



HAL
open science

Euclid: Construction du catalogue d'amas de galaxies détectés par effet de lentille gravitationnelle faible

Gavin Leroy

► **To cite this version:**

Gavin Leroy. Euclid: Construction du catalogue d'amas de galaxies détectés par effet de lentille gravitationnelle faible. Cosmologie et astrophysique extra-galactique [astro-ph.CO]. Université Paris-Saclay, 2023. Français. NNT : 2023UPASP158 . tel-04536654

HAL Id: tel-04536654

<https://theses.hal.science/tel-04536654>

Submitted on 8 Apr 2024

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.

Euclid : Construction du catalogue d'amas de galaxies détectés par effet de lentille gravitationnelle faible

*Galaxy Cluster detection with Weak Lensing : Towards a
Euclid weak lensing selected galaxy cluster catalogue*

Thèse de doctorat de l'université Paris-Saclay

École doctorale n°127, Astronomie et Astrophysique d'Île-de-France (AAIF)
Spécialité de doctorat : Astronomie et Astrophysique
Graduate School : Physique. Référent : Faculté des sciences d'Orsay

Thèse préparée dans l'unité de recherche **Astrophysique, Instrumentation et Modélisation de Paris Saclay** (Université Paris-Saclay, CNRS, CEA) , sous la direction de **Gabriel W. PRATT**, Directeur de recherche, et la co-direction de **Sandrine PIRES**, Ingénieure-Chercheuse

Thèse soutenue à Paris-Saclay, le 7 décembre 2023, par

Gavin LEROY

Composition du jury

Membres du jury avec voix délibérative

Sophie MAUROGORDATO Directrice de recherche, OCA, Université Côte d'Azur, CNRS	Présidente
Frédéric COURBIN Professeur, École polytechnique fédérale de Lausanne	Rapporteur & Examineur
Étienne POINTECOUTEAU Directeur de recherche, IRAP, Université de Toulouse, CNRS	Rapporteur & Examineur
James BARTLETT Professeur, APC, Université Paris Cité	Examineur
Eric SLEZAK Astronome, OCA, Université Côte d'Azur	Examineur

Titre : Euclid : Construction du catalogue d'amas de galaxies détectés par effet de lentille gravitationnelle faible

Mots clés : Amas de galaxies ; Halos ; Cosmologie ; Lentille gravitationnelle faible

Résumé : Dans l'Univers, la matière s'agrège sous l'influence de la gravité, formant des structures de plus en plus massives. Aboutissement de ce processus hiérarchique, les amas de galaxies trônent au sommet de cette évolution et jouent ainsi un rôle crucial pour mieux comprendre l'origine et l'évolution des structures dans l'Univers. Composés à 85 % de matière noire, les amas de galaxies sont également des sondes formidables pour cartographier la matière noire et de par leurs propriétés, contraindre les paramètres du modèle cosmologique standard. C'est pourquoi leur étude revêt une grande importance pour affiner notre compréhension fondamentale de l'Univers et motive alors tout particulièrement leur détection.

Étant directement sensible au potentiel gravitationnel, l'effet de lentille gravitationnelle faible est l'outil idéal pour la détection des amas de galaxies par la totalité de la matière qui les compose incluant la matière noire. Avec le développement rapide des détecteurs, la sensibilité et l'étendue de la zone couverte par les relevés optiques grand champ actuels et futurs, l'effet de lentille gravitationnelle peut désormais identifier un nombre croissant d'amas de galaxies. Cela motive considérablement le développement de nouvelles méthodes de détection pour analyser plus rapidement et plus efficacement le large volume de données à venir.

Cette thèse présente une nouvelle méthode de détection multi-échelle des amas de galaxies par l'effet de lentille faible, afin de produire le catalogue d'amas pour EUCLID. En premier lieu, nous introduisons cette nouvelle approche multi-échelle basée sur l'application de la décomposition en ondelettes aux cartes de matière noire. Ensuite, nous nous intéressons à sa validation sur des données simulées avec des propriétés s'approchant de nos attentes avec le télescope spatial EUCLID. Nous comparons ainsi notre méthode avec les méthodes les plus couramment utilisées et les plus performantes. Pour cette comparaison, nous introduisons une nouvelle méthode d'association qui prend en compte le rapport signal sur bruit théorique de la détection par effet de lentille faible et la taille du filtre considéré. Enfin, après ces comparaisons sur ces données simulées, nous présentons les premiers résultats prometteurs sur les données réelles du relevé Hyper-Suprime Cam et les défis que l'application à ces données apporte.

Par cette étude, nous montrons que notre méthode multi-échelle est plus performante que les autres méthodes standards et qu'elle est aussi plus adaptée pour le traitement des relevés optiques actuels et futurs. Cette étude ouvre ainsi la voie à la production de catalogues d'amas de galaxies sélectionnés par l'effet de lentille gravitationnelle faible.

Title : Galaxy Cluster detection with Weak Lensing : Towards a Euclid weak lensing selected galaxy cluster catalogue

Keywords : Galaxies clusters ; Haloes ; Cosmology ; Weak gravitational lensing

Abstract : In the Universe, matter aggregates hierarchically under the effect of gravity, forming progressively more massive structures. At the pinnacle of this process are galaxy clusters, the formation and evolution of which play a pivotal role in unraveling the origins and development of structures within the Universe. Comprising 85% dark matter, galaxy clusters also serve as potent probes for mapping dark matter and, through the evolution of their properties over time, constraining the parameters of the standard cosmological model. Their study is therefore essential for deepening our fundamental understanding of the Universe, and makes their detection crucial.

Being directly sensitive to the gravitational potential, weak gravitational lensing is an ideal tool for detecting galaxy clusters via the total matter content, including dark matter. With the increasing sensitivity and the expanded area covered by on-going and future wide-field optical surveys, and the development of ever-larger detectors, deep wide-field imaging observations of large areas of the sky are now available. Gravitational lensing can now be used to identify an ever-growing number of galaxy clusters in these images. This strongly motivates the development of

faster and more efficient detection methods to analyze the huge volume of upcoming data.

This thesis presents a new multiscale detection algorithm for galaxy clusters using the weak lensing effect, developed with the ultimate aim to produce the cluster catalog for EUCLID. First, we introduce this multi-scale approach based on the application of wavelet decomposition to dark matter maps. Then, we validate it on simulated data with properties approaching our expectations from the EUCLID space telescope. We compare our method to the most commonly used and best-performing methods in the literature. For this comparison, we introduce a new association procedure that takes into account the theoretical signal-to-noise ratio of weak lensing detections and the size of the filter. Finally, after these comparisons on simulated data, we present the first promising results on real data from the Hyper-Suprime Cam survey and the challenges that application to this data brings.

Our findings demonstrate that this multi-scale method outperforms the standard methods and thus, is well-suited to on-going and future optical surveys. This study paves the way for the production of weak gravitational lensing selected galaxy clusters catalogues from new and upcoming surveys.

Remerciements

Parmi les nombreux rêves d'enfance qui m'habitaient, l'alpinisme était l'un des plus étonnants pour le petit gars du nord que j'étais. Je voyais dans les grands explorateurs comme Ed. Hillary qui conquiert l'Everest, le dépassement de soi, le courage et la détermination. Mon vertige aura finalement raison de ce rêve jusqu'à ce que cette pensée me revienne durant la rédaction de ce manuscrit. Si Ed. Hillary a conquis l'Everest c'est par l'aide extraordinaire qu'il a reçu et notamment par son guide, Tenzing Norgway. Cette thèse a été pour moi à l'image d'une grande ascension : faite de hauts et de bas, et parsemée d'obstacles à surmonter. Ce voyage n'aurait pas été possible sans l'aide de mes Tenzing Norgway à moi, vous mes guides et compagnons de voyage.

C'est avec une reconnaissance profonde que je remercie donc mes guides et directeurs pour cette thèse : Gabriel W. Pratt pour avoir orienté ce travail avec une clairvoyance et une vision du domaine que j'espère acquérir un jour, et Sandrine Pires pour avoir su me guider dans le monde des ondelettes avec compétence, pour son investissement constant et pour son enthousiasme communicatif. Je les remercie également pour leur disponibilité malgré les difficultés, leur soutien et pour avoir poussé ce travail vers une analyse toujours plus complète et fine. Enfin, je les remercie vivement d'avoir corrigé avec patience les nombreux élans lyriques dans les premières versions de ce manuscrit.

Je remercie ensuite les rapporteurs de cette thèse, Frédéric Courbin et Etienne Pointecouteau, pour avoir accepté de jouer ce rôle et pour les remarques et conseils afin d'améliorer ce manuscrit. Je tiens également à remercier Sophie Maurogordato, James Bartlett et Eric Slezack, pour avoir accepté d'être membre de mon jury et pour leur retours et discussions plus qu'appréciés.

Je remercie ceux qui ont permis par leur expertise de faire de cette thèse ce qu'elle est aujourd'hui : merci à Jérôme B., Fabio A. et Jean-Baptiste M. pour leur retours qui ont permis de constamment améliorer les résultats présentés dans ce manuscrit.

Je remercie infiniment toute ma famille pour leur soutien, leur confiance et leur compréhension. En particulier, je remercie mes parents Sabine et Vincent, ainsi que ma soeur, Emilie, pour leur présence et leur aide précieuse tout au long de ces trois années, notamment pour la relecture de ce manuscrit (malgré son contenu qualifié par moments d'énigmatique). Je tiens à remercier tout particulièrement Dr. Madeline F. pour son soutien indéfectible en toutes circonstances (même lors des nuits blanches d'écriture), pour sa compréhension infinie et pour m'avoir permis de devenir une meilleure personne.

Je remercie aussi tous mes amis sans qui cette thèse n'aurait pas eu la même saveur. Merci à mes amis de toujours, Florencia D. (thank you so much), Eline B., Béryl D., Raphaël B., Louis D., Gautier D., Chloé S., Sylvain O., Marion B., Dr. Laura D., Marine B., Antoine D. et mes amis de Faidherbes pour leur encouragements et la confiance qu'ils m'ont témoigné avant même de commencer cette aventure. Merci à mes amis de Telecom, Dr. Pierre D., Pauline B., Tom B., Noha B., Dr. Raphaël B., Bastien A., Bastien Z., Paul V., Marguerite A., Maxime M., Louise L., Jessica M. et Dr. Grégoire P. pour tous les

moments d'éclats, de joies et de rires qui ont su éclairer mon voyage. Enfin, un grand merci à mes amis rencontrés le long du chemin, Dr. Sara M., Sarah A., Jeremy G., Sébastien H., Achille P., Bertrand A., Aurore D., Baptiste B., Paul E., Elise S., Marina D., Dr. Louise S., Dr. Agathe V., et Axelle F. pour leur compréhension pour toutes les invitations refusées malgré moi, leur joie et leur bienveillance. Je remercie aussi Dr. Camille M. et Dr. Yaël P. pour leur relecture assidue du manuscrit. Ces remerciements ne seraient pas complets sans une mention spéciale pour Misty, Ruby et Myrtille pour avoir rempli de bonheur ces trois années.

Je remercie également mes amis docteurs ou futurs docteurs du DAP tant pour m'avoir montré la voie que pour l'avoir illuminée : Nicolas C. pour sa co-direction avec brio du célèbre Cluster Club, Loris C., Alexandre H., Fabian, Felipe M., Jack B., Timothée T., Valentin, Adnan AA., Dr. Achrene D., pour nos échanges, nos soirées et repas doctorants, Dr. Denise L., Dr. Benjamin R., Dr. Tobias L., pour leur ambiance chaleureuse lors de mon passage à Cosmostat, Dr. Fadi N., pour notre humour unique, et mes compagnons de courses Dr. Anne-Cécile B., Clara P., Dr. Hugo A., Maxime T., qui ont embelli et supporté mes ambitions de courses parfois ridicules et mes amis de l'école d'été Euclid. Je tiens à remercier tout particulièrement mes amies et camarades de bureau Leila G. pour sa joie de vivre, ses fromages et tous les thés et biscuits qui m'ont nourri dans les moments de stress et Julia L. sans qui les travaux de ma thèse aurait été fini 1 an plus tôt, et qui par leur joie et leur bonne humeur ont fait de ces années de thèse des années de bonheur passées presque un peu trop vite.

Je remercie tous les doctorants, post-docs et chercheurs avec qui j'ai partagé les petits déjeuners, mes pauses, et je tiens aussi à mentionner le plaisir que j'ai eu de travailler au sein du département d'Astrophysique et du Département d'Electronique des Détecteurs et d'Informatique et dont je remercie sincèrement tous les membres. Je remercie l'école doctorale ED 127 Astronomie et Astrophysique pour avoir répondu aux nombreuses questions que j'ai pu avoir.

Je remercie kinés et médecins pour avoir rendu possible ces trois années qui furent aussi pour moi celles où je me suis le plus blessé, paradoxalement en étant le moins actif.

C'est grâce à vous tous que j'ai eu la chance de gravir mon propre Everest.

*
* *

Liste des acronymes

Sont présentés ici les acronymes utilisés tout au long de ce manuscrit. À la version en français s'ajoute la version anglaise selon l'usage. Nous avons privilégié l'usage courant pour le choix entre l'acronyme français et l'acronyme anglais.

CFHTLS : Canada-France-Hawaii Telescope Legacy Survey,

HSC : Hyper Suprime-Cam,

HSC-SSP : Hyper Suprime-Cam Subaru Strategic Program,

Jo4 : Filtre défini par [Jarvis et al. \(2004\)](#),

Λ CDM : Modèle cosmologique considérant une matière noire froide, (de l'anglais, Lambda Cold Dark Matter),

LSS : Structures à grande échelle, abrégées d'après l'anglais Large scale structures,

LSST : Legacy Survey of Space and Time,

M18 : Filtre défini par [Miyazaki et al. \(2018\)](#)

MMD : Distance maximale d'association, abrégée d'après l'anglais Maximal Matching Distance,

MO : méthode de la masse d'ouverture,

S96 : Filtre défini par [Schneider \(1996\)](#),

S98 : Filtre défini par [Schneider et al. \(1998\)](#),

SNR : Rapport signal-sur-bruit, abrégé de l'anglais signal to noise ratio,

W234 : Notre méthode de détection multi-échelle,

WL : Effet lentille gravitationnelle faible, abrégé d'après l'anglais Weak Lensing.

Notations, nomenclatures et échelles caractéristiques

Sont présentées ici les différentes notations utilisées tout au long de ce manuscrit, ainsi que les échelles caractéristiques des éléments qui y sont décrits.

Notations et quantités usuelles

- c , la célérité de la lumière dans le vide : $2.998 \times 10^{10} \text{ cm.s}^{-1}$,
- G , la constante de la gravitation : $6.67 \times 10^{-11} \text{ m}^3.\text{kg}^{-1}.\text{s}^{-2}$,
- z , le redshift, (en français, le décalage vers le rouge),
- M_{\odot} , la masse solaire : $1.988 \times 10^{30} \text{ kg}$,
- H_0 , la constante de Hubble : $70 \pm 15 \text{ km.s}^{-1}.\text{Mpc}^{-1}$,
- h , le paramètre de Hubble sans dimension,
- λ , la longueur d'onde,
- Domaine de l'infrarouge : $0.01 \text{ cm} - 800 \text{ nm}$,
- Domaine du visible : $800 \text{ nm} - 100 \text{ nm}$,
- Domaine des rayons X : $10 \text{ nm} - 0.03 \text{ nm}$,
- $1 \text{ minute d'arc} = 1 \text{ arcmin} = 1'$,
- ∇_{\perp} , gradient dans la direction orthogonale.

Dans ce manuscrit, nous utilisons la nomenclature anglo-saxonne pour les nombres décimaux avec le choix du point comme séparateur. Ce choix est motivé par la volonté d'harmoniser les notations avec les figures générées en Python, utilisant cette même nomenclature et pour s'accrocher aux conventions courantes du domaine. Dans les équations, nous faisons le choix d'identifier les vecteurs par une flèche, conformément à la nomenclature française. Les angles sont représentés survolés par un accent circonflexe en géométrie plane (2D) et par une flèche en géométrie dans l'espace (3D) pour conserver une clarté dans la lecture des équations.

Échelles caractéristiques

- $1 \text{ Megaparsec (Mpc)} = 3.08 \times 10^{24} \text{ cm}$,
- Masse des galaxies : $10^6 - 10^{13} M_{\odot}$,
- Température du gaz intra-amas : $1 - 3 \text{ keV}$,
- Masse des amas de galaxies : $5.10^{13} - 10^{16} M_{\odot}$,
- Rayon d'un amas de galaxie : $1 - 10 \text{ Mpc}$.

Table des matières

I	Première partie : La compréhension de l'Univers sombre par la détection des amas de galaxies	5
1	L'évolution des structures dans la cosmologie moderne	7
1.1	Le modèle standard de la cosmologie	8
1.1.1	Le cadre mathématique d'un Univers en expansion	8
1.1.2	Le redshift et les distances	10
1.1.3	Le côté sombre de l'Univers	11
1.2	Comment évoluent les grandes structures?	12
1.2.1	Origine et évolution des structures	13
1.2.2	La théorie de Press et Schechter et la fonction de masse	14
1.3	Les amas de galaxies, des sondes cosmologiques	16
1.3.1	Définition, masse et rayon des amas de galaxies	16
1.3.2	La détection des amas	18
1.3.3	La caractérisation des amas	22
1.3.4	Pourquoi maintenant : HSC, EUCLID et Vera C. Rubin	25
1.4	Conclusions	27
2	Une observation optimale par l'effet de lentillage faible?	29
2.1	L'effet de lentille gravitationnelle	30
2.1.1	La vie chaotique des photons soumis à la gravitation	30
2.1.2	L'équation lentille	32
2.1.3	Le potentiel effectif projeté et la convergence	34
2.1.4	La distortion des images et le cisaillement	35
2.1.5	Inversion de masse et convergence	37
2.2	Données de lentille gravitationnelle faible	38
2.2.1	Mesure de cisaillement et catalogue de galaxies sources	38
2.3	Données simulées	40
2.3.1	Description de la simulation	40
2.3.2	Le catalogue de halos de matière noire	41
2.3.3	Le catalogue de galaxies sources	41
2.4	Conclusions	42

II	Deuxième partie : Méthodes de détections	43
3	L'algorithme de détection	45
3.1	Quelles données d'entrée pour la détection?	46
3.1.1	Les données d'entrée : 3 implémentations possibles	46
3.1.2	Impact de l'utilisation d'une grille régulière sur le signal	48
3.1.3	Impact de l'utilisation des cartes de cisaillement ou de convergence	50
3.1.4	Impact du choix du traitement des données en matière de détection	52
3.1.5	Conclusion	52
3.2	Les méthodes de filtrage	53
3.2.1	Les enjeux du filtrage	54
3.2.2	Filtrage gaussien	55
3.2.3	Approche standard	58
3.2.4	Fonctions utilisées	59
3.2.5	Conclusion	61
3.3	La procédure de seuillage et l'identification des détections	62
3.3.1	Le seuillage	62
3.3.2	L'identification des pics	63
3.3.3	Récapitulatif de l'algorithme de détection	63
3.4	Conclusions	64
4	La méthode multi-échelle par les ondelettes	67
4.1	De l'art de transformer le signal	67
4.1.1	La transformée de Fourier et ses limites	68
4.1.2	Une solution : la transformée en ondelettes	69
4.2	La méthode de détection multi-échelle	71
4.2.1	Le formalisme des ondelettes...	71
4.2.2	... au service de la détection des amas	72
4.2.3	La procédure multi-échelle	73
4.3	Avantages des ondelettes comparés aux autres filtres	74
4.3.1	Comparaison avec les filtres MO	75
4.3.2	Comparaison en termes de détections	77
4.4	Conclusions	79
III	Troisième partie : Méthodes d'association	81
5	La procédure d'association	83
5.1	Enjeux de la procédure d'association	84
5.1.1	Associer mais jusqu'à quel point?	84
5.1.2	Le catalogue utilisé pour l'association	86
5.2	Présélection des halos selon leur détectabilité	87
5.2.1	Estimation du SNR théorique	87
5.2.2	Application à la simulation DUSTGRAIN-pathfinder	89
5.3	Une nouvelle méthode d'association	89

5.3.1	La distance maximale d'association optimale : MMD	90
5.3.2	Notre procédure d'association adaptative	93
5.3.3	Adaptation au cas multi-échelle	93
5.4	Conclusions	95
6	Résultats des méthodes et analyse	97
6.1	Deux outils d'analyse : la complétude et la pureté	97
6.1.1	Évolution de la complétude	98
6.1.2	Pureté vs Complétude	99
6.2	Analyse des détections	101
6.2.1	Caractérisation des détections associées dans le plan redshift-masse	101
6.2.2	Fausse détections	103
6.3	Classification des détections associées et non-associées	104
6.3.1	Classification des détections associées	104
6.3.2	Analyse des détections non-associées	106
6.4	Conclusions	108
IV	Quatrième partie : Vers l'application à des données réelles	109
7	Préparation pour les données d'EUCLID	111
7.1	EUCLID, l'observation de l'Univers Sombre	112
7.1.1	Un télescope et un projet européen	112
7.1.2	Les instruments et les données lentilles	114
7.2	Le challenge WLC	115
7.2.1	Contexte du challenge WLC	115
7.2.2	Caractéristiques du jeu de données	115
7.2.3	Caractéristiques des catalogues de détections	116
7.2.4	Procédure d'association	117
7.3	Les différentes méthodes de détection	119
7.3.1	Présentation des méthodes	119
7.3.2	Nouveautés pour l'approche multi-échelle	121
7.3.3	La distance maximale d'association MMD	122
7.4	Premiers résultats	124
7.4.1	Pureté versus complétude	124
7.4.2	Détections associées dans le plan z-M	128
7.4.3	Caractérisation des détections associées communes	130
7.5	Conclusions	132
8	Application à des données réelles	135
8.1	Un bond en avant pour la détection des amas : Les données lentilles HSC-SSP	136
8.1.1	Le télescope Subaru et Hyper Suprime-Cam	136
8.1.2	Des données adaptées pour la détection	138
8.2	Défi de l'application aux données réelles HSC	139
8.2.1	Les données manquantes	139

8.2.2	Le bruit non gaussien	141
8.2.3	Une distribution non uniforme des galaxies sources	142
8.3	Application de la méthode de détection multi-échelle	143
8.3.1	Pixellisation des données	144
8.3.2	Reconstruction des cartes de convergence filtrées et estimation du bruit	146
8.3.3	La procédure de seuillage	148
8.4	Résultats	151
8.4.1	Association des détections	152
8.4.2	Résultats des détections associées	153
8.5	Perspectives	156

Introduction Générale



Le modèle standard de la cosmologie moderne montre que l'effondrement gravitationnel est responsable de la formation de la structure cosmique au sein de l'Univers. Sous l'effet de la gravité, la matière s'agrège pour former des structures de plus en plus grandes. Au sommet de cette évolution, les amas de galaxies retracent directement la croissance des structures à travers la toile cosmique. Cette information peut être utilisée pour contraindre les paramètres du modèle cosmologique et affiner notre compréhension de l'Univers. En particulier, l'abondance des amas en fonction de la masse et du redshift est très sensible aux paramètres cosmologiques sous-jacents (e.g. [White and Rees, 1978](#); [Perrenod, 1980](#); [Voit, 2005](#); [Allen et al., 2011](#)). De même, la distribution de matière au sein des amas, composée de matière noire (85%), de gaz chaud ionisé dans le milieu intra-amas (12%) et d'étoiles (~ 3%), reflète la distribution de matière de l'Univers et en fait de bons traceurs de l'évolution cosmique.

Dans ce contexte, la détection des amas de galaxies est un enjeu important pour la cosmologie. Généralement, les amas sont détectés grâce à leurs composantes baryoniques soit par le gaz dans les relevés en rayons X ou par l'effet Sunyaev-Zeldovich (SZ); soit par les étoiles dans les relevés optiques. Cependant, pour pouvoir exploiter ces catalogues de détections pour la cosmologie, il est nécessaire de s'appuyer sur des relations d'échelles qui permettent de relier les observables baryoniques (richesse optique, luminosité des rayons X, et flux SZ) à la masse totale sous-jacente. Utilisant seulement une faible portion de la distribution de la masse totale des amas, ces relations d'échelle comportent des incertitudes significatives issues des effets systématiques sur les estimations de masse et des fonctions de sélection (e.g. [Pratt et al., 2019](#), et les références associées).

La déviation cohérente de la trajectoire de la lumière issue de sources lointaines par la matière au sein des amas offre une méthode unique pour détecter les amas de galaxies à partir de leur distribution totale de matière. À grande échelle, cette déviation, appelée lentille gravitationnelle (e.g. [Schneider et al., 1992](#); [Pyne and Birkinshaw, 1993](#)) n'est détectable que statistiquement, et est par conséquent qualifiée de faible. Directement sensible à la masse totale projetée le long de la ligne de visée (LOS), cette méthode est attrayante pour la détection des amas de galaxies. Cependant, la sensibilité de la détection d'amas par le signal de lentille gravitationnelle faible dépend fortement du nombre de sources qui sont observées en arrière-plan.

Ce n'est que récemment que des relevés profonds et à grand champ ont permis d'obtenir les densités de galaxies sources nécessaires à la détection d'un nombre important d'amas de galaxies. Ces relevés rendent possibles la construction de véritables catalogues d'amas de galaxies sélectionnés par leur distribution de masse totale.

Le travail exposé dans cette thèse est ainsi motivé par le potentiel, pour la détection des amas par effet de lentille gravitationnelle faible, des relevés optiques à grand champ actuels et à venir tels que le relevé Hyper Suprime-Cam (HSC) ([Aihara et al. 2018a](#)), le relevé Vera C. Rubin (anciennement, LSST¹; [Closson Ferguson et al. 2009](#); [Ivezic et al. 2019](#)), le relevé *Euclid*² ([Laureijs et al. 2011](#)), et le relevé du *Roman Space Telescope*³ (anciennement WFIRST; [Spergel et al. 2015](#)). En réunissant une grande

1. <https://www.lsst.org/>
2. <https://www.euclid-ec.org/>
3. <https://roman.gsfc.nasa.gov/>

densité de galaxies sources et un champ large, ces relevés font entrer l'astrophysique dans l'âge d'or des études par l'effet de lentille gravitationnelle faible. En cela, ils motivent le développement de nouvelles méthodes capables de pleinement saisir ce potentiel, en particulier dans le cadre de la détection des amas de galaxies par leur distribution totale de masse. Dans ce contexte, cette thèse introduit le développement et la mise en place d'une nouvelle méthode, qualifiée de multi-échelle, pour la détection des amas de galaxies par l'effet de lentille gravitationnelle sur des données simulées puis réelles. Ce manuscrit est divisé en 8 chapitres :

- Le chapitre 1 décrit le contexte de l'évolution des structures dans la cosmologie moderne et les principes à la base de l'observation des amas de galaxies en optique, en X et en millimétrique par l'effet SZ. Il permet d'introduire à la fois le rôle des amas de galaxies pour contraindre les paramètres du modèle cosmologique ainsi que les apports de l'effet de lentille gravitationnelle faible pour leur détection dans le contexte des relevés actuels et futurs.
- Le chapitre 2 présente en détails l'effet de lentille gravitationnelle faible et son formalisme appliqué à la détection des amas de galaxies. Nous décrivons comment les catalogues de galaxies sources sont construits à partir des observations. Puis, nous présentons les caractéristiques des simulations que nous utilisons dans les chapitres 3 à 7.
- Le chapitre 3 présente les différentes méthodes de détection qui s'appuient sur des filtrages standards qui sélectionnent une échelle unique. Nous introduisons aussi le filtrage gaussien et les filtres de masse d'ouverture les plus performants pour la détection des amas par effet de lentille gravitationnelle faible. Ce chapitre détaille l'algorithme de détection complet (pixelisation, filtrage, seuillage, identification des pics) que nous avons utilisé pour l'application de ces méthodes sur les données simulées afin de produire les catalogues de détection. Ces méthodes servent de base de comparaison pour notre nouvelle approche que nous décrivons au chapitre 4.
- Le chapitre 4 introduit notre nouvelle méthode de détection multi-échelle, baptisée W234, adaptée de la décomposition par ondelettes dont nous détaillons le formalisme associé. Une première comparaison entre notre approche et les méthodes standards est réalisée d'abord par leur représentation dans les espaces réel et de Fourier, puis par les catalogues de détections produits par chaque méthode.
- Le chapitre 5 décrit la procédure d'association développée durant cette thèse pour confirmer ou infirmer les détections obtenues par les différentes méthodes. Le développement d'une nouvelle méthode d'association était nécessaire dans le cas multi-échelle afin de s'assurer d'une caractérisation juste des performances des différentes méthodes en vue de leur comparaison. Nous présentons la sélection appliquée sur les catalogues de halos afin de réduire le nombre de halos à ceux potentiellement détectables. Puis, nous détaillons les spécificités de la procédure d'association que nous avons développée et utilisée dans le cadre de cette thèse pour tenir compte des propriétés des différents filtres appliqués.
- Le chapitre 6 est consacré à l'analyse et la présentation des performances en termes de détections associées et non associées obtenues sur les données simulées par toutes les méthodes présentées précédemment. La comparaison des différentes méthodes est réalisée à la fois à travers les outils de pureté et de complétude mais aussi par les détections associées dans le plan redshift-masse afin de comparer les comportements des méthodes en termes de sélection. Nous concluons ce chapitre par une classification complète des détections associées et non associées pour chacune des méthodes. Ce chapitre permet de présenter les premiers résultats de la méthode multi-échelle dans le cadre idéal d'un jeu de données simulées suffisamment large.
- Le chapitre 7 détaille notre participation et notre implication dans l'analyse des résultats du challenge WLC pour la détection des amas de galaxies par effet de lentille gravitationnelle faible,

organisé au sein du consortium *EUCLID*. Nous décrivons le contexte de ce challenge, et le jeu de données utilisé ainsi que toutes les méthodes soumises par les participants. Étant membre de l'équipe chargée de l'analyse des résultats, nous présentons les performances de chaque méthode en utilisant la procédure d'association définie en chapitre 5. Cette analyse se concentre sur la comparaison des méthodes en termes de pureté et complétude et dans le plan redshift-masse. Ce chapitre vise à caractériser l'efficacité des différentes méthodes de détection ainsi que leur complémentarité afin de définir la ou les méthodes qui seront appliquées sur les données *EUCLID*.

- Enfin, le chapitre 8 présente l'application de la méthode de détection multi-échelle aux données de la première année du relevé Hyper Suprime-Cam Subaru Strategic Program. Dans ce contexte, nous décrivons les défis de l'application à des données réelles tels que le bruit, les données manquantes et la distribution non uniforme des galaxies. Nous présentons les détails de l'application de la méthode multi-échelle et son adaptation pour tenir compte de ces défis. Ce chapitre permet de présenter les premiers résultats de la méthode multi-échelle en termes de détections sur six champs disjoints de données réelles et leur comparaison aux résultats publiés sur ces mêmes champs.

Nous achevons ce manuscrit par les conclusions du travail réalisé durant cette thèse ainsi que les perspectives.

*
* *

Première partie

Première partie : La compréhension de l'Univers sombre par la détection des amas de galaxies

1

L'évolution des structures dans la cosmologie moderne

A l'exception des enfants [...], peu d'entre nous passent beaucoup de temps à se demander pourquoi la nature est telle qu'elle est; d'où vient le cosmos ou s'il a toujours été là; si le temps fera un jour machine arrière et si les effets précéderont les causes ou s'il y a des limites extrêmes à la connaissance humaine.

Stephen Hawking
Une brève histoire du temps.

1.1	Le modèle standard de la cosmologie	8
1.1.1	Le cadre mathématique d'un Univers en expansion	8
1.1.2	Le redshift et les distances	10
1.1.3	Le côté sombre de l'Univers	11
1.2	Comment évoluent les grandes structures?	12
1.2.1	Origine et évolution des structures	13
1.2.2	La théorie de Press et Schechter et la fonction de masse	14
1.3	Les amas de galaxies, des sondes cosmologiques	16
1.3.1	Définition, masse et rayon des amas de galaxies	16
1.3.2	La détection des amas	18
1.3.3	La caractérisation des amas	22
1.3.4	Pourquoi maintenant : HSC, EUCLID et Vera C. Rubin	25
1.4	Conclusions	27



La cosmologie est un domaine de la physique dont l'objet premier est l'étude de la formation et de l'évolution de l'Univers. Au cours des dernières décennies, les progrès technologiques et les avancées théoriques ont permis à la cosmologie de passer d'une discipline spéculative à une science précise et observationnelle. Beaucoup de mystères entourent encore notre compréhension de l'Univers et notamment son évolution. Les amas de galaxies qui tracent la distribution de matière dans l'Univers forment un outil crucial pour la cosmologie. C'est dans ce contexte que s'inscrit cette thèse et l'enjeu de leur détection. Dans ce chapitre, nous décrivons en détail comment évoluent les grandes structures de l'Univers, du *Big Bang* à aujourd'hui, la place des amas de galaxies dans cette évolution et les enjeux de leur détection. Des questions cruciales sur la nature de la matière noire,

de l'énergie sombre, et sur les processus de formation des structures cosmiques sont également abordées dans ce chapitre. Pour ce faire, en section 1.2, nous introduisons tout d'abord le modèle cosmologique standard, baptisé Λ CDM (de l'anglais Cold Dark Matter), et l'évolution des structures dans la toile cosmique. Dans la section 1.3, nous détaillons ensuite la nature et le rôle des amas de galaxies dans l'avancée de notre compréhension de l'Univers. En particulier, nous introduisons les enjeux de la détection des amas de galaxies dans les différents domaines de la physique : en optique, en X et en millimétrique, la caractérisation des amas pour ces observations ainsi que les effets systématiques liés à chaque domaine. Nous concluons ce chapitre par un état de l'art de la détection des amas de galaxies par l'effet de lentille gravitationnelle faible et son importance dans le contexte des relevés d'aujourd'hui.

1.1 Le modèle standard de la cosmologie

Pour décrire l'Univers et son évolution, la cosmologie moderne s'appuie sur la théorie de la relativité générale, les équations de Friedmann et le principe cosmologique qui aboutissent à un modèle étoffé peu à peu par les différentes observations. Le modèle cosmologique vise à décrire et poser un cadre conceptuel, par un jeu d'équations et de paramètres, sur la structure, l'origine, l'évolution et le destin de l'Univers dans son ensemble. Dans cette section, nous introduisons les différentes notions clés pour la construction du modèle cosmologique Λ CDM et leur enjeux pour la cosmologie.

1.1.1 . Le cadre mathématique d'un Univers en expansion

Le modèle cosmologique est la pierre angulaire de la cosmologie moderne, et l'aboutissement d'années d'observation, de réflexion et de développement théorique visant à déchiffrer les mystères de l'Univers. Le modèle standard de la cosmologie, Λ CDM, repose sur trois piliers : la relativité générale pour sa géométrie, le principe cosmologique avec les équations de Friedmann associées et la présence de la matière noire et d'énergie noire comme composantes principales de l'Univers. Dans cette sous-section, nous introduisons ces différents piliers, le modèle cosmologique standard qui en découle et leur rôle dans la description de l'Univers connu.

★ La théorie de la relativité générale

La théorie de la relativité générale, formulée par Albert Einstein au début du XXe siècle, a révolutionné notre compréhension de la gravité. Cette théorie novatrice décrit comment la matière et l'énergie courbent l'espace-temps, créant ainsi une force de gravité. Elle nous a permis de dépasser la vision newtonienne de la gravité pour expliquer des phénomènes tels que la déviation de la lumière par des objets massifs (voir 2.1), la dilatation du temps en présence de champs gravitationnels forts, et même l'expansion de l'Univers lui-même. La relativité générale fournit le cadre mathématique permettant de comprendre la dynamique de l'Univers sur des échelles cosmiques. Elle décrit comment la matière et l'énergie affectent la géométrie de l'espace-temps au travers de l'équation (Einstein, 1915) :

$$G_{\mu\nu} - \Lambda g_{\mu\nu} = \frac{8\pi G}{c^4} T_{\mu\nu}, \quad (1.1)$$

où :

- $G_{\mu\nu}$ est le tenseur d'Einstein qui encode la géométrie de l'espace-temps,
- Λ est la constante cosmologique, initialement introduite par A. Einstein dans le cas d'un Univers plat avant d'être retirée puis de nouveau réintégrée (ce point est détaillé dans la partie ★ L'Univers sombre),
- $g_{\mu\nu}$ est le tenseur métrique de l'espace-temps de Minkowski de convention (+,-,-,-),
- G est la constante gravitationnelle newtonienne,

- c est la célérité de la lumière dans le vide,

- $T_{\mu\nu}$ est le tenseur énergie-impulsion qui encode la dynamique des composants de l'Univers.

Sous l'hypothèse que le contenu de l'Univers se comporte comme un fluide parfait comobile, le tenseur énergie-impulsion, $T_{\mu\nu}$, s'écrit alors :

$$T_{\mu\nu} = -pg_{\mu\nu} + \left(\frac{p}{c^2} + \rho\right)u^\mu u^\nu, \quad (1.2)$$

avec p , la pression du fluide, ρ , la densité d'énergie, et, $u^\mu u^\nu$, la quadri-vitesse du fluide.

La résolution de l'équation 1.1 constitue un pilier fondamental de notre compréhension de la dynamique cosmique. Le modèle standard postule que l'Univers est homogène et isotrope à grande échelle, c'est-à-dire que ses propriétés sont les mêmes dans toutes les directions et en tout point de l'espace. Cette hypothèse, appelée le principe cosmologique, a jeté les bases de notre compréhension moderne de la structure cosmique et aboutit à une solution dynamique de l'équation 1.1. Cette solution appelée la métrique Friedmann-Lemaître-Robertson-Walker, FLRW, s'écrit sous la forme (Friedmann, 1924; Lemaître, 1927; Robertson, 1935; Walker, 1935) :

$$ds^2 = g_{\mu\nu}dx^\mu dx^\nu = c^2 dt^2 - a^2(t) \left(\frac{dr^2}{1-kr^2} + r^2 d\Omega^2 \right), \quad (1.3)$$

en considérant $d\Omega^2 = d\theta^2 + \sin^2\theta d\phi^2$ pour les coordonnées sphériques (r, θ, ϕ) , avec t , le temps propre et $k \in \{-1, 0, 1\}$, le paramètre de courbure spatiale de l'Univers, supposé nul aujourd'hui. La fonction a est le facteur d'échelle de l'Univers qui caractérise son expansion.

★ Les équations de Friedmann

La combinaison des équations 1.1, 1.2 et 1.3 nous permet d'isoler l'évolution du facteur d'échelle, a , et nous aboutissons au système d'équations différentielles, dites équations de Friedmann (Friedmann, 1924).

$$H^2 \equiv \left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\rho - \frac{kc^2}{a^2} + \frac{\Lambda c^2}{3}, \quad (1.4)$$

et

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}\left(\rho + \frac{3p}{c^2}\right) + \frac{\Lambda c^2}{3}, \quad (1.5)$$

où :

- H , est le paramètre de Hubble,

- \dot{a} est la dérivée temporelle du paramètre d'échelle a ,

- ρ est la densité volumique de masse totale d'énergie de l'Univers,

- p est la pression totale de l'Univers.

Ces équations, élaborées dans les années 1920 par le mathématicien russe Alexander Friedmann, lient la dynamique de l'Univers à la densité d'énergie et la pression de ces composantes. L'évolution temporelle du facteur d'échelle est régie par la première équation de Friedmann, qui relie le taux de variation de a à la densité totale d'énergie de l'Univers. L'une des conséquences les plus importantes des équations de Friedmann est la prédiction de l'expansion de l'Univers. Le paramètre d'échelle, a , augmente au fil du temps, ce qui signifie que l'Univers se dilate. Cette prédiction a été confirmée par les observations, notamment la découverte de l'expansion des galaxies faite par Edwin Hubble dans les années 1920. C'est pourquoi, à partir de cette première équation, on définit notamment le paramètre de Hubble, H , qui quantifie l'accélération de l'expansion de l'Univers. Notons que le paramètre de Hubble varie en fonction du temps et on définit H_0 comme sa valeur actuelle.

La première équation de Friedmann peut également se réécrire en fonction du contenu de l'Univers. Dans le cas d'un Univers plat (sans courbure, $k = 0$), la combinaison des deux équations de Friedmann nous permet tout d'abord de retrouver l'équation de conservation locale de l'énergie :

$$\dot{\rho} = -3\frac{\dot{a}}{a}\left(\rho + \frac{p}{c^2}\right). \quad (1.6)$$

En considérant une équation d'état de la forme $p = w\rho c^2$ avec w le paramètre d'équation d'état w , la résolution de l'équation 1.6 nous donne :

$$\rho(a) \propto a^{-3(1+w)}. \quad (1.7)$$

L'équation 1.4 se réécrit alors :

$$H^2(a) = \frac{8\pi G}{3}\rho_i(a) - \frac{kc^2}{a^2} + \frac{\Lambda c^2}{3}, \quad (1.8)$$

avec ρ_i la densité volumique de masse des différentes composantes i du contenu de l'Univers. Il est courant de définir chaque composante sous la forme de la densité actuelle, $\Omega_i \equiv \rho_i/\rho_{\text{crit}}$, à partir de ρ_{crit} , appelée densité critique de l'Univers et définie comme :

$$\rho_{\text{crit}} \equiv \frac{3H^2(a)}{8\pi G}. \quad (1.9)$$

Alors, la première équation de Friedmann en fonction des densités relatives de l'Univers s'écrit :

$$H^2(a) = H_o^2 \sum_i \Omega_i a^{-3(1+w)}, \quad (1.10)$$

avec Ω_i incluant $\Omega_\Lambda \equiv \frac{\Lambda c^2}{3H^2}$ et $\Omega_k \equiv \frac{kc^2}{aH^2}$, qui est égal à 0 sous l'hypothèse d'un Univers plat.

Cette réécriture de la première équation de Friedmann permet de relier certains de ses paramètres clefs. Ces différentes équations (Eqs. 1.5, 1.9 et 1.10) fournissent un cadre théorique essentiel pour décrire l'évolution de l'Univers à grande échelle. Ceci constitue la base mathématique du modèle cosmologique standard.

1.1.2 . Le redshift et les distances

Les équations de Friedmann, couplées avec les prédictions de [Lemaître \(1927\)](#) et les observations [Hubble \(1929\)](#) ont mené à l'idée révolutionnaire que l'Univers est en constante expansion depuis son origine. La première idée correspond notamment au modèle du Big Bang pour expliquer le passé de l'Univers en remontant cette expansion au cours du temps. La seconde idée montre que l'espace-temps se dilate et que les distances entre les objets cosmiques augmentent.

Ainsi, en tenant compte de l'expansion de l'Univers et si on considère un photon provenant d'une galaxie lointaine à la position radial, r , qui parvient à un observateur sur terre en suivant l'axe polaire, l'équation 1.3 s'écrit :

$$ds^2 = 0 = c^2 dt^2 - a^2(t) dr^2. \quad (1.11)$$

Si on considère deux instants consécutifs t_1 et t_2 pour lesquels nous notons λ_1 et λ_2 les longueurs d'ondes mesurées à ces deux instants pour ce photon, en intégrant sur la trajectoire du photon, on obtient alors :

$$\frac{a(t_1)}{a(t_2)} = \frac{\lambda_1}{\lambda_2} \equiv 1 + z. \quad (1.12)$$

Cet effet est appelé décalage vers le rouge ou redshift et est caractérisé par le paramètre z (dans la suite, nous privilégions le terme redshift). Enfin, en normalisant le facteur d'échelle actuelle pour se ramener à 1, le paramètre d'échelle permet de définir le redshift cosmologique :

$$a = \frac{1}{1 + z}. \quad (1.13)$$

Cette nouvelle équation nous permet de réécrire la première équation de Friedmann une dernière fois pour retrouver sa forme la plus courante :

$$H^2(z) = H_o^2 \sum_i \Omega_i (1 + z)^{3(1+w)}. \quad (1.14)$$

Le redshift est une propriété importante de tous les objets cosmiques et peut être utilisé à la fois pour définir l'âge d'un objet mais aussi sa distance par rapport à l'observateur. Ce décalage observé est une composante essentielle de l'étude de l'évolution des structures dans l'Univers. Le même phénomène affecte également les distances entre deux positions dans l'Univers ou qui ne respectent plus les règles de la géométrie euclidienne. C'est pourquoi les astrophysiciens utilisent notamment la distance comobile ou la distance angulaire de diamètre apparent qui tiennent compte du redshift. Nous décrivons ici les distances qui sont utilisées dans ce manuscrit.

- La **distance comobile** est la distance parcourue entre deux instants t_1 et t_2 correspondant aux positions z_1 et z_2 dans une unité qui suit le facteur d'échelle. En d'autres termes, la distance comobile fait abstraction de l'expansion de l'Univers et ne varie pas au cours du temps. Sa formule s'écrit :

$$D(z_1, z_2) = \int_{t_1}^{t_2} \frac{c}{a(t')} dt'. \quad (1.15)$$

- Le **diamètre angulaire apparent** est utilisé dans l'étude de l'effet de lentille gravitationnelle (voir Sect. 2.1) et correspond au ratio entre le diamètre d'un objet source d et l'angle d'observation α sous lequel il est observé, $r = d/\alpha$. Le calcul est impacté par la cosmologie sous-jacente, de telle sorte que la distance angulaire d'une source à un redshift, z peut se réécrire en fonction de la distance comobile, r :

$$D(z) = \frac{r}{1+z}. \quad (1.16)$$

Notons que les non-linéarités de l'espace-temps défini pour un Univers en expansion impliquent que les distances en jeu ne sont pas additives. Pour trois objets célestes aux redshifts z_1 , z_2 et z_3 , $D(z_1, z_3) \neq D(z_1, z_2) + D(z_2, z_3)$.

1.1.3 . Le côté sombre de l'Univers

Maintenant que nous avons établi le cadre mathématique du modèle, intéressons-nous au dernier pilier du modèle Λ CDM, à savoir le contenu de l'Univers. Ce modèle considère que l'Univers est constitué à la fois de matière baryonique, de densité Ω_b , mais aussi de matière noire de densité Ω_{CDM} , et d'énergie sombre de densité Ω_Λ . Dans cette sous-section, nous décrivons les composantes sombres de l'Univers dans le modèle standard cosmologique.

- La matière noire qui est au centre du modèle cosmologique est originellement théorisée pour répondre au problème de la masse manquante. En effet, dès les années 1930, les astronomes tels que Zwicky (1937) constatent des vitesses plus élevées pour les galaxies de l'amas de la chevelure de Bérénice que celles théoriquement induites par la masse lumineuse de l'amas. En 1980, cette observation est étendue par Rubin et al. (1980) aux étoiles au sein des galaxies spirales. Elle observe notamment des vitesses de rotation trop élevées et pratiquement constantes indifféremment de la distance au centre de la galaxie. Ces vitesses ne sont pas explicables directement par la masse lumineuse observable dans les galaxies et les astronomes postulent l'existence d'une matière non observable entourant les structures cosmiques et responsable de la rotation rapide des étoiles. Cette matière invisible est alors baptisée matière noire.

Dans le modèle cosmologique standard, la matière noire représente 85% du contenu en matière de l'Univers. Interagissant pas ou faiblement avec le reste de la matière et avec le rayonnement électromagnétique, la matière noire est très difficile à détecter et à caractériser. Elle n'est ainsi sensible qu'à la force de gravitation. Cette matière serait à la base de la formation des structures dans l'Univers.

- L'énergie sombre est postulée pour répondre au problème de l'accélération de l'expansion de l'Univers. En considérant uniquement le contenu usuel (rayonnement, poussière, atomes) de

l'Univers dans l'équation 1.5, la théorie prédit que $\ddot{a} < 0$, à savoir une décélération de l'expansion de l'Univers. Cependant, les études menées par [Riess et al. \(1998\)](#) et [Perlmutter et al. \(1999\)](#) sur les supernovæ de type Ia corrélant la magnitude apparente au redshift montrent que cette expansion s'accélère. Par la suite, d'autres sondes cosmologiques comme le fond diffus cosmologique et les amas de galaxies ont motivé l'introduction d'une nouvelle composante de l'Univers constituant près de 75% de son contenu total. C'est la naissance de l'énergie noire, dite énergie du vide et considérée responsable de l'accélération de l'expansion de l'Univers ([Huterer and Turner, 1999](#); [Astier et al., 2006](#); [Amendola and Tsujikawa, 2010](#)).

Dans la pratique, cette énergie est liée dans les équations de Friedmann à la constante cosmologique Λ . Cette constante est introduite initialement par A. Einstein dans les équations pour obtenir un Univers statique. N'étant plus cohérente avec un Univers en expansion, elle est retirée dans les années 1930. Cependant, dans les équations d'Einstein et Friedmann, cette constante correspondait précisément à un fluide de pression négative et de paramètre d'état $w = -1$ très semblable aux propriétés attendues par l'énergie sombre. C'est pourquoi, la constante cosmologique Λ est de nouveau introduite et l'équation 1.1 contient alors le terme $\Lambda g_{\mu\nu}$ désormais associé à l'énergie sombre. Dans le modèle Λ CDM, la constante cosmologique Λ est conservée comme présenté dans l'équation 1.1 et correspond à l'intégration de l'énergie du vide dans le cadre mathématique.

Dans le cas du modèle Λ CDM, la densité de matière ordinaire, de matière noire et d'énergie sombre, ainsi que la courbure spatiale de l'Univers, influencent toutes l'évolution du facteur d'échelle a . Les équations de Friedmann tiennent compte de ces composantes en incorporant des termes spécifiques pour chaque type de matière et d'énergie. L'expression finale de l'équation de Friedmann (Eq. 1.10) est alors donnée par :

$$H^2(z) = H_0^2 \left(\Omega_\Lambda + (1 - \Omega_o)(1 + z)^2 + \Omega_m(1 + z)^3 + \Omega_R(1 + z)^4 \right), \quad (1.17)$$

avec :

- Ω_m , la densité de matière totale, incluant la matière noire et la matière baryonique, elle-même composée des densités de matière noire Ω_{CDM} , et de matière baryonique Ω_b ,
- Ω_R , la densité de radiation, négligeable aujourd'hui,
- Ω_Λ , la densité d'énergie noire,
- $\Omega_o = \Omega_\Lambda + \Omega_R + \Omega_m$.

En résumé, le modèle standard de la cosmologie, Λ CDM, basé sur la relativité générale, le principe cosmologique et les équations de Friedmann, constitue une base théorique solide pour notre compréhension de l'évolution de l'Univers. Ces équations intègrent les contributions de la matière noire, de la matière baryonique et de l'énergie sombre. Elles nous permettent de sonder les origines et le devenir de l'Univers à travers l'étude de la dynamique cosmique à grande échelle.

1.2 Comment évoluent les grandes structures ?

Lorsque l'on combine l'homogénéité et l'isotropie, on obtient une vision de l'Univers dans laquelle les lois physiques sont les mêmes partout et à toutes les époques. Cependant, la présence des structures dans l'Univers actuel est hautement hétérogène aux petites échelles. L'existence des structures et en particulier les amas de galaxies va de pair avec la croissance d'inhomogénéité de densité au sein de l'Univers. Dans cette section, nous décrivons comment nous sommes passés d'un Univers homogène isotrope à l'Univers hétérogène aux petites échelles que nous observons aujourd'hui.

1.2.1 . Origine et évolution des structures

La détection des amas de galaxies étant le sujet de cette thèse, il est également important d'introduire comment de telles structures se forment. Dans cette sous-section, nous questionons l'évolution des structures à partir de petites fluctuations primordiales.

Dans la théorie du Big Bang, à son origine, l'Univers était parfaitement homogène et isotrope. Puis, de petites fluctuations quantiques de densité se sont formées. L'un des concepts fondamentaux pour comprendre l'évolution de l'Univers est l'origine et la croissance de ces perturbations de matière.

Les fluctuations quantiques de densité à l'origine de l'Univers remontent à la période très précoce, appelée l'ère de Planck, qui se situe à une fraction extrêmement courte de seconde après le Big Bang. Cette ère va de pair avec des températures et des énergies extrêmement élevées. L'origine des fluctuations quantiques pendant l'ère de Planck est directement liée au principe d'incertitude quantique d'Heisenberg qui stipule qu'il est impossible de connaître simultanément avec une précision infinie la position et la quantité de mouvement d'une particule. Ainsi, même dans le vide absolu, des fluctuations quantiques spontanées de l'énergie peuvent se produire.

Ces fluctuations quantiques ont été amplifiées pendant une phase appelée inflation. L'inflation, théorisée pour expliquer certains problèmes cosmologiques, est une période d'expansion exponentielle extrêmement rapide de l'Univers. Durant cette phase, les fluctuations quantiques ont été étendues à des échelles cosmologiques observables (voir [Achúcarro et al., 2022](#), comme revue).

À mesure que l'Univers s'expandait, il refroidissait également. Lorsque la température a atteint environ 3000 K, soit environ 300 000 ans après le Big Bang, l'énergie moyenne des photons a diminué au point où ils n'étaient plus suffisamment énergétiques pour maintenir les électrons et les protons séparés. À ce stade, les électrons ont commencé à se combiner avec les protons pour former des atomes neutres d'hydrogène (e.g. [Olive et al., 2000](#)). Après cette étape de recombinaison, les photons ne sont plus constamment diffusés par les électrons et voyagent à travers l'Univers de manière plus libre. Par ce découplage matière-rayonnement, l'Univers devient plus transparent à la lumière, ce qui marque le début de la domination du rayonnement dans la composition de l'Univers. L'expansion de l'Univers s'est poursuivie, mais la domination du rayonnement a commencé à décliner à mesure que l'Univers s'expandait et se refroidissait (voir [Steigman, 2007](#), comme revue).

Finalement, une autre transition majeure s'est produite, appelée la transition matière-radiation. Cette transition s'est produite lorsque l'énergie du rayonnement a diminué au point où la densité de la matière est devenue prépondérante par rapport à la densité du rayonnement. Lors de cette transition, la gravité a commencé à jouer un rôle de plus en plus important dans l'évolution de l'Univers.

Ainsi, les fluctuations quantiques ont évolué sous l'influence de la gravité pour former les structures massives que nous observons aujourd'hui, telles que les galaxies, les amas de galaxies et les superamas. La compréhension de la croissance des perturbations de matière est essentielle pour interpréter les observations cosmologiques, telles que le rayonnement fossile (CMB), les cartes de la distribution de galaxies, et les observations des amas de galaxies.

Dans le régime linéaire ($\delta \ll 1$), ces fluctuations peuvent s'écrire sous la forme :

$$\rho(\vec{r}) = \rho_m(1 + \delta(\vec{r})), \quad (1.18)$$

soit,

$$\delta(\vec{r}) = \frac{\rho(\vec{r}) - \rho_m}{\rho_m}, \quad (1.19)$$

avec ρ_m qui représente la densité de masse moyenne de l'Univers. En prolongeant la description de l'Univers comme un fluide parfait, celui-ci vérifie alors les équations d'Euler et de la continuité :

$$\frac{\partial \vec{v}}{\partial t} + (\vec{v} \cdot \vec{\nabla})\vec{v} = -\frac{1}{\rho}\vec{\nabla}P - \vec{\nabla}\Phi, \quad (1.20)$$

$$\frac{\partial \rho}{\partial t} + \vec{\nabla} \cdot (\rho \vec{v}) = 0, \quad (1.21)$$

avec p la pression, v la vitesse, telle que $v \ll c$, et Φ , le potentiel gravitationnel newtonien, qui vérifie :

$$\nabla^2 \Phi = 4\pi G\rho. \quad (1.22)$$

et avec $\Phi \ll c^2$.

La combinaison des équations 1.20 et 1.21 avec l'équation de Poisson du potentiel gravitationnel, Eq. 1.22 nous permet de revenir à l'évolution des perturbations, δ , nommée équation de croissance linéaire :

$$\frac{d^2 \delta}{dt^2} + 2H \frac{d\delta}{dt} - 4\pi G\rho_m \delta = 0, \quad (1.23)$$

dans le cas où les forces de pression sont négligeables ($p = 0$) devant la gravitation. Cette équation relie la croissance des fluctuations de densité à l'expansion de l'Univers par le paramètre de Hubble, H , et à la force de gravité. Lorsque nous nous intéressons à des fluctuations de densité relativement petites par rapport à la densité moyenne de l'Univers, l'utilisation du développement de Taylor sur l'équation permet d'aboutir à la solution simple :

$$\delta(t) = \delta_o D(t), \quad (1.24)$$

où D représente le facteur de croissance, et δ_o , la valeur initiale des fluctuations. Cette solution montre comment de petites fluctuations de densité évoluent de manière linéaire avec le temps et amène à l'écriture de l'équation 1.23 :

$$\frac{d^2 D}{dt^2} + 2H \frac{dD}{dt} - \frac{3}{2} \Omega_m H^2 D = 0. \quad (1.25)$$

De manière générale, la résolution de l'équation de croissance linéaire exacte par une approche numérique met en valeur une solution comportant deux modes : un mode décroissant δ_- , puis un mode croissant δ_+ . L'expression de la solution est alors de la forme : $\delta(\vec{x}, t) = A(\vec{x})\delta_+(t) + A(\vec{x})\delta_-(t)$. Après calcul, lorsque la densité de matière domine la densité d'énergie de l'Univers, le mode croissant domine l'équation de croissance et les fluctuations augmentent de manière exponentielle pour former des structures. Puis, quand la densité de matière devient dominée par la densité d'énergie sombre, comme à l'heure actuelle, les fluctuations sont stables et la formation des structures est très ralentie.

La croissance des fluctuations décrite précédemment reste valable dans le régime linéaire, c'est-à-dire tant que $\delta \ll 1$. Cette hypothèse n'est plus remplie dès lors que des régions de l'Univers deviennent de plus en plus denses, comme par exemple dans le cas des amas de galaxies où la densité peut atteindre 10^4 fois la densité moyenne de l'Univers. Des processus non-linéaires complexes apparaissent et dans ce nouveau cadre, le régime linéaire et l'équation de croissance ne sont plus suffisants pour expliquer l'évolution des structures observées.

En résumé, la croissance des fluctuations de densité a conduit peu à peu à la formation des structures cosmiques. Les régions les plus denses sont devenues à la fois les sites de formation des phénomènes non-linéaires ainsi que des premières étoiles, des galaxies et des amas de galaxies sous l'effet de l'effondrement gravitationnel.

1.2.2 . La théorie de Press et Schechter et la fonction de masse

Afin d'étudier la croissance non-linéaire des grandes structures comme les amas de galaxies, il est nécessaire de s'appuyer sur une approche phénoménologique. L'approche la plus classique pour cela est la théorie formulée par William H. Press et Paul Schechter dans les années 1970 (Press and Schechter, 1974) permettant d'affiner la cosmologie à partir des observations. Dans cette sous-section, nous décrivons la théorie de Press et Schechter et l'utilisation de la fonction de masse qui en découle.

Pour faire face à la complexité du traitement des processus physiques non-linéaires intervenant dans la croissance des structures, Press and Schechter (1974) proposent une approche focalisée sur l'observation. Cette théorie repose sur l'idée que l'intégralité de la matière actuelle ($z = 0$) dans l'Univers est contenue dans des objets stables en équilibre hydrostatique, appelés halos virialisés. Ainsi,

l'évolution des structures est telle que les régions de surdensités dans l'Univers s'effondrent gravitationnellement de manière sphérique pour former des halos de matière noire. Ces halos de matière noire servent ensuite de noyaux pour la formation des galaxies et des amas de galaxies, créant ainsi les structures cosmiques. Notons que, dans les chapitres suivants, le terme "halo" fait référence à la distribution de masse théorique tandis que le terme "amas de galaxies" désigne généralement par convention la distribution de masse qui est observé.

Selon le modèle de Press et Schechter, la distribution des halos pour une masse M et un redshift z peut être calculée par la fonction de masse $\frac{dn}{dM}$. Cette fonction de masse, pour un redshift donné, peut s'écrire sous la forme (Voit, 2005) :

$$\frac{dn}{dM} = f(\Omega_m, \sigma_8, M) \frac{\rho_m}{M} \frac{d \ln \sigma^{-1}}{dM}, \quad (1.26)$$

où Ω_m est la densité de matière de l'Univers, σ_8 est la normalisation des fluctuations de densité à l'échelle de $8, h^{-1}$, Mpc, et ρ_m est la densité de matière critique. Enfin, la fonction f décrit la fraction de la masse contenue dans un halo de masse M par rapport à la masse de l'Univers et s'écrit :

$$f(\Omega_m, \sigma, M) = \sqrt{\frac{2}{\pi}} \frac{\rho_o \Omega_m}{M} \frac{\delta_c}{\sigma} \exp\left(-\frac{\delta_c^2}{2\sigma^2}\right). \quad (1.27)$$

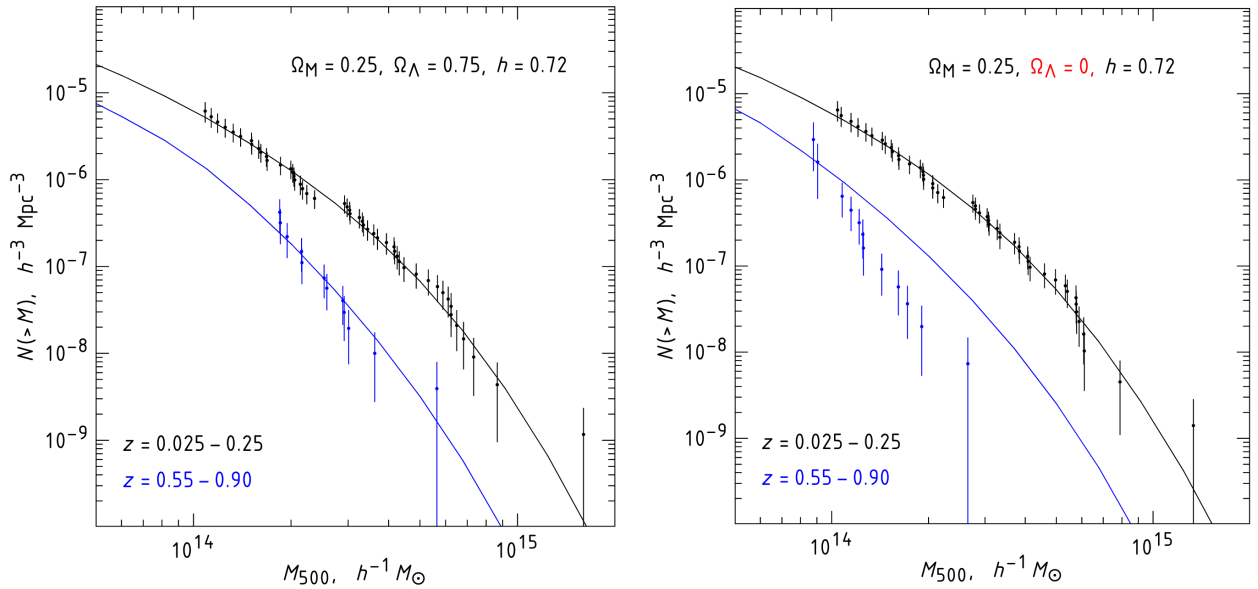


Figure 1.1 – Comparaison entre l'abondance d'amas à haut redshift (bleu) et bas redshift (noir) obtenues à partir des observations de *Chandra* (points) et celles prédites pour différentes cosmologies à partir de simulation à N corps (lignes solides avec à gauche : $\Omega_\Lambda = 0.75$, et droite : $\Omega_\Lambda = 0$) (Figure extraite de Vikhlinin et al. 2009).

L'équation de la fonction de masse des halos, $\frac{dn}{dM}$, est directement reliée à l'abondance des halos pour une masse et un redshift donné. La fonction de masse dépend de la distribution de matière dans l'Univers et fait donc le lien entre cette distribution et les paramètres cosmologiques par unité de masse et de redshift. Elle croît à la fois avec la densité de matière Ω_m et avec σ_8 . En d'autres termes, plus il y a de matière dans l'Univers et plus les fluctuations sont importantes, plus il y a de halos formés. Les résultats présentés par Vikhlinin et al. (2009) mettent en avant cette dépendance en fonction des paramètres cosmologiques. Cette dépendance est illustrée dans la figure 1.1.

De par le lien étroit entre l'évolution des amas de galaxies et le modèle cosmologique, les amas de galaxies et leurs halos associés sont des sondes cruciales dans l'étude de la cosmologie. Leur étude fournit des informations précieuses sur l'Univers à grande échelle et permet de contraindre les paramètres fondamentaux du modèle cosmologique (Allen et al., 2011; Kravtsov and Borgani, 2012).

Les amas de galaxies et les halos associés constituent des sondes essentielles pour ces mesures et contraindre les paramètres cosmologiques afin d'affiner notre compréhension de l'Univers. En observant la fonction de masse des halos dans l'Univers il est alors possible de contraindre, en particulier, les paramètres cosmologiques Ω_m , σ_8 . C'est pourquoi la détection d'un grand nombre d'amas à différents redshifts et pour différentes masses est essentielle pour améliorer les contraintes offertes par l'évolution de l'abondance des amas.

1.3 Les amas de galaxies, des sondes cosmologiques

Dans le modèle hiérarchique de formation des structures, la matière s'agrège pour former des structures de plus en plus grandes sous l'influence de la gravité. Au sommet de ce processus hiérarchique trône les amas de galaxies, regroupements de centaines de galaxies liées par l'effondrement gravitationnel. Dans cette section, nous décrivons plus en détail la nature des amas et leur définition en termes de masse et de rayon. Ensuite, nous détaillons les approches par lesquels les amas de galaxies sont observés : par leur matière lumineuse dans le visible et l'Infrarouge, par le gaz dans les rayons X, par leur gaz dans le millimétrique ou en optique par l'effet de lentille gravitationnelle faible puis la caractérisation qu'ils apportent pour les amas. Enfin, nous concluons cette section par un bref état de l'art de la détection des amas de galaxies par l'effet de lentille gravitationnelle faible.

1.3.1 . Définition, masse et rayon des amas de galaxies

Nous avons vu que l'évolution des structures est un processus par lequel des halos de matière noire se forment par effondrement gravitationnel puis attirent continuellement la matière autour sous l'effet de la gravité. En s'agréant peu à peu, la matière se rassemble et forme des structures de plus en plus massives. Les amas de galaxies sont le stade final de cette évolution, c'est pourquoi il est courant de parler de processus hiérarchique dans lequel les amas sont "les structures liées par la gravitation les plus massives de l'Univers". Dans cette sous-section, nous décrivons les propriétés des amas de galaxies et les définitions que nous utilisons concernant leur masse et leur rayon.

Au sein de la toile cosmique, les amas de galaxies correspondent aux nœuds, c'est-à-dire aux intersections des filaments de matière. Les amas de galaxies correspondent ainsi au regroupement de centaines ou milliers de galaxies avec entre elles du gaz chaud, et se sont formés lorsque la densité de matière dominait la densité d'énergie de l'Univers (voir Sect. 1.2.1). Ces structures se situent donc à bas redshift, $z < 3$ et peuvent s'étendre sur plusieurs millions d'années lumière, soit $\sim 1 - 10$ Mpc.

La composition des amas de galaxies est particulièrement intéressante car elle reflète la distribution de la matière dans l'Univers (Pratt et al., 2019). La distribution de masse des amas est ainsi dominée par la matière noire qui constitue $\sim 85\%$ de la masse des amas. La seconde composante en termes de masse est le gaz chaud entre les galaxies qui est appelé milieu intra amas (ICM, de l'anglais Intra Cluster Medium). Ce plasma chaud peut atteindre des températures correspondant à une énergie de 2 à 15 keV. L'ICM constitue à son tour $\sim 12\%$ de la distribution de masse des amas de galaxies. Les $\sim 3\%$ restants correspondent à la masse des galaxies, cette dernière étant principalement constituée par leur composante stellaire. Il apparaît immédiatement que l'observation des amas par leur matière baryonique ne sonde qu'une portion très faible de la distribution de masse totale de l'amas.

Nous avons vu dans la section 1.2.2 que la masse des amas est un point crucial de leur utilisation pour la cosmologie. Cependant, par définition, les amas de galaxies sont des objets étendus dans l'espace. C'est pourquoi les définitions normalisées de masse, de rayon et de taille sont des notions clés pour leur étude. L'approche classique est alors de définir une masse en fonction de propriétés caractéristiques à tous les amas.

La définition courante du rayon caractéristique des amas se focalise sur leur contraste de densité $\Delta \equiv \rho(r < R_\Delta)/\rho_{\text{crit}}$. Ainsi, sous l'hypothèse d'un amas de galaxies à distribution sphérique, le rayon

caractéristique R_Δ est défini comme le rayon de la sphère homogène contenant Δ fois la densité critique de l'Univers, ρ_{crit} (voir Eq. 1.9). Dans la pratique, les valeurs couramment utilisées pour le contraste de densité sont 200, 500 et 2500, qui délimitent des régions pour lesquelles les amas de galaxies sont virialisés (Peebles, 1993; Peacock, 1999). La masse associée à ce rayon est ensuite calculée par la formule :

$$M_\Delta = \frac{4}{3}\pi \Delta \rho_{\text{crit}} R_\Delta^3. \quad (1.28)$$

Rappelons que dans la pratique cette définition est dépendante du redshift et de la masse. La figure 1.2 trace ainsi la taille apparente de différents amas en fonction du redshift et de la masse. Pour donner un ordre de grandeur, la masse M_{200c} , à savoir la masse contenue dans une sphère dont la densité est de 200 fois la densité critique de l'Univers, des halos de matière noire que nous étudions dans ce manuscrit est de l'ordre de $5 \cdot 10^{13} - 5 \cdot 10^{15} M_\odot$, ce qui correspond à un rayon R_{200c} de l'ordre de 1 – 6 Mpc.

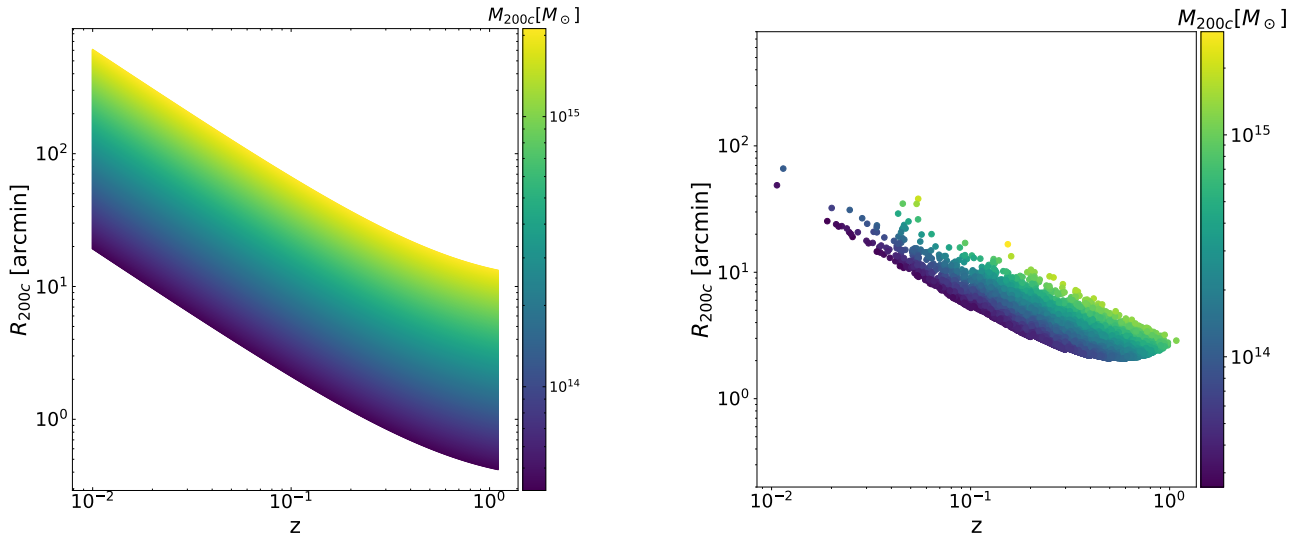


Figure 1.2 – Rayons apparents des amas de galaxies théoriques (gauche) et issus de la simulation DUSTGRAIN-Patfinder (droite, voir Sect. 2.3) en fonction du redshift et de la masse.

Dans le domaine théorique, pour décrire la distribution de matière noire dans les amas, il est courant d'utiliser des profils de masse. Le profil de masse le plus largement utilisé est le profil de densité de masse universel défini par Navarro et al. (1996) à partir de simulations à N corps et baptisé NFW. Ce profil s'écrit sous la forme :

$$\rho_{\text{NFW}}(R) = \frac{\rho_s}{(R/R_s)(1 + (R/R_s))^2}, \quad (1.29)$$

où ρ_s est le paramètre de densité central et R_s est le rayon spécifique d'échelle et caractérisent l'époque de formation de l'amas. Dans la pratique, le profil NFW d'un amas est entièrement caractérisé par sa masse M_Δ et la concentration associée, c_Δ :

$$c_\Delta = \frac{R_\Delta}{R_s}, \quad (1.30)$$

$$\rho_s \equiv \rho_{\text{crit}}(z) \delta_c, \quad \text{tel que} \quad (1.31)$$

$$\delta_c = \frac{\Delta}{3 \ln(1 + c_\Delta) - c_\Delta/(1 + c_\Delta)}. \quad (1.32)$$

La concentration c_Δ nous fournit ainsi une loi d'échelle supplémentaire à travers la relation c-M pour les amas détectés.

D'autres définitions pour l'estimation de la masse ou du rayon des amas de galaxies existent à partir des propriétés thermodynamiques du ICM. Ces définitions ne sont pas utilisées pour le travail exposé dans ce manuscrit mais sont décrites dans la revue de Pratt et al. (2019).

En résumé, les amas étant des objets diffus, différentes définitions de masse et de rayon existent. Dans la suite de ce manuscrit, nous utilisons les rayons et masses caractéristiques R_{200c} et M_{200c} définis pour une densité de contraste égale à 200.

1.3.2 . La détection des amas

Nous avons montré le rôle crucial des amas de galaxies pour la cosmologie et, tout particulièrement pour mettre à l'épreuve le modèle cosmologique. Pour contraindre les paramètres cosmologiques, il nous faut cependant un grand nombre d'amas et une estimation fiable de leur masse et de leur redshift. Dans cette sous-section, nous décrivons les spécificités de l'observation des amas selon leur différentes composantes : la matière lumineuse en optique, le gaz en rayons X ou par l'effet Sunayev-Zeldovich et en optique par l'effet de lentille gravitationnelle faible.

★ Observation en optique et Infrarouge par la matière lumineuse

Historiquement, les premières observations des amas ont été faites en optique grâce à la lumière émise par les étoiles de leurs galaxies membres.

L'observation et l'identification des amas de galaxies en optique n'ont rien de récent. Les amas de galaxies correspondent à une concentration notable de luminosité dans une région du ciel (voir Fig 1.3, gauche). Dès la fin du XVIIIe siècle, Charles Messier (1784) et William Herschel (1785) repèrent des regroupements de galaxies curieux au cœur de deux constellations : la Vierge et la Chevelure de Bérénice (Messier, 1784). Ces regroupements de galaxies, connus aujourd'hui sous les noms d'amas de la Vierge et d'amas de la Chevelure de Bérénice sont ainsi les premiers amas de galaxies identifiés. Au cours des deux siècles suivants, de plus en plus d'amas de galaxies sont identifiés grâce au développement de télescopes de plus en plus puissants. Ces recensements aboutissent dès la seconde moitié du XXe siècle aux premiers catalogues d'amas (Abell, 1957; Zwicky et al., 1961; Abell et al., 1989). Ces catalogues regroupent ainsi la majeure partie des amas de galaxies proches connus et constituent les fondements d'une grande partie de notre compréhension moderne des amas.

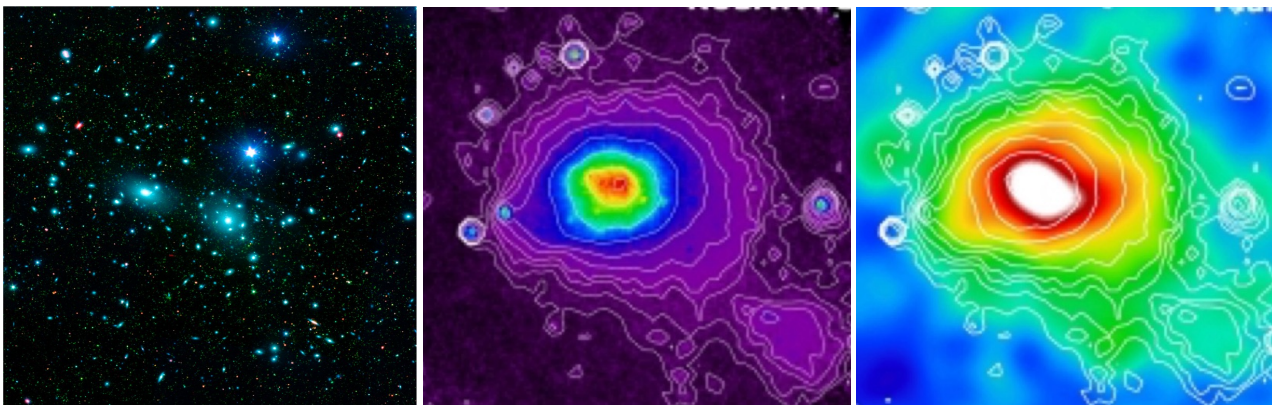


Figure 1.3 – Observations de l'amas de Coma (Abell 1656) dans le domaine visible et proche infrarouge (gauche), le domaine des rayons X (centre) et le domaine millimétrique par l'effet S-Z (droite). L'observation dans le visible combine les images visibles et infrarouges réalisées par le Sloan Digital Sky Survey avec le télescope spatial Spitzer. L'observation en rayons X provient du télescope ROSAT. L'observation par l'effet S-Z provient du télescope PLANCK avec superposition des contours en rayons X pour la comparaison des deux images. (Crédit : NASA/JPL-Caltech/L. Jenkins (GSFC); milieu : ESA/ LFI and HFI Consortia; Droite : Max-Planck-Institut für extraterrestrische Physik)

En utilisant des télescopes optiques et proche infrarouge au sol ou dans l'espace, les astronomes repèrent ces concentrations de lumière et identifient les amas de galaxies par la présence d'un grand nombre de galaxies. Le nombre de galaxies membres, appelé richesse est la première observable

dans le cas de la détection en optique. Dans la pratique, la position de ces galaxies est directement identifiée sur les images optiques.

La difficulté de la détection des amas de galaxies est alors d'estimer les spécificités des amas identifiés et notamment différencier les galaxies membres des galaxies d'avant et d'arrière plans. Étant sensible aux effets de projection, la détection en optique des amas s'appuie sur une information supplémentaire permettant leur identification et leur étude : la couleur des galaxies en leur sein, à savoir la distribution spectrale de la lumière que chaque galaxie émet ou réfléchit. Les galaxies observées sont réparties dans le plan couleur-magnitude afin d'identifier les galaxies membres de l'amas dans le champ. Dans ce plan, les galaxies d'un amas ont tendance à s'aligner le long d'une séquence de couleurs restreintes, généralement vers des couleurs plus rouges et des magnitudes plus lumineuses. On parle de la "séquence rouge". C'est sur ce principe que repose certains algorithmes de détection des amas dans le visible, tels que redMapper (Rykoff et al., 2014) ou CAMIRA (Oguri, 2014). Notons que les galaxies situées dans la séquence rouge sont principalement des galaxies elliptiques et des galaxies lenticulaires qui sont plus anciennes et contiennent généralement des populations stellaires plus anciennes et plus riches en métaux. Leur présence est représentative d'un amas régulier et stable.

Lors des observations optiques, l'identification des amas doit tenir compte d'effets systématiques tels que les effets de projections ou le contraste. Dans le cas des effets de projection, plusieurs amas de galaxies le long de la ligne de visée peuvent se superposer, ce qui complique la séparation et l'étude précise des amas individuels. Cette superposition peut impacter la distinction des galaxies membres de chaque amas.

Les possibilités qu'offre l'identification des amas dans le visible et l'infrarouge sont doubles. L'observation des regroupements de galaxies couplé aux méthodes de séquence rouge permet d'identifier les amas et de distinguer leurs galaxies membres des galaxies d'arrière-plans. Aujourd'hui, les observations des amas dans le visible et l'infrarouge entame une période florissante avec la mise en place de deux grands relevés : le relevé *EUCLID* (Laureijs et al., 2011) par le télescope spatiale du même nom, et le relevé Legacy Survey of Space and Time (Ivezić et al., 2019) par l'observatoire Vera Rubin. En effet, ces relevés devraient détecter jusqu'à 1000 fois plus d'amas de galaxies que le nombre total d'amas connus à ce jour et permettre d'améliorer notre compréhension de l'Univers (Sartoris et al., 2016).

Observation en X par le gaz

Historiquement, les amas de galaxies sont également observés par les rayons X dès les années 1970 (Forman et al., 1972) avec la détection de sources X au sein des amas de Persée et de Coma. L'observation des amas de galaxies par les rayons X offre également un regard profond sur la distribution de la matière et les processus thermodynamiques à l'œuvre en leur sein. Ce moyen d'observation se focalise sur la composante majoritaire (12%) baryonique au cœur des amas, à savoir le gaz dans le milieu intra-amas (ICM). Dans cette sous-section, nous résumons comment les amas de galaxies sont observés en X (voir Böhringer and Schartel (2013); Clerc and Finoguenov (2023) pour des revues).

Les amas de galaxies sont également des sources de rayons X ce qui permet leur détection mais l'observation des amas par les rayons X est limitée par l'atmosphère qui bloque ces longueurs d'ondes. Cette observation nécessite donc des télescopes spatiaux tels que *XMM-Newton* (Jansen et al., 2001) et *Chandra* (Weisskopf et al., 2000), qui sont à l'heure actuelle les télescopes les plus utilisés pour les relevés d'amas en rayons X.

Les émissions de rayons X des amas de galaxies proviennent principalement de l'émission Bremsstrahlung du gaz chaud de l'ICM. Tombant dans le potentiel gravitationnel formé par la matière noire, le gaz accélère et se réchauffe à hauteur de $\sim 1 - 10$ keV ce qui crée des sources étendues bien visibles (voir Fig 1.3 - panneau central). L'émission Bremsstrahlung, également appelée rayonnement de freinage, est un phénomène physique dans lequel des particules chargées, telles que les électrons, ralentissent en interagissant avec d'autres particules chargées, généralement des noyaux atomiques.

Lorsque ces particules chargées sont déviées par les champs électromagnétiques des noyaux, elles perdent de l'énergie cinétique. Cette perte d'énergie est émise sous forme de rayonnement électromagnétique, qui peut inclure des photons dans une certaine gamme de longueurs d'onde, y compris les rayons X. Physiquement, la brillance de surface de l'émission X est liée à la densité électronique (n_e) et à la température (T_e) du gaz par son flux, F_X :

$$F_X \propto \frac{1}{(1+z)^4} n_e^2 \cdot \sqrt{T_e} \cdot V, \quad (1.33)$$

où V est le volume du gaz dans l'amas. Ainsi, le flux mesuré d'un amas permet d'estimer sa densité électronique, nécessaire pour le calcul de la masse. Il est intéressant de noter que l'équation 1.33 présente une très forte dépendance avec le redshift, z . Cette dépendance est propre à l'utilisation du flux comme observable et résulte de la forte atténuation de la luminosité en fonction de la distance. Il est donc plus difficile de détecter les amas de galaxies à haut redshift à partir de leur rayonnement en X.

En parallèle, l'émission en X nous renseigne également sur les éléments présents dans le gaz à travers la distribution en énergie du spectre du milieu intra-amas (Böhringer and Werner, 2010). L'étude des raies spectrales illustre les transitions entre les niveaux d'énergie pour chaque élément du gaz et permet d'en déterminer l'abondance et la métallicité du gaz. La connaissance du spectre permet également d'estimer la température du gaz. Cependant, l'observation par l'émission X a montré une grande sensibilité vis à vis de l'état thermodynamique du gaz de l'amas (Rossetti et al., 2017). Cette sensibilité est forte car l'émission X dépend du carré de la densité électronique, n_e^2 . Cette dépendance peut impacter significativement la sélection des amas de galaxies et grandement affecter l'estimation précise de leur masse (Willis et al., 2021).

En résumé, nous avons vu que l'observation des émissions en X permet une détection des amas à partir de leur émission Bremsstrahlung.

★ Observation en millimétrique par l'effet Sunayev-Zeldovich

La troisième approche pour l'observation des amas est dans le domaine millimétrique à travers l'effet Sunyaev-Zeldovich (Zeldovich and Sunyaev, 1969; Sunyaev and Zeldovich, 1972). Dans cette sous-section, nous décrivons cette approche et ses enjeux.

L'effet Sunayev-Zeldovich est découvert et formalisé au début des années 1970 par R. A Sunyaev et Y. B. Zeldovich. Le phénomène, baptisé effet Sunayev-Zeldovich (SZ), à l'origine des anisotropies observées dans le spectre du fond diffus cosmologique (CMB), correspond physiquement à la diffusion Compton inverse des photons du CMB.

En d'autres termes, les photons provenant du CMB et passant à travers le gaz du milieu intra-amas diffusent sur les électrons libres du gaz. Cette diffusion s'observe alors par une modification de la forme du spectre du CMB le long de la ligne de visée en présence d'amas de galaxies. La Fig 1.3 (panneau droit) présente ainsi l'observation de l'effet SZ pour l'amas de Coma, nommé Abell 1656, observé par le télescope PLANCK (The Planck Collaboration, 2006). Sur cette image, les courbes du rayonnement en X sont également superposées.

Différentes causes de cet effet existent selon le processus physique à l'origine de l'énergie cinétique électrons qui est ensuite transmise aux photons. Parmi celles-ci, l'effet SZ thermique (tSZ) est le plus courant pour la détection des amas de galaxies et résulte de l'interaction entre les photons du CMB et les électrons chauds en mouvement à l'intérieur des amas de galaxies. La diffusion Compton inverse provoque un décalage du spectre du CMB à très haute énergie et entraîne un décalage des photons affectés par l'effet tSZ vers les grandes fréquences. Le spectre observé présente alors un excès au niveau des hautes énergies et déficit au niveau des basses énergies. Ces distortions forment un spectre caractéristique qui permet la détection des amas de galaxies. Cet effet correspond à un ordre de grandeur de 10^{-5} comparé au niveau de température du CMB et est ainsi du même ordre de

grandeur que les anisotropies du CMB. Sous l'hypothèse que le milieu est un gaz parfait, la mesure de l'effet SZ peut se faire à travers le paramètre de Compton, y :

$$y = \frac{\sigma_T}{m_e c^2} \int P_e dl, \quad (1.34)$$

où σ_T est la section efficace de diffusion Thompson, P_e la pression du gaz, et m_e la masse de l'électron. Cette mesure nous permet alors de remonter à nouveau à la pression du gaz, P_e qui permet de remonter à la masse de l'amas détecté. Il est intéressant de noter que le redshift des amas de galaxies n'a pas de réel impact sur leur détectabilité car l'amplitude de l'effet SZ est indépendante du redshift.

Comme mentionné précédemment d'autres composantes participent également à l'effet SZ. On trouve notamment deux autres composantes :

- L'effet SZ cinétique (kSZ), qui est dû au mouvement d'ensemble du gaz intra-amas par rapport au référentiel du CMB. Le déplacement global du gaz le long de la ligne de visée entraîne alors un décalage Doppler caractéristique dans le spectre observé du CMB (Birkinshaw, 1999).
- L'effet SZ relativiste (rSZ), qui correspond aux corrections relativistes sur les effets tSZ et kSZ. Ces corrections peuvent être utilisées pour contraindre la température du gaz et la masse des amas (Pointecouteau et al., 1998).

La différence entre ces effets transparaît par des dépendances spectrales différentes.

En résumé, dans le cas des relevés millimétriques, la détection des amas se fait par le spectre du CMB et son analyse, afin d'identifier les effets SZ en jeu. Cette approche n'est pas significativement impactée par le redshift ou les effets de projections.

★ Observation par l'effet de lentille gravitationnelle faible

En parallèle, l'observation des amas en optique peut se faire également par l'étude de l'effet de lentille gravitationnelle qu'ils induisent. Directement lié à la relativité générale et la courbure de l'espace-temps par une masse, cet effet correspond à la déviation de la trajectoire de la lumière par le champ gravitationnel de l'amas de galaxies présent. La déformation des images de galaxies sources lointaines nous renseigne alors sur la distribution totale de masse le long de la ligne de visée. Cet effet peut être fort et provoque l'observation d'images multiples d'une même galaxie d'arrière plan, ou faible et provoque de légères déformations de la forme des galaxies.

Dans le cas de l'effet de lentille gravitationnelle faible, également appelé lentillage faible, cette déformation est faible et nécessite un très grand nombre de galaxies sources lointaines pour être caractériser. La densité de galaxies sources est une composante essentielle des relevés de lentillage faible et conditionne en partie la possibilité de détecter les amas de galaxies. Cette approche permet ainsi de détecter les amas de galaxies mais aussi d'estimer précisément leur masse à partir de la déviation des rayons lumineux (Foëx et al., 2012; Miyazaki et al., 2018). Ainsi par définition, la détection par l'effet de lentille gravitationnelle faible dépend de la présence de la masse et n'est pas sensible à l'état thermodynamique du gaz. Étant le sujet principal de cette thèse, cet effet et son formalisme sont détaillés en profondeur dans la section 2.1.

Reposant sur des observations optiques, l'effet de lentille gravitationnelle faible est également sensible aux effets de projection et au redshift. En effet, le signal mesuré par l'effet de lentille gravitationnelle faible peut être notamment diluer par la présence de galaxies en avant-plan. Cependant, il est généralement possible de séparer le signal des différents amas alignés le long de la ligne de visée en tenant compte du redshift des galaxies sources par des approches dites tomographiques (Hamana et al., 2020). La dépendance de cet effet en fonction du redshift est très forte car le signal est proportionnel à une combinaison des redshifts des sources et de l'amas (voir section 2.1).

En résumé, nous avons vu que les amas de galaxies pouvaient être observés par l'effet de lentille gravitationnelle qu'ils induisent. Cet effet étant directement lié à la distribution totale de masse de l'amas, il est le seul permettant de détecter les amas en tenant compte de la composante de matière noire qui y domine.

1.3.3 . La caractérisation des amas

Afin de revenir à la cosmologie, il est essentiel de caractériser les amas détectés par les différentes observations. Dans cette sous-section, nous décrivons la caractérisation des amas en termes de redshift et de masse à partir des différentes observations.

★ La mesure du redshift

La mesure du redshift des amas est essentielle pour la fonction de masse et comprendre l'évolution des structures dans l'Univers. Pour un amas de galaxies composé de galaxies membres à des redshifts différents, la détermination du redshift de l'amas dans son ensemble peut être un processus plus complexe. Généralement, le redshift de l'amas de galaxies est associé au redshift moyen ou médian des galaxies contenues au sein de l'amas.

Indépendamment du domaine d'identification de l'amas, le redshift des galaxies membres est généralement obtenu par un relevé optique. Pour ce faire, deux techniques existent : la spectroscopie et la photométrie. Ces deux approches reposent sur l'estimation du spectre lumineux des galaxies afin d'estimer le redshift des galaxies membres et donc de l'amas de galaxies.

L'approche la plus précise pour déterminer le redshift des galaxies membres de l'amas est la spectrométrie. Cette méthode consiste à mesurer précisément l'ensemble du spectre lumineux et l'étudier en fonction des longueurs d'ondes de raies spectrales connues. En mesurant les décalages des raies spectrales caractéristiques, on peut déterminer le redshift des galaxies membres de l'amas.

La spectrométrie étant coûteuse en temps, son utilisation est limitée dans le cas des relevés contenant un grand nombre de galaxies. Seulement une faible portion des galaxies peuvent être étudiées par spectroscopie. C'est pourquoi les grands relevés privilégient également en parallèle souvent l'application de la méthode photométrique pour l'estimation du redshift des galaxies (Aihara et al., 2018a).

L'approche par photométrie mesure l'intensité lumineuse d'un objet dans différentes bandes spectrales, généralement à travers des filtres optiques. Ainsi, la photométrie peut être utilisée pour identifier les galaxies individuelles à l'intérieur de l'amas en mesurant leur luminosité dans différentes bandes sans obtenir le spectre complet (Mobasher et al., 2007; Gilbank et al., 2011). Cette approche a l'avantage d'être significativement plus rapide que la spectroscopie et peut être appliquée à un grand nombre de galaxies simultanément. Cette avantage est particulièrement pertinent dans les relevés actuels tel que EUCLID pour lesquels on s'attend à étudier deux milliards de galaxies. Dans la pratique, les redshifts estimés par photométrie sont une mesure approchée du redshift pour laquelle une calibration est souvent nécessaire à partir de redshifts spectrométriques (Newman and Gruen, 2022). Cependant, la photométrie peut être complémentaire à la spectrométrie pour des études de grande envergure.

Notons que l'effet de lentille gravitationnelle faible peut également proposer une première estimation du redshift d'un amas de galaxies à partir d'une approche tomographique. Dans la pratique, cet effet étant dépendant des redshifts des galaxies sources et de l'amas de galaxies, la sélection des galaxies sources pour différents intervalles de redshifts permet d'estimer un intervalle de redshift pour l'amas de galaxies.

★ L'estimation de la masse

Pour contraindre les paramètres cosmologiques à travers la fonction de masse, il est essentiel de caractériser les amas et de mesurer la masse individuelle de chaque amas observé. Cependant, dans la pratique peu de relevés d'amas permettent de donner précisément accès à cette masse individuelle. Généralement, les relevés se focalisent uniquement sur une observable liée à la masse baryonique et les relations d'échelle permettent de retrouver une estimation de la masse associée à cette observable.

- **Les mesures individuelles**

- i) **Par la dispersion des vitesses des galaxies**

La mesure de la dispersion des vitesses des galaxies membres peut être utilisée pour estimer la masse individuelle des amas (Andreon and Hurn, 2010; Munari et al., 2013; Ntampaka et al., 2015). Grâce au théorème du Viriel, la masse peut être reliée à la dispersion de vitesse circulaire σ_v par l'équation :

$$\sigma_v = \frac{GM}{3R}, \quad (1.35)$$

où R est le rayon du halo (Limber and Mathews, 1960; Spitzer, 1969). C'est notamment par cette approche qu'en 1933, Fritz Zwicky montra qu'il y avait un écart entre la masse des galaxies dans l'amas de Coma et le potentiel gravitationnel expliquant leur vitesse de rotation. Sous l'hypothèse que l'équilibre dynamique est atteint dans l'amas, le profil de masse M peut se calculer à partir des dispersions de vitesse radiale, σ_r , et tangentielle, σ_t (Zwicky, 1933). L'équation de Jeans pour un système sphérique s'écrit alors sous la forme :

$$GM(r) = -\frac{r^2}{n} \frac{d(n\sigma_r^2)}{dr} - 2r(\sigma_r^2 - \sigma_t^2), \quad (1.36)$$

avec n la densité de galaxies dans l'amas. Notons que cette mesure nécessite un nombre significatif de galaxies observées pour estimer la dispersion suffisamment précisément afin d'être efficace, voire de remonter à la masse de l'amas par le théorème du viriel. Ainsi, cette mesure dépend de la densité de galaxies membres observées et peut donc être d'autant plus complexe pour les amas à haut redshift pour lesquels l'observation des galaxies membres les plus lointaines est difficile. Notons que cette approche se montre plus efficace que l'estimation de la masse de l'amas de galaxies en considérant la définition du théorème du Viriel selon les équations décrites dans la section 1.3.1 (Andreon and Hurn, 2010; Munari et al., 2013).

- ii) **Par l'effet de lentille gravitationnelle faible**

L'effet de lentille gravitationnelle est sensible à toute la distribution de masse (matière noire et matière baryonique) et permet ainsi d'obtenir une estimation fiable de la masse des amas (Miyazaki et al., 2018).

Estimer la masse des amas de galaxies à partir de l'effet de lentille gravitationnelle faible consiste à réaliser un ajustement du cisaillement induit par l'amas par un modèle théorique tel que celui proposé par Navarro et al. (1996). Dans la pratique, la masse peut être estimée à partir du profil radial du cisaillement tangentiel moyenné dans un anneau (Bartelmann and Schneider, 2001; Miyazaki et al., 2018). Puis, on calcule la densité de masse surfacique différentielle en fonction de la distance radiale projetée, R :

$$\Delta\Sigma(R) = \bar{\Sigma}(< R) - \Sigma(R), \quad (1.37)$$

avec $\bar{\Sigma}(< R)$, la densité de masse surfacique moyennée à l'intérieur du rayon, et,

$$\Sigma(R) = 2 \int_{\text{inf}}^0 \rho(R, x) dx. \quad (1.38)$$

$\Delta\Sigma(R)$ est directement proportionnelle au cisaillement tangentiel. Son estimation pour différents rayons, R , est ensuite ajustée par rapport à un modèle théorique. Cette approche permet de contraindre les paramètres du modèle utilisé, et d'estimer la masse de l'amas associé.

La mesure de la masse par l'effet de lentille gravitationnelle faible nécessite cependant la présence d'un grand nombre de galaxies sources lointaines le long de la ligne de visée pour estimer précisément le cisaillement observé. Cette condition est remplie pour les amas détectés par effet de lentille gravitationnelle faible ce qui permet de construire un catalogue d'amas de galaxies pour lequel on a un signal suffisamment fort pour garantir une mesure plus précise

de la masse, généralement réalisée aussi par effet de lentillage faible (Miyazaki et al., 2018). C'est pourquoi, la masse estimée par lentillage faible peut être utilisée pour calibrer les masses estimées par les autres approches (Murata et al., 2018).

iii) Par le gaz en rayons X ou l'effet SZ

Pour estimer la masse de l'amas détecté, on fait l'hypothèse que l'équilibre hydrostatique est atteint pour l'amas. Cette hypothèse d'équilibre hydrostatique implique que la pression interne, exercée par les particules en mouvement, contre la force gravitationnelle qui résulte de la masse totale de l'amas et crée une situation stable. Cet équilibre prévient l'amas de s'effondrer sous l'effet de la gravité ou à l'inverse de se dilater en raison d'une pression excessive. Dans le cas d'un amas à symétrie sphérique, la pression électronique du gaz, P_e s'écrit :

$$\frac{dP_e}{dr} = -\rho g, \quad (1.39)$$

avec ρ est la masse volumique du gaz et g l'accélération gravitationnelle. La masse est ensuite estimée à partir de l'expression de g pour une masse contenue dans une sphère de rayon r :

$$g = \frac{GM(< r)}{r^2}. \quad (1.40)$$

Le milieu intra-amas présentant une densité électronique faible (10^{-2} électrons par centimètre cube), on peut le considérer comme un gaz parfait. Dans la pratique, les mesures de densité et de température permettent alors de revenir à la pression électronique du gaz à travers l'équation d'état des gaz parfaits :

$$P_e = n_e k_B T_e = \frac{\rho}{\mu m_{\text{proton}}} k_B T_e, \quad (1.41)$$

où k_B est la constante de Boltzmann, m_{proton} est la masse atomique du proton 1.66×10^{-27} kg, et μ est le poids moléculaire moyen.

La combinaison des équations 1.39, 1.40 et 1.41 permet alors d'estimer le profil de masse de l'amas détecté. Ce profil s'écrit généralement sous la forme :

$$M(r) = -\frac{r P_e}{\mu m_{\text{proton}} G n_e} \frac{d \ln(P_e)}{d \ln(r)}. \quad (1.42)$$

Les amas sont en vérité des systèmes dynamiques complexes et l'hypothèse d'équilibre hydrostatique introduit alors un biais significatif sur l'estimation de la masse. Ce biais est d'autant plus fort pour les amas perturbés (Pratt et al., 2019).

Dans le cas des relevés par l'effet Sunayev-Zeldovich (SZ), l'observable est le paramètre de Compton intégré, Y_Δ , qui retrace la pression thermique du milieu intra-amas (Planck Collaboration et al., 2014). Ce paramètre permet notamment de se ramener à la luminosité de l'émission X et à la masse (Arnaud et al., 2007; Melin et al., 2011).

$$Y_\Delta = \int_0^{R_\Delta} \int_0^{2\pi} y dS. \quad (1.43)$$

En revenant à la pression du gaz grâce aux équations 1.34 et 1.41, on peut estimer la masse individuelle de l'amas de galaxies.

• Par les relations d'échelle

La mesure de masse individuelle pour des grands échantillons d'amas de galaxies est cependant très coûteuse en temps et fastidieuse. Il est donc intéressant d'utiliser des relations d'échelle statistiquement robustes, qui sont capables de relier la masse totale des amas à des quantités plus facilement observables (voir Giardini et al., 2013, comme revue).

Ces relations d'échelle s'appuient sur la physique de la formation et de l'évolution des amas de galaxies et les observables de la masse : la richesse en optique, la luminosité de l'émission X,

la température du milieu intra-amas, et le paramètre de Compton intégré par l'effet SZ. Les différentes observables physiques, O , sont ainsi reliées à la masse totale, M_{Δ} , par une expression de la forme (Pratt et al., 2019) :

$$O \propto A(z) M_{\Delta}^{\alpha}, \quad (1.44)$$

qui tient compte du redshift et de la cosmologie à travers le facteur $A(z)$.

Dans le cas des relevés par l'effet Sunayev-Zeldovich et pour une distribution de masse sphérique, cette relation prend la forme :

$$Y_{\Delta} \propto H(z)^{2/3} M_{\Delta}^{5/3}. \quad (1.45)$$

Dans le cas des relevés en rayons X, la luminosité en X (bolométrique), L_X , est reliée à la masse telle que (Böhringer et al., 2014; Pacaud et al., 2018) :

$$L_X \propto H(z)^{-7/3} M_{\Delta}^{4/3}. \quad (1.46)$$

Ces relations d'échelle sont alors calibrées sur des sous-échantillons (généralement quelque dizaines d'amas) pour lesquels la masse individuelle a été déterminée par l'une des méthodes de mesure précédentes. En parallèle, il est également essentiel d'avoir une calibration fine de la dispersion intrinsèque de la distribution statistique étudiée afin de réduire les biais lors de l'estimation de la masse (Murata et al., 2018; Pratt et al., 2019; Chiu et al., 2022).

En résumé, nous avons vu la caractérisation des amas en termes de redshift et de masse selon la composante par laquelle ils sont observés. La méthode de lentille gravitationnelle faible montre des complémentarités avec les autres approches et un avantage direct pour la détection en permettant de cibler la masse totale de l'amas. Pour autant, les observations optiques et en rayons X fournissent des informations complémentaires sur le contenu baryonique et les propriétés physiques des amas de galaxies.

1.3.4 . Pourquoi maintenant : HSC, EUCLID et Vera C. Rubin

Nous avons montré que l'effet de lentille gravitationnelle faible est la seule approche capable de détecter et caractériser des amas de galaxies par leur distribution totale de masse. Dans cette sous-section, nous présentons un bref historique de la détection des amas de galaxies par effet de lentille gravitationnelle faible et son importance dans le contexte des relevés optiques actuels.

Historiquement, les premiers relevés optiques ayant permis la détection des amas de galaxies par effet de lentille gravitationnelle faible datent du début des années 2000 (Wittman et al., 2006; Gavazzi and Soucail, 2006; Schirmer et al., 2007). Cependant, ces premiers relevés présentent seulement quelques détections d'amas de galaxies. Ce faible nombre de détections s'explique par les caractéristiques de ces relevés (petite taille de champ, faible densité de galaxies sources) qui ne permettent pas une détection optimale des amas de galaxies par l'effet de lentille gravitationnelle faible. Le cisaillement gravitationnel étant très faible par rapport à la forme intrinsèque des galaxies. Il faut un grand nombre de galaxies derrière l'amas afin d'avoir une mesure du cisaillement précise et pouvoir mieux calibrer les biais associés à cette mesure. Le signal lentille étant dépendant de la position et du nombre des galaxies sources derrière l'amas. Il faut une certaine densité de galaxies sources pour permettre la détection des amas de galaxies.

Nous pouvons résumer les conditions optimales pour la production de catalogues d'amas de galaxies détectés par effet de lentille gravitationnelle faible en trois points clés :

- La qualité des images doivent être suffisantes pour détecter des galaxies sources proches et lointaines et pouvoir estimer précisément leurs ellipticités,
- Une densité de galaxies sources, $n_g > 20$ galaxies.arcmin⁻², afin de permettre la détection du signal lentille (seuil pour environ 1 amas détecté par deg²),

Table 1.1 – Etat de l’art des performances des différents relevés d’amas par l’effet de lentille gravitationnelle faible et leurs caractéristiques associées. Les nombres entre parenthèses dans la quatrième colonne correspondent aux détections d’amas de galaxies qui ont été confirmées à partir de correspondances trouvées dans d’autres relevés.

Relevé/Télescope	$n_g [\text{arcmin}^{-2}]$	Aire [deg^2]	Détections	Publication
Mayall	19	7.5	2 (2)	Wittman et al. (2006)
CFHTLens	35	3.6	2 (2)	Gavazzi and Soucail (2006)
MPG/ESO	12	19	17 (11)	Schirmer et al. (2007)
Subaru	46	2.2	4 (3)	Miyazaki et al. (2007)
CFHTLens	11.2	55	51 (30)	Shan et al. (2012)
HSC/Subaru	34.5	~ 3	7 (7)	Utsumi et al. (2014)
HSC/Subaru	24	9	8 (8)	Hamana et al. (2015)
HSC/Subaru	20.9	2.3	8 (8)	Miyazaki et al. (2015)
HSC/Subaru	21.8	~ 160	65 (63)	Miyazaki et al. (2018)
HSC/Subaru	19.3	~ 120	125 (107)	Hamana et al. (2020)
HSC/Subaru	~ 15	~ 510	187 (182)	Oguri et al. (2021)

- Une large taille de relevé pour avoir un échantillon suffisamment large car les amas sont des objets "rares".

La table 1.1 met en lumière les résultats de plus en plus prometteurs des méthodes de détections des amas de galaxies par effet de lentille gravitationnelle faible. Ces résultats sont quantifiés à travers le nombre de détections obtenues (quatrième colonne) avec entre parenthèses, le nombre de détections confirmées par l’intermédiaire d’autres relevés. Nous observons notamment que la mise en place du relevé HSC-SSP (Aihara et al., 2018a) a permis une augmentation prodigieuse du nombre d’amas de galaxies détectés. Ces nouveaux relevés ont multiplié de deux ordres de grandeur le nombre d’amas détectés. Les données du relevé HSC seront bientôt suivies par celles issues du télescope *EUCLID* (Laureijs et al., 2011) lancé en Juillet 2023, et celles du Vera C. Rubin (Ivezic et al., 2019) et du Nancy Grace Roman Telescope (Anciennement WFIRST Spergel et al., 2015). Les caractéristiques attendues pour ces différents relevés sont présentées dans la figure 1.4 sur laquelle nous rajoutons pour la comparaison les relevés COSMOS (Scoville et al., 2007), Dark Energy Survey, DES (Dark Energy Survey Collaboration et al., 2016) et Kilo-Degree Survey, Kids (Kuijken et al., 2015). La couleur des points fait référence aux caractéristiques des relevés vis à vis des conditions optimales pour la détection des amas de galaxies par effet de lentille faible. Le vert correspond aux relevés pour lesquels les trois conditions précédentes sont atteintes tandis que les points rouges illustrent respectivement les relevés pour lesquels les conditions sont partiellement atteintes et non atteintes.

Dans ce contexte, il est important de mentionner que cette thèse s’inscrit comme un travail préparatoire aux données à venir du télescope *EUCLID*. Comme le montre la figure 1.4, les données attendues pour *EUCLID* devraient remplir parfaitement les conditions pour permettre d’augmenter d’un facteur 10 le nombre d’amas de galaxies détectés par effet de lentille gravitationnelle que HSC. Nous présenterons plus en détails le télescope *EUCLID* et les données dans le chapitre 7.

Dans cette section, nous avons vu que les amas de galaxies sont des structures complexes et dont les paramètres caractéristiques, que sont la masse, le rayon et le redshift, peuvent être envisagées sous différentes définitions. Nous avons montré que les amas de galaxies pouvaient être observés et identifiés par trois approches : dans le visible, dans les rayons X par l’émission Bremsstrahlung et dans le domaine millimétrique par l’effet S-Z. Les trois approches principales que nous avons présentées se focalisent toutes sur la matière baryonique (étoiles, gaz) composants les amas de galaxies afin, par la suite, d’en estimer leur masse. Cependant, nous avons vu que les amas de galaxies sont composés de 15% seulement de matière baryonique. La détection dans le visible par l’effet de lentille gravitationnelle faible est la seule méthode qui permet de tenir compte également des 85% de matière noire

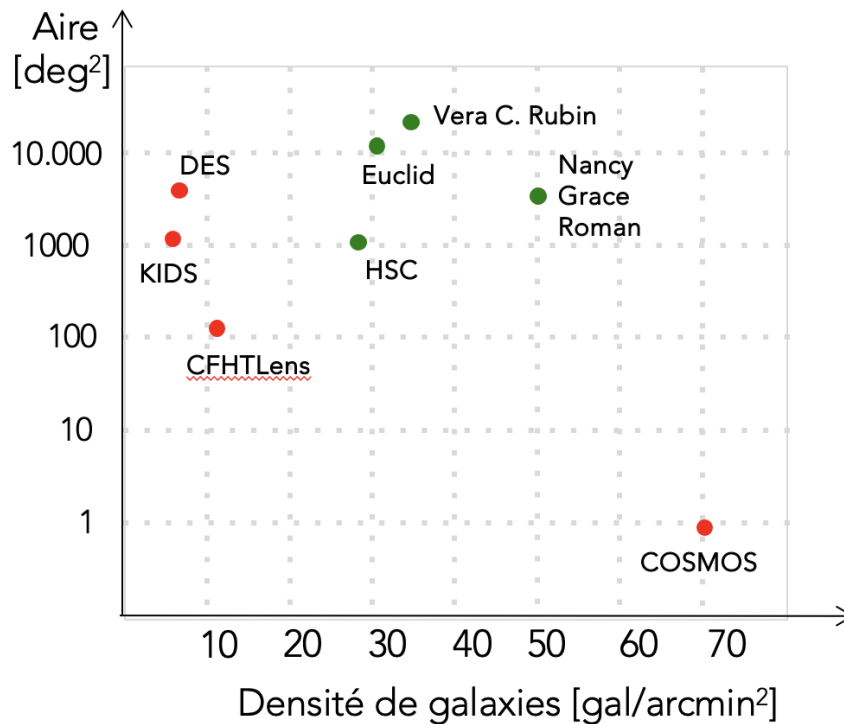


Figure 1.4 – Illustration des propriétés des différents grands relevés optique pour l'étude de l'effet de lentille gravitationnelle faible. Les points verts correspondent aux relevés remplissant les conditions optimales pour la détection des amas. Les points rouges correspondent respectivement aux relevés pour lesquels les conditions optimales pour la détection des amas sont partiellement remplies et pas remplies.

qui compose également les amas de galaxies.

1.4 Conclusions

En tant que toile de fond de la cosmologie, le modèle standard Λ CDM décrit la dynamique et la géométrie de l'Univers. Dans ce chapitre, nous avons vu que ce modèle, reposant sur la relativité générale et les équations de Friedmann permettait de décrire un Univers homogène et isotrope en expansion au sein duquel l'énergie sombre et la matière noire dominent. Localement, les fluctuations de densité ont entraîné la formation et l'évolution des grandes structures. Aujourd'hui, la distribution des grandes structures reste encore un sujet d'étude très actif et se révèle dépendante des paramètres du modèle cosmologique. Étant la phase finale de cette évolution, les amas de galaxies constituent d'excellents traceurs pour la cosmologie. En particulier, la distribution des amas de galaxies permet de remonter à l'état dynamique de l'Univers récent à travers la fonction de masse. Celle-ci relie la distribution des amas à une masse et un redshift donnés en fonction de la cosmologie.

Pour tirer pleinement partie de cette fonction de masse et sonder la cosmologie à partir des amas de galaxies, il est nécessaire de détecter un grand nombre d'amas représentatif de la population totale. Nous avons ainsi présenté les trois domaines des longueurs d'ondes dans lesquels les amas sont observés :

- en optique, à travers les étoiles représentant $\sim 3\%$ du contenu en matière des amas ou par l'effet de lentille gravitationnelle induit par la masse totale;
- en rayons X, par l'émission du gaz du milieu intra-amas, représentant $\sim 12\%$ du contenu en matière des amas;
- en millimétrique, par l'effet Sunayev-Zeldovich qui trace la pression électronique dans le milieu intra-amas.

La méthode de détection par l'effet de lentille gravitationnelle est très prometteuse car elle permet

de détecter les amas de galaxies directement à partir de leur distribution totale de masse. Par ailleurs, cette méthode est d'autant plus performante pour lier à termes le signal détecté à la masse.

Les relevés les plus récents et à venir (HSC, *EUCLID*, Vera C. Rubin, Nancy Grace Roman) font entrer la détection des amas de galaxies par effet de lentille gravitationnelle faible dans son âge d'or. À partir de ces relevés, nous pouvons espérer détecter un nombre suffisant d'amas de galaxies par l'effet de lentille gravitationnelle pour produire des catalogues capables de mieux caractériser les masses des amas et d'affiner, à termes, les contraintes cosmologiques.

*
* *

2

Une observation optimale par l'effet de lentillage faible?

*Il y a plus de mystères dans le ciel, Horatio
Que n'en peut contenir toute notre philosophie.*

William Shakespeare, Hamlet

2.1	L'effet de lentille gravitationnelle	30
2.1.1	La vie chaotique des photons soumis à la gravitation	30
2.1.2	L'équation lentille	32
2.1.3	Le potentiel effectif projeté et la convergence	34
2.1.4	La distortion des images et le cisaillement	35
2.1.5	Inversion de masse et convergence	37
2.2	Données de lentille gravitationnelle faible	38
2.2.1	Mesure de cisaillement et catalogue de galaxies sources	38
2.3	Données simulées	40
2.3.1	Description de la simulation	40
2.3.2	Le catalogue de halos de matière noire	41
2.3.3	Le catalogue de galaxies sources	41
2.4	Conclusions	42



Au sein d'un Univers en expansion et dominé par les composantes sombres (énergie noire et matière noire), les amas de galaxies ont une place toute particulière. Dans le chapitre précédent, nous avons décrit les différentes approches permettant la détection des amas de galaxies avec un focus tout particulier sur les effets de sélection de ces approches comparées à l'effet de lentille gravitationnelle. Dans ce chapitre, nous détaillons l'effet de lentille gravitationnelle de l'observation à la production des catalogues de galaxies sources par les relevés. En section 2.1, nous présentons l'effet de lentille gravitationnelle et son formalisme dans le cadre de la détection des amas de galaxies. Dans la section 2.2, nous introduisons les catalogues de cisaillement et leur production. Enfin, dans la section 2.3, nous présentons la simulation cosmologique et les données simulées qui en sont dérivées et qui sont à la base des études dans ce manuscrit (chapitres 3 à 7).

2.1 L'effet de lentille gravitationnelle

L'effet de lentille gravitationnelle est la seule méthode d'observation capable de détecter les amas à partir de la distribution totale de masse, c'est-à-dire en intégrant aussi la matière noire. Ce qui était autrefois un simple phénomène observé étonnant de l'Univers est devenu aujourd'hui un puissant outil pour la cosmologie (e.g [Munshi et al., 2008](#)). Dans cette section, nous détaillons ce que représente l'effet de lentille gravitationnelle faible tant sur le plan théorique que pour l'observation des amas. Nous détaillons en particulier le formalisme de l'effet de lentille gravitationnelle et ses composantes caractéristiques que sont la convergence et le cisaillement. Enfin, nous présentons l'algorithme d'inversion de masse de [Kaiser and Squires \(1993\)](#) qui relie la convergence et le cisaillement observé. Plus de détails peuvent être trouvés dans les revues : ([Narayan and Bartelmann, 1996](#); [Bartelmann and Schneider, 2001](#); [Umetsu, 2020](#)).

2.1.1 . La vie chaotique des photons soumis à la gravitation

Nous avons vu dans le chapitre précédent que l'Univers est régi par la gravité. La gravité, omniprésente, dessine la toile cosmique et gouverne la propagation de toutes les particules libres. La théorie de la Relativité Générale d'Albert Einstein (1911) prédit que cette gravité affecte aussi les photons et a pour effet de dévier leur trajectoire. Dit autrement, le champ gravitationnel des objets massifs, comme les amas de galaxies, affecte la trajectoire de la lumière dans leur voisinage. Les objets massifs agissent ainsi telles des lentilles qui courbent la trajectoire de la lumière qui passent au travers; c'est pourquoi on parle d'effet de lentille gravitationnelle. Dans cette sous-section, nous décrivons le cadre théorique de cet effet et les hypothèses pour la mise en place de son formalisme.

Dans le cadre des équations de Friedmann pour régir la géométrie de l'Univers (voir Sect 1.1) et en considérant que les inhomogénéités de matière sont des perturbations locales, il est possible d'étudier en détails l'effet de lentille gravitationnelle en séparant l'espace en trois zones de l'espace :

- Le photon part de sa source et suit une trajectoire non perturbée jusqu'à proximité d'un objet massif, la lentille,
- Le photon entre dans le champ gravitationnel de l'objet massif dans lequel il suit des géodésiques ($ds = 0$) et sa trajectoire est ainsi courbée.
- Le photon quitte le champ gravitationnel de la lentille et suit une trajectoire non perturbée jusqu'à nous.

La géométrie du système lentille est illustrée dans la figure 2.1 dans le cas d'une masse ponctuelle, M . Cette géométrie définit un système optique composé alors d'une lentille, la masse ponctuelle et d'une image source observée, par analogie avec les conventions en optique. Ce modèle s'accompagne de deux hypothèses :

- On considère que la longueur du système optique est très grande devant la taille de la lentille. Ce faisant, on se place alors dans l'approximation des lentilles minces. Cette approximation est vérifiée en particulier dans le cas des amas de galaxies dont la taille d'environ quelques Mpc est petite devant la taille du système d'environ quelques Gpc.
- Pour étudier finement l'effet de lentille gravitationnelle au voisinage de la lentille, on considère que l'espace-temps plat est faiblement perturbé par le potentiel gravitationnel, Φ de la lentille et que la vitesse de la lentille est faible devant c . Cette hypothèse permet de se placer dans un espace-temps localement perturbé où on néglige la courbure et l'expansion de l'Univers lors de la déviation. Dans le cas des amas de galaxies, la vitesse relative est d'environ $10^3 \text{ km.s}^{-1} \ll c$ et le potentiel gravitationnel est tel que $\frac{|\Phi|}{c^2} \ll 10^{-4}$.

Dans cette géométrie plane pour commencer, le photon émis par une source lointaine suit une trajectoire non perturbée jusqu'à atteindre au paramètre d'impact, b , le champ gravitationnel induit par la lentille (L) assimilée à une masse ponctuelle. La trajectoire de la particule est alors courbée ce qui induit un angle de déviation, $\hat{\alpha}$, entre la trajectoire d'entrée (trait en pointillés) et la trajectoire de

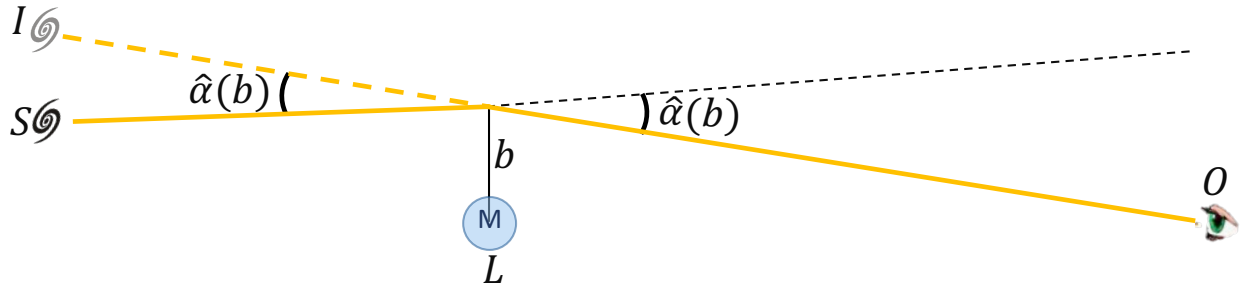


Figure 2.1 – Schéma de la géométrie typique de l'effet de lentille gravitationnelle formée par la source (S), la lentille (L) ponctuelle de masse M , l'image observée (I) et l'observateur (O). La trajectoire du photon (trait jaune) est déviée aux environs du paramètre d'impact b , et on note $\hat{\alpha}$, l'angle de déviation induit par le champ gravitationnel de la masse M . L'angle de déviation, $\hat{\alpha}(b)$ mesure l'écart entre la trajectoire d'entrée sans lentille (trait en pointillés) et la trajectoire observée (trait jaune).

l'image observée (trait jaune). Cette déviation peut être dérivée dans le cadre de l'espace-temps de Minkowski (voir Sect 1.1), pour un photon le long de l'axe polaire :

$$ds^2 = -\left(1 + \frac{2}{c^2}\Phi\right)c^2dt^2 + \left(1 + \frac{2}{c^2}\Phi\right)^{-1}dr^2 = 0, \quad (2.1)$$

où Φ est le potentiel Newtonien, et c la célérité de la lumière dans le vide. La vitesse effective du rayon lumineux dans le champ gravitationnel s'écrit ainsi :

$$v^2 = \frac{dr^2}{dt^2} = c^2 \left(1 + \frac{2}{c^2}\Phi\right)^2. \quad (2.2)$$

Par analogie avec l'optique, on introduit ainsi l'indice de réfraction de la lentille, n , qui est responsable de cet angle de déviation, $\hat{\alpha}$, (e.g. [Schneider et al., 1992](#)) tel que :

$$\left|\frac{dr}{dt}\right| = \frac{c}{n} \text{ soit } n = \left(1 + \frac{2}{c^2}\Phi\right)^{-1} \simeq \left(1 - \frac{2}{c^2}\Phi\right), \quad (2.3)$$

Ainsi, en intégrant le long de la trajectoire, nommée T , l'angle de déviation, $\hat{\alpha}$ équivaut à :

$$\hat{\alpha} = \int_T \vec{\nabla}_\perp n \cdot d\vec{l} = \frac{2}{c^2} \int_T \vec{\nabla}_\perp \Phi \cdot d\vec{l}. \quad (2.4)$$

où $\vec{\nabla}_\perp$ est la composante de l'opérateur gradient, perpendiculaire au rayon lumineux. Dans le cas où le paramètre d'impact $b = |\vec{b}|$ est grand devant le rayon de Schwarzschild de la lentille $R_s \equiv \frac{2GM}{c^2}$, le potentiel Newtonien de la lentille s'exprime simplement comme (e.g. [Bartelmann and Schneider, 2001](#)) :

$$\Phi(b, l) = -\frac{GM}{(b^2 + \Delta l^2)^{1/2}}, \quad (2.5)$$

avec b , le paramètre d'impact et Δl correspond à la variation progressive de la courbure le long de la trajectoire selon la distance au point le plus proche de la lentille. Dans l'expression du potentiel Newtonien, ce paramètre rend compte de l'espace-temps courbe au voisinage de la masse. Notons que $\hat{\alpha}$ peut également être exprimé en fonction du rayon de Schwarzschild de la masse ponctuelle, tel que prédit dans la théorie de la Relativité Générale :

$$\hat{\alpha}(b) = \frac{4GM}{c^2 b} = \frac{2R_s}{b}. \quad (2.6)$$

Notons que ces résultats restent valables en considérant une géométrie dans un espace tridimensionnel et avec une lentille à symétrie sphérique. Il est intéressant de noter également que l'angle de

dévié est proportionnel à la masse de lentille et inversement proportionnel au paramètre d'impact, ce qui indique que la déviation est d'autant plus faible que le photon est loin de la masse lors de son passage.

Historiquement, cette formule (Eq. 2.6) constitue le premier succès de la théorie de la Relativité Générale lorsque Arthur Eddington (1919) a confirmé par l'observation l'angle de déviation de rayons lumineux frôlant le Soleil prédit par Einstein.

Si on considère maintenant un objet avec une distribution de masse M , telle que $M(b)$ est la masse comprise dans une sphère de rayon b , alors par symétrie du système, l'angle de déviation α au paramètre d'impact b s'écrit :

$$\hat{\alpha}(b) = \frac{4GM(b)}{c^2 b}. \quad (2.7)$$

Cette déviation peut ainsi être causée par les différents objets dans le cosmos à différents niveaux. Dans le cas des étoiles, comme le Soleil, l'angle de déviation est alors de l'ordre de quelques microsecondes d'arc. Dans le cas des galaxies, l'angle de déviation atteint quelques arcsecondes. Dans le cas des amas de galaxies, qui sont les lentilles gravitationnelles les plus puissantes observées, l'angle de déviation est de quelques arcminutes. L'importance de l'angle de déviation et de la distorsion observée nous permet ainsi de qualifier l'effet de lentille gravitationnelle de faible ou de fort. Notons que les structures à grandes échelles peuvent aussi agir comme des lentilles gravitationnelles, cependant l'approximation des lentilles minces n'est plus valable dans leur cas (voir [Narayan and Bartelmann 1996](#) comme revue). Ainsi les effets de lentilles sont visibles à toutes les échelles de l'Univers, différentes structures, telles que les amas de galaxies, les filaments ou les galaxies individuelles, peuvent agir comme des lentilles et créer cet effet de déformation.

2.1.2 . L'équation lentille

L'effet de lentille gravitationnelle se caractérise par un écart entre la position de la source observée et réelle. Pour étudier cet effet, il est essentiel de revenir à la position exacte de la source, à partir de l'angle de déviation induit par le champ gravitationnel de la lentille. Dans cette sous-section, nous définissons cette relation, appelée l'équation lentille.

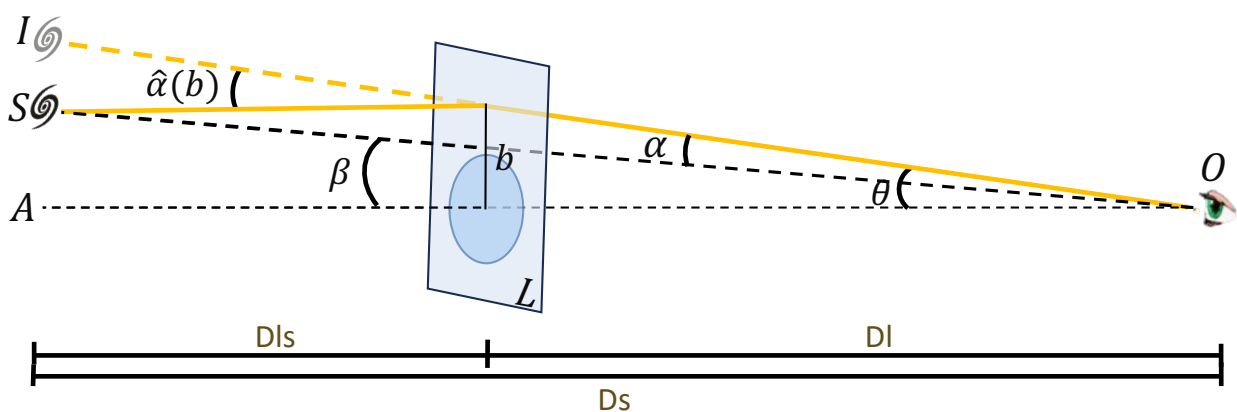


Figure 2.2 – Schéma de la géométrie dans l'espace de l'effet de lentille gravitationnelle formée par la source (S), le plan lentille (L), l'image observée (I), l'observateur (O) et l'axe optique (A). La trajectoire du photon (trait jaune) est déviée aux environs du paramètre d'impact \vec{b} , et on note $\hat{\alpha}(\vec{b})$, l'angle de déviation induit par le champ gravitationnel dans le plan lentille et qui mène à l'angle de déviation réduit α du point de vue de l'observateur. L'angle θ mesure la séparation angulaire entre l'axe optique du système et la position de l'image observée du point de vue de l'observateur tandis que l'angle β correspond à la séparation angulaire entre l'axe optique et la position réelle. Les distances D_{ls} , D_l et D_s correspondent respectivement aux distances angulaires apparentes entre la lentille et la source, l'observateur et la lentille, et l'observateur et la source.

Pour revenir à la position exacte de la source, nous replaçons le système dans une géométrie dans l'espace tel qu'illustré dans la figure 2.2. On peut alors définir les angles $\vec{\beta}$ et $\vec{\theta}$ qui correspondent à l'angle entre l'axe optique (A) et respectivement la position réelle de la source et la position observée de la source. De même, nous définissons les distances angulaires apparentes (angular diameter distance en anglais) entre la source et la lentille, notée D_{ls} ; entre la lentille et l'observateur, notée D_l ; et entre la source et l'observateur, notée D_s . Ces différentes données forment alors une géométrie au sein de laquelle nous pouvons estimer la relation liant les angles $\vec{\alpha}$, $\vec{\beta}$ et $\vec{\theta}$. Nous avons défini précédemment l'angle de déviation $\vec{\alpha}(\vec{b})$, qui dans l'espace dépend du paramètre d'impact \vec{b} , et note $\vec{\alpha}$, l'angle réduit de déviation :

$$\vec{\alpha}(\vec{b}) = \frac{D_{ls}}{D_s} \vec{\alpha}(\vec{b}). \quad (2.8)$$

Pour une lentille stationnaire et dans l'approximation des lentilles minces, nous trouvons que :

$$\vec{\theta} D_s = \vec{AO} = \vec{AS} + \vec{SO} = \vec{\beta} D_s + \vec{\alpha}(\vec{b}) D_{ls} = \vec{\beta} D_s + \vec{\alpha}(D_l \vec{\theta}) D_s. \quad (2.9)$$

Ainsi, en simplifiant par D_s , nous obtenons l'équation Eq. 2.10, dite équation lentille :

$$\vec{\beta} = \vec{\theta} - \frac{D_{ls}}{D_s} \vec{\alpha}(\vec{b}) \equiv \vec{\theta} - \vec{\alpha}(\vec{\theta}). \quad (2.10)$$

Cette équation lie la position observée de la source en fonction de sa position exacte. Elle est non-linéaire et admet alors plusieurs solutions $\vec{\theta}$ pour un même angle $\vec{\beta}$ ce qui traduit l'idée qu'une multitude de trajectoire peuvent atteindre l'observateur depuis le même angle. La résolution de cette équation permet de trouver la relation entre le plan source et le plan de la lentille. On observe alors deux régimes se distinguant par le nombre d'images sources qui en résultent en fonction de l'angle de déviation. Si l'angle de déviation induit par le plan lentille est faible, on associe à chaque source une seule image avec un angle d'observation légèrement dévié. C'est ce qu'on appelle l'effet de lentille gravitationnelle faible. Si l'angle de déviation par la lentille devient trop important, la relation n'est plus bijective et une même source engendre plusieurs images observées. C'est ce qu'on appelle l'effet de lentille gravitationnelle fort et ce phénomène est plus rare.

L'Anneau d'Einstein

Dans le cas particulier où la source est alignée avec la lentille et l'observateur ($\vec{\beta} = 0$) et pour une lentille à symétrie sphérique, l'équation lentille (Eq 2.10) admet une multitude de solutions : l'anneau d'Einstein, $\vec{\theta}_E$,

$$\vec{\theta}_E = \frac{D_{ls}}{D_s} \vec{\alpha}(\vec{b}). \quad (2.11)$$

On montre facilement à partir des équations 2.7 et $\vec{b} = D_l \vec{\theta}$ que les solutions peuvent s'écrire sous la forme de l'anneau de rayon :

$$\theta_E = \sqrt{\frac{D_{ls}}{D_s D_l} \frac{4GM(\theta_E)}{c^2}}. \quad (2.12)$$

Ainsi, la taille de l'anneau dépend de la masse de la lentille et des distances angulaires relatives au système lentille. La figure 2.3 présente 2 anneaux d'Einstein (LRG 3-757 et SDSSJ0146-0929)^a observés par le télescope spatial Hubble (HST). Notons que même si cette formation porte le nom d'anneau d'Einstein, le premier à le mentionner est Orest Khvolson sous le nom d'effet de halo, en 1924. Ce n'est qu'en 1988 que les radioastronomes du Very Large Array découvrent le premier "anneau d'Einstein", baptisé MG 1131+0456, créé par l'effet de lentille gravitationnelle

prédit par Einstein en 1936.



Figure 2.3 – Photographies de deux anneaux d'Einstein observés par le Hubble Space Telescope.

a. ESA/Hubble

2.1.3 . Le potentiel effectif projeté et la convergence

Nous avons vu que la position de la lentille et sa masse ont un impact significatif sur l'effet de lentille gravitationnelle. Cet impact nous permet alors de définir la puissance de la lentille et les propriétés des distorsions observées sur l'image de la source. Nous présentons dans cette sous-section la théorie de l'effet de lentille gravitationnelle par les amas de galaxies en termes de puissance de lentille.

L'effet de lentille gravitationnelle permet notamment d'étudier les propriétés de la lentille qui l'induit. C'est pourquoi nous étudions également du point de vue de la lentille et tout particulièrement à travers le potentiel effectif de la lentille ou potentiel projeté, $\psi(\vec{\theta})$:

$$\psi(\vec{\theta}) = \frac{2}{c^2} \frac{D_{1s}}{D_1 D_s} \int \Phi(D_1 \vec{\theta}, z) dz. \quad (2.13)$$

Ce potentiel effectif est défini en fonction de la position des sources à la distance D_s de l'observateur, à la distance D_{1s} de la lentille et la distance D_1 de l'observateur à la lentille et du potentiel newtonien Φ . Par définition, ce potentiel nous permet de revenir tout d'abord par son gradient en fonction de $\vec{\theta}$ à l'angle de déviation réduit, $\vec{\alpha}$ (voir Eq 2.8) :

$$\vec{\nabla}_{\theta} \psi(\vec{\theta}) = \frac{2}{c^2} \frac{D_{1s}}{D_1 D_s} \int \vec{\nabla}_{\theta} \Phi(D_1 \vec{\theta}, z) dz = \frac{2}{c^2} \frac{D_{1s}}{D_s} \int \vec{\nabla}_{\perp} \Phi dz = \frac{D_{1s}}{D_s} \vec{\alpha} = \vec{\alpha}. \quad (2.14)$$

Puis, par son laplacien, nous lions le potentiel effectif à la masse de la lentille :

$$\nabla_{\theta}^2 \psi(\vec{\theta}) = \frac{2}{c^2} \frac{D_{1s}}{D_1 D_s} \int D_1^2 \nabla_{\perp}^2 \Phi dz = \frac{2}{c^2} \frac{D_{1s} D_1}{D_s} \int \nabla_{\perp}^2 \Phi dz = \frac{8\pi G}{c^2} \frac{D_1 D_{1s}}{D_s} \Sigma(\vec{\theta}), \quad (2.15)$$

où $\Sigma(\vec{\theta})$ est la densité de masse projetée, intégrée à partir de sa densité de masse 3D, $\rho(\vec{\theta}, z)$, le long de la ligne de visée :

$$\Sigma(\vec{\theta}) = \int_0^{+\infty} \rho(\vec{\theta}, z) dz. \quad (2.16)$$

De même, nous définissons ainsi la densité de masse projetée critique associée Σ_{crit} :

$$\Sigma_{\text{crit}} = \frac{c^2}{4\pi G} \frac{D_s}{D_1 D_{1s}}. \quad (2.17)$$

Sa valeur est propre au système lentille et dépend uniquement des différentes distances angulaires et des propriétés de la lentille. Physiquement, la densité de masse projetée critique correspond à la

densité à partir de laquelle on observe des images multiples. À partir des deux équations précédentes, nous trouvons facilement :

$$\nabla_{\vec{\theta}}^2 \psi(\vec{\theta}) = \frac{\Sigma(\vec{\theta})}{\Sigma_{\text{crit}}} \equiv 2\kappa(\vec{\theta}), \quad (2.18)$$

par laquelle nous introduisons la convergence $\kappa(\vec{\theta})$ qui correspond à la densité de masse projetée normalisée par la densité de masse critique. La convergence est directement liée à la distribution de masse et permet alors de distinguer le régime du système lentille : faible ou fort. En particulier, si la convergence est supérieure à 1 alors le système engendre plusieurs images pour une même source et on est dans un régime d'effet de lentille gravitationnelle fort. À l'inverse, si la convergence est inférieure à 1 alors on se place dans le régime de l'effet de lentille gravitationnelle faible où la source apparaît faiblement déformée.

Notons que l'équation 2.18 montre que le potentiel effectif projeté, ψ , est solution d'une équation de Poisson à deux dimensions. Ainsi, il peut également s'écrire :

$$\psi(\vec{\theta}) = \frac{1}{\pi} \int_{\mathbb{R}^2} \kappa(\vec{\theta}') \ln(|\vec{\theta}' - \vec{\theta}|) d^2\vec{\theta}'. \quad (2.19)$$

Le potentiel effectif ψ peut ainsi directement s'écrire en fonction des paramètres du système lentille et de la convergence κ . Le gradient de ce potentiel (Eq. 2.19) nous permet de trouver la relation entre la convergence et l'angle de déviation réduit en suivant l'équation 2.14 :

$$\vec{\alpha}(\vec{\theta}) = \frac{1}{\pi} \int_{\mathbb{R}^2} \kappa(\vec{\theta}') \frac{\vec{\theta}' - \vec{\theta}}{|\vec{\theta}' - \vec{\theta}|^2} d^2\vec{\theta}'. \quad (2.20)$$

En résumé, nous venons d'étudier le formalisme de l'effet de lentille gravitationnelle avec un focus tout particulier sur le potentiel effectif de la lentille à partir de sa distribution de masse projetée. Cela nous a permis d'introduire la notion de convergence, directement liée à la distribution de masse et qui régit le régime du système lentille : faible ($\kappa < 1$) ou fort ($\kappa > 1$). Dans notre étude sur les amas de galaxies, nous nous focalisons sur l'effet de lentille gravitationnelle faible qui permet de contraindre le potentiel gravitationnel total de l'amas. Le régime faible a l'avantage d'être observable au voisinage de tous les amas mais il est plus difficile à quantifier que le régime fort qui n'est présent qu'au niveau du cœur des amas.

2.1.4 . La distortion des images et le cisaillement

Dans le cas du régime faible, l'effet de lentille gravitationnelle induit une déviation faible de la trajectoire de la lumière qui se traduit par une distortion de l'image des sources. Dans cette sous-section, nous décrivons plus en profondeur les notions de convergence et de cisaillement pour caractériser ces distorsions.

Comme nous l'avons précisé, l'effet de lentille gravitationnelle dévie la trajectoire de la lumière provenant des sources lointaines. En général, pour les sources étendues telles que les galaxies, l'observateur reçoit une multitude de rayons lumineux provenant de points différents de la source. Ces rayons vont atteindre le plan de la lentille en différentes positions, c'est-à-dire à différentes distances du cœur du potentiel, et ainsi subir une déviation légèrement variable. La recombinaison de ces différents rayons lumineux sur le capteur de l'observateur montre alors que l'image d'une source étendue apparaît déformée par rapport au cas sans lentille.

Pour estimer les distorsions observées sur l'image des sources, il nous faut résoudre l'équation lentille (voir Eq 2.10) pour tout point de la source étendue. Dans le régime faible ($\kappa < 1$), l'équation admet une unique solution et il est ainsi possible de faire du tracé de rayon pour tout point de la source (voir Sect. 2.3). Alternativement, la résolution la plus simple du problème consiste à calculer le plan de la lentille puis le linéariser. Cette linéarisation est possible dès lors que la source est significativement plus petite que la lentille. En d'autres termes, le potentiel gravitationnel de la lentille et donc l'angle

de déviation ne varie pas considérablement pour les différents rayons lumineux. Cette hypothèse est notamment vérifiée dans le cas des amas de galaxies dont le diamètre angulaire de quelques arcminutes est supérieure de 3 ordres de grandeur au diamètre angulaire des galaxies sources.

Les déformations locales dans le plan lentille sont ainsi caractérisables en calculant la dérivée de l'équation lentille et décrites par sa matrice Jacobienne A, dite matrice de distorsion :

$$A \equiv \frac{\partial \vec{\beta}}{\partial \vec{\theta}} = \delta_{ij} - \frac{\partial^2 \psi(\vec{\theta})}{\partial \theta_i \partial \theta_j} = \begin{pmatrix} 1 - \kappa - \gamma_1 & -\gamma_2 \\ -\gamma_2 & 1 - \kappa + \gamma_1 \end{pmatrix}. \quad (2.21)$$

pour laquelle on introduit les composantes du cisaillement $\gamma \equiv \gamma_1 + i\gamma_2 = |\gamma|e^{2i\tau}$,

$$\gamma_1 = \frac{1}{2}(\partial_1^2 - \partial_2^2)\psi \text{ et} \quad (2.22)$$

$$\gamma_2 = \partial_1 \partial_2 \psi. \quad (2.23)$$

En suivant la même écriture, on peut écrire la convergence κ (Eq. 2.18) en fonction des composantes du laplacien :

$$\kappa = \frac{1}{2}(\partial_1^2 + \partial_2^2)\psi. \quad (2.24)$$

La matrice A nous permet de donner davantage de sens aux composantes caractérisant la distorsion : κ et γ . En particulier, la décomposition :

$$A \equiv (1 - \kappa) \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} - \gamma \begin{pmatrix} \cos(2\tau) & \sin(2\tau) \\ \sin(2\tau) & -\cos(2\tau) \end{pmatrix}, \quad (2.25)$$

nous permet de mettre en lumière leurs différents impacts sur la distorsion. La convergence κ , proportionnelle à la matrice identité, correspond à la contraction ou la dilatation isotrope de la forme de l'image, sans en modifier la forme intrinsèque. En parallèle, le cisaillement γ correspond au caractère anisotrope de la lentille et permet de quantifier l'orientation de la distorsion par son argument (l'angle τ) et selon deux axes. Par ces deux composantes, on caractérise la distorsion des images des sources en termes de taille, de forme et d'orientation. Notons que le cisaillement a aussi un effet de dilatation ou de contraction sur la forme de l'image.

APARTE

L'effet de magnification

L'effet de lentille gravitationnelle va de pair avec un autre effet, dit de magnification et qui correspond à une augmentation de la luminosité apparente de la source. Physiquement, la lentille n'a pas d'effet sur l'intensité des rayons lumineux reçus et conserve la brillance de surface de la source. Cependant, la déviation induite par la lentille focalise les rayons et modifie l'angle solide sous lequel les rayons lumineux sont observés ce qui impacte le flux lumineux reçu par l'observateur. Le flux lumineux étant le produit de la brillance de surface de la source par l'angle solide d'observation, sa variation peut se calculer par l'expression $\frac{d\Omega_L}{d\Omega_S}$ où $d\Omega_i$ est l'angle solide sous influence de la lentille (L) et sans (S). Cet effet de magnification associé à la distorsion, noté μ , se calcule également à partir de la distorsion du plan lentille par l'inverse du déterminant de la matrice de distorsion, A :

$$\mu = \frac{1}{\det(A)} = \frac{1}{(1 - \kappa)^2 - |\gamma|^2}. \quad (2.26)$$

La nécessité de connaître le flux intrinsèque de chaque source pour analyser pleinement l'effet de magnification le rend moins efficace que l'utilisation des données de cisaillement. Cependant, cela peut être utilisé afin de normaliser la distribution de masse [Umetsu and Broadhurst \(2008\)](#).

2.1.5 . Inversion de masse et convergence

Nous avons vu qu'en pratique, nous observons une image déformée, à partir de laquelle nous mesurons les ellipticités des galaxies et estimons les composantes de cisaillement, γ_1 et γ_2 . Pour autant même si le cisaillement correspond à l'observable directe, c'est la convergence qui est directement liée à la distribution de masse. [Kaiser and Squires \(1993\)](#) a ainsi introduit la technique d'inversion de masse qui permet le calcul de la carte de convergence κ à partir du champ de cisaillement mesuré. Dans cette sous-section, nous détaillons le principe de cette technique, dit algorithme KS.

Dans la section 2.1.3, nous avons vu que la convergence et le cisaillement étaient tous deux liés au potentiel effectif de la lentille, ψ . Plus précisément, les équations 2.24, 2.22 et 2.23 sont des combinaisons linéaires de dérivées partielles secondes du potentiel effectif. Le principe de l'algorithme de Kaiser et Squires consiste alors à utiliser ces relations pour reconstruire la carte de convergence à partir de la mesure des champs de cisaillement γ_1 et γ_2 .

Pour cela, nous utilisons la notation complexe pour représenter le cisaillement, $\gamma = \gamma_1 + i\gamma_2$, et la convergence, $\kappa = \kappa_E + i\kappa_B$, avec κ_E et κ_B , appelés modes E et B. Puis, nous nous plaçons dans le domaine de Fourier pour transformer les équations 2.24, 2.22 et 2.23 afin d'en réduire l'ordre. La transformée de Fourier (dénnotée par le symbole d'accent circonflexe ici) des trois équations, nous donne alors :

$$\hat{\kappa} = \hat{\kappa}_E + i\hat{\kappa}_B = -\frac{1}{2}(k_1^2 + k_2^2)\hat{\psi}, \quad (2.27)$$

$$\hat{\gamma}_1 = -\frac{1}{2}(k_1^2 - k_2^2)\hat{\psi}, \quad (2.28)$$

$$\hat{\gamma}_2 = -k_1^2 k_2^2 \hat{\psi}, \quad (2.29)$$

où k_i est le nombre d'onde à la position angulaire $\vec{\theta}_i$. Ces relations linéaires peuvent facilement être recombinaées afin de revenir à une expression de la forme $\hat{\gamma} = \hat{P} \hat{\kappa}$ avec $\hat{P} = \hat{P}_1 + i\hat{P}_2$ l'opérateur complexe de la transformation. À partir du système d'équations précédent (Eq. 2.29), nous obtenons les paramètres \hat{P}_1 et \hat{P}_2 correspondant :

$$\hat{P}_1 = \frac{k_1^2 - k_2^2}{k_1^2 + k_2^2}, \quad (2.30)$$

$$\hat{P}_2 = \frac{2k_1^2 k_2^2}{k_1^2 + k_2^2}. \quad (2.31)$$

Ainsi, nous obtenons :

$$\hat{\kappa}_E = \hat{P}_1 \hat{\gamma}_1 + \hat{P}_2 \hat{\gamma}_2, \quad (2.32)$$

$$\hat{\kappa}_B = -\hat{P}_2 \hat{\gamma}_1 + \hat{P}_1 \hat{\gamma}_2. \quad (2.33)$$

$$(2.34)$$

Ces relations permettent de calculer les différents modes de la convergence dans le domaine de Fourier à partir des mesures des composantes du cisaillement. Alternativement, cet algorithme mène aussi à une relation entre κ et γ en termes de convolution par le noyau complexe, D tel que :

$$\kappa(\vec{\theta}) = \frac{1}{\pi} \int_{\mathbb{R}^2} \text{Re}[D^*(\vec{\theta} - \vec{\theta}')\gamma(\vec{\theta}')]d\theta^2 \quad (2.35)$$

avec Re désignant la partie réelle, et D^* , le conjugué du noyau D qu'on peut écrire selon les composantes du vecteur position $\vec{\theta}$:

$$D(\vec{\theta}) = \frac{(\theta_2^2 - \theta_1^2) - i2\theta_1\theta_2}{\theta^4}. \quad (2.36)$$

Cette décomposition met en lumière un point clef sur la convergence et en particulier ses différents modes. Étant donné que ψ est un potentiel scalaire, on peut montrer qu'un faible effet de lentille

ne produit en principe pas de modes B. Les modes B peuvent donc être utilisés pour estimer le niveau de bruit dans les données.

Enfin, la figure 2.4 montre un exemple de cartes de cisaillement et de convergence simulées, dérivées de la simulation *DUSTGRAIN-pathfinder* (décrite dans la section 2.3) couvrant un champ de $5^\circ \times 5^\circ$. Les panneaux de gauche et du milieu correspondent aux deux composantes du cisaillement, γ_1 et γ_2 , et le panneau de droite à la carte de convergence en mode E associée, κ_E . Bien que le cisaillement et la convergence puissent tous deux être utilisés pour la détection des amas, la convergence est plus facile et moins coûteuse à analyser car il s'agit d'un champ scalaire proportionnel à la distribution de matière projetée.

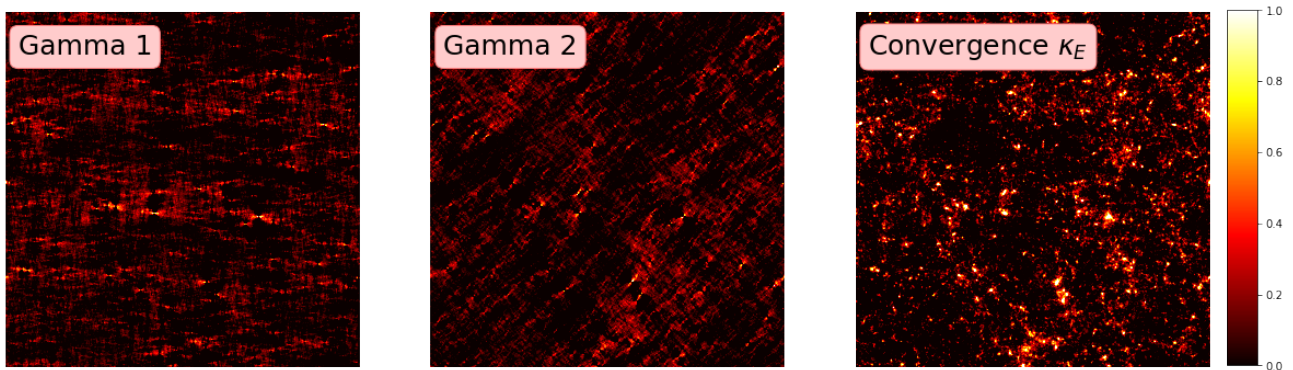


Figure 2.4 – Cartes des composantes caractéristiques de l’effet de lentille gravitationnelle : le cisaillement $\gamma = \gamma_1 + i\gamma_2$ et la convergence κ .

Dans cette section, nous avons détaillé le formalisme de l’effet de lentille gravitationnelle. Cet effet se caractérise par la déviation des rayons lumineux au voisinage d’un objet massif. Cette déviation est régie par l’équation lentille et la distortion qu’elle engendre est caractérisée par deux composantes essentielles : la convergence κ (dilatation ou contraction) et le cisaillement (distorsions biaxiales). Pour notre étude de la détection des amas de galaxies, nous nous plaçons dans le régime faible ($\kappa < 1$). Grâce aux traces qu’elles nous laissent dans les images des galaxies lointaines, les lentilles gravitationnelles faibles offrent une opportunité unique pour la détection des amas de galaxies à partir de leur distribution totale de masse de long de la ligne de visée.

2.2 Données de lentille gravitationnelle faible

Il a fallu attendre les années 1979 pour que les études de [Walsh et al. \(1979\)](#) confirment l’observation de l’effet de lentille gravitationnelle par une galaxie sur un quasar lointain. Pour les amas de galaxies, ces premières observations de l’effet de lentille gravitationnelle faible, vont de pair avec la possibilité de leur détection par des télescopes. À partir des photographies prises par les télescopes dans le domaine du visible, on génère un catalogue de galaxies. Dans cette section, nous présentons tout d’abord comment ce catalogue de galaxies est construit à partir des mesures de cisaillement observées. Enfin, nous parlerons du contexte, tout particulièrement en termes de relevés, dans lequel s’inscrit cette thèse.

2.2.1 . Mesure de cisaillement et catalogue de galaxies sources

Dans le cas des amas de galaxies, la lumière de galaxies sources lointaines est déviée lorsqu’elle se dirige vers nous et leurs images apparaissent déformées. Ces distorsions sont directement observables et constituent une empreinte de la distribution de matière à grande échelle. L’un des objectifs des relevés en optique étudiant l’effet de lentille gravitationnelle consiste à produire des catalogues contenant la mesure de ces distorsions pour les galaxies sources observées. Dans cette sous-section,

nous détaillons la construction de ces catalogues de sources à partir des observations de l'effet de lentille gravitationnelle faible.

Dans le cas des relevés réels, les données acquises par le télescope sont des photographies d'une portion du ciel. Ces photographies sont ensuite prétraitées afin de corriger les effets systématiques induits par l'instrument sur les images. Ensuite, différentes méthodes (catalogue préexistant, algorithmes de détections) existent pour identifier les galaxies présentes sur la portion du ciel observée.

Pour produire les catalogues de galaxies sources du relevé, on estime des déformations observées pour chacune des galaxies. En effet, l'observable direct des relevés n'est pas le cisaillement mais la forme de chaque galaxie source observée et la mesure d'ellipticités qui en découle.

Il existe de nombreuses approches pour mesurer les ellipticités des galaxies soit à partir des moments (Kaiser et al., 1995; Melchior et al., 2011), par ajustement de modèle (Bridle et al., 2002; Miller et al., 2007) ou encore par décomposition (Refregier, 2003; Hirata and Seljak, 2003). Cette dernière méthode est notamment appliquée pour l'estimation des ellipticités dans le cas des données Hyper Suprime-Cam (Aihara et al., 2018a), présentées dans le chapitre 8. Les ellipticités, e_{s1} et e_{s2} , des galaxies sont définies (Hirata and Seljak, 2003; Mandelbaum et al., 2018) sous la forme simple :

$$(e_{s1}, e_{s2}) = \frac{1 - (r_b/r_a)^2}{1 + (r_b/r_a)^2} (\cos(2\phi), \sin(2\phi)), \quad (2.37)$$

où r_b/r_a est le ratio entre les axes mineurs (b) et majeurs (a) de la galaxie et ϕ est l'angle entre l'axe majeur et l'axe équatorial.

Les ellipticités observées e_o sont reliées au cisaillement réduit par l'équation (Seitz and Schneider, 1997) :

$$e_o = \frac{e_s + g}{1 + g^* e_s}, \quad (2.38)$$

où e_s correspond aux ellipticités (e_1, e_2) intrinsèques des galaxies définies selon l'équation 2.37 et g est le cisaillement réduit, exprimé par la formule :

$$g \equiv \frac{\gamma}{1 - \kappa}. \quad (2.39)$$

Dans le régime du lentillage faible ($\kappa \ll 1$), le cisaillement réduit tend vers le cisaillement, γ . De plus, sous l'hypothèse que les formes des galaxies n'ont pas d'orientation privilégiée en l'absence de lentille, on trouve que $\langle e_s \rangle = 0$. Ainsi, l'ellipticité est un bon estimateur du cisaillement réduit :

$$\langle e_o \rangle \simeq g. \quad (2.40)$$

Ces méthodes sont sujettes à différents effets systématiques :

- des incertitudes liées au bruit de fond sur les données;
- des incertitudes additives liées à l'anisotropie de la PSF (point spread function), qui correspond à l'étalement potentiel d'une source ponctuelle par l'instrument;
- des biais liés au modèle choisi pour représenter et estimer la forme des galaxies;
- des biais de sélection si les galaxies sources sont indirectement sélectionnées en fonction de propriétés associées à leur forme.

La caractérisation de ces biais et effets systématiques pour la méthode considérée est essentielle pour la production du catalogue de galaxies sources. Généralement, cette caractérisation passe par une phase de calibration de la méthode à partir soit de données simulées, soit des données obtenues par le relevé avec application de tests de nullité (Mandelbaum et al., 2018). Le programme STEP (Shear Testing Programme, Heymans et al., 2006; Massey et al., 2007, 2013) et le challenge GREAT8 (Gravitational LEnsing Accuracy Testing 2008 Bridle et al., 2010) ont étudié en détails les méthodes de mesure du cisaillement des prochains et actuels grands relevés ainsi que les défis que cette mesure apporte. Ils ont montré que la prise en compte des biais pouvait se faire simplement à travers l'expression corrigée du cisaillement réduit :

$$g_i^{obs} = (1 + m) g_i^{vrai} + c_i, \quad (2.41)$$

avec m un biais multiplicatif et c_i un biais additif pour $i = 1$ ou 2 . L'utilisation des biais additifs c_1, c_2 dans le catalogue permet en particulier de tenir compte de la systématique additive due à une correction de l'anisotropie de la PSF. À chaque galaxie, on peut également ajouter un poids w , qui correspond à la fiabilité associée à la mesure de forme pour la galaxie considérée.

Pour produire le catalogue de galaxies sources, ces valeurs d'ellipticités sont accompagnées par la position observée des galaxies, les différents biais c_1, c_2 et m liés à l'estimation des ellipticités, le poids w , mais aussi par le redshift estimé des galaxies. Ce redshift est généralement obtenu par photométrie compte tenu du grand nombre de galaxies que comportent les relevés optiques. Notons que le redshift des galaxies sources est non seulement un point essentiel de certains algorithmes de détection (Hamana et al., 2020) mais qu'il est aussi nécessaire à termes, pour lier le signal lentille à la masse à partir des détections obtenues.

En résumé, à partir de l'observation des déformations des galaxies sur les images, il est possible de calculer les ellipticités associées aux galaxies. Dans le régime faible, ces ellipticités constituent un bon estimateur du cisaillement réduit dont l'expression est corrigée pour tenir compte des différents biais associés à cette estimation.

Dans ce chapitre, nous avons brièvement introduit la complexité de la production des catalogues de galaxies sources à partir des données collectées par les télescopes.

2.3 Données simulées

Au moment de l'écriture de ce manuscrit, le télescope *EUCLID* pour lequel les études présentées ici ont été développées est en route pour le point Lagrange L2. En attendant d'avoir accès à ces données, nous avons à disposition une simulation se rapprochant des attentes pour les données du télescope *EUCLID*. Dans cette section, nous présentons ainsi la simulation *DUSTGRAIN-pathfinder* (Giocoli et al., 2017) et les différents catalogues associés qui sont utilisés pour les études des chapitres 3, 4, 5, 6 et 7 de ce manuscrit.

2.3.1 . Description de la simulation

La simulation *DUSTGRAIN-pathfinder* (Dark Universe Simulations to Test GRAvity In the presence of Neutrinos) est une simulation cosmologique à N corps. Elle vise à produire des cartes ou des catalogues simulés aussi larges que détaillés pour l'étude de l'effet de lentille gravitationnelle faible en particulier. Pour ce faire, cette simulation retrace l'évolution sans collision de 783^3 particules de matière noire de masse $m_{\text{CDM}} = 8.1 \times 10^{10} h^{-1} M_{\odot}$ contenues dans une boîte cosmologique périodique de $750 h^{-1}$ Mpc de côté. Dans cette boîte, les paramètres du modèle cosmologique ont été pris en accord avec les observations de Planck 2015 (The Planck Collaboration, 2006) : $\Omega_M = \Omega_{\text{CDM}} + \Omega_b + \Omega_{\nu} = 0.31345$, $\Omega_b = 0.0481$, $\Omega_{\Lambda} = 0.68655$, $h = 0.6731$, et $\sigma_8 = 0.847$ (voir Sect. 1.1). Dans le cas de l'étude présentée dans ce manuscrit, nous utilisons des simulations contenant uniquement des particules de matière noire, soit $\Omega_{\nu} = 0$.

La simulation est ensuite utilisée pour produire une série de réalisations à partir de la boîte cosmologique préalablement découpée en 34 clichés instantanés (snapshot) comobiles sauvegardés pour différents redshifts. Dans la pratique, les différentes réalisations ont été obtenues en changeant l'orientation de la boîte cosmologique comobile en tenant compte des 34 clichés instantanés comobiles et par des combinaisons des procédures suivantes : (i) changer le signe des coordonnées cartésiennes, (ii) redéfinir la position de l'observateur, et (iii) modifier l'ordre des axes dans le système de coordonnées. Par construction, ces variations préservent les propriétés de regroupement de la distribution des particules pour chaque cliché instantané de la simulation. A partir de chaque réalisation obtenue par ces variations, on a accès à un catalogue de halos de matière noire (voir Sect. 2.3.2) et un catalogue de galaxies sources (voir Sect. 2.3.3)

Comme nous l'avons précisé, nous utilisons dans ce manuscrit cette simulation. Dans le cas des chapitres 3, 4, 5 et 6, nous utilisons un sous-ensemble des séries complètes de DUSTGRAIN-*pathfinder* et composé de 256 réalisations de taille $5^\circ \times 5^\circ$. Dans le chapitre 7, nous avons utilisé un second sous-ensemble issu de DUSTGRAIN-*pathfinder* et composé de 12 réalisations de taille $10^\circ \times 10^\circ$.

2.3.2 . Le catalogue de halos de matière noire

Dans cette simulation, les catalogues de halos de matière nous est accessible pour chacune des réalisations. Nous décrivons dans cette sous-section comment ils sont construits et leurs spécificités.

Les catalogues de halos de matière noire simulés ont été générés en identifiant les halos de matière noire dans la simulation DUSTGRAIN-*pathfinder* à l'aide d'un algorithme "friends-of-friends" (FOF, Davis et al. 1985) avec une distance de liaison $\lambda = 0.16 \times d$, où d est la distance moyenne de séparation entre les particules. L'algorithme SUBFIND (Springel et al. 2001) a ensuite été utilisé pour évaluer les paramètres standards de chaque halo identifié par FOF : le redshift z , la masse caractéristique, M_{200c} , et le rayon associé R_{200c} , correspondant à la masse et au rayon d'une région sphérique autour du centre de chaque halo englobant 200 fois la densité critique de l'Univers.

Ainsi, pour chacune des réalisations, le catalogue de halos est composé très exactement des positions (c'est-à-dire l'ascension droite, RA, et la déclinaison, Dec) des halos de matière noire identifiés, de leur redshift estimé, z_i , de leur masse caractéristique, M_{200c} , et de leur rayon caractéristique, R_{200c} . Ce catalogue de halos de matière noire est un élément essentiel de notre analyse car il sert de base à la procédure d'association (voir Sects. 5.3 et 7.2.4) lors de notre étude. Il permet de quantifier la pureté et la complétude de chaque catalogue de détection obtenu sur les différentes réalisations.

2.3.3 . Le catalogue de galaxies sources

En parallèle des catalogues de halos de matière noire, la simulation est accompagnée pour chaque réalisation du catalogue de galaxies sources pour l'étude de l'effet de lentille gravitationnelle faible. Nous détaillons dans cette sous-section, la création et le contenu de ce catalogue.

Les catalogues de galaxies sources sont construits à partir des cônes de lumières des particules de matière noire dans la boîte cosmologique de la simulation. En pratique, les particules stockées dans 21 clichés instantanés différents sont utilisées pour construire ces cônes de lumière allant de $z = 0$ à $z = 4$ en utilisant la routine MAPsim (Giocoli et al., 2015, 2017). La routine extrait les positions de chaque particule de matière noire pour recréer le cône de lumière passé et en parallèle, les particules sont distribuées dans 27 plans lentilles différents afin de recomposer la distribution de densité de matière projetée. Ensuite, le cisaillement de chaque galaxie est calculé par projection, en utilisant l'approximation de Born, qui suppose des trajectoires lumineuses non perturbées pour intégrer les distorsions du signal lentille (Bartelmann and Schneider, 2001). La distribution des redshifts des galaxies, $n(z)$, a été construite pour suivre une distribution réaliste proche de celle attendue pour le relevé *Euclid* (par exemple Cropper et al. 2013). Dans les chapitres 3 à 6, chaque réalisation contient environ 3 millions de galaxies dans un champ d'observation de $5^\circ \times 5^\circ$, avec une distribution de galaxies s'étendant jusqu'à $z = 4$. Cela conduit à une densité de galaxies d'environ $n_g = 30 \text{ gal.arcmin}^{-2}$, se rapprochant des attentes d'*EUCLID*. Pour le chapitre 7, le champ simulé est quatre fois plus grand. Pour chacune des réalisations, le catalogue de galaxies est ainsi composé : (i) des positions (c'est-à-dire l'ascension droite et la déclinaison) des galaxies, (ii) des deux composantes du cisaillement, et (iii) du décalage vers le rouge. Ces catalogues de galaxies sont ensuite utilisés pour évaluer les algorithmes de détection décrits ci-dessous.

Du bruit a également été ajouté au cisaillement dans le catalogue de galaxies sources pour simuler des études réalistes. Ce bruit imite les incertitudes dans le cisaillement (appelées "bruit de forme") qui proviennent d'une combinaison de la forme intrinsèque inévitable des galaxies (appelée "bruit de forme intrinsèque") et des erreurs de mesure qui comprennent, entre autres facteurs, les incer-

titudes dans la mesure de la forme des galaxies et la correction de la fonction d'étalement du point (couramment appelée PSF). La forme intrinsèque des galaxies étant la composante dominante et les galaxies étant distribuées au hasard, le bruit de forme peut être modélisé comme un bruit additif. Il peut être bien approximé par une distribution gaussienne avec une moyenne de $\mu = 0$ et un écart type de $\sigma_\epsilon = 0.26$ (e.g. [Leauthaud et al., 2007](#); [Schrabback et al., 2015, 2018](#)) par composante.

En résumé, la simulation *DUSTGRAIN-pathfinder* que nous avons présenté dans cette section est la base de la procédure d'association des chapitres 5, 6 et 7. Cette simulation est adaptée pour l'étude de l'effet de lentille gravitationnelle faible car on a un champ large par le nombre de réalisations et une densité de galaxies nous assurant un signal lentille détectable. D'autre part, les paramètres de la simulation (distribution de redshift, densité de galaxies) se rapprochent des attentes d'*EUCLID*, nous permettant d'en préparer les méthodes dans des conditions réalistes.

2.4 Conclusions

Dans ce chapitre, nous avons détaillé le formalisme mathématique permettant d'étudier l'effet de lentille gravitationnelle faible. Cet effet correspond à la déviation de la trajectoire des rayons lumineux par la distribution de masse le long de la ligne de visée. Il se traduit sur les images des télescopes par une déformation de la forme des galaxies sources lointaines observées. La caractérisation et l'étude de cette déformation nous permet de remonter à la distribution de masse sous-jacente responsable de cette déviation.

Nous avons vu que cette déformation pouvait être caractérisée par deux composantes : le cisaillement γ et la convergence κ , reliés par l'inversion de masse de [Kaiser and Squires](#). La convergence est tout particulièrement intéressante car elle est directement liée à la distribution de masse de la lentille. Ainsi, cet effet est parfaitement adapté pour la détection des amas de galaxies par leur distribution totale de masse et non simplement baryonique.

Dans les données issues de relevés réels, nous avons accès aux valeurs d'ellipticités des galaxies sources observées qui permet d'estimer le cisaillement qui leur est associé. Ces valeurs sont essentielles pour produire les catalogues de galaxies sources nécessaires à la détection des amas par effet de lentille gravitationnelle faible. Un nombre élevé de galaxies sources lointaines permet d'améliorer nettement la précision sur le cisaillement.

Dans ce chapitre, nous avons également introduit les données simulées qui sont à la base de notre étude dans les chapitres suivants. Ces données simulées ont l'avantage d'approcher les caractéristiques attendues pour les données du télescope *EUCLID*. Pour ce jeu de données simulées, nous avons à dispons à la fois les catalogues de galaxies sources mais aussi les catalogues de halos de matière noire qui permettent d'analyser les résultats obtenus par les méthodes de détections considérées.

*
* *

Deuxième partie

Deuxième partie : Méthodes de détections

3

L'algorithme de détection

C'est à partir de l'information reçue par l'observateur et transformée en signal que s'établissent les classements, les modèles physiques, les prévisions qui finalement constituent cette science qu'est l'astrophysique.

Pierre Léna, L'observation en astrophysique

3.1	Quelles données d'entrée pour la détection?	46
3.1.1	Les données d'entrée : 3 implémentations possibles	46
3.1.2	Impact de l'utilisation d'une grille régulière sur le signal	48
3.1.3	Impact de l'utilisation des cartes de cisaillement ou de convergence	50
3.1.4	Impact du choix du traitement des données en matière de détection	52
3.1.5	Conclusion	52
3.2	Les méthodes de filtrage	53
3.2.1	Les enjeux du filtrage	54
3.2.2	Filtrage gaussien	55
3.2.3	Approche standard	58
3.2.4	Fonctions utilisées	59
3.2.5	Conclusion	61
3.3	La procédure de seuillage et l'identification des détections	62
3.3.1	Le seuillage	62
3.3.2	L'identification des pics	63
3.3.3	Récapitulatif de l'algorithme de détection	63
3.4	Conclusions	64



Nous avons vu précédemment l'importance des amas pour la cosmologie puis leurs diverses possibilités d'observation et tout particulièrement par l'effet de lentille gravitationnelle faible. Attardons-nous maintenant sur comment détecter les amas à partir de leur effet de lentille faible. Nous décrivons ici les différentes approches qui sont habituellement utilisées pour la détection d'amas à une échelle donnée par effet de lentille gravitationnelle faible. Nous présentons également l'algorithme général de détection (Fig 3.1) que nous avons choisi d'adopter pour toutes les méthodes afin

de faciliter la comparaison des méthodes. L'objectif de l'algorithme de détection consiste à transformer les données d'entrée bruitées en catalogues de détections correspondant aux amas candidats. L'algorithme de détection prend en entrée des données **(A)**, au sein desquelles le signal des amas est noyé. Le catalogue de forme des galaxies sources est la forme la plus simple de données d'entrée pour l'algorithme de détection. Nous verrons en section 3.1 que les données d'entrée peuvent se présenter sous d'autres formes pour accélérer l'algorithme. Nous filtrons ensuite les données d'entrée pour isoler les structures d'intérêt **(B)**. Nous présentons en section 3.2, les méthodes de filtrage standard à une échelle donnée et au chapitre 4, notre filtrage multi-échelle. L'étape suivante consiste à réaliser le seuillage des cartes filtrées **(C)** et à détecter les amas candidats parmi les pixels au-delà du seuil fixé **(D)**, ce que nous détaillons en section 3.3.

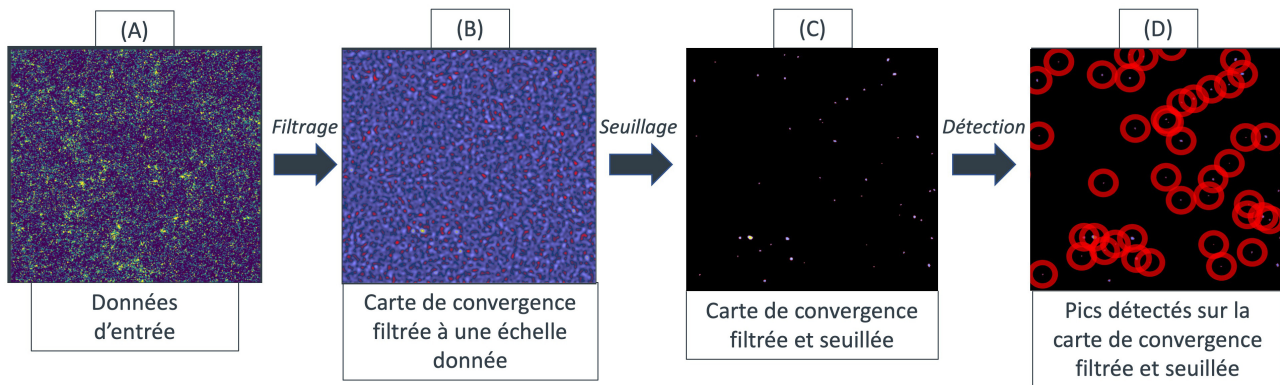


Figure 3.1 – Schéma des étapes clés de l'algorithme de détection.

3.1 Quelles données d'entrée pour la détection ?

La première étape de l'algorithme de détection consiste à choisir sous quelle forme traiter les données d'entrée. L'objectif de cette étape est d'adapter les données brutes afin de les préparer au filtrage. Ce prétraitement des données brutes s'étudie de pair avec le filtrage, qu'on peut généralement décrire comme une convolution des données par un filtre. Pour mener à des détections fiables, l'enjeu est de mettre les données sous une forme qui facilite la convolution. Le prétraitement doit être précis pour ne pas perdre de signal lentille, mais aussi rapide pour être applicable à des jeux de données de taille faramineuse. Dans cette section, nous détaillons les données d'entrée possibles et les comparons en termes de temps de calcul et de signal. Nous présentons ainsi une nouvelle forme pour les données d'entrée que nous utilisons dans l'algorithme de détection.

3.1.1 . Les données d'entrée : 3 implémentations possibles

Comme nous l'avons décrit en section 2.3, les données de lentille faible en entrée de l'algorithme de détection se présentent sous la forme d'un catalogue de galaxies sources bruitées. Il contient les observables directes, les ellipticités observées pour chaque galaxie source ainsi que leur position et leur redshift. Ce catalogue peut être utilisé comme tel pour conserver toute l'information sur le signal ou encore transformé pour gagner en vitesse de calcul. Nous présentons ici les 3 approches possibles :

- Utiliser le catalogue brut contenant le signal de cisaillement à la position (RA, Dec) des galaxies sources ;
- Transformer le catalogue en cartes de cisaillement contenant le signal de cisaillement réparti sur une grille régulière ;
- Reconstruire à partir des cartes de cisaillement, les cartes de convergence en modes E et B.

En pratique, l'approche la plus courante est de considérer en entrée le catalogue brut de galaxies sources, avec pour chacune, la mesure des ellipticités observées et bruitées. Puis, la procédure de filtrage est effectuée directement à la position des galaxies (e.g. [Maturi et al., 2005](#); [Wittman et al., 2006](#); [Miyazaki et al., 2018](#); [Hamana et al., 2020](#)). Cette première approche est privilégiée car le catalogue de galaxies sources est directement relié aux observables directes. Comme précisé dans le chapitre 2, l'observable directe est le cisaillement réduit, dérivé des ellipticités observées. Dans le régime du lentillage faible ($\kappa \ll 1$), c'est une bonne approximation du cisaillement. Par la suite, nous parlerons uniquement de cisaillement pour désigner les mesures présentes dans le catalogue pour chaque galaxie. Cette première approche est aussi privilégiée pour sa précision en termes de signal lentille car on travaille avec les positions exactes des galaxies.

Cependant, cette précision a une lourde contrepartie. Cette approche est très coûteuse en temps de calcul car toutes les galaxies sont traitées indépendamment lors du filtrage. Typiquement, pour un champ de $5^\circ \times 5^\circ$, dans le cas de la simulation *DUSTGRAIN-pathfinder* ([Giocoli et al., 2017](#)), qui a une densité de galaxies égale à $n_g = 30 \text{ gal. arcmin}^{-2}$, cela correspond à environ 3 000 000 de galaxies. Un tel volume de données entraîne un temps de calcul très important si la convolution est réalisée à la position de chaque galaxie et le temps de calcul ne peut qu'augmenter avec le volume croissant de données à venir (voir Sect 1.3.4).

Une alternative consiste alors à mettre le signal lentille d'entrée sous forme d'une grille régulière (e.g. [Hennawi and Spergel, 2005](#); [Leonard et al., 2015](#); [Oguri et al., 2021](#)), plus simple à traiter par la convolution. Pour ce faire, on définit une grille régulière et on combine le signal de cisaillement des galaxies qui tombent sur un même pixel de la grille pour créer des cartes de cisaillement. Cela équivaut à approximer les positions des galaxies par le point de la grille le plus proche. A partir d'un catalogue, on produit deux cartes de cisaillement en simultanément, qu'on vient ensuite convoluer par le filtre.

La troisième forme possible pour les données d'entrée correspond aux cartes de convergence (e.g. [White et al., 2002](#); [Hamana et al., 2004](#); [Tang and Fan, 2005](#); [Miyazaki et al., 2007](#)). Ces cartes sont également définies sur une grille régulière. Cette troisième approche découle directement de la seconde car il faut construire d'abord les cartes de cisaillement à partir du catalogue de galaxies. Puis, on leur applique la méthode d'inversion de masse de Kaiser & Squires (voir Sect 2.1.5) pour reconstruire les cartes de convergence en modes E et B associées. Là où les cartes de cisaillement correspondent à l'observable directe, le mode E des cartes de convergence donne accès directement à la distribution de masse projetée le long de la ligne de visée. La détection des amas de galaxies se fait donc tout naturellement sur la composante E de la carte de convergence. C'est sur cette idée que certaines méthodes sont développées et tout particulièrement la méthode que nous avons implémentée et présentons dans le chapitre suivant.

Ces trois approches ont pour objectif de permettre un traitement optimal du signal par le filtre. Mais chacune a ses limites : la première implémentation est plus précise mais plus coûteuse en temps de calcul et les suivantes sont plus rapides mais rajoutent des étapes de calculs pouvant impacter leur précision. Nous avons évalué le temps de calcul pour les trois différentes approches. Pour cela, nous convoluons le même catalogue de données brutes bruitées d'un champ de $5^\circ \times 5^\circ$ par un même filtre en suivant les trois approches pour produire une carte de masse filtrée sur une grille régulière à trois tailles différentes : 512×512 , 1024×1024 , et 2048×2048 pixels. Le filtre que nous avons choisi pour cette étude est celui défini par [Jarvis et al. \(2004\)](#) dont nous détaillerons l'implémentation dans la section suivante. On mesure alors le temps de calcul nécessaire à la convolution selon chaque approche dans la Table 3.1. Ces mesures sont des valeurs moyennes obtenues à partir de 20 réalisations. Le temps de calcul tient compte uniquement du temps de la convolution, non des temps de reconstruction des cartes qui sont peu significatifs devant le temps de la convolution. Le temps de calcul pour la première approche est de plus d'une heure (catalogue brut) tandis qu'il est de quelques secondes pour les autres implémentations. La troisième approche, qui passe par la carte de convergence, est la plus rapide. On mesure un facteur ~ 1500 par rapport à l'implémentation sur le catalogue brut et

Table 3.1 – Mesure du **temps de calcul de la convolution** en secondes par un même filtre pour les trois différentes approches pour un champ $5^\circ \times 5^\circ$, pour différentes tailles de grille.

Taille de la grille	Catalogue brut	Cartes de cisaillement	Cartes de convergence
512	5738 ± 59 sec	16 ± 1.1 sec	4 ± 0.8 sec
1024	8448 ± 78 sec	100 ± 6.3 sec	20 ± 2.7 sec
2048	31529 ± 340 sec	422 ± 27 sec	84 ± 12 sec

4 par rapport à celle qui utilise les cartes de cisaillement. Ce gain considérable en temps de calcul s’explique par le nombre d’opérations bien plus faible à réaliser lors de la convolution sur la carte de convergence dont le signal est compressé dans la composante en mode E. Pour un relevé comme *EUCLID*, dont la surface totale est estimée à un tiers du ciel, soit $15\,000 \text{ deg}^2$, l’écart entre les méthodes sur des cartes régulières et sur le catalogue, correspondrait à des mois de calcul.

De plus, si on travaille sur une grille régulière, l’application de la transformée de Fourier pour l’implémentation de la convolution est la méthode optimale en termes de temps de calcul. Elle permet de réduire au minimum d’un facteur 10 additionnel le temps de calcul. Ceci est un avantage supplémentaire en faveur des seconde et troisième implémentations mais ceci peut avoir un impact direct sur le signal qu’on traite. L’approche de Fourier a pour inconvénient l’apparition d’artefacts à la limite du champ en raison de l’hypothèse de périodicité de la transformée de Fourier discrète. Ces effets peuvent être considérablement atténués en supprimant les bords de l’image ou en traitant correctement les bords pendant l’inversion de masse, comme proposé dans [Pires et al. \(2020\)](#).

3.1.2 . Impact de l’utilisation d’une grille régulière sur le signal

À l’aube des grands relevés dédiés à l’effet de lentille gravitationnelle faible et face au volume important de données que ceux-ci apporteront, il est évident que le temps de calcul est un critère clef pour le filtrage et la détection des amas de galaxies. Mais l’objectif est de préparer les données pour un filtrage optimal, il faut donc également s’assurer que l’implémentation utilisée n’affecte pas le signal. Ainsi, le traitement des données doit-il être un compromis entre le temps de calcul et la préservation du signal. Nous avons vu que la première approche, qui préserve au mieux le signal, impliquait un temps de calcul significativement bien plus important que les approches sur une grille régulière. Cependant, l’utilisation d’une grille régulière engendre une perte de résolution dans la carte de cisaillement, qui peut affecter l’efficacité de la méthode de détection. On peut ainsi s’interroger sur l’impact de l’utilisation de cette grille régulière sur le signal. Cette sous-section a pour objectif d’analyser l’impact de chacune des méthodes par rapport au signal d’intérêt.

Nous avons vu trois approches possibles pour la convolution par le filtre. Dans le cas de données d’entrée sous forme de catalogue, la convolution est réalisée dans l’espace réel aux positions exactes des galaxies. Dans le cas de données d’entrée sous forme de cartes de cisaillement ou de convergence, on peut réaliser la convolution dans l’espace réel, ou dans l’espace de Fourier pour un gain de temps supplémentaire. Sur les deux dernières implémentations qui utilisent une grille régulière, on observe une perte de puissance du signal dont l’impact dépend de la résolution de la carte (c’est-à-dire le nombre de pixels de la carte). Cet impact peut être atténué en adoptant une grille adaptée aux données. En effet, il y a un compromis à faire sur la résolution de la carte. Une taille de pixel trop grande va générer une sorte de lissage qui engendre une perte de puissance des structures que l’on cherche à détecter. A l’inverse, une taille de pixel trop petite peut introduire des données manquantes dans la carte.

Les approches sur grille irrégulière (catalogue) en espace réel et sur grille régulière dans l’espace réel ou l’espace de Fourier sont toutes trois utilisées dans la littérature (par exemple [Hennawi and Spergel, 2005](#); [Gavazzi and Soucail, 2006](#); [Miyazaki et al., 2018](#); [Hamana et al., 2020](#); [Oguri et al., 2021](#)). Les approches sur une grille régulière reposent sur la pixellisation des données du catalogue. La mé-

thode standard pour répartir le cisaillement sur une grille régulière consiste à calculer simplement le cisaillement moyen par pixel (voir par exemple [Kaiser and Squires, 1993](#); [Van Waerbeke, 2000](#); [Pires et al., 2020](#)). Ce moyennage peut avoir pour effet de lisser le signal lentille global selon la résolution de la carte.

Dans ce manuscrit, nous avons introduit une nouvelle approche pour calculer le signal lentille sur grille régulière. Notre but est de nous rapprocher des performances de l'approche sur grille irrégulière en espace réel. Cette stratégie consiste à additionner le cisaillement de toutes les galaxies qui tombent dans chaque pixel. Ainsi, le cisaillement dans chaque pixel est pondéré par le nombre de galaxies dans ce pixel. Cette stratégie de pixellisation est plus adaptée à la détection par l'effet de lentille faible car chaque pixel contient une information supplémentaire sur la densité des galaxies. La carte de cisaillement, ainsi obtenue, peut ensuite être normalisée par le nombre de galaxies par pixel moyen estimé sur toute la carte afin de faciliter la comparaison avec d'autres approches. Cette normalisation globale n'a pas d'impact sur la détection car elle affecte les modes E et B de la même manière.

Pour étudier l'impact de l'utilisation d'une grille régulière sur la conservation du signal, nous avons ainsi considéré trois implémentations différentes auxquelles nous avons appliqué la convolution :

1. Le catalogue brut avec les données de cisaillement aux positions des galaxies, directement dans l'espace réel ;
2. Les cartes de cisaillement pixellisé obtenues en faisant la moyenne du signal des galaxies qui tombent dans chaque pixel (MOY).
3. Les cartes de cisaillement pixellisé obtenues en additionnant le signal des galaxies qui tombent dans chaque pixel (SOM).

Ces trois implémentations pour la convolution sont illustrées par le schéma de la figure 3.2. Le même catalogue brut de galaxies sources est mis en entrée des trois implémentations, et apporte ainsi l'identifiant des galaxies dans la table (Id gal), la position des galaxies (RA,Dec), et une des composantes du cisaillement (γ_1) pour cet exemple. Les galaxies (points noirs) sont convoluées (cercles rouges) à partir de leur position exacte dans le catalogue brut (gauche), ou convoluées à partir des positions sur la grille régulière et des valeurs de cisaillement moyennées (MOY - milieu) ou sommées (SOM - droite).

Tout d'abord, nous avons étudié les différences entre les trois implémentations à travers leurs spectres de puissance dans le domaine de Fourier. Pour ce test, nous avons à nouveau utilisé le filtre proposé par [Jarvis et al. \(2004\)](#) parce qu'il peut être facilement appliqué au cisaillement, à la fois à la position des galaxies, ou sur une grille régulière. À partir du catalogue de cisaillement d'un champ de $5^\circ \times 5^\circ$ sans bruit, nous avons utilisé les trois différentes implémentations pour générer une carte filtrée à partir de laquelle nous dérivons le spectre de puissance correspondant. Notons que la convolution dans le cas des implémentations sur grille régulière se fait ici en utilisant la transformée de Fourier. De plus, pour le calcul des spectres de puissance, on retire les bords de l'image où le signal peut être perturbé.

Les résultats sont présentés dans le panneau de gauche de la Figure 3.3, où les lignes bleue, verte en pointillés et orange en tirets montrent les spectres de puissance pour les implémentations 1, 2 et 3, respectivement. Les spectres de puissance sont très similaires, ce qui montre qu'en moyenne le signal est bien préservé par les trois approches. Pour comparer quantitativement les implémentations, nous avons calculé l'intégrale des spectres de puissance. Par rapport à l'approche sur le catalogue brut, nous constatons que l'intégrale du spectre de puissance de l'implémentation de l'espace de Fourier souffre d'une perte de 1.2% pour la pixellisation SOM et d'une perte de 1.5% pour la pixellisation MOY.

La table 3.1 montrait déjà l'efficacité de la grille régulière quant au temps de calcul. Face à cet énorme avantage s'ajoute une perte de signal globale peu significative si la taille de la grille est bien adaptée pour ne pas introduire de données manquantes. En effet, cette perte de signal est réduite si la taille de la grille régulière augmente, dans la mesure où on se rapproche davantage de la position réelle des galaxies. Cependant, une augmentation de la taille de la grille va de pair avec une augmen-

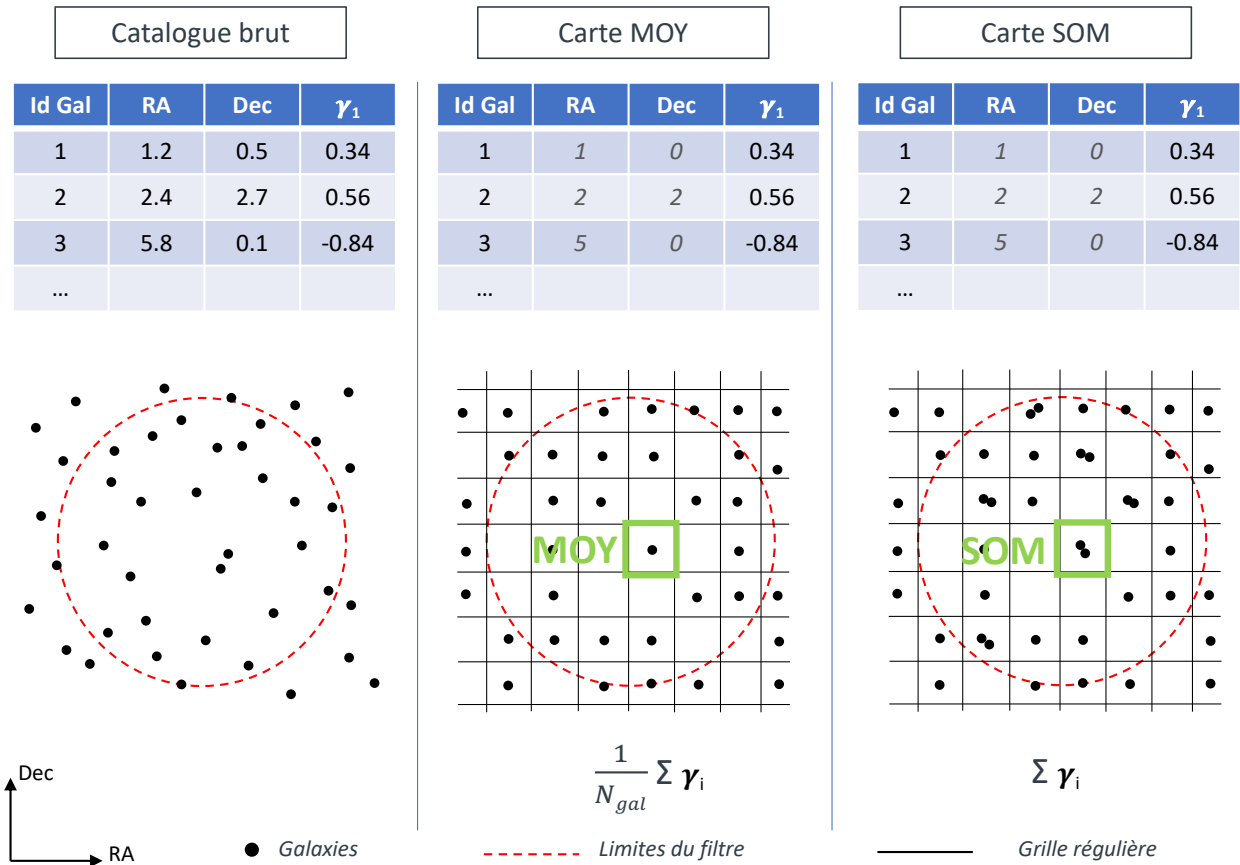


Figure 3.2 – Schéma des trois implémentations différentes (Catalogue, MOY et SOM) pour l’application de la convolution par un filtre (cercle rouge). *Gauche* : La convolution est appliquée sur toutes les galaxies (points noirs) dans les limites du filtre (cercle rouge) à partir de leur position exacte dans le catalogue de galaxies sources brut. *Milieu* : La convolution est appliquée sur une carte pixellisée à partir des galaxies dont la position est arrondie au pixel et dont les valeurs de cisaillement sont moyennées au sein du pixel. *Droite* : La convolution est appliquée sur une carte pixellisée à partir des galaxies dont la position est arrondie au pixel et dont les valeurs de cisaillement sont sommées au sein du pixel.

tation du nombre de données manquantes dans la grille régulière et un temps de calcul plus grand. Dans le cas de la simulation *DUSTGRAIN-pathfinder*, nous choisissons une taille de grille 512×512 pixels. Ce choix est motivé par un compromis entre précision et temps de calcul. De plus, l’utilisation des cartes de cisaillement obtenues en additionnant le signal des galaxies qui tombent dans chaque pixel permet de s’approcher au plus près du cas où on considère le signal à la position des galaxies. Cette analyse est complétée dans la section 3.1.4, par une comparaison du nombre de détections obtenues par chacune des implémentations. Il nous reste maintenant à étudier les différences entre les deux grilles régulières accessibles à travers les données : les cartes de cisaillement et les cartes de convergence.

3.1.3 . Impact de l’utilisation des cartes de cisaillement ou de convergence

Dans cette sous-section, nous comparons l’impact sur le signal de l’utilisation des cartes de cisaillement ou des cartes de convergence pour la détection des amas.

Le cisaillement et la convergence peuvent tous deux donner des indications sur les propriétés statistiques du champ de lentillage faible, et l’on peut même montrer qu’ils contiennent exactement les mêmes informations (e.g. [Schneider et al., 2002](#); [Pires et al., 2020](#)). Le cisaillement est une obser-

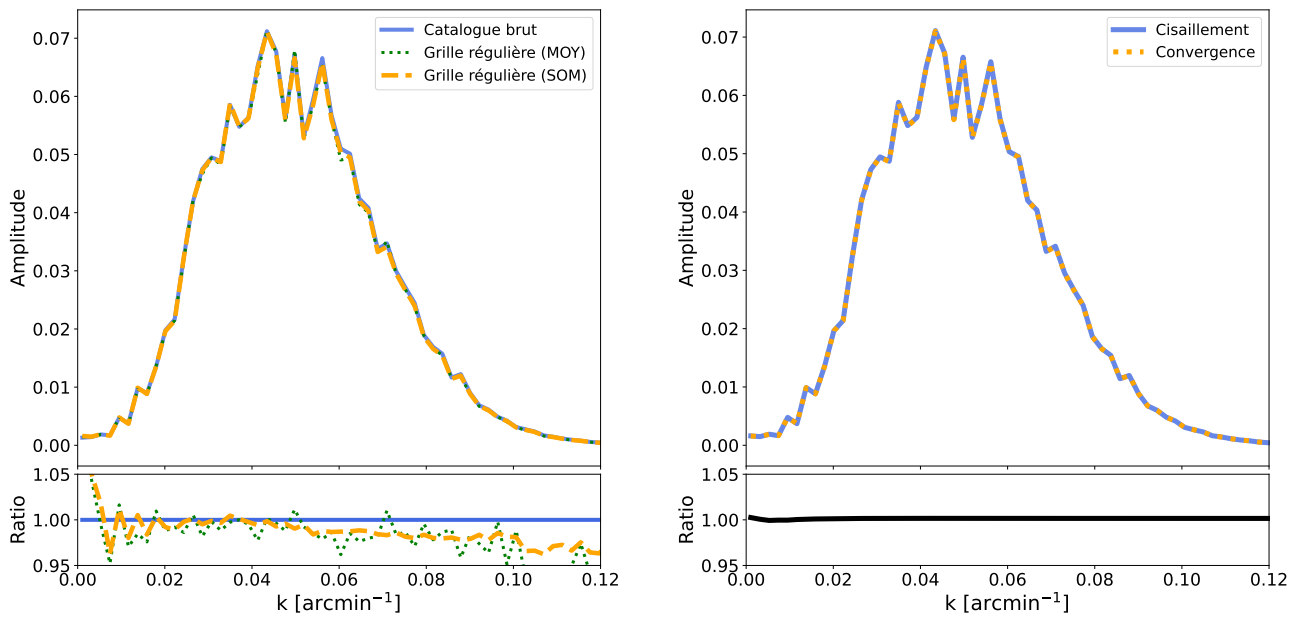


Figure 3.3 – Spectres de puissance des cartes convoluées dans le cas des implémentations sur grille irrégulière (espace réel) et sur grille régulière (espace de Fourier) pour un champ de $5^\circ \times 5^\circ$ avec 512×512 pixels. Gauche : Spectres de puissance obtenus en utilisant l’implémentation de l’espace réel (bleu), et en utilisant les implémentations de l’espace de Fourier (lignes orange pour MOY et verte pour SOM). Par rapport à l’implémentation sur l’espace réel, on observe une perte de puissance, respectivement, de 1.5% et de 1.2% pour l’implémentation avec la pixellisation MOY et SOM. Les ratios entre les différents spectres sont calculés par rapport au spectre issu de l’implémentation sur le catalogue brut. Droite : Spectres de puissance obtenus en appliquant le filtre de [Jarvis et al. \(2004\)](#) sur les cartes de cisaillement (ligne bleue) et sur la carte de convergence en mode E (points orange) calculée à partir des mêmes données que celles de la figure de gauche. Les spectres de puissance sont parfaitement concordants. Le ratio correspond au rapport entre les deux spectres.

vable directe et est généralement préféré pour des raisons de simplicité. Cependant, la convergence présente l’avantage essentiel que κ_E englobe tout le signal de lentille, alors qu’il est inévitablement partagé entre γ_1 et γ_2 dans le cisaillement. Dans ce contexte, la convergence est plus adaptée à la détection des amas de galaxies car elle trace explicitement la distribution totale de matière intégrée le long de la ligne de visée. Un autre avantage de l’utilisation de la convergence est son coût moindre en termes de temps de calcul (voir table 3.1).

Bien que la reconstruction du champ de convergence à partir des données observées soit une tâche difficile, [Pires et al. \(2020\)](#) montrent que le signal de lentille peut être préservé dans les cartes de convergence. Cela nécessite que l’inversion de masse soit effectuée sans régularisation du bruit, et que les effets systématiques tels que ces données manquantes et la géométrie complexe du relevé soient bien contrôlées.

Ainsi, le filtrage peut être réalisé soit sur les cartes de cisaillement, soit sur les cartes de convergence. Nous allons maintenant évaluer l’écart entre ces deux approches. Comme précisé en section 3.1.2, nous les avons d’abord comparées dans l’espace de Fourier. Pour chaque catalogue de cisaillement bruité, nous avons produit des cartes filtrées en utilisant à nouveau le filtre Jo4 défini par [Jarvis et al. \(2004\)](#) appliqué soit sur les cartes de cisaillement, soit sur les cartes de convergence. Nous avons ensuite calculé et comparé les spectres de puissance.

Le résultat de cette comparaison pour une réalisation est présenté dans le panneau de droite de la figure 3.3, où l’on voit que les spectres de puissance obtenus à partir des filtrages de la carte de cisaillement et de la carte de convergence sont en très bon accord. En faisant la moyenne des 256 réalisations, nous mesurons un écart de 0.03% dans l’intégrale du spectre de puissance entre les approches sur le cisaillement et sur la convergence. Cette différence peut s’expliquer par des effets de bord résiduels qui ne sont pas totalement supprimés malgré le retrait des bords de l’image pour le calcul du spectre.

Nous avons montré que le filtrage des cartes de cisaillement ou de convergence engendrait des spectres de puissance significativement similaires. Dans la sous-section suivante, nous comparons ces approches en termes de détections

3.1.4 . Impact du choix du traitement des données en matière de détection

Pour finaliser l'analyse des différents traitements des données, nous regardons comment les 3 approches que nous avons identifiées se comparent vis à vis des détections.

Pour chacune des trois implémentations (catalogue brut, MOY, SOM), nous avons convolué les données d'entrées à nouveau par le filtre défini par [Jarvis et al. \(2004\)](#) et appliqué la procédure de détection décrite dans la section 3.3.3. Nous nous contentons ici de comparer les détections obtenues par les différentes implémentations sur la même procédure.

Le panneau de gauche de la Fig. 3.4 compare les détections obtenues avec les trois différentes implémentations pour l'une des réalisations, où les cercles bleus indiquent les détections dans l'espace réel, et les cercles en pointillés verts et dorés indiquent les détections dans l'espace de Fourier avec la pixellisation MOY et SOM, respectivement. En faisant la moyenne de ces résultats sur les 256 réalisations de la simulation *DUSTGRAIN-pathfinder*, nous avons constaté que le nombre moyen de détections utilisant l'approche de l'espace réel (49.2 ± 0.6) est très proche de celui de l'approche de l'espace de Fourier avec la pixellisation SOM (47.9 ± 0.5). En revanche, l'approche dans l'espace de Fourier avec la pixellisation MOY donne un nombre moyen de détections plus faible (43.4 ± 0.5). De plus, un examen plus approfondi des détections effectuées par les différentes approches révèle des différences notables. Nous avons constaté que $28.2\% \pm 2.1\%$ des détections obtenues par l'approche de pixellisation MOY diffèrent de celles obtenues par l'approche de l'espace réel. En revanche, avec l'approche de pixellisation SOM, seules $7.6\% \pm 1.3\%$ des détections sont différentes.

Ces résultats montrent que la stratégie de prétraitement a un impact non négligeable sur l'efficacité de la détection, à la fois en termes de nombre et de distribution des détections. À la lumière de ces résultats, nous concluons que l'approche de l'espace de Fourier avec la stratégie de pixellisation SOM est en très bon accord avec l'approche de l'espace réel, produisant des résultats statistiquement similaires. De plus, ces résultats restent valables pour tous les filtres.

Nous comparons maintenant les approches sur le cisaillement et sur la convergence en termes de détections. Le formalisme de la masse d'ouverture peut être exprimé soit sur les cartes de cisaillement, soit sur les cartes de convergence. Pour ce test, nous avons à nouveau suivi les étapes réalisées dans la Sect. 3.1.3. Pour chaque réalisation du catalogue de cisaillement bruité, nous avons produit les cartes de convergence filtrées en mode E et B en utilisant le filtre défini par [Jarvis et al. \(2004\)](#) appliqué soit sur les cartes de cisaillement, soit sur les cartes de convergence. Nous avons ensuite comparé les deux implémentations en termes de détections, en appliquant la procédure de détection décrite dans la section précédente. Les résultats, présentés dans le panneau de droite de la Fig. 3.4, mettent en évidence la parfaite concordance entre les deux approches. Ainsi, les implémentations sur le cisaillement et sur la convergence sont équivalentes en termes de détections.

Notre choix de travailler sur la convergence n'a donc aucun impact sur l'efficacité des méthodes comparée à leur application sur le cisaillement. Ainsi, même si notre ambition est, à terme, de nous comparer avec notre approche uniquement applicable sur la convergence, cela n'altère en rien les performances des méthodes de filtres.

3.1.5 . Conclusion

Ainsi, nous avons vu trois implémentations possibles pour la mise en forme des données d'entrées de l'algorithme de détection. Il apparaît que la mise en forme des données d'entrée sous forme d'une grille régulière avant le filtrage permet de réduire drastiquement le temps de calcul. Ce résultat est d'autant plus important que le volume des données à traiter ne va cesser d'augmenter avec les

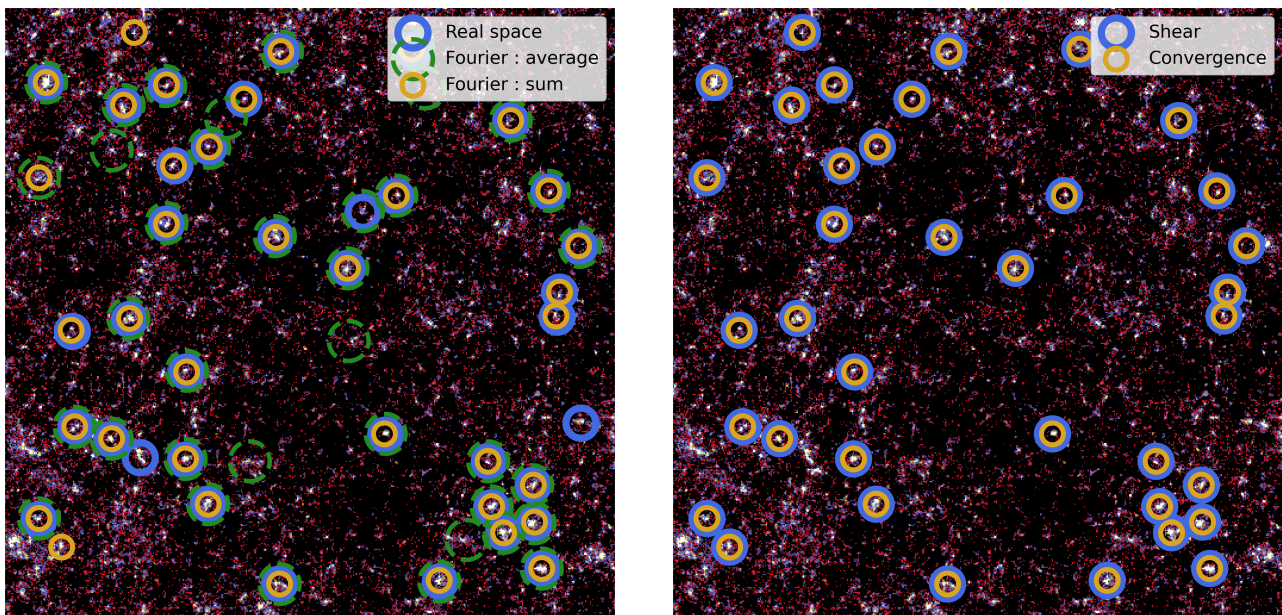


Figure 3.4 – Comparaison des implémentations dans l’espace réel et dans l’espace de Fourier puis des approches de cisaillement et de convergence. Gauche : Detections obtenues sur le même champ lorsque le calcul est effectué dans l’espace réel (cercles bleus) ou dans l’espace de Fourier en utilisant soit la pixellisation SOM (cercles dorés), soit la pixellisation MOY (cercles verts). L’implémentation dans l’espace de Fourier avec la pixellisation SOM est très proche de l’implémentation dans l’espace réel en termes de detections. Lorsque la pixellisation est effectuée en faisant la moyenne des valeurs de cisaillement, il y a moins de detections, et environ 30% de ces detections sont différentes de celles obtenues avec l’implémentation de l’espace réel. Droite : Detections obtenues sur le même champ lorsque la convolution par le filtre Jo4 est effectuée sur la carte de cisaillement (cercles bleus) et sur la carte de convergence (cercles dorés). Il y a une correspondance parfaite entre les detections.

prochains relevés. Les méthodes de detections doivent se montrer à la hauteur pour analyser dans un temps raisonnable les jeux de données à venir. Donc, cette ambition est à double tranchant car elle implique également que le temps de calcul accéléré ne soit pas obtenu au détriment de la conservation du signal. En ce sens, nous avons montré que l’utilisation de la grille régulière, dans laquelle les galaxies au sein d’un même pixel sont additionnées, permet d’approcher significativement le signal reconstruit à la position exacte des galaxies. De la même manière, l’utilisation des cartes de cisaillement ou des cartes de convergence pour l’algorithme de détection ne présente pas de différence significative. La différence majeure reste alors le temps de calcul, qui est 4 fois plus important dans le cas des cartes de cisaillement que des cartes de convergence (voir table 3.1). Dans notre étude, nous privilégions la mise en forme des données d’entrée en tant que carte de convergence. Ce résultat ne dépend pas de la méthode de filtrage utilisée. Dans un second temps, nous avons étudié l’impact de ces approches en matière de détection et montré que le traitement à partir des cartes de convergence ou de cisaillement mène aux mêmes detections pour une même pixellisation. Maintenant, nous allons nous intéresser aux différentes méthodes de filtrage.

3.2 Les méthodes de filtrage

Le filtrage est le point critique majeur de l’algorithme de détection aussi bien du point de vue des detections que du temps de calcul. Nous avons vu comment réaliser un filtrage rapide et précis en passant par la pixellisation sans perte de signal. Dans cette section, nous étudions les différents filtrages couramment utilisés que nous comparons ensuite à notre approche multi-échelle dans les chapitres suivants. Nous nous intéressons au filtrage gaussien pour sa primauté en traitement du signal et à la méthode de masse d’ouverture qui est aujourd’hui la plus couramment utilisée.

3.2.1 . Les enjeux du filtrage

En traitement du signal, le filtrage peut prendre différentes formes mais conserve un unique objectif : isoler le signal d'intérêt par rapport à d'autres signaux parasites. Dans notre cas, le signal parasite principal est le bruit, omniprésent dans les données astronomiques. Pour isoler le signal, il faut alors comprendre ce que représente le bruit et comment il se manifeste dans nos données.

Le bruit provient des erreurs de mesure, et surtout de la forme intrinsèque des galaxies. L'enjeu du filtrage consiste donc à isoler le signal des amas alors même qu'il est faible devant le niveau du bruit.

APARTE

Les fréquences spatiales

Nous venons ici d'introduire une notion fondamentale pour le traitement du signal : les fréquences spatiales. Dans cet aparté, nous souhaitons clarifier cette notion en illustrant précisément chaque composante au sein d'une image. Dans la Figure 3.5, on isole sur la même image ^a les basses fréquences (milieu) puis les hautes fréquences (droite). Les textures les plus fines correspondent aux hautes fréquences spatiales car elles varient très rapidement (détails, contour, bruit). Les basses fréquences correspondent aux variations lentes donc étendues dans l'image.

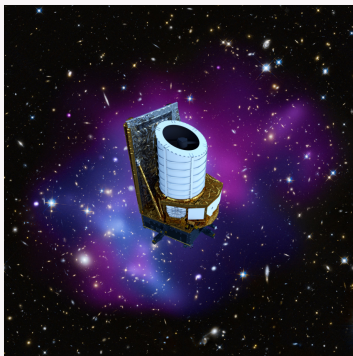


Image originelle



Basses fréquences



Hautes fréquences

Figure 3.5 – Illustration de la distribution des fréquences spatiales dans une image. De l'image originelle (gauche), on extrait les basses fréquences (milieu) puis les hautes fréquences (droite).

Les images basses et hautes fréquences ont respectivement été obtenues en appliquant un filtre passe-bas gaussien et un filtre passe-haut discret sur la décomposition RGB de l'image originelle, puis recombinaison.

^a. ESA/ATG medialab (spacecraft); NASA, ESA, CXC, C. Ma, H. Ebeling and E. Barrett (University of Hawaii/IfA), et al. and STScI (background)

Le bruit constitue des perturbations locales et non-corrélées entre elles qui affectent les galaxies puis les pixels une fois les cartes de cisaillement construites. Dans nos images, cela se traduit par une variation brusque des valeurs d'un pixel à l'autre. En d'autres termes, le bruit dans les images est présent à toutes les fréquences mais est largement dominant aux hautes fréquences spatiales, c'est-à-dire des signaux qui varient très rapidement. Pour réduire grandement le niveau de bruit, on utilise ainsi un filtre passe-bas qui ne laisse passer que les basses fréquences.

Mais pour la détection d'amas, le filtrage ne se limite pas à réduire le bruit. Comme nous l'avons mentionné, les données contiennent également du signal parasite venant d'autres structures de l'univers. En particulier, dans le cas de la détection des amas de galaxies, nous sommes sensibles à la présence des structures à grande échelle (ci-après LSS) (voir [White et al., 2002](#); [Hamana et al., 2004](#)). Ce signal parasite dû aux LSS correspond à un signal très étendu donc présent aux très basses fréquences; ce qui contamine le signal des amas qu'on cherche à détecter. La méthode de filtrage optimal doit alors réduire autant les basses fréquences que les hautes fréquences.

3.2.2 . Filtrage gaussien

Le filtrage gaussien est l'un des plus couramment utilisés en mathématiques tout comme en traitement du signal. C'est tout naturellement qu'il est aussi devenu l'une des premières approches utilisées pour la détection d'amas par effet de lentille gravitationnelle faible (e.g. [White et al., 2002](#); [Miyazaki et al., 2002](#); [Tang and Fan, 2005](#); [Gavazzi and Soucail, 2006](#); [Miyazaki et al., 2007](#); [Shan et al., 2012](#)). Comme son nom l'indique, il repose sur l'utilisation d'une fonction gaussienne bidimensionnelle :

$$F_{\text{gauss}}(\theta) = \frac{1}{\sqrt{2\pi}\theta_G} \exp^{-\frac{\theta^2}{2\theta_G^2}}, \quad (3.1)$$

avec θ la position et θ_G l'écart-type. Notons que l'écart-type définit la largeur de la fonction gaussienne et est ainsi assimilable à la taille du filtre (voir Fig. 3.6).

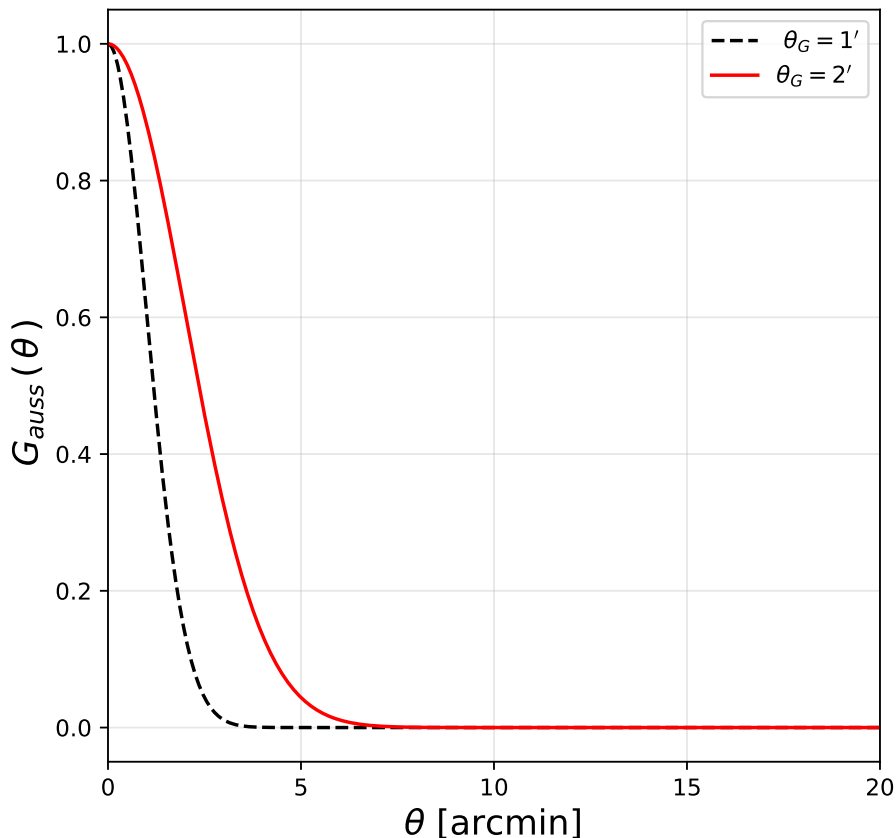


Figure 3.6 – Exemple de deux fonctions gaussiennes de largeur $\theta_G = 1'$ (rouge) et $\theta_G = 2'$ (noir).

Dans la pratique, le filtrage gaussien est réalisé en convoluant les données de lentillage faible par le filtre gaussien F_{gauss} . Cette convolution correspond à la moyenne pondérée des données D , autour du centre θ_0 :

$$C_{\text{gauss}}(\theta_0) = \int_{\mathbb{R}^2} D(\theta) F_{\text{gauss}}(|\theta - \theta_0|) d^2\theta. \quad (3.2)$$

La carte résultante C_{gauss} filtrée permet alors d'isoler les surdensités.

Afin d'adapter le filtre à la taille des amas que nous voulons détecter, θ_G est pris égal à $1'$ (voir Fig. 1.2). Pour ce choix, nous avons suivi les analyses de [Hamana et al. \(2004\)](#) qui montrent que cette valeur est optimale pour la détection des amas. Certaines études mettent également en avant $\theta_G = 2'$ (voir [White et al., 2002](#); [Shan et al., 2012](#)) mais [Tang and Fan \(2005\)](#) ont montré que les effets de projection sont plus significatifs avec un filtre plus large.

La figure. 3.7 illustre visuellement l'effet du filtrage gaussien ($\theta_G = 1'$) sur le bruit. Les pics candidats sont représentés par les structures en surbrillance. En comparant la carte sans bruit (gauche) à la carte



Figure 3.7 – Carte de convergence en mode E simulée pour un champ de $5^\circ \times 5^\circ$ sans le bruit (gauche), avec le bruit (milieu), bruitée puis filtrée par une gaussienne de largeur $\theta_G = 1'$ (droite). La carte de la convergence en mode E trace directement la distribution de matière projetée, les surdensités apparaissent plus brillantes.

filtrée (droite), on retrouve qualitativement les amas les plus massifs. La colorimétrie étant conservée sur les 3 cartes, on note que le signal de certains amas projetés a été lissé par le filtrage mais reste cohérent avec la carte sans bruit.

Le filtrage gaussien offre plusieurs avantages. Le premier avantage est le côté pratique et maîtrisable de la fonction gaussienne. C'est une fonction classique dont on connaît les propriétés et le comportement dans les espaces réel et de Fourier. Le second avantage est la réduction effective du bruit dans les données par la suppression des hautes fréquences.

Cependant, comme le soulignent [White et al. \(2002\)](#) et [Hamana et al. \(2004\)](#), la détection des amas par le filtrage gaussien est très sensible à la contamination par d'autres signaux dus aux LSS. En effet, ils observent un nombre très important de fausses détections qui dégradent l'efficacité de la méthode à bas SNR. Cela affecte la précision des détections, leur nombre, leur SNR et l'estimation de la masse des amas (voir également [Maturi et al., 2005](#)).

Dans le signal lentille, les LSS dominent les très basses fréquences spatiales et celles-ci ne sont pas exclues par le filtre gaussien, ce qui diminue les performances de l'algorithme de détection. Dans la figure. 3.8, nous illustrons cette limite du filtrage gaussien. Nous présentons des cartes de convergence simulées de $5^\circ \times 5^\circ$, originellement bruitées puis filtrées par deux filtres gaussiens de taille différente et un filtre "idéal". La colonne de droite présente le spectre des filtres associés dans le domaine de Fourier. Pour obtenir le spectre d'un filtre, on génère une carte de convergence de valeur nulle où on place un signal au centre (distribution de Dirac) égal à 1. Puis on calcule la transformée de Fourier de la carte convoluée par le filtre qu'on veut étudier. Ce spectre permet de voir le comportement du filtre en fonction des fréquences spatiales k . La courbe du spectre trace les fréquences qui sont acceptées par le filtre en question. Pour illustrer, nous avons superposé sur les spectres le signal dominant selon les fréquences : structures à grande échelle, amas ou bruit. Dans le cas du filtrage gaussien de $\theta_G = 1'$ (haut), la carte de convergence filtrée (en haut à gauche) présente des surdensités avec un niveau de bruit faible sur le spectre. On visualise que le filtre diminue bien le bruit aux plus hautes fréquences. En choisissant un filtre plus large avec $\theta_G = 6'$ (milieu), on voit apparaître sur la carte de convergence (au milieu à gauche) les structures à grande échelle sous-jacentes; le signal des amas est beaucoup plus faible mais présent. Sur le spectre (au milieu à droite), on voit que le filtre sélectionne des fréquences plus basses qui correspondent à des structures plus larges que les amas. Ce signal est un contaminant que nous pouvons soustraire à la carte pour avoir un filtrage idéal et ne conserver que les surdensités qui correspondent aux amas comme on peut le voir dans la carte de convergence en bas à gauche. Sur le spectre correspondant (en bas à droite), le comportement du filtre en est changé : on a un filtre qui n'accepte qu'une gamme précise de fréquences. Ce filtre est défini comme idéal pour la détection des amas car il réduit la contamination du signal par le bruit et les LSS.

Pour la suite de notre étude, nous avons choisi de conserver le filtre gaussien pour rendre compte

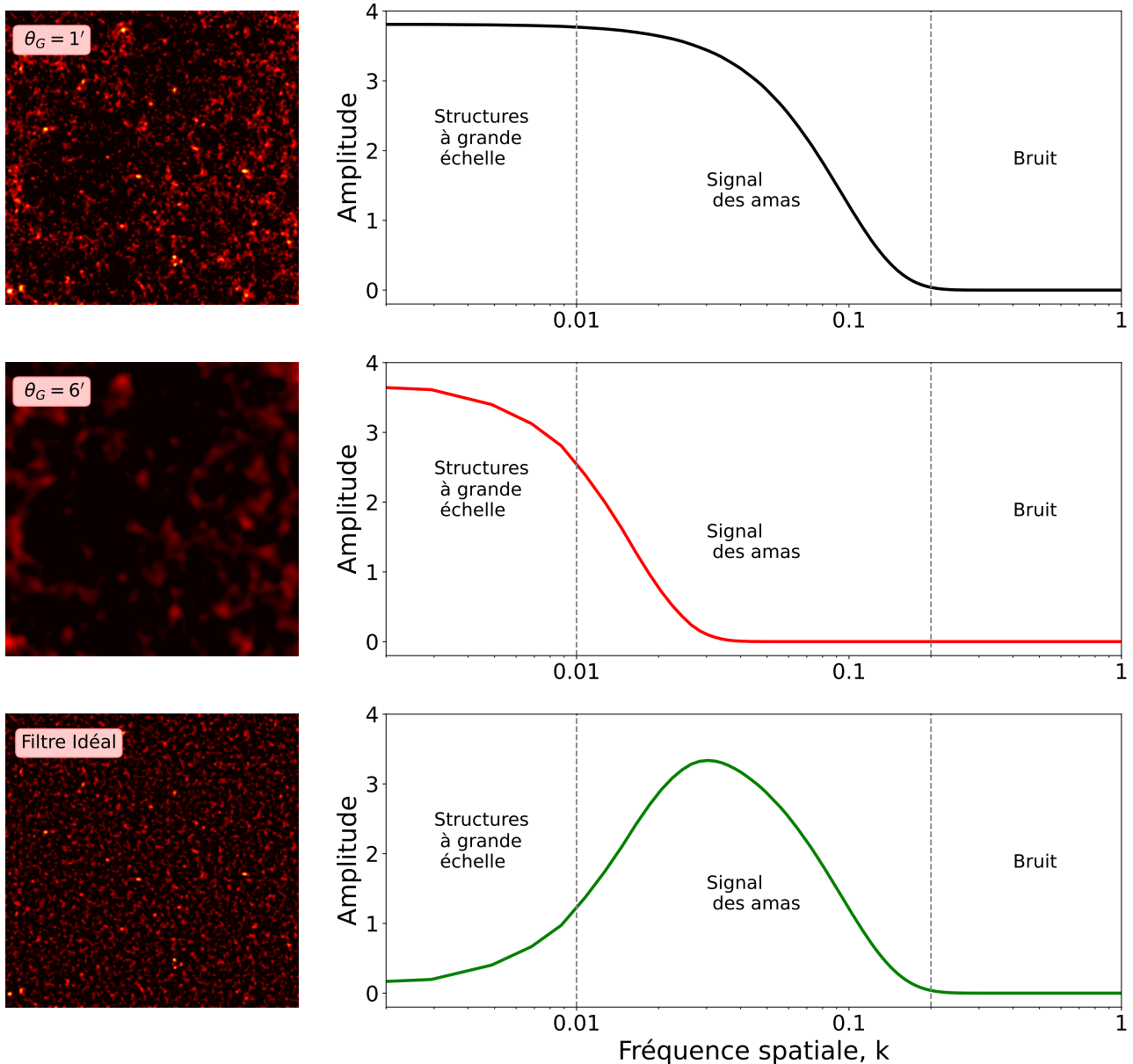


Figure 3.8 – Illustration schématique de la limite du filtrage gaussien à partir des cartes de convergence de $5^\circ \times 5^\circ$ simulées et filtrées et des spectres des filtres correspondants dans le domaine de Fourier. Nous avons représenté la carte filtrée par un filtre gaussien de largeur $\theta_G = 1'$ (haut) et son spectre (noir), la carte filtrée par un filtre gaussien de largeur $\theta_G = 6'$ (milieu) et son spectre (rouge), et finalement la différence des deux cartes précédentes dite 'Filtre Optimal' (bas) et le spectre qui en résulte (orange). Nous avons défini sur chaque intervalle de fréquences approximativement le signal qui domine : structures à grande échelle, amas, bruit.

de la dégradation des performances obtenues lorsque le filtre ne filtre pas le signal des structures à grande échelle. Pour autant, de nombreux autres filtres (Wiener, Ondelette) et méthodes de filtrage (MRLens, [Starck et al. 2006, 2011b](#); GLIMPSE, [Leonard et al. 2015](#)) ont été utilisés pour la détection des amas de galaxies et ne tiennent pas compte de ce problème. Comme l'utilisation de ces filtres, notamment les filtres passe-bas, n'est pas optimale pour l'application actuelle, nous concentrons notre analyse sur les filtres passe-bande qui sont mieux à même de cibler le signal autour d'une échelle donnée.

En résumé, la présence de structures à grande échelle a un impact lourd sur les performances du filtrage gaussien en termes de détections, leur nombre ou encore leur véracité. Par conséquent, le filtrage gaussien est sous-optimal pour la détection des amas par son incapacité à supprimer les très basses fréquences. C'est pourquoi une solution alternative a été envisagée en utilisant des filtres ca-

pables de supprimer les très hautes fréquences (dues au bruit) mais aussi les très basses fréquences (dues aux structures à grande échelle). On parle de filtres passe-bandes, car ils laissent passer seulement une bande déterminée de fréquences. On parle de détection à une échelle donnée. La notion d'échelle fait ici référence au rayon du filtre et donc au rayon apparent des structures qu'on vise à détecter (voir Fig. 1.2). Par exemple, un filtrage à faible échelle ($1'$) fait apparaître dans la carte filtrée des structures dont le rayon est compris entre $0.5'$ et $2'$.

3.2.3 . Approche standard

L'alternative au filtrage gaussien la plus couramment utilisée est la méthode de masse d'ouverture (MO) introduite par [Schneider \(1996\)](#). Nous en détaillons ici le formalisme et présentons 5 familles de filtres MO implémentées et utilisées pour notre comparaison. [Leonard et al. \(2012\)](#) montrent que la méthode de masse d'ouverture est comparable à notre approche par les ondelettes (voir Sect 4.2.1) en termes de comportement, qui permet ainsi une comparaison plus juste entre les méthodes. En particulier, nous avons sélectionné les filtres ayant déjà fait leurs preuves sur des données réelles (e.g. [Schirmer et al., 2007](#); [Miyazaki et al., 2018](#)).

Formalisme

La méthode de masse d'ouverture permet de reconstruire des cartes de matière projetée à une échelle donnée à partir des données de lentilles faibles. La carte de masse d'ouverture se calcule à la position θ_0 du ciel en convoluant la distribution de la convergence κ par un filtre U tel que :

$$M_{\text{ap}}(\theta_0) = \int_{\mathbb{R}^2} \kappa(\theta) U(|\theta - \theta_0|) d^2\theta, \quad (3.3)$$

où U est un filtre compensé ; c'est-à-dire que dans l'ouverture de rayon θ_{ap} ,

$$\int_0^{\theta_{\text{ap}}} \theta' U(\theta') d\theta' = 0. \quad (3.4)$$

En parallèle, on peut également définir la carte de masse d'ouverture à partir de la composante tangentielle du cisaillement γ_t , que l'on calcule directement avec γ_1 et γ_2 :

$$\gamma_t = \gamma_1 \cos(\phi) + \gamma_2 \sin(\phi), \quad (3.5)$$

où ϕ est l'angle polaire $\phi(\theta, \theta_0)$ par rapport au centre de l'ouverture θ_0 . La carte de masse d'ouverture peut alors être obtenue en convoluant la distribution de cisaillement tangentiel γ_t avec un filtre Q tel que :

$$M_{\text{ap}}(\theta_0) = \int_{\mathbb{R}^2} \gamma_t(\theta) Q(|\theta - \theta_0|) d^2\theta. \quad (3.6)$$

Les filtres U et Q sont liés par les relations ([Schneider and Bartelmann, 1997](#)) :

$$Q(\theta) = \frac{2}{\theta^2} \int_0^\theta \theta' U(\theta') d\theta' - U(\theta), \quad (3.7)$$

et

$$U(\theta) = 2 \int_\theta^{+\infty} \frac{Q(\theta')}{\theta'} d\theta' - Q(\theta). \quad (3.8)$$

L'efficacité de la méthode de masse d'ouverture pour détecter les amas de galaxies dépend du choix du filtre utilisé et de sa capacité à isoler le signal lentille autant du bruit que des structures à grande échelle.

Les filtres optimaux U et Q doivent remplir deux propriétés :

- Dans l'espace réel, les filtres doivent être définis "localement", c'est-à-dire que les filtres tendent vers 0 à un rayon fini. Idéalement, les filtres reproduisent la forme du signal lentille des amas.
- Dans l'espace de Fourier, les filtres doivent aussi être définis "localement" pour cibler une échelle angulaire spécifique. Idéalement, la fréquence centrale du filtre dans l'espace de Fourier doit correspondre à la taille du signal des amas.

Dès lors qu'un filtre remplit ces deux propriétés, il est considéré comme optimal pour isoler le signal provenant des amas. Notons que le filtre gaussien ne remplit aucune de ces deux propriétés. En effet, dans l'espace réel il tend vers 0 à l'infini et dans l'espace de Fourier, il admet toutes les fréquences jusqu'à la fréquence de coupure.

3.2.4 . Fonctions utilisées

De nombreuses familles de fonctions remplissent les deux conditions définies précédemment et peuvent ainsi servir à définir des filtres adaptés pour les amas. Dans les dernières décennies, différents filtres ou familles de filtres ont été proposés (e.g. [Schneider, 1996](#); [Schirmer, 2004](#); [Schirmer et al., 2007](#); [Jarvis et al., 2004](#); [Hamana et al., 2012](#); [Miyazaki et al., 2018](#)). Dans cette sous-section, nous introduisons les filtres MO les plus couramment utilisés pour la détection d'amas et qui constituent la base de comparaison pour notre méthode (voir Chapitre 4).

- La famille originelle de filtres définie par [Schneider \(1996, ci-après S96\)](#) :

$$U_{S96}(\theta) = \begin{cases} 1 & x \in [0, \nu_1 R], \\ \frac{1}{1-c} \left[\frac{\nu_1 R}{\sqrt{(x-\nu_1 R)^2 + (\nu_1 R)^2}} \right] & x \in [\nu_1 R, \nu_2 R], \\ \frac{b}{R^3} (R-x)^2 (x-\alpha R) & x \in [\nu_2 R, R], \\ 0 & x > R, \end{cases} \quad (3.9)$$

où les paramètres α , b et c assurent que le filtre est compensé et continu dans l'espace réel. Les paramètres ν_1 , ν_2 , R permettent de définir la taille angulaire du filtre et ainsi la taille des objets qui sont ciblés. Ces paramètres doivent être adaptés à la taille des amas qu'on veut détecter. Nous avons implémenté les différentes paramétrisations présentées dans [Schneider \(1996\)](#) et nous avons utilisé les valeurs $\nu_1 = 0.1$, $\nu_2 = 0.9$, $R = 9'$, $\alpha = 0.8531$, $b = -329.8$ et $c = 0.2415$, car elles maximisent les résultats du filtre en termes de pureté et complétude pour la détection d'amas (voir Sect. 6.1). Le filtre Q_{S96} associé peut être calculé analytiquement en utilisant l'Eq. 3.7. Son expression étant lourde et peu utile pour la suite, nous ne l'écrivons pas ici.

- Une seconde famille de filtres définie par [Schneider et al. \(1998, ci-après S98\)](#) :

$$U_{S98}(\theta) = \frac{(1+2)^2}{\pi R^2} \left(1 - \frac{\theta^2}{R^2} \right)^1 \left(\frac{1}{l+2} - \frac{\theta^2}{R^2} \right) H(R-\theta), \quad (3.10)$$

$$Q_{S98}(\theta) = \frac{(1+1)(2+1)\theta^2}{\pi R^4} \left(1 - \frac{\theta^2}{R^2} \right)^1 H(R-\theta) \quad (3.11)$$

avec R le rayon de troncation, l l'ordre du filtre et H la fonction de Heaviside, qui prend la valeur 1 si $\theta < R$ et 0 si $\theta > R$. La fonction de Heaviside assure que le filtre est compensé et défini localement dans les espaces réel et de Fourier. Par la suite, nous utilisons $l = 3$ comme suggéré par [Schneider et al. \(1998\)](#) et $R = 8'$ pour obtenir un filtre très compact, afin d'éviter les oscillations dans le domaine de Fourier. Grâce à une telle paramétrisation, le filtre S98 cible des amas avec un rayon angulaire apparent d'environ $2'$.

- Le filtre développé par [Schirmer \(2004, ci-après TANH\)](#) :

$$Q_{TANH}(\theta) = \frac{1}{1 + e^{6-150 \frac{\theta}{R}} + e^{-47+50 \frac{\theta}{R}}} \frac{\tanh\left(\frac{\theta}{x_c R}\right)}{\frac{\theta}{x_c R}}, \quad (3.12)$$

où R est le rayon de troncature et x_c (sans dimension) définit la largeur du filtre. Ce filtre est directement dérivé de la forme du cisaillement tangentiel du profil NFW (Navarro et al., 1996). Conformément à Schirmer et al. (2007), nous utilisons par la suite $x_c = 0.1$. Nous avons fixé le rayon de troncature à $R = 7'$ comme le choix optimal pour maximiser le nombre de détections, en suivant la procédure décrite par Hetterscheidt et al. (2005). Le filtre U_{TANH} associé peut être calculé numériquement à l'aide de l'Eq. 3.8. L'utilisation du modèle des amas selon un profil NFW vise à modéliser le filtre le plus précis possible. Notons que cette approche a également été suivie par Maturi et al. (2005) dans le but de retirer plus précisément la contribution des structures à grande échelle.

- Le filtre proposé par Jarvis et al. (2004, ci-après J04) :

$$U_{J04}(\theta) = \frac{1}{2\pi\sigma^2} \left(1 - \frac{\theta^2}{2\sigma^2} \right) \exp\left(-\frac{\theta^2}{2\sigma^2}\right), \quad (3.13)$$

$$Q_{J04}(\theta) = \frac{1}{2\pi\sigma^2} \frac{\theta^2}{2\sigma^2} \exp\left(-\frac{\theta^2}{2\sigma^2}\right), \quad (3.14)$$

où U_{J04} a été défini selon Van Waerbeke (1998) comme la dérivée seconde d'une fonction gaussienne avec un écart-type σ . Le U_{J04} est également connu sous le nom de filtre en ondelette du chapeau mexicain (Mexican Hat) ou ondelette de Ricker. Dans le calcul de la masse d'ouverture qui suit, nous utilisons $\sigma = 4'$, correspondant à un rayon angulaire apparent d'environ $4'$, et un rayon de troncature $R = 20'$, comme suggéré par Leonard et al. (2012) pour minimiser les oscillations et la contamination par les modes à hautes fréquences.

- Le filtre introduit par Hamana et al. (2012) et, utilisé par Miyazaki et al. (2018, ci-après M18) sur les données HSC, est défini comme un filtre gaussien tronqué :

$$U_{M18}(\theta) = \frac{1}{\pi\theta_s^2} \exp\left(-\frac{\theta^2}{\theta_s^2}\right) - U_0, \quad (3.15)$$

où U_0 est un paramètre qui assure que le filtre est compensé et

$$Q_{M18}(\theta) = \frac{1}{\pi\theta^2} \left[1 + \left(1 - \frac{\theta}{\theta_s} \right)^2 \right] \exp\left(-\frac{\theta^2}{\theta_s^2}\right), \quad (3.16)$$

avec θ_s défini comme la taille angulaire de l'ouverture. Dans la pratique, le choix de θ_s influence le rayon angulaire des amas ciblés par le filtre, mais la relation n'est pas linéaire ou directe. En suivant la paramétrisation de Miyazaki et al. (2018), nous utilisons $\theta_s = 1.5'$, et $U_{M18}(\theta) = 0$ si $\theta > R$, avec un rayon de troncature $R = 15'$. Cette valeur de R est fixée pour réduire l'impact de la troncature sur le comportement du filtre étant donné qu'il est dérivé d'un filtre gaussien défini sur \mathbb{R} . En effet, le filtre U_{M18} atteint des valeurs négatives en R , mais la troncature force $U_{M18}(R) = 0$. Ceci entraîne donc un saut dans les valeurs du filtre qui peut légèrement contaminer le signal dans l'ouverture. En parallèle, le choix du rayon de troncature a un impact lourd sur le temps de calcul. Il y a donc un compromis entre le temps de calcul et le comportement du filtre à prendre en compte. Le filtre M18 est actuellement utilisé pour le traitement des données HSC. Son application sur les données de la première année du relevé a permis de détecter 65 amas de galaxies avec $\text{SNR} > 4.7$

Dans la figure 3.9, nous comparons les différentes fonctions de filtrage Q et U , dans l'espace réel. Cette comparaison met en évidence les différences dans la forme des filtres. En particulier, nous portons notre attention sur leurs propriétés compensées et locales. Les courbes bleues en pointillés correspondent aux fonctions MO définies dans Schneider (1996). Le filtre U_{S96} est compensé et local par définition. Les courbes bleues continues correspondent aux fonctions MO définies dans Schneider et al. (1998). Le filtre U_{S98} est également compensé et local par définition. Les courbes rouges en

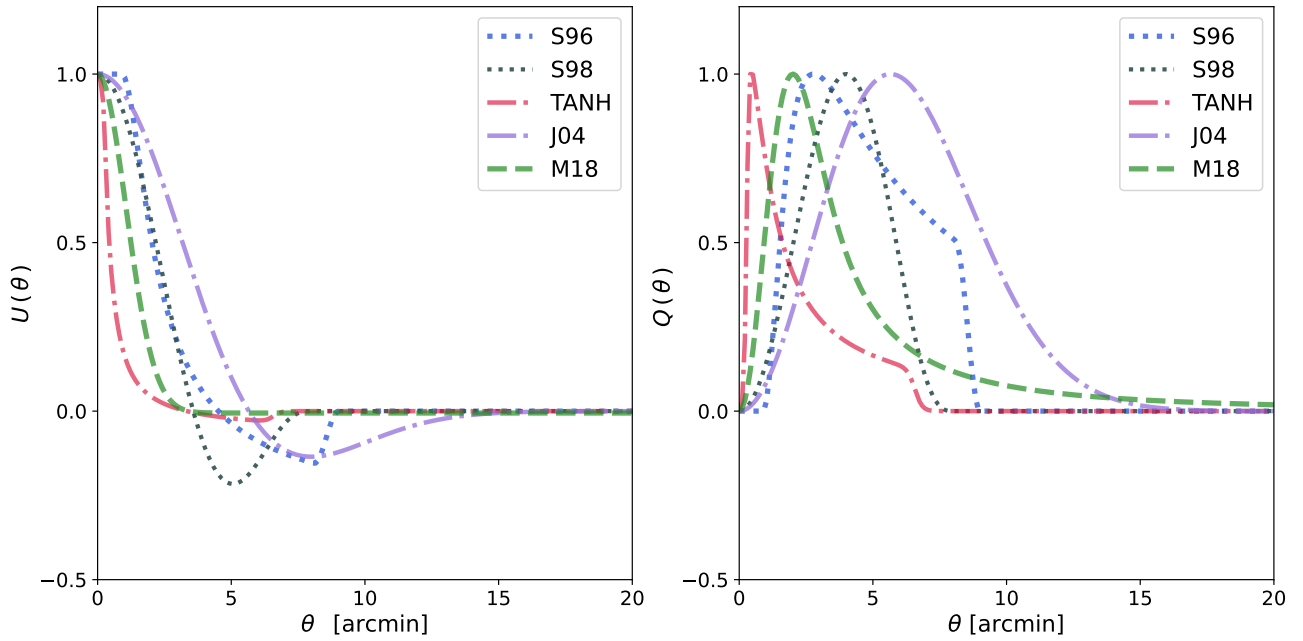


Figure 3.9 – Comparaison de différentes fonctions de filtrage U (gauche) et Q (droite) pour la reconstruction de la carte de masse d’ouverture en espace réel. La figure montre en bleu (lignes pointillées) les fonctions de filtrage de la masse d’ouverture définies par [Schneider \(1996, S96\)](#) avec $R = 9'$; en bleu clair (lignes solides), les fonctions de filtrage de la masse d’ouverture définie par [Schneider et al. \(1998, S98\)](#) avec $R = 8'$; en rouge (lignes pointillées), les fonctions de filtrage de la masse d’ouverture définies par [Schirmer \(2004, TANH\)](#) avec $R = 7'$; en violet (lignes pointillées), les fonctions de filtrage de la masse d’ouverture définies dans [Jarvis et al. \(2004, J04\)](#) avec $\sigma = 4'$; en vert (lignes pointillées), les fonctions de filtrage de la masse d’ouverture définies par [Miyazaki et al. \(2018, M18\)](#). Le filtre U_{M18} est également un filtre compensé grâce au terme U_0 , mais en raison du rayon de troncature élevé que nous avons utilisé, cela n’est pas distinctement visible. Tous les filtres sont normalisés pour avoir une amplitude maximale de 1.

pointillés correspondent aux fonctions MO définies dans [Schirmer \(2004\)](#). Le filtre U_{TANH} tombe rapidement à zéro, mais il n’est pas local et n’est compensé qu’à l’infini. La localisation est imposée par un rayon de troncature à $R = 7'$. On remarque que S96, et TANH présentent des ruptures abruptes qui témoignent d’une volonté de grandement réduire la contribution des objets en dehors d’une ouverture relativement fine. Les courbes violettes en pointillés correspondent aux fonctions MO définies dans [Jarvis et al. \(2004\)](#). Ce filtre U_{J04} n’est pas non plus local et n’est compensé qu’à l’infini puisqu’il tend vers zéro lorsque θ tend vers $+\infty$. En pratique, le filtre est tronqué à $R = 20'$ pour avoir un support local. Les courbes vertes en pointillés correspondent aux fonctions MO définies dans [Miyazaki et al. \(2018\)](#). Dans la définition de U_{M18} , U_0 garantit que le filtre est compensé dans une ouverture de rayon R . Cependant, la fonction Q_{M18} est non-locale, ce qui peut introduire des erreurs lors de la troncature de l’ouverture.

Notons que le comportement des filtres dans l’espace réel vise à approcher la forme du cisaillement tangentiel des amas qu’on cible. Leurs différences reposent sur la présence ou non d’une partie négative dans le cas des filtres U , qui est la traduction visuelle de la condition de compensation des filtres. Pour compléter cette analyse, il est essentiel de regarder le comportement des filtres dans le domaine de Fourier détaillé en section. 4.3.

3.2.5 . Conclusion

Dans cette section, nous avons introduit les différents filtres à une échelle donnée couramment utilisés pour la détection des amas de galaxies à partir de leur signal lentille. Le filtre gaussien est le premier utilisé pour le traitement du signal et la réduction du bruit dans de nombreux domaines. Cependant, dans notre étude, il conserve également le signal provenant des structures à grande échelle

et n'est donc pas optimal pour la détection des amas de galaxies par effet de lentille faible. A l'inverse, la méthode de masse d'ouverture (MO) a été développée pour être optimale pour la détection des amas. Nous avons ainsi détaillé les filtres MO les plus couramment utilisés (S96, S98, TANH, J04, M18) et qui forment la base de comparaison par rapport à notre méthode multi-échelle décrite en Chapitre 4.

3.3 La procédure de seuillage et l'identification des détections

Après avoir vu en détail les enjeux du filtrage, nous décrivons maintenant l'algorithme complet de détection des amas que nous avons utilisé dans notre étude pour la comparaison des différentes méthodes. Dans cette section, nous présentons les deux dernières étapes de l'algorithme de détection que nous avons mis en place, dans le cas des méthodes qui ciblent une échelle unique, soit utilisant un filtre unique.

3.3.1 . Le seuillage

Une fois que les cartes de convergence en mode E et B filtrées à une échelle spécifique sont produites, la procédure de détection continue par une étape de seuillage. Le seuillage sur les cartes de convergence filtrées est identique quelle que soit la méthode utilisée pour les produire. Dans cette sous-section, nous détaillons comment le bruit est estimé et comment les cartes sont seuillées.

Dans la pratique, l'effet de lentille gravitationnelle faible direct ne produit que des modes E. Les modes B sont, quant à eux, largement dominés par le bruit (voir par exemple [Hilbert et al., 2009](#); [Pereira et al., 2016](#)) et par conséquent, ils peuvent être utilisés pour évaluer le niveau de bruit. Comme proposé par [Hamana et al. \(2012\)](#); [Miyazaki et al. \(2018\)](#), nous utilisons la valeur maximale des pixels dans les cartes filtrées en mode B pour définir le seuil qui est ensuite appliqué à la carte filtrée en mode E. L'estimation du seuil par cette méthode est cependant assez instable et c'est pourquoi en pratique, nous calculons ce seuil en moyennant les valeurs des maxima des 256 cartes de convergence filtrées en mode B pour chacune des méthodes. La moyenne permet d'assurer un nombre de fausses détections moyen sur les simulations. Ceci est en particulier possible car nous travaillons sur des données simulées et semblables. Nous ne conservons alors que les pixels de la carte filtrée en mode E qui sont supérieurs à ce seuil, en mettant les autres pixels à zéro. En d'autres termes, la valeur moyenne des SNR ν_B maximaux dans les cartes en mode B est le seuil minimal autorisé sur la carte en mode E :

$$\min(\nu_E) > \text{moy}(\max(\nu_B)) = \frac{\max(M_B(\theta_o))}{\sigma_{\text{bruit}}}. \quad (3.17)$$

L'écart-type σ_{bruit} dépend du filtre considéré et est calculé à partir de la carte filtrée en mode B. Ce choix de seuil très conservatif permet de réduire grandement le nombre moyen de fausses détections. Mais ce faisant, il réduit le nombre global de détections. Il s'agit d'un compromis pour garantir un nombre de fausses détections, dues au bruit, le plus faible possible. Nous avons pu voir que le choix du seuil a un impact énorme sur les performances d'une méthode : une réduction de 5% du seuil peut entraîner jusqu'à 50% de détections en plus (voir Sect. 6.1) dont potentiellement des fausses détections.

Notons que la valeur du seuil peut grandement varier selon le filtre utilisé. Par la suite, nous présentons les résultats de l'algorithme de détection appliqué sur les 256 réalisations de la simulation DUSTGRAIN-*pathfinder* (voir [Giocoli et al., 2017](#)). Pour cette simulation, les seuils moyens pour chacun des différents filtres varient entre 4.01σ et 4.59σ . Ces valeurs sont précisées dans la table 3.2. À titre

illustratif, en utilisant les deux seuils moyens extrémaux sur la même carte filtrée, on obtient une différence jusqu'à 45% du nombre de détections obtenues.

Table 3.2 – Seuils moyens pour chacun des différents filtres, calculés à partir des 256 réalisations.

	Gauss	S96	S98	TANH	Jo4	M18
Seuil	4.59	4.09	4.36	4.37	4.01	4.35

Bien évidemment, la valeur de seuil dépend également des données et du bruit. Dans leur étude des données du relevé HSC, [Miyazaki et al. \(2018\)](#) trouvent un seuil (donné par le $\max(v_B)$) correspondant à 4.7σ . Mais en appliquant le même filtre sur la simulation DUSTGRAIN-*pathfinder* avec un bruit gaussien, nous trouvons un seuil légèrement différent et égal à 4.35σ . Ceci est cohérent avec le niveau de bruit plus élevé dans le cas de données réelles.

3.3.2 . L'identification des pics

La dernière étape de l'algorithme consiste à identifier les pics dans chacune des 256 cartes en mode E filtrées et seuillées. Dans cette sous-section, nous décrivons la définition choisie pour identifier un pic de signal sur les cartes filtrées et seuillées.

Nous définissons un pic comme un pixel dont la valeur est supérieure à celle de ses huit voisins les plus proches. On parle de maxima locaux sur la carte filtrée. Les pics situés en bordure de la carte sur une largeur égale au rayon du filtre sont écartés pour éviter toute contamination due aux effets de bord introduits par l'étape de convolution. En suivant cette procédure, nous obtenons une liste de détections candidates et les coordonnées des pics associés.

Cette approche pour l'identification des pics est l'une des plus simples mais aussi l'une des plus courantes. Des alternatives existent pour tenir compte de la taille des détections. En particulier, [Miyazaki et al. \(2018\)](#) et [Hamana et al. \(2020\)](#) choisissent par exemple, des pics pour lesquels 3 pixels sont supérieurs à un seuil donné. Cette idée permet de tenir compte de la largeur des amas pour avoir des détections plus fiables. Mais ceci dépend directement de la largeur du filtre et pénalise les filtres de taille plus fine ou à une échelle plus petite. À terme, avec l'objectif de comparer les méthodes entre elles, l'identification par les maxima locaux est la plus juste.

L'algorithme complet décrit ici suit l'implémentation que nous avons appliquée pour le traitement sur les données simulées. La même procédure est appliquée à chacun des filtres afin de les comparer de manière juste. Au-delà d'une comparaison entre eux, ces filtres forment la base de comparaison pour nous permettre d'évaluer les performances de notre approche vis à vis des méthodes les plus performantes à l'heure actuelle. Cet algorithme est en accord avec les implémentations des filtres proposés lors de leur publication originale.

3.3.3 . Récapitulatif de l'algorithme de détection

Nous avons introduit les 4 étapes clés de l'algorithme de détections dans la figure 3.1. Puis, ces différentes étapes ont été décrites en détail dans les sections précédentes. Nous proposons ici un récapitulatif global de l'algorithme tel qu'il est implémenté pour les méthodes à une échelle donnée, et sur lequel reposent les résultats présentés dans ce manuscrit (voir Sect. 6.1).

- **(A) : Prétraitement des données**

À partir des catalogues de forme des galaxies sources bruités, nous construisons les cartes de cisaillement γ_1 et γ_2 , de 512×512 pixels, en additionnant les valeurs d'ellipticités qui tombent dans un même pixel. Puis, à partir de ces cartes de cisaillement γ_1 et γ_2 , nous reconstruisons les cartes de convergence en modes E et B par l'algorithme d'inversion de masse de [Kaiser and Squires 1993](#).

- **(B) : Filtrage**

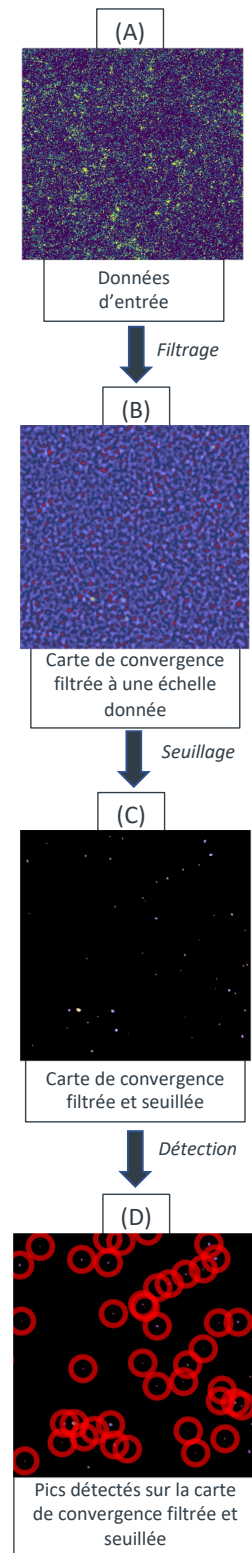
Une fois les cartes de convergence de modes E et B reconstruites, nous les convoluons par le filtre en suivant l'approche gaussienne ou MO pour obtenir les modes E et B filtrés.

- **(C) : Seuillage**

À partir de toutes les cartes de convergence en mode B filtrée, nous calculons le seuil $\nu\sigma$, pris comme la moyenne des maximums mesurés sur les 256 réalisations de la simulation *DUSTGRAIN-pathfinder*. Puis, ce seuil fixe, calculé pour chacun des filtres, est appliquée sur la carte de convergence en mode E filtrée.

- **(D) : Détection des pics**

Nous identifions les pics par les maxima locaux dans chaque carte de convergence de mode E filtrée et seuillée.



3.4 Conclusions

Dans ce chapitre, nous avons entrepris de présenter l'algorithme de détection des amas à partir de leur signal de lentille. Cette procédure se compose de 4 étapes essentielles : le traitement des données d'entrée, le filtrage des cartes de convergence, le seuillage des cartes filtrées et la détection des pics de surdensités. Chacune des étapes va de pair avec des difficultés qu'il a fallu surmonter :

- Afin de concevoir une approche rapide et sans perte de signal, nous avons montré que travailler sur des grilles régulières en additionnant les ellipticités des galaxies qui tombent dans un même

pixel permet de conserver le plus de signal tout en ayant une implémentation rapide. Ce faisant, on tient également compte du nombre de galaxies sources présentes dans chaque pixel. Nous avons choisi de travailler sur les cartes de convergence, que l'on reconstruit par l'inversion de [Kaiser and Squires \(1993\)](#), car elles présentent le même niveau de signal que les cartes de cisaillement tout en étant 4 fois plus rapides à traiter lors du filtrage.

- Afin de nous comparer aux méthodes les plus couramment utilisées pour la détection des amas de galaxies, nous avons implémenté 6 méthodes de filtrages. Le premier filtrage utilisé est un filtre passe-bas gaussien, très souvent utilisé en traitement du signal pour isoler le signal d'intérêt par rapport au bruit. Cependant, nous avons montré qu'il n'est pas adapté pour retirer la participation des LSS, contrairement au filtrage passe-bande. C'est pourquoi nous avons également présenté la méthode de masse d'ouverture et 5 filtres ou familles de filtres MO aussi complémentaires que d'usage courant dans la littérature (S96,S98,TANH,J04,M18).
- Pour le seuillage, nous avons choisi d'être conservatif en considérant le maximum des cartes filtrées de mode B comme seuil minimal de détection des amas. Dans le cas des données simulées, nous calculons le seuil moyen sur les 256 réalisations des cartes filtrées, puis nous seuillons chacune des cartes filtrées en mode E.
- Enfin, pour la détection des pics candidats, nous identifions les maxima locaux de la carte filtrée et seuillée en mode E. Puis, ces pics candidats sont regroupés en catalogues de détections.

Dans ce chapitre, nous avons présenté l'algorithme de détection à une échelle donnée et les différentes familles de filtres couramment utilisées qui représentent l'état de l'art dans la détection des amas de galaxies par effet de lentille gravitationnelle faible. Ces différentes méthodes de filtrage forment ainsi une excellente base de comparaison pour notre approche multi-échelle que nous introduisons dans le chapitre suivant.

Notons que les méthodes de filtrage sélectionnées dans ce manuscrit constituent une liste représentative des familles de filtres utilisées mais non exhaustive. En effet, d'autres méthodes de filtrage ont été développées pour s'adapter aux spécificités de lentillage faible et des amas. Comme nous l'avons précisé, la détection des amas est fortement influencée par la distribution en redshift des galaxies sources. Pour tenir compte de cette réalité et améliorer encore les performances des filtres, des méthodes ont été développées pour sélectionner les galaxies sources selon leur redshift.

*
* *

4

La méthode multi-échelle par les ondelettes

Diviser pour mieux régner.

attribuée à Phillippe II de Macédoine

4.1	De l'art de transformer le signal	67
4.1.1	La transformée de Fourier et ses limites	68
4.1.2	Une solution : la transformée en ondelettes	69
4.2	La méthode de détection multi-échelle	71
4.2.1	Le formalisme des ondelettes...	71
4.2.2	... au service de la détection des amas	72
4.2.3	La procédure multi-échelle	73
4.3	Avantages des ondelettes comparés aux autres filtres	74
4.3.1	Comparaison avec les filtres MO	75
4.3.2	Comparaison en termes de détections	77
4.4	Conclusions	79



travers le chapitre précédent, nous avons présenté les enjeux de l'algorithme de détection et en particulier du filtrage. Nous avons montré que l'idée principale de la détection était de transformer le signal lentille pour l'analyser et l'isoler. Cette transformation va de pair avec une perte d'information car l'utilisation d'un filtre à une échelle donnée, dans les méthodes les plus courantes, ne tient pas compte des différentes tailles des amas. Dans ce chapitre, nous introduisons une nouvelle approche dite multi-échelle pour la détection des amas, qui se fonde sur la transformation en ondelettes qui permet de décomposer un signal en différentes ondelettes. Dans la section 4.1, nous commençons par une introduction sur les limites de la transformée de Fourier qui a introduit les bases mathématiques de la transformée en ondelettes. Dans la section 4.2, nous décrivons notre méthode multi-échelle qui repose sur cette transformée en ondelettes. Enfin, en section 4.3, nous comparons les filtres ondelettes aux filtres introduits dans le chapitre précédent, dans l'espace réel et l'espace de Fourier et en termes de détections.

4.1 De l'art de transformer le signal

Nous avons vu précédemment que l'algorithme de détection à une échelle donnée cherche à transformer le signal des amas pour l'isoler plus facilement du bruit. Cette transformation du signal doit

idéalement en changeant la forme sans perdre d'information utile. La transformée de Fourier est celle qui a été la plus souvent utilisée pour analyser des signaux, notamment en permettant d'en révéler les gammes de fréquences qui les composent. Dans cette section, nous présentons la transformée de Fourier, ses limites, et l'avènement de la transformée en ondelettes qui en découle.

4.1.1 . La transformée de Fourier et ses limites

La volonté de transformer le signal pour pouvoir l'analyser n'est pas récente. Avant même que la question ne se pose sur les images, le problème s'est posé pour l'analyse des signaux physiques : Comment isoler et analyser les différentes composantes d'un signal ?

Au *XIX^{ème}* siècle, le français Jean-Baptiste Joseph Fourier démontre que toute fonction périodique f peut être décomposée en une somme infinie de signaux cosinus et sinus de fréquences et d'amplitudes variables. Cette décomposition, appelée série de Fourier, permet d'obtenir le spectre discret d'une fonction périodique en calculant les coefficients de Fourier $c_n(f)$ associée à la décomposition :

$$\forall t \in [0, 1], f(t) = \sum_{n=-\infty}^{+\infty} c_n(f) e^{i2\pi nt} = \sum_{n=-\infty}^{+\infty} c_n(f) (\cos(2n\pi t) + i \sin(2n\pi t)), \quad (4.1)$$

avec

$$c_n(f) = \int_{-1/2}^{1/2} f(t) e^{-i2\pi nt} dt. \quad (4.2)$$

Originellement, cette série fut définie uniquement pour les fonctions périodiques. En transformant cette somme discrète en somme continue, on peut étendre la décomposition aux fonctions intégrables. La transformée de Fourier est née et avec elle, les possibilités infinies de l'analyse du signal à partir d'un spectre continu. On peut alors décomposer toute fonction intégrable en une somme de ces différentes composantes et les isoler dans le spectre. Cette révolution mathématique a permis le développement des télécommunications, de la téléphonie à l'informatique en passant par le développement de l'analyse numérique.

APARTE

La transformée de Fourier bidimensionnelle

L'application de la transformée de Fourier ne se limite pas aux signaux 1D. Dans le cas bidimensionnel, la transformée de Fourier $\mathcal{F}(x)$ d'un signal numérique $x(k, l)$ s'écrit :

$$\mathcal{F}_{f,g}(x) = \sum_{k=-\infty}^{+\infty} \sum_{l=-\infty}^{+\infty} x(k, l) e^{i2\pi(fk+gl)}. \quad (4.3)$$

Cette transformée de Fourier est une fonction complexe qu'on peut exprimer sous forme de module et d'argument :

$$\mathcal{F}_{f,g}(x) = |\mathcal{F}_{f,g}(x)| \exp(i \arg[\mathcal{F}_{f,g}(x)]). \quad (4.4)$$

On appelle spectre d'amplitude le module $|\mathcal{F}_{f,g}(x)|$ car il décrit la distribution des fréquences dans le plan (f, g) de l'amplitude du signal $x(k, l)$ à savoir des pixels de l'image. Le second terme $\arg[\mathcal{F}_{f,g}(x)]$ représente le spectre de phase du signal 2D. Notons qu'ici, les variables, k , et l , représentent des distances. La temporalité du signal 1D devient la distance dans l'espace 2D. C'est pourquoi on parle de fréquences spatiales, dans le cas bidimensionnel.

Cependant, cette analyse très efficace admet des limites. En 1D, la décomposition du signal est indépendante du temps et ne se focalise que sur les fréquences. Si une fonction varie dans le temps, alors la transformée de Fourier permet uniquement de dériver le spectre du signal global. Ce problème vient notamment de l'utilisation des fonctions cosinus et sinus dont l'information de phase est perdue lors de la décomposition.

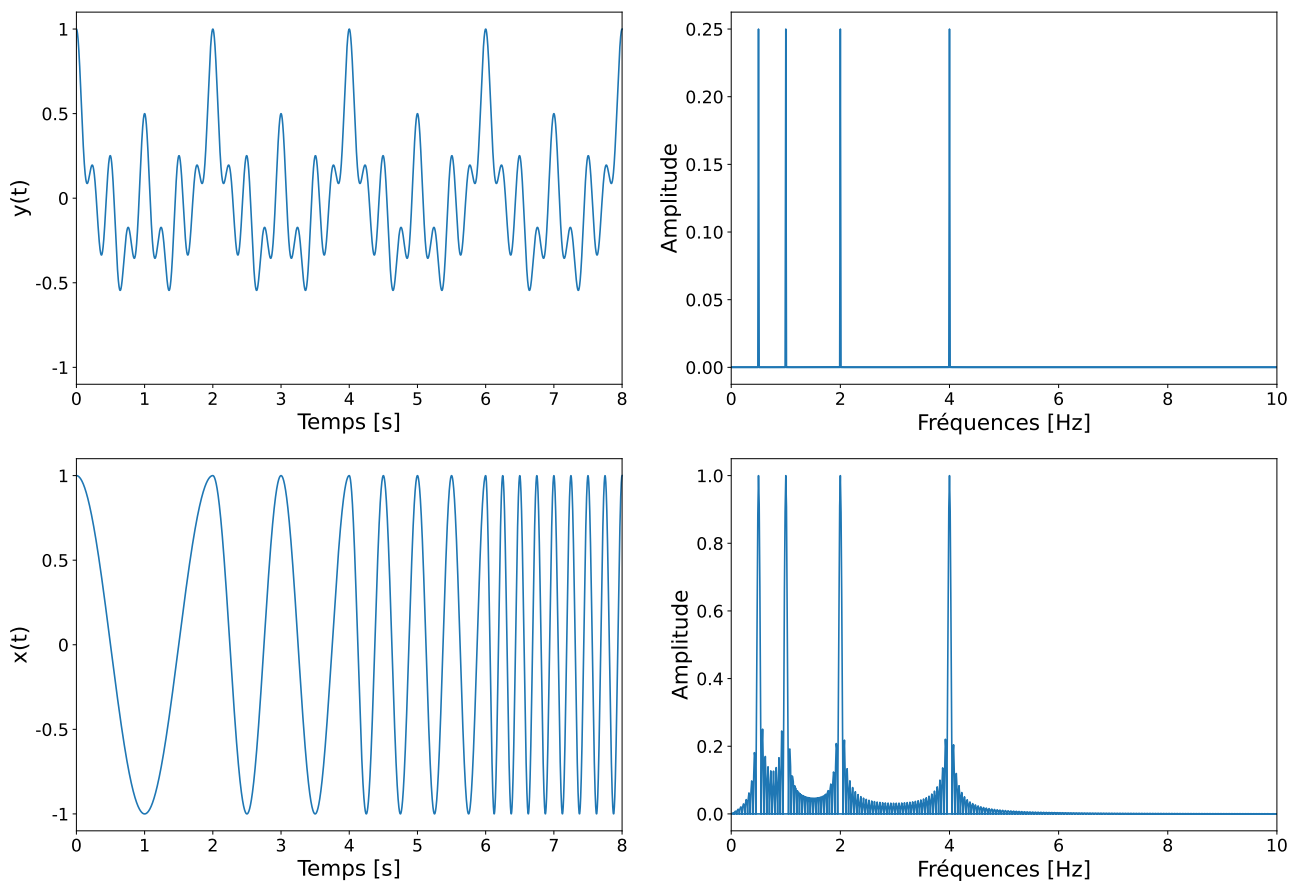


Figure 4.1 – Exemple de décomposition par la transformée de Fourier de deux signaux distincts dont le spectre est similaire. $x(t)$ est la suite de 4 fonctions sinusoïdales dont la période croît dyadiquement et $y(t)$ est la somme de ces 4 fonctions. La décomposition nous permet bien de retrouver les 4 fréquences des fonctions cosinus mais pas la succession temporelle.

Pour illustrer cette limite de la transformée de Fourier, on peut s'intéresser à deux signaux dont les fréquences sont identiques mais pas la temporalité. Dans la figure 4.1, on calcule la transformée de Fourier de deux signaux $x(t)$ et $y(t)$ différents mais dérivés des mêmes 4 fonctions cosinusoïdales. Dans le cas de $y(t)$, les 4 fréquences sont mélangées alors que pour $x(t)$, elles se succèdent dans le temps. On remarque que les pics fréquentiels principaux correspondent bien alors même que les signaux sont différents. Mais le moment d'apparition des fréquences n'apparaît pas dans le spectre de la transformée de Fourier. L'information temporelle est aussi dissimulée dans l'espace de phase. Il est donc possible de confondre les deux fonctions à la vue seule des fréquences centrales dans l'espace de Fourier.

Dans le cas de l'analyse des images, le problème est d'autant plus significatif avec la transformée de Fourier 2D. Cette incertitude temporelle des fréquences du signal devient une incertitude spatiale. L'application de la Transformée de Fourier aux images permet seulement de voir le spectre de l'image et les différentes composantes qu'elle contient. Même si le spectre dans l'espace de Fourier est un outil très efficace pour l'analyse, il se révèle limité pour la détection. A l'instar de l'information temporelle pour le signal 1D, dans le cas des images c'est l'information spatiale qui disparaît. Or, cette information spatiale est à la base de l'idée même de détection. En d'autres termes, la décomposition de l'image nous permet de savoir qu'il y a des amas mais pas de connaître leur position.

4.1.2 . Une solution : la transformée en ondelettes

On cherche alors une transformation capable de décomposer un signal à la fois en composantes fréquentielles et temporelles.

La transformée de Fourier locale ou à fenêtre glissante

L'une des solutions envisagées pour retrouver l'information temporelle en même temps que l'information fréquentielle fut l'introduction de la transformée de Fourier glissante, transformée de Fourier à court terme ou Transformée de Gabor en 1946. La transformée de Fourier locale (TFL) divise le signal en segments temporels et calcule la transformée de Fourier de chaque segment en appliquant la fenêtre glissante, w . La TFL d'une fonction x est donnée par l'expression :

$$TFL(x(t)) \equiv X(t, f) = \int_{-\infty}^{\infty} x(\tau) w(t - \tau) e^{-j2\pi f\tau} d\tau. \quad (4.5)$$

Cette décomposition permet d'observer comment les composantes fréquentielles évoluent dans le temps, offrant ainsi une représentation temps-fréquence d'un signal. Mais cette approche présente un certain nombre de limites telles que l'hypothèse d'un signal est stationnaire dans les fenêtres, les effets de fenêtres. De plus, lorsque la fenêtre glissante chevauche les bords des segments temporels, cela peut entraîner des artéfacts dans le résultat, car les valeurs du signal peuvent ne pas être continues à travers les segments. Enfin, la limite la plus contraignante est que les résolutions en temps et en fréquence sont inversement proportionnelles. Une fenêtre large offre une meilleure résolution en fréquence mais une mauvaise résolution en temps, tandis qu'une fenêtre étroite offre une meilleure résolution en temps mais une mauvaise résolution en fréquence.

C'est le travail combiné de plusieurs mathématiciens qui va permettre de résoudre ce problème. En 1909, le mathématicien Alfred Haar construit une base de fonctions continues par morceaux, chacune est une suite d'une impulsion courte positive puis d'une impulsion courte négative (voir Fig. 4.2, panneau supérieur gauche). Ces fonctions sont considérées comme les premières fonctions ondelettes. Initialement, ces fonctions ont été définies pour former un système orthonormal dans l'espace des fonctions de carré intégrables et non pour décomposer un signal. Notons que ces fonctions vont attirer l'attention en traitement du signal car elles ont l'avantage de ne pas être continues ce qui permet de repérer plus facilement les variations brusques dans une image ou un signal.

L'analyse par les ondelettes fait ses débuts en 1981 avec le travail de Jean Morlet et ses ondelettes à forme constante à qui il donnera son nom. Dans sa collaboration avec le mathématicien Alex Grossmann, il introduit pour la première fois le mot "ondelette". Les ondelettes sont alors des fonctions dont la forme est similaire à une vaguelette qui débute à 0, oscille et revient à 0 (voir Fig. 4.2, panneau supérieur droit). On se rapproche de la définition que nous avons aujourd'hui, à savoir des fonctions de carré sommable à moyenne nulle et qui oscillent localement. En 1985, le mathématicien Yves Meyer définit pour la première fois une base orthogonale de fonctions ondelettes (voir Fig. 4.2, panneau inférieur gauche). Par cette découverte, Yves Meyer a permis la mise en place de la transformée par ondelettes dans la continuité de la transformée en Fourier, ce qui lui vaudra le prestigieux prix Abel en 2017. En 1986, son ancien élève Stéphane Mallat reprend le formalisme des ondelettes pour les appliquer dans le cas de l'analyse multi-résolution et la discrétise. Enfin, Ingrid Daubechies conçoit des ondelettes à support compact qui peuvent alors être implémentées comme des filtres digitaux (voir Fig. 4.2, panneau inférieur droit). L'aboutissement final de cette collaboration est la mise en place d'une transformée en ondelettes (basée sur la transformation de Fourier) mais capable de décomposer simultanément le signal en composantes fréquentielles et temporelles. Alors que la transformée de Fourier a comme base de fonctions des sinusoides, la transformée en ondelettes utilise plutôt des ondelettes.

Les fonctions ondelettes sont aujourd'hui utilisées dans de nombreux domaines : dans les télécommunications à travers le stockage numérique ou le format jpeg2000 ; dans l'imagerie médicale, la détection des ondes gravitationnelles ou l'étude des grandes structures de l'Univers et même dans les films d'animations. Cette profusion d'applications s'explique par l'efficacité écrasante des ondelettes pour la compression du signal, son analyse, sa transmission et surtout sa différentiation par rapport

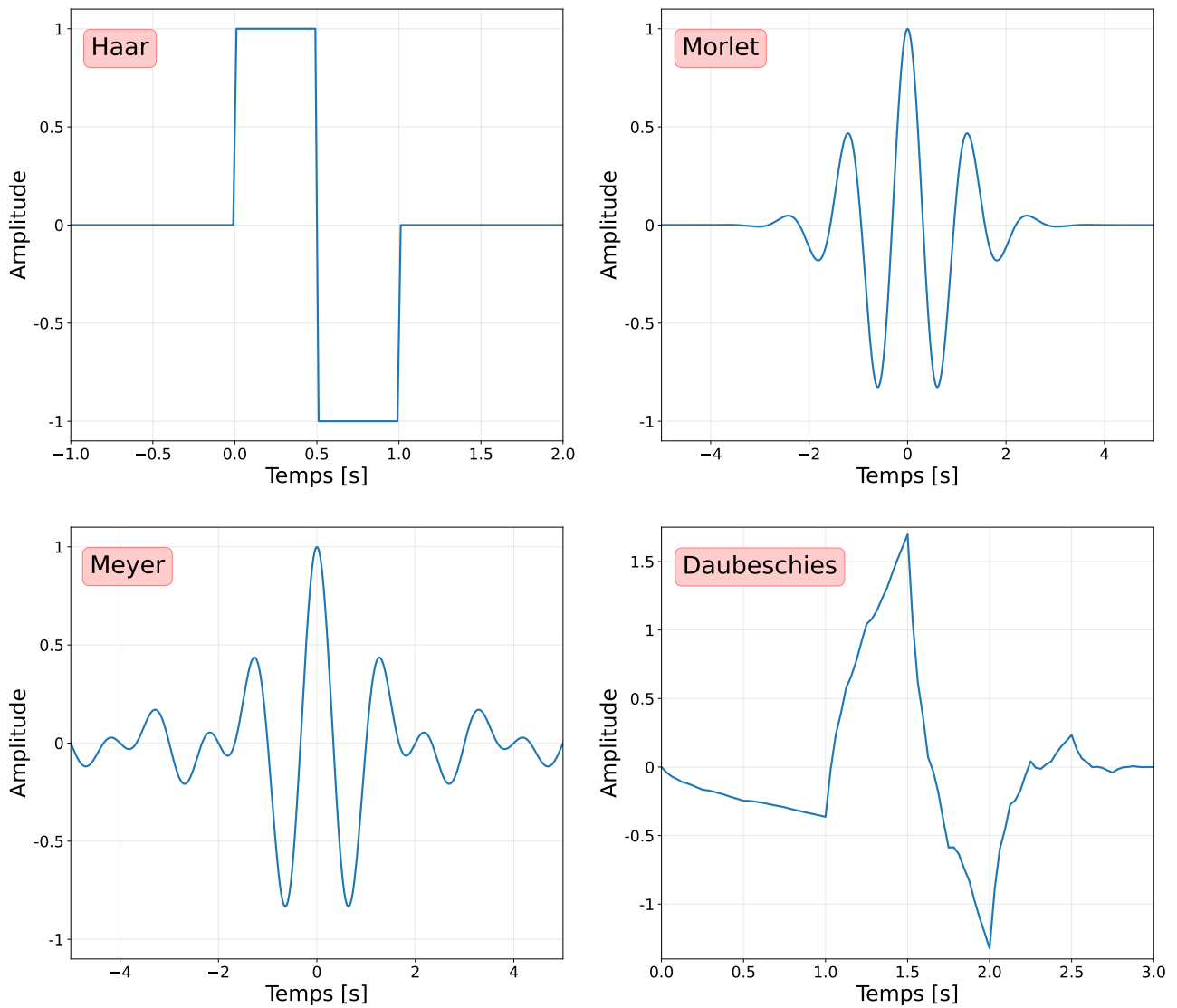


Figure 4.2 – Représentation dans l'espace réel des familles de fonctions ondelettes. Tous les filtres sont normalisés pour avoir une amplitude maximale égale à 1.

au bruit.

4.2 La méthode de détection multi-échelle

La transformée en ondelettes permet de décomposer un signal 1D ou 2D en échelles dyadiques complémentaires. Dans cette section, nous introduisons comment les ondelettes sont formalisées et appliquées à la détection. En particulier, nous détaillons en sous-section 4.2.3, l'algorithme de détection dans le cas de l'analyse multi-échelle pour la détection des amas par effet de lentille gravitationnelle faible.

4.2.1 . Le formalisme des ondelettes...

La transformée en ondelettes a été développée dans le but de décomposer un signal ou une image en différentes échelles complémentaires. La notion d'échelle, pour les ondelettes, fait référence à leur dimensionnement. Nous nous attardons ici sur cette notion importante pour la compréhension de l'analyse des résultats. Une ondelette dite mère est compressée ou dilatée pour former des copies de tailles plus petites ou plus grandes. Ainsi, les ondelettes sont des copies dont seule la taille diffère

et le paramètre d'échelle caractérise mathématiquement le coefficient de dilatation ou compression. Dans la pratique, l'échelle peut être assimilée à l'étalement ou à la largeur de la fonction ondelette. Ce faisant, les ondelettes s'adaptent à la taille des structures ciblées et les fréquences spatiales associées. Par exemple, les grandes échelles correspondent aux structures plus étendues donc aux variations lentes dans l'image caractérisées par les basses fréquences et à l'inverse, les échelles très fines ciblent les variations brusques.

Le formalisme de la transformée en ondelettes peut également être utilisé pour reconstruire des cartes de la distribution de matière projetée à une échelle donnée (voir par exemple [Leonard et al., 2012](#); [Pires et al., 2012](#)). Dans [Leonard et al. \(2012\)](#), les auteurs ont montré que les fonctions de filtrage ondelettes à une échelle donnée sont formellement identiques aux fonctions de filtrage MO à cette même échelle. Comme pour les fonctions MO, il existe de nombreuses fonctions d'ondelettes différentes, telles que Starlet, Mexican hat, Morlet, Biorthogonal, etc. (voir e.g. [Starck et al., 1998](#); [Starck et al., 2006](#)). Parmi elles, certaines fonctions ondelettes, comme la starlet, ont été conçues pour être locales dans les espaces réel et de Fourier, en d'autres termes correspondre à un filtre fini et défini sur un segment. Cette propriété permet aux ondelettes d'être des filtres passe-bandes par définition ce qui leur confère un avantage certain par rapport à l'approche MO.

Dans la pratique, la décomposition en ondelettes est appliquée sur les cartes de convergence qui sont directement liées à la distribution de matière projetée. Une carte ondelette sur la convergence κ à la position θ et à l'échelle a peut être calculée par :

$$W_a(\theta) = \frac{1}{\sqrt{a}} \int \kappa(t) \Psi_a(t - \theta) dt, \quad (4.6)$$

où Ψ_a est le filtre ondelette d'échelle a . Tous les filtres ondelettes sont définis de manière à respecter la condition de moyenne nulle. Ils sont donc compensés par définition.

Un autre avantage du formalisme des ondelettes est qu'il permet de décomposer une image donnée en plusieurs composantes d'échelle complémentaires. Ceci est possible grâce à la transformée en ondelettes, pour laquelle il existe des algorithmes rapides. Pour notre étude, nous utilisons la transformée en ondelettes isotrope non décimée, également appelée transformée en starlets (e.g. [Starck et al., 1998](#); [Starck et al., 2006](#)), qui est capable de calculer simultanément plusieurs cartes d'ondelettes sur des échelles dyadiques. Les fonctions d'ondelettes sont isotropes, ce qui les rend bien adaptées à l'extraction des amas de galaxies, qui apparaissent grossièrement circulaires dans les cartes de convergence. Comme ces fonctions décomposent une image en plusieurs carte d'échelles complémentaires (voir Sect. 4.3), elles sont idéales pour entreprendre une analyse multi-échelle et pour cibler des structures de différentes tailles angulaires sur la même carte.

La transformée en ondelettes (starlets) décompose la convergence de sorte que :

$$\kappa(\theta) = C_J(\theta) + \sum_{i=1}^J W_i(\theta) \quad (4.7)$$

où J est le nombre d'échelles dans la décomposition, C_J est la version lissée correspondante de la convergence κ , et W_i sont les cartes ondelettes ciblant les amas de galaxies avec un rayon angulaire apparent de 2^{i-1} fois la taille du pixel.

4.2.2 au service de la détection des amas

La nouvelle méthode de détection que nous présentons dans ce chapitre tire partie de cette décomposition multi-échelle, la transformée en ondelettes étant capable de décomposer rapidement la convergence en différentes échelles complémentaires deux à deux. Dans le cas de la transformée starlet que nous utilisons ici, les fonctions ondelettes sont définies comme la différence entre deux fonctions B^3 -spline à différentes résolutions.

Les filtres ondelettes à l'échelle i , utilisés dans la transformée en starlets, peuvent s'écrire dans le formalisme MO de la manière suivante (voir [Leonard et al., 2012](#)) :

$$U_{W_i}(u) = \frac{1}{9} \left[93 |u|^3 - 64 \left(\left| \frac{1}{2} - u \right|^3 + \left(\frac{1}{2} + u \right)^3 \right) + 18 (|1 - u|^3 + |1 + u|^3) - \frac{1}{2} (|2 - u|^3 + |2 + u|^3) \right]. \quad (4.8)$$

pour lequel

$$u = \frac{\theta}{2^i x}, \quad (4.9)$$

où θ est le paramètre de la fonction et x est la taille du pixel. Par la suite, cette expression nous permet de comparer simplement toutes les fonctions de filtrage dans l'espace réel (voir Sect. 4.3).

Dans la méthode de détection basée sur la transformée en ondelettes, nous choisissons de décomposer le signal lentille en $J = 5$ échelles (voir l'équation 4.7), afin d'englober un maximum de tailles angulaires apparentes possibles pour les amas de galaxies. Au-delà de $J = 5$, les échelles sont trop larges par rapport à la taille des amas de galaxies.

La Figure 4.3 montre le résultat d'une telle décomposition appliquée à une carte de convergence bruitée d'un champ simulé de $5^\circ \times 5^\circ$. Le panneau supérieur gauche montre la version bruitée de la carte de convergence du mode E présentée sur le panneau central de la Fig. 3.7, et les autres panneaux montrent les cartes ondelettes correspondant aux échelles $i = 1$ à 5 (W_1 - W_5). Chaque carte ondelette donne des détails sur la carte de convergence originale à différentes échelles. Notons que la première échelle (W_1) est principalement dominée par le bruit. En revanche, les échelles de W_2 à W_5 mettent en lumière différents signaux dans la carte de convergence du mode E d'origine. Dans notre analyse, nous décidons de ne conserver que les échelles W_2 , W_3 et W_4 . Nous supprimons l'échelle W_1 parce qu'elle est dominée par le bruit, et l'échelle W_5 parce que le rayon des signaux sur cette carte dépasse le rayon caractéristique projeté des amas, θ_{200c} , que nous voulons détecter (voir Fig. 1.2).

Par l'application de cette transformée en ondelettes sur la carte de convergence, nous définissons la première étape de notre algorithme de détection multi-échelle. Les cartes de convergence sont ainsi décomposées par les ondelettes et nous conservons les échelles W_2 , W_3 et W_4 pour la détection des amas de galaxies.

4.2.3 . La procédure multi-échelle

Dans cette sous-section, nous détaillons l'algorithme de détection multi-échelle (ci-après W_{234}) reposant sur la décomposition par ondelettes.

Comme nous l'avons précisé, la décomposition par ondelettes s'applique sur la carte de convergence. A partir du catalogue de galaxies, nous reconstruisons les cartes de convergence en utilisant la procédure décrite dans la section 3.1. Après décomposition des cartes de convergence en mode E et mode B par les fonctions ondelettes, nous conservons les échelles d'intérêts W_2 , W_3 et W_4 . Chaque échelle présente des spécificités uniques en termes de niveaux de bruit, de seuils, de taille du filtre et de précision. C'est pourquoi nous avons choisi de considérer les différentes échelles individuellement pour le seuillage et la détection. Ainsi, sur les échelles conservées, nous appliquons le seuillage et l'identification décrits dans la procédure de détection à échelle unique (voir Sect. 3.3). Nous extrayons les maxima locaux sur chacune des cartes de convergence en mode E filtrées et seuillées pour chaque échelle individuelle W_2 , W_3 , et W_4 .

Ce choix nous permet d'avoir un contrôle complet des détections à chaque échelle mais nous oblige à traiter un problème spécifique à l'approche multi-échelle. Un signal provenant d'un objet peut être détecté à plusieurs échelles, et avec une position qui peut varier légèrement d'une échelle à l'autre. C'est ce que nous appelons les détections communes. Une fois que nous avons obtenu une liste de pics candidats pour chaque échelle, il est important d'identifier ces détections communes et

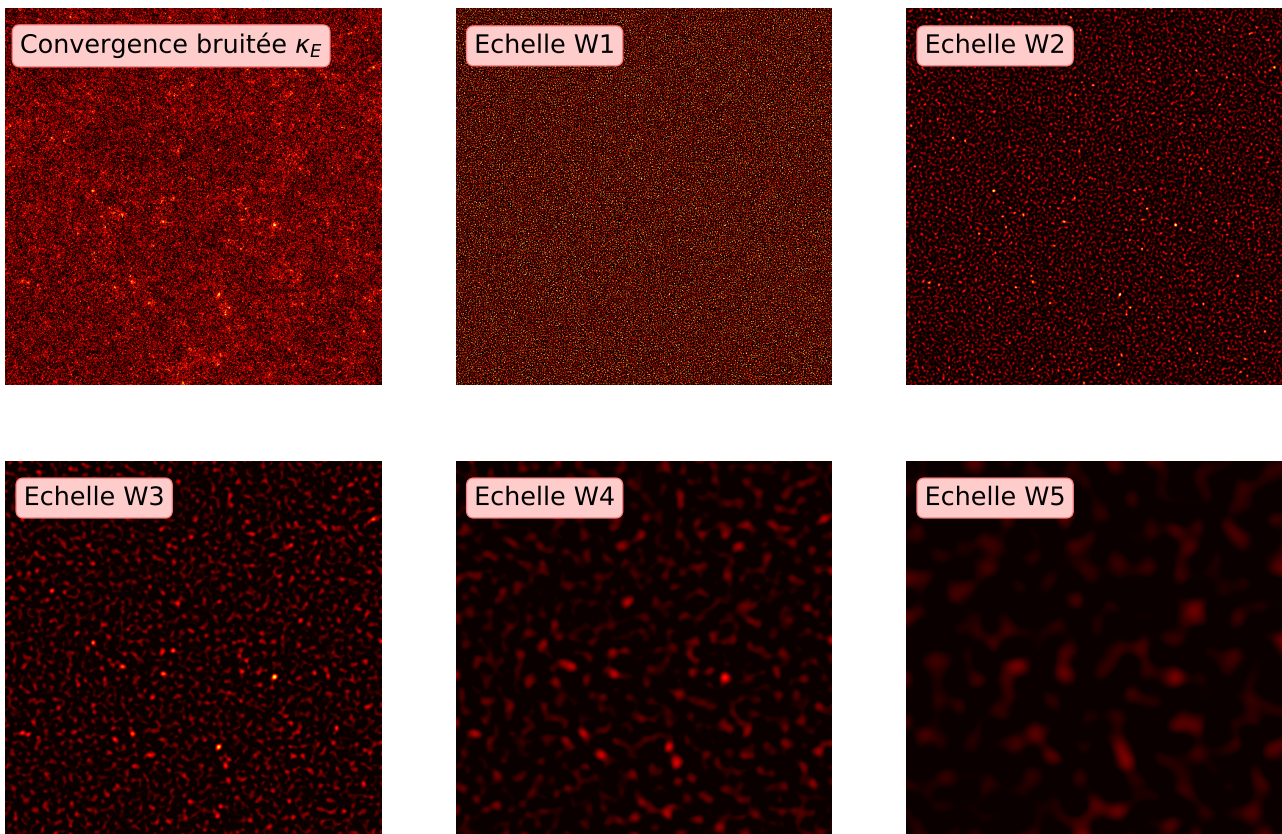


Figure 4.3 – Exemple de décomposition par la transformée en ondelettes sur une carte de convergence bruitée et simulée d'un champ de $5^\circ \times 5^\circ$.

de les regrouper. Pour ce faire, nous calculons la distance de séparation entre toutes les paires de détections sur des échelles consécutives. Si deux pics détectés sur des échelles consécutives ont une distance de séparation inférieure au plus grand des deux rayons du filtre, nous considérons qu'il s'agit d'une détection commune à ces deux échelles. Ces détections sont alors regroupées et la position du pic recombinaé est définie par la position de la détection à l'échelle la plus fine. S'il y a plus de deux pics, nous répétons cette procédure pour tous les pics de détection et toutes les échelles consécutivement.

Nous gardons trace de l'indice i de l'échelle la plus fine où la détection recombinaée apparaît, car il est considéré comme l'échelle à laquelle le signal a été détecté pour la procédure d'association décrite dans la section 5.3. Nous sauvegardons également les détections à chaque échelle ondelette individuelle, afin d'étudier leur complémentarité avant la combinaison, et de permettre la comparaison de leurs performances individuelles par rapport aux filtres MO correspondants.

En résumé, nous avons présenté ici la production du catalogue de détections dans le cas de l'approche multi-échelle W234. Les échelles d'intérêt, W2, W3, et W4, sont considérées individuellement pour tenir compte de leur spécificité. Puis les détections communes entre les différentes échelles sont regroupées pour obtenir un catalogue de détections par la méthode multi-échelle pour chaque carte de convergence.

4.3 Avantages des ondelettes comparés aux autres filtres

Maintenant que nous avons détaillé la procédure de détection multi-échelle, nous la comparons avec la procédure à une échelle donnée, décrite dans le chapitre précédent. Par construction, les deux approches partagent les mêmes données d'entrée, la même méthodologie pour le seuillage et la même approche pour l'identification. La comparaison met ainsi en lumière principalement le comportement des filtres et leur impact sur le signal. Nous comparons tout d'abord les filtres dans les

espaces réel et de Fourier. Puis, nous regardons les caractéristiques des filtres en termes de détections pour ceux ciblant des échelles similaires.

4.3.1 . Comparaison avec les filtres MO

Les performances de chaque méthode de détection sont étroitement liées à la capacité des fonctions de filtrage à réduire à la fois le bruit et la contribution des LSS, tout en minimisant simultanément la perte de signal. Dans ce qui suit, nous analysons les propriétés des filtres ondelettes et des filtres décrits dans le chapitre précédent à la fois dans l'espace réel et dans l'espace de Fourier afin de faciliter l'interprétation des résultats présentés dans la section 6.1.

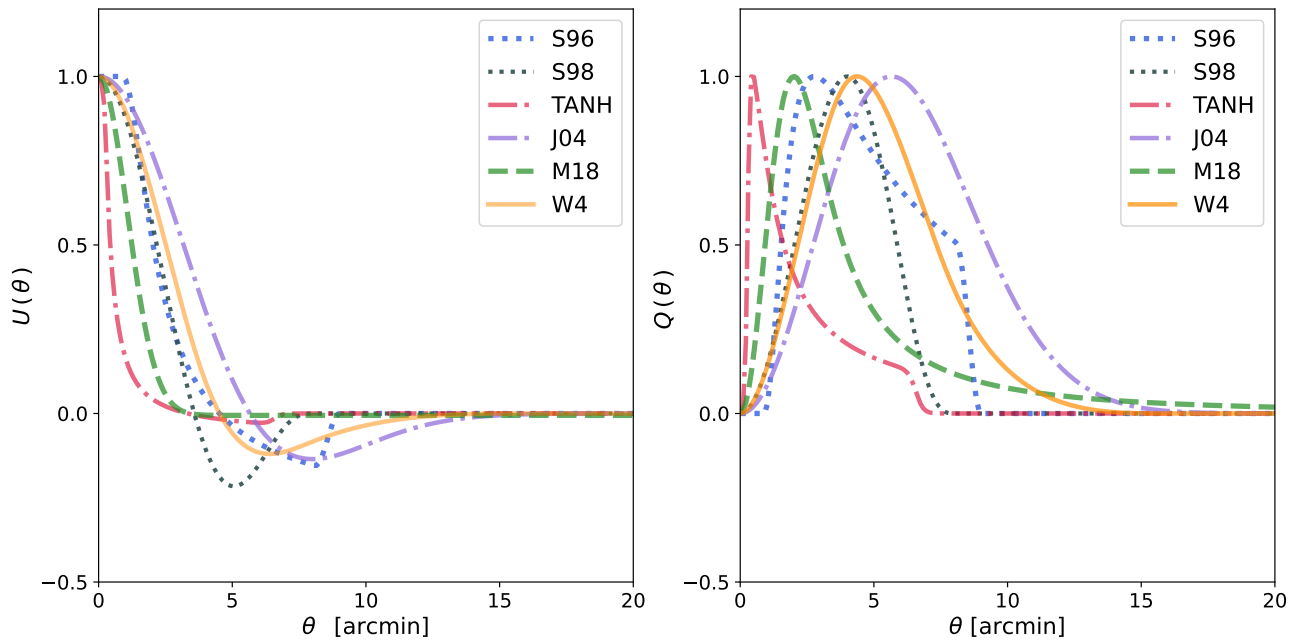


Figure 4.4 – Comparaison des différents filtres dans l'espace réel. Cette figure est similaire à la Fig. 3.9 à laquelle s'ajoutent les filtres U et Q correspondant à l'échelle W_4 (lignes orange) des ondelettes (voir notamment Eq. 4.8).

Dans la figure 4.4, nous comparons les différentes fonctions de filtrage Q et U , définies dans les sections 3.2.3 et 4.2.1, dans l'espace réel. Cette comparaison met en évidence les différences de forme des filtres MO (courbes bleues, bleu clair, rouges et violettes), dont l'analyse est présentée en section 3.2.3 et les filtres MO équivalents aux ondelettes. Les courbes orange correspondent à la fonction de filtrage par ondelettes U_{W_4} définie dans l'équation 4.8 pour $i = 4$ (ci-après W_4). Le filtre ondelette W_4 est compensé dans un rayon d'ouverture égal à deux fois le rayon du filtre. Cette comparaison met en avant les similarités de forme des filtres S98, J04 et W_4 . Notons que par la suite, nous appelons par extension, filtre ondelette W_i le filtre qui correspond à l'échelle W_i afin d'en faciliter la compréhension.

Il est également instructif d'étudier la représentation des fonctions de filtrage dans l'espace de Fourier, car cela nous permet de comprendre pleinement comment les filtres sont conçus pour localiser les structures. Dans l'espace de Fourier, un filtre passe-bande optimal doit être local, afin de ne sélectionner qu'une gamme spécifique de fréquences autour de l'échelle d'intérêt et de réduire ainsi le bruit et la contribution des LSS. La conception des filtres MO peut être délicate, et certains de ces filtres souffrent d'oscillations dans l'espace de Fourier en raison des troncatures appliquées dans l'espace réel qui détériorent la capacité passe-bande du filtre (voir par exemple [Leonard et al., 2012](#), pour plus de détails). En revanche, de nombreuses fonctions ondelettes ont été spécifiquement conçues pour avoir un support compact à la fois dans l'espace réel et l'espace de Fourier.

Pour analyser précisément les filtres dans l'espace de Fourier, nous avons étudié leur réponse

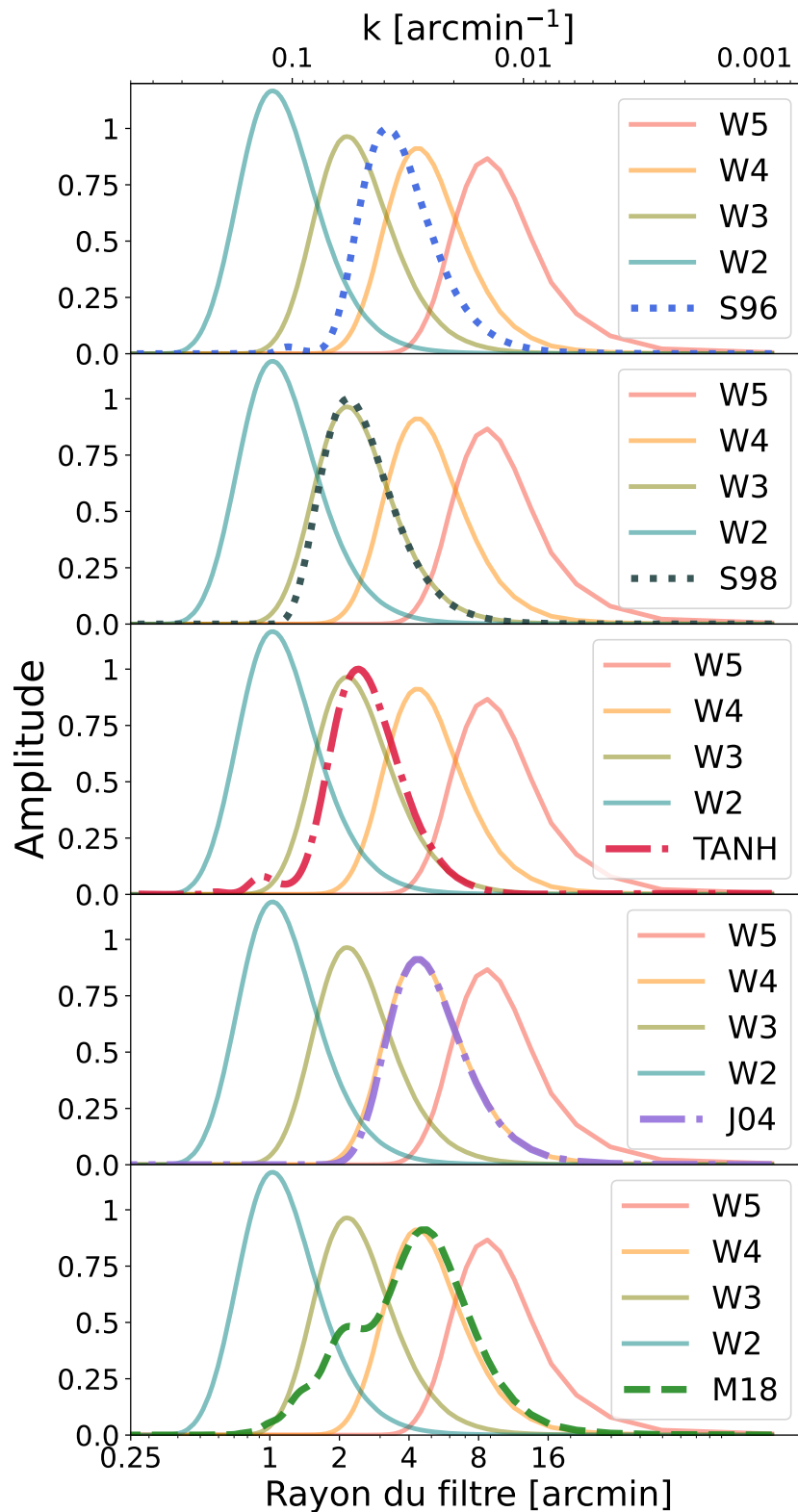


Figure 4.5 – Comparaison dans l’espace de Fourier de plusieurs filtres MO et filtres ondelettes, exprimée en fonction du nombre d’onde k en arcmin^{-1} et du rayon du filtre en arcmin. Les lignes bleue et grise en pointillés (panneaux supérieurs) représentent les filtres S96 et S98, les lignes rouges et violettes en pointillés (panneaux du milieu) représentent les filtres TANH et J04, et la ligne verte en pointillés (panneau inférieur) correspond au filtre M18. Les lignes continues bleue, verte, orange et rouge clair correspondent aux filtres à ondelettes W2, W3, W4 et W5 avec des rayons de 1:17, 2:32, 4:68, 9:37 respectivement. À des fins de comparaison, les filtres J04 et M18 sont forcés à la même amplitude que la fonction W4 pour illustrer leur comportement en termes de rayon de filtre.

impulsionnelle. Pour ce faire, nous avons simulé une carte de convergence nulle avec un seul pic au centre. Nous avons ensuite convolué cette carte de convergence par les quatre filtres MO et chacune des fonctions ondelettes, et nous avons calculé les spectres de puissance correspondants dans l'espace de Fourier. Les spectres de puissance obtenus sont comparés sur la figure 4.5. Cette représentation met en évidence le rayon effectif de chaque filtre, correspondant à la taille caractéristique des structures ciblées par le filtre. Dans les quatre panneaux, les lignes pleines bleue, verte, orange et rouge clair correspondent aux filtres ondelettes avec des rayons de 1', 1.7', 2', 3.2', 4', 6.8', 9', 37' respectivement. La ligne bleue en pointillés représente le filtre S96 correspondant à un rayon de filtre d'environ 3', qui se situe entre les filtres ondelettes W3 et W4. La ligne rouge en pointillés trace le filtre TANH, avec un rayon d'environ 2', qui est proche du filtre ondelettes W3. La ligne violette en pointillés représente le filtre J04, dont le rayon est d'environ 4', ce qui se rapproche du filtre ondelettes W4. Dans le panneau inférieur, la ligne verte en pointillés trace le filtre M18, qui peut être comparé à une combinaison des filtres ondelettes W3 et W4. Dans le domaine de Fourier, nous pouvons clairement voir le comportement passe-bande des différentes fonctions. On remarque que les filtres S96 et TANH présentent un comportement oscillatoire faible, qui peut s'expliquer par leur conception mais qui n'est pas optimale. Ces oscillations sont présentes à hautes fréquences spatiales $k \sim 0.1$ et par conséquent intègrent davantage de bruit dans la carte filtrée.

Cette représentation dans l'espace de Fourier montre en outre que les filtres ondelettes semblent bien définies et ne présentent aucune oscillation. De plus, il transparaît que les fonctions ondelettes sont bien dyadiques et complémentaires.

Leur comparaison avec les quatre filtres MO montre qu'un filtre ondelette à une échelle donnée peut être directement comparé à un filtre MO. Cependant, contrairement aux filtres MO, qui se concentrent essentiellement sur une échelle, le formalisme des ondelettes nous permet d'extraire simultanément le signal complémentaire à chaque échelle. En général, les méthodes MO standards ciblent les amas situés à un redshift relativement faible avec une grande taille angulaire apparente. Ainsi, les rayons typiques des filtres MO sont compris entre 1' et 5' (e.g. [Hamana et al., 2012](#)) afin de maximiser le rapport signal/bruit pour les amas ayant une étendue angulaire similaire. En revanche, le signal de lentille, provenant d'amas ayant des tailles angulaires apparentes différentes, peut facilement être ciblé en utilisant le formalisme des ondelettes.

A travers cette comparaison, nous voyons que les ondelettes sont optimales pour la détection simultanée des amas de galaxies à différentes tailles angulaires. Dans l'espace réel, l'analyse des filtres est complexe mais elle met tout de même en avant une grande similarité entre les filtres S96 et TANH, et S98, J04 et W4. En complément, nous avons approfondi notre analyse par une étude des filtres MO comparés aux ondelettes dans le domaine de Fourier. Là où la sélectivité des filtres MO est limitée par leur largeur dans l'espace de Fourier, la complémentarité des ondelettes permet une sélectivité maximale. À l'inverse de certains filtres MO, les ondelettes ne présentent pas de caractère oscillatoire qui pourrait affecter le calcul du seuil et la sélectivité des filtres. On observe également que le filtre ondelette W3 englobe les résultats de S98, et W4 semble reproduire J04. Nous avons montré ainsi que les ondelettes traitées individuellement sont déjà de très bons filtres passe-bandes.

4.3.2 . Comparaison en termes de détections

La représentation dans l'espace de Fourier (voir Fig. 4.5) des différents filtres MO face aux filtres ondelettes met en lumière des similarités en termes de rayon effectif et de sélectivité. En particulier, les paires de filtres S98 et W3, TANH et W3, W4 et J04, et M18 et W3-W4 partagent en majorité leur intervalle de fréquences spatiales. Ces similarités semblent indiquer qu'un seul filtre ondelette peut être aussi performant qu'un filtre MO originellement conçu pour la détection des amas. Dans cette sous-section, nous comparons les pics identifiés par la procédure de détection sur des cartes convoluées par des filtres différents mais similaires.

Dans cette étude, on s'intéresse aux cartes de convergence filtrées, seuillées et les détections

obtenues pour les 4 paires définies plus haut. Le nombre de détections dépend de 4 paramètres : (i) la largeur du filtre dans l'espace de Fourier, (ii) l'échelle ciblée par le filtre (pic dans la Fig. 4.5), (iii) le seuil de détection et (iv) la taille des structures dans la carte de convergence.

Le point (i) s'explique par la dépendance directe entre le caractère sélectif du filtre et l'intervalle de fréquences spatiales que couvre sa représentation dans l'espace de Fourier. Similairement, le point (ii) provient de la capacité du filtre à extraire des structures d'une taille spécifique, ce qui est directement lié à leur présence dans le champ étudié (iv). Finalement, le point (iii) est lié à la capacité du filtre à réduire le niveau de bruit tout en isolant le signal des structures. En fixant un seuil de détection identique à chaque paire, et en traitant les mêmes cartes on se limite ainsi à la comparaison pure des caractéristiques des filtres (i-ii) à travers leur détections.

Dans le cas des filtres MO, on applique la procédure de détection, décrite en Sect 3.3, sur les champs de $5^\circ \times 5^\circ$ de la simulation DUSTGRAIN-*pathfinder*. Dans le cas des filtres ondelettes, les mêmes cartes de convergence sont décomposées par la procédure multi-échelle définie dans la section précédente (voir Sect. 4.2.3). Mais pour la comparaison, nous traitons les cartes ondelettes issues de la décomposition de manière individuelle afin d'étudier chaque filtre à une échelle donnée. Une fois filtrées, les cartes de convergence sont seuillées en utilisant la même valeur de SNR pour chaque paire. Dans la pratique, on estime le niveau de bruit propre à chaque méthode de filtrage puis on calcule le SNR pour chaque pixel des cartes de convergence filtrées en mode E. On s'assure ainsi d'avoir la même définition pour le seuil de détection de chaque carte, tout en tenant compte des spécificités de chaque filtre pour une comparaison juste. On compare pour chaque paire les détections qui en résultent.

La Figure 4.6 représente les détections obtenues pour chacune des 4 paires de filtres présentant des similarités dans la représentation de Fourier (voir Fig. 4.5). En fond, on retrouve la carte de convergence débruitée présentée à la figure 3.7 où on peut voir en surbrillance la présence du signal. La comparaison des détections des filtres S98 avec l'ondelette W3 (panneau supérieur gauche) met en lumière que presque toutes les détections de S98 (cercles bleus) sont aussi obtenues par l'ondelette W3 (cercles dorés). En généralisant cette comparaison sur les 256 simulations, on constate que $92.7\% \pm 0.6\%$ des détections faites par S98 sont partagées par W3 lequel obtient plus de 80% de détections en plus. La comparaison des filtres (panneau supérieur droit) TANH (cercles bleus) et W3 (cercles dorée) met en lumière un comportement très similaire. Les écarts entre les détections s'expliquent grandement par la différence d'intervalles de fréquences de sélection dans l'espace de Fourier entre les deux filtres. Cependant, le filtre W3 partage $72.0\% \pm 0.9\%$ des détections du filtre TANH et en obtient 31% supplémentaires. Conformément aux observations faites dans le domaine de Fourier (voir Fig. 4.5), nous observons que $95.1\% \pm 1.0\%$ des détections de J04 (cercles bleus) se superposent avec celles obtenues par W4 (panneau inférieur gauche) pour un nombre de détections similaire. Dans le cas de la comparaison (panneau inférieur droit) avec le filtre M18 (cercles bleus), on utilise les détections combinées des filtres W3 et W4 (cercles dorées) car le filtre M18 dans son design ingénieux parvient à approcher la notion de détection multi-échelles. En effet, son comportement dans le domaine de Fourier englobe les filtres ondelettes W3 et W4. Nous pouvons ainsi voir que la somme des détections résultant de W3 et W4 parvient à reconstruire 85% des détections du filtre M18, avec environ 45% de détections en plus.

En résumé, cette comparaison souligne un avantage net des filtres ondelettes pour la détection des amas. Nous avons montré qu'un seul filtre ondelette est déjà capable d'égaliser ou dépasser les filtres MO au design simple. Par construction, les filtres ondelettes sont tout aussi adaptés à la détection des amas que les filtres MO. Leur combinaison permet d'extraire d'autant plus de structures liées à du signal. Parmi les détections des 4 panneaux, on note que certaines détections ne semblent pas visuellement reliées à un signal dans la carte de convergence non bruitée. Cette observation va de pair avec la nécessité de pousser plus loin l'analyse des résultats que le nombre pur de détections. Maintenant que nous avons accès à un catalogue de détections, il est essentiel de savoir si ces détections correspondent ou non à du signal sous-jacent.

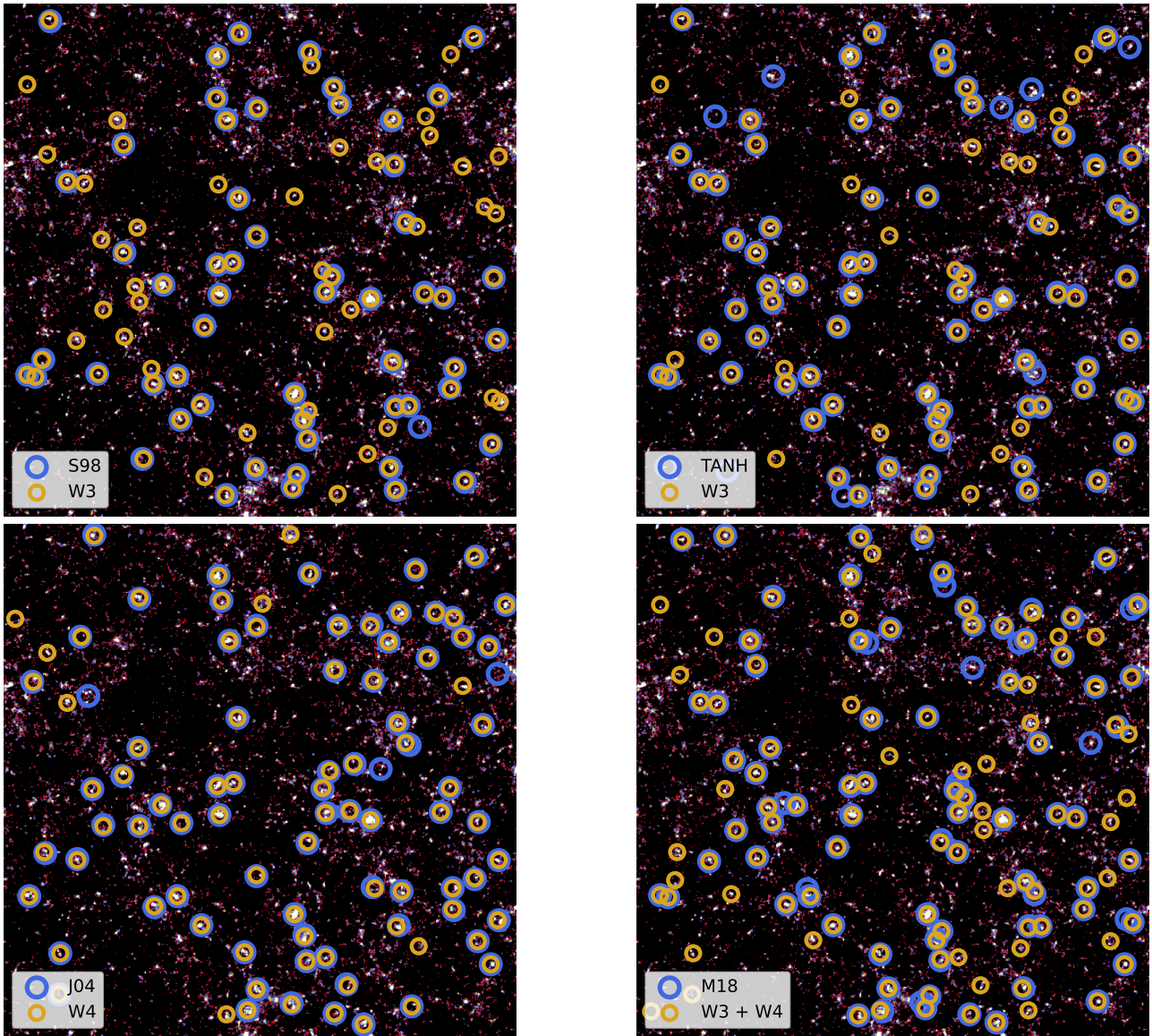


Figure 4.6 – Comparaison en termes de détections pour les différentes paires de filtres présentant des similarités dans leur représentation en Fourier. On compare ainsi les filtres S98 et W3 (panneau supérieur gauche), les filtres TANH et W3 (panneau supérieur droit), les filtres J04 et W4 (panneau inférieur gauche) et les filtres M18 et W3-W4 (panneau inférieur droit). Les détections obtenues par les filtres MO sont représentées par de larges cercles bleus et celles obtenues par les filtres ondelettes par de petits cercles dorés.

4.4 Conclusions

Dans ce chapitre, nous avons abordé les bases et les limites de la transformée de Fourier pour le traitement du signal. La transformée en ondelettes, introduite plus récemment, permet de dépasser ces limites en proposant une décomposition du signal autant fréquentielle que temporelle.

Notre approche pour la détection des amas de galaxies par effet de lentille gravitationnelle faible utilise une telle décomposition car leur résolution croissante et leur complémentarité permet de cibler des structures de tailles différentes au sein des données. Cet avantage les rend tout à fait adaptées pour la détection.

Nous avons alors détaillé la procédure suivie pour l'approche multi-échelle W234 pour la détection des amas de galaxies. Nous avons montré notamment comment les détections communes à plusieurs filtres sont regroupées en tenant compte du rayon des filtres.

Nous avons ensuite comparé les différentes échelles ondelettes avec les différentes fonctions de filtrages MO introduites dans le chapitre précédent. Nous avons montré que chaque filtre ne cible

pas la même taille de structures. De plus, certains parmi eux souffrent de légères oscillations par construction ce qui peut nuire à leur précision *in fine*.

Puis nous avons conclu cette analyse par une comparaison en termes de détections. La comparaison des détections obtenues en appliquant les méthodes sur une échelle individuelle démontre que les filtres ondelettes sont naturellement adaptés à la détection des amas. En nombre de détections, les filtres ondelettes pris individuellement égalent et même dépassent les filtres MO tout en incluant jusqu'à $92.7\% \pm 0.6\%$ de leurs détections.

Dans le chapitre suivant, nous nous intéressons à la caractérisation des performances des méthodes pour l'association des détections avec les halos fournis par le catalogue de halos de matière noire.

*
* *

Troisième partie

Troisième partie : Méthodes d'association

5

La procédure d'association

La mesure d'une erreur est en même temps la mesure de la vérité correspondante.

Louis-Claude de Saint-Martin, De l'esprit des choses.

5.1	Enjeux de la procédure d'association	84
5.1.1	Associer mais jusqu'à quel point?	84
5.1.2	Le catalogue utilisé pour l'association	86
5.2	Présélection des halos selon leur détectabilité	87
5.2.1	Estimation du SNR théorique	87
5.2.2	Application à la simulation DUSTGRAIN-pathfinder	89
5.3	Une nouvelle méthode d'association	89
5.3.1	La distance maximale d'association optimale : MMD	90
5.3.2	Notre procédure d'association adaptative	93
5.3.3	Adaptation au cas multi-échelle	93
5.4	Conclusions	95

A partir des cartes de convergence filtrées et seuillées, nous avons produit les catalogues de détections pour toutes les méthodes présentées (gaussienne, masse d'ouverture et multi-échelle). Attardons-nous maintenant à la validation des détections obtenues. Cette étape de validation, dite procédure d'association, repose sur la comparaison des positions des détections et des halos de matière noire dans le champ. La méthode d'association est déterminante pour la mesure des performances de chaque méthode. Dans ce chapitre, nous abordons comment implémenter la procédure d'association pour permettre ensuite une comparaison juste entre les méthodes de filtrage qui tient compte de leurs spécificités. En section 5.1, nous introduisons les défis que soulève la mise en place d'une méthode d'association commune, ainsi que ses limites. En section 5.2, nous présentons comment présélectionner les halos d'intérêt au sein du catalogue en étudiant la détectabilité des amas en fonction de leur masse et redshift. Dans la section 5.3, nous introduisons un nouveau critère d'association qui s'adapte aux différents filtres utilisés, et notamment à l'approche multi-échelle.

5.1 Enjeux de la procédure d'association

Dans notre étude, la comparaison des méthodes se fait essentiellement en termes de détections associées à un halo et de détections non-associées (voir chapitre 6). Les performances et résultats des méthodes sont donc conditionnés par la procédure d'association et ses paramètres. Dès lors, les résultats de chaque méthode peuvent être biaisés si l'association n'est pas effectuée correctement. Pour notre étude, nous avons identifié deux points critiques, qui peuvent impacter de manière artificielle et significative le nombre d'associations : les caractéristiques des halos de matière noire (nombre, détectabilité) et la distance d'association. Ce dernier point est particulièrement vrai lorsque l'échelle caractéristique des différents filtres se distinguent de manière significative entre les méthodes. Dans cette section, nous décrivons ces enjeux clefs relatifs à l'association.

5.1.1 . Associer mais jusqu'à quel point ?

La procédure d'association consiste à valider ou invalider les pics détectés dans les cartes de convergence filtrées et seuillées. La procédure d'association est une tâche complexe mais essentielle pour caractériser la performance des différentes méthodes. Dans cette sous-section, nous décrivons les méthodes d'association couramment utilisées et les enjeux autour de la distance d'association.

Typiquement, la procédure d'association consiste à évaluer la distance entre les positions des détections et les positions des objets dans le catalogue de halos ou d'amas à notre disposition. Dans le cas de données simulées, on peut avoir accès à un catalogue de halos qui sert de base de comparaison pour la position des détections (Hamana et al., 2004; Hennawi and Spergel, 2005; Gavazzi and Soucail, 2006). En ce qui concerne les données réelles, les détections sont comparées avec des catalogues d'amas déjà existants et partageant le même champ d'observation (Miyazaki et al., 2018; Hamana et al., 2020; Oguri et al., 2021). Dans ces deux cas, nous utilisons les positions estimées ou réelles des amas de galaxies pour l'évaluation de la distance. La paire halo-détection qui en résulte est considérée comme associée si leur distance est inférieure à une distance maximale d'association fixée (ci-après MMD, de l'anglais maximum matching distance). La problématique est alors de savoir quelle MMD choisir et comment la calculer.

Généralement, la procédure d'association est effectuée dans un rayon physique, comobile ou angulaire centré sur le pic candidat aussi bien sur des données simulées (e.g. Hamana et al., 2004; Hennawi and Spergel, 2005) que sur des données réelles (e.g. Gavazzi and Soucail, 2006; Miyazaki et al., 2018; Hamana et al., 2020; Oguri et al., 2021) à partir de catalogues d'amas pré-existants. Dans la pratique, la plupart des méthodes d'association font le choix de la distance d'association angulaire (e.g. Hamana et al., 2004; Hennawi and Spergel, 2005; Miyazaki et al., 2018; Hamana et al., 2020). On trace un cercle d'un rayon angulaire apparent donné autour du pic candidat et on l'associe au halo ou amas le plus proche, qui tombe dans le cercle. Ce rayon donné correspond à la MMD, qui est habituellement optimisée en fonction du filtre utilisé. Jusqu'à présent, sur des données réelles ou simulées, le choix de cette distance d'association était fixé à partir du rayon du filtre et de la taille des amas ou halos ciblés. Mais il n'existe pas de formule ou de relation exacte permettant de fixer cette distance. Par exemple, Hamana et al. (2004) font le choix d'une distance angulaire de 2.88 arcmin, car supérieure au rayon de 1 arcmin du filtre utilisé mais plus faible que le rayon du viriel angulaire d'un halo massif à faible redshift. Sur des données réelles, Hamana et al. (2020) vont jusqu'à 5 arcmin de tolérance pour un filtre de 1.5 arcmin de rayon, en se comparant à un catalogue réel de détection optique (CAMIRA - voir Sect. 8.4.1). Leur MMD tient également compte des imprécisions des positions du catalogue mais il est intéressant de noter que pour le même filtre et une association très similaire, Miyazaki et al. (2018) fixent cette même distance à 6 arcmin.

Pour aller plus loin dans l'association, Hamana et al. (2004) répètent la procédure du point de vue des halos du catalogue afin de pouvoir compter les détections multiples. Cette approche complémentaire a le mérite d'approfondir la caractérisation des détections obtenues, surtout si leur nombre est

égal ou grand devant le nombre de halos dans le catalogue. De plus, cette double association (halos du catalogue vers pics et pics vers halos du catalogue) permet de mettre en lumière les détections multiples d'un halo ou les associations multiples d'une détection. Ces deux cas de correspondances multiples peuvent être réduits en adaptant la méthode d'association en fonction du rayon du filtre.

Alternativement, certaines procédures proposent une association plus raffinée, en calculant la distance comobile. En traçant un cercle d'un rayon fixe (en Mpc) autour du halo, Oguri et al. (2021) estime la distance comobile entre le halo et les détections les plus proches. On associe ensuite le halo avec la détection la plus proche dans ce rayon. Désormais, le calcul du rayon est motivé autant par la taille des amas que par la taille du filtre (e.g. Miyazaki et al., 2018; Oguri et al., 2021). Cette dernière méthode est plus précise et s'adapte aux filtres, mais nécessite d'avoir accès au redshift des halos du catalogue et des détections soit par estimation soit en faisant l'hypothèse .

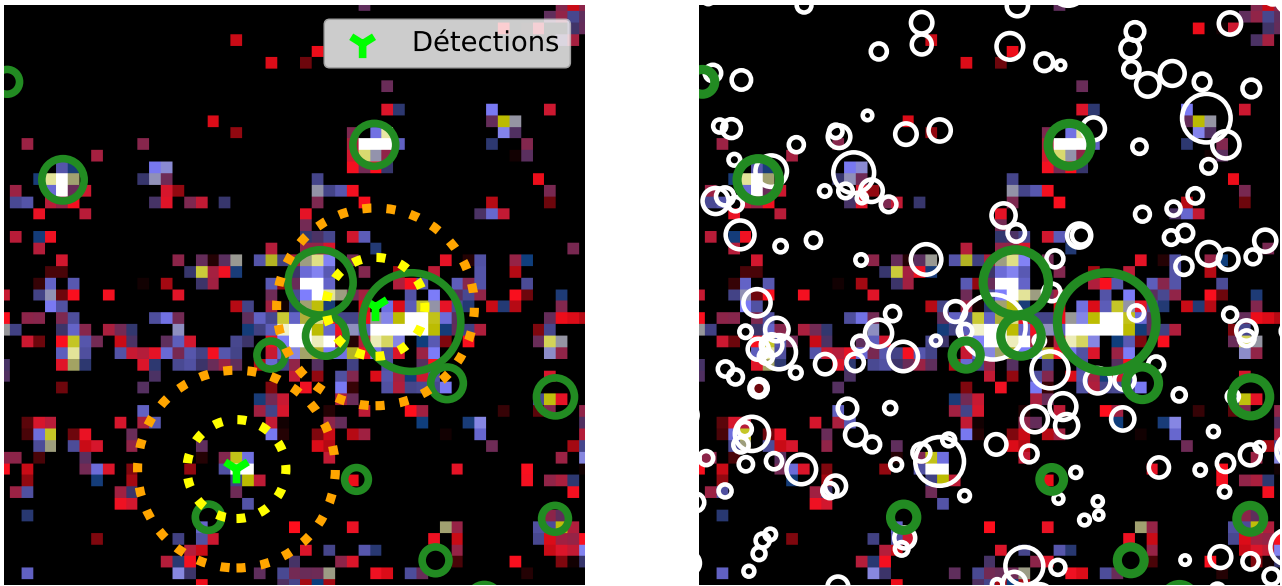


Figure 5.1 – Superposition des halos du catalogue sur une carte de convergence simulée, non bruitée. *Gauche* Sur un champ $25' \times 25'$, les cercles verts représentent les halos dont la masse $M_{200c} > 1.10^{14} h^{-1} M_{\odot}$ et leur rayon correspond directement au rayon apparent du halo. En vert clair, on superpose deux détections obtenues par toutes les méthodes sur ce patch. Les cercles jaunes et orange tracent la distance d'association maximale respectivement la plus petite utilisée ($2'$) et la plus grande ($5'$). *Droite* Les cercles blancs représentent les halos dont $M_{200c} > 2.10^{13} h^{-1} M_{\odot}$ et les cercles verts ceux dont la masse $M_{200c} > 1.10^{14} h^{-1} M_{\odot}$ sur un champ de $25' \times 25'$.

La distance d'association apparait comme un compromis entre les imprécisions sur les positions dues au filtre utilisé et le nombre de fausses associations. Toutefois, lorsque l'on compare différents filtres, leur imprécision en termes de positions des détections varient grandement. C'est pourquoi conserver une MMD commune à toutes les méthodes augmente ou réduit artificiellement le nombre de détections associées et biaise alors les mesures. Le panneau de gauche de la figure 5.1 illustre l'effet du choix de la distance d'association sur une carte de convergence simulée et non bruitée, zoomée sur un champ de $25' \times 25'$. Sur ce champ, les halos dont la masse $M_{200c} > 1.10^{14} h^{-1} M_{\odot}$ sont représentés par des cercles verts et la largeur des cercles correspond au rayon caractéristique apparent des halos, θ_{200c} . Les cercles jaunes et orange en pointillés tracent les distances d'association, respectivement la plus fine utilisée (2 arcminutes) et la plus large centrée (6 arcminutes). Ces cercles sont centrés sur deux détections (croix verte) communes à toutes les méthodes. Dans le cas de la détection au milieu à droite, nous observons que la distance d'association (cercle orange) peut aller jusqu'à inclure 4 halos du catalogue tandis que le choix de la distance d'association la plus fine ne permet pas d'associations même si elle se trouve à la limite du halo présentant le plus grand rayon apparent. Pour la seconde détection, le pic correspond à un halo dont la masse $M_{200c} < 1.10^{14} h^{-1} M_{\odot}$. La distance d'association la plus fine ne conduit alors à aucune association tandis que la distance la plus large associe la détection avec un halo lointain qui n'est pas lié au signal détecté (voir Sect. 5.1.2).

En résumé, pour ces deux détections, la distance d'association la plus fine ne parvient pas à associer avec les halos présents dans le catalogue. À l'inverse, la distance d'association la plus large entraîne de multiples associations et peut associer à des halos trop éloignés pour être responsable du signal et éviter les fausses associations. La MMD doit être adaptée à chaque filtre et méthode de détection pour permettre une comparaison équitable.

En résumé, la procédure d'association permet de valider les détections en comparant leur position avec celle des halos. Cette comparaison se fait essentiellement sur la distance qui les sépare ; si cette distance est inférieure à la distance maximale d'association, la paire est associée. Prendre une distance trop large par rapport au filtre peut entraîner une association fautive des détections compte tenu du nombre de structures tolérées sur la carte. Cette MMD dépend grandement des imprécisions des rayons du filtre et des imprécisions des positions du catalogue. Afin d'éviter les associations hasardeuses, il est essentiel d'avoir une distance maximale d'association adaptée pour chacun des filtres utilisés.

5.1.2 . Le catalogue utilisé pour l'association

Comme mentionné précédemment, le catalogue de halos ou d'amas utilisé a aussi un impact sur la procédure d'association. Dans le cas des données réelles, [Hamana et al. \(2020\)](#) mentionnent l'imprécision des positions issues du catalogue et son impact sur la distance d'association. Mais le contenu même du catalogue de halos est aussi un point majeur dont nous détaillons ici l'enjeu pour la procédure d'association.

Le catalogue peut contenir bien plus que les halos auxquels le lentillage faible est sensible. Or, la présence d'objets non détectables dans le catalogue peut engendrer à la fois des fausses et des mauvaises associations et biaiser ainsi les mesures de performances des méthodes. Dans le cas de la simulation *DUSTGRAIN-pathfinder*, le catalogue de halos de matière noire contient des objets jusqu'à une masse de $M_{200c} = 10^{12} h^{-1} M_{\odot}$ et jusqu'à un redshift de $z = 3.6$. Pour la cosmologie considérée, un catalogue de halos présentant ces caractéristiques contient environ 15 000 halos dans un champ de $5^{\circ} \times 5^{\circ}$. Devant un nombre si important de halos, tout pic détecté est à moins d'un arcminute d'un halo du catalogue. Or, il est évident que seule une petite fraction de ces 15 000 halos est détectable par l'effet de lentille faible, qui n'est sensible qu'aux halos les plus massifs ayant suffisamment de galaxies sources derrière eux.

Avec un tel catalogue brut de halos de matière noire, on a un nombre très important de fausses associations dues au surnombre des halos. Le panneau de droite de la figure 5.1 illustre cette idée sur une partie zoomée d'une des réalisations de la simulation. On superpose la position de tous les halos du catalogue sur une carte de convergence simulée, non bruitée et zoomée sur un champ de $25' \times 25'$. Les cercles blancs et verts représentent les halos issus du catalogue pour lesquels on a respectivement $M_{200c} > 2.10^{13} h^{-1} M_{\odot}$ et $M_{200c} > 1.10^{14} h^{-1} M_{\odot}$. La largeur des cercles correspond au rayon caractéristique apparent des halos, θ_{200c} . Notons que la détection présentée en bas du panneau gauche a maintenant une correspondance évidente parmi les halos de masse plus faible. Ceci permet de confirmer que la distance d'association la plus large conduit à une association hasardeuse. Par ailleurs, la détection n'est pas pour autant fautive car elle correspond bien à un halo. Cependant, en considérant tous les halos (cercles blancs et verts) du catalogue, toutes les détections se retrouvent associées. En considérant uniquement les halos de masse $M_{200c} > 1.10^{14} h^{-1} M_{\odot}$ (cercles verts), plus proche de la définition des amas de galaxies, on réduirait le nombre d'associations hasardeuses.

Pour quantifier le nombre de fausses associations hasardeuses, nous considérons des positions aléatoires, prises comme détections, et évaluons leur distance angulaire avec l'amas le plus proche. Pour un catalogue issu d'une réalisation de taille $5^{\circ} \times 5^{\circ}$, et contenant environ 15 000 halos, toute position aléatoire est associée à un halo dans un rayon apparent de 1.13 arcmin. Cette distance est inférieure au plus petit rayon des filtres utilisés mais ne garantit en rien la véracité de ces associations. Par conséquent, il est nécessaire de procéder à une présélection des halos considérés comme

déTECTABLES par effet de lentille faible au sein du catalogue de halos.

En résumé, nous avons vu que le catalogue de halos, bien que nécessaire à la procédure d'association, doit être considéré avec précaution. Il peut contenir des halos qui vont au-delà de la sensibilité de l'effet de lentille gravitationnelle faible ou de la définition même d'un halo de galaxies en termes de masse. Pour minimiser le nombre de fausses associations, il est essentiel de réduire le nombre d'halos en ne gardant que ceux qui sont détectables par l'effet de lentille gravitationnelle faible.

La procédure d'association est la clef de voûte de l'analyse des performances des différentes méthodes. Dans cette section, nous avons mis en évidence les deux piliers sur lesquels elle s'appuie : le choix de la distance d'association et le contenu du catalogue avec lequel on associe. Il convient de tenir compte de ces deux points critiques pour s'assurer que la procédure d'association est à la fois juste et fiable. Nous avons montré que le choix de la distance d'association devait être adapté au rayon de chacun des filtres pour ne pas biaiser leurs performances. En parallèle, le catalogue de halos, à la base de l'association, doit aussi être réduit aux objets d'intérêt afin de réduire les fausses associations.

5.2 Présélection des halos selon leur détectabilité

Nous avons vu que les catalogues de halos de matière noire pouvaient contenir plus d'objets que ceux potentiellement détectables par effet de lentille faible. Ce surnombre augmente le nombre de fausses associations. Afin de ne conserver que les halos dont le signal est suffisant pour être détecté, nous calculons leur rapport signal sur bruit théorique. Dans cette section, nous détaillons tout d'abord ce calcul, sous l'hypothèse d'un profil NFW (Navarro et al., 1996). Puis, nous présentons notre sélection des halos détectables dans le cas de la simulation DUSTGRAIN-*pathfinder* (voir Sect. 2.3.3).

5.2.1 . Estimation du SNR théorique

En supprimant les halos qui ont peu de chances d'être détectés par effet de lentille faible, on réduit de manière significative le nombre de fausses ou mauvaises associations. L'imposition d'une présélection sur le catalogue des halos de matière noire permet alors de palier au second défi de la procédure d'association. Nous avons défini notre présélection en dérivant une mesure de la détectabilité du halo sous la forme du SNR théorique des amas détectés par lentillage faible (e.g. Hamana et al., 2004; Hetterscheidt et al., 2005; Bergé et al., 2010; Andreon and Bergé, 2012).

Pour le calcul du SNR théorique, nous avons suivi l'approche proposée par Bergé et al. (2010). En supposant que le filtre représente parfaitement le signal κ , le rapport signal sur bruit d'un halo de masse M_{200c} et de redshift z_l peut être exprimé comme :

$$\nu = \frac{\sqrt{n_g}}{\sigma_\epsilon} \sqrt{\iint \kappa(\theta)^2 d\theta^2}, \quad (5.1)$$

où σ_ϵ est l'écart-type sur la mesure de cisaillement par composante et issu de la forme intrinsèque des galaxies sources et des erreurs de mesures, et n_g , leur densité moyenne. En supposant que la masse de l'amas suive une distribution NFW (Navarro et al., 1996), son profil de densité est donné par :

$$\rho(r) = \frac{\rho_s}{(r/r_s)(1+r/r_s)^2}, \quad (5.2)$$

où ρ_s est la densité caractéristique et $r_s = R_{200c}/c$ est le rayon "d'échelle", défini à partir de la concentration, c . La densité de masse projetée Σ le long de la ligne de visée, peut être écrite sous la forme (voir Bartelmann and Schneider, 2001) :

$$\Sigma(x) = 2 r_s \rho_s g(x), \quad (5.3)$$

où $x = r/r_s$ est un rayon sans dimension. La fonction g est définie par rapport à la distance au centre du halo et à la concentration, c . Si $c > 1$, g peut s'écrire (Wright and Brainerd, 2000) :

$$g(x) = \frac{1}{1+c} \begin{cases} -\frac{\sqrt{c^2-x^2}}{(1-x^2)(1+c)} + \frac{1}{(1-x^2)^{3/2}} \operatorname{arccosh}\left(\frac{x^2+c}{x(1+c)}\right) & (x < 1), \\ \frac{\sqrt{c^2+1}}{3} \left(1 + \frac{1}{1+c}\right) & (x = 1), \\ -\frac{\sqrt{c^2-x^2}}{(1-x^2)(1+c)} - \frac{1}{(x^2-1)^{3/2}} \operatorname{arccos}\left(\frac{x^2+c}{x(1+c)}\right) & (1 < x < c), \\ 0 & (x > c). \end{cases} \quad (5.4)$$

Le paramètre de concentration c est dérivé en utilisant le modèle semi-analytique introduit par Diemer and Joyce (2019), tel qu'il est mis en œuvre dans le package COLLOSSUS (Diemer, 2018). Le rapport signal sur bruit, ν , du halo peut alors être écrit comme (Bergé et al., 2010) :

$$\nu = \langle Z \rangle \frac{2\sqrt{2}\pi}{\sigma_\epsilon} \frac{\sqrt{n_g}}{D_d \Sigma_\infty} r_s^2 \rho_s \sqrt{\int_0^c x g(x)^2 dx}. \quad (5.5)$$

Le paramètre de pondération $\langle Z \rangle$ est utilisé pour tenir compte de l'impact de la distribution des galaxies sources (e.g. Seitz and Schneider, 1997). La distance angulaire à la lentille est décrite par D_d , et Σ_∞ est la valeur de Σ_{crit} pour une source à un redshift à l'infini.

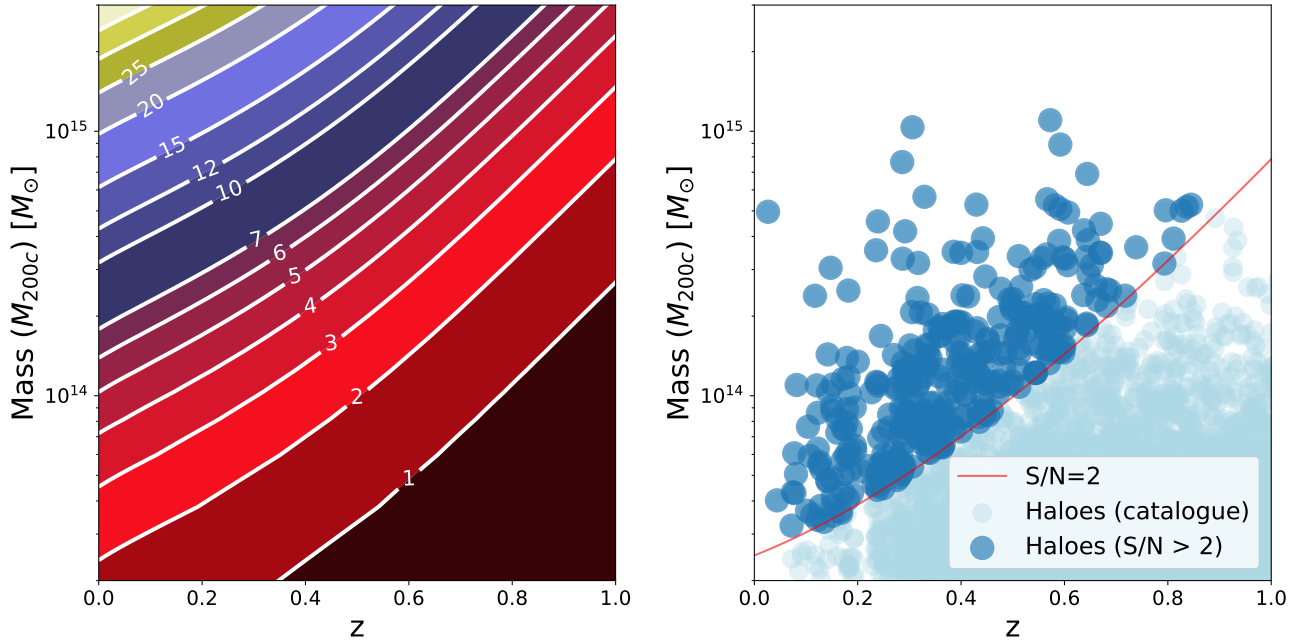


Figure 5.2 – Application d'une présélection théorique, par le SNR, au catalogue des halos. *gauche* : fonction de sélection de l'effet de lentille faible pour les amas de galaxies, dans le plan redshift-masse. La fonction de sélection a été calculée en supposant une distribution de redshifts similaire aux estimations pour *Euclid* avec $n_g = 30 \text{ gal. arcmin}^{-2}$ et $\sigma_\epsilon = 0.26$. Les contours indiquent le rapport SNR théorique pour un profil NFW. *Droite* : Exemple de présélection d'un catalogue de halos à l'aide du SNR théorique, pour un $5^\circ \times 5^\circ$. Les points bleu clair correspondent à tous les halos du champ, et les points bleu foncé sont les halos sélectionnés avec un SNR > 2.

Pour chaque halo du catalogue, nous calculons le SNR théorique, ν , en supposant une distribution de redshift des sources similaire à ce qui est attendu pour *Euclid*, p_{z_s} (dérivée de la simulation), une densité de galaxies de $n_g = 30 \text{ gal. arcmin}^{-2}$ et un écart-type sur la mesure de cisaillement par composante de $\sigma_\epsilon = 0.26$. Ce calcul reste également dépendant des caractéristiques du catalogue de halos de matière noire simulé à travers la masse et le redshift des halos. Nous montrons la fonction de sélection idéale de détection par effet de lentille faible qui résulte de ces caractéristiques, dans le panneau de gauche de la Fig. 5.2. Cette figure est obtenue en générant une grille fine de 300×300 points, dans le plan masse-redshift, et en calculant la valeur du SNR pour chaque point de la grille à l'aide de

l'équation 5.5. La colorimétrie met en évidence l'augmentation du rapport SNR vers le coin supérieur gauche, correspondant à la détection privilégiée d'amas massifs à faible redshift. Les lignes blanches continues correspondent aux contours de SNR de niveau égal à [0,1,2,3,4,5,6,7,10,12,15,20,25,30,35] qui ont été interpolés à partir des points de grille (z, M).

Notons que ceci ne donne qu'une valeur moyenne pour le rapport signal sur bruit théorique car on suppose une distribution de redshifts moyenne des sources et un filtre optimal. En pratique, les filtres ne sont jamais parfaitement adaptés à la forme des amas. En outre, le rapport signal sur bruit détecté dépend de la distribution des redshifts des sources situées derrière l'amas. Si les galaxies derrière l'amas sont peu nombreuses, de petites variations dans leur redshift peuvent artificiellement augmenter ou diminuer le signal de lentille faible observé.

Pour autant, le calcul du SNR théorique permet d'avoir une estimation moyenne de la détectabilité des halos dans le catalogue que nous avons utilisé pour réduire le nombre d'objets dans les catalogues et le nombre de fausses associations.

5.2.2 . Application à la simulation DUSTGRAIN-pathfinder

A partir du calcul du SNR théorique, on peut sélectionner les halos utilisés pour l'association. Nous nous intéressons ici à l'application de cette étape de présélection aux halos des 256 catalogues de la simulation DUSTGRAIN-*pathfinder*.

Nous avons appliqué un seuil conservateur aux catalogues de halos de matière noire en conservant tous les halos dont le SNR, $\nu > 2$. Bien qu'un halo avec $\nu = 2$ soit peu susceptible d'être détecté, nous avons choisi cette valeur pour tenir compte de la variance dans la détectabilité due à la distribution des galaxies sources et à l'utilisation d'un filtre idéal. En effet, la détectabilité théorique est estimée sous les hypothèses que le signal lentille est parfaitement imité par le filtre, et que la distribution en redshift des sources est idéale pour tout halo. Dans la pratique, ces deux hypothèses peuvent introduire un écart entre le SNR théorique et le SNR observé. La valeur de SNR, $\nu = 2$ est donc choisie comme compromis entre le nombre de halos dans le catalogue et l'impact de ces deux hypothèses sur les estimations. Ce choix permet de réduire le nombre de halos dans le catalogue, tout en conservant la majorité de ceux dont le signal est susceptible d'être détecté. À l'inverse, en dessous de $\nu = 2$, nous ne nous attendons pas à détecter beaucoup d'amas.

Dans le cas des données de la simulation, sur un champ type de $5^\circ \times 5^\circ$, seuls 2 – 3% des 15000 halos du catalogue perdurent avec cette présélection. Les 300 – 400 halos restants constituent alors la base de correspondances potentielles pour la procédure d'association. Le panneau de droite de la Fig. 5.2 montre les halos ainsi sélectionnés, dans le plan masse-redshift, pour un champ de $5^\circ \times 5^\circ$. Notons que cette présélection (cercles bleu foncé) est focalisée sur les amas massifs à moyen et faible redshift, en accord avec la sélectivité de l'effet de lentille faible. D'autre part, comparé à la définition en masse simple des amas de galaxies massifs ($M_{200c} > 10^{14} M_\odot$), notre présélection inclut des objets de masse plus faible à partir de $M = 3 \cdot 10^{13} M_\odot$ mais à faible redshift.

Grâce à cette présélection, nous avons pu résoudre le second défi de la procédure d'association afin de limiter le risque de fausses associations avec des sous-structures présentes dans le catalogue de halos. Son application aux catalogues issus de la simulation permet de drastiquement réduire le nombre de structures considérées pour l'association. L'impact de cette présélection sur les résultats est commenté plus en détail dans le chapitre suivant (voir Sect. 6.1).

5.3 Une nouvelle méthode d'association

Nous avons abordé comment présélectionner les halos au sein du catalogue pour se limiter, Attardons-nous maintenant à l'adaptation de l'association en fonction des différents filtres utilisés pour obtenir une comparaison juste. Dans cette section, nous proposons une procédure plus juste

pour l'estimation de la distance maximale d'association, MMD, inspirée de la préparation aux données Planck 2014 (Planck Collaboration et al., 2014). Puis, nous décrivons en détail la procédure d'association que nous appliquons pour chaque méthode, sur les données de la simulation et l'adaptation de cette procédure au cas multi-échelle.

5.3.1 . La distance maximale d'association optimale : MMD

Comme vu au début de ce chapitre, le choix de la MMD est crucial et doit tenir compte de la précision sur la position de la méthode de détection. Plus précisément, des filtres plus grands entraînent une perte de précision sur les coordonnées du centre du pic candidat. La MMD doit donc être optimisée et adaptée à la méthode de détection en question. Dans cette section, nous présentons notre procédure pour estimer la MMD tout en intégrant les spécificités et imprécisions de chaque filtre.

Si la MMD choisie est trop grande, de fausses associations peuvent être introduites, même après avoir présélectionné les halos du catalogue. A l'inverse, des associations peuvent être manquées si la MMD est trop petite. Pour cette étude, nous avons donc développé une méthode empirique capable d'adapter la MMD à la méthode de détection. Notre approche dérive la MMD à partir de l'incertitude sur la position de la détection, introduite par la taille du filtre. Pour la calculer, on force l'association entre les détections et le halo du catalogue le plus proche, sur l'ensemble des 256 champs. Dans la pratique, nous calculons la distance angulaire apparente, D , entre chaque pic candidat et chaque halo du catalogue de halos de matière noire, après présélection du rapport signal-sur-bruit (voir Sect. 5.2). Puis, on calcule cette même distance angulaire, D , normalisée par le rayon apparent du halo θ_{200c} . On considère chaque pic candidat associé avec le halo qui minimise la distance angulaire normalisée. Nous obtenons alors une association pour chaque pic candidat et parmi toutes ces associations, un certain nombre sont fausses. Nous cherchons alors à différencier clairement les cas où cette distance d'association correspond à une fausse association, afin de fixer notre MMD. Pour accentuer cette différenciation, il est possible de diminuer le seuil lors du filtrage afin d'augmenter artificiellement le nombre de détections, et donc, d'associations correctes et fausses.

Nous obtenons pour toutes les détections sur chaque réalisation, un halo correspondant et leurs distances angulaires absolues, D , et normalisées, D/θ_{200c} . Ces deux valeurs nous permettent de mettre en évidence la distribution des associations, sans application de MMD. La figure 5.3 montre un exemple de distribution des distances d'association correspondantes pour l'ensemble des 256 réalisations (environ 75 000 halos après présélection du catalogue) pour le filtre W_3 , pour lequel le rayon du filtre est égal à $2'32$. Chaque point bleu correspond à une association entre la détection d'un pic candidat et son halo le plus proche en suivant notre procédure d'association. Le panneau supérieur-gauche montre la distribution des distances d'association absolues D , exprimée en minutes d'arc, en fonction de la distance d'association D normalisée par le rayon angulaire du halo θ_{200c} . Nous différencions clairement deux groupes distincts d'associations. Dans le groupe supérieur-droit, les détections sont largement en dehors du rayon caractéristique du halo. Nous pouvons donc considérer ce groupe comme de fausses associations. A l'inverse, le groupe inférieur-gauche est composé d'associations pour lesquelles la position du pic candidat se situe dans le rayon θ_{200c} du halo; elles peuvent donc être considérées comme de vraies associations. Puis, afin d'isoler les associations correctes, et d'en définir une distance d'association maximale, nous utilisons une visualisation différente de ces résultats. Dans le panneau inférieur-droit de la Fig. 5.3, les points bleus représentent la décomposition en termes de RA et Dec des distances d'association, D , entre un pic candidat et son halo le plus proche (nous ne montrons que les 20×20 arcminutes centraux). La région centrale, à haute densité, correspond aux pics détectés qui ont été correctement associés à un halo sous-jacent. Dans le plan RA-Dec, la dispersion dans cette région centrale met en évidence les incertitudes dans la position de la détection du pic introduites par la taille du filtre.

Pour mettre en évidence les propriétés de la dispersion des distances d'association, nous avons projeté la distribution sur les axes RA et Dec. Ces projections sont affichées respectivement dans les

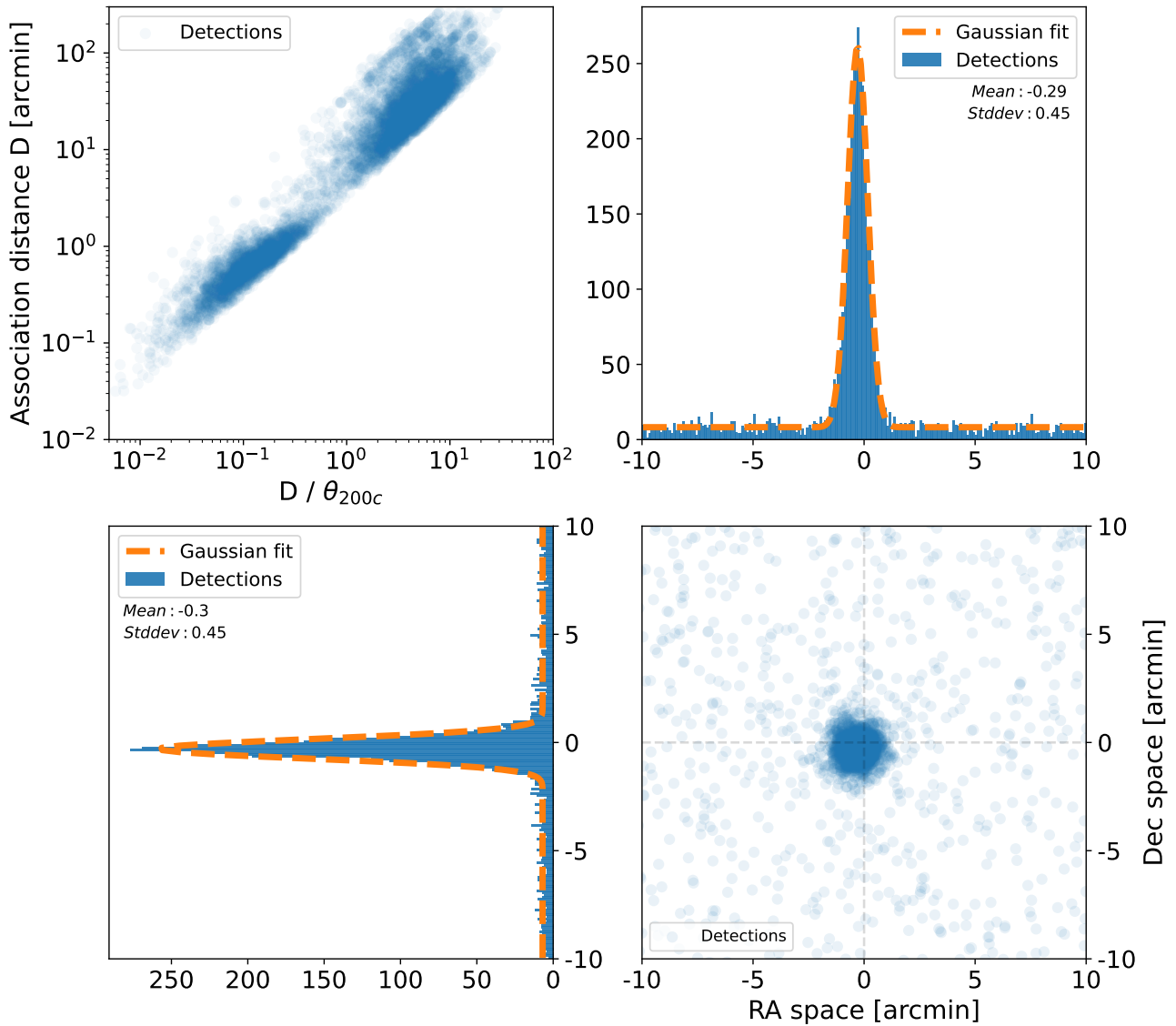


Figure 5.3 – Distributions de la distance d’association (sans distance maximale d’association) pour le filtre ondelette W3, sur les 256 réalisations. Dans tous les panneaux, les points bleus correspondent à une association entre la détection d’un pic candidat et son halo le plus proche. *En haut à gauche* : Distance d’association absolue D en fonction de la distance d’association normalisée D par rapport au rayon caractéristique du halo θ_{200c} . *En bas à droite* : Zoom de la distribution centrale de 20×20 arcminutes des distances d’association absolues exprimées en RA et Dec, pour mettre en évidence leur dispersion. Les deux autres panneaux montrent la projection de la distribution des distances d’association absolues sur les axes RA et Dec : la projection de *panneau inférieur gauche* sur l’axe Dec et la projection de *panneau supérieur droit* sur l’axe RA. Les axes RA et Dec sont zoomés autour de la zone centrale. Sur les deux panneaux, la ligne pointillée orange montre le résultat de l’interpolation d’une fonction gaussienne dont les paramètres (μ , σ) sont étiquetés ‘Mean’ et ‘Stddev’.

panneaux en haut à droite et en bas à gauche de la Fig. 5.3. Nous avons ensuite approximé l’incertitude sur la position par une distribution gaussienne; laquelle montre une très bonne correspondance avec la distribution. En pratique, nous obtenons des paramètres similaires si nous ajustons une distribution gaussienne 2D sur le plan RA-Dec ou une distribution gaussienne 1D sur chacune des composantes RA et Dec. A partir des paramètres de la distribution gaussienne obtenue, nous avons extrait l’écart-type σ , qui correspond à la dispersion de l’incertitude sur la position de la détection par rapport à la position réelle du halo. L’ajustement gaussien montre un décalage sur l’axe y qui peut être expliqué par la fraction des halos associés au hasard. L’ajustement gaussien montre également un décalage sur l’axe x correspondant à la moyenne de la distribution gaussienne μ , qui est due à la pixellisation. À partir de ces paramètres, nous avons défini la MMD comme étant $5\sigma + |\mu|$.

Table 5.1 – Estimation de la distance maximale d’association (MMD) pour les différents filtres, évaluée en utilisant la procédure décrite en Sect 5.3.1. Les rayons des filtres MO sont estimés à partir de la figure 4.5.

Filtre	σ	$ \mu $	MMD [arcmin]	Rayon des filtres [arcmin]
Gauss	0.32	0.17	1.77	1
Sg6	0.83	0.26	4.4	~ 3
Sg8	0.53	0.3	2.9	~ 2.3
TANH	0.47	0.28	2.6	~ 2.5
Jo4	0.87	0.31	4.7	~ 4.6
M18	0.60	0.28	3.3	~ 4.5
W2	0.23	0.27	1.7	1.17
W3	0.45	0.30	2.5	2.32
W4	0.89	0.29	4.7	4.68

Nous avons appliqué cette procédure à tous les filtres et les MMD obtenues pour chaque méthode de détection sont présentées dans le tableau 5.1. En comparant ces valeurs avec les rayons des filtres, obtenus pour rappel à partir de la figure 4.5, nous constatons que la MMD est directement proportionnelle au rayon du filtre, comme attendu. De plus, les valeurs obtenues font écho aux similarités déjà observées entre les filtres, notamment Jo4 et W4. Notons que cette approche a été appliquée à chacun des filtres ondelettes W2, W3, W4 individuellement pour estimer leur MMD respective et affiner la procédure d’association multi-échelle, que nous détaillons en Sect. 5.3.3.

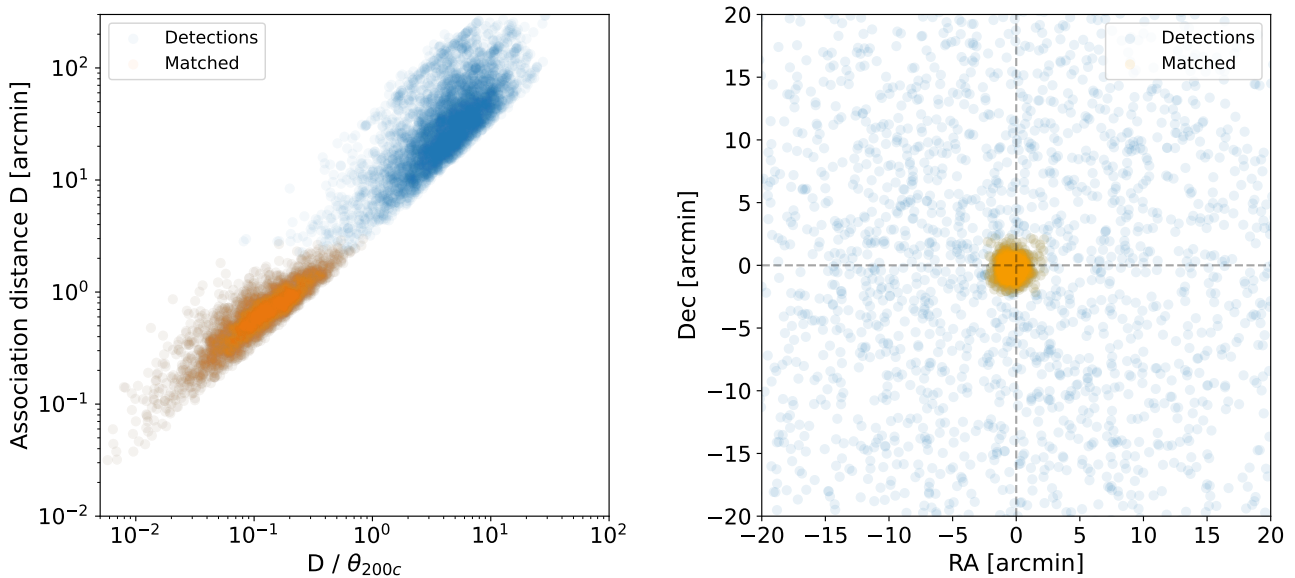


Figure 5.4 – Détections associées pour le filtre W3. Les deux panneaux sont similaires à ceux de la Fig. 5.3. Les points bleus correspondent aux associations entre un pic candidat et le halo le plus proche. Les points orange se réduisent aux mêmes associations, mais dont la distance d’association absolue est inférieure à la MMD définie pour W3.

Une fois les MMD définies, nous pouvons appliquer la procédure d’association définie dans ce chapitre, et différencier les deux groupes d’associations de la Fig. 5.3. Ainsi, la figure 5.4 montre les résultats de l’association pour le filtre W3. Les points bleus correspondent toujours aux associations de toutes les détections et de leur halo le plus proche. Les points orange correspondent aux détections dont l’association est conforme à la MMD fixée. Le fait de ne considérer que les paires pour lesquelles la distance d’association absolue, D , est inférieure à la MMD nous permet de séparer distinctement les deux groupes d’associations. Cela nous permet également de construire deux catalogues : l’un de détections associées, et l’autre, de détections non associées.

Cette procédure est appliquée pour chacune des méthodes de filtrage à une échelle donnée et chaque filtre ondelette, chaque échelle étant traitée individuellement. En résumé, notre approche pour le calcul des distances maximales d'association pour chaque filtre permet d'avoir une procédure d'association optimale en différenciant les associations correctes et incorrectes. Cette distance s'adapte d'un filtre à l'autre en tenant compte des imprécisions sur les positions des centres des détections obtenues. Nous présentons ici la procédure dans le cas des filtres à une échelle donnée; dans le cas multi-échelle, la procédure est affinée à cause des détections communes aux échelles.

5.3.2 . Notre procédure d'association adaptative

Comme nous l'avons vu en Sect. 5.1, le choix de la MMD affecte significativement la validation des détections du fait de l'incertitude sur la position due aux rayons des filtres. La grande variabilité de l'échelle des différents filtres utilisés rendent nécessaire la mise en place d'une procédure adaptative pour les comparer.

Nous avons donc implémenté une méthode adaptative, en nous inspirant des procédures d'association utilisées pour l'étude des amas en Sunyaev-Zeldovich de la mission Planck (voir par exemple [Planck Collaboration et al., 2014](#)). Dans notre étude, l'association est entreprise en calculant la distance angulaire apparente, D , entre chaque pic candidat et chaque halo du catalogue de halos de matière noire, après présélection du rapport signal-sur-bruit (voir Sect. 5.2). Cette distance angulaire, D , est ensuite normalisée par le rayon apparent du halo θ_{200c} . Cette distance normalisée, D/θ_{200c} , ainsi que la MMD, sont nos critères définitifs d'association. Ensemble, ils tiennent compte de la taille du filtre et ont pour effet de favoriser les amas les plus proches et les plus massifs lors de l'association. Cela nous permet de réduire encore davantage les associations aléatoires avec des amas de faibles redshift et masse, moins susceptibles d'être détectés par effet de lentille gravitationnelle faible.

Dans la pratique, la procédure d'association est mise en œuvre de manière itérative, comme illustré dans la Fig. 5.5. Pour N détections et M halos, nous produisons deux tableaux de taille $N \times M$ comportant respectivement les distances d'association, D , et les distances angulaires normalisées, D/θ_{200c} , entre chaque candidat et chaque halo. A partir du tableau des distances normalisées, nous commençons le processus itératif en sélectionnant la distance normalisée minimale parmi toutes les distances normalisées du tableau. La détection et le halo correspondant à cette distance normalisée minimale sont considérés comme associés. En parallèle, la ligne et la colonne correspondantes sont supprimées des tableaux. La procédure est ainsi répétée jusqu'à ce que la distance d'association absolue atteigne la MMD définie pour le filtre utilisé. A travers ces deux tableaux, nous nous assurons que l'association se fasse en tenant compte de la distance au halo; de ses caractéristiques; tout en limitant les associations hasardeuses.

Cette procédure est parfaitement adaptée à l'association des détections avec des catalogues de halos simulés et donne d'excellents résultats par rapport aux méthodes habituelles (voir Chapitre 6). Cependant, notre approche contient deux limites intrinsèques pour son succès. D'une part, pour l'application de cette procédure d'association à partir de catalogues d'amas réels (par exemple optique, X-ray,...) l'accès à une estimation du R_{200c} (ou M_{200c}) des halos est nécessaire. D'autre part, l'estimation de la MMD doit être optimale vis-à-vis de la méthode de filtrage, voire le rayon du filtre considéré.

5.3.3 . Adaptation au cas multi-échelle

Dans l'approche multi-échelle, un même halo peut être détecté à plusieurs échelles successives, ce qui complique la procédure d'association. Comme décrit en Sect. 4.2, pour évaluer les performances de l'approche multi-échelle, il est important de recombinaison les détections communes et de les associer à une seule échelle (l'échelle la plus fine, et donc la plus précise). Dans cette sous-section, nous présentons comment la procédure d'association est adaptée au filtrage multi-échelle.

En pratique, lors de la recombinaison des détections communes, nous sauvegardons en mémoire

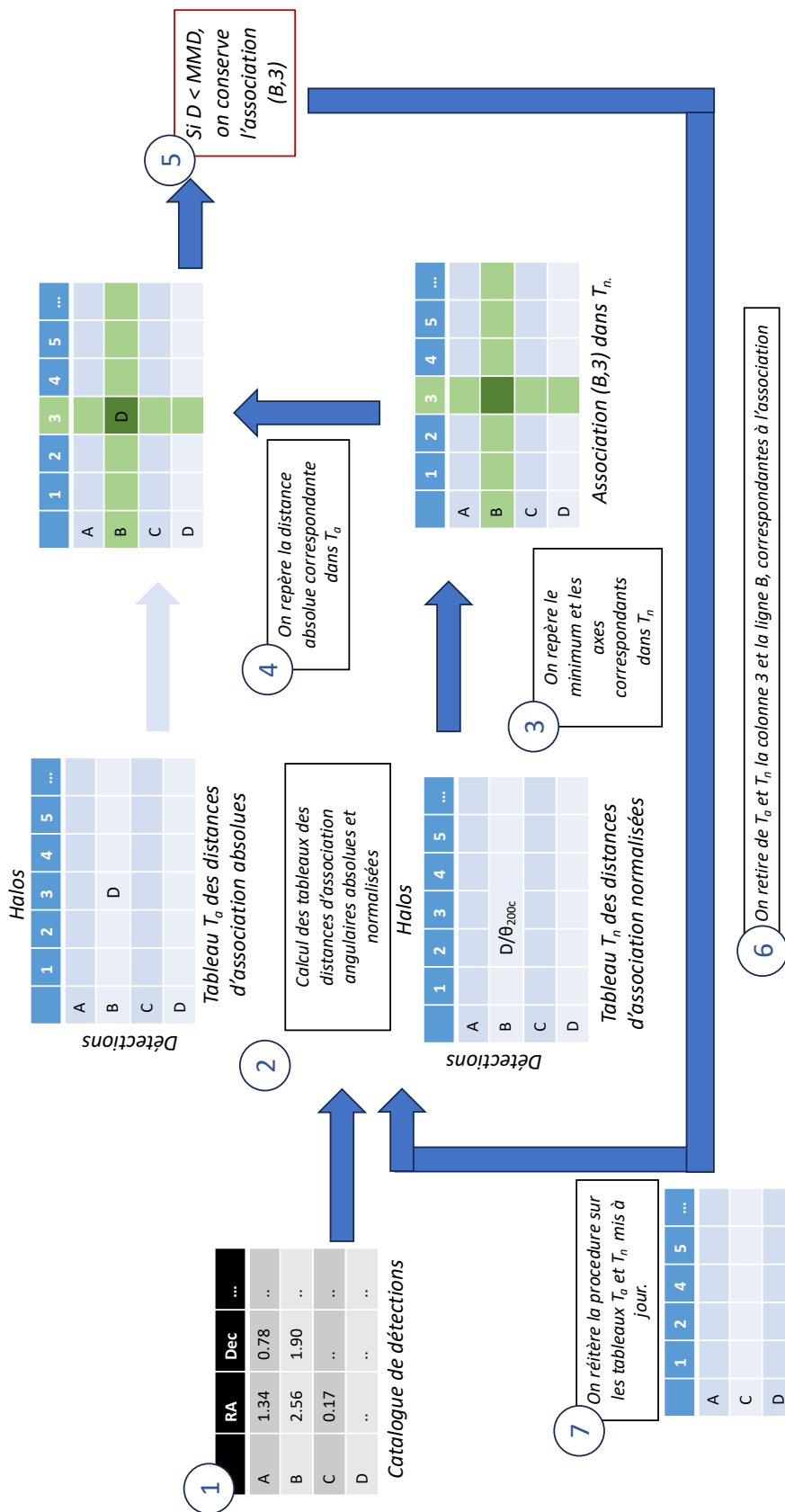


Figure 5.5 – Schéma des différentes étapes (1 à 7) de la nouvelle méthode d'association utilisée pour chaque échelle.

l'échelle la plus fine, et donc la plus précise, sur laquelle la détection est faite. De même, lors du calcul de la MMD pour une échelle donnée, nous avons également traité les ondelettes individuellement. Ainsi, il nous est possible d'utiliser la MMD, adaptée à l'échelle la plus fine de détection, pour affiner la procédure d'association dans le cas multi-échelle.

Dans la pratique, on associe les détections issues du filtrage multi-échelle en fonction de leur échelle de première apparition. En commençant par l'échelle la plus fine, la procédure d'association décrite précédemment, est exécutée en utilisant la distance d'association appropriée définie dans le tableau 5.1. Les associations à cette échelle sont considérées comme uniques et ces halos sont supprimés du catalogue utilisé pour les échelles suivantes. Cette procédure est répétée pour chaque échelle suivante, avec les détections restantes.

Le résultat final est double. Comme pour les autres méthodes à une échelle donnée, on obtient un catalogue de détections associées et un catalogue de détections non associées. En plus de ces deux catalogues pour la méthode W234 complète, nous gardons les catalogues de détections associées et non associées pour chaque échelle ondelette la constituant. Ces données sont importantes pour mesurer les performances individuelles de chaque échelle ondelette et pour comparer leur contribution globale en termes de détections et d'associations.

En résumé, en conservant le détail des échelles de détection relatives aux détections communes, nous pouvons adapter notre procédure d'association au filtrage multi-échelle. La procédure multi-échelle reste ainsi juste vis-à-vis des autres méthodes avec lesquelles elle se compare, en tenant compte de la précision respective de chaque échelle ondelette.

5.4 Conclusions

Dans ce chapitre, nous avons abordé les enjeux de la procédure d'association dans le cadre de la validation et la comparaison des différentes méthodes de détections. Cette comparaison exige que la procédure d'association ne pénalise ou n'avantage aucune méthode. Ainsi, nous avons développé une procédure d'association qui s'adapte à la méthode de filtrage utilisée et qui tient compte des deux enjeux de l'association :

1. D'une part, le nombre d'objets au sein du catalogue de halos de matière noire utilisé, auxquels les détections sont associées, peut avoir un effet dramatique sur la procédure d'association. Nous avons choisi d'appliquer une présélection des halos aux 256 catalogues simulés, à partir de leur rapport signal sur bruit théorique. Avec une présélection fixée à un SNR, $\nu > 2$, nous avons grandement réduit le nombre d'objets, de 15 000 à seulement 300-400; à savoir ceux théoriquement détectables par l'effet de lentille gravitationnelle faible. La procédure d'association avec ce catalogue est ensuite appliquée en calculant la distance entre chaque pic candidat et le halo le plus proche.
2. D'autre part, le choix de la distance maximale d'association, MMD, face à l'imprécision des positions obtenues selon le rayon des filtres peut réduire ou augmenter artificiellement les résultats d'une méthode. Nous avons estimé la distance maximale d'association afin de réduire le nombre de fausses associations. A partir de la dispersion des distances d'association et leurs imprécisions, nous avons calculé la distance maximale d'association pour chacun des filtres. Nous avons ainsi défini une méthode adaptative pour chaque méthode de filtrage à une échelle donnée, et qui s'adapte aux spécificités de celle-ci.
3. Enfin, nous avons mis en place une méthode d'association qui tient ainsi compte des différents filtres utilisés, puis nous l'avons appliquée à la méthode multi-échelle en utilisant la distance d'association maximale propre à chaque échelle. En tenant compte de l'échelle d'où proviennent les détections communes, nous avons pu adapter précisément la procédure d'association aux défis de l'approche multi-échelle.

Ainsi, la mise en place de ces solutions nous a permis de définir une méthode d'association optimale aussi bien pour les méthodes de filtrages à une échelle donnée que pour notre filtrage multi-échelle. L'analyse des résultats issus de cette nouvelle procédure d'association est présentée dans le chapitre 6 avec une analyse sur les associations et les non-associations. De plus, nous y proposons une caractérisation de l'efficacité de la procédure d'association par son taux de fausses associations intrinsèques.

*
* *

6

Résultats des méthodes et analyse

Et nous avons des nuits plus belles que vos jours,
Jean Racine, Lettre XIV à M.Vitart

6.1	Deux outils d'analyse : la complétude et la pureté	97
6.1.1	Évolution de la complétude	98
6.1.2	Pureté vs Complétude	99
6.2	Analyse des détections	101
6.2.1	Caractérisation des détections associées dans le plan redshift-masse	101
6.2.2	Fausse détections	103
6.3	Classification des détections associées et non-associées	104
6.3.1	Classification des détections associées	104
6.3.2	Analyse des détections non-associées	106
6.4	Conclusions	108



Dans ce chapitre, nous allons évaluer et comparer les performances de toutes les méthodes de détection, définies dans les chapitres 3 et 4, en utilisant la procédure d'association définie dans le chapitre 5. Cette étude est réalisée à partir des données de la simulation *DUSTGRAIN-pathfinder* (voir Sect. 2.3). Dans la section 6.1, nous commençons par une analyse des performances des méthodes en s'appuyant sur les outils statistiques usuels que sont la complétude et la pureté. Ensuite, en section 6.2, nous analysons les distributions des détections en termes de redshift, de masse et d'échelle. En parallèle, nous quantifions le nombre de fausses détections dues au bruit de forme pour chaque méthode de détection. Enfin en section 6.3, nous entreprenons une caractérisation et une classification des détections associées (Sect. 6.2.1) et non associées (Sect. 6.3.2), pour chacune des méthodes.

6.1 Deux outils d'analyse : la complétude et la pureté

Pour notre analyse, nous commençons par quantifier la qualité et l'exhaustivité des catalogues de détection des méthodes par rapport aux catalogues de halos de matière noire. Comme le font généralement d'autres études (e.g. [Miyazaki et al., 2018](#); [Euclid Collaboration et al., 2019](#); [Hamana et al., 2020](#)), notre analyse se fait par deux outils statistiques : la pureté et la complétude. Dans cette section,

nous étudions tout d'abord la complétude des catalogues obtenues par les différentes méthodes en utilisant différentes valeurs de SNR théorique utilisé pour la présélection du catalogue de halos. Dans un second temps, nous mesurons la pureté et l'évolution de la complétude en fonction de la pureté.

6.1.1 . Évolution de la complétude

En appliquant la procédure d'association à chaque catalogue de détections, nous obtenons les détections associées et non associées pour chaque méthode. Dans cette sous-section, pour en quantifier les performances et les comparer entre elles, nous utilisons tout d'abord la mesure standard de l'exhaustivité des détections vis à vis des halos connus dans le champ, à savoir la **complétude C**, définie comme :

$$C = \frac{\text{Nombre de détections associées}}{\text{Nombre total de halos dans le catalogue (SNR donné)}}. \quad (6.1)$$

La complétude correspond à la mesure de la qualité externe du catalogue de détections, à savoir sa capacité à reproduire le catalogue de halos de manière exhaustive. Etant donné que cette mesure nécessite de connaître les halos sur le champ observé, cette mesure se prête particulièrement bien à l'étude de données simulées. Notons que la complétude dépend alors des caractéristiques du catalogue de halos utilisé pour l'association. La modification du SNR utilisé pour la présélection dans le catalogue de halos modifie les valeurs de la complétude des catalogues de chaque méthode. Cependant, cela n'affecte pas significativement la comparaison entre les méthodes.

Pour débiter notre analyse, nous avons calculé la complétude pour chaque méthode de détection en utilisant différents seuils de SNR de présélection pour le catalogue du halo. Pour chaque catalogue de détections, nous avons calculé la complétude sur les 256 réalisations en ne considérant que les halos du catalogue dont le SNR théorique est supérieur à une valeur donnée. Nous avons répété cette opération pour différentes valeurs de SNR, ν : 2, 5, 7 et 10. Les résultats, moyennés sur les 256 réalisations, sont présentés dans la Table 6.1. Nous constatons que la présélection des halos avec $\nu > 2$ conduit à une complétude inférieure à 25% pour toutes les méthodes de détection. Cette observation s'explique par le faible nombre de détections (voir table 6.2) devant le nombre de halos dans le catalogue (300-400), du fait du seuillage conservatif utilisé. Comme prévu, la complétude augmente avec le seuil de SNR de la présélection du halo, et presque tous les halos à SNR élevé sont détectés par les différentes méthodes de détection. Même si le comportement global est comparable, cette quantification montre déjà que les méthodes de détection basées sur les filtres W234 et M18 sont plus performantes que celles basées sur les filtres GAUSS, S96, S98, TANH et Jo4 en termes de complétude.

Table 6.1 – Complétudes pour les différentes méthodes de détections pour différentes valeurs de présélection en SNR, affichées en pourcentage. La complétude est calculée sur chaque champ de $5^\circ \times 5^\circ$ en utilisant des présélections de SNR, ν , différents lors de la procédure d'association pour $\nu > 2, 5, 7$ et 10. Les résultats montrent la moyenne et les incertitudes associées dans la complétude (en pourcentage), estimée à partir de la moyenne des 256 réalisations.

Filtre	Complétude ($\nu > 2$)	Complétude ($\nu > 5$)	Complétude ($\nu > 7$)	Complétude ($\nu > 10$)
GAUSS	15.1 ± 0.4	54.7 ± 0.2	85.8 ± 0.1	88.0 ± 0.1
S96	13.3 ± 0.3	65.6 ± 0.1	89.3 ± 0.1	98.1 ± 0.1
S98	10.6 ± 0.2	55.5 ± 0.1	85.1 ± 0.1	94.7 ± 0.1
TANH	11.9 ± 0.2	62.3 ± 0.1	85.3 ± 0.1	95.8 ± 0.1
Jo4	12.9 ± 0.2	59.7 ± 0.2	87.7 ± 0.1	97.4 ± 0.1
M18	17.1 ± 0.2	71.5 ± 0.1	93.9 ± 0.1	98.1 ± 0.1
W234	21.0 ± 0.2	75.3 ± 0.1	95.7 ± 0.1	99.2 ± 0.1

6.1.2 . Pureté vs Complétude

Bien que la complétude pour une présélection en SNR donnée soit un bon indicateur de la performance des différentes méthodes de détection, elle ne fournit aucune information sur la proportion de détections non associées. C'est pourquoi une mesure de la pureté de l'échantillon est également nécessaire pour affiner la comparaison. Dans cette sous-section, nous calculons pour chaque catalogue de détection, la **pureté**, P :

$$P = \frac{\text{Nombre de détections associées}}{\text{Nombre totale de détections}}. \quad (6.2)$$

Cette quantité mesure la qualité interne du catalogue de détections (e.g. [Euclid Collaboration et al., 2019](#)), par la proportion de détections associées et peut être utilisée pour évaluer le taux de fausse détection. Nous avons calculé la pureté des catalogues de détections obtenus par chaque méthode, pour un seuil de présélection du catalogue fixé à $\nu > 2$. Ce seuil a été choisi de manière à conserver les halos détectables sans modifier artificiellement le nombre d'associations et donc, la pureté, comme expliqué dans la Sect. 5.2.2. Les résultats des différentes méthodes de détection, moyennés sur les 256 réalisations de taille $5^\circ \times 5^\circ$ de la simulation, sont résumés dans le tableau 6.2. La pureté estimée pour toutes les méthodes de détection est comparable, à environ 85%. Cela nous a conduit à effectuer une analyse au cas par cas des détections non associées détaillées dans la suite de ce chapitre (voir Sect. 6.3.2). En particulier, nous avons observé que pour certaines détections, la procédure d'association a échoué ou est trop conservatrice. Bien que cela ait évidemment un impact sur la pureté estimée, cela n'affecte pas la comparaison entre les méthodes car cela les affecte toutes dans la même mesure.

Table 6.2 – Performances des méthodes de détection en termes de détections, pureté et complétude. Les valeurs en gras présentent les maximums de chaque colonne.

Filtre	Nombre de détections	Pureté P [%]	Complétude C [%]
GAUSS	85.7 ± 1.2	74.1 ± 0.4	15.1 ± 0.4
S96	57.6 ± 1	83.7 ± 0.3	13.3 ± 0.3
S98	43.8 ± 0.7	87.2 ± 0.3	10.6 ± 0.3
TANH	50.9 ± 0.9	84.5 ± 0.4	11.9 ± 0.2
Jo4	47.9 ± 0.5	88.3 ± 0.3	12.9 ± 0.2
M18	66.6 ± 0.7	83.1 ± 0.3	17.1 ± 0.2
W234	83.2 ± 0.9	82.7 ± 0.3	21.0 ± 0.2

En comparant les résultats du tableau 6.2, nous constatons que les méthodes de détection basées sur les filtres S96, S98, TANH et Jo4 conduisent à une pureté plus élevée mais à une complétude plus faible. Bien que la complétude ne soit pas directement liée au nombre de détections, mais plutôt au nombre de détections associées, la complétude plus faible pour ces filtres est cohérente avec leur nombre moyen de détections inférieur à celui des deux autres méthodes. Leur représentation dans l'espace de Fourier, dont la largeur de la bande passante est plus étroite, explique ce nombre moindre de détections. Inversement, la complétude est plus élevée et la pureté est plus faible pour les filtres GAUSS, M18 et W234, qui utilisent des filtres dont la bande passante est plus large. Cependant, l'approche multi-échelle couvre une plus large gamme d'échelles, et donc de tailles d'amas possibles, et nous permet donc d'atteindre une plus grande complétude. Par conséquent, l'approche multi-échelle permet d'obtenir un nombre moyen de détections plus élevé pour une même pureté. Par rapport aux filtres M18 et GAUSS, par exemple, l'approche multi-échelle permet d'obtenir 25% de détections en plus pour une pureté similaire ou supérieure. Le filtre GAUSS a le plus haut nombre de détections, ce qui peut expliquer sa complétude élevée. Cependant, comparé aux filtres M18 et W234, il montre une complétude plus faible et la pureté la plus faible. Ces performances soulignent d'ores et déjà son caractère

non optimal pour la détection des amas de galaxies par effet de lentille gravitationnelle faible (voir Sect. 3.2.2).

Le tableau 6.2 montre également que certaines méthodes de détection sont plus efficaces en termes de complétude, tandis que d'autres sont plus efficaces en termes de pureté. Comme la pureté et la complétude ne peuvent être comparées séparément, une comparaison équitable entre les méthodes de détection nécessite une analyse de l'évolution de la complétude en fonction de la pureté. Nous avons donc calculé cette évolution pour chaque méthode de détection en utilisant le catalogue $SNR > 2$ présélectionné pour l'association. Dans la pratique, nous avons fait varier le seuil dans la procédure de détection décrite dans la Sect. 3.3.3 par pas de 5% et avons obtenu de nouvelles valeurs de complétude et de pureté pour chaque nouvelle valeur du seuil de détection. Ce faisant, on fait varier le taux de fausses détections qu'on accepte, ce qui modifie la pureté et permet de calculer la complétude qui en découle.

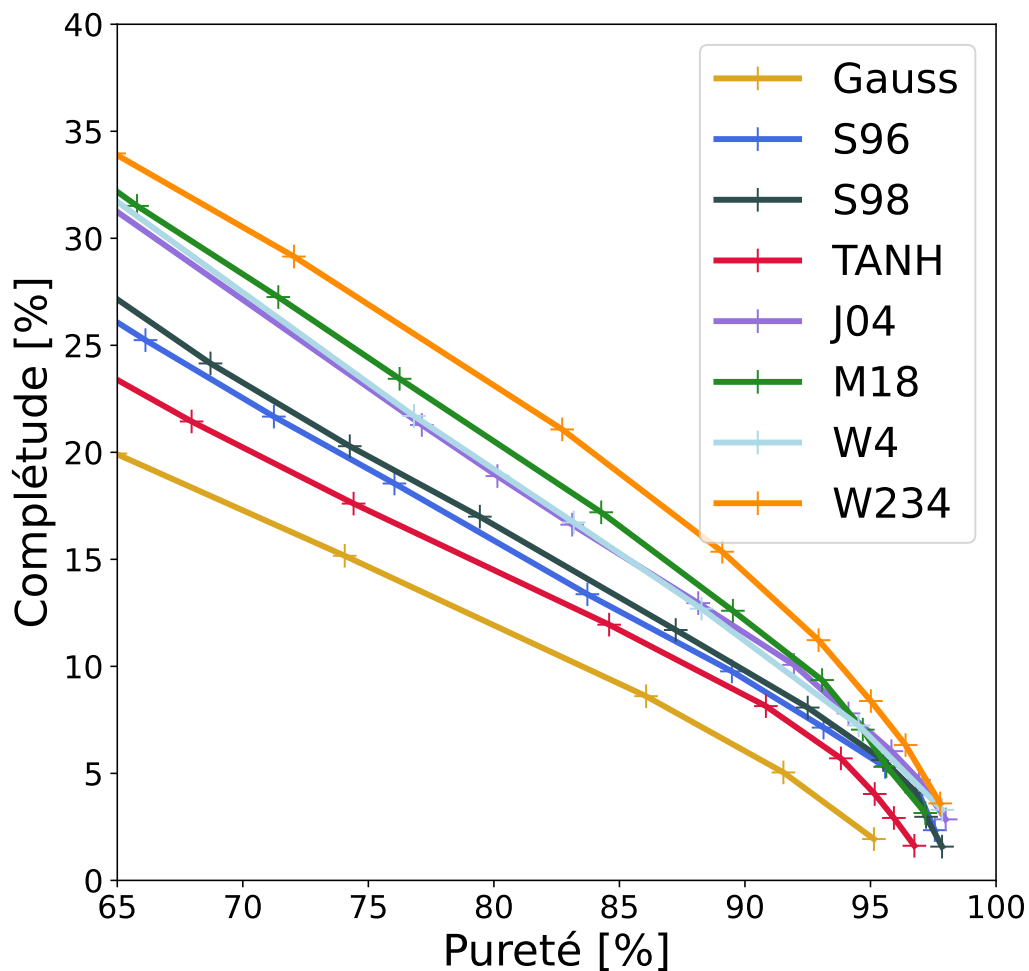


Figure 6.1 – Complétude en fonction de la pureté pour les méthodes de détection basées sur les filtres MO S96 (bleu), S98 (TBD), TANH (rouge), J04 (violet) et M18 (vert), le filtre ondelettes W4 (bleu clair) et le filtrage ondelettes W234 (orange). Les croix représentent précisément l'endroit où les méthodes sont évaluées en termes de pureté et de complétude. Ces valeurs sont calculées en utilisant une présélection de halo $\nu > 2$ et en calculant la moyenne sur les 256 réalisations.

Les résultats de cette évolution, moyennés sur les 256 réalisations, sont présentés dans la figure 6.1. Nous constatons que l'approche multi-échelle est plus performante que les autres méthodes de détection pour toutes les valeurs de pureté. Dans le cas de la comparaison avec le filtrage gaussien qui présente les performances les plus faibles, il y a une augmentation de la complétude allant jusqu'à 74% de plus, à faible pureté. Par rapport à la méthode de détection de masse d'ouverture basée sur le filtre TANH, cette augmentation de la complétude atteint jusqu'à 65% de plus, à faible pureté.

L'augmentation de la complétude est plus faible si on se compare à la méthode de détection basée sur le filtre M18, mais elle reste significative dans la gamme des puretés étudiées. Nous observons également que les performances du filtre ondelettes W4 seul et du filtre à une échelle donnée Jo4 sont comparables.

En résumé, les outils usuels, que sont la pureté et la complétude, permettent de caractériser et de comparer les méthodes entre elles. L'étude de la complétude et son évolution selon la pureté mettent en avant la supériorité de l'approche multi-échelle comparée aux autres méthodes, en obtenant une complétude plus élevée pour toute valeur de pureté. Pour autant, malgré un seuil artificiellement augmenté la pureté n'atteint la valeur escomptée de 100% pour aucune des méthodes. Cette observation entraîne le besoin de caractériser plus en détails à la fois les détections associées et non associées.

6.2 Analyse des détections

Après avoir évalué la complétude et la pureté de chaque méthode de détections, nous souhaitons expliquer la pureté obtenue. Dans cette section, nous affinons notre étude des détections associées en termes de redshift et de masse. Puis, nous continuons l'analyse des performances des méthodes en étudiant le taux de fausses détections propre à chacune.

6.2.1 . Caractérisation des détections associées dans le plan redshift-masse

Une fois l'association réalisée, il nous est possible d'étudier la distribution des associations du point de vue du catalogue de halos. Cela permet de mettre en valeur la sensibilité des différentes méthodes en fonction des propriétés de masse et de redshift des halos associées. Dans cette sous-section, nous caractérisons les détections associées en traçant leur distribution dans le plan redshift-masse. La figure 6.2 montre la distribution dans le plan redshift-masse des amas détectés pour chacune des sept méthodes de détection. La figure est obtenue à partir d'une seule réalisation, mais elle est représentative de l'ensemble de l'échantillon de 256 réalisations (voir Table 6.2). L'association est effectuée en utilisant uniquement les halos dont $\nu > 2$. Dans tous les panneaux, les points bleu clair correspondent aux halos du catalogue des halos avant la présélection en SNR, et les points bleu foncé aux halos avec $\nu > 2$. Les croix orange identifient les amas détectés et associés par les différentes méthodes de détection.

Comme prévu, toutes les méthodes de détection sont plus sensibles au signal des amas de masse élevée se trouvant à faible redshift. La comparaison montre que la méthode de détection multi-échelle W234 a plus de détections associées que les autres, en accord avec les résultats de complétude présentés dans le Tableau 6.1. En outre, la quasi-totalité des amas détectés par les méthodes de détection GAUSS, S96, S98, TANH, Jo4 et M18 sont également détectés par la procédure de détection multi-échelle. Par rapport aux 6 autres méthodes de détection, l'approche multi-échelle détecte généralement plus d'amas à des masses plus faibles, et jusqu'à des redshifts plus élevés.

Dans la figure 6.3, nous entreprenons une analyse plus approfondie de l'approche multi-échelle, en étudiant la contribution de chaque échelle ondelette au nombre total de détections associées. Nous montrons la distribution des amas détectés et associés dans le plan $z - M$ pour le filtre ondelette W2 (panneau supérieur gauche), le filtre ondelette W3 (panneau supérieur droit) et le filtre ondelettes W4 (panneau inférieur gauche). Le panneau inférieur droit montre la distribution des amas détectés et associés dans le plan $z - M$ pour la méthode de détection multi-échelle W234. Les couleurs nous permettent de voir la contribution de chaque échelle après recombinaison des détections multiples. Les différents panneaux mettent également en évidence les différences en termes d'amas ciblés par les différentes échelles ondelettes. Si nous regardons le nombre de détections associées pour chaque échelle ondelette après recombinaison, nous voyons que la proportion de détections faites à chaque échelle ne varie pas beaucoup d'une réalisation à l'autre. En moyenne, le filtre onde-

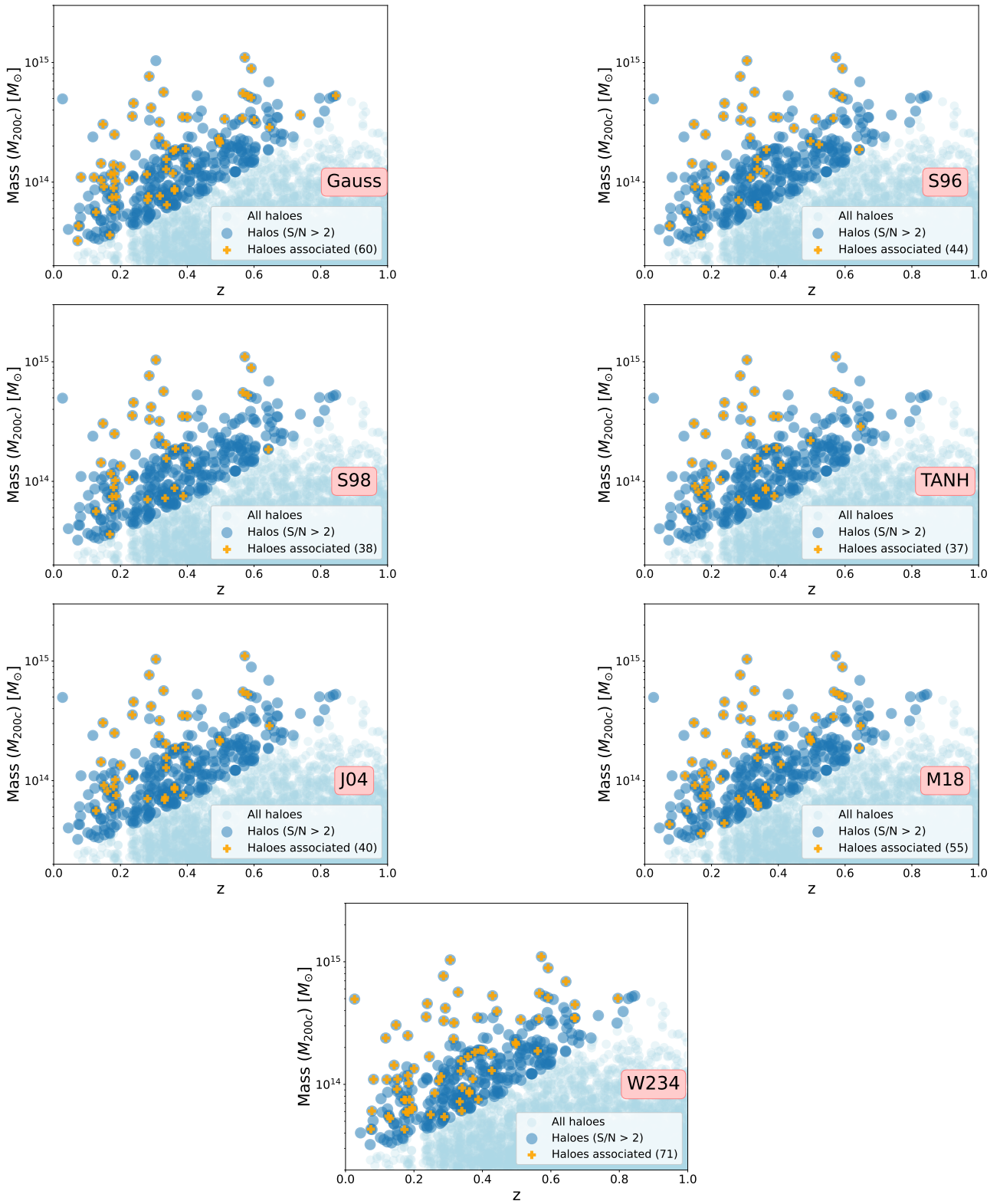


Figure 6.2 – Distribution des halos associés dans le plan $z - M$ pour un patch de $5^\circ \times 5^\circ$, en utilisant le filtre gaussien dans le panneau supérieur gauche, le filtre S96 dans le panneau supérieur droit, le filtre S98 dans le panneau central gauche, le filtre TANH dans le panneau central droit, le filtre J04 dans le panneau inférieur gauche, le filtre M18 dans le panneau inférieur droit, et les filtres ondelettes W234 dans le panneau inférieur central. Les points bleu clair correspondent à tous les halos du catalogue de halos, et les points bleu foncé aux halos que nous avons présélectionnés pour avoir un SNR > 2 . Les croix orange surlignées correspondent aux halos qui ont fait l'objet d'une détection, et entre parenthèses, nous notons le nombre d'associations pour chaque méthode sur une réalisation. Cette réalisation a été choisie pour être représentative de la performance des différentes méthodes en termes de nombre de détections.

lette W2 fournit jusqu'à $29\% \pm 0.39$, W3 environ $33\% \pm 0.57$, et W4 environ $38\% \pm 0.42$ des détections correspondantes. Par conséquent, les trois échelles ondelettes sont utiles et contribuent de manière similaire au nombre total de détections correspondantes. Cela confirme le caractère complémentaire des détections associées obtenues par les différentes échelles.

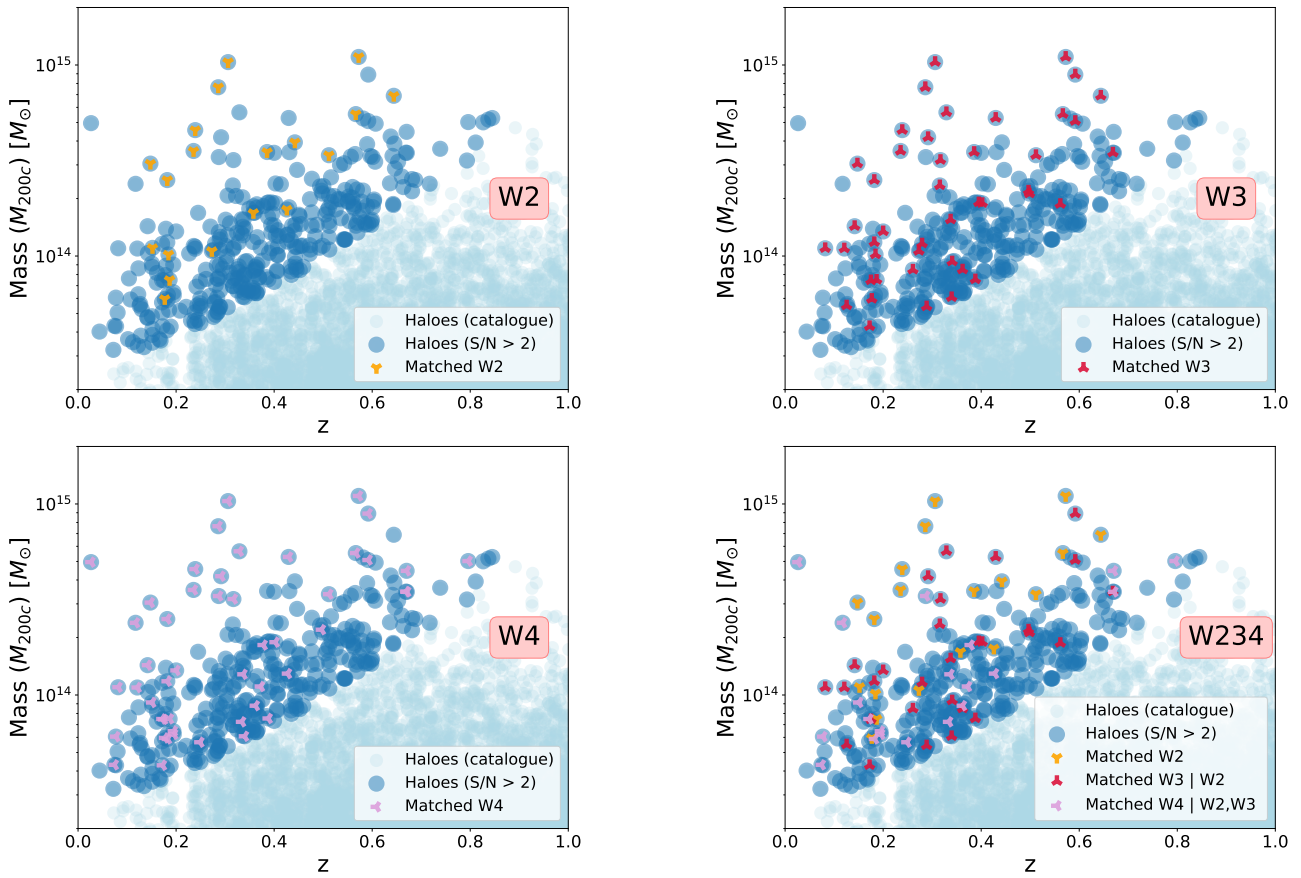


Figure 6.3 – Distribution des amas associés dans le plan z - M pour un champ de $5^\circ \times 5^\circ$, obtenue en utilisant le filtre ondelette W2 (panneau supérieur gauche), le filtre ondelette W3 (panneau supérieur droit), le filtre ondelette W4 (panneau inférieur gauche), et l'approche multi-échelle (panneau inférieur droit).

Un aspect frappant des Figs. 6.2 et 6.3 est que certains halos avec un SNR prédit très élevé ne sont pas détectés, quelle que soit la méthode utilisée. Nous avons effectué une analyse au cas par cas de ces halos et nous avons trouvé plusieurs explications à cela. Dans certains cas, la détection est manquée à cause du bruit de forme. Dans d'autres cas, la détection est manquée parce que le signal de lentille est diminué par la distribution de redshift des sources derrière l'amas (à travers les Eqs. 2.18 et 2.17).

6.2.2 . Fausses détections

Nous avons montré dans la table 6.2 que la pureté atteint une valeur de $\sim 85\%$ pour les différentes méthodes même avec un seuillage très conservatif. Dans cette sous-section, nous étudions le taux de fausses détections intrinsèque à ce seuillage pour comprendre le niveau de pureté mesuré.

Dans ce qui suit, nous définissons les fausses détections comme les détections qui ne sont pas liées à un signal de lentille (c'est-à-dire une surdensité) et qui sont donc uniquement dues au bruit. Nous quantifions le nombre de fausses détections dues au bruit pour chaque méthode de détection sur les données simulées. Pour ce faire, nous avons randomisé l'orientation des sources dans les catalogues de galaxies bruités afin de supprimer le signal de lentille et de générer des catalogues

de galaxies identiques aux réalisations mais purement bruités. Nous avons ensuite dérivé les cartes de convergence associées et appliqué les différentes méthodes de détection, comme décrit dans les sections 3.3.3 et 4.2.3.

Pour chaque méthode, nous avons ensuite calculé le taux moyen de fausses détections sur les 256 réalisations. Ce taux a été obtenu à partir du rapport entre le nombre de détections purement dues au bruit (obtenues sur les cartes purement bruitées) et le nombre de détections obtenues lorsque le signal n'est pas supprimé (voir colonne 2, Table 6.2). Le taux de fausses détections dues au bruit est indiqué, pour chaque méthode de détection, dans la deuxième colonne de la table 6.4.

Le taux de fausses détections dû au bruit de forme, le plus faible est de 1.1% pour la méthode de détection J04. Les filtres S96, S98, TANH et W234 comptent environ 1.8% de fausses détections, et ce taux monte à 3% pour la méthode de détection M18. L'impact du bruit dépend du filtre. Le niveau de bruit est plus faible à grande échelle qu'à petite échelle, ce qui explique pourquoi la méthode de détection J04 est moins sensible au bruit que les autres méthodes. En revanche, la méthode de détection M18 est la plus sensible au bruit car elle intègre les petites échelles en même temps que les grandes échelles sans parvenir à une distinction (voir Fig. 4.5). De même, le comportement oscillatoire des filtres S96, S98 et TANH augmente leur sensibilité au bruit, en intégrant des échelles plus petites. La méthode de détection W234 intègre également de petites échelles, mais l'impact du bruit de forme est réduit grâce à sa procédure de détection multi-échelle les traitant individuellement. Malgré ces différences, le nombre de fausses détections dues au bruit de forme est faible, quelle que soit la méthode de détection.

En résumé, dans cette section, nous avons mis en avant la dominance de la méthode multi-échelle pour la détection des amas de galaxies. En particulier, l'analyse du plan redshift-masse souligne que la méthode multi-échelle permet de détecter la majorité des amas massifs à faible redshift que détecterait les autres méthodes. À cela se rajoute la détection de halos à basse masse et à bas redshift que les autres méthodes manquent. Nous avons également étudié le taux de fausses détections propre à chaque méthode de détection, et montré que ce taux était très faible ($< 3\%$). Ce qui est cohérent avec le choix du seuil utilisé mais qui ne permet pas d'expliquer la pureté d'environ 85% des différentes méthodes.

6.3 Classification des détections associées et non-associées

Ainsi, la caractérisation des taux de fausses détections propres à chaque méthode ne suffit pas à expliquer le niveau de pureté mesurée ($\sim 85\%$). Il apparaît que les notions de détections associées et non-associées doivent être approfondies pour compléter notre analyse. Dans cette section, nous entamons cette idée par la classification des détections associées selon deux catégories : vraies détections correctement associées et fausses détections incorrectement associées. Puis, nous suivons la même approche et identifions pour les détections non-associées trois catégories : fausses détections correctement non associées, signal boosté et incorrectement non-associé et signal de plusieurs sources incorrectement non-associé.

6.3.1 . Classification des détections associées

Pour affiner notre analyse des détections, nous nous intéressons aux détections associées. Dans cette sous-section, nous classons les détections associées en deux catégories pour les étudier en détails. Pour ce classement, nous avons identifié deux catégories :

- **Vraies détections correctement associées**

La première catégorie correspond évidemment aux détections associées qui correspondent bien à un signal dans le champ observé. Compte tenu du taux de fausses détections, cette catégorie représente la grande majorité des détections associées. Cependant, au sein de cette

dénomination se distingue deux situations : les associations simples et les associations multiples. Presque toutes les détections associées correspondent à des associations avec un seul halo du catalogue ('Single matched' dans la Fig. 6.4). Cependant, il est inévitable que certaines détections réelles puissent être correctement associées mais soient en fait associées à deux halos (ou peut-être plus) dont le signal de lentille est fusionné ou proche géométriquement. Cela peut se produire lorsque deux halos sont alignés dans la ligne de visée (L.O.S) ou lorsque leur distance angulaire apparente est faible par rapport à la taille du filtre. Dans ce cas, la procédure d'association ne fait pas de différence et l'association se fait simplement avec le halo le plus proche. Nous estimons qu'environ 5% des détections associées appartiennent à cette catégorie en recherchant des halos multiples à proximité des détections associées ("Multiple Matched (L.O.S)" dans la Fig. 6.4). Ces associations imparfaites peuvent réduire légèrement la complétude estimée, mais toutes les méthodes sont affectées de la même manière par cet effet.

- **Fausse détections incorrectement associées**

Parmi les détections associées, une petite fraction est en fait constituée de fausses détections associées par chance. Il s'agit de détections dues au bruit de forme qui sont associées parce qu'elles sont proches d'un halo dans le catalogue. Leur nombre dépend à la fois du nombre de fausses détections (voir la sous-section 6.2.2) et du taux de fausses associations. Pour quantifier le taux de fausses associations, nous avons considéré des positions aléatoires sur le champ pour les détections de chaque méthode et leur avons appliqué la procédure d'association. Cette analyse nous permet de voir le taux d'association pour les différentes méthode de détection. Ces résultats, moyennés sur les 256 réalisations, sont donnés dans la table 6.3. Nous détaillons également le taux de fausses associations obtenues pour les échelles W2, W3 et W4 individuellement pour mettre en évidence l'impact de l'utilisation d'une MMD optimale.

Le taux de fausses détections incorrectement associées est alors la multiplication des taux de fausses détections et de fausses associations. Dans notre étude, nous avons intentionnellement maintenu ce chiffre très bas pour toutes les méthodes de détection (moins de 0.3%). Cependant, il faut être prudent, car les procédures de détection et d'association peuvent toutes deux faire augmenter ce chiffre. Il est donc essentiel de toujours calculer ce nombre pour évaluer pleinement les performances des méthodes de détection.

Table 6.3 – Taux de fausses association pour les différents filtres.

Filtre	Taux de fausses associations [%]
Gauss	3.4 ± 0.3
S96	12.8 ± 0.3
S98	9.1 ± 0.4
TANH	9.0 ± 0.3
J04	14.2 ± 0.4
M18	10.6 ± 0.4
W234	10.8 ± 0.4
W2	2.8 ± 0.3
W3	8.7 ± 0.4
W4	13.6 ± 0.4

Cette classification en deux catégories met en avant que la procédure d'association est bien mise en place. En effet, parmi les détections associées, plus de 95% le sont avec un halo unique et seulement moins de 0.3% des détections sont des fausses détections mais associées par chance.

6.3.2 . Analyse des détections non-associées

Après avoir étudié les détections associées, nous nous intéressons aux détections restantes non-associées. Dans cette sous-section, nous nous concentrons sur les détections qui ne sont pas associées à un halo, afin de compléter l'analyse des différentes méthodes de détection. Ces détections non associées peuvent être classées en trois catégories :

- **Fausses détections correctement non associées**

La catégorie la plus intuitive pour les détections non associées est celle composée des détections dues au bruit (voir Sect. 6.2.2). Au-delà de la petite fraction qui est incorrectement associée parce qu'elle est située près d'un halo dans le catalogue (voir Fausses détections incorrectement associées), presque toutes ces fausses détections ne sont pas associées. Leur nombre correspond donc approximativement au taux de fausses détections, et est donné dans la deuxième colonne du Tableau 6.4. Ce nombre est très faible par rapport à la pureté estimée dans la section 6.1.2, et d'autres explications doivent être trouvées pour expliquer les $\sim 15\%$ restants des détections non associées.

- **Un signal lentille boosté et incorrectement non associé**

Une explication possible pour les détections non associées restantes est que certaines de ces détections devraient être associées à des halos qui ont été retirés du catalogue de halos après la présélection SNR. Ceci pourrait se produire parce que, lors de la sélection, le SNR que nous estimons pour chaque halo du catalogue est théorique. Il est calculé en supposant une densité théorique moyenne des sources et une distribution théorique, p_{z_s} , pour le redshift des sources. En pratique, les sources qui génèrent le signal d'un amas sont peu nombreuses. Par conséquent, leur nombre et leur distribution de redshift peuvent s'écarter de manière significative des valeurs théoriques, et peuvent dans certains cas augmenter le signal de lentille d'un amas, indépendamment de sa masse et de son redshift. Cependant, la sélection de halos avec un SNR beaucoup plus faible dans le catalogue de halos augmente grandement le taux de fausses associations, et donc le nombre de fausses détections mal associées (voir fausses détections incorrectement associées). La table 6.3 donne le taux de fausses associations lors de l'utilisation d'une présélection SNR > 2 dans le catalogue du halo. Ce taux reste inférieur à 15% pour toutes les méthodes de détection. Ainsi, le choix d'imposer une présélection SNR > 2 dans le catalogue du halo est un compromis entre le taux de fausses associations et le taux d'associations manquées. Nous avons estimé la fraction de détections non associées due à un signal de lentille boosté en appliquant la procédure d'association une seconde fois pour les 256 réalisations, mais avec une présélection des halos de SNR > 1.5 . Les résultats pour chaque méthode sont donnés dans la troisième colonne de la table 6.4.

- **Le signal lentille de plusieurs sources mélangé et incorrectement non associé**

Une autre explication pour les détections non associées restantes est le cas de deux halos contribuant à la même détection. D'une part, ces deux halos peuvent être alignés le long de la ligne de visée. Cela signifie que la détection provient de deux halos avec SNR < 2 qui sont alignés le long de la ligne de visée et dont les signaux de lentille ne peuvent pas être séparés sans tenir compte de la distribution des redshifts des galaxies sources lors de l'étape même de détection.

D'autre part, ces deux halos peuvent également avoir leur signal mélangé à cause de la taille du filtre. En effet, certaines détections non associées peuvent provenir de deux halos dont la distance apparente est faible par rapport au rayon du filtre. Le signal de lentille de ces deux halos étant fusionné dans la carte filtrée, il y a une seule détection associée aux deux halos, et dont la position dépend de leur masse et de leur redshift. Dans ce cas, la détection n'est pas associée lorsque le SNR théorique des deux halos est inférieur à 2, ou lorsque la position de la détection (au barycentre) est trop éloignée du centre des deux halos.

Table 6.4 – Classification des **detections**.

Filtre	Associations [%]	Fausse détactions [%]	Objets boostés (p_z) [%]	Signaux mélangés [%]	Autres [%]
Gauss	74.1 ± 0.4 (64)	6.8 ± 0.4 (6)	3.9 ± 0.2 (3)	5.7 ± 0.3 (5)	~ 9.5 (8)
S96	83.7 ± 0.3 (48)	2.0 ± 0.2 (2)	3.5 ± 0.2 (2)	5.8 ± 0.3 (3)	~ 5.0 (3)
S98	87.2 ± 0.3 (38)	1.4 ± 0.2 (1)	2.2 ± 0.1 (1)	6.2 ± 0.5 (2)	~ 3.0 (2)
TANH	84.5 ± 0.3 (43)	1.7 ± 0.3 (1)	4.3 ± 0.3 (2)	6.2 ± 0.3 (3)	~ 3.3 (2)
Jo4	88.3 ± 0.3 (42)	1.1 ± 0.2 (1)	3.6 ± 0.2 (1)	5.9 ± 0.4 (2)	~ 1.1 (1)
M18	83.1 ± 0.3 (56)	3.0 ± 0.3 (2)	4.0 ± 0.1 (3)	5.6 ± 0.3 (3)	~ 4.3 (3)
W234	82.7 ± 0.3 (67)	1.8 ± 0.3 (1)	6.4 ± 0.2 (6)	5.5 ± 0.2 (5)	~ 3.6 (4)

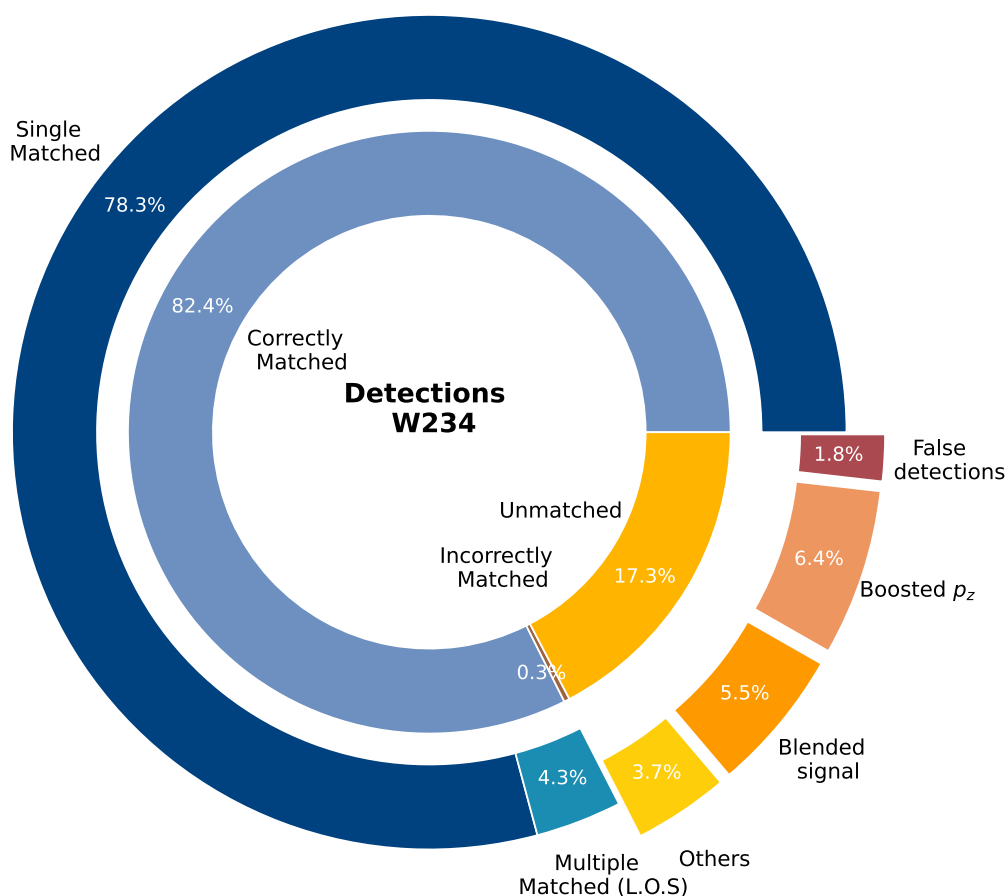


Figure 6.4 – Caractérisation des détactions de la méthode W234. Les détactions associées (segments bleus) sont classées comme correctement associées avec des halos simples ou multiples, ou incorrectement associées comme défini dans la section 6.2.1. Les détactions non associées sont sous-catégorisées en signaux mélangés, signaux boostés par la distribution des sources, fausses détactions, ou autres, comme défini dans la section 6.3.2. La moyenne des résultats est calculée sur les 256 réalisations.

En résumé, nous avons montré que les détactions associées et non-associées pouvaient être approfondies et classées en différentes catégories. Dans la figure 6.4, nous avons présenté un récapitulatif des différentes catégories étudiées dans cette section. La figure met en avant les différents pourcentages obtenus de chaque catégorie dans le cas de la méthode multi-échelle W234. L'analyse de cette méthode souligne qu'une partie des détactions non-associées correspond à un signal lentille présent sur le champ étudié. Cela nous permet de corriger la pureté mesurée pour la méthode, pour atteindre 94.3 dans le cas de la méthode multi-échelle.

6.4 Conclusions

Dans ce chapitre, nous avons présenté la comparaison des méthodes définies dans les chapitres 3 et 4 grâce à la procédure du chapitre 5. Nous avons divisé cette comparaison en 3 axes :

- La mesure de la complétude et de la pureté ont permis une première comparaison des méthodes selon différents seuils SNR de présélection du catalogue de halos. La complétude dépend du seuil de présélection du catalogue du halo. Cependant, quel que soit le seuil exact de présélection du rapport signal sur bruit, les filtres GAUSS, S96, S98, TANH et Jo4 présentent une complétude plus faible que les filtres M18 et W234. En revanche, le filtre Jo4 présente une pureté légèrement supérieure au détriment de la complétude. Ainsi, la méthode multi-échelle W234 a montré une complétude supérieure aux 6 autres méthodes à la fois pour les différents seuils SNR mais aussi pour toute valeur de pureté mesurée.
- La méthode multi-échelle détecte pratiquement toutes les détections des approches gaussienne et MO individuelles, avec l'ajout de détections supplémentaires, en particulier à des masses plus faibles et à des redshifts plus bas.
- La quantification du taux de fausses détections mesurées pour chaque méthode est très faible ($< 3\%$) et ne suffit pas à expliquer toutes les détections non associées.

Puis, nous avons aboutis à une catégorisation, la plus précise possible, des détections selon 5 cas de figures. Cette classification permet de nuancer les performances des différentes méthodes et en particulier la méthode multi-échelle.

- Nous avons montré que plus de 82% des détections sont correctement associées, et parmi celles-ci, l'écrasante majorité (95%) est due à un seul halo. Pour notre procédure d'association, les détections correctement associées dues à des halos multiples le long de la ligne de visée représentent moins de 5% des associations. De plus, comme nos méthodes de détection et d'association visent une grande pureté, le nombre de halos incorrectement associés est négligeable ($< 1\%$).
- Pour la méthode multi-échelle, environ 17% des détections ne sont pas associées. Pourtant seule une petite fraction de ces détections (10%) est due à des fausses détections. Le reste peut être expliqué principalement par des halos en dessous du seuil SNR dont le signal a été boosté en raison d'une distribution en redshift heureuse des galaxies sources (37%), ou d'un signal de lentille mélangé à partir d'alignements le long de la ligne de visée, ou mélangé à partir d'une distance angulaire de séparation plus petites que la taille du filtre (32%). Les dernières détections non associées (20%) peuvent probablement être associées à des halos avec un SNR < 1.5 qui n'apparaissent pas dans le catalogue en raison de la présélection.

Dans cette étude, nous avons montré qu'une approche multi-échelle donne de meilleurs résultats, en termes de pureté et de complétude, que les méthodes de détection actuellement utilisées.

*
* *

Quatrième partie

Quatrième partie : Vers l'application à des données réelles

7

Préparation pour les données d'*EUCLID*

[*EUCLID*].. c'est le projet d'observation cosmologique le plus ambitieux jamais conçu par son ampleur. En un mot, une véritable aventure scientifique et humaine, avec ses rebondissements, ses moments d'euphorie comme de doute, ses coups de théâtres.

Y. Mellier, Euclid, Lumière sur l'énergie noire

7.1	EUCLID, l'observation de l'Univers Sombre	112
7.1.1	Un télescope et un projet européen	112
7.1.2	Les instruments et les données lentilles	114
7.2	Le challenge WLC	115
7.2.1	Contexte du challenge WLC	115
7.2.2	Caractéristiques du jeu de données	115
7.2.3	Caractéristiques des catalogues de détections	116
7.2.4	Procédure d'association	117
7.3	Les différentes méthodes de détection	119
7.3.1	Présentation des méthodes	119
7.3.2	Nouveautés pour l'approche multi-échelle	121
7.3.3	La distance maximale d'association MMD	122
7.4	Premiers résultats	124
7.4.1	Pureté versus complétude	124
7.4.2	Détections associées dans le plan z-M	128
7.4.3	Caractérisation des détections associées communes	130
7.5	Conclusions	132



Notre ambition pour la méthode multi-échelle est son application à des données réelles. À terme, nous visons son application sur les données à venir du télescope *EUCLID*. Lancé en Juillet 2023, ce télescope européen fournira des données d'observation d'une qualité sans précédent mais hélas, après la rédaction de ce manuscrit. L'enjeu de notre travail est alors de nous préparer aux données d'*EUCLID* à venir dans les conditions les plus proches de celles attendues et d'affiner l'analyse à venir sur les données une fois disponibles. Pour ce faire, un challenge de détection des

amas par effet de lentille gravitationnelle faible (Weak Lensing Cluster Challenge, challenge WLC) a été organisé par le groupe de travail scientifique qui s'intéresse à la détection des amas de galaxies par l'effet de lentille gravitationnelle faible au sein de la collaboration *EUCLID*. Le but de ce challenge, auquel nous avons activement participé tant par les méthodes de détection que par l'analyse, est la comparaison des différentes méthodes de détection qui parcourront un jour les données *EUCLID*. Dans ce chapitre, nous présentons notre participation dans ce challenge, et notre implication dans l'étude des méthodes pour en produire une première comparaison. En section 7.1, nous décrivons plus en détails le télescope *EUCLID* ainsi que, son contexte scientifique, les instruments à son bord et les données qui seront produites. Puis, en section 7.2, nous présentons le contexte du challenge WLC auquel nous avons pris part et sa mise en place (caractéristiques propres, procédure d'association). En section 7.3, nous détaillons les différentes méthodes appliquées et comparées dans le cadre de ce challenge. Enfin, en section 7.4, nous montrons les premiers résultats obtenus par chaque approche implémentée en termes de pureté et complétude, définis au chapitre précédent.

Nous présentons dans ce chapitre uniquement les résultats et les analyses que nous avons conduits en vue d'une publication pour laquelle des analyses complémentaires seront nécessaires.

7.1 *EUCLID*, l'observation de l'Univers Sombre

En section 1.3.4, nous avons brièvement présenté le télescope *EUCLID* (Laureijs et al., 2011) et ses objectifs. Dans cette section, nous détaillons les ambitions et le contexte scientifique de ce projet européen. Puis, nous présentons les instruments et les données associées.

7.1.1 . Un télescope et un projet européen



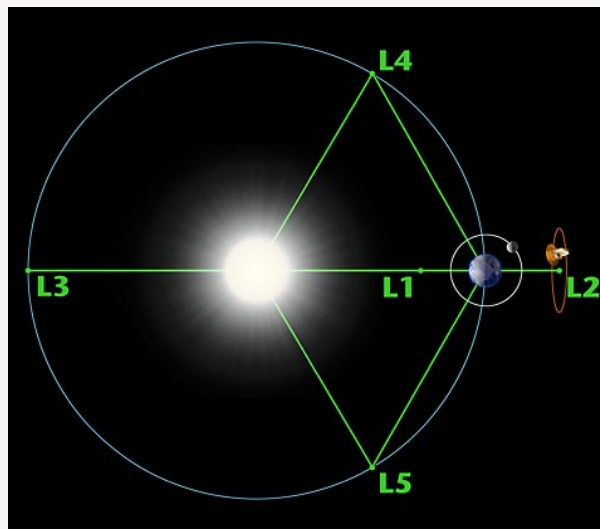
Figure 7.1 – Photographies avant et pendant le lancement du télescope spatial *EUCLID*. (Photographies réalisées par SpaceX et augmentées en résolution par J.C Cuillandre.)

Il est 17h12, heure française, nous sommes le 1^{er} Juillet 2023 et la fusée Falcon 9 décolle avec succès du pas de tir de Cap Canaveral. A son bord, le télescope européen *EUCLID* embarque pour sa grande aventure sous le regard des scientifiques du monde entier. Nous présentons ici l'histoire et l'importance de ce succès.

EUCLID (European Cosmic all Sky Investigator of the Dark Universe) est une mission de classe moyenne de l'Agence Spatiale Européenne (ESA). Elle s'inscrit comme la seconde mission du programme Cosmic Vision pour la décennie 2015-2025, qui vise à cartographier, grâce aux galaxies, la géométrie de l'Univers sombre en réalisant la première carte 3D du cosmos. Pour ce faire, ce télescope de 1.2m de diamètre observe le ciel depuis l'espace (point de Lagrange L2) avec, en son foyer, deux instruments imageurs : le VIS (Visual instrument) en optique et le NISP (Near-infrared spectrometer and photometer) en proche infrarouge (voir Sect 7.1.2). Ces deux instruments complémentaires sont utilisés pour réaliser un relevé grand champ (15 000 deg^2) et un relevé profond (40 deg^2). Le relevé grand champ d'*EUCLID* devrait couvrir un tiers du ciel, soit la fraction totale du ciel non saturée par la lumière provenant de notre galaxie, ce qui en fait l'un des relevés les plus ambitieux et les plus larges au monde.

Les points de Lagrange

Les points de Lagrange dont fait partie le point L2, font référence à des positions dans l'espace qui sont fixes par rapport à la Terre et au Soleil. Physiquement dans le système solaire, ce sont des positions où les forces de gravité et la force de Coriolis se compensent. On compte ainsi 5 positions possibles, L1 à L5, qui correspondent à une région de l'espace pour laquelle le satellite est à une distance fixe tout en étant dans un mouvement parfaitement maîtrisé par la gravitation. Ces points sont calculés comme les solutions des équations du problème à trois corps entre la Terre, le Soleil et un objet de masse négligeable, qui resterait immobile par rapport aux deux autres.



Le point L2 est tout particulièrement intéressant pour le positionnement des télescopes car il est éloigné du Soleil et de la Terre tout en restant à une distance raisonnable pour faciliter la transmission des données prises. C'est pourquoi ce point a été choisi par de nombreuses missions : WMAP^a (Hinshaw et al., 2013), Planck^b (The Planck Collaboration, 2006), Herschel^c (Pilbratt et al., 2010), Gaia^d (Gaia Collaboration et al., 2016), JWST^e (Gardner et al., 2006), et maintenant *EUCLID*^f (Laureijs et al., 2011).

a. <https://map.gsfc.nasa.gov>

b. https://www.esa.int/Science_Exploration/Space_Science/Planck

c. https://www.esa.int/Science_Exploration/Space_Science/Herschel

d. https://www.esa.int/Science_Exploration/Space_Science/Gaia

e. <https://jwst.nasa.gov>

f. <https://www.euclid-ec.org/>

Les objectifs scientifiques principaux de la mission *EUCLID* sont : l'étude de l'énergie noire et la matière noire, la mesure de l'accélération de l'expansion de l'Univers et l'étude de l'évolution des structures cosmiques. Dans le cas du relevé grand champ, la réalisation de ces objectifs se fait par la photographie de 1,5 milliard de galaxies pour un total d'environ 40 000 clichés du ciel. Ce nombre prodigieux de galaxies va également permettre la détection des amas de galaxies en optique par l'observation directe de ces galaxies. Par ailleurs, la détection des amas de galaxies par l'effet de lentille gravitationnelle faible constitue une sonde supplémentaire de premier ordre au vu de la qualité d'image du télescope. Cela va permettre d'étudier l'évolution des structures cosmiques et estimer le biais de sélection en optique (voir Sect. ??).

7.1.2 . Les instruments et les données lentilles

Le télescope *EUCLID* photographie l'espace à l'aide de deux instruments complémentaires et fonctionnant en simultané en son cœur : le VIS et le NISP. Nous en décrivons ici leurs spécificités.

Composé de 36 capteurs CCDs de 4096×4096 pixels pour une surface de 609 Mégapixels, l'instrument VIS (Copper et al., 2016) observe dans le visible afin de mesurer la forme des 1,5 milliards de galaxies ayant un redshift tel que $0 < z < 3$. Il offre une qualité d'image excellente avec une résolution angulaire de 0,18 arcsec pour un champ de vue de 0.57 deg^2 . La résolution est similaire à celle du HSC (0.16 arcsec) pour un champ angulaire 3 fois plus faible, ce qui est cohérent avec la fraction utilisée de CCDs. Mais ces caractéristiques similaires à HSC sont d'autant plus avantageuses dans le cas d'*EUCLID* qui s'est libéré des contraintes atmosphériques. Depuis le point L2, VIS photographie le ciel à travers la bande large *riz* (550-920 nm) allant du bleu au proche infrarouge afin de photographier les galaxies proches comme éloignées et la déformation de leur image pour en déduire l'effet de lentille gravitationnelle faible.

L'instrument NISP (voir Euclid Collaboration et al., 2022) est un spectrophotomètre à infrarouge dont le but est de mesurer la distance aux galaxies observées. Équipé de 16 détecteurs, il opère dans deux modes simultanés : spectrométrique et photométrique. Le mode spectrométrique opère dans la plage de fréquences du proche infrarouge allant de 1100 à 2000 nm. Cette approche permet de mesurer avec une grande précision la distance à une galaxie à partir de l'analyse de son spectre lumineux. Cependant, cette méthode présente des limites (bruit, sensibilité, temps d'observation) qui réduisent son application à un nombre restreint de galaxies, typiquement quelques millions. En parallèle, le mode photométrique permet d'estimer le redshift pour toutes les galaxies observées à partir de leur intensité lumineuse prise à travers différents filtres. Cette estimation est plus rapide mais moins précise que la spectrométrie. Les premières comparaisons des méthodes de photométrie envisagées pour le NISP (voir Euclid Collaboration et al., 2020) montrent des résultats prometteurs, avec une faible incertitude sur le redshift estimé ($< 5\%$) et un nombre de redshifts catastrophiques de 10%.

Ces deux instruments vont permettre d'estimer la forme des 1,5 milliards de galaxies sources ainsi que leur redshift, essentiel à notre étude. Dans le cas du relevé grand champ, la densité de galaxies sources observée devrait atteindre $30 \text{ gal.arcmin}^{-2}$ pour un champ de $15\,000 \text{ deg}^2$. À ceci s'ajoute une qualité d'image excellente avec une résolution fine d'environ 0,2 arcsec. De telles caractéristiques devraient permettre de réunir les conditions optimales (voir Sect 1.3.4) pour une détection des amas de galaxies par effet de lentille faible. En comparaison aux relevés actuels comme le Hyper Suprime-Cam Subaru Strategic Program ($n_g = 22 \text{ gal.arcmin}^{-2}$ sur 1600 deg^2 , voir Sect 8.1), les caractéristiques bien supérieures d'*EUCLID* devraient permettre de produire le plus grand catalogue d'amas de galaxies détectés par effet de lentille faible.

En résumé, le télescope spatial *EUCLID* devrait être une révolution pour l'étude de l'Univers sombre. Ces deux instruments imageurs, le VIS et le NISP en font l'un des télescopes les plus performants pour la mesure de l'effet de lentille gravitationnelle faible. Par ailleurs, les performances attendues pour ce relevé grand champ du télescope spatial *EUCLID* en font le meilleur relevé de données pour la

détection des amas par l'effet de lentille gravitationnelle faible.

7.2 Le challenge WLC

Afin de se préparer à l'étude des données *EUCLID* pour la production d'un catalogue d'amas détectés par l'effet de lentille faible, il convient de bien définir les méthodes de détection qui leur seront appliquées. Au sein de la collaboration *EUCLID*, le challenge WLC vise ainsi à étudier les différentes méthodes de détection des amas de galaxies par l'effet de lentille gravitationnelle faible à partir des données simulées réalistes. Dans cette section, nous détaillons tout d'abord le contexte du challenge, les caractéristiques des données du challenge puis la mise en place de la procédure d'association pour la validation des résultats.

7.2.1 . Contexte du challenge WLC

Le challenge WLC a débuté en Novembre 2021 avec pour objectif la comparaison des méthodes de détection des amas par effet de lentille gravitationnelle faible. Dans cette sous-section, nous décrivons le contexte dans lequel ce challenge s'inscrit.

Des challenges sont organisés au sein d'*EUCLID*, pour la préparation du traitement des différents objectifs scientifiques (Euclid Collaboration et al., 2019, 2020, 2023). Le challenge WLC s'inscrit ainsi dans un contexte plus global où chaque traitement pour les futures données du télescope est implémenté et analysé en amont. Les challenges sont une étape clef afin d'unifier les idées et les procédures à mettre en place lors de l'application aux données réelles. Nous avons vu qu'un des objectifs scientifiques de la mission est la détection des amas de galaxies en optique. Euclid Collaboration et al. (2019) ont ainsi mis en place un premier challenge pour la comparaison des méthodes de détection des amas en optique sur des données simulées approchant les caractéristiques des données *EUCLID*.

C'est dans ce contexte qu'un challenge pour la détection directe des amas par l'effet de lentille faible a également été organisé au sein du consortium de la mission *Euclid*. L'objectif du challenge WLC est ainsi d'analyser et de définir la ou les méthodes de détection à privilégier. Notons que ce challenge n'intègre pas l'estimation des masses des objets détectés par l'effet de lentille gravitationnelle faible.

Dans la pratique, le challenge WLC s'est déroulé en deux temps. Tout d'abord, différents groupes ont pu appliquer leur méthode de détection sur des données simulées similaire aux caractéristiques des données *EUCLID* sans autre connaissance que les catalogues de galaxies bruitées (voir sous-section ci-dessous). Une fois les détections en aveugle obtenues pour chaque méthode, les catalogues de halos associés aux données d'entrée ont été rendu accessibles.

En m'appuyant sur l'expertise développée durant la première partie de ma thèse j'ai coordonné avec un autre membre du groupe de travail, Alberto Manjón-García l'analyse des résultats en procédant à la comparaison des méthodes et l'évaluation de leurs performances (voir Sect 7.4).

7.2.2 . Caractéristiques du jeu de données

Dans cette sous-section, nous décrivons les caractéristiques du jeu de données d'entrée distribué pour ce challenge. Pour une analyse pertinente, il est essentiel que le challenge WLC reproduise au mieux l'application aux données réelles attendues pour *EUCLID*.

Pour ce faire, il a été fourni à tous participants du challenge un jeu de données généré à partir de la simulation DUSTGRAIN (Giocoli et al., 2017, voir Sect 2.3). Ce jeu de données se compose de 12 catalogues de galaxies sources bruitées couvrant chacun un champ de $10^\circ \times 10^\circ$, soit des champs 4 fois plus grand que ceux utilisés lors de l'analyse conduite sur les chapitres 3 à 6 de ce manuscrit. Ces catalogues contiennent les positions des galaxies RA et Dec, les ellipticités mesurées et le redshift photométrique. Comme précédemment, ces catalogues ont une densité de galaxies de $30 \text{ gal.arcmin}^{-2}$,

avec un redshift allant jusqu'à $z = 3$, suivant la distribution en redshift attendue pour *EUCLID*. Le bruit sur la mesure d'ellipticités pour chacune des galaxies sources suit une loi gaussienne à moyenne nulle et d'écart-type 0,26 (Leauthaud et al., 2007). De même, le redshift des galaxies intègre l'erreur statistique sur la mesure de 5% (Euclid Collaboration et al., 2020) et 10% d'erreurs catastrophiques, conformément aux attentes pour les données photométriques d'*EUCLID*. Le jeu de données comporte également des masques pour approcher l'application à des données réelles. Ces masques simulent $\sim 20\%$ de données manquantes dues à la présence d'objets brillants sur le champ, conformément aux prévisions pour *EUCLID*, .

Ces différents ajouts par rapport aux chapitres précédents nous permettent d'avoir des données simulées qui se rapprochent encore plus des données réelles attendues. En comparaison aux travaux présentés dans les chapitres 3 à 6 de ce manuscrit et conduits sur des données simulées, le challenge nous permet de nous confronter aux effets statistiques qui découlent d'un jeu de données plus réduit. Notons que les principales limites des simulations utilisées pour ce challenge sont que le bruit est parfaitement gaussien et que la distribution des galaxies sources est parfaitement homogène. Nous parlerons de l'impact de ces points dans le chapitre suivant où nous appliquons notre méthode sur des vraies données.

En résumé, pour ce challenge les données d'entrée fournies aux participants sont 12 catalogues de galaxies sources incluant leur position, leur redshift et leur forme, couvrant $10^\circ \times 10^\circ$ avec possibilité d'ajouter un masque intégrant 20% de données manquantes. Ces données contiennent à la fois les incertitudes sur la forme des galaxies, les incertitudes sur le redshift photométrique des galaxies et le niveau de données manquantes attendus pour *EUCLID*.

7.2.3 . Caractéristiques des catalogues de détections

Une fois le jeu de données filtré et les détections identifiées, les participants fournissent un catalogue de détections pour chacun des 12 catalogues de galaxies sources bruitées du challenge. Dans cette sous-section, nous décrivons le format sous lequel les catalogues de détections soumis sont attendus.

Pour ce challenge, il a été demandé aux différents participants de produire obligatoirement un catalogue de détections à partir de chaque catalogue de galaxies sources bruitées et sans application des masques de données manquantes. Il s'agit du seul type de catalogue obligatoire pour participer au challenge et permet de comparer les méthodes dans leur application la plus simple. C'est sur ces catalogues que se focalise notre étude dans la section 7.4.

En outre, il a été proposé aux différents participants de produire des catalogues optionnels qui seront intégrés dans la comparaison des méthodes dans un second temps. Ces catalogues optionnels correspondent à l'application des méthodes de filtrage sur les catalogues de galaxies sources bruitées :

- avec l'application des masques intégrant $\sim 20\%$ de données manquantes, et/ou,
- en considérant une approche tomographique sur les catalogues de galaxies sources bruitées, à partir de 3 intervalles de redshift fournis aux participants : $[0, 0.65[$, $[0.65, 1.05[$, et $[1.05, 3.]$.

En effet, un dernier aspect du challenge qui n'a pas été abordé dans la thèse est l'approche tomographique. Dans le cas de détection des amas par effet de lentille faible, les méthodes peuvent être améliorées à travers une approche tomographique en considérant différents intervalles de redshift pour les galaxies sources (Hamana et al., 2020; Oguri et al., 2021). Hamana et al. (2020) ont montré, en particulier, l'avantage significatif de cette approche pour diminuer l'effet de dilution par les galaxies membres (voir Sect. 2.3) et pour augmenter les performances (nombre de détections sans augmenter les fausses détections) de leur méthode de filtrage.

Ces deux dernières applications étant optionnelles, leur mise en place était laissée librement adaptable par les participants et peuvent être appliquées séparément ou simultanément. Pour le challenge, nous avons développé une procédure tomographique très simple pour notre méthode, qui donnera

lieu à une publication.

Pour tous les catalogues de détections obligatoires et optionnelles, il a été demandé de fournir pour chaque détection :

- La position des galaxies en termes de RA et Dec,
- La taille de l'échelle équivalente à laquelle la détection est identifiée,
- Le rapport signal-sur bruit (SNR) de la détection calculé selon la définition standard du rapport entre l'amplitude du signal à la position de la détection dans la carte filtrée, et l'écart-type du bruit mesurée sur la carte.

Il a été demandé aux participants de fixer des seuils bas afin d'inclure aussi des candidats ayant un SNR faible afin de pouvoir évaluer la méthode pour différentes valeurs de pureté. Dans le cas de l'approche multi-échelle, nous détaillons le calcul du SNR dans la sous-section 7.3.2. Notons que le SNR est un élément important notamment pour la comparaison des méthodes lors du tracé des performances des méthodes en termes de pureté et de complétude (voir Fig 7.4).

En résumé, pour ce challenge, les catalogues produits par les participants sont 12 catalogues de détections identifiés après application de la méthode filtrage avec pour chaque détection la position, l'échelle de détection et le SNR qui servent de base pour la procédure d'association.

7.2.4 . Procédure d'association

Afin de comparer les résultats soumis par les différents participants, la procédure d'association avec le catalogue de halos simulés doit être équitable pour les différentes méthodes afin de ne pas introduire de fausses associations, et donc de biais, lors de la comparaison des résultats. En section 5.1, nous avons montré que la procédure d'association était impactée par deux facteurs, la distance maximale d'association (MMD) et le nombre d'objets dans le catalogue de halo de matière noire.

Dans le cas du challenge nous faisons face aux mêmes difficultés :

- Le **nombre de halos dans le catalogue**

Comme le challenge WLC utilise également la simulation *DUSTGRAIN-pathfinder*, qui inclue un grand nombre de structures sur le champ qui ne correspondent pas à des halos détectables (voir Sect 5.1). Pour réduire le nombre de halos à ceux détectables par l'effet de lentille gravitationnelle faible, il a été décidé d'utiliser une coupure diagonale sur le plan redshift-masse ($z - M_{200c}$). Cette coupure diagonale correspond à

$$\log_{10}(M_{200c}) > 13.3 + 1.049z + 0.489z^2, \quad (7.1)$$

exprimée en $h^{-1}.M_{\odot}$ et définie selon [Andreon and Bergé \(2012\)](#).

Cette approche est similaire à celle proposée en section 5.2 et vise également à sélectionner les halos dont le SNR est suffisant pour permettre sa détection. Le seuil minimal choisi ($10^{13.3}$) pour la masse est à nouveau un compromis entre le nombre de halos au sein du catalogue et leur détectabilité (voir Sect 5.1.2). Sur les 12 réalisations de $10^{\circ} \times 10^{\circ}$, cela permet de réduire le nombre de halos dans le catalogue à une valeur moyenne d'environ 2000 halos par champ. Ce nombre est cohérent avec celui obtenu avec notre propre pré-sélection selon le SNR théorique appliqué sur des champs de taille $5^{\circ} \times 5^{\circ}$ (300-400 halos, voir Sect 5.2.2). Cette coupure diagonale est illustrée dans la figure 7.2 pour une des réalisations.

- La **distance d'association**

En sections 5.1.1 et 5.3, nous avons montré que la distance d'association était critique pour l'analyse des performances des méthodes de détection. Cette étape clef peut entraîner des fausses associations et biaiser les résultats présentés. Pour mettre en lumière l'effet de la distance d'association sur les résultats, il a été décidé d'utiliser deux procédures d'association lors de l'analyse des résultats.

- ★ La première méthode d'association, conduite par Alberto Manjón-García, considère une distance d'association angulaire où les détections sont associées à partir du rayon caractéris-

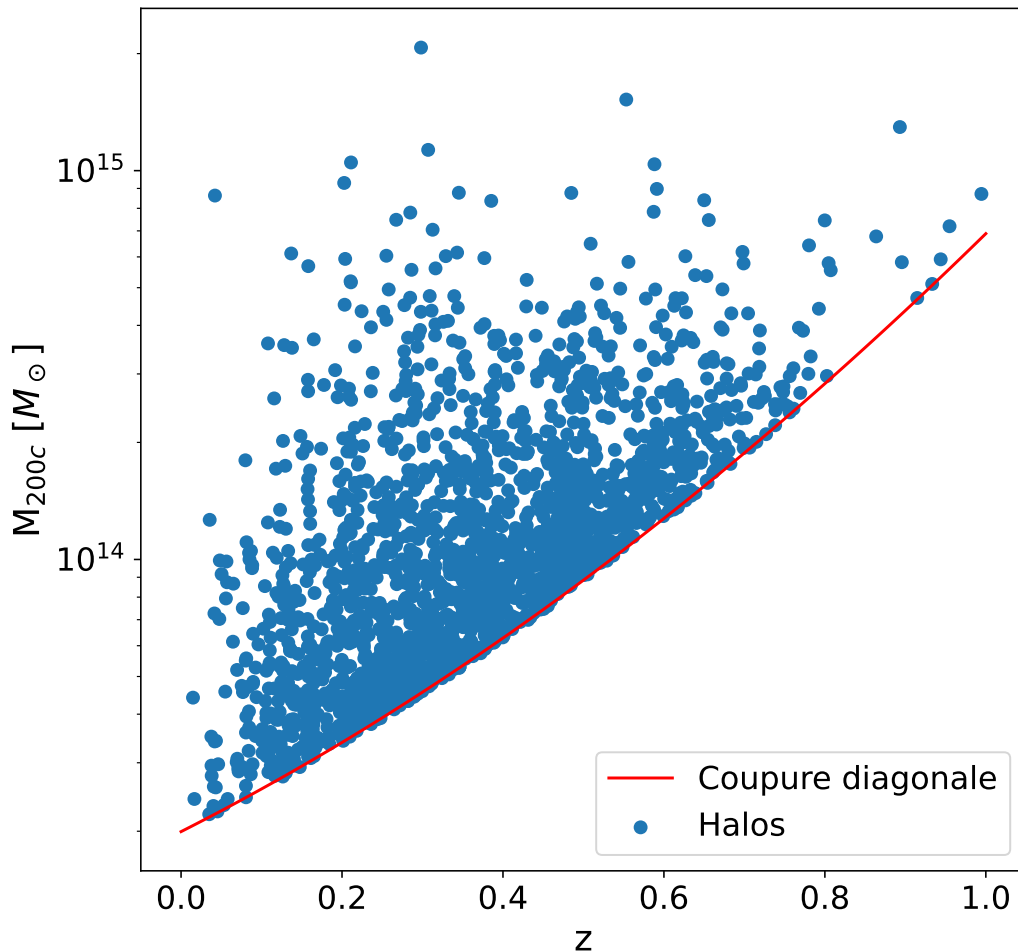


Figure 7.2 – Distribution des halos dans le plan redshift-masse pour un champ de $10^\circ \times 10^\circ$ avec application de la coupure diagonale définie par l'équation 7.1.

tique, R_{200c} , des halos. Dans la pratique, on calcule la distance entre une détection et son halo le plus proche. La détection est ensuite considérée associée au halo si la distance les séparant est inférieure au rayon caractéristique, R_{200c} , donné dans le catalogue de halos. On ne conserve ainsi que les détections à l'intérieur du rayon caractéristique du halo, et sans tenir compte de la taille du filtre. En cas d'associations à plusieurs halos candidats, celui ayant le plus grand rayon est privilégié car c'est aussi le plus massif. N'étant pas de notre fait, les résultats de cette procédure ne sont pas présentés dans ce manuscrit.

- ★ La seconde approche est celle définie en section 5.3 de ce manuscrit où la procédure d'association s'adapte aux spécificités de la méthode de détection utilisée. Cette approche permet de tenir compte des imprécisions sur la position des différents filtres utilisés, en utilisant pour chaque filtre la distance maximale d'association (MMD) qui correspond à la distance d'association optimale vis-à-vis de la fiabilité de la détection. La procédure d'association consiste ensuite à calculer les distances projetées, entre les amas et les détections, et normalisées par le rayon apparent de l'amas associé. Nous sélectionnons alors l'amas présentant la distance normalisée la plus faible si la distance projetée qui correspond est inférieure à la MMD. Cette procédure permet de sélectionner les amas les plus probables de correspondre au signal selon leur redshift et leur rayon.

En résumé, nous avons présenté ici les trois étapes de la mise en place du challenge : le jeu de données, le format des catalogues de détections et la procédure d'association. Comparées aux jeux de données utilisés dans les chapitres précédents, nous avons ici moins de réalisations mais des champs plus larges. La validation des résultats se fait à travers deux méthodes d'association dont une

que nous avons implémentée et décrite dans le chapitre 5 et appliquée dans le cadre du challenge WLC.

7.3 Les différentes méthodes de détection

Dans le cadre du challenge WLC, dix méthodes de détection ont été appliquées par les participants pour produire les catalogues de détections soumis. Dans cette section, nous commençons par introduire ces différentes méthodes et présentons également les nouveautés apportées à l'implémentation de la méthode de détection multi-échelle dans le cadre du challenge. Puis, nous calculons la distance maximale d'association pour chacune des méthodes.

7.3.1 . Présentation des méthodes

Dans cette sous-section, nous présentons succinctement les spécificités des différentes méthodes de détections qui ont été utilisées.

Parmi les dix méthodes utilisées pour le challenge, 4 ont déjà été introduites précédemment dans ce manuscrit : trois méthodes de masse d'ouverture et la méthode multi-échelle. En effet, les méthodes de détection MO utilisant les filtres Sg6, Jo4 et TANH (voir Sect 3.2.4) et la méthode multi-échelle (voir Sect 4.2) ont été appliquées sur le jeu de données du challenge.

Ainsi, les méthodes de filtrage appliquées pour le challenge sont :

- Le **filtre Sg6** proposé par Gavin Leroy et détaillé dans le chapitre 3 (voir Sect. 3.2.4) et en suivant la procédure de détection détaillée en section 3.3.3.
- Le **filtre TANH** proposé par Gavin Leroy et détaillé dans le chapitre 3 (voir Sect. 3.2.4) et en suivant la procédure de détection détaillée en section 3.3.3.
- Le **filtre Jo4** proposé par Gavin Leroy et détaillé dans le chapitre 3 (voir Sect. 3.2.4) et en suivant la procédure de détection détaillée en section 3.3.3.
- La **méthode multi-échelle W234** proposée par Gavin Leroy et suivant l'implémentation décrite dans le chapitre 4 (voir Sect. 4.2).
- Le **filtre optimal adaptif** (AMICO-WL, Optimal matched filter) proposé par Matteo Maturi et Gabriele Panebianco.

Cette méthode, baptisée AMICO-WL, a été premièrement introduite par [Maturi et al. \(2005\)](#) puis affiné selon le principe utilisé pour l'algorithme de détection optique AMICO ([Bellagamba et al., 2018](#); [Maturi et al., 2019](#)), qui est ici adapté pour l'effet de lentille gravitationnelle faible. De manière générale, l'approche repose sur la combinaison d'un modèle décrivant les caractéristiques d'observation des amas ciblés auquel s'ajoute le calcul du spectre de puissance du bruit mesuré directement sur les données.

Dans la pratique, le filtre utilisé pour convoluer les données est optimisé itérativement pour maximiser l'estimation du rapport signal sur bruit des détections. Pour l'estimation du signal, le filtre s'appuie en particulier sur le profil radial moyen du cisaillement d'un halo suivant un profil NFW ([Navarro et al., 1996](#)) de masse $M = 1 \times 10^{15} h^{-1} M_{\odot}$. Dans le cas du bruit, le spectre de puissance du bruit est estimé selon les trois composantes qui y contribuent : le bruit de forme intrinsèque, le bruit de fond poissonien des galaxies et le signal des structures à grande échelle.

Additionnellement, seules les galaxies dont le redshift est supérieur à $z > 0.6$ sont utilisées. Cette sélection des sources d'arrière-plan minimise la dilution du signal par les galaxies présentes dans l'avant-plan.

- Le **MRLens** proposé par Sandrine Pires.
Cette approche utilise la procédure décrite dans [Starck et al. \(2011a\)](#) pour filtrer les cartes de

convergence et se révèle similaire à la méthode multi-échelle dans la mesure où elle repose sur la décomposition en ondelettes (Starck et al., 2006). Les catalogues de cisaillement bruités sont pixellisés en cartes 1024×1024 à partir desquelles on dérive les cartes de convergence associées, par la méthode d'inversion de masse de Kaiser and Squires (1993). Sur la carte de convergence en mode E reconstruite, une décomposition en ondelettes est appliquée. Dans la pratique, chaque carte ondelette est ensuite filtrée en fixant le taux de fausses détections, puis les cartes d'échelles seuillées sont ensuite sommées. Les détections sont faites sur la carte sommée qui en résulte. Notons que les différences avec la méthode multi-échelle W234 sont le seuillage des cartes en fonction d'un taux de fausses détections, sans utiliser les cartes filtrées en modes B et la sommation des cartes avant le seuillage.

- Le **filtrage gaussien multiple** (MSGF, Multiple-smoothing Gaussian filter) proposé par Sandrine Pires.
Ce filtrage suit la logique des premières approches de la détection (e.g. Miyazaki et al., 2002; Gavazzi and Soucail, 2006; Miyazaki et al., 2007; Shan et al., 2012), où différentes fonctions gaussiennes filtrent les cartes de convergence. Dans cette méthode, les cartes de convergence bruitées sont reconstruites comme décrit en section 3.3.3 où les données de cisaillement sont redistribuées sur une grille régulière par la méthode de pixellisation SUM puis inversées par la méthode de Kaiser and Squires (1993). Ces cartes de convergence sont ensuite convoluées par 4 filtres gaussiens d'écart-types respectifs ($\sigma_G = 1.17', 2.34', 4.68'$ et $9.37'$). Ces filtres sont inspirés de l'approche ondelette et visent à sélectionner des amas/halos de taille différentes. Cette méthode a été soumise comme base pour la comparaison mais tout en anticipant un grand nombre de fausses détections, comme précisé en section 3.2.2.
- Le **filtre de masse d'ouverture O21** (O21 Mass Aperture filter) proposé par Masamune Oguri. Cette approche est conforme à la méthodologie développée par Oguri et al. (2021) où les cartes de cisaillement sont convoluées par le filtre spatiale 'TIO5', qui correspond au filtre S96, défini par Schneider (1996) et également introduit dans ce manuscrit dans la section 3.2.3. Pour cette étude, le filtre est paramétré tel que $\nu_1 = 0.025$ et $R = 20'$. Oguri et al. (2021) estiment que ces paramètres permettent de définir un filtre optimal pour la détection, capable de maximiser le niveau de signal sur bruit des amas massifs tout en réduisant la contribution des LSS. Dans la pratique, les cartes de masse d'ouverture sont reconstruites à partir des cartes de cisaillement convoluées par le filtre TIO5. Ces cartes de cisaillement sont obtenues en répartissant sur une grille régulière les galaxies sources issues du catalogue, en considérant uniquement celles dont le redshift est supérieure à 0.3. Pour réduire le temps de calcul, cette opération est réalisée grâce à la transformée de Fourier. Ceci est d'autant plus intéressant pour le calcul du niveau de bruit qui est estimé en convolant de la même manière 500 cartes de cisaillement pour lesquelles les ellipticités ont été randomisées. Les bords de la carte sont traités en tenant compte de la densité de galaxie.
- Le **UPCT** proposé par Alberto Manjón-García et Anastasio Díaz-Sánchez. Cette approche consiste dans un premier temps à lisser les catalogues de cisaillement bruités par un filtre gaussien de taille angulaire $\sigma = 1'$, en suivant la méthode définie par Seitz and Schneider (1995). Une fois ce lissage gaussien réalisé, on reconstruit les cartes de convergence associées par l'inversion de masse de Kaiser and Squires (1993) en considérant une résolution de $0,8'$ soit une taille de 750×750 pixels. Lors de cette reconstruction, un second filtre gaussien de taille angulaire $\sigma = 1'$ est appliqué sur les données afin d'obtenir des cartes de convergence filtrées sur lesquelles les détections sont identifiées. Cette méthode se rapproche des méthodes MO en sélectionnant une bande de fréquence dans l'espace de Fourier.

Une dernière méthode de détection, nommée WSLAP+ a été utilisé pour le challenge WLC mais ne fait pas partie de ce chapitre car non disponible dans sa version finale lors de l'écriture.

Une fois chacune des méthodes de filtrages appliquées, chaque groupe a fourni une liste de dé-

tections pour chacune des réalisations. Pour toutes les approches, les détections sont considérées à partir des maximums locaux présents sur les cartes filtrées. Ces détections sont caractérisées par leur position sur les champs (RA et Dec), l'échelle angulaire à laquelle la détection est faite et le SNR mesuré pour la détection. Les différents groupes ont été libres de définir le seuil minimal pour leur catalogue de détection puis de communiquer le seuil qu'il juge optimal pour leur méthode. Cette liberté de choix sur un seuil minimal permet d'étendre le catalogue de détection à des détections à plus faible SNR afin d'améliorer la comparaison fine des méthodes.

En résumé, 10 méthodes ont été appliquées dans le cadre du challenge WLC et 9 sont comparées dans ce manuscrit.

7.3.2 . Nouveautés pour l'approche multi-échelle

En comparaison à ce qui est décrit dans le chapitre 4, la méthode multi-échelle appliquée pour le challenge présentent deux nouveautés : des seuils observés plus élevés et le calcul du SNR des détections. Dans cette sous-section, nous détaillons ces deux points dans le cadre du challenge.

La première différence que nous observons lors de l'application de la méthode multi-échelle dans le cas du challenge est une augmentation des seuils optimaux pour chaque échelle. Pour rappel, ces seuils sont calculés à partir de la moyenne des maxima mesurées sur les différentes cartes de convergence de mode B filtrées. Dans les chapitres précédents, cette moyenne est réalisée sur 256 cartes de convergence filtrées de taille $5^\circ \times 5^\circ$ (voir Sect. 4.2.3). Dans le cas du challenge, sur douze cartes ($10^\circ \times 10^\circ$) issue de la même simulation, nous mesurons les seuils optimaux de 4.81 (au lieu de 4.57) pour W2, 4.57 (au lieu de 4.35) pour W3, et 4.16 (au lieu de 4.04) pour W4. Cette observation est également vérifiée pour les autres filtres Sg6, TANH et Jo4. Pour autant, cette différence n'a pas d'impact direct sur les résultats et les analyses de ce chapitre car le seuil pour le challenge doit être fixé artificiellement plus bas afin de permettre la comparaison des méthodes à différentes valeurs de pureté. Cette augmentation de $\sim 5\%$ en moyenne observée sur les seuils peut s'expliquer par la taille des champs qui est plus grande ($10^\circ \times 10^\circ$) et le faible nombre de réalisations. Cette observation introduit le caractère instable de l'utilisation du maximum comme seuil en dehors de conditions optimales : taille des cartes adaptée et un grand nombre de réalisations. Nous détaillons ce point dans le chapitre suivant (voir Sect. 8.3.3) où il prend toute son importance.

La seconde nouveauté concerne l'ajout du SNR au sein des catalogues de détection produits. Pour rappel, la méthode multi-échelle décompose les cartes de convergence par les différents filtres ondelettes ; puis les seuille ; et identifie les détections à chaque échelle individuellement. Les détections obtenues à chaque échelle sont regroupées pour ne garder que la position correspondant à l'échelle la plus fine. En traitant les cartes filtrées individuellement, cette procédure nous permet de garantir une plus grande précision sur les détections. Cependant, le signal est également décomposé entre les échelles ce qui complexifie grandement le calcul d'un SNR pour les détections communes et constitue l'inconvénient majeur de la méthode multi-échelle. À cela s'ajoute la décroissance du niveau de bruit au sein des cartes filtrées pour les échelles les plus larges.

Nous avons fait une première tentative pour l'estimation du SNR de chaque détection. Elle consiste à attribuer un SNR égal à une combinaison linéaire des SNR de chaque échelle. Les facteurs de pondérations sont définis en tenant compte du caractère dyadique des ondelettes. Cette méthode ne tient malheureusement pas compte de l'étape de recombinaison et de ce fait n'est pas toujours un estimateur fiable.

L'estimation du SNR d'une détection par la méthode multi-échelle est une tâche délicate. Pour une future publication, il est prévu de calibrer le SNR de chaque échelle à partir de la pureté mesurée avant l'étape de recombinaison.

Table 7.1 – Estimation de la distance maximale d’association (MMD) pour les différentes méthodes d’association du challenge WLC, évaluée en utilisant la procédure décrite en Sect 5.3.1.

Filtre	σ	$ \mu $	MMD [arcmin]
AMICO-WL	0.42	0.02	2.1
Jo4	1.09	0.14	5.5
MRLens	0.56	0.16	2.9
MSGF	0.73	0.16	3.8
O21	0.70	0.03	3.5
S96	0.80	0.16	4.3
TANH	0.51	0.16	2.7
UPCT	0.71	0.22	3.8
W2	0.36	0.18	2.0
W3	0.58	0.15	3.1
W4	1.04	0.04	5.2

7.3.3 . La distance maximale d’association MMD

Comme nous l’avons précisé dans le chapitre 5, la comparaison des méthodes est une tâche complexe. Pour éviter de désavantager ou d’avantager une méthode par rapport à une autre, il est essentiel d’en comprendre les spécificités, en particulier lors de la procédure d’association. Dans cette sous-section, nous calculons pour chaque filtre la MMD correspondante en suivant la procédure définie dans la section 5.3.1.

Nous étudions et comparons les différents filtres à travers leur imprécision sur la position des détections à travers leur MMD. Cette donnée est essentielle pour assurer une procédure d’association juste dans le cadre du challenge car elle dépend des caractéristiques des différents filtres. Pour calculer les MMD de chacune des méthodes de détection, nous avons utilisé directement les catalogues de détections fournis par les différents groupes pour le challenge WLC. Puis, nous avons appliqué la procédure décrite en section 5.3.1 qui permet d’estimer la MMD pour chaque filtre. Nous présentons les MMD calculées pour chacune des méthodes dans la table 7.1 à partir des 12 catalogues de galaxies sources distribués pour le challenge.

Par rapport aux valeurs de MMD présentées dans la table 5.1 pour les méthodes MO et multi-échelle, (S96 :4.4, TANH :2.6, Jo4 :4.7, W2 :1.7, W3 :2.5, et W4 :4.7), nous constatons des écarts avec les valeurs de MMD calculées dans le cadre du challenge. Nous mesurons ainsi un écart relatif pour les filtres S96, TANH d’environ 1.3%, entre la MMD sur les 256 réalisations de taille $5^\circ \times 5^\circ$ et la MMD sur les 12 réalisations du challenge WLC de taille $10^\circ \times 10^\circ$. Dans le cas de W2 et W3, nous calculons un écart de 17% et de 24% qui s’expliquent par le faible nombre de détections dans le cas des échelles les plus fines. L’écart est également marqué aux grandes échelles comme le montre les filtres Jo4 et W4 pour lesquels on mesure un écart relatif de 19% et 10%. Ces écarts peuvent être expliqués majoritairement par le nombre de détections plus faibles dans le cas du challenge WLC qui augmente les incertitudes sur les mesures de moyennes et d’écart-types. En effet, pour le challenge, nous utilisons un jeu de données cinq fois plus petit que celui utilisé dans la première partie de ce manuscrit ce qui réduit la précision de la MMD calculée.

La figure 7.3 montre la distribution des détections associées et non-associées selon la MMD pour la méthode de filtrage AMICO-WL (haut) et la méthode de filtrage Jo4 (bas). Similairement à la figure 5.4, la distribution des détections associées et non-associées (gauche) est présentée en termes de distance d’association, D , en fonction de cette même distance D normalisée par le rayon caractéristique apparent du halo associé, θ_{200c} . Cette distance d’association, D , est ensuite projetée sur les axes RA et Dec (droite). On constate que les deux groupes d’associations que l’on s’attend à avoir sont moins facilement dissociables. Dans le cas de la distribution des détections associées du filtre Jo4, les

deux groupes sont à peine discernables et révèlent clairement le manque de statistiques dû au faible nombre de détections. Ainsi, cela explique la valeur de la MMD du filtre Jo4 qui, calculée dans le cas du challenge WLC, est bien supérieure à la valeur calculée sur les 256 jeux de données simulées des chapitres précédents. Plus généralement, il apparaît que les filtres correspondant à une plus grande échelle ont besoin de plus de statistiques pour distinguer les deux et avoir une mesure précise de la MMD. Le manque de cartes dans le cas du challenge WLC entraîne ainsi une surestimation de la MMD associées au filtre qui s'accroît avec le rayon caractéristique du filtre.

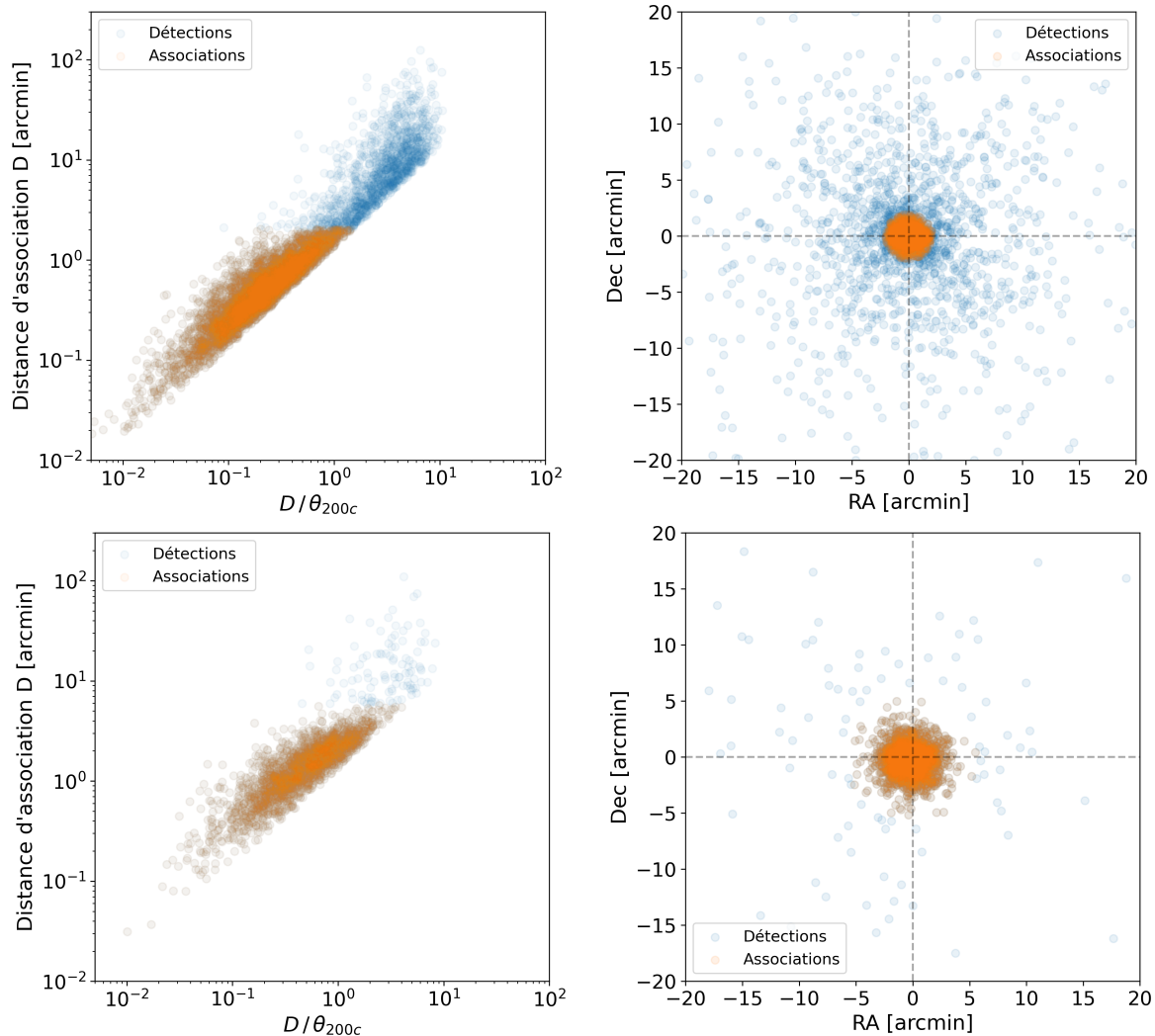


Figure 7.3 – Distribution des détections associées pour le filtre AMICO-WL (haut) et le filtre Jo4 (bas). Les panneaux sont similaires à ceux de la Fig. 5.3. Les points bleus correspondent aux associations entre une détection et le halo le plus proche. Les points orange se réduisent aux mêmes associations, mais dont la distance d'association absolue est inférieure à la MMD définie pour le filtre AMICO-WL.

Dans cette section, nous avons introduit 9 des 10 méthodes appliquées pour le challenge WLC avec leurs différentes implémentations. Nous avons focalisé notre attention sur les points spécifiques à l'application de la méthode multi-échelle au jeu de données du challenge : des seuils plus élevés et le calcul du SNR. Ensuite, pour toutes les méthodes, nous avons calculé la MMD pour la procédure d'association pour chaque méthode appliquée sur le jeu de données du challenge WL. Le manque de statistiques introduit des écarts entre les MMD calculées dans le cas du challenge WLC et les valeurs mesurées dans le chapitre 5. Cette observation est à prendre en compte lors de l'analyse des premiers résultats de la comparaison entre les méthodes.

7.4 Premiers résultats

Après avoir défini la seconde procédure d'association mise en place dans le cadre du challenge et la MMD associée à chacune des neuf méthodes implémentées dans le cadre du challenge, nous analysons leurs associations. Les résultats que nous présentons dans cette section ne concernent que l'application des méthodes sur le jeu de données sans les masques. Nous entamons cette analyse par l'étude de l'évolution de la pureté en fonction de la complétude des détections faites par chacune des méthodes. Puis, nous nous intéressons à la répartition des détections associées dans le plan redshift-masse pour chacune des méthodes. Enfin, nous caractérisons la complémentarité des différentes méthodes à travers les détections associées qu'elles présentent en commun.

7.4.1 . Pureté versus complétude

Pour notre analyse, nous utilisons de nouveau les outils statistiques de pureté et de complétude présentés en section 6.1. Dans cette sous-section, nous évaluons et comparons les différentes méthodes soumises pour le challenge à travers ces deux outils de mesure.

Après application de la procédure d'association définie en section 7.2.4, nous commençons notre analyse par les mesures de pureté et de complétude pour chacune des méthodes. Pour rappel, la pureté est le rapport entre le nombre de détections associées et le nombre de détections totales, tandis que la complétude correspond au rapport entre le nombre de halos associés et le nombre de halos dans le catalogue utilisé pour la procédure d'association. Le nombre de halos associés et le nombre de détections associées étant les mêmes pour notre procédure d'association, ces deux quantités sont directement reliées mais de manière non linéaire. C'est pourquoi toutes deux sont nécessaires afin d'analyser les performances des différentes méthodes.

★ Table des résultats des différentes méthodes.

Nous présentons les mesures moyennes et les incertitudes de ces deux rapports dans la table 7.2 pour les différentes méthodes, calculés à partir des 12 catalogues de galaxies du challenge WLC. Pour la comparaison des méthodes dans la table 7.2, nous conservons les détections au-dessus d'un seuil optimal permettant d'atteindre 90% de pureté pour chaque méthode à l'exception de la méthode MSGF. En effet, la méthode MSGF montre un nombre de détections, une pureté et une complétude bien plus faible que les autres méthodes et ne parvient pas à atteindre 90% de pureté pour un nombre raisonnable de détections. Comme nous l'avons précisé en section 6.1.1, la valeur de la complétude dépend grandement du catalogue de halos et de sa taille ce qui explique que la complétude soit relativement faible pour toutes les méthodes. Cependant, les méthodes W234, J04 et UPCT se distinguent par une complétude plus élevée. Notons que les incertitudes sur la moyenne calculées sur les douze champs de $10^\circ \times 10^\circ$ sont légèrement plus élevées comparée aux résultats du précédent chapitre (voir table 6.2) du fait du nombre réduit de réalisations pour le jeu de données du challenge WLC.

Pour chacune des méthodes, nous calculons également le taux d'associations hasardeuses (TAH), à savoir le taux d'associations qui sont faites par chance ce qui avantage certaines méthodes. Pour estimer l'impact du TAH sur les performances des méthodes, nous simulons un nombre de détections, égal au nombre de détections obtenues par chacune des méthodes, avec des positions aléatoires et nous calculons la pureté obtenue pour ces détections. Les valeurs du TAH pour les différentes méthodes sont ainsi présentées dans la dernière colonne de la table 7.2 et nuancent les résultats des différentes méthodes. Ces valeurs sont impactées par le nombre de halos dans le catalogue de halos de matière noire, par le nombre de détections obtenues par la méthode et par la valeur estimée de la MMD. Dans la pratique, nous observons que la majorité des méthodes ont un TAH proche de $\sim 20\%$, ce qui est cohérent avec les valeurs obtenues dans la section 6.3.1 à l'exception de J04 dont le TAH atteint $\sim 34\%$. Il est intéressant de noter que les méthodes ayant une MMD faible (AMICO-WL, TANH,

MRLens) présentent aussi un TAH plus faible car l'impact de la MMD est significativement réduit.

Table 7.2 – Classification des performances moyennes avec incertitudes des méthodes sur un champ. Les meilleurs résultats de chaque colonne sont mis en évidence en gras. Les incertitudes sont calculées sur les douze cartes.

Méthodes	Détections	Associations	Pureté [%]	Complétude [%]	TAH [%]
AMICO-WL	193.3 ± 4.5	173 ± 3.7	90.0 ± 0.7	8.7 ± 0.2	7.4 ± 0.5
Jo4	305.0 ± 4.4	275.3 ± 3.1	90.0 ± 0.5	13.7 ± 0.2	34.4 ± 0.9
MRLens	204.0 ± 4.4	183.6 ± 4.5	90.0 ± 0.5	8.8 ± 0.2	9.1 ± 0.5
MSGF	71.0 ± 4.2	58.8 ± 3.6	84.1 ± 1.2	2.9 ± 0.2	19.5 ± 1.6
O21	205.8 ± 5.5	185 ± 4.6	90.0 ± 0.6	9.3 ± 0.2	14.6 ± 0.6
Sg6	199.0 ± 4.4	179.2 ± 3.6	90.0 ± 0.7	8.9 ± 0.2	23.1 ± 1.0
TANH	151.8 ± 4.5	136.4 ± 3.7	90.0 ± 0.7	6.8 ± 0.2	11.6 ± 0.8
UPCT	273.3 ± 4.6	245.3 ± 2.3	90.0 ± 0.5	12.2 ± 0.2	20.5 ± 1.2
W234	320.2 ± 5.1	287.2 ± 4.3	90.0 ± 0.5	14.3 ± 0.2	21.4 ± 0.6

★ Évolution de la complétude en fonction de la pureté.

Pour affiner notre analyse, nous nous intéressons ensuite à l'évolution de la complétude en fonction de la pureté pour chacune des méthodes. Cette approche nous permet de comparer plus précisément le caractère sélectif des différents filtres et leurs performances pour des niveaux de pureté équivalents.

Conformément, à l'approche qui nous a permis de tracer la figure 6.1, dans cette section nous évaluons les différentes valeurs de Pureté-Complétude associées à chaque filtre et représentatives de leurs performances. Ces valeurs dessinent alors l'évolution de la complétude en fonction de la pureté pour une implémentation des filtres donnée. Précédemment sur les données simulées des chapitres précédents (voir Sect. 6.1.2), nous avons la possibilité de produire des catalogues de détections pour différentes valeurs de seuils pour chaque filtre. Ce qui nous a permis de tracer l'évolution de la complétude en fonction de la pureté pour chacune des méthodes.

Dans le cas du challenge WLC, les participants ont fournis des catalogues de détections comprenant notamment les positions des détections, et le SNR associé. Dans le cas des méthodes à un seul filtre, les valeurs de pureté et de complétude associées au filtre sont déterminées par le seuil qui permet de sélectionner les détections avec un SNR supérieur à ce seuil. C'est une relation bijective telle qu'une valeur de seuil donnée correspond à une solution (Pureté-Complétude) unique. En d'autres termes, pour ces méthodes, modifier le seuil de sélection des pics candidats revient à se déplacer le long de la courbe traçant l'évolution de la complétude en fonction de la pureté.

Dans le cas de la méthode multi-échelle, cette approche est plus complexe à mettre en place car le SNR d'une détection passe par une étape de calibration (voir Sect. 7.3.2) et n'est plus directement liée au seuil. De manière générale, dans le cas de l'utilisation de plusieurs filtres et le regroupement des détections communes en termes de positions impactent le calcul de l'évolution de la complétude en fonction de la pureté. C'est pourquoi, pour mesurer l'évolution de la complétude en fonction de la pureté de manière rigoureuse, nous avons fait varier les différents seuils par le même facteur de proportionnalité simultanément et avant l'étape de regroupement des détections communes.

La figure 7.4 montre ainsi l'évolution de la complétude en fonction de la pureté pour les différentes méthodes de détection dans le plan [75, 100] × [0, 35] (gauche) puis en zoomant sur la région [92, 100] × [0, 12] (droite). Ces deux panneaux permettent de comparer les méthodes sur leur évolution générale (gauche) mais aussi à très haute pureté, $P > 92\%$ (droite). Plus une courbe est proche du coin supérieur droit, meilleures sont les performances de la méthode associée. Concernant l'évolution générale, on observe en particulier des performances proches pour les méthodes AMICO-WL, MRLens, O21 et Sg6;

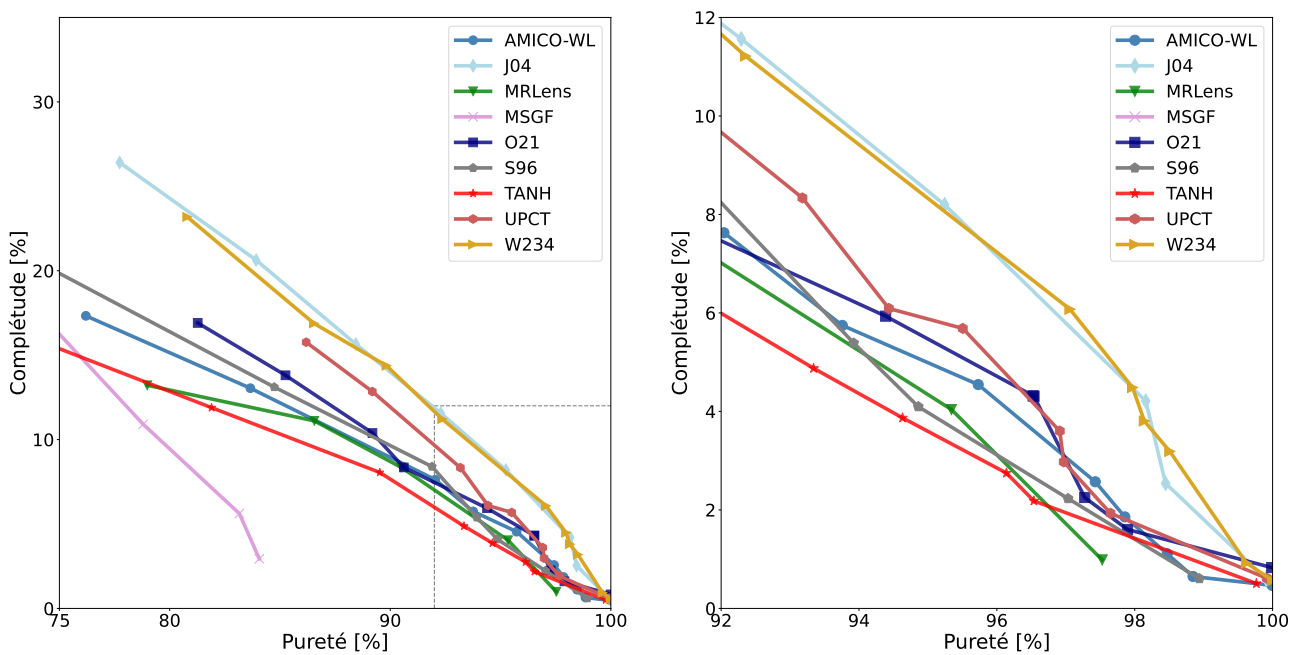


Figure 7.4 – Complétude en fonction de la pureté pour les filtres O21 (bleu foncé), AMICO-WL (bleu), MRLens (vert), UPCT (rouge foncé), MSGF (rose), J04 (bleu clair), S96 (gris), TANH (rouge) et W234 (orange) pour une pureté > 75% (gauche) et zoomé sur pureté > 92% (droite). L’encadré noir sur le panneau de gauche indique la région zoomée illustrée par le panneau de droite.

pour les méthodes J04 et W234. La méthode MSGF montre les performances les plus faibles avec une complétude qui chute à 3% pour une pureté P supérieure à 85%, soit une complétude 4 à 8 fois plus faibles que les autres approches. Cela semble indiquer pour cette méthode qu’un nombre significatif de détections sont des fausses détections ayant pour autant un SNR élevé. À l’inverse, les méthodes présentant les meilleurs résultats sont W234 et J04 avec une complétude plus grande pour toute valeur de pureté.

Dans le cas des puretés élevées, les méthodes J04 et W234 sont très proches et présentent de nouveaux les meilleurs résultats suivis par UPCT puis O21. Nous observons une chute de la complétude plus marqué pour les méthodes MRLens et TANH à pureté élevée. Notons que l’estimation des complétudes et puretés au-delà de $P=99\%$ est assez imprécise car le nombre de détections est très faible (environ une dizaine).

Dans la figure 7.4, la similarité entre les performances des méthodes J04 et W234 peut sembler surprenante étant donné que le filtre W_4 utilisé dans la méthode multi-échelle est presque identique au filtre J04 (voir Fig. 4.5). Ainsi, on s’attend à ce que les détections de J04 soient incluses en très grande partie dans les détections de W_4 , amenant à des résultats par définition inférieure. Cette observation est également appuyée par la comparaison entre W234 et J04, dans la section 6.1 qui montre que la méthode W234 est théoriquement égale ou plus efficace que J04 car elle intègre toutes ses détections. Ceci reste vraie dès lors que les conditions d’applications des filtres J04 et W_4 sont les mêmes. En effet, les résultats de la méthode J04 peut être biaisée par sa MMD artificiellement augmentée car plus imprécise sur le jeu de données restreint. Cette MMD plus large entraîne aussi un taux d’associations hasardeuses plus grand, comme le prouve la dernière colonne de la table 7.2. Le TAH de la méthode J04 est fortement impacté par sa MMD élevée ce qui peut expliquer l’écart marqué avec la méthode W234.

En résumé, la similarité que l’on observe entre les courbes de J04 (ligne bleu clair) et W234 (ligne jaune) peut être expliquée par la MMD et les associations hasardeuses qui favorisent J04.

★ Évolution corrigée de la complétude en fonction de la pureté.

Bien que par la suite, nous utilisons les valeurs de MMD imprécises estimées à partir du jeu de données du challenge WLC, dans cette sous-section nous étudions l'impact de cette imprécision sur l'évolution de la pureté et la complétude. Afin d'affiner notre analyse et confirmer les performances des différentes méthodes, nous utilisons les valeurs plus précises de MMD pour chacun de nos filtres (S96, TANH, J04, W2, W3, et W4), présentées dans la table 5.1. Bien que cette correction sorte du cadre du challenge WLC, elle nous paraît essentielle pour donner une idée de la comparaison la plus correcte possible entre les méthodes en corrigeant le biais introduit sur la MMD dû au jeu de données plus restreint. Cette comparaison est d'autant plus pertinente pour conclure cette section que l'objectif premier du challenge vise à termes à définir les méthodes de détections optimales pour traiter les données réelles du relevé *EUCLID* pour lesquelles nous n'aurons pas de catalogues de halos.

Nous avons reproduit la procédure d'association appliquée précédemment sur les catalogues de détections produits par les différentes méthodes en ajustant la valeur de MMD en fonction des valeurs présentes dans la table 5.1. Ces valeurs plus précises nous permettent de corriger les courbes dans les cas des filtres TANH, S96, J04, W2, W3 et W4. Nous avons ainsi repris les différents seuils estimés précédemment et tracé l'évolution corrigée de la complétude en fonction de la pureté. La figure 7.5 montre l'évolution de la complétude en fonction de la pureté en utilisant les valeurs de MMD présentées dans la table 5.1. Pour la comparaison, les anciennes performances mesurées dans la figure 7.4 pour nos méthodes (J04, S96, TANH et W234) sont représentées par les courbes en pointillées.

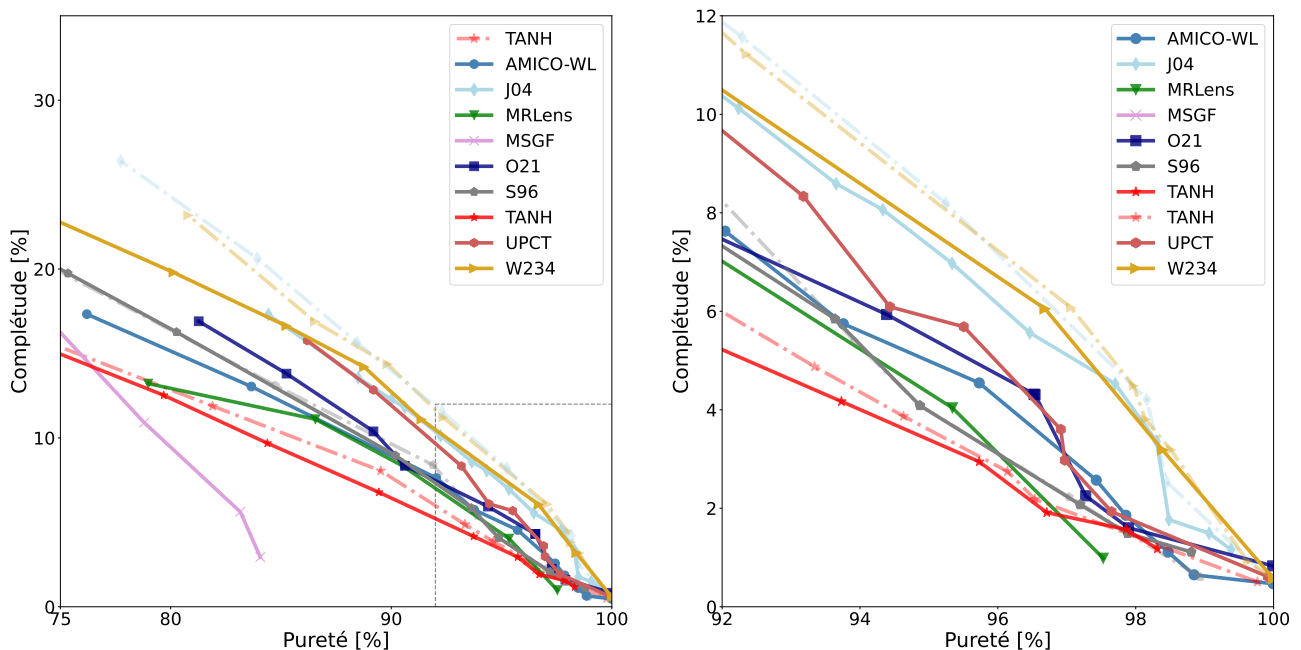


Figure 7.5 – Complétude en fonction de la pureté estimée à partir de la MMD corrigée pour les filtres O21 (bleu foncé), AMICO-WL (bleu), MRLens (vert), UPCT (rouge foncé), MSGF (rose), J04 (bleu clair), S96 (gris), TANH (rouge) et W234 (orange) pour une pureté > 75% (gauche) et zoomé sur pureté > 92% (droite). L'encadré noir sur le panneau de gauche indique la région zoomée illustrée par le panneau de droite. Les courbes en pointillés correspondent aux performances présentées dans la figure 7.4 avec les MMD estimées sur les jeux de données du challenge.

Les valeurs de MMD étant surestimées, la correction diminue la complétude mesurée pour une même pureté en comparaison à la figure 7.4. Nous observons désormais que le niveau du filtre J04 (ligne bleu clair) est plus faible que la méthode multi-échelle W234 ce qui confirme l'impact de la MMD dans les résultats observés précédemment. En comparant les valeurs de complétude obtenues pour la même pureté avant et après correction, nous mesurons une diminution de 15% des valeurs mesu-

rées pour le filtre Jo4 et une diminution de 9% des valeurs mesurées pour le filtre W234. Cependant, les méthodes présentant les meilleurs résultats restent inchangées avec W234 en tête, suivie par Jo4 et UPCT qui montrent des résultats très proches. Cette analyse nous permet de confirmer les observations faites sur les jeux de données du challenge WLC malgré le biais introduit sur les MMDs.

En résumé, les mesures de pureté et de complétude nous permettent de mettre en lumière les différents caractères sélectifs des différentes méthodes appliquées au challenge. On distingue ainsi que la méthode MSGF présente les résultats les moins convaincants en termes de détections, de pureté et de complétude. A l'inverse, les méthodes Jo4, O21 et W234 présentent de meilleures performances en termes de complétude et de pureté. Nous avons complété cette analyse par l'évolution de la complétude pour différentes valeurs de pureté afin de comparer au même niveau toutes les méthodes. Il en ressort que, comparées aux autres, les méthodes W234, Jo4, et UPCT ont une complétude significativement plus grande pour une même pureté. Pour la suite, nous considérons les détections associées par la MMD calculée sur les données propres au challenge WLC (sans correction).

7.4.2 . Détections associées dans le plan z-M

Nous avons vu que ces méthodes présentent des caractères sélectifs différents. Nous étudions maintenant plus en détails les effets de sélections de chaque méthode en regardant les halos détectés. Dans cette sous-section, nous analysons les distributions des halos détectés dans le plan redshift-masse pour chacune des méthodes. Puis, nous affinons cette analyse en décomposant le plan redshift-masse en différents intervalles de redshifts et de masses, au sein desquels nous calculons une complétude locale.

★ Détections associées pour une réalisation dans le plan z-M.

Pour étudier la sélectivité des méthodes vis à vis des halos du champ, nous observons la distribution des halos détectés dans le plan redshift-masse pour une pureté donnée de 90%. Nous affichons ainsi pour les différentes méthodes la répartition des associations dans la figure 7.6. Pour la comparaison, nous utilisons les mêmes seuils que pour la table 7.2 afin d'obtenir un niveau de pureté similaire. Les halos présents dans le catalogue y sont représentés par les cercles bleus et les croix orange correspondent aux halos associés pour la méthode considérée. La majorité des approches parviennent à détecter les halos les plus massifs à bas redshifts, qui sont ceux pour lesquels l'effet de lentille faible est le plus sensible. Comme attendu, la méthode MSGF est en dessous des autres méthodes en termes de résultats mais retrouve la majorité des halos massifs malgré son faible nombre de détections. Il est intéressant de noter que les halos à très bas redshift ($z < 0.2$) sont aussi détectés de manière significative par O21 et W234. En particulier, il apparait que les méthodes W234, UPCT et MR-Lens sont plus performantes à haut redshift. Plus globalement, on observe que certaines méthodes (Jo4, MSGF et TANH) ne détectent pas les halos à plus haut redshift alors qu'ils sont détectés par les autres méthodes.

★ Distribution spatiale des détections associées dans le plan z-M.

Nous étudions également la complétude spatiale de chaque méthode pour différents intervalles du plan redshift-masse. Pour ce faire, nous découpons le plan en une grille régulière en masse et en redshift de taille 11×11 , définissant ainsi 121 zones, et nous regroupons les halos appartenant à la même zone. Puis, nous calculons la complétude, dite locale, de chacune de ces zones pour les différentes méthodes. Cette approche permet de compléter la représentation précédente et de distinguer plus en détails les performances moyennes des méthodes en particulier à basse masse et bas redshift mesurées sur les 12 champs.

Les résultats de cette approche sont présentés dans la figure 7.7 où les détections associées des 12 champs sont empilées pour chaque méthode. Ces différents panneaux mettent en avant la répartition spatiale (dans le plan $z - M$) des halos détectés pour une pureté similaire $P \sim 90\%$. Tout d'abord, la méthode AMICO-WL se distingue notamment par une répartition spatiale plus regroupée que toutes les autres méthodes avec un manque de halos détectés à bas redshift et basse masse. A l'inverse,

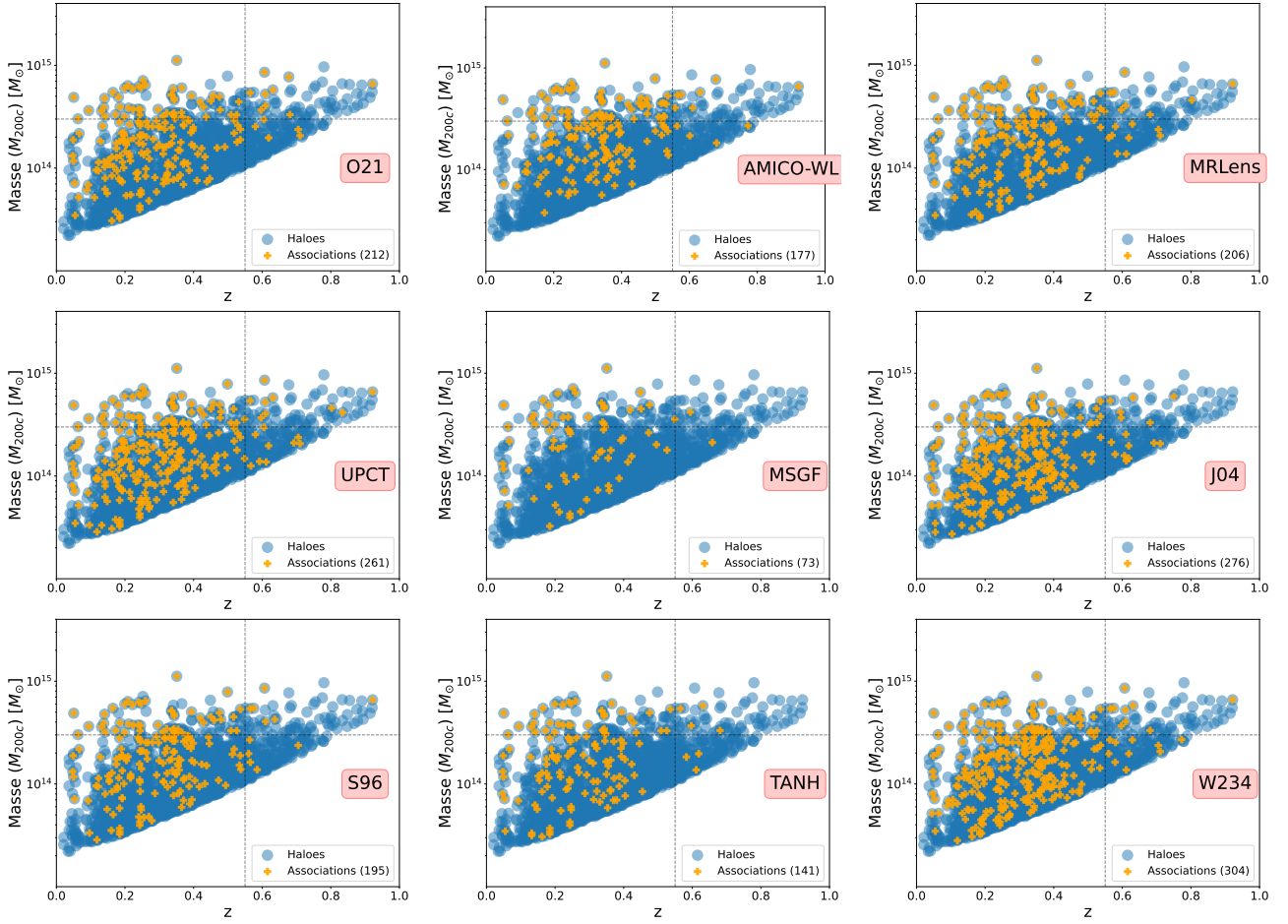


Figure 7.6 – Distribution des halos associés dans le plan $z - M$ pour une réalisation de $10^9 \times 10^9$ pour les différentes méthodes de détection du challenge pour une pureté similaire ($P \sim 90\%$). Les points bleu clair correspondent à tous les halos du catalogue de halos fournis et les croix orange surlignées correspondent aux halos qui ont fait l'objet d'une détection, et entre parenthèses, nous notons le nombre d'associations pour chaque méthode sur cette réalisation. Les lignes en pointillés ($M = 3 \cdot 10^{14} M_{\odot}$ et $z = 0.55$) sont fixées arbitrairement pour mieux se repérer entre les différentes figures.

nous observons que les méthodes W234, UPCT et J04 parviennent à détecter davantage de halos à basse masse et très bas redshift ($z < 0.2$) que les autres méthodes. Les méthodes UPCT, MRLens et O21 présentent les meilleurs résultats en termes de complétude pour les halos très massifs ($M_{200c} > 3 \cdot 10^{14}$) à bas redshift ($z < 0.55$). Ces performances sont presque égalées par les méthodes UPCT et MRLens qui se distinguent en particulier à haut redshift ($z > 0.55$). Cette répartition met ainsi en avant les performances supérieures des méthodes MRLens, UPCT et W234 pour la détection à la fois sur les halos massifs à bas redshift et les halos peu massifs à bas et haut redshift.

Il est intéressant de noter qu'aucune méthode ne surpasse les autres ou n'inclue toutes les autres détections. Ainsi, les méthodes présentent une complémentarité entre elles en termes de détections à la fois pour les halos à bas redshift qu'à haut redshift.

En résumé, l'analyse des associations des différentes méthodes dans le plan $z - M$ pour différents intervalles de redshift et de masse montre une dominance des méthodes O21, UPCT et W234. Ces méthodes se distinguent par leur capacité à détecter les halos massifs en même temps que les halos moins massifs à plus bas redshifts que les autres méthodes ratent. Il ressort également de cette comparaison que les méthodes sont complémentaires entre elles en termes de détections.

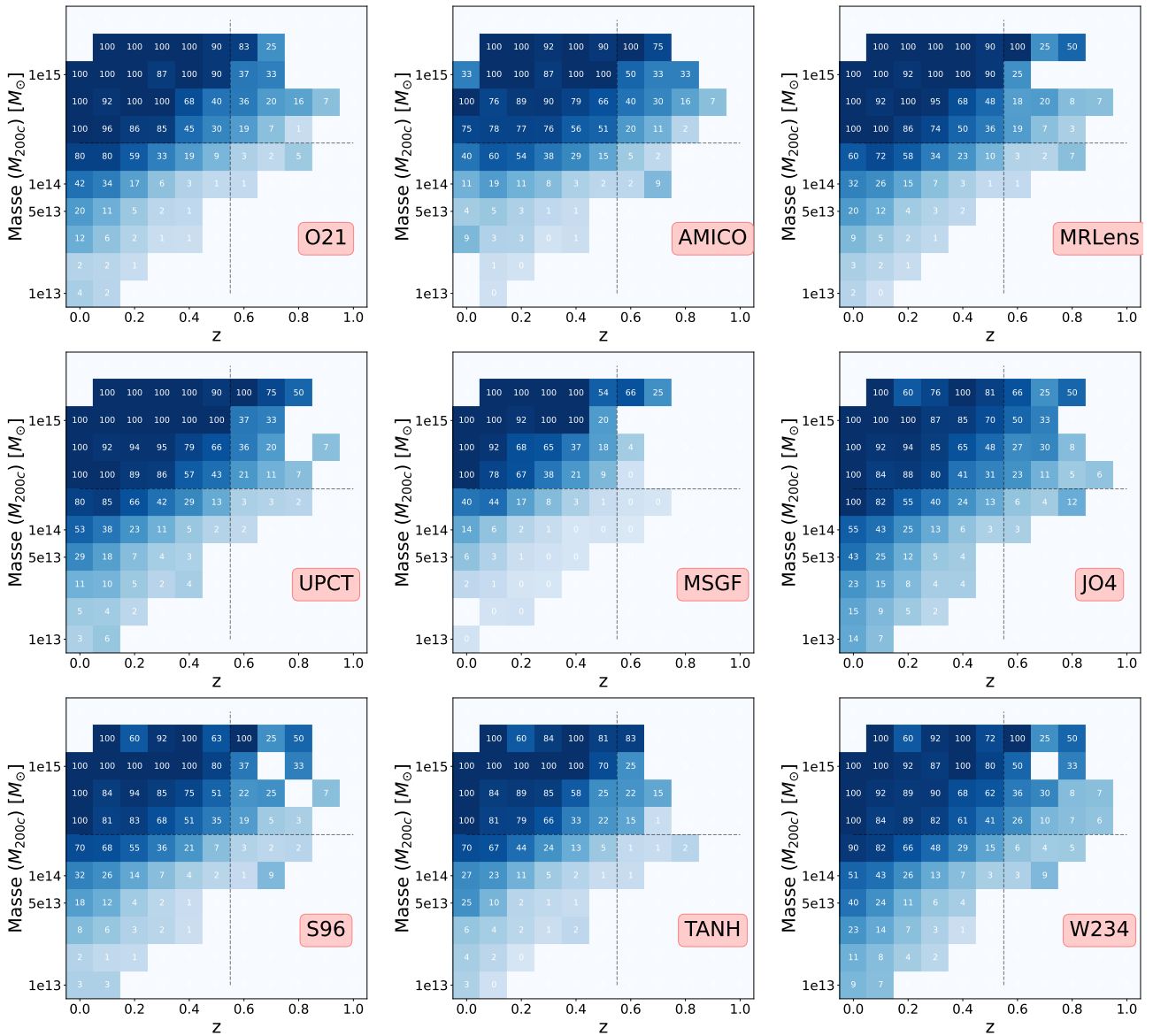


Figure 7.7 – Distribution de la complétude locale dans le plan $z - M$ empilée à partir des 12 réalisations de $10^\circ \times 10^\circ$ pour les différentes méthodes de détection du challenge. Les nombres inscrits dans chaque carré bleu correspondent à la complétude locale (en pourcentage).

7.4.3 . Caractérisation des détections associées communes

Afin de finaliser notre analyse des résultats et pour aller plus loin dans l'étude de la complémentarité des détections associées, obtenues par chacune des méthodes proposées, dans cette sous-section, nous nous intéressons aux associations communes entre les méthodes.

Pour ce faire, nous nous plaçons dans le cas où les seuils sont optimaux pour chaque méthode et mène à une pureté proche de $P \sim 90\%$. Puis, nous calculons le nombre de détections associées qui sont partagées entre chaque méthode que nous normalisons par le nombre de détections totales obtenues par la méthode inscrite en bas de chaque colonne de l'image. Cela nous permet de comparer les associations communes du point de vue des deux méthodes comparées. Cette comparaison est ainsi essentielle pour mettre en avant les complémentarités et les équivalences en termes d'associations des différentes méthodes.

La figure 7.8 présente ainsi la comparaison en pourcentages des associations communes obtenues entre les différentes méthodes. Chaque ligne correspond aux pourcentages d'associations communes avec la méthode de la colonne correspondante. Comme attendu compte tenu de son faible

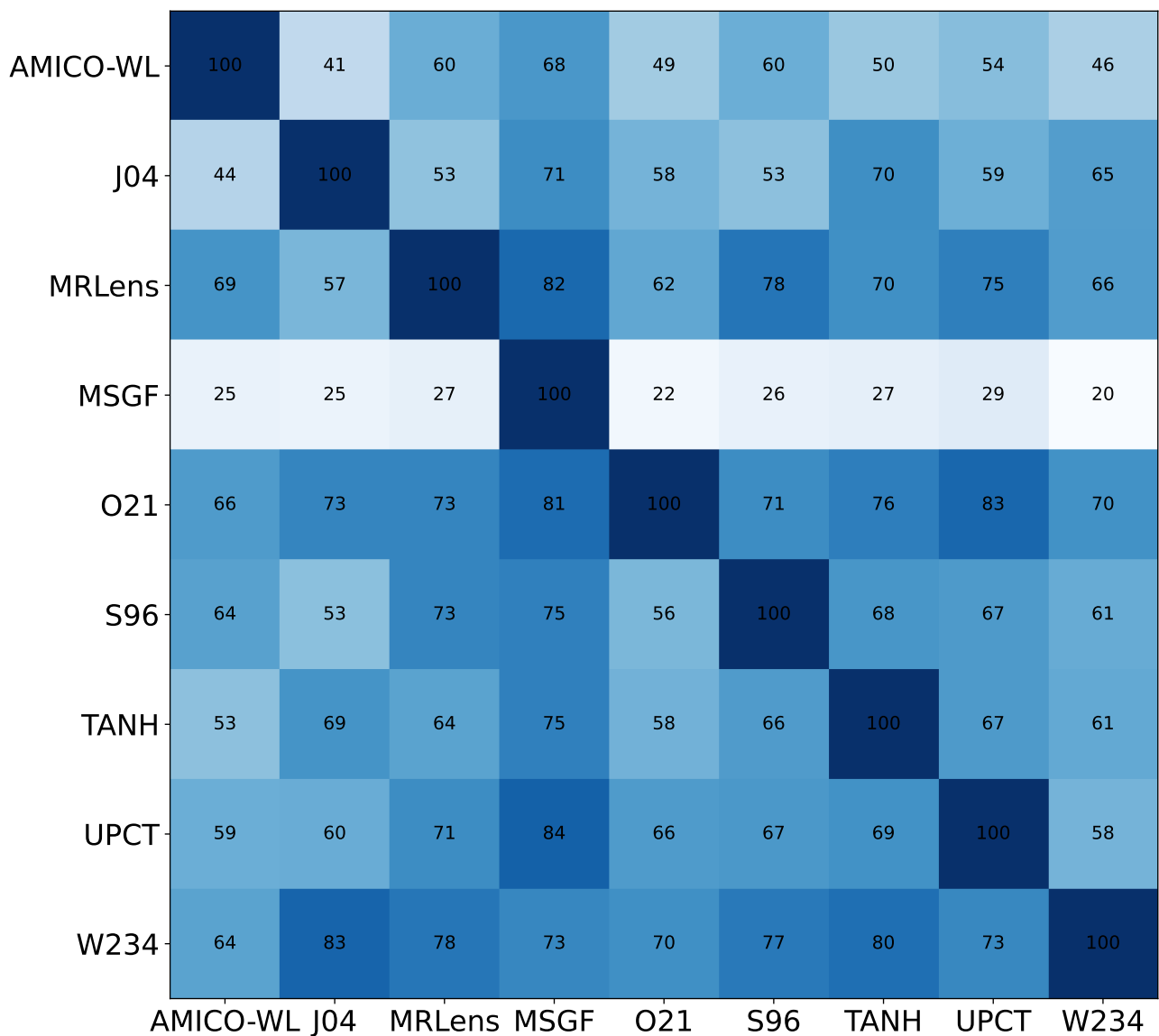


Figure 7.8 – Comparaison en pourcentage des associations communes entre les différentes méthodes utilisées pour le challenge et moyennées sur les 12 champs. Chaque ligne correspond au pourcentage de détections associées qui sont partagées avec chaque méthode inscrite dans la colonne. Ainsi la matrice n'est pas symétrique car les lignes sont normalisées par le nombre de détections de chaque méthode en colonne.

nombre de détections, les associations de la méthode MSGF incluent très peu (~ 30%) de détections associées obtenues par les autres méthodes qui à l'inverse, contiennent jusqu'à 80% des associations obtenues par MSGF. En comparant les autres méthodes, nous observons que la méthode AMICO-WL est celle qui présente le moins d'associations communes avec les autres méthodes alors même que ses associations sont incluses jusqu'à 76% par la méthode MRLens. Nous constatons également que les associations obtenues par les méthodes O21 et W234 permettent de retrouver chacune environ 75% des associations de toutes les autres méthodes. Cette observation nous pousse vers l'idée qu'une combinaison de certaines méthodes pourrait être l'approche la plus efficace pour le traitement des données à venir du télescope *EUCLID*.

En résumé, notre analyse complète met en lumière qu'aucune méthode ne surpasse directement toutes les autres tant sur le plan des détections que de la pureté et la complétude. Certaines méthodes montrent cependant des prédispositions pour la détection des halos soit à moyen et haut redshift soit pour les hautes masses. Leur combinaison nous permettrait d'obtenir une complétude bien plus grande sur les champs observés. Compte tenu de ces résultats, cette analyse nous permet cependant

d'exclure la méthode MSGF du fait de son nombre important de fausses détections qui entraîne une pureté très faible comparée aux différentes méthodes.

7.5 Conclusions

Dans ce chapitre, nous avons détaillé l'application de notre approche multi-échelle dans le cadre de la préparation aux données du télescope européen *EUCLID* par le challenge WLC. Ce challenge, organisé au sein du groupe de travail sur la détection des amas de galaxies par effet de lentille faible, vise à déterminer les méthodes à privilégier pour l'étude des données *EUCLID* à venir. Dans cet optique, le challenge était constitué de 12 champs simulés de $10^\circ \times 10^\circ$ avec des caractéristiques proches de celles attendues pour *EUCLID* avec pour chaque champ : le bruit sur la mesure de forme des galaxies sources, une densité de galaxie proche de 30 galaxies/arcmin, une erreur statistique de 5% sur les mesures de redshift des galaxies sources et 10% de redshift catastrophique. A cela s'ajoutait un masque réaliste de 20% de données manquantes et la donnée de 3 intervalles de redshift de même densité, pour permettre une approche tomographique.

Nous avons activement participé au déroulement de ce challenge :

- Parmi les 10 méthodes implémentées pour ce challenge, nous avons présenté 3 méthodes de détection MO (S96, TANH, Jo4) et l'approche multi-échelle. Pour ces 4 méthodes, nous avons proposé une approche intégrale telle que présentée dans les chapitres précédents (Chapitres 3 à 6). Ces deux implémentations alternatives ont été appliquées dans le cas de données d'entrée avec et sans les masques simulant des objets brillants sur le champ.
- Nous avons également utilisé notre expertise développée sur les données simulées (Chapitres 3 à 6) pour proposer une approche alternative à l'analyse des performances des différentes méthodes. Nous avons ainsi appliqué la procédure d'association sur les 10 méthodes afin de les comparer en termes d'associations, de complétude et de pureté. Nous avons par ailleurs approfondi cette étude par l'analyse des détections communes associées.

Notre participation au sein de ce challenge nous a tout d'abord permis de nous comparer avec d'autres méthodes que l'approche de masse d'ouverture. Comparées au chapitre précédent, les résultats sont obtenus sur un jeu de données plus restreint mais plus réalistes. Notre analyse des résultats a montré tout d'abord que la taille du jeu de données était trop faible pour pouvoir estimer une MMD avec précision. Nous mesurons des erreurs sur la MMD qui peuvent aller jusqu'à 20% sur les méthodes que nous avons déjà étudié dans les chapitres précédents. Cette MMD reste toutefois plus fiable que l'utilisation d'un rayon fixe pour toutes les méthodes mais nos résultats sont à pondérer par ces incertitudes.

Malgré ce biais, les méthodes W234, UPCT et Jo4 se distinguent en termes de complétude. En particulier, nous observons que pour la même valeur de pureté (90%), la méthode W234 montre la complétude la plus élevée (14.3%) en comparaison aux autres méthodes. En allant plus loin, l'évolution de la complétude en fonction de la pureté met en lumière la dominance des méthodes W234, UPCT et Jo4 sur la gamme de pureté que nous avons analysé. Cette dominance se maintient quand on regarde plus en détails les détections dans le plan redshift-masse pour une réalisation. En particulier, les méthodes W234, UPCT et Jo4 sont davantage sensibles aux halos à basse masse et très bas redshift ($z < 0.2$). À l'inverse, pour les halos les plus massifs, la méthode O21 montre les meilleures performances. Enfin, la comparaison des méthodes en termes de détections met en avant que la méthode W234 englobe à elle seule en moyenne $\sim 75\%$ des détections réalisées par les autres méthodes.

Cette première analyse nous permet d'isoler trois méthodes plus performantes que les autres pour la détection des amas de galaxies sur les futures données du télescope *EUCLID* : W234, UPCT, et o21. Le filtre Jo4 présente également des bons résultats mais au vu de sa similarité avec le filtre W4, il peut être remplacé par le filtre multi-échelle. Ce point est confirmé également par le fait que 83% des détections de Jo4 sont incluses dans les détections de W234. Le choix de ces méthodes s'impose

d'autant plus que les méthodes semblent complémentaires et sont capables ensemble de reproduire une grande partie des résultats obtenues par les autres méthodes.

*
* *

8

Application à des données réelles

L'astre d'or brille sur un tableau de couleurs : du bleu, du jaune, puis de l'orange, du rose. Les teintes changeantes surplombent une mer de coton. Figés sur le toit du monde, nous sommes bercés par nulle autre mélodie que le chant du vent. Nous sommes émerveillés, fascinés même, par cette vision de la terre qui s'endort, témoins privilégiés du rituel de la nature.

The Wondering Wanderer, Mauna Kea : la porte des étoiles

8.1	Un bond en avant pour la détection des amas : Les données lentilles HSC-SSP	136
8.1.1	Le télescope Subaru et Hyper Suprime-Cam	136
8.1.2	Des données adaptées pour la détection	138
8.2	Défi de l'application aux données réelles HSC	139
8.2.1	Les données manquantes	139
8.2.2	Le bruit non gaussien	141
8.2.3	Une distribution non uniforme des galaxies sources	142
8.3	Application de la méthode de détection multi-échelle	143
8.3.1	Pixellisation des données	144
8.3.2	Reconstruction des cartes de convergence filtrées et estimation du bruit	146
8.3.3	La procédure de seuillage	148
8.4	Résultats	151
8.4.1	Association des détections	152
8.4.2	Résultats des détections associées	153
8.5	Perspectives	156



Dans les chapitres précédents 4, 5, 6, et 7, nous avons développé et appliqué notre méthode multi-échelle sur des données simulées. Nous en avons quantifié les résultats et les performances dans un cadre idéal et maîtrisé : sans données manquantes, avec une distribution homogène des galaxies sources, avec un bruit parfaitement gaussien et en ayant le catalogue de halos associé pour l'analyse des résultats. Nous appliquons maintenant cette méthode dans le cas de données

réelles obtenues par le relevé Hyper Suprime-Cam Subaru Strategic Program (abrégé HSC-SSP par la suite). Dans ce chapitre, en section 8.1, nous introduisons tout d'abord les données HSC-SSP et leur pertinence pour notre étude. Ensuite, dans la section 8.2, nous commençons par détailler les différents défis qu'apportent l'application à des données issues de télescopes : l'effet des données manquantes, l'impact d'un bruit réaliste (non gaussien), et l'effet de la distribution non homogène des galaxies. Puis, dans la section 8.3, nous décrivons les différentes étapes que nous avons suivi pour faire face à ces défis lors de l'application de notre méthode de détection multi-échelle aux données du relevé. Enfin, nous présentons en section 8.4, nos résultats sur les données HSC issue de la première année d'observation et nous les comparons aux résultats déjà publiés (Miyazaki et al., 2018; Hamana et al., 2020). Des résultats qui s'appuient sur les données des trois premières années ont récemment été publiées (e.g. Dalal et al., 2023; Li et al., 2023; Miyatake et al., 2023; More et al., 2023; Sugiyama et al., 2023); mais ces données ne sont pas encore publiques au moment de l'écriture de ce manuscrit.

8.1 Un bond en avant pour la détection des amas : Les données lentilles HSC-SSP

Comme nous l'avons introduit en section 1.3.4, le relevé Hyper Suprime-Cam Subaru Strategic Program (Aihara et al., 2018a; Li et al., 2022), qui a débuté en mars 2014, est actuellement le relevé ayant permis le plus grand nombre de détections d'amas de galaxies par effet de lentille faible. Dans cette section, nous décrivons plus finement ce relevé et ses caractéristiques techniques. En particulier, nous détaillons ici ses différentes propriétés utiles pour la détection d'amas par effet de lentille gravitationnelle faible.

8.1.1 . Le télescope Subaru et Hyper Suprime-Cam

Au sommet du volcan endormi Mauna Kea à Hawaï siège le télescope optique-infrarouge Subaru, symbole de la collaboration scientifique entre le Japon, Taiwan et les Etats-Unis (voir Fig. 8.1). Hyper Suprime-Cam (HSC¹, Miyazaki et al., 2018) est l'une des caméras de ce géant de 8.2 mètres de diamètre. Le plan focal de cette caméra grand angle se compose de 116 CCDs de 2048 × 4096 pixels soit un total de 870 Mégapixels. Ses 116 CCDs couvrent un champ angulaire de 1.8 deg² de diamètre, pour une résolution effective de 0.168 arcsecondes, offrant ainsi une superbe qualité d'image. A cela s'ajoute une large surface collectrice de photons. Ces caractéristiques font de HSC une des caméras les plus performantes au monde pour l'observation de l'effet de lentille gravitationnelle faible.

Cette caméra est utilisée dans le cadre du programme stratégique Subaru Hyper Suprime-Cam (HSC-SSP). Ce projet a trois objectifs scientifiques principaux : l'étude de la cosmologie par l'analyse de l'effet de lentille gravitationnelle faible, l'étude des galaxies à haut redshift et l'évolution des galaxies sur le temps cosmique. La détection des amas de galaxies, en particulier par leur masse totale, s'inscrit dans les objectifs scientifiques additionnels du relevé HSC (Miyazaki et al., 2015).

Le programme HSC-SSP se compose de trois relevés du ciel : grand champ, profond et ultra-profond. Ces relevés sont réalisés durant un total d'environ 330 nuits d'observation pour couvrir un total de 1400 deg² pour le relevé grand champ, 27 deg² pour le relevé profond et 3.5 deg² pour le relevé ultra-profond (voir Table 8.1). D'un côté, les champs profonds et ultra-profonds sont orientés vers l'étude des supernovæ, des AGNs, et l'évolution des galaxies à haut et très haut redshift ($z > 1.5$). De l'autre, le champ large vise l'étude du lentillage faible, l'étude des galaxies à faible redshift et les amas

1. Voir également Komiyama et al. (2018) pour les détails de la conception du système et de l'environnement isolé des CCDs, Kawanomoto et al. (2018) pour le détail des filtres de caméra utilisés, Furusawa et al. (2018) pour les détails de l'estimation automatique de la qualité d'observation sur le site.



Figure 8.1 – Photographie du télescope Subaru.

Table 8.1 – Caractéristiques estimées des trois relevés grand champ, profond et ultra-profond du programme HSC-SSP.

Relevé	Taille [deg ²]	Bandes
Grand champ	1400	grizy ($mag_i \sim 26$)
Profond	27	grizy + 3 NBs ($mag_i \sim 27$)
Ultra-profond	3.5	grizy + 3 NBs ($mag_i \sim 28$)

de galaxies. Dans la pratique, ces observations se font à travers cinq bandes du spectre électromagnétique *grizy* qui couvrent les longueurs d'ondes 300 - 1100 nm. Le relevé HSC-SSP utilise ainsi une série de filtres optiques afin d'observer différentes bandes spectrales. La bande *i* (700 - 850 nm) en particulier, est utilisée pour la mesure de forme des galaxies pour le lentillage faible (voir [Mandelbaum et al., 2018](#)). Les 4 autres bandes restantes, *grzy*, sont utilisées pour la mesure du redshift des galaxies par photométrie (voir [Aihara et al., 2018a](#)). Dans le cas des relevés profond et ultra-profond, les cinq bandes précédentes sont également complétées par un total de 4 bandes très étroites (Narrow-Bands 387, 816, 921 et 101).

Couvrant 1400 *deg*², le relevé grand champ remplit les conditions nécessaires pour la détection des amas par effet de lentille faible comme décrit en section 1.3.4. La taille du champ permet de détecter un grand nombre d'amas et la profondeur du relevé, avec une magnitude de mag_i de ~ 26 , permet d'observer une grande densité de galaxies sources.

Jusqu'à présent, le programme HSC-SSP a connu un succès dans différents domaines de l'astrophysique. Même si le relevé HSC-SSP et le traitement des données sont aujourd'hui toujours en cours, les données récoltées sur la première année ont été rendues publiques ([Aihara et al., 2018b](#); [Mandelbaum et al., 2018](#)) et les résultats associés ont donné lieu à des premières publications ([Miyazaki et al., 2018](#); [Sugiyama et al., 2022](#)). Puis, en 2023, les résultats scientifiques obtenus à partir des données collectées sur les trois premières années (416 *deg*²) ont été publiés ([Dalal et al., 2023](#); [Li et al., 2023](#); [Miyatake et al., 2023](#); [More et al., 2023](#); [Sugiyama et al., 2023](#)).

Ces différentes propriétés du relevé nous ont naturellement porté vers HSC pour l'application de notre méthode pour la première fois à des données réelles.

Cependant, ces possibilités d'études ne sont pas uniques à HSC. En effet, il n'est pas le seul relevé

multi-bandes orienté précisément sur l'effet de lentille faible. Comme nous l'avons mentionné en section 1.3.4, nous entrons dans l'âge d'or de l'étude de la cosmologie par l'effet de lentille faible. D'autres relevés terrestres multibandes performants et même plus larges existent pour le lentillage faible, Kilo-Degree Survey (KiDS, [de Jong et al., 2013](#)) et le Dark Energy Survey (DES, [Dark Energy Survey Collaboration et al., 2016](#)). Mais HSC est unique par la combinaison de son champ large avec une haute-résolution d'image et une profondeur bien supérieure aux autres relevés. Cette profondeur assure ainsi une densité de galaxies bien supérieure aux autres relevés actuels.

8.1.2 . Des données adaptées pour la détection

Dans cette section, nous utilisons le premier catalogue de cisaillement, nommé S16A, produit par HSC-SSP et décrit par [Mandelbaum et al. \(2018\)](#). Ce catalogue contient les mesures d'ellipticités (voir Sect. 2.2) de six champs disjoints de tailles variables et nommés XMM, GAMA09H, WIDE12H, GAMA15H, HECTOMAP et VVDS qui couvrent un total de 136.9 deg². Les cartes correspondantes aux six champs sont présentées dans la figure 8.2. Des masques ont été appliqués pour masquer les objets brillants proches (étoiles).

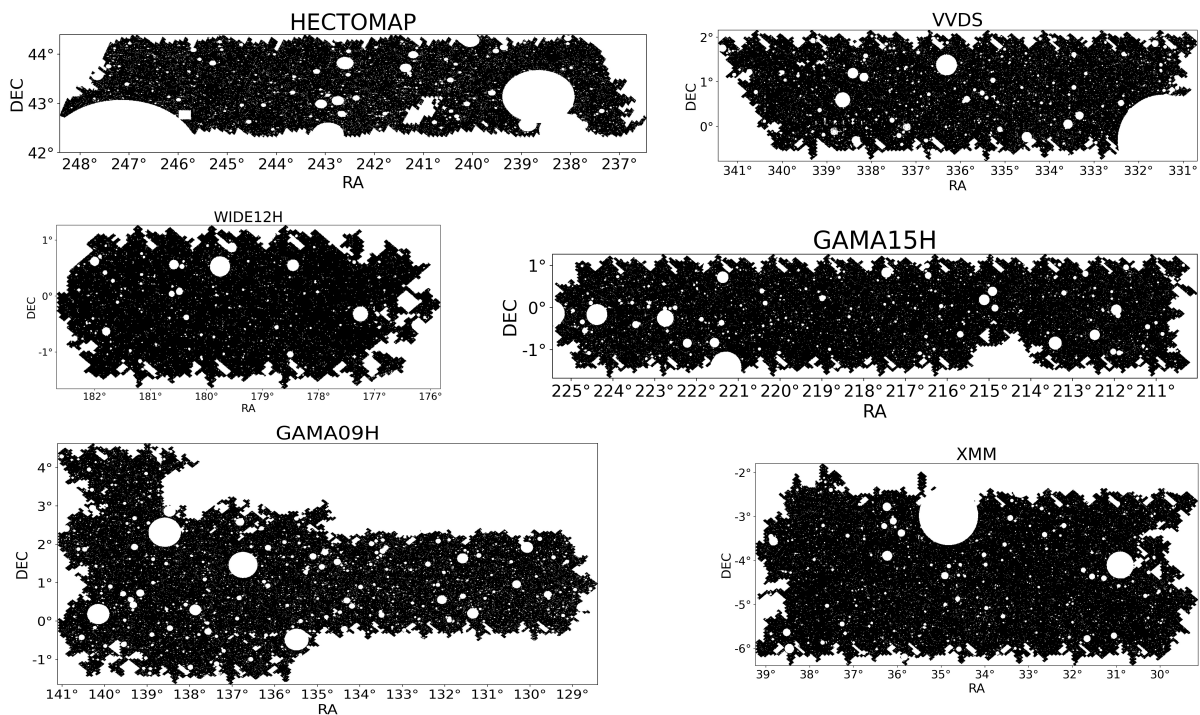


Figure 8.2 – Cartographie des 6 champs couverts par le relevé HSC. Les trous (partie blanche) présents dans les champs correspondent à la position des objets brillants dans le masque.

Le catalogue que nous utilisons contient environ 12 millions de galaxies avec pour chacune leur position RA, DEC, et les ellipticités mesurées des galaxies (e_1 , e_2 , qui est un estimateur pour γ_1 et γ_2). À cela s'ajoute les biais multiplicatifs m et additifs (c_1, c_2) dont le rôle est de corriger la mesure de forme des différents biais de la méthode dus au bruit et aux effets systématiques introduits par la chaîne de traitement (e.g. [Mandelbaum et al., 2018](#)). Un poids, w , accompagne également chaque galaxie, pour rendre compte de la fiabilité globale de la mesure de forme de cette galaxie par la méthode. Sur ce catalogue, on mesure une densité moyenne de galaxie de $n_g = 24.6 \text{ gal.arcmin}^{-2}$ sans les poids w en dehors des masques, et une densité de galaxies moyenne effective $n_g = 21.8 \text{ gal.arcmin}^{-2}$ en incluant les poids w , ce qui le rend adapté pour la détection des amas.

Pour chaque galaxie, nous disposons également du redshift photométrique estimé en utilisant différentes méthodes. Dans notre cas, nous adoptons le redshift photométrique obtenu grâce à la

méthode d'ajustement de modèle *Mizuki* (Tanaka, 2015; Tanaka et al., 2018) faisant le même choix que Miyazaki et al. (2018) pour faciliter notre comparaison avec les résultats publiés.

L'application de la méthode de masse d'ouverture sur les premières données S16A HSC a montré des résultats très prometteurs avec un catalogue de 65 amas de galaxies détectés avec un SNR > 4.7 (Miyazaki et al., 2018). Puis, Hamana et al. (2020) ont amélioré ces performances en ajoutant une approche tomographique et étendent ainsi le catalogue à 107 amas de galaxies. Cela conforte notre choix pour les données HSC afin d'appliquer la méthode ondelettes sur des données réelles.

Les résultats déjà publiés (Miyazaki et al., 2018; Hamana et al., 2020; Oguri et al., 2021) sont très intéressants pour nous et constituent un premier avantage de ce relevé car ils servent de base pour la comparaison avec les performances de la méthode multi-échelle. Ce point est d'autant plus important que l'application de notre méthode à des données réelles soulève comme défi la validation des détections. Notons que les amas de galaxies détectés par Miyazaki et al. (2018) et Hamana et al. (2020) sont obtenus par l'application de la masse d'ouverture (voir Sect. 3.2.3) directement sur le catalogue de cisaillement, avec le filtre M18.

Un second avantage de ce relevé pour l'application de notre approche et son évaluation est sa superposition avec d'autres relevés disposant de catalogues d'amas détectés dans différentes longueurs d'ondes. En particulier, nous avons accès au catalogue de détections optiques CAMIRA (Oguri et al., 2018) contenant plus de 1900 amas répartis sur les six champs et obtenus grâce à l'application de l'algorithme de détection CAMIRA (Cluster-finding Algorithm based on Multi-band Identification of Red-sequence gAlaxies, Oguri, 2014) sur les données obtenues par HSC. Le relevé se superpose également avec d'autres relevés tels que COSMOS (Scoville et al., 2007), VVDS (Le Fèvre et al., 2013), GAMA (Driver et al., 2011) et HECTOMAP (Hwang et al., 2016), et les relevés rayons X, eROSITA (Merloni et al., 2012), XMM-XXL (Pierre et al., 2016). Ces différents relevés pourront servir de base de comparaison dans un second temps pour les détections n'ayant pas de correspondances dans le catalogue optique CAMIRA.

En résumé, ces différents points font du relevé HSC-SSP un jeu de données adapté à la fois pour la détection des amas par lentillage faible et pour la validation des performances de notre méthode multi-échelle sur des données réelles. Cette conclusion est appuyée par le fait que ce jeu de données a déjà été utilisé pour produire un catalogue d'amas de galaxies détectés par effet de lentille gravitationnelle faible.

8.2 Défi de l'application aux données réelles HSC

Après avoir analysé des jeux de données simulées dans les chapitres précédents, nous avons désormais à traiter des données réelles. Dans le cas de l'application aux données HSC, nous avons identifié trois problèmes à traiter :

- Les données manquantes dues aux masques des étoiles brillantes proches et la complexité des relevés astronomiques dans leur mise en place.
- Le bruit sur la mesure de cisaillement qui, en pratique, n'est pas parfaitement gaussien comme supposé dans l'analyse des simulations ce qui complexifie le filtrage des données.
- La distribution des galaxies dans le ciel qui n'est pas homogène et ceci est amplifié par le déroulement des observations dont la qualité n'est pas égale d'un champ à l'autre.

Dans cette section, nous détaillons ces problèmes liés à l'application de la méthode multi-échelle aux données réelles. Le défi de la validation des détections est abordé séparément dans la section 8.4.

8.2.1 . Les données manquantes

Lorsque l'on travaille sur les données réelles d'un télescope terrestre, plusieurs défis se posent à travers les imprécisions de mesures, et les conditions d'observations. Le premier problème qui trans-

paraît directement dans le catalogue est la présence de données manquantes dans la carte. Dans cette sous-section, nous expliquons l'origine de ces données manquantes et leur impact sur la détection.

Les données manquantes ont différentes origines. Premièrement, il peut s'agir d'une défaillance technique au niveau des capteurs CCDs du télescope. La seconde et principale raison est la présence d'étoiles brillantes sur le champ. Ces objets lumineux saturent les capteurs et forment ainsi des tâches lumineuses autour de leur position. Il faut alors masquer ces régions saturées faute de ne pouvoir voir au travers. La troisième explication possible pour les données manquantes vient du nombre de galaxies détectées dans le champ. Ce nombre étant fini, il peut y avoir des zones comportant peu ou pas de galaxies. Ainsi, la résolution des cartes, à savoir la largeur d'un pixel, a aussi un impact sur la présence de données manquantes (voir Sect. 8.3.1). Dans la pratique, l'origine multiple des données manquantes se traduit sur les cartes par des trous dont la taille varie du pixel à des cercles de dix pixels de rayons.

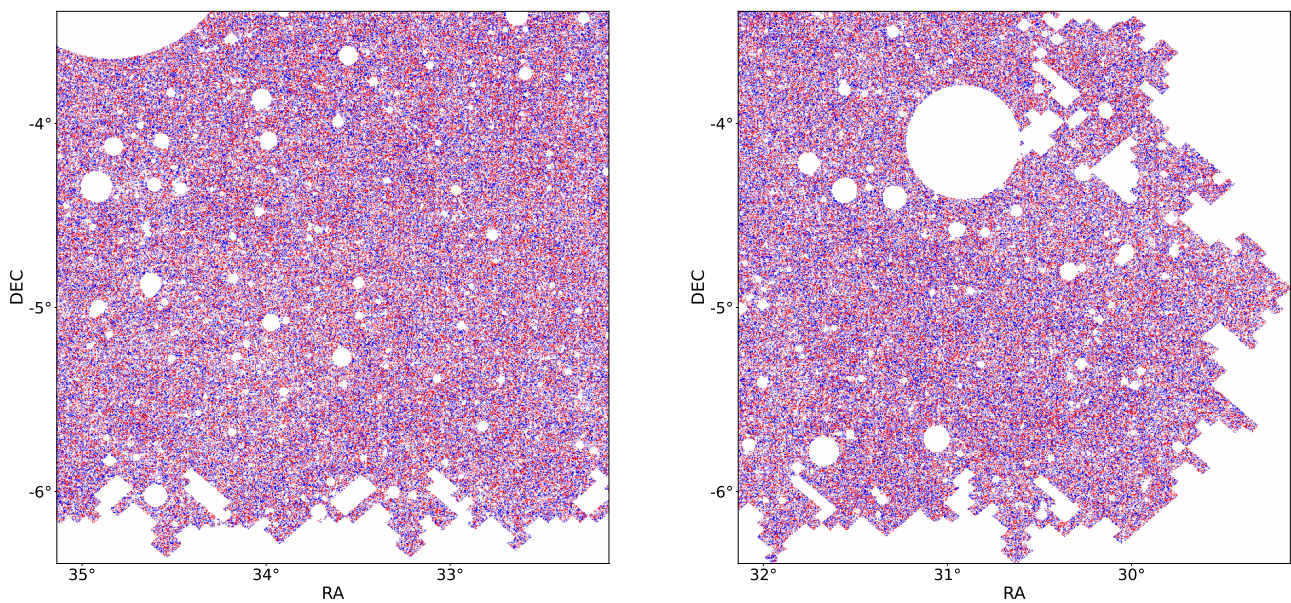


Figure 8.3 – Zooms (cartes de cisaillement γ_2) de $3 \times 3 \text{ deg}^2$ sur les données S16A de HSC (patch XMM) illustrant la répartition et la taille variable des données manquantes (pixels blancs).

Pour les données HSC, en tenant compte des données manquantes uniquement dues aux masques, la surface couverte théorique de 160 deg^2 se réduit à 136.9 deg^2 . Cela correspond à environ 15% de régions masquées par des étoiles brillantes sur l'ensemble des six patches. Dans la figure 8.3, nous illustrons la présence des données manquantes ainsi que leur répartition sur un zoom du champ XMM de taille $3 \times 3 \text{ deg}^2$ pour lequel le blanc correspond à l'absence de données. On constate que les données manquantes dans le cas de HSC peuvent couvrir des zones très larges (droite) ou majoritairement très fines (gauche) tout en suivant une répartition aléatoire.

Jusqu'à présent, nous avons travaillé sur des champs complets simulés avec des bords parfaitement droits. Le seul traitement nécessaire était alors la gestion des bords sur les différentes cartes. Pour rappel, nous avons choisi comme traitement de retirer les bords lors de l'identification des pics. Ce traitement est tout à fait courant car les bords sont des régions définies et donc facilement isolables. Pour le traitement des données manquantes dans le relevé HSC, nous avons suivi cette même approche. Nous avons exclu les bords lorsque la taille des zones masquées est plus grande que 4 pixels.

8.2.2 . Le bruit non gaussien

Le second défi que nous avons identifié dans le cas de l'application à des données réelles est le bruit. Dans cette sous-section, nous décrivons l'impact du bruit et sa nature dans le cas des données réelles.

Dans les données de lentille faible, nous avons vu que le bruit sur le cisaillement était dû aux erreurs de mesure et surtout à la forme intrinsèque des galaxies. Dans les chapitres précédents, ce bruit suivait une distribution gaussienne. Cette hypothèse de bruit gaussien est à la fois raisonnable (Leauthaud et al., 2007) et permet d'estimer le niveau de bruit sur les cartes de convergence lors de la procédure de seuillage (voir Sect. 3.3.1). Le bruit étant gaussien et la distribution homogène, il est possible de définir un seuil global car tous les pixels ont la même distribution de bruit.

Dans le cas des données réelles, le bruit peut contenir des fluctuations non-gaussiennes induites essentiellement par les effets systématiques lors de la mesure des ellipticités (voir Sect. 2.2). La présence de ces non-gaussianités complexifie l'estimation du niveau de bruit au sein des différentes cartes filtrées.

Étant dominée par le bruit, la distribution des valeurs d'ellipticités, soit γ_1 et γ_2 , au sein du catalogue de galaxies sources nous permet de caractériser la distribution du bruit globale dans les données. Ainsi, la figure 8.4 montre les distributions des ellipticités γ_1 (gauche), γ_2 (droite) pour le champ XMM du relevé HSC, corrigées par les biais additifs et multiplicatifs, puis multipliées par le poids, w . Sur chacune de ces distributions nous ajustons une distribution gaussienne (courbe orange). Nous observons que les distributions des ellipticités ne suivent pas une distribution parfaitement gaussienne.

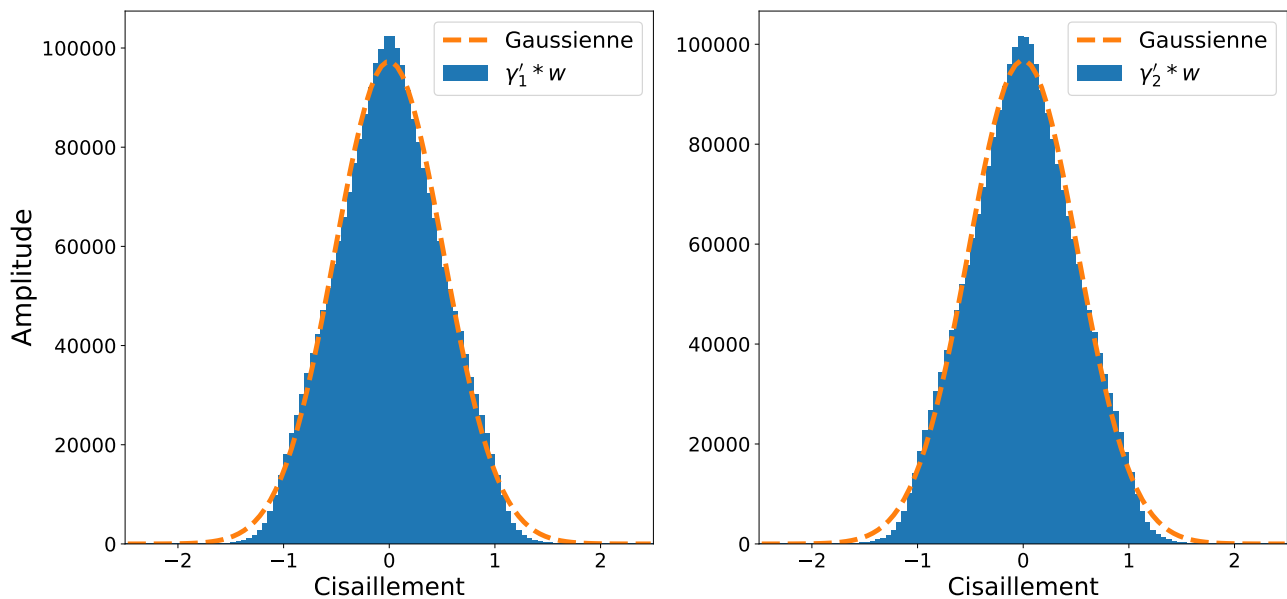


Figure 8.4 – Distributions des ellipticités corrigées γ_1 (gauche) et γ_2 (droite) pour le catalogue de galaxies sources du champ XMM. Ces deux distributions sont chacune interpolées par une distribution gaussienne (courbe orange).

Pour quantifier les écarts observés avec les distributions gaussiennes obtenues par ajustement, nous calculons la fonction de distribution cumulée, CDF. Ces résultats sont présentés dans la figure 8.5 pour la distribution de γ_1 (gauche) et de γ_2 (droite). Nous calculons également la différence entre les CDFs des distributions de bruit et des distributions gaussiennes associées. Nous observons des écarts très marqués pour le début des CDF qui s'expliquent par le faible nombre de valeurs pour cette zone qui correspond à une des extrémités de la gaussienne. Dans la région centrale pour les deux distributions, il est intéressant de noter que les écarts significatifs persistent et témoignent de fluctuations non-gaussiennes au sein des distributions. Ces écarts confirment la présence de non-gaussianités au sein des catalogues de galaxies sources pour le champ XMM. Nous constatons également des écarts

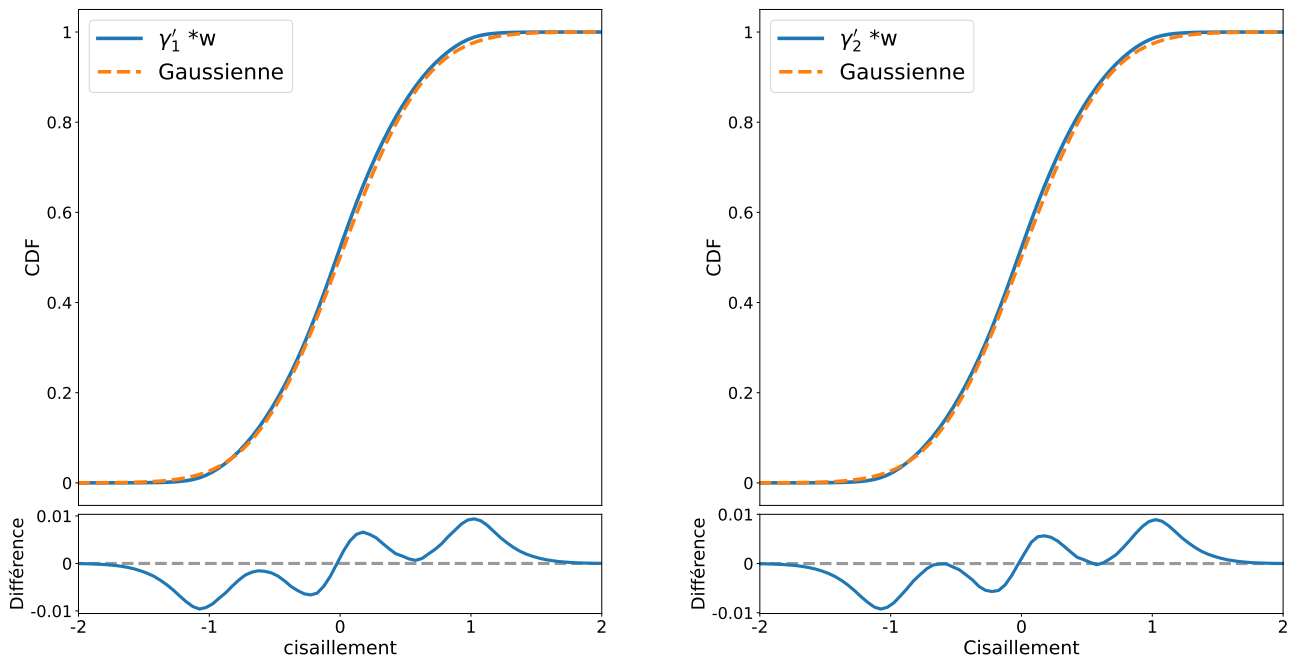


Figure 8.5 – Fonctions de distribution cumulée (CDF) du catalogue de galaxies sources du champ XMM avec les données brutes de γ_1 (gauche), et γ_2 (droite) corrigées par les biais multiplicatifs et le poids des galaxies. L’interpolation par une distribution gaussienne (courbe orange) réalisée avec la figure 8.4 sert de base de comparaison pour illustrer les non-gaussianités. La différence entre les CDFs de la distribution de γ_i corrigée et de la distribution gaussienne sont présentées en dessous des figures pour accentuer les écarts (courbe bleue).

à la gaussianité pour les autres champs. Il est intéressant de noter que [Mandelbaum et al. \(2018\)](#) soulignent les conditions d’observation variables d’un champ à l’autre, les meilleures conditions observées correspondant aux champs HECTOMAP et VVDS. À l’inverse, les pires conditions d’observation sont mesurées sur le champ XMM.

En résumé, l’application de notre méthode à des données réelles doit aller de pair avec la prise en compte de ces non-gaussianités lors de la procédure d’estimation du bruit et du seuillage qui en découle.

8.2.3 . Une distribution non uniforme des galaxies sources

Enfin, le troisième défi que nous avons identifié porte sur la distribution non uniforme et non homogène des galaxies sources au sein du catalogue. Dans cette sous-section, nous décrivons comment les galaxies sont distribuées dans le jeu de données, et l’impact de cette distribution sur le filtrage.

Précédemment, nous avons vu que dans le cas de la simulation DUSTGRAIN (voir Sect. 2.3), les galaxies sources sont distribuées de manière uniforme et homogène derrière les amas de galaxies. Cette distribution idéale est impossible à mettre en place dans les relevés réels du fait tout d’abord de l’inhomogénéité de la distribution des galaxies dans l’univers et de la qualité des relevés qui varie en fonction des observations. Dans le jeu de données, cela se traduit par une variation de la densité moyenne de galaxies à la fois entre les différents champs d’observation, présentée dans la table 8.2 mais aussi la densité de galaxies entre les différents régions ou pixels d’un même champ. La figure 8.6 illustre ainsi cette variation au sein du champ XMM en traçant la densité de galaxies pour une taille de pixel de 1 arcmin/pixel. Les variations de la densité de galaxies sont telles que certaines régions sont plus denses ou plus faibles que la moyenne de la carte, $n_g = 22.2 \text{ gal.arcmin}^{-2}$. Par la variation de la colorimétrie au sein de la carte, nous observons également que la densité de galaxies est plus faible aux environs des bords et des régions masquées. Dans la pratique, une distribution non uniforme des galaxies sources augmente l’apparition de données manquantes sur le champ d’observation. Ces variations importantes de la densité de galaxies d’un pixel à l’autre ont un impact significatif sur la

reconstruction des cartes de convergence et sur leur niveau de bruit. Pour notre approche, l'impact est d'autant plus important que les différents filtres ondelettes correspondent à des échelles différemment affectées par le niveau de bruit.

Table 8.2 – Densité de galaxies effective sur les différents champs du relevé HSC.

Champ observé	ng
XMM	22.2
GAMA09H	21.6
WIDE12H	24.2
GAMA15H	23.2
HECTOMAP	17.4
VVDS	22.9

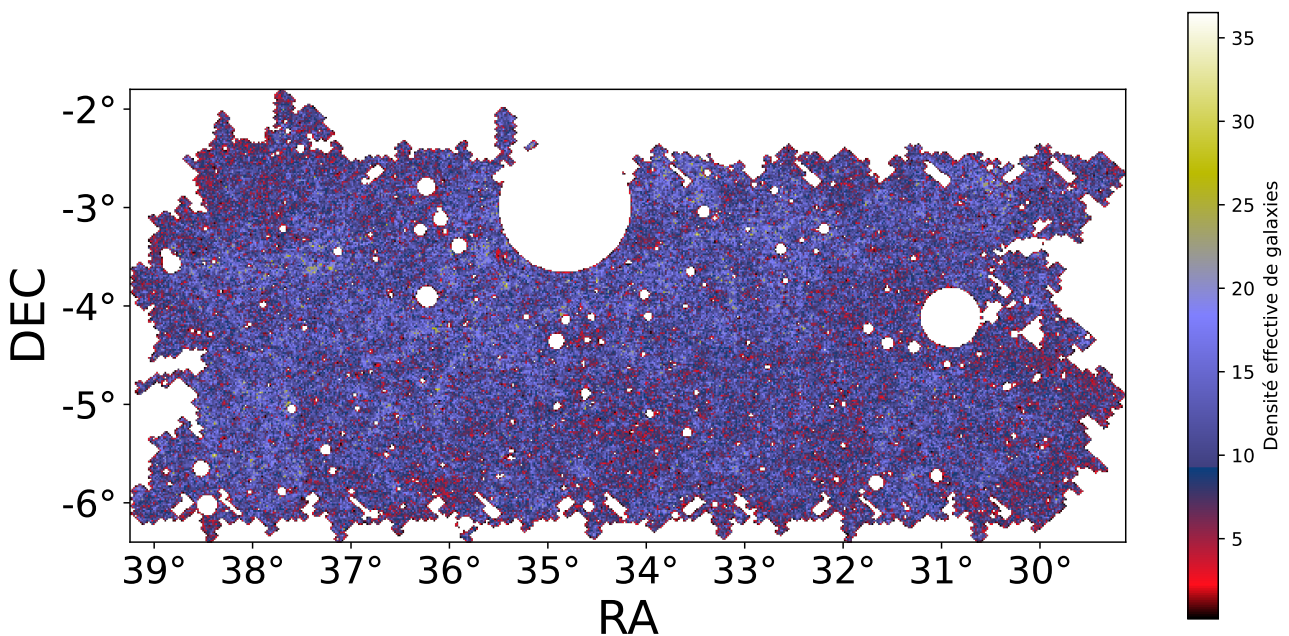


Figure 8.6 – Distribution de la densité effective de galaxies dans le champ XMM à partir du catalogue de galaxies sources corrigées par leur poids, w .

En résumé, nous avons identifié trois défis majeurs que pose l'application aux données réelles HSC : les données manquantes, la distribution non-gaussienne du bruit et la densité de galaxies non uniforme. Ces trois points ont un impact significatif sur la procédure de seuillage car ils affectent tout particulièrement l'estimation du niveau du bruit. Il est essentiel de réduire ou résoudre ces défis lors de l'application de la méthode multi-échelle.

8.3 Application de la méthode de détection multi-échelle

Nous avons introduit les différents défis qui vont de pair avec l'application de notre méthode aux données réelles HSC. Ces défis ont un impact significatif sur la procédure que nous avons mis en place sur les données simulées (voir Sect. 4.2) notamment lors de l'estimation du bruit et du seuillage. Dans cette section, nous détaillons l'adaptation de notre algorithme multi-échelle sur les données S16A collectées lors de la première année du relevé HSC pour faire face à ces défis.

8.3.1 . Pixellisation des données

Dans le cas des données réelles, les défis auxquels nous faisons face sont la présence de données manquantes (voir Sect. 8.2.1) au sein des catalogues et la distribution non uniforme des galaxies (voir Sect. 8.2.3). Ces défis demandent d'adapter le traitement des données afin d'éviter les pertes de signal et les fausses détections. Dans cette sous-section, nous présentons comment nous reconstruisons les cartes de convergence à partir des catalogues de galaxies sources bruités et masqués correspondant aux différents champs.

Comme décrit en section 3.1, nous commençons par transformer les catalogues de cisaillement en cartes de convergence avant de procéder à la procédure de filtrage. Dans notre cas, nous appliquons cette procédure sur les catalogues de galaxies sources en tenant compte pour chaque galaxie de la correction par les biais multiplicatif et additif, et de leur poids, w . Dans la pratique, cela revient à reconstruire les catalogues de cisaillement corrigés, γ'_i , sous la forme proposée par [Mandelbaum et al. \(2018\)](#) :

$$\gamma'_i = \frac{e_i - c_i}{1 + m} * w, \quad (8.1)$$

avec $i \in \{1, 2\}$, m le biais multiplicatif, c_i les biais additifs et w le poids associé à l'ellipticité e_i . Nous pixellisons ensuite ces catalogues corrigés pour reconstruire les cartes de cisaillement associées.

Dans le cas des données réelles, la pixellisation et la reconstruction des différentes cartes peuvent être grandement impactées par les données manquantes et la distribution non uniforme des galaxies. Ainsi, il est essentiel de tenir compte de la présence des masques et de la densité variable des galaxies sources lors de la reconstruction des cartes de convergence. Notons que les données manquantes ont également un impact différent en fonction de l'approche qui est utilisée pour la pixellisation des catalogues.

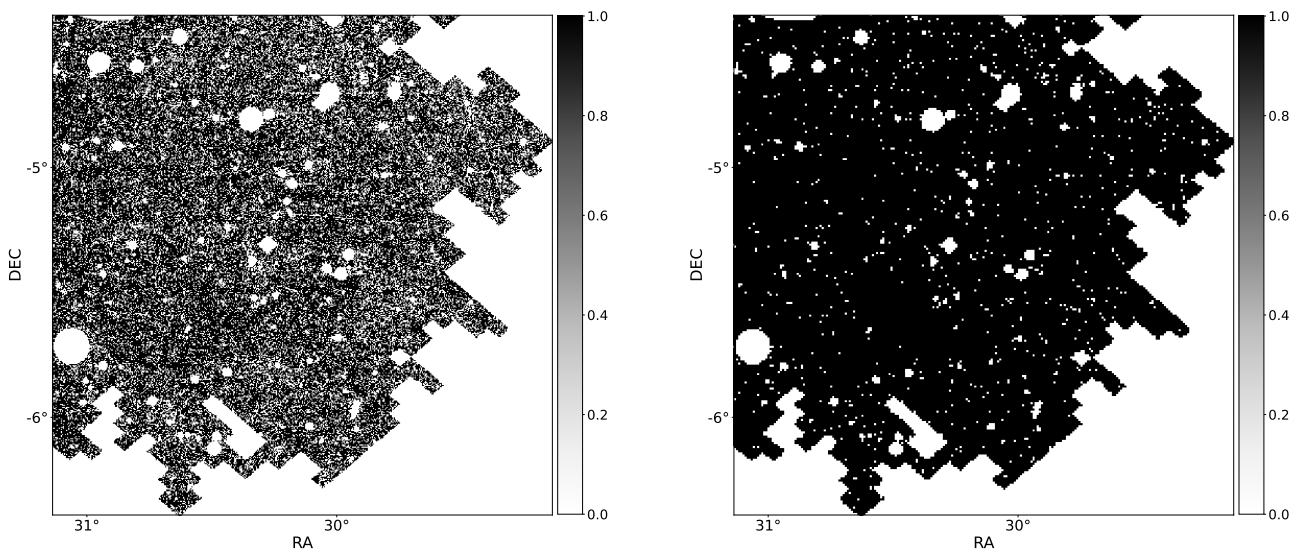


Figure 8.7 – Extrait de $2^\circ \times 2^\circ$ de la carte de cisaillement du champ XMM pixelisée à partir d'une résolution de 0.25 arcmin/pixel (gauche) et d'une résolution de 0.50 arcmin/pixel (droite).

Pour l'approche qui utilise directement le catalogue de cisaillement aux positions exactes des galaxies, les données manquantes ont un effet moindre car elle réduit le nombre effectif de galaxies. Cette approche peut être facilement adaptée en normalisant par le nombre de galaxies prises en compte lors de la convolution ou en utilisant un filtre plus large ([Miyazaki et al., 2018](#)). Mais cette approche est coûteuse à mettre en œuvre en termes de temps de calcul compte tenu de la taille des données du relevé actuel et à venir.

Dans la section 3.1, nous avons introduit trois approches possibles pour la pixellisation des données et quantifié leur impact sur le signal. Nous complétons cette analyse maintenant en l'adaptant à la présence des données manquantes dans les relevés réels. Dans le cas des approches (MOY et SOM)

qui consistent à pixeliser les catalogues de cisaillement avant de convoluer par le filtre, l'impact des données manquantes est dépendant de la résolution choisie. Dans le chapitre 3, nous avons montré que lorsque la position des galaxies est approximée au pixel le plus proche, la pixellisation n'entraîne qu'une légère perte de signal de $\sim 1.5\%$ pour MOY et de $\sim 1.2\%$ pour SOM. La pixellisation la plus fine qui est celle qui offre la réduction la plus faible en termes de perte de signal et cependant celle qui ajoute le plus grand nombre de données manquantes qui ont un lourd impact sur le signal filtré. Il apparaît logiquement que la pixellisation est un compromis entre données manquantes et perte du signal.

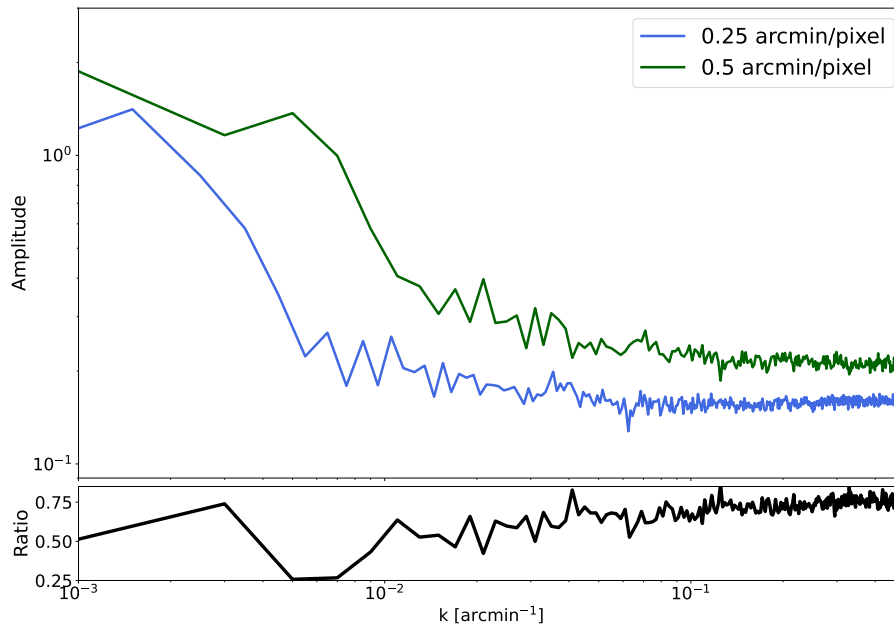


Figure 8.8 – Spectre des cartes de convergence du champ XMM reconstruites à partir des résolutions 0.25 arcmin/pixel (ligne bleue) et 0.50 arcmin/pixel (ligne verte).

La figure 8.7 permet de visualiser la distribution des données pour une portion de $2^\circ \times 2^\circ$ du champ XMM pour deux résolutions différentes. La résolution restant inchangée, les observations sur un sous-champ se maintiennent alors également sur le champ total. Le panneau gauche correspond à une résolution de 0.25 arcmin/pixel, comme utilisé par Miyazaki et al. (2018) et Hamana et al. (2020) pour leurs cartes filtrées tandis que le panneau droit correspond à une résolution de 0.50 arcmin/pixel. Sur les deux cartes, les pixels vides sont pris égaux à 0 et les pixels contenant au moins une galaxie source sont pris égaux à 1. On peut alors mesurer pour l'ensemble des 6 champs, le taux de données manquantes par le pourcentage de pixels à zéro à l'intérieur de la zone d'observation.

En moyennant les résultats sur les 6 champs, la résolution la plus fine, 0.25 arcmin/pixel, conduit à environ $21.3\% \pm 1.6\%$ de pixels nuls. Ce taux de données manquantes chute à $4.9\% \pm 0.7\%$ quand on utilise la résolution plus large de 0.50 arcmin/pixel. Notons que cette valeur faible s'explique par le fait que nous avons calculé uniquement les pixels à 0 dus à la distribution des galaxies sources en retirant ceux dus aux masques des étoiles dont l'impact est le même sur les deux résolutions et parties intégrantes des données. Ainsi, la résolution de 0.5 arcmin/pixel permet de grandement réduire les données manquantes dues à la distribution non uniforme des galaxies.

Pour pousser au-delà cette analyse, nous avons étudié l'effet de la résolution sur le signal lors de la reconstruction des cartes de convergence à partir des cartes de cisaillement. Pour cela, on reconstruit, pour chacune des deux résolutions, les cartes de convergence associées pour chacun des champs d'observation. Ensuite, nous avons calculé le spectre de puissance dans l'espace de Fourier des cartes de convergence en mode E. Dans la Figure 8.8, nous comparons le spectre du signal obtenu par les deux résolutions (courbes solides bleue et verte). La résolution de 0.25 arcmin/pixel correspond à un signal plus faible et dégradé par rapport à 0.50 arcmin/pixel qui est relié directement à la fraction

de données manquantes. Le ratio (ligne noire) entre les deux courbes a une valeur de 0.76, ce qui montre que la résolution la plus fine entraîne une perte de 24% du signal comparée à une résolution de 0.5 arcmin/pixel. Ces résultats restent valables sur les 6 champs des données S16A dans la mesure où l'impact vient de la résolution.

La pixellisation est un compromis entre les données manquantes introduites et la perte du signal. Cette comparaison indique que la résolution de 0.50 arcmin/pixel est optimale pour traiter les données d'entrées avec données manquantes car elle assure une perte de signal minimale ($\sim 1.2\%$, voir Fig. 3.3) tout en garantissant un nombre de pixels manquants très faible. Une résolution plus élevée réduirait encore un peu le nombre de pixels nuls mais cela aurait un impact plus significatif sur la réduction du signal.

En résumé, la première étape de l'application à des données réelles est la production des cartes de cisaillement à partir des catalogues de galaxies sources en tenant compte des données masquées. Pour minimiser les données manquantes, nous fixons alors une résolution de 0.50 arcmin/pixel pour la pixellisation des cartes de cisaillement.

8.3.2 . Reconstruction des cartes de convergence filtrées et estimation du bruit

Comme nous l'avons précisé dans les chapitres précédents, notre méthode multi-échelle par les ondelettes s'applique sur les cartes de convergence. Dans cette sous-section, nous détaillons ainsi la reconstruction des cartes de convergence, leur filtrage par la méthode multi-échelle et l'estimation du niveau de bruit.

★ Construction des cartes de convergence filtrées

Nous commençons par produire les cartes de cisaillement à partir des catalogues de cisaillement de chaque champ, en suivant la procédure décrite dans la section 3.1. Comme nous l'avons précisé dans la section 8.3.1, nous corrigeons les valeurs de cisaillement au sein des catalogues à partir des biais additifs et multiplicatifs et nous appliquons le poids associé à chaque galaxie. Une fois la résolution fixée à 0.50 arcmin/pixel, nous utilisons la pixellisation SOM qui consiste à sommer les valeurs des galaxies tombant dans le même pixel pour pixelliser les catalogues de cisaillement. Pour la reconstruction des cartes de convergence dans l'espace de Fourier, nous augmentons ensuite la taille des cartes en ajoutant des pixels nuls (zéro-padding) afin de former des cartes de cisaillement de largeur et longueur identiques. Ainsi par la méthode de Kaiser et Squires ([Kaiser and Squires, 1993](#)), nous reconstruisons les cartes de convergence de modes E et B à partir de ces cartes désormais carrées. Puis, nous inversons le zero-padding en revenant aux dimensions des cartes originelles.

Une fois les cartes de convergence de modes E et B reconstruites, nous appliquons la méthode de filtrage par ondelettes décrite dans la section 4.2.3. Dans la pratique, on filtre les modes E et B des cartes de convergence par les filtres ondelettes W_2 , W_3 et W_4 pour chacun des 6 champs. Les cartes de convergence de mode E filtrées qui en résultent pour le champ XMM sont présentées dans la figure 8.9. Nous répétons cette procédure pour chacun des six champs d'observation disjoints pour obtenir les cartes de convergence de mode E et B filtrées associées.

Nous observons que la reconstruction des cartes de convergence et leur filtrage créent des artefacts au niveau des bords des données manquantes car le signal s'étend aussi sur les zones masquées. Il est important de tenir compte de ces artefacts lors des étapes de seuillages et d'identification des pics candidats afin d'éviter les fausses détections.

★ Estimation du niveau de bruit

Comme nous l'avons souligné précédemment, les différents défis que soulève l'application à des données réelles impactent grandement l'estimation du niveau de bruit (voir Sect. 8.2). Celle-ci étant essentielle pour reconstruire les cartes de signal sur bruit et, à termes, identifier les pics candidats. Cependant, son estimation sur une seule réalisation n'est pas une tâche facile d'autant plus que le niveau de bruit est pondéré par le nombre de galaxies qui tombent dans un pixel qui varie fortement d'un pixel à l'autre.

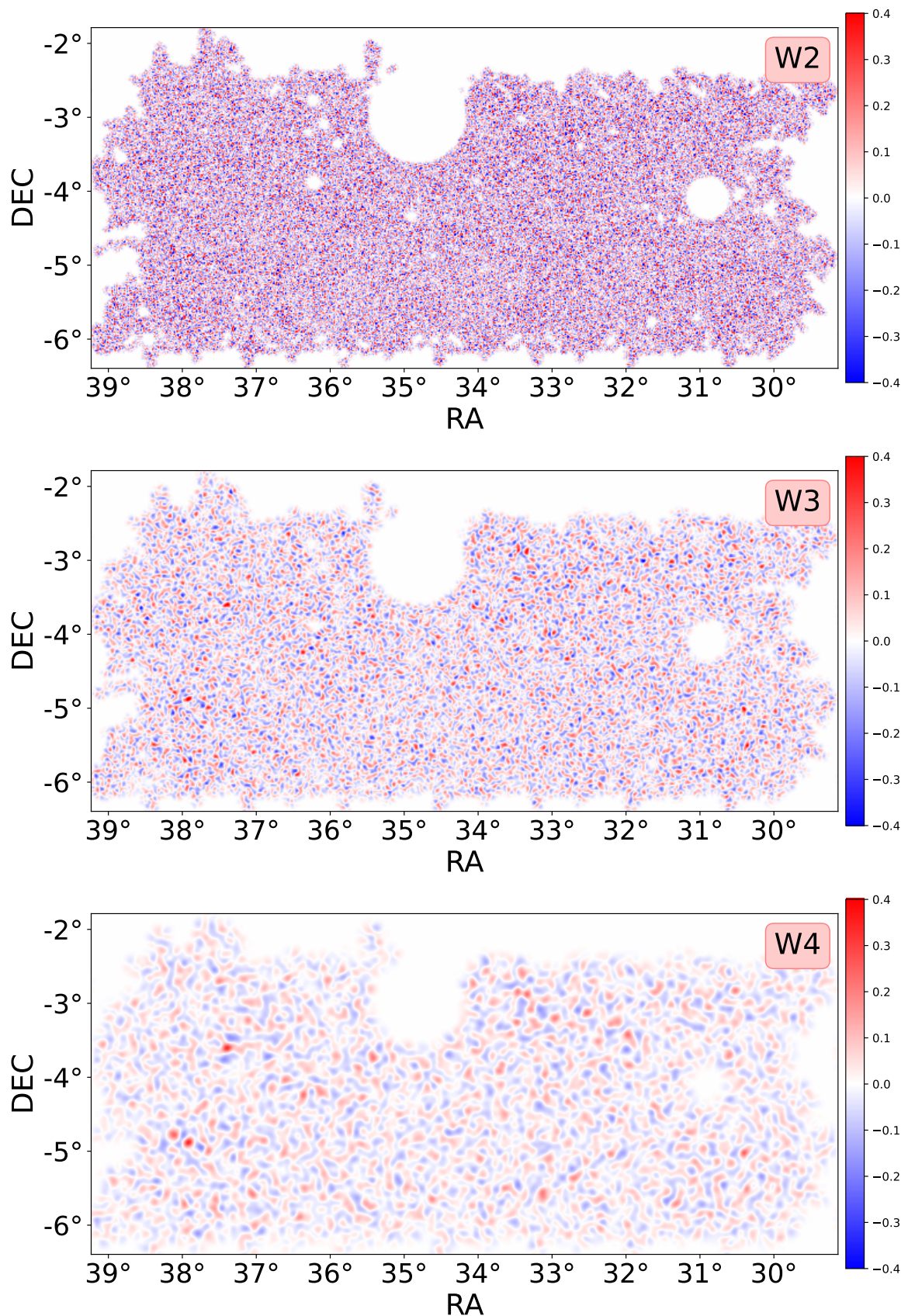


Figure 8.9 – Cartes de convergence de mode E du champ XMM des données S16A de HSC, filtrées par la méthode multi-échelle avec les filtres ondelettes W2, W3 et W4.

Pour pallier ce problème, nous avons choisi de créer des cartes de bruit pour chaque champ par la méthode de randomisation des ellipticités. La technique de 'randomisation des ellipticités' (Miyazaki

et al., 2018) consiste, à partir du catalogue de source, à annihiler le signal en appliquant une orientation aléatoire au cisaillement mesuré pour chacune des galaxies du catalogue. Ce faisant, nous constituons un catalogue contenant purement du bruit mais avec la position des galaxies, le poids associé, la densité et les effets systématiques inhérents au catalogue originel. Dans la pratique, nous générons plusieurs catalogues de bruit à partir de cette méthode puis nous reconstruisons les cartes de convergence de modes E et B bruitées associées. Ces cartes de convergence de modes E et B sont ensuite filtrées par les filtres ondelettes pour reconstruire des cartes de convergence filtrées contenant purement du bruit. Pour chacun des six champs, nous avons répété cette procédure 400 fois afin d'estimer avec précision le niveau de bruit dans chaque pixel des cartes de convergence en mode E filtrées.

Nous disposons alors de 400 points par pixels par champ qui y tracent la distribution du bruit. L'estimation du bruit au sein de chacun des pixels est calculée par l'écart-type sur la profondeur, à savoir pour chaque même pixel sur les 400 cartes de bruit.

Cette approche nous permet d'estimer l'écart-type du bruit reconstruire la carte de niveau de bruit dans chaque pixel de la carte après filtrage qui dépend du nombre de galaxies au sein de ce pixel. En d'autres termes, le niveau de bruit mesuré est d'autant plus grand que le nombre de galaxies dans ce pixel est élevé, suite au choix de la pixellisation SOM. Estimer le niveau de bruit par pixel nous permet ainsi de tenir compte de la distribution non uniforme des galaxies et des variations des caractéristiques du bruit d'un pixel à l'autre.

Afin de reconstruire les cartes de signal sur bruit, nous divisons les cartes de convergence de mode E filtrées par les cartes de niveau de bruit. Nous répétons cette procédure pour chacun des 6 champs.

En résumé, après pixellisation des cartes de cisaillement et reconstruction des cartes de convergence de modes E et B, nous les filtrons par les filtres ondelettes W_2 , W_3 et W_4 . En parallèle, nous utilisons la randomisation des ellipticités pour reconstruire une carte donnant le niveau de bruit (i.e. l'écart-type par pixel) associée à chaque champ et estimer le niveau de bruit dans chaque pixel du champ. Ces deux procédures nous permettent de produire les cartes de signal sur bruit de chaque champ sur laquelle se fait le seuillage qui est décrit dans la sous-section suivante.

8.3.3 . La procédure de seuillage

Une fois le niveau de bruit estimé au sein des différents champs, nous entamons la procédure de seuillage. Pour l'application aux données réelles, nous avons étudié et affiné la procédure détaillée dans le chapitre 3 (voir Sect. 3.3.1) pour tenir compte à la fois de la variation de la densité de galaxie et des caractéristiques du bruit. Dans cette sous-section, nous présentons l'implémentation des solutions qui nous ont permis de faire face à ces problèmes et l'identification des détections qui en découlent. En particulier, nous présentons l'application et les limites de la procédure de seuillage utilisant le maximum du mode B comme seuil sur les données HSC. Puis, nous introduisons une procédure de seuillage plus stable.

★ La moyenne des maximums

Dans les sections précédentes, pour isoler les pixels correspondant à des amas de galaxies, nous avons utilisé un seuil de détection différent pour chacun des filtres. Nous appuyant sur la méthodologie utilisée dans Hamana et al. (2012); Miyazaki et al. (2018); Hamana et al. (2020), ce seuil était obtenu en faisant la moyenne des maximums obtenus sur les 256 réalisations de cartes de convergence de mode B filtrées.

Dans le cas des données réelles, ce type d'approche est grandement complexifiée par le fait que la densité de galaxies n'est pas homogène et par les conditions d'observations qui sont variables d'un champ à l'autre. Pour en tenir compte, il est nécessaire de considérer l'écart-type du bruit par pixel et

d'avoir un seuil propre à chaque champ.

La méthode de randomisation des ellipticités que nous avons introduit précédemment permet de générer des cartes de bruit à partir d'un catalogue de galaxies. Nous disposons ainsi de 400 cartes de bruit pour chaque champ du relevé HSC-SSP qui vont permettre une estimation d'un seuil global pour chaque champ.

Table 8.3 – Moyenne des maximums mesurés pour chacun des champs d'observations de HSC présents dans les données S16A et pour les filtres ondelettes W2, W3, W4.

Champ observé	Aire [deg ²]	Max moyen W2	Max moyen W3	Max moyen W4
XMM	26.3	4.61	4.49	4.23
GAMA09H	28.52	6.47	4.61	4.28
WIDE12H	11.45	4.42	4.32	4.03
GAMA15H	27.50	4.61	4.52	4.26
HECTOMAP	9.48	4.48	4.34	4.00
VVDS	16.77	4.75	4.44	4.15

En suivant cette approche, nous dérivons ainsi des seuils pour les filtres W2, W3 et W4, pour chacun des champs. Les seuils sont ainsi calculés à partir de la moyenne du maximum mesuré sur chaque carte de bruit après filtrage, normalisée par la carte des écart-types du bruit à chaque pixel. Pour un calcul fiable des niveaux de bruits sur les cartes, nous ne considérons ni les zones originellement masquées, ni les zones autour des masques. La procédure ne change pas par rapport à la section 3.3.1, exception faite que l'écart-type varie d'un pixel à l'autre. Le détail de ces seuils est présenté dans la table 8.3. Nous présentons également la taille effective des différents champs. En utilisant cette approche, nous observons une augmentation des seuils par rapport à l'application aux données simulées (voir Sect. 3.3) ainsi qu'une corrélation des valeurs de seuils avec la taille des champs.

Pour approfondir cette observation, nous avons étudié la distribution de maximums pour un grand nombre de cartes de bruit en fonction de leur taille. Dans la pratique, nous avons généré 400 cartes de bruit correspondant à des distributions de bruit gaussiennes de moyenne nulle et d'écart-type 1 et de taille croissante : 128×128 , 256×256 , 512×512 , 1024×1024 , 2048×2048 et 4096×4096 . Puis, nous avons calculé la moyenne de chacune des distributions des maximums pour chaque taille de carte. Les différentes distributions des maximums sont présentées dans la figure 8.10.

Nous observons que les moyennes correspondant aux différentes distributions de maximums augmentent linéairement avec la taille des cartes. Ceci s'explique par le fait que la distribution gaussienne sur laquelle repose l'estimation du niveau de bruit est infinie et n'admet pas de maximums. Ainsi, plus la carte est grande et plus la probabilité d'y mesurer une valeur élevée est importante. Ce point est essentiel et explique les instabilités observées lors de l'utilisation des maximums comme seuil à la fois entre les six champs HSC mais aussi dans le cas du challenge (voir Sect. 7.3.2) pour lequel des seuils plus élevés sont observés sur des cartes plus larges comparées aux résultats des chapitres 3 à 6. Ces instabilités sont d'autant plus marquées dans les données HSC que la taille des cartes varie du simple au triple parmi les six champs. Dans le cas des simulations, le seuil calculé par la moyenne des maximums admet les mêmes limitations. Cependant, son impact est moindre du fait premièrement de la taille fixe des cartes qui nous assure un niveau de confiance similaire entre les différentes réalisations. De plus, une carte de 512×512 pixels correspond à un seuil moyen autour de 4.5 qui est par chance, proche de la valeur optimale trouvée par [Miyazaki et al. \(2018\)](#).

Pour corriger ces instabilités entre les champs, il convient d'étudier non pas le maximum mais l'entière de la distribution des cartes de bruit et de fixer le même niveau de confiance dans l'estimation des seuils qui est reliée directement au nombre de fausses détections.

En résumé, l'utilisation du maximum n'est pas optimale pour fixer le même niveau de confiance et de consistance sur les différents champs qui varient grandement en taille.

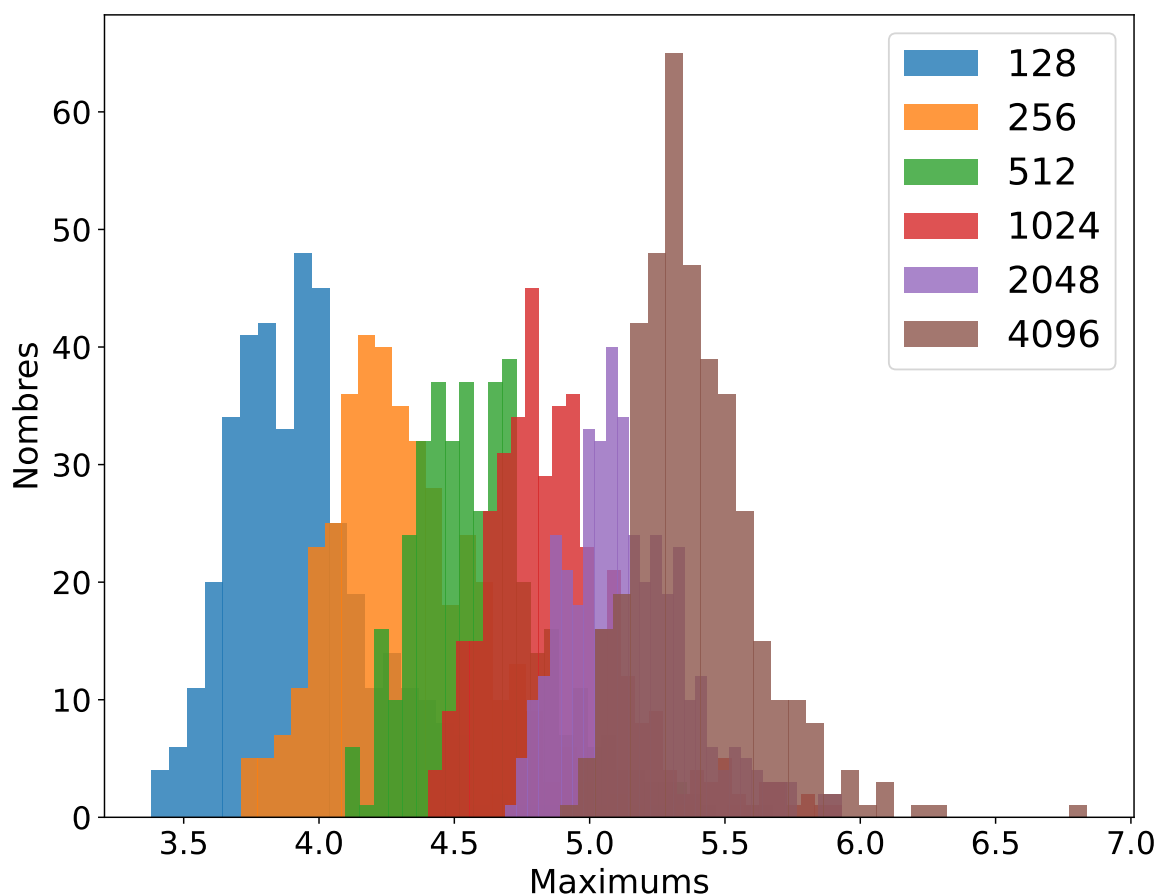


Figure 8.10 – Distribution des maximums obtenus sur 400 cartes de bruit suivant une distribution gaussienne pour des cartes de taille 128 × 128 (bleu), 256 × 256 (orange), 512 × 512 (vert), 1024 × 1024 (rouge), 2048 × 2048 (violet) et 4096 × 4096 (marron).

★ Seuillage par valeur-p

Afin de pouvoir réduire les instabilités entre les différents champs dans les données HSC, il est essentiel de fixer un niveau de confiance commun entre les différents champs. Pour cela, à la place du maximum nous calculons les seuils correspondant à une même valeur-p sur l'ensemble des 400 cartes de bruits pour chaque champ et chaque filtre. La valeur-p donne la probabilité d'obtenir une valeur (après soustraction de la moyenne) plus grande que le seuil défini s'il n'y avait que du bruit. Par exemple, le maximum correspond par définition à une valeur-p proche de 0% et pour une distribution gaussienne une valeur-p de 2.7% correspond à un seuil de 3σ .

Dans la pratique, nous estimons la distribution des valeurs au sein de chaque carte de bruit pour chaque champ et pour chaque filtre. Pour le filtre W2, nous choisissons la valeur-p correspondant au seuil que nous avons obtenu sur les données simulées. Pour le filtre W3 et W4, nous adaptons la valeur-p pour tenir compte de leur caractère dyadique. Nous avons ainsi $p_{W2} = 2.9 \times 10^{-7}$, $p_{W3} = 1.1 \times 10^{-6}$, $p_{W4} = 4.4 \times 10^{-6}$. Nous faisons ensuite la moyenne des seuils obtenus à partir des valeurs-p de chaque carte de bruit. Les seuils finaux correspondant à cette méthode de seuillage sont présentés dans la table 8.4 avec les propriétés (ng, aire effective) des différents champs.

Grâce à cette approche, les valeurs de seuils ne sont plus corrélés à la taille des champs. Ceci nous permet d'affirmer que nous avons gagné en stabilité sur le calcul des seuils. Ceci nous permet d'affirmer que nous avons gagné en stabilité sur le calcul des seuils.

Cependant, même si l'approche de la valeur-p permet d'augmenter le niveau de confiance sur les seuils, nous restons en partie dépendant de la taille des champs car le nombre de valeur au dessus des seuils que l'on fixe est faible, ce qui rend le calcul de la valeur-p encore un peu instable. L'application de cette approche sur des champs plus larges dans le cas des données sur les trois ans du relevé,

Table 8.4 – Récapitulatif des propriétés et des seuils par valeur-p mesurés pour chacun des champs d’observations de HSC présents dans les données S16A et pour les filtres ondelettes W2, W3, W4.

Champ observé	ng	Taille [deg ²]	Seuil W2	Seuil W3	Seuil W4
XMM	22.2	26.3	4.55	4.45	4.2
GAMA09H	21.6	28.52	5.99	4.51	4.23
WIDE12H	24.2	11.45	4.47	4.29	4.05
GAMA15H	23.2	27.50	4.63	4.49	4.19
HECTOMAP	17.4	9.48	4.48	4.35	4.09
VVDS	22.9	16.77	4.0	4.41	4.12

nous permettra de gagner davantage en précision sur les seuils.

En résumé, pour la procédure de seuillage sur les données réelles du relevé HSC, pour chaque champ les seuils sont calculés à partir d’une même valeur-p (voir table 8.4) qui nous assure un niveau de confiance cohérent entre les différents champs.

★ Identification des pics candidats

La dernière étape de notre procédure est l’identification des pics candidats. Une fois les cartes de convergence filtrées et seuillées pour les 6 champs, nous détectons les amas candidats en considérant les pixels dont la valeur est supérieure à la fois au seuil défini pour l’échelle et à leurs huit voisins (maximums locaux). Notons que cette approche diffère légèrement de celle présentée par [Miyazaki et al. \(2018\)](#), et au sein de laquelle les pics candidats correspondent à 3 pixels contigus dont la valeur dépasse le seuil fixé. Leur approche a le mérite de tenir compte de la taille des amas détectables au sein de la carte filtrée. Cependant, leur approche est hautement dépendante de la résolution choisie pour la carte filtrée. Dans le cas de [Miyazaki et al. \(2018\)](#), la résolution de la carte de sortie est de 0.25 arcmin/pixel, soit la moitié de notre résolution. Ainsi, notre approche pour l’identification des pics se révèle similaire.

En considérant les pixels dont la valeur est supérieure à ces huit voisins (maximums locaux), nous sommes impactés par la présence de données manquantes. Les zones de la carte en bordure de données manquantes étant peu fiables pour la détection, nous trions ensuite les détections en deux temps. Premièrement, nous ne considérons que les maximums locaux qui ne sont pas sur un pixel originellement masqué ou près des bords. Pour cela, nous érodons les cartes de convergence filtrées et masquées pour ne considérer que les régions non affectées par les données manquantes et en limiter l’impact. La méthode d’érosion consiste à masquer volontairement les bords dans un rayon donné. Dans un second temps, nous regroupons les détections qui sont plus proches que le rayon caractéristique du filtre à l’échelle où ils sont détectés. Cela permet d’éviter les détections multiples dues à la présence de données manquantes pouvant générer des maximums locaux multiples pour des amas de galaxies larges.

En résumé, nous avons adapté l’implémentation de la méthode multi-échelle afin de relever les défis posés par l’application aux données réelles HSC. Pour la procédure de seuillage, notre approche finale repose sur l’utilisation de la valeur-p adaptée entre les différentes échelles pour évaluer les seuils de chaque champ. Après avoir appliqué notre procédure sur les 6 champs disjoints des données S16A, nous nous intéressons maintenant à l’association des résultats qui forme le point critique de l’analyse des résultats.

8.4 Résultats

Une fois identifiées sur les cartes filtrées, érodées et seuillées, les détections doivent être validées. Sur les données simulées, nous avons vu en chapitre 5 la nécessité et la complexité de la pro-

cédure d'association même en disposant du catalogue des halos. Heureusement, nous disposons du catalogue d'amas optique CAMIRA sur lequel nous appuyer lors de l'étape de validation des résultats. Mais nous avons besoin d'une méthodologie rigoureuse afin d'éviter toutes fausses associations. Dans cette section, nous décrivons les détails de la procédure d'association utilisant le catalogue de détections optique CAMIRA. Puis, nous présentons nos premiers résultats en termes de détections en comparaison à la fois aux détections optique de CAMIRA et aux détections publiées par [Miyazaki et al. \(2018\)](#).

8.4.1 . Association des détections

L'association des détections est une étape clef mais délicate de la validation des détections obtenues dans le cas des données réelles. Contrairement au cas des données simulées, nous ne disposons plus du catalogue de halos de matière noire correspondant aux détections. Dans cette sous-section, nous présentons la procédure d'association que nous avons adopté pour la validation de notre méthode sur des données réelles.

L'approche la plus simple consiste à comparer les détections obtenues par effet de lentille gravitationnelle faible avec des catalogues d'amas existants, pouvant être obtenus par d'autres sondes ou à d'autres longueurs d'ondes. Dans le cas du relevé HSC-SSP, l'un des objectifs scientifiques principaux était la détection optique des amas de galaxies. Ainsi, comme l'on fait [Miyazaki et al. \(2018\)](#); [Hamana et al. \(2020\)](#), nous avons comparé nos détections avec le catalogue de détections optiques CAMIRA ([Oguri, 2014](#)) réalisé sur le même jeu de données à partir des amas détectés optiquement par leur séquence rouge (voir Sect. 1.3.2).

Dans la pratique, nous complétons notre procédure d'association à l'aide d'un second catalogue de détections optiques CAMIRA, baptisé S20A. Ce catalogue est légèrement différent car il est obtenu sur les données des trois premières années du relevé et est également plus complet. Étant donné que les données peuvent avoir évolué nous utilisons ce catalogue uniquement pour vérifier dans un second temps les détections qui n'ont pas de correspondances au sein du catalogue CAMIRA S16A.

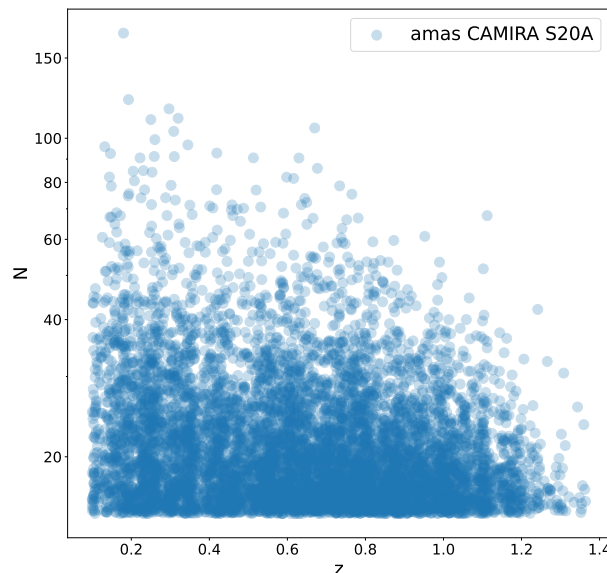


Figure 8.11 – Distribution des amas du catalogue CAMIRA S20A dans le plan redshift-richesse.

Pour la procédure d'association, nous adoptons la procédure décrite en section 5.3 en utilisant les catalogues CAMIRA. Nous sélectionnons au sein de ces catalogues uniquement les amas dont le redshift est inférieur à 1.1 pour réduire les associations hasardeuses. Contrairement aux catalogues de halos de la simulation DUSTGRAIN, ces catalogues ne contiennent pas le rayon caractéristique R_{200c} des amas. Nous disposons par contre de la richesse et du redshift des amas, comme illustrée par la

figure 8.11. Dans la pratique, nous associons les détections avec l'amas le plus proche en termes de distance projetée. La procédure d'association est appliquée ainsi pour chacune des cartes ondelettes W_2 , W_3 et W_4 et pour les 6 champs. Cette procédure pourra également être affinée en revenant à la masse des amas.

Notons aussi que cette approche peut être élargie à d'autres catalogues externes. En particulier, on s'attend à avoir des biais de sélection différents entre les méthodes de détection adaptées de l'effet de lentille faible et les méthodes de détection optiques. Miyazaki et al. (2018) utilisent également le catalogue d'amas détectés par rayons X, XXL (Pierre et al., 2016) pour valider les détections qui n'auraient pas de correspondances avec CAMIRA. Ainsi, dans une étude complémentaire en vue d'une publication, nous comptons utiliser des catalogues d'amas de galaxies complémentaires détectés à d'autres longueurs d'ondes.

8.4.2 . Résultats des détections associées

Dans la section précédente, nous décrivons comment nous avons filtré et seuillé la carte de convergence de mode E pour chacun des champs d'observations, puis détecté avec précaution les pics candidats. En suivant la procédure d'association que nous venons de décrire et à partir des amas des catalogues CAMIRA S16A puis S20A, nous associons les détections de la méthode multi-échelle sur les données HSC. Dans cette sous-section, nous présentons les détections obtenues que nous comparons avec les détections déjà publiées (Miyazaki et al., 2018).

Comme nous l'avons précisé, des algorithmes très performants ont déjà été utilisés sur ces données HSC pour la détection des amas par leur effet de lentille faible. En particulier, l'étude menée par Miyazaki et al. (2018) qui utilisent les mêmes données en appliquant la méthode de masse d'ouverture directement aux positions des galaxies. Leur publication contient à la fois les cartes de convergence filtrée et le catalogue des amas détectés avec leur position que nous utilisons pour la comparaison présentée dans cette sous-section.

Ainsi, nous affichons dans les figures 8.12 à 8.17, nos résultats. Pour les six champs, les détections obtenues par la méthode multi-échelle sont représentées soit par un cercle rouge en cas d'association avec un amas CAMIRA. Les amas détectés par CAMIRA sont représentés par un triangle dont la couleur indique le redshift. Les carrés noirs correspondent aux détections sans contreparties optiques au sein des deux catalogues. Enfin, les détections publiées par Miyazaki et al. (2018) sont représentées par un cercle vert. Dans la légende, les nombres entre parenthèses dénotent le nombre de détections obtenues avec la méthode multi-échelle dans chaque catégorie.

Les figures 8.12 à 8.17 mettent en lumière les détections obtenues sur les 6 champs disjoints des données S16A du relevé HSC. Les résultats d'associations pour ces détections sont ensuite résumés dans la Table 8.5. Notons que la seconde colonne correspond à la taille exploitable en deg^2 (après érosion) de chacun des 6 champs. Ce calcul retire les zones masquées par les étoiles brillantes mais aussi le voisinage de ces zones où le signal est sujet aux effets de bords. Ensuite, nous détaillons les résultats de l'association avec les catalogues CAMIRA S16A puis S20A (quatrième colonne) et la pureté relative à ces catalogues (cinquième colonne). Enfin, nous comparons pour chaque champ observé, la taille, le nombre de détections obtenues (troisième colonne) que nous mettons en parallèle du nombre de détections publiées (sixième colonne).

Sur les 6 champs, nous observons que la méthode multi-échelle a produit significativement plus de détections (120) que la méthode utilisée pour les détections publiées (65) sur le même jeu de données. Malgré ce nombre très important de détections, la méthode multi-échelle recouvre 80% (52) des détections publiées. Cet écart s'explique à la fois par l'érosion de la carte qui empêche la détection de certains pics au niveau des bords ou des données manquantes (7) et par les différences entre les méthodes de filtrage utilisées.

Parmi les 120 détections obtenues par la méthode multi-échelle, 98 ont une correspondance optique dans les catalogues CAMIRA ce qui se traduit par une pureté de 82% pour nos détections. Cette

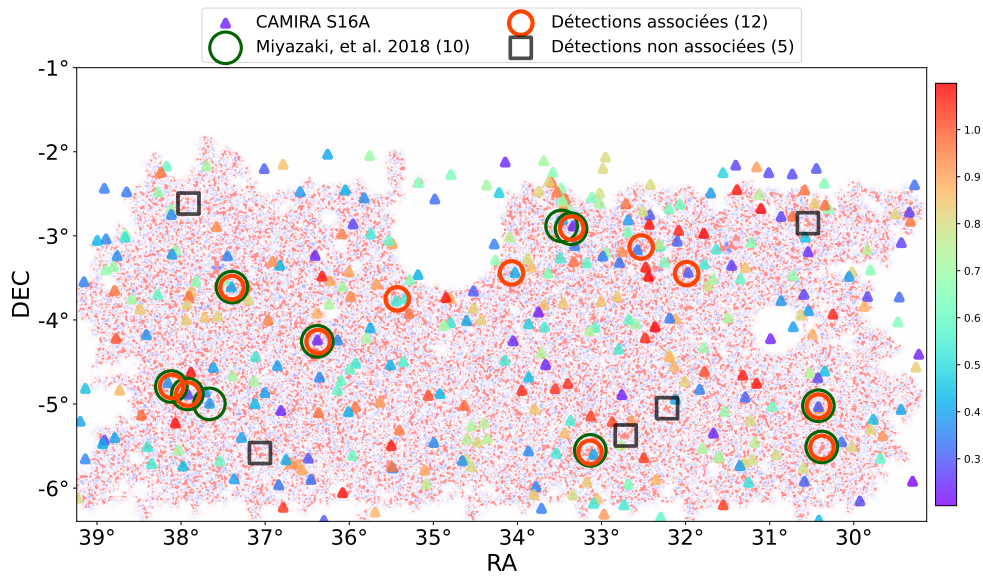


Figure 8.12 – Mode E de la carte de convergence du champ XMM des données HSC, filtrée par la méthode multi-échelle avec les ondelettes W_2 , W_3 et W_4 . Les détections réalisées par l'approche multi-échelle sont représentées soit par un cercle rouge quand elles sont associées, soit par un carré noir quand elles n'ont pas de correspondances CAMIRA. Les triangles correspondent aux amas du catalogue CAMIRA et leur couleur fait référence au redshift estimé entre 0.2 et 1.1, comme l'illustre le gradient de couleur.

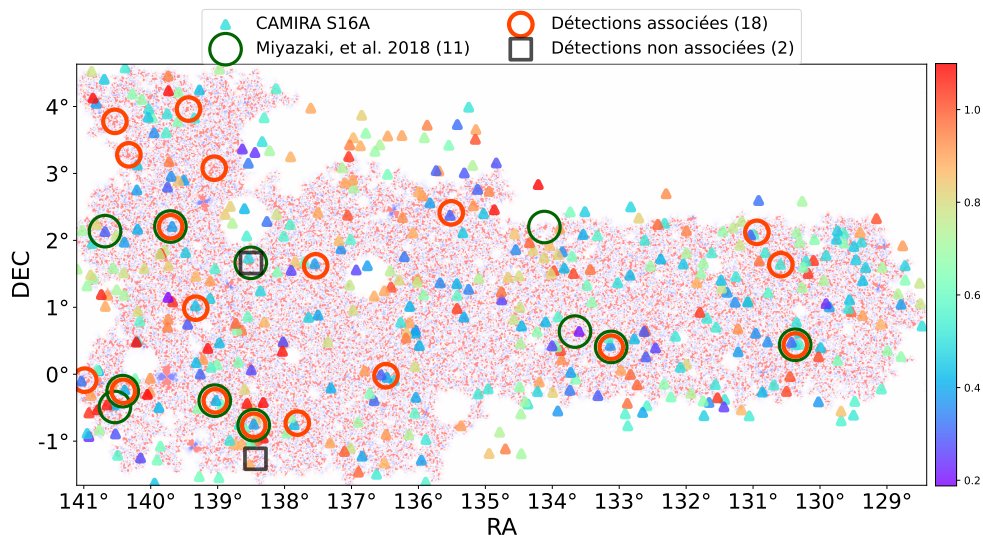


Figure 8.13 – Identique à la Figure 8.12 mais sur le champ GAMA09H.

pureté est en partie sous-estimée car nous trouvons que sur les champs GAMA09H et VVDS, 3 détections non associées avec CAMIRA correspondent à des détections publiées par [Miyazaki et al. \(2018\)](#) car associées avec des amas du catalogue XXL. Il est intéressant de noter que cette pureté globale reste relativement cohérente avec la pureté mesurée dans le cas des données simulées (85%) présenté dans le chapitre 6 laquelle est par ailleurs également sous-estimée.

Comme dans le chapitre 6, la pureté globale peut à nouveau être considérée comme faible compte tenu de l'utilisation des seuils très conservatifs. Nous affinons cette analyse par le calcul de la pureté obtenue pour chacun des champs (quatrième colonne). On constate des écarts importants entre les différents champs qui expliquent que notre pureté globale mesurée soit relativement faible. En effet, les champs d'observations HECTOMAP, XMM et VVDS présentent un grand nombre de détections non associées et une pureté plus faible (67%, 71% et 78%) comparées aux autres champs d'observations dont la pureté surpasse 82%. La qualité inégale des détections entre les différents champs peut s'expliquer par trois facteurs concomitants :

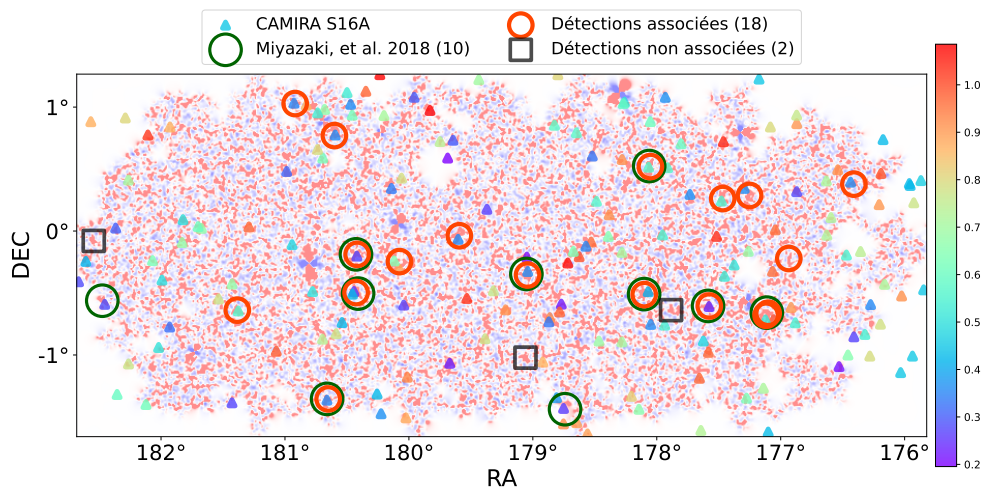


Figure 8.14 – Identique à la Figure 8.12 mais sur le champ WIDE12H.

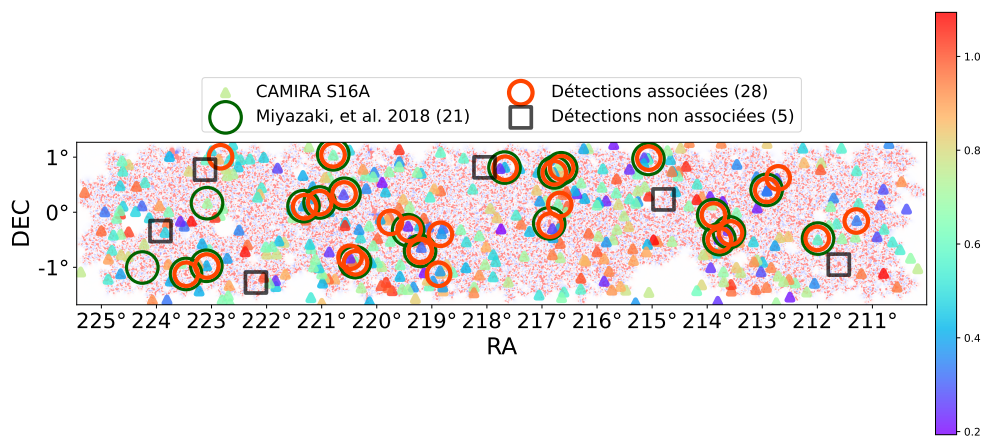


Figure 8.15 – Identique à la Figure 8.12 mais sur le champ GAMA15H.

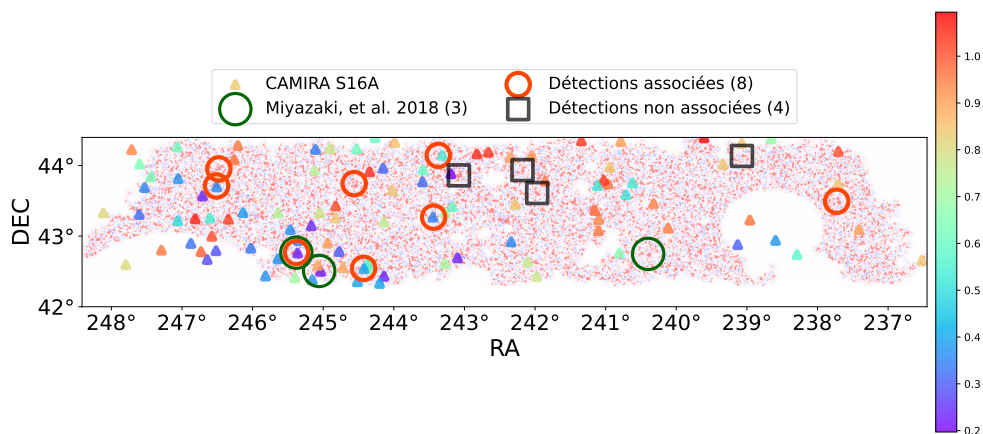


Figure 8.16 – Identique à la Figure 8.12 mais sur le champ HECTOMAP.

- La faible taille des champs entraîne des effets statistiques qui réduisent la précision des seuils calculés. Ce premier facteur s'observe tout particulièrement dans la table 8.4 où les seuils W_2 mesurés sont plus faibles pour les champs VVDS, XMM et HECTOMAP, introduisant par conséquent des fausses détections qui affectent la pureté.
- Les conditions d'observations, la densité de galaxies et la proportion de données manquantes des différents champs peuvent augmenter ou diminuer les performances sur les champs. Ainsi le champ XMM qui a bénéficié des pires conditions d'observations d'après [Mandelbaum et al.](#)

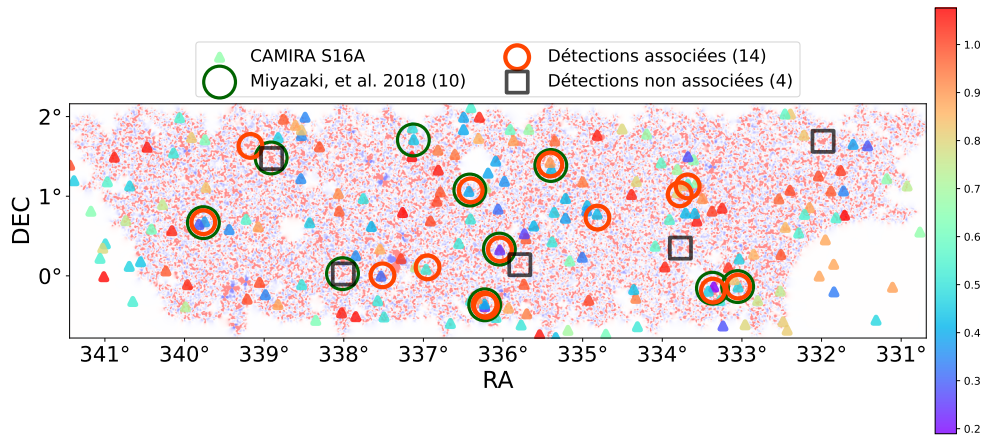


Figure 8.17 – Identique à la Figure 8.12 mais sur le champ VVDS.

Table 8.5 – Detections obtenues pour chacun des champs d’observations de HSC présents dans les données S16A et pour les filtres ondelettes W2, W3, W4. Ces detections sont associées avec les catalogues d’amas optiques CAMIRA S16A puis S20A et les résultats sont comparés avec les detections publiées par Miyazaki et al. (2018).

Champ observé	Taille [deg ²]	Detections	Associées	Pureté [%]	Publiées
XMM	26.30	17	12	71	10
GAMA09H	28.52	20	18	90	11
WIDE12H	11.45	20	18	90	10
GAMA15H	27.50	33	28	85	21
HECTOMAP	9.48	12	8	67	3
VVDS	16.77	18	14	78	10
Total	120.02	120	98	82	65

(2018) présente une pureté assez faible au contraire de WIDE12H qui a bénéficié de bonnes conditions d’observations.

- La complétude des catalogues CAMIRA peut également entraîner un biais sur les résultats des méthodes. On peut voir notamment que certaines detections n’ont pas de correspondances optiques au sein des catalogues CAMIRA mais ont une correspondance avec des amas détectés dans XXL.

Finalement, nous avons montré que la méthode multi-échelle proposait des résultats prometteurs sur les données réelles en atteignant une pureté surpassant 84% sur les champs XMM, GAMA09H, et GAMA15H. Par ailleurs, nous obtenons également environ le double du nombre des detections publiées sur l’ensemble des 6 champs. Ces résultats confirment que les défis soulevés dans la section 8.2 ont été approchés. Cependant, une marge d’amélioration est laissée sur la procédure de seuillage, significativement impactée par la taille du champ et les caractéristiques du bruit. Cet impact se retrouve dans les performances mesurées des champs de plus petites tailles offrant une pureté inférieure aux attentes.

8.5 Perspectives

Dans ce chapitre, nous avons présenté les résultats préliminaires de la méthode multi-échelle sur les données HSC publiquement accessibles, à savoir la première année de données.

Les données HSC que nous utilisons se décomposent en six champs disjoints pour lesquels nous mesurons une densité de galaxie moyenne de ~ 22 galaxies/arcmin, permettant la détection des amas de galaxies par effet de lentille gravitationnelle faible. Pour ces six champs d’observation, nous avons

accès aux catalogues de galaxies sources contenant les ellipticités mesurées, les facteurs correctifs et les poids associés aux galaxies et leur redshift photométrique.

Nous avons tout d'abord, introduit les différents défis que soulève l'application de la méthode multi-échelle aux données réelles. Nous avons identifié trois défis principaux qui ont un impact significatif sur la reconstruction des cartes de convergence, et sur la procédure de seuillage et en conséquence les détections obtenues :

- Les données manquantes,
- Le bruit et ses non-gaussianités,
- La distribution non uniforme des galaxies.

Pour faire face à ces défis, nous avons adapté l'implémentation de la méthode multi-échelle à la fois lors de la reconstruction des cartes de convergence en tenant compte du masque mais également dans la procédure de seuillage. Nous avons ensuite estimé pour chaque champ par la méthode de randomisation des ellipticités (400 réalisations), le niveau de bruit dans chacun des pixels de la carte. Le seuillage a alors été conduit sur des cartes de signal sur bruit filtrées, tenant compte ainsi de la densité de galaxie présente dans chaque pixel.

Les détections obtenues sur les cartes de signal sur bruit filtrés et seuillées montrent que la méthode multi-échelle permet d'obtenir 120 détections dont 98 avec une correspondance optique dans les catalogues CAMIRA S16A et S20A. Nous obtenons ainsi presque le double des amas de galaxies détectés par effet de lentille gravitationnelle publiés sur ce même jeu de données. En calculant la pureté mesurée sur chacun des six champs disjoints des données HSC, nous retrouvons des performances similaires à l'application de la méthode de détection multi-échelle sur des données simulées. Les résultats présentés sont une première ébauche de nos performances sur des données réelles. Notre approche doit encore être affinée en particulier concernant l'estimation des seuils en utilisant une valeur-p pour calculer plus finement le seuil à chaque échelle. Par ailleurs, les détections sans correspondances optiques au sein des catalogues CAMIRA obtenues devront être comparées à d'autres catalogues d'amas à partir d'autres longueurs d'onde (e.g. eROSITA, XXL).

Pour la suite, quatre directions sont envisagées en parallèle pour faire progresser notre approche :

- Notre étude a montré que le calcul des seuils à partir du calcul des maximums qui est utilisé par de nombreux auteurs [Hamana et al. \(2012\)](#); [Miyazaki et al. \(2018\)](#); [Oguri et al. \(2021\)](#) est très instable et se révèle être impacté par la taille des champs. Nous avons amélioré la méthode de seuillage en définissant le seuil ($k\sigma$), non pas à partir du maximum de la carte de mode B, mais comme le pourcentage de valeur qui doit se situer à plus de $k\sigma$ de la moyenne. Cette méthode est plus stable par rapport au maximum et permet d'avoir un seuil plus robuste face à la présence de non-gaussianités. Cependant, la définition précise de ce pourcentage (appelé valeur-p) pour chaque échelle à besoin d'être améliorée. Le calcul du seuil qui en découle devra se faire sur des champs de plus grandes tailles pour réduire les effets statistiques.
- Ensuite, le traitement des données manquantes lors de la pixellisation et lors de la reconstruction peut être approfondi. En particulier, l'application de la méthode dite d'inpainting permettrait d'augmenter notre précision ([Pires et al., 2009](#); [Pires et al., 2020](#)). Le but de cette méthode est de reconstruire par interpolation les pixels manquants et ainsi permettre de compenser l'effet des masques sur le signal.
- Nous aimerions tenir compte du redshift des galaxies afin de réduire l'effet de dilution, comme décrit par [Hamana et al. \(2020\)](#). Une première approche de cette idée a été développée dans le cadre de la préparation aux données d'EUCLID.
- A leur sortie, nous appliquerons cette procédure aux données des trois premières années du relevé HSC. À ce titre, les résultats présentés ici peuvent être vues comme une prémisse pour l'étude HSC et plus généralement pour les données réelles à venir, et notamment le relevé *Euclid*. *Euclid* et HSC étant comparables en termes de profondeur et de qualité d'images, les approches développées dans ce chapitre pourront facilement s'étendre aux données *EUCLID*.

*
* *

Conclusions



Excellents traceurs de l'évolution des structures dans la toile cosmique, les amas de galaxies se sont imposés comme de puissants outils pour contraindre le modèle cosmologique. Leur abondance pour une masse et un redshift donnés, connue sous le nom de fonction de masse, nous donne une méthode directe pour contraindre les paramètres cosmologiques. Pour tirer profit de leur potentiel, il est nécessaire de développer des méthodes capables de les détecter. L'effet de lentille gravitationnelle faible est la seule méthode permettant la détection des amas selon leur distribution totale de masse, se libérant ainsi des effets de sélection propres à la matière baryonique.

Pour autant, l'utilisation de cet effet est restée limitée du fait du manque de relevés optiques alliant à la fois une forte densité de galaxies et un large champ. Aujourd'hui, la sensibilité des relevés optiques à grand champ permet une détection aveugle et efficace des amas de galaxies grâce à l'effet de lentille gravitationnelle faible. Le lancement en Juillet 2023 du télescope spatial *EUCLID* et l'arrivée prochaine de ses données et de celles de l'observatoire Vera C. Rubin, et du futur télescope spatial Nancy Grace Roman permettront d'améliorer davantage cette réalité. Cependant, la construction de catalogues d'amas de galaxies sélectionnés par lentillage faible nécessite l'amélioration ou le développement de nouveaux algorithmes de détection d'amas de galaxies capables de traiter le volume de données sans précédents de ces prochains relevés.

Ce manuscrit résume le travail effectué durant ces trois années de thèse sur la détection des amas de galaxies par effet de lentille gravitationnelle faible. Dans ce cadre, nous avons développé une nouvelle approche multi-échelle rapide et basée sur la décomposition par ondelettes. Le développement de cette méthode, baptisée W234, s'est composé en trois actes :

- ★ Implémentation et caractérisation de l'algorithme de détection multi-échelle et des méthodes de détection comparément utilisées;
- ★ Développement d'une procédure d'association pour permettre une comparaison impartiale de ces méthodes;
- ★ Application de la méthode à des données simulées puis à des données réelles.

Le premier enjeu de cette thèse porte sur la mise en place d'un algorithme complet de détection des amas de galaxies par leur signal de lentillage faible. Nous avons tout d'abord mené une étude portant sur la manière optimale de traiter les données issues des catalogues de galaxies sources avant le processus de filtrage. Il en ressort que distribuer les données sur une grille régulière en additionnant les valeurs des ellipticités permet un traitement à la fois rapide des catalogues et sans impact significatif sur le niveau de signal dans les cartes reconstruites. De plus, nous avons démontré que l'utilisation des cartes de convergence offre le même niveau de signal que les cartes de cisaillement, tout en étant 4 fois plus rapides à traiter lors du processus de filtrage. Afin de pouvoir comparer et évaluer précisément les performances de notre nouvelle méthode de détection multi-échelle, nous avons implémenté deux méthodes couramment utilisées : le filtrage gaussien et la méthode de masse d'ouverture. Après avoir introduit les limitations du filtre gaussien sur la contribution au signal des structures à grandes échelles, nous avons opté pour l'utilisation de 5 filtres de masse d'ouverture S96, S98, TANH, J04, M18. Ces filtres ont l'avantage d'être à la fois complémentaires en termes de sélection et d'avoir déjà prouvé leur potentiel. Afin de comparer les différentes méthodes de filtrage,

nous avons utilisé le même algorithme de détection, en ne changeant que le filtre. Dans le cas de notre méthode multi-échelle, nous avons affiné cette algorithme en l'appliquant à chaque échelle ondelette avant de recombinaison les détections obtenues. Notre première comparaison des différentes méthodes de filtrage en termes de détection a montré qu'à seuillage identique, la méthode multi-échelle W₂₃₄ parvient déjà à récupérer $92.7\% \pm 0.6\%$ des détections obtenues par les filtres dont le comportement est proche d'une ondelette (Sg6, J04).

Le second enjeu de cette thèse porte sur la mise en place d'une procédure d'association impartiale aussi bien pour les méthodes de filtrages à une échelle donnée que pour notre filtrage multi-échelle. Cette procédure est essentielle afin de valider et comparer les détections issues des différentes méthodes. Pour cela, nous avons implémenté une procédure d'association adaptative qui ne pénalise ou n'avantage aucune méthode en s'adaptant aux spécificités de celle-ci. Cette approche calcule la distance maximale d'association correspondant à chaque filtre afin de tenir compte de la dispersion et des imprécisions sur les positions des détections obtenues. Puis, en tenant compte de l'échelle d'où proviennent les détections communes, nous avons pu adapter la procédure d'association aux défis de l'approche multi-échelle. De plus, dans le cas de notre application à des données simulées pour lesquelles nous avons les catalogues de halos de matière noire à notre disposition, les caractéristiques du catalogue peuvent avoir un effet considérable sur le nombre de fausses associations. C'est pourquoi nous avons mis en place une méthode de présélection des halos basée sur leur détectabilité. Nous avons choisi d'accéder à cette détectabilité par le calcul du rapport signal-sur-bruit (SNR) théorique des halos puis en conservant uniquement les halos dont le SNR est supérieur à 2.

Enfin, dans cette thèse, l'application et l'analyse des performances de la méthode multi-échelle se décomposent en trois parties :

- Nous avons d'abord comparé les performances des méthodes de détection sur 256 champs simulés de $5^\circ \times 5^\circ$ issus de la simulation DUSTGRAIN-*pathfinder* en utilisant la procédure d'association que nous avons développé. À travers les outils statistiques de pureté et de complétude, nous avons réalisé une première comparaison des performances moyennes obtenues sur les 256 champs par chacune des méthodes. Pour notre méthode W₂₃₄, 82% des détections sont correctement associées, presque exclusivement avec un unique halo, tout en assurant un nombre de halos incorrectement associés négligeable ($< 1\%$). Ces résultats valident ainsi l'efficacité de notre procédure d'association adaptative. Nous avons montré ensuite que parmi les détections non associées ($\sim 17\%$), seulement 10% correspondent à des fausses détections et les $\sim 90\%$ restants correspondent bien à du signal. En affinant l'analyse, nous avons montré que la méthode multi-échelle détecte pratiquement tous les halos identifiés par les autres approches, et obtient même au moins 25% de détections supplémentaires, en particulier à des masses plus faibles et à des redshifts plus élevés. En comparaison aux méthodes de détection standards, nous avons démontré avec succès que l'approche multi-échelle donne strictement de meilleurs résultats en termes de pureté et de complétude tout en recouvrant les détections.
- Dans un second temps, nous avons complété cette analyse par l'application de notre approche dans le cadre du challenge WLC organisé dans le consortium *EUCLID* dans le but de définir les méthodes de détections d'amas par les lentilles faibles qui seront utilisées sur les données *EUCLID*. Dans ce cadre, nous avons à la fois soumis 4 méthodes de détection parmi les 10 méthodes puis nous avons contribué à la partie : analyse des résultats. Co-chargé de l'analyse des résultats, nous avons appliqué notre procédure d'association adaptative sur toutes les méthodes de détections pour les caractériser et les comparer en termes d'associations, de complétude, de pureté mais aussi de détections communes. Cela nous a permis d'observer l'apparition de biais sur le calcul de la distance maximale d'association ($\sim 20\%$ d'écart dus aux manques de statistiques) et d'en comprendre les limitations. Cette analyse a montré à nouveau que la méthode de détection W₂₃₄ surpasse les autres approches en matière de complétude (14.3%) pour une même valeur de pureté (90%) et cette dominance se maintient également dans l'évolution de la complétude en fonction de la pureté. Par ailleurs, les résultats comparés des différentes

méthodes dans le plan redshift-masse fait ressortir le caractère complémentaire de certaines méthodes. Ainsi, pour l'application aux données *EUCLID*, nous avons préconisé l'application des méthodes de détection W234, UPCT et O21.

- Nous concluons la présentation des performances de notre nouvelle approche par son application aux données réelles de la première année d'observation du relevé Hyper Suprime-Cam, réparties en six champs disjoints. Lors de cette application, nous avons identifié et détaillé les différents défis apportés par l'utilisation de données réelles : les données manquantes, le bruit non-gaussien et la distribution non homogène de galaxies. Ces défis nous ont permis de confronter et d'étudier la robustesse statistique des différentes étapes de notre algorithme de détection. En conséquence, nous avons adapté la méthode multi-échelle lors de la reconstruction des cartes de convergence en tenant compte du masque. En parallèle, nous avons amélioré la procédure de seuillage en estimant le niveau de bruit par pixel sur chacun des six champs par la méthode de randomisation des ellipticités puis en adaptant le seuil en conséquence. Notre méthode de détection identifie 120 pics de surdensités dont 98 ont une correspondance optique directe avec les catalogues CAMIRA S16A et S20A. En comparaison aux résultats publiés sur le même jeu de données, l'approche multi-échelle double le nombre de détections obtenues tout en maintenant des performances similaires à celles observées lors de l'application de la méthode de détection multi-échelle sur des données simulées.

Ces trois analyses complémentaires nous ont permis de montrer le potentiel de notre nouvelle méthode de détection multi-échelle à la fois sur des données simulées idéales mais aussi sur des données réelles. Les résultats que nous avons montré durant ces analyses rendent compte du caractère très prometteur de la méthode que nous avons introduit ici.

*
* *



Dans cette section, nous présentons les pistes de réflexion et travaux personnels motivés par les résultats de ces trois années de thèse. Ces pistes, partiellement introduites tout au long de ce manuscrit, sont organisées selon trois axes de recherche :

- Amélioration de notre approche multi-échelle ;
- Application aux données *EUCLID* et adaptation aux données réelles ;
- Lier le signal à la masse.

Amélioration de la méthode de détection multi-échelle

★ La procédure de seuillage

L'une des premières pistes d'amélioration possibles porte sur la procédure de seuillage. En suivant les analyses présentées par [Hamana et al. \(2012\)](#); [Miyazaki et al. \(2018\)](#), nous avons fixé le maximum de la carte de convergence de mode B filtrée pour définir un seuil conservatif. Dans le chapitre 8, notre étude a montré que ce choix de seuil est instable et varie en fonction de la taille des champs étudiés. Nous avons alors mis en place une première amélioration utilisant le seuil ($k\sigma$) défini, non pas à partir du maximum des cartes de mode B, mais à partir du pourcentage de valeur qui doit se situer à plus de $k\sigma$ de la moyenne. Cette valeur-p a montré des résultats préliminaires prometteurs mais son calcul doit encore être affiné par une procédure de calibration. Pour réaliser cette calibration, nous comptons élargir notre étude à des champs de plus grande taille soit à partir de relevés plus récents soit en augmentant le nombre de réalisations de bruit.

★ La définition d'un rapport signal-sur-bruit associé aux détections ondelettes

Au chapitre 7, nous avons montré que la principale limite actuelle de la méthode de détection multi-échelle réside dans le manque de définition précise de rapport signal-sur-bruit (SNR) pour les détections réalisées. Ce problème est intrinsèque à l'utilisation de plusieurs filtres qui intègrent chacun une partie du signal d'intérêt tout en ayant un niveau de bruit apparent qui varie entre les échelles. Pour définir correctement un SNR, il est essentiel de prendre en compte cette répartition du signal entre les échelles. La recombinaison des détections communes aux échelles accentue encore ce problème. Afin de pallier à ce manque, nous envisageons de calibrer le SNR à partir de la pureté mesurée avant l'étape de recombinaison. En d'autres termes, nous souhaitons utiliser la pureté à chaque échelle pour redéfinir le niveau du SNR des détections présentes à cette échelle. Cette calibration nous permettrait de définir un SNR pour chaque détection tout en restant cohérent entre les différentes échelles. Pour quantifier ces effets de sélection, la calibration pourrait être couplée avec l'utilisation d'amas de galaxies simulés pour lesquels nous définissons le SNR théorique et les étudions après filtrage. Une autre possibilité est de calculer le SNR à partir d'amas de galaxies simulés après filtrage.

★ L'approche tomographique

Dans ce manuscrit, notre attention s'est focalisée majoritairement sur les méthodes et les filtres les plus couramment utilisés dans leur approche usuelle. D'autres implémentations alternatives ont récemment montré leur preuve dans la détection des amas de galaxies par l'effet de lentille gravitationnelle faible ([Hamana et al., 2020](#)). L'approche tomographique, en particulier, qui permet de sélection-

ner les sources les plus lointaines des amas s'est révélée très performante pour réduire l'impact de l'effet de dilution. Cet effet fait référence à la présence de galaxies à bas redshift non-membres des amas et dont le signal vient contaminer le signal lentille des amas. Le signal des amas est dilué par le signal de la galaxie situé à un redshift proche. Pour faire face à ce problème, [Hamana et al. \(2020\)](#) proposent une approche tomographique qui intègre l'information sur le redshift des galaxies sources en considérant différents intervalles de redshift des sources. Sur le même jeu de données HSC qui est présenté dans le chapitre 8, cette approche tomographique a poussé les résultats de 65 pics identifiés à un catalogue de 124 candidats avec $\nu > 5$ dont 107 avec une correspondance. Cela correspond à une amélioration des performances de [Miyazaki et al. \(2018\)](#) de presque 85% sur les données réelles. Nous sommes confiants que l'implémentation d'une approche tomographique dans le cas de la méthode multi-échelle puisse se révéler également très prometteuse. Tenir compte de cet effet de dilution est d'autant plus important qu'il constitue l'une des sources principales d'incertitudes systématiques lors de l'évaluation de la masse des amas ([Pratt et al., 2019](#)).

Application à *EUCLID* et adaptation aux données réelles

Sur la base des résultats obtenus durant ces trois années, nous souhaitons appliquer notre approche multi-échelle aux données actuelles et futures des relevés grands champs, ce qui permettra de construire de grands catalogues d'amas sélectionnés par effet de lentille gravitationnelle faible. Ainsi, tout d'abord, nous comptons appliquer notre méthode aux données des trois premières années du relevé HSC dès leur sortie. Pour la procédure d'association, il serait intéressant d'élargir la liste des catalogues utilisés et notamment en incluant d'autres longueurs d'ondes : redMaPPer ([Rykoff et al., 2014](#)), XXL ([Pierre et al., 2016](#)), et eRosita ([Merloni et al., 2012](#); [Boller et al., 2016](#)). En parallèle, nous aimerions appliquer également notre méthode multi-échelle sur les données provenant du télescope *EUCLID* en suivant l'implémentation que nous avons utilisé pour traiter les données HSC dont la profondeur et la qualité d'images sont similaires aux attentes d'*EUCLID*. Pour pleinement faire face aux défis de l'application aux données réelles à travers ces deux relevés, il est nécessaire d'adapter encore davantage notre approche pour le traitement des données manquantes. En particulier, nous souhaitons améliorer notre précision par l'application de la méthode dite d'inpainting ([Pires et al., 2009](#); [Pires et al., 2020](#)). Cette méthode permet de reconstruire les pixels manquants par interpolation et ainsi permettre de compenser l'effet des données manquantes sur les champs. La mise en place de cette approche pourrait nous permettre de détecter notamment les amas de galaxies aux bords des données manquantes, que nous ratons sur les six champs disjoints des données HSC alors qu'ils sont identifiés par [Miyazaki et al. \(2018\)](#).

Lier le signal à la masse

La masse des amas est le pilier essentiel pour sonder la cosmologie par les amas de galaxies. L'abondance des amas à une certaine masse pour un redshift donné est liée à la cosmologie sous-jacente et permet donc d'en contraindre les paramètres (voir Sect. 1.2.2). C'est pourquoi, notre prochaine ambition est d'utiliser les détections obtenues par la méthode multi-échelle afin de mesurer la masse des amas de galaxies et sonder la cosmologie. Pour lier le signal à la masse, nous souhaitons ajuster le profil de densité de masse surfacique sous l'hypothèse d'un profil NFW. Dans un premier temps, cette approche serait implémentée et étudiée sur les détections réalisées sur les données simulées. Dans un second temps, nous aimerions estimer les masses des détections obtenues sur les données HSC à notre disposition (1^{ère} ou 3^{ème} années). De plus, une telle approche ayant déjà été appliquée sur les données HSC ([Miyazaki et al., 2018](#); [Oguri et al., 2021](#)), les résultats publiés pourraient également servir de base de comparaison pour les masses mesurées par notre approche. Ensuite, nous envisageons d'appliquer cette estimation de la masse sur les données *EUCLID* et caractériser

les masses mesurées par effet de lentille faible en comparaison aux masses provenant des relations d'échelle des autres méthodes de détections (AMICO, PZWav). Enfin, à partir des masses obtenues pour les différents relevés, nous aimerions remonter à la cosmologie par la fonction de masse pour les détections par effet de lentille gravitationnelle faible.

*
* *

Mes publications

- **Leroy, G., Pires, S., Pratt, G. W., & Giocoli, C.** 2023, *Fast multiscale galaxy cluster detection with weak lensing : towards a mass-selected sample*, arXiv :2304.01812.
- **Manjon-Garcia, A., Leroy, G., Andreon, S, et al.** *Weak lensing selection of galaxy clusters : comparing the performance of several algorithms*, In prep.

Mes contributions

- **Conférence à France for EUCLID France - Symposium 10**, LPSC, Grenoble 2021, 17 - 19 novembre 2021.
- **Poster à la conférence STScI's 2022 Spring Symposium** : Galaxy Clusters 2022 : Challenging Our Cosmological Perspectives, 25 - 29 avril, 2022.
- **EUCLID Summer School**, Laboratoire Arago - Observatoire océanologique, Banyuls-sur-mer, France, 16 - 29 août 2022.
- **Conférence au Euclid Clusters SWG Meeting Tenerife**, Tenerife, Espagne, 26-29 septembre 2022.
- **Conférence à France for EUCLID - Symposium 11**, Institut d'Astrophysique de Paris, Paris, 30 novembre - 2 décembre 2022.
- **Conférence à Journées de la Société Française d'Astronomie et d'Astrophysique**, Strasbourg, 20 - 23 juin 2023.

Mes séminaires

- **Séminaire à Max Planck Institute for Extraterrestrial Physics** : "Fast multiscale weak lensing galaxy cluster detection : Towards a mass selected sample.", Garching, Allemagne, 20 février 2023.
- **Séminaire à l'Institut d'Astrophysique Spatiale** : "Fast multiscale weak lensing galaxy cluster detection : Towards a mass selected sample.", Orsay, France, 31 mai 2023.

Mes expérience d'enseignement

- **Stand grand Public Explor'espace**, Montrouge, France, 5 - 7 novembre 2021.
- **Projet Savanturiers** : "Pourquoi la nuit est noire?", Orsay, France, mars - avril 2021.

- **Encadrement de Travaux Pratiques de métrologie, d'électricité et d'électronique de 1^{ère} et 2^{ème} années**, 2021- 2022, Université Paris-Saclay - IUT d'Orsay, Gif-sur-Yvette.

*
* *

Bibliographie

- Abell, G. O. *The distribution of rich clusters of galaxies. A catalogue of 2712 rich clusters found on the National Geographic Society Palomar Observatory Sky Survey.* 1957.
- Abell, G. O., Corwin, Harold G., J., & Olowin, R. P. A Catalog of Rich Clusters of Galaxies. *ApJS*, 70 :1, May 1989. doi : 10.1086/191333.
- Achúcarro, A., Biagetti, M., Braglia, M., et al. Inflation : Theory and Observations. *arXiv e-prints*, art. arXiv :2203.08128, Mar. 2022. doi : 10.48550/arXiv.2203.08128.
- Aihara, H., Arimoto, N., Armstrong, R., et al. The Hyper Suprime-Cam SSP Survey : Overview and survey design. *PASJ*, 70 :S4, Jan. 2018a. doi : 10.1093/pasj/psx066.
- Aihara, H., Armstrong, R., Bickerton, S., et al. First data release of the Hyper Suprime-Cam Subaru Strategic Program. *PASJ*, 70 :S8, Jan. 2018b. doi : 10.1093/pasj/psx081.
- Allen, S. W., Evrard, A. E., & Mantz, A. B. Cosmological parameters from observations of galaxy clusters. *Annual Review of Astronomy and Astrophysics*, 49(1) :409–470, Sep 2011. ISSN 1545-4282. doi : 10.1146/annurev-astro-081710-102514. URL <http://dx.doi.org/10.1146/annurev-astro-081710-102514>.
- Amendola, L. & Tsujikawa, S. *Dark Energy : Theory and Observations.* 2010.
- Andreon, S. & Bergé, J. Richness-mass relation self-calibration for galaxy clusters. *A&A*, 547 :A117, nov 2012. doi : 10.1051/0004-6361/201220115. URL <https://doi.org/10.1051%2F0004-6361%2F201220115>.
- Andreon, S. & Hurn, M. A. The scaling relation between richness and mass of galaxy clusters : a Bayesian approach. *MNRAS*, 404(4) :1922–1937, June 2010. doi : 10.1111/j.1365-2966.2010.16406.x.
- Arnaud, M., Pointecouteau, E., & Pratt, G. W. Calibration of the galaxy cluster $M_{500}-Y_{\text{X}}$ relation with XMM-Newton. *A&A*, 474(3) :L37–L40, Nov. 2007. doi : 10.1051/0004-6361:20078541.
- Astier, P., Guy, J., Regnault, N., et al. The Supernova Legacy Survey : measurement of Ω_M , Ω_Λ and w from the first year data set. *A&A*, 447(1) :31–48, Feb. 2006. doi : 10.1051/0004-6361:20054185.
- Bartelmann, M. & Schneider, P. Weak gravitational lensing. *Physics Reports*, 340(4–5) :291–472, Jan 2001. ISSN 0370-1573. doi : 10.1016/s0370-1573(00)00082-x. URL [http://dx.doi.org/10.1016/S0370-1573\(00\)00082-X](http://dx.doi.org/10.1016/S0370-1573(00)00082-X).
- Bellagamba, F., Roncarelli, M., Maturi, M., & Moscardini, L. AMICO : optimized detection of galaxy clusters in photometric surveys. *MNRAS*, 473(4) :5221–5236, Feb. 2018. doi : 10.1093/mnras/stx2701.
- Bergé, J., Amara, A., & Réfrégier, A. Optimal Capture of Non-Gaussianity in Weak-Lensing Surveys : Power Spectrum, Bispectrum, and Halo Counts. *ApJ*, 712(2) :992–1002, Apr. 2010. doi : 10.1088/0004-637X/712/2/992.

- Birkinshaw, M. The Sunyaev-Zel'dovich effect. *Phys. Rep.*, 310(2-3) :97–195, Mar. 1999. doi : 10.1016/S0370-1573(98)00080-5.
- Böhringer, H. & Schartel, N. X-ray galaxy cluster studies for astrophysics and cosmology. *Astronomische Nachrichten*, 334(4-5) :482, Apr. 2013. doi : 10.1002/asna.201211886.
- Böhringer, H. & Werner, N. X-ray spectroscopy of galaxy clusters : studying astrophysical processes in the largest celestial laboratories. *A&A Rev.*, 18(1-2) :127–196, Feb. 2010. doi : 10.1007/s00159-009-0023-3.
- Böhringer, H., Chon, G., & Collins, C. A. The extended ROSAT-ESO Flux Limited X-ray Galaxy Cluster Survey (REFLEX II). IV. X-ray luminosity function and first constraints on cosmological parameters. *A&A*, 570 :A31, Oct. 2014. doi : 10.1051/0004-6361/201323155.
- Boller, T., Freyberg, M. J., Trümper, J., et al. Second ROSAT all-sky survey (2RXS) source catalogue. *A&A*, 588 :A103, Apr. 2016. doi : 10.1051/0004-6361/201525648.
- Bridle, S., Balan, S. T., Bethge, M., et al. Results of the GREAT08 Challenge : an image analysis competition for cosmological lensing. *MNRAS*, 405(3) :2044–2061, July 2010. doi : 10.1111/j.1365-2966.2010.16598.x.
- Bridle, S. L., Kneib, J. P., Bardeau, S., & Gull, S. F. Bayesian Galaxy Shape Estimation. In Natarajan, P., editor, *The Shapes of Galaxies and their Dark Halos*, pages 38–46, Mar. 2002. doi : 10.1142/9789812778017_0006.
- Chiu, I. N., Ghirardini, V., Liu, A., et al. The eROSITA Final Equatorial-Depth Survey (eFEDS). X-ray observable-to-mass-and-redshift relations of galaxy clusters and groups with weak-lensing mass calibration from the Hyper Suprime-Cam Subaru Strategic Program survey. *A&A*, 661 :A11, May 2022. doi : 10.1051/0004-6361/202141755.
- Clerc, N. & Finoguenov, A. X-ray cluster cosmology. In *Handbook of X-ray and Gamma-ray Astrophysics*, pages 1–52. Springer Nature Singapore, 2023. doi : 10.1007/978-981-16-4544-0_117-1. URL https://doi.org/10.1007%2F978-981-16-4544-0_117-1.
- Closson Ferguson, H., Armus, L., Borne, K., et al. The LSST Galaxies Science Collaboration : Nearby Groups and Clusters. In *American Astronomical Society Meeting Abstracts #213*, volume 213 of *American Astronomical Society Meeting Abstracts*, page 460.07, Jan. 2009.
- Cropper, M., Hoekstra, H., Kitching, T., et al. Defining a weak lensing experiment in space. *MNRAS*, 431 (4) :3103–3126, June 2013. doi : 10.1093/mnras/stt384.
- Cropper, M., Pottinger, S., Niemi, S., et al. VIS : the visible imager for euclid. In MacEwen, H. A., Fazio, G. G., Lystrup, M., et al., editors, *SPIE Proceedings*. SPIE, jul 2016. doi : 10.1117/12.2234739. URL <https://doi.org/10.1117%2F12.2234739>.
- Dalal, R., Li, X., Nicola, A., et al. Hyper Suprime-Cam Year 3 Results : Cosmology from Cosmic Shear Power Spectra. *arXiv e-prints*, art. arXiv :2304.00701, Apr. 2023. doi : 10.48550/arXiv.2304.00701.
- Dark Energy Survey Collaboration, Abbott, T., Abdalla, F. B., et al. The Dark Energy Survey : more than dark energy - an overview. *MNRAS*, 460(2) :1270–1299, Aug. 2016. doi : 10.1093/mnras/stw641.
- Davis, M., Efstathiou, G., Frenk, C. S., & White, S. D. M. The evolution of large-scale structure in a universe dominated by cold dark matter. *ApJ*, 292 :371–394, May 1985. doi : 10.1086/163168.
- de Jong, J. T. A., Verdoes Kleijn, G. A., Kuijken, K. H., & Valentijn, E. A. The Kilo-Degree Survey. *Experimental Astronomy*, 35(1-2) :25–44, Jan. 2013. doi : 10.1007/s10686-012-9306-1.

- Diemer, B. COLOSSUS : A Python Toolkit for Cosmology, Large-scale Structure, and Dark Matter Halos. *ApJS*, 239(2) :35, Dec. 2018. doi : 10.3847/1538-4365/aace8c.
- Diemer, B. & Joyce, M. An Accurate Physical Model for Halo Concentrations. *ApJ*, 871(2) :168, Feb. 2019. doi : 10.3847/1538-4357/aafad6.
- Driver, S. P., Hill, D. T., Kelvin, L. S., et al. Galaxy and Mass Assembly (GAMA) : survey diagnostics and core data release. *MNRAS*, 413(2) :971–995, May 2011. doi : 10.1111/j.1365-2966.2010.18188.x.
- Einstein, A. Die Feldgleichungen der Gravitation. *Sitzungsberichte der Königlich Preussischen Akademie der Wissenschaften*, pages 844–847, Jan. 1915.
- Euclid Collaboration, Adam, R., Vannier, M., et al. Euclid preparation. III. Galaxy cluster detection in the wide photometric survey, performance and algorithm selection. *A&A*, 627 :A23, July 2019. doi : 10.1051/0004-6361/201935088.
- Euclid Collaboration, Desprez, G., Paltani, S., et al. Euclid preparation. X. The Euclid photometric-redshift challenge. *A&A*, 644 :A31, Dec. 2020. doi : 10.1051/0004-6361/202039403.
- Euclid Collaboration, Schirmer, M., Jahnke, K., et al. Euclid preparation. XVIII. The NISP photometric system. *A&A*, 662 :A92, June 2022. doi : 10.1051/0004-6361/202142897.
- Euclid Collaboration, Bretonnière, H., Kuchner, U., et al. Euclid preparation. XXVI. The Euclid Morphology Challenge : Towards structural parameters for billions of galaxies. *A&A*, 671 :A102, Mar. 2023. doi : 10.1051/0004-6361/202245042.
- Foëx, G., Soucail, G., Pointecouteau, E., et al. The dark matter distribution in $z \sim 0.5$ clusters of galaxies. I. Determining scaling relations with weak lensing masses. *A&A*, 546 :A106, Oct. 2012. doi : 10.1051/0004-6361/201218973.
- Forman, W., Kellogg, E., Gursky, H., Tananbaum, H., & Giacconi, R. Observations of the Extended X-Ray Sources in the Perseus and Coma Clusters from UHURU. *ApJ*, 178 :309–316, Dec. 1972. doi : 10.1086/151791.
- Friedmann, A. Über die Möglichkeit einer Welt mit konstanter negativer Krümmung des Raumes. *Zeitschrift für Physik*, 21(1) :326–332, Dec. 1924. doi : 10.1007/BF01328280.
- Furusawa, H., Koike, M., Takata, T., et al. The on-site quality-assurance system for Hyper Suprime-Cam : OSQAH. *PASJ*, 70 :S3, Jan. 2018. doi : 10.1093/pasj/psx079.
- Gaia Collaboration, Prusti, T., de Bruijne, J. H. J., et al. The Gaia mission. *A&A*, 595 :A1, Nov. 2016. doi : 10.1051/0004-6361/201629272.
- Gardner, J. P., Mather, J. C., Clampin, M., et al. The James Webb Space Telescope. *Space Science Reviews*, 123(4) :485–606, Apr 2006. doi : 10.1007/s11214-006-8315-7. URL <https://doi.org/10.1007/s11214-006-8315-7>.
- Gavazzi, R. & Soucail, G. Weak lensing survey of galaxy clusters in the CfHTLS Deep. *Astronomy & Astrophysics*, 462(2) :459–471, Nov 2006. ISSN 1432-0746. doi : 10.1051/0004-6361:20065677. URL <http://dx.doi.org/10.1051/0004-6361:20065677>.
- Gilbank, D. G., Gladders, M. D., Yee, H. K. C., & Hsieh, B. C. The Red-sequence Cluster Survey-2 (RCS-2) : Survey Details and Photometric Catalog Construction. *AJ*, 141(3) :94, Mar. 2011. doi : 10.1088/0004-6256/141/3/94.

- Giocoli, C., Metcalf, R. B., Baldi, M., et al. Disentangling dark sector models using weak lensing statistics. *MNRAS*, 452(3) :2757–2772, Sept. 2015. doi : 10.1093/mnras/stv1473.
- Giocoli, C., Di Meo, S., Meneghetti, M., et al. Fast weak-lensing simulations with halo model. *MNRAS*, 470(3) :3574–3590, Sept. 2017. doi : 10.1093/mnras/stx1399.
- Giodini, S., Lovisari, L., Pointecouteau, E., et al. Scaling Relations for Galaxy Clusters : Properties and Evolution. *Space Sci. Rev.*, 177(1-4) :247–282, Aug. 2013. doi : 10.1007/s11214-013-9994-5.
- Hamana, T., Takada, M., & Yoshida, N. Searching for massive clusters in weak lensing surveys. *MNRAS*, 350(3) :893–913, May 2004. doi : 10.1111/j.1365-2966.2004.07691.x.
- Hamana, T., Oguri, M., Shirasaki, M., & Sato, M. Scatter and bias in weak lensing selected clusters. *MNRAS*, 425(3) :2287–2298, Sept. 2012. doi : 10.1111/j.1365-2966.2012.21582.x.
- Hamana, T., Sakurai, J., Koike, M., & Miller, L. Cosmological constraints from Subaru weak lensing cluster counts. *PASJ*, 67(3) :34, June 2015. doi : 10.1093/pasj/psv034.
- Hamana, T., Shirasaki, M., & Lin, Y.-T. Weak-lensing clusters from hsc survey first-year data : Mitigating the dilution effect of foreground and cluster-member galaxies. *Publications of the Astronomical Society of Japan*, 72(5), Jul 2020. ISSN 2053-051X. doi : 10.1093/pasj/psaa068. URL <http://dx.doi.org/10.1093/pasj/psaa068>.
- Hennawi, J. & Spergel, D. Shear selected cluster cosmology i : Tomography and optimal filtering. *The Astrophysical Journal*, 624, 04 2005. doi : 10.1086/428749.
- Hettterscheidt, M., Erben, T., Schneider, P., et al. Searching for galaxy clusters using the aperture mass statistics in 50 VLT fields. *A&A*, 442(1) :43–61, Oct. 2005. doi : 10.1051/0004-6361:20053339.
- Heymans, C., Van Waerbeke, L., Bacon, D., et al. The Shear Testing Programme - I. Weak lensing analysis of simulated ground-based observations. *MNRAS*, 368(3) :1323–1339, May 2006. doi : 10.1111/j.1365-2966.2006.10198.x.
- Hilbert, S., Hartlap, J., White, S. D. M., & Schneider, P. Ray-tracing through the Millennium Simulation : Born corrections and lens-lens coupling in cosmic shear and galaxy-galaxy lensing. *A&A*, 499(1) : 31–43, May 2009. doi : 10.1051/0004-6361/200811054.
- Hinshaw, G., Larson, D., Komatsu, E., et al. Nine-year Wilkinson Microwave Anisotropy Probe (WMAP) Observations : Cosmological Parameter Results. *ApJ*, 208(2) :19, Oct. 2013. doi : 10.1088/0067-0049/208/2/19.
- Hirata, C. & Seljak, U. Shear calibration biases in weak-lensing surveys. *MNRAS*, 343(2) :459–480, Aug. 2003. doi : 10.1046/j.1365-8711.2003.06683.x.
- Hubble, E. A Relation between Distance and Radial Velocity among Extra-Galactic Nebulae. *Proceedings of the National Academy of Science*, 15(3) :168–173, Mar. 1929. doi : 10.1073/pnas.15.3.168.
- Huterer, D. & Turner, M. S. Prospects for probing the dark energy via supernova distance measurements. *Phys. Rev. D*, 60(8) :081301, Oct. 1999. doi : 10.1103/PhysRevD.60.081301.
- Hwang, H. S., Geller, M. J., Park, C., et al. HectoMAP and Horizon Run 4 : Dense Structures and Voids in the Real and Simulated Universe. *ApJ*, 818(2) :173, Feb. 2016. doi : 10.3847/0004-637X/818/2/173.
- Ivezić, Ž., Kahn, S. M., Tyson, J. A., et al. LSST : From Science Drivers to Reference Design and Anticipated Data Products. *ApJ*, 873(2) :111, Mar. 2019. doi : 10.3847/1538-4357/ab042c.

- Jansen, F., Lumb, D., Altieri, B., et al. XMM-Newton observatory. I. The spacecraft and operations. *A&A*, 365 :L1–L6, Jan. 2001. doi : 10.1051/0004-6361:20000036.
- Jarvis, M., Bernstein, G., & Jain, B. The skewness of the aperture mass statistic. *Monthly Notices of the Royal Astronomical Society*, 352(1) :338–352, Jul 2004. ISSN 1365-2966. doi : 10.1111/j.1365-2966.2004.07926.x. URL <http://dx.doi.org/10.1111/j.1365-2966.2004.07926.x>.
- Kaiser, N. & Squires, G. Mapping the Dark Matter with Weak Gravitational Lensing. *ApJ*, 404 :441, Feb. 1993. doi : 10.1086/172297.
- Kaiser, N., Squires, G., & Broadhurst, T. A Method for Weak Lensing Observations. *ApJ*, 449 :460, Aug. 1995. doi : 10.1086/176071.
- Kawanomoto, S., Uraguchi, F., Komiyama, Y., et al. Hyper Suprime-Cam : Filters. *PASJ*, 70(4) :66, Aug. 2018. doi : 10.1093/pasj/psy056.
- Komiyama, Y., Obuchi, Y., Nakaya, H., et al. Hyper Suprime-Cam : Camera dewar design. *PASJ*, 70 :S2, Jan. 2018. doi : 10.1093/pasj/psx069.
- Kravtsov, A. V. & Borgani, S. Formation of Galaxy Clusters. *ARA&A*, 50 :353–409, Sept. 2012. doi : 10.1146/annurev-astro-081811-125502.
- Kuijken, K., Heymans, C., Hildebrandt, H., et al. Gravitational lensing analysis of the Kilo-Degree Survey. *MNRAS*, 454(4) :3500–3532, Dec. 2015. doi : 10.1093/mnras/stv2140.
- Laureijs, R., Amiaux, J., Arduini, S., et al. Euclid definition study report, 2011. URL <https://arxiv.org/abs/1110.3193>.
- Le Fèvre, O., Cassata, P., Cucciati, O., et al. The VIMOS VLT Deep Survey final data release : a spectroscopic sample of 35 016 galaxies and AGN out to $z \sim 6.7$ selected with $17.5 \leq i_{AB} \leq 24.75$. *A&A*, 559 :A14, Nov. 2013. doi : 10.1051/0004-6361/201322179.
- Leauthaud, A., Massey, R., Kneib, J.-P., et al. Weak Gravitational Lensing with COSMOS : Galaxy Selection and Shape Measurements. *ApJS*, 172(1) :219–238, Sept. 2007. doi : 10.1086/516598.
- Lemaître, G. Un Univers homogène de masse constante et de rayon croissant rendant compte de la vitesse radiale des nébuleuses extra-galactiques. *Annales de la Société Scientifique de Bruxelles*, 47 :49–59, Jan. 1927.
- Leonard, A., Pires, S., & Starck, J.-L. Fast calculation of the weak lensing aperture mass statistic. *Monthly Notices of the Royal Astronomical Society*, 423(4) :3405–3412, May 2012. ISSN 0035-8711. doi : 10.1111/j.1365-2966.2012.21133.x. URL <http://dx.doi.org/10.1111/j.1365-2966.2012.21133.x>.
- Leonard, A., Lanusse, F., & Starck, J.-L. Weak lensing reconstructions in 2D and 3D : implications for cluster studies. *MNRAS*, 449(1) :1146–1157, May 2015. doi : 10.1093/mnras/stv386.
- Li, X., Miyatake, H., Luo, W., et al. The three-year shear catalog of the Subaru Hyper Suprime-Cam SSP Survey. *PASJ*, 74(2) :421–459, Apr. 2022. doi : 10.1093/pasj/psac006.
- Li, X., Zhang, T., Sugiyama, S., et al. Hyper Suprime-Cam Year 3 Results : Cosmology from Cosmic Shear Two-point Correlation Functions. *arXiv e-prints*, art. arXiv :2304.00702, Apr. 2023. doi : 10.48550/arXiv.2304.00702.
- Limber, D. N. & Mathews, W. G. The Dynamical Stability of Stephan’s Quintet. *ApJ*, 132 :286, Sept. 1960. doi : 10.1086/146928.

- Mandelbaum, R., Miyatake, H., Hamana, T., et al. The first-year shear catalog of the Subaru Hyper Suprime-Cam Subaru Strategic Program Survey. *PASJ*, 70 :S25, Jan. 2018. doi : 10.1093/pasj/psx130.
- Massey, R., Heymans, C., Bergé, J., et al. The Shear Testing Programme 2 : Factors affecting high-precision weak-lensing analyses. *MNRAS*, 376(1) :13–38, Mar. 2007. doi : 10.1111/j.1365-2966.2006.11315.x.
- Massey, R., Hoekstra, H., Kitching, T., et al. Origins of weak lensing systematics, and requirements on future instrumentation (or knowledge of instrumentation). *MNRAS*, 429(1) :661–678, Feb. 2013. doi : 10.1093/mnras/sts371.
- Maturi, M., Meneghetti, M., Bartelmann, M., Dolag, K., & Moscardini, L. An optimal filter for the detection of galaxy clusters through weak lensing. *A&A*, 442(3) :851–860, Oct 2005. ISSN 1432-0746. doi : 10.1051/0004-6361:20042600. URL <http://dx.doi.org/10.1051/0004-6361:20042600>.
- Maturi, M., Bellagamba, F., Radovich, M., et al. Amico galaxy clusters in kids-dr3 : sample properties and selection function. *Monthly Notices of the Royal Astronomical Society*, 485(1) :498–512, 2019.
- Melchior, P., Viola, M., Schäfer, B. M., & Bartelmann, M. Weak gravitational lensing with DEIMOS. *MNRAS*, 412(3) :1552–1558, Apr. 2011. doi : 10.1111/j.1365-2966.2010.17875.x.
- Melin, J. B., Bartlett, J. G., Delabrouille, J., et al. The galaxy cluster $Y_{SZ} - L_X$ and $Y_{SZ} - M$ relations from the WMAP 5-yr data. *A&A*, 525 :A139, Jan. 2011. doi : 10.1051/0004-6361/201013999.
- Merloni, A., Predehl, P., Becker, W., et al. eROSITA Science Book : Mapping the Structure of the Energetic Universe. *arXiv e-prints*, art. arXiv :1209.3114, Sept. 2012. doi : 10.48550/arXiv.1209.3114.
- Messier, C. *Catalogue des nébuleuses et amas d'étoiles* », *La Connaissance des temps, ou connaissance des mouvements célestes*. Paris, Impression royale, 1784. URL <https://gallica.bnf.fr/ark:/12148/bpt6k6514280n/f235>.
- Miller, L., Kitching, T. D., Heymans, C., Heavens, A. F., & van Waerbeke, L. Bayesian galaxy shape measurement for weak lensing surveys - I. Methodology and a fast-fitting algorithm. *MNRAS*, 382 (1) :315–324, Nov. 2007. doi : 10.1111/j.1365-2966.2007.12363.x.
- Miyatake, H., Sugiyama, S., Takada, M., et al. Hyper Suprime-Cam Year 3 Results : Cosmology from Galaxy Clustering and Weak Lensing with HSC and SDSS using the Emulator Based Halo Model. *arXiv e-prints*, art. arXiv :2304.00704, Apr. 2023. doi : 10.48550/arXiv.2304.00704.
- Miyazaki, S., Hamana, T., Shimasaku, K., et al. Searching for Dark Matter Halos in the Suprime-Cam 2 Square Degree Field. *ApJ*, 580(2) :L97–L100, Dec. 2002. doi : 10.1086/345613.
- Miyazaki, S., Hamana, T., Ellis, R. S., et al. A Subaru Weak-Lensing Survey. I. Cluster Candidates and Spectroscopic Verification. *ApJ*, 669(2) :714–728, Nov. 2007. doi : 10.1086/521621.
- Miyazaki, S., Oguri, M., Hamana, T., et al. Properties of Weak Lensing Clusters Detected on Hyper Suprime-Cam's 2.3 deg² field. *ApJ*, 807(1) :22, July 2015. doi : 10.1088/0004-637X/807/1/22.
- Miyazaki, S., Komiyama, Y., Kawanomoto, S., et al. Hyper Suprime-Cam : System design and verification of image quality. *PASJ*, 70 :S1, Jan. 2018. doi : 10.1093/pasj/psx063.
- Miyazaki, S., Oguri, M., Hamana, T., et al. A large sample of shear-selected clusters from the hyper suprime-cam subaru strategic program s16a wide field mass maps. *Publications of the Astronomical Society of Japan*, 70(SP1), Dec 2018. ISSN 2053-051X. doi : 10.1093/pasj/psx120. URL <http://dx.doi.org/10.1093/pasj/psx120>.

- Mobasher, B., Capak, P., Scoville, N. Z., et al. Photometric Redshifts of Galaxies in COSMOS. *ApJS*, 172 (1) :117–131, Sept. 2007. doi : 10.1086/516590.
- More, S., Sugiyama, S., Miyatake, H., et al. Hyper Suprime-Cam Year 3 Results : Measurements of Clustering of SDSS-BOSS Galaxies, Galaxy-Galaxy Lensing and Cosmic Shear. *arXiv e-prints*, art. arXiv :2304.00703, Apr. 2023. doi : 10.48550/arXiv.2304.00703.
- Munari, E., Biviano, A., Borgani, S., Murante, G., & Fabjan, D. The relation between velocity dispersion and mass in simulated clusters of galaxies : dependence on the tracer and the baryonic physics. *MNRAS*, 430(4) :2638–2649, Apr. 2013. doi : 10.1093/mnras/stt049.
- Munshi, D., Valageas, P., van Waerbeke, L., & Heavens, A. Cosmology with weak lensing surveys. *Phys. Rep.*, 462(3) :67–121, June 2008. doi : 10.1016/j.physrep.2008.02.003.
- Murata, R., Nishimichi, T., Takada, M., et al. Constraints on the Mass-Richness Relation from the Abundance and Weak Lensing of SDSS Clusters. *ApJ*, 854(2) :120, Feb. 2018. doi : 10.3847/1538-4357/aaaab8.
- Narayan, R. & Bartelmann, M. Lectures on Gravitational Lensing. *arXiv e-prints*, art. astro-ph/9606001, June 1996. doi : 10.48550/arXiv.astro-ph/9606001.
- Navarro, J. F., Frenk, C. S., & White, S. D. M. The Structure of Cold Dark Matter Halos. *ApJ*, 462 :563, May 1996. doi : 10.1086/177173.
- Newman, J. A. & Gruen, D. Photometric Redshifts for Next-Generation Surveys. *ARA&A*, 60 :363–414, Aug. 2022. doi : 10.1146/annurev-astro-032122-014611.
- Ntampaka, M., Trac, H., Sutherland, D. J., et al. A Machine Learning Approach for Dynamical Mass Measurements of Galaxy Clusters. *ApJ*, 803(2) :50, Apr. 2015. doi : 10.1088/0004-637X/803/2/50.
- Oguri, M. A cluster finding algorithm based on the multiband identification of red sequence galaxies. *MNRAS*, 444(1) :147–161, Oct. 2014. doi : 10.1093/mnras/stu1446.
- Oguri, M., Lin, Y.-T., Lin, S.-C., et al. An optically-selected cluster catalog at redshift $0.1 < z < 1.1$ from the Hyper Suprime-Cam Subaru Strategic Program S16A data. *PASJ*, 70 :S20, Jan. 2018. doi : 10.1093/pasj/psx042.
- Oguri, M., Miyazaki, S., Li, X., et al. Hundreds of weak lensing shear-selected clusters from the hyper suprime-cam subaru strategic program s19a data. *Publications of the Astronomical Society of Japan*, 73(4) :817–829, May 2021. ISSN 2053-051X. doi : 10.1093/pasj/psab047. URL <http://dx.doi.org/10.1093/pasj/psab047>.
- Olive, K. A., Steigman, G., & Walker, T. P. Primordial nucleosynthesis : theory and observations. *Phys. Rep.*, 333 :389–407, Aug. 2000. doi : 10.1016/S0370-1573(00)00031-4.
- Pacaud, F., Pierre, M., Melin, J. B., et al. The XXL Survey. XXV. Cosmological analysis of the C1 cluster number counts. *A&A*, 620 :A10, Nov. 2018. doi : 10.1051/0004-6361/201834022.
- Peacock, J. A. *Cosmological Physics*. 1999.
- Peebles, P. J. E. *Principles of Physical Cosmology*. 1993. doi : 10.1515/9780691206721.
- Pereira, T. S., Pitrou, C., & Uzan, J.-P. Weak-lensing B-modes as a probe of the isotropy of the universe. *A&A*, 585 :L3, Jan. 2016. doi : 10.1051/0004-6361/201527258.
- Perlmutter, S., Aldering, G., Goldhaber, G., et al. Measurements of Ω and Λ from 42 High-Redshift Supernovae. *ApJ*, 517(2) :565–586, June 1999. doi : 10.1086/307221.

- Perrenod, S. C. The evolution of cluster X-ray sources. IV - The luminosity function. *ApJ*, 236 :373–387, Mar. 1980. doi : 10.1086/157755.
- Pierre, M., Pacaud, F., Adami, C., et al. The XXL Survey. I. Scientific motivations - XMM-Newton observing plan - Follow-up observations and simulation programme. *A&A*, 592 :A1, June 2016. doi : 10.1051/0004-6361/201526766.
- Pilbratt, G. L., Riedinger, J. R., Passvogel, T., et al. Herschel Space Observatory. An ESA facility for far-infrared and submillimetre astronomy. *A&A*, 518 :L1, July 2010. doi : 10.1051/0004-6361/201014759.
- Pires, S., Starck, J. L., Amara, A., et al. FASt STatistics for weak Lensing (FASTLens) : fast method for weak lensing statistics and map making. *MNRAS*, 395(3) :1265–1279, May 2009. doi : 10.1111/j.1365-2966.2009.14625.x.
- Pires, S., Leonard, A., & Starck, J.-L. Cosmological constraints from the capture of non-Gaussianity in weak lensing data. *MNRAS*, 423(1) :983–992, June 2012. doi : 10.1111/j.1365-2966.2012.20940.x.
- Pires, S., Vandenbussche, V., Kansal, V., et al. Euclid : Reconstruction of weak-lensing mass maps for non-gaussianity studies. *A&A*, 638 :A141, jun 2020. doi : 10.1051/0004-6361/201936865. URL <https://doi.org/10.1051/0004-6361/201936865>.
- Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. Planck 2013 results. XXIX. The Planck catalogue of Sunyaev-Zeldovich sources. *A&A*, 571 :A29, Nov. 2014. doi : 10.1051/0004-6361/201321523.
- Pointecouteau, E., Giard, M., & Barret, D. Determination of the hot intracluster gas temperature from submillimeter measurements. *A&A*, 336 :44–48, Aug. 1998. doi : 10.48550/arXiv.astro-ph/9712271.
- Pratt, G. W., Arnaud, M., Biviano, A., et al. The galaxy cluster mass scale and its impact on cosmological constraints from the cluster population. *Space Science Reviews*, 215(2), Feb 2019. ISSN 1572-9672. doi : 10.1007/s11214-019-0591-0. URL <http://dx.doi.org/10.1007/s11214-019-0591-0>.
- Press, W. H. & Schechter, P. Formation of Galaxies and Clusters of Galaxies by Self-Similar Gravitational Condensation. *ApJ*, 187 :425–438, Feb. 1974. doi : 10.1086/152650.
- Pyne, T. & Birkinshaw, M. Null Geodesics in Perturbed Spacetimes. *ApJ*, 415 :459, Oct. 1993. doi : 10.1086/173178.
- Refregier, A. Shapelets – i. a method for image analysis. *Monthly Notices of the Royal Astronomical Society*, 338(1) :35–47, jan 2003. doi : 10.1046/j.1365-8711.2003.05901.x. URL <https://doi.org/10.1046/j.1365-8711.2003.05901.x>.
- Riess, A. G., Filippenko, A. V., Challis, P., et al. Observational Evidence from Supernovae for an Accelerating Universe and a Cosmological Constant. *AJ*, 116(3) :1009–1038, Sept. 1998. doi : 10.1086/300499.
- Robertson, H. P. Kinematics and World-Structure. *ApJ*, 82 :284, Nov. 1935. doi : 10.1086/143681.
- Rossetti, M., Gastaldello, F., Eckert, D., et al. The cool-core state of Planck SZ-selected clusters versus X-ray-selected samples : evidence for cool-core bias. *MNRAS*, 468(2) :1917–1930, June 2017. doi : 10.1093/mnras/stx493.
- Rubin, V. C., Ford, W. K., J., & Thonnard, N. Rotational properties of 21 SC galaxies with a large range of luminosities and radii, from NGC 4605 (R=4kpc) to UGC 2885 (R=122kpc). *ApJ*, 238 :471–487, June 1980. doi : 10.1086/158003.

- Rykoff, E. S., Rozo, E., Busha, M. T., et al. redMaPPer. i. ALGORITHM AND SDSS DR8 CATALOG. *The Astrophysical Journal*, 785(2) :104, apr 2014. doi : 10.1088/0004-637x/785/2/104. URL <https://doi.org/10.1088%2F0004-637x%2F785%2F2%2F104>.
- Sartoris, B., Biviano, A., Fedeli, C., et al. Next generation cosmology : constraints from the Euclid galaxy cluster survey. *MNRAS*, 459(2) :1764–1780, June 2016. doi : 10.1093/mnras/stw630.
- Schirmer. *Weak gravitational lensing : Detection of mass concentrations in wide field imaging data*. PhD thesis, Rheinische Friedrich-Wilhelms-Universität Bonn, 2004. URL <https://hdl.handle.net/20.500.11811/2017>.
- Schirmer, M., Erben, T., Hettterscheidt, M., & Schneider, P. GaBoDS : the Garching-Bonn Deep Survey. IX. A sample of 158 shear-selected mass concentration candidates. *A&A*, 462(3) :875–887, Feb. 2007. doi : 10.1051/0004-6361:20065955.
- Schneider, P. Detection of (dark) matter concentrations via weak gravitational lensing. *Monthly Notices of the Royal Astronomical Society*, 283(3) :837–853, Dec 1996. ISSN 1365-2966. doi : 10.1093/mnras/283.3.837. URL <http://dx.doi.org/10.1093/mnras/283.3.837>.
- Schneider, P. & Bartelmann, M. Aperture multipole moments from weak gravitational lensing. *MNRAS*, 286(3) :696–708, Apr. 1997. doi : 10.1093/mnras/286.3.696.
- Schneider, P., Ehlers, J., & Falco, E. E. *Gravitational Lenses*. 1992. doi : 10.1007/978-3-662-03758-4.
- Schneider, P., Van Waerbeke, L., Jain, B., & Kruse, G. A new measure for cosmic shear. *MNRAS*, 296(4) : 873–892, June 1998. doi : 10.1046/j.1365-8711.1998.01422.x.
- Schneider, P., van Waerbeke, L., Kilbinger, M., & Mellier, Y. Analysis of two-point statistics of cosmic shear. I. Estimators and covariances. *A&A*, 396 :1–19, Dec. 2002. doi : 10.1051/0004-6361:20021341.
- Schrabback, T., Hilbert, S., Hoekstra, H., et al. CFHTLenS : weak lensing constraints on the ellipticity of galaxy-scale matter haloes and the galaxy-halo misalignment. *MNRAS*, 454(2) :1432–1452, Dec. 2015. doi : 10.1093/mnras/stv2053.
- Schrabback, T., Schirmer, M., van der Burg, R. F. J., et al. Precise weak lensing constraints from deep high-resolution K_s images : VLT/HAWK-I analysis of the super-massive galaxy cluster RCS2J 232727.7-020437 at $z = 0.70$. *A&A*, 610 :A85, Feb. 2018. doi : 10.1051/0004-6361/201731730.
- Scoville, N., Aussel, H., Brusa, M., et al. The Cosmic Evolution Survey (COSMOS) : Overview. *ApJS*, 172 (1) :1–8, Sept. 2007. doi : 10.1086/516585.
- Seitz, C. & Schneider, P. Steps towards nonlinear cluster inversion through gravitational distortions II. Generalization of the Kaiser and Squires method. *A&A*, 297 :287, May 1995. doi : 10.48550/arXiv.astro-ph/9408050.
- Seitz, C. & Schneider, P. Steps towards nonlinear cluster inversion through gravitational distortions. III. Including a redshift distribution of the sources. *A&A*, 318 :687–699, Feb. 1997.
- Shan, H., Kneib, J.-P., Tao, C., et al. Weak Lensing Measurement of Galaxy Clusters in the CFHTLS-Wide Survey. *ApJ*, 748(1) :56, Mar. 2012. doi : 10.1088/0004-637X/748/1/56.
- Spergel, D., Gehrels, N., Baltay, C., et al. Wide-field infrared survey telescope-astronomy focused telescope assets wfirst-afta 2015 report, 2015. URL <https://arxiv.org/abs/1503.03757>.
- Spitzer, Lyman, J. Equipartition and the Formation of Compact Nuclei in Spherical Stellar Systems. *ApJ*, 158 :L139, Dec. 1969. doi : 10.1086/180451.

- Springel, V., White, S. D. M., Tormen, G., & Kauffmann, G. Populating a cluster of galaxies - I. Results at $z=0$. *MNRAS*, 328(3) :726–750, Dec. 2001. doi : 10.1046/j.1365-8711.2001.04912.x.
- Starck, J., Murtagh, F., & Bijaoui, A. *Image Processing and Data Analysis : The Multiscale Approach*. Cambridge University Press, 1998. ISBN 9780521590846. URL <https://books.google.fr/books?id=lpfgtAEACAAJ>.
- Starck, J. L., Pires, S., & Réfrégier, A. Weak lensing mass reconstruction using wavelets. *A&A*, 451(3) : 1139–1150, June 2006. doi : 10.1051/0004-6361:20052997.
- Starck, J.-L., Pires, S., & Refregier, A. MRLENS : Multi-Resolution methods for gravitational LENSing. Astrophysics Source Code Library, record ascl :1102.005, Feb. 2011a.
- Starck, J.-L., Pires, S., & Refregier, A. MRLENS : Multi-Resolution methods for gravitational LENSing. Astrophysics Source Code Library, record ascl :1102.005, Feb. 2011b.
- Steigman, G. Primordial Nucleosynthesis in the Precision Cosmology Era. *Annual Review of Nuclear and Particle Science*, 57(1) :463–491, Nov. 2007. doi : 10.1146/annurev.nucl.56.080805.140437.
- Sugiyama, S., Takada, M., Miyatake, H., et al. HSC Year 1 cosmology results with the minimal bias method : HSC \times BOSS galaxy-galaxy weak lensing and BOSS galaxy clustering. *Phys. Rev. D*, 105(12) : 123537, June 2022. doi : 10.1103/PhysRevD.105.123537.
- Sugiyama, S., Miyatake, H., More, S., et al. Hyper Suprime-Cam Year 3 Results : Cosmology from Galaxy Clustering and Weak Lensing with HSC and SDSS using the Minimal Bias Model. *arXiv e-prints*, art. arXiv :2304.00705, Apr. 2023. doi : 10.48550/arXiv.2304.00705.
- Sunyaev, R. A. & Zeldovich, Y. B. The Observations of Relic Radiation as a Test of the Nature of X-Ray Radiation from the Clusters of Galaxies. *Comments on Astrophysics and Space Physics*, 4 :173, Nov. 1972.
- Tanaka, M. Photometric Redshift with Bayesian Priors on Physical Properties of Galaxies. *ApJ*, 801(1) : 20, Mar. 2015. doi : 10.1088/0004-637X/801/1/20.
- Tanaka, M., Coupon, J., Hsieh, B.-C., et al. Photometric redshifts for Hyper Suprime-Cam Subaru Strategic Program Data Release 1. *PASJ*, 70 :S9, Jan. 2018. doi : 10.1093/pasj/psx077.
- Tang, J. Y. & Fan, Z. H. Effects of the Complex Mass Distribution of Dark Matter Halos on Weak-Lensing Cluster Surveys. *ApJ*, 635(1) :60–72, Dec. 2005. doi : 10.1086/497285.
- The Planck Collaboration. The Scientific Programme of Planck. *arXiv e-prints*, art. astro-ph/0604069, Apr. 2006. doi : 10.48550/arXiv.astro-ph/0604069.
- Umetsu, K. Cluster-galaxy weak lensing. *A&A Rev.*, 28(1) :7, Dec. 2020. doi : 10.1007/s00159-020-00129-w.
- Umetsu, K. & Broadhurst, T. Combining Lens Distortion and Depletion to Map the Mass Distribution of A1689. *ApJ*, 684(1) :177–203, Sept. 2008. doi : 10.1086/589683.
- Utsumi, Y., Miyazaki, S., Geller, M. J., et al. Reducing systematic error in weak lensing cluster surveys. *The Astrophysical Journal*, 786(2) :93, Apr 2014. ISSN 1538-4357. doi : 10.1088/0004-637x/786/2/93. URL <http://dx.doi.org/10.1088/0004-637X/786/2/93>.
- Van Waerbeke, L. Scale dependence of the bias investigated by weak lensing. *A&A*, 334 :1–10, June 1998.

- Van Waerbeke, L. Noise properties of gravitational lens mass reconstruction. *MNRAS*, 313(3) :524–532, Apr. 2000. doi : 10.1046/j.1365-8711.2000.03259.x.
- Vikhlinin, A., Kravtsov, A. V., Burenin, R. A., et al. Chandra Cluster Cosmology Project III : Cosmological Parameter Constraints. *ApJ*, 692(2) :1060–1074, Feb. 2009. doi : 10.1088/0004-637X/692/2/1060.
- Voit, G. M. Tracing cosmic evolution with clusters of galaxies. *Reviews of Modern Physics*, 77(1) :207–258, Apr 2005. ISSN 1539-0756. doi : 10.1103/revmodphys.77.207. URL <http://dx.doi.org/10.1103/RevModPhys.77.207>.
- Walker, A. G. On the formal comparison of Milne's kinematical system with the systems of general relativity. *MNRAS*, 95 :263–269, Jan. 1935. doi : 10.1093/mnras/95.3.263.
- Walsh, D., Carswell, R. F., & Weymann, R. J. 0957+561 A, B : twin quasistellar objects or gravitational lens? *Nature*, 279 :381–384, May 1979. doi : 10.1038/279381a0.
- Weisskopf, M. C., Tananbaum, H. D., Van Speybroeck, L. P., & O'Dell, S. L. Chandra X-ray Observatory (CXO) : overview. In Truemper, J. E. & Aschenbach, B., editors, *X-Ray Optics, Instruments, and Missions III*, volume 4012 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, pages 2–16, July 2000. doi : 10.1117/12.391545.
- White, M., van Waerbeke, L., & Mackey, J. Completeness in Weak-Lensing Searches for Clusters. *ApJ*, 575(2) :640–649, Aug. 2002. doi : 10.1086/341351.
- White, S. D. M. & Rees, M. J. Core condensation in heavy halos : a two-stage theory for galaxy formation and clustering. *MNRAS*, 183 :341–358, May 1978. doi : 10.1093/mnras/183.3.341.
- Willis, J. P., Oguri, M., Ramos-Ceja, M. E., et al. Understanding X-ray and optical selection of galaxy clusters : a comparison of the XXL and CAMIRA cluster catalogues obtained in the common XXL-HSC SSP area. *MNRAS*, 503(4) :5624–5637, June 2021. doi : 10.1093/mnras/stab873.
- Wittman, D., Dell'Antonio, I. P., Hughes, J. P., et al. First results on shear-selected clusters from the deep lens survey : Optical imaging, spectroscopy, and x-ray follow-up. *The Astrophysical Journal*, 643 (1) :128–143, May 2006. ISSN 1538-4357. doi : 10.1086/502621. URL <http://dx.doi.org/10.1086/502621>.
- Wright, C. O. & Brainerd, T. G. Gravitational Lensing by NFW Halos. *ApJ*, 534(1) :34–40, May 2000. doi : 10.1086/308744.
- Zeldovich, Y. B. & Sunyaev, R. A. The Interaction of Matter and Radiation in a Hot-Model Universe. *Ap&SS*, 4(3) :301–316, July 1969. doi : 10.1007/BF00661821.
- Zwicky, F. Die Rotverschiebung von extragalaktischen Nebeln. *Helvetica Physica Acta*, 6 :110–127, Jan. 1933.
- Zwicky, F. On the Masses of Nebulae and of Clusters of Nebulae. *ApJ*, 86 :217, Oct. 1937. doi : 10.1086/143864.
- Zwicky, F., Herzog, E., Wild, P., Karpowicz, M., & Kowal, C. T. *Catalogue of galaxies and of clusters of galaxies, Vol. I*. 1961.