similar manner, described in Appendix A. Figure 4 shows the resulting tree-order cumulants as a function of the smoothing scale in dashed black, compared to the measurements in the simulation at different redshifts (as shown with different colours). We see a good agreement at almost all smoothing scales and redshifts as the prediction is almost always within the error bars. For large smoothing scales, the error bars increase due to the finite volume of the Quijote simulation which thus misses large wave modes. As expected, at low redshift and small smoothing scales, departures from tree-order predictions are seen as the non-linearities increase. In this work, we focus on mildly non-linear scales (about $10 \text{ Mpc } h^{-1}$ and above) where a perturbative treatment is accurate, as illustrated by the values of the cumulants depicted in this figure.

4.2. Cosmic web abundances at next-to-leading order

We now turn to the explicit computation of the probability of different cosmic environments at next-to-leading order in the Gram-Charlier formalism. Let us first rewrite the Gram-Charlier expansion of the joint distribution of the rotational invariants of the tidal

¹ We remind the reader that we are working with fields renormalised by their variance, so that $\langle J_1^3 \rangle = \langle \delta^3 \rangle / \sigma_0^3$.

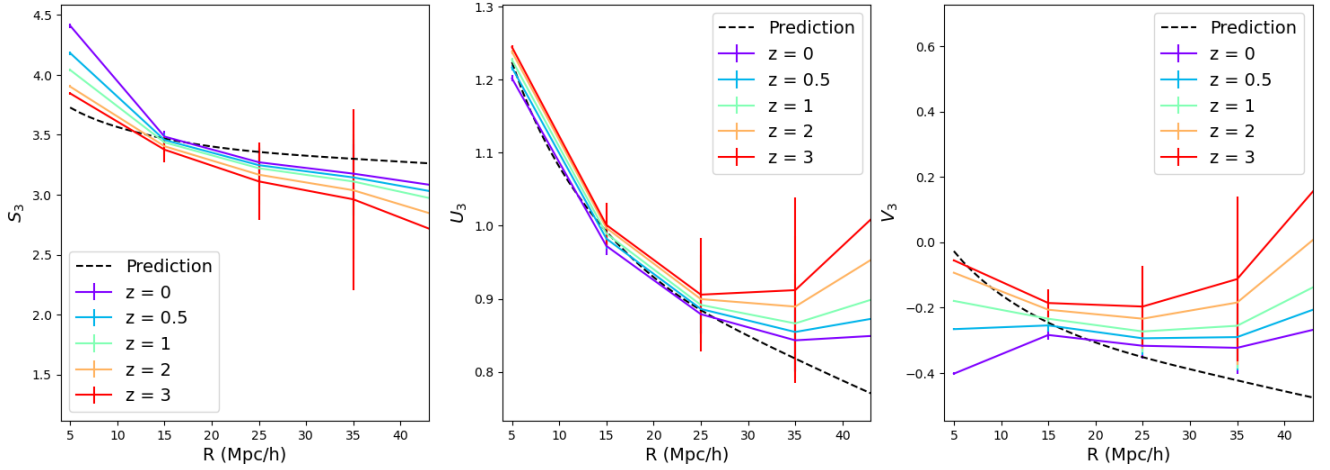


Fig. 4. Cumulants S_3 , U_3 , and V_3 as a function of the smoothing scale R . The dashed black line is the analytical prediction. The coloured lines are the measurements from the simulations at different redshifts.

tensor as

$$\mathcal{P}(J_1, J_2, J_3) = \mathcal{P}_G(J_1, J_2, J_3) + \sigma_0 \mathcal{P}^{(3)}(J_1, J_2, J_3) + O(\sigma_0^2), \quad (31)$$

where \mathcal{P}_G is the Gaussian part and $\mathcal{P}^{(3)}$ is the first non-Gaussian corrective term

$$\mathcal{P}^{(3)}(J_1, J_2, J_3) = \frac{25\sqrt{5}}{12\pi} \exp\left[-\frac{J_1^2}{2} - \frac{5J_2}{2}\right] \left(\frac{1}{6}(J_1^3 - 3J_1)S_3 + \frac{5}{2}J_1(J_2 - 1)U_3 + \frac{25}{21}J_3V_3\right). \quad (32)$$

Integrating this distribution function over the constraints given in Eqs. (18)–(21) will give us the expression of the probability of a given environment (E) as

$$\begin{aligned} P_{(E)}(\lambda_{\text{th}}) &= P_{G,(E)}(\lambda_{\text{th}}) + \sigma_0 \mathcal{P}^{(3)}_{(E)}(\lambda_{\text{th}}) \\ &= P_{G,(E)}(\lambda_{\text{th}}) + \sigma_0 [s_{(E)}(\lambda_{\text{th}})S_3 + u_{(E)}(\lambda_{\text{th}})U_3 + v_{(E)}(\lambda_{\text{th}})V_3], \end{aligned} \quad (33)$$

where s , u and v are threshold-dependent functions and where the Gaussian probability of the environments was derived in Eqs. (22)–(25).

Again, two degrees of freedom can be integrated out and one remains to be integrated numerically. The resulting expressions for $s(\lambda_{\text{th}})$, $u(\lambda_{\text{th}})$, and $v(\lambda_{\text{th}})$ in each environment are given below (written in terms of 1D integrals over J_2).

4.2.1. Void

For voids, we get

$$\begin{aligned} s_{\text{void}}(\lambda_{\text{th}}) &= \int_0^\infty \frac{25\sqrt{5}}{144\pi} e^{\frac{1}{2}(-9)(J_2 + \lambda_{\text{th}}^2)} \left[-3\sqrt{2\pi} e^{2J_2 + \frac{9J_2^2}{2}} \operatorname{erf}\left(\frac{2\sqrt{J_2} - 3\lambda_{\text{th}}}{\sqrt{2}}\right) + 3\sqrt{2\pi} e^{2J_2 + \frac{9J_2^2}{2}} \operatorname{erf}\left(\frac{\sqrt{J_2} - 3\lambda_{\text{th}}}{\sqrt{2}}\right) \right. \\ &\quad + \sqrt{J_2} e^{3\sqrt{J_2}\lambda_{\text{th}}} \left[e^{3\sqrt{J_2}\lambda_{\text{th}}} (-48J_2^{3/2}\lambda_{\text{th}} + 16J_2^2 + 36J_2\lambda_{\text{th}}^2 - 27\sqrt{J_2}\lambda_{\text{th}} \right. \\ &\quad \left. \left. + 14J_2 + 12) - 6e^{\frac{3J_2}{2}} \right] \right] + \frac{25\sqrt{5}}{36\pi} J_2^{3/2} e^{6\sqrt{J_2}\lambda_{\text{th}} - \frac{9J_2}{2} - \frac{9\lambda_{\text{th}}^2}{2}} \left(1 - (2\sqrt{J_2} - 3\lambda_{\text{th}})^2\right) dJ_2, \end{aligned} \quad (34)$$

$$\begin{aligned} u_{\text{void}}(\lambda_{\text{th}}) &= \int_0^\infty -\frac{125\sqrt{5}}{96\pi} (J_2 - 1) e^{\frac{1}{2}(-9)(J_2 + \lambda_{\text{th}}^2)} \left[-3\sqrt{2\pi} e^{2J_2 + \frac{9J_2^2}{2}} (J_2 - 9\lambda_{\text{th}}^2 - 1) \times \operatorname{erf}\left(\frac{2\sqrt{J_2} - 3\lambda_{\text{th}}}{\sqrt{2}}\right) + 3\sqrt{2\pi} e^{2J_2 + \frac{9J_2^2}{2}} (J_2 - 9\lambda_{\text{th}}^2 - 1) \operatorname{erf}\left(\frac{\sqrt{J_2} - 3\lambda_{\text{th}}}{\sqrt{2}}\right) \right. \\ &\quad \left. - 2e^{6\sqrt{J_2}\lambda_{\text{th}}} (4J_2^{3/2} + 6\sqrt{J_2} + 9\lambda_{\text{th}}) + 6e^{3\sqrt{J_2}\lambda_{\text{th}} + \frac{3J_2}{2}} (\sqrt{J_2} + 3\lambda_{\text{th}}) \right] - \frac{125\sqrt{5}}{12\pi} (J_2 - 1) J_2^{3/2} e^{6\sqrt{J_2}\lambda_{\text{th}} - \frac{9J_2}{2} - \frac{9\lambda_{\text{th}}^2}{2}} dJ_2, \end{aligned} \quad (35)$$

$$\begin{aligned} v_{\text{void}}(\lambda_{\text{th}}) &= \int_0^\infty \frac{625\sqrt{5}}{4032\pi} e^{-\frac{5J_2}{2}} \left[-\sqrt{2\pi} \left[-4J_2^3 + 9J_2^2 + 243(5 - 2J_2)\lambda_{\text{th}}^4 + 81((J_2 - 4)J_2 + 5)\lambda_{\text{th}}^2 - 18J_2 + 729\lambda_{\text{th}}^6 + 15 \right] \right. \\ &\quad \times \operatorname{erf}\left(\frac{3\lambda_{\text{th}} - 2\sqrt{J_2}}{\sqrt{2}}\right) + \sqrt{2\pi} \left[-4J_2^3 + 9J_2^2 + 243(5 - 2J_2)\lambda_{\text{th}}^4 + 81((J_2 - 4)J_2 + 5)\lambda_{\text{th}}^2 - 18J_2 + 729\lambda_{\text{th}}^6 + 15 \right] \operatorname{erf}\left(\frac{3\lambda_{\text{th}} - \sqrt{J_2}}{\sqrt{2}}\right) \\ &\quad + 2e^{-\frac{1}{2}(\sqrt{J_2} - 3\lambda_{\text{th}})^2} \left[J_2^{3/2} (- (45\lambda_{\text{th}}^2 + 13)) + 12J_2^2\lambda_{\text{th}} + 4J_2^{5/2} - 9J_2\lambda_{\text{th}} (15\lambda_{\text{th}}^2 + 7) + 3\sqrt{J_2} (27\lambda_{\text{th}}^4 + 36\lambda_{\text{th}}^2 + 5) \right. \\ &\quad \left. + 9\lambda_{\text{th}} (27\lambda_{\text{th}}^4 + 42\lambda_{\text{th}}^2 + 11) \right] - 2e^{-\frac{1}{2}(2\sqrt{J_2} - 3\lambda_{\text{th}})^2} \left[J_2^{3/2} (4 - 36\lambda_{\text{th}}^2) + 3J_2^2\lambda_{\text{th}} + 2J_2^{5/2} + J_2 (18\lambda_{\text{th}} - 54\lambda_{\text{th}}^3) \right. \\ &\quad \left. + 6\sqrt{J_2} (27\lambda_{\text{th}}^4 + 36\lambda_{\text{th}}^2 + 5) + 9\lambda_{\text{th}} (27\lambda_{\text{th}}^4 + 42\lambda_{\text{th}}^2 + 11) \right] \Big] dJ_2, \end{aligned} \quad (36)$$

where $\operatorname{erf}(z)$ is, again, the error function. Once that analytical expression is obtained, we can perform the numerical integration over J_2 . We display $s(\lambda_{\text{th}})$, $u(\lambda_{\text{th}})$, and $v(\lambda_{\text{th}})$ as a function of the threshold in Fig. B.1.

4.2.2. Wall

For walls, the same procedure gives

$$s_{\text{wall}}(\lambda_{\text{th}}) = \int_0^\infty \frac{25\sqrt{5}}{48\pi} e^{-\frac{9J_2}{2}} \left[-\sqrt{2\pi} e^{2J_2} \operatorname{erf}\left(\frac{3\lambda_{\text{th}} - 2\sqrt{J_2}}{\sqrt{2}}\right) + \sqrt{2\pi} e^{2J_2} \operatorname{erf}\left(\frac{\sqrt{J_2} + 3\lambda_{\text{th}}}{\sqrt{2}}\right) + \sqrt{J_2} \left(-e^{\frac{1}{2}(-3)\lambda_{\text{th}}(2\sqrt{J_2} + 3\lambda_{\text{th}})}\right) \right. \\ \left. \times \left(e^{9\sqrt{J_2}\lambda_{\text{th}}} (-9\sqrt{J_2}\lambda_{\text{th}} + 6J_2 + 4) + 2e^{\frac{3J_2}{2}} \right) \right] dJ_2, \quad (37)$$

$$u_{\text{wall}}(\lambda_{\text{th}}) = \int_0^\infty \frac{125\sqrt{5}}{32\pi} e^{-\frac{9J_2}{2}} (J_2 - 1) \left[-\sqrt{2\pi} e^{2J_2} (J_2 - 9\lambda_{\text{th}}^2 - 1) \operatorname{erf}\left(\frac{\sqrt{J_2} + 3\lambda_{\text{th}}}{\sqrt{2}}\right) + \sqrt{2\pi} e^{2J_2} (J_2 - 9\lambda_{\text{th}}^2 - 1) \operatorname{erf}\left(\frac{3\lambda_{\text{th}} - 2\sqrt{J_2}}{\sqrt{2}}\right) \right. \\ \left. - 2e^{\frac{1}{2}(-3)\lambda_{\text{th}}(2\sqrt{J_2} + 3\lambda_{\text{th}})} \left[e^{9\sqrt{J_2}\lambda_{\text{th}}} (2\sqrt{J_2} + 3\lambda_{\text{th}}) + e^{\frac{3J_2}{2}} (\sqrt{J_2} - 3\lambda_{\text{th}}) \right] \right] dJ_2, \quad (38)$$

$$v_{\text{wall}}(\lambda_{\text{th}}) = \int_0^\infty -\frac{625\sqrt{5}}{4032\pi} e^{-\frac{5J_2}{2}} \left[-\sqrt{2\pi} \left[-4J_2^3 + 9J_2^2 + 243(5 - 2J_2)\lambda_{\text{th}}^4 + 81((J_2 - 4)J_2 + 5)\lambda_{\text{th}}^2 - 18J_2 + 729\lambda_{\text{th}}^6 + 15 \right] \operatorname{erf}\left(\frac{3\lambda_{\text{th}} - 2\sqrt{J_2}}{\sqrt{2}}\right) \right. \\ \left. + \sqrt{2\pi} \left[-4J_2^3 + 9J_2^2 + 243(5 - 2J_2)\lambda_{\text{th}}^4 + 81((J_2 - 4)J_2 + 5)\lambda_{\text{th}}^2 - 18J_2 + 729\lambda_{\text{th}}^6 + 15 \right] \operatorname{erf}\left(\frac{\sqrt{J_2} + 3\lambda_{\text{th}}}{\sqrt{2}}\right) \right. \\ \left. - 2e^{-\frac{1}{2}(2\sqrt{J_2} - 3\lambda_{\text{th}})^2} \left[J_2^{3/2} (4 - 36\lambda_{\text{th}}^2) + 3J_2^2\lambda_{\text{th}} + 2J_2^{5/2} + J_2 (18\lambda_{\text{th}} - 54\lambda_{\text{th}}^3) + 6\sqrt{J_2} (27\lambda_{\text{th}}^4 + 36\lambda_{\text{th}}^2 + 5) + 9\lambda_{\text{th}} (27\lambda_{\text{th}}^4 + 42\lambda_{\text{th}}^2 + 11) \right] \right. \\ \left. + 2e^{-\frac{1}{2}(\sqrt{J_2} + 3\lambda_{\text{th}})^2} \left[J_2^{3/2} (45\lambda_{\text{th}}^2 + 13) + 12J_2^2\lambda_{\text{th}} - 4J_2^{5/2} - 9J_2\lambda_{\text{th}} (15\lambda_{\text{th}}^2 + 7) - 3\sqrt{J_2} (27\lambda_{\text{th}}^4 + 36\lambda_{\text{th}}^2 + 5) + 9\lambda_{\text{th}} (27\lambda_{\text{th}}^4 + 42\lambda_{\text{th}}^2 + 11) \right] \right] dJ_2. \quad (39)$$

4.2.3. Filament

For filaments, we get

$$s_{\text{filament}}(\lambda_{\text{th}}) = \int_0^\infty -\frac{25\sqrt{5}}{144\pi} e^{\frac{1}{2}(-3)(4\sqrt{J_2}\lambda_{\text{th}} + 3J_2 + 3\lambda_{\text{th}}^2)} \left[3\sqrt{2\pi} e^{\frac{1}{2}(2\sqrt{J_2} + 3\lambda_{\text{th}})^2} \times \operatorname{erf}\left(\frac{2\sqrt{J_2} + 3\lambda_{\text{th}}}{\sqrt{2}}\right) + 3\sqrt{2\pi} e^{\frac{1}{2}(2\sqrt{J_2} + 3\lambda_{\text{th}})^2} \operatorname{erf}\left(\frac{\sqrt{J_2} - 3\lambda_{\text{th}}}{\sqrt{2}}\right) \right. \\ \left. - 3\sqrt{J_2} \left(2e^{9\sqrt{J_2}\lambda_{\text{th}} + \frac{3J_2}{2}} + 9\sqrt{J_2}\lambda_{\text{th}} + 6J_2 + 4 \right) \right] dJ_2, \quad (40)$$

$$u_{\text{filament}}(\lambda_{\text{th}}) = \int_0^\infty -\frac{125\sqrt{5}}{48\pi} (J_2 - 1) \left[3\sqrt{\frac{\pi}{2}} e^{-\frac{5J_2}{2}} (J_2 - 9\lambda_{\text{th}}^2 - 1) \times \operatorname{erf}\left(\frac{3\lambda_{\text{th}} - \sqrt{J_2}}{\sqrt{2}}\right) - \frac{3}{2} e^{\frac{1}{2}(-3)(4\sqrt{J_2}\lambda_{\text{th}} + 3J_2 + 3\lambda_{\text{th}}^2)} \right. \\ \left. \times \left[\sqrt{2\pi} e^{\frac{1}{2}(2\sqrt{J_2} + 3\lambda_{\text{th}})^2} (J_2 - 9\lambda_{\text{th}}^2 - 1) \times \operatorname{erf}\left(\frac{2\sqrt{J_2} + 3\lambda_{\text{th}}}{\sqrt{2}}\right) + 6\lambda_{\text{th}} \left(e^{9\sqrt{J_2}\lambda_{\text{th}} + \frac{3J_2}{2}} - 1 \right) + 2\sqrt{J_2} \left(e^{9\sqrt{J_2}\lambda_{\text{th}} + \frac{3J_2}{2}} + 2 \right) \right] \right] dJ_2, \quad (41)$$

$$v_{\text{filament}}(\lambda_{\text{th}}) = \int_0^\infty \frac{625\sqrt{5}}{4032\pi} e^{-\frac{5J_2}{2}} \left[-\sqrt{2\pi} \left[-4J_2^3 + 9J_2^2 + 243(5 - 2J_2)\lambda_{\text{th}}^4 + 81((J_2 - 4)J_2 + 5)\lambda_{\text{th}}^2 - 18J_2 + 729\lambda_{\text{th}}^6 + 15 \right] \operatorname{erf}\left(\frac{3\lambda_{\text{th}} - \sqrt{J_2}}{\sqrt{2}}\right) \right. \\ \left. + \sqrt{2\pi} \left[-4J_2^3 + 9J_2^2 + 243(5 - 2J_2)\lambda_{\text{th}}^4 + 81((J_2 - 4)J_2 + 5)\lambda_{\text{th}}^2 - 18J_2 + 729\lambda_{\text{th}}^6 + 15 \right] \operatorname{erf}\left(\frac{2\sqrt{J_2} + 3\lambda_{\text{th}}}{\sqrt{2}}\right) - 2e^{-\frac{1}{2}(\sqrt{J_2} - 3\lambda_{\text{th}})^2} \right. \\ \left. \times \left[J_2^{3/2} (-45\lambda_{\text{th}}^2 + 13) + 12J_2^2\lambda_{\text{th}} + 4J_2^{5/2} - 9J_2\lambda_{\text{th}} (15\lambda_{\text{th}}^2 + 7) + 3\sqrt{J_2} (27\lambda_{\text{th}}^4 + 36\lambda_{\text{th}}^2 + 5) + 9\lambda_{\text{th}} (27\lambda_{\text{th}}^4 + 42\lambda_{\text{th}}^2 + 11) \right] \right. \\ \left. + e^{-\frac{1}{2}(2\sqrt{J_2} + 3\lambda_{\text{th}})^2} \left[8J_2^{3/2} (9\lambda_{\text{th}}^2 - 1) + 6J_2^2\lambda_{\text{th}} - 4J_2^{5/2} - 36J_2\lambda_{\text{th}} (3\lambda_{\text{th}}^2 - 1) - 12\sqrt{J_2} (27\lambda_{\text{th}}^4 + 36\lambda_{\text{th}}^2 + 5) + 18\lambda_{\text{th}} (27\lambda_{\text{th}}^4 + 42\lambda_{\text{th}}^2 + 11) \right] \right] dJ_2. \quad (42)$$

4.2.4. Knots

Finally, for knots:

$$s_{\text{knot}}(\lambda_{\text{th}}) = \int_0^\infty \frac{25\sqrt{5}}{144\pi} e^{\frac{1}{2}(-3)(4\sqrt{J_2}\lambda_{\text{th}} + 3J_2 + 3\lambda_{\text{th}}^2)} \left[-3\sqrt{2\pi} e^{\frac{1}{2}(2\sqrt{J_2} + 3\lambda_{\text{th}})^2} \times \operatorname{erf}\left(\frac{\sqrt{J_2} + 3\lambda_{\text{th}}}{\sqrt{2}}\right) + 3\sqrt{2\pi} e^{\frac{1}{2}(2\sqrt{J_2} + 3\lambda_{\text{th}})^2} \operatorname{erf}\left(\frac{2\sqrt{J_2} + 3\lambda_{\text{th}}}{\sqrt{2}}\right) \right. \\ \left. - \sqrt{J_2} \left[48J_2^{3/2}\lambda_{\text{th}} + 16J_2^2 + 36J_2\lambda_{\text{th}}^2 - 6e^{3\sqrt{J_2}\lambda_{\text{th}} + \frac{3J_2}{2}} + 27\sqrt{J_2}\lambda_{\text{th}} + 14J_2 + 12 \right] + \frac{25\sqrt{5}}{36\pi} J_2^{3/2} e^{\frac{1}{2}(-3)(4\sqrt{J_2}\lambda_{\text{th}} + 3J_2 + 3\lambda_{\text{th}}^2)} (12\sqrt{J_2}\lambda_{\text{th}} + 4J_2 + 9\lambda_{\text{th}}^2 - 1) \right] dJ_2, \quad (43)$$

$$u_{\text{knot}}(\lambda_{\text{th}}) = \int_0^\infty \frac{125\sqrt{5}}{96\pi} (J_2 - 1) e^{\frac{1}{2}(-3)(4\sqrt{J_2}\lambda_{\text{th}} + 3J_2 + 3\lambda_{\text{th}}^2)} \times \left[-3\sqrt{2\pi} e^{\frac{1}{2}(2\sqrt{J_2} + 3\lambda_{\text{th}})^2} (J_2 - 9\lambda_{\text{th}}^2 - 1) \operatorname{erf}\left(\frac{2\sqrt{J_2} + 3\lambda_{\text{th}}}{\sqrt{2}}\right) \right. \\ \left. + 3\sqrt{2\pi} e^{\frac{1}{2}(2\sqrt{J_2} + 3\lambda_{\text{th}})^2} (J_2 - 9\lambda_{\text{th}}^2 - 1) \operatorname{erf}\left(\frac{\sqrt{J_2} + 3\lambda_{\text{th}}}{\sqrt{2}}\right) - 8J_2^{3/2} - 18\lambda_{\text{th}} \left(e^{3\sqrt{J_2}\lambda_{\text{th}} + \frac{3J_2}{2}} - 1 \right) + 6\sqrt{J_2} \left(e^{3\sqrt{J_2}\lambda_{\text{th}} + \frac{3J_2}{2}} - 2 \right) \right. \\ \left. + \frac{125\sqrt{5}}{12\pi} (J_2 - 1) J_2^{3/2} e^{\frac{1}{2}(-3)(4\sqrt{J_2}\lambda_{\text{th}} + 3J_2 + 3\lambda_{\text{th}}^2)} \right] dJ_2, \quad (44)$$

$$v_{\text{knot}}(\lambda_{\text{th}}) = \int_0^\infty -\frac{625\sqrt{5}}{4032\pi} e^{-\frac{5J_2}{2}} \left[-\sqrt{2\pi} \left[-4J_2^3 + 9J_2^2 + 243(5 - 2J_2)\lambda_{\text{th}}^4 + 81((J_2 - 4)J_2 + 5)\lambda_{\text{th}}^2 - 18J_2 + 729\lambda_{\text{th}}^6 + 15 \right] \operatorname{erf}\left(\frac{\sqrt{J_2} + 3\lambda_{\text{th}}}{\sqrt{2}}\right) \right. \\ \left. + \sqrt{2\pi} \left[-4J_2^3 + 9J_2^2 + 243(5 - 2J_2)\lambda_{\text{th}}^4 + 81((J_2 - 4)J_2 + 5)\lambda_{\text{th}}^2 - 18J_2 + 729\lambda_{\text{th}}^6 + 15 \right] \operatorname{erf}\left(\frac{2\sqrt{J_2} + 3\lambda_{\text{th}}}{\sqrt{2}}\right) + e^{-\frac{1}{2}(2\sqrt{J_2} + 3\lambda_{\text{th}})^2} \left[8J_2^{3/2} (9\lambda_{\text{th}}^2 - 1) \right. \right. \\ \left. \left. + 6J_2^2\lambda_{\text{th}} - 4J_2^{5/2} - 36J_2\lambda_{\text{th}} (3\lambda_{\text{th}}^2 - 1) - 12\sqrt{J_2} (27\lambda_{\text{th}}^4 + 36\lambda_{\text{th}}^2 + 5) + 18\lambda_{\text{th}} (27\lambda_{\text{th}}^4 + 42\lambda_{\text{th}}^2 + 11) \right] \right] dJ_2.$$

$$\begin{aligned}
& + 6J_2^2\lambda_{\text{th}} - 4J_2^{5/2} - 36J_2\lambda_{\text{th}}(3\lambda_{\text{th}}^2 - 1) - 12\sqrt{J_2}(27\lambda_{\text{th}}^4 + 36\lambda_{\text{th}}^2 + 5) + 18\lambda_{\text{th}}(27\lambda_{\text{th}}^4 + 42\lambda_{\text{th}}^2 + 11) \Big] + 2e^{-\frac{1}{2}(\sqrt{J_2}+3\lambda_{\text{th}})^2} \Big[J_2^{3/2}(-45\lambda_{\text{th}}^2 + 13) - 12J_2^2\lambda_{\text{th}} \\
& + 4J_2^{5/2} + 9J_2\lambda_{\text{th}}(15\lambda_{\text{th}}^2 + 7) + 3\sqrt{J_2}(27\lambda_{\text{th}}^4 + 36\lambda_{\text{th}}^2 + 5) - 9\lambda_{\text{th}}(27\lambda_{\text{th}}^4 + 42\lambda_{\text{th}}^2 + 11) \Big] \Big] dJ_2. \tag{45}
\end{aligned}$$

Plugging the previous expressions for the s, u, v functions into Eq. (33) finally allows us to compute the next-to-leading-order correction to the (threshold-dependent) probability of the different cosmic environments in the T-web classification.

As for the Gaussian case, we choose to measure a rarity threshold in the simulations, and therefore we plug in the non-linear variance measured in the simulation as the normalisation factor appearing in the threshold λ_{th} . An alternative option to try and keep a meaningful threshold selection in terms of abundance or rarity could be to work with a filling factor, as mentioned at the end of Sect. 3.

The probabilities of the different environments as functions of redshift are displayed in Fig. 2; these were already partially described in Sect. 3 for the Gaussian case. We now focus on the solid lines, which represent the Gram-Charlier correction described in this section. For every environment, redshift, and smoothing scale, the results obtained with the Gram-Charlier always improve the prediction over the Gaussian case. The agreement between the simulation and our non-linear model is almost perfect; only at the most non-linear scales (very low redshift and small smoothing scale) do the predictions start to become slightly different from the measurements, although the NLO correction always improves upon the Gaussian case. We can also look at the abundance of voids, walls, filaments, and knots as a function of the threshold in Fig. 3, which was also already partially described in Sect. 3. We see that the Gram-Charlier correction not only improves the prediction over the Gaussian case but provides us with an extremely accurate prediction at all the relevant thresholds and for almost the whole spectrum of field values. Let us look at the contribution of each function that appears in our Gram-Charlier expansion: $s_{(\text{E})}(\lambda_{\text{th}})$, $u_{(\text{E})}(\lambda_{\text{th}})$, and $v_{(\text{E})}(\lambda_{\text{th}})$ in Eq. (33) as shown in Fig. B.1. We note that S_3 is always larger than U_3 and V_3 , implying that $s(\lambda_{\text{th}})$ is typically the dominant correction to the Gaussian case (and is actually related to the non-linear evolution of the threshold). This is the reason why our Gaussian model for the evolution of the cosmic web abundances already gives a fairly accurate prescription. The higher-order non-linear corrections are then modulated by the functions s, u, v depending on the chosen threshold.

5. Discussion and conclusion

Above, using the T-web classification with a rarity threshold $\lambda_{\text{th}} = (\Lambda_{\text{th}} = 0.01)/\sigma_{\text{NL}}$ (a commonly Λ_{th} value used in the literature), we describe a theoretical framework to accurately model the probabilities of voids, walls, filaments, and knots in the cosmic web. More precisely, as this computation requires knowledge of the joint probability distribution of the eigenvalues of the Hessian of the gravitational potential, we model an analogous object, which is the joint distribution between maximally decorrelated, rotational invariants that are polynomial in the Cartesian matrix components.

We first focused on the linear regime of structure formation where the density field is Gaussian (assuming Gaussian initial conditions) and where the distribution of the eigenvalues of the Hessian of the gravitational potential – and then a Gaussian field – are known. We find in that case that normalising the rarity threshold λ_{th} by the non-linear variance is surprisingly accurate with respect to measurements of the environment probabilities in the Quijote simulation in regimes where $\sigma \lesssim 0.1$ and captures most of the redshift and scale dependence (see Fig. 2).

To probe the mildly non-linear regime, we then accounted for corrections to the Gaussian case in a perturbative manner relying on a Gram-Charlier expansion at first non-linear order (Eq. (30)) for the joint distribution of our rotational invariants, and tree-order Eulerian perturbation theory to compute the Gram-Charlier coefficients. We showed that this correction to the Gaussian case increases the accuracy of predictions for the probability of cosmic environments at all considered redshifts and smoothing scales with threshold $\lambda_{\text{th}} = 0.01/\sigma_{\text{NL}}$ (see Fig. 2), but also when varying the threshold at fixed redshift and smoothing scale (see Fig. 3 for a reference case at $z = 0$ and $R = 15 \text{ Mpc } h^{-1}$).

In conclusion, all redshift- and scale-dependent evolution of T-web cosmic abundances previously observed in numerical simulations are shown to be predictable from first principles. In addition, having a precise theoretical model not only allows us to benchmark simulations but also to understand how these features depend on choices of threshold or background cosmological model for example. This information can be readily extracted from the predictions we provide here (without the need to run simulations or post-process them, etc.). This is notably important when one wants to use these cosmic web elements as a cosmological probe (to go typically beyond the power spectrum), which requires a cost-effective inference pipeline. We leave investigations along this line for future works. Let us also emphasise that we assume Gaussian initial conditions in this work, but primordial non-Gaussianities could easily be added to the formalism (following e.g., Codis et al. 2013), in which case a careful study could be performed of the extent to which cosmic web abundance depends on the physics of the primordial Universe.

Looking at Fig. 3 for the direct result of the threshold-dependent modulation of the non-Gaussian cumulants appearing in the Gram-Charlier expansion of Fig. B.1, we observe that the different probabilities can fluctuate significantly when changing the threshold, as was pointed out for instance by Forero-Romero et al. (2009), who studied the impact of the choice of the threshold on the percolation and fragmentation of the cosmic web. With the present formalism, all this information can be described from first principles, which can be very useful in defining a physically motivated threshold rather than relying on visual inspection. Working with the abundance or rarity of an environment or with a filling factor could definitely help in that direction.

Another possible extension of this work could be to perform a multi-scale analysis in order to study the properties of halos depending on their environment. The T-web description developed here would provide the constrained environment (as often used in simulations) while the fields on a smaller scale would characterise the halo. However, let us note that the T-web classification is a rather rough description of the cosmic web (being sensitive not only to the mass distribution but also its larger-scale environment) and more sophisticated frameworks have already been developed in the literature. Interestingly though, most analyses studying environmental effects have shown that the precise definition of the environment has typically (and maybe surprisingly) little effect on

the result (Libeskind et al. 2017) which is additional motivation for the development of theoretical predictions for even rudimentary cosmic web estimators such as that described here. In essence, the physical effects at play in this context are mostly tidal effects, which are naturally captured in the T-web description (Regaldo-Saint Blancard et al. 2021).

Acknowledgements. E.A. is partly supported by a PhD Joint Programme between the CNRS and the University of Arizona. AB's work is supported by the ORIGINS excellence cluster. S.C. acknowledges financial support from Fondation MERAC and by the SPHERES grant ANR-18-CE31-0009 of the French Agence Nationale de la Recherche. This work has made use of the Infinity Cluster hosted by Institut d'Astrophysique de Paris. We thank Stephane Rouberol for running this cluster smoothly for us.

References

- Abel, T., Hahn, O., & Kaehler, R. 2012, *MNRAS*, 427, 61
- Aragón-Calvo, M. A., Jones, B. J. T., van de Weygaert, R., & van der Hulst, J. M. 2007a, *A&A*, 474, 315
- Aragón-Calvo, M. A., van de Weygaert, R., Jones, B. J. T., & van der Hulst, J. M. 2007b, *ApJ*, 655, L5
- Aragón-Calvo, M. A., Platen, E., van de Weygaert, R., & Szalay, A. S. 2010a, *ApJ*, 723, 364
- Aragón-Calvo, M. A., van de Weygaert, R., & Jones, B. J. T. 2010b, *MNRAS*, 408, 2163
- Arnold, V. I., Shandarin, S. F., & Zeldovich, I. B. 1982, *Geophys. Astrophys. Fluid Dyn.*, 20, 111
- Bernardeau, F., Colombi, S., Gaztañaga, E., & Scoccimarro, R. 2002, *Phys. Rep.*, 367, 1
- Biswas, R., Alizadeh, E., & Wandelt, B. D. 2010, *Phys. Rev. D*, 82, 023002
- Bond, J. R., Kofman, L., & Pogosyan, D. 1996, *Nature*, 380, 603
- Bonnaire, T., Aghanim, N., Kuruvilla, J., & Decelle, A. 2022, *A&A*, 661, A146
- Carlesi, E., Knebe, A., Lewis, G. F., Wales, S., & Yepes, G. 2014, *MNRAS*, 439, 2943
- Castorina, E., Paranjape, A., Hahn, O., & Sheth, R. K. 2016, arXiv e-prints [arXiv:1611.03619]
- Codis, S., Pichon, C., Devriendt, J., et al. 2012, *MNRAS*, 427, 3320
- Codis, S., Pichon, C., Pogosyan, D., Bernardeau, F., & Matsubara, T. 2013, *MNRAS*, 435, 531
- Codis, S., Jindal, A., Chisari, N. E., et al. 2018a, *MNRAS*, 481, 4753
- Codis, S., Pogosyan, D., & Pichon, C. 2018b, *MNRAS*, 479, 973
- Colombi, S., Pogosyan, D., & Souradeep, T. 2000, *Phys. Rev. Lett.*, 85, 5515
- Cui, W., Knebe, A., Yepes, G., et al. 2017, *MNRAS*, 473, 68
- Cui, W., Knebe, A., Libeskind, N. I., et al. 2019, *MNRAS*, 480, 2898
- de Lapparent, V., Geller, M. J., & Huchra, J. P. 1986, *ApJ*, 302, L1
- Desjacques, V., Jeong, D., & Schmidt, F. 2018, *J. Cosmol. Astropart. Phys.*, 2018, 017
- Dome, T., Fialkov, A., Sartorio, N., & Mocz, P. 2023, *MNRAS*, 525, 348
- Doroshkevich, A. G. 1970, *Astrophysics*, 6, 320
- Falck, B. L., Neyrinck, M. C., & Szalay, A. S. 2012, *ApJ*, 754, 126
- Fard, M. A., Taamoli, S., & Baghran, S. 2019, *MNRAS*, 489, 900
- Feldbrugge, J., van de Weygaert, R., Hidding, J., & Feldbrugge, J. 2018, *J. Cosmol. Astropart. Phys.*, 2018, 027
- Feldbrugge, J., van Engelen, M., van de Weygaert, R., Pranav, P., & Vegter, G. 2019, *J. Cosmol. Astropart. Phys.*, 2019, 052
- Fisher, J. D., Faltenbacher, A., & Johnson, M. S. T. 2016, *MNRAS*, 458, 1517
- Forero-Romero, J. E., Hoffman, Y., Gottläber, S., Klypin, A., & Yepes, G. 2009, *MNRAS*, 396, 1815
- Gay, C., Pichon, C., & Pogosyan, D. 2012, *Phys. Rev. D*, 85, 023011
- Gott, J. R., III, Weinberg, D. H., & Melott, A. L. 1987, *ApJ*, 319, 1
- Hahn, O., Carollo, C. M., Porciani, C., & Dekel, A. 2007b, *MNRAS*, 381, 41
- Hahn, O., Porciani, C., Carollo, C. M., & Dekel, A. 2007a, *MNRAS*, 375, 489
- Hasan, F., Burchett, J. N., Abeyta, A., et al. 2023, *ApJ*, 950, 114
- Hellwing, W. A., Cautun, M., van de Weygaert, R., & Jones, B. T. 2021, *Phys. Rev. D*, 103, 063517
- Hoffman, Y., Metuki, O., Yepes, G., et al. 2012, *MNRAS*, 425, 2049
- Juszkiewicz, R., Bouchet, F. R., & Colombi, S. 1993, *ApJ*, 412, L9
- Klypin, A. A., & Shandarin, S. F. 1983, *MNRAS*, 204, 891
- Kraljic, K., Arnouts, S., Pichon, C., et al. 2018, *MNRAS*, 474, 547
- Leclercq, F., Lavaux, G., Jasche, J., & Wandelt, B. 2016, *J. Cosmol. Astropart. Phys.*, 2016, 027
- Lee, J., & Park, D. 2009, *ApJ*, 696, L10
- Lewis, A., Challinor, A., & Lasenby, A. 2000, *ApJ*, 538, 473
- Libeskind, N. I., Hoffman, Y., Forero-Romero, J., et al. 2013, *MNRAS*, 428, 2489
- Libeskind, N. I., van de Weygaert, R., Cautun, M., et al. 2017, *MNRAS*, 473, 1195
- Lokas, E. L., Juszkiewicz, R., Weinberg, D. H., & Bouchet, F. R. 1995, *MNRAS*, 274, 730
- Matsubara, T. 2003, *ApJ*, 584, 1
- Metuki, O., Libeskind, N. I., Hoffman, Y., Crain, R. A., & Theuns, T. 2015, *MNRAS*, 446, 1458
- Nuza, S. E., Kitaura, F.-S., Heß, S., Libeskind, N. I., & Müller, V. 2014, *MNRAS*, 445, 988
- Peebles, P. J. E. 1980, *The Large-scale Structure of the Universe* (Princeton: Princeton University Press)
- Pogosyan, D., Gay, C., & Pichon, C. 2009a, *Phys. Rev. D*, 80, 081301
- Pogosyan, D., Pichon, C., Gay, C., et al. 2009b, *MNRAS*, 396, 635
- Poudel, A., Heinämäki, P., Tempel, E., et al. 2017, *A&A*, 597, A86
- Regaldo-Saint Blancard, B., Codis, S., Bond, J. R., & Stein, G. 2021, *MNRAS*, 504, 1694
- Rossi, G. 2012, *MNRAS*, 421, 296
- Shandarin, S. F. 2011, *J. Cosmol. Astropart. Phys.*, 2011, 015
- Sousbie, T. 2011, *MNRAS*, 414, 350
- Sousbie, T., Pichon, C., Colombi, S., Novikov, D., & Pogosyan, D. 2008, *MNRAS*, 383, 1655
- Sousbie, T., Colombi, S., & Pichon, C. 2009, *MNRAS*, 393, 457
- Suárez-Pérez, J. F., Camargo, Y., Li, X.-D., & Forero-Romero, J. E. 2021, *ApJ*, 922, 204
- van de Weygaert, R., & Bertschinger, E. 1996, *MNRAS*, 281, 84
- Villaescusa-Navarro, F., Hahn, C., Massara, E., et al. 2020, *ApJS*, 250, 2
- Wang, X., & Szalay, A. 2014, arXiv e-prints [arXiv:1411.4117]
- Wang, X., Szalay, A., Aragón-Calvo, M. A., Neyrinck, M. C., & Eyrink, G. L. 2014, *ApJ*, 793, 58
- Zel'dovich, Y. B. 1970, *A&A*, 5, 84

Appendix A: Computation of cumulants

As discussed in section 4.1 and shown in equation (30), three cumulants are needed to predict cosmic web abundances at NLO. To obtain the value of these three cumulants, we use tree-order Eulerian perturbation theory.

Given the scaling of the tree-order cumulants with the linear variance, and as we are working with normalised variables, there is no redshift dependence of the reduced cumulants, that is the normalised cumulants appearing in equation (30) divided by the linear standard deviation σ_0 (Bernardeau et al. 2002). In this Appendix, we propose to derive the required cumulants at tree-order starting from the well-known skewness, before extending this result to the other cumulants that are needed.

A.1. $S_3 = \langle J_1^3 \rangle / \sigma_0$

We note that $J_1 = \nu = \delta / \sigma_0$, meaning that $\langle J_1^3 \rangle = \langle \delta^3 \rangle / \sigma_0^3 = S_3 \sigma_0$, where S_3 is the usual redshift-independent reduced skewness parameter characterising the asymmetry of the probability distribution function. Its computation is a well-established result of Eulerian perturbation theory (Peebles 1980; Juszkiewicz et al. 1993) and we repeat its derivation here, on which we then base the extension to other cumulants.

The core hypothesis of the Eulerian perturbation scheme is to assume that the density field can be expanded at every order as a function of the initial (linear) density so that it can be written as a series of the form $\delta(\mathbf{x}, t) = \sum_n \delta^{(n)}(\mathbf{x}, t)$. At leading order, we thus obtain $\langle \delta^3 \rangle \approx 3 \langle (\delta^{(1)})^2 \delta^{(2)} \rangle$ and where

$$\delta^{(2)}(\mathbf{x}, t) = \int \frac{d^3 \mathbf{k}_1}{(2\pi)^{3/2}} \frac{d^3 \mathbf{k}_2}{(2\pi)^{3/2}} \delta^{(1)}(\mathbf{k}_2, t) \delta^{(1)}(\mathbf{k}_2, t) F_2(\mathbf{k}_1, \mathbf{k}_2) e^{i(\mathbf{k}_1 + \mathbf{k}_2) \cdot \mathbf{x}} \quad (\text{A.1})$$

with

$$F_2(\mathbf{k}_1, \mathbf{k}_2) = \frac{5}{7} + \frac{1}{2} \frac{\mathbf{k}_1 \cdot \mathbf{k}_2}{k_1 k_2} \left(\frac{k_1}{k_2} + \frac{k_2}{k_1} \right) + \frac{2}{7} \frac{(\mathbf{k}_1 \cdot \mathbf{k}_2)^2}{k_1^2 k_2^2}. \quad (\text{A.2})$$

The computation of $\langle \delta^3 \rangle$ therefore leads to the appearance of ensemble averages of the form $\langle \delta^{(1)}(\mathbf{k}_1) \delta^{(1)}(\mathbf{k}_2) \delta^{(1)}(\mathbf{k}_3) \delta^{(1)}(\mathbf{k}_4) \rangle$, which for Gaussian initial conditions can be estimated using Wick's theorem only considering pairs of wave vectors. Taking into account a smoothing window function $W(kR)$ and after some algebra, we obtain

$$\langle \delta^3 \rangle = 6 \int d^3 \mathbf{k}_1 \int d^3 \mathbf{k}_2 P(k_1) P(k_2) F_2(\mathbf{k}_1, \mathbf{k}_2) W(k_1 R) W(k_2 R) W(|\mathbf{k}_1 + \mathbf{k}_2| R), \quad (\text{A.3})$$

with $P(k)$ being the linear power spectrum.

In this paper, we chose to work with Gaussian smoothing so that

$$W(kR) = \exp\left(-\frac{1}{2} k^2 R^2\right), \quad (\text{A.4})$$

$$W(|\mathbf{k}_1 + \mathbf{k}_2| R) = \exp\left(-\frac{1}{2} (k_1^2 + k_2^2) R^2\right) \sum_{l=0}^{\infty} (-1)^l (2l+1) P_l\left(\frac{\mathbf{k}_1 \cdot \mathbf{k}_2}{k_1 k_2}\right) I_{l+1/2}(k_1 k_2 R^2) \sqrt{\frac{\pi}{2 k_1 k_2 R^2}}, \quad (\text{A.5})$$

and R is our smoothing scale, P_l are the Legendre polynomials, and I_α are the modified Bessel functions of the first kind. The angular integration of (A.3) is then performed, also decomposing F_2 on the basis of Legendre polynomials,

$$F_2 = \frac{17}{21} P_0 + \frac{1}{2} \left(\frac{k_1}{k_2} + \frac{k_2}{k_1} \right) P_1 + \frac{4}{21} P_2, \quad (\text{A.6})$$

and using the Legendre polynomial orthogonality relation,

$$\int_{-1}^1 P_m(x) P_n(x) dx = \frac{2}{2m+1} \delta_{m,n}^{\text{Dirac}}. \quad (\text{A.7})$$

Noting that the angular part of (A.3) only depends on the angle between \mathbf{k}_1 and \mathbf{k}_2 , we obtain

$$\langle \delta^3 \rangle = \int \frac{24}{\sqrt{R}} \sqrt{2\pi}^{5/2} k_1^{3/2} k_2^{3/2} P(k_1) P(k_2) e^{R(-k_1^2 - k_2^2)} \times \left(\frac{4 \sqrt{\frac{2}{\pi}} \left(\left(\frac{6}{k_1^2 k_2^2 R^2} + 2 \right) \sinh(k_1 k_2 R) - \frac{6 \cosh(k_1 k_2 R)}{k_1 k_2 R} \right)}{21 \sqrt{k_1 k_2 R}} + \frac{34 \sqrt{\frac{2}{\pi}} \sinh(k_1 k_2 R)}{21 \sqrt{k_1 k_2 R}} - \frac{\left(\frac{k_1}{k_2} + \frac{k_2}{k_1} \right) \left(2 \cosh(k_1 k_2 R) - \frac{2 \sinh(k_1 k_2 R)}{k_1 k_2 R} \right)}{\sqrt{2\pi} \sqrt{k_1 k_2 R}} \right) dk_1 dk_2. \quad (\text{A.8})$$

The final 2D integrations over k_1 and k_2 are then usually obtained numerically for a general power spectrum. However, this step can be performed analytically in the special case of a power-law linear power spectrum $P(k) \propto k^n$, where n is the spectral index. In such a case, it yields

$$\begin{aligned} \langle \delta^3 \rangle / \sigma_0^4 = & -\frac{1}{14n^2\Gamma^2(\frac{3+n}{2})} 3\Gamma^2\left(\frac{1+n}{2}\right) \left[3(16+7n)_2F_1\left(\frac{1+n}{2}, \frac{1+n}{2}, -\frac{1}{2}, \frac{1}{4}\right) \right. \\ & \left. + (-32+n(2+n)(23+28n))_2F_1\left(\frac{1+n}{2}, \frac{1+n}{2}, \frac{1}{2}, \frac{1}{4}\right) - 7n(1+n)(3+2n)_2F_1\left(\frac{1+n}{2}, \frac{3+n}{2}, \frac{1}{2}, \frac{1}{4}\right) \right], \end{aligned} \quad (\text{A.9})$$

where ${}_2F_1$ is the hypergeometric function and Γ the gamma function. This result is equivalent to the result obtained originally by [Lokas et al. \(1995\)](#) (see their equation (38)).

For the more general case of a generic linear matter power spectrum, we resort to numerical integration of equation (A.8). For the results shown in this paper (e.g. in figure 4), we compute the linear power spectrum with CAMB and for the cosmological parameters used in the Quijote simulation.

A.2. $U_3 = \langle J_1 J_2 \rangle / \sigma_0$

We now compute $U_3 = \langle J_1 J_2 \rangle / \sigma_0$. With the previous notations, we notice that $\langle J_1 J_2 \rangle = \langle \nu(\nu^2 - 3I_2) \rangle = \langle \nu^3 \rangle - 3\langle \nu I_2 \rangle$, where the first term is the previously computed skewness. This leaves only $\langle \nu I_2 \rangle = \langle I_1 I_2 \rangle$ to be computed. Thanks to isotropy, this latter can be written as

$$\begin{aligned} \langle I_1 I_2 \rangle = & 3\langle \delta\phi_{11}\phi_{22} \rangle - 3\langle \delta\phi_{12}^2 \rangle \\ = & 3\langle (\phi_{11} + \phi_{22} + \phi_{33})\phi_{11}\phi_{22} \rangle - 3\langle (\phi_{11} + \phi_{22} + \phi_{33})\phi_{12}^2 \rangle \\ = & 3(2\langle \phi_{11}^2\phi_{22} \rangle + \langle \phi_{11}\phi_{22}\phi_{33} \rangle - 2\langle \phi_{11}\phi_{12}^2 \rangle - \langle \phi_{33}\phi_{12}^2 \rangle). \end{aligned} \quad (\text{A.10})$$

At leading order in standard perturbation theory (SPT), this reads

$$\langle I_1 I_2 \rangle = 3\left(2\langle \phi_{11}^{(1)2}\phi_{22}^{(2)} \rangle + 4\langle \phi_{11}^{(2)}\phi_{11}^{(1)}\phi_{22}^{(1)} \rangle\right) + 3\langle \phi_{11}^{(2)}\phi_{22}^{(1)}\phi_{33}^{(1)} \rangle - 2\left(2\langle \phi_{11}^{(1)}\phi_{12}^{(1)}\phi_{12}^{(2)} \rangle + \langle \phi_{11}^{(2)}\phi_{12}^{(1)2} \rangle\right) - \left(2\langle \phi_{33}^{(1)}\phi_{12}^{(1)}\phi_{12}^{(2)} \rangle + \langle \phi_{33}^{(2)}\phi_{12}^{(1)2} \rangle\right). \quad (\text{A.11})$$

The computation of each of the above ensemble averages can be performed in a similar manner to the case of the skewness. Indeed, using a Fourier representation, the Poisson equation, and Wick's theorem, one has to integrate terms of the generic form

$$2 \int d^3\mathbf{k}_1 \int d^3\mathbf{k}_2 P(k_1)P(k_2)F_2(\mathbf{k}_1, \mathbf{k}_2) \frac{K}{(\mathbf{k}_1 + \mathbf{k}_2)^2} W(k_1 R_1)W(k_2 R_1)W(|\mathbf{k}_1 + \mathbf{k}_2| R_2), \quad (\text{A.12})$$

where (i) the easiest way to proceed is to distinguish between two scales R_1 and R_2 as an intermediate step while in the end they will both be taken equal to our smoothing scale R and (ii) K is a polynomial in \mathbf{k}_1 and \mathbf{k}_2 , which depends on which second derivatives of the gravitational potential are considered.

The term $1/(\mathbf{k}_1 + \mathbf{k}_2)^2$ unfortunately prevents direct analytical integration of the angular part of the previous form but differentiation by R_2^2 under the integral sign (sometimes referred to as *Feynman's trick*) turns out to be a viable solution. Applied to the first term of equation (A.11), for example, we obtain

$$\begin{aligned} \frac{\partial \langle \phi_{11}^2\phi_{22} \rangle}{\partial R_2^2} = & -\frac{1}{2}(\mathbf{k}_1 + \mathbf{k}_2)^2 \langle \phi_{11}^2\phi_{22} \rangle \\ = & -\int d^3\mathbf{k}_1 \int d^3\mathbf{k}_2 P(k_1)P(k_2)F_2(\mathbf{k}_1, \mathbf{k}_2) \frac{k_{1(1)}^2 \left(k_{2(1)}^2 (k_1 + k_2)_{(2)}^2 + 2k_{2(2)}^2 (k_1 + k_2)_{(1)}^2 \right)^2}{k_1^2 k_2^2} W(k_1 R_1)W(k_2 R_1)W(|\mathbf{k}_1 + \mathbf{k}_2| R_2), \end{aligned} \quad (\text{A.13})$$

where $k_{i(j)}$ is the j -th Cartesian component of \mathbf{k}_i for $i \in \{1, 2\}$. This form is much more suited to integration. Using the same decomposition of the Gaussian filter and the perturbation theory kernel into the basis of Legendre polynomials as in the skewness computation, we perform integration on the angle between \mathbf{k}_1 and \mathbf{k}_2 , and a final integration on R_2^2 . We finally obtain

$$\begin{aligned} \langle I_1 I_2 \rangle = & \int \frac{1}{70R^4(k_1 k_2 R^2)^{5/2}} \pi^2 P(k_1)P(k_2) \sqrt{\frac{1}{k_1 k_2 R^2}} e^{-\frac{1}{2}R^2(k_1^2 + k_2^2)} \left(-5R^{10}(k_1^2 - k_2^2)^4(6k_1^2 - k_2^2) \left(\text{Ei}\left(-\frac{1}{2}(k_1 - k_2)^2 R^2\right) - \text{Ei}\left(-\frac{1}{2}(k_1 + k_2)^2 R^2\right) \right) \right. \\ & + e^{-\frac{1}{2}R^2(k_1^2 + k_2^2)} \left(-8k_1 k_2 R^2 \left(R^2(5R^4(k_1^2 - k_2^2)^2(6k_1^2 - k_2^2) - 708k_1^2 - 20R^2(6k_1^4 + 19k_1^2 k_2^2 - k_2^4) - 372k_2^2) - 960 \right) \cosh(k_1 k_2 R^2) \right. \\ & \left. - 7680 \sinh(k_1 k_2 R^2) \right) - 4R^2 e^{-\frac{1}{2}R^2(k_1^2 + k_2^2)} \left(5R^6(k_1 - k_2)^2(k_1 + k_2)^2(6k_1^2 - k_2^2)(k_1^2 + k_2^2) + 1416k_1^2 + 40R^2(6k_1^4 + 35k_1^2 k_2^2 - k_2^4) \right. \\ & \left. + 2R^4(-30k_1^6 + 301k_1^4 k_2^2 + 84k_1^2 k_2^4 + 5k_2^6) + 744k_2^2 \right) \sinh(k_1 k_2 R^2) \left. \right) dk_1 dk_2, \end{aligned} \quad (\text{A.14})$$

where Ei is the exponential integral function $\text{Ei}(z) = -\int_z^\infty e^{-t}/tdt$.

Finally, we perform a numerical integration of equation (A.14) with the Quijote linear power spectrum as for the skewness. Our results are displayed in figure 4.

A.3. $V_3 = \langle J_3 \rangle / \sigma_0$

We now compute $V_3 = \langle J_3 \rangle / \sigma_0$. With the previous notations we note that $\langle J_3 \rangle = \langle \nu^3 \rangle - 9\langle I_1 I_2 \rangle / 2 + 27\langle I_3 \rangle / 2$, where the first two terms are computed as part of S_3 and U_3 in the above subsections. The third term can then be written as

$$\langle I_3 \rangle = \langle \phi_{11} \phi_{22} \phi_{33} \rangle + 2\langle \phi_{12} \phi_{23} \phi_{13} \rangle - 3\langle \phi_{11} \phi_{23}^2 \rangle, \quad (\text{A.15})$$

which, using isotropy and at leading order in SPT, can be rewritten as

$$\langle I_3 \rangle = 3\langle \phi_{11}^{(2)} \phi_{22}^{(1)} \phi_{33}^{(1)} \rangle + 6\langle \phi_{12}^{(2)} \phi_{23}^{(1)} \phi_{13}^{(1)} \rangle - 3\left(2\langle \phi_{11}^{(1)} \phi_{23}^{(1)} \phi_{23}^{(2)} \rangle + \langle \phi_{11}^{(2)} (\phi_{23}^{(1)})^2 \rangle\right). \quad (\text{A.16})$$

The exact same steps as in the computation of $\langle I_1 I_2 \rangle$ are then performed to evaluate every ensemble average appearing in the previous equation (A.16), most notably the differentiation under the integral sign and the decomposition in Legendre polynomials of the Gaussian and perturbation theory kernels. We obtain

$$\begin{aligned} \langle I_3 \rangle = & \int \frac{1}{140(k_1 k_2 R^2)^{7/2}} \pi^2 k_1 k_2 P(k_1) P(k_2) (k_1 - k_2)(k_1 + k_2) \sqrt{\frac{1}{k_1 k_2 R^2}} e^{-\frac{1}{2}R^2(k_1^2 + k_2^2)} \left(-5R^8(k_1^2 - k_2^2)^4 \text{Ei}\left(-\frac{1}{2}(k_1 - k_2)^2 R^2\right) \right. \\ & + 5R^8(k_1^2 - k_2^2)^4 \text{Ei}\left(-\frac{1}{2}(k_1 + k_2)^2 R^2\right) + e^{-\frac{1}{2}R^2(k_1^2 + k_2^2)} \left(8k_1 k_2 R^2 \left(-5R^4(k_1^2 - k_2^2)^2 + 20R^2(k_1^2 + k_2^2) + 48 \right) \cosh(k_1 k_2 R^2) \right. \\ & \left. \left. - 4\left(5R^6(k_1^2 - k_2^2)^2(k_1^2 + k_2^2) + 40R^2(k_1^2 + k_2^2) - 2R^4(5k_1^4 - 26k_1^2 k_2^2 + 5k_2^4) + 96 \right) \sinh(k_1 k_2 R^2) \right) \right) dk_1 dk_2, \quad (\text{A.17}) \end{aligned}$$

where Ei is the exponential integral function $\text{Ei}(z) = -\int_z^\infty e^{-t}/t dt$.

Again, we perform a numerical integration of equation (A.17) with the Quijote linear power spectrum as for the skewness and $\langle I_1 I_2 \rangle$. Our results are displayed in figure 4.

Appendix B: Behaviour of the Gram-Charlier

The NLO prediction for the cosmic web abundances are derived in equation (33), which involves three functions of the threshold per environment:

$$P_{(E)}(J_1, J_2, J_3) = P_{G,(E)}(J_1, J_2, J_3) + \sigma_0 [s_{(E)}(\lambda_{th}) S_3 + u_{(E)}(\lambda_{th}) U_3 + v_{(E)}(\lambda_{th}) V_3]. \quad (\text{B.1})$$

For the sake of completeness, in figure B.1 we show the behaviour of these functions for the voids, walls, filaments, and knots in blue, green, yellow, and red, respectively, and as a function of the chosen threshold. The threshold chosen in our main analysis is the dashed black vertical line. We note that for a threshold of between -2 and 2 , the s , u , and v functions fluctuate a lot, thus potentially strongly modulating the environment probabilities depending on the values of the cumulants—that is depending on how non-Gaussian the field is—and most importantly on the value of the chosen threshold. This emphasises the importance of the choice of the threshold value.

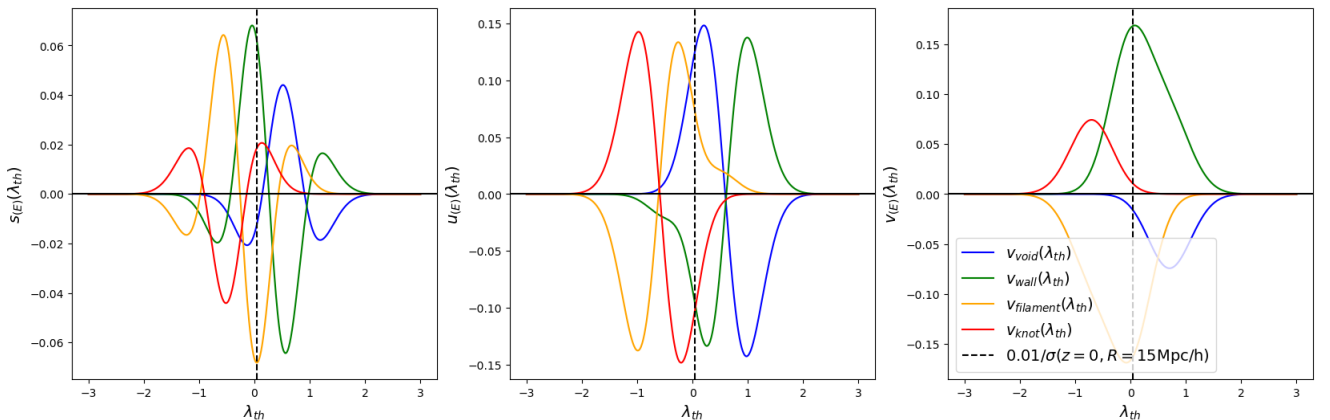


Fig. B.1. $s_{(E)}(\lambda_{th})$, $u_{(E)}(\lambda_{th})$, and $v_{(E)}(\lambda_{th})$ as a function of threshold predicted by the Gram-Charlier development. Results are shown for the four different environments: voids in blue, walls in green, filaments in orange, and knots in red. The vertical black line is the threshold used in the paper.

CHAPTER 6

PEAK COUNTS STATISTICS IN WEAK LENSING

6.1	Overview	110
6.2	"UNIONS: The impact of systematic errors on weak-lensing peak counts" (article)	110

In Chapters 4 and 5, we have explored how the tSZ effect and the cosmic web provide insights into the large-scale structures of the Universe. The tSZ effect, sensitive primarily to highest-mass halos, and the study of the cosmic web allow to investigate non-linearities in the Universe. In this chapter, I will investigate another method to probe these non-linearities: peak count statistics in weak lensing. I will begin with an overview in Sect. 6.1, then I will describe and copy the article, of which I am the primary author, focusing on the impact of systematics on cosmological parameters when using peak count statistics in Sect. 6.2.

6.1 Overview

Gravitational lensing, particularly weak lensing, was discussed in Sect. 2.2.3. It is a powerful tool for probing the large-scale structures of the Universe, as it is sensitive to the total matter content, including dark matter. In this regard, weak lensing is complementary to the tSZ effect, which is only sensitive to baryonic matter. This complementarity is why cross-correlation between weak-lensing and tSZ is a powerful method for investigating the large-scale structures of the Universe (e.g., Tröster et al., 2022; Van Waerbeke et al., 2014).

While weak-lensing has been widely used in cosmological analyses through methods such as the two-point correlation function and power spectrum, I specifically focus on its application in higher-order statistics, such as peak counts. Peak counts involve identifying local maxima in a field and are sensitive to non-Gaussianities within that field. For example, studies have demonstrated that weak lensing peak counts can be employed to constrain parameters like neutrino masses (e.g., Ajani et al., 2020) or modified gravity theories (e.g., Davies et al., 2024). It has also been shown that it can constrain cosmological parameters more robustly than traditional two-point statistics (e.g., Euclid Collaboration et al., 2023).

6.2 "UNIONS: The impact of systematic errors on weak-lensing peak counts" (article)

In this section, I will briefly describe the method and main result of the article (Ayçoberry et al., 2023) and attach it at the end of the section.

Method In this article, we present a study on the impact of different uncertainties and systematics on cosmological parameters when using peak count statistics in weak lensing. We perform a qualitative analysis to assess the influence of different systematics and identify the dominant ones.

To perform our analysis, we use the weak-lensing data from the ongoing UNIONS¹ survey. Our predictions are derived using the MassiveNuS simulations (Liu et al., 2018), with peak counts computed following the methodology of Ajani et al. (2020). The peak counts are determined on signal-to-noise ratio (SNR) maps, defined as the noisy convergence map smoothed with a Gaussian filter over the standard deviation of the noise. The peak count distribution is then obtained by counting the number of pixels with values higher than their neighbouring pixels, and the counts are used in our parameter inference.

Another important aspect of our work is the calibration of the shear. Our calibration approach is based on the "metacalibration" method (Huff & Mandelbaum, 2017), which involves measuring the response matrix \mathbf{R} of a shape measurement algorithm. The relation between the true and estimated shear (g_j^{true} and g_i^{obs}) is:

$$g_i^{\text{obs}} = \sum_{j=1}^2 R_{ij} g_j^{\text{true}} + c_i, \quad (6.1)$$

where i and j are the components of the shear, and c_i is the additive bias. The response matrix R_{ij} can be described as $R_{ij} = \partial g_i^{\text{obs}} / \partial g_j^{\text{true}}$, and $\text{tr}(\mathbf{R}) = 2(1 + m)$ where m is the multiplicative bias. The response matrix is composed of the shear response matrix and selection response matrix ($\mathbf{R} = \mathbf{R}^{\text{shear}} + \mathbf{R}^{\text{selection}}$). Calibration is typically performed at the catalogue level, but we will investigate the impact of a local calibration.

We consider various systematics in our analysis:

- Local shear calibration: we perform a local calibration by applying the metacalibration method on subsamples of the catalogue of a few degrees (ranging from 0.5 deg² to 4 deg²). We investigate the impact on the response matrix and additive bias. This method can account for local effects of the point spread function, galaxy size, etc.

¹<https://www.skysurvey.cc/>

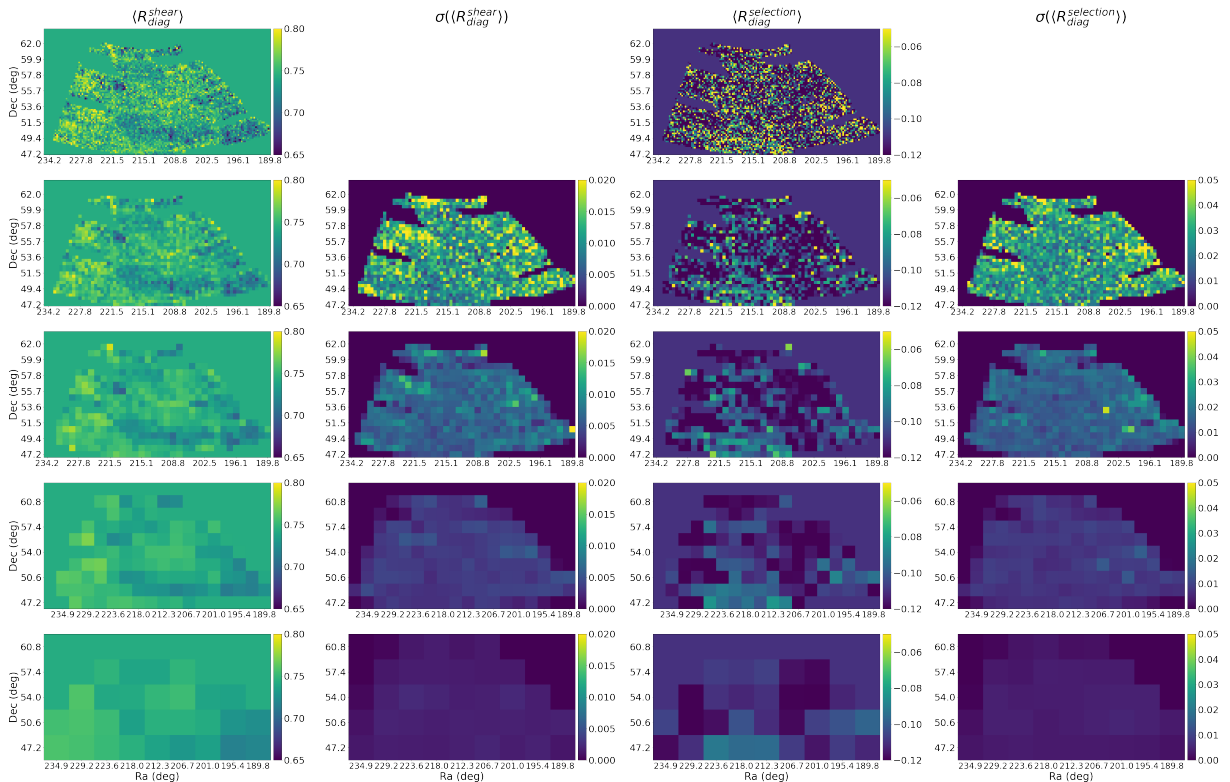


Figure 6.1: $\langle R_{diag} \rangle$ depending on the size of the local calibration. From top to bottom, calibration is done on $0.25, 0.5, 1, 2, 4 \text{ deg}^2$. The four columns are $\langle R_{diag}^{shear} \rangle$, its standard deviation, $\langle R_{diag}^{selection} \rangle$ and its standard deviation.

- Residual multiplicative shear bias: as the shear calibration is not perfect, we account for a residual bias. We employ a residual bias $\Delta m = 0.007$ as reported by Guinot et al. (2022).
- Redshift uncertainties: the mean redshift of the survey ($\bar{z} = 0.65$) has a level of uncertainty. We investigate the impact of a redshift uncertainty of $\Delta z = 0.03$.
- Baryonic feedback: the MassiveNuS simulations are dark matter only simulations, while baryonic feedback affects the large-scale structures probed by the UNIONS survey. To mitigate this effect, we use the results of Coulton et al. (2020) to correct our predictions, applying three strengths of feedback.
- Intrinsic alignment and boost factor: these two effects can bias the weak-lensing signal. According to Harnois-Déraps et al. (2021), the effect is small for peaks with $\text{SNR} < 3$ (< 4) for intrinsic alignment (boost factor). We thus employ a restrictive cut on $\text{SNR} < 3$ (instead of going up to $\text{SNR} = 6$) to mitigate these biases.

Results The first main result focuses on the local shear calibration. We performed a local calibration on sub-patches of $0.5, 1, 2, 4 \text{ deg}^2$, allowing us to obtain the shear matrix, response matrix, and additive bias on these sizes. In Fig. 6.1, we present the spatial variation of the average of the diagonal of the response matrix $\langle R_{diag}^{shear} \rangle$ and its standard deviation, as well as the average of the diagonal of the selection response matrix $\langle R_{diag}^{selection} \rangle$ and its standard deviation. From top to bottom, the quantities are computed on sub-patches of $0.25, 0.5, 1, 2, 4 \text{ deg}^2$. The 0.25 deg^2 sub-patches are only used to produce the standard deviation of the 0.5 deg^2 calibration.

We observe that the larger sub-patches result in a calibration closer to the global one, with reduced variations. The standard deviation of the shear matrix remains small, and the standard deviation of the selection matrix is small for a calibration performed on sub-patches of at least 1 deg^2 . From this analysis, we conclude that local calibration can capture local effects and improve the calibration, but it should not be done at too small scale.

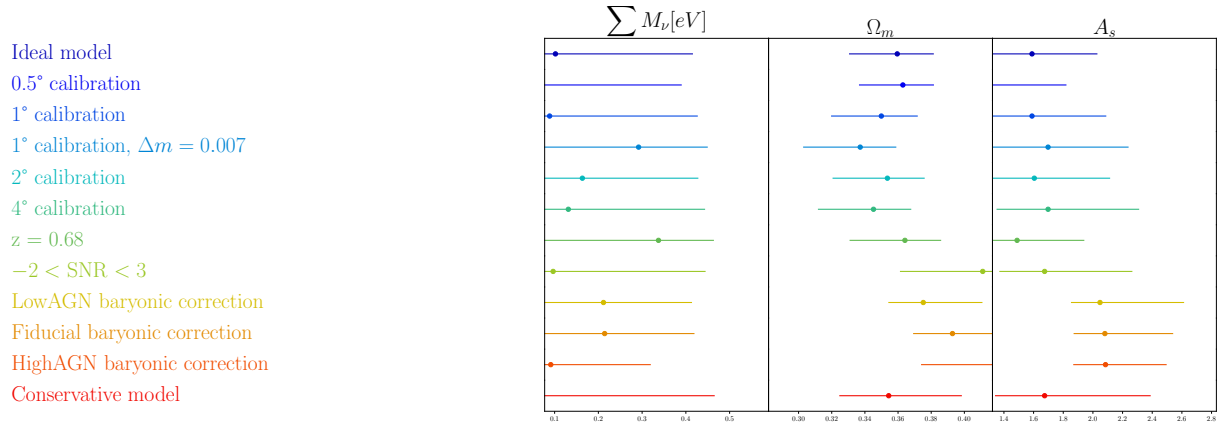


Figure 6.2: Marginalised distributions for the 68% confidence interval for the three cosmological parameters for the different cases, including uncertainties and systematic effects.









We then focus on the impact of the different uncertainties and systematics on the constraining power of cosmological parameters. The **MassiveNuS** simulations allow us to probe M_ν , Ω_m , and A_s . This study aims to identify trends and obtain qualitative results on the uncertainties, rather than focusing on the precise values of the parameters.

The results for the different uncertainties and systematics are summarised in Fig. 6.2, where we show the marginalized distribution for the 68% confidence interval. The "ideal case" represents an unrealistic best-case scenario where the data are perfectly known, with no uncertainties or systematics considered. The "conservative model" includes different mitigation schemes: no redshift uncertainty, local calibration on 1 deg^2 , a residual multiplicative shear bias $\Delta m = 0.007$, correction of the simulation predictions with a fiducial baryonic correction, and a cut on the $\text{SNR} < 3$ to mitigate intrinsic alignment and boost factor.

The mass of the neutrino is not well constrained, and the lower bound of A_s reaches the prior in most cases. We thus focus on the constraining power and shift of Ω_m . Local calibration and the addition of the multiplicative shear bias can shift Ω_m by -0.5σ . Baryonic correction results in a shift of $+0.5\sigma$. Reducing the SNR range leads to larger constraints, reducing the constraining power. Finally, the redshift uncertainty has a small impact on the constraining power, shifting Ω_m by only $+0.02\sigma$.

This analysis emphasizes the need for careful shear calibration and the importance of accounting for the different systematics to obtain robust constraints. For future analyses, it will be crucial to determine the smallest calibration size that can be trusted. Additionally, it will be necessary to use hydrodynamical simulations and properly account for intrinsic alignment and boost factor to avoid losing constraining power.

UNIONS: The impact of systematic errors on weak-lensing peak counts[★]

Emma Ayçoberry^{1,2,3}, Virginia Ajani^{2,4}, Axel Guinot⁵, Martin Kilbinger², Valeria Pettorino², Samuel Farrens², Jean-Luc Starck², Raphaël Gavazzi^{3,6}, and Michael J. Hudson^{7,8,9}

¹ Observatoire de Paris, Université PSL, 61 avenue de l'Observatoire, 75014 Paris, France
e-mail: emma.aycoberry@iap.fr

² Université Paris-Saclay, Université Paris-Cité, CEA, CNRS, Astrophysique, Instrumentation et Modélisation Paris-Saclay, 91191 Gif-sur-Yvette, France

³ Institut d'Astrophysique de Paris, UMR 7095, CNRS & Sorbonne Université, 98 bis boulevard Arago, 75014 Paris, France

⁴ Institute for Particle Physics and Astrophysics, Department of Physics, ETH Zürich, Wolfgang Pauli Strasse 27, 8093 Zürich, Switzerland

⁵ Université de Paris, CNRS, Astroparticule et Cosmologie, 75013 Paris, France

⁶ Laboratoire d'Astrophysique de Marseille, Aix-Marseille Univ., CNRS, CNES, 13388 Marseille Cedex 13, France

⁷ Department of Physics and Astronomy, University of Waterloo, Waterloo, ON N2L 3G1, Canada

⁸ Waterloo Centre for Astrophysics, Waterloo, ON N2L 3G1, Canada

⁹ Perimeter Institute for Theoretical Physics, 31 Caroline St. N., Waterloo, ON N2L 2Y5, Canada

Received 29 April 2022 / Accepted 28 October 2022

ABSTRACT

Context. The Ultraviolet Near-Infrared Optical Northern Survey (UNIONS) is an ongoing deep photometric multiband survey of the northern sky. As part of UNIONS, the Canada-France Imaging Survey (CFIS) provides *r*-band data, which we use to study weak-lensing peak counts for cosmological inference.

Aims. We assess systematic effects for weak-lensing peak counts and their impact on cosmological parameters for the UNIONS survey. In particular, we present results on local calibration, metacalibration shear bias, baryonic feedback, the source galaxy redshift estimate, intrinsic alignment, and cluster member dilution.

Methods. For each uncertainty and systematic effect, we describe our mitigation scheme and the impact on cosmological parameter constraints. We obtain constraints on cosmological parameters from Monte Carlo Markov chains using CFIS data and MassiveNuS N-body simulations as a model for peak counts statistics.

Results. Depending on the calibration (local versus global, and the inclusion or not of the residual multiplicative shear bias), the mean matter density parameter, Ω_m , can shift by up to -0.024 (-0.5σ). We also see that including baryonic corrections can shift Ω_m by $+0.027$ ($+0.5\sigma$) with respect to the dark-matter-only simulations. Reducing the impact of the intrinsic alignment and cluster member dilution through signal-to-noise cuts leads to larger constraints. Finally, with a mean redshift uncertainty of $\Delta\bar{z} = 0.03$, we see that the shift in Ω_m ($+0.001$, which corresponds to $+0.02\sigma$) is not significant.

Conclusions. This paper investigates, for the first time with UNIONS weak-lensing data and peak counts, the impact of systematic effects. The value of Ω_m is the most impacted and can shift by up to ~ 0.03 , which corresponds to 0.5σ depending on the choices for each systematics. We expect constraints to become more reliable with future (larger) data catalogs, for which the current pipeline will provide a starting point. The code used to obtain the results is available on GitHub.

Key words. large-scale structure of Universe – gravitational lensing: weak – methods: data analysis

1. Introduction

Weak gravitational lensing has been used as a cosmological probe in recent years with great success, for example with Dark Energy Survey¹ (DES), Kilo-Degree Survey² (KiDS), Hyper Suprime-Cam³ (HSC), and Canada France Hawaii Lensing Survey⁴ (CFHTLenS). It corresponds to the small distortions we observe in the images of background sources (such as high-redshift galaxies) due to the deflection of photons as they pass

through tidal fields in the large-scale structure (LSS) in the Universe (Bartelmann & Schneider 2001).

Second-order statistics of weak lensing, such as the two-point correlation function or the power spectrum, only capture the Gaussian part of the LSS (Weinberg et al. 2013). Its non-Gaussian part, which is induced by the nonlinear evolution of structure on small scales and low redshifts, contains, however, a wealth of information about cosmology. Several higher-order statistics, such as Minkowski functionals (Kratochvil et al. 2012; Parroni et al. 2020), higher-order moments (for example, Petri et al. 2016; Gatti et al. 2020), the bispectrum (Takada & Jain 2004; Coulton et al. 2019), peak counts (Kruse & Schneider 1999; Dietrich & Hartlap 2010; Liu et al. 2015b; Lin & Kilbinger 2015; Peel et al. 2017; Martinet et al. 2017; Li et al. 2019; Ajani et al. 2020, and references therein), the starlet ℓ_1 norm (Ajani et al. 2021), the

[★] <https://github.com/CosmoStat/shear-pipe-peaks.git>

¹ <https://www.darkenergysurvey.org/>

² <https://kids.strw.leidenuniv.nl/>

³ <https://www.naoj.org/Projects/HSC/>

⁴ <https://www.cfhtlens.org/>

scattering transform (Cheng et al. 2020), wavelet phase harmonic statistics (Allys et al. 2020), and machine learning-based methods (Fluri et al. 2018, 2022; Shirasaki et al. 2021, among others), have been introduced to account for non-Gaussian information.

In this work we choose to focus on peak counts extracted from real data, using the pipeline developed to perform the study presented in Ajani et al. (2020), which had previously only been tested on simulations. Peaks in weak-lensing convergence maps are tracers of overdense regions. They are the local maxima defined as a pixel that is larger than all eight of its neighbors. The peak function – that is, the number of peaks as a function of peak height (in a convergence map) or signal-to-noise ratio (S/N) – depends on the nonlinear and non-Gaussian part of the LSS. This higher-order weak-lensing statistic can be used to constrain cosmological parameters. Peak counts are complementary to second-order shear statistics (Jain & Waerbeke 2000), and by combining both parameters degeneracies can be removed (Dietrich & Hartlap 2010).

Weak-lensing peaks are an indirect tracer of dark-matter halos: large peaks are strongly correlated with massive halos, whereas lower-amplitude peaks are generally created by multiple smaller halos along the line of sight (Yang et al. 2011). Low-amplitude peaks can also be caused by mass outside dark-matter halos, or by galaxy shape noise (Liu & Haiman 2016; Yang et al. 2011).

Explicit expressions or complete theoretical predictions of peak counts are still an active area of research. It is, however, possible to generate weak-lensing simulations densely sampled in cosmological parameter space in order to interpolate them and use the interpolation as a prediction. The advantage of simulations is the possibility to incorporate the exact survey mask and shape noise. For example, Dietrich & Hartlap (2010) created a set of N -body simulations in the (Ω_m, σ_8) plane for 158 cosmologies, and Liu et al. (2015b) and Kacprzak et al. (2016) used ray-tracing N -body simulations in the (Ω_m, σ_8, w) plane for 91 cosmologies. Here we use the MassiveNuS simulations (Liu et al. 2018) to predict peak counts as in Li et al. (2019) and Ajani et al. (2020). These simulations are described in more detail in Sect. 3.

The first cosmological constraints from peak counts were obtained on real data by Liu et al. (2015a) using the Canada-France-Hawaii Telescope (CFHT) Stripe 82 Survey, and by Liu et al. (2015b), using CFHTLenS data. DES Science Verification data have been analyzed by Kacprzak et al. (2016). KiDS 450 deg² data have been studied by Shan et al. (2017) and Martinet et al. (2017). The first tomographic analysis was performed by Harnois-Déraps et al. (2021b) for the DES-Y1 data release. Recently, Zürcher et al. (2022) analyzed the DES-Y3 data release using an emulator approach. These analyses all use peak counts and complement second-order statistics analyses, constraining cosmological parameters.

Weak-lensing observables have to be corrected for systematic effects, which can have an observational and astrophysical origin and can induce biases into the cosmological constraints if not properly taken into account. These artifacts can easily be created by the atmosphere, the telescope, and the detector, and during the data analysis. Astrophysical correlations such as intrinsic galaxy alignments add to the lensing correlations in a nontrivial way. Furthermore, to be able to interpret weak-lensing observables in a cosmological context, the physics of small scales needs to be reliably estimated.

Extensive studies of many of these systematics for second-order statistics exist, for example Massey et al. (2012),

Troxel & Ishak (2015), Hildebrandt et al. (2016), and Mandelbaum (2018). For current and future surveys, precise requirements on the instrument and data processing are routinely derived for those statistics. However, how to properly include systematic effects in the context of higher-order statistics is an ongoing research topic (Kacprzak et al. 2016; Coulton et al. 2020; Zürcher et al. 2021; Harnois-Déraps et al. 2021a; Pyne & Joachimi 2021).

This paper is the first to study several weak-lensing systematics and uncertainty, their effect on peak counts, and the resulting constraints on cosmological parameters from the Ultraviolet Near-Infrared Optical Northern Survey (UNIONS)/Canada-France Imaging Survey (CFIS) galaxy survey. In addition, we develop a novel method for locally calibrating the measured shear, including shear and selection biases. We investigate mitigation schemes for these effects and quantify biases in cosmological parameters.

This paper is organized as follows: In Sects. 2 and 3 we describe the data and simulations used in this work, respectively. In Sect. 4 we describe the measurement of weak-lensing peak counts. In Sect. 5 we introduce local shear calibration and compare it to the standard global calibration. In Sect. 6 we present results for each of the studied biases and uncertainties and discuss mitigation methods. Finally, we draw our conclusions in Sect. 7.

2. Data

2.1. Image processing

We used weak-lensing data from UNIONS⁵. This ongoing survey will provide 4800 deg² of multiband photometric images in the Northern Hemisphere. Founded in 2018, UNIONS is a collaboration of several groups and surveys, each providing data in different bands. These participating surveys are CFIS, the Panoramic Survey Telescope And Rapid Response System (Pan-STARRS), Wide Imaging with Subaru HSC of the Euclid Sky (WISHES), and the Waterloo Hawai'i IfA G -band Survey (WHIGS).

For the CFIS part of the survey r - and u -band images are taken at the CFHT with MegaCAM, a wide-field optical imaging facility. The CFIS r -band images have a median seeing of 0.65 arcsec reflecting the extremely stable atmosphere at Manua Kea together with the excellent CFHT optical system. Each observed sky location is covered by at least three single exposures, which are dithered by one-third of the focal plane, or 0.33°. The exposure time varies between 100 and 300 s, where smaller exposure times are chosen in better observing conditions. This survey strategy provides images of a very homogeneous depth. In this work, we only use r -band data from CFIS for our weak-lensing peak count study. In particular, the weak-lensing data from P3, a patch of size 34.7×17.7 deg² is analyzed. This patch overlaps with CFHTLenS, which is used to infer the redshift distribution (see Sect. 2.3). These data have been processed and validated (see Guinot et al. 2022 for details). The pre-processing of the CFHT data consists of a calibration step with the MegaPipe pipeline (Gwyn 2008). The single exposures are first calibrated before they are combined with SWARP⁶ to build stacked images. MegaPipe provides an astrometric calibration of the survey using *Gaia* Data Release 2 (Gaia Collaboration 2018), and photometric calibration of the r -band data relative to the Pan-STARRS1 survey (Chambers et al. 2016). Both astrometric

⁵ <https://www.skysurvey.cc/>

⁶ <https://github.com/astromatic/swarp>

Table 1. P3-CFIS survey specifications.

Number density of galaxies	n_{gal}	7 arcmin^{-2}
Pixel size	A_{pix}	$0.4^2 \text{ arcmin}^2 \text{ px}^{-2}$
Global ellipticity dispersion	σ_e	0.44
Size of the field		$34.7 \times 17.7 \text{ deg}^2$

and photometric calibration are excellent so we do not expect calibration errors to significantly impact weak-lensing measurements (Guinot et al. 2022).

2.2. The weak-lensing catalog

Since the stacks have a larger S/N than the single-exposure images, we use the former to detect galaxy candidates. Due to the large dithers, the point-spread function (PSF) on the stacks is very inhomogeneous and discontinuous. For this reason, we detect stars and construct the PSF model on single exposures. Galaxy shapes are then obtained using the multi-epoch model-fitting method `ngmix` (Sheldon & Huff 2017)⁷. Those measurements are calibrated with `metacalibration` (Huff & Mandelbaum 2017) to provide shear estimates. The creation and validation of the shear catalog are fully described in Guinot et al. (2022). We note that this is a preliminary version, called “version 0”, of the catalog where the source density is conservative and does not reflect future versions of the CFIS shear data. Specifics of the shear catalog are presented in Table 1.

2.3. Redshift distribution of source galaxies

Multiband data of UNIONS are still sparse, and photometric redshifts have not been obtained yet. Using the r -band data only, redshifts have been obtained for a population of galaxies with two different methods, as follows.

For both methods, CFIS galaxies were matched to the deeper CFHTLenS on the 50 deg^2 W3 field. First, Guinot et al. (2022) approximated the CFIS redshift distribution as the histogram of the best-fit photometric redshifts from CFHTLenS of that matched subsample. Photo- z 's for CFHTLenS had been obtained in Hildebrandt et al. (2012) from u, g, i, r, z multiband data, calibrated with various spectroscopic deep data sets. Second, employing the direct calibration technique, Spitzer et al. (in prep.) re-weighted the Deep Extragalactic Evolutionary Probe 2 (DEEP2) spectroscopic sample (Newman et al. 2013) in a 5D space spanned by the u, g, i, r, z photometric bands, to match the density of the matched subsample in that space. The re-weighted DEEP2 spectroscopic redshift distribution is an estimate of the CFIS $n(z)$. This distribution was fit in Spitzer et al. (in prep.) by an analytical function with two components. The first, exponential, component accounts for the bulk of the distribution, whereas the second, Gaussian, term, models the tail at $z > 2$. However, the addition of this second term somewhat overestimates the re-weighted DEEP2 redshift distribution between $z = 1.5$ and 2.

The mean redshift is obtained as $\bar{z} = \int_0^{z_{\text{max}}} dz z n(z)$, where the integral over the normalized redshift distribution $n(z)$ is carried out nominally up to the limiting redshift of the survey, z_{max} . Both redshift distributions have a significantly non-vanishing probability at high redshifts $z \gtrsim 2$. This is most likely not a physical

Table 2. Mean redshift, \bar{z} , of CFIS weak-lensing galaxies for different redshift distributions, $n(z)$, and maximum redshifts, z_{max} .

$n(z)$	z_{max}	\bar{z}	Comment
Guinot et al. (2022)	2	0.65	Fiducial
Guinot et al. (2022)	∞	0.68	
Spitzer et al. (in prep.) one term	2	0.68	
Spitzer et al. (in prep.) both terms	2	0.68	
Spitzer et al. (in prep.) both terms	∞	0.73	Likely biased

feature, since we do not expect a large number of galaxies in the CFIS sample at those redshifts.

Our best estimate of \bar{z} is obtained by integrating over the CFHTLenS-matched $n(z)$ with a limit of $z_{\text{max}} = 2$, resulting in $\bar{z} = 0.65$. Table 2 presents alternative estimates with varying $n(z)$ and z_{max} . Three further, reasonable combinations yield a slightly higher \bar{z} of 0.68. We use this alternative value in Sect. 6.3 to test the impact of the estimated redshift uncertainty.

3. Simulations

To get the predictions for the summary statistics that we use to perform cosmological inference, we employed the MassiveNuS simulations, a suite of cosmological dark-matter-only N -body simulations that explore different cosmologies including massive neutrinos in the range $\sum m_\nu = 0\text{--}0.62 \text{ eV}$. The simulations have a $512 \text{ Mpc } h^{-1}$ box size with 1024^3 cold dark matter particles. The pixel size is 0.4 arcmin. The implementation is performed using a modified version of the public tree-Particle Mesh (tree-PM) code Gadget2⁸ with a neutrino patch, describing the effect of massive neutrinos on the growth of structures up to $k = 10h \text{ Mpc}^{-1}$. A complete description of the implementation and the products is provided in Liu et al. (2018). The cosmological parameters vary across the simulations within the range $M_\nu \in [0, 0.62]$, $\Omega_m \in [0.18, 0.42]$, and $A_s \in [1.29, 2.91] \times 10^{-9}$. We thus worked on the constraints on these three cosmological parameters, which are well sampled by the simulations. They include the effects of radiation on the background expansion and the impact of massive neutrinos is included with a linear-responses method: neutrinos are evolved perturbatively, while their clustering is caused by the nonlinear dark-matter evolution.

The simulations assume a flat universe with Hubble constant $H_0 = 70 \text{ km s}^{-1} \text{ Mpc}^{-1}$. The primordial power-spectrum scalar index is $n_s = 0.97$, the baryon density $\Omega_b = 0.046$, and the dark-energy equation of state $w = -1$. A fiducial cosmology is set to $[M_\nu, \Omega_m, 10^9 \times A_s] = [0.1, 0.3, 2.1]$. Simulations are available for 101 cosmologies, with 10 000 realizations for each cosmology, obtained by randomly rotating and shifting the lensing potential planes.

Following Ajani et al. (2020), the peak counts were computed for each of the MassiveNuS cosmologies from the simulated convergence maps, averaged over the 10 000 realizations for each model. A model with massless neutrinos corresponding to $[M_\nu, \Omega_m, 10^9 \times A_s] = [0.0, 0.3, 2.1]$ is also provided, and we used it to compute the covariance matrix.

3.1. Effective redshift distribution

For each of the 101 cosmological models, five source redshifts, $z_s = \{0.5, 1.0, 1.5, 2.0, 2.5\}$, are present. To match the simulations

⁷ <https://github.com/esheldon/ngmix>

⁸ <https://wwwmpa.mpa-garching.mpg.de/gadget/>

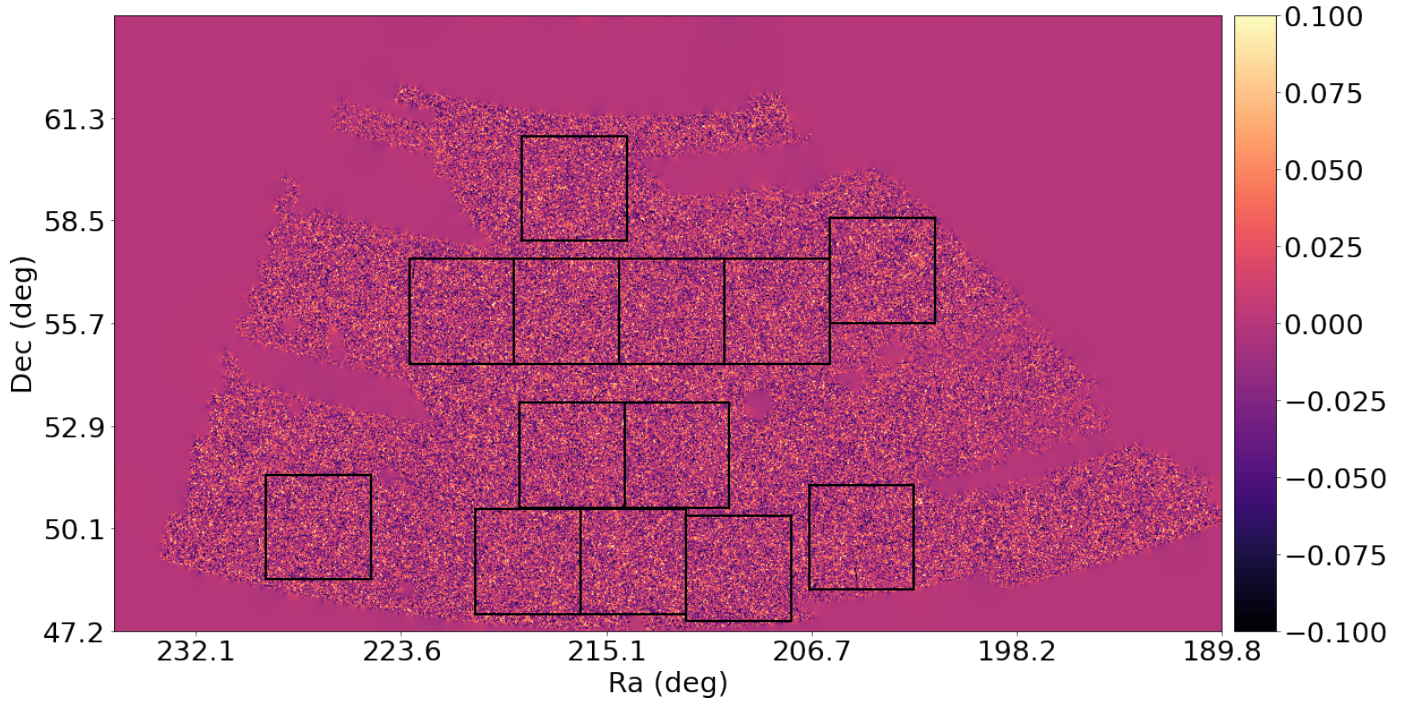


Fig. 1. Convergence map of the CFIS-P3 patch. The squares indicate the regions free of large masks, which were used to compute the peak count. The total peak count is the mean of the peaks over the 13 patches.

to the observed redshift distribution of CFIS, we made the following approximations. We matched the mean redshift between MassiveNus and CFIS (Sect. 2.3) and neglected the shape of the redshift distribution. For that, we interpolated the two convergence maps closest in redshift to \bar{z} and obtained the new effective map as

$$\kappa_{\bar{z}} = \kappa_{z=0.5}\lambda + \kappa_{z=1}(1 - \lambda). \quad (1)$$

This effectively defines the new redshift distribution by a weighted sum of two Dirac delta distributions,

$$n(z) = \lambda\delta_D(z - 0.5) + (1 - \lambda)\delta_D(z - 1). \quad (2)$$

To match our best estimate, $\bar{z} = 0.65$, we set $\lambda = 0.7$. As anticipated in Sect. 2.3, we also employed $\bar{z} = 0.68$ in order to estimate the uncertainty related to the redshift estimation. To do so, we used the relation in Eq. (1) to the new effective maps at $\kappa_{\bar{z}=0.65}$ and $\kappa_{z=1}$ and imposed $\bar{z} = 0.68$. This results in an interpolation parameter $\lambda' = 0.91$.

3.2. Noise

To include the CFIS shape noise in the simulations that we employ to perform inference, we add Gaussian noise. First, we compute the global ellipticity dispersion σ_e^2 . The ellipticity noise per smoothed pixel is then

$$\sigma_{\text{pix}}^2 = \frac{\sigma_e^2}{2n_{\text{gal}}A_{\text{pix}}}, \quad (3)$$

where n_{gal} is the galaxy number density and A_{pix} is the pixel size. For CFIS data, we used the value listed in Table 1. We applied the Gaussian noise to the convergence maps because we do not have direct access to the simulated shear.

4. Weak-lensing peak counts

We employed weak-lensing peak counts as summary statistics for our analysis, measured from the weak-lensing convergence maps.

4.1. Convergence maps

The reduced shear g is the main observable in weak gravitational lensing by galaxies, it is estimated from galaxy ellipticities. We can define it as

$$g = \frac{\gamma}{1 - \kappa}, \quad (4)$$

where γ is the shear and κ the convergence. For the data, the E - and B -mode convergence maps are built through Kaiser-Squires inversion (Kaiser & Squires 1993) from the reduced shear provided by the weak-lensing catalog. The CFIS-P3 map is quite large for a projection but peaks are local, and any distortion will rather affect large scales. In our analysis we worked with the E -mode of the convergence map as the B -mode contain mostly noise (Guinot et al. 2022). Hereafter, when we speak about convergence, we mean the E -mode of the convergence map. The CFIS-P3 convergence map is shown in Fig. 1. Every pixel has a size of 0.4 arcmin, as in the simulations. The 13 black squares are boxes of 512×512 pixels, which correspond to the size of the simulation convergence maps. These boxes are placed such that they do not overlap with larger masked or missing areas.

For the simulations, we employed the already existing maps from the MassiveNus⁹ suite, obtained with the LensTools¹⁰ (Petri 2016) ray-tracing package. We mimicked the CFIS shape noise by adding the noise introduced in Sect. 3.2 to the maps

⁹ <http://columbialensing.org/#massivenus>

¹⁰ <https://github.com/apetri/LensTools/blob/master/docs/source/index.rst>

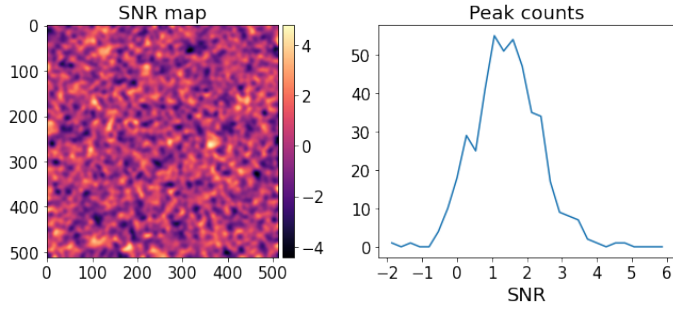


Fig. 2. S/N map and peak counts histogram. The left plot is the S/N map of a part of the 512×512 pixels of CFIS-UNIONS data. The right plot is the peak counts computed on that S/N map.

coming from the simulations, and we smoothed them with a Gaussian kernel of width $\sigma = 2$ arcmin. Li et al. (2019) find that for the same simulations, using peak counts, the optimal smoothing to get tighter constraints is 2 arcmin. As a first work, we chose this smoothing scale to be consistent with the simulations. Moreover, Hildebrandt et al. (2016) find that a smoothing scale of 2 arcmin is a good choice for KiDS data, which have the same number of galaxies per pixel as us. For future work, it can be interesting to adapt simulations to the data to determine the optimal smoothing scale for future releases. We thus computed signal-to-noise ratio (S/N) maps, where the S/N is defined as the noisy convergence map, smoothed with a Gaussian filter over the standard deviation of the noise defined in Eq. (3). Then we computed the peak counts with the `lenspack`¹¹ python package on the S/N maps collecting the local maxima, namely computing the pixels with higher values with respect to their neighboring pixels. In our analysis, we consider linearly spaced bins in the range $S/N = [-2, 6]$. The peak counts distribution used for parameter inference (Sect. 6) corresponds to the mean of the peak counts from the 13 patches. This corresponds to an area of $13 \times 12.25 \text{ deg}^2 \approx 160 \text{ deg}^2$. Figure 2 presents the S/N map and the peak counts histogram from one patch of the CFIS-P3 data.

4.2. Modeling and parameter inference

We modeled the peak function with the MassiveNuS N -body and ray-tracing simulations (Liu et al. 2018). In these simulations the matter density parameter, Ω_m , the primordial power-spectrum normalization amplitude, A_s , and the total mass of neutrinos, $M_\nu = \sum m_\nu$, were varied (see Sect. 3 for more details). We constructed a likelihood function as follows. First, the theoretical model was obtained by numerically computing the peak function for each simulated parameter combination. These functions were then interpolated to arbitrary parameters within the parameter boundaries using a Gaussian process. The error of prediction was always below the CFIS statistical error and on the order of a few per cent. The covariance matrix was computed from the variations of realizations at the fiducial model and is thus assumed to be parameter-independent. As explained in the previous section, for the data, the peaks were computed as the mean of the peaks on 13 mask-free patches; thus, the re-scaling factor of the covariance when we infer parameters from the data is 1/13. The likelihood was taken as a multivariate Gaussian as a function of the data vector. As prior on the parameters, we used a flat prior for the three parameters: $M_\nu \in [0.06, 0.62]$, $\Omega_m \in [0.18, 0.42]$ and $A_s \in [1.29, 2.91] \times 10^9$. The prior of Ω_m and A_s

are the bounds of the simulation. For M_ν , 0.06 is the minimum from oscillation experiments (Particle Data Group 2016) and 0.62 is the upper bound of the simulations. We explored the posterior distribution with Monte Carlo Markov chain (MCMC) sampling using the python package `emcee`. Specifically, we employed 250 walkers initialized in a tiny Gaussian ball of radius 10^{-3} around the fiducial cosmology $[M_\nu, \Omega_m, 10^9 \times A_s] = [0.1, 0.3, 2.1]$ and estimated the posteriors using 6500 sampling steps and 200 burn-in steps.

5. Local shear calibration

In this section, we describe and measure a spatially varying calibration of the estimated shear. For the multiplicative shear bias, including both shear and selection biases, we use the technique of “metacalibration” (Huff & Mandelbaum 2017), which we briefly describe in the following subsection.

5.1. Metacalibration

This method consists in measuring the response matrix \mathbf{R} of a shape measurement algorithm to a shear artificially applied to an image. The i th component of the observed ellipticity of a galaxy, ε_i , is an estimator of the galaxy shear, $\langle \varepsilon_i \rangle = g_i^{\text{obs}}$, however, not an unbiased estimator in general. The linearized relation between true and estimated shear for a given galaxy can be written as

$$g_i^{\text{obs}} = \sum_{j=1}^2 R_{ij} g_j^{\text{true}} + c_i, \quad (5)$$

where the response matrix can be described as $R_{ij} = \partial g_i^{\text{obs}} / \partial g_j^{\text{true}}$ and c_i is the additive bias. The trace of the shear response matrix is also parameterized as $\text{tr}(\mathbf{R}) = 2(1 + m)$ where m is the multiplicative shear bias.

We numerically compute the shear response matrix by replacing the derivative with finite differences. For this, we create four new images for each detected galaxy. These images are deconvolved with the model PSF at that position, then sheared in two directions for both shear components, and re-convolved with a circularized PSF that is slightly larger than the original one. We note that a fifth image is created, which is unsheared but has the same re-convolved PSF. This image is used to measure shapes that are consistent with the metacalibration correction.

The response matrix has two additive components, the shear response matrix and the selection response matrix, both of which can be computed via metacalibration, as follows. First, the shear response matrix is computed for each galaxy individually. Denoting with $g_i^{\text{obs}, \pm, j}$ the observed shear of the galaxy image sheared by $\pm \Delta g$ in the j direction, we get

$$R_{ij}^{\text{shear}} = \frac{g_i^{\text{obs}, +, j} - g_i^{\text{obs}, -, j}}{2\Delta g}. \quad (6)$$

We used $\Delta g = 0.01$. Since this is a very noisy measurement, often resulting in singular matrices, we compute the mean value over all galaxies,

$$\langle R_{ij}^{\text{shear}} \rangle = \left\langle \frac{g_i^{\text{obs}, +, j} - g_i^{\text{obs}, -, j}}{2\Delta g} \right\rangle. \quad (7)$$

Second, the selection response matrix quantifies selection biases originating from correlations between shear and selection-dependent image properties. For example, if a cut is applied on the S/N of the galaxy sample and the S/N varies with shear, the

¹¹ <https://github.com/CosmoStat/lenspack>

effective sample shear after the cut is modified, and no longer representative of the underlying population shear, which introduces a bias.

The selection matrix cannot be obtained for a single galaxy, but only for a sample of galaxies. For that, we first apply our selection S to the four samples of sheared galaxies. This provides us with a selection mask $S^{\pm,j}$ for each of the four cases. These four masks are now in turn applied to the fifth sample without added shear. We compute the mean observed shear for each of the four masked samples, denoted as $\langle g_i^{\text{obs}} \rangle_{S^{\pm,j}}$. Any difference in these mean values is due to shear-dependent selection criteria, and can be used to define the selection response matrix as

$$\langle R_{ij}^{\text{selection}} \rangle = \frac{\langle g_i^{\text{obs}} \rangle_{S^{+,j}} - \langle g_i^{\text{obs}} \rangle_{S^{-,j}}}{2\Delta g}. \quad (8)$$

The total mean response matrix is

$$\mathbf{R} = \langle \mathbf{R}^{\text{shear}} \rangle + \langle \mathbf{R}^{\text{selection}} \rangle. \quad (9)$$

The estimated shears are calibrated by matrix multiplication with \mathbf{R}^{-1} , to obtain g_i^{true} for all i following Eq. (5).

In the following sections of the paper, we will need the use of the diagonal and off-diagonal terms of the $\mathbf{R}^{\text{shear}}$ and $\mathbf{R}^{\text{selection}}$ matrix. They are defined as

$$R_{\text{diag}} = (R_{11} + R_{22})/2 \quad (10)$$

$$R_{\text{off-diag}} = (R_{12} + R_{21})/2. \quad (11)$$

5.2. Global values of the multiplicative shear bias and additive bias

Usually, the metacalibration method is applied to the entire catalog, providing us with global values for the shear calibration quantities. We reproduce those results for CFIS-P3 here. The response matrices for CFIS-P3 are

$$\mathbf{R}^{\text{shear}} = \begin{pmatrix} 0.74397 & 0.00001 \\ -0.00049 & 0.74122 \end{pmatrix} \pm \begin{pmatrix} 0.00020 & 0.00014 \\ 0.00022 & 0.00024 \end{pmatrix};$$

$$\mathbf{R}^{\text{selection}} = \begin{pmatrix} -0.10872 & -0.00105 \\ -0.00001 & -0.11154 \end{pmatrix} \pm \begin{pmatrix} 0.00050 & 0.00046 \\ 0.00046 & 0.00048 \end{pmatrix}. \quad (12)$$

The additive shear bias components are

$$\mathbf{c} = \begin{pmatrix} -0.00104 \\ 0.00043 \end{pmatrix} \pm \begin{pmatrix} 0.00011 \\ 0.00011 \end{pmatrix}. \quad (13)$$

The errors are computed via jackknife resampling of all galaxies.

5.3. Local calibration

The multiplicative shear bias depends on quantities that vary spatially, such as the PSF properties (Paulin-Henriksson et al. 2008), galaxy size (for example Spindler 2018; Kuchner et al. 2017) and magnitude (Miller et al. 2013), or the local galaxy density (Hoekstra et al. 2017). These spatial variations may be correlated with shear: in some cases, they both vary with the LSS environment, such as galaxy density and shear. In other cases, residual errors may create cross-correlations, for example, an imperfect PSF calibration influences both calibration and shear.

The understanding and mitigation of spatial patterns in shear calibration is an active field of research for future weak-lensing surveys (Kitcing et al. 2021; Cragg et al. 2022). Here,

we investigate local shear calibration, the dependence of calibration on observed quantities, and the impact on cosmological parameters.

In the context of weak-lensing mass maps as tracers of the LSS, Van Waerbeke et al. (2013) carry out a local calibration in pixels of 1 arcmin size. Their multiplicative shear bias is computed as a smooth, two-parameter fitting function from image simulations. This is in contrast to our case of metacalibration, which provides very noisy calibration estimates for individual galaxies. These estimates need to be averaged over substantially larger areas to reduce the uncertainty such that the corresponding response matrices are numerically stable enough for inversion.

In this work, we employed a series of square patches with a size of $d_s = 4, 2, 1, 0.5$, and 0.25 degrees over which the shear calibration estimates are averaged in turn. To easily divide the observed sky area into an integer number of square-sized sub-patches, we first projected the data into a Cartesian plane. Next, we extended this area via zero padding to size $N_x \times N_y \text{ deg}^2$, such that N_x and N_y are multiples of the largest sub-patch size $d_{s,\text{max}} = 4 \text{ deg}$.

We computed local versions of $\mathbf{R}^{\text{shear}}$ (7) and $\mathbf{R}^{\text{selection}}$ (8) by carrying out the averages per sub-patch. If the number of galaxies in a sub-patch was smaller than a threshold $n_{\text{gal},0}$, we replaced the local response matrices in that sub-patch by their global mean, to avoid numerical instabilities. We chose $n_{\text{gal},0} = \bar{n}_{\text{gal}}/2$. This occurs mainly at the edges of the field where sub-patches overlap with the area of zero padding. We calibrated the estimated shear of each galaxy in a given sub-patch by the total local response matrix, in analogy to (9). Similarly, we computed the local shear bias \mathbf{c} as the average per sub-patch.

We computed the error of the local response matrices by creating jackknife resamples from the smallest sub-patch size $d_{s,\text{min}} = 0.25 \text{ deg}$. Each sub-patch of size $d_s > d_{s,\text{min}}$ was thus split into $(d_s/d_{s,\text{min}})^2$ jackknife subsamples; for the smallest sub-patch size, we could not compute the error. The additive bias is easily obtained locally by computing jackknife errors over galaxies per sub-patch.

5.4. Results of the local calibration

5.4.1. Variation of the shear and selection response

We show the spatial variation of $\mathbf{R}^{\text{shear}}$ and $\mathbf{R}^{\text{selection}}$ in sub-patches of different size in Fig. 3. From top to bottom the shear and selection responses are computed on sub-patches with increasing size of $d_s = 0.25, 0.5, 1, 2$ and 4 degrees. The smaller the sub-patch size, the more the response shows variations corresponding to the local spatial environment. The larger the sub-patch size, the more the calibration approximates the global calibration. The errors of the shear response are small, even in the case of the 0.5 degree sub-patch size. Concerning the selection matrix, we can see that on 0.5 deg^2 , the errors have the same order of magnitude as $\mathbf{R}^{\text{selection}}$. On 1 deg^2 , the errors are smaller than $\mathbf{R}^{\text{selection}}$ indicating that the computation of the local selection calibration is precise when done on 1 deg^2 or more. To emphasize these conclusions, the 1D distributions of $\mathbf{R}^{\text{shear}}$ and $\mathbf{R}^{\text{selection}}$ are plotted in Fig. A.1.

5.4.2. Variation of the additive bias

The additive bias and its standard deviation are shown in Fig. 4. The standard deviation is as large as the value itself for local calibration on small scales. Except for the largest sub-patch sizes,

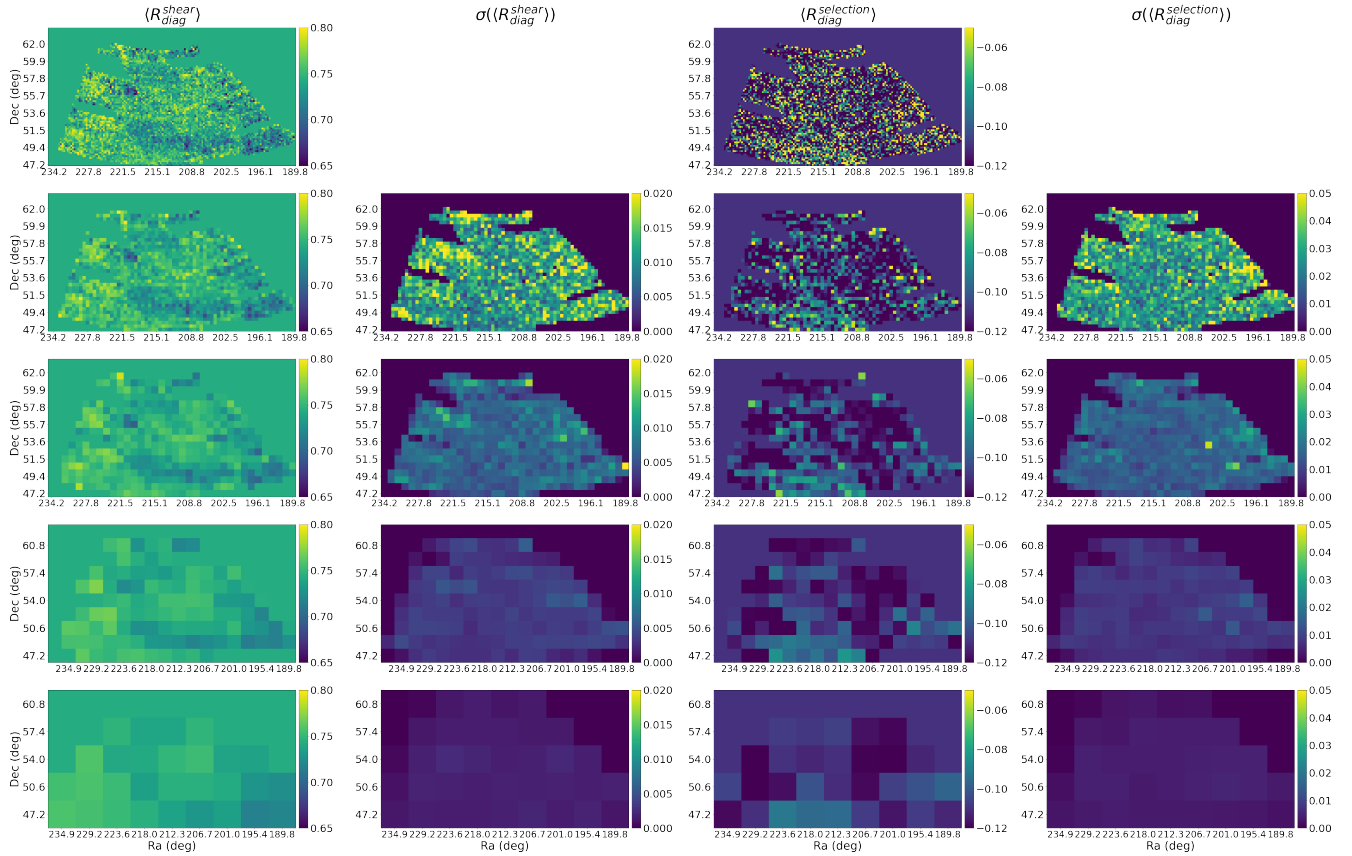


Fig. 3. $\langle R_{diag} \rangle$ depending on the size of the local calibration. From top to bottom, calibration is done on 0.25, 0.5, 1, 2, 4 deg². The four columns are $\langle R_{diag}^{shear} \rangle$, its standard deviation, $\langle R_{diag}^{selection} \rangle$, and its standard deviation.

the values of c_1 and c_2 per pixel are consistent with zero. This reflects the large fluctuations and randomness of the additive bias on small scales. To emphasize these conclusions, the 1D distributions of c_1 and c_2 are plotted in Fig. A.2.

5.4.3. Relative errors

We computed the errors of the response matrices at different sub-patch sizes, averaged over all sub-patches (blue lines), compared to the relative errors of the global calibration (black lines). The results are shown in Fig. 5.

The relative error of the shear matrix is below 1% in all cases but always above the relative error of the global calibration, meaning that there are fluctuations at all the studied scales, but these are small. The fractional error of the selection matrix is higher. It reaches 17% at small sub-patch sizes but decreases up to 3% at 4 deg². For the shear and selection matrix, we see that when we go on a smaller calibration size, the relative errors asymptotically approach the one of the global calibration.

5.5. Parameter correlation matrix

The calculation of the local shear bias allows us to further explore possible origins of shear bias, and their influence on other quantities obtained from the galaxy sample. To that end, we computed the correlation matrix between different quantities, for which we used patches of size $d_s = 1$ deg. We also considered correlations between quantities other than shear bias. The correlation matrix is calculated using the pandas *DataFrame.corr*

function¹²; it uses the Pearson method, which computes the standard correlation coefficient. The correlation matrix between the different quantities listed in Table 3 is shown in Fig. 6. To better see the correlation terms by terms, another correlation matrix is shown in Fig. B.1.

First, we see that the E - and B -mode smoothed convergence, $\kappa_{E,sm}$ and $\kappa_{B,sm}$, are not strongly correlated. This indicates that there is no significant systematic effect that mixes both modes. Neither mode is strongly correlated to other quantities. Also, the number of peaks n_{peak} is not correlated to observational effects.

Further, there is no visible correlation between the PSF size and the additive bias. This is evidence for the correct estimation of PSF size since a bias in the PSF size typically leads to an additive shear bias. We also note the absence of a correlation between the PSF ellipticity and the shear selection matrix. This gives us confidence that the PSF deconvolution in the metacalibration process works well. There is, however, a 20–30% negative correlation between the selection response elements and the size of the PSF. A plausible explanation is that in areas with larger seeing, fewer small and faint galaxies make it into the weak-lensing sample. This leads to a stronger (more negative) selection bias, which is reflected in the anticorrelation.

A negative correlation can be seen between the number of epochs, n_{epoch} , and the PSF ellipticity modulus $|e_{PSF}|$. Since the former is the mean over all contributing epochs, a reason for this anticorrelation might be the PSF ellipticity gets more circular when averaged over more independent observations.

¹² <https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.corr.html>

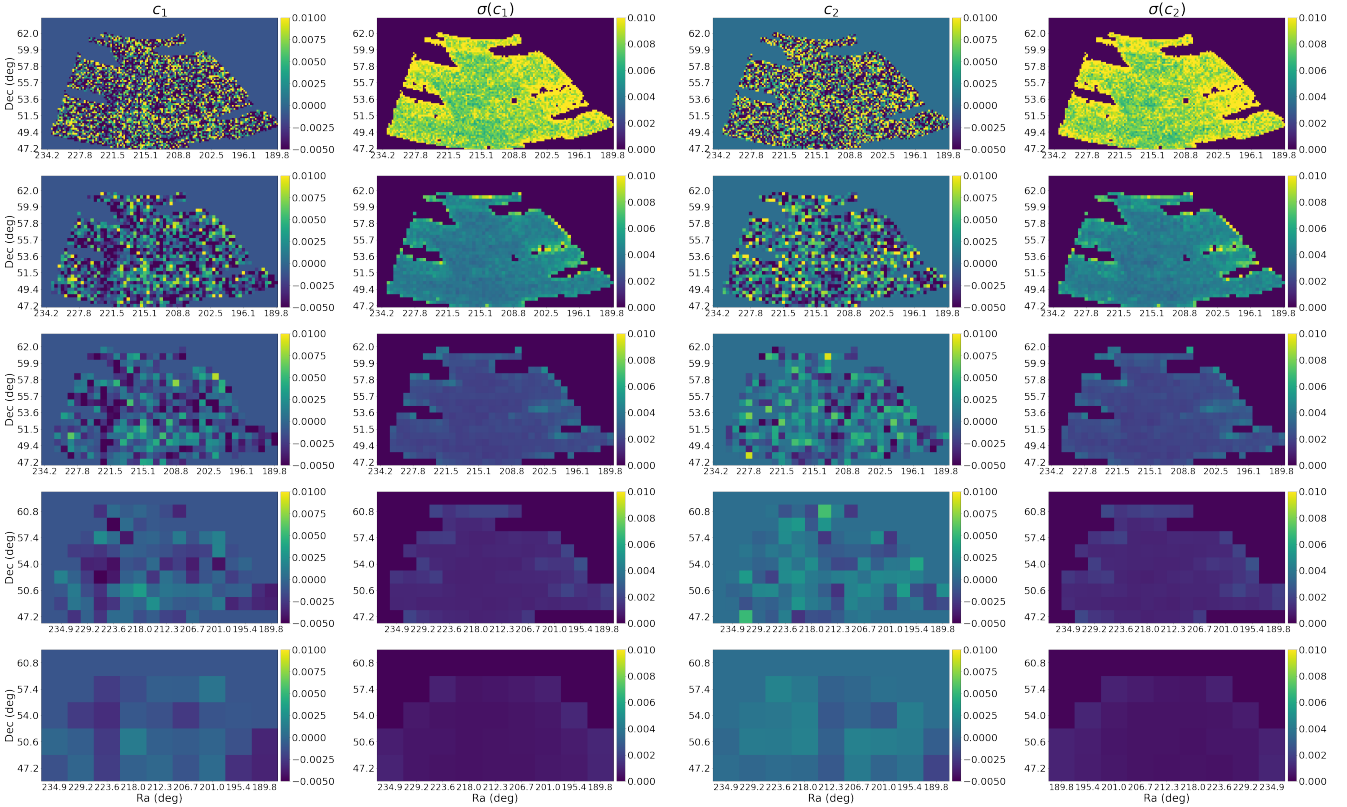


Fig. 4. c_1 and c_2 depending on the size of the local calibration. From top to bottom, calibration is done on 0.25, 0.5, 1, 2, 4 deg². The four columns are c_1 , its standard deviation, c_2 , and its standard deviation.

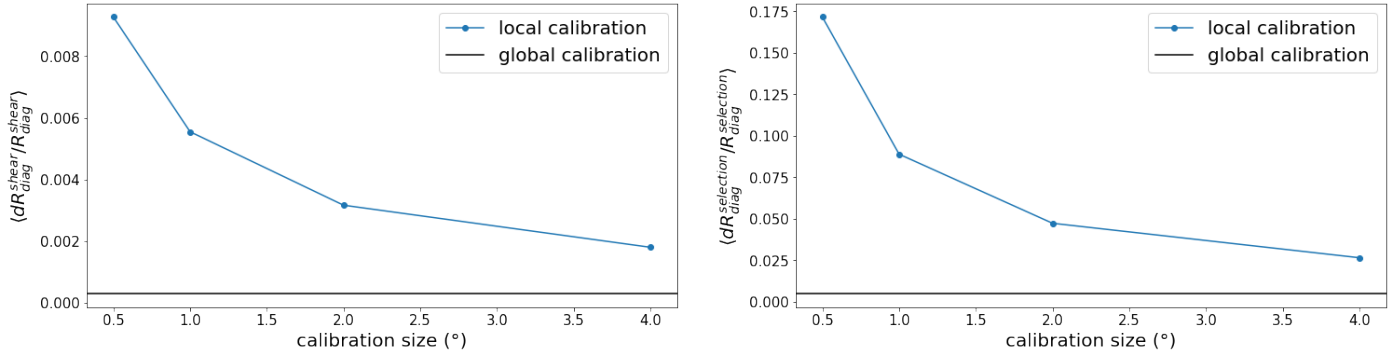


Fig. 5. Relative errors of shear matrix (left panel) and selection matrix (right panel). In both cases, the blue line is the relative error of the diagonal terms of the matrix depending on the calibration size. The black line is the relative error of the diagonal terms of the matrix computed with the value obtained on the global calibration case.

There is a negative correlation of the r -band magnitude with galaxy weight w , S/N, and size T_{gal} , as well as a positive correlation with n_{epoch} . This reflects the expectation that faint galaxies have a lower weight and S/N, and are easier observed with more exposures.

6. Impact of uncertainty and systematic effects on cosmological parameters

This section explores different systematic effects and uncertainties, and their impact on cosmological parameter constraints. We quantify the potential biases on cosmology from the spatially varying shear calibration (Sect. 6.1), the redshift uncertainty (Sect. 6.3), a residual shear bias (Sect. 6.2), baryonic feedback

(Sect. 6.4), intrinsic galaxy alignment and cluster member dilution (Sect. 6.5). All these effects are studied jointly in Sect. 6.6.

For each effect, we will show constraints on cosmological parameters as compared to those obtained with the “ideal” case. This case, which is represented in blue in all the following figures represents the constraints obtained when we use the data calibrated globally without residual shear bias, a mean redshift of $\bar{z} = 0.65$, without baryonic correction, cluster member dilution or intrinsic alignment. This assumes an unrealistic best-case scenario of vanishing spatial variation of shear calibration and residual shear bias, and no biases from baryons, intrinsic alignment or cluster members. When the parameters are well-constrained, we specify the shift compared to the ideal case, which is the case for Ω_m . The parameters M_r and the lower bound for A_s are in most cases not well constrained within the prior.

Table 3. Symbols and description of quantities used in the cross-correlation matrix (Fig. 6).

Symbol	Description
$\kappa_{E,sm}, \kappa_{B,sm}$	Smoothed E - and B -mode convergence, respectively
n_{gal}	Number of galaxies per pixel
n_{peak}	Number of peaks
$ e_{psf} $	Modulus of the point spread function
T_{gal}	Galaxy size
n_{epoch}	Number of single-exposure epochs
r	r -band galaxy magnitude
SNR_{gal}	Galaxy signal-to-noise ratio
T_{psf}	psf size
w	Galaxy weight
$\overline{R}_{diag}^{shear}, \overline{R}_{off-diag}^{shear}$	Average of the diagonal and off-diagonal shear response matrix element, respectively, Eq. (7)
$\overline{R}_{diag}^{selection}, \overline{R}_{off-diag}^{selection}$	Average of the diagonal and off-diagonal selection response matrix element, respectively, Eq. (8)
c	Average of the additive bias element

6.1. Local calibration

To study the impact on cosmology of spatially varying shear biases, we performed a local shear calibration using different scales, following the method explained in Sect. 5.3 for the multiplicative shear bias. For the additive bias, we used the value obtained with the global calibration because the local calibration value is very noisy. The result is shown in Fig. 7. We see that a local calibration will always shift Ω_m to a lower value except for the calibration on 0.5 deg^2 . These variations are all within the statistical error bars. There is no systematic variation that becomes evident when going to smaller calibration sizes. For A_s and $\sum M_v$, the trend is not clear. Ω_m shifts by -0.015 , which corresponds to -0.3σ (i.e., 0.3 times the statistical uncertainty) when the calibration goes from global to local on 4 deg^2 for example.

From this work, it is not clear which calibration size is the best one because, depending on the size of the calibration, we will capture more or less the local effect. Future work has to be done to determine the criteria to choose the size of the calibration. In this work, when we need a local calibration, we use the one done on 1 deg^2 because the relative error of the selection matrix (which is the limiting one) is below 10%.

6.2. Residual multiplicative shear bias

The shear bias m computed earlier with metacalibration is not perfect. A residual bias $\Delta m = m^{\text{metacal}} - m^{\text{true}}$ remains and generally has to be quantified with image simulations. We used the value $\Delta m = 0.007$ based on the results from Guinot et al. (2022), who found $\Delta m = 0.007 \pm 0.03$, estimated using CFIS-like image simulations of isolated galaxies and ignoring the effect of blending. In comparison, other surveys also find residual biases at around the per cent level or below; for example, MacCrann et al. (2021) and Gatti et al. (2021) state a value of $\Delta m = -0.0208 \pm 0.0012$ for DES-Y3.

We modeled the effect of the residual multiplicative shear bias by adding the residual bias Δm to the response matrix in the local case. This corresponds to a conservative, worst-case scenario where the residual bias is constant in space. Thus, Eq. (9) is modified to

$$\mathbf{R} = \langle \mathbf{R}^{\text{shear}} \rangle + \mathbf{I} \times \Delta m + \langle \mathbf{R}^{\text{selection}} \rangle, \quad (14)$$

where \mathbf{I} is the identity matrix. The impact of this bias on the cosmological constraints is shown in Fig. 8. In blue, the calibration is global whereas it is done on 1 deg^2 for the green and red cases. Moreover, the red case includes the multiplicative shear bias. Including the residual shear bias changes the values of the parameters in the same direction as the local calibration compared to the global calibration. Ω_m shifts by -0.024 , which corresponds to -0.5σ with respect to the global calibration. A positive Δm effectively reduces the estimated shear, resulting on average in a smaller number of peaks, and thus smaller clustering parameters. The 1D marginalized contours for A_s shift to the right, contrary to what we would expect. However, the joint $\Omega_m - A_s$ 2D contours clearly reflect the smaller calibrated shear and shift toward smaller clustering amplitudes.

6.3. Redshift uncertainty

As discussed in Sect. 2.3, $\bar{z} = 0.65$ is our best estimate of the redshift of the data observed from CFIS. This estimate has however a level of uncertainty. To quantify the impact of this uncertainty on our results, we compared cosmological constraints using the simulated convergence maps interpolated to $\bar{z} = 0.65$ and $\bar{z} = 0.68$, which corresponds to the two mean redshift estimates discussed in Sect. 2.3. The result is shown in Fig. 9. Fitting the data with the model at a higher mean redshift only slightly shifts the posterior: Ω_m shifts by $+0.001$, which corresponds to $+0.02\sigma$ when the redshift goes from 0.65 to 0.68. These shifts are well within the statistical uncertainties of the two parameters. Such a shift is, however, expected to have a significant influence on analyses using a larger survey area and/or redshift tomography. We also performed a test using only simulations, where the data vector was the mean of the fiducial simulation at $z = 0.65$ or at $z = 0.68$. The result is presented in Appendix C. We note, however, that when using simulations only, the shift appears to be larger, this may indicate some residual systematics that impacts the results on the data shown in this section.

6.4. Baryonic feedback

The impact of baryons on the LSS is important for cosmological analyses with weak-lensing peak counts (Osato et al. 2015; Harnois-Déraps et al. 2021b; Coulton et al. 2020). The redistribution of matter due to baryonic processes tends to reduce the number of high S/N peaks and augment that of smaller S/N values.

We used the results from Coulton et al. (2020) who model the fractional difference of the number of peaks, $\Delta N_{\text{peaks}}/N_{\text{peaks}}$, for three different baryon physics scenarios based on the BAHAMAS hydrodynamical simulations (McCarthy et al. 2016) compared to MassiveNuS dark-matter-only simulations (Liu et al. 2018). This assumes that baryonic processes are independent of the underlying cosmology. The three scenarios are denoted as LowAGN, Fiducial, and HighAGN, where the amount of baryonic feedback increases in that order. The fiducial baryonic correction comes from the BAHAMAS simulations with a feedback model designed to best match the observations. The LowAGN and HighAGN corrections are simulations where the active galactic nucleus (AGN) heating is lowered or raised by 0.2 dex, respectively, and thus the simulations skirt the lower and upper bounds of the observed gas fraction.

We modified the model predictions of peak counts by multiplying the number of peaks from the MassiveNuS (dark-matter-only) simulations by $N_{\text{peaks}}/N_{\text{peaks,DM}}$, shown in Fig. 10.

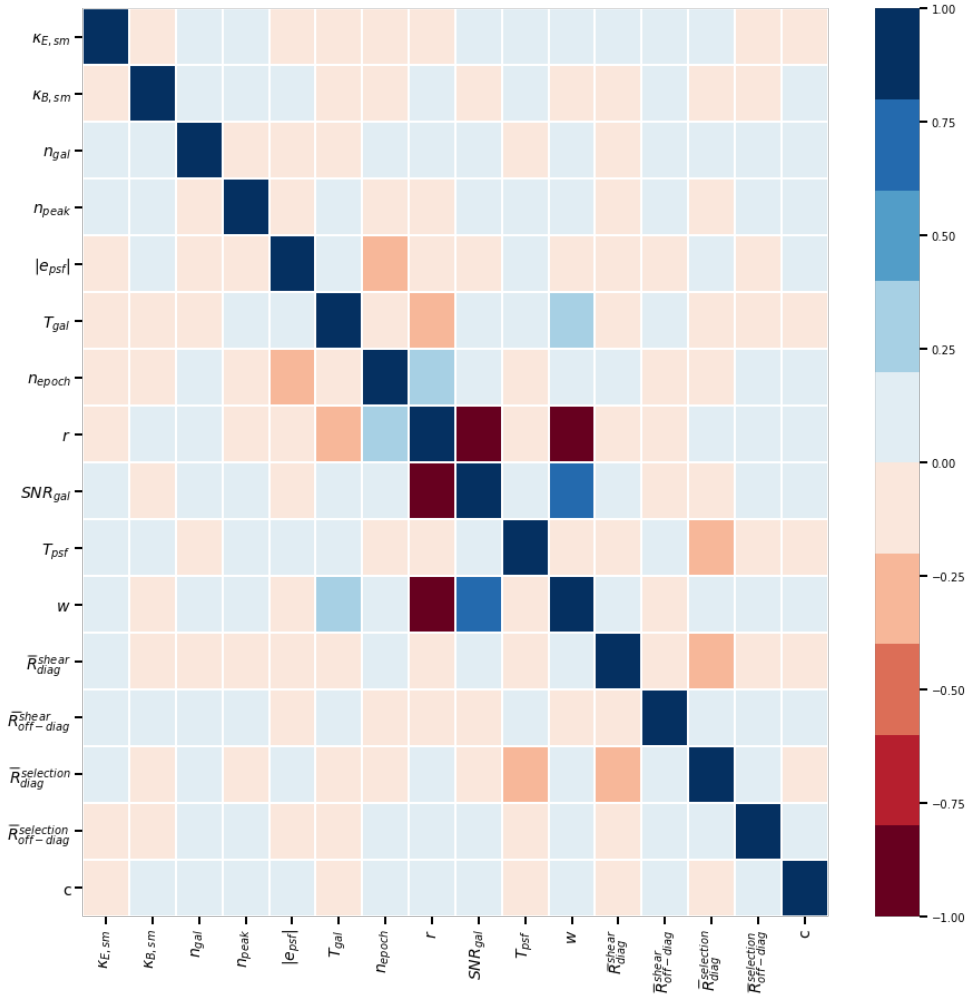


Fig. 6. Correlation matrix of the quantities listed in Table 3. The colors indicate the amplitude of the correlation, ranging between -1 and 1 . All quantities are mean values.

We can see how the constraints evolve when we use real CFIS-P3 data and corrected simulations. To see the effect of the baryonic correction only, we use the global calibration, simulations at $z = 0.65$, no intrinsic alignment or cluster member dilution correction. We do not know how the baryonic feedback influences the observed data thus we correct the MassiveNuS simulations with the three different baryonic feedback corrections. In Fig. 11 we show the results of these corrections (LowAGN in green, Fiducial in red, and HighAGN in pink) compare to the case of the dark-matter-only simulation (blue).

Baryonic feedback has a significant impact on the best-fit parameters. Ω_m and A_s are shifted to higher values with increasing baryonic modifications. For example, Ω_m shifts by $+0.027$, which corresponds to $+0.5\sigma$ when the fiducial goes from dark-matter-only to LowAGN correction. Li et al. (2019) found that it is the high S/N peaks that dominate the constraints and in Fig. 10 we see that the number of high peaks is suppressed by the baryonic feedback. This results in a shift toward higher values in Ω_m , compensating for the suppression of peaks for $S/N > 3$. We also see that the strength of the feedback is not as important as just taking it into account.

6.5. Further systematic effects

At least two further effects can affect constraints: intrinsic alignment and cluster member dilution. As we briefly explain below,

both effects can be suppressed, in first approximation, following Harnois-Déraps et al. (2021a), with a cut in S/N.

6.5.1. Intrinsic alignment

Due to the radial alignment of satellite galaxies within dark-matter halos, galaxies are not randomly oriented. Thus, their shape and alignments are affected by their environment and tidal fields. The intrinsic alignment has two components: the intrinsic-intrinsic correlations caused by the alignment of galaxies that are physically linked together and the gravitational-intrinsic correlations, which is the alignment of halo galaxies.

6.5.2. Cluster member dilution

The source density is not homogeneous and increases around foreground clusters. Around a cluster at a given redshift, there are more galaxies but, as we do not know their redshift, they are included in the signal but may not be lensed. This effect leads to a coupling between the peak positions and the amplitude of the measured shear relative to the expected shear (Kacprzak et al. 2016). Moreover, these regions of clusters have a larger blending rate, and thus galaxies behind clusters are more likely to be missed. These effects can result in a miscalibration between data and simulations and lead to a reduction in the mean shear signal (as we count more galaxies for the signal). An analysis of

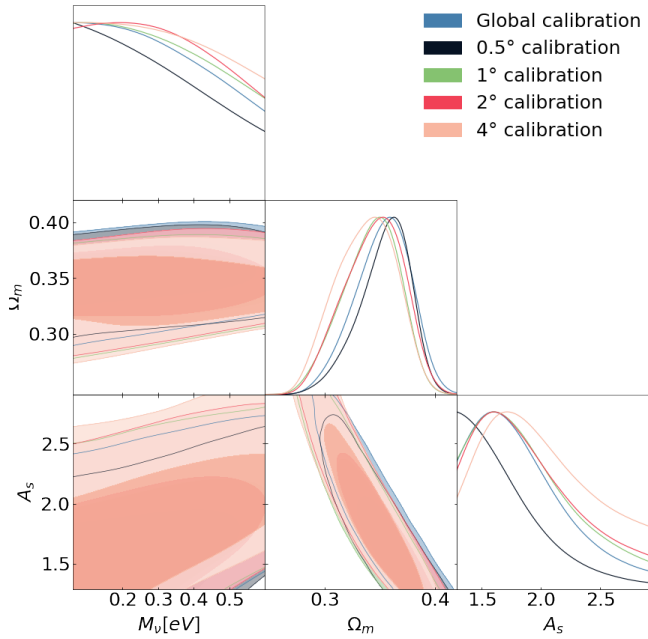


Fig. 7. 1D and 2D marginal posteriors for CFIS-P3 using different calibration sizes in the metacalibration. The 2D inner and outer contours show the 68% and 95.5% credible region, respectively. The case when the calibration is done globally (blue) is compared to the case where the calibration is done on 0.5 (black), 1 (green), 2 (red), and 4 deg² (pink). Ω_m shifts by -0.015 , which corresponds to -0.3σ when the calibration goes from global to local on 4 deg².

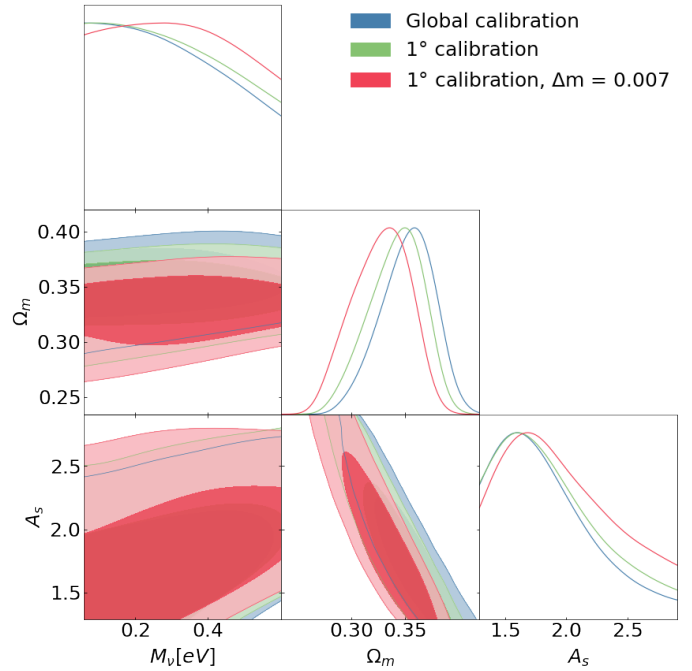


Fig. 8. 1D and 2D marginal posteriors for CFIS-P3, with the residual multiplicative shear bias added. The 2D inner and outer contours show the 68% and 95.5% credible region, respectively. Ω_m shifts by -0.01 , which corresponds to -0.2σ when the calibration goes from global (blue contour) to local on 1 deg² (green contour). When Δm is added (red contour), Ω_m shifts by -0.024 , which corresponds to -0.5σ with respect to the global calibration.

cluster member dilution in Subaru HSC weak-lensing mass maps is available, for example, in [Oguri et al. \(2021\)](#).

6.5.3. Reducing intrinsic alignment and cluster member dilution with a cut on S/N

[Harnois-Déraps et al. \(2021a\)](#) find that the effect of intrinsic alignment is small for peaks with $S/N < 3$ and that the effect of the cluster member dilution is small for $S/N < 4$. The local calibration might capture part of the cluster dilution effect, but as we have no way of knowing it without simulations, we chose to cut the S/N range to be conservative. Thus, as a first approximation of how these effects impact the cosmological parameters, we computed the constraints using only the range for peak counts of $S/N < 3$. This selection should minimize the impact of both systematic contaminations. We tested both cases ($-2 < S/N < 3$ and $-2 < S/N < 6$) on mock peaks from MassiveNuS and find consistent results.

The comparison of the resulting constraints obtained with $S/N < 3$ (red) or $S/N < 6$ (blue) is shown in Fig. 12. Constraints are slightly larger when we cut the S/N range, which is expected. For both M_v and A_s parameters, the change is small.

The high-S/N peaks are affected by intrinsic alignment and cluster member dilution, which leads to a reduction in peak S/N ([Harnois-Déraps et al. 2021a](#)). When using the full S/N range and not accounting for those two effects, the reduction in peak S/N results in a lower inferred clustering amplitude and a lower Ω_m . This can be seen in the $\Omega_m - A_s$ 2D posterior distribution. We also perform a test using simulations only, with the range for peak counts of $S/N < 3$. As in previous tests, the data vector is the mean of the fiducial simulation at $z = 0.65$ and the result is presented in Appendix C.

6.6. Parameter constraints that combine all systematic effects

Our conservative model, which is the most suitable to represent the data, combines different mitigation schemes, and uses the following parameters and settings: The mean redshift is set to $\bar{z} = 0.65$. We calibrate the shear locally at a scale of 1 deg². This accounts for spatial variations of shear bias and is the smallest scale for which the estimate of the selection is not dominated by noise. The residual shear bias is set to the value estimated from image simulations, $\Delta m = 0.007$. Baryonic feedback is accounted for by using the “fiducial baryon” case of [Coulton et al. \(2020\)](#). The data vector is composed of peak counts with $-2 < S/N < 3$ to minimize the systematic errors from intrinsic alignment and cluster member dilution.

In Fig. 13 we show constraints using this conservative model, and compare those to the ones obtained under the ideal (optimistic) case described at the beginning of this section. Ω_m shifts by $+0.008$, which corresponds to $+0.2\sigma$ when we go from the ideal to the conservative model. As expected, the conservative model results in wider constraints. In Fig. 14, we can see the marginalized distributions for the 68% confidence interval of the different parameters depending on the different cases we have tested.

7. Conclusions

This study highlights the importance of properly accounting for systematics, such as the local calibration, residual multiplicative shear bias, intrinsic alignment and cluster member dilution, redshift uncertainties, and baryonic corrections, in the context of a peak-count cosmological analysis with CFIS data. We performed

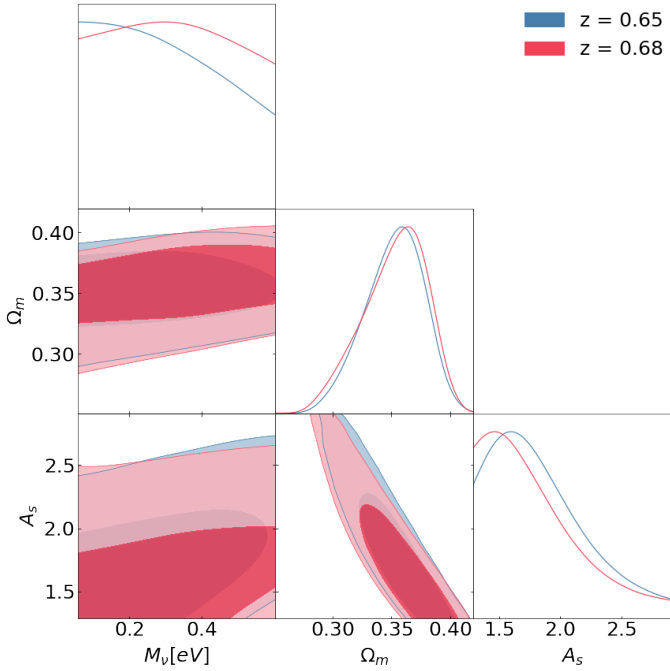


Fig. 9. 1D and 2D marginal posteriors using simulations at different redshift. The 2D inner and outer contours show the 68% and 95.5% credible region, respectively. The shear calibration is global. Ω_m shifts by $+0.001$, which corresponds to $+0.02\sigma$ when the redshift goes from 0.65 (blue) to 0.68 (red).

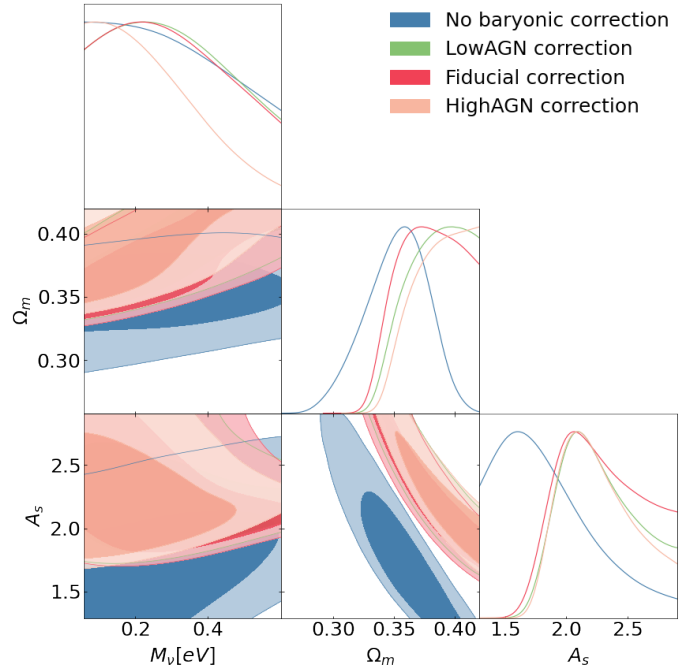


Fig. 11. 1D and 2D marginal posteriors for CFIS-P3, with the peak count predictions corrected with different baryonic feedback. The 2D inner and outer contours show the 68% and 95.5% credible region, respectively. The case without baryonic correction (blue) is compared to the three baryonic scenarios: LowAGN (green), Fiducial (red), and HighAGN (pink). Ω_m shifts by $+0.027$, which corresponds to $+0.5\sigma$ when the fiducial goes from dark-matter-only to LowAGN correction.

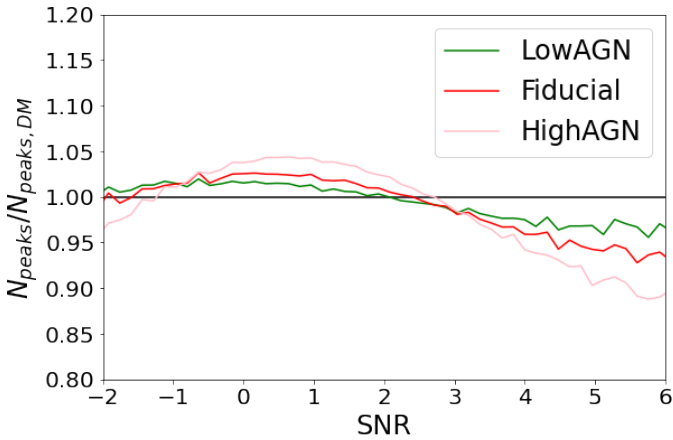


Fig. 10. Fractional difference in peak count obtained with Coulton et al. (2020) data. The difference between simulations with LowAGN, Fiducial, or HighAGN and dark-matter-only simulations is presented in green, red, and pink, respectively.

a likelihood analysis on the sum of neutrino masses, $\sum M_\nu$, the matter density parameter, Ω_m , and the amplitude of the primordial power spectrum, A_s . For this purpose, we used the CFIS-P3 catalog and the MassiveNuS N -body simulations because they have the advantage of providing us with a large set of cosmological models and different tomographic bins. First, we obtained constraints with simulations only to validate the methodology used to model the redshift uncertainty, baryonic feedback, intrinsic alignment, and cluster member dilution. We then used the CFIS-P3 data to quantify the effect of systematics on cosmological constraints. We remark that Ω_m is the parameter that is constrained the most and is the most impacted by the systematics. As a summary plot, Fig. 14, which shows the marginalized

distributions for the 68% confidence interval of the parameters per studied case. On the one hand, the baryonic corrections, the intrinsic alignment, and the cluster member dilution shift this parameter to higher values. On the other hand, a local calibration and a multiplicative shear bias $\Delta m = 0.007$ shift Ω_m to smaller values. Concerning A_s , the baryonic correction and the addition of Δm shifts this parameter to higher values, whereas cutting the high S/N peaks shifts A_s to a lower value. For $\sum M_\nu$, the parameter is not constrained enough to allow for any conclusions to be drawn. More specifically, concerning the shear calibration, it is probable that a local one is preferred over a global one, as long as the error and standard deviation are small, because it accounts for the local effects of the catalog. Nevertheless, some work has to be done to determine which calibration size is the better one. We notice that using a calibration on 1 deg^2 with a value of the residual multiplicative shear bias $\Delta m = 0.007$ shifts Ω_m by -0.024 (-0.5σ) compared to the case in which a global calibration is applied. Choosing a reasonable calibration size and having a robust estimate of Δm is important for having a better estimate of the constraints. We note that the residual bias used here computed from image simulations neglected effects such as galaxy image blending. A more realistic estimate of Δm is necessary for future analyses of CFIS weak-lensing data.

Concerning the uncertainty on the mean redshift estimate, for a bias of $\Delta \bar{z} = 0.03$, as considered in this study, the impact on the constraints is small enough to not be considered: the shift in Ω_m is only $+0.001$ ($+0.02\sigma$). Nevertheless, we are aware that further work is needed to have a more complete description of the redshift distribution and obtain more accurate constraints. Using simulations only, the shift in Ω_m is larger, 0.7σ . This indicates that other effects in the data might lead to an underestimation of the actual shift.

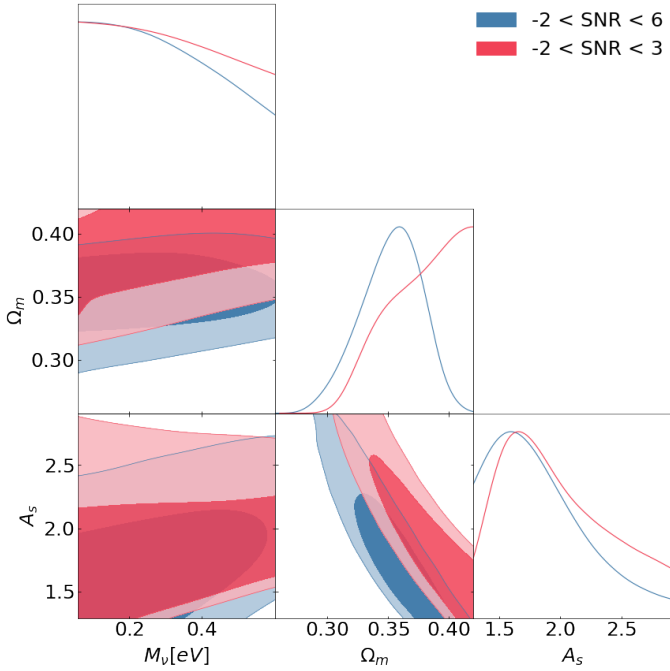


Fig. 12. 1D and 2D marginal posteriors for CFIS-P3 using different ranges in S/N. The 2D inner and outer contours show the 68% and 95.5% credible region, respectively. The shear calibration is global. The case with the full S/N range (blue) is compared to the case where we use $-2 < S/N < 3$ to minimize the effect of residual bias (red). Ω_m shifts by $+0.027$, which corresponds to $+0.5\sigma$ when we cut the high S/N peaks.

To account for baryonic corrections, we considered three different flavors of baryonic feedback, labeled in this study as HighAGN, Fiducial, and LowAGN correction. We conclude that the specific flavor of baryonic feedback is less important than the difference between baryonic feedback and pure dark-matter models. Applying the LowAGN baryonic correction shifts Ω_m by $+0.027$ ($+0.5\sigma$), showing the importance of modeling the baryonic feedback. Using a prediction of peak counts based on hydrodynamical simulations to model the baryonic feedback can help in getting more accurate constraints.

Then, we see that when we minimize the effect of the cluster member dilution and intrinsic alignment by only considering peaks with $S/N < 3$, the posterior on Ω_m is subject to an offset of 0.027 (0.5σ) toward higher values. Using simulations that model these effects will significantly improve the constraints since a cut in the S/N range will no longer be necessary.

Finally, we computed a conservative model where we used the parameters and settings that are most representative of the data. This gives larger constraints on M_v , Ω_m , and A_s .

We noticed that the value of Ω_m in particular can shift a lot due to different systematics. Our aim in this paper is not to estimate the final reference cosmological parameters, but rather to investigate, for the first time with UNIONS and peak counts, the impact of some of the different systematics at play for this survey. Among all the systematics considered in this paper, we have shown that the one with the highest impact on Ω_m is related to how baryonic corrections are implemented (Fig. 11); the choice of cut on the S/N (Fig. 12) also causes a substantial shift. Other systematics not considered here may further shift the final parameters. It is necessary to further pursue this effort to include other systematic effects, as well as to extend the way we incorporate the baryonic feedback in the analysis and the choice of the calibration size. The constraints will be more accu-

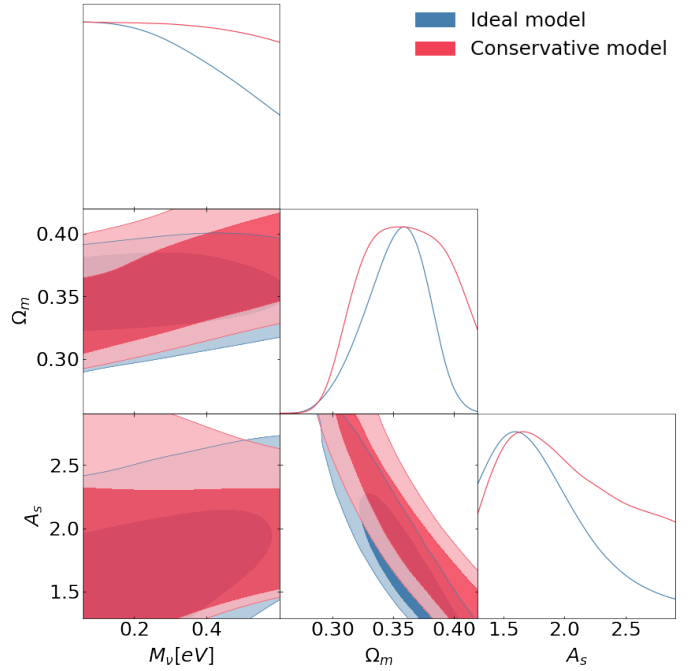


Fig. 13. 1D and 2D marginal posteriors for CFIS-P3 using the ideal or conservative model. The 2D inner and outer contours show the 68% and 95.5% credible region, respectively. The ideal one (blue) is obtained with global calibration, no baryonic correction, no intrinsic alignment, no boost factor, and on $-2 < S/N < 6$. The conservative model (red) is obtained with local calibration on 1 deg^2 , with residual bias, modeled under the fiducial baryonic correction, and on $-2 < S/N < 3$. Both models are computed at $\bar{z} = 0.65$. The 2D contours show the 95.5%, and the 1D filled area corresponds to the constraints within the 1 sigma confidence level. Ω_m shifts by $+0.008$, which corresponds to $+0.2\sigma$ between the reference (blue) and the fiducial model (red).

rate when simulations are able to also model intrinsic alignment and cluster member dilution; hydrodynamical simulations that include baryonic feedback could further improve the robustness of the results. In the future, having access to simulations built for UNIONS will allow us to make more precise investigations of various sources of bias. We also expect constraints to become more reliable with future (larger) data catalogs, such as the full UNIONS data set, for which the current pipeline will provide a starting point.

Acknowledgements. This work is based on data obtained as part of the Canada-France Imaging Survey, a CFHT large program of the National Research Council of Canada and the French Centre National de la Recherche Scientifique. Based on observations obtained with MegaPrime/MegaCam, a joint project of CFHT and CEA Saclay, at the Canada-France-Hawaii Telescope (CFHT) which is operated by the National Research Council (NRC) of Canada, the Institut National des Science de l'Univers (INSU) of the Centre National de la Recherche Scientifique (CNRS) of France, and the University of Hawai'i. Pan-STARRS is a project of the Institute for Astronomy of the University of Hawai'i, and is supported by the NASA SSO Near Earth Observation Program under grants 80NSSC18K0971, NNX14AM74G, NNX12AR65G, NNX13AQ47G, NNX08AR22G, and by the State of Hawai'i. This work has made use of the CANDIDE Cluster at the Institut d'Astrophysique de Paris and was made possible by grants from the PNCG, CNES, and the DIM-ACAV. This work was supported in part by the Canadian Advanced Network for Astronomical Research (CANFAR) and Compute Canada facilities. We would like to thank the anonymous referee for their helpful suggestions that improved the manuscript. We would also like to thank Joachim Hernois-Déraps for various discussions, particularly about the local metacalibration method. We are further grateful to Jia Liu and William R. Coulton for useful discussions and for providing us with the baryonic corrections employed in the analysis. We would like to thank Santiago Casas for useful discussion. Finally, we would thank

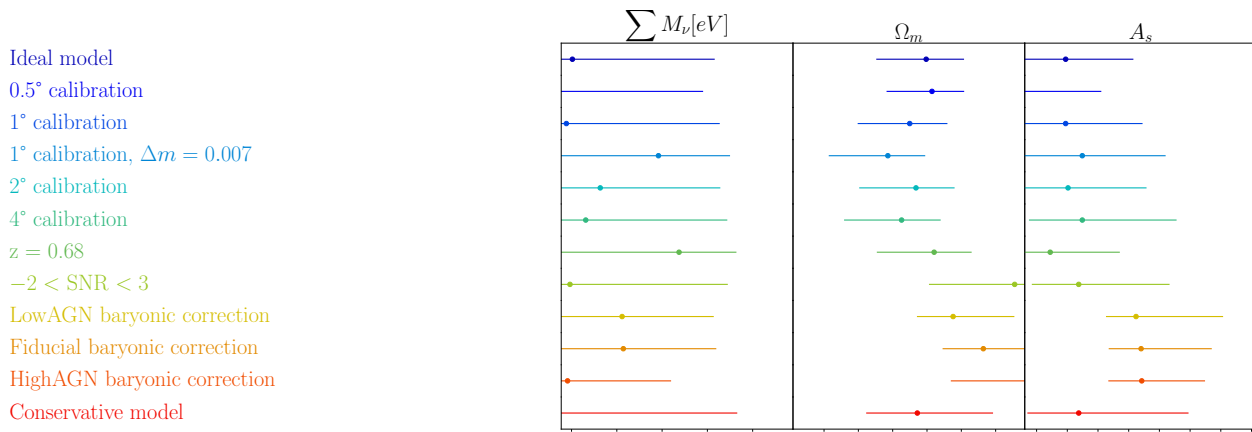


Fig. 14. Marginalized distributions for the 68% confidence interval for the three cosmological parameters for the different cases that include uncertainties and systematic effects.

the Cosmostat members for discussions about different aspects of this project. We thank the Columbia Lensing group (<http://columbialensing.org>) for making their simulations available. The creation of these simulations is supported through grants NSF AST-1210877, NSF AST-140041, and NASA ATP-80NSSC18K1093. We thank New Mexico State University (USA) and Instituto de Astrofísica de Andalucía CSIC (Spain) for hosting the Skies & Universes site for cosmological simulation products. Finally, this work uses the following software packages: `astropy`, `numpy`, `jupyter`, `matplotlib`, `sklearn`, `scipy`, `emcee`, `chainconsumer`, `getdist`, `os` and `sys`.

References

- Ajani, V., Peel, A., Pettorino, V., et al. 2020, *Phys. Rev. D*, **102**, 103531
- Ajani, V., Starck, J.-L., & Pettorino, V. 2021, *A&A*, **645**, L11
- Allys, E., Marchand, T., Cardoso, J.-F., et al. 2020, *Phys. Rev. D*, **102**, 103506
- Bartelmann, M., & Schneider, P. 2001, *Phys. Rep.*, **340**, 291
- Chambers, K. C., Magnier, E. A., Metcalfe, N., et al. 2016, ArXiv e-prints [arXiv:1612.05560]
- Cheng, S., Ting, Y.-S., Ménard, B., & Bruna, J. 2020, *MNRAS*, **499**, 5902
- Coulton, W. R., Liu, J., Madhavacheril, M. S., Böhm, V., & Spergel, D. N. 2019, *J. Cosmol. Astropart. Phys.*, **2019**, 043
- Coulton, W. R., Liu, J., McCarthy, I. G., & Osato, K. 2020, *MNRAS*, **495**, 2531
- Cragg, C., Duncan, C. A. J., Miller, L., & Alonso, D. 2022, *MNRAS*, **518**, 4909
- Dietrich, J. P., & Hartlap, J. 2010, *MNRAS*, **402**, 1049
- Fluri, J., Kacprzak, T., Refregier, A., et al. 2018, *Phys. Rev. D*, **98**, 123518
- Fluri, J., Kacprzak, T., Lucchi, A., et al. 2022, *Phys. Rev. D*, **105**, 083518
- Gaia Collaboration (Brown, A. G. A., et al.) 2018, *A&A*, **616**, A1
- Gatti, M., Chang, C., Friedrich, O., et al. 2020, *MNRAS*, **498**, 4060
- Gatti, M., Sheldon, E., Amon, A., et al. 2021, *MNRAS*, **504**, 4312
- Guinot, A., Kilbinger, M., Farrens, S., et al. 2022, *A&A*, **666**, A162
- Gwyn, S. D. J. 2008, *PASP*, **120**, 212
- Harnois-Déraps, J., Martinet, N., & Reischke, R. 2021a, *MNRAS*, **509**, 3868
- Harnois-Déraps, J., Martinet, N., Castro, T., et al. 2021b, *MNRAS*, **506**, 1623
- Hildebrandt, H., Erben, T., Kuijken, K., et al. 2012, *MNRAS*, **421**, 2355
- Hildebrandt, H., Viola, M., Heymans, C., et al. 2016, *MNRAS*, **465**, 1454
- Hoekstra, H., Viola, M., & Herbonnet, R. 2017, *MNRAS*, **468**, 3295
- Huff, E., & Mandelbaum, R. 2017, ArXiv e-prints [arXiv:1702.02600]
- Jain, B., & Waerbeke, L. V. 2000, *ApJ*, **530**, L1
- Kacprzak, T., Kirk, D., Friedrich, O., et al. 2016, *MNRAS*, **463**, 3653
- Kaiser, N., & Squires, G. 1993, *ApJ*, **404**, 441
- Kitching, T. D., Deshpande, A. C., & Taylor, P. L. 2021, *Open J. Astrophys.*, **4**, 1
- Kratochvil, J. M., Lim, E. A., Wang, S., et al. 2012, *Phys. Rev. D*, **85**, 103513
- Kruse, G., & Schneider, P. 1999, *MNRAS*, **302**, 821
- Kuchner, U., Ziegler, B., Verdugo, M., Bamford, S., & Häußler, B. 2017, *A&A*, **604**, A54
- Li, Z., Liu, J., Matilla, J. M. Z., & Coulton, W. R. 2019, *Phys. Rev. D*, **99**, 063527
- Lin, C.-A., & Kilbinger, M. 2015, *A&A*, **576**, A24
- Liu, J., & Haiman, Z. 2016, *Phys. Rev. D*, **94**, 043533
- Liu, X., Pan, C., Li, R., et al. 2015a, *MNRAS*, **450**, 2888
- Liu, J., Petri, A., Haiman, Z., et al. 2015b, *Phys. Rev. D*, **91**, 063507
- Liu, J., Bird, S., Matilla, J. M. Z., et al. 2018, *J. Cosmol. Astropart. Phys.*, **2018**, 049
- MacCrann, N., Becker, M. R., McCullough, J., et al. 2021, *MNRAS*, **509**, 3371
- Mandelbaum, R. 2018, *ARA&A*, **56**, 393
- Martinet, N., Schneider, P., Hildebrandt, H., et al. 2017, *MNRAS*, **474**, 712
- Massey, R., Hoekstra, H., Kitching, T., et al. 2012, *MNRAS*, **429**, 661
- McCarthy, I. G., Schaye, J., Bird, S., & Le Brun, A. M. C. 2016, *MNRAS*, **465**, 2936
- Miller, L., Heymans, C., Kitching, T. D., et al. 2013, *MNRAS*, **429**, 2858
- Newman, J. A., Cooper, M. C., Davis, M., et al. 2013, *ApJS*, **208**, 5
- Oguri, M., Miyazaki, S., Li, X., et al. 2021, *PASJ*, **73**, 817
- Osato, K., Shirasaki, M., & Yoshida, N. 2015, *ApJ*, **806**, 186
- Parroni, C., Cardone, V. F., Maoli, R., & Scaramella, R. 2020, *A&A*, **633**, A71
- Paulin-Henriksson, S., Amara, A., Voigt, L., Refregier, A., & Bridle, S. L. 2008, *A&A*, **484**, 67
- Particle Data Group (Patrignani, C., et al.) 2016, *Chin. Phys. C*, **40**, 100001
- Peel, A., Lin, C.-A., Lanusse, F., et al. 2017, *A&A*, **599**, A79
- Petri, A. 2016, *Astron. Comput.*, **17**, 73
- Petri, A., May, M., & Haiman, Z. 2016, *Phys. Rev. D*, **94**, 063534
- Pyne, S., & Joachimi, B. 2021, *MNRAS*, **503**, 2300
- Shan, H., Liu, X., Hildebrandt, H., et al. 2017, *MNRAS*, **474**, 1116
- Sheldon, E. S., & Huff, E. M. 2017, *ApJ*, **841**, 24
- Shirasaki, M., Moriwaki, K., Oogi, T., et al. 2021, *MNRAS*, **504**, 1825
- Spindler, A. 2018, Ph.D. Thesis, The Open University
- Takada, M., & Jain, B. 2004, *MNRAS*, **348**, 897
- Troxel, M. A., & Ishak, M. 2015, *Phys. Rep.*, **558**, 1
- Van Waerbeke, L., Benjamin, J., Erben, T., et al. 2013, *MNRAS*, **433**, 3373
- Weinberg, D. H., Mortonson, M. J., Eisenstein, D. J., et al. 2013, *Phys. Rep.*, **530**, 87
- Yang, X., Kratochvil, J. M., Wang, S., et al. 2011, *Phys. Rev. D*, **84**, 043529
- Zürcher, D., Fluri, J., Sgier, R., Kacprzak, T., & Refregier, A. 2021, *J. Cosmol. Astropart. Phys.*, **2021**, 028
- Zürcher, D., Fluri, J., Sgier, R., et al. 2022, *MNRAS*, **511**, 2075

Appendix A: Variation of the response matrix and the additive bias

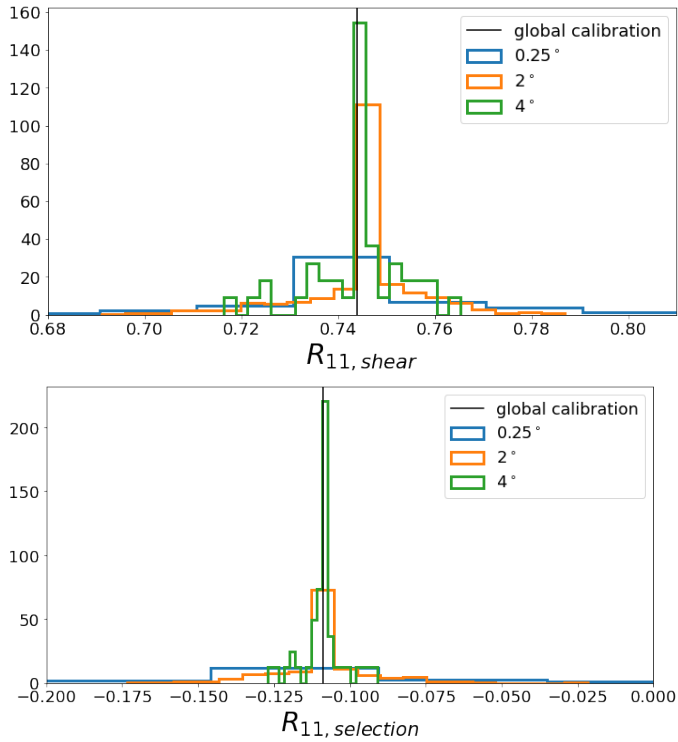


Fig. A.1. Histogram of $R_{11, \text{shear}}$ (top) and $R_{11, \text{selection}}$ (bottom) depending on the size of the local calibration.

To have a better visualization of the variation of the response matrix and the additive bias, we compute the histogram of their distribution in Figs. A.1 and A.2, respectively. In every his-

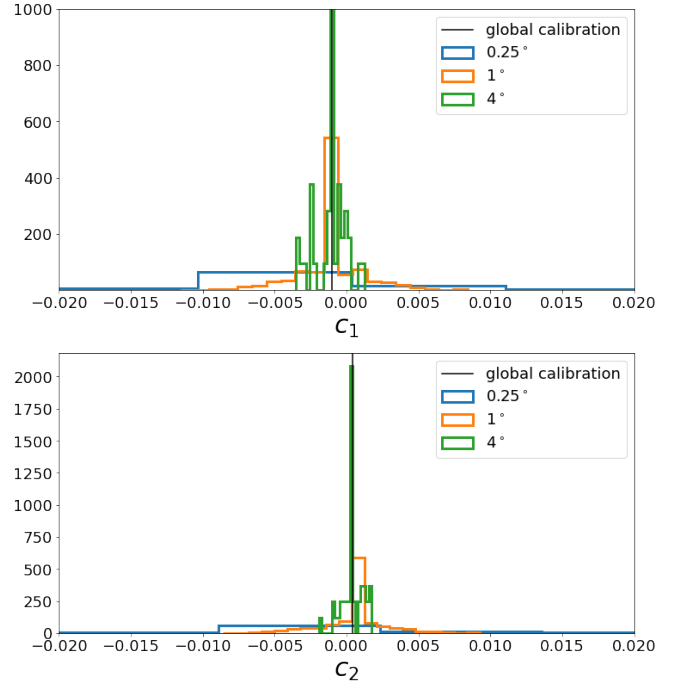


Fig. A.2. Histogram of c_1 (top) and c_2 (bottom) depending on the size of the local calibration.

togram, we add as a black line the value found by the global calibration. From those histograms, we can clearly see that when the pixels are smaller the dispersion is larger, which is because the number of galaxies in one patch is lower and therefore there is more noise. In all cases, we see that the values are spread around the mean one.

We confirm that the values are spread around the mean one, with larger dispersion at smaller scale of calibration.

Appendix B: Correlation matrix

The correlation matrix presented in Fig. B.1 is presented to see the correlation between the individual elements of the matrices.

We can observe a strong correlation (60 - 80%) between R_{11}^{shear} and R_{22}^{shear} , which is expected because if the shear com-

ports a bias, both component will be impacted. A correlation of 20 - 40% between $R_{11}^{\text{selection}}$ and $R_{22}^{\text{selection}}$ is seen for the same reason as the correlation between R_{11}^{shear} and R_{22}^{shear} . The $R_{11}^{\text{selection}}$ and $R_{22}^{\text{selection}}$ are anticorrelated (20 - 40%) with the size of the psf T_{psf} because if the psf is larger, we will miss more objects. Thus, the correlation effect is stronger.

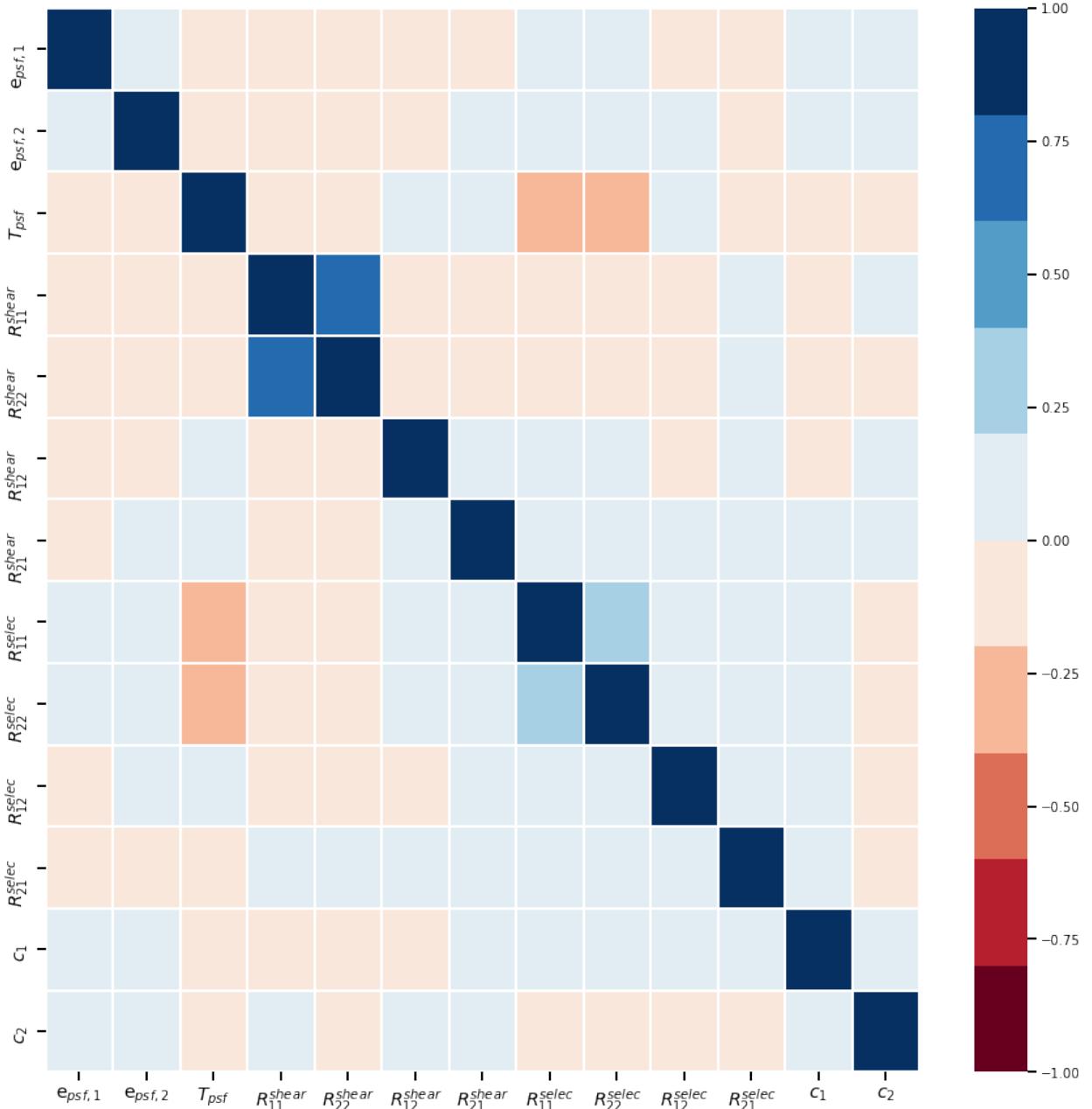


Fig. B.1. Correlation matrix of the quantities described in Sect. 5.1 or listed in Table 3. The colors indicate the amplitude of the correlation, ranging between -1 and 1 .

Appendix C: Impact of the systematics using only simulations

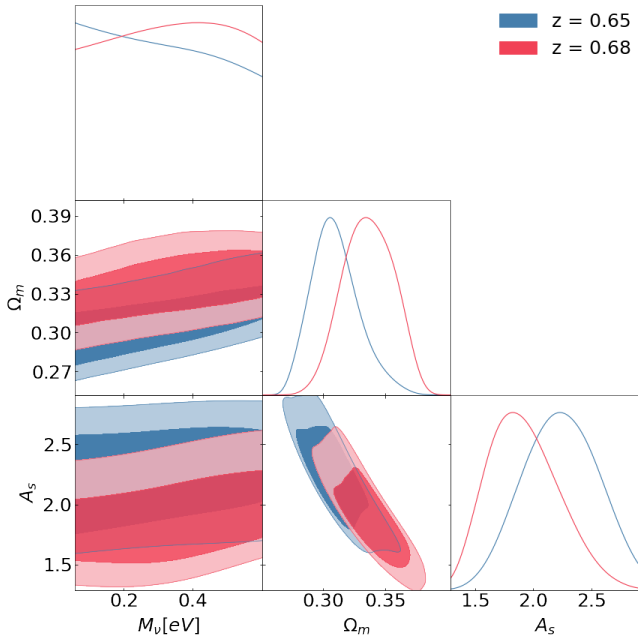


Fig. C.1. 1D and 2D marginal posteriors using as data the mean fiducial from the simulations at different redshifts. The 2D inner and outer contours show the 68% and 95.5% credible region, respectively. Ω_m shifts by $+0.026$, which corresponds to $+0.7\sigma$ when the redshift goes from 0.65 (blue) to 0.68 (red).

As the data contain different systematics that are sometimes hard to model precisely, we perform some tests, using simulations only. The data vector is the mean of peak counts at the fiducial cosmology, which we modified to include systematics. The model used in the MCMC is the simulations at $z = 0.65$ without any modification. For every test, we compare the result to a data vector without any modification (blue of every figure).

C.1. Redshift uncertainty

The impact of the redshift uncertainty is shown in Fig. C.1, where we compute the data vector at $z = 0.65$ (blue) or $z = 0.68$

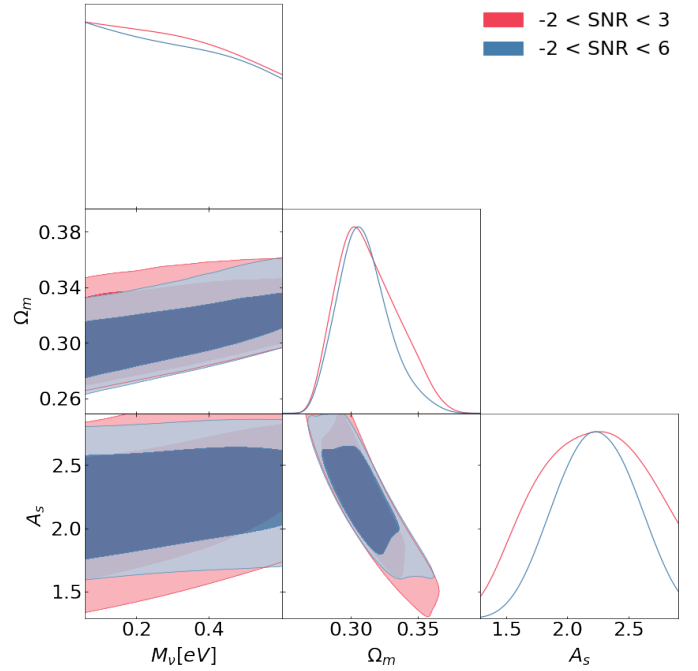


Fig. C.2. 1D and 2D marginal posteriors using as data the mean fiducial from the simulations and different ranges in S/N. The 2D inner and outer contours show the 68% and 95.5% credible region, respectively. The shear calibration is global. The case with the full S/N range (blue) is compared to the case where we use $-2 < S/N < 3$ to minimize the effect of residual bias (red). The constraints are larger when we cut the high S/N peaks.

(red), whereas the model is kept at $z = 0.65$ in both cases. Modeling with lower redshift shifts Ω_m by $+0.7\sigma$.

C.2. Intrinsic alignment and cluster member dilution

The method used here to decrease the impact of intrinsic alignment and cluster member dilution is to cut the high S/N range. In Fig. C.2, we show the constraints when we use a S/N range between -2 and 6 (blue) or a reduced S/N range between -2 and 3 (red). When the reduced range is used, the constraints are larger, which is expected because there is less information.

CONCLUSIONS AND PERSPECTIVES

My work has focused on studying the complexities of large-scale structures to enhance the constraining power of the different observables. In this final chapter, I summarise the main results obtained on the tSZ effect, the cosmic web, and peak count statistics. I then discuss the perspectives of this work and potential future directions².

Thermal Sunyaev-Zel’dovich effect The tSZ effect is a sensitive probe of baryonic matter in the Universe. Cross-correlating it with other probes, as demonstrated in Fang et al. (2024), can improve the constraining power on cosmological and astrophysical parameters. However, robust modelling of the tSZ effect is essential, and my work contributes to this objective. I have undertaken a comparison between the HMx halo model and measurements from simulations (Horizon-AGN, Horizon-noAGN, Horizon-Large and Magneticum) at different levels: angular power spectrum, power spectrum, profile, and intrinsic properties such as the polytropic index. This study highlights that the HMx halo model, while predicting the matter power spectrum correctly, fails to accurately reproduce the tSZ effect, with discrepancies increasing with redshift. This may be due to the assumptions of a halo model, such as all the matter being within symmetric spherical halos or from parameters that do not fully capture the physics. Adjusting the parameter’s value in HMx based on fits from simulations also does not allow for recovery of the measured power spectrum. The challenge arises from the degeneracies between parameters and the contribution of the intracluster medium at higher redshift. Additionally, the tSZ effect is sensitive to the cosmological model, particularly the dark energy, which affects the growth of structures. I have thus pursued an analysis of the tSZ properties within simulations run under a w CDM model. The first results indicate that the tSZ is indeed sensitive to the value of w . These results encourage continuing efforts to model the tSZ effect within different cosmologies, as it can serve as a robust probe of dark energy and baryonic matter in the Universe.

Cosmic web The distribution of matter in the Universe has evolved to form different environments: voids, walls, filaments, and knots, constituting the cosmic web structures. These structures can be classified using different methods, such as the T-web formalism employed in my work. This method relies on the number of eigenvalues of the tidal shear tensor, which is the second derivative of the gravitational potential. Using perturbation theory and Gram-Charlier expansion to account for a non-Gaussian density field, I developed a theoretical model that predicts the abundances of these different environments as a function of redshift and smoothing scale. By comparing these results with measurements from the Quijote simulations, I find a remarkable level of agreement, even at the most non-linear scale studied.

Peak counts Another method for extracting non-Gaussian information is through higher-order statistics. I focused on the peak counts statistics applied to weak lensing surveys, in particular, the UNIONS survey. I investigated the impact of different systematics and uncertainties on the constraining power of the weak lensing peak count statistics. My study revealed that several factors could influence the constraints on cosmological parameters, particularly Ω_m . These factors include the method of shear calibration, the baryonic correction applied, and the signal-to-noise ratio. These systematics and uncertainties can shift the inferred value of Ω_m in different directions. This study is a preliminary step towards understanding the impact of systematics on peak count statistics. It shows the necessity for careful calibration and consideration of systematics in future analyses.

²This chapter was expanded after the defense, following the jury’s comments.

Perspectives

From this work, different perspectives can be identified for future research. Specifically, regarding the tSZ effect, continued efforts to achieve a robust modelisation are of great interest.

One promising avenue involves developing a model capable of accurately reproducing the expected tSZ power spectrum from a given pressure profile, with or without incorporating the intrinsic properties of halos. Extending the approach of Mead et al. (2020) to jointly fit the pressure profile and specific properties on the power spectrum is particularly interesting. Additionally, exploring the impact of adopting other pressure profiles, such as the one proposed by Battaglia et al. (2012), could provide valuable insights. The overall objective is to identify optimal parameters for predicting the power spectrum while maintaining the physical coherence of the model. This procedure could also serve as a tool to investigate the correlations between different model parameters. Given that simulations include different properties, such a model could be adapted to fit a broad range of simulations and might also be useful for inferring the power spectrum from observational data. For instance, using the pressure profile of NIKA2 or those reconstructed from X-ray observations to estimate the tSZ power spectrum could offer significant benefits for cosmological analyses. Such tests could be implemented using the python package `pyhalomodel`, which allows to generate a power spectrum from a given profile.

Incorporating more realistic features, such as non-spherical halos, environmental effects, and parameters that vary with mass and redshift, could significantly improve prediction accuracy. Adding mass and redshift dependence to the parameters is a relatively straightforward extension of the model and may address some observed discrepancies. However, including environmental impacts and non-spherical halos poses greater challenges and requires more complex modelling approaches.

Another promising avenue involves incorporating the effects of the intracluster medium into the model. Developing a new framework based on cosmic web analysis, but applied to the pressure field rather than solely the density field, could be an option. Such improvements may help to reduce the discrepancies observed at higher redshifts, where the contribution of the intracluster medium becomes more significant. As an initial step, it is important to investigate the contribution of the different cosmic web environments to the total power spectrum across different redshifts. For instance, while filaments constitute a large portion of the Universe's content, the pressure scales with mass, and filaments are inherently less massive than cluster. Understanding these contributions would help prioritize efforts to enhance the model's accuracy.

Further investigation into the tSZ effect under different cosmologies, such as the w CDM model studied here, as well as more complex dark energy equations of state or modified gravity models, is also worthwhile. Analysing the key quantities that describe the modifications observed in the power spectrum is an interesting avenue. Additionally studying the constraining power of the tSZ effect on these different cosmological parameters will be beneficial. This includes probing potential degeneracies between cosmological and astrophysical parameters, which could refine our understanding of their interplay and improve the accuracy of parameter inference in cosmological analyses.

Globally, predicting the tSZ power spectrum from the basic properties of halos remains challenging but crucial. Exploring machine learning techniques could provide significant advancements in this area. For example, training models to learn halo properties and establish mappings between dark matter-only and hydrodynamic simulations could enable faster and more accurate predictions of tSZ properties.

Enhancing the constraining power of the tSZ effect through cross-correlating with other probes, as demonstrated by the 10x2pt analysis in Fang et al. (2024), presents a promising avenue. Refining models with improved priors on halo parameters could further strengthen these constraints. Applying such a framework to upcoming or existing observational data would be valuable, for instance, this approach could be used with datasets from `Euclid` and `Planck` or `SPT`.

Finally, to refine our modeling efforts, it would be interesting to run large-scale simulations with zoom-in capabilities. Such simulations would provide access to both a large volume, enabling the study of massive halos, and high resolution zoom-in regions, allowing for a more detailed analysis of the thermodynamical state of halos.

In conclusion, a comprehensive understanding of the tSZ effect is crucial for using it as a powerful cosmological and astrophysical probe, enabling the extraction of maximum information from observations. In particular, further investigation into its sensitivity to baryonic matter is crucial, as the tSZ effect provides valuable insights into the impact of baryons on large-scale structures.

Concerning the cosmic web analysis, as mentioned earlier, it would be valuable to use this framework to classify the Universe into distinct environments and study their specific properties, such as the pressure.

To further advance this work, a multi-scale analysis could be conducted to examine the properties of halos in relation to their surrounding environment.

Finally, peak count statistics represents an interesting higher-order statistics with significant potential for extracting cosmological information. Continued efforts to accurately model all relevant systematics are essential for effectively constraining cosmological parameters with observational data. However, given the lack of a theoretical framework, this approach remains computationally expensive, in particular is using hydrodynamical simulations.

Through this thesis, I have investigated different methods to study the properties of the large-scale structures of the Universe. Each method has its own advantages and limitations, making it particularly suited to addressing specific problems. In all cases, my work focused on utilizing the complexities of the large-scale structures by exploring the non-Gaussianities of the Universe and employing higher-order statistics. Understanding better these complexities is crucial for obtaining more robust cosmological and astrophysical constraints, ultimately leading to a more accurate description of the Universe's content and evolution.

APPENDIX A

PHYSICAL CONSTANTS & CONVENTIONS

Physical constant

Table A.1 lists all the physical constants used in this thesis.

Table A.1: Physical constants used in this thesis.

Quantity	Notation	Value	Unit
Speed of light	c	2.99792458×10^8	$\text{m} \cdot \text{s}^{-1}$
Proton mass	m_{p}	1.6726×10^{-27}	kg
Electron mass	m_{e}	9.1094×10^{-31}	kg
Boltzmann constant	k_{B}	1.380649×10^{-23}	$\text{J} \cdot \text{K}^{-1}$
Planck constant	h	6.62607×10^{-34}	$\text{m}^2 \cdot \text{kg} \cdot \text{s}^{-1}$
Gravitational constant	G	6.6743×10^{-11}	$\text{m}^3 \cdot \text{kg}^{-1} \cdot \text{s}^{-2}$
Solar mass	M_{\odot}	1.9884×10^{30}	kg
Parsec	pc	3.0856×10^{16}	m

Conventions

$\dot{f}(\mathbf{x}, t) \equiv \partial f(\mathbf{x}, t) / \partial t$: derivative with respect to cosmic time t .

\mathbf{x} : vector \mathbf{x} .

\mathbf{R} : matrix \mathbf{R} .

APPENDIX B

POWER SPECTRUM IN THE HORIZON-NOAGN AND HORIZON-LARGE SIMULATIONS

In Sect. 4.3.3, I compared the matter, pressure, and matter-pressure power spectra measured in the simulations to the HMx predictions. For these three power spectra, we presented the results for the `Horizon-AGN` and `Magneticum` simulations as a function of redshift. Additionally, we included the results of the `Horizon-noAGN` and `Horizon-Large` simulations at $z = 0$ to highlight the differences of physics while noting that the trend with redshift is similar. In this appendix, I present the results for the matter, pressure, and matter-pressure power spectra in the `Horizon-noAGN` and `Horizon-Large` simulations compared to the HMx prediction as a function of redshift. I also include the results of the `Horizon-AGN` and `Magneticum` simulations at $z = 0$ to support our conclusions.

In Fig. B.1, I show the matter auto-power spectrum as a function of redshift. The left panel shows the results for simulations with a box size of $100 h^{-1}$ Mpc, thus `Horizon-noAGN` in dotted line compared to HMx in solid line, and `Horizon-AGN` at $z = 0$ in dashed line. The right panel shows the results for simulations with a box size of $896 h^{-1}$ Mpc, thus `Horizon-Large` in dotted line compared to HMx in solid line, and `Magneticum` at $z = 0$ in dashed line. The power spectra go from $z = 0$ in dark blue to $z = 4.25$ in yellow. Similarly to Fig. 4.5, all simulations and redshifts show a good agreement with the prediction. At low k , there is less power than predicted, which can be explained by the cosmic variance and the size of the simulated box, whereas at high k , the differences can be explained by the resolution.

I then compare the pressure auto-power spectrum as a function of redshift in Fig. B.2. The left panel shows the results for simulations with a box size of $100 h^{-1}$ Mpc, thus `Horizon-noAGN` in dotted line compared to HMx in solid line, and `Horizon-AGN` at $z = 0$ in dashed line. The right panel shows the results for simulations with a box size of $896 h^{-1}$ Mpc, thus `Horizon-Large` in dotted line compared to HMx in solid line, and `Magneticum` at $z = 0$ in dashed line. The power spectra go from $z = 0$ in dark blue to $z = 4.25$ in yellow. Similarly to Fig. 4.6, all the simulations show a relatively good agreement at low redshift with the predictions from HMx (up to $z \sim 1$ for `Horizon-noAGN` and `Horizon-Large`). However, as the redshift increases, discrepancies become more evident. HMx predicts an excess of power at high redshift, indicating that the model's physics fails to capture the nuances present in the simulations. We also notice that the measured power spectra are flatter than the predictions.

Finally, Fig. B.3 represents the matter-pressure power spectrum as a function of redshift. The left panel shows the results for simulations with a box size of $100 h^{-1}$ Mpc, thus `Horizon-noAGN` in dotted line compared to HMx in solid line, and `Horizon-AGN` at $z = 0$ in dashed line. The right panel shows the results for simulations with a box size of $896 h^{-1}$ Mpc, thus `Horizon-Large` in dotted line compared to HMx in solid line, and `Magneticum` at $z = 0$ in dashed line. The power spectra go from $z = 0$ in dark blue to $z = 4.25$ in yellow. Similarly to Fig. 4.7, we observe a better agreement up to a comparable redshift ($z = 1.18$ for both simulations) than for the pressure auto-power spectrum. This outcome is expected as HMx has been calibrated on the matter-pressure power spectrum up to $z = 1$. Moreover, the matter auto-power spectrum agrees well across all redshifts, mitigating the discrepancies in the pressure auto-power spectrum. At higher redshifts, the discrepancies caused by the pressure auto-power spectra persist.

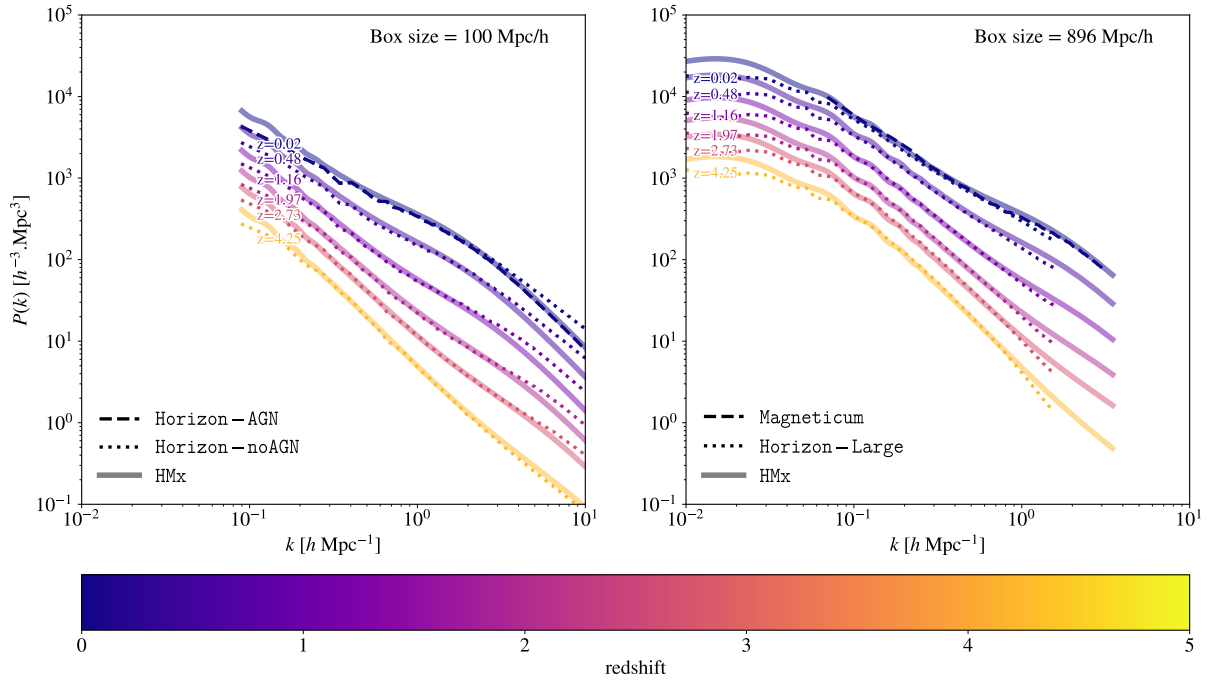


Figure B.1: Matter auto-power spectrum as a function of redshift. The left panel shows the results for simulations with a box size of $100 h^{-1}$ Mpc, thus Horizon-noAGN in dotted line compared to HMx in solid line, and Horizon-AGN at $z = 0$ in dashed line. The right panel shows the results for simulations with a box size of $896 h^{-1}$ Mpc, thus Horizon-Large in dotted line compared to HMx in solid line, and Magneticum at $z = 0$ in dashed line. The power spectra go from $z = 0$ in dark blue to $z = 4.25$ in yellow.

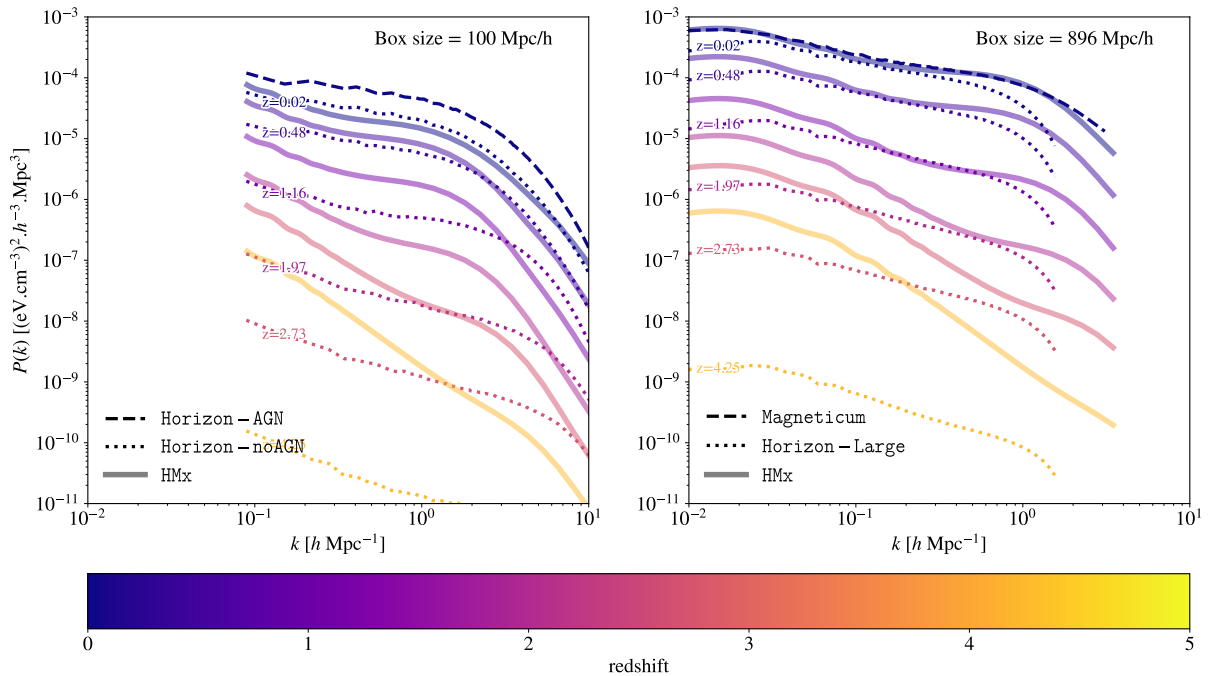


Figure B.2: Pressure auto-power spectrum as a function of redshift. The left panel shows the results for simulations with a box size of $100 h^{-1}$ Mpc, thus Horizon-noAGN in dotted line compared to HMx in solid line, and Horizon-AGN at $z = 0$ in dashed line. The right panel shows the results for simulations with a box size of $896 h^{-1}$ Mpc, thus Horizon-Large in dotted line compared to HMx in solid line, and Magneticum at $z = 0$ in dashed line. The power spectra go from $z = 0$ in dark blue to $z = 4.25$ in yellow.

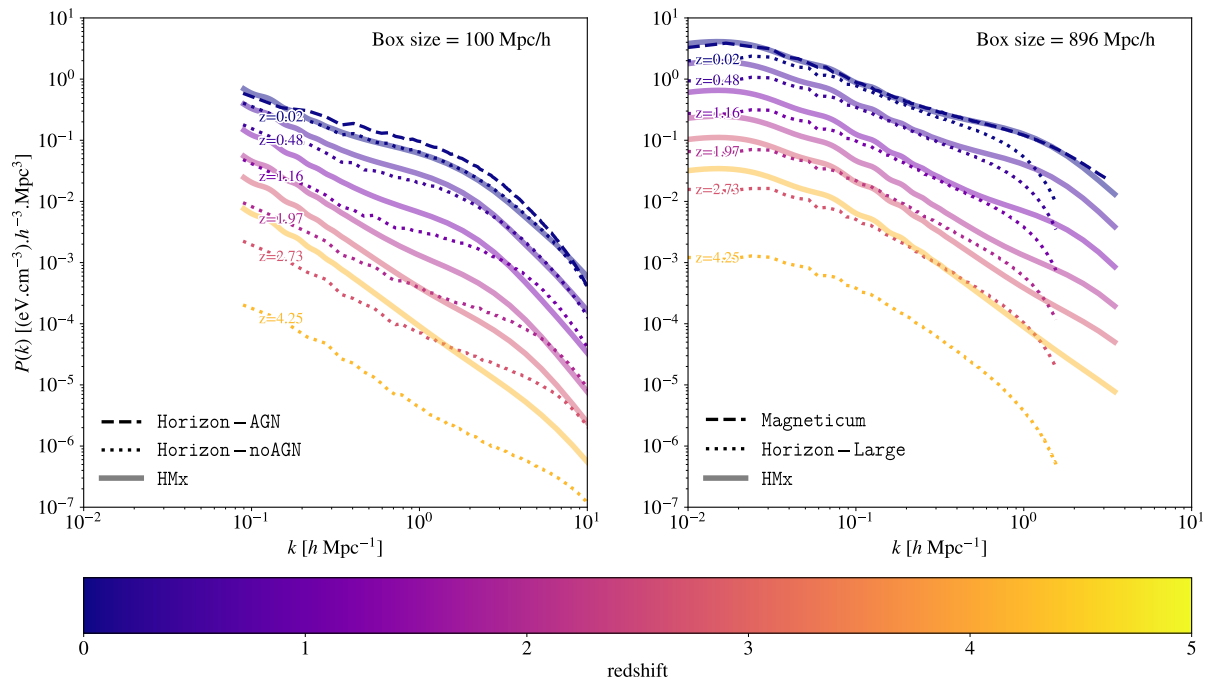


Figure B.3: Matter-pressure power spectrum as a function of redshift. The left panel shows the results for simulations with a box size of $100 h^{-1}$ Mpc, thus **Horizon-noAGN** in dotted line compared to **HMx** in solid line, and **Horizon-AGN** at $z = 0$ in dashed line. The right panel shows the results for simulations with a box size of $896 h^{-1}$ Mpc, thus **Horizon-Large** in dotted line compared to **HMx** in solid line, and **Magneticum** at $z = 0$ in dashed line. The power spectra go from $z = 0$ in dark blue to $z = 4.25$ in yellow.

ABBREVIATIONS

ΛCDM	Lambda Cold Dark Matter
2LPT	Second-order Lagrangian Perturbation Theory
ACT	Atacama Cosmology Telescope
AGN	Active Galactic Nuclei
AMR	Adaptive Mesh Refinement
BAO	Baryon Acoustic Oscillations
CDM	Cold Dark Matter
CFHTLens	Canada-France-Hawaii Telescope Lensing Survey
CIB	Cosmic Infrared Background
CIC	Cloud-In-Cell
CMB	Cosmic Microwave Background
COBE	Cosmic Background Explorer
DM	Dark Matter
DR	Data Release
GUT	Grand Unified Theory
FLRW	Friedmann-Lemaitre-Robertson-Walker
FoF	Friends-of-Friends
FoM	Figure of Merit
GAMA	Galaxy And Mass Assembly
GW	Gravitational Wave
ISW	Integrated Sachs-Wolfe
kSZ	Kinetic Sunyaev-Zel'dovich
LSS	Large Scale Structures
MFF	Mass Filling Fraction
MHD	Magneto-hydrodynamics
MMF	Multiscale Morphology Filter
MST	Minimum Spanning Tree
NFW	Navarro-Frenk-White
NIKA2	New IRAM Kids Arrays
PDF	Probability Distribution Function
RSD	Redshift Space Distortion
SDSS	Sloan Digital Sky Survey
SNR	Signal-to-Noise Ratio
SO	Spherical Overdensity
SPT	South Pole Telescope
SPH	Smooth Particle Hydrodynamic
SW	Sachs-Wolfe
SZ	Sunyaev-Zel'dovich
tSZ	Thermal Sunyaev-Zel'dovich
VFF	Volume Filling Fraction
WDM	Warm Dark Matter
WL	Weak Lensing
WMAP	Wilkinson Microwave Anisotropy Probe

- Abbott, T. M. C., Agüena, M., Alarcon, A., et al. 2023, *Phys. Rev. D* , 107, 083504
- Ajani, V., Peel, A., Pettorino, V., et al. 2020, *Phys. Rev. D* , 102, 103531
- Alcock, C. & Paczynski, B. 1979, *Nature*, 281, 358
- Alpaslan, M., Robotham, A. S. G., Driver, S., et al. 2014a, *MNRAS* , 438, 177
- Alpaslan, M., Robotham, A. S. G., Obreschkow, D., et al. 2014b, *MNRAS* , 440, L106
- Aragón-Calvo, M. A., Jones, B. J. T., van de Weygaert, R., & van der Hulst, J. M. 2007a, *Astron. Astrophys.* , 474, 315
- Aragón-Calvo, M. A., Platen, E., van de Weygaert, R., & Szalay, A. S. 2010a, *Ap. J.* , 723, 364
- Aragón-Calvo, M. A., van de Weygaert, R., & Jones, B. J. T. 2010b, *MNRAS* , 408, 2163
- Aragón-Calvo, M. A., van de Weygaert, R., Jones, B. J. T., & van der Hulst, J. M. 2007b, *Ap. J. Lett.* , 655, L5
- Arnaud, M., Pratt, G. W., Piffaretti, R., et al. 2010, *Astron. Astrophys.* , 517, A92
- Asgari, M., Mead, A. J., & Heymans, C. 2023, *The Open Journal of Astrophysics*, 6, 39
- Aubert, D., Pichon, C., & Colombi, S. 2004, *MNRAS* , 352, 376
- Ayçoberry, E., Ajani, V., Guinot, A., et al. 2023, *Astron. Astrophys.* , 671, A17
- Ayçoberry, E., Barthelemy, A., & Codis, S. 2024a, *Astron. Astrophys.* , 686, A276
- Ayçoberry, E., Pranjal R., S., Benabed, K., et al. 2024b, *arXiv e-prints*, arXiv:2409.11472
- Battaglia, N., Bond, J. R., Pfrommer, C., & Sievers, J. L. 2012, *Ap. J.* , 758, 75
- Baumann, D. 2018, *arXiv e-prints*, arXiv:1807.03098
- Beck, A. M., Murante, G., Arth, A., et al. 2016, *MNRAS* , 455, 2110
- Beckmann, R. S., Devriendt, J., Slyz, A., et al. 2017, *MNRAS* , 472, 949
- Berger, M. J. & Colella, P. 1989, *Journal of Computational Physics*, 82, 64
- Berger, M. J. & Olinger, J. 1984, *Journal of Computational Physics*, 53, 484
- Bernardeau, F. 2007, *Cosmologie - Des fondements théoriques aux observations* (EDP Sciences)
- Bernardeau, F., Colombi, S., Gaztañaga, E., & Scoccimarro, R. 2002, *Physics Reports*, 367, 1–248
- Bisnovatyi-Kogan, G. S. & Tsupko, O. Y. 2017, *Universe*, 3, 57
- Biswas, R., Alizadeh, E., & Wandelt, B. D. 2010, *Phys. Rev. D* , 82, 023002
- Bleem, L. E., Crawford, T. M., Ansarinejad, B., et al. 2022, *Ap. J. Suppl.* , 258, 36

- Bleem, L. E., Stalder, B., de Haan, T., et al. 2015, *Ap. J. Suppl.* , 216, 27
- Bolliet, B., Kusiak, A., McCarthy, F., et al. 2023, *arXiv e-prints*, arXiv:2310.18482
- Bond, J. R., Cole, S., Efstathiou, G., & Kaiser, N. 1991, *Ap. J.* , 379, 440
- Bonnaire, T., Aghanim, N., Kuruvilla, J., & Decelle, A. 2022, *Astron. Astrophys.* , 661, A146
- Bryan, G. L. & Norman, M. L. 1998, *Ap. J.* , 495, 80
- Cadiou, C., Dubois, Y., & Pichon, C. 2019, *Astron. Astrophys.* , 621, A96
- Carlesi, E., Knebe, A., Lewis, G. F., Wales, S., & Yepes, G. 2014, *Monthly Notices of the Royal Astronomical Society*, 439, 2943
- Carlstrom, J. E., Holder, G. P., & Reese, E. D. 2002, *Annu.Rev.Astron.Astrophys.*, 40, 643
- Cautun, M., van de Weygaert, R., & Jones, B. J. T. 2013, *MNRAS* , 429, 1286
- Chandran, J., Remazeilles, M., & Barreiro, R. B. 2023, *arXiv e-prints*, arXiv:2310.13516
- Chisari, N. E., Richardson, M. L. A., Devriendt, J., et al. 2018, *MNRAS* , 480, 3962
- Codis, S., Pichon, C., Devriendt, J., et al. 2012, *MNRAS* , 427, 3320
- Codis, S., Pichon, C., Pogosyan, D., Bernardeau, F., & Matsubara, T. 2013, *MNRAS* , 435, 531
- Codis-Decara, S. 2015, From cosmology to galaxy formation : what can we learn from the large-scale structure of the Universe ?, Astrophysics [astro-ph]. Université Pierre et Marie Curie - Paris VI, 2015. English. NNT : 2015PA066343. tel-01264538
- Cole, S. & Kaiser, N. 1989, *MNRAS* , 237, 1127
- Colombi, S., Pogosyan, D., & Souradeep, T. 2000, *Phys. Rev. Lett.*, 85, 5515
- Contarini, S., Verza, G., Pisani, A., et al. 2022, *Astron. Astrophys.* , 667, A162
- Coulton, W. R., Liu, J., McCarthy, I. G., & Osato, K. 2020, *Monthly Notices of the Royal Astronomical Society*, 495, 2531–2542
- Cui, W., Knebe, A., Libeskind, N. I., et al. 2019, *Monthly Notices of the Royal Astronomical Society*
- Cui, W., Knebe, A., Yepes, G., et al. 2017, *Monthly Notices of the Royal Astronomical Society*, 473, 68–79
- Davies, C. T., Harnois-Déraps, J., Li, B., et al. 2024, *arXiv e-prints*, arXiv:2406.11958
- Davis, O., Devriendt, J., Colombi, S., Silk, J., & Pichon, C. 2011, *MNRAS* , 413, 2087
- Debackere, S. N. B., Schaye, J., & Hoekstra, H. 2020, *MNRAS* , 492, 2285
- Dehnen, W. & Aly, H. 2012, *MNRAS* , 425, 1068
- Dodelson, S. & Schmidt, F. 2020, *Modern Cosmology* (Elsevier)
- Dolag, K., Jubelgas, M., Springel, V., Borgani, S., & Rasia, E. 2004, *Ap. J. Lett.* , 606, L97
- Dolag, K., Komatsu, E., & Sunyaev, R. 2016, *MNRAS* , 463, 1797
- Dolag, K. & Stasyszyn, F. 2009, *MNRAS* , 398, 1678
- Dolag, K., Vazza, F., Brunetti, G., & Tormen, G. 2005, *MNRAS* , 364, 753
- Doroshkevich, A. G. 1970, *Astrophysics*, 6, 320
- Dubois, Y. & Commerçon, B. 2016, *Astron. Astrophys.* , 585, A138
- Dubois, Y., Commerçon, B., Marcowith, A., & Brahim, L. 2019, *Astron. Astrophys.* , 631, A121
- Dubois, Y., Devriendt, J., Slyz, A., & Teyssier, R. 2012, *MNRAS* , 420, 2662
- Dubois, Y., Peirani, S., Pichon, C., et al. 2016, *MNRAS* , 463, 3948

- Dubois, Y., Pichon, C., Welker, C., et al. 2014, *MNRAS* , 444, 1453
- Dubois, Y., Rodríguez Montero, F., Guerra, C., et al. 2024, [arXiv e-prints](#), arXiv:2402.18515
- Dubois, Y. & Teyssier, R. 2008a, *Astron. Astrophys.* , 477, 79
- Dubois, Y. & Teyssier, R. 2008b, in Astronomical Society of the Pacific Conference Series, Vol. 390, Pathways Through an Eclectic Universe, ed. J. H. Knapen, T. J. Mahoney, & A. Vazdekis, 388
- Duffy, A. R., Schaye, J., Kay, S. T., & Dalla Vecchia, C. 2008, *MNRAS* , 390, L64
- Eifler, T., Fang, X., Krause, E., et al. 2024, [arXiv e-prints](#), arXiv:2411.04088
- Eifler, T., Krause, E., Schneider, P., & Honscheid, K. 2014, *MNRAS* , 440, 1379
- Einasto, M., Lietzen, H., Tempel, E., et al. 2014, *Astron. Astrophys.* , 562, A87
- Euclid Collaboration, Ajani, V., Baldi, M., et al. 2023, *Astron. Astrophys.* , 675, A120
- Falck, B. L., Neyrinck, M. C., & Szalay, A. S. 2012, *Ap. J.* , 754, 126
- Fang, X., Krause, E., Eifler, T., et al. 2024, *MNRAS* , 527, 9581
- Fard, M. A., Taamoli, S., & Baghran, S. 2019, *MNRAS* , 489, 900
- Fedeli, C. 2014, *JCAP*, 2014, 028
- Forero-Romero, J. E., Hoffman, Y., Gottlöber, S., Klypin, A., & Yepes, G. 2009, *Monthly Notices of the Royal Astronomical Society*, 396, 1815–1824
- Fromang, S., Hennebelle, P., & Teyssier, R. 2006, *Astron. Astrophys.* , 457, 371
- Gao, Z., Vlah, Z., & Challinor, A. 2024, *JCAP*, 2024, 003
- Gay, C., Pichon, C., & Pogosyan, D. 2012, *Physical Review D*, 85
- Gil-Marín, H., Schmidt, F., Hu, W., Jimenez, R., & Verde, L. 2011, *JCAP*, 2011, 019
- Gnedin, N. Y. 1995, *Ap. J. Suppl.* , 97, 231
- Guinot, A., Kilbinger, M., Farrens, S., et al. 2022, *Astronomy & Astrophysics*, 666, A162
- Haardt, F. & Madau, P. 1996, *Ap. J.* , 461, 20
- Hahn, O., Carollo, C. M., Porciani, C., & Dekel, A. 2007b, *Monthly Notices of the Royal Astronomical Society*, 381, 41
- Hahn, O., Porciani, C., Carollo, C. M., & Dekel, A. 2007a, *Monthly Notices of the Royal Astronomical Society*, 375, 489
- Hanser, C., Adam, R., Ade, P., et al. 2023, [arXiv e-prints](#), arXiv:2310.07400
- Harnois-Déraps, J., Martinet, N., & Reischke, R. 2021, *Monthly Notices of the Royal Astronomical Society*, 509, 3868
- Hasan, F., Burchett, J. N., Abeyta, A., et al. 2023, *Ap. J.* , 950, 114
- Hill, J. C. & Pajer, E. 2013, *Phys. Rev. D* , 88, 063526
- Hinshaw, G., Larson, D., Komatsu, E., et al. 2013, *Ap. J. Suppl.* , 208, 19
- Hirschmann, M., Dolag, K., Saro, A., et al. 2014, *MNRAS* , 442, 2304
- Hockney, R. W. & Eastwood, J. W. 1981, *Computer Simulation Using Particles* (CRC Press)
- Hoffman, Y., Metuki, O., Yepes, G., et al. 2012, *Monthly Notices of the Royal Astronomical Society*, 425, 2049–2057
- Hoffmann, K., Bel, J., & Gaztañaga, E. 2017, *MNRAS* , 465, 2225
- Holz, D. E. & Perlmutter, S. 2012, *Ap. J. Lett.* , 755, L36
- Huchet, A. & Melin, J.-B. 2024, [arXiv e-prints](#), arXiv:2402.18346

- Huff, E. & Mandelbaum, R. 2017, [arXiv e-prints](#), [arXiv:1702.02600](#)
- Jasche, J., Leclercq, F., & Wandelt, B. D. 2015, *JCAP*, 2015, 036
- Jenkins, A. 2010, *MNRAS*, 403, 1859
- Juszkiewicz, R., Bouchet, F. R., & Colombi, S. 1993, *The Astrophysical Journal*, 412, L9
- Kaiser, N. 1984, *Ap. J. Lett.*, 284, L9
- Khokhlov, A. 1998, *Journal of Computational Physics*, 143, 519
- Kilbinger, M. 2015, *Reports on Progress in Physics*, 78, 086901
- Kimm, T. & Cen, R. 2014, *Ap. J.*, 788, 121
- Komatsu, E. & Kitayama, T. 1999, *Ap. J. Lett.*, 526, L1
- Komatsu, E. & Seljak, U. 2001, *MNRAS*, 327, 1353
- Komatsu, E. & Seljak, U. 2002, *MNRAS*, 336, 1256
- Komatsu, E., Smith, K. M., Dunkley, J., et al. 2011, *Ap. J. Suppl.*, 192, 18
- Kraljic, K., Arnouts, S., Pichon, C., et al. 2018, *MNRAS*, 474, 547
- Krause, E. & Eifler, T. 2017, *MNRAS*, 470, 2100
- Kravtsov, A. V., Klypin, A. A., & Khokhlov, A. M. 1997, *Ap. J. Suppl.*, 111, 73
- Laureijs, R., Amiaux, J., Arduini, S., et al. 2011, [arXiv e-prints](#), [arXiv:1110.3193](#)
- Lavaux, G. & Wandelt, B. D. 2010, *MNRAS*, 403, 1392
- Le Brun, A. M. C., McCarthy, I. G., & Melin, J.-B. 2015, *MNRAS*, 451, 3868
- Le Brun, A. M. C., McCarthy, I. G., Schaye, J., & Ponman, T. J. 2014, *MNRAS*, 441, 1270
- Leclercq, F., Jasche, J., Lavaux, G., Wandelt, B., & Percival, W. 2017, *JCAP*, 2017, 049
- Leclercq, F., Lavaux, G., Jasche, J., & Wandelt, B. 2016, *Journal of Cosmology and Astroparticle Physics*, 2016, 027
- Lee, B. K. K., Coulton, W. R., Thiele, L., & Ho, S. 2022, *MNRAS*, 517, 420
- Lee, J. & Park, D. 2009, *Ap. J. Lett.*, 696, L10
- Libeskind, N. I., van de Weygaert, R., Cautun, M., et al. 2017, *Monthly Notices of the Royal Astronomical Society*, 473, 1195–1217
- Limber, D. N. 1954, *Ap. J.*, 119, 655
- Liu, J., Bird, S., Matilla, J. M. Z., et al. 2018, *Journal of Cosmology and Astroparticle Physics*, 2018, 049–049
- Lokas, E. L., Juszkiewicz, R., Weinberg, D. H., & Bouchet, F. R. 1995, *Monthly Notices of the Royal Astronomical Society*, 274, 730
- Ma, Y.-Z., Van Waerbeke, L., Hinshaw, G., et al. 2015, *JCAP*, 2015, 046
- Maniyar, A., Béthermin, M., & Lagache, G. 2021, *Astron. Astrophys.*, 645, A40
- Martel, H. & Shapiro, P. R. 1998, *MNRAS*, 297, 467
- Massonneau, W., Volonteri, M., Dubois, Y., & Beckmann, R. S. 2023, *Astron. Astrophys.*, 670, A180
- McCarthy, I. G., Le Brun, A. M. C., Schaye, J., & Holder, G. P. 2014, *MNRAS*, 440, 3645
- McCarthy, I. G., Salcido, J., Schaye, J., et al. 2023, *MNRAS*, 526, 5494
- McCarthy, I. G., Schaye, J., Bird, S., & Le Brun, A. M. C. 2017, *MNRAS*, 465, 2936

- Mead, A. 2015, HMcode: Halo-model matter power spectrum computation, Astrophysics Source Code Library, record ascl:1508.001
- Mead, A. J., Tröster, T., Heymans, C., Van Waerbeke, L., & McCarthy, I. G. 2020, *Astron. Astrophys.* , **641**, A130
- Mead, A. J. & Verde, L. 2021, *MNRAS* , **503**, 3095
- Metuki, O., Libeskind, N. I., Hoffman, Y., Crain, R. A., & Theuns, T. 2015, *MNRAS* , **446**, 1458
- Mishra, S. S., Sahni, V., & Starobinsky, A. A. 2021, *JCAP* , **2021**, 075
- Mo, H. J. & White, S. D. M. 1996, *MNRAS* , **282**, 347
- Moser, E., Battaglia, N., Nagai, D., et al. 2022, *Ap. J.* , **933**, 133
- Navarro, J. F., Frenk, C. S., & White, S. D. M. 1997a, *Ap. J.* , **490**, 493
- Navarro, J. F., Frenk, C. S., & White, S. D. M. 1997b, *Ap. J.* , **490**, 493
- Nojiri, S., Odintsov, S. D., & Tsujikawa, S. 2005, *Phys. Rev. D* , **71**, 063004
- Novikov, D., Colombi, S., & Doré, O. 2006, *MNRAS* , **366**, 1201
- Nuza, S. E., Kitaura, F.-S., Heß, S., Libeskind, N. I., & Müller, V. 2014, *MNRAS* , **445**, 988
- Oguri, M. 2010, *Publications of the Astronomical Society of Japan* , **62**, 1017
- Osato, K., Shirasaki, M., Miyatake, H., et al. 2020, *MNRAS* , **492**, 4780
- Osato, K. & Takada, M. 2021, *Phys. Rev. D* , **103**, 063501
- Page, L., Hinshaw, G., Komatsu, E., et al. 2007, *Ap. J. Suppl.* , **170**, 335
- Paillass, E., Cuesta-Lazaro, C., Percival, W. J., et al. 2024, *MNRAS* , **531**, 898
- Pandey, S., Lehman, K., Baxter, E. J., et al. 2023, *MNRAS* , **525**, 1779
- Partridge, R. B. & Peebles, P. J. E. 1967, *Ap. J.* , **148**, 377
- Peebles, P. J. E. 1973, *Ap. J.* , **185**, 413
- Peebles, P. J. E. 1980, *The large-scale structure of the universe* (Princeton University Press)
- Pen, U.-L. 1995, *Ap. J. Suppl.* , **100**, 269
- Penzias, A. A. & Wilson, R. W. 1965, *Ap. J.* , **142**, 419
- Perotto, L., Adam, R., Ade, P., et al. 2023, *arXiv e-prints*, arXiv:2310.04553
- Peter, P. & Uzan, J.-P. 2013, *Primordial Cosmology* (Oxford Graduate Texts)
- Pfister, H., Volonteri, M., Dubois, Y., Dotti, M., & Colpi, M. 2019, *MNRAS* , **486**, 101
- Planck Collaboration, Adam, R., Ade, P. A. R., et al. 2016a, *Astron. Astrophys.* , **594**, A10
- Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2014, *Astron. Astrophys.* , **571**, A16
- Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2013, *Astron. Astrophys.* , **550**, A131
- Planck Collaboration, Aghanim, N., Akrami, Y., et al. 2020a, *Astron. Astrophys.* , **641**, A1
- Planck Collaboration, Aghanim, N., Akrami, Y., et al. 2020b, *Astron. Astrophys.* , **641**, A6
- Planck Collaboration, Aghanim, N., Arnaud, M., et al. 2016b, *Astron. Astrophys.* , **594**, A22
- Planck Collaboration, Akrami, Y., Arroja, F., et al. 2020c, *Astron. Astrophys.* , **641**, A10
- Pogosyan, D., Gay, C., & Pichon, C. 2009, *Phys. Rev. D* , **80**, 081301
- Pogosyan, D., Pichon, C., Gay, C., et al. 2009, *MNRAS* , **396**, 635
- Poudel, A., Heinämäki, P., Tempel, E., et al. 2017, *Astron. Astrophys.* , **597**, A86

- Press, W. H. & Schechter, P. 1974, *Ap. J.* , 187, 425
- Prunet, S., Pichon, C., Aubert, D., et al. 2008, *Ap. J. Suppl.* , 178, 179
- Puget, J. L., Abergel, A., Bernard, J. P., et al. 1996, *Astron. Astrophys.* , 308, L5
- Rasera, Y. & Teyssier, R. 2006, *Astron. Astrophys.* , 445, 1
- Refregier, A., Komatsu, E., Spergel, D. N., & Pen, U.-L. 2000, *Phys. Rev. D* , 61, 123001
- Refregier, A. & Teyssier, R. 2002, *Phys. Rev. D* , 66, 043002
- Rosdahl, J., Blaizot, J., Aubert, D., Stranex, T., & Teyssier, R. 2013, *MNRAS* , 436, 2188
- Rudd, D. H., Zentner, A. R., & Kravtsov, A. V. 2008, *Ap. J.* , 672, 19
- Ruppin, F., Mayet, F., Pratt, G. W., et al. 2018, *Astron. Astrophys.* , 615, A112
- Sachs, R. K. & Wolfe, A. M. 1967, *Ap. J.* , 147, 73
- Sadeh, S., Rephaeli, Y., & Silk, J. 2007, *MNRAS* , 380, 637
- Schaye, J., Dalla Vecchia, C., Booth, C. M., et al. 2010, *MNRAS* , 402, 1536
- Schneider, A. & Teyssier, R. 2015, *JCAP*, 2015, 049
- Sefusatti, E., Crocce, M., Pueblas, S., & Scoccimarro, R. 2006, *Phys. Rev. D* , 74, 023522
- Seljak, U., Burwell, J., & Pen, U.-L. 2001, *Phys. Rev. D* , 63, 063001
- Shandarin, S. F. & Zeldovich, Y. B. 1989, *Reviews of Modern Physics*, 61, 185
- Sheth, R. K., Mo, H. J., & Tormen, G. 2001, *MNRAS* , 323, 1
- Sheth, R. K. & Tormen, G. 1999, *MNRAS* , 308, 119
- Sousbie, T. 2011, *Monthly Notices of the Royal Astronomical Society*, 414, 350
- Sousbie, T., Pichon, C., Colombi, S., Novikov, D., & Pogosyan, D. 2008, *MNRAS* , 383, 1655
- Spacek, A., Richardson, M. L. A., Scannapieco, E., et al. 2018, *Ap. J.* , 865, 109
- Springel, V. & Hernquist, L. 2003, *MNRAS* , 339, 289
- Springel, V., White, M., & Hernquist, L. 2001, *Ap. J.* , 549, 681
- Springel, V., White, S. D. M., Jenkins, A., et al. 2005, *Nature*, 435, 629
- Sunyaev, R. A. & Zeldovich, Y. B. 1970, *Astrophysics and Space Science*, 7, 3
- Sunyaev, R. A. & Zeldovich, Y. B. 1980, *MNRAS* , 190, 413
- Sutherland, R. S. & Dopita, M. A. 1993, *Ap. J. Suppl.* , 88, 253
- Takada, M. & Jain, B. 2004, *MNRAS* , 348, 897
- Tegmark, M., Blanton, M. R., Strauss, M. A., et al. 2004, *Ap. J.* , 606, 702
- Tempel, E., Stoica, R. S., Martínez, V. J., et al. 2014, *MNRAS* , 438, 3465
- Teyssier, R. 2002, *Astron. Astrophys.* , 385, 337
- Teyssier, R., Fromang, S., & Dormy, E. 2006, *Journal of Computational Physics*, 218, 44
- Tinker, J., Kravtsov, A. V., Klypin, A., et al. 2008, *Ap. J.* , 688, 709
- Tinker, J. L., Robertson, B. E., Kravtsov, A. V., et al. 2010, *Ap. J.* , 724, 878
- Tornatore, L., Borgani, S., Springel, V., et al. 2003, *MNRAS* , 342, 1025
- Tornatore, M., Maier, G., & Pattavina, A. 2006, *Journal of Optical Networking*, 5, 858
- Tröster, T., Mead, A. J., Heymans, C., et al. 2022, *Astron. Astrophys.* , 660, A27

- Van Waerbeke, L., Hinshaw, G., & Murray, N. 2014, *Phys. Rev. D* , **89**, 023508
- Villaescusa-Navarro, F. 2018, Pylians: Python libraries for the analysis of numerical simulations, Astrophysics Source Code Library, record ascl:1811.008
- Villaescusa-Navarro, F., Hahn, C., Massara, E., et al. 2020, *Ap. J. Suppl.* , **250**, 2
- Waizmann, J. C. & Bartelmann, M. 2009, *Astron. Astrophys.* , **493**, 859
- Wambsganss, J. 1998, *Living Reviews in Relativity*, **1**, 12
- Wiersma, R. P. C., Schaye, J., Theuns, T., Dalla Vecchia, C., & Tornatore, L. 2009, *MNRAS* , **399**, 574
- Zel'dovich, Y. B. 1970, *Astron. Astrophys.* , **5**, 84
- Zentner, A. R., Rudd, D. H., & Hu, W. 2008, *Phys. Rev. D* , **77**, 043507
- Zhang, Y., Wang, W., Han, J., et al. 2024, *Ap. J.* , **961**, 223