



**HAL**  
open science

# Clustering redshift : une nouvelle fenêtre sur l'univers

Vivien Scottez

► **To cite this version:**

Vivien Scottez. Clustering redshift : une nouvelle fenêtre sur l'univers. Astrophysique galactique [astro-ph.GA]. Université Pierre et Marie Curie - Paris VI, 2015. Français. NNT : 2015PA066210 . tel-01227831

**HAL Id: tel-01227831**

**<https://theses.hal.science/tel-01227831>**

Submitted on 12 Nov 2015

**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.



Université Pierre et Marie Curie

ED 127 - Astronomie & Astrophysique

*Institut d'Astrophysique de Paris*

Thèse de doctorat  
Spécialité : Cosmologie

**Clustering redshift :  
une nouvelle fenêtre sur l'Univers**

Par Vivien L. Scottez

*En vue de l'obtention du grade de :  
Docteur ès sciences*

Présentée et soutenue publiquement le 21 septembre 2015  
devant un jury composé de :

Présidente :	Francoise COMBES,	Professeur au Collège de France	– LERMA
Rapporteurs :	Alain BLANCHARD, Mathieu LANGER,	Professeur à l'université Paul Sabatier Maître de conférences	– LATT – IAS
Examineurs :	Laurence TRESSE, Olivier ILBERT,	Astronome Astronome adjoint	– LAM – LAM
Directeurs de thèse :	Yannick MELLIER, Brice MENARD,	Astronome Assistant Professor	– IAP – JHU



---

## Résumé

Les principaux objectifs de cette thèse sont de valider, consolider et développer une nouvelle méthode permettant de mesurer la distribution en redshift d'un échantillon de galaxies. Là où les méthodes actuelles - redshifts *spectroscopiques* et *photométriques* - sont toutes liées à l'étude de la distribution d'énergie spectrale des sources extragalactiques, l'approche ici présentée repose sur les propriétés d'agrégation des galaxies entre elles. En effet l'agrégation (*clustering* en anglais) des galaxies due à la gravité leur confère une distribution spatiale - et angulaire - particulière.

La méthode des *clustering redshifts* utilise cette propriété particulière d'agrégation entre une population de galaxies dont le redshift est *inconnu* et un échantillon d'objets de *référence* afin de déprojeter l'information et de reconstruire la distribution en redshift de la population *inconnue*. On peut s'attendre à ce que les systématiques de cette approche soient différentes de celles des méthodes existantes qui elles s'intéressent à la distribution spectrale d'énergie (SED) des galaxies.

Ce type d'approche répond à un réel besoin de la part de la communauté scientifique dans le cadre des grands projets d'observations tels que la mission *Euclid* de l'Agence Spatiale Européenne (ESA). Après avoir situé le contexte scientifique général et avoir mis en évidence le rôle crucial de la mesure des distances en astronomie, je présente les outils statistiques généralement utilisés dans le cadre de l'étude de la répartition de la matière dans l'Univers ainsi que leur modification afin de pouvoir mesurer des distributions en redshift.

Après avoir validé cette approche sur un type d'objets extragalactiques particuliers, j'ai ensuite étendu son application à l'ensemble des galaxies existantes. J'ai ensuite exploré la précision et les systématiques affectant ces mesures dans un cas idéal. Puis, je m'en suis éloigné de façon à me trouver en situation réelle. J'ai également poussé plus loin cette analyse et montré que les objets de *référence* utilisés lors de la mesure n'ont pas besoin de constituer un échantillon dont la magnitude limite est représentative de celle de la population de redshift *inconnu*. Cette propriété constitue un avantage considérable pour l'utilisation de cette approche dans le cadre des futurs grands projets observationnels comme la mission spatiale *Euclid*.

Pour finir, je résume mes principaux résultats et présente certains de mes futurs projets.

---



## Abstract

The main goals of this thesis are to validate, consolidate and develop a new method to measure the redshift distribution of a sample of galaxies. Where current methods - *spectroscopic* and *photometric* redshifts - rely on the study of the spectral energy distribution of extragalactic sources, the approach presented here is based on the clustering properties of galaxies. Indeed clustering of galaxies caused by gravity gives them a particular spatial - and angular - distribution.

In this *clustering* redshift approach, we use this particular property between a galaxies sample of *unknown* redshifts and a galaxies sample of *reference* to reconstruct the redshift distribution of the *unknown* population. Thus, possible systematics in this approach should be independent of those existing in other methods.

This new method responds to a real need from the scientific community in the context of large dark imaging experiments such as the *Euclid* mission of the European Space Agency (ESA). After introducing the general scientific context and having highlighted the crucial role of distance measurements in astronomy, I present the statistical tools generally used to study the large scale structure of the Universe as well as their modification to infer redshift distributions.

After validating this approach on a particular type of extragalactic objects, I generalized its application to all types of galaxies. Then, I explored the precision and some systematic effects by conducting an ideal case study. Thus, I performed a real case study. I also pushed further this analysis and found that the *reference* sample used in the measurement does not need to have the same limiting magnitude than the population of *unknown* redshift. This property is a great advantage for the use of this approach in the context of large imaging dark energy experiments like the *Euclid* space mission.

Finally, I summarize my main results and present some of my future projects.

---



---

## Remerciements

**J**e tiens à exprimer ma reconnaissance envers tous ceux qui ont rendu possible le travail présenté dans cet ouvrage.

Elle se dirige tout d'abord vers Brice Ménard et Yannick Mellier, qui m'ont suivi depuis les balbutiements de ce projet. Brice, tes précieuses remarques sont toujours d'une effroyable pertinence et m'ont permis de me dépasser et cela mérite toute ma gratitude. J'ai appris beaucoup sur moi-même en travaillant avec toi. Yannick, je te remercie de croire autant à ce projet et d'avoir toujours été à l'écoute de mes interrogations malgré tes très lourdes responsabilités. Plus généralement, merci à vous deux de m'avoir permis de travailler sur un projet aussi novateur et prometteur.

Un grand merci à Martin Kilbinger pour avoir toujours pris le temps de m'aider et de dispenser ses précieux conseils avec bienveillance et sympathie. Je tiens également à remercier Françoise Combes, Laurence Tresse et Olivier Ilbert qui m'ont fait l'honneur d'accepter de faire partie de mon jury et plus particulièrement Alain Blanchard et Mathieu Langer pour en être les rapporteurs. Vos remarques m'ont permis d'améliorer la qualité de ce manuscrit.

Je remercie l'ensemble des membres de l'Institut d'Astrophysique de Paris pour m'avoir accueilli pendant ces trois années et plus particulièrement : Henry Joy McCracken, Valérie de Lapparent, Raphaël Gavazzi, Karim Benabed, Florence Durret, Stéphane Colombi, Matt Lehnert, Gary Mamon, Brigitte Rocca-Volmerange pour leur soutien et leur disponibilité sur les différentes questions auxquelles j'ai dû faire face durant ma thèse.

Un énorme merci à mes parents qui m'ont toujours soutenu et qui me donnent l'impression d'être le meilleur physicien du monde. Merci également à mon frère Julien qui croit plus en moi que je ne le fais moi-même.

Merci également à tous les doctorants de l'IAP grâce à qui venir au labo était un véritable plaisir. En particulier merci à Nicolas pour ne jamais refuser une bière et surtout pour ne pas parler boulot à la pause déjeuner ; à Alba, Julia et Mélanie pour apporter un peu de douceur - et de n'importe nawak - dans un monde de brutes ; ainsi qu'à Pierre pour nos discussions de physique mais aussi pour m'avoir fait découvrir la méthode de Feynman qu'il nous faudra bien essayer de mettre en application un jour ou l'autre ne serait-ce que pour vérifier sa validité. Nous avons la chance d'avoir une très bonne ambiance au sein de l'IAP, cela est extrêmement précieux.

Je tiens très particulièrement à remercier Alexis pour son amitié et pour nos innombrables soirées à parler de physique. Faire des calculs sur un carton de pizza devant le film *Le Cinquième Élément* pour vérifier ce que les personnages viennent d'affirmer n'a pas de prix.

Un très grand merci à tous mes "amis de Lille", pour ne pas dire Fourmies! La distance et les nombreux mois passés sans se voir laissent inchangé le plaisir de passer du temps avec vous. Je remercie en particulier Nicolas pour être toujours présent. En dix-huit ans d'amitié les raisons sont trop nombreuses et nécessiteraient une thèse pour être énumérées. Bon courage avec ton manuscrit de thèse, j'ai hâte d'assister à ta soutenance.

Enfin je souhaite terminer ces remerciements en mentionnant les enseignants qui m'ont transmis plus que des connaissances : Marie-Hélène Avisse pour les sciences physiques et Alain Bergmann pour les mathématiques. Avoir des enseignants passionnés et passionnants est une chance incroyable.

# Table des matières

<b>Résumé</b>	<b>i</b>
<b>Abstract</b>	<b>iii</b>
<b>Remerciements</b>	<b>v</b>
<b>Table des matières</b>	<b>vii</b>
<b>Introduction</b>	<b>1</b>
<b>Distances measurements in astronomy : abstract</b>	<b>7</b>
<b>1 Mesures des distances en astronomie</b>	<b>9</b>
1 Des étoiles aux galaxies proches ( $< 500$ M a.l.) . . . . .	9
2 Pour les galaxies lointaines ( $> 500$ M a.l.) . . . . .	14
3 Un œil vers l'avenir . . . . .	32
4 Synthèse . . . . .	45
<b>Introducing the method : abstract</b>	<b>47</b>
<b>2 Présentation de la méthode</b>	<b>49</b>
1 Concepts . . . . .	50
2 Formalisme . . . . .	55
3 Calcul de $\bar{\omega}(z)$ . . . . .	59
4 Synthèse . . . . .	65
<b>Validating the method : abstract</b>	<b>67</b>
<b>3 Validation de la méthode</b>	<b>69</b>
1 Evolution du biais et estimation du rapport signal sur bruit . . . . .	69
2 Mesure de $\bar{\omega}_{ur}(z)$ sur les données SDSS . . . . .	73

3	Validation du redshift clustering sur les données SDSS . . . . .	78
4	Généralisation à tous types de populations de galaxies . . . . .	81
5	Synthèse . . . . .	83
	<b>Ideal case study : abstract</b>	<b>85</b>
<b>4</b>	<b>Clustering redshift : cas idéal</b>	<b>87</b>
1	Sélection des données et mesure du biais de référence . . . . .	87
2	Comparaison <i>cluster/spectro-z</i> et étude du bruit . . . . .	92
3	Synthèse . . . . .	97
	<b>Real case study : abstract</b>	<b>99</b>
<b>5</b>	<b>Clustering redshift : cas réel</b>	<b>101</b>
1	Les données <i>CFHTLS &amp; VIPERS</i> . . . . .	102
2	Présélection tomographique . . . . .	110
3	Présélection dans l'espace des couleurs . . . . .	117
4	Synthèse . . . . .	127
	<b>Conclusions et perspectives</b>	<b>129</b>
	<b>A Fonctions de corrélation</b>	<b>133</b>
	<b>B Publication</b>	<b>137</b>
	<b>Bibliographie</b>	<b>151</b>

# Introduction

---

L'étude de l'Univers, en tant qu'objet physique, est l'une des disciplines les plus fondamentales et passionnantes qui soit. Comprendre le monde, s'interroger sur la nature de la réalité et essayer de la décrire le plus fidèlement possible est le quotidien du physicien. Que ce soit l'étude de l'infiniment petit ou de l'infiniment grand la physique a ceci de passionnant qu'elle transporte, qu'elle fait rêver, qu'elle révèle.

L'astrophysique a ceci de particulier qu'elle ne peut pas s'appuyer sur la reproduction, en laboratoire, d'expériences rigoureusement identiques. Pour des raisons évidentes cette approche, qui a fait la force de la démarche scientifique dans de nombreux domaines, n'est pas envisageable. Il faut donc apprendre à se passer d'une connaissance quasi parfaite des conditions initiales d'un système et faire tout son possible pour extraire le maximum d'informations complémentaires à partir d'observations différentes. Pour cela, les outils statistiques dont nous disposons aujourd'hui sont d'une aide précieuse et d'une efficacité redoutable.

Pendant longtemps, la question de l'origine de l'Univers ne fut abordée que par les religions, chacune développant sa propre cosmogonie. De nos jours, notre connaissance de la physique - notamment de la relativité générale - ainsi que les avancés remarquables de la technologie permettent d'aborder cette question de façon scientifique.

Portés par ces innovations technologiques, les modèles cosmologiques furent de plus en plus contraints pour finalement donner naissance au *modèle standard cosmologique* ou *modèle de concordance*. Bien que ce modèle soit en perpétuelle évolution, il a d'ores et déjà permis de mettre en évidence et d'expliquer des phénomènes jusqu'alors insoupçonnés.

Depuis le rayonnement fossile - *Cosmic Microwave background* (CMB) - présentant un Univers extrêmement homogène et isotrope ; jusqu'à aujourd'hui où l'Univers est dans une phase d'expansion accélérée ; en passant par l'époque de la réionisation avec la formation des premières étoiles, la *grande histoire cosmique* semble être assez bien décrite par ce modèle.

Nous entrons aujourd'hui dans une ère dite de *cosmologie de précision*. Il s'agit ici d'atteindre une précision de l'ordre du pourcent sur chacun des paramètres cosmologiques. Pour cela, la communauté scientifique va devoir réaliser, traiter et analyser des Peta-Bytes de données issues d'innombrables heures d'observations. De grands projets d'observations virent alors le jour tels que la mission *Euclid* de l'Agence Spatiale Européenne (ESA) dont l'objectif principal est de percer les mystères de l'actuelle accélération de l'expansion de l'Univers.

L'accélération de l'expansion de l'Univers est une des grandes découvertes de la physique de la fin du XXIème siècle. Très largement consolidée par de multiples observations astronomiques elle ne fait plus vraiment l'objet d'un débat et c'est donc vers la question fondamentale de la source de cette accélération que se portent les interrogations des physiciens et des astrophysiciens. À ce jour c'est une véritable énigme et l'objet d'une des grandes quêtes de la physique fondamentale du début de ce siècle. Sa nature étant totalement inconnue c'est sous le terme générique d'énergie sombre qu'elle est identifiée.

Parmi les hypothèses les plus immédiates, l'énergie sombre pourrait traduire une interaction nouvelle de la nature associée à un champ scalaire dont les propriétés sont encore totalement inconnues. Alternativement, il est aussi tout à fait possible que l'accélération de l'expansion soit l'expression d'une déviation de la gravitation aux prédictions de la relativité générale lorsque l'on aborde des phénomènes portant sur des très grandes échelles de distance dans l'Univers - de l'ordre du Gpc ou plus - et pour lesquelles il n'existe à ce jour aucune expérience qui la valide. Ces deux grandes directions de recherche portent en elles des bouleversements pour la physique et notre compréhension des lois fondamentales de la nature mais aussi des scénarios décrivant le passé et l'avenir de l'Univers et même de la place de notre univers dans les théories cosmologiques modernes.

Comprendre la véritable nature de cette énergie sombre fait donc partie des questions les plus fondamentales et passionnantes de la physique et de la cosmologie contemporaines. C'est la raison pour laquelle les physiciens et astrophysiciens déploient aujourd'hui tout une panoplie de "sondes cosmologiques" qui cherchent à détecter et caractériser des signatures de l'énergie sombre dans des observables astronomiques qui caractérisent l'histoire de l'expansion de l'Univers et l'histoire de la formation des structures cosmiques dans l'Univers. Les sondes cosmologiques qui sont aujourd'hui les plus prometteuses en matière d'observation astronomique, de mesures et d'interprétation physique dans un contexte cosmologique sont les supernovae du type Ia, les oscillations acoustiques des baryons, les distorsions gra-

vitationnelles faibles, les distorsions espace-redshift, les amas de galaxies et l'effet Sachs-Wolf intégré. Malheureusement, les différences des signatures portées par ces sondes cosmologiques et que l'on prédit pour les modèles d'énergie sombre les plus vraisemblables sont toutes infimes. Les écarts sont tellement infimes que seule des mesures de grande précision fondées sur une analyse statistique d'un très grand nombre d'objets astronomiques pourraient apporter une réponse décisive à la question de la nature de l'énergie sombre.

Par ailleurs certaines sondes sont plus sensibles aux propriétés géométriques de l'Univers et à son taux d'expansion alors que d'autres voient préférentiellement des manifestations de la gravitation et donc du taux de croissance des structures. Pour aborder le problème de l'origine de l'accélération de l'expansion de l'Univers il est donc indispensable de construire des expériences utilisant des sondes multiples, indépendantes et complémentaires. C'est notamment ce qui a conduit à la conception de la mission spatiale *Euclid*. Cette mission est dotée d'un télescope illuminant trois instruments d'imagerie et de spectroscopie visible et infrarouge permettant d'exploiter 5 sondes cosmologiques (toutes celles énumérées ci-dessus, sauf les supernovae) et de reconstruire l'histoire de l'expansion et celle de la croissance des structures au cours des 10 derniers milliards d'années.

Une des grandes spécificités de la mission *Euclid* est qu'elle est conçue prioritairement pour la mesure des effets de lentille gravitationnelle faibles avec une très haute précision sur un échantillon de 1.5 milliards de galaxies. Cette méthode est fondée sur des mesures très précises de morphométrie des galaxies pour reconstruire le champ cohérent de déformation gravitationnelle par la matière noire localisée dans les structures cosmiques (le cisaillement cosmique). Cependant, les mesures de forme des galaxies ne suffisent pas car l'amplitude des effets de lentilles gravitationnelles dépend aussi des distances relatives des sources et des lentilles à l'observateur. Par ailleurs le champ cohérent de cisaillement cosmique est contaminé par les alignements cohérents intrinsèques des galaxies résultant des processus de formation des galaxies et des interactions de marée entre les galaxies. Pour nettoyer cette contamination il est indispensable de connaître les distances des galaxies. La connaissance des distances de plusieurs centaines de millions de galaxies se trouve donc au coeur de l'enjeu scientifique de la mission *Euclid*.

Il se pose alors la question de savoir comment mesurer les distances d'environ un milliard de galaxies. En effet aucun instrument à ce jour et aucun spectrographe en projet dans la décennie à venir ne seront capables de mesurer les redshifts d'autant de galaxies et cela sur tous les types de galaxies. L'utilisation du redshift spectroscopique en tant que mesure principale est donc compromise pour les galaxies d'*Euclid*. Néanmoins, on peut toutefois noter que pour les mesures de lentilles gravitationnelles et pour les alignements intrinsèques, il n'est pas nécessaire de disposer de distances mesurées avec une grande précision. *Euclid* se satisfait d'une précision de  $0.05(1+z)$  mais a besoin de mesurer ces redshifts avec une bonne précision, donc sans erreur systématique. La solution évidente est donc de s'orienter vers les redshifts photométriques, à la condition que les erreurs systématiques soient parfai-

tement contrôlées et que le taux de redshift catastrophique soit inférieur à 10%. Pour parvenir à ces niveaux de précision et d'exactitude, *Euclid* a besoin de mesurer des redshifts photométriques à partir des 3 bandes photométriques infrarouges que fournira son télescope et d'au moins 4 bandes photométriques visibles. C'est un enjeu considérable puisqu'il faut donc observer 15 000 deg<sup>2</sup> du ciel boréal et austral à des magnitudes typiques de  $AB = 24$ . La quête des données photométriques de la mission *Euclid* s'avère donc un projet en soi qui demande des centaines de nuits d'observation avec des grands télescopes terrestres répartis sur les deux hémisphères et un suivi spectroscopique complémentaire pour étalonner les redshifts photométriques. Le coût des observations sols est de plusieurs dizaines de millions d'euros et la mise à disposition de plusieurs télescopes pendant plusieurs centaines des nuits ainsi que les ressources humaines et matérielles nécessaires au traitement de ces données sont techniquement et opérationnellement des obstacles difficiles. Par conséquent, l'acquisition de ces observations n'est pas encore garantie ce qui place les redshifts photométriques au coeur des difficultés de la mission tant sur le plan de la précision et de l'exactitude des mesures de redshifts que sur celui de la production de ces données et du surcoût induit sur le projet.

Ces interrogations et les problèmes de mesure des distances des galaxies aux échelles cosmologiques motivent le coeur de mon projet de recherche : trouver un moyen de déterminer les distances d'un milliard de galaxies qui permettrait d'éviter un lourd et cher programme d'observation avec les télescopes au sol comme celui qu'il est nécessaire de conduire avec les redshifts photométriques. Si cet outil existe, il faut alors évaluer ses performances, ses limites, ses sensibilités aux conditions de sélection et d'observation, déterminer de potentielles sources d'erreurs systématiques, le comparer à la méthode des redshifts photométriques et discuter comment ces méthodes pourraient être utilisées conjointement puis développer ce nouvel outil de mesure pour en faire un outil opérationnel pour *Euclid*, le valider et le distribuer à la communauté.

C'est dans ce contexte que s'inscrit le travail présenté dans cet ouvrage. Durant ces trois ans, je me suis intéressé à une toute nouvelle façon de mesurer le redshift - la distance - des galaxies lointaines. Cette approche, très peu utilisée à l'heure actuelle, va permettre une petite révolution dans notre façon de faire de la cosmologie simplement parce qu'elle nous permet d'adopter un nouveau point de vue sur les données, elle ouvre une nouvelle fenêtre sur l'Univers.

Afin de bien comprendre la motivation derrière le travail accompli tout au long de ces trois années nous commencerons par présenter les différentes méthodes existantes pour mesurer la distance d'un objet lointain. Une attention toute particulière sera portée aux redshifts photométriques qui sont devenus indispensables aujourd'hui. Nous verrons leurs forces et leurs limitations, notamment dans le cadre de la mission *Euclid* dans lequel s'inscrit ce travail de thèse.

Puis, nous aborderons plus en détail le formalisme propre à la méthode dévelop-

pée dans cet ouvrage. Nous verrons également la validation de cette méthode que ce soit en simulation ou sur de vraies données issues d'observations.

Enfin, nous discuterons deux analyses la première correspondant à un cas idéal, la seconde à un cas réel via l'utilisation des données provenant du *Canada-France-Hawaii Telescope Legacy Survey (CFHTLS)* dont la stratégie d'observation ainsi que les filtres utilisés sont relativement proches de ceux du télescope spatial *Euclid*.



---

## Distances measurements in astronomy : abstract

This chapter presents the crucial role of distance measurements in astronomy. I review some of the methods used to infer distances from stars to distant galaxies. Also, I introduce the key concept of redshift as a distance indicator and present current methods allowing its measurements. In particular, I present a unified picture of *photometric redshifts*, as a function whose aim is to reproduce the mapping between photometric space - including : magnitude, position, ellipticity... - and redshift space based on spectral energy distribution (SED) inference from magnitudes or using training sets, Figure 1.

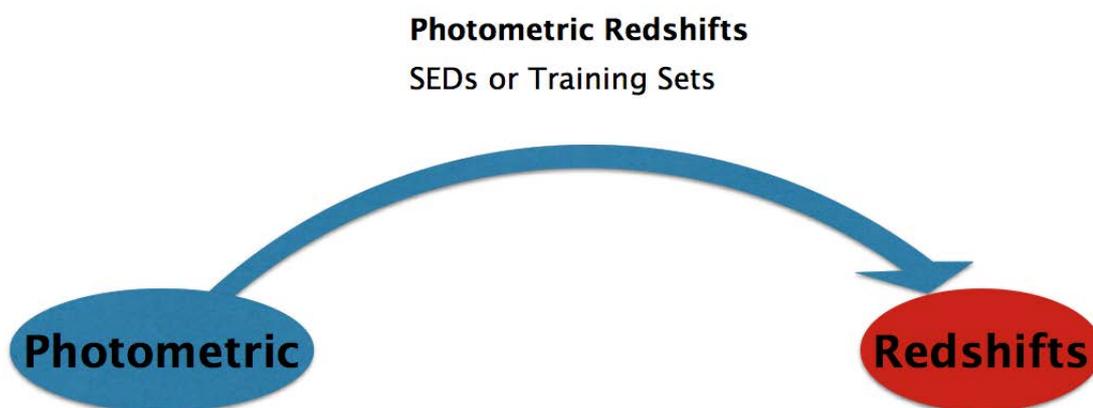


FIGURE 1 – This picture summarizes the main idea of *photometric redshifts*.

Then, I introduce the particular scientific context of upcoming large imaging dark energy experiments like the *Euclid* space telescope. Finally, looking at cosmological probes used to understand the cosmological accelerated expansion I present the key requirements on the redshift measurements introducing then the scientific challenge in which this work takes place.

---



# Chapitre 1

## Mesures des distances en astronomie

---

Lorsque l'on regarde le ciel, et les objets qui s'y trouvent, on peut naturellement s'interroger sur la distance qui les séparent de nous. Cette simple question est des plus fondamentales. En effet, si le soleil et la lune échangeaient leur place, notre vision du système solaire serait radicalement différente - et incompatible - avec celle que nous avons aujourd'hui. Nous évoluerions alors dans un paradigme totalement différent. La connaissance de la distance des astres est donc la pierre angulaire de notre compréhension de l'Univers.

### 1 Des étoiles aux galaxies proches ( $< 500$ M a.l.)

#### 1.1 Pour les étoiles ( $< 100$ M a.l.)

##### Mesure de parallaxe

La connaissance de l'éloignement, par rapport au soleil, des étoiles se trouvant dans notre galaxie passe par la mesure de leur parallaxe annuelle. Par définition, la parallaxe est l'angle sous lequel on voit le demi-grand axe de l'orbite terrestre depuis une étoile, c'est-à-dire l'angle  $\theta$  sur le schéma Figure 1.1 . Il s'agit en réalité d'un simple calcul de trigonométrie.

Cette méthode fut notamment utilisée en 1989 par le satellite *Hipparcos* (*High Precision PARallax Collection Satellite*) de l'Agence Spatiale Européenne (ESA). *Hipparcos* fit passer le nombre de mesures de parallaxe d'environ 8 000 à plus de 100 000 et cela avec une précision de l'ordre de la milliseconde d'arc. En 2013, le lancement réussi du satellite *Gaïa* de l'ESA qui localisera un milliard d'étoiles - soit dix mille fois plus que le satellite *Hipparcos*, Figure 1.2 - marque un pas de plus dans le développement et l'utilisation de la méthode des parallaxes. En cosmologie, les observations de *Gaïa* permettront de ré-étalonner les mesures de parallaxe. Ce ré-étalonnage se propagera vers les distances cosmologiques dans le régime où parallaxes et céphéides sont toutes deux applicables.

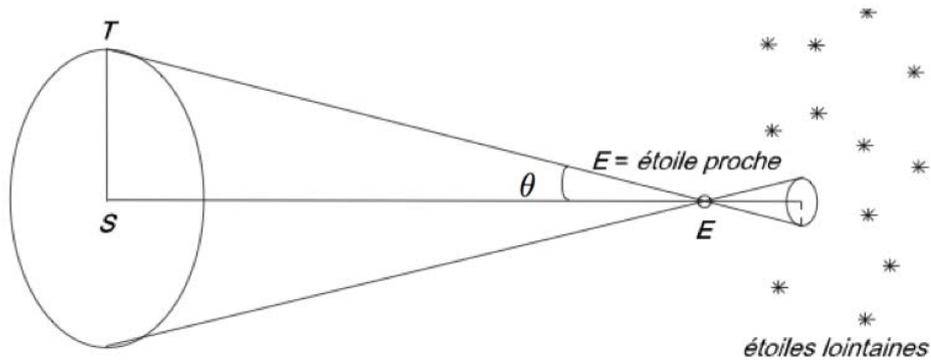


FIGURE 1.1 – Soit :  $S$  le soleil,  $T$  la terre et  $E$  une étoile proche. La parallaxe de  $E$  est notée  $\theta$ . Les angles  $\widehat{ETS}$  et  $\theta$  étant complémentaires la mesure de  $\widehat{ETS}$  permet de déterminer  $\theta$ . Finalement la distance  $SE$  est obtenue via :  $SE = \frac{ST}{\tan(\theta)}$ . La parallaxe d'une étoile est généralement inférieure à  $1''$ .

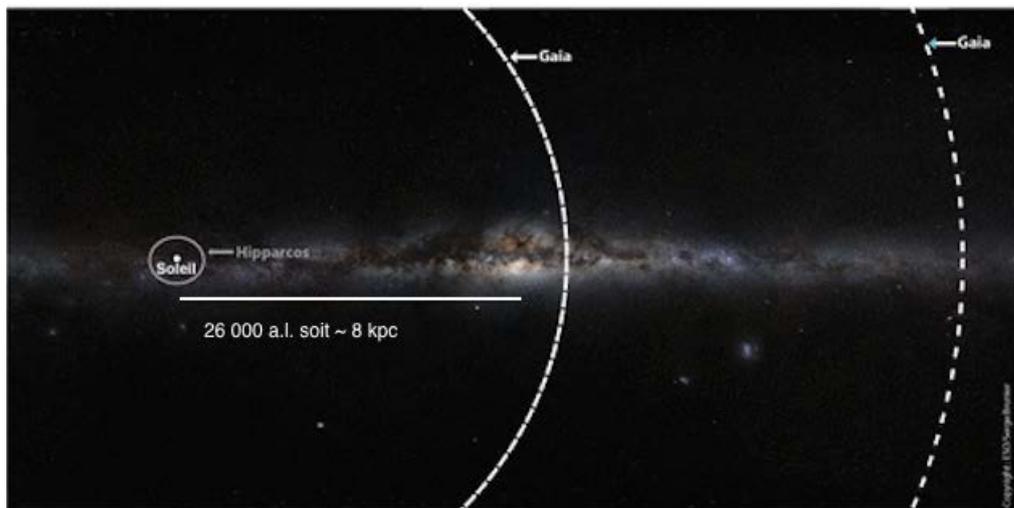


FIGURE 1.2 – *Hipparcos* : mesure de la distance des étoiles avec une erreur  $< 10\%$ .

*Gaïa* : mesure de la distance des étoiles avec une erreur  $< 10\%$ .

*Gaïa* : mesure de la vitesse des étoiles avec une précision  $< 1$  km/s. Figure extraite de <http://endirect.univ-fcomte.fr/download/en-direct/image/ed-240-janvier-2012/gaia-2.gif>

## Diagramme de Hertzsprung-Russell

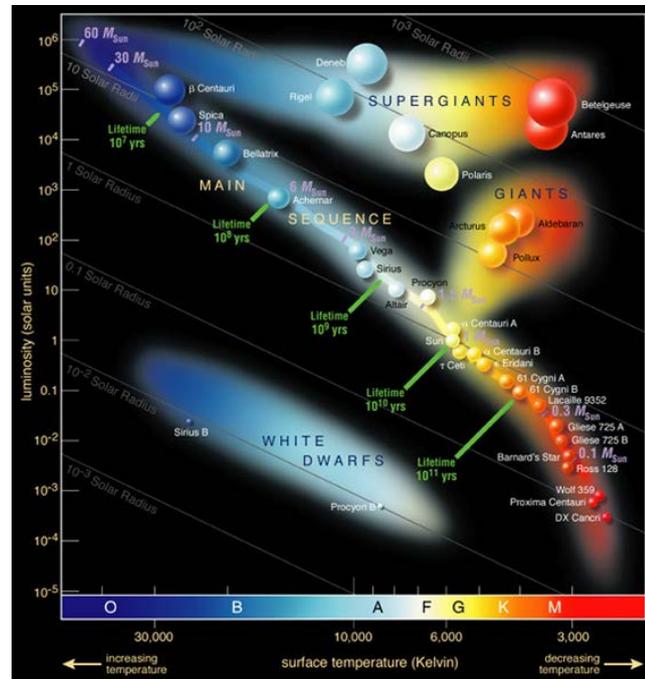


FIGURE 1.3 – Diagramme H-R montrant des étoiles connues de la Voie Lactée. On distingue clairement la séquence principale, la branche des géantes, celle des supergéantes ainsi que la zone des naines blanches. Copyright ©2008 Pearson Education, Inc., publishing as Pearson Addison-Wesley.

Le diagramme de Hertzsprung-Russell présente la luminosité d'un ensemble d'étoiles en fonction de leur température effective, Figure 1.3. Il s'agit d'une méthode de mesure indirecte présentée pour la première fois dans Russel (1914). La mesure de la température, d'une étoile de la séquence principale, permet via le diagramme H-R de déterminer sa magnitude absolue. Connaissant la magnitude apparente de l'objet on peut alors, via l'équation (1.1), obtenir la distance  $d$  en parsec.

$$d = 10 \times 10^{(m-M)/5}, \quad (1.1)$$

avec :

- $m$ , la magnitude apparente.
- $M$ , la magnitude absolue.
- $d$ , la distance en parsec.

Remarquons que cette méthode ne fonctionne pas pour les galaxies à cause des étoiles peu brillantes, de la poussière ainsi que de l'effet de l'âge et de la métallicité sur la luminosité.

## 1.2 Pour les galaxies proches ( $< 500$ M a.l.)

### Méthode des céphéides

Les céphéides, Figure 1.4, sont des étoiles variables, brillantes dont la luminosité varie périodiquement sur une durée de quelques jours (Leavitt, 1908). On dit que ces étoiles possèdent une relation période/luminosité (Mo et Van den Bosch, 2010).



FIGURE 1.4 – Exemple d’une céphéide. Il s’agit d’un cliché de *RS Puppis* pris par le télescope spatial Hubble (HST).

Cette relation est généralement utilisée pour déterminer la distance des galaxies proches. La mesure de la période de l’étoile permet de déterminer sa magnitude absolue. Une fois de plus l’utilisation de l’équation (1.1) permet de déterminer la distance de la galaxie hôte. Elles ont notamment joué un rôle crucial dans l’étalonnage des distances utilisées par Hubble lorsqu’il mit en évidence la fuite des galaxies.

### Lois de Tully-Fisher et de Faber-Jackson

**La relation de Tully-Fisher** relie la luminosité intrinsèque d’une galaxie spirale à l’amplitude de sa courbe de rotation.

Cette relation (Tully et Fisher, 1977) permet de calculer la magnitude absolue d’une galaxie spirale grâce à la vitesse de rotation des étoiles autour du centre. La vitesse de rotation de la galaxie est elle-même mesurable par effet Doppler. Une fois la magnitude absolue de la galaxie connue, on peut une fois encore utiliser l’équation (1.1) et en déduire sa distance, Figure 1.5.

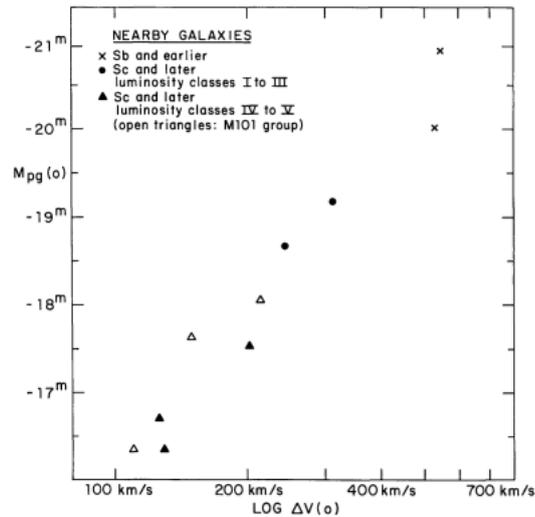


FIGURE 1.5 – Magnitude absolue en fonction de la vitesse de rotation pour des galaxies proches. Les croix correspondent à M31 et M81, les points sont M33 et NGC 2403, les triangles pleins sont des systèmes plus petits dans le groupe M81 et les triangles creux sont des systèmes du groupe M101. Extrait de [Tully et Fisher \(1977\)](#)

*La relation de Faber-Jackson* relie la magnitude absolue à la dispersion des vitesses  $\sigma$  des étoiles centrales des galaxies elliptiques, Figure 1.6. Elle permet la détermination des distances pour ce type de galaxies et s'exprime mathématiquement ainsi :  $L \propto \sigma^\gamma$ , où l'on observe que l'exposant  $\gamma$  est approximativement de 4, selon la catégorie des galaxies considérées.

La relation de [Faber et Jackson \(1976\)](#), équivaut pour les galaxies elliptiques à la loi de Tully-Fisher - qui s'applique uniquement dans le cas des galaxies spirales.

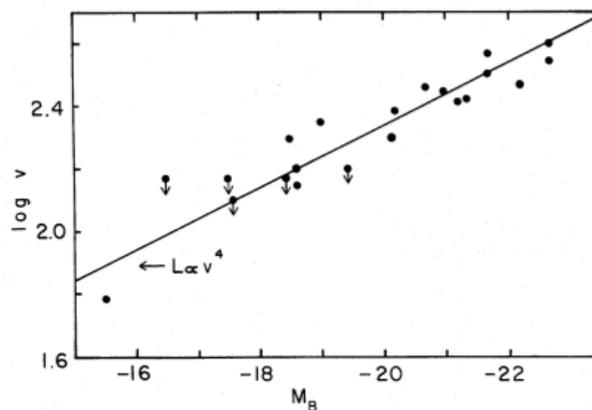


FIGURE 1.6 – Dispersion des vitesses dans la ligne de visée en fonction de la magnitude absolue. Le point avec la plus petite vitesse correspond à M32. Extrait de [Faber et Jackson \(1976\)](#)

## 2 Pour les galaxies lointaines ( $> 500$ M a.l.)

En cosmologie, la distance des galaxies lointaines est communément exprimée en termes de décalage spectral. Avant de présenter les différentes méthodes permettant de mesurer ce décalage, nous commencerons par rappeler brièvement les grands principes ainsi que le cadre mathématique de la cosmologie moderne puis nous nous attarderons sur la notion de redshift.

### 2.1 Principe cosmologique et métrique FLRW

La formulation du principe cosmologique se fait en supposant que l'Univers est homogène et isotrope à très grande échelle. En effet à l'échelle du système solaire, de la galaxie ou même du groupe local, l'Univers n'est ni homogène ni isotrope. Les deux principales raisons qui ont conduit à l'acceptation du principe cosmologique sont d'une part le principe copernicien qui stipule qu'il n'y a pas de point de vue privilégié dans l'Univers ainsi que l'isotropie autour de chaque point. Au cours des dernières décennies les observations astronomiques ont permis d'explorer la validité de ce principe et d'en constater la robustesse. Par exemple, l'isotropie de l'Univers a été confirmée par l'impressionnante isotropie du fond diffus cosmologique (CMB) observé par les satellites *COBE* (Smoot et collab., 1992), *WMAP* (Bennett et collab., 2003; Spergel et collab., 2003) et *PLANCK* (Planck Collaboration, 2014). Les relevés de galaxies à bas redshifts comme le *2-degree Field galaxy redshift survey* (Colless et collab., 2001) ou encore le *Sloan Digital Sky Survey* (York et collab., 2000) ont montré que l'homogénéité de l'Univers est également une hypothèse raisonnable.

La première description de la gravité fut donnée par Newton puis fut généralisée par Einstein avec la relativité générale en 1915, dans laquelle il met en évidence le lien intrinsèque entre temps et espace. Comme présenté dans Schneider (2015), l'espace peut être décrit par une sphère en expansion où la position des objets est décrite par leurs coordonnées comobiles (constantes s'il n'y a pas de mouvements propres) et leurs coordonnées physiques. Le lien entre ces deux coordonnées est appelé le facteur d'échelle  $a(t)$  :

$$d(t) = a(t)x , \quad (1.2)$$

avec  $a(t_0) = 1$  aujourd'hui par convention. La vitesse de récession est définie par :

$$v(d, t) = \frac{d}{dt}d(t) = \frac{da}{dt}x = \frac{\dot{a}}{a}d \equiv H(t)d , \quad (1.3)$$

où  $H(t)$  est le paramètre de Hubble. Dans l'Univers local, cette relation devient :

$$v(d, t_0) = H(t_0)d = H_0x , \quad (1.4)$$

également connue sous le nom de loi de Hubble et où  $H_0$  est la constante de Hubble.

En relativité générale l'espace est déformé par la distribution en masse de l'Univers. Cela implique que l'Univers n'est pas forcément plat comme dans la géométrie

Euclidienne mais qu'il peut être courbé par le potentiel gravitationnel des objets qu'il contient. La gravité peut alors être vue comme une propriété de l'espace-temps. Dans ce contexte le mouvement d'une particule suit simplement les distorsions de l'espace-temps. La distance entre deux évènements d'un tel espace est alors décrite par le tenseur métrique  $g_{\mu\nu}$  selon :

$$ds^2 = g_{\mu\nu} dx^\mu dx^\nu , \quad (1.5)$$

où  $x^\mu$  sont les coordonnées d'espace-temps.

Le principe cosmologique contraint la métrique de façon à ce que l'Univers soit homogène et isotrope. On peut montrer (Weinberg, 1972) que ces conditions sont vérifiées par la métrique de Fridmann-Lemaître-Robertson-Walker (FLRW) :

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

où  $a(t)$  est le facteur d'échelle,  $k$  est la courbure spatiale qui décrit la topologie de l'Univers. Si  $k = 0$  l'Univers est plat, si  $k = -1$  il est hyperbolique, si  $k = 1$  il est sphérique.

Toute particule se déplaçant à la vitesse de la lumière (comme un photon) suit une trajectoire appelée géodésique satisfaisant  $ds^2 = 0$ . Si sa direction est radiale, on a :

$$c^2 dt^2 = a^2(t) \frac{dr^2}{1 - kr^2} , \quad (1.7)$$

et :

$$\int_{t_e}^{t_o} \frac{c}{a(t)} dt = \int_{x_e}^{x_o} \frac{dr}{\sqrt{1 - kr^2}} . \quad (1.8)$$

Ici,  $t_e$  est le temps d'émission du photon et  $t_o$  le temps d'observation.  $x_e$  et  $x_o$  sont les positions respectives de l'objet émetteur et de l'observateur. Comme la distance comobile (terme de droite) est constante dans le temps, il vient :

$$\frac{c dt_o}{a(t_o)} = \frac{c dt_e}{a(t_e)} . \quad (1.9)$$

En substituant  $c dt_e$  et  $c dt_o$  par les longueurs d'onde  $\lambda_e$  et  $\lambda_o$ , il est possible d'exprimer le décalage spectral d'un objet distant en fonction du facteur d'échelle :

$$\frac{\lambda_o}{\lambda_e} = \frac{a(t_o)}{a(t_e)} . \quad (1.10)$$

Le décalage spectral résultant est alors défini par :

$$z = \frac{\lambda_o}{\lambda_e} - 1 , \quad (1.11)$$

avec :

- $\lambda_e$ , la longueur d'onde émise.
- $\lambda_o$ , la longueur d'onde observée.
- $z$ , le décalage spectral.

D'après l'équation (1.10) il vient :

$$1 + z = \frac{a_0}{a(t)} , \quad (1.12)$$

avec :

- $a_0$ , le facteur d'échelle aujourd'hui choisi pour être  $a_0 = 1$ .
- $a(t)$ , le facteur d'échelle au temps  $t$  correspondant au redshift  $z$ .

On voit immédiatement que le décalage spectral  $z$ , est relié à l'information sur la distance des objets. Étant une quantité observable cette expression fait du redshift un indicateur privilégié dans l'estimation de la distance des galaxies.

## 2.2 Notion de redshift

Comme nous venons de le voir, la distance des galaxies lointaines est reliée au décalage spectral de la lumière. En effet pour différentes raisons la longueur d'onde de la lumière émise par un objet lointain (comme une galaxie) peut-être décalée vers les grandes longueurs d'onde - vers le rouge - ou même dans certains cas vers les petites longueurs d'onde - vers le bleu. Nous présentons maintenant les différents effets pouvant engendrer un tel décalage.

### Effet Doppler

Ce décalage peut être dû au mouvement de la source, Figure 1.7. Il s'agit de l'effet Doppler. On observe alors un décalage vers le rouge - redshift - si la source s'éloigne de l'observateur et vers le bleu - blueshift - si elle s'en rapproche. On calcule le décalage spectral selon l'équation (1.11).

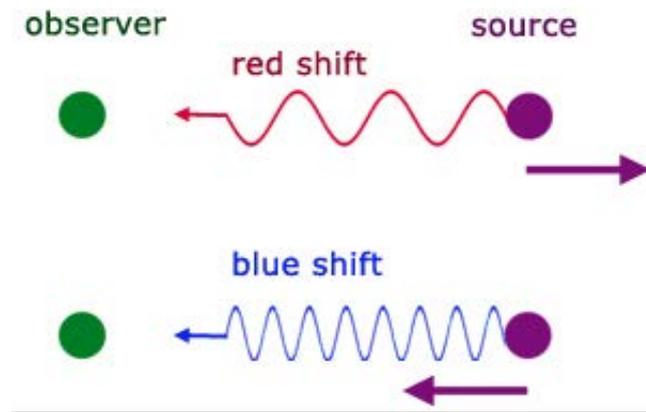


FIGURE 1.7 – Illustration d'un décalage spectral dû à un effet Doppler. Un observateur, immobile, reçoit un signal émis par une source en mouvement. Le sens de déplacement de la source affecte la longueur d'onde du signal. Cette dernière est allongée - raccourcie - lorsque la source s'éloigne - se rapproche. Source : <http://a-levelphysicstutor.com/wav-doppler.php>.

### Effet cosmologique

L'expansion de l'Univers est également une source de décalage spectral mais cette fois il s'agit systématiquement d'un décalage vers le rouge, Figure 1.8. On parle alors de redshift cosmologique. Dans un univers homogène et isotrope décrit par une métrique FLRW son expression en fonction du facteur d'échelle est donnée par l'équation (1.12).

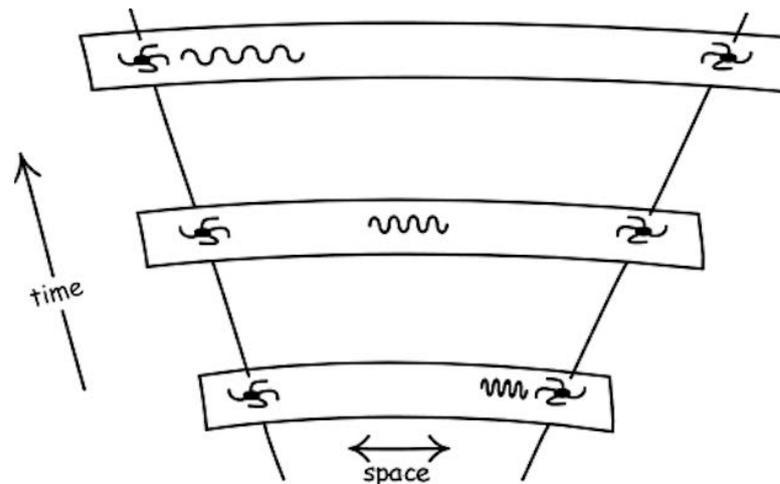


FIGURE 1.8 – Illustration de l'agrandissement de la longueur d'onde entre le temps d'émission et le temps de réception dû à l'expansion de l'Univers. Source : <http://www.pitt.edu>.

Pour comprendre la différence entre ces deux définitions il convient de distinguer les effets cosmologiques des effets locaux. La chose importante à remarquer est que les objets ne bougent plus lorsque l'on se place en coordonnées comobiles. Cependant il existe tout de même un décalage vers le rouge dû à l'expansion de l'espace lui-même. C'est le redshift cosmologique.

### Effet gravitationnel

La relativité générale prédit que la lumière se déplaçant dans un champ de gravitation non homogène subit un décalage vers le rouge ou vers le bleu selon que la lumière est émise ou reçue par l'astre, Figure 1.9. Cet effet est appelé le décalage d'Einstein ou redshift gravitationnel. Sur la Terre, il est faible mais a été mesuré par [Pound et Rebka \(1959\)](#). Pour les galaxies cet effet est de l'ordre de  $10^{-5}$  ([Wojtak et collab., 2015](#)).

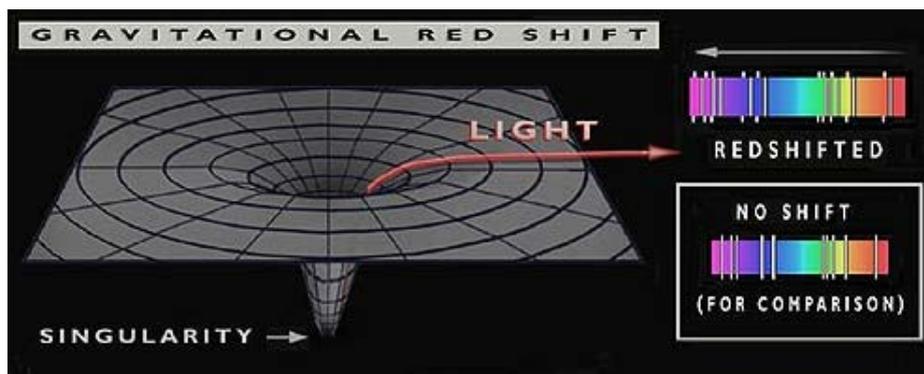


FIGURE 1.9 – Illustration de l'agrandissement de la longueur d'onde d'un signal s'échappant du potentiel gravitationnel d'un objet massif. Source : <http://archive.ncsa.illinois.edu/Cyberia/NumRel/EinsteinTest.html>.

### 2.3 Redshift spectroscopique

La classification des galaxies se réfère toujours à la séquence de Hubble visible à la Figure 1.10. Cette classification est basée sur des critères morphologiques des galaxies dont les composants principaux sont :

- le bulbe central, dont le profil de brillance est en général décrit par une loi de de Vaucouleurs :

$$I(r) = I_e \times 10^{-3.33((r/r_e)^{0.25}-1)} , \quad (1.13)$$

où  $r_e$  représente le rayon à demi-luminosité et  $I_e$  est la brillance de surface à  $r = r_e$ .

- le disque, dont le profil de brillance est décrit par la fonction :

$$I(r) = I_0 \times \exp^{-r/r_0} , \quad (1.14)$$

où  $r_0$  est la longueur caractéristique du disque et  $I_0$  est la brillance centrale à  $r = 0$ .

- le halo de matière baryonique constitué d'étoiles âgées.
- le halo de matière noire qui représente la majorité de la masse.

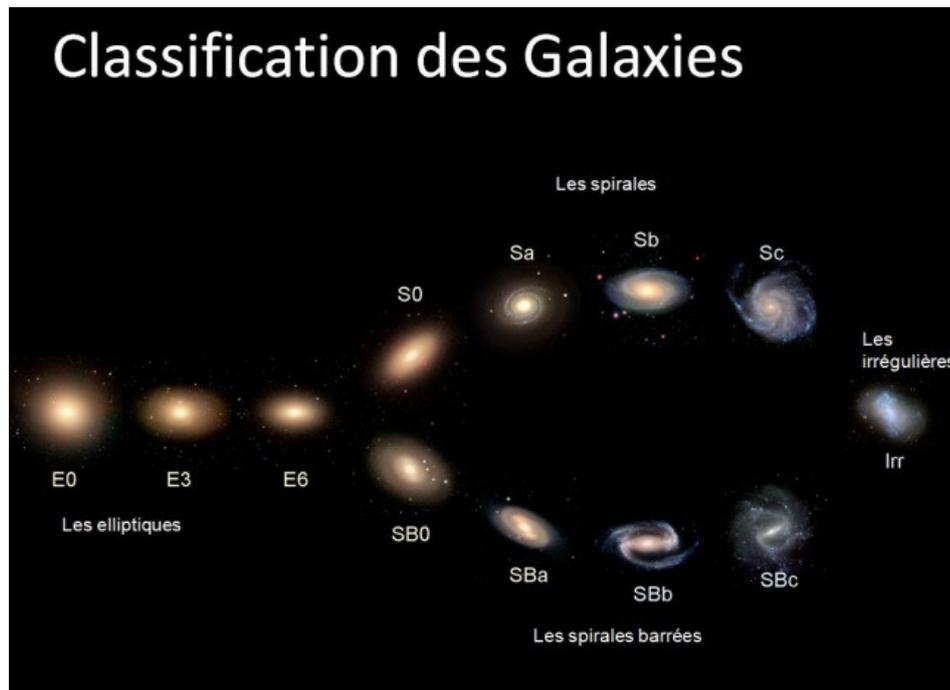


FIGURE 1.10 – Séquence morphologique de Hubble. Les galaxies elliptiques sont notées E, les spirales sont notées S, les lenticulaires sont notées S0 et les irrégulières sont notées Irr. Source : <https://moulindesetoiles.wordpress.com/les-galaxies/>.



En parcourant la séquence de Hubble de droite à gauche la morphologie évolue du type tardif au type précoce ce qui correspond globalement à l'évolution d'une galaxie. En effet, les galaxies jeunes sont des objets actifs dont la formation d'étoiles dans leur disque est en cours. Elles sont généralement bleues. Puis à un certain moment dans l'évolution de la galaxie, la formation d'étoiles s'arrête, la galaxie devient passive avec une population stellaire vieillissante. La galaxie devient rouge. Les galaxies inhabituelles comme celles possédant un noyau très actif (*active galactic nucleus*, AGNs), celles à fort taux de formation stellaire ou encore celles en interaction ne sont pas représentées sur cette séquence. Les mécanismes contrôlant l'arrêt de la formation d'étoiles font toujours l'objet de recherches.

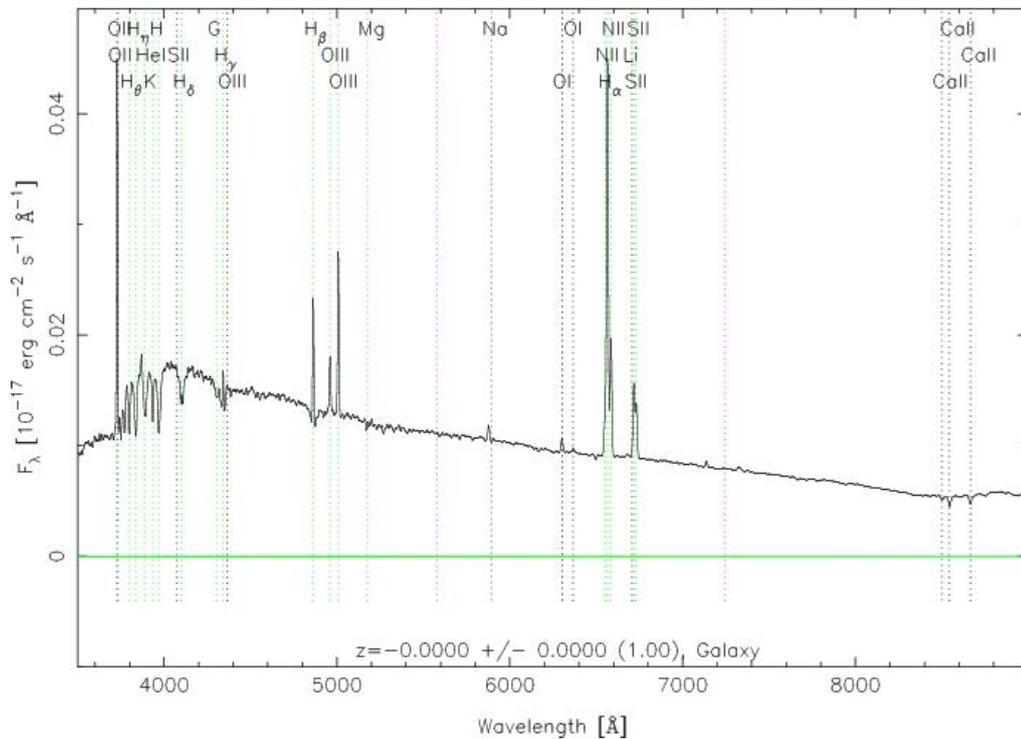


FIGURE 1.12 – Spectre d'une galaxie de type tardif. Source : <http://classic.sdss.org/dr7/algorithms/spectemplates/>.

Mesurer le redshift revient à identifier au moins une caractéristique connue dans le spectre lumineux et d'en mesurer le décalage spectral via l'équation (1.11). Cela nécessite une bonne connaissance du spectre des galaxies c'est à dire de leur distribution spectrale d'énergie (SED). L'acquisition du spectre se fait en séparant le faisceau de lumière incidente à l'aide d'un prisme ou d'une fente.

La composition et le mécanisme d'émission des étoiles sont à présent bien connus : le spectre est proche de celui d'un corps noir, avec des raies d'absorptions et d'émissions qui sont la signature de sa composition, en fonction de son âge et de sa taille. Les étoiles jeunes émettent fortement dans l'ultraviolet, sont très énergétiques, mais

le flux diminue rapidement à mesure que les étoiles évoluent et leur spectre subit alors une forte diminution du flux à 912 Å. Cette cassure, qui correspond à l'absorption de l'hydrogène, est appelée discontinuité de Lyman. Les étoiles vieillissantes sont de plus en plus rouges et créent une discontinuité supplémentaire, celle de Balmer, à  $\sim 4000$  Å. L'évolution en couleur des étoiles jeunes est très rapide et devient de plus en plus rouge avec le temps.

Si l'évolution des étoiles est assez bien comprise, le spectre lumineux total d'une galaxie n'est cependant pas toujours facile à prédire car les populations d'étoiles ne sont pas forcément homogènes et le flux total dépend du taux de formation stellaire. Une manière simple de modéliser le spectre d'une galaxie consiste à définir une population initiale d'étoiles (*Initial Mass Function*, IMF) et une évolution qui dépend du taux de formation stellaire (*Star Formation Rate*, SFR) et de l'évolution chimique de la galaxie (la métallicité). En plus de dépendre du type de l'objet, les caractéristiques présentes dans le spectre d'une galaxie peuvent également dépendre de l'environnement la séparant de l'observateur.

En considérant une dispersion de vitesse typique dans un amas de  $\sim 500 \text{ km.s}^{-1}$  pour des galaxies proches de moins de quelques dizaines de mégaparsecs (Mpc) l'erreur relative sur la mesure du redshift par spectroscopie est de  $\frac{\Delta z}{z} \sim 0.1\%$ . Pour de tels objets il est préférable d'utiliser la méthode des céphéides ou encore la relation de Tully-Fisher.

Au-delà de cette distance, la spectroscopie est le meilleur moyen, en termes de précision, pour mesurer le redshift. En effet, une résolution  $R = \frac{\lambda}{\Delta\lambda}$ , de quelques centaines permet d'obtenir une précision de  $\sim 5.10^{-3}(1+z)$ .

Les spectrographes multi-objets (MOS) existants permettent de mesurer simultanément des centaines de redshifts spectroscopiques. Cependant, cela n'est pas suffisant pour traiter les dizaines de millions d'objets déjà répertoriés ou les quelques milliards que l'on prévoit pour des relevés comme *Euclid*. Bien que la spectroscopie soit la méthode la plus précise, elle reste donc une méthode longue et coûteuse qu'il n'est pas envisageable d'utiliser pour estimer le redshift de l'ensemble des objets à notre disposition.

## 2.4 Redshift photométrique

La photométrie est une façon de déterminer le redshift d'un objet à partir des mesures de sa magnitude apparente à travers différents filtres. Cette technique a été développée pour la première fois par [Baum \(1962\)](#). Il observa ainsi la SED de six galaxies elliptiques dans l'amas Virgo dans neuf bandes allant de 3730 Å à 9875 Å. [Koo \(1985\)](#) suivit une approche différente. Il n'utilisa que quatre filtres et au lieu d'utiliser une SED empirique il utilisa un spectre simulé issu du modèle de [Bruzual A. \(1983\)](#) qui ne prenait pas en compte l'évolution des galaxies. La méthode de la coupure ultra-violette présentée dans [Madau et collab. \(1996\)](#); [Steidel et collab. \(1996\)](#) utilise la coupure à 912 Å dans le spectre des galaxies due à la limite de Lyman de l'atome d'hydrogène. Le passage de cette coupure à travers la bande  $u$  (généralement centrée à  $\sim 3\,000$  Å) permet de mettre une limite basse sur le redshift :  $z > 2.25$ . Cette méthode est utilisée pour préselectionner des galaxies à haut redshift avant d'en faire la mesure de façon spectroscopique. Vers la fin des années 1990 et au début de années 2000 de grands relevés astronomiques comme le *SDSS* ont grandement participé au développement des redshifts photométriques ([Csabai et collab., 2003](#)).

On peut distinguer deux types d'approches permettant de mesurer le redshift photométrique : les méthodes utilisant des modèles (ajustement de modèles, approche Bayésienne...) et les méthodes empiriques de type « apprentissage automatique » (réseaux de neurones, forêts d'arbres décisionnels...).

### (i) Méthodes utilisant des modèles

*L'ajustement de modèles* est une procédure qui permet d'obtenir le redshift photométrique. Elle est basée sur la détection de motifs particuliers dans le spectre des galaxies à partir des mesures de magnitude dans différents filtres et sur l'ajustement de l'allure globale du spectre par rapport à une bibliothèque de spectres de références. La Figure 1.13 montre quelques exemples d'ajustement pour plusieurs modèles de galaxies.

Afin d'obtenir des résultats plus précis, l'ensemble de filtres utilisé doit être choisi de façon à encadrer certaines caractéristiques du spectre des galaxies. On utilise généralement la brisure à 4000 Å (due à l'accumulation des raies d'absorptions des métaux premièrement ionisé tel que le doublet du CaII [lignes H et K]); mais aussi les raies de Balmer (correspondant à la série de Balmer de l'hydrogène) et la brisure de Lyman à 912 Å. Les magnitudes photométriques observées sont alors comparées à celles obtenues via l'échantillon de spectres de référence. Ces spectres de référence sont soit observés soit synthétiques. La méthode d'ajustement de modèles consiste à réduire l'immensité des différents types de galaxies en sous-échantillons de plus en plus petits en calculant leurs différentes valeurs de magnitudes à différents redshifts.

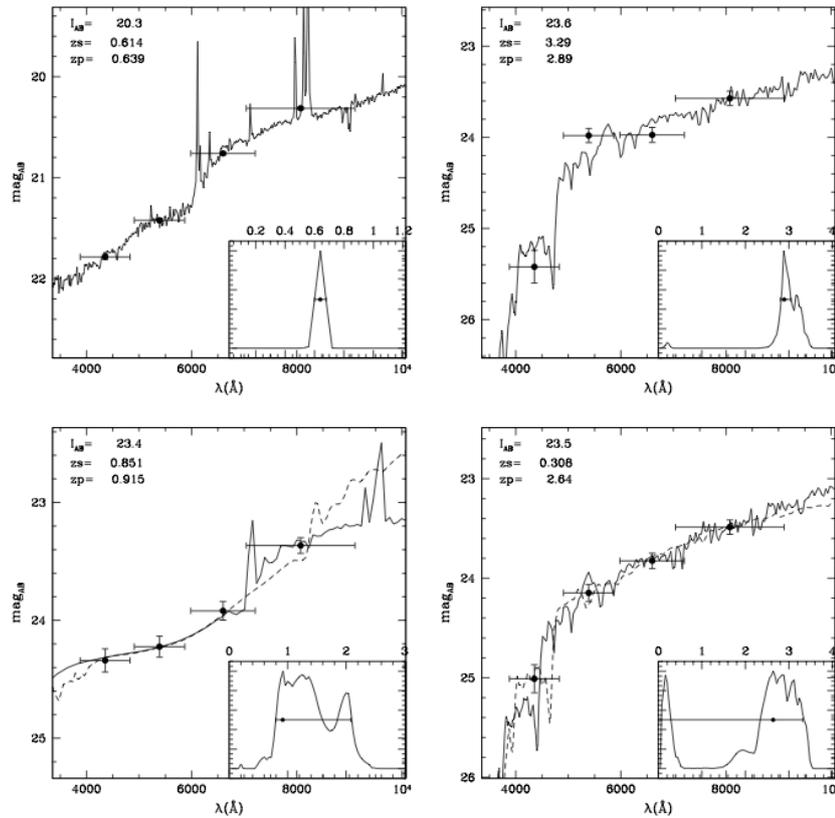


FIGURE 1.13 – Exemple d’ajustement des templates de galaxies sur les magnitudes apparentes  $B$ ,  $V$ ,  $R$  et  $I$  du  $VVDS - F02$  (points noirs). Le spectre en trait continu correspond au meilleur template ajusté. Si une seconde solution est probable, le template est représenté en tirets. La fonction de distribution en probabilité du redshift (PDFz) est présentée dans un cadre réduit pour chaque galaxie.  $z_p$  est le redshift photométrique,  $z_s$  est le redshift spectroscopique. Source : [Ilbert \(2004\)](#).

En résumé, cela revient à créer une grille de galaxies « standard » à différents redshifts puis à associer un ensemble de valeurs de flux à chaque point de la grille. Enfin, pour une galaxie test, il s’agit de localiser le point de la grille qui ressemble le plus à l’objet étudié. La valeur du redshift est soit directement la valeur de ce point sur la grille soit interpolée à partir des points les plus proches selon le meilleur ajustement de la SED photométrique via la minimisation du  $\chi^2$  :

$$\chi^2(z, T, A) = \sum_{f=1}^{N_f} \left[ \frac{F_{obs}^f - A \times F_{pred}^f(z, T)}{\sigma_{obs}^f} \right]^2, \quad (1.15)$$

avec :

- $N_f$ , le nombre total de filtres.
- $A$ , un facteur permettant de renormaliser le flux du modèle au flux de la galaxie.
- $F_{obs}^f$ , le flux observé de l'objet dans la bande large  $f$ .
- $F_{pred}^f$ , le flux dérivé du spectre théorique en prenant en compte le type d'objet  $T$  et le redshift  $z$ .
- $\sigma_{obs}^f$ , l'incertitude sur le flux mesuré.

La minimisation du  $\chi^2$  fournit l'ensemble de valeurs (redshift, type, brillance...) le plus probable.

Bien que cette méthode soit largement utilisée et permette d'accéder à une estimation du redshift pour de très nombreux objets, on peut noter certaines limitations d'une telle approche. En effet si les modèles ne sont pas représentatifs de l'échantillon de galaxies alors l'estimation du redshift de certains objets sera biaisée. On peut également remarquer que les étoiles sont traitées comme des galaxies, à moins que l'on ne fasse une sélection indépendante ou que l'on introduise des modèles de spectres d'étoiles.

Cette approche est également très sensible à l'étalonnage photométrique de l'échantillon utilisé. En effet il arrive que la pente du spectre soit parfois très prononcée. Dans ce cas, si la pente se trouve dans un filtre dont le décalage photométrique est aussi faible que 0.1 cela peut provoquer des décalages en redshift suffisant pour expliquer les systématiques observées dans l'estimation des redshifts photométriques.

Enfin, on peut noter divers effets tels que :

- La dégénérescence entre redshift et brillance : une galaxie peu brillante à bas redshift peut avoir la même signature qu'une galaxie lointaine et brillante.
- La confusion entre les discontinuités Lyman et Balmer. Les erreurs de redshifts, très importantes, sont alors qualifiées de catastrophiques.

*L'approche Bayésienne* pour l'estimation des redshifts photométriques développée par Benítez (2000), utilise également des modèles mais est aussi basée sur l'utilisation d'un *a priori* et d'une marginalisation Bayésienne au lieu de la minimisation du  $\chi^2$ . L'utilisation du théorème de Bayes (1.16) permet de prendre en compte de précieuses informations dans l'*a priori* comme la distribution en redshift ou la proportion des différents types de galaxies. La probabilité  $P(z|C)$  d'obtenir le redshift  $z$  pour des couleurs  $C$  est donnée par le théorème de Bayes.

$$P(z|C) = \frac{P(C|z)P(z)}{P(C)}, \quad (1.16)$$

avec :

- $P(C|z) = \mathcal{L}(z) = \exp(-\frac{\chi^2}{2})$ , la vraisemblance soit la probabilité d'obtenir les données (les couleurs) pour un redshift donné.
- $P(C) = \int P(C|z)P(z)dz$ , la normalisation de l'*a posteriori*.

Pour un ensemble de modèles, l'équation précédente se réécrit :

$$P(z, T|C) \propto P(C|z, T)P(z, T), \quad (1.17)$$

avec :

- $P(z, T|C)$ , la probabilité de redshift pour un modèle de galaxie, connaissant les données.

Un *a priori* très utilisé est celui de la probabilité du redshift connaissant la magnitude apparente de l'objet  $m_0$ . Il s'agit simplement de la distribution en redshift  $n(z)$  à une magnitude donnée. Cette distribution est généralement tirée des sondages de redshifts spectroscopiques. On a alors :

$$P(z|C, m_0) \propto P(C|z)P(z|m_0). \quad (1.18)$$

En prenant la somme des différentes probabilités, c'est-à-dire en marginalisant, on obtient :

$$P(z|C, m_0) \propto \sum_T P(z, T|C, m_0) \propto \sum_T P(C|z, T)P(z, T|m_0). \quad (1.19)$$

Le résultat s'exprime sous la forme d'une fonction de probabilité notée PDF( $z$ ). Les étapes sont résumées et illustrées à la Figure 1.14.

Comme pour l'ajustement de modèles, il existe une dégénérescence modèle/redshift, mais la magnitude apparente de l'objet permet de distinguer les deux cas. Un objet brillant a beaucoup plus de chances d'être à bas redshift qu'à haut redshift. La PDF contient toute l'information nécessaire à l'estimation finale du redshift (en supposant avoir pris en compte tous les modèles). En général, le redshift retenu est celui du maximum de vraisemblance, qui correspond au plus grand pic - le plus petit  $\chi^2$  - et la forme de la PDF permet de définir un degré de confiance de la mesure.

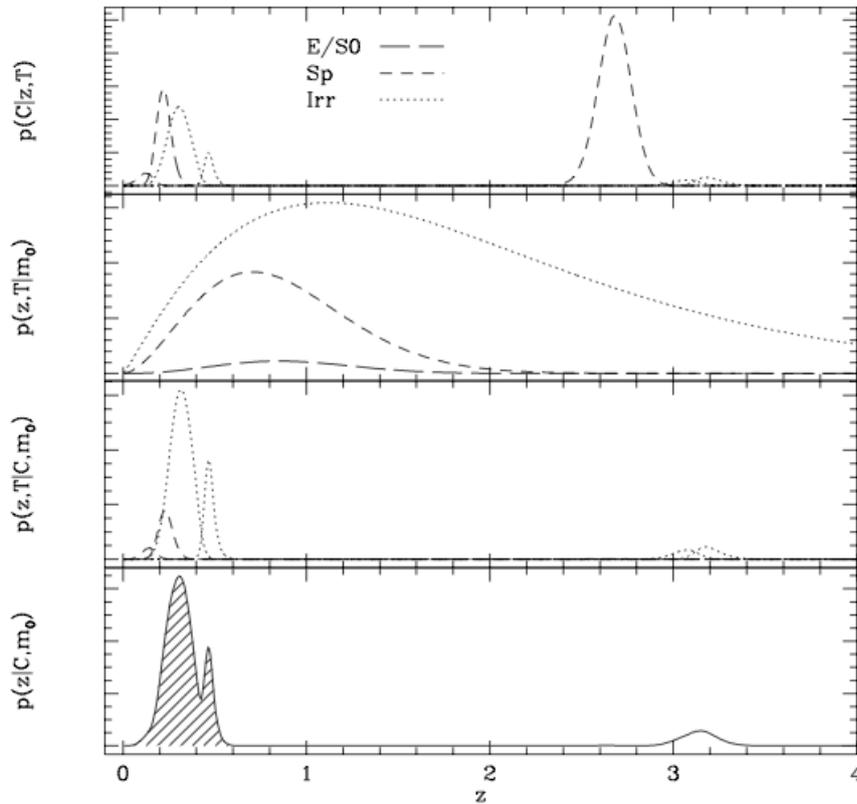


FIGURE 1.14 – Estimation bayésienne du redshift pour une galaxie située à  $z = 0.28$  de type spectral *Irr* et  $I = 26$ .

De haut en bas : *i*) vraisemblance pour différents types de galaxies donnés. *ii*) Apriori sachant la magnitude de l'objet pour chaque type spectral. *iii*) à *posteriori* pour chaque type de galaxie. *iv*) à *posteriori* marginalisé dont le maximum se trouve à  $z = 0.305$ . La région hachurée correspond à la région où la fiabilité du redshift atteint 91%. Source : [Benítez \(2000\)](#).

De façon générale, les désavantages des méthodes utilisant des modèles sont :

- Ces méthodes dépendent de la qualité des modèles et de la précision de la photométrie.
- Les résultats peuvent être biaisés et sont souvent plus bruités que les méthodes empiriques.

## (ii) Méthodes empiriques : apprentissage automatique

*Les réseaux de neurones* utilisent la puissance de calcul des ordinateurs modernes pour s'entraîner à reconnaître, et à reproduire, les relations entre les objets brillants pour lesquels à la fois la spectroscopie et la photométrie sont possibles. On peut alors utiliser ces relations pour des objets moins brillants pour lesquels seule la photométrie est possible.

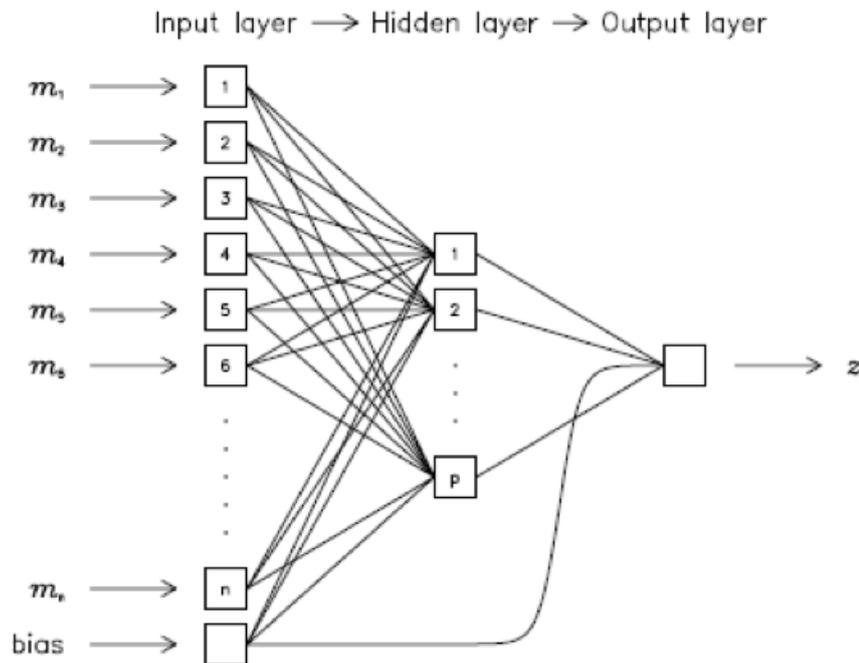


FIGURE 1.15 – Diagramme schématique d'un réseau de neurones avec : en entrée les magnitudes dans différents filtres, une étape cachée et une sortie fournissant le redshift. Chaque ligne de connexion porte un poids. Le noeud biais, permet de prendre en compte certains décalages systématiques. Source : [Collister et Lahav \(2004\)](#)

À partir d'un échantillon de référence le programme s'entraîne, seul, à retrouver la bonne réponse, Figure 1.15. Cela ressemble à l'approche Bayésienne à ceci près qu'ici le programme construit lui-même l'a priori à partir de l'échantillon de référence. C'est l'étape d'apprentissage. En pratique, il est nécessaire d'utiliser un échantillon de redshifts spectroscopiques que l'on divise en deux. Le premier sert à l'entraînement du programme, le second est utilisé pour vérifier les résultats.

Il y a trois conditions nécessaires pour justifier l'utilisation d'un réseau de neurones :

- les données utilisées pour l'étalonnage du programme doivent être des redshifts spectroscopiques.
- le sous-échantillon doit être représentatif de l'échantillon.
- les données doivent avoir été observées avec le même instrument et étalonnées exactement de la même façon.

*La forêt d'arbres décisionnels* est générée à partir de ré-échantillonnages successifs, de type *bootstrap*, d'un jeu de données de référence afin d'obtenir un ensemble d'arbres de prédiction.

Ces arbres sont construits en divisant l'échantillon de référence en deux branches selon une liste de critères consécutifs (couleurs, magnitudes...). Les sous ensembles obtenus sont à nouveau divisés jusqu'à l'obtention d'un sous ensemble particulier (appelé feuille) lorsqu'un critère d'arrêt est rencontré. Ce critère d'arrêt peut être par exemple, une taille de feuille minimale. Pour une galaxie donnée, le ré-échantillonnage de ses paramètres est fait en perturbant les valeurs mesurées conformément à l'erreur sur chaque variable qui est supposée suivre une distribution normale.

Les prédictions de ces arbres sont ensuite combinées en une PDF( $z$ ) finale qui peut être utilisée pour estimer le redshift d'une galaxie ainsi que son erreur. La forêt d'arbres décisionnels est une méthode d'apprentissage automatique d'ensemble qui génère plusieurs arbres de prédiction puis, combine leurs prédictions de façon à obtenir une PDF( $z$ ) pour chaque galaxie, voir Carrasco Kind et Brunner (2013) pour plus de détails.

De façon générale les méthodes empiriques possèdent deux principaux désavantages :

- Elles sont spécifiques à un jeu de données.
- L'échantillon d'entraînement doit couvrir la même gamme de couleurs, redshift, types que les objets cibles. Ce qui est très difficile à réaliser en pratique.

Bien que ces méthodes (modèles et empiriques) soient basées sur des concepts et des formalismes différents, leur rôle consiste à reproduire la relation existant entre l'espace photométrique - incluant brillances, positions, ellipticité... - et l'espace des redshifts, Figure 1.16. Dans le cas des redshifts photométriques, la reproduction de cette relation se base sur la corrélation entre les magnitudes mesurées et des SED ou des jeux de données servant à l'entraînement du programme.

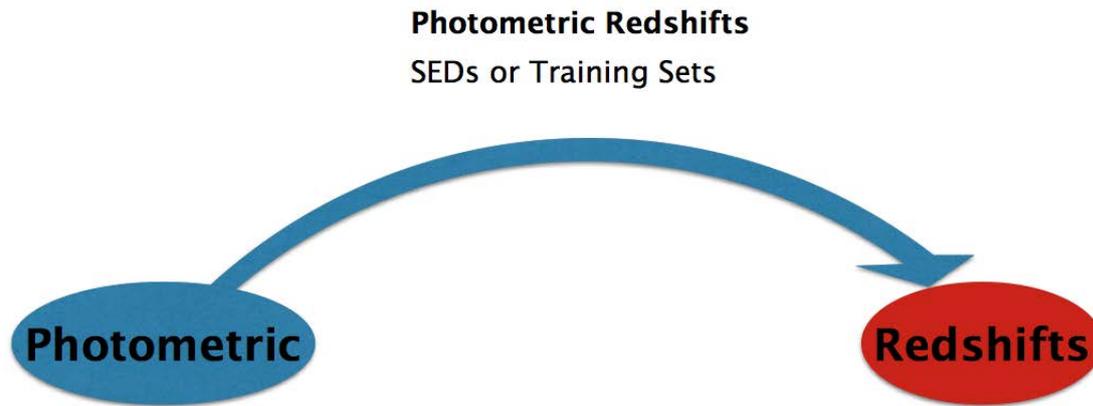


FIGURE 1.16 – Schéma résumant et unifiant les différents types de redshift photométrique. De façon générale, il s’agit d’une méthode permettant de reproduire la relation existante entre l’espace photométriques - brillances, positions, ellipticité... - et l’espace des redshifts.

Cependant, la multitude de codes existants - BPZ (Benítez, 2000), EASY (Brammer et collab., 2008), GALEV (Kotulla et collab., 2009), GOODZ (Dahlen et collab., 2007, 2005), Hyperz (Bolzonella et collab., 2000), Le Phare (Arnouts et collab., 1999; Ilbert et collab., 2006), LRT (Assef et collab., 2008, 2010), ZEBRA (Feldmann et collab., 2006), ANNz (Collister et Lahav, 2004), BDT (Gerdes et collab., 2010) etc... - soulève la question de l’unicité de la réponse.

En effet selon la méthode utilisée, un même objet peut avoir des redshifts photométriques différents. Ce point est développé dans l’analyse PHATz (Hildebrandt et collab., 2010) qui compare les résultats obtenus par de nombreux codes de redshifts photométriques, Figure 1.17. La dispersion autour de la droite  $z_p = z_s$  ainsi que le taux de redshifts catastrophiques pour un même échantillon varient d’un code à l’autre. Ces différents estimateurs fournissent en fait des mesures différentes de la même quantité fondamentale : la PDF( $z$ ) d’une galaxie.

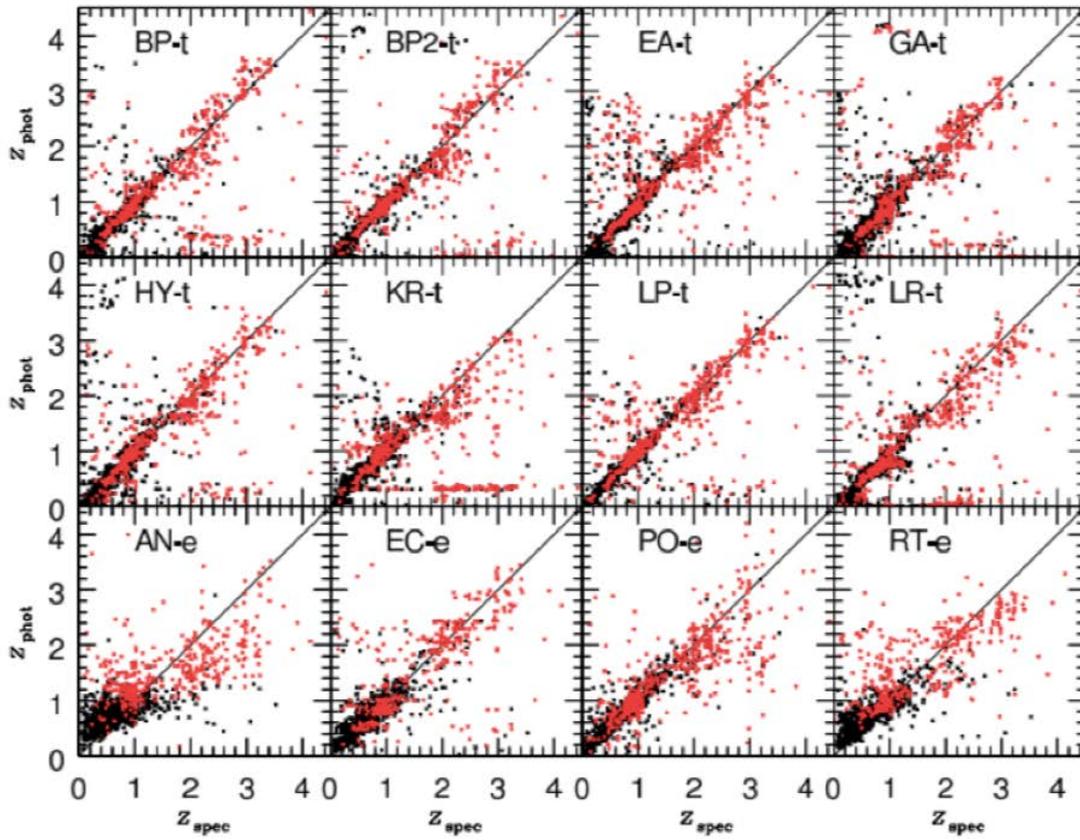


FIGURE 1.17 – Résultat d’une étude comparant différents codes de redshift photométrique réalisée par [Hildebrandt et collab. \(2010\)](#). Le même échantillon de galaxies tests est utilisé lors de l’estimation du redshift pour chacun des codes. La dispersion autour de la droite  $z_p = z_s$  ainsi que le taux de redshifts catastrophiques pour un même échantillon varient d’un code à l’autre. Les points rouges correspondent aux objets ayant une magnitude en bande  $R \geq 24$ , les noirs correspondent à  $R \leq 24$ .

### 3 Un œil vers l'avenir

Le travail effectué durant cette thèse s'inscrit dans le cadre des relevés grands champs à venir dont l'objectif est, principalement, de percer les mystères de l'énergie sombre. Dans cette section je présente les relevés *DES*, *LSST* et *Euclid* qui sont assez proches. Chacun d'eux est décrit plus en détails ci-dessous.

#### 3.1 Les futurs grands relevés et les sondes cosmiques

##### (i) *DES : Dark Energy Survey*

Ce sondage ([The Dark Energy Survey Collaboration, 2005](#)) vise à observer 5 000 deg<sup>2</sup> dans les cinq bandes  $g, r, i, z, Y$  ( $i_{AB} < 24.3$ ) pour mesurer les distorsions gravitationnelles, dériver les redshifts photométriques ([Banerji et collab., 2008](#)) et compter les amas de galaxies afin de contraindre l'équation d'état de l'énergie noire. Les observations sont effectuées au télescope de 4 m Blanco du *Cerro Tololo Inter-American Observatory (CTIO)* situé à 2 207 mètres au Chili où le seeing median est de 0.9 arcsec, Figure 1.18.

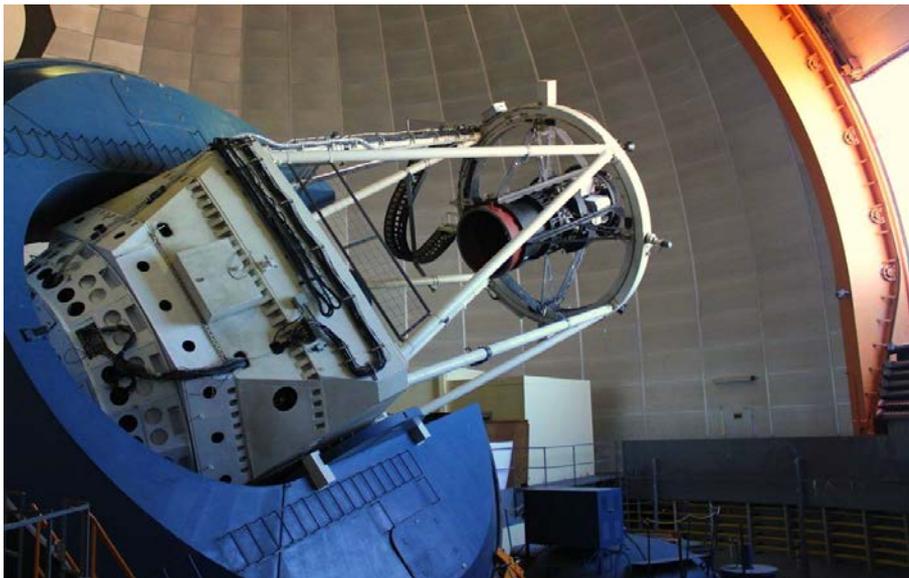


FIGURE 1.18 – Photo du télescope utilisé pour le *Dark Energy Survey*.  
Source : <https://news.slac.stanford.edu>.

(ii) *LSST : Large Synoptic Survey Telescope*

Le relevé ([LSST Dark Energy Science Collaboration, 2012](#)) observera 20 000 deg<sup>2</sup> avec son miroir de 8.4 m. Il est prévu d'équiper la caméra, au champ de vue de 10 deg<sup>2</sup>, de six bandes photométriques  $u, g, r, i, z, Y$  ( $r < 24.5$ ). Le télescope se situera sur le Cerro Pachón, une montagne nord-Chilienne située à 2 680 mètres d'altitude où le seeing médian est de 0.67 arcsec.

(iii) *Euclid*

L'Agence Spatiale Européenne (ESA) a l'ambitieux projet ([Laureijs et collab., 2011](#)) de concevoir et lancer un satellite équipé d'un télescope de 4.5 mètres afin de réaliser un relevé profond de 40 deg<sup>2</sup> ( $AB_{VIS} = 26.5$ ,  $AB_{Y,J,H} = 26.0$ ), et un relevé large de 15 000 deg<sup>2</sup> ( $AB_{VIS} = 24.5$ ,  $AB_{Y,J,H} = 24.0$ ) du ciel pour observer environ un milliard de galaxies avec une très haute résolution, équivalente à celle du télescope spatial Hubble, Figure 1.19. Le satellite sera équipé d'un miroir de 1.2 m couvrant un champ de 0.5 deg<sup>2</sup>. *Euclid* sera doté de deux caméras, l'une de 600 millions de pixels (visible) et l'autre de 65 millions de pixels (proche IR). Ce projet est l'une des missions spatiales majeures des vingt prochaines années. Ses caractéristiques principales sont résumées dans la Table 1.1.



FIGURE 1.19 – Le satellite *Euclid* de l'ESA. Source : <http://www.euclid-ec.org>.

---

**Caractéristiques de la mission**


---

Organisation	ESA
Objectif scientifique principal	Cosmologie et physique fondamentale : comprendre les propriétés et la nature de l'énergie sombre
Date de lancement	Q1 2020
Orbite	Second point de Lagrange (L2)
Durée de la mission	6.25 années
Masse totale	2 100 kg
Charge utile	855 kg
Taille	longueur 4.5 m, diamètre 3.1 m
Télescope	miroir 1.2m Korsch Silicon Carbide
Distance focale	24.5 m
Domaine de longueur d'onde couvert	Visible (550-900 nm) Infrarouge proche (900-2000 nm)
Téléométrie	855 Gbit/jour
Mode d'observation	<i>Step and stare</i> , temps d'exposition jusqu'à 4 500 sec/champ
Lanceur	ESA/AE : Soyuz-Fregat
Site de lancement	ESA/CSG : Kourou spaceport
Satellite et module de service	Thales Alenia Space
<i>Payload module</i> et télescope	Airbus Defense and Space division
Instruments	Imageur visible (VIS) Spectromètre et photomètre proche infrarouge (NISP)
Taille des pixels	0.1 arcseconde (VIS) ; 0.3 arcsecond (NISP)
Champ de vue commun VIS et NISP	0.53 deg <sup>2</sup>
Filtres	R, I et Z (VIS), Y, J et H (NISP)
Couverture spectrale des grisms	1 "bleu" (920 nm - 1250 nm) 3 "rouge" (1250 nm - 1850 nm)
Résolution des grisms	380 pour une source de 0.5 arcseconde
Détecteurs CCD visible	36, 4096×4096 pixels
Détecteurs proche IR	16, 2048×2048 pixels
Relevé champ large	15 000 deg <sup>2</sup> , AB <sub>VIS</sub> = 24.5, AB <sub>YJH</sub> = 24.0
Relevé champ profond	40 deg <sup>2</sup> , AB <sub>VIS</sub> = 26.5, AB <sub>YJH</sub> = 26.0
Sondes cosmologique principale	<i>Weak lensing, Galaxy Clustering</i>
Opération	ESOC
Segment sol	ESAC et Euclid Consortium

TABLE 1.1 – Table présentant les caractéristiques principales de la mission spatiale *Euclid*.  
Source : [http://www.euclid-ec.org/?page\\_id=2540](http://www.euclid-ec.org/?page_id=2540)

Son objectif scientifique principal est de comprendre la source de l'accélération de l'expansion de l'Univers et de découvrir la véritable nature de ce que les physiciens et astrophysiciens appellent l'énergie sombre. *Euclid* soulèvera donc les questions suivantes : l'énergie sombre est-elle une constante cosmologique comme celle introduite par Einstein ? Correspond-elle à un nouveau champ évoluant de façon dynamique avec l'expansion de l'Univers ? Correspond-elle à une déviation des lois de la gravité et indique-t-elle les limites de la Relativité Générale ? Quelles sont la nature et les propriétés de la matière noire ? Quelles sont les conditions initiales qui ont mené à la formation des grandes structures de l'Univers ? Quelle sera l'évolution future de l'Univers dans les prochains 10 milliards d'années ?

Les empreintes de l'énergie sombre et de la gravité seront détectées à partir de leurs signatures sur le taux d'expansion de l'Univers et sur la croissance des structures cosmiques à partir des effets de lentilles gravitationnelles faibles (*Weak Lensing*, WL) et des propriétés d'agglomération des galaxies via les oscillations acoustiques des baryons (BAOs) et la distorsion espace-redshift (RSD). Les BAOs permettent de sonder directement la relation distance-redshift afin d'explorer le taux d'expansion de l'Univers, alors que le WL est une sonde quasi directe de la matière noire mais combine les distances angulaires (qui sondent le taux d'expansion) et le contraste de densité massique (qui sonde le taux de croissance des structures et la gravité). Combinées ensemble ces trois sondes robustes et complémentaires permettent de sonder les effets de l'énergie sombre.

Ces observations seront complétées par des observations indépendantes venant également des données *Euclid* sur les amas de galaxies et l'effet Sachs-Wolfs Intégré. Ces observations permettront de vérifier les résultats obtenus à partir du WL, des BAOs et du RSD et permettront également de mieux comprendre et contrôler les erreurs systématiques.

Parallèlement à la science principale d'*Euclid*, le relevé large couvrant 15 000 deg<sup>2</sup> jusqu'à  $AB_{VIS} = 24.5$  et  $AB_{YJH} = 24.0$  détectera plusieurs milliards de galaxies permettant l'emploi de sondes additionnelles et complémentaires au WL et aux amas de galaxies. De même, le champ profond de 40 deg<sup>2</sup> allant jusqu'à  $AB_{VIS} = 26.5$  et  $AB_{YJH} = 26.0$  fournira un échantillon unique de plusieurs milliards de sources pour l'astronomie stellaire et extragalactique.

Les relevés larges et profonds fourniront également de nombreuses cibles d'un grand intérêt pour *ALMA JWST*, *E-ELT* et *TMT* et seront en synergie avec *LSST*, *GAIA*, *e-ROSITA* ou encore *SKA*.

L'instrument visible d'*Euclid* (VIS) a pour rôle d'imager toutes les galaxies du relevé *Euclid* avec une très haute qualité d'image. Il sera utilisé pour mesurer les formes des galaxies afin d'en déduire les effets de lentilles gravitationnelles induits les grandes structures de l'Univers sur l'image des galaxies. Cela permettra de savoir de quelle façon la matière noire est distribuée ainsi que la façon dont cette distribution a évolué ces 10 derniers milliards d'années.

Le plan focal de VIS se compose d'une matrice de 6×6 4096×4096 e2v CCDs couvrant un champ de vue de 0.57 deg<sup>2</sup> avec des pixels de 0.1". Il sera équipé d'un

filtre à bande large couvrant le domaine de longueur d'onde allant de 550 à 900 nm.

L'instrument s'occupant de la spectrométrie et de la photométrie en infrarouge proche (NISP) a pour rôle de fournir la photométrie en infrarouge proche (de 1000 à 2000 nm) de toutes les galaxies également observées avec VIS ainsi que les spectres basse résolution en infrarouge proche et les redshifts de millions de galaxies. La photométrie en infrarouge proche sera combinée avec les données de VIS pour l'obtention des redshifts photométriques.

Le plan focal de NISP se compose d'une matrice de  $4 \times 4$   $2048 \times 2048$  détecteurs Teledyne TIS H2RG couvrant un champ de vue de  $0.53 \text{ deg}^2$  partagé avec VIS avec une taille de pixels de  $0.3''$ . Le canal photométrique sera équipé de 3 filtres à bande larges (Y, J et H) couvrant le domaine de longueur d'onde allant de 1000 à 2000 nm. Le canal spectroscopique sera équipé de 4 grisms à basse résolution ( $R = 380$ ) dans le proche infrarouge, 3 rouges allant de 1250 nm à 1850 nm et un bleu allant de 920 nm à 1250 nm.

Durant ses 6 ans d'opération la mission *Euclid* collectera plus de 500 000 images visibles (VIS) et infrarouges (NISP imagerie et spectroscopie) correspondant à environ 1 Teraoctet de données par jour qui seront quotidiennement transférés sur Terre.

Les images spatiales d'*Euclid* seront complétées par des observations dans le visible réalisées à partir de télescopes au sol couvrant la même région du ciel qu'*Euclid* pour l'obtention des redshifts photométriques. Le coût des observations sols, la mise à disposition de plusieurs télescopes pendant plusieurs centaines des nuits ainsi que les ressources humaines et matérielles nécessaires au traitement de ces données sont techniquement et opérationnellement des obstacles difficiles.

Environ 10 milliards de sources seront observés par *Euclid* parmi lesquelles plus d'un milliard seront utilisées pour l'étude de WL et environ 35 millions de spectres.

La capacité d'observation unique d'*Euclid* en fait donc un outil de choix dans l'étude de l'Univers et promet des retombées dans la majorité des domaines de l'astrophysique.

#### (iv) *Les sondes cosmiques*

Le principal objectif de ces relevés est d'approfondir notre connaissance de l'accélération de l'expansion de l'Univers. Cette expansion accélérée peut s'expliquer de quatre façons différentes : l'existence d'une constante cosmologique, une nouvelle interaction s'exerçant à grande échelle, une modification des lois de la gravité telle que nous les connaissons actuellement, ou encore si nous vivons dans un univers inhomogène (Martin, 2012).

Afin de pouvoir mesurer l'équation d'état de l'énergie sombre et de pouvoir la distinguer des modèles de gravité modifiée, ces expériences combineront plusieurs sondes cosmiques comme : les supernovae de type Ia (SN), les oscillations acoustiques

des baryons (BAO), les amas de galaxies (GC), la distorsion dans l'espace des redshifts (RSD), l'effet Sachs-Wolfe intégré (ISW) et l'effet de lentille gravitationnelle faible (WL).

### *Les supernovae de type Ia*

Les *SNIa* sont des explosions thermonucléaires résultant de l'explosion d'une naine blanche qui a atteint la limite de Chandrasekhar (dépassement de la pression de dégénérescence électronique) après avoir accreté suffisamment de matière de la part d'un compagnon, Figure 1.20. Voir Branch et collab. (1997); Branch (1998); Branch et collab. (2001) pour plus de détails.

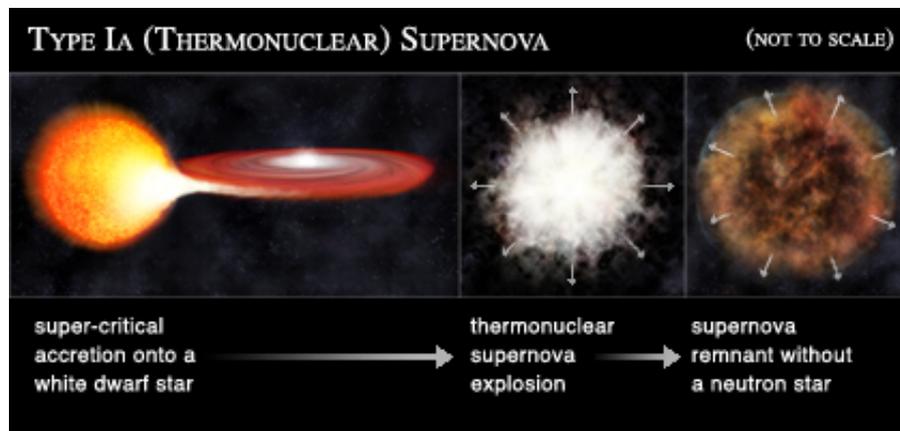


FIGURE 1.20 – Schéma décrivant l'accrétion que subit un compagnon par un objet compact. Cette accrétion mène à une explosion thermonucléaire de l'objet compact lorsqu'il dépasse la limite de Chandrasekhar. L'objet compact expulse ainsi la quasi-totalité de sa masse qui se propage dans l'espace sous la forme de reste de supernovae.

Source : [http://chandra.harvard.edu/xray\\_sources/supernovas.html](http://chandra.harvard.edu/xray_sources/supernovas.html). Illustration : NASA/CXC/M.Weiss

Comme le processus d'amorce de l'explosion est très spécifique - survenant dans des conditions bien particulières - la magnitude atteinte et la courbe de décroissance de sa luminosité sont caractéristiques de ce type de supernova. C'est pourquoi ces supernovae sont utilisées comme *chandelles standards* afin de déterminer les distances extragalactiques. On peut alors tracer un diagramme distance-luminosité aussi appelé diagramme de Hubble comme montré à la Figure 1.21.

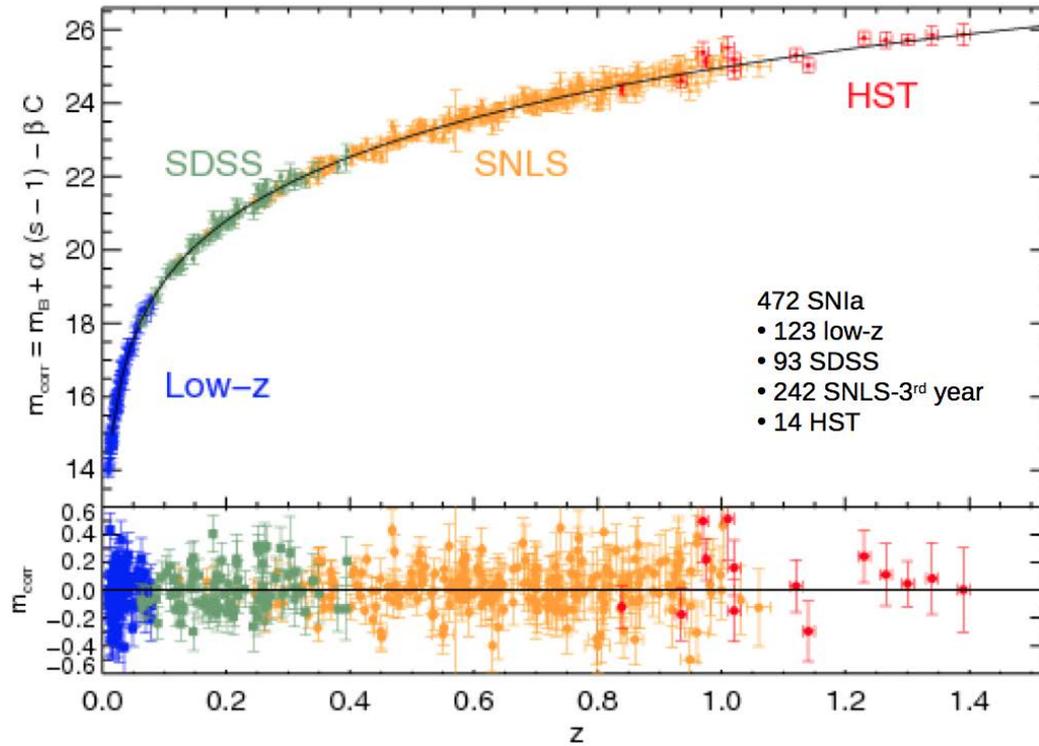


FIGURE 1.21 – Diagramme de Hubble réalisé par [Conley et collab. \(2011\)](#), montrant l'accord entre les observations - points de couleurs - et le modèle cosmologique  $\Lambda$ CDM, c'est-à-dire avec énergie sombre, en noir.  $s$  est le facteur d'étendue utilisé pour standardiser temporellement la courbe de lumière d'une supernova.  $C$  est la couleur mesurée pour une supernova.  $\alpha$  et  $\beta$  sont des paramètres de nuisance caractérisant les relations étendue-luminosité et couleur-luminosité.

La méthode des supernovae est une des plus célèbres puisqu'elle a permis à [Perlmutter et collab. \(1999\)](#); [Riess et collab. \(1998\)](#) de prouver l'accélération de l'expansion de l'Univers et d'obtenir le prix Nobel en 2011.

### *Les oscillations acoustiques des baryons*

Après la nucléosynthèse primordiale, l'Univers se compose d'un plasma -  $T \sim 10$  keV - contenant principalement des photons, des électrons non relativistes, des noyaux d'hydrogène et d'hélium également non relativistes. Le plasma est à l'équilibre thermique - équilibre cinétique - via diffusion élastique d'un photon sur un électron libre.

$$\gamma + e^- \rightarrow \gamma + e^- . \quad (1.20)$$

Les surdensités initiales issues de l'inflation se propagent dans le plasma sous la forme d'ondes acoustiques. Le phénomène est analogue, à une dimension, à celui d'une corde sur laquelle on frappe brusquement. La surdensité alors appliquée se propage de proche en proche le long de la corde selon deux ondes de sens opposés.

Cela se produit jusqu'à la recombinaison, lors de laquelle les ondes acoustiques de pression sont « gelées ». La distance alors parcourue par ces ondes est appelée l'horizon sonore à la recombinaison. Cette distance est simplement égale à la vitesse du son dans le plasma multipliée par l'âge de l'Univers lors de la recombinaison. De la même façon qu'on observe une augmentation locale de la densité de l'air lors du passage d'une onde sonore classique, on observe une surdensité de matière au niveau de l'horizon sonore et donc une légère augmentation de la probabilité de trouver des galaxies séparées par cette distance. Le rayon de l'horizon sonore est d'environ  $105 \text{ Mpc} \cdot h^{-1}$ , ce qui correspond à une distance d'environ 450 millions d'années-lumière aujourd'hui. Cette distance caractéristique fournit un *étalon standard* pour la mesure des distances cosmologiques.

Les BAOs correspondent donc à un pic dans l'espace réel. Ce pic a été mesuré pour la première fois par Eisenstein et collab. (2005), voir Figure 1.22. Le terme « oscillation » vient du fait que dans l'espace de Fourier ce pic correspond à une oscillation du spectre de puissance. Pour plus de détails sur les oscillations acoustiques des baryons, voir la revue de Bassett et Hlozek (2010).

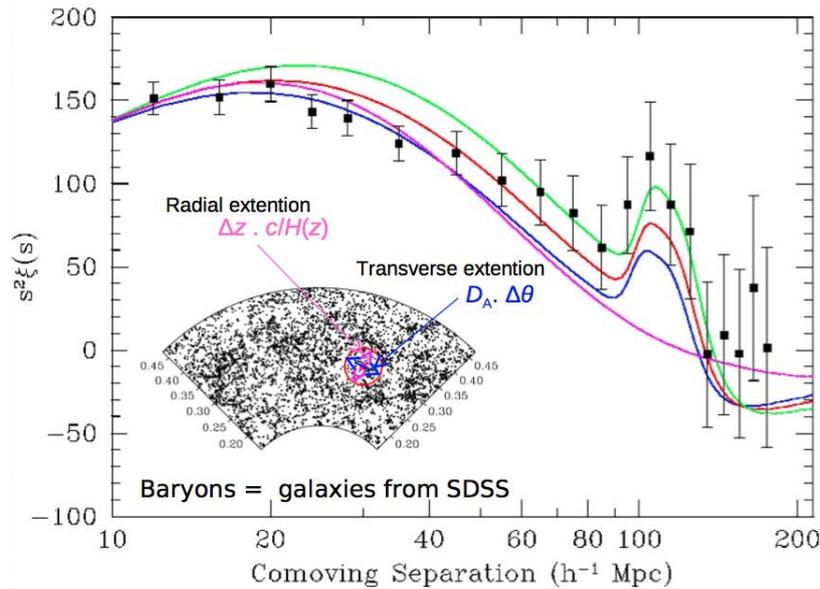


FIGURE 1.22 – Fonction de corrélation dans l’espace des redshifts, réalisée par [Eisenstein et collab. \(2005\)](#), représentant l’excès de probabilité, par rapport à une distribution Poissonienne de trouver une galaxie  $A$  à une distance  $s$  d’une autre galaxie  $B$ . Les trois courbes du haut montrent différents scénarios du modèle  $CDM$  avec des  $BAO$ . La courbe du bas correspond à un modèle  $CDM$  sans pic acoustique. Les mesures, en noir, sont en accord avec les modèles possédant un pic acoustique.

### *Les amas de galaxies*

Le nombre d’amas d’une masse donnée dans un certain volume de l’Univers ainsi que l’évolution de cette quantité avec le temps dépendent de l’énergie sombre de deux façons. Premièrement, elle influence la façon dont l’Univers grandit, c’est à dire qu’elle influence la façon dont le volume varie avec le temps. D’un autre côté, la formation de ces amas dépend de la compétition entre la gravité et l’énergie sombre. La Figure 1.23 illustre cette dépendance. En mesurant l’évolution du nombre d’amas au cours du temps, il est possible de sonder l’intensité relative de ces deux forces. Pour plus d’informations sur la formation des amas de galaxies, voir [Kravtsov et Borgani \(2012\)](#).

Contrairement aux supernovae et aux  $BAO$ , qui ne sont sensibles qu’au taux d’expansion de l’Univers, les amas de galaxies permettent de sonder à la fois le taux d’expansion ainsi que le taux de croissance des structures. La comparaison des résultats de ces deux types de sondes permet de savoir si la relativité générale est suffisante pour expliquer l’accélération cosmique.

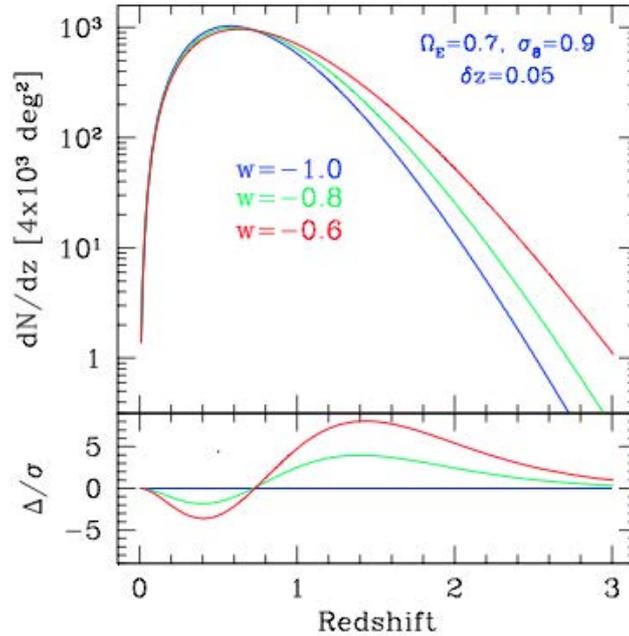


FIGURE 1.23 – Haut : distribution attendue du nombre d'amas détectés en fonction du redshift et pour différentes valeurs du paramètre  $w$ .

Bas : signification statistique de la différence entre les différents modèles.

Les modèles ont été normalisés pour reproduire les comptages à  $z \sim 0$ . Figure extraite de [Mohr et collab. \(2003\)](#).

### *Les distorsions de l'espace des redshifts*

Les distorsions de l'espace des redshifts sont un effet visible lorsque l'on trace la distribution spatiale des galaxies dans l'espace des redshifts au lieu de l'espace réel. Cet effet est dû aux vitesses particulières des galaxies causant un décalage Doppler en plus du redshift cosmologique. Cette distorsion se compose de deux effets : les "doigts de Dieu" et l'effet Kaiser.

Les "doigts de Dieu" correspondent à l'élongation de la distribution spatiale des galaxies dont l'axe d'élongation est pointé vers l'observateur ([Jackson, 1972](#)). Cela est dû à un effet Doppler associé aux vitesses particulières des galaxies dans les amas.

L'effet Kaiser ([Kaiser, 1987](#)) correspond à une distorsion causée par le mouvements cohérents des galaxies tombant vers le centre de la structure par exemple un amas. Cet effet ne produit pas une élongation mais produit un aplatissement apparent de la structure. Cet effet est bien plus faible que les doigts de Dieu et se produit aux grandes échelles.

La mesure du RSD dans les relevés de galaxies permet de sonder la formation

des structures (Percival et White, 2009) et la façon dont la gravité se comporte aux grandes échelles (Raccanelli et collab., 2013).

**L'effet Sachs-Wolfe intégré** L'effet Sachs-Wolfe intégré (Sachs et Wolfe, 1967) est une anisotropie secondaire du CMB qui se produit à cause de la variation dans le temps du potentiel gravitationnel cosmique entre un observateur et la surface de dernière diffusion des photons du CMB (*last scattering surface*). Le potentiel peut être tracé par les grands relevés de galaxies comme *Euclid* et l'effet ISW est donc une sonde reliant le CMB ( $z \sim 1100$ ) à la distribution de matière à bas redshift et peut être détecté en mesurant la corrélation croisée entre les deux. En tant que sonde cosmique l'effet ISW est directement sensible à l'énergie sombre, la courbure de l'Univers ou les théories de gravité modifiée.

### **L'effet de lentille gravitationnelle faible**

Comme les amas de galaxies, l'effet de lentille gravitationnelle permet de sonder à la fois la croissance des structures dans l'Univers en fonction du temps et, le taux d'expansion. La façon dont les grandes structures se forment dans l'histoire de l'Univers peut nous en apprendre beaucoup sur l'interaction entre la gravité et l'énergie sombre. Cependant, la plupart de ces structures sont composées de matière noire, que l'on ne peut détecter via les outils standard de l'astronomie.

Toutes les formes de matière dans l'Univers sont affectées par la gravité, et l'effet de lentille gravitationnelle est causé par l'action de la gravité sur la lumière. La lumière en provenance des galaxies lointaines est donc déviée lorsqu'elle passe à proximité d'un objet massif. Dans certains cas, la lumière est tellement déviée que l'on peut observer des images multiples d'un même objet. On parle alors d'effet de lentille gravitationnelle fort. Dans les autres cas, quand la déviation est faible, l'image des galaxies est légèrement déformée, étirée et amplifiée. Dans ce cas on parle d'effet de lentille gravitationnelle faible, voir Figure 1.24.

Cette légère distorsion de l'image d'une galaxie est appelé cisaillement cosmique et est généralement de l'ordre de 2% de l'étirement visible d'une image. Les galaxies n'étant pas parfaitement circulaires et parce qu'on regarde des galaxies différentes sous différents angles, on ne peut pas mesurer l'effet de lentille gravitationnelle faible pour un objet individuel. Par chance, la distorsion de la lumière par la gravité affecte de nombreuses galaxies dans une même région du ciel et cela de la même façon. On peut alors mener l'étude de façon statistique en cherchant des alignements dans ces déformations et donc, mesurer cet effet. Voir Hoekstra et Jain (2008), pour une revue sur l'effet de lentille gravitationnelle faible.

## **3.2 Exigence sur les *photo-z* dans le cadre de la mission *Euclid***

Le travail effectué durant cette thèse s'inscrit dans le cadre de la mission spatiale *Euclid* et, plus particulièrement, dans la partie effet de lentille gravitationnelle faible.

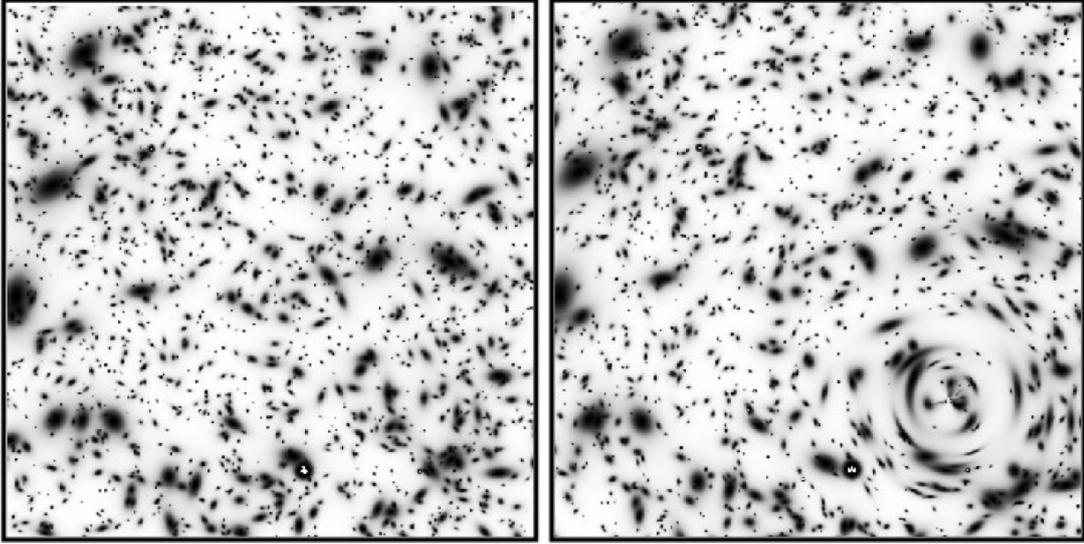


FIGURE 1.24 – Gauche : une image simulée du champ profond du télescope spatial Hubble (*HST*).

Droite : la même image simulée mais en ajoutant une lentille gravitationnelle à  $z = 0.3$ . Près de la lentille on observe : un déplacement, un grossissement et une distorsion des sources lumineuses. Figure extraite de Mellier (1999).

Afin de bien saisir le rôle crucial de la mesure des distances pour l'étude de cet effet il convient de rappeler le formalisme utilisé. Voir Dodelson (2003); Peter et Uzan (2012) pour plus de détails.

On caractérise la distorsion d'une image par sa matrice de déformation,  $\mathcal{A}$ , dont l'expression est :

$$\mathcal{A}_{ab} = \delta_{ab} - \frac{\partial^2 \psi(\vec{\theta})}{\partial \theta^a \partial \theta^b} . \quad (1.21)$$

Cette matrice, symétrique, peut être décomposée en fonction de deux paramètres : la convergence  $\kappa$  et le cisaillement  $\vec{\gamma} = (\gamma_1, \gamma_2)$ , selon :

$$\mathcal{A} = \begin{pmatrix} 1 - \kappa - \gamma_1 & \gamma_2 \\ \gamma_2 & 1 - \kappa + \gamma_2 \end{pmatrix} . \quad (1.22)$$

La convergence  $\kappa$  induit une focalisation isotrope du faisceau lumineux et est reliée au potentiel gravitationnel projeté  $\psi$  ainsi qu'aux densités surfacique  $\Sigma$  et surfacique critique  $\Sigma_{crit}$  de masse par :

$$2\kappa(\vec{\theta}) = \nabla^2 \psi(\vec{\theta}) = \frac{\Sigma}{\Sigma_{crit}} . \quad (1.23)$$

Afin de détecter l'effet de lentille faible, on mesure la corrélation entre la forme des galaxies dont l'ellipticité observée,  $\epsilon_{obs}$ , est reliée à leur ellipticité intrinsèque,  $\epsilon_s$ ,

et au cisaillement gravitationnel que l'on souhaite extraire,  $\gamma$ .

$$\epsilon_{obs} = \epsilon_s + \gamma . \quad (1.24)$$

La fonction de corrélation angulaire à deux points observée,  $\langle \epsilon_{obs}^i(\vec{\theta}_1) \epsilon_{obs}^j(\vec{\theta}_2) \rangle = \xi(\theta)_{\gamma_i \gamma_j}$ , avec,  $\theta = |\vec{\theta}_1 - \vec{\theta}_2|$ , peut être reliée à  $\langle \gamma^i \gamma^j \rangle$ , via :

$$\langle \epsilon_{obs}^i \epsilon_{obs}^j \rangle = \langle \epsilon_s^i \epsilon_s^j \rangle + \langle \epsilon_s^i \gamma^j \rangle + \langle \gamma^i \epsilon_s^j \rangle + \langle \gamma^i \gamma^j \rangle , \quad (1.25)$$

où on considère généralement que les trois premiers termes sont négligeables.

Le spectre de puissance du cisaillement pour un nombre d'onde angulaire  $l$  est la transformée de Fourier de  $\xi(\theta)_{\gamma_i \gamma_j}$ . Cette quantité est identique au spectre de puissance de la convergence et peut s'exprimer comme une projection du spectre de puissance de la densité de matière,  $P_\delta$ . Pour des galaxies sources se trouvant dans les bins en redshift  $i$  et  $j$ , il s'écrit :

$$C_{\gamma_i \gamma_j}(\ell) = \int_0^\infty dz \frac{W_i(z) W_j(z)}{\chi^2(z) H(z)} P_\delta \left( \frac{\ell}{\chi(z)}, z \right) , \quad (1.26)$$

où les indices  $i$  et  $j$  couvrent tous les bins en redshifts. Cette équation montre de quelle façon le spectre de puissance du cisaillement observable est sensible, à la fois, aux facteurs géométriques contenus dans  $W_i(z)$  et  $W_j(z)$ , ainsi qu'à la croissance des structures contenues dans le spectre de puissance,  $P_\delta$ . Les deux sont sensibles à l'énergie sombre. Mesurer le spectre de puissance ainsi que son évolution possible à partir de la forme des galaxies permet donc d'en savoir plus sur les propriétés de l'énergie sombre.

Enfin, le terme géométrique  $W$  s'écrit :

$$W(\chi) = \frac{3}{2} \Omega_m^0 H_0^2 a^{-1}(\chi) \chi \int d\chi_s n_s(\chi_s) \frac{\chi_s - \chi}{\chi_s} , \quad (1.27)$$

avec :

- $d\chi = -dz/H(z)$ , l'élément de distance comobile avec  $c = 1$ .
- $H(z)$ , le paramètre de Hubble au redshift  $z$ .
- $n_s(\chi_s)$ , la fonction de sélection - la distribution - en redshift des galaxies sources.

Si toutes les galaxies sources sont au même redshift  $z_s$ , alors :  $n_s(\chi) = \delta_D(\chi - \chi_s)$ . Afin de pouvoir extraire l'information sur l'énergie sombre de la mesure du spectre de puissance angulaire du cisaillement, il est donc nécessaire de connaître la distance des galaxies sources. Évidemment, la précision finale sur les paramètres de l'énergie sombre dépendra en partie de la précision sur la mesure du redshift - de la distribution en redshift - que l'on utilisera dans l'équation (1.26).

On peut alors faire une estimation de la précision nécessaire sur le redshift afin d'obtenir une précision finale donnée sur les paramètres de l'énergie sombre. Il a été montré (Huterer et collab., 2006; Kitching et collab., 2009; Ma et collab., 2006) que le biais et la diffusion - les quantités  $\langle z_p - z_s \rangle$  et  $\langle (z_p - z_s)^2 \rangle / 2$  - dans chaque bin  $\delta z = 0.1$ , doivent être contrôlés, au moins, à  $\sim 0.003$  de façon à entraîner une dégradation de moins de  $\sim 50\%$  sur la précision des paramètres cosmologiques. Ce qui est énorme ! Cela illustre bien l'importance de la mesure du redshift dans l'exploration de l'énergie sombre puisque actuellement ces exigences ambitieuses ne sont pas totalement atteintes par les redshifts photométriques. Il est donc crucial d'améliorer les redshifts photométriques et/ou de développer de nouvelles méthodes de mesure capable d'atteindre ces objectifs seules ou lorsqu'elles sont utilisées conjointement aux redshifts photométriques.

Les redshifts spectroscopiques permettraient d'atteindre ces exigences en termes de précision mais il n'est tout simplement pas réaliste d'envisager d'obtenir le spectre complet de chacune des 2 milliards de galaxies que détectera un télescope spatial comme *Euclid*. Cela ne peut être fait pour de multiples raisons comme : le temps d'observation nécessaire même en utilisant des MOS, le coût que de telles observations représenteraient ainsi que la difficulté d'obtenir les spectres des objets très peu brillants. L'approche photométrique du problème est elle aussi limitée d'une part par la limite de précision de la méthode. Pour des redshifts photométriques en cinq bandes, comme ceux provenant du *CFHTLS* la précision moyenne est de :  $\sim 0.04$ . D'autre part, notre incompréhension des erreurs catastrophiques affectant ce type de mesure ne permet pas de corriger ces effets qui créent un biais important au sein de l'échantillon. Le challenge à relever par la communauté scientifique est donc de taille.

## 4 Synthèse

Dans ce chapitre, j'ai présenté différentes méthodes de mesure de la distance nous séparant des objets célestes. J'ai notamment introduit la notion de redshift qui est un indicateur privilégié de la distance à l'échelle cosmologique. Après avoir présenté les méthodes couramment utilisées pour l'estimation du redshift (spectroscopie, photométrie) j'ai présenté la mission spatiale *Euclid* dont l'objectif scientifique principal est de percer les mystères de l'énergie sombre en déterminant ses propriétés et sa na-

ture. La précision et le contrôle des mesures de redshifts pour un très grand nombre d'objets peu brillants sont des enjeux majeurs pour le succès de cette mission.

Durant ces trois années, je me suis intéressé à cette problématique concernant la mesure de redshifts, en abordant le problème sous un tout nouvel angle. Là où les précédentes mesures de redshifts étaient toutes fondées sur l'information contenue dans la lumière émise par les objets - spectre, couleurs - l'approche développée dans ce manuscrit se base, elle, sur l'agrégation sur le ciel entre des galaxies de redshift *inconnu* et un échantillon de *référence*. Comme nous le verrons dans les prochains chapitres, cette rupture totale avec les méthodes précédentes permet de s'affranchir des systématiques propres à l'étude de la distribution spectrale d'énergie (SED) des sources.

---

## Introducing the method : abstract

The main goal of this chapter is to introduce the key concepts and the mathematical tools needed to carry out a clustering redshift analysis. First, I remind the reader that the main goal behind photometric redshift as well as clustering redshift techniques is to reproduce the mapping between the photometric space - including : colors, positions, ellipticity... - and the redshift space. But instead of using colors and spectral energy distributions (SEDs) templates, the clustering redshift procedure is based on the spatial cross-correlation between a galaxy sample of *unknown* redshift distribution and a *reference* set. The fundamental quantity in such a measurement is then the variation of the number of *unknown* objects around those of *reference*. Then, I present the correlation function and I introduce the *cluster-z* formalism as well as a new quantity named the integrated cross-correlation function,  $\bar{\omega}_{ur}(z)$ . This quantity is defined as :

$$\bar{\omega}_{ur}(z) = \int d\theta W(\theta) \omega_{ur}(\theta, z) , \quad (1.28)$$

where :

- $W(\theta)$ , is a weighted function aimed at optimizing the overall signal-to-noise ratio, following a matched filter procedure.
- $\omega_{ur}(\theta)$ , is the angular cross-correlation between the *unknown* and *reference* samples.

I show that considering a narrow redshift distribution for the *unknown* sample  $dN_u/dz$ , it can be written as :

$$dN_u/dz(z) \propto \bar{\omega}_{ur}(z) \left( \frac{1}{\bar{b}_r(z) \bar{\omega}(z)} \right) , \quad (1.29)$$

with :

- $\bar{b}_r(z)$ , the galaxy-matter bias of the *reference* sample.
- $\bar{\omega}(z)$ , the integrated matter correlation function.

Then I build an estimator to directly measure  $\bar{\omega}_{ur}(z)$ , given by :

$$\bar{\omega}_{ur} = \frac{\langle \bar{\Sigma}_{ur} \rangle_z}{\Sigma_R} - 1 , \quad (1.30)$$

where :

- $\bar{\Sigma}_{ur}$ , is the local surface density of *unknown* objects around the one of *reference*.
- $\Sigma_R$ , is the equivalent random surface density of the *unknown* sample.

Finally, I show how to compute the integrated matter correlation function from the matter power spectrum.

---



# Chapitre 2

## Présentation de la méthode

---

Ce chapitre présente la méthode de mesure du redshift développée durant cette thèse et qui repose sur les propriétés d'agrégation (*clustering* en anglais) des galaxies. On parle alors de *clustering redshifts* ou encore de *cluster-z*. Cette approche se distingue des méthodes généralement employées pour obtenir le redshift des galaxies par deux aspects. D'une part elle permet de mesurer la *distribution* en redshift d'un ensemble d'objets. Donc, pour des objets de même type et possédant des propriétés similaires cette distribution peut également être vue comme la PDF( $z$ ) de chaque objet. D'autre part elle diffère des redshifts photométriques par le fait qu'elle ne repose pas directement sur les propriétés de la lumière émise par les galaxies mais sur leurs propriétés statistique d'agrégation mutuelle qui décrit leur distribution spatiale particulière résultant de l'histoire de leur interaction gravitationnelle. La méthode du *cluster-z*, bien qu'absente de la majorité des études actuelles, est connue depuis plusieurs décennies et repose fondamentalement sur la variation du nombre de galaxies  $\mathcal{N}_{gal}$  autour d'objets de *référence* situés à différents redshifts.

[Seldner et Peebles \(1979\)](#) présentèrent pour la première fois cette approche pour étudier la distribution en redshift des quasars en mesurant leur corrélation angulaire en fonction du redshift avec un échantillon de galaxies de *référence* de redshifts spectroscopiques connus. Quelques années plus tard [Phillipps et Shanks \(1987\)](#), comptèrent les galaxies peu brillantes autour de sources de *référence* brillantes dont les redshifts étaient connus pour obtenir une estimation de la fonction de luminosité des galaxies à magnitude plus faible. Dix ans plus tard, [Landy et collab. \(1996\)](#) montrèrent qu'il est possible de combiner des mesures d'auto-corrélation et de corrélation croisée entre deux populations de galaxies dont l'une de distribution en redshift connue afin de déterminer la fraction des objets de l'échantillon inconnu qui possède un recouvrement en redshift avec le second.

Plus récemment, [Newman \(2008\)](#); [Matthews et Newman \(2010\)](#); [Matthews et Newman \(2012\)](#) présentèrent et testèrent en simulation une méthode itérative afin de mesurer la distribution en redshift d'une population *inconnue* à l'aide d'un échan-

tillon de *référence*. Plus récemment, [McQuinn et White \(2013\)](#) ont proposé un estimateur optimal pour une telle mesure.

La démarche présentée dans l'ensemble de ces travaux peut être qualifiée de *globale* dans le sens où aucune hypothèse n'est faite sur la distribution en redshift de la population *inconnue*. À l'inverse, les travaux développés dans cette thèse se basent sur une approche *locale*. En effet, nous nous intéresserons au cas où la population *inconnue* possède une distribution étroite en redshift. Ce point sera plus largement présenté dans la section 2.

Dans ce chapitre, je commence par présenter les concepts généraux sur lesquels repose l'estimation de la distribution en redshift par corrélation croisée. Puis j'aborderai le formalisme mathématique qui fonde la mesure de *cluster-z*. Enfin, je discuterai comment calculer la fonction de corrélation de la matière (noire) qui apparaît dans les équations.

## 1 Concepts

Comme nous l'avons vu dans le chapitre précédent, les redshifts photométriques tentent de caractériser la relation existante entre l'espace photométrique - couleurs, position, ellipticité... - et l'espace des redshifts. Pour cela les *photo-z* utilisent d'une part les mesures de la luminosité et des rapports de flux entre différentes bandes spectrales de chaque objet et d'autre part des modèles de distributions spectrales d'énergies (SED). Les clustering redshifts caractérisent aussi la relation entre l'espace photométrique et l'espace des redshifts. Cependant, le moyen mis en oeuvre est, lui, radicalement différent. Les *cluster-z* utilisent le principe que des galaxies qui se trouvent angulairement proches l'une de l'autre sur la voûte céleste ont de grandes chances d'être au même redshift par regroupement progressif sous l'effet de la gravitation. Les *cluster-z* utilisent donc la corrélation spatiale d'un échantillon *inconnu* avec une population de *référence* situés dans une même région du ciel, Figure 2.1.

Dans une région donnée du ciel et à un redshift donné, la distribution de matière et ses propriétés d'agrégation sont spécifiques et caractéristiques du redshift considéré. On peut donc mesurer la corrélation croisée entre un échantillon de *référence* - situé à un redshift donné - et un échantillon *inconnu* afin d'étudier son redshift. L'étude de cette corrélation peut se faire via la fonction de corrélation à deux points. Cette fonction peut être spatiale - à trois dimensions - notée  $\xi(r)$ , ou angulaire - à deux dimensions - notée  $\omega(\theta)$ .

La fonction de corrélation spatiale à deux points  $\xi(r)$  représente l'excès de probabilité, par rapport à une distribution de Poisson, de trouver un objet A à une certaine distance  $r$  d'un objet B.

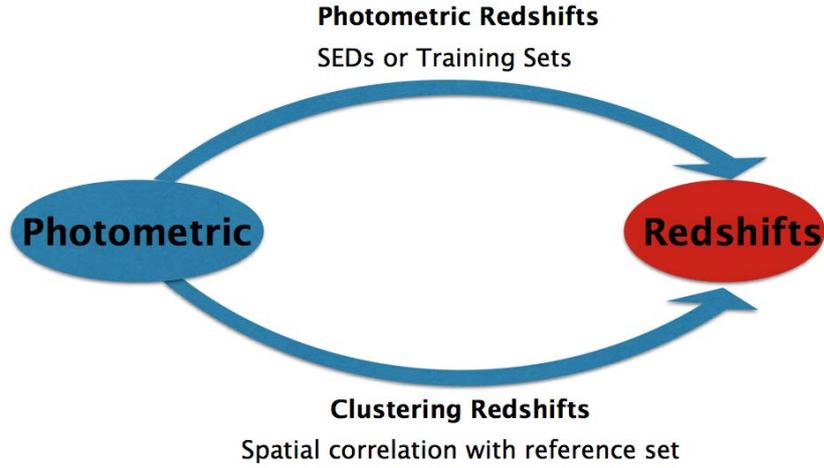


FIGURE 2.1 – Schéma présentant la différence entre les clustering redshifts et les redshifts photométriques.

Si les galaxies sont distribuées selon la statistique de Poisson avec une densité volumique moyenne  $n_0$ , la probabilité  $dN(r)$  de trouver une galaxie A dans le volume  $dV_A$  à une distance  $r$  de la galaxie B dans le volume  $dV_B$  est :

$$dN(r) = n_0^2 dV_A dV_B . \quad (2.1)$$

Par distribution poissonnienne on entend un processus aléatoire stationnaire - dont les propriétés moyennes sont identiques quel que soit l'endroit et la taille de l'échantillon considéré.

Si la distribution des galaxies s'écarte d'une distribution poissonnienne, cette probabilité sera :

$$dN(r) = n_0^2 (1 + \xi(r)) dV_A dV_B , \quad (2.2)$$

où  $\xi(r)$  est la fonction de corrélation à deux points des galaxies.  $\xi$  exprime donc à l'écart par rapport à une distribution poissonnienne. Dans le cas des galaxies, cet écart résulte principalement de leur agrégation - *clustering* - sous l'effet de la gravité.

Si les galaxies sont distribuées selon une distribution de Poisson alors, par définition, la fonction de corrélation est nulle. Elle peut être positive ou négative selon que les objets sont respectivement plus ou moins proches les uns des autres relativement à une distribution Poissonnienne.

Afin de mesurer cette quantité on peut utiliser l'estimateur suivant, (Davis et Peebles, 1983a) :

$$\hat{\xi}_{ij}^{DP}(r) = \frac{\langle D_i D_j \rangle_r}{\langle D_i R_j \rangle_r} - 1, \quad (2.3)$$

avec :

- $\{i, j\}$ , des jeux de données.
- $\xi_{ij}$ , la fonction de corrélation. On parle d'auto-corrélation lorsque  $i = j$ , et de corrélation croisée lorsque  $i \neq j$ .
- $D_i D_j$ , le nombre de paires d'objets distants de  $r$  entre les jeux  $i$  et  $j$ .
- $D_i R_j$ , le nombre de paires d'objets entre le jeu de données  $i$  et une distribution poissonnienne de même géométrie que le jeu  $j$ . Par géométrie on entend même surface et même forme du relevé sur le ciel incluant la prise en compte d'éventuelles régions masquées.

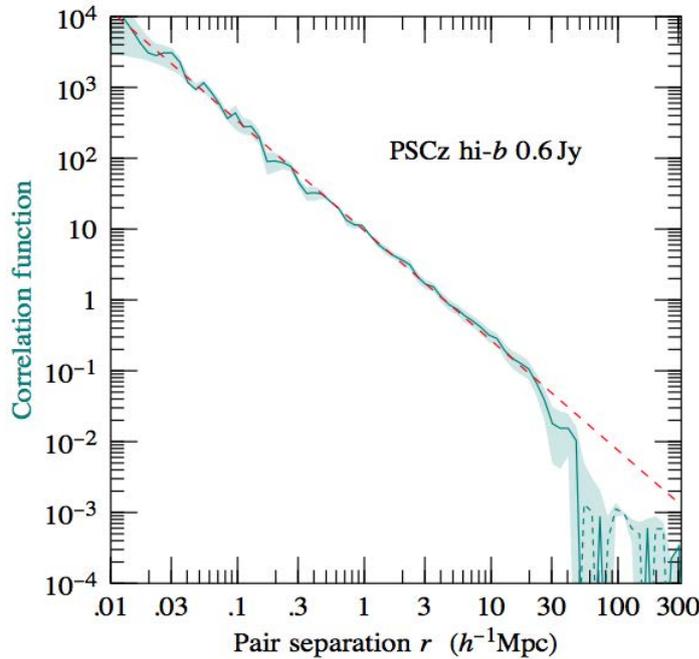


FIGURE 2.2 – Fonction de corrélation tridimensionnelle  $\xi(r)$  de PSCz extrait de [Hamilton et Tegmark \(2002\)](#) obtenue par transformé de Fourier du spectre de puissance à partir des galaxies du catalogue *PSCz* d'*IRAS*. Les pointillés cyan correspondent aux valeurs négatives pour des séparations de paires entre 50 – 200  $\text{Mpc} \cdot \text{h}^{-1}$ . La région ombragée correspond à l'enveloppe définie par la transformée de Fourier du spectre de puissance pour ses  $\pm 1\sigma$  extrema. Les pointillés rouges correspondent à une loi de puissance  $(r/4.27 \text{ Mpc} \cdot \text{h}^{-1})^{-1.55}$ .

La mesure de la fonction de corrélation - 3D et 2D - est l'un des estimateurs statistiques les plus simples et immédiats permettant de caractériser la distribution des galaxies et de conduire l'analyse et l'interprétation des grands catalogues en cosmologie ([Davis et Peebles, 1983b](#)). La Figure 2.2, montre la fonction de corrélation spatiale,  $\xi(r)$  qui est en accord avec une loi de puissance.

De la même façon que pour  $\xi(r)$  on peut définir  $\omega(\theta)$ , la fonction de corrélation 2D. Cette fonction, dont des exemples sont montrés sur la Figure 2.3, représente l'excès de probabilité de trouver un objet A à une distance angulaire  $\theta$  d'un objet B sur le ciel. Cette figure illustre l'accord de la pente de  $\omega(\theta)$  avec une loi de puissance et l'augmentation de son amplitude pour les objets brillants.

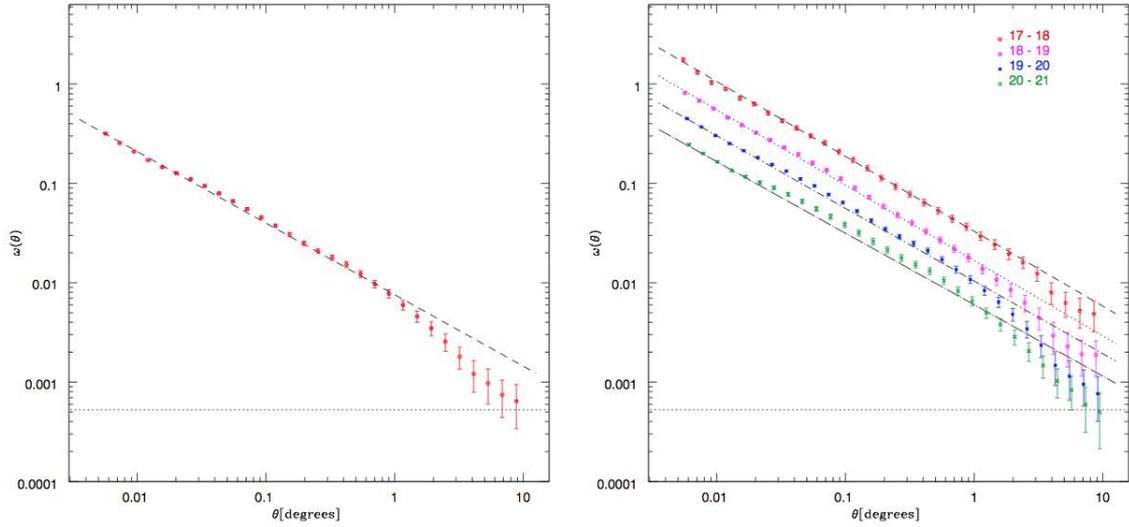


FIGURE 2.3 – Gauche : la fonction de corrélation angulaire des galaxies du *SDSS* pour des magnitudes comprises entre :  $17 < r \leq 21$ . La mesure est en accord avec une loi de puissance de la forme  $\omega(\theta) = A_\omega \theta^{(1-\gamma)}$  et  $\gamma \simeq 1.72$  sur la gamme  $0.005 - 10$  deg.

Droite : la fonction de corrélation angulaire triée par bin de magnitude. Cette figure montre l'augmentation du clustering (l'amplitude de la fonction de corrélation) pour des objets brillants en accord avec le fait que les objets brillants sont plus massifs.

Figure extraite de Wang et collab. (2013).

La fonction de corrélation angulaire est reliée à la probabilité  $dP_{2D}$  d'observer sur la surface du ciel deux galaxies chacune dans un angle solide  $d\Omega_i$ ,  $i \in \{A, B\}$  séparées de l'angle  $\theta$ , par :

$$dP_{2D} = \Sigma^2 d\Omega_A d\Omega_B (1 + \omega(\theta)) , \quad (2.4)$$

où,  $\Sigma$  est la densité de surface des galaxies.

L'estimateur de Davis et Peebles (1983a) pour la fonction de corrélation angulaire s'écrit alors :

$$\hat{\omega}_{ij}^{DP}(\theta) = \frac{\langle D_i D_j \rangle_\theta}{\langle D_i R_j \rangle_\theta} - 1, \quad (2.5)$$

avec :

- $\{i, j\}$ , des jeux de données.
- $\omega_{ij}$ , la fonction de corrélation.
- $D_i D_j$ , le nombre de paires d'objets distants de  $\theta$  sur le ciel.
- $D_i R_j$ , le nombre de paires d'objets entre le jeu  $i$  et une distribution Poissonnienne de même géométrie que le jeu  $j$ .

La fonction de corrélation angulaire,  $\omega(\theta)$ , n'est autre que la projection sur le ciel de  $\xi(r)$  et présente l'avantage de ne pas nécessiter la connaissance de la distance physique nous séparant des objets étudiés. Si on néglige une possible évolution temporelle du champs de densité de galaxies, les fonctions de corrélation 2D et 3D sont reliées entre elles par l'équation (2.6), (Peebles, 1980) :

$$\omega(\theta) = \int dr_1 \int dr_2 \phi_1(r_1) \phi_2(r_2) \xi(R), \quad (2.6)$$

avec :

- $R = |\vec{r}_1 - \vec{r}_2| = \sqrt{r_1^2 + r_2^2 - 2r_1 r_2 \cos \theta}$ .
- $r_i$   $i \in \{1, 2\}$ , la distance de l'objet  $i$ .
- $\phi_i$ , la fonction de sélection de  $i$ . Elle exprime le fait que les galaxies contenues dans un échantillon incluent des effets observationnels ou instrumentaux qui affectent les propriétés du catalogue (limite en flux, homogénéité sur la voute céleste, discrimination en type morphologique, en type spectral ou en redshift, etc...).

Intéressons nous maintenant à ce que peut nous apprendre la fonction de corrélation. Les Figures 2.2 & 2.3 montrent que  $\xi(r)$  et  $\omega(\theta)$  sont positives et décroissantes. Cela signifie que la distribution des galaxies ne suit pas une distribution de Poisson dans le ciel. Cette distribution particulière se trouve principalement aux petites échelles, là où l'amplitude de la fonction de corrélation est la plus importante. Cela est le résultat de l'évolution au cours du temps des fluctuations primordiales de matière issues de l'inflation. Le spectre de puissance décrivant l'évolution de ces fluctuations est présenté plus en détails dans la Section 3. Plus on s'éloigne d'une galaxie donnée plus la distribution des galaxies se rapproche d'une distribution de Poisson. Autrement dit, quiconque chercherait une galaxie dans l'immensité de l'espace devrait commencer ses recherches en regardant juste à coté d'une autre galaxie. C'est exactement ce point qui motive l'exploitation du redshift clustering.

Pour plus d'informations sur les fonctions de corrélation voir Annexe A et pour une revue complète sur la structure à grande échelle de l'Univers voir [Bernardeau et collab. \(2002\)](#).

## 2 Formalisme

Après une première approche plus qualitative des *clustering redshifts*, intéressons-nous maintenant plus en détails au formalisme mathématique. Comme nous l'avons vu précédemment nous aurons besoin de deux échantillons de galaxies :

- un échantillon de *référence*, dont les redshifts spectroscopiques sont connus et qui sera désigné par un indice  $r$  dans le reste du manuscrit.
- un échantillon dont le redshift est *inconnu* qui sera désigné par un indice  $u$ , ( $u = unknown$ ).

La densité de surface des objets de l'échantillon *inconnu* situés à une distance  $\theta$  sur le ciel des objets de l'échantillon de *référence* se trouvant au redshift  $z$  est donnée par :

$$\langle \Sigma_{ur}(\theta, z) \rangle = \Sigma_R [1 + \omega_{ur}(\theta, z)] , \quad (2.7)$$

où :

- $\Sigma_R$ , est la densité de surface d'un champ Poissonien de même géométrie que la population *inconnue*.
- $\omega_{ur}(\theta, z)$ , est la fonction de corrélation croisée des objets de l'échantillon *inconnu* situés à une distance  $\theta$  sur le ciel des objets de *référence* se trouvant au redshift  $z$ .

On définit une nouvelle quantité appelée la fonction de corrélation croisée *intégrée*, notée  $\bar{\omega}_{ur}(z)$ , où la barre désigne la somme sur  $\theta$ . Cette quantité permet de quantifier la valeur du *clustering* dans une coquille en redshift. On pose :

$$\bar{\omega}_{ur}(z) = \int d\theta W(\theta) \omega_{ur}(\theta, z) , \quad (2.8)$$

où :

- $W(\theta)$ , est un filtre adapté afin de maximiser l'extraction du signal. Cette technique, connue sous le nom de *matched filter* en analyse du signal consiste à appliquer un poids dont la forme est similaire au signal que l'on souhaite extraire, c'est-à-dire dans notre cas  $W(\theta) \propto \theta^{-0.8}$  et  $\int W(\theta) d\theta = 1$ . Cette pondération particulière - correspondant à la pente de la fonction de corrélation angulaire - nous permet d'attribuer une plus grande importance aux objets se trouvant à faible distance, et cela, *de la même façon que la fonction de corrélation*.
- $\omega_{ur}(\theta)$ , est la fonction de corrélation croisée angulaire entre les échantillons *inconnu* et de *référence*.

La fonction de corrélation croisée angulaire peut également s'écrire :

$$\omega_{ur}(\theta) = \int dz' \phi_u(\theta, z') \phi_r(\theta, z') b_u(\theta, z') b_r(\theta, z') \omega(\theta, z') , \quad (2.9)$$

avec :

- $\phi_i(\theta, z)$ , la fonction de sélection de l'échantillon  $i \in \{u, r\}$ .
- $b_i(\theta, z)$ , le biais galaxie-matière de l'échantillon  $i \in \{u, r\}$ .
- $\omega(\theta, z)$ , la fonction de corrélation des halos de matière noire.

Dans cette expression on peut remarquer le produit  $\phi_u(\theta, z') \times \phi_r(\theta, z')$ , qui traduit le fait que les populations *inconnue* et de *référence* doivent se trouver dans la même région du ciel. Si ce n'est pas le cas la corrélation entre les deux populations sera nulle quel que soit le recouvrement en redshift. Ainsi, dans une même région du ciel, seul un recouvrement des deux échantillons dans la profondeur - c'est-à-dire en redshift - donnera une corrélation non nulle. L'intégration de cette expression selon l'équation 2.8 se fait en considérant que les fonctions de sélection ainsi que les biais galaxie-matière des deux populations ne dépendent pas de  $\theta$ .

On obtient alors l'expression suivante pour la fonction de corrélation intégrée :

$$\bar{\omega}_{ur} = \int dz' \frac{dN_u}{dz'}(z') \frac{dN_r}{dz'}(z') \bar{b}_u(z') \bar{b}_r(z') \bar{\omega}(z') , \quad (2.10)$$

où les fonctions de sélection des échantillons sont identifiées avec les distributions en redshift.

Si on considère un échantillonnage fin de la distribution en redshift spectroscopique alors  $dN_r/dz' = N_r \delta_D(z' - z)$ , et on peut simplement inverser l'équation précédente pour obtenir :

$$\frac{dN_u}{dz}(z) \propto \bar{\omega}_{ur}(z) \times \frac{1}{\bar{b}_u(z)} \times \frac{1}{\bar{b}_r(z) \bar{\omega}(z)} , \quad (2.11)$$

où :

- $\bar{\omega}_{ur}(z)$ , est la fonction de corrélation croisée intégrée. Cette quantité peut être mesurée sur les données.
- $\bar{b}_r(z)$ , est le biais galaxie-matière de l'échantillon de *référence*. Il peut être mesuré sur l'échantillon de *référence* via  $\bar{\omega}_{rr}(z) = \bar{b}_r^2(z) \bar{\omega}(z)$ .
- $\bar{\omega}(z)$ , est la fonction de corrélation intégrée de la matière. Elle peut être calculée à partir du spectre de puissance de la matière, voir Section 3.
- $dN_u/dz(z)$ , est la distribution en redshift de l'échantillon *inconnu* que l'on cherche.
- $\bar{b}_u(z)$ , est le biais galaxie-matière de l'échantillon *inconnu*, la seule quantité inconnue.

Le problème qui se pose ici concerne notre manque de connaissance de l'évolution, avec le redshift, du biais galaxie-matière de l'échantillon *inconnu*. En effet contrairement à l'échantillon de *référence*, il n'est pas possible de mesurer  $\bar{b}_u(z)$ . Pour cela il faudrait posséder l'information sur le redshift de la population *inconnue*... Ce qui est précisément ce que nous cherchons à mesurer !

C'est à cette étape qu'intervient la notion d'approche *locale* mentionnée au début du chapitre. Si l'on ne fait aucune hypothèse sur la largeur de la distribution en redshift de la population *inconnue* - approche *globale* - alors la quantité mesurée est  $\bar{b}_u(z) \times dN_u/dz(z)$ . Afin de surmonter cette difficulté, considérons une distribution de Dirac pour la population *inconnue*. Dans ce cas  $\bar{b}_u(z) = cste$ , et on peut écrire :

$$dN_u/dz(z) \propto \bar{\omega}_{ur}(z) \left( \frac{1}{\bar{b}_r(z) \bar{\omega}(z)} \right). \quad (2.12)$$

On peut alors obtenir la distribution en redshift de l'échantillon *inconnu* via la normalisation suivante :

$$\int dz \frac{dN_u}{dz} = N_u. \quad (2.13)$$

Dans le cas limite où la distribution en redshift de l'échantillon *inconnu* est une fonction de Dirac, cette approche *locale* permet de retrouver la solution exacte. En pratique, l'échantillon *inconnu* possède une largeur  $\Delta z$ . Dans le cas où  $\Delta z$  est petit on se rapproche du cas précédent. Plus  $\Delta z$  sera petit - approche *locale* - plus la variation relative de  $dN_u/dz$  sera grande et plus la variation relative des autres quantités sera petite. Autrement dit, pour  $\Delta z$  petit, on a selon [Ménard et collab. \(2013\)](#) :

$$\frac{d \log dN_u/dz}{dz} \gg \frac{d \log \bar{b}_u}{dz}, \quad (2.14)$$

Il est crucial de pouvoir présélectionner des populations d'objets *inconnus* localisés en redshift. Pour cela on pourra par exemple utiliser le redshift photométriques des objets ou encore faire des sélections dans l'espace des couleurs. Ces approches que j'ai toutes les deux explorées, seront développées dans les prochains chapitres.

Enfin, afin d'estimer directement la fonction de corrélation croisée intégré  $\bar{\omega}_{ur}(z)$  on intègre l'expression de la fonction de corrélation croisée angulaire présenté à l'équation (2.7). En considérant que  $\Sigma_R$  ne dépend pas de  $\theta$ , on obtient l'estimateur suivant :

$$\bar{\omega}_{ur}(z) = \frac{\langle \bar{\Sigma}_{ur} \rangle_z}{\Sigma_R} - 1, \quad (2.15)$$

avec :

- $\bar{\Sigma}_{ur}(z)$ , la densité locale d'objets *inconnus* autour d'un objet de *référence* se trouvant au redshift  $z$ , où la barre désigne la somme sur  $\theta$  en accord avec l'équation (2.8). Afin de maximiser l'extraction du signal lors de la mesure de la densité de surface du champ de galaxies *inconnues* autour des objets de *référence*, on utilise le même filtre adapté que précédemment  $W(\theta) \propto \theta^{-0.8}$ . Pour un objet de *référence* donné possédant  $n$  proches voisins *inconnus*, distants de  $\theta$ , répartis sur une surface  $\mathcal{A}$ , la mesure de  $\bar{\Sigma}_{ur}$  se fait selon :

$$\bar{\Sigma}_{ur} = \frac{\sum_{i=1}^n W(\theta_i)}{\mathcal{A}}, \quad (2.16)$$

- $\Sigma_R$ , la densité d'un champ Poissonien - mais de même géométrie que  $u$  - - correspond simplement au nombre de galaxies  $\mathcal{N}_{gal}$  réparties uniformément sur la surface non masquée  $\mathcal{A}_{tot}$ , soit :

$$\Sigma_R = \frac{\mathcal{N}_{gal}}{\mathcal{A}_{tot}}. \quad (2.17)$$

L'application directe de l'équation (2.17) pour le calcul de  $\Sigma_R$  nécessite une connaissance précise des régions masquées c'est-à-dire, la prise en compte de leurs surfaces ainsi que de leurs formes. Cela permet de s'affranchir de l'étape de génération de catalogues aléatoires. Techniquement, cela revient à créer une carte du ciel de la région observée dont on retire les régions masquées. Les galaxies sont ensuite placées sur la carte où sont réalisées toutes les mesures de densité. En plus d'être plus simple, cette approche pour le calcul de  $\Sigma_R$  présente l'avantage d'être plus rapide. De cette façon, il n'est pas nécessaire de générer des millions d'objets par ordinateur - de façon à pouvoir négliger le bruit de Poisson - puis de mesurer la densité du champ résultant. Cette dernière étape peut s'avérer assez longue puisque le temps de calcul d'une telle mesure évolue, au mieux, en  $N \log(N)$ , où  $N$  est le nombre d'objet.

Cet estimateur - équation (2.15) - possède les mêmes propriétés que sa version *angulaire* présentée dans l'équation (A.9). Cela signifie que son incertitude à grande échelle est, elle aussi, limitée par l'incertitude sur la densité moyenne. Dans notre cas, on peut négliger cet effet puisque nous nous intéressons justement aux petites échelles, de l'ordre de quelques Megaparsecs. L'erreur sur  $\bar{\omega}_{ur}(z)$  peut être estimée via la procédure standard de propagation des erreurs pour une distribution de Poisson, que l'on rappelle brièvement.

L'erreur sur la mesure d'un nombre  $N$  d'objets est :

$$\sigma_N = \sqrt{N} . \quad (2.18)$$

Ce qui donne pour la densité de surface,  $\Sigma$  :

$$\sigma_\Sigma^2 = \left( \frac{\Sigma}{\sqrt{N}} \right)^2 . \quad (2.19)$$

Considérons l'estimateur de [Davis et Peebles \(1983a\)](#) pour la fonction de corrélation croisée angulaire :  $\omega_{12} = \frac{D_1 D_2}{D_1 R_2} - 1$ . La procédure standard de propagation des erreurs donne :

$$\sigma_{\omega_{12}}^2 = \left( \frac{\partial \omega_{12}}{\partial D_1 D_2} \right)^2 \sigma_{D_1 D_2}^2 + \left( \frac{\partial \omega_{12}}{\partial D_1 R_2} \right)^2 \sigma_{D_1 R_2}^2 , \quad (2.20)$$

et donc :

$$\sigma_{\omega_{12}}^2 = \left( \frac{\omega_{12} + 1}{\sqrt{D_1 D_2}} \right)^2 + \left( \frac{\omega_{12} + 1}{\sqrt{D_1 R_2}} \right)^2 . \quad (2.21)$$

C'est de cette façon que sont obtenues les barres d'erreurs relatives à la mesure de la fonction de corrélation croisée intégrée dans toutes les études que j'ai réalisées et qui sont présentées dans ce manuscrit

Afin de pouvoir estimer  $dN_u/dz$  selon l'équation (2.12), il ne reste plus qu'à être en mesure de calculer la fonction de corrélation intégrée de la matière  $\bar{\omega}(z)$ . Le détail de la procédure que j'ai suivie ainsi que les tests des codes que j'ai développés sont présentés dans la section suivante.

### 3 Calcul de $\bar{\omega}(z)$

La fonction de corrélation intégrée de la matière,  $\bar{\omega}(z)$ , est obtenue par application de l'équation (2.8) sur la fonction de corrélation angulaire de la matière. Afin d'obtenir cette quantité, il faut tout d'abord générer le spectre de puissance de la matière noire,  $P(k, z)$ . Sa transformée de Fourier permettra d'obtenir la fonction de corrélation 3D,  $\xi(r, z)$  qui, une fois projetée sur le ciel donnera la fonction de corrélation angulaire,  $\omega(\theta, z)$ . Finalement l'intégration sur  $\theta$ , suivant l'équation(2.8) fournira la quantité désirée,  $\bar{\omega}(z)$ . Bien qu'il existe de nombreux codes permettant de générer le spectre de puissance, j'ai entrepris d'écrire mon propre programme pour différentes raisons comme : acquérir de nouvelles compétences en calcul numérique, comprendre le spectre de puissance ainsi que les différents facteurs le composant, pouvoir facilement intégrer le calcul de  $\bar{\omega}(z)$  dans l'estimation de  $dN_u/dz$ . Cette section présente en détails la démarche que j'ai suivie.

## Obtention du spectre de puissance non linéaire de la matière

Le spectre de puissance de la matière décrit le contraste de densité de l'Univers - le rapport entre la densité locale et la densité moyenne - en fonction de l'échelle physique considérée. Il s'agit de la transformée de Fourier de la fonction de corrélation de la matière.

Dans les premiers instants de l'Univers, la densité de rayonnement domine totalement la densité de matière. Durant cette époque le spectre de puissance primordial est une loi de puissance  $P_0(k) \propto k^n$ , où  $n \sim 1$  est l'indice spectral.

Avec l'expansion de l'Univers la densité de rayonnement diminue plus rapidement que celle de la matière jusqu'à atteindre l'égalité  $\Omega_r = \Omega_m (1 + z_{eq})$ , à  $z_{eq} \sim 3\,200$ . Dans un premier temps si on considère un spectre de puissance linéaire, alors les modes évoluent de manière indépendante les uns des autres. L'évolution des perturbations lors de la transition entre l'ère dominée par le rayonnement et l'ère dominée par la matière est alors décrite par la fonction de transfert  $T(k)$ , visible à la Figure 2.4. Elle permet de passer d'un spectre initial  $P_0(k)$  à un spectre tenant compte de l'évolution des perturbations aux différentes échelles :

$$P(k) = T(k)^2 P_0(k) . \quad (2.22)$$

La fonction de transfert prend en compte le fait que certaines échelles sont entrées dans la sphère de Hubble et d'autres pas, lors de la transition rayonnement-matière. L'action de la fonction de transfert se résume à considérer que l'amplitude des perturbations en dehors de la sphère de Hubble, lors de l'égalité rayonnement-matière, n'est pas affectée. On a donc  $T(k) = 1$ . Par contre l'amplitude des perturbations à l'intérieur de la sphère, lors de l'égalité rayonnement-matière, est diminuée proportionnellement à leurs entrées, c'est-à-dire  $T(k) < 1$ . Les petites échelles, entrées les premières, sont plus atténuées.

Afin de d'évaluer numériquement cette fonction, je me suis basé sur [Eisenstein et Hu \(1998\)](#) en utilisant les paramètres cosmologiques de *WMAP* ([Komatsu et collab., 2011](#)) suivant :  $\Omega_m = 0.27$  ;  $w_0 = -1$  ;  $w_a = 0$  ;  $h = 0.71$  ;  $\sigma_8 = 0.8$  ;  $\Omega_b = 0.045$ ,  $n = 0.96$  ; and  $\Omega_{DE} = 0.73$ . La Figure 2.4 compare la fonction de transfert que j'obtiens (en noir) à celle disponible sur le site de W. Hu (en rouge). L'accord entre les deux est de l'ordre du pourcent.

L'évolution des surdensités au cours du temps peut s'écrire comme la somme de deux solutions, une solution croissante et une solution décroissante. Dans notre cas on ne considère que la solution croissante puisque l'on s'intéresse au surdensité de matière donnant naissance aux galaxies. Si on suppose que cette solution peut s'écrire comme le produit d'un terme temporel et d'un terme spatial, on a alors :

$$\delta(\vec{x}, t) = D_+(t) \delta(\vec{x}) , \quad (2.23)$$

où le facteur de croissance  $D_+(t)$ , décrit l'évolution temporelle, alors que  $\delta(\vec{x})$  décrit l'évolution spatiale.

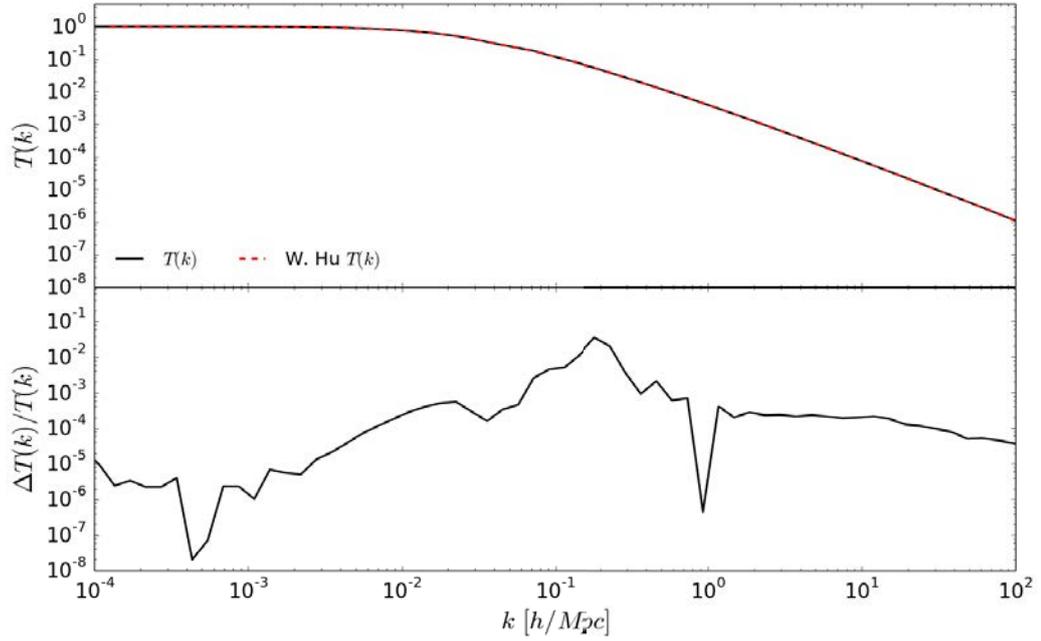


FIGURE 2.4 – Comparaison entre la fonction de transfert utilisée dans cette analyse et celle disponible sur le site web de W. Hu. L'accord entre les deux est, dans le pire des cas, de l'ordre du pourcent.

Le facteur de croissance est défini de façon à être normalisé aujourd'hui (Carroll et collab., 1992; Heath, 1977) ce qui donne :

$$D_+(a) = \frac{g(a)}{g(a_0)}, \quad (2.24)$$

et :

$$g(a) = \frac{5\Omega_M}{2a} \frac{da}{d\tau} \int_0^a \left( \frac{da'}{d\tau} \right)^{-3} da', \quad (2.25)$$

avec :

$$\left( \frac{da'}{d\tau} \right)^2 = 1 + \Omega_M \left( \frac{1}{a} - 1 \right) + \Omega_\Lambda (a^2 - 1), \quad (2.26)$$

où :  $\tau = H_0 t$ , est une variable temporelle sans dimension qui exprime le temps en unité de temps de Hubble - voir Carroll et collab. (1992) équation (9).

Finalement, le spectre de puissance linéaire est donné par :

$$P(k, a) = A k^n T^2(k) D_+^2(a), \quad (2.27)$$

avec :

- $a$ , le facteur d'échelle.

- $n$ , l'indice spectral  $\sim 1$ .
- $T(k)$ , la fonction de transfert (Eisenstein et Hu, 1998).
- $D_+(a)$ , le facteur de croissance (Carroll et collab., 1992; Heath, 1977).
- $A = \frac{\sigma_8^2}{\sigma^2(R=8 \text{ Mpc.h}^{-1})}$ , un facteur de normalisation où :

$$\sigma^2(R) = \int_0^\infty \frac{dk}{k} \frac{k^3}{2\pi^2} P(k) \left[ \frac{3j_1(kR)}{kR} \right]^2, \quad (2.28)$$

et :  $j_1(x) = (\sin x - x \cos x) / x^2$ , la fonction de Bessel du premier ordre en coordonnées sphériques.  $\sigma_8$  est, ici, la variance de l'amplitude des fluctuations de masse dans une sphère de  $8 \text{ Mpc.h}^{-1}$ .

La Figure 2.5 présente, pour deux redshifts, la comparaison entre le spectre de puissance linéaire que j'obtiens et celui obtenu avec le code *nicaea 2.4* (Kilbinger et collab., 2009).

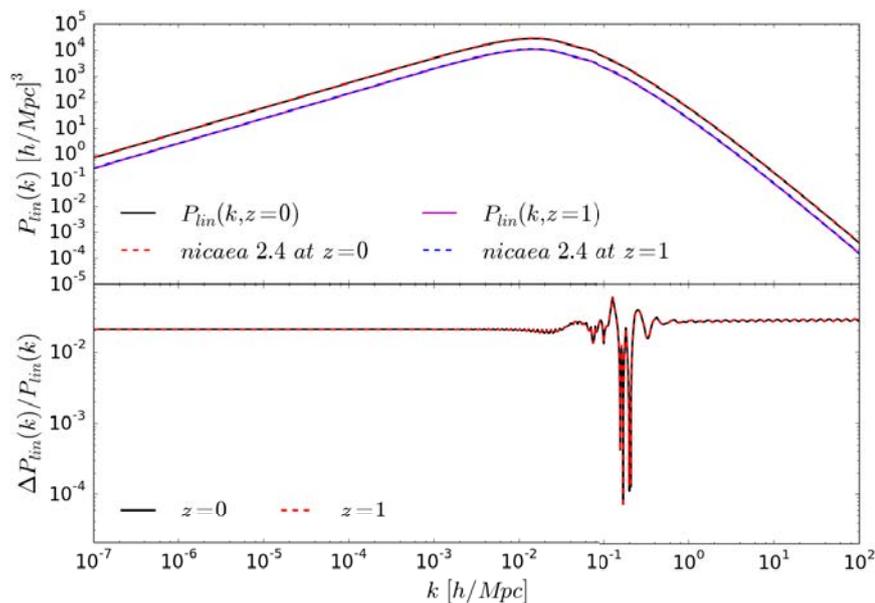


FIGURE 2.5 – Ce graphique illustre la comparaison entre le spectre de puissance linéaire utilisé dans cette analyse et celui obtenu avec le code *nicaea 2.4* (Kilbinger et collab., 2009) pour deux redshifts différents. À chaque fois l'accord est de l'ordre du pourcent.

Le calcul du spectre de puissance non linéaire est hautement non perturbatif. Pour cette raison de nombreuses relations paramétriques reliant le spectre linéaire au spectre non linéaire ont été développées, notamment : [Peacock et Dodds \(1996\)](#) ; [Smith et collab. \(2003\)](#) ; [McDonald et collab. \(2006\)](#) dont la précision est de l'ordre de 10 – 20%. Les paramètres d'ajustement de ces modèles sont basés sur des simulations à N-corps. Récemment [Takahashi et collab. \(2012\)](#), ont fourni une nouvelle estimation des paramètres pour le code halofit ([Smith et collab., 2003](#)) en se basant sur les toutes dernières simulations de ce type. Ces nouveaux paramètres augmentent la précision de ce modèle d'un facteur deux.

Le régime non linéaire ne concerne que les modes subissant l'effet non linéaire de la gravité aux petites échelles physiques, c'est-à-dire les grands  $k$  ( $[1-100] \text{ h.Mpc}^{-1}$ ). Aux échelles physiques plus grandes on retrouve la pente du spectre de puissance primordial ( $k < 0.01 \text{ h.Mpc}^{-1}$ ) ainsi que la brisure correspondant à l'égalité matière rayonnement ( $k = 0.01 \text{ h.Mpc}^{-1}$ ). Ces trois régimes sont visibles sur la Figure 2.6 qui compare le spectre de puissance non linéaire que j'obtiens avec celui issu de *nicaea 2.4*.

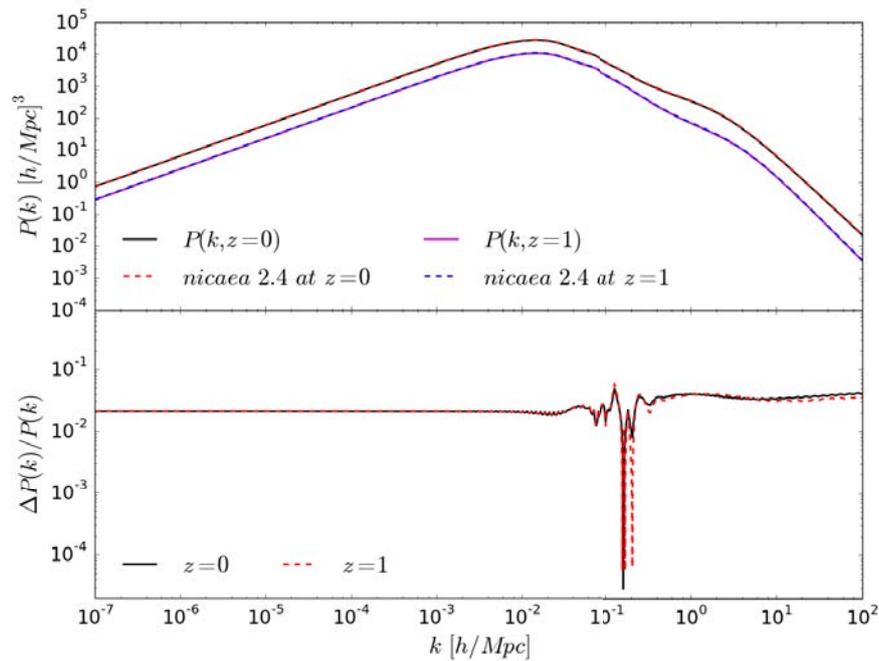


FIGURE 2.6 – Même type de comparaison que pour la Figure 2.5 avec cette fois le spectre de puissance non linéaire.

### Calcul de la fonction de corrélation 3D de la matière, $\xi(r, z)$

L'objectif est, maintenant, de calculer la fonction de corrélation 3D de la matière,  $\xi(r)$ . Pour cela on prend la transformée de Fourier du spectre de puissance suivant l'équation (2.29) :

$$\xi(r, z) = \int_0^\infty \Delta^2(k, z) j_0(kr) \frac{dk}{k}, \quad (2.29)$$

avec :

- $\Delta^2(k) = \frac{k^3 P(k)}{2\pi^2}$ , le spectre de puissance réduit.
- $j_0(k) = \frac{\sin(k)}{k}$ , la fonction de Bessel d'ordre 0 en coordonnées sphériques.

### Obtention de $\omega(\theta, z)$ et intégration sur $\theta$

Afin d'obtenir la fonction de corrélation angulaire,  $\omega(\theta, z)$ , on peut simplement projeter  $\xi(r, z)$  sur le ciel. En partant du spectre de puissance, cela revient à écrire selon [Dodelson \(2003\)](#) :

$$\omega(\theta, z) = \int_0^\infty dk k P(k, z) K(k\theta). \quad (2.30)$$

Le noyau,  $K(k\theta)$ , est donné par :

$$K(k\theta) = \frac{1}{2\pi} \int_0^\infty dz J_0(k\theta\chi(z)) \left( \frac{dn}{dz} \right)^2 \left( \frac{dz}{d\chi} \right) F(z), \quad (2.31)$$

avec :

- $\chi(z) = \frac{c}{H_0} \int_0^z \frac{dz'}{E(z')}$ , la distance comobile.
- $J_0$ , la fonction de Bessel d'ordre zero.
- $E(z) = \frac{dz}{d\chi} = \sqrt{\Omega_m(1+z)^3 + \Omega_K(1+z)^2 + \Omega_\Lambda}$ .
- $F(z)$ , est une fonction qui décrit l'évolution avec le redshift des densités de fluctuations au sein d'un bin en magnitude. En principe cette fonction est assez complexe et dépend également de l'échelle considérée. Cependant pour toutes les applications pratiques les données actuelles ne sont pas sensible à cette évolution, on prend donc :  $F(z) = 1$ . Voir [Dodelson et collab. \(2002\)](#) et [Maller et collab. \(2005\)](#) pour plus de détails.

Finalement on intègre sur  $\theta$  afin d'obtenir  $\bar{\omega}(z)$ . Afin de pouvoir mesurer la distribution en redshift de l'échantillon *inconnu* selon l'équation (2.12), il ne reste plus qu'à mesurer  $\bar{\omega}_{ur}(z)$  et estimer  $\bar{b}_r(z)$  ce qui sera présenté dans les chapitres suivants.

## 4 Synthèse

Après avoir présenté les concepts clefs de la mesure de la distribution en redshift par corrélation croisée, j'ai introduit le formalisme mathématique propre à ce type de mesure. J'ai alors mis en évidence l'importance de la capacité à sélectionner des populations d'objets localisés en redshift de façon à minimiser la variation du biais galaxie-matière de l'échantillon *inconnu* lors de l'obtention de l'équation (2.12). L'influence de la variation de  $\bar{b}_u(z)$  constitue une des limitations fondamentales de l'approche du redshift clustering présentée dans ce manuscrit. Dans les prochains chapitres, nous explorerons son influence et nous développerons différentes méthodes afin de réduire son effet. Enfin, j'ai présenté les différentes étapes nécessaires pour le calcul de la fonction de corrélation intégrée de la matière  $\bar{\omega}(z)$  à partir du spectre de puissance. Dans le prochain chapitre nous traiterons de la validation des *cluster-z* à la fois en simulation et sur des données.



## Validating the method : abstract

This chapter summarizes the main results from simulations studies. Key points are :

- there is more clustering information on scales smaller than 1 Mpc.
- the redshift recovery using linear clustering assumptions well into the nonlinear regime is only reliable for redshift distributions of narrow redshift width.

This suggests that the clustering based redshift estimation works better when inferring narrow distributions because it allows us to look at small scales where there is more information to be extracted. Then, I present in details a first application to real data using Luminous Red Galaxies (LRGs) as *unknown* sample and quasars (QSOs) for *reference* reproducing the analysis done in [Ménard et collab. \(2013\)](#). I demonstrate the validity of the *clustering redshift* approach by comparing the first two moments of the recovered distribution with the ones obtained through *photometric redshift*.

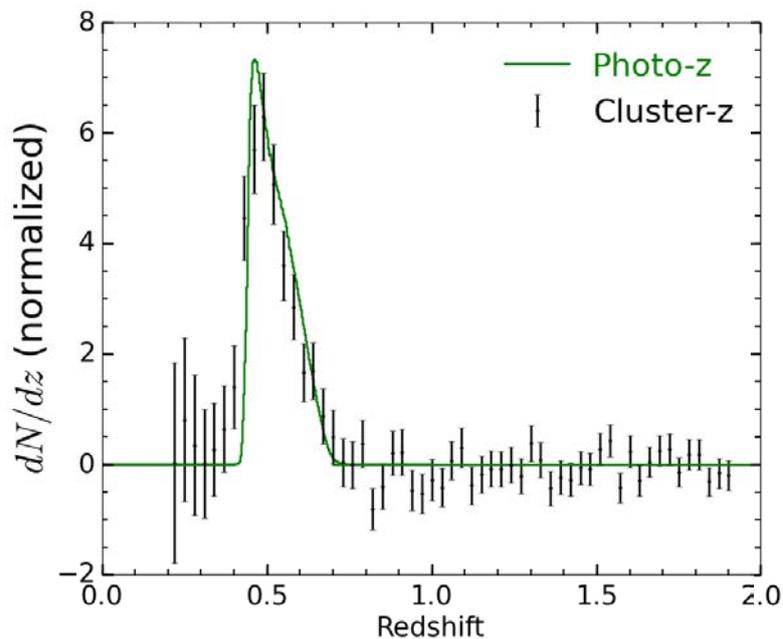


FIGURE 2.7 – Overplot between the redshift distribution of the *unknown* sample from *clustering redshift* and the one obtained using *photometric redshift*. Both distributions are in agreement.

Since *LRGs* and *QSOs* are known to be highly clustered, I generalize this approach to any type of objects by conducting a *cluster-z* analysis on an *unknown* data set with no selection on galaxy type. Finally, using the *spectroscopic* and *photometric redshifts* of the *unknown* sample I show that *clustering redshifts* are in better agreement with *spectro-z* than are *photo-z*, Figure 3.11.



# Chapitre 3

## Validation de la méthode

---

Ce chapitre présente l'ensemble des travaux que j'ai effectués pour implémenter et valider la méthode de mesure de la distribution en redshift décrite précédemment. Nous verrons tout d'abord le résultat des tests conduits avec des simulations numériques. Puis, je présenterai en détails ma première analyse réalisée sur des données pour reproduire des résultats de [Ménard et collab. \(2013\)](#). Elle utilise deux populations d'objets extragalactique, des Galaxies Rouges Lumineuses (LRGs) et des quasars (QSOs) issues d'observations du *SDSS*. Enfin, je montrerai que la mesure de *cluster-z* reste valide quelle que soit la nature des objets utilisés.

### 1 Evolution du biais et estimation du rapport signal sur bruit

#### 1.1 Tests en simulations et évolution du biais

Les premiers tests en simulations, réalisés par [Schmidt et collab. \(2013\)](#), explorent l'évolution de la méthode de reconstruction de la distribution en redshift lorsque l'on inclut les petites échelles physiques, c'est-à-dire quand les champs de densité de galaxies et de matière atteignent le régime non-linéaire. Le point principal de cette étude concerne la capacité à découpler l'amplitude de la distribution en redshift de l'influence de l'évolution du biais galaxie-matière. En se basant sur des simulations numériques explorant différents scénarii pour l'évolution du biais, ces auteurs montrent la fiabilité de la méthode de reconstruction de la distribution en redshift en considérant une évolution linéaire du clustering en régime non linéaire pour des distributions en redshift étroites, c'est-à-dire en ayant une approche *locale* en redshift.

Le fait que l'on puisse considérer une évolution linéaire du clustering en régime non linéaire lorsque l'on considère des distributions en redshift étroites est très intéressant. En effet pour une distribution d'objets non localisés en redshift et pour des

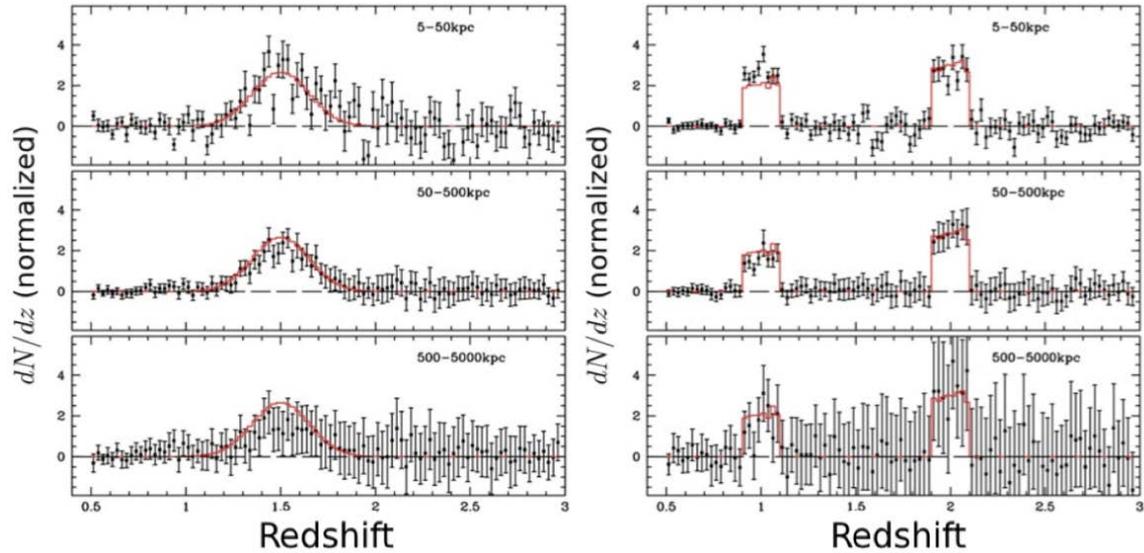


FIGURE 3.1 – Influence de l'échelle sur l'intensité du signal. Le signal est correctement retrouvé quelle que soit sa forme (gaussienne ou carrée) pour les échelles plus petites et se dégrade lorsque l'on considère des échelles physiques de plus en plus grandes. Figure extraite de [Schmidt et collab. \(2013\)](#).

petites séparations physiques - de l'ordre de quelques Mpc - le clustering des objets augmente et le biais devient non-linéaire ([Benson et collab., 2000](#)). L'augmentation du clustering (l'amplitude de la fonction de corrélation) nous indique qu'une quantité importante de l'information est contenue dans les petites échelles. Cette information était inexploitée par [Newman \(2008\)](#) dont l'étude s'intéressait aux échelles physiques allant de 2 à 10  $\text{Mpc.h}^{-1}$ . L'approche *locale* en redshift permet de contourner cette difficulté de façon à accéder à cette information tout en conservant une évolution linéaire du biais.

La diminution du clustering avec l'augmentation de la distance physique signifie qu'inclure de trop grandes échelles dans l'analyse peut dégrader le rapport signal sur bruit de la mesure,  $\mathcal{S}/\mathcal{N}$ . Il y a donc un compromis à trouver entre un fort rapport signal sur bruit et une faible évolution du biais. Ces effets sont illustrés sur la Figure 3.1.

L'échelle physique optimale dépend à la fois du clustering de l'échantillon *inconnu* et de sa densité. En montrant que l'hypothèse d'une évolution linéaire du biais reste valide pour des distributions étroites en redshift, cette étude suggère que la meilleure stratégie pour l'application des *cluster-z* est l'approche *locale* en redshift. Elle permet en effet d'avoir accès à plus d'information tout en réduisant la sensibilité de l'analyse à l'évolution du biais.

Afin de quantifier cet effet, considérons une distribution gaussienne centrée en  $z_0$  de largeur  $\sigma_z$  :  $G(z_0, \sigma_z)$ . Si le biais galaxie-matière de cette population évolue selon une loi de puissance,  $\bar{b}_u(z) \propto z^\alpha$ , et que lors de la mesure de *clustering redshift* on considère qu'il n'a pas d'évolution du biais (c'est-à-dire  $d\bar{b}_u/dz = 0$ ) alors l'écart entre le redshift moyen estimé  $\langle z \rangle$  et le vrai redshift moyen  $z_0$  s'écrit :

$$\langle z \rangle - z_0 = \int dz z^{\alpha+1} G(z_0, \sigma_z) - \int dz z G(z_0, \sigma_z) . \quad (3.1)$$

On peut alors s'intéresser à l'évolution de cet écart en fonction du redshift pour différentes largeurs  $\sigma_z$  et pour différents scenarios d'évolution du biais. La Figure 3.2 montre que dans le cas d'une distribution dont la largeur est  $\sigma_z = 0.05$ , l'erreur due à l'évolution du biais sur le redshift moyen à  $z = 0.5$  est de l'ordre du pourcent ce qui nous informe immédiatement sur le domaine de validité de l'approche *locale* et de son formidable intérêt potentiel.

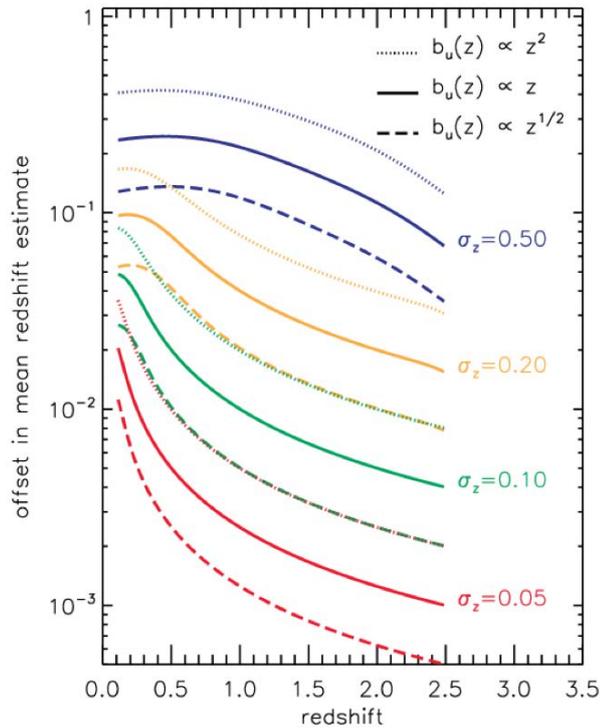


FIGURE 3.2 – Evolution de l'écart entre le redshift moyen mesuré et le vrai redshift moyen d'une population. Figure extraite de [Ménard et collab. \(2013\)](#).

## 1.2 Estimation du rapport signal sur bruit de la mesure

Avant d'entamer une analyse sur des catalogues, il est important connaître le signal-sur-bruit attendu pour la mesure de la corrélation croisée afin d'établir le domaine d'application de la méthode pour des données réelles.

Pour cela, considérons que la matière s'aggrège jusqu'à une certaine échelle  $r_c$  ou  $\delta z_c$  dans l'espace des redshifts. Dans ce cas, le rapport signal sur bruit de la mesure de la corrélation croisée peut être estimé de la façon suivante.

Soit un objet de référence  $j$  qui mesure le signal,  $\mathcal{S}_j$ . Fondamentalement pour cet objet, le signal est proportionnel au nombre de ses voisins  $N_u$  mais également à l'échelle du clustering  $\delta z_c$ . En effet pour un nombre de voisins donnés plus l'échelle du clustering sera grande plus la distribution des objets s'écartera d'une distribution de Poisson. On a donc :

$$\mathcal{S}_j \propto \delta z_c N_u . \quad (3.2)$$

En multipliant le terme de droite par :  $\pi \theta_{\max}^2 / \pi \theta_{\max}^2 = 1$ , où  $\theta_{\max}$  est l'angle correspondant à la plus grande échelle physique utilisée lors de la mesure. Il vient :

$$\mathcal{S}_j \propto \delta z_c \sum_u \pi \theta_{\max}^2 . \quad (3.3)$$

Pour  $N_{r,i}$  objets de référence dans le bin  $i$  :

$$\mathcal{S}_{r,i} \propto \delta z_c \sum_u \pi \theta_{\max}^2 N_{r,i} . \quad (3.4)$$

Pour tous les bins de référence, on a :

$$\mathcal{S} = \sum_i \mathcal{S}_{r,i} , \quad (3.5)$$

et pour une largeur de bin de référence  $\delta z_i$  constante :

$$N_r = \delta z_i \sum_i N_{r,i} . \quad (3.6)$$

On obtient :

$$\mathcal{S} \propto \frac{\delta z_c}{\delta z_i} \sum_u \pi \theta_{\max}^2 N_r . \quad (3.7)$$

Le bruit de Poisson associé à cette mesure dans un bin de référence peut s'écrire :

$$\mathcal{N}_{r,i} \propto \sqrt{\sum_u N_{r,i}} \sqrt{\pi} \theta_{\max} , \quad (3.8)$$

ce qui donne :

$$\mathcal{N} \propto \frac{1}{\sqrt{\delta z_i}} \sqrt{\sum_u N_r} \sqrt{\pi} \theta_{\max} . \quad (3.9)$$

On obtient donc l'expression suivante pour le rapport signal sur bruit :

$$\frac{\mathcal{S}}{\mathcal{N}} \propto \frac{\delta z_c}{\sqrt{\delta z_i}} \sqrt{N_r \Sigma_u} \sqrt{\pi} \theta_{\max}, \quad (3.10)$$

avec :

- $\theta_{\max}$ , l'angle correspondant à la plus grande échelle physique utilisée lors de la mesure.
- $\delta z_c$ , l'échelle où l'on considère le clustering dans l'espace des redshifts.
- $\delta z_i$ , la taille du  $i^{\text{ème}}$  bin de *référence*.
- $N_r$ , le nombre d'objets total dans l'échantillon de *référence*.
- $\Sigma_u = N_u/\pi\theta_{\max}$ , la densité de surface de l'échantillon *inconnu*.

Cette expression permet de prédire la faisabilité de la mesure et peut être utilisée pour déterminer la taille des bins en redshift de *référence*  $\delta z_i$  et l'échelle physique maximum exploitable. Comme on pouvait s'y attendre pour la mesure de la fonction de corrélation croisée, la quantité qui contrôle  $\mathcal{S}/\mathcal{N}$  dans un catalogue est le nombre de *paire* d'objets. On peut également noter que cette expression approximative ne prend pas en compte l'influence de la luminosité des objets sur l'amplitude du clustering, plus faible pour les objets peu brillants. Elle ne prend pas non plus en compte le fait que  $\theta_{\max}$  est limité par le choix de l'échelle physique maximale utilisée qui ne doit pas excéder quelques mégaparsecs de façon à réduire l'importance de l'évolution du biais comme nous l'avons vu précédemment.

## 2 Mesure de $\bar{\omega}_{ur}(z)$ sur les données SDSS

Dans cette partie j'entreprends la reconstruction de la distribution en redshift d'une population *inconnue* à partir du formalisme présenté précédemment afin de redémontrer la validité de cette approche *cluster-z* sur des données. Pour cela j'ai reproduit en partie l'analyse présentée dans [Ménard et collab. \(2013\)](#). La reproduction de cette étude a constitué une partie importante du travail effectué durant cette thèse. Cela m'a permis d'acquérir les connaissances et les compétences nécessaires afin de maîtriser la mesure de redshift clustering et de valider les codes et les outils que j'ai développés et implémentés tout au long de ma thèse pour conduire mes propres analyses.

Dans cette étude j'utilise des quasars (QSOs) du *SDSS* ([Schneider et collab., 2010](#)) comme échantillon de *référence* car d'une part ce sont des objets résidant dans des halos de matière noire massifs - pour lesquels on s'attend à avoir un signal fort - et d'autre part parce qu'ils couvrent une large gamme de redshift allant de 0 à 5.

J'utilise le catalogue mega-z LRG (Collister et collab., 2007) comme échantillon *inconnu*. Les LRGs - pour galaxies rouges lumineuses - sont également des objets résidants dans des halos massifs ; on peut donc s'attendre à ce que le signal de clustering associé à ces objets soit aussi important. De plus, les redshifts photométriques de cette population de galaxies sont à notre disposition de façon à pouvoir vérifier les résultats.

### Sélection des données

J'ai appliqué les critères suivants :

- population *inconnue*, LRGs : pas de doublon avec les objets de *référence*. Les deux échantillons sont mutuellement exclusifs.
- population de *référence*, QSOs : couvre les mêmes régions du ciel que celles des LRGs, en incluant les masques du *SDSS*.

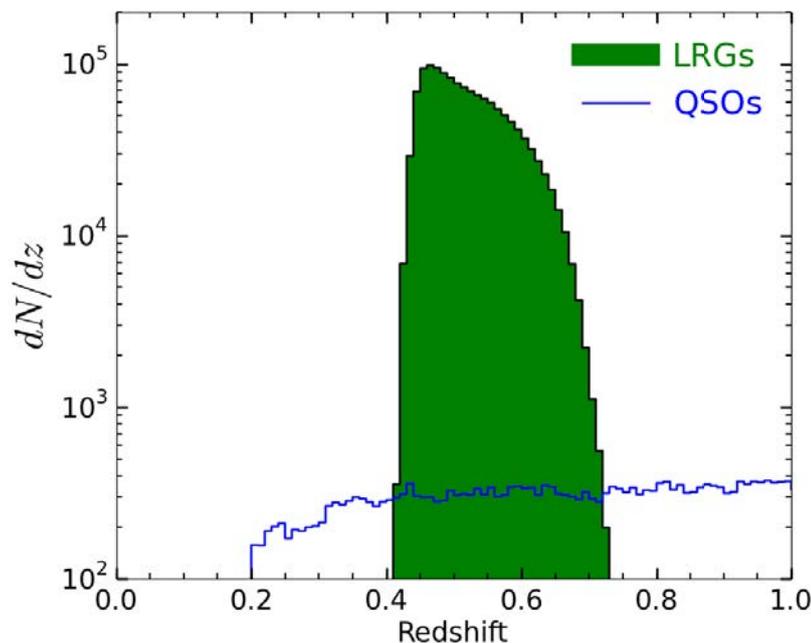


FIGURE 3.3 – Distribution en redshift des deux échantillons que j'ai obtenue à partir des données du *SDSS*. La taille des bins de l'histogramme est  $\Delta z = 0.01$ . Les redshifts des objets de *référence* sont des mesures spectroscopiques, ceux de l'échantillon *inconnu* sont des redshifts photométriques. La distribution en vert correspond au signal que l'on veut retrouver par la méthode de corrélation croisée.

L'échantillon de *référence* se compose d'environ 90 000 objets et l'échantillon *inconnu* contient environ 1 300 000 LRGs. Les distributions en redshift des deux populations ainsi que leurs positions sur le ciel avant et après l'application des masques, sont visibles sur les Figures 3.3 & 3.4.

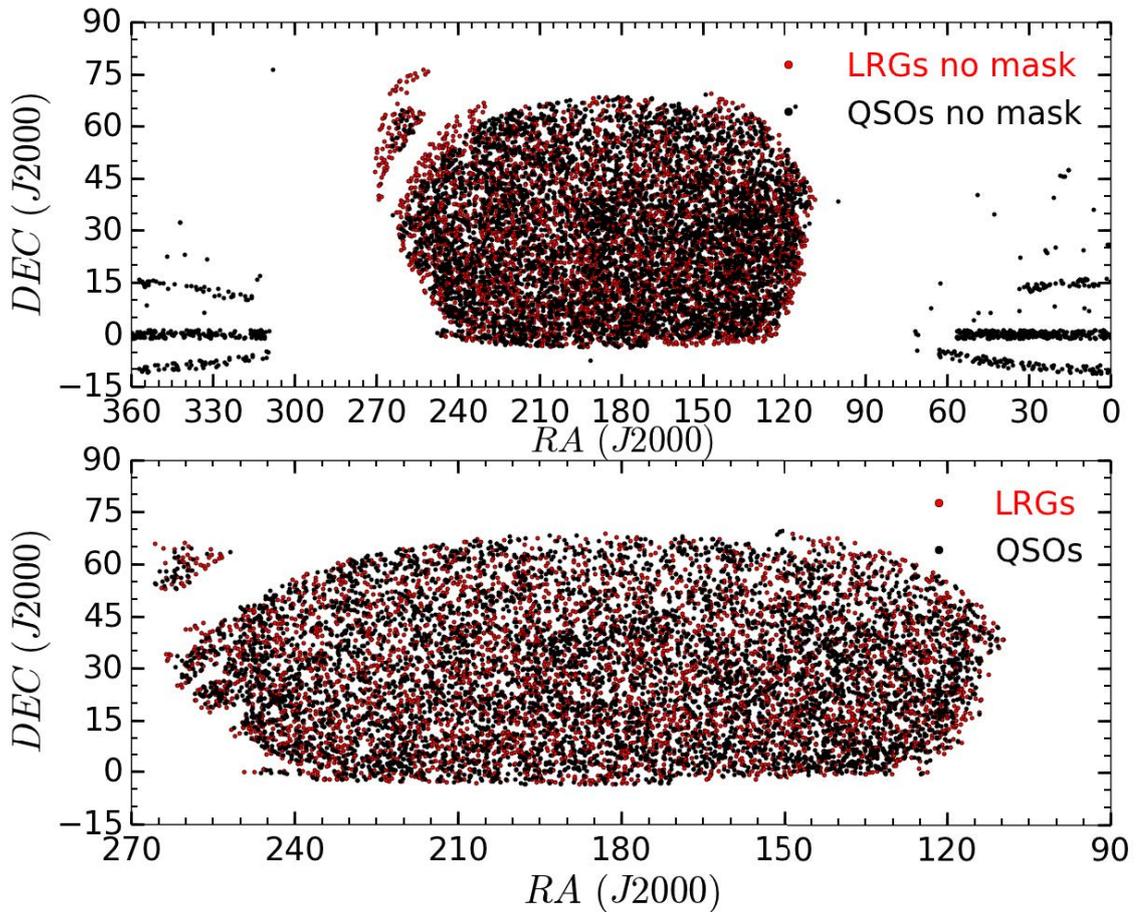


FIGURE 3.4 – Haut : position sur le ciel des objets des deux échantillons que j'ai sélectionnés avant l'application des masques.

Bas : la même chose, après l'application des masques. Les 1.3 millions d'objets *inconnus* et les 90 mille objets de *référence* couvrent une surface totale d'environ 7 400 deg<sup>2</sup>.

En considérant les paramètres suivants :  $\delta z_c = 0.001$ ,  $\delta z_i = 0.03$ ,  $N_r = 90\,000$ ,  $\Sigma_u = 1.3M/7\,400 \sim 176$ ,  $\theta_{\max} \sim 0.1$  deg pour une échelle physique de 1 Mpc à redshift 0.2, l'estimation du rapport signal sur bruit de la mesure calculé suivant l'équation (3.10) est de  $\mathcal{S}/\mathcal{N} \sim 4$ .

### Mesure directe de $\bar{\omega}_{ur}(z)$

Dans le calcul de la corrélation à partir des catalogues LRGs & QSOs, les galaxies sont assimilées à des points alors qu'en réalité elles possèdent une certaine taille et donc certaines galaxies se chevauchent. De plus la largeur de la fonction d'étalement (*point spread function*, PSF) de l'instrument ne permet pas de distinguer des objets trop proches l'un de l'autre. Il n'est donc pas possible d'exploiter des objets du catalogue localisés dans un certain rayon autour d'une galaxie de *référence* donnée. L'estimation du nombre d'objets *inconnus* dans un cercle de rayon  $\theta_{\max}$  centré sur un objet de *référence* est donc biaisée si la mesure est réalisée de  $\theta = 0$  à  $\theta = \theta_{\max}$ . Afin de prendre en compte ces effets j'ai considéré une échelle minimale  $\theta_{\min}$  pour la mesure de corrélation correspondant à une échelle physique de 100 kpc autour de chaque objet de *référence*.

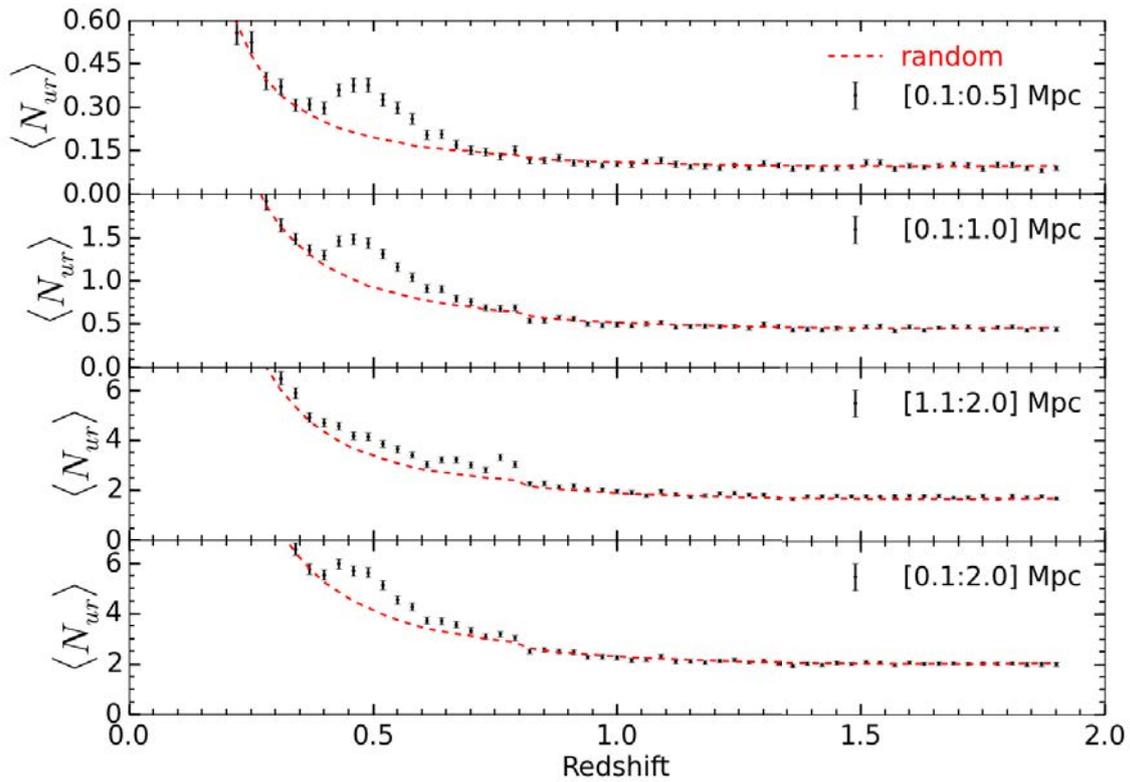


FIGURE 3.5 – Évolution avec le redshift du nombre moyen d'objets *inconnus* autour des objets de *référence* que j'ai mesurée pour différentes échelles physiques. On observe une augmentation du nombre d'objets entre  $0.4 < z < 0.7$ . On peut également remarquer que la majorité de l'information à extraire se situe aux petites échelles comme le montre la disparition du pic lorsque l'on retire les échelles  $< 1$  Mpc. J'ai obtenu les barres d'erreurs en supposant une simple statistique de Poisson.

La Figure 3.5 montre la mesure de l'évolution avec le redshift du nombre moyen d'objets *inconnus* autour des objets de *référence*  $\langle N_{ur} \rangle$ . J'ai effectué ces mesures pour trois intervalles d'échelle physique. On observe une augmentation du nombre d'objets entre  $0.4 < z < 0.7$ , en accord avec l'intervalle des redshifts photométriques des objets de l'échantillon *inconnu*. En dehors de cet intervalle, la quantité mesurée est en accord avec une distribution aléatoire. Les barres d'erreurs sont obtenues via la statistique de Poisson.

Je me suis alors intéressé à l'évolution avec le redshift de la densité moyenne mesurée sur les données  $\langle \bar{\Sigma}_{ur} \rangle$  et je l'ai comparée avec celle obtenue pour un champ Poissonien, voir Figure 3.6. On peut y voir à l'oeil où se trouve le signal. J'ai ensuite calculé les  $\bar{\omega}_{ur}(z)$  qui sont montrés sur la Figure 3.7 à partir de l'équation (2.15).

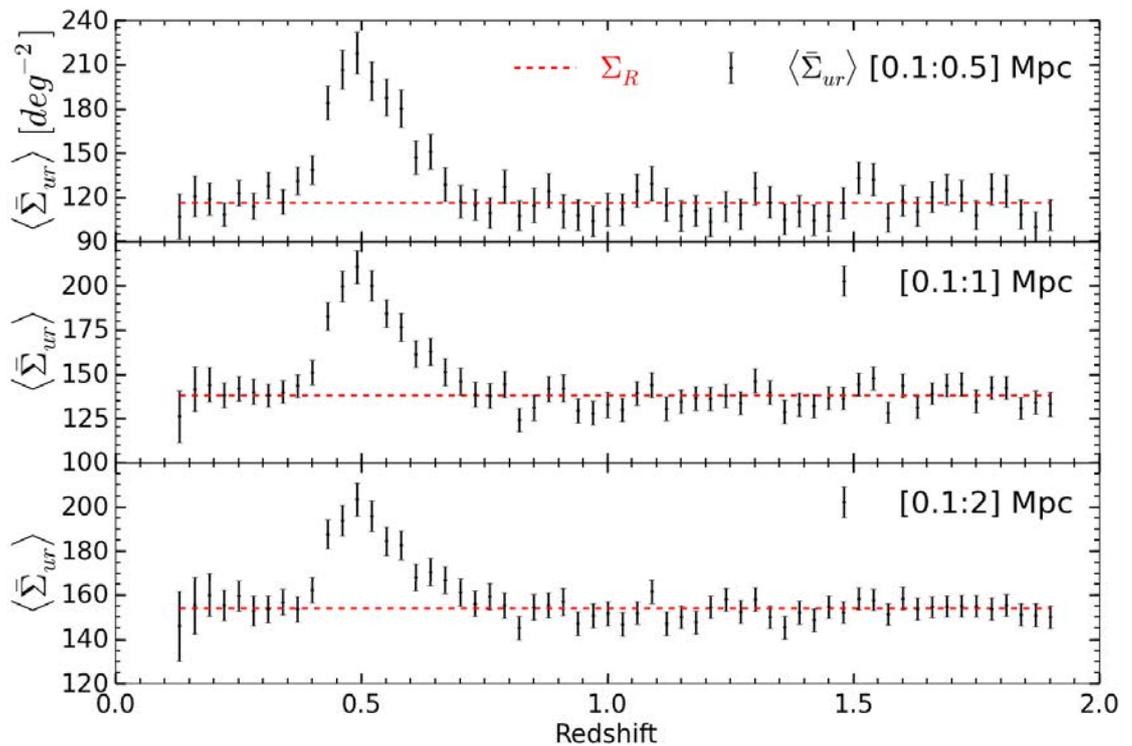


FIGURE 3.6 – Évolution avec le redshift de la densité surfacique moyenne mesurée sur les données - en noir - et pour un champ aléatoire - en rouge. Le rapport de ces deux quantités selon l'équation (2.15) fournit la fonction de corrélation intégrée :  $\bar{\omega}_{ur}(z)$ . J'ai obtenu les barres d'erreurs en suivant la procédure standard de propagation des erreurs.

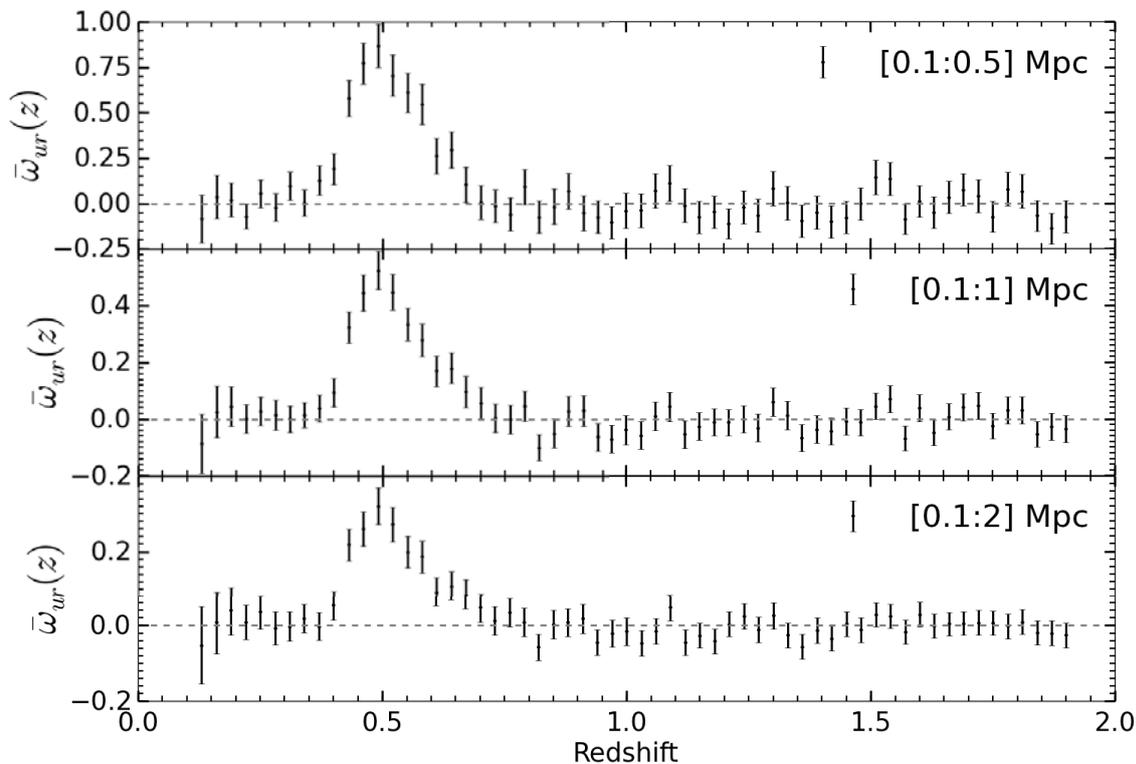


FIGURE 3.7 – Évolution avec le redshift de la fonction de corrélation croisée intégrée  $\bar{\omega}_{ur}(z)$  telle que j’ai pu la mesurer directement sur les données pour plusieurs échelles physiques. Les pointillés montrent le niveau zéro sur toute la gamme de mesure. Les barres d’erreurs sont obtenues en suivant la procédure standard de propagation des erreurs.

### 3 Validation du redshift clustering sur les données SDSS

En accord avec l’équation (2.12) il ne me reste plus qu’à déterminer l’évolution avec le redshift du biais galaxie-matière de l’échantillon de *référence*. Pour cela, j’ai utilisé l’équation (3.11), selon Porciani et Norberg (2006) :

$$\bar{b}_{QSOs}(z) = \frac{1}{\sigma_8} \left[ 1 + \left( \frac{1+z}{2.5} \right)^\gamma \right], \quad (3.11)$$

avec :  $\gamma = 4$ , selon la correction de Shen et collab. (2013).

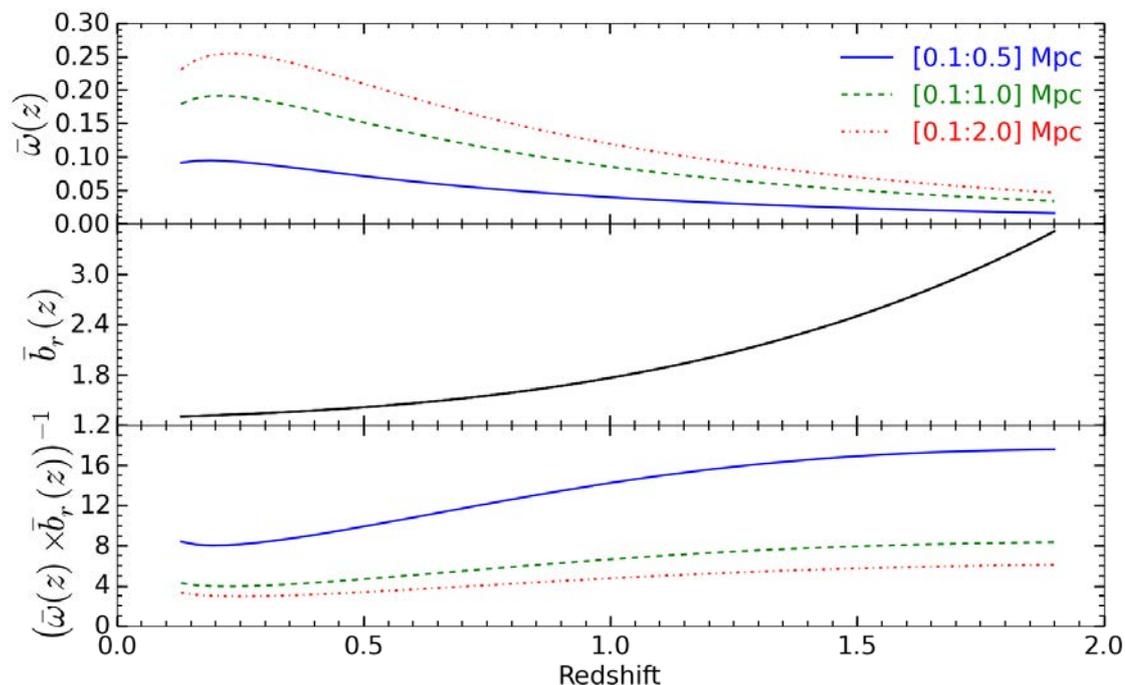


FIGURE 3.8 – Haut : évolution avec le redshift de la fonction de corrélation intégrée de la matière pour trois échelles différentes, que j’ai obtenue à partir du spectre de puissance comme présenté au Chapitre 2.

Milieu : évolution avec le redshift du biais galaxie-matière de l’échantillon de *référence* selon Porciani et Norberg (2006).

Bas : second terme de l’équation (2.12).

La Figure 3.8 résume l’évolution de  $\bar{\omega}(z)$  et de  $\bar{b}_r(z)$  avec le redshift. A partir de la mesure de  $\bar{\omega}_{ur}$ , j’ai ensuite appliqué l’équation (2.11) et j’ai pu reconstruire la distribution en redshift de l’échantillon *inconnu* montré sur la Figure 3.9. Le résultat est en très bon accord avec celui de Ménard et collab. (2013). Cette mesure de la distribution en redshift à partir de l’information spatiale du clustering des objets constitue une partie importante du travail que j’ai effectué durant la première partie de ma thèse. Les principales différences entre ces deux analyses sont le choix du filtre  $W(\theta)$  utilisé lors de la mesure ( $W(\theta) \propto \theta^{-1}$  dans Ménard et collab. (2013)) et l’estimation des barres d’erreurs qui dans cet article sont obtenues via une procédure de ré-échantillonnage de type Jackknife.

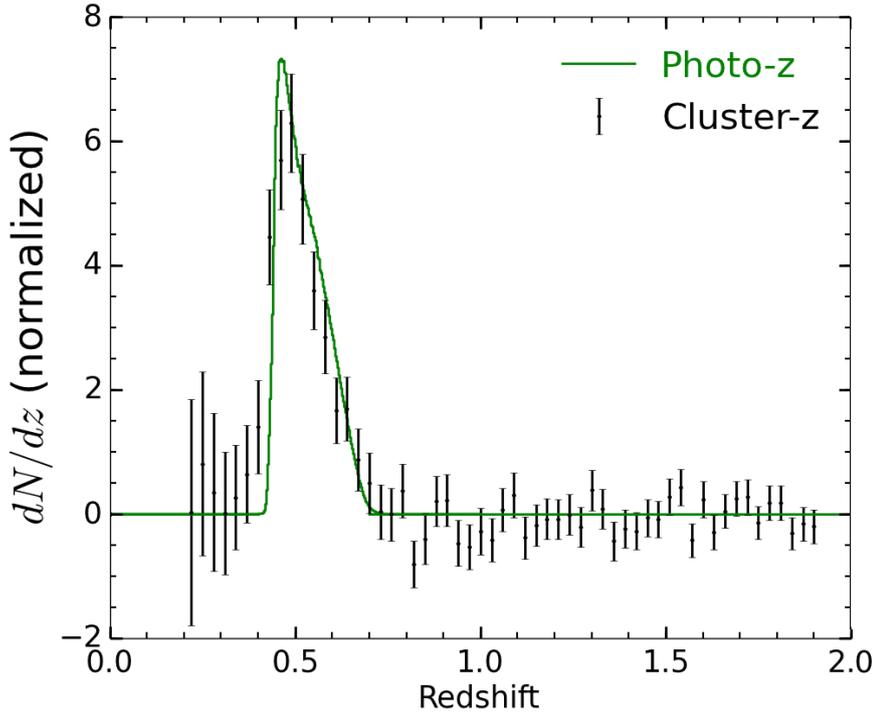


FIGURE 3.9 – Comparaison entre la distribution en redshift de l'échantillon *inconnu* que j'ai mesurée en utilisant la méthode de corrélation croisée entre 100 kpc et 1 Mpc et celle obtenue par les redshifts photométriques. Les deux distributions sont en accord, voir Table 3.1. La taille des bins est :  $\delta z = 0.03$ . Ces résultats sont en accord avec Ménard et collab. (2013). Les barres d'erreurs sont obtenues via la statistique de Poisson.

Afin de quantifier l'accord entre les *clustering redshift* et les *photo-z*, j'ai comparé les deux premiers moments de ces distributions, Table 3.1.

Pour rappel la moyenne  $\langle x \rangle$  et la variance  $\sigma^2$  d'une distribution  $y = f(x)$ , sont :

$$\langle x \rangle = \frac{\sum_i x_i y_i}{\sum_i y_i}, \quad (3.12)$$

et :

$$\sigma^2 = \frac{\sum_i y_i (x_i - \langle x \rangle)^2}{\sum_i y_i}. \quad (3.13)$$

La validation de la méthode du *clustering redshift* sur des données est très encourageante. Il reste cependant des points à développer pour en estimer son potentiel et ses limites éventuelles. Il convient notamment de rappeler que cette analyse est un cas particulier favorable dans le sens où les objets utilisés - LRGs & QSOs - sont connus pour résider dans des halos de matière noire très massifs. On peut donc s'attendre à ce que le signal ressorte davantage que pour d'autres types d'objets.

	Photo-z	Cluster-z
$\langle z \rangle$	0.51	0.50
$\sigma$	0.06	0.07

TABLE 3.1 – Moyenne et variance pour la distribution en redshift de l'échantillon *inconnu*. Les *clustering redshifts* et les redshifts photométriques sont en accord.

## 4 Généralisation à tous types de populations de galaxies

Dans cette partie, j'entreprends le même type d'étude que précédemment mais en remplaçant les LRGs par un catalogue de galaxies comportant tous les types. Je construis ce nouvel échantillon *inconnu* qui contient environ 66 000 objets à partir des critères suivants appliqués aux données *SDSS DR10* :

- l'objet est une galaxie.
- l'objet possède à la fois un redshift photométrique et un redshift spectroscopique. Ces informations seront utiles à la fin de l'analyse afin de comparer l'accord entre ces trois méthodes de mesure du redshift.
- le  $z_{phot}$  de l'objet est :  $0.45 < z_{phot} < 0.50$  et  $\sigma_{z_{u,phot}} < 3 \cdot 10^{-2}$ .

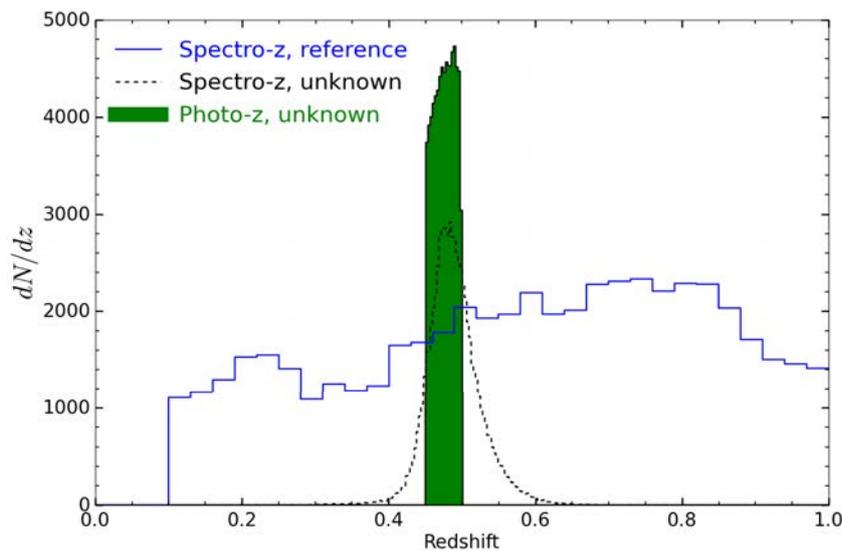


FIGURE 3.10 – Distribution en redshift de l'échantillon de référence que j'ai sélectionné à partir de la base de données du *SDSS DR10*. Pour des raisons de clarté la largeur des bins de cet histogramme est choisie telle que  $\Delta z_r = 0.03$  et  $\Delta z_u = 0.003$ .

Le nouvel échantillon de *référence* contient environ 47 000 objets et est obtenu en sélectionnant des QSOs observés spectroscopiquement et dont le redshift se situe entre  $0.1 < z_r < 1.0$ , Figure 3.10. Les deux échantillons se recouvrent sur une région du ciel d'environ  $4\,300 \text{ deg}^2$ . Afin de préserver un rapport signal sur bruit comparable à l'analyse précédente, j'ai choisi de prendre :  $\delta z_i = 0.015$  et  $r_{\text{max}} = 3 \text{ Mpc}$ , ce qui correspond à  $z = 0.1$  à  $\theta_{\text{max}} \sim 0.45 \text{ deg}$ . On a alors,  $\mathcal{S}/\mathcal{N} \sim 5.5$ .

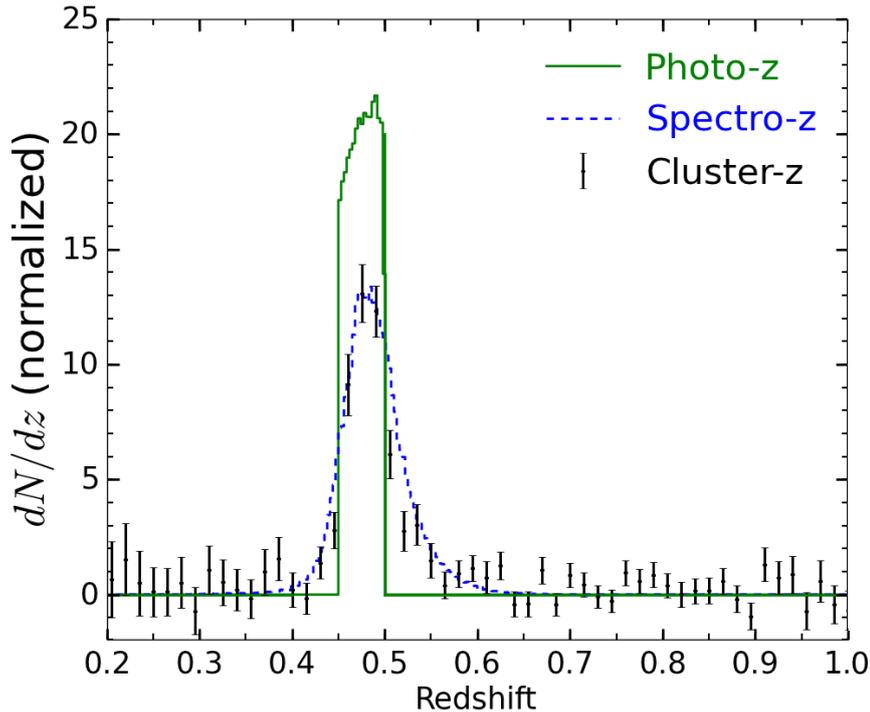


FIGURE 3.11 – Superposition des trois mesures de redshift pour l'échantillon *inconnu* contenant des galaxies de tous types. La distribution que j'ai mesurée par corrélation croisée présente un meilleur accord avec les observations spectroscopiques que les redshifts photométriques. Il semble que les *clusters-z* soient plus sensibles au vrai redshift que les *photos-z*. Le bin des mesures vaut  $\delta z = 0.015$ . Toutes les distributions sont normalisées à l'unité.

La distribution en *clustering redshift*, ainsi que les réponses photométriques et spectroscopiques de l'échantillon *inconnu* que j'ai mesurées sur ces données sont visibles à la Figure 3.11. On peut remarquer que la distribution obtenue par corrélation croisée est plus conforme à la distribution *spectroscopique* que de la distribution *photométrique*. En plus de valider les redshifts clustering quel que soit le type de galaxie, cela suggère également que les *cluster-z* peuvent apporter une mesure plus proche des *vrais* redshifts que les *photos-z*. Il convient cependant de préciser que dans cette analyse j'ai négligé la dispersion propres aux redshifts *photométriques*. Il est aussi possible que la précision et l'exactitude des *cluster-z* dépendent du contenu du catalogue de référence. Cet accord a priori meilleur entre cluster et spectro-z est

donc à nuancer et nécessite une investigation plus approfondie pour en connaître la portée. Elle sera présentée au Chapitre 4.

Comme précédemment j'ai comparé ces distributions en calculant leurs deux premiers moments, Table 3.2. Les trois valeurs moyennes sont en accord, cependant dans ce cas les *clusters-z* sont en meilleur accord que les redshifts photométriques pour retrouver la largeur de la distribution.

	Spectro-z	Photo-z	Cluster-z
$\langle z \rangle$	0.48	0.48	0.48
$\sigma$	0.04	0.02	0.04

TABLE 3.2 – Les trois valeurs moyennes sont en accord, cependant dans ce cas les *clusters-z* sont meilleurs que les redshifts photométriques pour retrouver la largeur de la distribution. Comme précédemment, ces valeurs sont obtenues à partir des équations (3.12) et (3.13).

## 5 Synthèse

Après avoir étudié les résultats des simulations, mon travail a consisté à confirmer la validité de la mesure de redshifts par corrélation croisée sur des données en reproduisant les résultats obtenus par [Ménard et collab. \(2013\)](#). Cette étape nécessaire m'a permis de construire et de valider tous mes outils d'analyse, d'acquérir les connaissances et les compétences requises pour réaliser une mesure *cluster-z*. J'ai ensuite étendu cette analyse en généralisant l'application des *cluster-z* à n'importe quel type de galaxies. J'ai alors montré qu'en plus de fonctionner pour tous les types de galaxies les *cluster-z* semble également mieux reproduire les propriétés de la distribution des *vrai* redshifts que les *photos-z*. Bien que j'ai ici négligé la dispersion des redshifts *photométriques* ce point mérite d'être étudié plus en détails, ce qui sera fait dans le Chapitre 4. Je me placerai alors dans un cas idéal de façon à explorer les limites de cette technique en comparant directement les *cluster-z* aux redshifts *spectroscopiques*.



## Ideal case study : abstract

This chapter explores the accuracy of the *clustering redshift* procedure in an ideal case. This is done by selecting *unknown* subsamples in a very narrow redshift region. While this selection is based on the *spectroscopic redshift* of *unknown* objects and is then irrelevant for a real case study, it allows us to completely neglect the evolution with redshift of the galaxy bias  $\bar{b}_u(z)$ . I show that this approach reaches an accuracy of  $\sim 10^{-4}$  on the bias  $\langle \Delta \bar{z} \rangle$ , and the scatter  $\sigma$ , where  $\Delta \bar{z} = \bar{z}_{clust} - \bar{z}_{spec}$ , Figure 3.12.

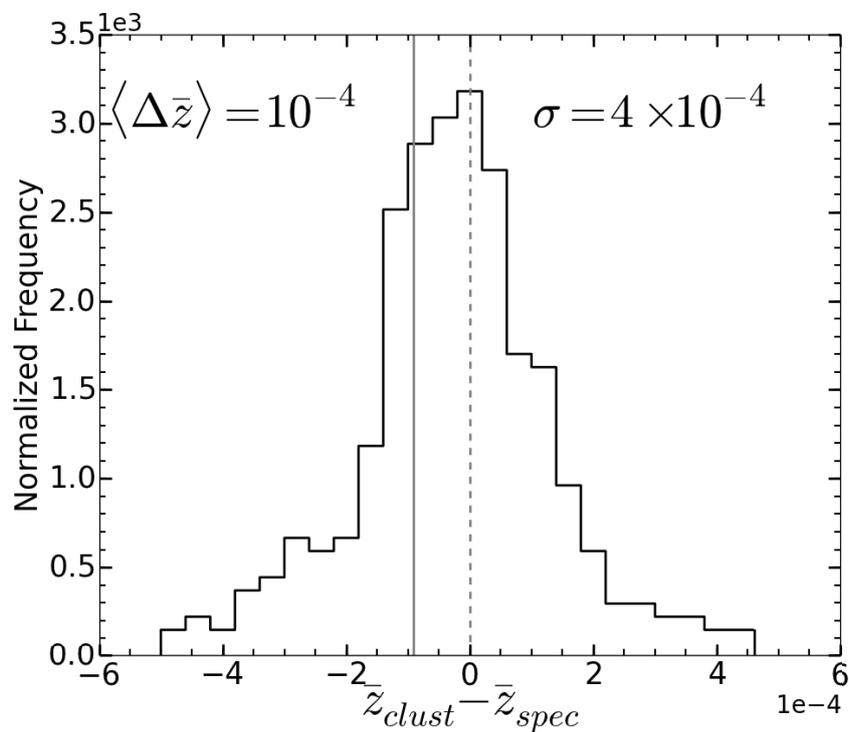


FIGURE 3.12 – Distribution of the residues normalized to unity. The vertical grey line shows the mean of the histogram while the dashed line is centered on zero.

Then, I present and discuss different sources of systematic effects affecting this measurement.



# Chapitre 4

## Clustering redshift : cas idéal

---

Dans ce chapitre, je me concentre sur une application idéale des *cluster-z*. Il est important de préciser que l'analyse qui y est présentée utilise le redshift spectroscopique des objets de l'échantillon *inconnu* lors du prétraitement des données. Cette étape permet de fortement réduire l'influence de la variation avec le redshift du biais galaxie-matière de la population *inconnue*. C'est en ce sens qu'il s'agit d'un cas idéal. Cette étude me permet d'explorer l'accord entre les *cluster-z* et les redshifts *spectroscopiques* et de discuter certains effets systématiques qui limitent les performances de la méthode des redshifts clustering. Des développements permettant de se passer de l'étape de pré-sélection utilisant des *spectro-z* seront présentés au chapitre suivant.

### 1 Sélection des données et mesure du biais de référence

#### 1.1 Sélection des données

Comme nous l'avons vu précédemment, la détermination de la distribution en redshift par la méthode des *cluster-z* nécessite la connaissance de l'évolution avec le redshift du biais galaxie-matière,  $\bar{b}_u(z)$ , équation (2.11). J'ai alors montré que ce terme pouvait être considéré comme constant dans le cas où la largeur de la distribution *inconnue* est faible. Afin de constater l'influence de la variation de  $\bar{b}_u(z)$  je m'intéresse, dans un premier temps, à une distribution bimodale.

Considérons une telle distribution pour laquelle chaque mode contient le même nombre d'objets et se trouve à un redshift moyen différent  $z_1$  et  $z_2$ . La valeur du biais  $\bar{b}_u(z)$  qui sera prise en compte lors de la normalisation décrite par l'équation (2.13) correspondra à  $\bar{b}_u(z = (z_1 + z_2)/2)$ .

Le biais galaxie-matière est une fonction croissante avec le redshift (Fry, 1996; Tegmark et Peebles, 1998). Par conséquent la valeur du biais du premier mode

est surestimée alors que celle du second est sous-estimée. D'après l'équation (2.11) cela se répercute sur la mesure du clustering de façon inverse, c'est-à-dire que le clustering sera respectivement sous-estimé et surestimé. Je montre sur la Figure 4.1 qu'en séparant ces modes, lors d'une étape de prétraitement, cet effet est grandement atténué. Dans le cas d'une distribution large, cet effet se traduit par une distorsion de la distribution en redshift mesurée par rapport à la véritable distribution.

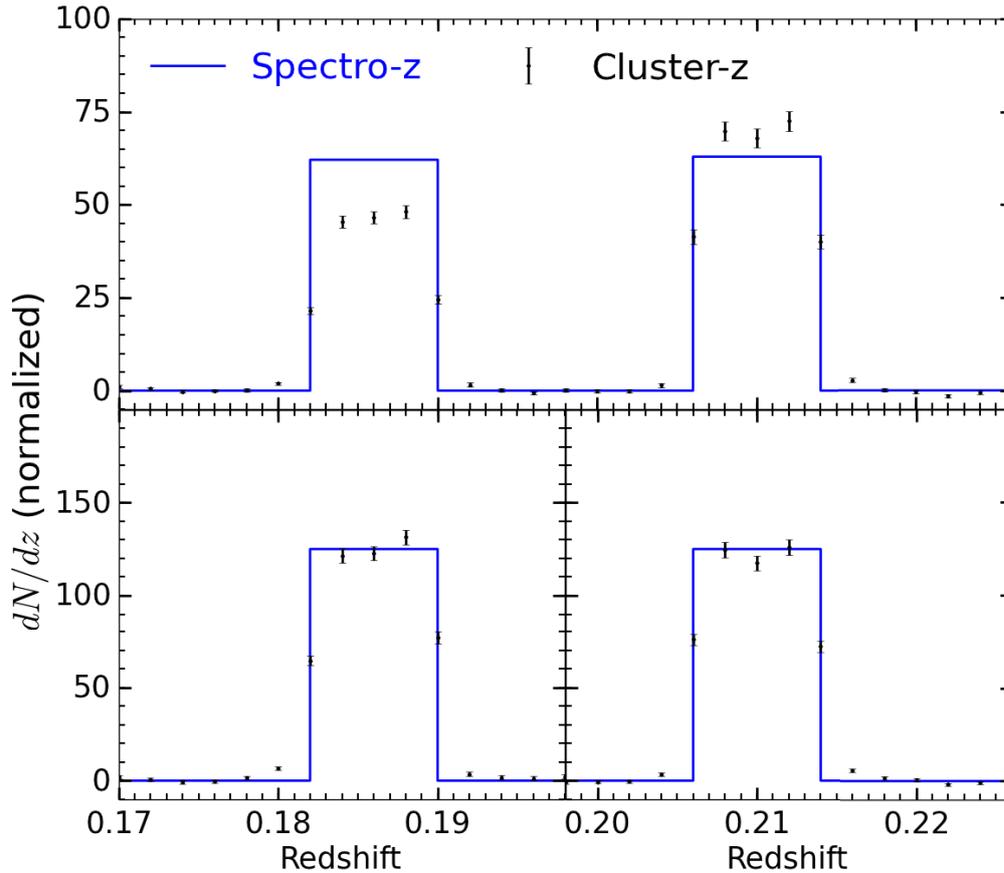


FIGURE 4.1 – Mesure de redshift par la méthode du clustering pour une distribution bimodale (haut) et pour chacun des modes séparément (bas). Sur cette figure je montre que séparer les modes permet de grandement diminuer l'effet de l'évolution de  $\bar{b}_u(z)$ .

Comme nous l'avons vu au Chapitre 3 la meilleure stratégie pour l'application des *cluster-z* consiste à utiliser l'approche *locale* en redshift. Dans un premier temps, j'ai choisi de travailler avec une population *inconnue* pour laquelle le redshift spectroscopique est connu. J'utilise cette information lors d'une étape de prétraitement des données afin de créer des sous-échantillons localisés dans l'espace des redshifts. Je pourrai alors utiliser la méthode des *cluster-z* pour mesurer ces distributions. Ce

prétraitement permet de négliger l'effet de la variation de  $\bar{b}_u(z)$  et de me placer dans le cas idéal.

J'ai sélectionné dans la base de données du *SDSS-DR7* toutes les galaxies qui possèdent un redshift spectroscopique compris entre  $z = 0.01$  et  $z = 0.45$  et dont la précision est meilleure que  $10^{-4}$ . On obtient alors un échantillon d'environ 700 000 objets répartis sur 5 000  $\text{deg}^2$  du ciel et qui constituent la population de *référence*.

J'ai construit l'échantillon *inconnu* en sélectionnant aléatoirement sur le ciel un sous-ensemble de 100 000 objets dont la distribution en redshift est uniforme, c'est-à-dire telle que  $\frac{d^2N}{dz^2} = 0$ . Ces objets sont ensuite retirés du catalogue de *référence* dont la taille est alors de 600 000 objets. Les deux distributions que j'ai obtenues après cette sélection sont visibles sur la Figure 4.2.

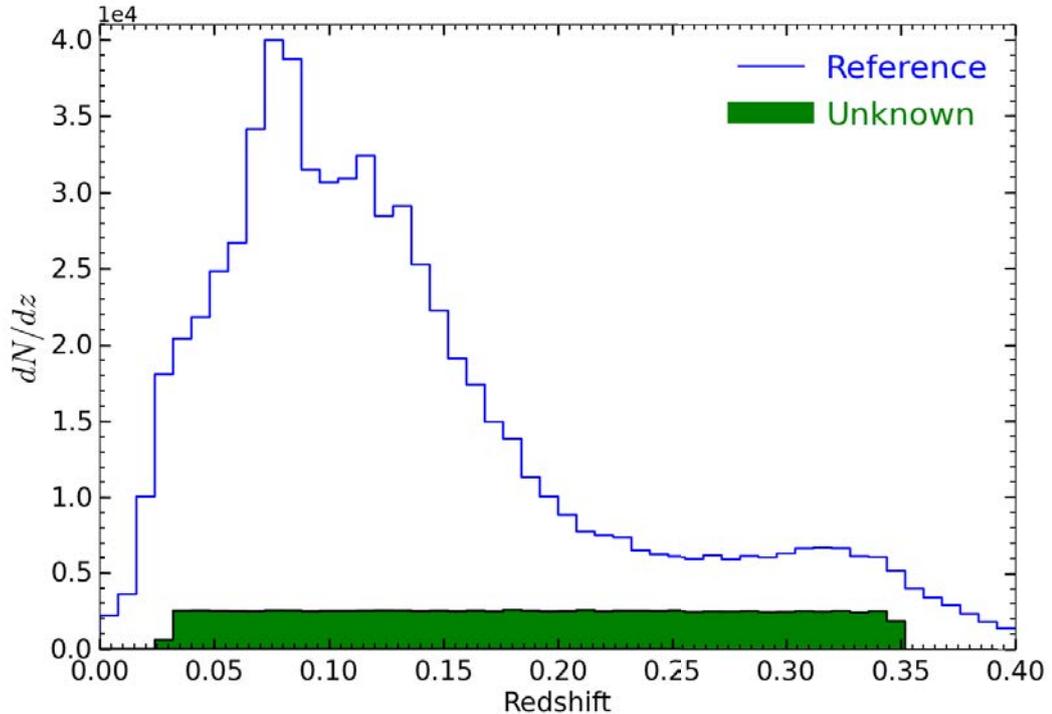


FIGURE 4.2 – Distributions en redshift des populations de *référence/inconnue*, respectivement en bleu/vert.  $\Delta z = 0.008$  pour les deux histogrammes.

Le rayon de collision des fibres du *SDSS* correspond au rayon des fibres utilisées lors de l'acquisition des spectres. La région correspondante sur le ciel ne peut physiquement pas être couverte par une autre fibre. Cette contrainte induit un biais dans la détection des objets. La valeur de ce rayon est de  $55''$ . À  $z = 0.4$  elle impose une échelle physique minimale de 300 kpc pour la mesure de clustering.

Dans cette analyse en considérant :

- $\Sigma_u = 100\,000 / 5\,000 = 20$  .
- $N_r = 600\,000$  .
- $\delta z_i = 2 \cdot 10^{-4}$  .
- $\delta z_c = 10^{-3}$  .
- $\theta_{\max} \sim 1$  deg, pour  $r_{\max} = 3$  Mpc à  $z = 0.05$ .

On obtient, pour l'estimation du rapport signal sur bruit :  $\mathcal{S}/\mathcal{N} \sim 400$ . Afin de diminuer l'influence de l'évolution de  $\bar{b}_u(z)$  j'ai alors choisi de faire 400 sous-ensembles *inconnu* de largeur en redshift  $\Delta z_u = 8 \cdot 10^{-4}$ . Cela réduit le  $\mathcal{S}/\mathcal{N}$  pour chaque analyse à environ 20, que l'on peut comparer au  $\mathcal{S}/\mathcal{N}$  de l'analyse avec les LRGs et les QSOs du chapitre précédent qui valait environ 4. On peut donc s'attendre à obtenir des mesures moins bruitées.

Puisque l'échantillon de *référence* n'est plus constitué uniquement de QSOs, on ne peut plus se baser sur l'équation (3.11) pour l'estimation du biais galaxie-matière de la population de *référence*. Je me suis alors intéressé à la mesure de ce biais directement sur les données.

## 1.2 Mesure de l'amplitude du clustering en régime non-linéaire

La mesure de distribution en redshift via la méthode du clustering dépend du redshift du biais galaxie-matière de la population de *référence* selon la relation :

$$dN_u/dz \propto \bar{\omega}_{ur}(z)/\bar{b}_r(z) , \quad (4.1)$$

où, comme vu précédemment, on a considéré :  $\Delta z_u \rightarrow \delta_D(z_u)$ .

La mesure de  $\bar{b}_r(z)$  via  $\bar{\omega}_{rr}(z) = \bar{b}_r(z) \bar{\omega}(z)$  est problématique pour plusieurs raisons. D'une part cette expression est uniquement valide dans le régime où les champs de densité de galaxies et de matière noire sont, en moyenne, linéaires. C'est-à-dire aux grandes échelles physiques. Or, nous nous intéressons justement aux petites échelles. D'autre part l'utilisation de cette expression dépend des paramètres cosmologiques via  $\bar{\omega}(z)$ . Cette approche pour la mesure de  $\bar{b}_r(z)$  dépend donc du modèle d'univers considéré.

Afin de s'affranchir de ces limites on définit,  $\beta_r(z)$ , l'amplitude du clustering normalisé à un redshift arbitraire  $z_0$  :

$$\beta_r(z) = \sqrt{\frac{\bar{\omega}_{rr}(z)}{\bar{\omega}_{rr}(z_0)}} . \quad (4.2)$$

Il est important de réaliser que cette quantité est différente du biais galaxie-matière linéaire qui n'est défini qu'à grande échelle. Cette définition de l'amplitude

du clustering inclut la contribution des petites échelles pour lesquelles les champs de galaxies et de matière sont reliés de façon non linéaire car elle est le ratio de deux quantités mesurées directement sur les données qui contiennent cette information. Par contre, cette quantité ne fait plus intervenir le modèle cosmologique par l'intermédiaire de  $\bar{\omega}(z)$ . En adoptant une approche locale en redshift, on a :

$$\beta_r(z) = \sqrt{\frac{\bar{\omega}_{rr}(z)}{\bar{\omega}_{rr}(z_0)}} \propto \frac{\bar{b}_r(z)}{\bar{b}_r(z_0)}. \quad (4.3)$$

Il vient :

$$dN_u/dz \propto \bar{\omega}_{ur}(z)/\beta_r(z). \quad (4.4)$$

J'ai alors déterminé  $\beta_r(z)$  en mesurant la fonction d'auto-corrélation intégrée de l'échantillon de *référence* entre 300 kpc et 3 Mpc. L'évolution de ma mesure de l'amplitude du clustering est montrée sur la Figure 4.3. La figure montre la mesure point par point ainsi qu'une version lissée via l'application d'un filtre de Hann d'intervalle de lissage  $\Delta_H z = 0.02$ .

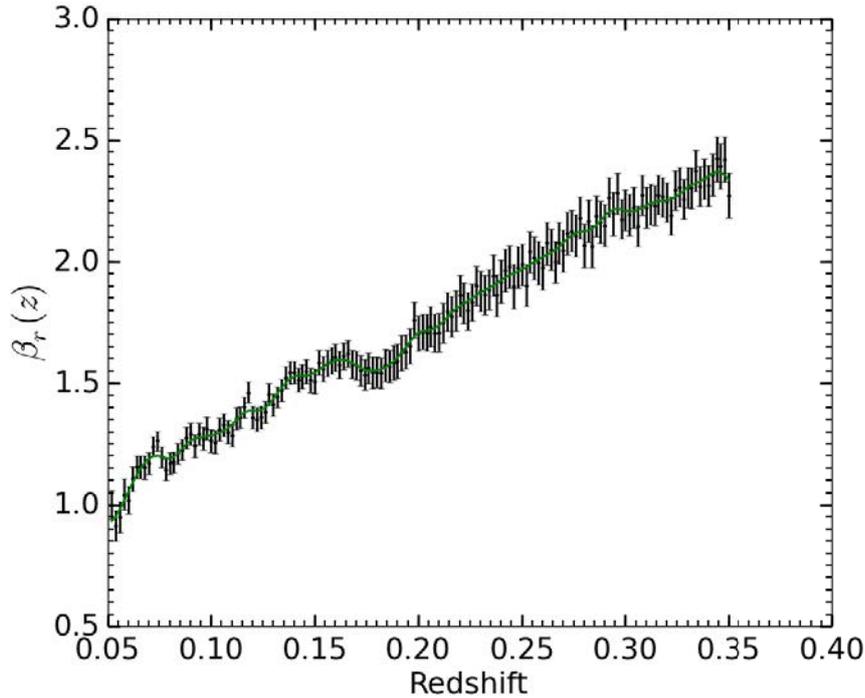


FIGURE 4.3 – Évolution de l'amplitude du clustering de la population de *référence* entre 300 kpc et 3 Mpc, normalisée en  $z_0 = 0.05$ . J'ai mesuré cette quantité directement sur les données. La courbe verte correspond à la même quantité après application d'un filtre de Hann d'intervalle de lissage  $\Delta_H z = 0.02$ .

Maintenant que  $\beta_r(z)$  est connu, j'ai mesuré la corrélation croisée entre les deux échantillons  $\bar{\omega}_{ur}(z)$  puis appliqué l'équation (4.4) afin d'obtenir la distribution en redshift.

## 2 Comparaison *cluster/spectro-z* et étude du bruit

### 2.1 Comparaison *cluster/specro-z*

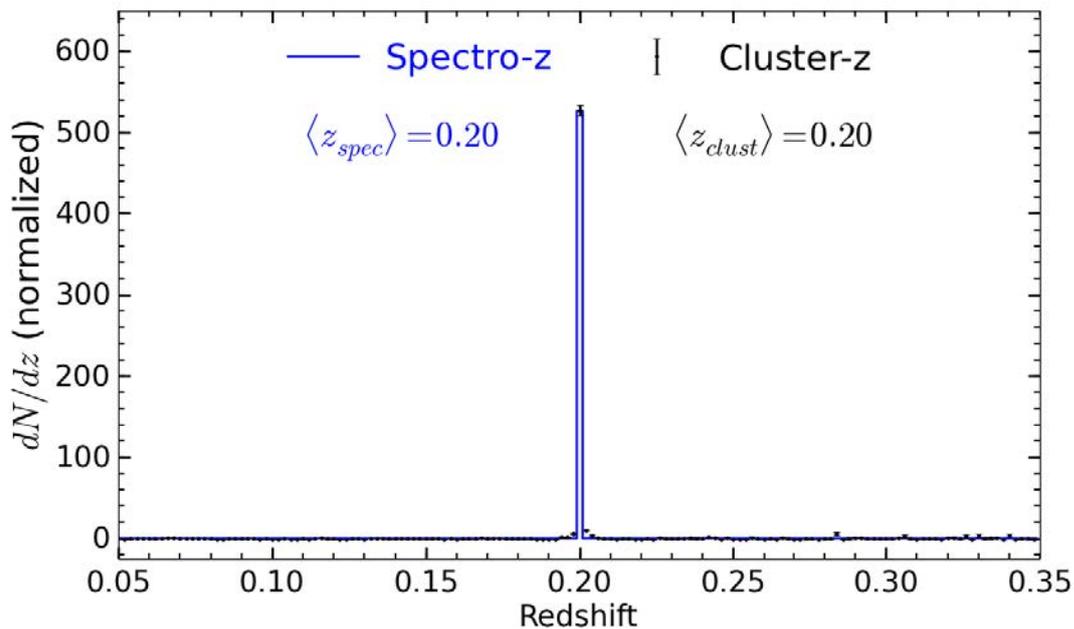


FIGURE 4.4 – Distributions en redshift de l'un des 400 sous-échantillons *inconnus* de largeur  $\Delta z_u = 8.10^{-4}$ . Pour des raisons de clarté, j'ai choisi de montrer la distribution avec  $\delta z_r = 2.10^{-3}$  sur la figure au lieu de  $2.10^{-4}$ .

Cette section présente le résultat de la mesure des *cluster-z* entre l'échantillon de *référence* et les 400 sous-ensembles *inconnus*. La Figure 4.4 montre la distribution en redshift que j'obtiens pour un seul de ces sous-ensembles. La distribution en redshift de l'échantillon *inconnu* est correctement retrouvée alors que le reste des mesures est bien en accord avec zéro.

J'ai alors effectué la même mesure pour chacun des 400 sous-ensembles *inconnus*. Le résultat de l'ensemble de ces mesures est visible sur la Figure 4.5. Elle rassemble les  $400 \times 1\,600 = 640\,000$  mesures de clustering effectuées dans cette analyse et les présente sous la forme d'un plan  $(z_{clust}; z_{spec})$  dans lequel la valeur en chaque point est donnée par un code couleur. Chaque colonne correspond à une distribution en redshift pour laquelle  $\delta z_r = 2.10^{-4}$  et chaque pixel de cette carte correspond à une

mesure. Pour des raisons évidentes de clarté les barres d'erreur de ces points sont omises. La distribution montrée sur la Figure 4.4 correspond à la ligne verticale d'abscisse  $\bar{z}_{spec} = 0.2$ .

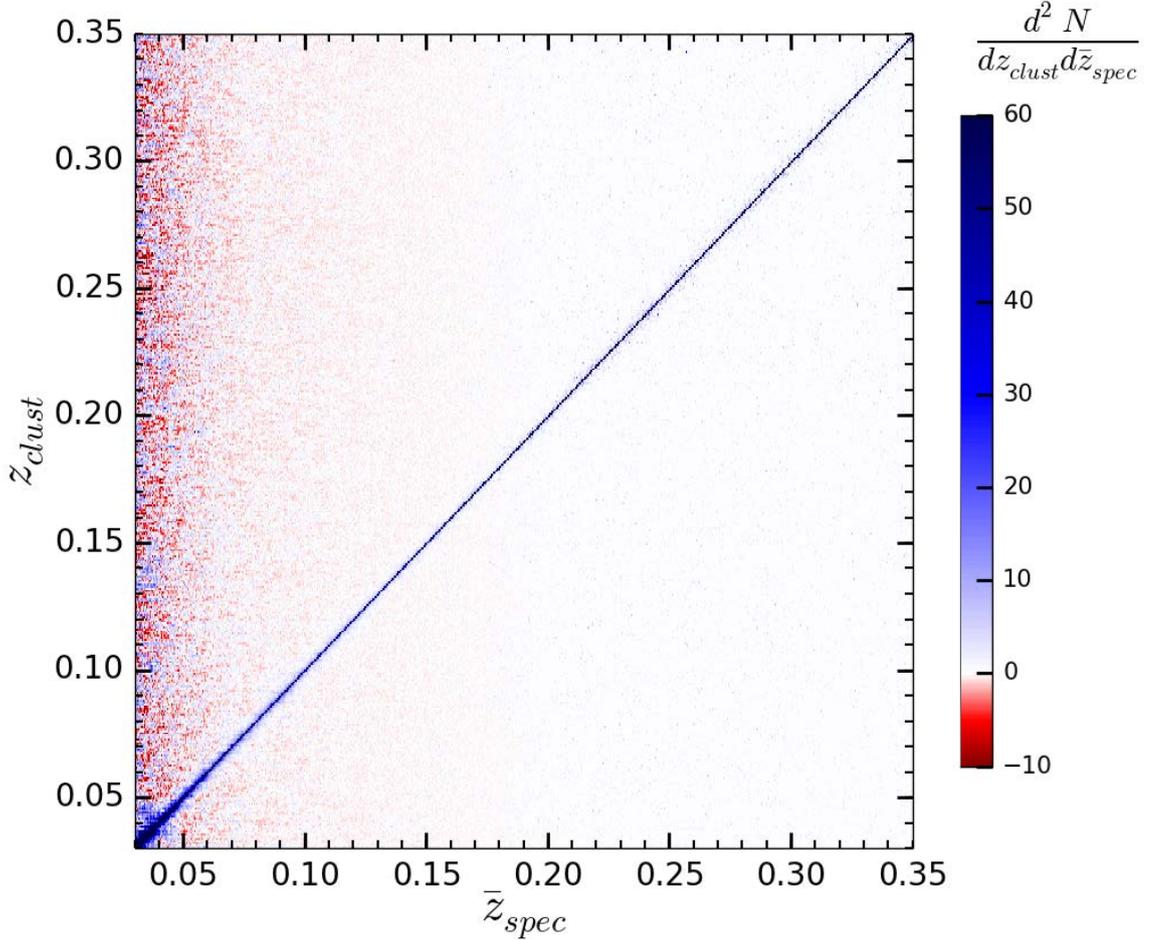


FIGURE 4.5 – Carte de densité montrant les 400 distributions en *cluster-z* de largeur :  $\Delta z_u = 8.10^{-4}$ . La largeur des bins de *référence* est de  $\delta z_r = 2.10^{-4}$ . L'accord avec les redshifts spectroscopiques est remarquable. On note toutefois un niveau de bruit plus élevé pour  $z < 0.07$  provenant à la fois de la variance cosmique et d'une légère surestimation de  $\Sigma_R$  ce qui explique la surdensité de valeurs négatives. On peut également noter la présence d'une large structure autour de  $z = 0.04$ .

La Figure 4.5 montre un accord remarquable avec les redshifts spectroscopiques le long de la droite  $z_{clust} = z_{spec}$ . Le niveau de bruit plus élevé pour  $z < 0.07$  provient à la fois d'une légère surestimation de  $\Sigma_R$  et de l'amplitude de la variace cosmique. La section suivante discute les erreurs systématiques de cette mesure de façon plus détaillée. Les résultats que j'obtiens sont en accord avec [Rahman et collab. \(2015b\)](#)

dont l'étude a été réalisée simultanément à ce travail. Ce graphique peut être comparé à la Figure 1.17 du premier chapitre qui présente le même type de résultat mais pour le plan  $(z_{phot}; z_{spec})$ .

Pour chacun des 400 sous-ensembles, je calcule le résidu sur la mesure de la valeur centrale,  $\bar{z}_{clust} - \bar{z}_{spec}$ . La distribution de ces résidus est montrée à la Figure 4.6.

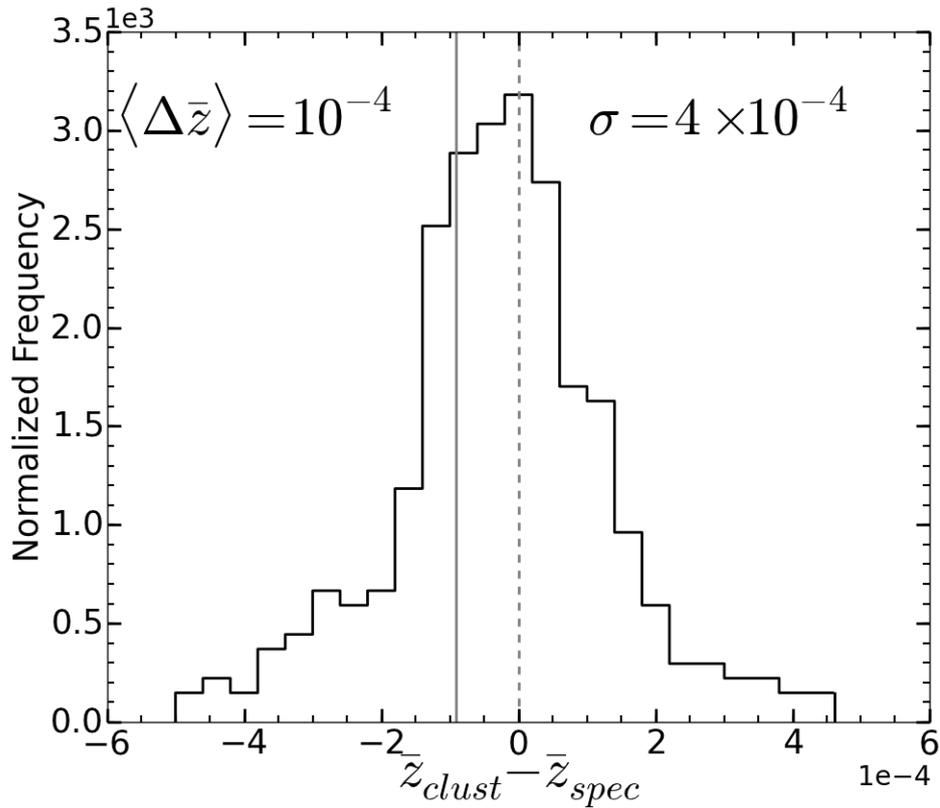


FIGURE 4.6 – Distribution normalisée des résidus  $\Delta \bar{z} = \bar{z}_{clust} - \bar{z}_{spec}$  dérivée de mon analyse. Le trait plein gris montre la moyenne  $\langle \Delta \bar{z} \rangle$  de l’histogramme alors que le trait en pointillés est centré sur zéro.

## 2.2 Étude du bruit visible à $z < 0.15$

À redshift  $z < 0.15$ , le bruit - les fluctuations en dehors de la droite  $z_{clust} = z_{spec}$  - est dominé par une surestimation de  $\Sigma_r$  et par la variance cosmique.

La surestimation de  $\Sigma_r$  explique le nombre plus élevé de valeurs négatives (en rouge) que de valeurs positives (en bleu) dans les régions où l'on n'attend aucun signal. Cette surestimation de  $\Sigma_R$  peut provenir d'une légère surestimation des masques. À mesure que  $\theta_{max}$  diminue la fraction des masques affectant la mesure diminue également. Ainsi, si l'effet affecte l'ensemble des mesures il diminue progressivement avec le redshift.

Dans notre cas la variance cosmique correspond à des patchs cohérents de fausses corrélations sur le ciel. Ces patchs correspondent à la zone de recouvrement entre deux mesures réalisées autour de deux objets de *référence* différents. Les fausses corrélations se trouvant dans ces régions sont alors amplifiées. À mesure que  $\theta_{max}$  diminue, le taux de recouvrement des régions entourant deux objets de *référence* diminue également ce qui réduit l'effet des fausses corrélations pouvant s'y trouver. Ainsi, l'effet est plus important à bas redshift et diminue lorsque  $z$  augmente. Ces fausses corrélations peuvent être dues à trois effets :

- un biais causé par le sur-échantillonnage par les objets de *référence* dans certaines régions du ciel où se trouvent des grandes structures. Une région dense du ciel (amas, super amas) possédant plus d'objets de *référence*, on y fera donc plus de mesure ;
- une ambiguïté sur le redshift cosmologique due aux vitesses particulières des galaxies ;
- une superposition fortuite le long de la ligne de visée de deux grandes structures localisées à deux redshifts différents.

**L'amplitude du premier effet** diminue avec le redshift puisque l'homogénéité du ciel augmente. On peut cependant réduire cet effet en procédant à une étape de nettoyage de la population de *référence* afin de la rendre homogène à la fois sur le ciel et en redshift. Cette étape consiste à diviser le ciel en sous-régions, elles-mêmes triées en intervalles de redshifts de largeur  $10^{-3}$ , puis à mesurer la densité locale d'objets de *référence* pour chacun d'eux. Seuls les intervalles dont la densité se trouve à moins de  $2\sigma$  de la valeur moyenne sont conservés. Pour chaque intervalle au-delà de cette limite on retire tous les objets de *référence* se trouvant dans  $\Delta z = \pm 1.5 \times 10^{-3}$ . La fraction de régions masquées est de 30% à  $z = 0.03$  et diminue à moins de 5% à  $z = 0.3$ . L'échantillon de *référence* homogénéisé résultant se compose d'environ 440 000 objets.

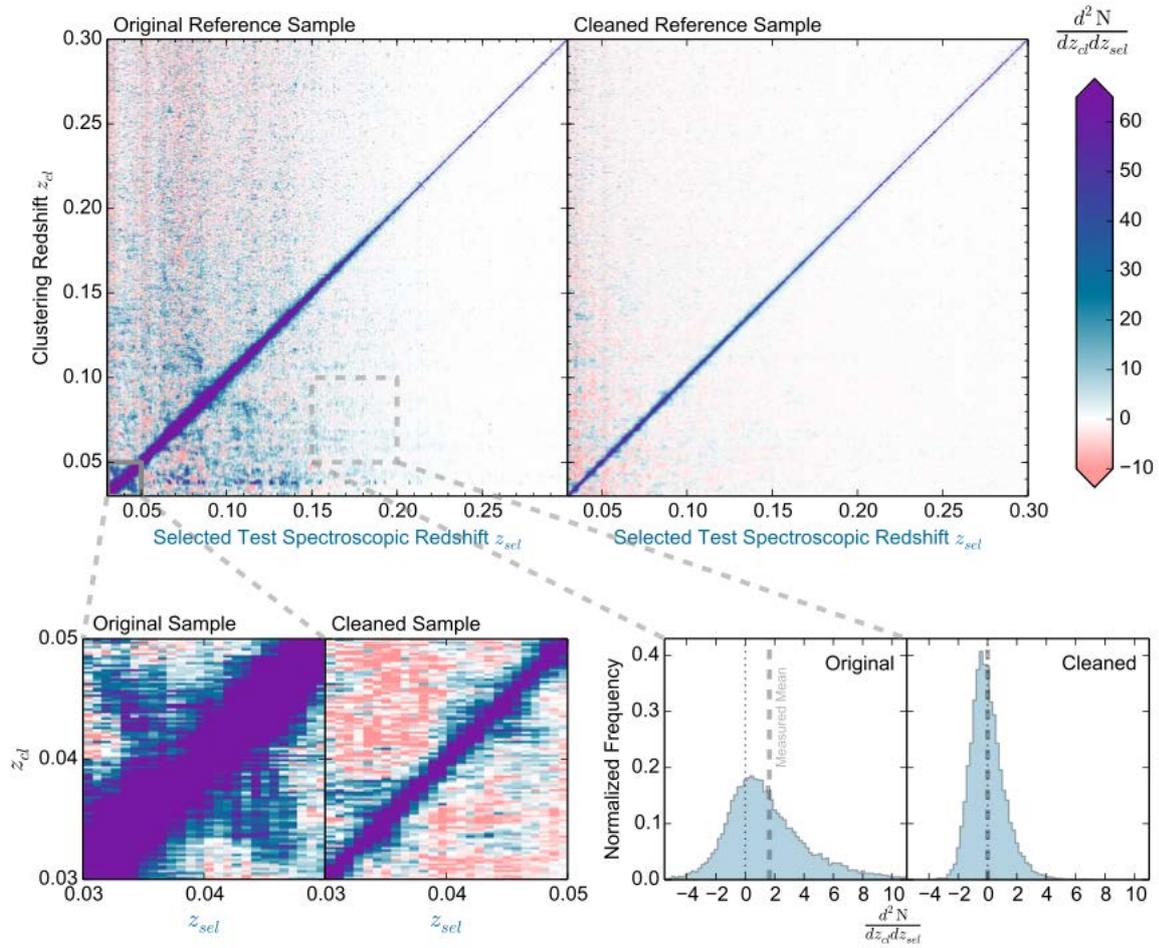


FIGURE 4.7 – Comparaison entre les cartes de densité obtenues avant et après homogénéisation de la population de *référence*. L’insert de droite montre la diminution du niveau de bruit dans une région où aucun signal n’est attendu. L’insert de gauche montre la disparition de la structure se trouvant autour de  $z = 0.4$ . Figure extraite de [Rahman et collab. \(2015b\)](#).

La Figure 4.7 montre les distributions en redshifts mesurées avec et sans l’étape d’homogénéisation. L’insert de droite montre la diminution du niveau de bruit dans une région où aucun signal n’est attendu. L’insert de gauche montre la disparition de la structure se trouvant autour de  $z = 0.04$ . Cette structure correspond au super amas d’Hercule, le plus gros super amas connu de l’univers local ([Barmby et Huchra, 1998](#)), situé entre  $z \sim 0.0304$  et  $z \sim 0.0414$  ([Kopylova et Kopylov, 2013](#)). La présence de ce super amas induit un sur-échantillonnage de la région du ciel correspondante, ce qui se traduit par une corrélation plus forte avec les sous-ensembles *inconnus* qui contiennent des surdensités dans cette même région.

**Le second effet** ne peut pas être pris en compte facilement et induit de fausses corrélations tangentielles à la droite  $z_{clust} = z_{spec}$ .

La vitesse radiale d'un objet extragalactique provient à la fois de l'entraînement de Hubble et de sa vitesse particulière. Puisque seul le redshift cosmologique est corrélé avec la distance, les régimes dans lesquels les vitesses particulières sont importantes par rapport au redshift cosmologique créent de faux signaux dans l'estimation du  $dN/dz$ .

Les objets se rapprochant de nous (dont la vitesse est négative dans la ligne de visée) auront un *cluster-z* légèrement plus grand que leur redshift spectroscopique. Leur redshift sera surestimé. Inversement, des objets s'éloignant de nous (dont la vitesse est positive dans la ligne de visée) auront un *cluster-z* légèrement plus faible que le redshift spectroscopique. Leur redshift sera sous-estimé.

Cet effet est encore plus fort lorsqu'un amas de galaxies très massif est présent. Pour un amas de galaxies possédant  $\Delta v = 10^3 \text{ km.s}^{-1}$ , cela correspond à un décalage sur le redshift de l'ordre de  $\Delta z \sim 10^{-3}$ . La contribution des vitesses particulières affecte notre capacité à mesurer le redshift cosmologique et se traduit par un signal perpendiculaire et symétrique par rapport à la droite  $z_{clust} = z_{spec}$ .

**La superposition fortuite** de grandes structures se trouvant à deux redshifts différents produit une corrélation artificielle qui apparaît symétrique autour de la droite  $z_{clust} = z_{spec}$  puisque des objets de *référence* se trouvent dans chacune des structures.

Ces trois effets expliquent pratiquement tous les effets visibles sur la Figure 4.5. L'origine des fausses corrélations étant principalement due à l'inhomogénéité de la population de *référence*, on peut fortement réduire le niveau du bruit en ajoutant une étape d'homogénéisation de l'échantillon de *référence* avant la mesure de corrélation croisée comme l'illustre la Figure 4.7. Rappelons que ces effets diminuent avec le redshift.

### 3 Synthèse

Cette approche tomographique dans le cas idéal m'a permis d'apprécier tout le potentiel de la mesure de redshift via l'agrégation des galaxies. En effet j'ai montré qu'il est possible d'obtenir des mesures très précises - de l'ordre de  $10^{-4}$  sur le redshift moyen - et cela sans faire intervenir des millions d'objets. De plus on peut remarquer l'absence de redshifts catastrophiques ce qui peut être très intéressant dans le cadre d'une analyse conjointe *cluster/photo-z*.

J'ai également présenté et discuté différents effets systématiques produisant de fausses corrélations. Ces effets sont principalement dus à l'inhomogénéité de la po-

pulation de *référence* et peuvent être grandement diminué en homogénéisant cette population à la fois sur le ciel et en redshift. Cette étape n'est cependant nécessaire que pour des études à bas redshift.

De plus, le fait d'utiliser le redshift spectroscopique des objets de la population *inconnue* lors du prétraitement des données n'est pas aussi éloigné d'un cas réel que ce que l'on pourrait penser. En effet il est possible d'adopter la même approche tomographique mais en utilisant des redshifts photométriques. Les *photo-z* et les *cluster-z* peuvent être utilisés de façon complémentaire puisqu'ils ne sont a priori pas soumis aux mêmes systématiques. C'est ce que nous verrons dans le prochain chapitre.

## Real case study : abstract

In this chapter I explore the *clustering redshift* accuracy in *CFHTLS & VIPERS* data in a real case scenario. I subsampled the *unknown* population using both a tomographic approach - based on *photo-z* - and colors criteria.

Using the tomographic approach I show that this method allows to measure the redshift distribution of *unknown* population fainter than the *reference* sample, Figure 5.12.

Then, subdividing the color space - at same magnitude limit - I show that *clustering redshift* are almost 20 times better than *photo-z* on the mean redshift recovery, Figure 4.8.

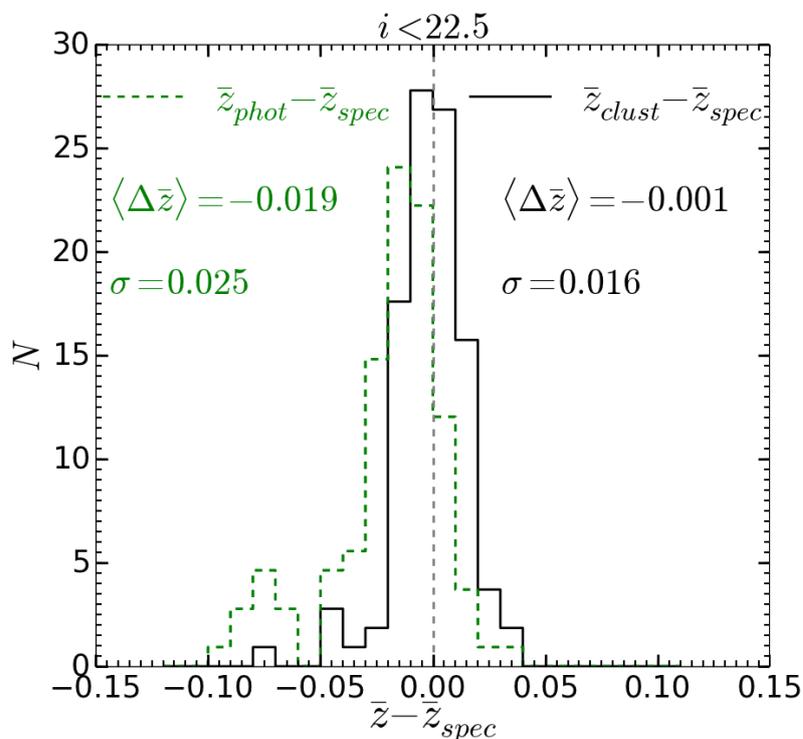


FIGURE 4.8 – Distributions of the residues for *cluster-z* (black) and *photo-z* (green). This plot shows that *cluster-z* are 20 times better than *photo-z* on the mean redshift recovery.

Finally, I use the last two points to explore the redshift distribution of *CFHTLS* sources beyond  $i_{AB} = 22.5$  and give their full redshift distribution, Figure 5.22.



# Chapitre 5

## Clustering redshift : cas réel

---

Dans ce chapitre, je présente une application originale très concrète des redshifts clustering que j'ai conduite sur des données d'observation qui me placent dans une situation réelle. L'analyse précédente utilisait le redshift spectroscopique des objets de l'échantillon *inconnu* lors d'une étape de présélection. Cette présélection ne remet pas en cause la validité de cette approche mais ne correspond pas à un cas réaliste de relevé cosmologique.

Le but de ce chapitre est de montrer que l'on peut se passer de l'information spectroscopique tout en continuant de réduire l'influence de l'évolution du biais galaxie-matière de la population *inconnue*. Pour cela, deux solutions sont envisageable :

- utiliser les redshifts photométriques. En effet ces derniers seront disponibles pour certains grands relevés qui disposent d'un grand nombre d'information photométrique multi-couleurs.
- faire des sélections dans l'espace des couleurs afin de construire des sous-échantillons de la population *inconnue*. Par construction, ces sous-échantillons auront des distributions en redshifts moins larges que l'échantillon initial.

J'explore ces deux approches dans ce chapitre qui fera l'objet d'une publication (Scottéz et collab. (in prep), voir Annexe B). Nous verrons que les sélections en couleurs présentent l'avantage de s'affranchir des biais systématiques liés à la dispersion des photos-z.

# 1 Les données *CFHTLS* & *VIPERS*

## 1.1 Présentation

### *CFHTLS*

Le *Canada-France-Hawaii Telescope Legacy Survey* (*CFHTLS*) est un projet franco-canadien composé de trois sondages photométriques de profondeur et de taille différentes. Mon étude utilise la version finale - T0007 - de ce relevé. Cette version des données provient d'observations réalisées entre le 26 Mai 2003 et le 2 Février 2009 à l'aide du télescope de 3.6 m (*CFHT*) équipé de la caméra MegaCam ([Boulade et collab., 2000](#)) montée au foyer primaire MegaPrime du CFHT.

MegaCam est constituée d'une mosaïque de 36 détecteurs CCD EEV, disposées sur quatre lignes et neuf colonnes. Chaque CCD possède  $2\,080 \times 4\,644$  pixels dont l'échelle est de  $0.186''/\text{pixel}$ , donnant ainsi une couverture angulaire totale de  $57.6 \times 58.8$  arcmin<sup>2</sup>, couvrant la totalité du plan focal utile disponible.

Les données *CFHTLS-T0007* comprennent deux parties :

- *CFHTLS Deep*, composé de quatre pointés ultra-profonds (D1,D2,D3,D4) et indépendants du ciel noir extragalactique. Chacun de ces pointés couvre une surface de  $1 \text{ deg}^2$  et atteint une limite de complétude à 80% en :  $u_{AB}^* = 26.3, g_{AB} = 26.0, r_{AB} = 25.6, i_{AB} = 25.4, z_{AB} = 25.0$ .
- *CFHTLS Wide*, composé de 171 pointés profonds du ciel extragalactique qui représentent au final une surface de  $\sim 155 \text{ deg}^2$  répartie sur quatre champs (W1,W2,W3,W4). Les limites de complétude à 80% sont :  $u_{AB}^* = 25.2, g_{AB} = 25.5, r_{AB} = 25.0, i_{AB} = 24.8, z_{AB} = 23.9$ .

La Figure 5.1 montre les courbes de transmission pour les filtres utilisés ; la position de ces champs dans le ciel est visible sur la Figure 5.2. Pour plus d'informations voir : <http://terapix.iap.fr/cplt/T0007/doc/T0007-doc.pdf> .

Dans l'analyse présentée dans ce chapitre je me concentre uniquement sur le champ *CFHTLS* W1 dont la surface est de  $8 \times 9 = 72 \text{ deg}^2$ . Les redshifts photométriques disponibles pour les objets de ce champ sont obtenus via la méthode présentée dans : [Ilbert et collab. \(2006\)](#) ; [Coupon et collab. \(2009\)](#). Le code Le Phare ([Arnouts et collab., 1999, 2002](#)) qui y est utilisé est basé sur l'ajustement de modèle.

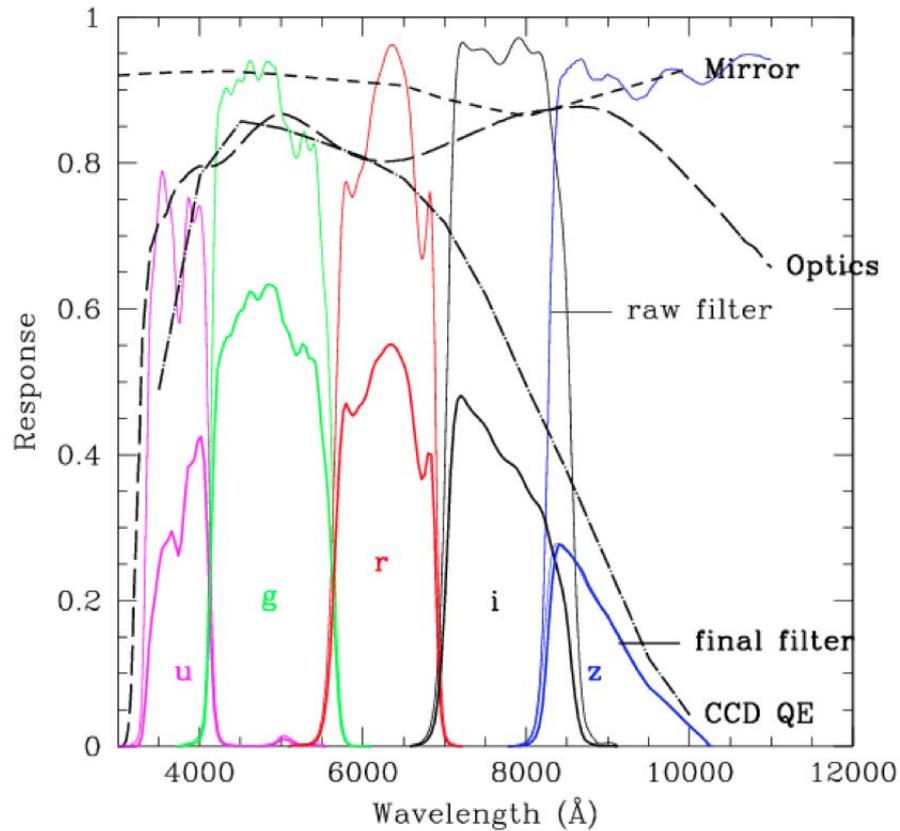


FIGURE 5.1 – Courbes de transmission des filtres utilisés pour le relevé *CFHTLS*. La figure montre la réponse du miroir, de l’optique, l’efficacité quantique des CCDs de la caméra et enfin les courbes de réponses effectives des filtres. Source : site web du *CFHTLS*.

La procédure suivie pour l’obtention des redshifts photométriques disponibles dans le catalogue que j’utilise est ici brièvement résumée :

Les modèles sont décalés vers le rouge et intégrés à travers la courbe de transmission de l’instrument. L’opacité du milieu intragalactique est prise en compte et une extinction interne peut être ajoutée en tant que paramètre libre pour chaque galaxie. Les redshifts photométriques sont obtenus en comparant les flux provenant des modèles et les flux observés via une fonction de mérite de type  $\chi^2$ . L’analyse retourne aussi une PDF - fonction de distribution de probabilité - associée à chaque redshift photométrique.

Le premier échantillon de modèles utilisés correspond aux quatre spectres observés (Ell,Sbc,Scd,Irr) de Coleman et collab. (1980). À ces quatre spectres s’ajoutent deux distributions spectrales d’énergie (SED) pour des galaxies starburst également

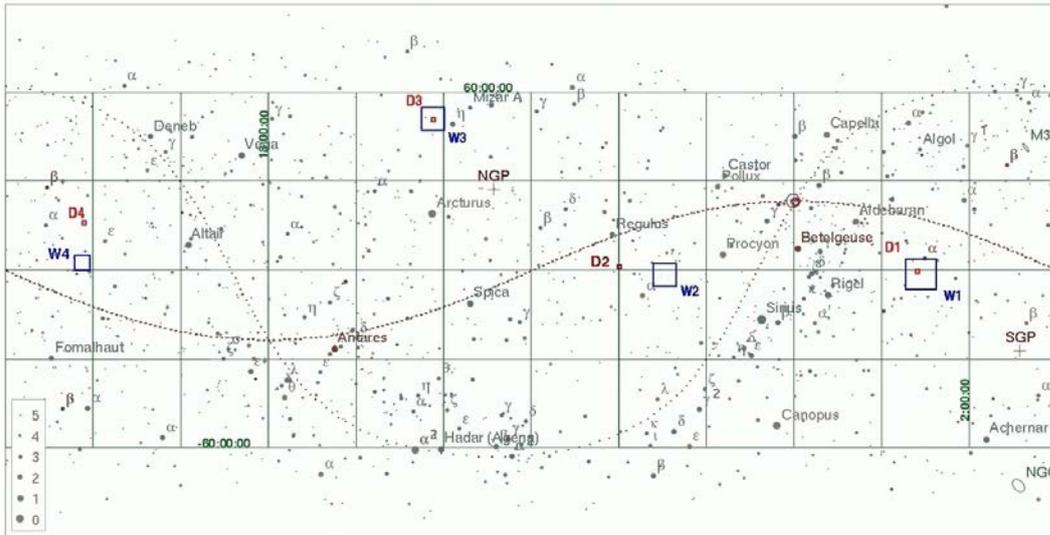


FIGURE 5.2 – Position des champs Deep et Wide du relevé *CFHTLS*. Source : site web du *CFHTLS*.

observées et provenant de [Kinney et collab. \(1996\)](#). Ces modèles ont ensuite été optimisés à partir de l'échantillon spectroscopique profond *VVDS*.

Un étalonnage automatique du point zéro est réalisée en utilisant les redshifts spectroscopiques présents dans les champs *CFHTLS* W1, W2 et W4. Cette étape consiste à rajouter un faible décalage sur le point zéro de la photométrie de façon uniforme à travers l'échantillon.

Pour les types spectraux de galaxie plus tardifs que Sbc, un rougissement  $E(B - V) = 0 - 0.35$  est introduit suivant la loi d'extinction de [Calzetti et collab. \(2000\)](#). Lorsqu'il n'y a pas de données disponibles dans le proche IR, un a priori sur la distribution en redshift est utilisé et l'inférence du redshift est conduite selon une approche bayésienne.

Aucune solution en redshift n'est admise si elle produit une galaxie de luminosité plus brillante que  $M(G) = -24$ . Plutôt que d'utiliser directement le redshift qui minimise la fonction de distribution du  $\chi^2$ , la médiane de la PDF est utilisée. Cela limite le risque d'une solution préférée sur un faible intervalle de redshift. La séparation étoile-galaxie est réalisée à partir d'une sélection jointe prenant en compte la compacité des objets et le modèle le mieux ajusté.

Finalement à partir du redshift photométrique, du modèle le mieux ajusté et des magnitudes apparentes observées dans une bande donnée, on peut directement en déduire la correction-k ainsi que la magnitude absolue dans n'importe quelle bande.

Sur la Figure 5.3 je montre la comparaison des *photo-z* du *CFHTLS* aux redshifts spectroscopiques issus de *VIPERS*. Pour plus de détails voir : <http://www.cfht.>

[hawaii.edu/~arnouts/LEPHARE/DOWNLOAD/lephare\\_doc.pdf](http://hawaii.edu/~arnouts/LEPHARE/DOWNLOAD/lephare_doc.pdf) .

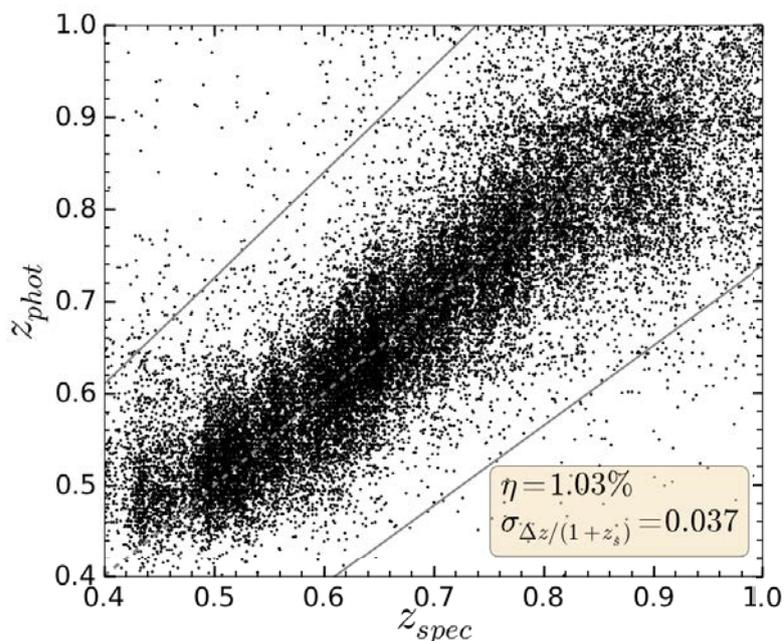


FIGURE 5.3 – Précision des redshifts photométriques du *CFHTLS-T0007* dans le champ W1 pour environ 26 000 objets communs avec *VIPERS*. J’ai vérifié ces redshifts à l’aide des données *VIPERS* qui fournissent un échantillon spectroscopique plus important que celui utilisé lors de la release *T0007* du *CFHTLS*. La dispersion est définie comme étant  $\sigma_{\Delta z / (1 + z_s)} = 1.48 \times \text{median}(|\Delta z| / (1 + z_s))$  et le taux d’erreurs catastrophiques  $\eta$  correspond aux objets dont  $|\Delta z| \geq 0.15 \times (1 + z_s)$ .

## *VIPERS*

*VIPERS* pour *VIMOS Public Extragalactic Redshift Survey*, est un relevé spectroscopique profond en cours de réalisation (Guzzo et collab., 2014) dont le but est de cartographier en détail la distribution spatiale des galaxies. Le relevé se compose de deux champs disjoints, sélectionnés dans les champs W1 et W4 du *CFHTLS*. La surface totale couverte par *VIPERS* est de 24 deg<sup>2</sup>.

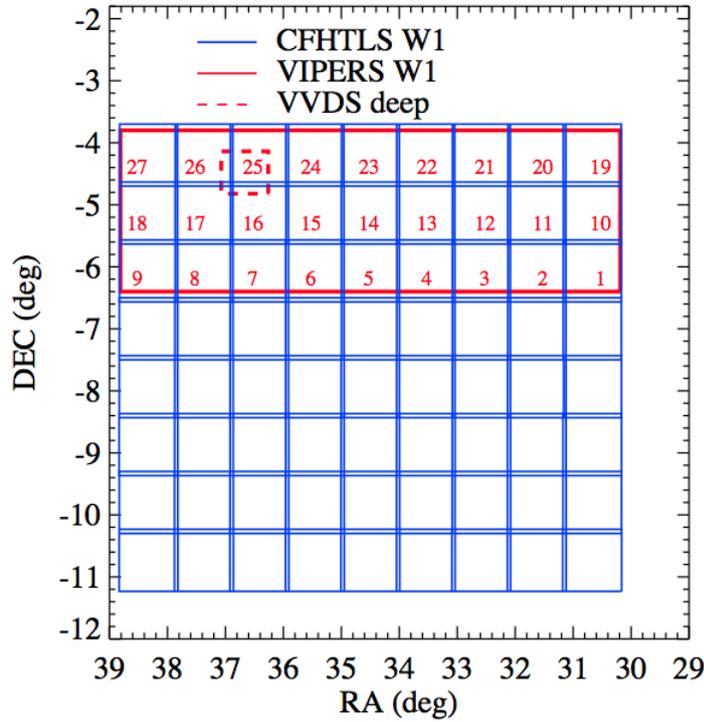


FIGURE 5.4 – Recouvrement entre les champs W1 de *VIPERS*, du *CFHTLS* et du *VVDS Deep*, Le Fèvre et collab. (2005). Figure extraite de Guzzo et collab. (2014).

La Figure 5.4 montre le recouvrement entre les champs W1 de *VIPERS* et du *CFHTLS*. Les spectres *VIPERS* sont le résultat d'environ 440h d'observation au *VLT - Very Large Telescope* - au Chili. La position des deux champs de *VIPERS* a été choisie de façon à tirer profit de la grande quantité de données photométriques collectées par le *CFHTLS*. Pour l'échantillon *VIPERS* une présélection en couleur, selon l'équation (5.1), évite de sélectionner les galaxies à bas redshift - en dessous de 0.4 - tout en conservant les objets à redshift intermédiaire  $z \sim 0.7$ .

$$(r - i) > 0.5(u - g) \text{ OU } (r - i) > 0.7 . \quad (5.1)$$

L'évolution de la complétude des données *VIPERS* dans l'intervalle de redshift correspondant à la sélection en couleurs est montrée sur la Figure 5.5. À terme l'équipe de *VIPERS* fournira un catalogue d'environ 100 000 sources disposant d'un redshift spectroscopique et d'informations spectrales, de magnitude limite  $i_{AB} < 22.5$ , et dans l'intervalle de redshift  $0.5 < z < 1.0$ .

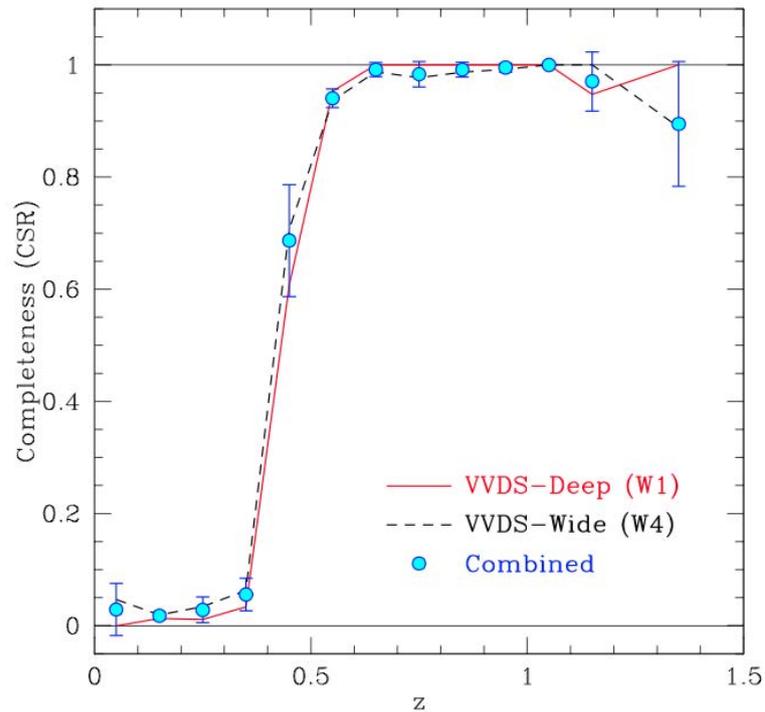


FIGURE 5.5 – Vérification de la complétude de la sélection en couleur utilisée pour le relevé *VIPERS* en fonction du redshift. Figure extraite de Guzzo et collab. (2014).

L'erreur sur la mesure du redshift est obtenue à partir des galaxies pour lesquelles une double mesure existe. Ces doubles mesures, réalisées par des pointés indépendants, concernent par exemple les objets se trouvant sur les bords du quadrant dans la région où deux quadrants contigus se superposent. Elles sont alors utilisées pour estimer la variance interne de l'erreur sur le redshift des galaxies *VIPERS*.

Le panneau du bas de la Figure 5.6 montre la distribution de la différence entre ces doubles mesures. Après avoir normalisé les redshifts par le facteur d'expansion  $1 + z$ , la distribution de ces mesures est en accord avec une gaussienne de largeur  $\sigma_2 = 200 \text{ km.s}^{-1}$ . Cela correspond à une erreur à  $1\sigma$  pour chaque objet de  $\sigma_v = \sigma_2/\sqrt{2} = 141 \text{ km.s}^{-1}$ , soit, dans l'espace des redshifts, une déviation standard sur la mesure du redshift de  $\sigma_z = 0.00047(1 + z)$ .

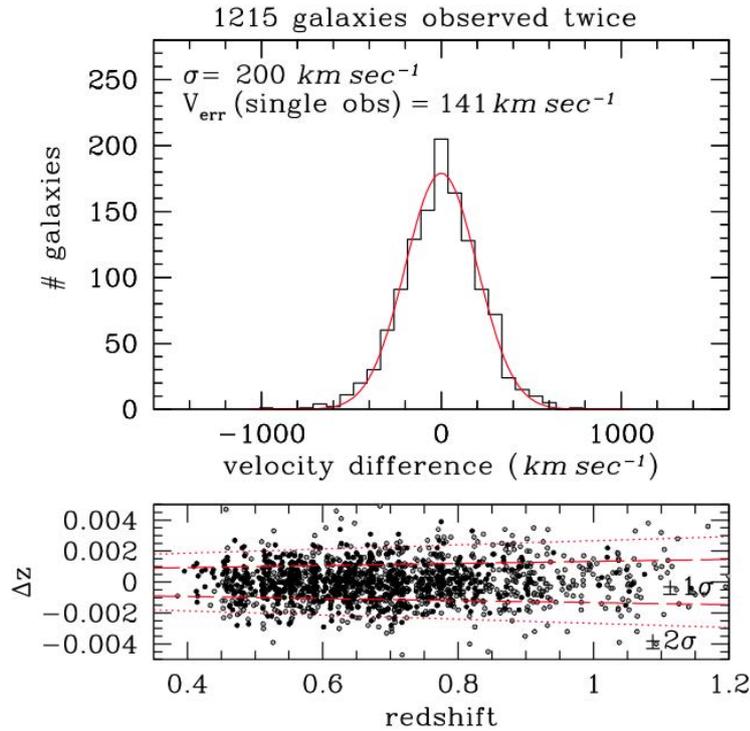


FIGURE 5.6 – Haut : distribution de la différence  $\Delta v = c\Delta z/(1+z)$ .

Bas : la même information mais ramenée dans l'espace des redshifts. Ces mesures donnent une erreur sur le redshift individuel des galaxies de  $\sigma_z = 0.00047(1+z)$ . Figure extraite de Guzzo et collab. (2014).

## 1.2 Sélection

*CFHTLS* : population *inconnue*.

Pour construire la population *inconnue* de mon étude, j'ai sélectionné les données *CFHTLS* de telle sorte que :

- les objets soient des galaxies ;
- les objets soient dans la même région du ciel que *VIPERS* ;
- les objets soient en dehors des masques *CFHTLS* ;
- les objets disposent d'un redshift photométrique compris dans l'intervalle  $0.4 < z_{\text{phot}} < 1.0$ .

Cette dernière étape me permet d'éviter de fausser la normalisation de la distribution en redshift. En effet puisque les données *VIPERS* ne couvrent que la gamme en redshift  $0.4 < z < 1.0$  la corrélation croisée n'est mesurable que sur cet intervalle. Un objet de l'échantillon *inconnu* se trouvant en dehors de ce domaine n'apporte aucun signal alors qu'il pourrait néanmoins être comptabilisé dans le nombre total

d'objets de l'équation (2.13), ce qui biaiserait la reconstruction.

**VIPERS** : population de *référence*.

J'ai sélectionné les données *VIPERS* de telle sorte que :

- le redshift spectroscopique de chaque objet soit mesuré avec un niveau de confiance supérieur à 95% ;
- les objets soient en dehors des masques *CFHTLS* ;
- les objets aient un  $z$ -*spec* compris dans l'intervalle  $0.4 < z < 1.0$ .

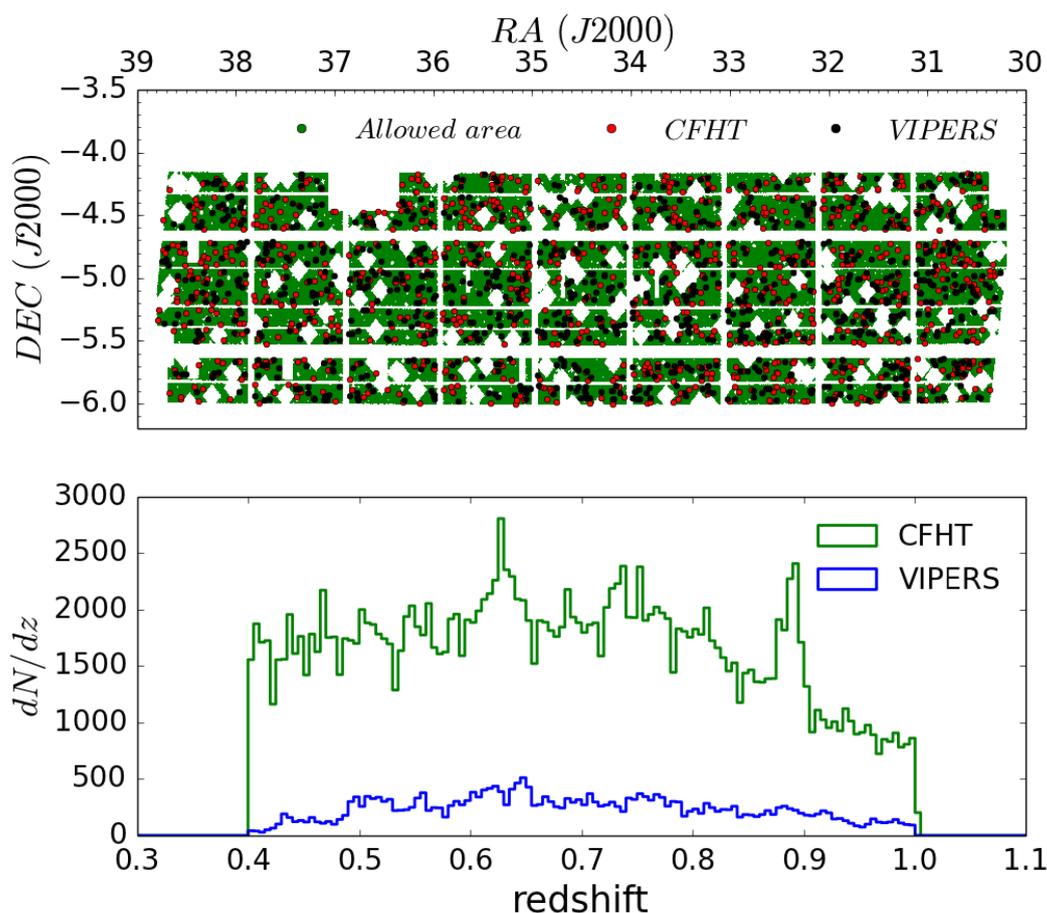


FIGURE 5.7 – Haut : position sur le ciel d'un sous-échantillon aléatoire de 1 000 objets des données *VIPERS* et *CFHT*. La surface en vert correspond aux régions non masquées et s'étend sur  $8.9 \text{ deg}^2$

Bas : distributions en redshift des deux séries de données après sélection. En bleu il s'agit du redshift spectroscopique pour les données *VIPERS* alors qu'en vert il s'agit du redshift photométrique des données *CFHT*.

L'échantillon de *référence* contient  $N_r \sim 30\,000$  avec  $i_{AB} < 22.5$ . La Figure 5.7 montre la superposition sur le ciel et en redshift des deux populations *inconnue* et *référence*.

## 2 Présélection tomographique

Dans cette section je m'écarte du cas idéal du chapitre précédent en substituant l'information spectroscopique par l'information photométrique. Il est important de noter que cette approche est tout à fait compatible avec une étude *réelle* puisque les relevés de galaxies fournissent généralement les redshifts photométriques. J'ai divisé la population *inconnue* en une quarantaine de sous-échantillons distribués dans tout le domaine de redshift photométrique disponible ( $0.4 < z < 1.0$ ). Chacun de ces sous-échantillons est réalisé de façon à contenir un nombre identique d'objets ce qui me permet de travailler à rapport signal sur bruit constant pour tous.

J'ai ensuite mesuré la corrélation croisée entre les deux populations en fonction du redshift pour des échelles physiques de 100 kpc à 2 Mpc. L'amplitude du clustering de la population de *référence* est directement mesurée sur les données et est montrée sur la Figure 5.8.

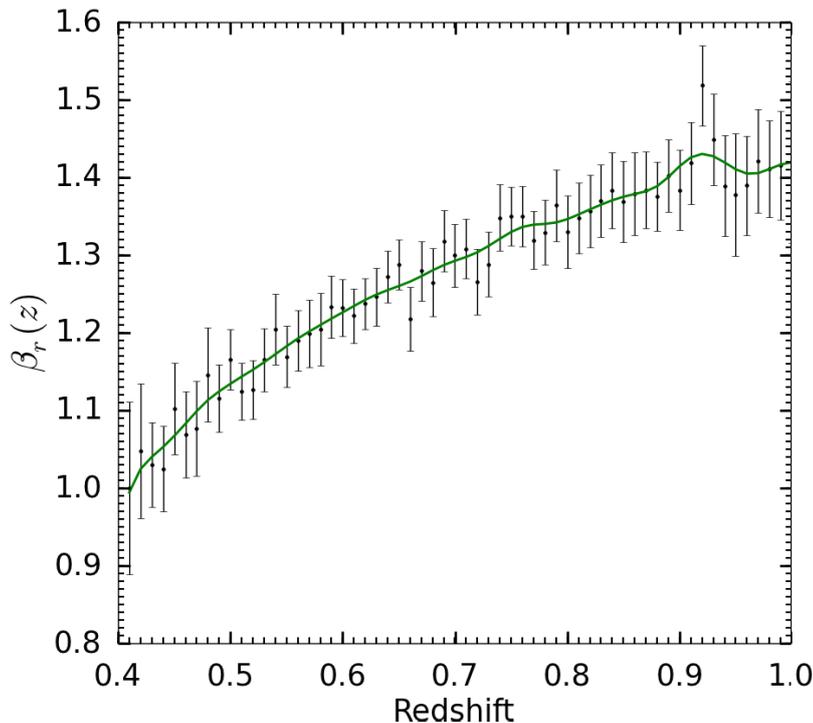


FIGURE 5.8 – Mesure de l'amplitude du clustering de l'échantillon de *référence* que j'obtiens à partir des données *VIPERS*. Les mesures sont normalisées à l'unité en  $z_0 = 0.41$ . La courbe en trait plein correspond à la version lissée des mesures à travers un filtre de Hann d'intervalle de lissage  $\Delta_H z = 0.02$ . C'est cette courbe qui est utilisée lors du calcul de la distribution en redshift.

Dans un premier temps, je me suis intéressé aux objets de la population *inconnue* avec  $i_{AB} < 22.5$  puis, dans un second temps, à ceux compris entre  $22.5 < i_{AB} < 23$ .

## 2.1 Même magnitude limite pour les deux populations

En sélectionnant la population *inconnue* de façon à avoir  $i < 22.5$ , mon échantillon "brillant" contient  $N_u \sim 130\,000$  galaxies. J'ai ensuite choisi de les répartir en 44 sous-échantillons de 3 000 objets chacun de façon à avoir un  $\mathcal{S}/\mathcal{N} \sim 8$  avec  $\delta z = 0.005$ .

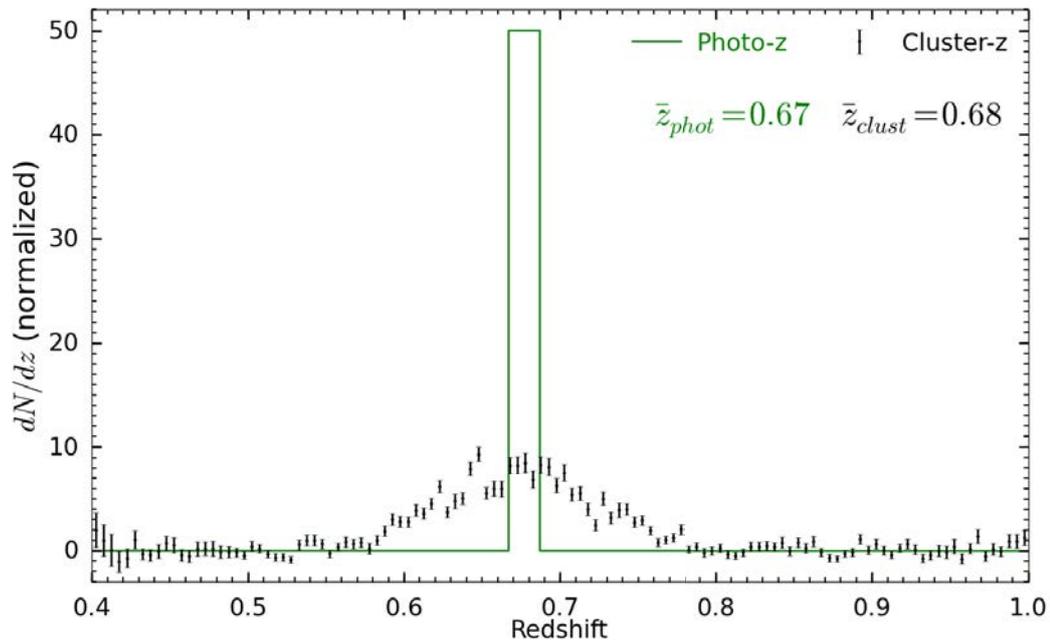


FIGURE 5.9 – Superposition des distributions *photo-z* (vert) et *cluster-z* (noir),  $\delta z = 0.005$ . Les deux mesures sont en accord pour la valeur moyenne mais pas pour la largeur des distributions. Cette différence provient de la dispersion des *photo-z* que je n'ai pas prise en compte lors de la sélection des sous-échantillons.

La Figure 5.9 montre à titre d'exemple une des distributions en redshift-clustering que j'obtiens pour l'un des sous-échantillons ainsi que la distribution en redshift photométrique correspondante. Les deux distributions sont en accord sur la valeur moyenne mais leur forme et leur largeur sont radicalement différentes. C'est un artefact qui provient essentiellement du fait que je n'ai pas pris en compte la dispersion des redshifts photométriques lors de la sélection des sous-échantillons. On peut aussi noter les barres d'erreur, plus élevées pour les mesures des *cluster-z* à bas redshift. C'est un effet du nombre d'objets de *référence* dans chaque bin, qui est plus faible à bas redshift.

J'ai réalisé le même type de mesure que l'exemple précédent en parcourant l'ensemble des sous-échantillons de la population *inconnue*. Ces  $44 \times 120 = 5\,280$  mesures de  $\bar{\omega}_{ur}$  sont combinées à  $\beta_r(z)$  suivant l'équation (4.4) et les 44 distributions en redshifts résultantes sont présentées sur la Figure 5.10. Chaque colonne correspond à une distribution en redshift.

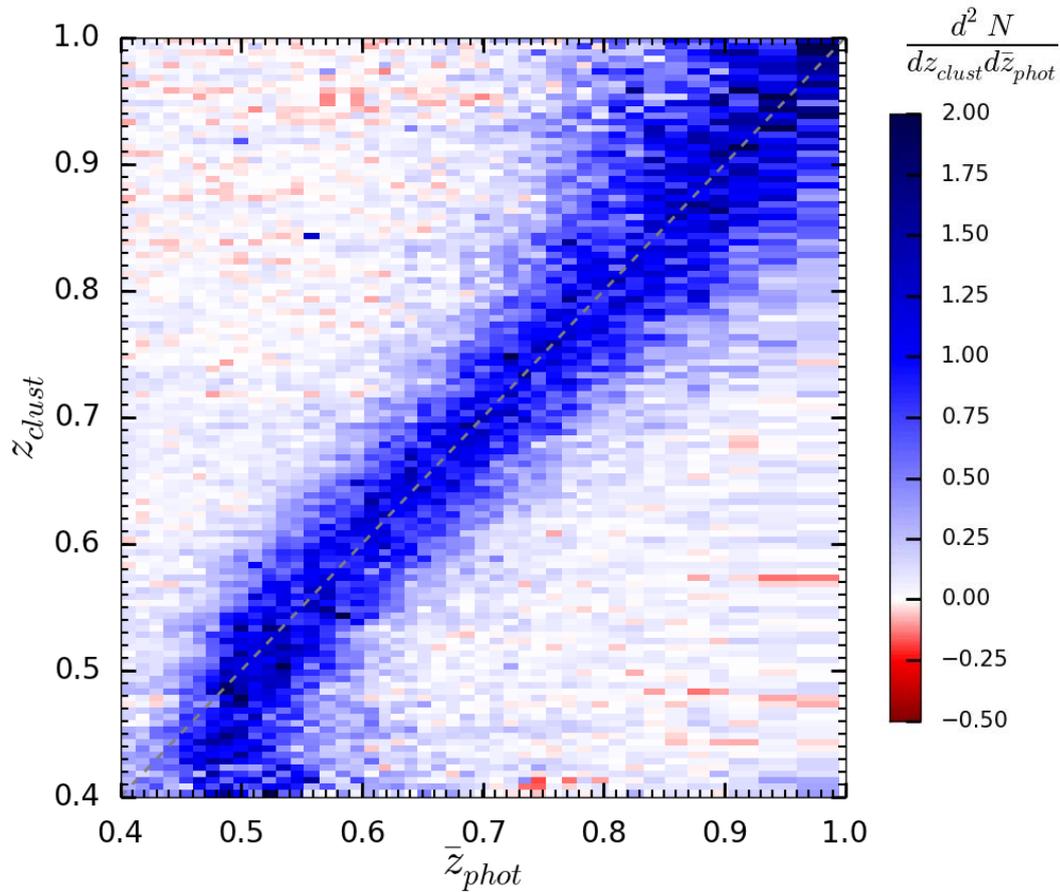


FIGURE 5.10 – Carte de densité regroupant les 5 280 mesures de clustering réalisées dans cette analyse. Contrairement au chapitre précédent, l'utilisation des redshifts photométriques ne permet pas de sélectionner des sous-échantillons très localisés en redshift avec une faible dispersion. Cela se retrouve directement sur la largeur des distributions en *cluster-z* mesurées.

Pour chacune des distributions en clustering redshift je calcule alors la valeur moyenne et je la compare au redshift moyen photométrique. La distribution des résidus est montrée sur la Figure 5.11.

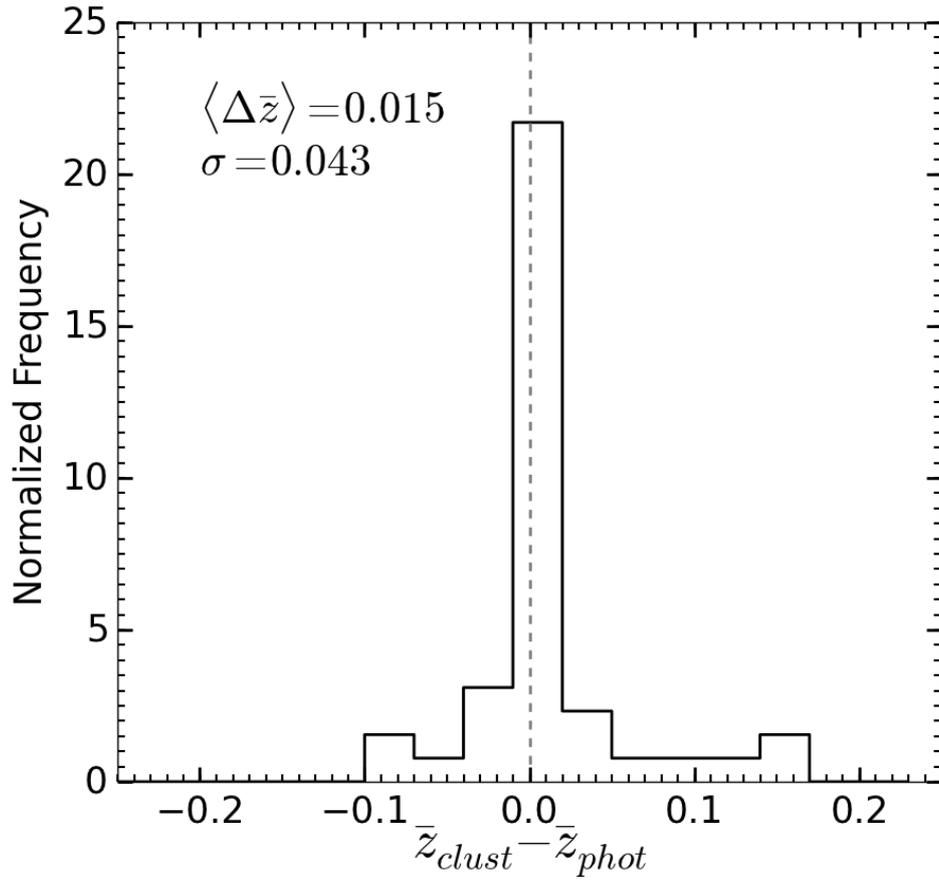


FIGURE 5.11 – Distribution des résidus sur le redshift moyen entre les *cluster-z* et les *photo-z*.

Dans cette étude j'ai démontré la validité de l'approche consistant à présélectionner les données à l'aide des redshifts photométriques et cela jusqu'à  $z = 1$ . La limitation de cette analyse, comparé à celle présentée au Chapitre 4, vient de la largeur des bins utilisés lors de l'étape de présélection c'est-à-dire ici  $\Delta z_u \sim 2 \cdot 10^{-2}$ . Cette limitation est directement liée à l'utilisation des redshifts photométriques. En effet une sélection plus fine en *photo-z* fournirait des sous-échantillons proportionnellement moins localisés en redshift à cause de la dispersion des redshifts photométriques.

## 2.2 Echantillon *inconnu* moins brillant

Le but de cette analyse est d'explorer les performances de la méthode des *cluster-z* lorsque l'échantillon *inconnu* est moins brillant que l'échantillon de *référence*. La population de *référence* est donc la même que précédemment, avec  $i_r < 22.5$ , mais je sélectionne cette fois les objets de la population *inconnue* dont les magnitudes sont comprises dans l'intervalle  $22.5 < i_u < 23$ .

À un redshift donné, les objets moins brillants sont en général moins massifs. Le signal attendu devrait donc être a priori plus faible. À cette source d'incertitude s'ajoute la dispersion des redshifts photométriques qui augmente lorsque les objets sont peu brillants, ce qui affecte directement la capacité des redshifts photométriques à sélectionner des sous-échantillons bien localisés en redshift.

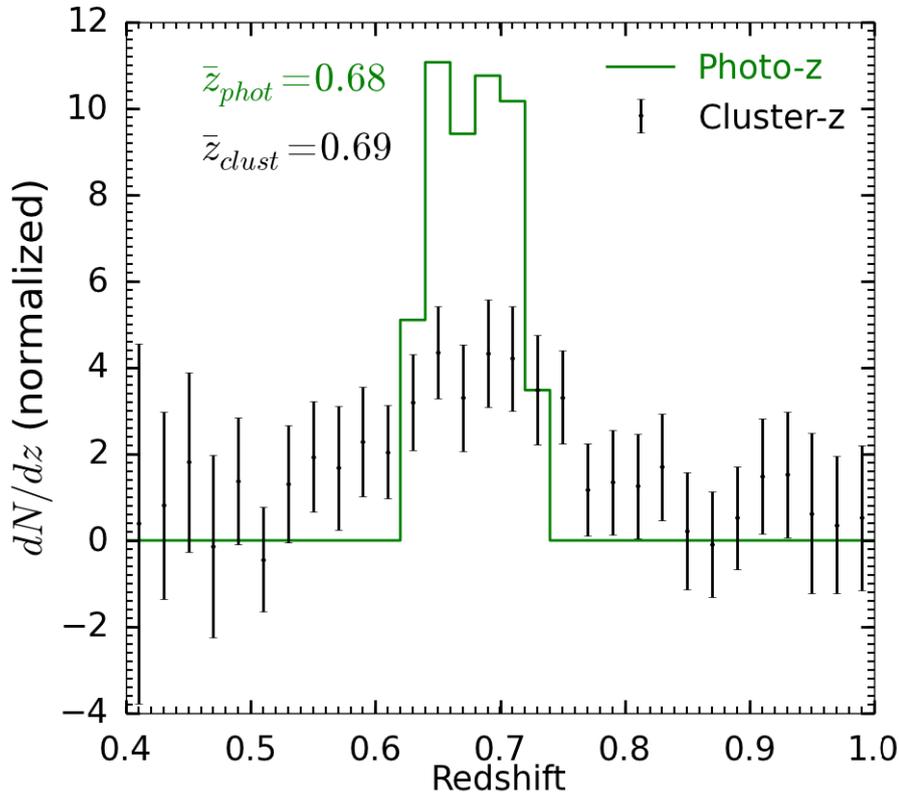


FIGURE 5.12 – Mesure de clustering redshift que j'ai mesuré pour une population *inconnue* se trouvant dans un bin en magnitude moins brillant que la population de *référence*.

Pour préserver la capacité de détection d'un signal, j'ai donc décidé de construire un sous-échantillon contenant  $N_u = 10\,000$  galaxies. La position de ce sous-échantillon en redshift est choisie de façon à avoir la plus faible largeur pour la distribution en *photo-z*. Le résultat est montré sur la Figure 5.12.

Comme attendu la mesure est plus bruitée que précédemment mais le signal est bien présent et parfaitement détecté. Les mesures des autres intervalles en redshift photométrique sont montrés dans à la Figure 5.13.

Cette détection permet d'affirmer que l'estimation de redshifts par mesure de clustering est tout à fait possible pour des populations plus faibles que la magnitude limite de la population de *référence*. Cela signifie également que dans le cadre d'une étude de *cluster-z*, il est envisageable de privilégier un relevé profond plutôt que complet pour la constitution du catalogue d'objets de *référence* ce qui est très intéressant par exemple, en matière de temps d'observation. À titre d'exemple l'utilisation de QSOs peut permettre de constituer un important catalogue d'objets de *référence* et cela jusqu'à  $z \sim 5$ . Pour des relevés comme *Euclid* pour lesquels la constitution d'échantillons spectroscopiques peu brillants et complets est requise - mais difficile à réaliser - afin de calibrer les redshifts photométriques cette propriété est extrêmement intéressante.

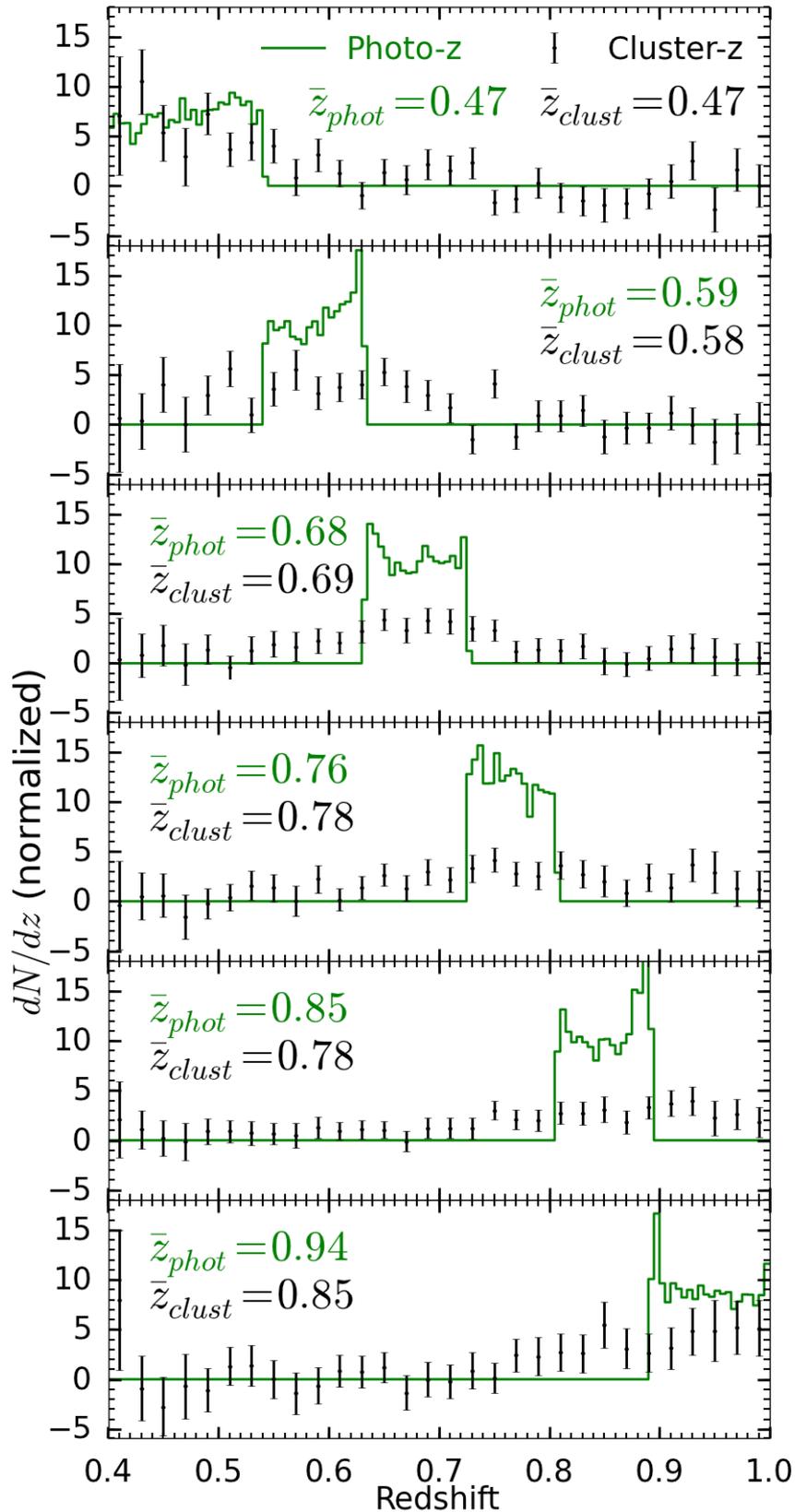


FIGURE 5.13 – Distributions en *clustering redshift* obtenues dans différents bins tomographiques en *photo-z* pour :  $22.5 < i < 23$ . La grande dispersion est due aux systématiques intrinsèques des redshifts photométriques au delà de  $i = 22.5$ , diminuant grandement l'efficacité de l'approche tomographique.

### 3 Présélection dans l'espace des couleurs

Dans cette section j'entreprends de m'affranchir des redshifts photométriques en présélectionnant les objets de l'échantillon *inconnu* en fonction de leurs couleurs. Le principe de cette approche est illustré sur la Figure 5.14.

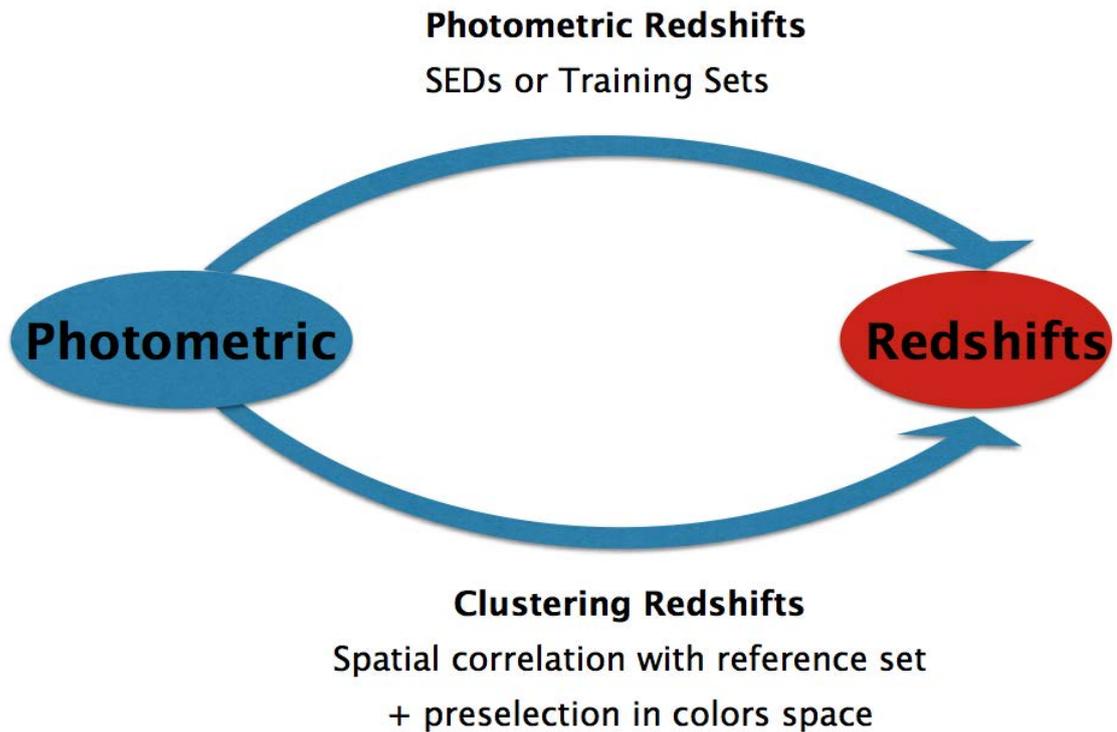


FIGURE 5.14 – Schéma résumant l'approche que j'ai développée dans ce chapitre. Elle combine les propriétés d'agrégation et de couleur des galaxies tout en s'affranchissant des systématiques propres aux redshifts photométriques.

Cette approche basée sur une présélection dans l'espace des couleurs, devrait permettre de s'affranchir d'un certain nombre de problèmes spécifiques aux redshifts photométriques comme les fausses détections de la brisure de Balmer, la confusion Lyman/Balmer ou encore les problèmes dûs à l'absorption de la poussière pour les galaxies en train de former des étoiles.

Cette analyse fut particulièrement délicate notamment à cause d'un problème d'étalonnage des magnitudes dans les données *CFHTLS*. En effet dans une analyse de *cluster-z*, le signal que l'on cherche à extraire provient de la surdensité de galaxies par rapport à une distribution de Poisson et cela, à un redshift donné. Des erreurs d'étalonnage du point zéro photométrique d'un pointé à l'autre de la caméra engendrent d'importantes fluctuations de densité sur le ciel lorsque l'on s'intéresse à une région particulière du plan couleur-couleur. Le panneau du haut de la Figure 5.15 montre ces fluctuations.

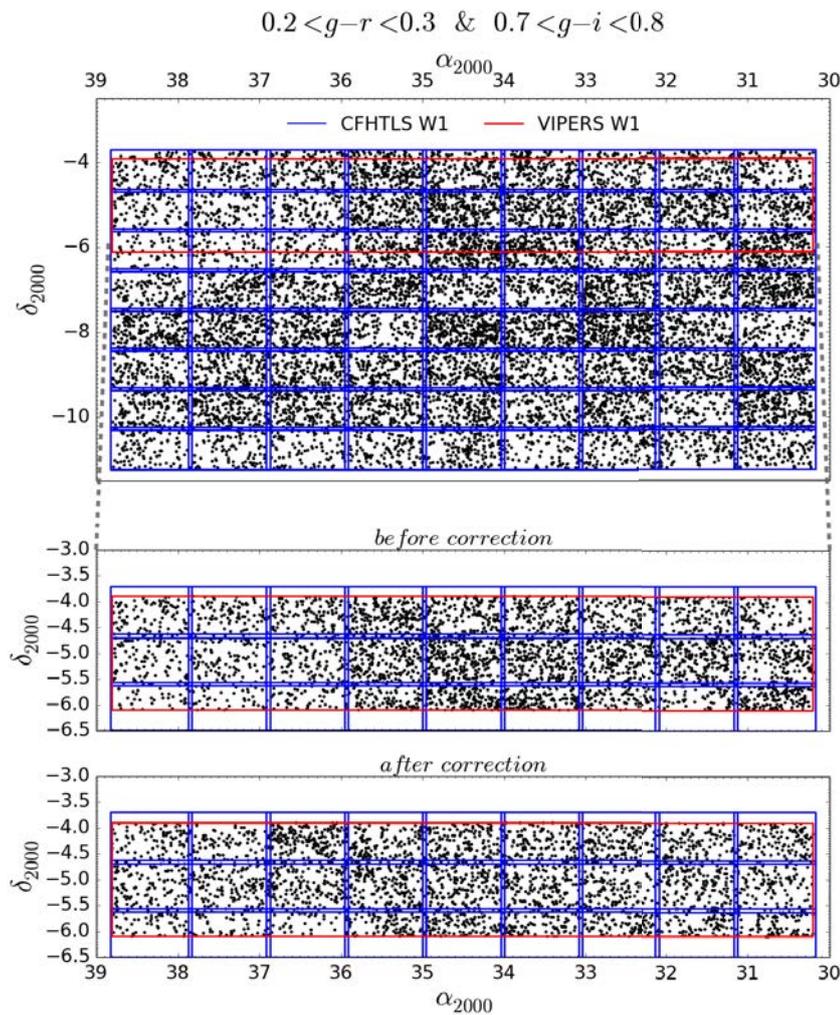


FIGURE 5.15 – Haut : fluctuation de densité dans une région du plan couleur-couleur dû à un problème d'étalonnage du point zéro photométrique dans le champ W1 du *CFHTLS*. On distingue des fluctuations de densité qui suivent parfaitement les pointés de la caméra MegaCam.

Milieu : région où est faite l'analyse avant correction.

Bas : région où est faite l'analyse après correction.

La correction appliquée consiste à recalibrer les flux ISO (mesurés pour une limite isophotal fixée) sur le flux total (AUTO) (mesuré pour un rayon donné) en calculant un facteur de correction moyen sur les bandes  $g,r,i$  et  $K_s$  pour chaque objet. La procédure complète est détaillée dans Moutard et al. 2015 (in prep).

De plus, l'étalonnage du point zéro photométrique est réalisée à partir des zones de recouvrement - environ 6% - entre deux pointés adjacents. A titre de comparaison le taux de recouvrement entre deux pointés adjacents du *Dark Energy Survey* est de 33%. Dans le cadre du *CFHTLS* ce recouvrement fut choisi pour être relativement faible de façon à maximiser la surface totale du champ. En contrepartie ce faible taux de recouvrement rend l'étalonnage du point zéro plus délicate à cause du plus faible nombre de sources communes. Cela peut également engendrer des fluctuations de densité d'un pointé à l'autre.

### 3.1 Même magnitude limite

Dans un premier temps, je me suis intéressé à une population *inconnue* dont la magnitude limite est la même que l'échantillon de *référence*. Cela m'a permis de quantifier la qualité des mesures des *cluster-z* bien que l'on ne possède pas le redshift spectroscopique de chacun des objets *inconnus*. En effet dans ce cas la population de *référence* constitue un sous-échantillon représentatif de la population *inconnue*. Cela signifie que ces deux populations doivent posséder (en moyenne) la même relation couleurs-redshift.

Afin de réduire l'influence du biais galaxies-matière de l'échantillon *inconnu* l'approche que j'ai développée consiste à sélectionner des sous-échantillons correspondant chacun à une région d'un plan couleur-couleur choisi. Ici, je me place dans le plan  $(g - i; g - r)$ . J'ai choisi de prendre  $\Delta g - r/i = 0.1$  de façon à avoir les plus petites cellules possibles tout en conservant un nombre d'objets minimum par cellule de 500 galaxies afin de pouvoir détecter un signal de corrélation croisée. Par construction, les distributions en redshift de chacune de ces cellules du plan couleur-couleur seront plus étroites que la distribution de l'échantillon initial.

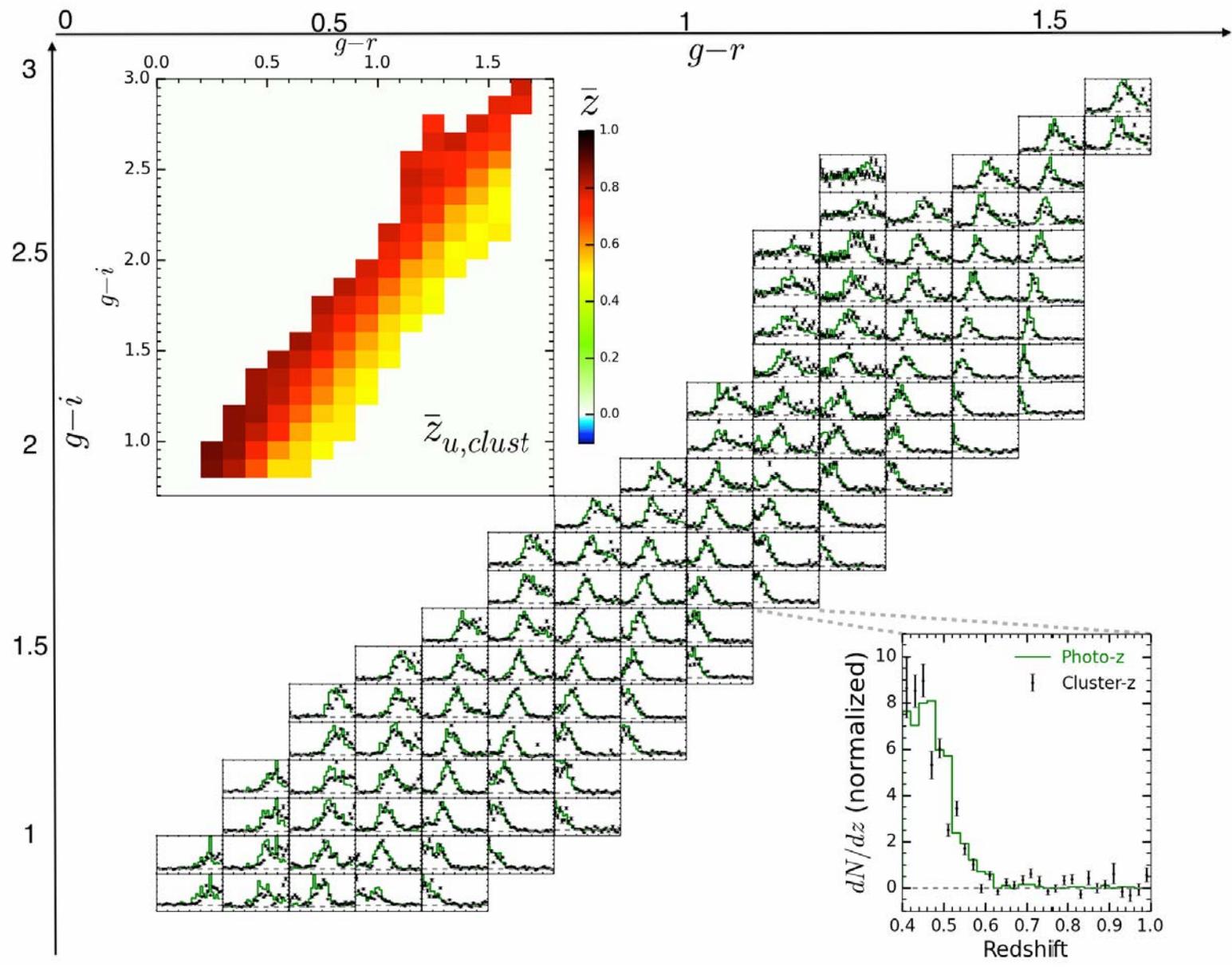


FIGURE 5.16 – Distribution en *cluster-z* (points noirs) et en *photo-z* (vert) à travers l'ensemble du plan couleur-couleur.

J'ai alors mesuré la distribution en clustering redshift dans chacune de ces cellules. L'ensemble des cellules utilisées, leurs distributions en *cluster-z* ainsi qu'en *photo-z* sont visibles sur la Figure 5.16. La partie centrale de cette figure montre l'évolution dans le plan couleur-couleur de la distribution en redshift.

Quand  $g-r$  diminue, le redshift augmente. Cela correspond au passage de la brisure à 4000 Å à travers la bande  $r$  entre redshifts 0.4 et 1. Le passage de cette brisure à travers la bande  $r$  induit une diminution du flux apparent avec l'augmentation du redshift. Par conséquent la magnitude en bande  $r$  augmente et  $g-r$  diminue.

Inversement  $g-i$  augmente avec le redshift. Cela est dû à l'arrivée de la brisure à 4000 Å dans la bande  $i$ . Avec l'approche de la brisure entre redshifts 0.4 et 1, le flux reçu en bande  $i$  ne cesse d'augmenter. La magnitude en bande  $i$  diminue et donc  $g-i$  augmente. Au-delà de  $z = 1$  on retrouve le même phénomène que celui présenté dans le paragraphe précédent et la magnitude diminue.

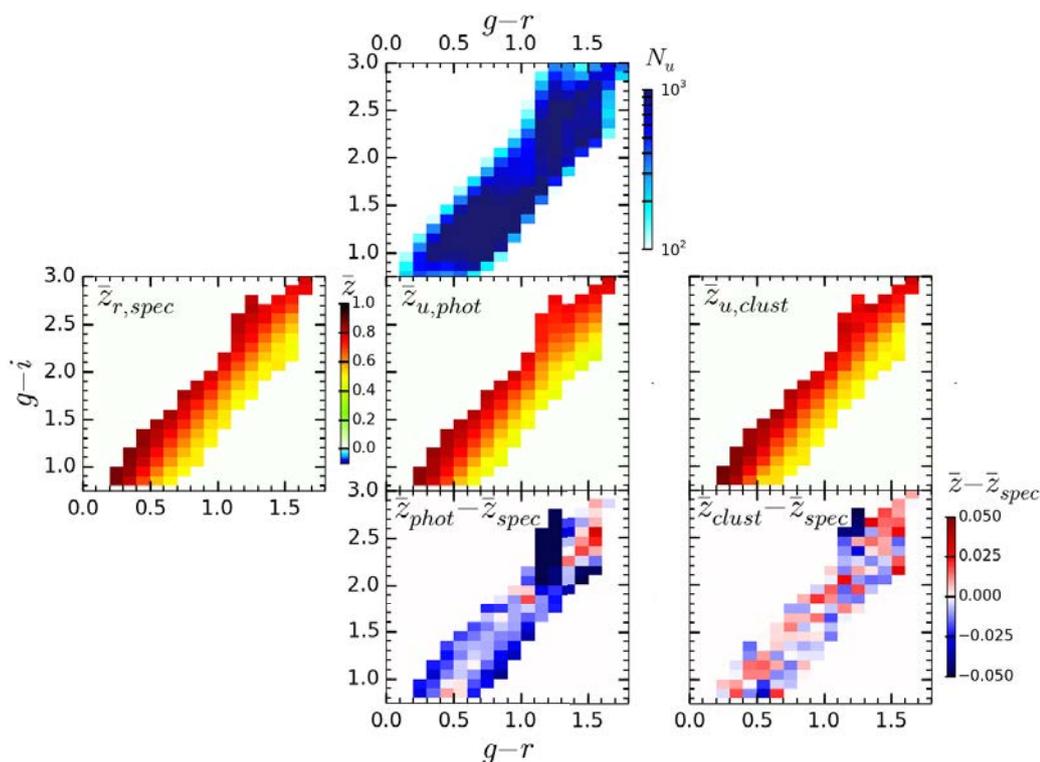


FIGURE 5.17 – Haut : carte montrant le nombre d'objets de l'échantillon *inconnu* dans le plan :  $(g-i, g-r)$ .

Milieu : relation couleurs-redshift pour l'échantillons de *référence* (gauche), obtenue par les *photos-z* pour l'échantillon *inconnu* (centre), obtenue par les *clusters-z* pour l'échantillon *inconnu* (droite).  
Bas : carte des résidus pour les *photos-z* (centre) et pour les *clusters-z* (droite).

L'insert en bas à droite montre un agrandissement de l'une des régions du plan couleur-couleur. L'insert en haut à gauche présente, à l'aide d'un code couleur, l'évolution du redshift moyen  $\bar{z}_{clust}$  mesuré à partir des distributions en *cluster-z* à travers l'ensemble du plan couleur-couleur. Il s'agit en fait de la carte montrant la relation couleurs-redshift reconstruite par les clustering redshifts. La résolution de cette carte est directement liée à la précision des magnitudes.

La Figure 5.17 reprend la carte couleurs-redshift mesuré par les *cluster-z* et présente également la même carte mais cette fois-ci obtenue à l'aide des redshifts photométriques ainsi que la carte couleurs-redshift de l'échantillon de *référence* dont le redshift est spectroscopique. La magnitude limite des deux populations étant la même, on peut considérer que l'échantillon de *référence* est représentatif de la population *inconnue*. De cette façon on peut considérer que la carte obtenue à partir des objets de *référence* correspond à la *vraie* relation couleurs-redshift. Il est donc possible d'utiliser cette carte pour vérifier et comparer les résultats obtenus par les redshifts clustering et photométriques, par exemple en calculant les quantités  $\bar{z}_{clust} - \bar{z}_{spec}$  et  $\bar{z}_{phot} - \bar{z}_{spec}$ , dans chaque cellule. Ces quantités sont également visibles à la dernière ligne de la Figure 5.17.

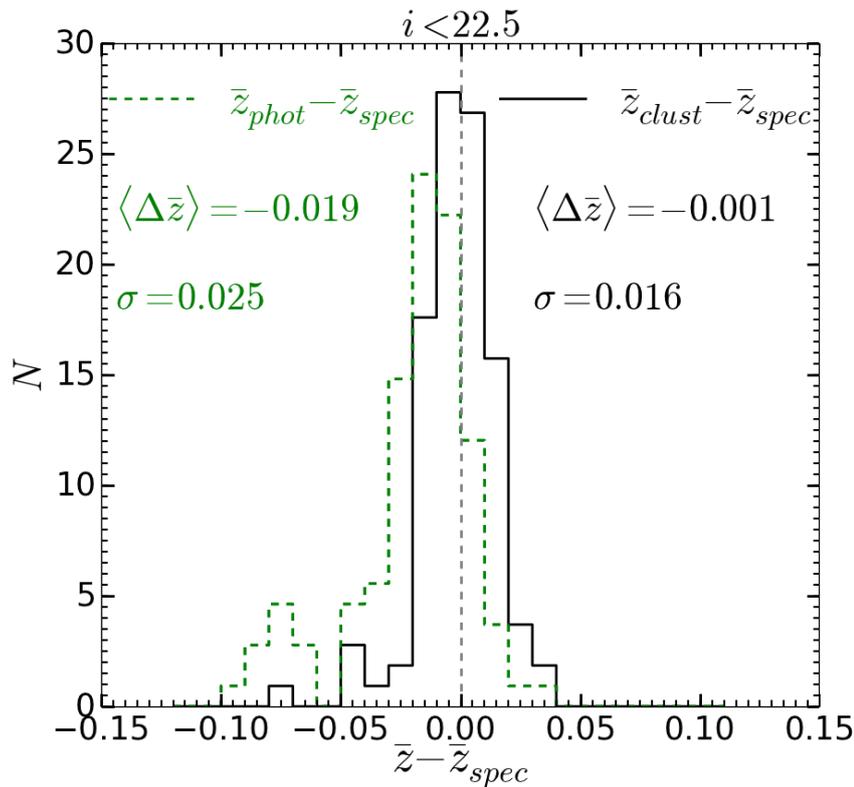


FIGURE 5.18 – Distribution des résidus pour les *cluster-z* (noir) et *photo-z* (vert). L'écart des *cluster-z* est environ 20 fois plus faible que celui des redshifts photométriques.

Visuellement on peut remarquer la présence d'une région particulièrement problématique pour les redshifts photométriques aux alentours de  $g-r = 1.2$  et  $g-i = 2.5$ . Ces objets, très rouges, ont un redshift photométrique sous-estimé. Cela peut être dû à une population particulière d'objets induisant des confusions dans l'estimation du redshift photométrique comme par exemple des galaxies en période de formation d'étoiles (*starforming galaxies*). Cela peut également s'expliquer si aucun modèle utilisé lors de l'estimation du redshift photométrique ne correspond à cette population d'objets.

Afin de comparer de façon plus quantitative la précision des *cluster-z* et des *photo-z* lorsque l'on fait une présélection en couleurs de l'échantillon *inconnu*, je me suis intéressé à la distribution des résidus, disponible à la Figure 5.18. Les *cluster-z* ont un écart moyen environ 20 fois moins important que les redshifts photométriques alors que leurs dispersions sont du même ordre de grandeur. La Figure 5.19 présente la distribution en redshift totale de la population *inconnue* pour  $i < 22.5$ , telle que mesurée par les *redshifts photométriques* ainsi que celle obtenue par la méthode de corrélation croisée. Les résultats de cette analyse concernant la précision des *cluster-z* me permettent d'affirmer que la distribution totale obtenue par les redshifts clustering est plus représentative de la *vrai* distribution.

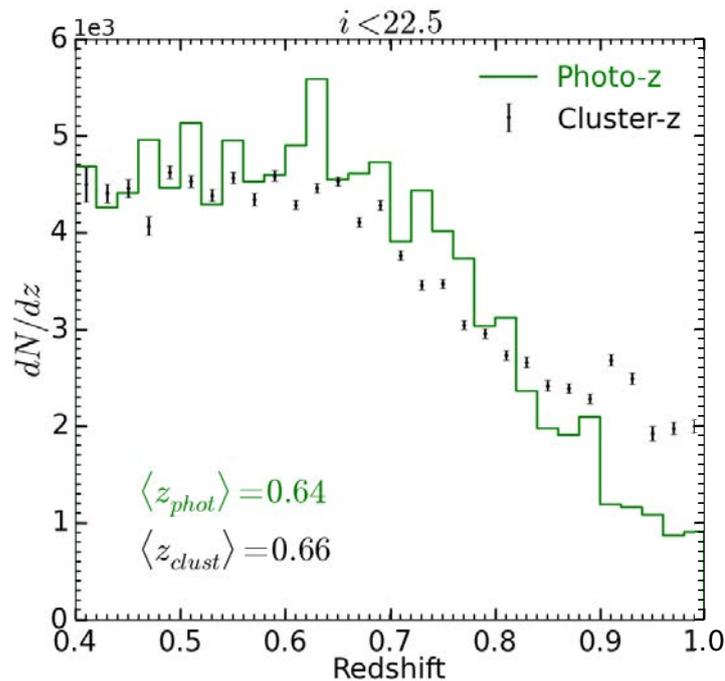


FIGURE 5.19 – Distribution en redshift globale de l'échantillon *inconnu* pour  $i < 22.5$ . Cette distribution est obtenue en combinant toutes les distributions visibles dans le plan couleur-couleur, Figure 5.16. La plus grande précision des *cluster-z* permet d'affirmer que la distribution totale obtenue par les *cluster-z* est plus représentative de la *vrai* distribution que celle des *photo-z*.

### 3.2 Echantillon *inconnu* moins brillant

Dans cette partie j'utilise le même type de sélections dans l'espace des couleurs que précédemment mais cette fois la population *inconnue* possède une magnitude comprise entre  $22.5 < i_u < 23$ . Les échantillons *inconnu* et de *référence* présentent cette fois-ci des relations couleurs-redshift qui ne sont plus identiques. Je ne pourrai donc pas vérifier le résultat comme nous l'avons fait dans la section précédente. Les relations couleurs-redshift obtenues par mesure de clustering et photométrie sont visibles, Figure 5.20. La relation reconstruite par les redshifts *photométriques* est étrangement semblable à celle de l'échantillon de *référence* alors que cela ne devrait pas être le cas. Cela vient probablement de la limite en magnitude à  $i = 21.5$  des modèles utilisés lors de l'estimation des redshifts photométriques. Les redshifts photométriques ne peuvent reproduire que des relations couleurs-redshift représentatives de ces modèles. L'évolution des distributions mesurées dans l'espace des couleurs est visible à la Figure 5.21.

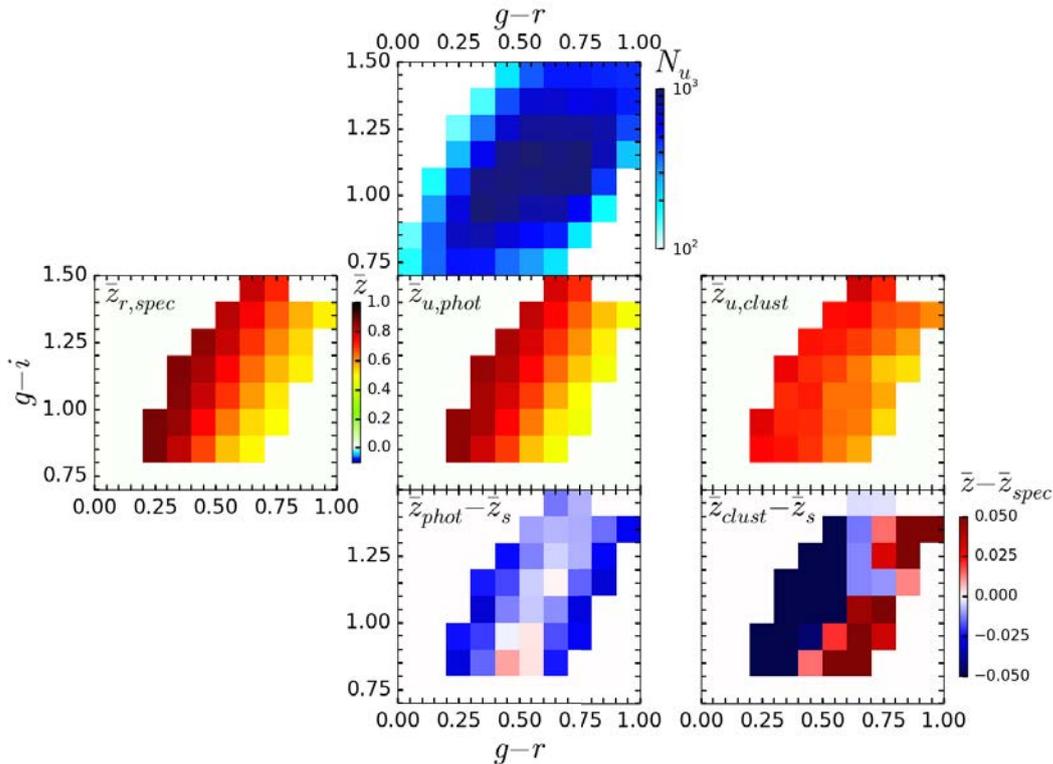


FIGURE 5.20 – Cette figure est similaire à la Figure 5.17 mais cette fois-ci avec :  $22.5 < i_u < 23$ . La forte différence entre la relation couleurs-redshift de la population de *référence* et les mesures de *cluster-z* était prévisible. Par contre l'accord entre les relations couleurs-redshifts des *photo-z* et l'échantillon de *référence* est en contradiction avec le fait que les magnitudes des deux populations soient différentes.

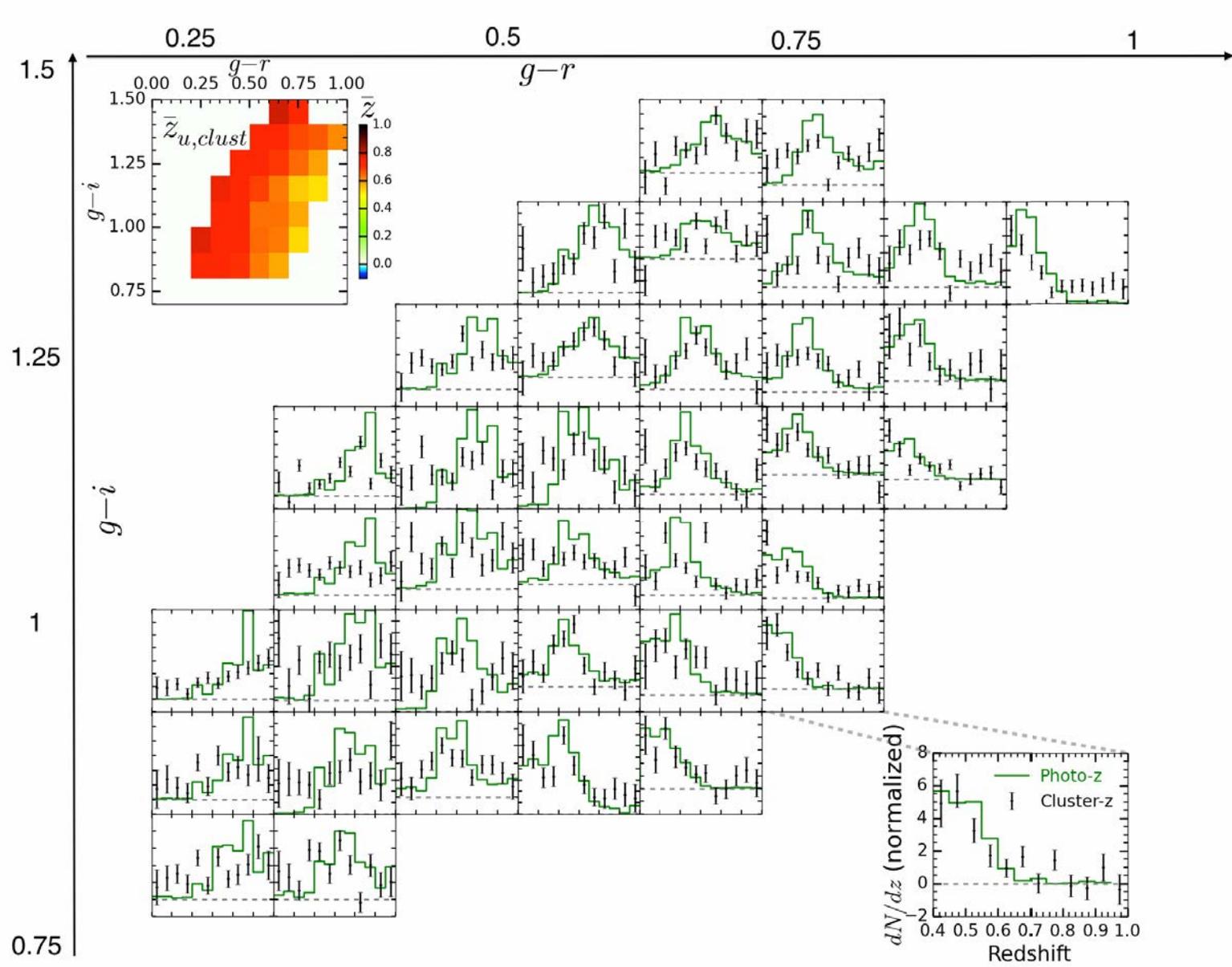


FIGURE 5.21 – Distribution en *cluster-z* (points noirs) et en *photo-z* (vert) à travers l'ensemble du plan couleur-couleur.  $22.5 < i_u < 23$

J'ai ensuite combiné toutes ces distributions en redshift, Figure 5.22. En me basant la plus grande précision des cluster-z à même magnitude vu à la Section 3 et sur le fait que l'échantillon de *référence* n'a pas besoin d'être représentatif de la population *inconnue* vu à la Section 2, j'affirme que la distribution obtenue par mesure de corrélation croisée (Figure 5.22) est plus représentative de la *vraie* distribution que celle obtenue avec les redshifts photométriques.

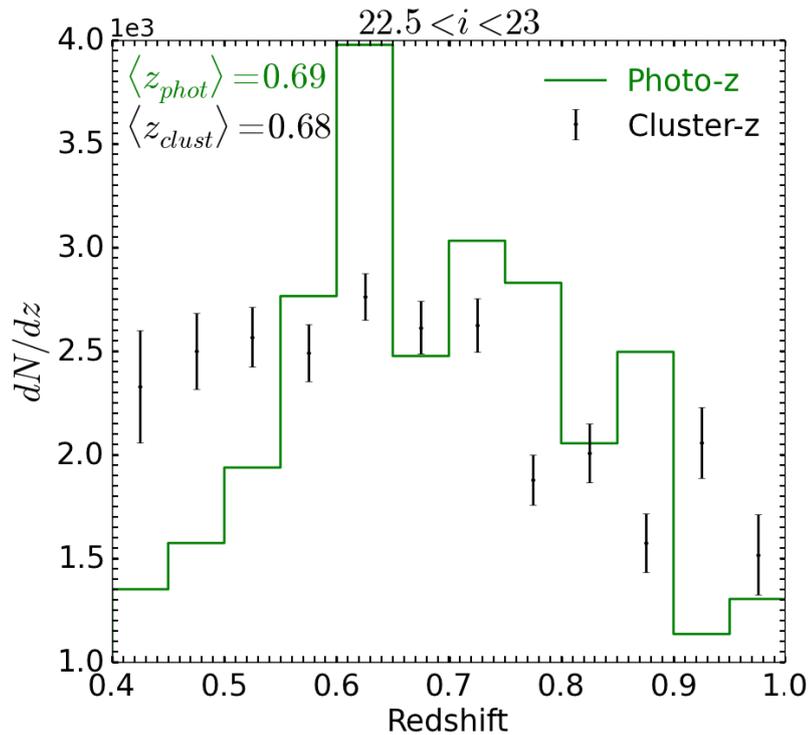


FIGURE 5.22 – Distributions en redshift globale de l'échantillon *inconnu* entre  $22.5 < i_u < 23$ . En me basant sur les résultats des Section 2 & 3 j'affirme que la distribution obtenue par les *cluster-z* est plus proche de la *vraie* distribution que celle obtenue avec les *photo-z*.

## 4 Synthèse

Dans ce chapitre je me suis éloigné du cas idéal vu au chapitre précédent. Pour cela, j'ai commencé par remplacer l'information spectroscopique par le redshift photométrique et j'ai validé cette approche tomographique sur une gamme de redshift jusqu'alors inexplorée, de 0.4 à 1.0. L'exploration de la robustesse de cette approche à plus grands redshifts est laissée à des études ultérieures.

En me basant sur la validité de l'approche tomographique j'ai ensuite montré que l'échantillon de *référence* ne doit pas nécessairement avoir la même limite en magnitude que la population *inconnue*. Cette propriété permet de constituer facilement un catalogue de *référence* très profond - QSOs - et peut permettre d'envisager un gain conséquent - temps d'observation, coût... - tant sur la photométrie que sur la spectroscopie dans le cadre d'un relevé utilisant les *cluster-z*. Pour des relevés comme *Euclid* pour lesquels la constitution d'échantillons spectroscopiques profonds et complet et un objectif à part entière, une telle propriété est des plus intéressante.

Dans le but de m'affranchir des limitations propres à l'utilisation des redshifts photométriques je me suis ensuite intéressé à un autre moyen de réduire l'influence du biais galaxie-matière de l'échantillon *inconnu* basé sur une présélection des données dans l'espace des couleurs. Cette approche est schématisée à la Figure 5.14. Enfin, après avoir validé cette approche basée sur la présélection en couleurs, j'ai utilisé la capacité des *cluster-z* à calibrer des échantillons moins brillants que la population de *référence* afin de donner une distribution en redshift pour un échantillon de galaxies peu brillantes.



# Conclusions et perspectives

---

**D**urant ces trois années de thèse, je me suis intéressé à une toute nouvelle méthode permettant de mesurer la distribution en redshift d'une population de galaxies. Cette approche est fondée sur la répartition spatiale particulière des galaxies. Les redshifts photométriques, largement utilisés en astronomie extragalactique et en cosmologie, ont pour but de caractériser la relation entre l'espace photométrique et l'espace des redshifts en utilisant certaines caractéristiques de la distribution spectrale en énergie de la lumière émise par ces objets. C'est une méthode éprouvée qui mesure les redshifts individuelles des galaxies. Ses performances sont remarquables lorsque l'on dispose de suffisamment de données photométriques réparties dans un grand nombre de filtres (une dizaine, typiquement) couvrant un domaine spectral aussi large que possible et d'un échantillon de redshifts spectroscopiques portant sur l'ensemble des populations de galaxies de l'échantillon photométrique. Cette méthode est sensible à certains effets systématiques comme la confusion entre les brisures de Lyman et de Balmer, l'extinction due à la poussière... L'approche développée dans ce manuscrit a également pour but de caractériser cette relation. Mais elle se distingue par le fait qu'elle utilise les propriétés statistiques d'agrégation mutuelle qui décrit la distribution spatiale particulière des galaxies résultant de l'histoire de leur interaction gravitationnelle. Il s'agit d'un nouveau type d'information. Avec seulement une bande photométrique fournissant les positions angulaires des galaxies d'une population dont les redshifts sont inconnus et un échantillon spectroscopique couvrant les mêmes zones du ciel, cette méthode retrouve la distribution en redshift des galaxies de l'échantillon inconnu par la corrélation croisée avec l'échantillon spectroscopique.

J'ai commencé par rappeler le rôle central de la mesure des distances en astronomie. J'ai particulièrement insisté sur les redshifts photométriques qui ont joué un rôle important en cosmologie ces vingt dernières années. ils ont notamment permis d'aborder les questions cosmologiques de façon statistique.

Mon travail de recherche s'est d'abord concentré sur la confirmation de la vali-

dation de la méthode de mesure de la distribution en redshift par corrélation croisée entre un échantillon de redshift *inconnu* et un échantillon de *référence*, en reproduisant les résultats de [Ménard et collab. \(2013\)](#). Cela m'a notamment permis de tester et de valider les codes que j'ai développés. Cette étude, faisait intervenir des populations particulières d'objets extragalactiques pour lesquels on pouvait s'attendre à un fort *clustering* et donc à une mesure ayant un fort rapport signal sur bruit. Une partie de mon travail a consisté à étendre cette étude à n'importe quel type d'objets. Cela m'a permis de mettre en évidence le fait que la mesure de la distribution en redshift par corrélation croisée semble plus sensible au *vrai* redshift que ne le sont les redshifts photométriques.

Afin d'explorer plus en détail ce point, j'ai réalisé une étude comparative entre les *clustering redshifts* et les redshifts spectroscopiques en me plaçant dans un cas idéal. J'ai alors pu confirmer ce point en obtenant une précision quasi spectroscopique sur la mesure du redshift moyen. J'ai ensuite exploré les différentes sources de bruits pouvant affecter ce type de mesure telles que le sur-échantillonnage de la population de référence ou encore l'influence des vitesses particulières des galaxies.

Dans une seconde étape j'ai réalisé une étude d'un cas plus réaliste. Pour cela, j'ai utilisé le lien existant entre l'espace des couleurs et l'espace des redshifts afin de sélectionner des sous-échantillons dont la dispersion en redshift est intrinsèquement plus faible que la population mère. Cette approche basée sur une présélection en couleurs fut d'abord validée pour des échantillons *inconnu* et de *référence* de même limite en magnitude. Les résultats de l'étude indiquent une mesure du redshift moyen environ vingt fois meilleure que celle obtenue par les redshifts photométriques. J'ai également exploré et validé la méthode du *cluster-z* lorsque la population *inconnue* est moins brillante que la population de *référence*. Cette propriété est extrêmement intéressante car la réalisation d'échantillons spectroscopiques d'objets faibles est longue et délicate et est l'un des enjeux majeurs des grands relevés panoramiques profonds de galaxies pour la calibration des redshifts photométriques.

Parmi les pistes ouvertes par ce travail, on peut mentionner la possibilité de combiner les *cluster-z* avec les redshifts photométriques. Les erreurs systématiques de ces deux méthodes étant différentes on peut espérer une diminution du taux de redshifts catastrophiques. Il serait intéressant de tester cette approche avec les simulations numériques *Euclid* et de quantifier le gain obtenu sur la précision des paramètres cosmologiques dans le cadre d'une analyse de *weak lensing* lorsqu'on utilise les *cluster-z* avec et sans les redshifts photométriques. *Euclid* semble un projet parfaitement adapté à l'exploitation des *cluster-z* car non seulement il disposera d'un catalogue d'un milliard de galaxies avec des positions angulaires très précises mais en plus il disposera d'un échantillon connu de 35 millions de redshifts spectroscopiques couvrant exactement ce catalogue de positions. Évidemment, il est d'un très grand intérêt de poursuivre l'exploration et le développement des redshifts clustering, que ce soit en investiguant ses possibles systématiques, en étudiant plus précisément

l'évolution du biais de la population *inconnue* mais aussi l'influence du nombre de bandes nécessaires pour cartographier la relation entre l'espace des couleurs et l'espace des redshifts ou encore en ajoutant de l'information supplémentaire portant sur les propriétés des galaxies mesurées dans les catalogues, comme la brillance de surface, la taille, l'ellipticité, le type morphologique ou quelques informations de couleurs, etc... Avec ces informations il est possible d'obtenir des mesures de PDF( $z$ ) pour chaque objet ( Figure 5.23). D'autres points que j'aimerais également développer concernent : la mesure de fonctions de luminosité, notamment l'influence des *cluster-z* sur la *faint end* ; ainsi que l'influence de la distorsion redshift-espace sur la mesure de la distribution en redshift.

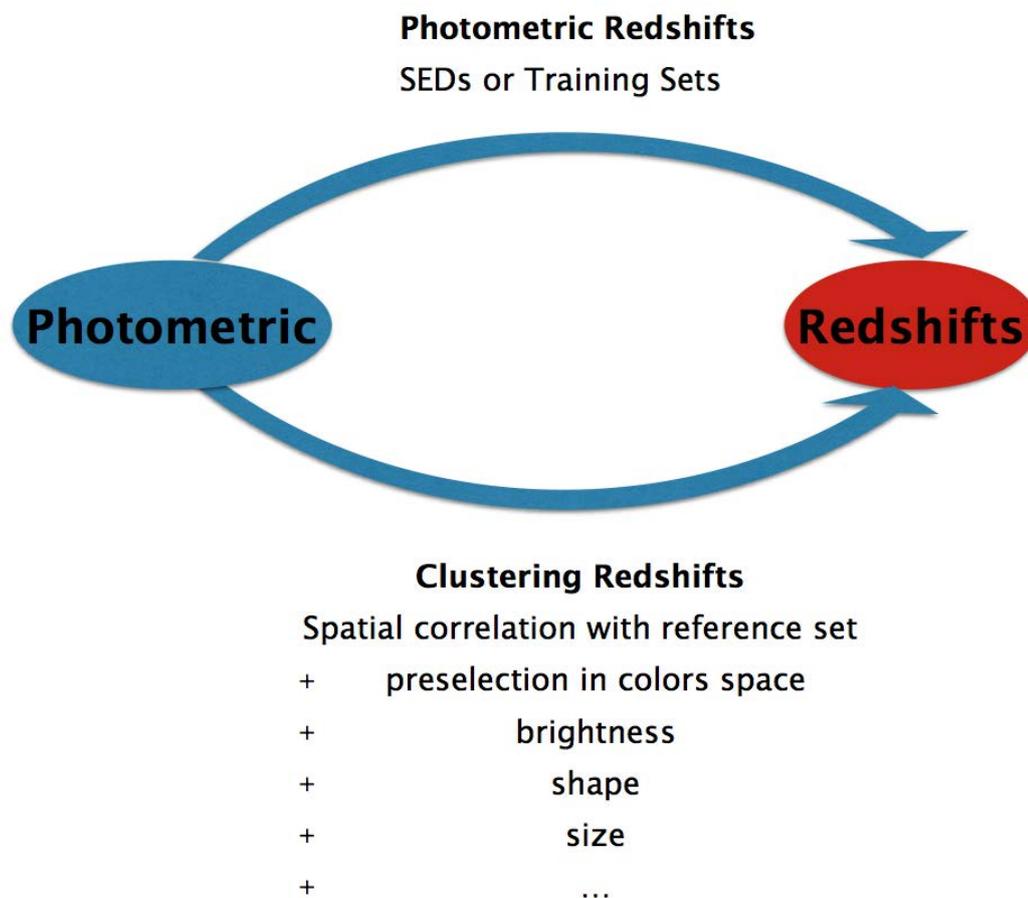


FIGURE 5.23 – Schéma illustrant la prochaine étape dans le développement de la méthode de mesure de redshifts par corrélation croisée. L'ajout d'informations supplémentaires comme par exemple des critères morphologiques devrait permettre de caractériser avec encore plus de précision la relation entre l'espace photométrique et l'espace des redshifts et retrouver non seulement la distribution en redshift mais aussi les redshifts individuels.

De façon plus générale, en ce qui concerne les *clustering redshifts*, il me semble qu'à la lecture des premières études leur performance et leur rôle en cosmologie de-

vraient considérablement augmenter dans les 20 prochaines années. Vingt ans, c'est à peu près le temps qu'il a fallu aux redshifts photométriques pour arriver à maturité et occuper une place centrale dans la majorité des études réalisées. L'exploration de l'ensemble du spectre électromagnétique et de tout un éventail de paramètres structuraux et morphologiques des galaxies constituera une étape indispensable dans le développement des *cluster-z*.

En particulier, les *cluster-z* présentent un fort potentiel pour l'exploitation de la photométrie en infrarouge proche (*near-infrared*, NIR) avec laquelle l'estimation des redshifts photométriques est difficile et peut conduire à des erreurs de l'ordre de 50%, (Jarrett, 2004; Kochanek et collab., 2003). L'ajout de données dans le domaine visible permet de réduire ces erreurs mais, outre le surcoût considérable des observations à conduire et de la production des données, il introduit un risque de biais de sélection puisqu'une partie des sources en NIR n'est pas détectée en optique. Ce domaine est particulièrement intéressant car il permet de très fortement réduire l'effet de l'extinction dû à la poussière que ce soit celle de la Voie Lactée ou bien celle interne aux objets extragalactiques. L'exploration de ce domaine avec la méthode des *cluster-z* a déjà commencé, voir Rahman et collab. (2015a).

L'estimation des redshifts photométriques uniquement à partir des observations dans le domaine des rayons-X peut se faire via la raie d'émission Fe K $\alpha$  et permet d'obtenir une précision sur le redshift inférieur à 5%. Pour des temps d'exposition de 100 ks, 300 ks et 1 Ms on obtient cette précision pour respectivement 20%, 40% et 60% des données, (Castelló-Mor et collab., 2011). Ainsi, l'estimation des *photo-z* dans ce domaine repose généralement sur une étude multi-longueur d'onde, (Hsu et collab., 2014; Zheng et collab., 2004). Comme nous l'avons vu pour le domaine infrarouge, cela introduit un risque de biais de sélections pour les échantillons résultants. L'utilisation du redshift clustering pourrait permettre de résoudre ce problème tout en réduisant le coût des observations. Leur utilisation semble donc également très intéressante pour ce domaine de longueur d'onde qui ouvre l'étude des AGNs et a donc également des implications sur la formation et l'évolution des galaxies.

Enfin, à plus long terme, j'attends avec évidemment beaucoup d'impatience les observations de la mission spatiale *Euclid*. C'est en fait cette mission qui a motivé tout mon travail de recherche sur les *cluster-z*. Comme je le rappelais quelques lignes plus haut, cette mission semble parfaitement adaptée à cette approche de mesure de redshift. C'est aussi un moyen de réduire les coûts et les risques associés aux observations avec les télescopes au sol puisqu'il ne devient plus nécessaire d'obtenir toute la photométrie visible dont a besoin la méthode des redshifts photométriques. Toutefois, c'est en combinant les deux approches que les gains devraient être spectaculaires car ils porteront sur la réductions des erreurs systématiques. Ces observations marqueront, je l'espère, une étape importante dans l'âge d'or de la cosmologie.

# Annexe A

## Fonctions de corrélation

---

### 1 Fonction de corrélation spatiale

La fonction de corrélation spatiale à deux points  $\xi(r)$  représente l'excès de probabilité, par rapport à une distribution aléatoire, de trouver un objet A à une certaine distance  $r$  d'un objet B.

Si les galaxies sont distribuées selon la statistique de Poisson avec une densité moyenne  $n_0$ , la probabilité  $dN(r)$  de trouver une galaxie A dans le volume  $dV_A$  à une distance  $r$  de la galaxie B dans le volume  $dV_B$  est :

$$dN(r) = n_0^2 dV_A dV_B . \quad (\text{A.1})$$

Par distribution poissonnienne on entend un processus aléatoire stationnaire - dont les propriétés moyennes sont identiques quel que soit l'endroit et la taille de l'échantillon considéré.

Si la distribution des galaxies s'écarte un peu d'une distribution poissonnienne, cette probabilité sera :

$$dN(r) = n_0^2 (1 + \xi(r)) dV_A dV_B , \quad (\text{A.2})$$

où  $\xi(r)$  est la fonction de corrélation des galaxies à deux points.  $\xi$  correspond donc à l'écart à une distribution poissonnienne qui, dans le cas des galaxies, est dû à leur agrégation - *clustering* - sous l'effet de la gravité.

Si les galaxies sont distribuées selon cette statistique, par définition, la fonction de corrélation est nulle. Elle peut être positive ou négative selon que les objets soient respectivement, plus ou moins proches les uns des autres qu'ils ne le seraient dans une distribution de Poisson.

Afin de mesurer cette quantité on peut utiliser l'estimateur suivant :

$$\hat{\xi}_{ij}^{PH}(r) = \frac{\langle D_i D_j \rangle_r}{\langle R_i R_j \rangle_r} - 1 . \quad (\text{A.3})$$

Avec :

- $\{i, j\}$ , des jeux de données.
- $\xi_{ij}$ , la fonction de corrélation. On parle d'auto-corrélation lorsque  $i = j$ , et de corrélation croisée - où cross-corrélation - lorsque  $i \neq j$ .
- $D_i D_j$ , le nombre de paires d'objets distants de  $r$  entre les jeux  $i$  et  $j$ .
- $R_i R_j$ , le nombre de paires d'objets distant de  $r$ , pour des distributions de Poisson de même géométrie - c'est-à-dire de même surface et de même forme - que les jeux  $i$  et  $j$ .

Cet estimateur est communément appelé, estimateur naturel ou encore estimateur de [Peebles et Hauser \(1974\)](#). Son principal défaut est que son incertitude à grande échelle est limitée par l'incertitude sur la densité moyenne.

- l'estimateur d' [Hewett \(1982\)](#).

$$\hat{\xi}_{HEW}(r) = \frac{\langle DD \rangle_r - \langle DR \rangle_r}{\langle RR \rangle_r} . \quad (\text{A.4})$$

- l'estimateur de [Davis et Peebles \(1983a\)](#).

$$\hat{\xi}_{DP}(r) = \frac{\langle DD \rangle_r}{\langle DR \rangle_r} - 1 . \quad (\text{A.5})$$

- l'estimateur d'[Hamilton \(1993\)](#), dont l'incertitude ne varie pas - ou peu - avec l'échelle et est limité par le nombre de paires.

$$\hat{\xi}_{HAM}(r) = \frac{\langle DD \rangle_r \langle RR \rangle_r}{\langle DR \rangle_r^2} - 1 . \quad (\text{A.6})$$

- l'estimateur de [Landy et Szalay \(1993\)](#), qui a le même comportement que l'estimateur d'Hamilton mais dont la variance est minimale dans le cas d'une distribution aléatoire, c'est-à-dire lorsque la fonction de corrélation est nulle.

$$\hat{\xi}_{LS}(r) = 1 + \frac{\langle DD \rangle_r}{\langle RR \rangle_r} - 2 \frac{\langle DR \rangle_r}{\langle RR \rangle_r} . \quad (\text{A.7})$$

Voir [Labatie et collab. \(2010\)](#) pour une comparaison plus détaillée.

Plus récemment [Vargas-Magaña et collab. \(2013\)](#) ont proposé un nouvel estimateur qui est une combinaison linéaire de tous les ratios possibles - jusqu'au second

ordre - du nombre de paires entre les données et l'échantillon aléatoire. Cette combinaison linéaire peut être optimisée afin de minimiser la variance de la fonction de corrélation pour une géométrie de relevé donnée.

## 2 Fonction de corrélation angulaire

De la même façon que pour  $\xi(r)$ , on peut définir  $\omega(\theta)$ , la fonction de corrélation 2D. Cette fonction représente l'excès de probabilité de trouver un objet A à une distance  $\theta$  sur le ciel d'un objet B.

On peut alors exprimer la probabilité  $dP_{2D}$  d'observer sur la surface du ciel - donc à deux dimensions - 2 galaxies chacune dans un angle solide  $d\Omega_i$ ,  $i \in \{A, B\}$  séparées de l'angle  $\theta$ , étant donnée une densité de surface moyenne de galaxies  $\Sigma$  :

$$dP_{2D} = \Sigma^2 d\Omega_A d\Omega_B (1 + \omega(\theta)) . \quad (\text{A.8})$$

L'estimateur de Peebles pour la fonction de corrélation angulaire s'écrit :

$$\hat{\omega}_{ij}^{PH}(\theta) = \frac{\langle D_i D_j \rangle_\theta}{\langle R_i R_j \rangle_\theta} - 1 , \quad (\text{A.9})$$

avec :

- $\omega_{ij}$ , la fonction de corrélation.
- $DD$ , le nombre de paires d'objets entre les jeux  $i$  et  $j$  distants de  $\theta$  sur le ciel.
- $RR$ , le nombre de paires d'objets distants de  $\theta$ , pour une distribution de Poisson.

## 3 Relation exacte entre $\xi$ et $\omega$ & relation de Limber

La fonction de corrélation angulaire,  $\omega(\theta)$ , n'est autre que la projection sur le ciel de  $\xi(r)$  et présente l'avantage de ne pas nécessiter la connaissance de la distance nous séparant des objets étudiés. Les fonctions de corrélation 2D et 3D sont reliées entre elles par l'équation (A.10) (Peebles, 1980) :

$$\omega(\theta) = \int dr_1 \int dr_2 \phi_1(r_1) \phi_2(r_2) \xi \left( R, \frac{r_1 + r_2}{2} \right) , \quad (\text{A.10})$$

avec :

- $R = |\vec{r}_1 - \vec{r}_2| = \sqrt{r_1^2 + r_2^2 - 2r_1 r_2 \cos \theta}$  .
- $r_i$   $i \in \{1, 2\}$ , la distance de l'objet  $i$ .
- $\phi_i$ , la fonction de sélection de  $i$ . Elle exprime le fait que les galaxies contenues dans un échantillon incluent des effets observationnels ou instrumentaux qui affectent les propriétés du catalogue (limite en flux, homogénéité sur la route

- céleste, discrimination en type morphologique, en type spectral ou en redshift, etc...).
- et où le second argument de  $\xi$  prend en compte une possible évolution temporelle du champs de densité de galaxies. Cet argument paramétrise le temps de regard en arrière  $t_b(r)$  auquel la fonction de corrélation est observée (Simon, 2007).

Pour obtenir cette relation *exacte* entre les fonctions de corrélations 3D et 2D on a seulement fait les hypothèses suivantes :

- les champs de densité sont statistiquement homogènes et isotropes.
- l'évolution temporelle de  $\xi$  est petite dans le domaine où le produit  $\phi_1(r_1) \phi_2(r_2)$  est non nul.

En posant :

$$\bar{r} = \frac{r_1 + r_2}{2}, \quad \Delta r = r_2 - r_1. \quad (\text{A.11})$$

En se plaçant dans l'approximation des petits angles, on obtient la relation de Limber (1953) :

$$\omega(\theta) = \int d\bar{r} \phi_1(\bar{r}) \phi_2(\bar{r}) \int d\Delta r \xi(R, \bar{r}), \quad (\text{A.12})$$

avec :  $R = \sqrt{\bar{r}^2 \theta^2 + \Delta r^2}$

Pour plus d'informations concernant la relation de Limber, voir Simon (2007).

Annexe B

Publication

---

# CLUSTERING-BASED REDSHIFT ESTIMATION : APPLICATION TO VIPERS/CFHTLS

Vivien Scottez<sup>1\*</sup>, Brice Ménard<sup>2,3,4</sup>, Yannick Mellier<sup>1</sup>, Martin Kilbinger<sup>1,5</sup>, VIPERS team

<sup>1</sup>*Institut d'Astrophysique de Paris, UMR7095 CNRS, Université Pierre & Marie Curie, 98 bis boulevard Arago, F-75014 Paris, France*

<sup>2</sup>*Department of Physics & Astronomy, Johns Hopkins University, 3400 N. Charles Street, Baltimore, MD 21218, USA*

<sup>3</sup>*Institute for the Physics and Mathematics of the Universe, Tokyo University, Kashiwa 277-8583, Japan*

<sup>4</sup>*Alfred P. Sloan Fellow*

<sup>5</sup>*CEA/Irfu/SAP Saclay, Laboratoire AIM, F-91191 Gif-sur-Yvette, France*

Accepted 2015 ... . Received 2015 ... ; in original form 2015 ...

## ABSTRACT

We explore the accuracy - in a real case scenario - of the clustering-based redshift estimation proposed by Ménard et al. (2013) and compared to spectroscopic redshift by Rahman et al. (2015). This method enables to reconstruct redshift distributions from measurement of the angular clustering of objects using a set of secure spectroscopic redshifts. We use state of the art spectroscopic measurements with  $i_{AB} < 22.5$  from the VIMOS Public Extragalactic Redshift Survey (VIPERS) as *reference* population to infer the redshift distribution of galaxies from the Canada-France-Hawaii Telescope Legacy Survey (CFHTLS) T0007 release. We show that this method enables to better reproduce - by a factor  $\sim 20$  - the *true* colors-redshift relation when both populations have the same magnitude limit. We also show that this technique allows the inference of redshift distributions for a population fainter than the one of *reference*.

**Key words:** redshift - clustering - methods: data analysis - extragalactic - surveys

## 1 INTRODUCTION

Large future and ongoing redshift surveys like the ESA Euclid space mission aims to probe dark energy with unprecedented accuracy. Many of the cosmological measurements to be performed with these surveys - e.g., tomographic weak lensing - will require extremely well characterized redshift distributions Albrecht et al. (2006); Huterer et al. (2006); Ma et al. (2006).

Since, it is impractical to measure spectroscopic redshifts for hundreds of millions of galaxies - especially extremely faint ones- these experiments are largely dependent upon photometric redshifts: i.e., estimates of the redshifts of objects based only on flux information obtained through broad filters. *Photo-z* also require large set of spectroscopic redshifts to create templates of how colors varies with redshift Connolly et al. (1995). However current and future spectroscopic surveys will be highly incomplete due to selection biases dependent on redshift and galaxy properties Cooper et al. (2006). Because of this, along with the catastrophic photometric errors that can occur at a significant ( $\sim 1\%$ ) rate Sun et al. (2009); Bernstein & Huterer (2010), photometric redshifts are not as well understood as redshifts

determined spectroscopically. If future dark energy experiments are to reach their goals, it is necessary to develop a method to infer, at least, the redshift distribution with high precision.

Current projections for cosmic shear measurements estimate that the true mean redshift of objects in each *photo-z* bin must be known to better than  $\sim 0.002(1+z)$  Zhan & Knox (2006); Zhan (2006); Knox et al. (2006) with stringent requirements on the fraction of unconstrained catastrophic outliers Hearin et al. (2010), while the width of the bin must be known to  $\sim 0.003(1+z)$ . Newman et al. (2013) investigated the spectroscopic needs for imaging dark experiments and insisted on the extremely high ( $\sim 99.9\%$ ) completeness required for calibration techniques. The idea of measuring redshift distributions using the apparent clustering of objects on the sky is not new. It was first developed by Seldner & Peebles (1979); Phillipps & Shanks (1987) and Landy et al. (1996). This was practically forgotten mainly due to the essort of photometric redshifts. To face the challenges of futur and ongoing imaging dark experiments Newman (2008); Matthews & Newman (2010) and Matthews & Newman (2012) resurrected this idea on simulations while McQuinn & White (2013) proposed an optimal estimator for such a measurement. In this paper we explore the clustering-based redshift estimation,

\* E-mail: scottez@iap.fr

i.e. *cluster-z*, via a local (i.e. within few Mpc) approach introduced and validated on simulations by: Schmidt et al. (2013) ; then succesfully applied on real data by Ménard et al. (2013) (M13) and compared to spectroscopic redshift in an ideal case study at limiting magnitude  $r_{model} < 19$  by Rahman et al. (2015) (R15). This works aims to explore the strenght of *cluster-z* in a real case study at fainter magnitude  $i_{AB} < 22.5$  and demonstrate our ability to recover the redshift distribution of an *unknown* sample with  $22.5 < i_{AB} < 23$  when the *reference* sample used for calibration has only:  $i_{AB} < 22.5$ .

This paper is organized as follows. In Section 2, we review the clustering-based redshift formalism while the data used in this work is described in Section 3. Then, in Section 4 we show our ability to measure the clustering redshift distribution in a real case scenario using a tomographic *photo-z* approach. We also show that this method allows the estimation of redshift distribution when the sample of *unknown* redshift is fainter than the *reference* one. Finally, we free *cluster-z* from the use of *photos-z* in Section 5 by selecting subsamples in colors-space. Conclusions are presented in Section 6.

## 2 CLUSTERING-BASED REDSHIFT: FORMALISM

The method used in this paper is based on the work of M13 and R15. We refer the reader to those papers for more details. In this section, we briefly review the formalism.

The main idea behind this approach is that galaxies are not randomly distributed on the sky. Indeed, due to gravity, galaxies clusters with each others. This information is encoded into the two-point correlation function as an excess probability - compare to a random distribution - to find two objects close together. This is valid in 3D and, by projection, on the sky. Using a *reference* sample of secure spectroscopic redshift - and by looking at cluster scale - it is then possible to extract the *excess probability* of finding a population of galaxies at a given redshift. Obviously the *reference* population and the *unknown* one - for which angular positions are known but redshifts are not - have to overlap on the sky.

The mean surface density of *unknown* objects at a distance  $\theta$  from a *reference* one which is at a redshift  $z$ , is:

$$\Sigma_{ur}(\theta, z) = \Sigma_R [1 + \omega_{ur}(\theta, z)] , \quad (1)$$

where  $\Sigma_R$  is the random surface density of the *unknown* sample and  $\omega_{ur}(\theta, z)$  the two-point angular cross-correlation function between the two samples. Then, one can define the integrated cross-correlation function:

$$\bar{\omega}_{ur}(z) = \int_{\theta_{min}}^{\theta_{max}} d\theta W(\theta) \omega_{ur}(\theta, z) , \quad (2)$$

where  $\theta$  ranges from  $\sim 100$  kpc to several megaparsecs and  $W(\theta)$  is a weight function -  $\propto \theta^{-0.8}$  - aimed at optimizing the overall  $S/N$  whose integral is normalized to unity. This integrated cross-correlation function translate the excess probability, compare to a random distribution, to find objects from the *unknown* sample at a redshift  $z$ .

One can also write this quantity as a function of redshifts selection functions of both samples,  $\frac{dN_i}{dz}(z)$ , as well as the galaxy-dark matter biases,  $\bar{b}_i(z)$ , and the dark matter correlation function,  $\bar{\omega}_m(z)$ .

$$\bar{\omega}_{ur} = \int dz' \frac{dN_u}{dz}(z') \frac{dN_r}{dz}(z') \bar{b}_u(z') \bar{b}_r(z') \bar{\omega}_m(z') . \quad (3)$$

Applying the narrow sample approximation for the *reference*  $\frac{dN_r}{dz}(z) = N_r \delta_D(z' - z)$  - with  $\delta_D$  the Dirac delta function - we can then simply invert the previous integral and get:

$$\frac{dN_u}{dz}(z) \propto \bar{\omega}_{ur}(z) \times \frac{1}{\bar{b}_u(z)} \times \frac{1}{\bar{b}_r(z) \bar{\omega}_m(z)} , \quad (4)$$

where  $\bar{\omega}_{ur}(z)$  can be directly measured on data,  $\bar{b}_r(z)$  can be measured on the *reference* sample,  $\bar{\omega}_m(z)$  is given by the cosmology and  $\bar{b}_u(z)$  is the only not known quantity.

Considering narrow redshift distribution for the *unknown* sample, we can neglect the variation of its galaxy-dark matter bias with respect to the variation of the number of objects:

$$\frac{d \log dN_u/dz}{dz} \gg \frac{d \log \bar{b}_u}{dz} , \quad (5)$$

we get:

$$\frac{dN_u}{dz}(z) \propto \bar{\omega}_{ur}(z) \left( \frac{1}{\bar{b}_r(z) \bar{\omega}_m(z)} \right) . \quad (6)$$

In the same way that we wrote equation(5), we can neglect the variation of  $\sqrt{\bar{\omega}_m}$  with redshift compare to  $\frac{dN_r}{dz}(z)$

$$\frac{d \log dN_r/dz}{dz} \gg \frac{d \log \sqrt{\bar{\omega}_m}}{dz} . \quad (7)$$

Thus introducing the clustering amplitude of the *reference* sample,  $\beta_r(z)$ , we can write:

$$\beta_r(z) = \sqrt{\frac{\bar{\omega}_{rr}(z)}{\bar{\omega}_{rr}(z_0)}} \propto \frac{\bar{b}_r(z)}{\bar{b}_r(z_0)} . \quad (8)$$

As explained in (R15) one should note that this quantity is different from the linear galaxy bias which is usually defined only on large-scales for which the galaxy and dark matter density fields are, on average, linearly related. This bias definition includes contributions from small scales over which the galaxy and matter fields are non-linearly related. We can then rewrite a model independant version of equation(6), and we get:

$$\frac{dN_u}{dz}(z) \propto \bar{\omega}_{ur}(z) / \beta_r(z) . \quad (9)$$

Finally, the redshift distribution is normalized to the number of objects in the *unknown* sample through:

$$\int dz dN_u/dz = N_u . \quad (10)$$

It is important to realise that to be able to write and use equation(9), we have to select *unknown* samples with relatively small redshift distributions. The ability of selecting subsamples with narrow redshift distributions is important.

To directly measure the integrated cross-correlation function, we can simply use the Peebles & Hauser (1974) estimator, also known as natural estimator:

$$\bar{\omega}_{ur}(z) = \frac{\langle \Sigma_{ur} \rangle_z}{\Sigma_R} - 1 . \quad (11)$$

The error on the measurement is then estimated through Poisson statistic, and is given by:

$$\sigma_{\bar{\omega}}^2 = \left( \frac{\bar{\omega} + 1}{\sqrt{N_{ur}}} \right)^2 + \left( \frac{\bar{\omega} + 1}{\sqrt{N_R}} \right)^2, \quad (12)$$

where  $N_{ur}$  is the neighbors number of *unknown* objects over  $[\theta_{\min}; \theta_{\max}]$  around *reference* galaxies and  $N_R$  is the corresponding number of neighbors for a random distribution.

### 3 DATA ANALYSIS

#### 3.1 The Datasets

##### 3.1.1 CFHTLS: unknown sample

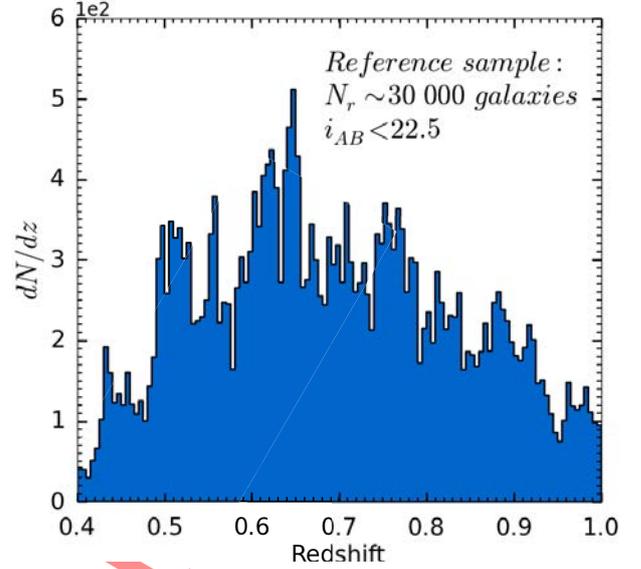
The Canada-France-Hawaii Telescope Legacy Survey (CFHTLS<sup>1</sup>) Wide includes four fields labelled W1, W2, W3 and W4.

Complete documentation of the CFHTLS-T0007 release can be found at the CFHT<sup>2</sup> site. In summary, the CFHTLS-Wide is a five-band survey of intermediate depth. It consists of 171 MegaCam deep pointings (of 1 deg<sup>2</sup> each) which, as a consequence of overlaps, consists of a total of  $\sim 155$  deg<sup>2</sup> in four independent contiguous patches, reaching a 80% completeness limit in AB of  $u^* = 25.2$ ,  $g = 25.5$ ,  $r = 25.0$ ,  $i = 24.8$ ,  $z = 23.9$  for point sources. The photometric catalogs we use here were constructed by running SExtractor<sup>3</sup>, Bertin & Arnouts (1996), in dual-image mode on each CFHTLS tile, using the *gri* images to create a  $\chi^2$  detection image Szalay et al. (1999). The final merged catalogue is constructed by keeping objects with the best signal-to-noise from the multiple objects in the overlapping regions. Objects are selected using *i*-band MAG\_AUTO total magnitudes Kron (1980) and object colours are measured using 3 *arcsec* aperture magnitudes.

In this work, we focus on the W1 field of the CFHTLS-T0007 release. We select all galaxies in the same region of the sky than VIPERS and which are outside CFHTLS masks resulting in an area of  $\sim 8.9$ deg<sup>2</sup>. Since VIPERS data only span the redshift range  $0.4 < z < 1$  we will not be able to measure any signal outside this interval. *unknown* objects outside this range will biased the overall redshift distribution since it is normalized to the total number of *unknown* galaxies following equation(10). To minimize this problem we select objects with a photometric redshift matching the range of redshift covered by the *reference* population. The resulting population correspond to the parent *unknown* sample - sample for which we try to recover the redshift distribution - used in this paper.

##### 3.1.2 VIPERS: reference sample

The VIMOS Public Extragalactic Redshift Survey (VIPERS<sup>4</sup>) Guzzo et al. (2014), is an on going spectroscopic deep survey whose aim is to map the detailed spatiale distribution of galaxies. The survey is made of two

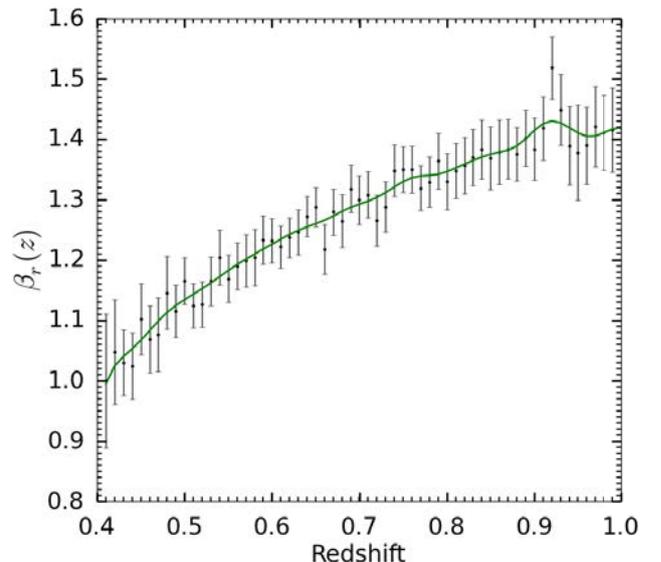


**Figure 1.** The redshift distribution of the *reference* sample built from VIPERS data with  $i_{AB} < 22.5$  and a bin width  $\delta z = 0.005$ .

distincts fields selected inside CFHTLS W1 et W4 fields. The total survey area is 24 deg<sup>2</sup>. VIPERS spectra are the results of 440h of observation at the Very Large Telescope (VLT) in Chili. The targets selection was performed to keep galaxies at redshift higher than 0.4 using the following colors criteria:

$$(r - i) > 0.5(u - g) \text{ OR } (r - i) > 0.7. \quad (13)$$

The one sigma standard deviation on the spectro-



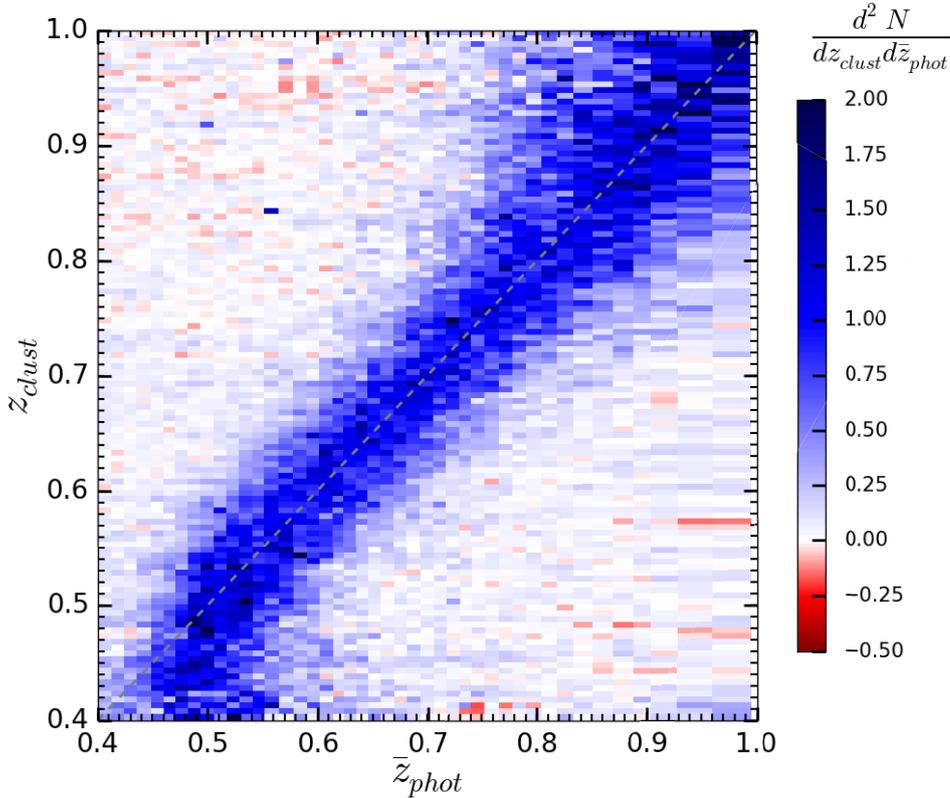
**Figure 2.** Clustering amplitude evolution of the *reference* sample normalized to 1 at  $z_0 = 0.41$ . The solid line is the smooth version of the clustering amplitude evolution used to remove this factor from the clustering redshift distributions in this paper.

<sup>1</sup> <http://www.cfht.hawaii.edu/Science/CFHLS/>

<sup>2</sup> <http://www.cfht.hawaii.edu/Science/CFHLS/T0007/>

<sup>3</sup> <http://www.astromatic.net/software/sextractor>

<sup>4</sup> <http://vipers.inaf.it>



**Figure 3.** Density map showing the 5 280 clustering measurements made in Section 4.2. Each vertical line correspond to a clustering redshift distribution measured in a tomographic sample of mean photometric redshift  $\bar{z}_{ph}$

scopic redshifts measurements for VIPERS objects is:  $\sigma_z = 0.00047(1+z)$ .

The *reference* sample is made of VIPERS objects outside CFHTLS masks and with secure spectroscopic redshifts ( $CL > 95\%$ ). The resulted *reference* sample is composed of  $N_r \sim 30\,000$  galaxies with  $i_{AB} < 22.5$  over  $\sim 8.9$  deg<sup>2</sup>. It correspond to the *reference* population used in all the analysis presented in this paper, see Figure 1.

### 3.1.3 Reference clustering amplitude measurement

As previously seen in Section 2, the determination of a clustering redshift distribution, require the knowledge of the evolution with redshift of the clustering amplitude of the *reference* population,  $\beta_r(z)$ . This quantity can be directly measured on data using equation(8) and is shown in Figure 2. We also show a smoothed version, for which we convolved the binned measurements with a Hann filter of width  $\Delta z = 0.02$

## 4 TOMOGRAPHIC SAMPLING:

As seen in Section 2, minimizing the variation of  $\bar{b}_u(z)$  is a key point in this method. In this section we aim to demonstrate our ability to measure the redshift distribution. To minimize the variation of  $\bar{b}_u(z)$  we choose to work with

tomographic subsamples of the *unknown* population. The tomography is done by selecting objects on their best-fit photometric redshift.

### 4.1 Photometric redshifts estimation

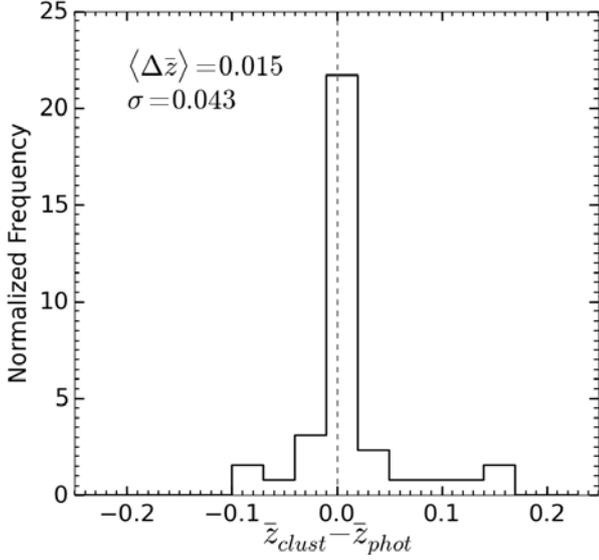
The photometric redshifts we use here have been provided as part of the CFHTLS-T0007 release. They are fully documented on the corresponding pages at the TERAPIX<sup>5</sup> site and the photometric redshifts were computed following Ilbert et al. (2006); Coupon et al. (2009).

In summary, the photometric redshifts are computed using the *Le Phare*<sup>6</sup> photometric redshift code which uses a standard template fitting procedure. These templates are redshifted and integrated through the transmission curves. The opacity of the intra-galactic medium (IGM) is accounted for and internal extinction can be added as a free parameter to each galaxy. The photometric redshifts are derived by comparing the modelled fluxes and the observed fluxes with a  $\chi^2$  merit function. In addition, a probability distribution function is associated to each photometric redshift.

The primary template sets used are the four observed

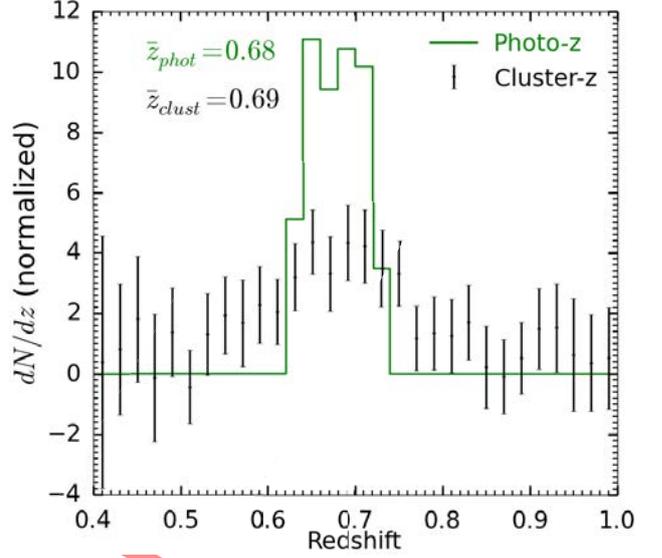
<sup>5</sup> <http://terapix.iap.fr/cplt/T0007/doc/T0007-doc.html>

<sup>6</sup> <http://www.cfht.hawaii.edu/arnouts/LEPHARE/lephare.html>



**Figure 4.** Histogram showing the distribution of the difference between the mean of the clustering redshift distribution and the mean of the tomographic photometric redshift distribution.

spectra (Ell, Sbc, Scd, Irr) from Coleman et al. (1980) complemented with two observed starburst spectral energy distribution (SED) from Kinney et al. (1996). These templates have been optimised using the VVDS deep spectroscopic sample Le Fèvre et al. (2005). Next, an automatic zero-point calibration has been carried out using spectroscopic redshifts present W1, W3 and W4 fields. For spectral type later than Sbc, a reddening  $E(B - V) = 0$  to 0.35 using the Calzetti et al. (2000) extinction law is applied. Star-galaxy separation is carried out using a joint selection taking into account the compactness of the object and the best-fitting templates.



**Figure 5.** Measurement of the redshift distribution of an unknown sample with  $22.5 < i_u < 23$  using a reference population with  $i_r < 22.5$ .

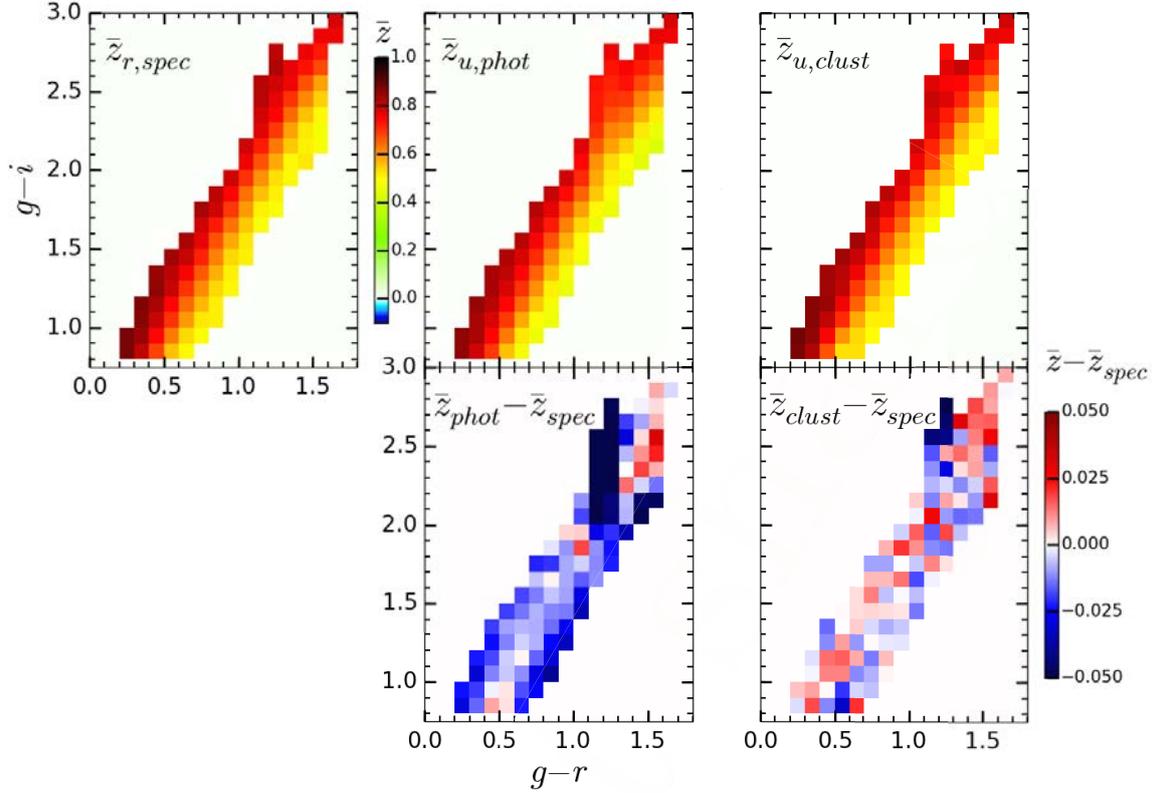
Finally, using the photometric redshift, the associated best-fitting template and the observed apparent magnitude in a given band, one can directly measure the k-correction and the absolute magnitude in any rest-frame band. Since at high redshifts the k-correction depends strongly on the galaxy SED, it is the main source of systematic errors in determining absolute magnitudes. To minimise the k-correction uncertainties, the rest-frame luminosity is derived at a wavelength  $\lambda$  using the objects apparent magnitude closest to  $\lambda(1 + z)$  according to the redshift of the galaxy (the procedure is described in Appendix A of Ilbert et al. (2005)). For this reason the bluest absolute magnitude selection takes full advantage of the complete observed magnitude set. However, as the  $u$ -band flux has potentially larger photometric errors  $M_g$ -band magnitudes were preferred.

#### 4.2 Magnitude limit for both samples: $i < 22.5$

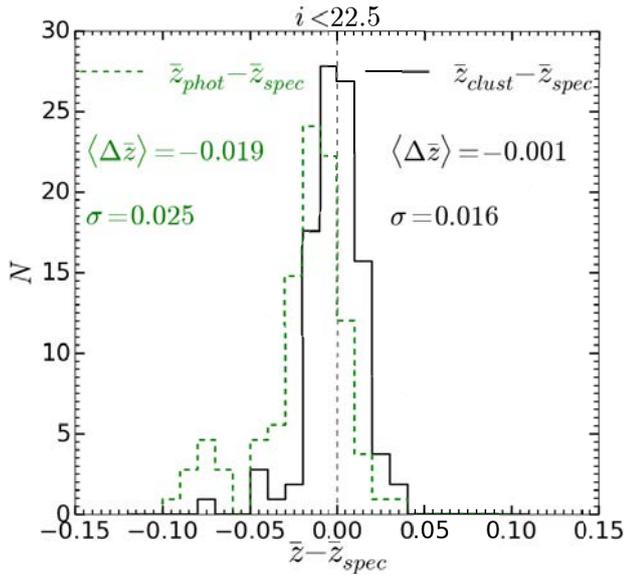
Here we select objects with  $i_{AB} < 22.5$  in the unknown population. The resulting sample contains  $N_u \sim 130\,000$  galaxies. We split them into 44 tomographic subsamples of 3 000 objects each to work at constant signal-to-noise ratio. Thus, we measure the integrated cross-correlation from 100 kpc to 2 Mpc in reference slices of width:  $\delta z = 0.005$ .

The results of these  $44 \times 120 = 5\,280$  measurements of  $\bar{\omega}_{ur}$  are translated into redshifts distributions following equation(9). Those distributions are finally summarized in Figure 3 where we show the  $(z_{cl}; z_{ph})$  plane. In this figure each vertical line correspond to a clustering redshift distribution measured in a tomographic sample of mean photometric redshift  $\bar{z}_{ph}$ .

This figure demonstrates our ability to measure the redshift distribution of a sample for which photometric redshifts are known and can be used to minimize the variation of  $\bar{b}_u$ . For each cluster- $z$  distribution the measured mean redshift



**Figure 6.** Left & middle: Mean redshift evolution through colors-space for both *reference* and *unknown* samples. Right: Difference between the estimated mean redshift - *cluster* or *photo-z* - and the *true* mean redshift from spectroscopic measurement.



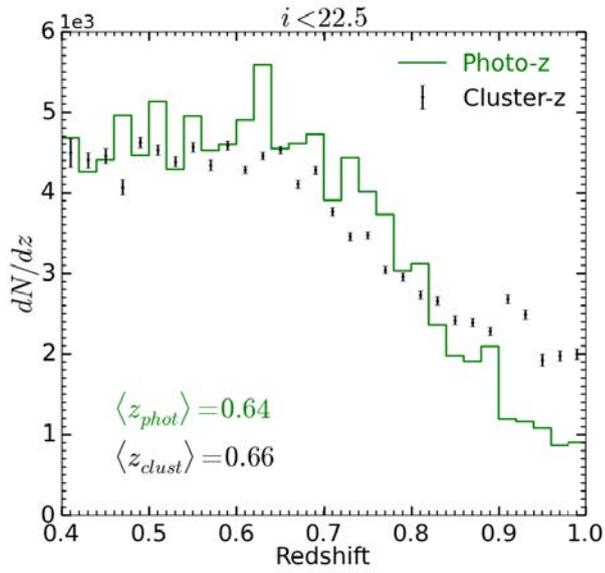
**Figure 7.** Histograms showing the residues distribution for *cluster-z* (black solid line) and *photo-z* (green dashed line). The *cluster-z* are almost 20 times less biased than the *photo-z* whereas their scatter are of the same order.

$\bar{z}_{cl}$  can be compared to the mean photometric redshift,  $\bar{z}_{ph}$ , see Figure 4. We see that both method are in relatively good agreement.

### 4.3 Fainter *unknown* sample: $22.5 < i < 23$

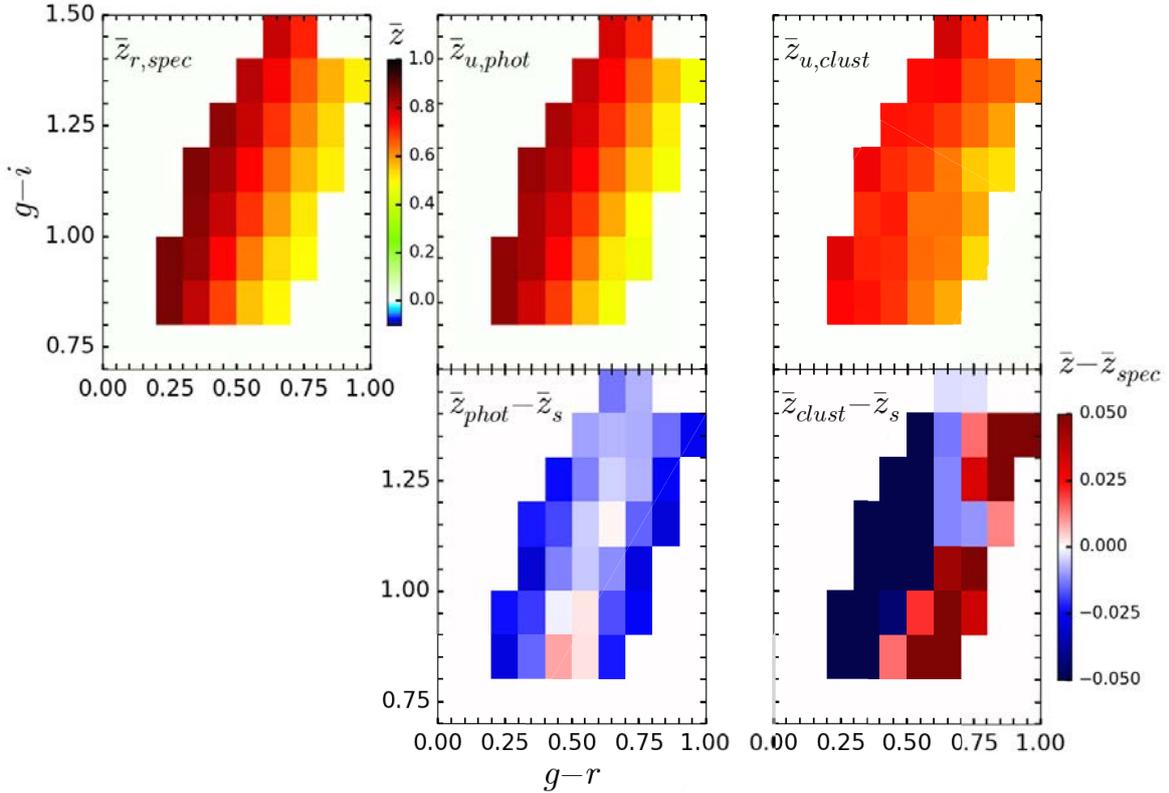
This section aims to show our ability to measure clustering redshifts when the *unknown* sample is fainter than the *reference* sample. Indeed, since we are not looking at the spectral energy distribution (SED) but on the clustering of objects and since all objects cluster with each others - regardless of their magnitude - we do expect a signal. Nevertheless, since at a given redshift, fainter objects are less massive we do expect the signal to be lower than in the previous case. To prove this, we use the same *reference* population than previously with  $i_r < 22.5$  but we select objects from the *unknown* sample with  $22.5 < i_u < 23$ . To detect a signal we build tomographic subsamples of  $N_u = 10\,000$  galaxies. The result is shown in Figure 5.

More over, beyond  $i = 22.5$  photometric redshift errors as well as the outlier fraction increase. This explain the relatively low signal-to-noise and the scatter of the distribution. As said previously, in the context of large imaging dark energy experiment, the requirements on *spectroscopic redshifts* are challenging. Such properties could be of great interest for cosmology in general and for large imaging dark



**Figure 8.** Overplot showing the global redshift distribution of the *unknown* sample measured by *cluster-z* (black points) and *photo-z* (green line).

energy experiments in particular. Measurements on others tomographic bins are available in Appendix A.



**Figure 9.** Same type of maps than Figure 6. Here the colors-redshift relation of the *reference* and *unknown* samples are expected to differ. The *photo-z* colors-redshift relation seems close to the *reference* sample. This is probably due to the limiting magnitude of the templates used in the *photometric redshifts* measurements, which is  $i < 22.5$ .

## 5 COLORS SAMPLING

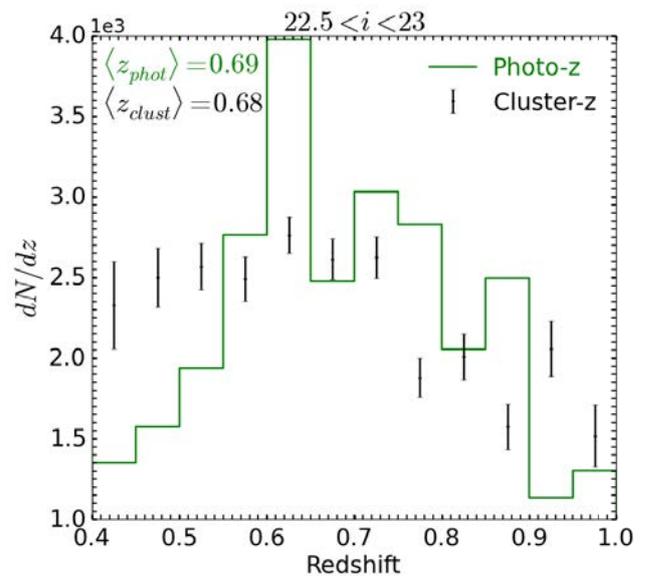
As we just seen, the tomographic sampling approach suffer from intrinsic systematics due to the use of the best-fit photometric redshift, e.g wrong balmer break detections, Lyman/Balmer confusion, extinctions etc... In this section we aim to free the clustering-based redshift estimation technique from the need of photometric redshifts and quantify the resulting accuracy.

### 5.1 Magnitude limit for both samples: $i < 22.5$

First, we look at an *unknown* population with same limiting magnitude than the *reference* sample. In this case, we expect the *reference* sample to be a representative sample of our *unknown population*. Then, colors-redshift relation of both samples should be the same in average.

To reduce the influence of the evolution with redshift of the galaxy-matter bias,  $\bar{b}_u(z)$ , we build subsamples in color-space. Working on the  $(g-i; g-r)$  plane we choose a binning size of  $\Delta_{g-r/i} = 0.1$  to get the smallest cells as possible containing at least 500 *unknown* galaxies. By construction the redshift distribution inside each of those cells will be more narrow than the one of the initial sample.

Then, we measure the clustering redshift distribution in each cell. All these distributions across the color space as



**Figure 10.** Overplot showing the global redshift distribution of the *unknown* sample measured by *cluster-z* (black points) and *photo-z* (green line).

well as their corresponding photometric redshift distributions are available in Appendix B. Figure 6 shows the mean redshift evolution through color space for both *reference* and *unknown* samples. This figure also shows for each cell the quantities:  $\bar{z}_{cl} - \bar{z}_{spec}$  and  $\bar{z}_{ph} - \bar{z}_{spec}$ .

To compare, in a more quantitative way, the ability of *cluster-z* to reproduce the colors-redshift relation compared to *photo-z* we look at the residues  $\bar{z}_{cl/ph} - \bar{z}_{spec}$  distributions, see Figure 7. *Cluster-z* are almost twenty times less biased than *photo-z* in the recovery of the *true* colors-redshift relation whereas their scatter are of the same order.

We can then combine all distributions of Appendix B and reconstruct the global clustering redshift distribution of the *unknown* population, see Figure 8. Based on results from Figure 7, we claim that the resulting clustering-redshift distribution is a better estimate of the *true* redshift distribution of the *unknown* sample than the one from photometric redshifts.

## 5.2 Fainter *unknown* sample: $22.5 < i < 23$

In this section we apply the same sampling approach in colors-space than previously but we use the fainter *unknown* sample defined in Section 4.3 with:  $22.5 < i_u < 23$ . This time, colors-redshift relation of both samples are expected to differ. Consequently we will not be able to check our results. Nevertheless, using the ability of *cluster-z* to well reconstruct redshift distributions of faint objects - see in Section 4.3 - and their higher accuracy in mapping the colors-redshift space we can reasonably be confident on those results.

Figure 9 shows the resulting colors-redshift maps. As expected the colors-redshift relation from *cluster-z* do not match with the one from the *reference* sample. Strangely, the *photo-z* colors-redshift relation seems close to the *reference* sample. This can be explained by the fact that the templates used for the photometric redshifts estimation are less representative of fainter objects. Clustering-redshift distributions in each of those cells are available in Appendix C.

Finally we combine all distributions of Appendix C and reconstruct the global clustering redshift distribution of the fainter *unknown* population, see Figure 10. As previously, based on results from previous sections we argue that the resulting clustering-redshift distribution is a better estimate of the *true* redshift distribution of the fainter *unknown* sample than the one from photometric redshifts.

## 6 SUMMARY

We have explored and quantified the ability of clustering-based redshift estimation in a real case study. The method used in this paper follows the one presented in M13 and R15. We have first demonstrated our ability to measure the clustering redshift distribution using a tomographic *photo-z* approach. Then we have investigated the potential of this method to estimate redshift distributions of samples fainter than the magnitude limit of the *reference* population. Then we have freed *cluster-z* from the need of *photo-z* by selecting subsamples in colors-space and shown that this method is  $\sim 20$  times better than photometric redshifts to recover the

*true* mean redshift of a distribution. Finally, we used the two last points to give a redshift distribution for a sample of faint objects -  $22.5 < i_{AB} < 23$  - and claim that this distribution is closer to the *true* distribution than the one from photometric redshifts.

## ACKNOWLEDGMENTS

We would like to thank Mubdi Rahman and Martin Kilbinger for useful discussions.

This paper uses data from the VIMOS Public Extragalactic Redshift Survey (VIPERS). VIPERS has been performed using the ESO Very Large Telescope, under the "Large Programme" 182.A-0886. The participating institutions and funding agencies are listed at <http://vipers.inaf.it>

This work is also based on observations obtained with MegaPrime/MegaCam, a joint project of CFHT and CEA/IRFU, at the Canada-France-Hawaii Telescope (CFHT) which is operated by the National Research Council (NRC) of Canada, the Institut National des Science de l'Univers of the Centre National de la Recherche Scientifique (CNRS) of France, and the University of Hawaii. This work is based in part on data products produced at Terapix available at the Canadian Astronomy Data Centre as part of the Canada-France-Hawaii Telescope Legacy Survey, a collaborative project of NRC and CNRS.

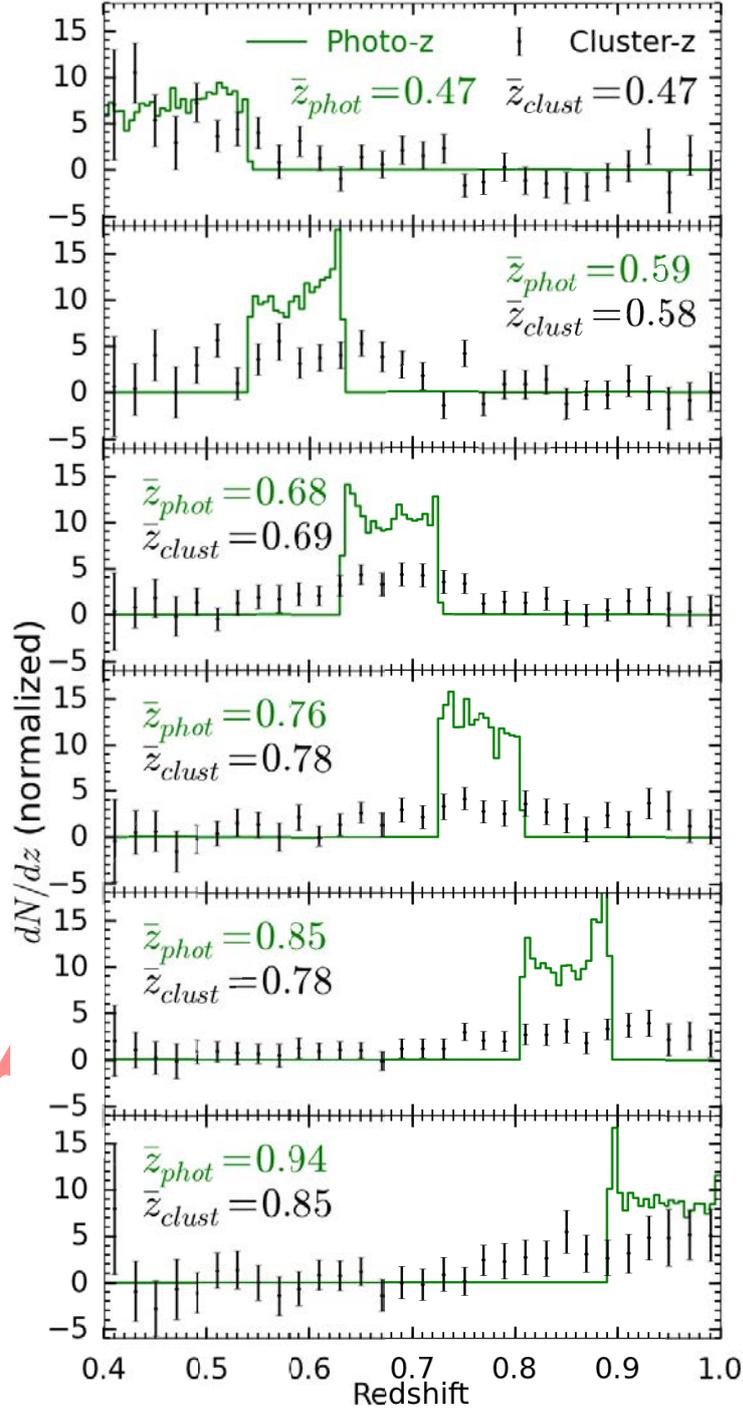
VS acknowledges funding from French Ministry for Research and Université Pierre et Marie-Curie (UPMC).

## REFERENCES

- Albrecht A., Bernstein G., Cahn R., Freedman W. L., Hewitt J., Hu W., Huth J., Kamionkowski M., Kolb E. W., Knox L., Mather J. C., Staggs S., Suntzeff N. B., 2006, ArXiv Astrophysics e-prints
- Bernstein G., Huterer D., 2010, MNRAS, 401, 1399
- Bertin E., Arnouts S., 1996, A&AS, 117, 393
- Calzetti D., Armus L., Bohlin R. C., Kinney A. L., Koornneef J., Storchi-Bergmann T., 2000, ApJ, 533, 682
- Coleman G. D., Wu C.-C., Weedman D. W., 1980, ApJS, 43, 393
- Connolly A. J., Csabai I., Szalay A. S., Koo D. C., Kron R. G., Munn J. A., 1995, AJ, 110, 2655
- Cooper M. C., Newman J. A., Croton D. J., Weiner B. J., Willmer C. N. A., Gerke B. F., Madgwick D. S., Faber S. M., Davis M., Coil A. L., Finkbeiner D. P., Guhathakurta P., Koo D. C., 2006, MNRAS, 370, 198
- Coupon J., Ilbert O., Kilbinger M., McCracken H. J., Mellier Y., Arnouts S., Bertin E., Hudelot 2009, A&A, 500, 981
- Guzzo L., Scodreggio M., Garilli B., Granett B. R., Fritz A., Abbas U., Adami C., Arnouts S., Bel J., Bolzonella M., Bottini D., Branchini E., Cappi A., Coupon J., Cucciati O., Davidzon I., De Lucia G., de la Torre 2014, A&A, 566, A108
- Hearin A. P., Zentner A. R., Ma Z., Huterer D., 2010, ApJ, 720, 1351
- Huterer D., Takada M., Bernstein G., Jain B., 2006, MNRAS, 366, 101

- Ilbert O., Arnouts S., McCracken H. J., Bolzonella M., Bertin E., Le Fèvre O., Mellier Y., 2006, *A&A*, 457, 841
- Ilbert O., Tresse L., Zucca E., Bardelli S., Arnouts S., Zamorani G., Pozzetti L., 2005, *A&A*, 439, 863
- Kinney A. L., Calzetti D., Bohlin R. C., McQuade K., Storchi-Bergmann T., Schmitt H. R., 1996, *ApJ*, 467, 38
- Knox L., Song Y.-S., Zhan H., 2006, *ApJ*, 652, 857
- Kron R. G., 1980, *ApJS*, 43, 305
- Landy S. D., Szalay A. S., Koo D. C., 1996, *ApJ*, 460, 94
- Le Fèvre O., Vettolani G., Garilli B., Tresse L., Bottini D., Le Brun V., Maccagni D., 2005, *A&A*, 439, 845
- Ma Z., Hu W., Huterer D., 2006, *ApJ*, 636, 21
- Matthews D. J., Newman J. A., 2010, *ApJ*, 721, 456
- Matthews D. J., Newman J. A., 2012, *ApJ*, 745, 180
- McQuinn M., White M., 2013, *MNRAS*, 433, 2857
- Ménard B., Scranton R., Schmidt S., Morrison C., Jeong D., Budavari T., Rahman M., 2013, *ArXiv e-prints*
- Newman J., Abate A., Abdalla F., Allam S., Allen S., Ansari R., Bailey S., Barkhouse W., Beers T., Blanton M., Brodwin M., Brownstein J., Brunner R., Carrasco-Kind M., Cervantes-Cota J., Chisari E., Colless 2013, *ArXiv e-prints*
- Newman J. A., 2008, *ApJ*, 684, 88
- Peebles P. J. E., Hauser M. G., 1974, *ApJS*, 28, 19
- Phillipps S., Shanks T., 1987, *MNRAS*, 229, 621
- Rahman M., Ménard B., Scranton R., Schmidt S. J., Morrison C. B., 2015, *MNRAS*, 447, 3500
- Schmidt S. J., Ménard B., Scranton R., Morrison C., McBride C. K., 2013, *MNRAS*, 431, 3307
- Seldner M., Peebles P. J. E., 1979, *ApJ*, 227, 30
- Sun L., Fan Z.-H., Tao C., Kneib J.-P., Jouvel S., Tilquin A., 2009, *ApJ*, 699, 958
- Szalay A. S., Connolly A. J., Szokoly G. P., 1999, *AJ*, 117, 68
- Zhan H., 2006, *J. Cosmology Astropart. Phys.*, 8, 8
- Zhan H., Knox L., 2006, *ApJ*, 644, 663

This paper has been typeset from a  $\text{\TeX}/\text{\LaTeX}$  file prepared by the author.

APPENDIX A: TOMOGRAPHIC SAMPLING:  $22.5 < i < 23$ 

**Figure A1.** Clustering redshift distributions for the tomographic bins build in Section 4.3. The large scatter are due to intrinsic errors of *photo-z* beyond  $i = 22.5$  reducing the tomographic approach efficiency. Nevertheless this shows that we are able to extract the signal.

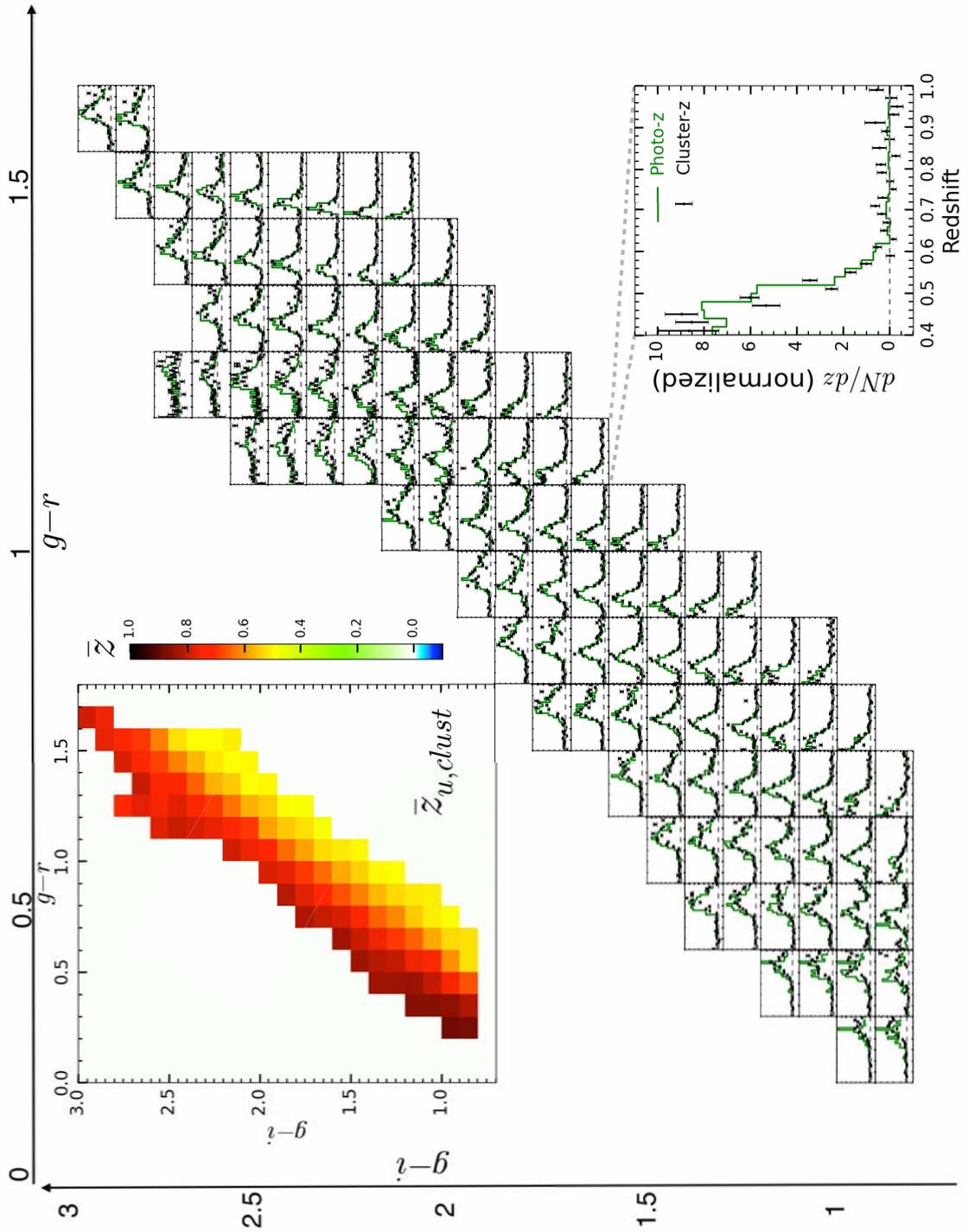
APPENDIX B: COLORS SAMPLING:  $i < 22.5$ 

Figure B1. Cluster- $z$  (black points) and  $photo-z$  distributions in each cells of the all colors-space.

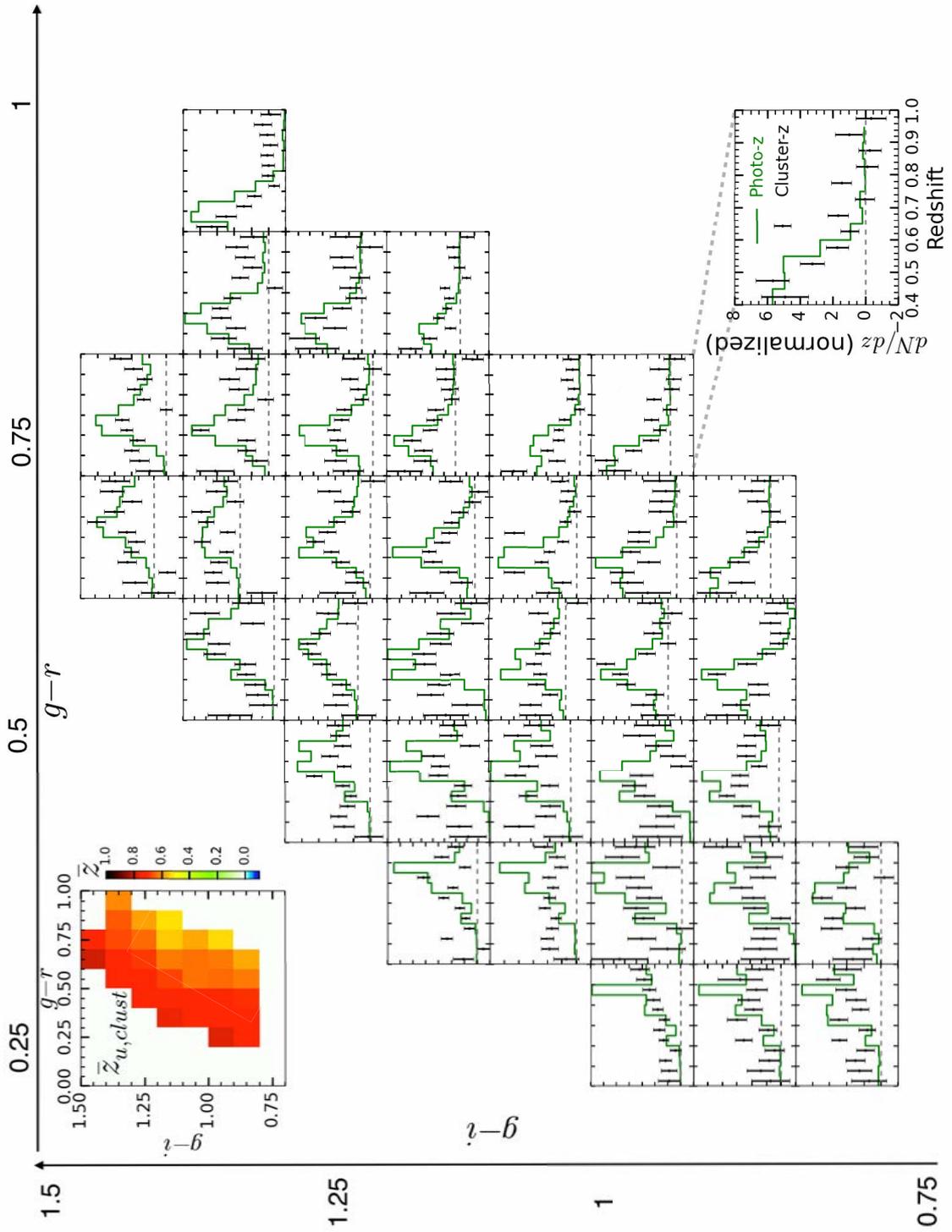
APPENDIX C: HUGE MAP  $22.5 < i < 23$ 


Figure C1. Cluster-z (black points) and photo-z distributions in each cells of the all colors-space.

# Bibliographie

---

- Arnouts, S., S. Cristiani, L. Moscardini, S. Matarrese, F. Lucchin, A. Fontana et E. Giallongo. 1999, «Measuring and modelling the redshift evolution of clustering : the Hubble Deep Field North», *MNRAS*, vol. 310, doi :10.1046/j.1365-8711.1999.02978.x, p. 540–556.
- Arnouts, S., L. Moscardini, E. Vanzella, S. Colombi, S. Cristiani, A. Fontana, E. Giallongo, S. Matarrese et P. Saracco. 2002, «Measuring the redshift evolution of clustering : the Hubble Deep Field South», *MNRAS*, vol. 329, doi : 10.1046/j.1365-8711.2002.04988.x, p. 355–366.
- Assef, R. J., C. S. Kochanek, M. Brodwin, M. J. I. Brown, N. Caldwell, R. J. Cool, P. Eisenhardt, D. Eisenstein, A. H. Gonzalez, B. T. Jannuzi, C. Jones, E. McKenzie, S. S. Murray et D. Stern. 2008, «Low-Resolution Spectral Templates for Galaxies from 0.2 to 10  $\mu\text{m}$ », *ApJ*, vol. 676, doi :10.1086/527533, p. 286–303.
- Assef, R. J., C. S. Kochanek, M. Brodwin, R. Cool, W. Forman, A. H. Gonzalez, R. C. Hickox, C. Jones, E. Le Floch, J. Moustakas, S. S. Murray et D. Stern. 2010, «Low-Resolution Spectral Templates for Active Galactic Nuclei and Galaxies from 0.03 to 30  $\mu\text{m}$ », *ApJ*, vol. 713, doi :10.1088/0004-637X/713/2/970, p. 970–985.
- Banerji, M., F. B. Abdalla, O. Lahav et H. Lin. 2008, «Photometric redshifts for the Dark Energy Survey and VISTA and implications for large-scale structure», *MNRAS*, vol. 386, doi :10.1111/j.1365-2966.2008.13095.x, p. 1219–1233.
- Barmby, P. et J. P. Huchra. 1998, «Kinematics of the Hercules supercluster», *AJ*, vol. 115, doi :10.1086/300179, p. 6.
- Bassett, B. et R. Hlozek. 2010, *Baryon Acoustic Oscillations*, Cambridge University Press, p. 246.
- Baum, W. A. 1962, «Photoelectric Magnitudes and Red-Shifts», dans *Problems of Extra-Galactic Research, IAU Symposium*, vol. 15, McVittie, G. C., p. 390.

- Benítez, N. 2000, «Bayesian Photometric Redshift Estimation», *ApJ*, vol. 536, doi : 10.1086/308947, p. 571–583.
- Bennett, C. L., M. Halpern, G. Hinshaw, N. Jarosik, A. Kogut, M. Limon, S. S. Meyer, L. Page, D. N. Spergel, G. S. Tucker, E. Wollack, E. L. Wright, C. Barnes, M. R. Greason, R. S. Hill, E. Komatsu, M. R. Nolta, N. Odegard, H. V. Peiris, L. Verde et J. L. Weiland. 2003, «First-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations : Preliminary Maps and Basic Results», *ApJS*, vol. 148, doi :10.1086/377253, p. 1–27.
- Benson, A. J., S. Cole, C. S. Frenk, C. M. Baugh et C. G. Lacey. 2000, «The nature of galaxy bias and clustering», *MNRAS*, vol. 311, doi :10.1046/j.1365-8711.2000.03101.x, p. 793–808.
- Bernardeau, F., S. Colombi, E. Gaztañaga et R. Scoccimarro. 2002, «Large-scale structure of the Universe and cosmological perturbation theory», *Phys. Rep.*, vol. 367, doi :10.1016/S0370-1573(02)00135-7, p. 1–248.
- Bolzonella, M., J.-M. Miralles et R. Pelló. 2000, «Photometric redshifts based on standard SED fitting procedures», *A&A*, vol. 363, p. 476–492.
- Boulade, O., X. Charlot, P. Abbon, S. Aune, P. Borgeaud, P.-H. Carton, M. Carty, D. Desforge, D. Epele, P. Gallais, L. Gosset, R. Granelli, M. Gros, J. de Kat, D. Loiseau, Y. Mellier, J. L. Ritou, J. Y. Rousse, P. Starzynski, N. Vignal et L. G. Vigroux. 2000, «Development of MegaCam, the next-generation wide-field imaging camera for the 3.6-m Canada-France-Hawaii Telescope», dans *Optical and IR Telescope Instrumentation and Detectors, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, vol. 4008, édité par M. Iye et A. F. Moorwood, p. 657–668.
- Brammer, G. B., P. G. van Dokkum et P. Coppi. 2008, «EAZY : A Fast, Public Photometric Redshift Code», *ApJ*, vol. 686, doi :10.1086/591786, p. 1503–1513.
- Branch, D. 1998, «Type IA Supernovae and the Hubble Constant», *ARA&A*, vol. 36, doi :10.1146/annurev.astro.36.1.17, p. 17–56.
- Branch, D., E. Baron et D. J. Jeffery. 2001, «Optical Spectra of Supernovae», *ArXiv Astrophysics e-prints*.
- Branch, D., P. Nugent et A. Fisher. 1997, «Type Ia supernovae as extragalactic distance indicators», dans *NATO Advanced Science Institutes (ASI) Series C, NATO Advanced Science Institutes (ASI) Series C*, vol. 486, Ruiz-Lapuente, P. and Canal, R. and Isern, J., p. 715.
- Bruzual A., G. 1983, «Spectral evolution of galaxies. I - Early-type systems», *ApJ*, vol. 273, doi :10.1086/161352, p. 105–127.

- Calzetti, D., L. Armus, R. C. Bohlin, A. L. Kinney, J. Koornneef et T. Storchi-Bergmann. 2000, «The Dust Content and Opacity of Actively Star-forming Galaxies», *ApJ*, vol. 533, doi :10.1086/308692, p. 682–695.
- Carrasco Kind, M. et R. J. Brunner. 2013, «TPZ : photometric redshift PDFs and ancillary information by using prediction trees and random forests», *MNRAS*, vol. 432, doi :10.1093/mnras/stt574, p. 1483–1501.
- Carroll, S. M., W. H. Press et E. L. Turner. 1992, «The cosmological constant», *ARA&A*, vol. 30, doi :10.1146/annurev.aa.30.090192.002435, p. 499–542.
- Castelló-Mor, N., X. Barcons et L. Ballo. 2011, «X-ray redshifts with the international x-ray observatory (ixo)», *Advances in Space Research*, vol. 48, n° 7, p. 1304–1310.
- Coleman, G. D., C.-C. Wu et D. W. Weedman. 1980, «Colors and magnitudes predicted for high redshift galaxies», *ApJS*, vol. 43, doi :10.1086/190674, p. 393–416.
- Colless, M., G. Dalton, S. Maddox, W. Sutherland, P. Norberg, S. Cole, J. Bland-Hawthorn, T. Bridges, R. Cannon, C. Collins, W. Couch, N. Cross, K. Deeley, R. De Propris, S. P. Driver, G. Efstathiou, R. S. Ellis, C. S. Frenk, K. Glazebrook, C. Jackson, O. Lahav, I. Lewis, S. Lumsden, D. Madgwick, J. A. Peacock, B. A. Peterson, I. Price, M. Seaborne et K. Taylor. 2001, «The 2dF Galaxy Redshift Survey : spectra and redshifts», *MNRAS*, vol. 328, doi :10.1046/j.1365-8711.2001.04902.x, p. 1039–1063.
- Collister, A., O. Lahav, C. Blake, R. Cannon, S. Croom, M. Drinkwater, A. Edge, D. Eisenstein, J. Loveday, R. Nichol, K. Pimbblet, R. de Propris, I. Roseboom, N. Ross, D. P. Schneider, T. Shanks et D. Wake. 2007, «MegaZ-LRG : a photometric redshift catalogue of one million SDSS luminous red galaxies», *MNRAS*, vol. 375, doi :10.1111/j.1365-2966.2006.11305.x, p. 68–76.
- Collister, A. A. et O. Lahav. 2004, «ANNz : Estimating Photometric Redshifts Using Artificial Neural Networks», *PASP*, vol. 116, doi :10.1086/383254, p. 345–351.
- Conley, A., J. Guy, M. Sullivan, N. Regnault, P. Astier, C. Balland, S. Basa, R. G. Carlberg, D. Fouchez, D. Hardin, I. M. Hook, D. A. Howell, R. Pain, N. Palanque-Delabrouille, K. M. Perrett, C. J. Pritchett, J. Rich, V. Ruhlmann-Kleider, D. Balam, S. Baumont, R. S. Ellis, S. Fabbro, H. K. Fakhouri, N. Fourmanoit, S. González-Gaitán, M. L. Graham, M. J. Hudson, E. Hsiao, T. Kronborg, C. Lidman, A. M. Mourao, J. D. Neill, S. Perlmutter, P. Riposte, N. Suzuki et E. S. Walker. 2011, «Supernova Constraints and Systematic Uncertainties from the First Three Years of the Supernova Legacy Survey», *ApJS*, vol. 192, doi : 10.1088/0067-0049/192/1/1, 1.

- Coupon, J., O. Ilbert, M. Kilbinger, H. J. McCracken, Y. Mellier, S. Arnouts, E. Bertin, P. Hudelot, M. Schultheis, O. Le Fèvre, V. Le Brun, L. Guzzo, S. Bardelli, E. Zucca, M. Bolzonella, B. Garilli, G. Zamorani, A. Zanichelli, L. Tresse et H. Aussel. 2009, «Photometric redshifts for the CFHTLS T0004 deep and wide fields», *A&A*, vol. 500, doi :10.1051/0004-6361/200811413, p. 981–998.
- Csabai, I., T. Budavári, A. J. Connolly, A. S. Szalay, Z. Gyóry, N. Benítez, J. Annis, J. Brinkmann, D. Eisenstein, M. Fukugita, J. Gunn, S. Kent, R. Lupton, R. C. Nichol et C. Stoughton. 2003, «The Application of Photometric Redshifts to the SDSS Early Data Release», *AJ*, vol. 125, doi :10.1086/345883, p. 580–592.
- Dahlen, T., B. Mobasher, M. Dickinson, H. C. Ferguson, M. Giavalisco, C. Kretchmer et S. Ravindranath. 2007, «Evolution of the Luminosity Function, Star Formation Rate, Morphology, and Size of Star-forming Galaxies Selected at Rest-Frame 1500 and 2800 Å», *ApJ*, vol. 654, doi :10.1086/508854, p. 172–185.
- Dahlen, T., B. Mobasher, R. S. Somerville, L. A. Moustakas, M. Dickinson, H. C. Ferguson et M. Giavalisco. 2005, «The Evolution of the Optical and Near-Infrared Galaxy Luminosity Functions and Luminosity Densities to  $z \sim 2$ », *ApJ*, vol. 631, doi :10.1086/432027, p. 126–144.
- Davis, M. et P. J. E. Peebles. 1983a, «A survey of galaxy redshifts. V - The two-point position and velocity correlations», *ApJ*, vol. 267, doi :10.1086/160884, p. 465–482.
- Davis, M. et P. J. E. Peebles. 1983b, «A survey of galaxy redshifts. V - The two-point position and velocity correlations», *ApJ*, vol. 267, doi :10.1086/160884, p. 465–482.
- Dodelson, S. 2003, *Modern Cosmology*, Academic Press, Academic Press. ISBN : 9780122191411.
- Dodelson, S., V. K. Narayanan, M. Tegmark, R. Scranton, T. Budavári, A. Connolly, I. Csabai, D. Eisenstein, J. A. Frieman, J. E. Gunn, L. Hui, B. Jain, D. Johnston, S. Kent, J. Loveday, R. C. Nichol, L. O’Connell, R. Scoccimarro, R. K. Sheth, A. Stebbins, M. A. Strauss, A. S. Szalay, I. Szapudi, M. S. Vogeley, I. Zehavi, J. Annis, N. A. Bahcall, J. Brinkman, M. Doi, M. Fukugita, G. Hennessy, Ž. Ivezić, G. R. Knapp, P. Kunszt, D. Q. Lamb, B. C. Lee, R. H. Lupton, J. A. Munn, J. Peoples, J. R. Pier, C. Rockosi, D. Schlegel, C. Stoughton, D. L. Tucker, B. Yanny et D. G. York. 2002, «The Three-dimensional Power Spectrum from Angular Clustering of Galaxies in Early Sloan Digital Sky Survey Data», *ApJ*, vol. 572, doi :10.1086/340225, p. 140–156.
- Eisenstein, D. J. et W. Hu. 1998, «Baryonic Features in the Matter Transfer Function», *ApJ*, vol. 496, doi :10.1086/305424, p. 605–614.

- Eisenstein, D. J., I. Zehavi, D. W. Hogg, R. Scoccimarro, M. R. Blanton, R. C. Nichol, R. Scranton, H.-J. Seo, M. Tegmark, Z. Zheng, S. F. Anderson, J. Annis, N. Bahcall, J. Brinkmann, S. Burles, F. J. Castander, A. Connolly, I. Csabai, M. Doi, M. Fukugita, J. A. Frieman, K. Glazebrook, J. E. Gunn, J. S. Hendry, G. Hennessy, Z. Ivezić, S. Kent, G. R. Knapp, H. Lin, Y.-S. Loh, R. H. Lupton, B. Margon, T. A. McKay, A. Meiksin, J. A. Munn, A. Pope, M. W. Richmond, D. Schlegel, D. P. Schneider, K. Shimasaku, C. Stoughton, M. A. Strauss, M. SubbaRao, A. S. Szalay, I. Szapudi, D. L. Tucker, B. Yanny et D. G. York. 2005, «Detection of the Baryon Acoustic Peak in the Large-Scale Correlation Function of SDSS Luminous Red Galaxies», *ApJ*, vol. 633, doi :10.1086/466512, p. 560–574.
- Faber, S. M. et R. E. Jackson. 1976, «Velocity dispersions and mass-to-light ratios for elliptical galaxies», *ApJ*, vol. 204, doi :10.1086/154215, p. 668–683.
- Feldmann, R., C. M. Carollo, C. Porciani, S. J. Lilly, P. Capak, Y. Taniguchi, O. Le Fèvre, A. Renzini, N. Scoville, M. Ajiki, H. Aussel, T. Contini, H. McCracken, B. Mobasher, T. Murayama, D. Sanders, S. Sasaki, C. Scarlata, M. Scodreggio, Y. Shioya, J. Silverman, M. Takahashi, D. Thompson et G. Zamorani. 2006, «The Zurich Extragalactic Bayesian Redshift Analyzer and its first application : COSMOS», *MNRAS*, vol. 372, doi :10.1111/j.1365-2966.2006.10930.x, p. 565–577.
- Fry, J. N. 1996, «The Evolution of Bias», *ApJ*, vol. 461, doi :10.1086/310006, p. L65.
- Gerdes, D. W., A. J. Sypniewski, T. A. McKay, J. Hao, M. R. Weis, R. H. Wechsler et M. T. Busha. 2010, «ArborZ : Photometric Redshifts Using Boosted Decision Trees», *ApJ*, vol. 715, doi :10.1088/0004-637X/715/2/823, p. 823–832.
- Guzzo, L., M. Scodreggio, B. Garilli, B. R. Granett, A. Fritz, U. Abbas, C. Adami, S. Arnouts, J. Bel, M. Bolzonella, D. Bottini, E. Branchini, A. Cappi, J. Coupon, O. Cucciati, I. Davidzon, G. De Lucia, S. de la Torre, P. Franzetti, M. Fumana, P. Hudelot, O. Ilbert, A. Iovino, J. Krywult, V. Le Brun, O. Le Fèvre, D. Maccagni, K. Małek, F. Marulli, H. J. McCracken, L. Paioro, J. A. Peacock, M. Polletta, A. Pollo, H. Schlegel, L. A. M. Tasca, R. Tojeiro, D. Vergani, G. Zamorani, A. Zanichelli, A. Burden, C. Di Porto, A. Marchetti, C. Marinoni, Y. Mellier, L. Moscardini, R. C. Nichol, W. J. Percival, S. Phleps et M. Wolk. 2014, «The VIMOS Public Extragalactic Redshift Survey (VIPERS). An unprecedented view of galaxies and large scale structure at  $0.5 < z < 1.2$ », *A&A*, vol. 566, doi :10.1051/0004-6361/201321489, A108.
- Hamilton, A. J. S. 1993, «Toward Better Ways to Measure the Galaxy Correlation Function», *ApJ*, vol. 417, doi :10.1086/173288, p. 19.
- Hamilton, A. J. S. et M. Tegmark. 2002, «The real-space power spectrum of the PSCz survey from 0.01 to  $300 h \text{ Mpc}^{-1}$ », *MNRAS*, vol. 330, doi :10.1046/j.1365-8711.2002.05033.x, p. 506–530.

- Heath, D. J. 1977, «The growth of density perturbations in zero pressure Friedmann-Lemaitre universes», *MNRAS*, vol. 179, p. 351–358.
- Hewett, P. C. 1982, «The estimation of galaxy angular correlation functions», *MNRAS*, vol. 201, p. 867–883.
- Hildebrandt, H., S. Arnouts, P. Capak, L. A. Moustakas, C. Wolf, F. B. Abdalla, R. J. Assef, M. Banerji, N. Benítez, G. B. Brammer, T. Budavári, S. Carliles, D. Coe, T. Dahlen, R. Feldmann, D. Gerdes, B. Gillis, O. Ilbert, R. Kotulla, O. Lahav, I. H. Li, J.-M. Miralles, N. Purger, S. Schmidt et J. Singal. 2010, «PHAT : PHoto-z Accuracy Testing», *A&A*, vol. 523, doi :10.1051/0004-6361/201014885, A31.
- Hoekstra, H. et B. Jain. 2008, «Weak Gravitational Lensing and Its Cosmological Applications», *Annual Review of Nuclear and Particle Science*, vol. 58, doi :10.1146/annurev.nucl.58.110707.171151, p. 99–123.
- Hsu, L.-T., M. Salvato, K. Nandra, M. Brusa, R. Bender, J. Buchner, J. L. Donley, D. D. Kocevski, Y. Guo, N. P. Hathi, C. Rangel, S. P. Willner, M. Brightman, A. Georgakakis, T. Budavári, A. S. Szalay, M. L. N. Ashby, G. Barro, T. Dahlen, S. M. Faber, H. C. Ferguson, A. Galametz, A. Grazian, N. A. Grogin, K.-H. Huang, A. M. Koekemoer, R. A. Lucas, E. McGrath, B. Mobasher, M. Peth, D. J. Rosario et J. R. Trump. 2014, «CANDELS/GOODS-S, CDFS, and ECDFS : Photometric Redshifts for Normal and X-Ray-Detected Galaxies», *ApJ*, vol. 796, doi :10.1088/0004-637X/796/1/60, 60.
- Huterer, D., M. Takada, G. Bernstein et B. Jain. 2006, «Systematic errors in future weak-lensing surveys : requirements and prospects for self-calibration», *MNRAS*, vol. 366, doi :10.1111/j.1365-2966.2005.09782.x, p. 101–114.
- Ilbert, O. 2004, *Formation et évolution de populations de galaxies à grand décalage spectral*, thèse de doctorat, Université de Provence - Aix-Marseille I, Laboratoire d'Astrophysique de Marseille.
- Ilbert, O., S. Arnouts, H. J. McCracken, M. Bolzonella, E. Bertin, O. Le Fèvre, Y. Mellier, G. Zamorani, R. Pellò, A. Iovino, L. Tresse, V. Le Brun, D. Bottini, B. Garilli, D. Maccagni, J. P. Picat, R. Scaramella, M. Scodeggio, G. Vettolani, A. Zanichelli, C. Adami, S. Bardelli, A. Cappi, S. Charlot, P. Ciliegi, T. Contini, O. Cucciati, S. Foucaud, P. Franzetti, I. Gavignaud, L. Guzzo, B. Marano, C. Marinoni, A. Mazure, B. Meneux, R. Merighi, S. Paltani, A. Pollo, L. Pozzetti, M. Radovich, E. Zucca, M. Bondi, A. Bongiorno, G. Busarello, S. de La Torre, L. Gregorini, F. Lamareille, G. Mathez, P. Merluzzi, V. Ripepi, D. Rizzo et D. Vergani. 2006, «Accurate photometric redshifts for the CFHT legacy survey calibrated using the VIMOS VLT deep survey», *A&A*, vol. 457, doi : 10.1051/0004-6361:20065138, p. 841–856.

- Jackson, J. C. 1972, «A critique of Rees's theory of primordial gravitational radiation», *MNRAS*, vol. 156, p. 1P.
- Jarrett, T. 2004, «Large scale structure in the local universe — the 2mass galaxy catalog», vol. 21, doi :10.1071/AS04050, p. 396–403, ISSN 1448-6083. URL [http://journals.cambridge.org/article\\_S132335800006640](http://journals.cambridge.org/article_S132335800006640).
- Kaiser, N. 1987, «Clustering in real space and in redshift space», *MNRAS*, vol. 227, p. 1–21.
- Kilbinger, M., K. Benabed, J. Guy, P. Astier, I. Tereno, L. Fu, D. Wraith, J. Coupon, Y. Mellier, C. Balland, F. R. Bouchet, T. Hamana, D. Hardin, H. J. McCracken, R. Pain, N. Regnault, M. Schultheis et H. Yahagi. 2009, «Dark-energy constraints and correlations with systematics from CFHTLS weak lensing, SNLS supernovae Ia and WMAP5», *A&A*, vol. 497, doi :10.1051/0004-6361/200811247, p. 677–688.
- Kinney, A. L., D. Calzetti, R. C. Bohlin, K. McQuade, T. Storchi-Bergmann et H. R. Schmitt. 1996, «Template Ultraviolet to Near-Infrared Spectra of Star-forming Galaxies and Their Application to K-Corrections», *ApJ*, vol. 467, doi : 10.1086/177583, p. 38.
- Kitching, T. D., A. Amara, F. B. Abdalla, B. Joachimi et A. Refregier. 2009, «Cosmological systematics beyond nuisance parameters : form-filling functions», *MNRAS*, vol. 399, doi :10.1111/j.1365-2966.2009.15408.x, p. 2107–2128.
- Kochanek, C. S., M. White, J. Huchra, L. Macri, T. H. Jarrett, S. E. Schneider et J. Mader. 2003, «Clusters of Galaxies in the Local Universe», *ApJ*, vol. 585, doi :10.1086/345896, p. 161–181.
- Komatsu, E., K. M. Smith, J. Dunkley, C. L. Bennett, B. Gold, G. Hinshaw, N. Jarosik, D. Larson, M. R.olta, L. Page, D. N. Spergel, M. Halpern, R. S. Hill, A. Kogut, M. Limon, S. S. Meyer, N. Odegard, G. S. Tucker, J. L. Weiland, E. Wollack et E. L. Wright. 2011, «Seven-year Wilkinson Microwave Anisotropy Probe (WMAP) Observations : Cosmological Interpretation», *ApJS*, vol. 192, doi : 10.1088/0067-0049/192/2/18, 18.
- Koo, D. C. 1985, «Optical multicolors - A poor person's Z machine for galaxies», *AJ*, vol. 90, doi :10.1086/113748, p. 418–440.
- Kopylova, F. G. et A. I. Kopylov. 2013, «Investigation of properties of galaxy clusters in the Hercules supercluster region», *Astronomy Letters*, vol. 39, doi :10.1134/S1063773712120043, p. 1–16.
- Kotulla, R., U. Fritze, P. Weilbacher et P. Anders. 2009, «GALEV evolutionary synthesis models - I. Code, input physics and web interface», *MNRAS*, vol. 396, doi :10.1111/j.1365-2966.2009.14717.x, p. 462–484.

- Kravtsov, A. V. et S. Borgani. 2012, «Formation of Galaxy Clusters», *ARA&A*, vol. 50, doi :10.1146/annurev-astro-081811-125502, p. 353–409.
- Labatie, A., J.-L. Starck, M. Lachièze-Rey et P. Arnalte-Mur. 2010, «Uncertainty in 2-point correlation function estimators and BAO detection in SDSS DR7», *ArXiv e-prints*.
- Landy, S. D. et A. S. Szalay. 1993, «Bias and variance of angular correlation functions», *ApJ*, vol. 412, doi :10.1086/172900, p. 64–71.
- Landy, S. D., A. S. Szalay et D. C. Koo. 1996, «Strong Angular Clustering of Very Blue Galaxies : Evidence of a Low-Redshift Population», *ApJ*, vol. 460, doi : 10.1086/176954, p. 94.
- Laureijs, R., J. Amiaux, S. Arduini, J. . Auguères, J. Brinchmann, R. Cole, M. Cropper, C. Dabin, L. Duvet, A. Ealet et et al. 2011, «Euclid Definition Study Report», *ArXiv e-prints*.
- Le Fèvre, O., G. Vettolani, B. Garilli, L. Tresse, D. Bottini, V. Le Brun, D. Maccagni, J. P. Picat, R. Scaramella, M. Scodeggio, A. Zanichelli, C. Adami, M. Arnaboldi, S. Arnouts, S. Bardelli, M. Bolzonella, A. Cappi, S. Charlot, P. Ciliegi, T. Contini, S. Foucaud, P. Franzetti, I. Gavignaud, L. Guzzo, O. Ilbert, A. Iovino, H. J. McCracken, B. Marano, C. Marinoni, G. Mathez, A. Mazure, B. Meneux, R. Merighi, S. Paltani, R. Pellò, A. Pollo, L. Pozzetti, M. Radovich, G. Zamorani, E. Zucca, M. Bondi, A. Bongiorno, G. Busarello, F. Lamareille, Y. Mellier, P. Merluzzi, V. Ripepi et D. Rizzo. 2005, «The VIMOS VLT deep survey. First epoch VVDS-deep survey : 11 564 spectra with  $17.5 \leq IAB \leq 24$ , and the redshift distribution over  $0 \leq z \leq 5$ », *A&A*, vol. 439, doi :10.1051/0004-6361:20041960, p. 845–862.
- Leavitt, H. S. 1908, «1777 variables in the Magellanic Clouds», *Annals of Harvard College Observatory*, vol. 60, p. 87–108.
- Limber, D. N. 1953, «The Analysis of Counts of the Extragalactic Nebulae in Terms of a Fluctuating Density Field.», *ApJ*, vol. 117, doi :10.1086/145672, p. 134.
- LSST Dark Energy Science Collaboration. 2012, «Large Synoptic Survey Telescope : Dark Energy Science Collaboration», *ArXiv e-prints*.
- Ma, Z., W. Hu et D. Huterer. 2006, «Effects of Photometric Redshift Uncertainties on Weak-Lensing Tomography», *ApJ*, vol. 636, doi :10.1086/497068, p. 21–29.
- Madau, P., H. C. Ferguson, M. E. Dickinson, M. Giavalisco, C. C. Steidel et A. Fruchter. 1996, «High-redshift galaxies in the Hubble Deep Field : colour selection and star formation history to  $z \sim 4$ », *MNRAS*, vol. 283, p. 1388–1404.

- Maller, A. H., D. H. McIntosh, N. Katz et M. D. Weinberg. 2005, «The Galaxy Angular Correlation Functions and Power Spectrum from the Two Micron All Sky Survey», *ApJ*, vol. 619, doi :10.1086/426181, p. 147–160.
- Martin, J. 2012, «Everything you always wanted to know about the cosmological constant problem (but were afraid to ask)», *Comptes Rendus Physique*, vol. 13, doi :10.1016/j.crhy.2012.04.008, p. 566–665.
- Matthews, D. J. et J. A. Newman. 2010, «Reconstructing Redshift Distributions with Cross-correlations : Tests and an Optimized Recipe», *ApJ*, vol. 721, doi : 10.1088/0004-637X/721/1/456, p. 456–468.
- Matthews, D. J. et J. A. Newman. 2012, «Improving Correlation Function Fitting with Ridge Regression : Application to Cross-correlation Reconstruction», *ApJ*, vol. 745, doi :10.1088/0004-637X/745/2/180, 180.
- McDonald, P., H. Trac et C. Contaldi. 2006, «Dependence of the non-linear mass power spectrum on the equation of state of dark energy», *MNRAS*, vol. 366, doi : 10.1111/j.1365-2966.2005.09881.x, p. 547–556.
- McQuinn, M. et M. White. 2013, «On using angular cross-correlations to determine source redshift distributions», *MNRAS*, vol. 433, doi :10.1093/mnras/stt914, p. 2857–2883.
- Mellier, Y. 1999, «Probing the Universe with Weak Lensing», *ARA&A*, vol. 37, doi :10.1146/annurev.astro.37.1.127, p. 127–189.
- Ménard, B., R. Scranton, S. Schmidt, C. Morrison, D. Jeong, T. Budavari et M. Rahman. 2013, «Clustering-based redshift estimation : method and application to data», *ArXiv e-prints*.
- Mo, H. et F. Van den Bosch. 2010, *Galaxy Formation and Evolution*, Cambridge, p. 820.
- Mohr, J. J., B. O’Shea, A. E. Evrard, J. Bialek et Z. Haiman. 2003, «Studying dark energy with galaxy cluster surveys», *Nuclear Physics B Proceedings Supplements*, vol. 124, doi :10.1016/S0920-5632(03)02078-4, p. 63–67.
- Newman, J. A. 2008, «Calibrating Redshift Distributions beyond Spectroscopic Limits with Cross-Correlations», *ApJ*, vol. 684, doi :10.1086/589982, p. 88–101.
- Peacock, J. A. et S. J. Dodds. 1996, «Non-linear evolution of cosmological power spectra», *MNRAS*, vol. 280, p. L19–L26.
- Peebles, P. J. E. 1980, *The large-scale structure of the universe*, Princeton University Press.

- Peebles, P. J. E. et M. G. Hauser. 1974, «Statistical Analysis of Catalogs of Extragalactic Objects. III. The Shane-Wirtanen and Zwicky Catalogs», *ApJS*, vol. 28, doi :10.1086/190308, p. 19.
- Percival, W. J. et M. White. 2009, «Testing cosmological structure formation using redshift-space distortions», *MNRAS*, vol. 393, doi :10.1111/j.1365-2966.2008.14211.x, p. 297–308.
- Perlmutter, S., G. Aldering, G. Goldhaber, R. A. Knop, P. Nugent, P. G. Castro, S. Deustua, S. Fabbro, A. Goobar, D. E. Groom, I. M. Hook, A. G. Kim, M. Y. Kim, J. C. Lee, N. J. Nunes, R. Pain, C. R. Pennypacker, R. Quimby, C. Lidman, R. S. Ellis, M. Irwin, R. G. McMahon, P. Ruiz-Lapuente, N. Walton, B. Schaefer, B. J. Boyle, A. V. Filippenko, T. Matheson, A. S. Fruchter, N. Panagia, H. J. M. Newberg, W. J. Couch et T. S. C. Project. 1999, «Measurements of  $\Omega$  and  $\Lambda$  from 42 High-Redshift Supernovae», *ApJ*, vol. 517, doi :10.1086/307221, p. 565–586.
- Peter, P. et J.-P. Uzan. 2012, *Cosmologie primordiale*, Belin.
- Phillipps, S. et T. Shanks. 1987, «On the variation of galaxy correlations with luminosity», *MNRAS*, vol. 229, p. 621–626.
- Planck Collaboration. 2014, «Planck 2013 results. I. Overview of products and scientific results», *A&A*, vol. 571, doi :10.1051/0004-6361/201321529, A1.
- Porciani, C. et P. Norberg. 2006, «Luminosity- and redshift-dependent quasar clustering», *MNRAS*, vol. 371, doi :10.1111/j.1365-2966.2006.10813.x, p. 1824–1834.
- Pound, R. V. et G. A. Rebka. 1959, «Gravitational red-shift in nuclear resonance», *Phys. Rev. Lett.*, vol. 3, doi :10.1103/PhysRevLett.3.439, p. 439–441. URL <http://link.aps.org/doi/10.1103/PhysRevLett.3.439>.
- Raccanelli, A., D. Bertacca, D. Pietrobon, F. Schmidt, L. Samushia, N. Bartolo, O. Doré, S. Matarrese et W. J. Percival. 2013, «Testing gravity using large-scale redshift-space distortions», *MNRAS*, vol. 436, doi :10.1093/mnras/stt1517, p. 89–100.
- Rahman, M., B. Ménard et R. Scranton. 2015a, «Exploring the 2MASS Extended and Point Source Catalogs with Clustering Redshifts», *ArXiv e-prints*.
- Rahman, M., B. Ménard, R. Scranton, S. J. Schmidt et C. B. Morrison. 2015b, «Clustering-based redshift estimation : comparison to spectroscopic redshifts», *MNRAS*, vol. 447, doi :10.1093/mnras/stu2636, p. 3500–3511.
- Riess, A. G., A. V. Filippenko, P. Challis, A. Clocchiatti, A. Diercks, P. M. Garnavich, R. L. Gilliland, C. J. Hogan, S. Jha, R. P. Kirshner, B. Leibundgut, M. M. Phillips, D. Reiss, B. P. Schmidt, R. A. Schommer, R. C. Smith, J. Spyromilio, C. Stubbs, N. B. Suntzeff et J. Tonry. 1998, «Observational

- Evidence from Supernovae for an Accelerating Universe and a Cosmological Constant», *AJ*, vol. 116, doi :10.1086/300499, p. 1009–1038.
- Russel, H. R. 1914, «The spectrum luminosity diagram», *Nature*, vol. 92, doi :<http://www.astrosurf.com/luxorion/Sciences/spectrum-luminosity-russell.pdf>, p. 252.
- Sachs, R. K. et A. M. Wolfe. 1967, «Perturbations of a Cosmological Model and Angular Variations of the Microwave Background», *ApJ*, vol. 147, doi :10.1086/148982, p. 73.
- Schmidt, S. J., B. Ménard, R. Scranton, C. Morrison et C. K. McBride. 2013, «Recovering redshift distributions with cross-correlations : pushing the boundaries», *MNRAS*, vol. 431, doi :10.1093/mnras/stt410, p. 3307–3318.
- Schneider, D. P., G. T. Richards, P. B. Hall, M. A. Strauss, S. F. Anderson, T. A. Boroson, N. P. Ross, Y. Shen, W. N. Brandt, X. Fan, N. Inada, S. Jester, G. R. Knapp, C. M. Krawczyk, A. R. Thakar, D. E. Vanden Berk, W. Voges, B. Yanny, D. G. York, N. A. Bahcall, D. Bizyaev, M. R. Blanton, H. Brewington, J. Brinkmann, D. Eisenstein, J. A. Frieman, M. Fukugita, J. Gray, J. E. Gunn, P. Hibon, Ž. Ivezić, S. M. Kent, R. G. Kron, M. G. Lee, R. H. Lupton, E. Malanushenko, V. Malanushenko, D. Oravetz, K. Pan, J. R. Pier, T. N. Price, III, D. H. Saxe, D. J. Schlegel, A. Simmons, S. A. Snedden, M. U. SubbaRao, A. S. Szalay et D. H. Weinberg. 2010, «The Sloan Digital Sky Survey Quasar Catalog. V. Seventh Data Release», *AJ*, vol. 139, doi :10.1088/0004-6256/139/6/2360, 2360.
- Schneider, P. 2015, *Extragalactic Astronomy and Cosmology*, Springer Berlin Heidelberg, doi :10.1007/978-3-642-54083-7.
- Seldner, M. et P. J. E. Peebles. 1979, «Statistical analysis of catalogs of extragalactic objects. XI - Evidence of correlation of QSOs and Lick galaxy counts», *ApJ*, vol. 227, doi :10.1086/156699, p. 30–36.
- Shen, Y., C. K. McBride, M. White, Z. Zheng, A. D. Myers, H. Guo, J. A. Kirkpatrick, N. Padmanabhan, J. K. Parejko, N. P. Ross, D. J. Schlegel, D. P. Schneider, A. Streblyanska, M. E. C. Swanson, I. Zehavi, K. Pan, D. Bizyaev, H. Brewington, G. Ebelke, V. Malanushenko, E. Malanushenko, D. Oravetz, A. Simmons et S. Snedden. 2013, «Cross-correlation of SDSS DR7 Quasars and DR10 BOSS Galaxies : The Weak Luminosity Dependence of Quasar Clustering at  $z \sim 0.5$ », *ApJ*, vol. 778, doi :10.1088/0004-637X/778/2/98, 98.
- Simon, P. 2007, «How accurate is Limber's equation ?», *A&A*, vol. 473, doi :10.1051/0004-6361:20066352, p. 711–714.
- Smith, R. E., J. A. Peacock, A. Jenkins, S. D. M. White, C. S. Frenk, F. R. Pearce, P. A. Thomas, G. Efstathiou et H. M. P. Couchman. 2003, «Stable clustering, the halo model and non-linear cosmological power spectra», *MNRAS*, vol. 341, doi :10.1046/j.1365-8711.2003.06503.x, p. 1311–1332.

- Smoot, G. F., C. L. Bennett, A. Kogut, E. L. Wright, J. Aymon, N. W. Boggess, E. S. Cheng, G. de Amici, S. Gulkis, M. G. Hauser, G. Hinshaw, P. D. Jackson, M. Janssen, E. Kaita, T. Kelsall, P. Keegstra, C. Lineweaver, K. Loewenstein, P. Lubin, J. Mather, S. S. Meyer, S. H. Moseley, T. Murdock, L. Rokke, R. F. Silverberg, L. Tenorio, R. Weiss et D. T. Wilkinson. 1992, «Structure in the COBE differential microwave radiometer first-year maps», *ApJ*, vol. 396, doi :10.1086/186504, p. L1–L5.
- Spergel, D. N., L. Verde, H. V. Peiris, E. Komatsu, M. R.olta, C. L. Bennett, M. Halpern, G. Hinshaw, N. Jarosik, A. Kogut, M. Limon, S. S. Meyer, L. Page, G. S. Tucker, J. L. Weiland, E. Wollack et E. L. Wright. 2003, «First-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations : Determination of Cosmological Parameters», *ApJS*, vol. 148, doi :10.1086/377226, p. 175–194.
- Steidel, C. C., M. Giavalisco, M. Dickinson et K. L. Adelberger. 1996, «Spectroscopy of Lyman Break Galaxies in the Hubble Deep Field», *AJ*, vol. 112, doi :10.1086/118019, p. 352.
- Takahashi, R., M. Sato, T. Nishimichi, A. Taruya et M. Oguri. 2012, «Revising the Halofit Model for the Nonlinear Matter Power Spectrum», *ApJ*, vol. 761, doi : 10.1088/0004-637X/761/2/152, 152.
- Tegmark, M. et P. J. E. Peebles. 1998, «The Time Evolution of Bias», *ApJ*, vol. 500, doi :10.1086/311426, p. L79–L82.
- The Dark Energy Survey Collaboration. 2005, «The Dark Energy Survey», *ArXiv Astrophysics e-prints*.
- Tully, R. B. et J. R. Fisher. 1977, «A new method of determining distances to galaxies», *A&A*, vol. 54, p. 661–673.
- Vargas-Magaña, M., J. E. Bautista, J.-C. Hamilton, N. G. Busca, É. Aubourg, A. Labatie, J.-M. Le Goff, S. Escoffier, M. Manera, C. K. McBride, D. P. Schneider et C. N. A. Willmer. 2013, «An optimized correlation function estimator for galaxy surveys», *A&A*, vol. 554, doi :10.1051/0004-6361/201220790, A131.
- Wang, Y., R. J. Brunner et J. C. Dolence. 2013, «The SDSS galaxy angular two-point correlation function», *MNRAS*, vol. 432, doi :10.1093/mnras/stt450, p. 1961–1979.
- Weinberg, S. 1972, *Gravitation and Cosmology : Principles and Applications of the General Theory of Relativity*, Wiley.
- Wojtak, R., T. M. Davis et J. Wiis. 2015, «Local gravitational redshifts can bias cosmological measurements», *J. Cosmology Astropart. Phys.*, vol. 7, doi :10.1088/1475-7516/2015/07/025, 025.

- York, D. G., J. Adelman, J. E. Anderson, Jr., S. F. Anderson, J. Annis, N. A. Bahcall, J. A. Bakken, R. Barkhouser, S. Bastian, E. Berman, W. N. Boroski, S. Bracker, C. Briegel, J. W. Briggs, J. Brinkmann, R. Brunner, S. Burles, L. Carey, M. A. Carr, F. J. Castander, B. Chen, P. L. Colestock, A. J. Connolly, J. H. Crocker, I. Csabai, P. C. Czarapata, J. E. Davis, M. Doi, T. Dombeck, D. Eisenstein, N. Ellman, B. R. Elms, M. L. Evans, X. Fan, G. R. Federwitz, L. Fiscelli, S. Friedman, J. A. Frieman, M. Fukugita, B. Gillespie, J. E. Gunn, V. K. Gurbani, E. de Haas, M. Haldeman, F. H. Harris, J. Hayes, T. M. Heckman, G. S. Hennessy, R. B. Hindsley, S. Holm, D. J. Holmgren, C.-h. Huang, C. Hull, D. Husby, S.-I. Ichikawa, T. Ichikawa, Ž. Ivezić, S. Kent, R. S. J. Kim, E. Kinney, M. Klaene, A. N. Kleinman, S. Kleinman, G. R. Knapp, J. Korienek, R. G. Kron, P. Z. Kunszt, D. Q. Lamb, B. Lee, R. F. Leger, S. Limmongkol, C. Lindenmeyer, D. C. Long, C. Loomis, J. Loveday, R. Lucinio, R. H. Lupton, B. MacKinnon, E. J. Mannery, P. M. Mantsch, B. Margon, P. McGehee, T. A. McKay, A. Meiksin, A. Merelli, D. G. Monet, J. A. Munn, V. K. Narayanan, T. Nash, E. Neilsen, R. Neswold, H. J. Newberg, R. C. Nichol, T. Nicinski, M. Nonino, N. Okada, S. Okamura, J. P. Ostriker, R. Owen, A. G. Pauls, J. Peoples, R. L. Peterson, D. Petravick, J. R. Pier, A. Pope, R. Pordes, A. Prosapio, R. Rechenmacher, T. R. Quinn, G. T. Richards, M. W. Richmond, C. H. Rivetta, C. M. Rockosi, K. Ruthmansdorfer, D. Sandford, D. J. Schlegel, D. P. Schneider, M. Sekiguchi, G. Sergey, K. Shimasaku, W. A. Siegmund, S. Smee, J. A. Smith, S. Snedden, R. Stone, C. Stoughton, M. A. Strauss, C. Stubbs, M. SubbaRao, A. S. Szalay, I. Szapudi, G. P. Szokoly, A. R. Thakar, C. Tremonti, D. L. Tucker, A. Uomoto, D. Vanden Berk, M. S. Vogeley, P. Waddell, S.-i. Wang, M. Watanabe, D. H. Weinberg, B. Yanny, N. Yasuda et SDSS Collaboration. 2000, «The Sloan Digital Sky Survey : Technical Summary», *AJ*, vol. 120, doi :10.1086/301513, p. 1579–1587.
- Zheng, W., V. J. Mikles, V. Mainieri, G. Hasinger, P. Rosati, C. Wolf, C. Norman, G. Szokoly, R. Gilli, P. Tozzi, J. X. Wang, A. Zirm et R. Giacconi. 2004, «Photometric Redshift of X-Ray Sources in the Chandra Deep Field-South», *ApJS*, vol. 155, doi :10.1086/424385, p. 73–87.