

# Les forêts Lyman- $\alpha$ du relevé eBOSS : comprendre les fonctions de corrélation et les systématiques

**Thèse de doctorat de l'Université Paris-Saclay**

École doctorale n° 576, Particules, Hadrons, Énergie, Noyau :  
Instrumentation, Imagerie, Cosmos et Simulation (PHENIICS)

Spécialité de doctorat : astroparticules et cosmologie

Unité de recherche : Université Paris-Saclay, CEA, Département de Physique  
des Particules, 91191, Gif-sur-Yvette, France.

Référent : faculté des sciences d'Orsay

**Thèse présentée et soutenue à l'Orme des merisiers,  
le 24 septembre 2020, par**

**Thomas ETOURNEAU**

## **Composition du jury :**

<b>Réza Ansari</b>	Président
Professeur, Université Paris-Saclay	
<b>Delphine Hardin</b>	Rapportrice
Professeure, Sorbonne Université	
<b>Matthew Pieri</b>	Rapporteur
Professeur, Université Aix-Marseille	
<b>Andreu Font-Ribera</b>	Examinateur
Docteur, Institut de Física d'Altes Energies	
<b>Jean-Marc Le Goff</b>	Directeur de thèse
Docteur, IRFU, CEA, Université Paris Saclay	

*Dedication*



# Remerciements

---

merci



# Préambule

---

Le siècle dernier a été extrêmement fécond pour la physique fondamentale : les théories de la mécanique quantique et de la relativité ont conduit au développement du modèle standard de la physique des particules. Par ailleurs, la théorie de la relativité générale a fourni le cadre pour le développement du modèle standard de la cosmologie. Ces deux branches ont révolutionné notre compréhension de l'univers en changeant complètement le paradigme dans lequel nous nous trouvions.

Au cours des cinquante dernières années, les observations se sont accumulées, renforçant encore et encore notre compréhension de la cosmologie. Nous pouvons citer en particulier la détection du fond diffus cosmologique en 1964 qui a confirmé l'expansion de l'univers. Ceci a encouragé des projets ambitieux d'observation du ciel, comme par exemple le satellite Cosmic Background Explorer (COBE) lancé en 1989, qui a détecté pour la première fois les anisotropies du fond diffus cosmologique correspondant aux perturbations primordiales de densité. En 1998, l'observation de l'accélération de l'expansion de l'univers a stimulé encore davantage les observations afin de comprendre l'origine de cette accélération. Le programme d'observation Sloan Digital Sky Survey (SDSS) s'inscrit dans ce cadre là. Ce programme entreprend l'observation d'un nombre faramineux d'objets célestes lointains : plusieurs millions de spectres de galaxies et de quasars sont mesurés. Grâce à ces projets de grande ampleur, de nouvelles sondes cosmologiques voient le jour. C'est le cas de l'échelle des oscillations acoustiques de baryon (BAO) qui devient mesurable grâce à l'énorme quantité de données produite par SDSS.

Ce manuscrit présente les travaux menés au long de ma thèse de doctorat. Celle-ci s'inscrit dans le cadre du relevé eBOSS de SDSS et du futur relevé DESI. Au cours de cette thèse, j'ai participé à l'analyse des données finale du relevé eBOSS, conduite au sein du groupe Lyman alpha ( $\text{Ly}\alpha$ ). Cette analyse utilise les forêts  $\text{Ly}\alpha$  présentes dans les spectres de quasars mesurés par eBOSS. Elle consiste à mesurer la fonction d'auto-corrélation du  $\text{Ly}\alpha$  et la fonction de corrélation croisée entre le  $\text{Ly}\alpha$  et les quasars. L'ajustement de ces fonctions de corrélation permet de mesurer l'échelle BAO à grand redshift, et ainsi mesurer les rapports  $D_{\text{H}}(z)/r_{\text{d}}$  et  $D_{\text{M}}(z)/r_{\text{d}}$ . La mesure de ces rapports permet de contraindre les quantités  $(\Omega_m, \Omega_\Lambda, H_0 r_{\text{d}})$ , relatives à l'expansion de l'univers.

Ma contribution à cette analyse est le développement de simulations, destinées à tester les procédures de mesure des fonctions de corrélation et d'estimation de leur covariance ainsi que les modèles utilisés pour ajuster ces fonctions de corrélation. Ces simulations reproduisent les données fournies par eBOSS : un relevé de quasars dont les forêts  $\text{Ly}\alpha$  présentes dans les spectres simulés possèdent les bonnes fonctions d'auto-corrélation et de corrélation croisée. Ces données synthétiques permettent de

---

tester l’analyse menée sur les vraies données et ainsi d’identifier si cette analyse est affectée par des systématiques. Ces simulations sont développées conjointement avec le groupe Ly $\alpha$  de la collaboration DESI. Elles seront aussi utilisées pour tester l’analyse des futures données de DESI.

En plus de tester l’analyse BAO menée sur les données d’eBOSS, j’ai utilisé les simulations développées durant ma thèse pour étudier la faisabilité d’une analyse des distorsions dans l’espace des redshifts (RSD) en utilisant les données Ly $\alpha$  du relevé eBOSS ou DESI. Ces analyses permettent de contraindre la gravité aux grandes échelles via la mesure du paramètre  $f\sigma_8$ . A l’aide de ces simulations, je m’intéresse donc à la stabilité de la mesure du biais  $b_{\text{Ly}\alpha}$  et du paramètre RSD  $\beta_{\text{Ly}\alpha}$  du Ly $\alpha$  dans la fonction d’auto-corrélation du Ly $\alpha$  issue des données eBOSS, dans le but d’une éventuelle analyse RSD.

Ce manuscrit s’articule comme suit : le chapitre premier est une introduction à la cosmologie moderne et aux éléments nécessaires à la compréhension de ce manuscrit. Puis, le relevé eBOSS qui fait partie de la troisième génération de SDSS est présenté. S’en suit la description du traitement de ces données, nécessaire à l’estimation des fonctions de corrélation. Le chapitre quatre présente le cœur de mon travail : le développement des simulations utilisées dans les analyses Ly $\alpha$  d’eBOSS et DESI. Puis, l’analyse et la validation de ces simulations sont exposées. Enfin, le dernier chapitre présente le résultat de l’analyse des données finale d’eBOSS. Une comparaison avec l’analyse des mocks y est proposée. Une critique de la modélisation utilisée et des propositions d’amélioration y sont également données.





# Table des matières

---

<b>1</b>	<b>Introduction à la cosmologie</b>	<b>5</b>
1.1	Qu'est-ce que la cosmologie ? . . . . .	6
1.2	Le modèle $\Lambda$ CDM . . . . .	9
1.2.1	Mise en place du modèle . . . . .	9
1.2.2	Description du modèle . . . . .	11
1.3	La fonction de corrélation de la matière . . . . .	20
1.3.1	Une analogie avec le son . . . . .	20
1.3.2	Le spectre de puissance . . . . .	22
1.3.3	La fonction de corrélation . . . . .	24
1.4	Les oscillations acoustiques de baryon . . . . .	27
1.4.1	La genèse . . . . .	27
1.4.2	Mesurer l'échelle BAO . . . . .	29
1.4.3	Contraintes cosmologiques . . . . .	30
1.5	Traceurs de la matière . . . . .	31
1.5.1	Traceurs et biais . . . . .	32
1.5.2	Distorsions dans l'espace des redshifts . . . . .	33
1.5.3	Les quasars . . . . .	34
1.5.4	La forêt Ly $\alpha$ . . . . .	37
1.5.5	Les contaminants . . . . .	41
<b>2</b>	<b>Le relevé eBOSS de SDSS</b>	<b>43</b>
2.1	Le Sloan Digital Sky Survey . . . . .	44
2.1.1	SDSS I-II-III . . . . .	44
2.1.2	SDSS-IV . . . . .	45
2.2	eBOSS . . . . .	46
2.2.1	Les traceurs de eBOSS . . . . .	46
2.2.2	Sélection des cibles . . . . .	47
2.2.3	Pavage du ciel . . . . .	48
2.2.4	Phase d'observation . . . . .	49
2.2.5	Caractéristiques techniques de l'instrument . . . . .	49
2.2.6	Résultats . . . . .	55
2.3	DESI . . . . .	55

## TABLE DES MATIÈRES

---

<b>3 Présentation des données</b>	<b>59</b>
3.1 Réduction des données . . . . .	60
3.2 Le catalogue de quasar . . . . .	60
3.3 La sélection des forêts . . . . .	62
3.4 Définition du champ d'absorption . . . . .	64
3.4.1 Calcul des $\delta$ . . . . .	64
3.4.2 Prise en compte de l'ajustement du continuum . . . . .	65
<b>4 Développement des mocks</b>	<b>67</b>
4.1 Objectifs des mocks . . . . .	68
4.2 Construction des mocks . . . . .	69
4.2.1 Les champs de densité . . . . .	69
4.2.2 Le relevé de quasars . . . . .	73
4.2.3 Création des lignes de visée . . . . .	77
4.2.4 De la densité à l'absorption . . . . .	79
4.2.5 Ajout des HCD . . . . .	88
4.3 Production des mocks . . . . .	90
4.4 Ajustement des paramètres . . . . .	95
4.5 Expansion des mocks . . . . .	100
<b>5 Validation des mocks</b>	<b>103</b>
5.1 Les estimateurs . . . . .	104
5.1.1 L'auto-corrélation Ly $\alpha$ × Ly $\alpha$ . . . . .	105
5.1.2 La corrélation croisée Ly $\alpha$ × QSO . . . . .	106
5.1.3 Le spectre de puissance à une dimension . . . . .	106
5.1.4 La fonction de corrélation à une dimension . . . . .	107
5.1.5 L'auto-corrélation QSO × QSO . . . . .	107
5.2 Les matrices de distorsion . . . . .	108
5.3 Les matrices de covariance . . . . .	109
5.4 Modélisation des fonctions de corrélation . . . . .	110
5.4.1 Modélisation des données . . . . .	110
5.4.2 Modélisation des mocks . . . . .	116
5.5 Analyse des mocks . . . . .	119
5.5.1 Analyse des raw mocks . . . . .	119
5.5.2 Analyse des mocks eboss-0.0 . . . . .	128
5.5.3 Analyse des mocks eboss-0.2 . . . . .	132
5.5.4 Analyse des mocks eboss-0.3 . . . . .	136
<b>6 Analyse des données et étude des systématiques</b>	<b>139</b>
6.1 L'analyse des données DR16 . . . . .	140
6.1.1 Résultats . . . . .	140
6.1.2 Stabilité des paramètres Ly $\alpha$ . . . . .	143

---

## TABLE DES MATIÈRES

6.2	Étude de la modélisation des HCD . . . . .	144
6.2.1	Comparaison des mocks . . . . .	145
6.2.2	Effet du masquage des HCD . . . . .	146
6.2.3	Une meilleure modélisation des HCD . . . . .	149
6.2.4	Mieux comprendre les HCD . . . . .	153
<b>7</b>	<b>Conclusion</b>	<b>163</b>
<b>Appendices</b>		<b>167</b>
<b>A</b>	<b>Fluctuating Gunn Peterson Approximation</b>	<b>168</b>



# 1

# Introduction à la cosmologie

---

## Sommaire

---

<b>1.1</b>	<b>Qu'est-ce que la cosmologie ?</b>	<b>6</b>
<b>1.2</b>	<b>Le modèle <math>\Lambda</math>CDM</b>	<b>9</b>
1.2.1	Mise en place du modèle	9
1.2.2	Description du modèle	11
La métrique		11
Le redshift		12
Les équations d'Einstein		13
Les équations de Friedmann-Lemaître		14
Évolution de l'univers		15
Les distances		18
Les paramètres du modèle		19
<b>1.3</b>	<b>La fonction de corrélation de la matière</b>	<b>20</b>
1.3.1	Une analogie avec le son	20
1.3.2	Le spectre de puissance	22
1.3.3	La fonction de corrélation	24
<b>1.4</b>	<b>Les oscillations acoustiques de baryon</b>	<b>27</b>
1.4.1	La genèse	27
1.4.2	Mesurer l'échelle BAO	29
1.4.3	Contraintes cosmologiques	30
<b>1.5</b>	<b>Traceurs de la matière</b>	<b>31</b>
1.5.1	Traceurs et biais	32
1.5.2	Distorsions dans l'espace des redshifts	33
1.5.3	Les quasars	34
1.5.4	La forêt Ly $\alpha$	37
1.5.5	Les contaminants	41

---

Ce premier chapitre présente la cosmologie moderne et explique brièvement sa construction au fil du siècle dernier. L'idée est de donner une vue d'ensemble du paradigme actuel, tout en détaillant davantage les points clés nécessaires à ce manuscrit. Pour une étude approfondie de la cosmologie moderne, nous référons le lecteur aux ouvrages suivants : RICH (2010), DODELSON (2003).

## 1.1 Qu'est-ce que la cosmologie ?

Le terme cosmogonie (du grec *cosmo-* : monde ; *gon-* : engendrer) désigne une conception et tentative d'explication de la naissance du monde, et parfois de l'Homme. Il existe une grande nombre de cosmogonies, très souvent d'origines religieuses. Nous pouvons citer par exemple la cosmogonie hindoue, dans laquelle le monde est vu comme un cycle : le dieu Brahma crée le monde lorsqu'il se réveille, et le détruit lorsqu'il s'endort. Notre univers correspond ainsi à une journée de Brahma, débutant lorsque Brahma ouvre les yeux et prenant fin lorsqu'il les referme. Le monde suit ainsi une suite de créations et de destructions. Nous pouvons aussi citer la cosmogonie abrahamique, décrite dans la Genèse. Cette cosmogonie est commune au judaïsme, au christianisme, et à l'islam. Dans cette cosmogonie, le dieu créateur, intemporel, conçoit le monde en 7 jours. Il commence par créer la lumière le premier jour. Il termine par créer l'Homme à son image le sixième jour, puis se repose le dernier jour.

Nous pourrions passer la totalité de ce manuscrit à décrire diverses cosmogonies. Mais celle qui nous intéresse et que nous allons détailler ici est la cosmogonie scientifique : la *cosmologie*. La cosmologie est donc l'étude de l'univers, son origine, ses constituants et son devenir, dans le cadre de la méthode scientifique. Même si aujourd'hui la cosmologie fait consensus au sein des scientifiques en ce qui concerne la compréhension de l'univers, cela n'a pas toujours été le cas. Pendant longtemps les croyances religieuses ont dominé, allant jusqu'à limiter voire interdire les avancées scientifiques. Il faut attendre le XVI<sup>e</sup> siècle pour que Copernic propose le modèle héliocentrique, soit presque 2000 ans après le modèle géocentrique d'Aristote, soutenu par l'église et les savants jusqu'alors. Par la suite, les observations de Galilée, les travaux de Kepler ainsi que l'émancipation des dogmes religieux ont permis au modèle héliocentrique, fondé sur les lois de Kepler, de s'imposer. Cela a aussi permis à Newton



FIGURE 1.1 – Gauche : illustration artistique de la cosmologie hindoue. Droite : couverture du livre de la Genèse, Bible de Saint-Paul-hors-des-Murs, vers 870.

de proposer sa théorie de la gravitation peu de temps après. Cette période marque la naissance de la physique et de la cosmologie.

Jusqu'au XIX<sup>e</sup> siècle, le modèle héliocentrique décrivant l'univers comme se limitant à notre système solaire fait consensus. Puis émerge l'idée que les étoiles sont d'autres systèmes solaires, notamment grâce aux premières mesures de distance d'étoiles proches<sup>1</sup>. L'idée de galaxie, un système rassemblant une multitude de systèmes solaires, fait aussi son apparition, nous conduisant vers un paradigme de moins en moins anthropocentrique.

La cosmologie moderne naît réellement au début du XX<sup>e</sup> siècle. En 1915, Einstein propose sa théorie de la gravitation : la *relativité générale*. Elle offre une vision radicalement différente de la théorie bien établie de Newton. La gravitation n'est plus vue comme une force instantanée entre les corps massifs mais comme une déformation de l'espace temps se propageant à la vitesse de la lumière. La théorie d'Einstein prédit correctement l'avance du périhélie de Mercure, dont la valeur était jusque-là incomprise. Puis en 1919 lors d'une éclipse de Soleil, la déviation de la lumière par un corps massif, prédiction directe de la relativité générale et non présente dans la théorie de Newton, est observée. Non seulement la déviation de la lumière est observée pour la première fois, de plus l'angle de déviation observé correspond à celui prédit par la théorie. Ceci assoit au sein de la communauté scientifique la théorie d'Einstein en tant que nouvelle théorie de la gravitation.

Par ailleurs, la cosmologie observationnelle connaît des avancées remarquables, notamment grâce à Edwin Hubble qui observe le décalage vers le rouge, ou *redshift*<sup>2</sup>, du spectre d'objets lointains, dû à leur vitesse d'éloignement. Il comprend aussi que les objets étendus, jusque-là interprétés comme des nuages de poussière et de gaz et appelés nébuleuses, sont d'autres galaxies semblables à la nôtre. Parallèlement, Alexandre Friedmann résout en 1922 les équations d'Einstein de la relativité générale pour un univers homogène et isotrope et trouve une solution d'univers en expansion, qui contraste avec l'idée d'un univers statique et éternel jusque-là ancrée dans les esprits. Enfin, Georges Lemaître effectue le lien entre tous ces éléments. En 1927, il publie un papier expliquant que l'éloignement des galaxies et le décalage vers le rouge de leur spectre pouvait être expliqué par une théorie d'univers en expansion, et donne la première estimation de la constante de Hubble<sup>3</sup>. En 1929, Edwin Hubble publie son célèbre article, exposant la loi de Hubble et favorisant très fortement le modèle d'univers en expansion.

Nous pouvons noter ici qu'après avoir publié sa théorie, Einstein ajoute dans ses équations une constante *ad hoc*, dite *constante cosmologique*, et noté  $\Lambda$ . Cette constante est rajoutée afin de rendre les solutions à ses équations capables de décrire un univers statique (idée dominante de l'époque). Puis, suite à la publication de Hubble, Einstein retire la constante cosmologique de ses équations et la qualifie de « plus grande bêtise de sa vie ». L'ironie fait qu'en 1998, la constante cosmologique est réintroduite dans les modèles afin d'expliquer l'observation de l'accélération de l'expansion de l'univers. Les mesures estiment que la densité d'énergie correspondant à cette constante cosmologique représente environ 70 % de l'énergie totale de notre univers aujourd'hui. Cependant, il existe des modèles plus

---

1. Par exemple la mesure de la distance de 61 Cygni par Bessel en 1838.

2. Voir explication du redshift section 1.2.2, paragraphe *Le redshift*.

3. Constante reliant proportionnellement la vitesse d'éloignement des galaxies à leur distance, voir section 1.2.2, paragraphe *Les équations de Friedmann-Lemaître*.

complexes qui rendent compte de l'accélération de l'expansion de l'univers. Le terme *énergie noire* est un terme générique employé pour désigner la cause de cette accélération, constante cosmologique  $\Lambda$  incluse.

Ces quinze années (1915 - 1929) très fertiles pour la cosmologie ont popularisé l'idée d'un univers en expansion. Si certains s'y opposent et défendent un univers statique, d'autres s'y intéressent et étudient en détail les conséquences de ces modèles théoriques. Si l'univers est en expansion, c'est qu'il a été dans le passé plus petit qu'il ne l'est aujourd'hui. L'étude des solutions aux équations d'Einstein montre que l'expansion dilue la matière dans l'univers, et conduit à son refroidissement. L'univers était donc plus chaud et plus dense dans le passé. Si l'on remonte suffisamment dans l'histoire de l'univers, celui-ci devient de plus en plus dense, jusqu'à être à l'origine infiniment chaud et dense. Ceci conduit à nommer ces classes de modèles *hot big bang models*, ou modèles de big bang chaud en français. Il est à noter que cet *instant zéro* est une extrapolation des modèles et reste hypothétique : au-delà d'une certaine température et densité, les effets quantiques ne peuvent plus être négligés, rendant alors impossible l'utilisation de la relativité générale. Cet instant est appelé mur de Planck. Afin de comprendre ce qu'il se passe entre le mur de Planck et l'instant zéro, une théorie traitant à la fois la gravitation et l'aspect quantique de la matière est nécessaire. C'est un domaine de recherche très dynamique aujourd'hui, dans lequel un grand nombre de théories de gravité quantique sont étudiées.

Suite notamment aux publications de Friedmann, Lemaître et Hubble, les défenseurs des modèles de big bang ont commencé à chercher des observables capables de prouver ces modèles. En 1948, George Gamow, Ralph Alpher et Robert Herman, reprenant les travaux de Georges Lemaître, prédisent l'existence du *fond diffus cosmologique* (CMB : Cosmic Microwave Background). Ce rayonnement fossile, si les modèles de big bang sont vérifiés, aurait été émis lorsque l'univers était encore dense et chaud. Il repose sur le raisonnement suivant : du fait de la température initialement très élevée, les particules possèdent trop d'énergie pour s'assembler et former les premières briques élémentaires. L'univers n'est alors qu'une soupe où toutes les particules s'entrechoquent constamment. Lorsque l'univers s'expand, la température baisse et l'énergie des particules aussi, autorisant ainsi la formation des premiers noyaux d'atomes. Mais la température et la densité sont toujours trop importantes pour laisser les premiers atomes se former : l'univers est alors un bain de noyaux, principalement d'hydrogène et d'hélium, d'électrons et de photons. Les photons sont diffusés constamment sur les électrons libres, rendant le plasma de l'univers primordial opaque. Puis, lorsque l'univers devient suffisamment froid, les électrons ne disposant plus de suffisamment d'énergie sont capturés par les noyaux, formant les premiers atomes de l'univers. Cet instant est appelé la recombinaison. Les atomes ainsi formés, neutres, ne diffusent pas les photons. Ces derniers peuvent alors se propager librement, et l'univers devient transparent. Ce sont ces photons, diffusés une dernière fois environ 380 000 ans après le big bang, qui forment le fond diffus cosmologique et que nous pouvons mesurer aujourd'hui. Les principales étapes sont résumées sur la figure 1.2, dont notamment la formation des premiers noyaux durant les premières secondes, puis le CMB vers 380 000 ans. Après l'apparition des premières étoiles, le rayonnement émis par celles-ci commence progressivement à réioniser l'hydrogène, qui était devenu neutre au moment de l'émission du CMB. Les dernières analyses estiment le redshift de la réionisation à  $5 \lesssim z_{\text{réionisation}} \lesssim 9$  (PLANCK COLLABORATION et al. 2018). Après la phase de réionisation, la quasi-

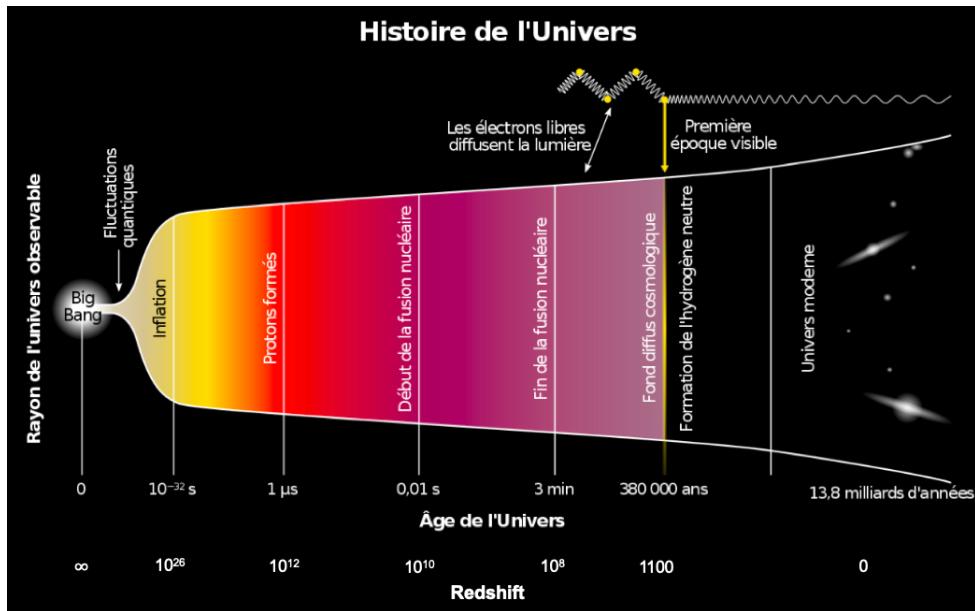


FIGURE 1.2 – Illustration de l'histoire de l'univers depuis ses origines jusqu'à aujourd'hui. Les principales étapes sont représentées : la formation des premiers protons et neutrons, puis des premiers noyaux, et enfin l'émission du CMB.

totalité des atomes d'hydrogène du milieu intergalactique sont ionisés. Seul 1 atome sur 10 000 à 1 atome sur 1000 demeure neutre.

En 1965, 17 ans après sa prédiction, le CMB est détecté par Penzias et Wilson, établissant ainsi le consensus sur les modèles de big bang. A partir de ce moment là, un certain nombre d'observations ont été menées par les cosmologistes afin de contraindre et distinguer les différents modèles de big bang. Nous retracons dans la section suivante comment le modèle  $\Lambda$ CDM s'est établi en tant que modèle standard de la cosmologie, puis nous le décrivons en détail.

## 1.2 Le modèle $\Lambda$ CDM

Le modèle  $\Lambda$ CDM est aujourd'hui le modèle cosmologique qui fait consensus dans la communauté scientifique. Il est souvent désigné comme le modèle standard de la cosmologie. C'est un modèle de big bang, décrivant un univers composé principalement d'énergie noire, aussi appelée constante cosmologique ( $\Lambda$ ), et de matière noire froide (CDM : Cold Dark Matter). La figure 1.3 présente la répartition de ses différentes composantes.

### 1.2.1 Mise en place du modèle

Le modèle s'est établi suite à un certain nombre d'observations. D'abord, la détection du CMB en 1965, qui confirme les modèles de big bang. Puis l'introduction de la matière noire dans les modèles au cours des années 70 et 80, notamment grâce aux travaux de Vera Rubin sur le problème de la masse manquante dans les galaxies. Déjà en 1933, Fritz Zwicky remarquait que la masse visible dans les amas n'était pas suffisante pour expliquer leur cohésion, et supposa donc l'existence d'une matière invisible.

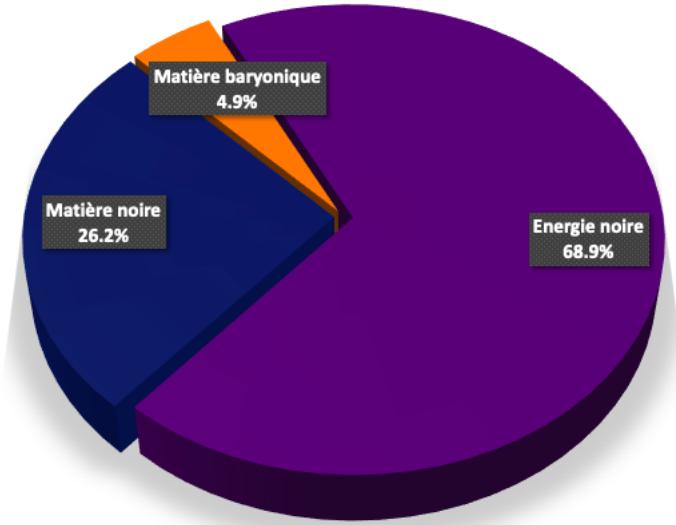


FIGURE 1.3 – Répartition des différentes composantes du modèle  $\Lambda$ CDM. L'énergie noire est montrée en violet, la matière noire en bleu et la matière baryonique en orange.

Une série d'observations fut menée dans les années 70 afin d'étudier les courbes de vitesse des étoiles au sein des galaxies. Les étoiles situées en périphérie furent mesurées avec une vitesse plus importante qu'attendue. La conclusion fut similaire à celle de Zwicky : la présence de masse invisible dans les halos de galaxies permet d'expliquer ces courbes de rotation. Ainsi la matière noire froide<sup>1</sup> fut introduite dans les modèles cosmologiques : environ 25 % de la masse de l'univers est sous la forme d'une matière non standard interagissant uniquement via la gravitation<sup>2</sup> avec la matière ordinaire. Nous présentons ici le modèle standard de la physique des particules. Il décrit les particules et leurs interactions, qui constituent la matière que nous appelons matière ordinaire. Le modèle standard divise les particules en deux catégories : les bosons et les fermions. Les bosons sont les particules vectrices des interactions. Le photon par exemple, est le boson vecteur de l'interaction électromagnétique. Les fermions quant à eux, constituent la matière. Ils sont classés en deux catégories : les leptons, et les quarks. Les leptons comprennent entre autre les électrons et les neutrinos. Les quarks ne peuvent exister isolément, ils sont regroupés par paires pour former des mésons, ou par trios pour former des baryons. Par exemple, le neutron et le proton sont deux baryons. Les autres baryons ainsi que les mésons sont instables, et possèdent une durée de vie courte. La masse des électrons étant négligeable, les protons et les neutrons représentent l'essentiel de la masse de la matière ordinaire, qui est ainsi appelée matière baryonique.

Plus tard, le satellite COBE fut envoyé dans l'espace afin de détecter les anisotropies du CMB. Selon les prédictions des modèles de big bang, le spectre du CMB suit une loi de corps noir, avec une température d'environ 3 K, et possède des anisotropies correspondant aux perturbations primordiales de densité. La mission fut un succès : les mesures de COBE ont permis d'identifier les anisotropies de température du CMB, mettant en évidence les fluctuations de densité de l'univers primordial. D'autre part, le spectre du CMB est mesuré avec une température  $T = (2,728 \pm 0,004)$  K (Fixsen et al. 1996),

1. Par opposition à la matière noire chaude, la matière noire froide est non relativiste.

2. Cependant certaines expériences recherchent des particules candidates à la matière noire, qui interagissent très faiblement via l'interaction faible, comme par exemple les *WIMPs* (Weakly Interactive Massive Particles).

ne déviant pas du spectre du corps noir de plus de 0,25 % (BENNETT et al. 1993). La détection des anisotropies du CMB constitue un des arguments les plus solides en faveur des modèles de big bang.

Jusque alors, les modèles cosmologiques n'incluaient pas d'énergie noire. Puis en 1998, deux équipes différentes publient l'analyse de distances de luminosité de supernovae de type 1a (SN1a), toutes les deux mettant en évidence l'accélération de l'expansion de l'univers et donc favorisant les modèles contenant de l'énergie noire. Ce sont ces dernières observations qui ancrent  $\Lambda$ CDM comme modèle de big bang préféré. Par la suite, le satellites WMAP puis le satellite Planck sont lancés en 2001 et en 2009 afin de mesurer avec une plus grande précision les anisotropies du CMB. Ces mesures successives sont effectuées avec une précision sans précédent, permettant de contraindre très fortement les paramètres cosmologiques. Les résultats finaux de la collaboration Planck ont été publiés en 2018 (PLANCK COLLABORATION et al. 2018) et fournissent les paramètres cosmologiques du modèle  $\Lambda$ CDM avec une précision inférieur au pourcent (tableau 1.1).

### 1.2.2 Description du modèle

Le modèle  $\Lambda$ CDM, et plus généralement les modèles de big bang, sont fondés sur le formalisme de la relativité générale. Cette théorie, élaborée par Einstein en 1915, est la généralisation de la relativité restreinte, proposée par Einstein 10 ans plus tôt. La relativité restreinte émet deux postulats :

- les lois de la physique sont les mêmes dans tous les référentiels inertiels<sup>1</sup>,
- la vitesse de la lumière dans le vide est la même dans tous les référentiels inertiels.

Cette théorie traite les mouvements des corps dans les référentiels inertiels, mais n'inclut pas la gravitation. Afin d'inclure la gravitation, le principe d'équivalence est supposé. Ce principe affirme que la masse inertielle et la masse grave sont équivalentes, et que les effets de la gravitation sont identiques aux effets de l'accélération du référentiel de l'observateur. Autrement dit, il n'existe pas d'expérience permettant à l'observateur de distinguer s'il se trouve dans un champ de gravitation uniforme ou dans un référentiel uniformément accéléré. La gravitation n'est alors plus vue comme une force, mais comme un effet géométrique, conséquence de la déformation de l'espace-temps.

Dans la suite de cette section, afin de simplifier les équations, nous nous plaçons dans un système d'unité dans lequel

$$c = \hbar = k_B = 1 . \quad (1.1)$$

#### La métrique

Le formalisme de la relativité générale s'appuie donc sur celui de la relativité restreinte. La géométrie de l'espace-temps est décrite par la métrique. Cet objet mathématique<sup>2</sup> permet de définir le produit scalaire sur l'espace-temps à 4 dimensions, et donc de mesurer les distances et les angles. Nous verrons plus loin dans ce manuscrit que la métrique dépend de la distribution de masse. Ainsi, et c'est le fondement de la relativité générale, la masse courbe l'espace temps et l'espace-temps indique à la masse, via la métrique, comment se déplacer au sein de celui-ci<sup>3</sup>.

---

1. Un référentiel dans lequel l'observateur n'est pas accéléré.

2. Un tenseur de rang 2.

3. Inspiré de la citation de John Wheeler : « Spacetime tells matter how to move; matter tells spacetime how to curve ».

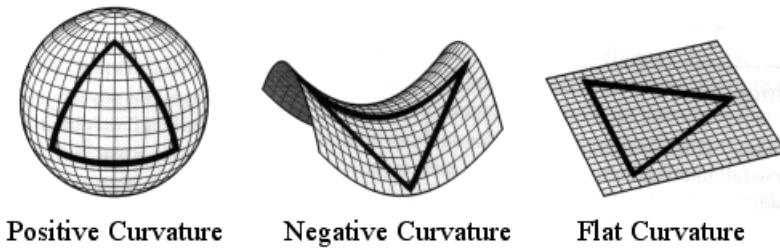


FIGURE 1.4 – Représentation de la courbure de l'univers : positive à gauche, négative au centre et nulle à droite.

Dans le cadre du modèle  $\Lambda$ CDM, la métrique utilisée est la métrique FLRW (pour Friedmann Lemaître Robertson Walker), elle s'exprime comme :

$$ds^2 = -dt^2 + R(t) \left[ \frac{dr^2}{1 - kr^2} + r^2 d\Omega^2 \right], \quad (1.2)$$

où  $d\Omega = d\theta \sin(\theta) d\phi$ ,  $R(t)$  rend compte de l'expansion de l'univers à l'instant  $t$ , et  $k$  vaut soit 1, 0 ou  $-1$  selon que l'univers possède une courbure positive, nulle ou négative (figure 1.4). A l'aide d'un changement de coordonnées, il est possible de se ramener à la formule suivante :

$$ds^2 = -dt^2 + a(t) [d\chi^2 + S_k^2(\chi) d\Omega^2], \quad (1.3)$$

où  $a(t) = R(t)/R(t_0)$ ,  $t_0$  est le temps présent et  $S_k$  est défini comme

$$S_k(\chi) = R(t_0) \begin{cases} \sin(\chi/R(t_0)) & \text{si } k = 1 \\ \chi/R(t_0) & \text{si } k = 0 \\ \sinh(\chi/R(t_0)) & \text{si } k = -1 \end{cases}. \quad (1.4)$$

Cette formulation permet de mettre en évidence le rapport  $a(t)$ , appelé facteur d'échelle. Par définition il vaut 1 aujourd'hui. Afin de rendre compte de l'expansion,  $a(t) < 1$  pour  $t < t_0$  (passé) et  $a(t) > 1$  pour  $t > t_0$  (futur).

## Le redshift

Le décalage vers le rouge, ou *redshift* en anglais, est la mesure du décalage du spectre des objets vers les grandes longueurs d'onde. Le redshift  $z$  est défini comme

$$1 + z = \frac{\lambda_o}{\lambda_e}, \quad (1.5)$$

où  $\lambda_e$  est la longueur d'onde émise et  $\lambda_o$  la longueur d'onde observée. Le redshift trouve son origine dans deux effets. Le premier : le redshift causé par la vitesse particulière de l'objet. Dans ce cas, nous parlons d'*effet Doppler*<sup>1</sup>. Le second : dans le cadre des modèles de big bang, l'expansion de l'univers

1. L'effet Doppler est l'augmentation ou la diminution de la longueur d'onde d'une onde lorsque l'émetteur de cette dernière s'approche ou s'éloigne de l'observateur. L'exemple le plus connu est celui de l'ambulance : le son entendu est

produit un décalage vers le rouge du spectre des objets distants. Dans ce cas, nous parlons de *redshift cosmologique*. Pour les objets lointains que nous considérons dans ce manuscrit, le redshift induit par les vitesses particulières est faible devant le redshift causé par l'expansion de l'univers. Nous confondons donc dans la suite redshift et redshift cosmologique. Mais ceci est une approximation, qui produit des distorsions (voir section 1.5.2) dans l'objet d'étude de ce manuscrit : les fonctions de corrélation (voir section 1.3.3).

Les objets distants s'éloignent de nous du fait de l'expansion. Similairement à l'effet Doppler, le spectre observé de ces objets est décalé vers les grandes longueurs d'onde. Mais contrairement à l'effet Doppler, le redshift n'est pas directement dû à la vitesse de récession de l'objet : à cause de l'expansion, les photons, lors de leur propagation, voient leur longueur d'onde augmenter. On peut montrer que

$$\frac{\lambda_o}{\lambda_e} = \frac{a(t_e)}{a(t_o)}, \quad (1.6)$$

où  $t_e$  et  $t_o$  sont les temps d'émission et d'observation du photon,  $\lambda_e$  et  $\lambda_o$  sa longueur d'onde lors de l'émission et de l'observation. Le redshift est donc relié au facteur d'échelle via la relation

$$1+z = \frac{1}{a(t)}. \quad (1.7)$$

Le redshift (cosmologique) est donc directement dû à l'expansion de l'univers. Il peut servir de mesure de temps (et aussi de distance, voir section 1.2.2 paragraphe *Les distances*) : le spectre d'un objet avec un redshift  $z=2$  est décalé vers le rouge d'un facteur 3. Il en découle que sa lumière observée aujourd'hui a été émise lorsque l'univers avait une taille 3 fois plus petite qu'aujourd'hui, soit il y a environ 12 milliards d'années.

## Les équations d'Einstein

Lorsque Einstein publie sa théorie en 1915, la façon d'introduire les équations d'Einstein, le cœur de la théorie, est différente de la façon de les introduire aujourd'hui. Nous nous proposons ici de suivre l'approche de la physique moderne, qui formule toutes les théories en termes d'un seul et même principe : le *principe de moindre action*. Ce principe stipule que l'action mis en œuvre lors de l'évolution d'un système entre deux instants est toujours extrémale<sup>1</sup>. L'action est une quantité caractérisant globalement un système, elle est définie comme

$$\mathcal{S} = \int_{t_0}^{t_1} L dt, \quad (1.8)$$

où  $L$  est le lagrangien du système. En mécanique newtonienne, ce dernier est défini comme la différence de l'énergie cinétique et de l'énergie potentiel. En relativité générale, tout comme dans les théories de champs<sup>2</sup>, le terme du lagrangien est représenté plutôt par une densité de lagrangien. Cette densité de lagrangien est alors intégrée sur l'espace-temps afin d'obtenir l'action. Dans le cas de la relativité

---

plus aigu lorsque l'ambulance s'approche, puis plus grave lorsqu'elle s'éloigne. Le même processus est à l'œuvre avec la lumière.

1. Elle est minimale dans la grande majorité des cas.

2. En particulier la théorie quantique des champs.

générale, l'action est définie comme

$$\mathcal{S} = \int d^4x \sqrt{-g} \frac{R}{4\pi G}, \quad (1.9)$$

où  $g$  est le déterminant de la métrique,  $R$  le scalaire de Ricci, et  $G$  la constante de Newton. Le scalaire de Ricci caractérise la courbure, il dépend des dérivées secondes de la métrique. Une fois l'action déterminé, sa minimisation conduit aux équations du mouvement du système. Dans notre cas, ce sont les équations d'Einstein :

$$R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu} + \Lambda g_{\mu\nu} = 8\pi GT_{\mu\nu}, \quad (1.10)$$

où  $g_{\mu\nu}$  est la métrique,  $R_{\mu\nu}$  le tenseur de Ricci,  $T_{\mu\nu}$  le tenseur énergie-impulsion, et  $\Lambda$  la constante cosmologique. Le tenseur de Ricci, dont la contraction donne le scalaire de Ricci  $R$ , dépend des dérivées secondes de la métrique. C'est donc un terme purement géométrique. Le tenseur énergie-impulsion quant à lui contient l'information de la distribution de masse et d'énergie. Ainsi il y a un lien direct entre la métrique, qui décrit la déformation de l'espace-temps, et la masse présente dans l'univers.

L'équation 1.10 regroupe en réalité plusieurs équations. Les indices  $\mu$  et  $\nu$  varient de 0 à 3, 0 représentant la coordonnée temporelle et 1 à 3 les coordonnées spatiales. Il existe donc une équation par couple  $(\mu, \nu)$ , produisant 16 équations. Par des arguments de symétrie, ce nombre se réduit à 6 équations indépendantes, que l'on nomme les équations d'Einstein.

### Les équations de Friedmann-Lemaître

Les équations d'Einstein forment un système d'équations différentielles, de second ordre et non linéaires, et de fait, difficile à résoudre. Afin de simplifier les équations et trouver des solutions, certaines hypothèses sont faites. Dans la plupart des modèles cosmologiques, l'univers est supposé homogène et isotrope à grande échelle. La métrique qui décrit un univers homogène et isotrope est la métrique FLRW (équation 1.3). Dans un tel cas, on peut calculer le membre de gauche de l'équation 1.10. Ce calcul, que nous ne détaillerons pas ici, est très bien détaillé dans la section 2.1.2 de DODELSON (2003). De plus, pour un fluide parfait, le tenseur énergie impulsion prend la forme

$$T_{\mu\nu} = \begin{pmatrix} -\rho & 0 & 0 & 0 \\ 0 & \mathcal{P} & 0 & 0 \\ 0 & 0 & \mathcal{P} & 0 \\ 0 & 0 & 0 & \mathcal{P} \end{pmatrix}, \quad (1.11)$$

où  $\rho$  est la densité du fluide, et  $\mathcal{P}$  est sa pression. Dans ces conditions, la partie temporelle de l'équation 1.10 donne

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\rho + \frac{\Lambda}{3} - \frac{k}{a^2} \quad (1.12)$$

et la partie spatiale

$$2\frac{\ddot{a}}{a} + \left(\frac{\dot{a}}{a}\right)^2 = -8\pi G\mathcal{P} + \Lambda - \frac{k}{a^2}, \quad (1.13)$$

où le point désigne la dérivé temporelle. On définit alors le taux d'expansion  $H$  comme  $H(t) = \frac{\dot{a}(t)}{a(t)}$ . Sa valeur actuelle, notée  $H_0$ , est appelée constante de Hubble. Elle relie proportionnellement la distance des galaxies à leur vitesse d'éloignement, via la loi de Hubble :

$$V = H_0 \times D . \quad (1.14)$$

$H_0$  est souvent donné comme  $H_0 = 100 h \text{ km s}^{-1} \text{ Mpc}^{-1}$ , où  $h$  est un paramètre sans dimension qui prend en compte l'incertitude sur  $H_0$ . D'après les mesures les plus récentes (PLANCK COLLABORATION et al. 2018; RIESS et al. 2019),  $h$  varie entre 0,67 et 0,75. L'équation 1.14 est nommée en l'honneur d'Edwin Hubble, après sa publication en 1929, même si Georges Lemaître fut sans doute le premier à interpréter le lien entre distance et vitesse d'éloignement des galaxies par l'expansion de l'univers. Suite à cette brève parenthèse, retournons à nos deux équations. Il est courant de récrire ces équations en injectant  $H(t)$ , ainsi qu'en remplaçant la seconde par une combinaison linéaire des deux précédentes :

$$H^2 = \frac{8\pi G}{3}\rho + \frac{\Lambda}{3} - \frac{k}{a^2}, \quad (1.15)$$

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}(\rho + 3\mathcal{P}) + \frac{\Lambda}{3}. \quad (1.16)$$

Ces deux équations sont appelées les équations de Friedmann-Lemaître. Elles découlent directement des équations d'Einstein pour un univers homogène et isotrope, et permettent de calculer l'évolution du facteur d'échelle en fonction de la densité des différentes composantes de l'univers. Nous pouvons noter que le membre de droite de l'équation 1.15 contient 3 entités : le fluide parfait ainsi que la courbure et la constante cosmologique. Même si cela reste un choix d'écriture et ne relève d'aucun argument mathématique, il permet de mettre en évidence le fait que ces deux dernières entités peuvent être considérées comme des composante énergétique de l'univers, avec leur propre densité d'énergie.

## Évolution de l'univers

Avant de résoudre les équations de Friedmann-Lemaître, nous devons relier  $\rho$  et  $\mathcal{P}$  au facteur d'échelle. Considérons un univers constitué de plusieurs fluides parfaits. Chaque fluide  $i$  est décrit par sa densité  $\rho_i$  et sa pression  $\mathcal{P}_i$ . Nous pouvons relier l'une à l'autre grâce à l'équation d'état du fluide :

$$\mathcal{P}_i = w_i \rho_i , \quad (1.17)$$

où  $w_i$  est le paramètre d'état du fluide  $i$ , ici supposé constant. De plus, pour chaque fluide, la conservation du tenseur énergie-impulsion,  $\partial_\mu T_i^{\mu\nu} = 0$ , nous donne

$$\dot{\rho}_i + 3H(\rho_i + \mathcal{P}_i) = 0 . \quad (1.18)$$

En intégrant cette équation et en utilisant l'équation d'état 1.17, nous obtenons donc l'évolution de  $\rho_i$  avec le facteur d'échelle

$$\rho_i = \rho_{i,0} a^{-3(1+w)} , \quad (1.19)$$

où  $\rho_{i,0}$  est la densité du fluide  $i$  aujourd'hui.

Selon le fluide, la valeur de  $w$  est différente. Nous pouvons déjà distinguer les particules relativistes<sup>1</sup> des particules non relativistes. La matière non relativiste ( $m$ ) ou simplement matière, se compose de la matière baryonique ( $b$ ) et de la matière noire froide ( $c$ ). La matière baryonique peut être vue comme un gaz de galaxies n'interagissant entre elles que via la gravitation. De la même manière, les constituants de la matière noire froide n'interagissent les uns avec les autres que via la gravitation. Ainsi, le fluide décrivant la matière possède une pression nulle, son paramètre d'état est donc  $w_m = 0$ . Nous avons alors :  $\rho_m \propto a^{-3}$ .

Concernant les particules relativistes, elles constituent ce qu'on appelle la radiation ( $r$ ). Celle-ci est composée des photons ( $\gamma$ ) et des neutrinos relativistes ( $\nu$ ). Son paramètre d'état est  $w_r = 1/3$ , ce qui donne  $\rho_r \propto a^{-4}$ . Nous pouvons remarquer que la densité de matière diminue proportionnellement au volume de l'univers, par simple effet de dilution. La densité de radiation possède un facteur  $1/a$  supplémentaire. Ce facteur provient du redshift des photons observés, et s'ajoute au  $1/a^3$  de la dilution.

Afin de travailler avec des quantités sans dimension et normalisées, il est courant d'introduire la densité critique  $\rho_{\text{crit}}$ . Cette densité est la densité limite pour laquelle l'univers est plat. Au delà de cette densité limite, l'univers est fermé, en deçà, l'univers est ouvert. Pour  $k = 0$ , l'équation 1.15 s'écrit

$$H^2 = \frac{8\pi G}{3}\rho, \quad (1.20)$$

où  $\rho$  désigne désormais la densité totale d'énergie de l'univers, incluant la contribution  $\rho_\Lambda = \Lambda/8\pi G$  de la constante cosmologique, d'où  $\rho_{\text{crit}} = 3H_0^2/8\pi G$ . En divisant cette équation par  $H_0^2$ , l'équation précédente s'écrit alors

$$\frac{H^2}{H_0^2} = \frac{\rho}{\rho_{\text{crit}}}. \quad (1.21)$$

A l'aide de l'équation 1.19, chaque composante peut être mise sous la forme

$$\frac{\rho_i}{\rho_{\text{crit}}} = \Omega_i a^{-3(1+w)}, \quad (1.22)$$

où  $\Omega_i$  est le ratio de la densité de l'espèce  $i$  par la densité critique aujourd'hui. Nous pouvons alors récrire l'équation de Friedmann-Lemaître 1.15 comme

$$\frac{H^2}{H_0^2} = \sum_i \Omega_i a^{-3(1+w)} - \frac{k}{a^2 H_0^2}, \quad (1.23)$$

où  $i$  court sur toutes les espèces contribuant à l'énergie totale de l'univers : la matière, la radiation et la constante cosmologique. Ces trois densités relatives valent

$$\Omega_r = \frac{8\pi G \rho_{r,0}}{3H_0^2} \quad ; \quad \Omega_m = \frac{8\pi G \rho_{m,0}}{3H_0^2} \quad ; \quad \Omega_\Lambda = \frac{\Lambda}{3H_0^2}. \quad (1.24)$$

Par ailleurs, nous remarquons que la densité d'énergie  $\rho_\Lambda = \Lambda/8\pi G$  est indépendante de  $a$ . Il en découle  $w_\Lambda = -1$  (équation 1.19) : la constante cosmologique peut être interprétée comme un fluide de densité d'énergie constante et de pression négative. Nous verrons dans la suite de ce paragraphe que sa

---

1. Une particule est dite relativiste lorsque sa vitesse est proche de celle de la lumière dans le vide.

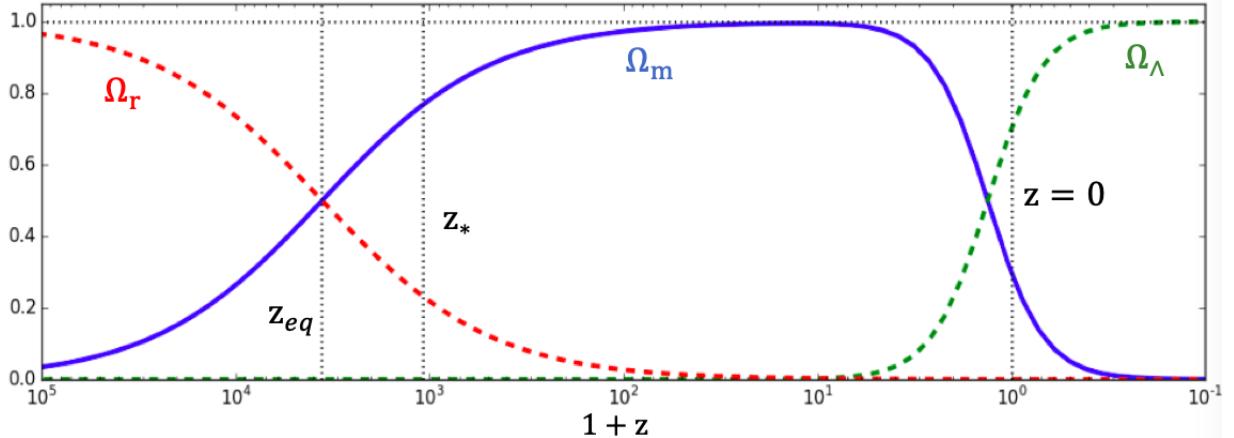


FIGURE 1.5 – Évolution en fonction du redshift des densités relatives d'énergie pour un univers plat. Sont indiqués en pointillés le redshift d'égalité radiation-matière,  $z_{eq} = 3387 \pm 21$ , le redshift d'émission du CMB,  $z_* = 1089,80 \pm 0,21$ , et le redshift aujourd'hui,  $z = 0$  (PLANCK COLLABORATION et al. 2018). Crédits : ZARROUK (2018).

domination dans le bilan énergétique de l'univers actuel est responsable de l'accélération de l'expansion. Enfin, en utilisant les définitions précédentes, nous obtenons l'évolution du taux d'expansion

$$H^2 = H_0^2 \left[ \Omega_m a^{-3} + \Omega_r a^{-4} + \Omega_\Lambda \right] - \frac{k}{a^2}. \quad (1.25)$$

En évaluant l'équation précédente pour  $t = 0$ , nous obtenons

$$1 + \frac{k}{a^2 H_0} = \Omega_m + \Omega_r + \Omega_\Lambda = \Omega_{\text{total}}. \quad (1.26)$$

Ainsi pour un univers plat, nous avons  $k = 0$ , et donc  $\Omega_{\text{total}} = 1$ . Nous retrouvons alors que  $\rho_{\text{crit}}$  correspond à la densité totale de l'univers.

Bien que les  $\Omega_i$  donnent les densités relatives aujourd'hui, nous pouvons comparer les différentes densités d'énergie à n'importe quel redshift, en considérant le rapport  $\Omega_i(z) = \rho_i(z)/\rho_{\text{crit}}(z)$ , où  $\rho_{\text{crit}}(z) = 3H(z)^2/8\pi G$ . La figure 1.5 présente l'évolution avec le redshift des densités relatives, et indique donc les différentes ères de domination. Pour  $z > z_{eq}$ , c'est la radiation qui domine. Puis vient l'époque de domination de la matière, jusqu'à  $z \sim 0,3$ . Enfin, c'est la constante cosmologique  $\Lambda$  qui domine aujourd'hui.

Lorsque l'univers est dominé par un seul fluide  $i$ , nous avons  $\rho_{\text{tot}} \sim \rho_i$ , et nous pouvons donc injecter l'équation 1.19 dans l'équation de Friedmann-Lemaître 1.15. Nous obtenons alors l'évolution de taux d'expansion avec le facteur d'échelle

$$H = H_0 a^{-3(1+w)/2}, \quad (1.27)$$

ce qui nous donne finalement, pour  $w \neq -1$ , l'évolution temporelle du facteur d'échelle

$$a(t) \propto t^{\frac{2}{3(1+w)}}. \quad (1.28)$$

Ainsi, dans chaque phase de domination,  $a$  évolue comme une loi de puissance :  $a \propto t^{\frac{1}{2}}$  durant l'aire de domination de la radiation, puis  $a \propto t^{\frac{2}{3}}$  lors de la domination de la matière. Cependant, l'univers est actuellement dominé par l'énergie noire, l'équation différentielle 1.27 devient  $H = H_0$ , et nous obtenons  $a \propto e^{H_0 t}$ . Nous pouvons noter que durant toutes les phases de domination,  $\dot{a} > 0$  : l'univers est en expansion. D'autre part, lors des phases de domination de la radiation et de la matière,  $\ddot{a} < 0$  : l'expansion de l'univers décélère. Cependant, pour  $z \lesssim 0,3$ , nous avons  $\ddot{a} > 0$ , l'univers est alors en expansion accélérée.

## Les distances

La notion de distance en relativité générale n'est pas très intuitive. Du fait de l'expansion, la distance qui sépare deux astres lointains au moment où ils ont émis leur lumière n'est pas la distance qui les sépare aujourd'hui. Depuis la terre, nous n'avons pas directement accès aux distances, mais seulement aux différences d'angles et de redshifts. Lorsque nous observons deux astres, nous observons une différence d'angles qui correspond à la distance qui les séparent lorsque la lumière que nous captions aujourd'hui a été émise. Entre ce moment et aujourd'hui, l'expansion a éloigné ces deux astres et leur distance de séparation a été multipliée par un facteur  $\frac{1}{a(t_e)} = 1 + z_e$ , où  $t_e$  est le temps correspondant à l'émission. Afin de simplifier les comparaisons de distances à différentes époques, nous définissons la distance *comobile* comme ceci : deux objets à un redshift  $z$  et séparés d'une distance physique  $D$  possèdent une distance comobile  $(1+z)D$ . C'est la distance physique, en incluant l'expansion jusqu'à aujourd'hui. Ainsi, la distance comobile séparant deux objets soumis à l'expansion reste la même au cours du temps.

Nous présentons ici les différentes distances utilisées en cosmologie. Elles sont très bien décrites dans Hogg (1999), dont nous suivons d'ailleurs les notations. Aussi, nous sortons désormais du cadre dans lequel  $c = \hbar = k_B = 1$  et nous reprenons le système d'unité usuel. Définissons premièrement la quantité

$$E(z) = \frac{H(z)}{H_0} = \sqrt{\Omega_m a^{-3} + \Omega_r a^{-4} + \Omega_\Lambda + 1 - \Omega_{\text{total}}}, \quad (1.29)$$

ainsi que la distance de Hubble aujourd'hui

$$D_H = \frac{c}{H_0}. \quad (1.30)$$

Nous pouvons alors définir les distances suivantes :

- la distance comobile le long de la ligne de visée  $D_C$ . C'est la distance comobile qui sépare un objet lointain de l'observateur. Elle est obtenue en intégrant chaque contribution infinitésimale de  $z = 0$  jusqu'à l'objet :

$$D_C = D_H \int_0^z \frac{dz'}{E(z')} . \quad (1.31)$$

- la distance comobile transverse  $D_M$  : deux objets à un redshift  $z$  et séparés par un angle  $\delta\theta$  sur le ciel sont à une distance comobile  $\delta\theta D_M$  l'un de l'autre. Dans le cas où l'univers n'est pas plat ( $\Omega_k \neq 0$ ), la distance comobile transverse  $D_M$  n'est pas la même que la distance comobile le long

de la ligne de visée  $D_C$ . Elle est reliée à  $D_C$  par

$$D_M = \begin{cases} D_H \frac{1}{\sqrt{\Omega_k}} \sin(\Omega_k D_C / D_H) & \text{si } \Omega_k < 0 \\ D_C & \text{si } \Omega_k = 0 \\ D_H \frac{1}{\sqrt{\Omega_k}} \sinh(\Omega_k D_C / D_H) & \text{si } \Omega_k > 0 \end{cases}. \quad (1.32)$$

- La distance de diamètre angulaire  $D_A$  : c'est la distance reliée à la taille apparente d'un objet. Deux objets à un redshift  $z$  et séparés par un angle  $\delta\theta$  sur le ciel possèdent une distance physique  $\delta\theta D_A$ . La distance de diamètre angulaire diffère de  $D_M$  du fait qu'elle considère la distance physique et non comobile entre les deux objets. Elle est donc reliée à  $D_M$  par

$$D_A = \frac{D_M}{1+z}. \quad (1.33)$$

- la distance de luminosité  $D_L$  : elle est définie via la relation qui exprime le flux d'une source lumineuse en fonction de sa luminosité

$$F = \frac{L}{4\pi D_L^2} \quad \rightarrow \quad D_L = \sqrt{\frac{L}{4\pi F}}. \quad (1.34)$$

Elle est reliée à la distance comobile transverse via

$$D_L = (1+z)D_M = (1+z)^2 D_A, \quad (1.35)$$

Les distances sont usuellement mesurées en  $h^{-1}$  kpc ou  $h^{-1}$  Mpc, le facteur  $h$  permettant de rendre ces distances indépendantes de l'incertitude sur la mesure de  $H_0$ . Un parsec vaut environ 3,2616 années lumières, soit environ  $3,0857 \times 10^{16}$  m. Dans ce manuscrit, les distances qui nous intéressent particulièrement sont  $D_C$  et  $D_M$ . Nous y ferons appel dans la section 1.4.2.

### Les paramètres du modèle

Le modèle  $\Lambda$ CDM est un modèle qui comporte 6 paramètres. Ils sont mesurés par le satellite Planck (PLANCK COLLABORATION et al. 2018) avec une précision d'environ 1 % et sont résumés dans le tableau 1.1. Les 6 paramètres mesurés par Planck sont

- $\Omega_b h^2$ , la densité de baryons multipliée par  $h^2$
- $\Omega_c h^2$ , la densité de matière noire multipliée par  $h^2$
- $\theta_{MC}$ , une approximation de  $\theta_*$  : l'angle sur le ciel de l'échelle acoustique
- $\tau$ , la profondeur optique totale, intégrée de  $z = 0$  jusqu'au CMB. La contribution provient essentiellement des électrons libres entre  $z = 0$  et  $z_{\text{réionisation}}$
- $A_s$ , l'amplitude du *spectre de puissance des fluctuations primordiales*, décrit dans la section suivante
- $n_s$ , l'indice spectral du spectre de puissance des fluctuations primordiales

TABLE 1.1 – Paramètres cosmologiques mesurés par le satellite Planck. La partie supérieure du tableau indique les six paramètres ajustés aux données. La partie inférieure donne d’autres paramètres déduits de ces six paramètres ajustés. Ces chiffres sont tirés de la table 1.1 de PLANCK COLLABORATION et al. (2018)

Parameters	Combined
$\Omega_b h^2 \dots$	$0.02233 \pm 0.00015$
$\Omega_c h^2 \dots$	$0.1198 \pm 0.0012$
$100\theta_{\text{MC}} \dots$	$1.04089 \pm 0.00031$
$\tau \dots$	$0.0540 \pm 0.0074$
$\ln(10^{10} A_s) \dots$	$3.043 \pm 0.014$
$n_s \dots$	$0.9652 \pm 0.0042$
<hr/>	
$\Omega_m h^2 \dots$	$0.1428 \pm 0.0011$
$H_0 [\text{km s}^{-1} \text{Mpc}^{-1}]$	$67.37 \pm 0.54$
$\Omega_m \dots$	$0.3147 \pm 0.0074$
$\text{Age [Gyr]} \dots$	$13.801 \pm 0.024$
$\sigma_8 \dots$	$0.8101 \pm 0.0061$
$S_8 \equiv \sigma_8 (\Omega_m / 0.3)^{0.5}$	$0.830 \pm 0.013$
$z_{\text{re}} \dots$	$7.64 \pm 0.74$
$100\theta_* \dots$	$1.04108 \pm 0.00031$
$r_{\text{drag}} [\text{Mpc}] \dots$	$147.18 \pm 0.29$

D’autres paramètres se déduisent de ces 6 paramètres, notamment les densités d’énergies aujourd’hui, dont nous venons de parler. Certains sont indiqués dans la seconde partie du tableau 1.1, dont notamment  $r_{\text{drag}}$ , la taille comobile de l’horizon acoustique au moment du découplage des baryons avec les photons, ou encore  $\Omega_m$  la densité relative de matière aujourd’hui. Les paramètres cosmologiques utilisés pour la confection des simulations présentées dans ce manuscrit et par le code d’analyse **picca**<sup>1</sup> sont légèrement différents de ceux présentés dans le tableau 1.1. Nous les donnons ici :

$$\Omega_m = 0,31457 \quad ; \quad \Omega_k = 0 \quad ; \quad \Omega_\Lambda = 0,68543 . \quad (1.36)$$

### 1.3 La fonction de corrélation de la matière

Dans le paragraphe précédent, nous parlions du spectre de puissance des fluctuations primordiales sans avoir auparavant défini ce dont il s’agissait. Nous donnons ici une explication de la notion de spectre de puissance, ainsi que de la fonction de corrélation, objet d’étude de ce manuscrit.

#### 1.3.1 Une analogie avec le son

Prenons l’exemple d’un phénomène simple : le son créé par un diapason. Le diapason est un outil utilisé par les musiciens pour accorder leurs instruments. Lorsqu’il est joué, le diapason produit un signal sonore très proche d’une sinusoïde. Le son produit correspond alors à une note particulière, d’une fréquence donnée. Par opposition au diapason, la corde de guitare par exemple, lorsqu’elle vibre,

1. **picca** est le code que nous utilisons pour analyser les données et nos simulations (voir chapitre 5).

### 1.3. LA FONCTION DE CORRÉLATION DE LA MATIÈRE

produit un son composé de plusieurs fréquences : la fréquence fondamentale, qui donne la hauteur de la note, et les fréquences harmoniques, des multiples de la fréquence fondamentale. Ces fréquences harmoniques participent à la richesse du son de l'instrument. L'outil mathématique permettant d'étudier ces phénomènes s'appelle la transformation de Fourier. Elle permet d'associer à un signal temporel, sa transformée de Fourier, un signal dans l'espace des fréquences.

Reprendons l'exemple du diapason. Comme dit précédemment, le signal sonore produit est très proche d'une sinusoïde. La figure 1.6 illustre la transformation de Fourier : à gauche se trouve le signal temporel, qui correspond au signal sonore, et à droite se trouve la transformée de Fourier de ce signal. La première ligne correspond au cas d'un diapason idéal : une sinusoïde dont la transformée de Fourier donne un dirac dans l'espace de Fourier, tandis que le cas de la corde de guitare ressemblerait plutôt au cas de la troisième ligne : une somme de sinusoïdes de différentes fréquences, la fréquence la plus basse étant la fréquence fondamentale. Dans ce cas, nous pouvons remarquer que le signal dans l'espace fréquentiel est relativement simple : une somme de dirac indiquant les fréquences issues de la décomposition du signal temporel en sinusoïdes.

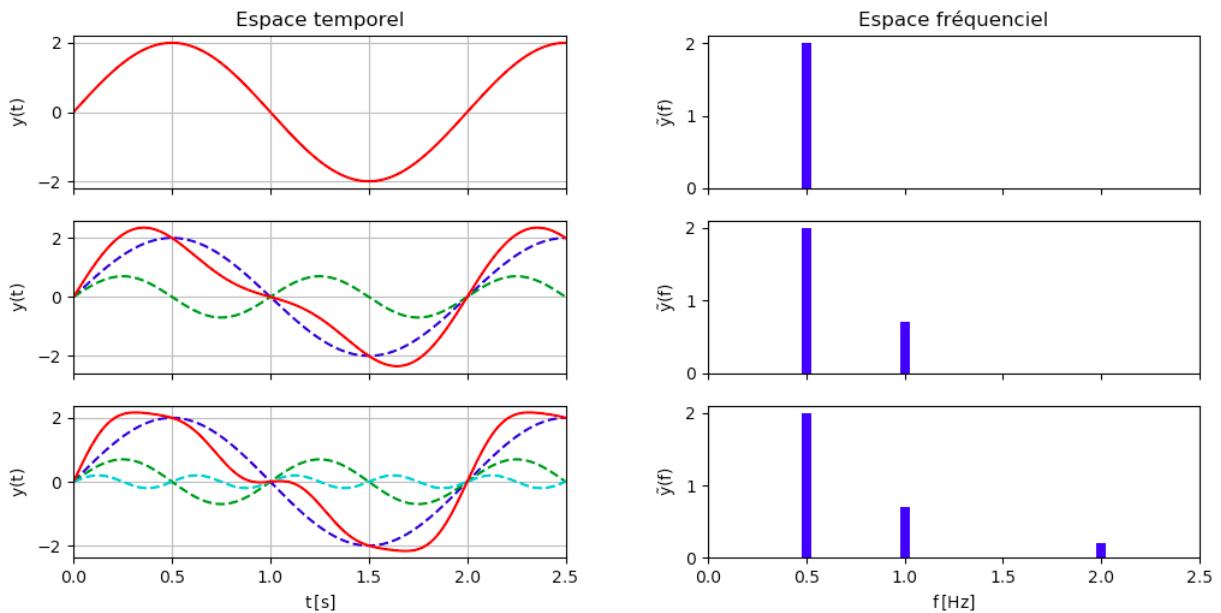


FIGURE 1.6 – Illustration de la transformation de Fourier. Le signal temporel à gauche en rouge est décomposé en somme de sinusoïdes. La transformée de Fourier correspond au signal fréquentiel, à droite, donnant la répartition des fréquences mises en jeu dans le signal temporel.

La transformation de Fourier permet donc de décomposer un signal temporel en une série de fonctions sinusoïdales, et d'indiquer la répartition entre les différentes fréquences. Pour un signal temporel  $f$ , la transformée de Fourier  $\tilde{f}$  associée à ce signal est donnée par

$$\tilde{f}(\omega) = \int_{-\infty}^{+\infty} f(t) e^{-i\omega t} dt, \quad (1.37)$$

où  $t$  est le temps en s, et  $\omega$  la pulsation en  $s^{-1}$ . Cette dernière est reliée à la fréquence par  $\omega = 2\pi f$ .

La transformation inverse est donnée par

$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \tilde{f}(\omega) e^{i\omega t} d\omega. \quad (1.38)$$

### 1.3.2 Le spectre de puissance

Le spectre de puissance est un outil mathématique utilisé afin d'étudier la répartition des modes présents dans un ensemble de données. Les modes sont la généralisation du concept de fréquence. Par exemple, dans le cas du diapason, les modes sont les différentes fréquences qui composent le signal temporel. En cosmologie, les modes sont associés à des fluctuations spatiales. Prenons l'exemple de la distribution de la matière. On peut définir le contraste de densité en  $\vec{x}$  comme

$$\delta(\vec{x}) = \frac{\rho(\vec{x}) - \bar{\rho}}{\bar{\rho}}, \quad (1.39)$$

où  $\rho(\vec{x})$  est la densité en  $\vec{x}$ , et  $\bar{\rho}$  la densité moyenne. Le spectre de puissance du contraste de densité de la matière renseigne donc sur la répartition des modes de fluctuations spatiales de la matière. La variable dans l'espace de Fourier associée au vecteur position  $\vec{x}$  est le vecteur d'onde  $\vec{k}$ . Les grands  $k$  correspondent aux modes à petites échelles, et les petits  $k$  aux modes à grandes échelles. Nous définissons le contraste de densité  $\delta(\vec{k})$  associé au vecteur d'onde  $\vec{k}$  comme étant la transformée de Fourier en trois dimensions du contraste de densité  $\delta(\vec{r})$

$$\delta(\vec{k}) = \int_{\mathbb{R}^3} d\vec{r} \delta(\vec{r}) e^{-i\vec{r}\cdot\vec{k}}. \quad (1.40)$$

Le spectre de puissance du contraste de densité de la matière est alors défini comme

$$P(\vec{k}) = \langle \delta(\vec{k})^2 \rangle, \quad (1.41)$$

où  $\langle . \rangle$  désigne la moyenne d'ensemble : c'est la moyenne prise sur plusieurs réalisations d'univers. Étant donné que l'isotropie est supposée en cosmologie, le spectre de puissance dépend uniquement de  $k$ , la norme de  $\vec{k}$ . La figure 1.9 montre le spectre de puissance de la matière à  $z = 0$ , elle est discuté dans la section suivante.

Similairement au spectre de puissance du contraste de densité, il est possible de calculer celui des fluctuations du CMB. Le formalisme est cependant légèrement différent : plutôt que de faire une transformation de Fourier, les fluctuations du CMB sont décomposées sur la sphère céleste à l'aide des harmoniques sphériques. Dans un tel cas, la variable analogue à  $k$  est le multipôle  $l$ . La figure 1.7 présente la carte des fluctuations en température mesurée par Planck, et la figure 1.8 le spectre de puissance de ces fluctuations. Ce dernier présente un pic à  $l \sim 200$ , qui correspond à une taille angulaire sur le ciel d'environ  $1^\circ$ . Cette taille caractéristique correspond à l'horizon acoustique au moment de la recombinaison, elle est expliquée dans la section 1.4.

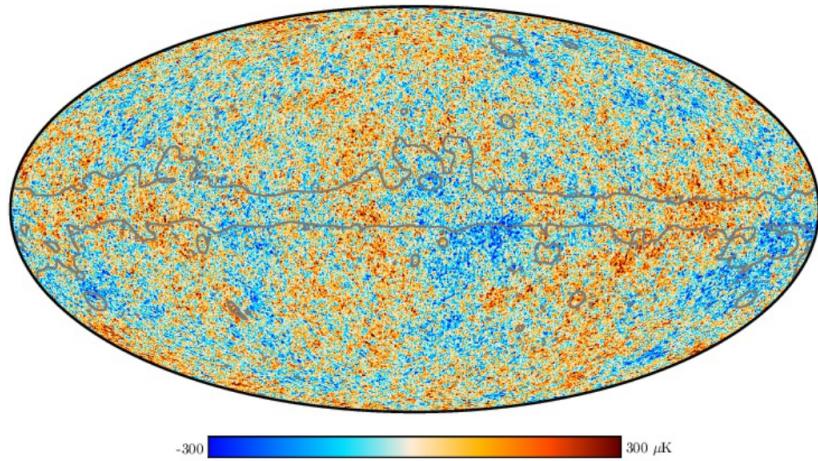


FIGURE 1.7 – Carte des fluctuations en température mesurées par le satellite Planck (PLANCK COLLABORATION et al. 2020). La zone grise délimite la zone contaminée par les émissions galactiques, qui est masquée dans l’analyse.

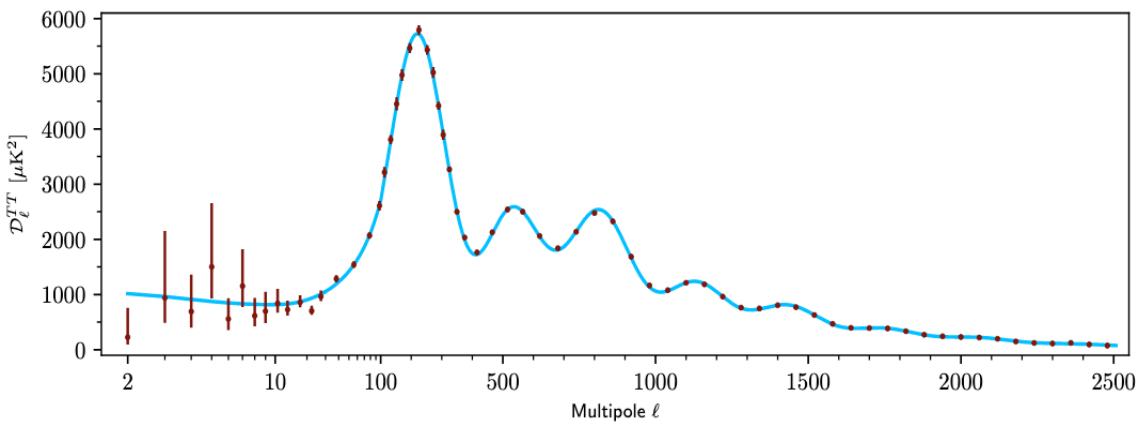


FIGURE 1.8 – Spectre de puissance des fluctuations en température en fonction du multipôle  $\ell$  mesuré par le satellite Planck (PLANCK COLLABORATION et al. 2020). Le premier pic correspond à l’horizon acoustique au moment de la recombinaison. Le plateau à bas  $\ell$  est appelé le plateau de Sachs-Wolfe. L’amortissement à grand  $\ell$  est appelé amortissement de Silk.

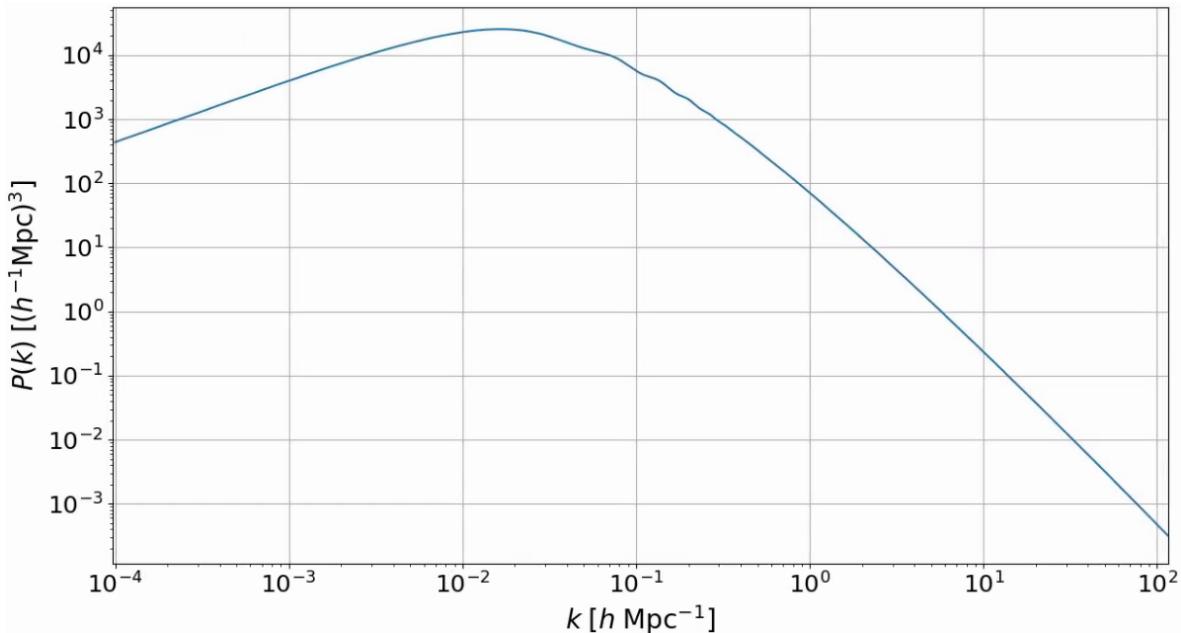


FIGURE 1.9 – Spectre de puissance de la matière à  $z = 0$ . Ce spectre de puissance est obtenu avec le code Camb (LEWIS et al. 1999). Il est calculé en considérant les perturbations linéaires. Les oscillations dues aux oscillations acoustiques de baryon sont visibles pour  $k \in [0,003; 0,3]h \text{ Mpc}^{-1}$ .

### 1.3.3 La fonction de corrélation

Maintenant que nous avons présenté le spectre de puissance, nous allons décrire la fonction de corrélation à deux points, l'objet d'étude de ce manuscrit. De la même manière que nous nous sommes intéressés précédemment au spectre de puissance de la matière, nous nous intéressons ici à la fonction de corrélation de la matière. Elle permet d'étudier de façon statistique la distribution de matière dans l'univers. Plus exactement, elle donne la corrélation de la distribution de matière entre 2 points de l'espace séparés d'une distance  $\vec{r}$ . La fonction de corrélation du contraste de densité de la matière  $\xi$  est définie comme

$$\xi(\vec{r}) = \langle \delta(\vec{r}') \delta(\vec{r} + \vec{r}') \rangle, \quad (1.42)$$

où  $\langle \cdot \rangle$  désigne la moyenne sur  $\vec{r}'$ , et  $\delta(\vec{r})$  est le contraste de densité. Du fait de l'isotropie de l'univers, la fonction de corrélation ne dépend que de la distance  $r$ . Elle peut aussi être vue comme un excès de probabilité :

$$dP(r_1, r_2) = \bar{\rho}^2 (1 + \xi(r_1 - r_2)) dV_1 dV_2, \quad (1.43)$$

où  $dP(r_1, r_2)$  donne la probabilité de trouver de la matière en  $r_1$  et  $r_2$ . Ainsi, si la fonction de corrélation  $\xi(r)$  est positive, alors il est plus probable de trouver de la matière en deux points de l'espace séparés par une distance  $r = r_1 - r_2$  que si celle-ci avait été distribuée de manière uniforme<sup>1</sup>. On peut montrer que la fonction de corrélation est reliée au spectre de puissance par la transformation de Fourier :

$$P(\vec{k}) = \int \xi(\vec{r}) e^{-i\vec{k}\vec{r}} d^3\vec{r}, \quad (1.44)$$

1. Pour une distribution de matière uniforme,  $\xi(r) = 0$  pour tout  $r$ .

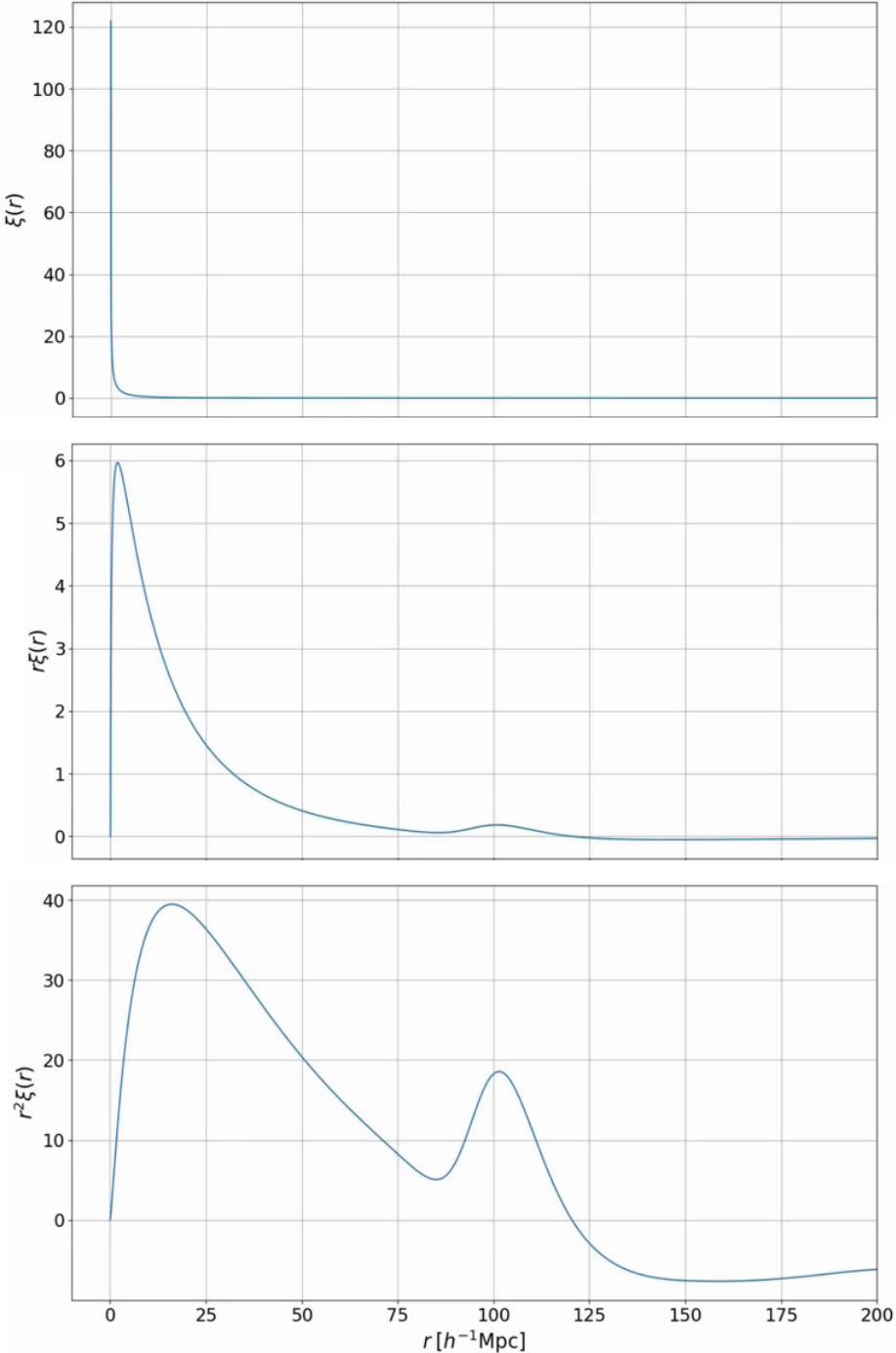


FIGURE 1.10 – Fonction de corrélation de la matière à  $z = 0$ . Elle est obtenue comme la transformée de Fourier du spectre de puissance de la matière, lui-même obtenu avec le code Camb (LEWIS et al. 1999). Les oscillations présentes dans le spectre de puissance et dues aux oscillations acoustiques de baryon se traduisent par un pic à  $r \sim 100 h^{-1} \text{ Mpc}$ . Ce pic est davantage visible lorsqu'on multiplie la fonction de corrélation  $\xi$  par  $r$  (graphique du milieu), ou encore mieux, par  $r^2$  (graphique du bas).

ce qui donne, une fois l'isotropie supposée,

$$P(k) = \frac{i}{4\pi^2 k} \int_{-\infty}^{+\infty} e^{-ikr} r \xi(r) dr. \quad (1.45)$$

Les figures 1.9 et 1.10 présentent respectivement le spectre de puissance et la fonction de corrélation de la matière aujourd'hui. Plusieurs choses sont à noter. Premièrement, le spectre de puissance aux grandes échelles (petits  $k$ ) se comporte comme  $P(k) \propto k^{n_s}$ , où  $n_s$  est l'indice spectral. Ces modes à grande échelle ne sont pas affectés par la physique qui se déroule durant la domination de la radiation. Ils sondent donc directement les fluctuations primordiales de densité. Pour les petites échelles (grands  $k$ ), le spectre de puissance est proportionnel à  $k^{-3}$ . Ce changement de comportement entre les grandes et petites échelles est dû au fait que les modes  $k > k_{eq}$ , avec  $k_{eq} \sim 0,01 h \text{ Mpc}^{-1}$ , possèdent une taille caractéristique plus petite que l'horizon<sup>1</sup> au moment de l'égalité radiation-matière : ces modes sont donc entrés dans l'horizon durant la phase de domination de la radiation, et ont donc été affectés par la physique qui s'y déroule. Les modes plus grands que l'horizon au moment de l'égalité radiation-matière ne sont pas affectés par cette physique et sont donc gelés : ils n'évoluent pas, le spectre de puissance reste donc semblable au spectre de puissance primordial. Les modes plus petits quant à eux croissent moins vite que les modes plus grands que l'horizon durant la phase de domination de la radiation. Plus le mode est petit, plus il entre rapidement dans l'horizon, et donc plus il est réduit comparé aux modes à l'extérieur de l'horizon, d'où le changement de comportement pour les  $k > k_{eq}$ . Après l'égalité, tous les modes croissent proportionnellement à  $G(z)$ .  $G$  est appelé le facteur de croissance des structures, et varie comme  $G(z) \propto (1+z)^{-1}$  à grand  $z$ , lorsque  $\Omega_m = 1$ . Ainsi, pour  $z \ll z_{eq}$ , le spectre de puissance varie comme

$$P(k, z) = G(z)^2 P(k, z=0). \quad (1.46)$$

Le facteur de croissance des structures ne dépendant pas de  $k$ , la fonction de corrélation est aussi proportionnelle à  $G(z)$  :

$$\xi(r, z) = G(z)^2 \xi(r, z=0). \quad (1.47)$$

Tout ceci est un bref résumé de l'évolution des inhomogénéités en cosmologie. Cette dernière est très bien décrite dans le chapitre 7 de DODELSON (2003) et nous référons le lecteur à cet ouvrage pour davantage d'explications.

Enfin, un point pertinent pour ce manuscrit sont les oscillations présentes dans le spectre de puissance de la matière pour  $k \in [0,003; 0,3] h \text{ Mpc}^{-1}$ . Ces oscillations sont dues aux *oscillations acoustiques de baryon* (BAO pour Baryonic Acoustic Oscillations) et sont la trace de la physique qui se déroulait avant l'émission du CMB. Ce mécanisme est décrit plus en détail dans la section suivante. Nous pouvons cependant déjà noter que ces oscillations caractéristiques dans le spectre de puissance correspondent au pic présent dans la fonction de corrélation de la matière à  $r \sim 100 h^{-1} \text{ Mpc}$ . Ce pic est davantage visible lorsque l'on représente  $r^2 \xi(r)$  en fonction de  $r$ .

---

1. L'horizon désigne la sphère causale de l'observateur : tout événement en dehors de l'horizon n'a pas de lien causal avec l'observateur, car l'information n'a pas eu le temps de se propager jusqu'à ce dernier.

## 1.4 Les oscillations acoustiques de baryon

Les BAO sont une empreinte laissée par la physique pré-recombinaison, et détectable aujourd’hui dans la distribution de matière. Cette empreinte correspond à un excès de corrélation de la matière, à une distance comobile d’environ  $100 h^{-1}$  Mpc. Cette distance, appelée *échelle BAO*, fournit une règle standard pour la cosmologie : après l’émission du CMB, la taille comobile de l’échelle BAO reste constante avec le temps. Ainsi, en mesurant l’évolution de la taille physique de l’échelle BAO au cours du temps, grâce notamment à la mesure d’angles et de différences de redshift, nous accédons à l’historique de l’expansion de l’univers. Dans cette section, nous décrivons les BAO, comment elles se forment, comment elles sont mesurées et les contraintes qu’elles permettent d’établir sur les modèles cosmologiques.

### 1.4.1 La genèse

Comme expliqué au début de ce manuscrit, l’univers avant la recombinaison est un plasma chaud et dense, qui présente de faibles inhomogénéités. Les baryons et les photons y sont couplés. De fait, la pression de radiation donne une pression non nulle au gaz et des ondes acoustiques peuvent s’y propager. Ainsi, chaque surdensité primordiale crée une surpression, qui produit une onde acoustique. Cette dernière se propage à la vitesse du son dans ce milieu, donnée par

$$c_s = c \sqrt{\frac{1}{3(1+R)}}, \quad (1.48)$$

où  $R = 3\rho_b/4\rho_\gamma$ . Étant donné que la densité de photons  $\rho_\gamma$  est très supérieure à la densité de baryons  $\rho_b$ , la vitesse du son  $c_s$  vaut  $c/\sqrt{3}$  en bonne approximation. Ces ondes acoustiques se propagent donc dans le plasma primordial depuis chaque surdensité. La figure 1.11 schématise le mécanisme pour une seule surdensité. A l’instant  $t = 0$ , nous considérons donc une surdensité en  $R = 0$  Mpc. Cette surdensité est composée de matière noire (CDM), de baryons et de photons. Grâce à la pression du milieu, une onde acoustique est initiée. Au fur et à mesure que le temps s’écoule (le redshift diminue), le front d’onde dans le plasma photon-baryon se propage. Puis, à un redshift de  $z_* \sim 1090$ , les photons se découpent des baryons. Ils se propagent donc librement. À un redshift  $z_{drag} \sim 1060$ , les baryons se découpent des photons<sup>1</sup>. La pression dans le milieu devient nulle, faisant chuter la vitesse du son à zéro. L’onde est alors gelée. Ainsi la surdensité de baryon ne se propage plus. Il n’y a alors plus que la gravitation qui affecte la distribution de chaque espèce. La surdensité de matière noire à  $R = 0$ , qui a continué de croître, attire les baryons par effet gravitationnel. Cependant, la surdensité de baryon à  $R \sim 150$  Mpc produit aussi un puits de potentiel, dans lequel la matière noire alentour tombe progressivement. À la fin, c’est à dire aujourd’hui, les distributions de la matière noire et des baryons sont quasiment identiques aux grandes échelles. Cette distance d’environ 150 Mpc est appelée *horizon acoustique* :

---

1. A cause de la grande asymétrie entre le nombre de baryons et le nombre de photons  $n_b/n_\gamma \sim 10^{-9}$ , les baryons se découpent des photons après que ces derniers se soient découpés des baryons.

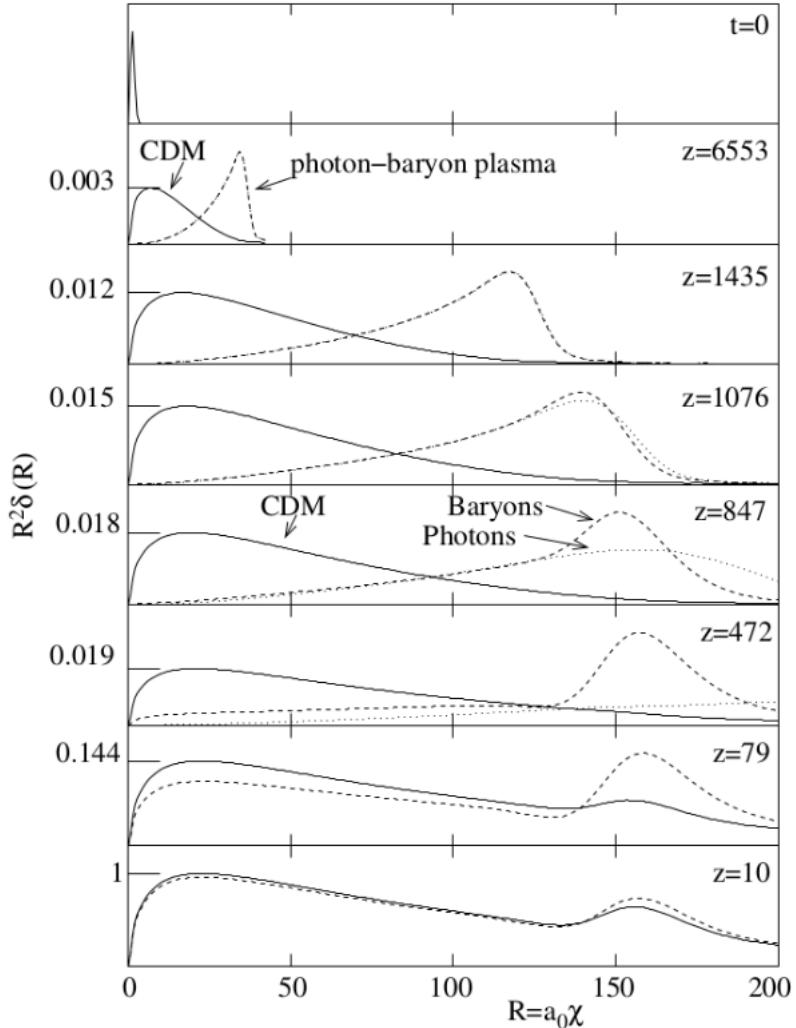


FIGURE 1.11 – Évolution d'une surdensité primordiale depuis  $t = 0$  jusqu'à  $z = 10$ . Sont représentés en lignes continues la surdensité liée à la matière noire, en lignes tiretées la surdensité liée aux baryons, et en lignes pointillées la surdensité liée aux photons. Les neutrinos ne sont pas représentés sur cette figure. L'explication de la figure est donnée dans le corps du texte.

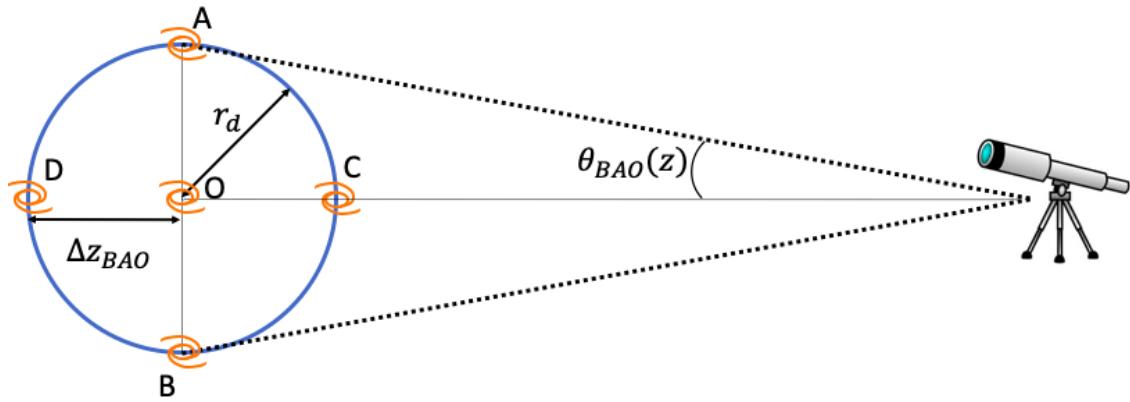


FIGURE 1.12 – Schéma illustratif des distances mises en jeu dans la mesure de l'échelle BAO. L'observateur est représenté par le télescope, à droite. L'horizon acoustique est représenté par le cercle bleu, centré en  $O$ . Des galaxies sont disposées en  $A$ ,  $B$ ,  $C$  et en  $O$ , comme exemple de traceur de la matière. L'angle sous lequel l'observateur identifie l'échelle BAO  $OA$  est  $\theta_{\text{BAO}}(z)$ . Enfin, la différence de redshift entre les objets situés en  $O$  et en  $D$  est  $\Delta z_{\text{BAO}}$ .

c'est la distance que l'onde sonore a pu parcourir avant d'être gelée. L'horizon acoustique vaut

$$r_d = \int_{z_{\text{drag}}}^{\infty} \frac{c_s}{H(z)} dz, \quad (1.49)$$

et sa valeur mesurée par PLANCK COLLABORATION et al. (2018) est  $r_d = (99,16 \pm 0,20) h^{-1} \text{ Mpc}$ .

Ce processus a laissé des traces dans la distribution de matière à grande échelle : à chaque surdensité primordiale est associée une sphère de surdensité de rayon comobile 150Mpc. Il y a donc un excès de probabilité de trouver deux traceurs de densité de matière, comme par exemple des galaxies, séparés par une distance comobile d'environ 150Mpc. Cet excès est traduit par le pic BAO présent dans la fonction de corrélation de la matière, montrée sur la figure 1.10

### 1.4.2 Mesurer l'échelle BAO

Comme nous allons le voir dans ce qui suit, les analyses BAO ne mesurent pas directement l'échelle BAO,  $r_d$ . En astronomie extragalactique, les observations ne permettent pas de mesurer des distances. Les informations auxquelles l'observateur a accès sont des différences de vitesse le long de la ligne de visée, via l'observation des spectres, ainsi que des angles, via les projections sur la sphère céleste. Nous considérons le schéma 1.12, où l'observateur identifie une surdensité primordiale et l'horizon acoustique qui lui correspond, via l'observation de galaxies. Le cercle bleu représente l'horizon acoustique, et le point  $O$  la surdensité primordiale. Grâce aux galaxies situées en  $O$ ,  $A$  et  $D$ , l'observateur a accès à deux informations : l'angle  $\theta_{\text{BAO}}(z)$  séparant  $O$  et  $A$ , et la différence de redshift  $\Delta z_{\text{BAO}}$  entre  $O$  et  $D$ . Comme décrit dans le paragraphe sur les distances, l'angle  $\theta_{\text{BAO}}(z)$  est relié à la distance comobile transverse  $D_M(z)$  par

$$\theta_{\text{BAO}}(z) = \frac{r_d}{D_M(z)}. \quad (1.50)$$

Ainsi, dans la direction transverse à la direction d'observation, l'observateur mesure le rapport  $D_M(z)/r_d$ . Le long de la ligne de visée, l'observateur peut comparer les spectres des galaxies en  $O$  et en  $D$ , et

déduire la différence de redshift  $\Delta z_{\text{BAO}}$  qui existe entre les deux, due à la distance  $r_d$  qui les sépare. La différence de redshift est proportionnelle à la différence de vitesse  $\Delta v_{\text{BAO}}(z)$ , qui s'exprime grâce à la loi de Hubble comme

$$\Delta z = \frac{\Delta v_{\text{BAO}}(z)}{c} = \frac{H(z)r_d}{c}, \quad (1.51)$$

et est donc reliée à la distance de Hubble  $D_H(z)$  par :

$$\Delta z = \frac{r_d}{D_H(z)}. \quad (1.52)$$

Les deux informations accessibles et pertinentes pour les analyses BAO sont donc les quantités  $D_M(z)/r_d$  et  $D_H(z)/r_d$ . Comme dans beaucoup d'analyses cosmologiques, les analyses BAO nécessitent de supposer une cosmologie, afin notamment de transformer les angles et différences de redshift en distance. Ces analyses ne mesurent alors pas directement les rapports  $D_M(z)/r_d$  et  $D_H(z)/r_d$  mais leur déviation par rapport à la cosmologie de référence utilisée dans l'analyse, que l'on nomme *cosmologie fiduciale*. Il est donc coutume de définir les quantités  $\alpha_{\parallel}$  et  $\alpha_{\perp}$  comme

$$\alpha_{\parallel}(z) = \frac{D_H(z)/r_d}{(D_H(z)/r_d)_{\text{fiduciale}}}, \quad (1.53)$$

$$\alpha_{\perp}(z) = \frac{D_M(z)/r_d}{(D_M(z)/r_d)_{\text{fiduciale}}}, \quad (1.54)$$

qui valent 1 si la cosmologie observée est la même que la cosmologie fiduciale.

Les premières observations du premier pic acoustique dans le CMB ont été faites à la fin des années 90. Ces observations ont été menées à l'aide de ballons volant en haute atmosphère (première observation par les expériences QMAP et MAT/TOCO (MILLER et al. 2001) et confirmée par les expériences BOOMERanG (MASI et al. 2002) et MAXIMA (HANANY et al. 2000)). Peu de temps après, l'observation des galaxies locales avec le projet SDSS permettait la première détection du pic BAO dans la distribution de matière (D. J. EISENSTEIN et al. 2005) à bas redshift. La mesure des rapports  $D_M(z)/r_d$  et  $D_H(z)/r_d$  s'est ensuite poursuivie à des redshifts intermédiaires à l'aide des galaxies distantes et des forêts Lyman- $\alpha$  (voir la description de ce traceur dans la section 1.5.4) toujours avec le projet SDSS. Le chapitre suivant présente ce projet, ainsi qu'une revue des différentes mesures des rapports  $D_M(z)/r_d$  et  $D_H(z)/r_d$  (voir notamment la figure 2.9 qui résume les mesures les plus récentes).

### 1.4.3 Contraintes cosmologiques

Indépendamment d'autres sondes cosmologiques, telles le CMB ou les supernovae, les BAO permettent de mesurer les rapports  $D_M(z)/r_d$  et  $D_H(z)/r_d$ . Ces rapports sont reliés aux paramètres

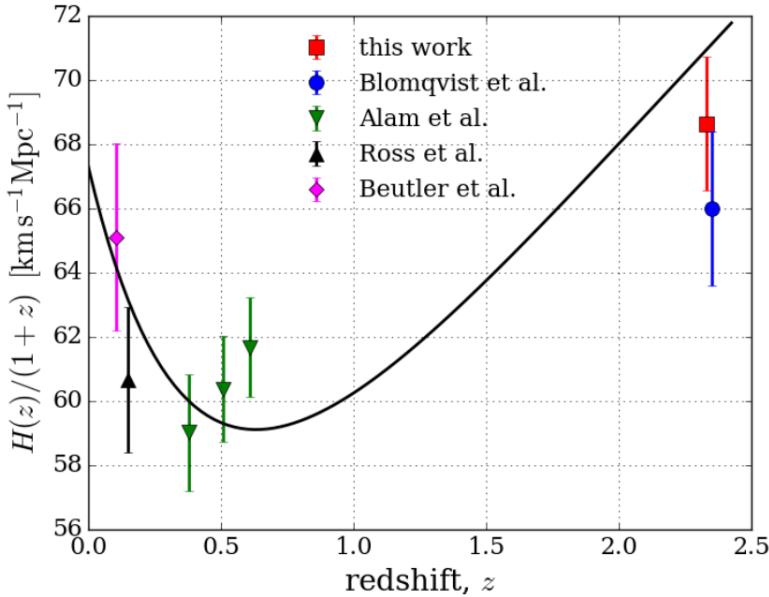


FIGURE 1.13 – Mesure de  $H(z)/(1+z)$  à l'aide des BAO. Les points roses (BEUTLER et al. 2011), noirs (ROSS et al. 2014) et verts (ALAM et al. 2016) présentent les mesures faites avec des galaxies. Le point rouge donne la mesure faite avec l'auto-corrélation du Ly $\alpha$  (DE SAINTE AGATHE et al. 2019) et le point bleu la mesure faite avec la corrélation croisée Ly $\alpha$ -QSO (BLOMQVIST et al. 2019). La ligne noire montre la valeur donnée par la cosmologie de Planck (PLANCK COLLABORATION et al. 2015) dans le cadre de  $\Lambda$ CDM plat. Crédits : DE SAINTE AGATHE et al. (2019).

cosmologiques par

$$D_H(z)/r_d = \frac{1}{H_0 E(z) r_d}, \quad (1.55)$$

$$D_M(z)/r_d = \frac{1}{H_0 E(z) r_d} \int_0^z \frac{dz'}{E(z')}, \quad (1.56)$$

où  $E(z) = \sqrt{\Omega_m(1+z)^3 + \Omega_\Lambda + 1 - \Omega_{\text{total}}}$ .

Ainsi, sans faire de supposition sur le modèle cosmologique, comme par exemple la platitude, les BAO permettent de contraindre les paramètres  $(\Omega_m, \Omega_\Lambda, H_0 r_d)$  du modèle  $\Lambda$ CDM ouvert<sup>1</sup>. La distinction entre ces trois paramètres n'est possible qu'avec des mesures de l'échelle BAO à différents redshifts (voir par exemple la figure 2.11). La figure 1.13 présente la mesure de  $H(z)/(1+z)$  à l'aide des BAO.

## 1.5 Traceurs de la matière

Jusqu'à maintenant, nous avons beaucoup mentionné la fonction de corrélation et le spectre de puissance de la matière. Pourtant, en pratique, ceux-ci ne sont pas directement accessibles. En effet, la matière est constituée à 85 % de matière noire qui, par définition, n'est pas visible. La seule matière observable est la matière baryonique, via l'observation de traceurs. Du fait de la nature de ces traceurs, le champ de matière sondé diffère du champ de matière total sous-jacent, et par conséquent, la fonction

1. Le modèle  $\Lambda$ CDM ouvert désigne le modèle  $\Lambda$ CDM dans lequel on n'impose pas une courbure nulle

de corrélation obtenue à l'aide de ces traceurs n'est pas la fonction de corrélation de la matière. Dans cette section, nous décrivons les différents traceurs utilisés dans l'analyse présentée dans ce manuscrit, et comment, à l'aide de leur fonction de corrélation, déduire la fonction de corrélation de la matière.

### 1.5.1 Traceurs et biais

Le moyen le plus évident auquel nous pouvons penser pour sonder la matière est l'observation de la matière baryonique via la lumière qu'elle émet. Les étoiles en sont un bon exemple. La lumière qu'elles émettent nous permet de les localiser dans notre galaxies, à la différence des planètes qui sont quasiment invisibles. C'est notamment grâce à la lumière émise par les milliards d'étoiles présentes dans les galaxies que nous pouvons observer ces dernières. Mais ceci présente un défaut : la distribution que tracent les étoiles ou les galaxies n'est pas la distribution de matière sous-jacente.

Prenons un exemple. Supposons que nous voulons reconstruire la distribution de la matière dans une région de l'espace. Pour ce faire, nous observons toutes les galaxies, que nous supposons identiques, dans cette région. Les galaxies, comme tous les objets effondrés, se forment dans les endroits les plus denses de l'univers. En dessous d'un certain seuil de densité, les galaxies ne peuvent se former<sup>1</sup>. Ainsi, sonder la distribution de matière via l'observation des galaxies nous fait manquer toute la distribution de matière pour laquelle la densité est inférieure à ce seuil. La distribution reconstruite sera donc plus structurée que la distribution sous-jacente (voir la section 5.7 de PADMANABHAN (1993)). Ce phénomène est décrit par ce qu'on nomme le *biais*. Le biais  $b_i(z)$  du traceur  $i$  au redshift  $z$  est défini par

$$\delta_i(\vec{r}, z) = b_i(z) \delta_{\text{matière}}(\vec{r}, z), \quad (1.57)$$

où  $\delta_i$  est le contraste de densité du traceur  $i$ , et  $\delta_{\text{matière}}$  le contraste de densité de la distribution de matière sous-jacente. Cette expression est valable uniquement aux grandes échelles, pour  $r \gtrsim 15 h^{-1} \text{ Mpc}$ . Le biais relie donc les fluctuations de densité du traceur  $i$ , aux fluctuations de densité de la matière. Pour les objets compacts, telles les galaxies, le biais est supérieur à 1. Plus le traceur se forme dans des régions denses, et plus le biais est grand. Grâce à la relation précédente, nous pouvons relier la fonction de corrélation du traceur  $i$  à celle de la matière par

$$\xi_i(r, z) = b_i^2(z) \xi_{\text{matière}}(r, z). \quad (1.58)$$

Ainsi, par l'intermédiaire du traceur  $i$ , nous pouvons mesurer la fonction de corrélation de la matière, à un facteur multiplicatif près. De plus, en comparant la fonction de corrélation du traceur  $i$  à celle de la matière prédite par les modèles, nous pouvons mesurer le biais  $b_i$  de ce traceur. Comme indiqué dans l'équation précédente, la fonction de corrélation associée au traceur  $i$  est amplifiée par un facteur  $b_i^2$ . Il est donc avantageux de choisir un traceur avec un biais important, afin d'obtenir une fonction de corrélation avec une amplitude importante, et ainsi un rapport signal sur bruit plus grand. La première détection du pic BAO dans la distribution de matière a ainsi été faite avec l'observation des LRG (des galaxies rouges) dont la mesure du biais donne  $b_{LRG} \sim 2$  (D. J. EISENSTEIN et al. 2005).

1. Pour être exact, la probabilité qu'une galaxie se forme dans une région où la densité moyenne est inférieure à ce seuil est extrêmement faible.

### 1.5.2 Distorsions dans l'espace des redshifts

Afin de construire la fonction de corrélation du traceur  $i$ , le redshift cosmologique de chaque traceur est utilisé. Dans le cas des traceurs booléens, comme par exemple les galaxies, le redshift est obtenu en mesurant le spectre de l'objet, puis en comparant les longueurs d'onde des raies d'émission présentes dans le spectre aux longueurs d'onde mesurées en laboratoire. Le redshift mesuré est donc le suivant :

$$z_{\text{mesure}} = z_{\text{cosmo}} + \Delta z_v + \delta z_{\text{sys}} + \delta z_{\text{stat}}, \quad (1.59)$$

où  $z_{\text{mesure}}$  est le redshift mesuré,  $\delta z_{\text{sys}}$  et  $\delta z_{\text{stat}}$  sont les erreurs statistiques et systématiques sur la mesure du redshift,  $z_{\text{cosmo}}$  est le redshift cosmologique, et  $\Delta z_v$  est le redshift induit par la vitesse particulière  $v_{\parallel}$  du traceur le long de la ligne de visée (effet Doppler). Celui-ci s'exprime comme (HOGG 1999) :

$$\Delta z_v = (1+z) \frac{v_{\parallel}}{c}. \quad (1.60)$$

Comme nous l'avons vu dans la section 1.2.2, l'effet Doppler induit par les vitesses particulières des traceurs s'ajoute au redshift cosmologique. Ces deux effets sont indiscernables, et nous n'avons donc pas accès au redshift cosmologique des traceurs observés. Les vitesses particulières des traceurs étant provoquées par la gravité, celles-ci sont corrélées avec le champ de matière sous-jacent : les traceurs ont tendance à se déplacer vers les surdensités par effet gravitationnel. La figure 1.14 illustre la situation : au centre se trouve une surdensité, représentée par un amas de galaxies. Quatre galaxies se trouvent autour de cette surdensité, leur vitesse particulière est représentée par une flèche noire, qui est dirigée vers le centre à cause du puits de potentiel créé par la surdensité. La galaxie se trouvant derrière cette surdensité se déplace donc vers l'observateur, son redshift mesuré est ainsi plus petit que son redshift cosmologique. Similairement, la galaxie se trouvant devant la surdensité est reconstruite avec un redshift plus grand. Les objets se déplaçant perpendiculairement à la ligne de visée ne sont pas affectés. La ligne en pointillés bleus indique la distribution reconstruite dans l'espace des redshifts : cette distribution n'est plus circulaire, elle est aplatie selon la direction de la ligne de visée. Cet effet est appelé *RSD* (Redshift Space Distortions), ou *distorsions dans l'espace des redshifts*.

Ces distorsions résultent d'un effet gravitationnel : les traceurs acquièrent leur vitesse en tombant dans les trous de potentiel créés par les surdensités. L'effet peut donc être modélisé. La formule de Kaiser (KAISER 1987) relie le contraste de densité  $\delta^s(\vec{k})$  dans l'espace des redshifts au contraste de densité  $\delta(\vec{k})$  dans l'espace réel

$$\delta^s(\vec{k}, z) = (1 + f(z)\mu_k^2)\delta(\vec{k}, z), \quad (1.61)$$

où  $f = \frac{d \ln G}{d \ln a}$  est le taux de croissance des structures, et  $\mu_k = \frac{\vec{k} \cdot \vec{u}}{k}$ , où  $\vec{u}$  est la direction de la ligne de visée. Le vecteur  $\vec{k}$  peut être décomposé comme  $\vec{k} = k_{\parallel}\vec{u} + \vec{k}_{\perp}$ , où  $\vec{k}_{\perp}$  est perpendiculaire à  $\vec{u}$ . La quantité  $\mu_k$  vaut alors  $k_{\parallel}/k$ . Lorsque le champ de matière est sondé dans l'espace des redshifts à l'aide d'un traceur  $i$ , l'équation précédente devient

$$\delta_i^s(\vec{k}, z) = (b_i(z) + f(z)\mu_k^2)\delta_{\text{matière}}(\vec{k}, z). \quad (1.62)$$

Nous remarquons ici que le biais ne s'applique pas au terme  $f(z)\mu_k^2$  : les vitesses ne sont pas biaisées.

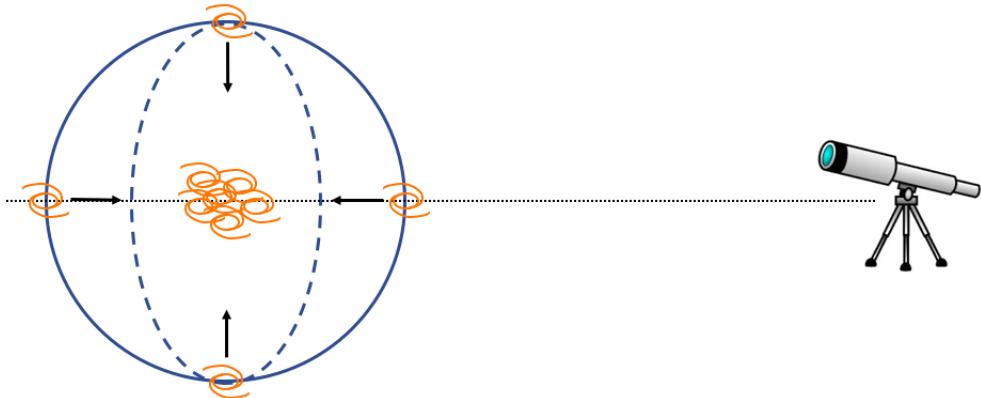


FIGURE 1.14 – Schéma explicatif des distorsions dans l'espace des redshifts. L'observateur est représenté à droite par le télescope. Le cercle bleu en trait plein indique une distribution de matière circulaire dans l'espace réel. Quatre galaxies sont placées sur cette distribution. Au centre se trouve une surdensité, représentée par un amas de galaxies. Le puits de potentiel créé par cette surdensité attire les galaxies par effet gravitationnel. La vitesse particulière de chaque galaxie est indiquée par une flèche noire. A cause de ces vitesses particulières, le redshift observé des galaxies n'est pas leur redshift cosmologique. Ainsi, la distribution reconstruite n'est pas circulaire, mais aplatie selon la direction de la ligne de visée. Cette distribution, dans l'espace des redshifts, est indiquée par une ligne tiretée bleue.

De l'équation 1.62, nous déduisons la relation entre le spectre de puissance  $P_i^s$  du traceur  $i$  dans l'espace des redshifts et le spectre de puissance de la matière :

$$P_i^s(k, \mu_k, z) = b_i^2(z)(1 + \beta_i(z)\mu_k^2)^2 P_{\text{matière}}(k, z), \quad (1.63)$$

où  $\beta_i = f/b_i$  est le paramètre RSD du traceur  $i$ . Du fait de l'isotropie dans la direction transverse à la ligne de visée,  $P_i^s$  ne dépend que de la norme de  $k_\perp$ . Il en va de même pour le vecteur  $\vec{r}$ , qui est décomposé en  $\vec{r} = r_\parallel \vec{u} + \vec{r}_\perp$ . Nous avons alors  $\mu = r_\parallel/r$ , et la fonction de corrélation  $\xi_i^s$  ne dépend que de  $r_\parallel$  et de la norme de  $\vec{r}_\perp$ . L'expression de  $\xi_i^s(r, \mu)$  est donnée dans la section 4.2.4.

L'analyse de la fonction de corrélation selon  $r_\parallel$  et  $r_\perp$  permet donc non seulement de mener une analyse BAO : mesurer la position du pic afin de déduire les quantités  $D_H/r_d$  et  $D_M/r_d$ , mais aussi de mener une analyse dite RSD : mesurer  $\xi_i^s(r_\parallel, r_\perp)$  afin de déduire  $b_i$  et  $\beta_i$  et ainsi mesurer  $f$ . Le taux de croissance  $f$  est une prédiction de la relativité générale, il vaut  $f \sim \Omega_m^{0,55}$  (LINDER et CAHN 2007). La mesure de  $f$  est donc un test direct de la relativité générale et des théories de gravité modifiée.

### 1.5.3 Les quasars

Les quasars, pour *quasi - star*, sont parmi les objets les plus lumineux de l'univers. Ils font partie des noyaux actifs de galaxie, abrégés AGN (*Active galactic nuclei*). Ces galaxies hébergent en leur centre un trou noir supermassif, de quelques millions à plusieurs milliards de masses solaires. La figure 1.15 schématise le noyau actif. Le trou noir supermassif, au centre, accrète la matière environnante. Sous forme d'un disque, cette dernière s'échauffe à cause des collisions et rayonne énormément d'énergie, dans toutes les longueurs d'ondes. Selon les cas, et souvent par cycle, le noyau actif peut émettre de

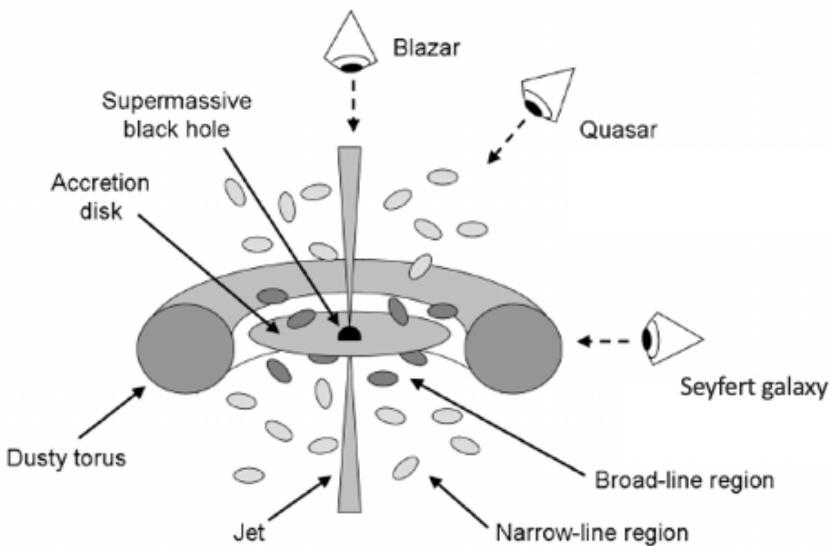


FIGURE 1.15 – Schéma explicatif d'un noyau actif de galaxie. Le trou noir supermassif central est entouré d'un disque d'accrétion. Ce disque d'accrétion, à cause des collisions, s'échauffe et émet puissamment dans toutes les longueurs d'onde. Des jets peuvent se former, et éjectent de la matière ultra-relativiste selon l'axe de rotation. Un tore, constitué de poussière entoure le disque. Lorsque ce tore obstrue la lumière qui atteint l'observateur, l'objet apparaît moins brillant et est classé en tant que galaxie *Seyfert*. Sinon, l'objet est dénommé *quasar*. Lorsque le jet pointe directement vers l'observateur, l'objet est appelé *blazar*. Cependant, les quasars restent mal compris. La formation et l'impact de leurs jets est au cœur de nombreuses études (voir par exemple CHABANIER et al. (2020)).

puissants jets de matière ultra-relativiste. La figure 1.16 montre de tels jets. Ces jets, d'une taille de plus de  $200 h^{-1}$  kpc, sont principalement constitués de noyaux ionisés, d'électrons et de positrons, éjectés à une vitesse proche de celle de la lumière. Ces phénomènes font partie des plus énergétiques de l'univers. Selon l'orientation du disque par rapport à l'observateur, le noyau actif est dénommé différemment. Lorsque nous observons le disque par une direction proche de la tranche, le tore constitué de poussière qui l'entoure, obstrue la lumière. Le noyau actif nous apparaît alors moins brillant, il est classé comme une galaxie *Seyfert*. Lorsque nous observons le disque de face, la lumière n'est pas obstruée, et le noyau actif est alors beaucoup plus brillant. Ces objets sont désignés en tant que *quasar* ou *QSO* (*Quasi Stellar Object*). Lorsque le jet pointe directement vers la terre, l'objet est classifié comme *blazar*.

La grande luminosité des quasars les rend observables à très grand redshift, et permet d'obtenir des spectres avec un bon rapport signal sur bruit à moindre coût. L'obtention du spectre de ces objets est nécessaire pour déterminer leur redshift avec une précision raisonnable. La figure 1.17 présente le spectre d'un quasar avec un grand rapport signal sur bruit et une haute résolution. Ce spectre présente un certain nombre de raies d'émission, parmi lesquelles figurent la raie Lyman- $\alpha$  (Ly $\alpha$ , voir section suivante) qui est la plus intense, ainsi que la raie Lyman- $\beta$  (Ly $\beta$ ), ou encore les raies associées au silicium ou au carbone. Les différentes raies sont référencées dans le tableau 1.2. La détermination du redshift du quasar passe par l'identification d'un certain nombre de ces raies.

Les quasars sont donc des objets idéaux pour construire un relevé spectroscopique étendu et à grand redshift, et ainsi sonder l'échelle BAO sur un grand volume d'univers. De plus, comme expliqué

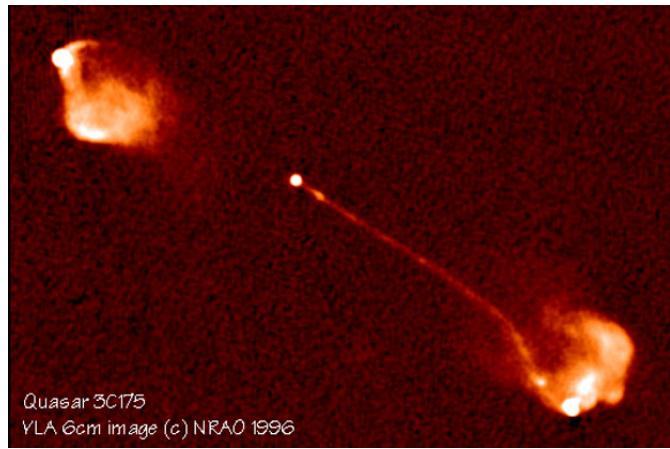


FIGURE 1.16 – Image radio du quasar 3C 175 prise avec le VLA. Les deux jets ultra-relativistes visibles ont une taille d'environ  $212 h^{-1}$  kpc. Crédits : BRIDLE et al. (1994)

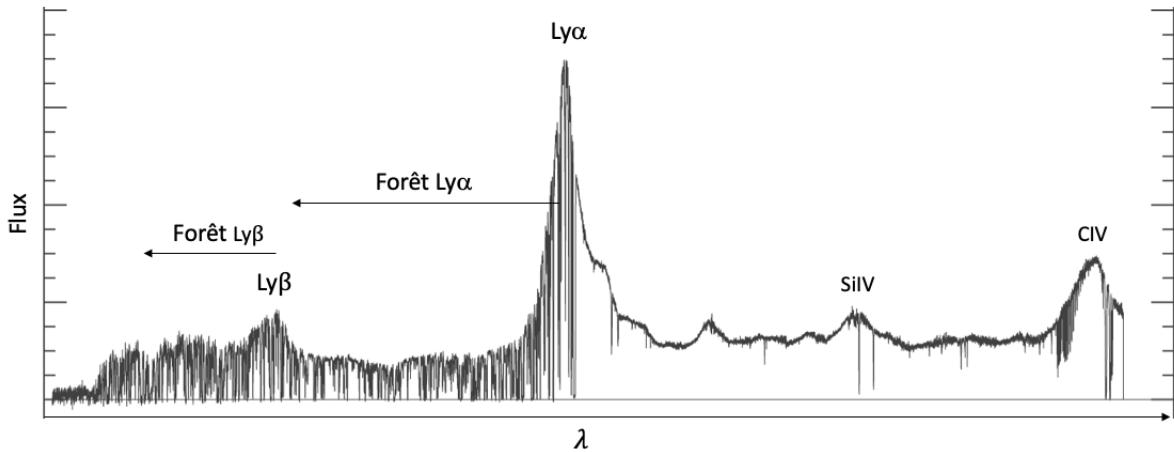


FIGURE 1.17 – Spectre d'un quasar à  $z = 3.5$  possédant un grand rapport signal sur bruit et une haute résolution. Un certain nombre de raies d'émission sont visibles, ainsi que leur forêt associée.

TABLE 1.2 – Liste non exhaustive des principales raies d'émissions présentes dans les spectres des quasars observés par eBOSS. La 3<sup>ème</sup> colonne donne la longueur d'onde de la raie dans le référentiel propre du quasar.

Raie	Notation	$\lambda_{RF} [\text{\AA}]$
Lyman- $\beta$	Ly $\beta$	1025,72
Lyman- $\alpha$	Ly $\alpha$	1215,67
Silicium IV	SiIV(1394)	1394,760 18
Silicium IV	SiIV(1403)	1402,772 91
Carbone IV	CIV(1548)	1548,2049
Carbone IV	CIV(1551)	1550,778 45
Magnesium II	MgII(2796)	2796,3511
Magnesium II	MgII(2804)	2803,5324

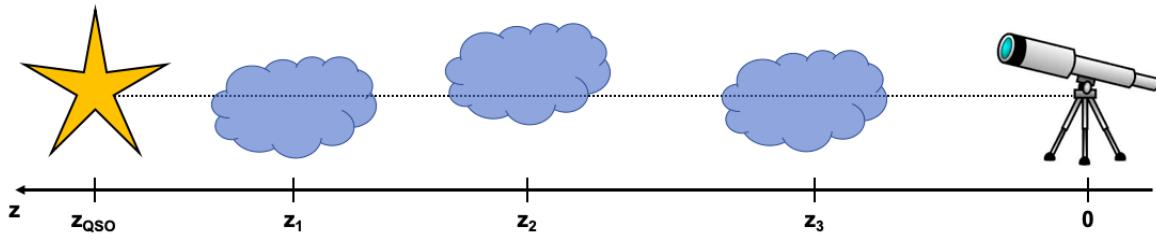


FIGURE 1.18 – Schéma illustratif de l'absorption le long de la ligne de visée d'un quasar. L'observateur, à  $z = 0$  est représenté par le télescope. Le quasar, à  $z_{\text{QSO}}$  est représenté par l'étoile jaune. Les absorbeurs le long de la ligne de visée sont représentés par les nuages, aux redshifts  $z_1$ ,  $z_2$  et  $z_3$ .

précédemment, les quasars sont des objets qui se forment dans des environnements très denses. Ils constituent donc un traceur très biaisé. Le biais des quasars est mesuré par LAURENT et al. (2017) dans quatre bins en redshift (voir table 2). Nous utilisons leurs valeurs ainsi que celles données dans CROOM et al. (2004) pour paramétriser le biais des quasars selon une loi de puissance. La paramétrisation que nous utilisons dans ce manuscrit est la suivante

$$b_{\text{QSO}}(z) = 3,7 \left( \frac{1+z}{1+2,33} \right)^{1,7}. \quad (1.64)$$

Le paramètre RSD des quasars est donné par  $\beta_{\text{QSO}} = f/b_{\text{QSO}}$ . En bonne approximation, pour  $z > 2$ ,  $f = \Omega_m^{0,55} \sim 1$ .

#### 1.5.4 La forêt Ly $\alpha$

Comme expliqué précédemment, le spectre des quasars possède une raie d'émission très intense : la raie Ly $\alpha$ . Cette raie résulte de la désexcitation des atomes d'hydrogène. Découverte par le physicien Theodore Lyman au début du XX<sup>e</sup> siècle, la série de Lyman regroupe les transitions électroniques des états excités de l'atome d'hydrogène vers son état fondamental. Dans son état fondamental, l'électron se trouve sur la couche électronique la plus proche du noyau, il possède un nombre quantique principal  $n = 1$ . Dans un état excité, l'électron se trouve sur une couche externe avec  $n > 1$ . Dans une telle configuration, l'atome d'hydrogène n'est pas stable, il tend à réduire son énergie. L'électron passe alors de la couche électronique sur laquelle il se trouve à une couche électronique de plus basse énergie, pour laquelle le nombre quantique principal est plus faible. La série de Lyman correspond aux transitions pour lesquelles la couche électronique finale est la couche fondamentale  $n = 1$ . Plus l'électron se trouvait initialement sur une couche éloignée du noyau, et plus la transition est énergétique. La raie Ly $\alpha$  correspond à la transition la moins énergétique : de  $n = 2$  à  $n = 1$ . La raie Ly $\beta$  est la seconde moins énergétique : de  $n = 3$  à  $n = 1$ . S'en suivent les autres transitions pour  $n > 3$ . Du fait de l'abondance de l'hydrogène neutre dans le disque d'accrétion, la raie d'émission Ly $\alpha$  très intense.

Grâce à leur luminosité importante dans toutes les longueurs d'onde, les quasars peuvent nous servir à sonder la distribution de matière à grande échelle, en « éclairant » le milieu intergalactique. Nous décrivons dans les lignes qui suivent comment utiliser l'absorption Ly $\alpha$  comme traceur. La figure 1.18 schématisse la situation. Un quasar, situé à un redshift  $z_{\text{QSO}}$  émet de la lumière jusqu'à un observateur situé sur terre, à  $z = 0$ . Trois absorbeurs sont disposés le long de la ligne de visée, à des redshifts  $z_1$ ,

$z_2$  et  $z_3$ . Ils appartiennent aux structures à grande échelle de l'univers. L'hydrogène neutre et non excité présent dans ces absorbeurs peut absorber les photons issus du quasar. D'autres absorptions peuvent avoir lieu. Mais, en première approximation, nous pouvons considérer que ces absorbeurs absorbent uniquement en Ly $\alpha$ . Les photons possédant une longueur d'onde  $\lambda_{\text{Ly}\alpha} = 1215,67\text{\AA}$  dans le référentiel de l'absorbeur sont donc absorbés par l'hydrogène neutre et l'électron de ce dernier effectue une transition électronique des couches  $n = 1$  à  $n = 2$ . Plus l'absorbeur est dense, plus le nombre de photons absorbés est important. Considérons à présent l'absorbeur 1. L'absorption Ly $\alpha$  s'effectue dans son référentiel à  $\lambda_{\text{RF}} = \lambda_{\text{Ly}\alpha}$ , ce qui correspond, dans le référentiel de l'observateur, à  $\lambda_{\text{obs}} = \lambda_{\text{Ly}\alpha}(1 + z_1)$ . De plus, la longueur d'onde correspondant à la raie d'émission du Ly $\alpha$  du quasar dans le référentiel de l'observateur est  $\lambda_{\text{obs}} = \lambda_{\text{Ly}\alpha}(1 + z_{\text{QSO}})$ . Ainsi, comme  $z_1 < z_{\text{QSO}}$ , l'observateur identifie une raie d'absorption à une longueur d'onde plus faible que la raie d'émission Ly $\alpha$  dans le spectre du quasar. Il en est de même pour les 2<sup>ème</sup> et 3<sup>ème</sup> absorbeurs : le spectre observé présente une raie d'absorption à  $\lambda_{\text{obs}} = \lambda_{\text{Ly}\alpha}(1 + z_2)$  et à  $\lambda_{\text{obs}} = \lambda_{\text{Ly}\alpha}(1 + z_3)$ . Le mécanisme se généralise pour tous les absorbeurs présents le long de la ligne de visée. Plus précisément, la distribution de l'hydrogène étant continue, elle produit un profil d'absorption continu le long de la ligne de visée. Cependant, la façon dont sont conçus les spectrographes, et plus précisément les pixels des capteurs CCD, rendent ce profil discret. Ainsi, la région du spectre à gauche de la raie Ly $\alpha$  contient toute une série de raies d'absorption. Cette région s'appelle la forêt Ly $\alpha$ . Elle est visible sur la figure 1.17. Ce même processus est à l'œuvre pour la raie Ly $\beta$ , et pour toutes les autres raies d'émission présentes dans le spectre. Cependant, il est beaucoup plus marqué pour la forêt Ly $\alpha$ . Ceci vient de l'abondance de l'hydrogène dans le milieu intergalactique, et aussi de la grande section efficace de la transition Ly $\alpha$ .

L'absorption Ly $\alpha$  peut donc être utilisée afin de tracer la matière à grande échelle. Contrairement aux quasars et aux galaxies, qui sont des objets ponctuels et donc des traceurs booléens, la forêt Ly $\alpha$  est un traceur continu. En effet, chaque absorbeur le long de la ligne de visée produit une raie d'absorption, dont l'intensité dépend de la densité de l'absorbeur. Ainsi, en plus de nous renseigner sur la distribution spatiale de la matière, chaque absorbeur nous renseigne sur la quantité de matière présente en chaque point sondé. Dans les lignes qui suivent, nous décrivons comment relier la densité d'hydrogène neutre le long de la ligne de visée à la quantité de flux absorbé. Nous négligeons ici et dans la suite de ce manuscrit l'élargissement thermique des raies d'absorption Ly $\alpha$ , la source dominante d'élargissement des ces profils d'absorption, car cet effet est environ cinq fois inférieur à la résolution des données que nous traitons. Considérons un absorbeur très fin, de largeur  $dl$ , à un redshift  $z_{\text{abs}}$ . Le flux absorbé  $dF$  par cet absorbeur est reliée à la densité d'hydrogène neutre  $n_{\text{HI}}$  par la relation

$$dF = n_{\text{HI}} \sigma_{\text{Ly}\alpha} F dl , \quad (1.65)$$

où  $\sigma_{\text{Ly}\alpha}$  est la section efficace de la transition Ly $\alpha$ , et  $F$  le flux émis par le quasar source. La fraction de flux transmis est alors donnée par

$$\frac{dF}{F} = -n_{\text{HI}} \sigma_{\text{Ly}\alpha} dl . \quad (1.66)$$

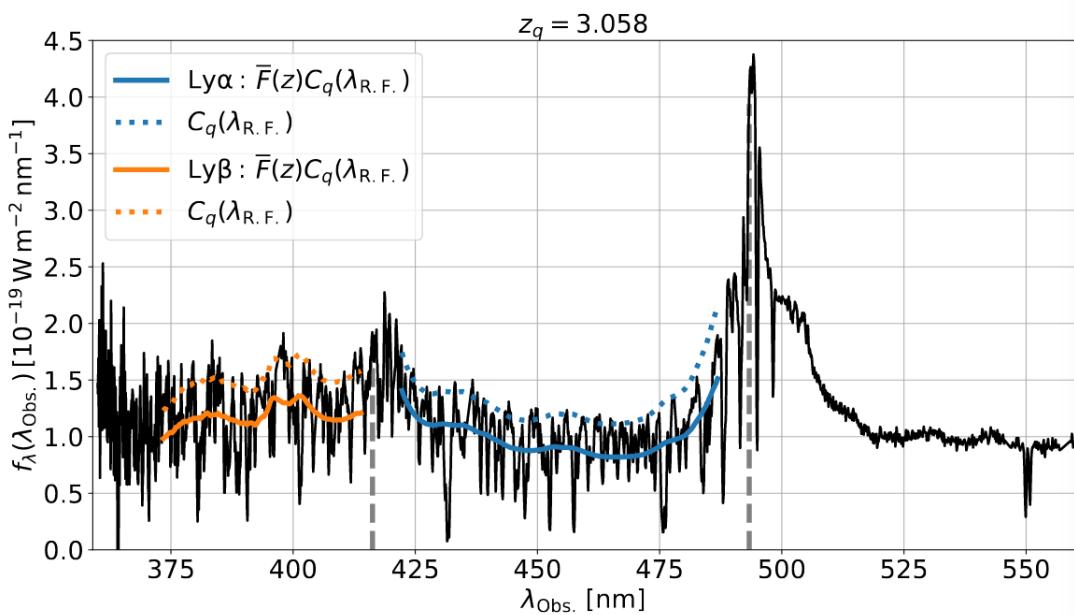


FIGURE 1.19 – Spectre d’un quasar eBOSS à haut rapport signal sur bruit. Le quasar se trouve à un redshift  $z = 3.058$ . Les pointillés gris à droite indique la raie Ly $\alpha$ , et celle de gauche la raie Ly $\beta$ . Les lignes continues donnent les deux meilleurs ajustements indépendants de  $C_q(\lambda_{\text{RF}})\bar{F}(z)$  (le continu du spectre sans absorption multiplié par l’absorption moyenne à ce redshift) pour la forêt Ly $\alpha$ (bleu) et Ly $\beta$  (orange). Les lignes continues donnent le continu  $C_q(\lambda_{\text{RF}})$  en supposant  $\bar{F}(z)$  mesuré par CALURA et al. (2012), pour la forêt Ly $\alpha$  et Ly $\beta$ . L’identifiant du quasar est Thing\_id = 498518806. Crédits : DU MAS DES BOURBOUX et al. (2020)

En intégrant cette équation, nous obtenons

$$F = \exp(-\tau), \quad (1.67)$$

où

$$\tau = \sigma_{\text{Ly}\alpha} \int n_{\text{HI}} dl \quad (1.68)$$

est la profondeur optique de l'absorbeur. Plus l'absorbeur est dense, plus la profondeur optique  $\tau$  est importante et plus la fraction de flux transmis  $F$  est faible. Lorsque tout le flux est absorbé,  $F = 0$ . A l'inverse, lorsque le flux n'est pas absorbé,  $F = 1$ .

Fort heureusement, grâce à l'importante ionisation du milieu intergalactique, et donc de la faible proportion d'hydrogène neutre vis à vis de l'hydrogène ionisé (variant entre 1 partie pour 10 000 et 1 partie pour 1 000), le flux provenant du quasar n'est pas totalement absorbé<sup>1</sup> et nous pouvons ainsi identifier dans le spectre des quasars les raies d'absorption caractéristiques de la forêt Ly $\alpha$ .

Définie précédemment, la fraction de flux transmis  $F$  est notre observable. Nous définissons donc le contraste de densité du Ly $\alpha$  à partir de celle-ci :

$$\delta_{\text{Ly}\alpha}(z) = \frac{F}{\bar{F}(z)} - 1, \quad (1.69)$$

où  $\bar{F}$  est la transmission moyenne au redshift  $z$ .

Du fait que le contraste du Ly $\alpha$  est défini en fonction de l'absorption observée (moins il y a de flux, plus il y a de matière), le biais  $b_{\text{Ly}\alpha}$  qui relie  $\delta_{\text{Ly}\alpha}$  à  $\delta_{\text{matière}}$  est négatif. De plus, étant donné que la forêt Ly $\alpha$  sonde l'hydrogène neutre du milieu intergalactique, composé de gaz très peu dense, les fluctuations de densité de matière tracées appartiennent au régime linéaire, par opposition aux traceurs booléens qui sondent des régions effondrées et non linéaires. Le biais de l'hydrogène est donc proche de 1. Cependant, la non-linéarité de l'équation 1.67 compresse les variations de  $\tau$ , le ramenant entre 0 et 1. Ceci a pour effet de réduire drastiquement le biais du Ly $\alpha$ . A un redshift  $z \sim 2,5$ , le biais de  $F$  est de l'ordre de 0,15. Il est nettement inférieur à celui des LRG, qui vaut environ 2, et aussi à celui des quasars, qui est de l'ordre de 3,5 à  $z \sim 2,5$ . Ce faible biais est compensé par deux effets. Premièrement, pour chaque spectre de quasar observé, la forêt Ly $\alpha$  fournit un grand nombre de pixels, qui chacun trace la densité du champ de matière contrairement aux traceurs booléens. Ceci augmente énormément la statistique. Deuxièmement, le paramètre RSD du Ly $\alpha$  est très important,  $\beta_{\text{Ly}\alpha} \sim 2$  contre  $\sim 0,3$  pour les quasars, ce qui augmente la fonction de corrélation le long de la ligne de visée d'un facteur  $\sim (1 + \beta)^2 = 9$  au lieu de  $\sim 2$ .

Un point important est à noter concernant les RSD du Ly $\alpha$ . L'équation 1.61 est obtenue en utilisant la conservation du nombre de traceurs entre l'espace réel et l'espace des redshifts. Cependant, ce n'est pas le cas du Ly $\alpha$ . En passant de l'espace réel à l'espace des redshifts,  $\tau$  est conservé, mais  $F$  ne l'est pas, à cause de la non-linéarité de l'équation 1.67. Nous pouvons tout de même définir le paramètre

1. L'observation de quasars à des redshifts supérieurs à  $\sim 6$  révèle des régions de l'univers pas encore réionisées : le flux dans une partie ou dans l'entièreté de la forêt Ly $\alpha$  est totalement absorbé (voir par exemple WANG et al. (2018))

RSD du Ly $\alpha$  comme

$$\delta_{\text{Ly}\alpha}^s(\vec{k}, z) = b_{\text{Ly}\alpha}(z)(1 + \beta_{\text{Ly}\alpha}(z)\mu^2)\delta_{\text{matière}}(\vec{k}, z). \quad (1.70)$$

Mais, dans ce cas,  $\beta_{\text{Ly}\alpha} \neq f/b_{\text{Ly}\alpha}$ . Il n'est donc pas possible de mener une analyse RSD uniquement avec le Ly $\alpha$ , car il n'est pas possible de déduire le taux de croissance  $f$  des paramètres  $b_{\text{Ly}\alpha}$  et  $\beta_{\text{Ly}\alpha}$ . Cependant, en combinant les ajustements des fonctions de corrélation Ly $\alpha$  $\times$ Ly $\alpha$  et Ly $\alpha$  $\times$ QSO, il est possible de mesurer les paramètres  $b_{\text{Ly}\alpha}$  et  $\beta_{\text{Ly}\alpha}$  grâce à l'auto-corrélation Ly $\alpha$  $\times$ Ly $\alpha$ , et ainsi distinguer les paramètres  $b_{\text{QSO}}$  et  $\beta_{\text{QSO}}$  des paramètres du Ly $\alpha$  dans la corrélation croisée Ly $\alpha$  $\times$ QSO. Finalement, la mesure de  $b_{\text{QSO}}$  et  $\beta_{\text{QSO}}$  permet de déduire le taux de croissance  $f$  et donc de mener une analyse RSD à grand redshift, grâce au Ly $\alpha$ . Nous présentons les mesures du biais et du paramètre RSD du Ly $\alpha$  dans la section 6.1.1. Nous y discutons la stabilité de ces mesures, et ainsi la faisabilité d'une analyse RSD avec le Ly $\alpha$ .

La forêt Ly $\alpha$  est donc une observable très avantageuse. Elle permet de sonder le régime linéaire des fluctuations de densité en traçant le milieu intergalactique diffus. Elle permet aussi de sonder l'univers à grands redshifts grâce à la grande luminosité des quasars. De plus, chaque spectre de quasar fournit un grand nombre de pixels d'absorption Ly $\alpha$ , qui sondent la densité du champ de matière. Enfin, le relevé spectroscopique de quasar Ly $\alpha$  ne requiert pas une sélection homogène, contrairement aux traceurs booléens. En effet, comme le Ly $\alpha$  est sensible à la densité, via l'observation du flux transmis, il est possible de construire un contraste (équation 1.69) similaire au contraste de densité de la matière (équation 1.39). Ce n'est pas le cas des traceurs booléens car ces derniers ne renseignent pas sur la densité sondée. Les analyses BAO et RSD qui utilisent ces traceurs construisent alors la fonction de corrélation différemment, en ayant recours à l'estimateur de Landy-Szalay. Cet estimateur compare la distribution de traceurs observés à une distribution aléatoire. Il est donc important d'avoir un échantillon de traceurs sélectionnés homogènement à travers le relevé.

### 1.5.5 Les contaminants

Le Ly $\alpha$  n'est pas le seul absorbeur présent dans le spectre des quasars. D'autres espèces peuvent être présentes le long de la ligne de visée, et ainsi causer de l'absorption. Par abus de langage, les métaux désignent en astronomie les éléments qui possèdent un numéro atomique supérieur à 2. À la différence de l'hydrogène et d'une partie de l'hélium qui est d'origine primordiale<sup>1</sup>, les métaux sont formés dans les étoiles, puis sont dispersés via l'explosion des supernovae.

Comme expliqué dans la section précédente, en première approximation, chaque absorption observée dans la forêt Ly $\alpha$  est causée par un absorbeur Ly $\alpha$ . Cependant, certaines absorptions sont causées par des métaux. Étant donné que le champ des métaux est corrélé avec le champ d'hydrogène neutre, les raies d'absorptions des métaux sont corrélées avec celles du Ly $\alpha$ . Supposons la présence d'un métal à un redshift  $z_{\text{met}}$ , absorbant dans son référentiel à une longueur d'onde  $\lambda_{\text{met}}$ . L'absorption produite est observée à  $\lambda_{\text{obs}1} = \lambda_{\text{met}}(1 + z_{\text{met}})$ , mais sera reconstruite à un redshift  $z_{\text{abs}} = \lambda_{\text{obs}1}/\lambda_{\text{Ly}\alpha} - 1 \neq z_{\text{met}}$ . De plus, la présence d'hydrogène neutre au redshift du métal  $z_{\text{met}}$  produit aussi une absorption, à une longueur d'onde  $\lambda_{\text{obs}2} = \lambda_{\text{Ly}\alpha}(1 + z_{\text{met}})$ . Ainsi, à chaque raie d'absorption de métal, interprétée comme

1. L'hélium est formé durant la nucléo-synthèse primordiale.

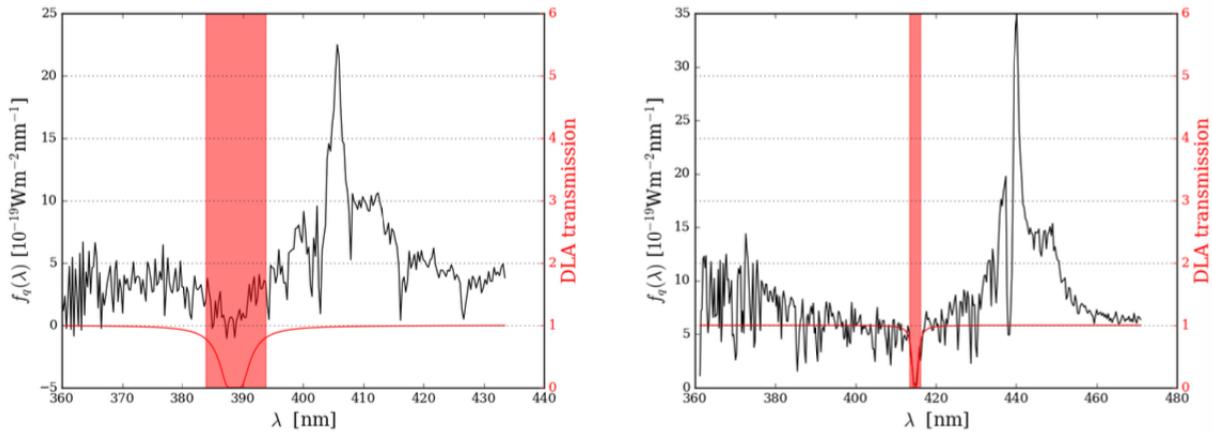


FIGURE 1.20 – La figure présente deux spectres de quasar, en noir, en fonction de la longueur d’onde observée. Chaque quasar possède un DLA. Pour chaque DLA, un profil de Voigt est ajusté. Il est montré en rouge. Les pixels pour lesquels la fraction de flux transmis est inférieure à 0,8 sont masqués. Ces pixels sont contenus dans les bandes rouges. Crédits : DE SAINTE AGATHE (p.d.).

de l’absorption Ly $\alpha$ , est associée une autre raie d’absorption Ly $\alpha$ . Ceci produit une fausse corrélation le long de la ligne de visée, à une séparation caractérisée par le rapport  $\lambda_{\text{met}}/\lambda_{\text{Ly}\alpha}$ . Cet effet est très visible dans la fonction de corrélation à une dimension : lorsque l’on corrèle des pixels d’une même forêt (voir figure 5.19). L’effet est aussi visible dans la fonction de corrélation à trois dimensions le long de la ligne de visée ( $r_{\perp} < 4 h^{-1} \text{ Mpc}$ ) (voir figure 6.1).

Un autre contaminant présent dans la forêt Ly $\alpha$  sont les HCD (High Column Density). Les HCD sont des absorbeurs possédant une densité de colonne<sup>1</sup>, notée  $\log n_{\text{HI}}$ , supérieure à  $1,6 \times 10^{17} \text{ atome/cm}^2$  (ROGERS et al. 2017). Ces absorbeurs très denses produisent de très fortes absorptions. Ils correspondent à des systèmes effondrés, comme le gaz présent dans ou autour des galaxies. Leur fonction de corrélation est donc beaucoup plus grande que celle du Ly $\alpha$ . De plus, du fait de leur très grande densité, le profil d’absorption des HCD est très élargi<sup>2</sup>, produisant une absorption au delà de la position physique de l’absorbeur. Ce profil d’absorption peut être modélisé par un profil de Voigt. Les DLA (pour *Damped Ly $\alpha$  Absorbeur*) sont les HCD pour lesquels la densité de colonne est supérieure à  $2 \times 10^{20} \text{ atome/cm}^2$ . La figure 1.20 montre le spectre de deux quasars possédant un DLA, ainsi que les profils de Voigt ajustés. Ces HCD sont tellement denses qu’ils peuvent être identifiés dans les spectres. Dans l’analyse DU MAS DES BOURBOUX et al. 2020, les DLA identifiés sont modélisés par un profil de Voigt, puis, tous les pixels pour lesquels la fraction de flux transmis est inférieure à 0,80 sont masqués, les autres sont corrigés grâce au profil de Voigt. Les HCD pour lesquels la densité de colonne est inférieure à  $2 \times 10^{20} \text{ atome/cm}^2$  ne sont pas identifiables. Ils ne peuvent donc pas être masqués. Cependant, leur présence doit être modélisée dans le modèle de la fonction de corrélation. Le traitement des HCD est détaillé dans la section 5.4.1.

1. La densité de colonne mesure la quantité de matière intégrée le long de la ligne de visée. Elle est mesurée en atome/cm<sup>2</sup>.

2. A l’inverse, l’élargissement des profils d’absorption du Ly $\alpha$  est dominé par l’élargissement thermique.

# 2

## Le relevé eBOSS de SDSS

---

### Sommaire

---

<b>2.1</b>	<b>Le Sloan Digital Sky Survey . . . . .</b>	<b>44</b>
2.1.1	SDSS I-II-III . . . . .	44
2.1.2	SDSS-IV . . . . .	45
<b>2.2</b>	<b>eBOSS . . . . .</b>	<b>46</b>
2.2.1	Les traceurs de eBOSS . . . . .	46
2.2.2	Sélection des cibles . . . . .	47
2.2.3	Pavage du ciel . . . . .	48
2.2.4	Phase d'observation . . . . .	49
2.2.5	Caractéristiques techniques de l'instrument . . . . .	49
Le télescope . . . . .	49	
La caméra . . . . .	50	
Le spectrographe . . . . .	51	
Les performances . . . . .	53	
2.2.6	Résultats . . . . .	55
<b>2.3</b>	<b>DESI . . . . .</b>	<b>55</b>

---

## 2.1 Le Sloan Digital Sky Survey

Le *Sloan Digital Sky Survey* (SDSS) est né dans les années 1990 de l’ambition de réaliser un relevé spectroscopique à la fois profond et étendu dans le but d’étudier, entre autres, la structuration de l’univers à grande échelle. Cette idée a été rendue possible par l’apparition des capteurs CCD, des spectrographes à fibres optiques ainsi que par les capacités à analyser des lots de données de plus en plus grands. En 1991, la fondation Sloan décide donc d’investir dans ce projet ambitieux, permettant sa construction et lui donnant ainsi son nom. Le télescope fut construit à l’Apache Point Observatory (APO) au Nouveau Mexique, et la première lumière prise en 1998. Ce télescope capable de mesurer le spectre de centaines d’objets à la fois, générant ainsi de grandes quantités de données, a marqué une nouvelle ère, à la fois en terme de relevé cosmologique et de manière d’analyser de telles quantités de données.

### 2.1.1 SDSS I-II-III

Les deux premières générations du SDSS, SDDS-I de 2000 à 2005 puis SDSS-II de 2005 à 2008, ont produit plusieurs résultats importants.

Premièrement, un relevé photométrique multi-bandes de plus de  $11\,600 \text{ deg}^2$ , couvert par les filtres  $u$ ,  $g$ ,  $r$ ,  $i$  and  $z$ <sup>1</sup>. Parmi ces  $11\,600 \text{ deg}^2$ , 7500 forment le relevé contigu appelé *the Legacy Survey*. Ce relevé photométrique a permis de sélectionner plus de 930 000 galaxies et 120 000 quasars candidats, afin d’en mesurer les spectres. Il a été rendu public lors de la septième publication de données de SDSS<sup>2</sup> et est détaillé dans ABAZAJIAN (2008). Son étude a conduit à la première détection à  $3,6\sigma$  du pic BAO (D. J. EISENSTEIN et al. 2005), mesurant ainsi l’échelle BAO moyennée sur le volume,  $D_V$ , à un redshift effectif de 0,35 et avec une précision de 5 %.

Un second relevé photométrique et spectroscopique de  $3500 \text{ deg}^2$ , contenant 460 000 étoiles de notre galaxie, a été produit dans le but d’étudier la structure et la composition de la Voie Lactée. Ce relevé est nommé *SEGUE* (Sloan Extension for Galactic Understanding and Exploration (THE SDSS-II SEGUE COLLABORATION et al. 2009)).

Enfin, le *SDSS Supernova Survey*, un relevé d’environ  $250 \text{ deg}^2$  observé à plusieurs reprises et visant à contraindre l’accélération de l’expansion de l’univers grâce à l’étude des supernovae de type 1a, a produit un échantillon d’environ 500 SN1a, toutes confirmées spectroscopiquement entre  $z = 0,1$  et  $z = 0,4$ .

Au vu du succès de SDSS-I et II, une troisième génération de SDSS débuta fin 2008. Elle se compose de 4 relevés différents (Daniel J. EISENSTEIN et al. 2011), avec le même télescope que les deux générations précédentes.

Parmi ces relevés figure *SEGUE-2*, suite du programme SEGUE. Il a produit un relevé d’étoiles de la Voie Lactée dans le but de mieux comprendre la formation et l’évolution de notre galaxie. Le relevé *MARVEL* (Multi Object APO Radial Velocity Exoplanet Large-ara Survey) a opéré de 2008 à 2012. Ce programme étudie les exo-planètes via deux spectrographes interférométriques de 60 fibres chacun,

---

1. Ces filtres couvrent les longueurs d’onde allant de 400 à 1000 nm. Voir figure 2.4

2. SDSS DR7 : <http://www.sdss.org/dr7>

en mesurant la vitesse radiale d'environ 11 000 étoiles brillantes. *APOGEE-1*, pour APO Galactic Evolution Experiment 1, vise à mieux comprendre la formation et l'évolution de notre galaxie. Il a produit un relevé spectrométrique dans l'infra rouge de très haute résolution et très haut rapport signal sur bruit d'environ 150 000 étoiles. Ce relevé ainsi que celui de SEGUE-2 jouent un rôle important dans la compréhension de la formation de notre galaxie et dans l'étude de la distribution de matière noire à petite échelle.

Enfin, le 4e relevé de SDSS-III est nommé *BOSS*, pour Baryon Oscillation Spectroscopic Survey. Il poursuit la mesure de l'échelle BAO faite avec le Legacy Survey et inclut un nouveau traceur : la forêt Lyman alpha dans le spectre de quasars lointains. Afin d'observer un plus grand nombre d'objets et d'avoir une mesure plus précise, le spectrographe initialement présent sur SDSS a été amélioré : la gamme de longueur d'onde observée est plus grande, le diamètre des fibres a été réduit, ce qui conduit à un nombre total de 1000 fibres. Les observations commencèrent en 2009 et se terminèrent, en avance, en 2014. Le temps d'observation restant permit de mener plusieurs programmes d'observation en dark time<sup>1</sup>, tel que *SEQUELS* (Sloan Extended QUasars, ELG and LRG Survey) qui a réalisé un relevé de 300deg<sup>2</sup> avec le spectrographe de BOSS, afin de tester la sélection des cibles pour SDSS-IV.

*BOSS* mena à de nombreux résultats importants. Il produisit un relevé de plus de 1,5 million de galaxies, qui donna une mesure de  $D_M$  avec une précision de 2,0 % et une mesure de  $D_H$  avec une précision de 3,2 % à un redshift  $z = 0,38$ , une mesure de  $D_M$  avec une précision de 1,9 % et une mesure de  $D_H$  avec une précision de 3,0 % à  $z = 0,51$ , ainsi qu'une mesure de  $D_M$  avec une précision de 1,9 % et une mesure de  $D_H$  avec une précision de 3,0 % à  $z = 0,61$  (ALAM et al. 2016). L'étude du Ly $\alpha$  conduisit à une mesure de  $D_M$  avec une précision de 4 % et une mesure de  $D_H$  avec une précision de 4,8 % à un redshift effectif  $z = 2,4$  (DU MAS DES BOURBOUX et al. 2017). L'ensemble des données de SDSS-III a été rendu public lors de la douzième publication de données de SDSS<sup>2</sup> (ALAM et al. 2015).

### 2.1.2 SDSS-IV

La quatrième génération de SDSS débute courant 2014. Elle comporte 3 programmes spectroscopiques majeurs (Michael R. BLANTON et al. 2017), dont nous donnons une brève description ici :

- *APOGEE-2*, suite d'*APOGEE-1*, effectue de la spectroscopie stellaire grâce à un spectrographe de haute résolution dans le proche infrarouge. En plus du relevé dans le *NGC* (North Galactic Cap) mené à l'APO, le programme comporte un relevé dans le *SGC* (South Galactic Cap). Le relevé sud est produit grâce au télescope du Pont, de 2,5 mètres, situé à Las Campanas au Chili. L'objectif final est d'obtenir le spectre de plus de 400 000 étoiles dans le proche infrarouge, afin d'accéder aux régions clés de la Voie Lactée non-observables en optique, et ainsi mieux comprendre la formation et l'évolution des galaxies.
- *MaNGA* est un nouveau programme de SDSS. C'est le premier à utiliser un spectrographe de champ intégré<sup>3</sup>. Contrairement aux programmes précédents qui obtenaient le spectre du centre de la galaxie cible, *MaNGA* va mesurer le spectre en plusieurs points répartis sur toute la surface

1. Le dark time correspond au moment où le ciel est le plus sombre et le plus propice à observer des objets peu lumineux.

2. SDSS DR12 : <http://www.sdss.org/dr12>

3. *integral field spectrograph* en anglais

de la galaxie observée. Ce procédé va permettre d'étudier la structure interne d'environ 10 000 galaxies proches ( $z < 0,03$ ), en mesurant notamment les cartes des vitesses dans le plan transverse à la ligne de visée et des dispersions de vitesse des étoiles, ainsi que leur métallicité.

- *eBOSS* pour Extended Baryon Oscillation Spectroscopic Survey, est le troisième programme majeur de SDSS-IV. C'est le programme qui fournit les données de ce manuscrit, il est décrit en détail dans la section suivante ainsi que dans DAWSON et al. (2015). En plus du programme principal, eBOSS inclut deux sous-programmes : *SPIDERS* et *TDSS*.

*SPIDERS* étudie la nature des émissions en X. Il utilise 5 % des fibres d'eBOSS afin de viser des sources émettant en X. Ces sources sont issues principalement du relevé *ROSAT All Sky Survey* ainsi que des sources observées par le satellite *eROSITA*.

*TDSS* vise à comprendre la nature des sources variables. Il utilise lui aussi 5 % des fibres d'eBOSS, afin de mesurer le spectre de sources variables présentes dans le relevé *Pan-STARRS1* (PS1) ou dans l'imagerie de SDSS. Parmi les sources identifiées par PS1, environ 60 % sont des quasars, et 40 % des étoiles variables. A la fin, TDSS aura produit un relevé d'environ 200 000 spectres, jusqu'à une magnitude<sup>1</sup> limite de 21.

## 2.2 eBOSS

Suite au succès de BOSS, la collaboration SDSS poursuit la mesure de l'échelle BAO avec le programme eBOSS. Celui-ci inclut deux nouveaux traceurs, afin de sonder une époque clé qui est l'époque de transition de la domination de la matière à celle de l'énergie noire. L'objectif principal est d'étendre la mesure de distance effectuée avec les galaxies de BOSS à un redshift  $0,6 < z < 1$ , ainsi que de permettre la première analyse RSD<sup>2</sup> à l'aide de quasars en tant que traceurs et la première mesure de l'échelle BAO à un redshift  $1 < z < 2$ . La précision visée sur la mesure de distance avec l'échantillon de galaxies est de 1 %, et de 2 % pour les nouveaux traceurs.

### 2.2.1 Les traceurs de eBOSS

Afin d'étendre la gamme de redshift sondée par BOSS, eBOSS inclut donc deux nouveaux traceurs : les ELG et les quasars en tant que traceurs de la matière. Auxquels s'ajoutent les traceurs déjà présents dans BOSS : les LRG et les forêts Ly $\alpha$  des quasars. Ces traceurs sont présentés ici :

- LRG : l'échantillon de *Luminous Red Galaxies* déjà observé lors du programme BOSS est étendu. Un total de 298 762 nouvelles cibles avec une densité de  $60/\text{deg}^2$  sont ajoutées. Ces cibles couvrent la gamme  $0,6 < z < 1$ , portant le redshift moyen de cet échantillon à  $z = 0,72$ .
- ELG (*Emission Line Galaxies*) : ce sont des galaxies bleues formant beaucoup d'étoiles, facilement identifiables par leurs raies d'émission. Leur relevé couvre les redshifts  $0,6 < z < 1,1$ . Le nombre de cibles de ce nouveau traceur est de 269 889, correspondant à une densité d'environ 130 cibles par  $\text{deg}^2$ .

---

1. La magnitude est une mesure de la luminosité des objets célestes. Elle est construite sur une échelle logarithmique inversée : plus la magnitude est importante, et moins l'objet est lumineux.

2. Les analyses RSD désignent les analyses qui visent à mesurer le taux de croissance  $f$  afin de contraire la gravité. Voir la section 1.5.2.

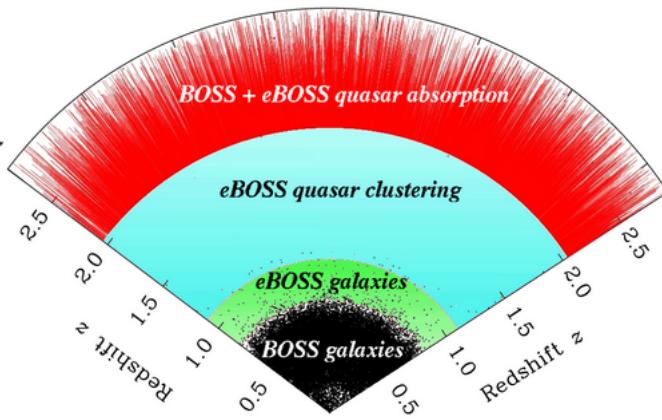


FIGURE 2.1 – Représentation des quatre différents traceurs d'eBOSS et leur répartition en redshift.

- QSO : nouveaux traceurs, ils sont utilisés afin de construire un relevé spectroscopique au même titre que les LRG, mais à plus grand redshift. Ils permettent ainsi de mener une analyse RSD et de mesurer l'échelle BAO dans une époque jusque là insondée ( $0,9 < z < 2,2$ ). Cet échantillon contient 434 820 cibles, avec une densité de  $90/\text{deg}^2$ . Du fait que ce relevé est dominé par le bruit de grenaille (*shot noise*), de futurs relevés pourront venir le compléter et ainsi augmenter la précision statistique de la mesure.
- Ly $\alpha$  QSO : ces quasars sont les objets dont les spectres fournissent la forêt Ly $\alpha$ . Contrairement aux traceurs directs, la quantité d'information contenue dans la forêt Ly $\alpha$  dépend du rapport signal sur bruit. Les quasars Ly $\alpha$  de BOSS possédant un faible rapport signal sur bruit sont donc réobservés dans eBOSS. Le gain attendu sur le rapport signal sur bruit grâce à cette réobservation est d'environ 40 %. La densité moyenne de ces objets est de l'ordre de  $8/\text{deg}^2$ . En plus de ces objets, eBOSS a observé de nouveaux quasars Ly $\alpha$ , avec une densité moyenne de  $11/\text{deg}^2$ .

La répartition en redshift de ces quatre traceurs est représentée sur la figure 2.1. Pour la suite, nous nous intéresserons uniquement aux quasars Ly $\alpha$ , qui constituent l'ensemble des données traitées dans cette étude. Ils seront simplement désignés par quasars ou QSO.

## 2.2.2 Sélection des cibles

La sélection des cibles s'effectue sur la base du relevé photométrique dans les bandes *ugriz* réalisé par SDSS-I et II et rendu publique lors de la neuvième publication de donnée SDSS (DR9<sup>1</sup>). Lors de la sélection des cibles de BOSS, la photométrie provenant de *UKIDSS* (UKIRT Infrared Deep Sky Survey, LAWRENCE et al. (2006)) et de *GALEX* (Galaxy Evolution Explorer, MARTIN et al. (2004)) a été utilisée. De la même manière, la photométrie de SDSS est complétée par plusieurs relevés afin de définir les nouveaux quasars à observer dans eBOSS :

- Les bandes W1 et W2 ( $3,4\mu\text{m}$  et  $4,6\mu\text{m}$ ) du relevé photométrique du satellite *WISE* (Wide-field Infrared Survey Explorer, WRIGHT et al. (2010))

1. <http://www.sdss.org/dr9>

- La photométrie multi-époques de *PTF* (Palomar Transient Factory, LAW et al. (2009))
- Les données de *FIRST* (Faint Images of the Radio Sky at Twenty-Centimeters, BECKER et al. (1995))

Pour BOSS, le relevé de QSO a été construit en utilisant d'une part l'algorithme **XDQSO** (BOVY et al. 2010) pour l'échantillon **QSO\_CORE**, ce qui a permis d'avoir un échantillon homogène, et d'autre part en incluant des QSO sélectionnés via différentes techniques afin d'augmenter au maximum la densité de quasars Ly $\alpha$ . Contrairement aux relevés d'objets utilisés en tant que traceurs directs pour la mesure de l'échelle BAO ou pour les analyses RSD, les quasars Ly $\alpha$  ne nécessitent pas un relevé homogène (voir l'introduction de la section 5.1). C'est pour cette raison que l'échantillon de QSO Ly $\alpha$  peut être complété sans se soucier de dégrader l'homogénéité. Cependant, la présence d'un échantillon homogène de quasars à grand redshift permet d'autres sciences que la mesure de l'échelle BAO. Par exemple, le papier LAURENT et al. (2016) étudie l'homogénéité cosmique en utilisant l'échantillon de quasars DR12 de BOSS.

Pour eBOSS, la présence d'un échantillon de quasars à  $0,9 < z < 2,2$  destiné à la fois à la mesure de l'échelle BAO et aux analyses RSD permet de relâcher le critère d'homogénéité imposé pour les quasars de BOSS, et ainsi d'augmenter le nombre de cibles de quasars Ly $\alpha$ . L'algorithme **XDQSO** est donc utilisé avec des paramètres moins strictes que pour BOSS, afin d'augmenter la densité de quasars Ly $\alpha$ . Ainsi, 6,6 nouveaux quasars par deg $^2$  sont ajoutés à l'échantillon de cette manière. Ensuite, les QSO de BOSS ayant un rapport signal sur bruit  $0,75 < S/R < 3$  et ne comprenant pas de BAL sont réobservés. Enfin, de nouveaux candidats quasars sont sélectionnés grâce aux données de PTF, avec une densité de 3,2 deg $^{-2}$ . Les catalogues FIRST fournissent eux aussi de nouveaux quasars potentiels, avec une densité de 1 deg $^{-2}$ . Ainsi, un total d'environ 8 QSO par deg $^2$  seront réobservés, accompagnés d'environ 11 nouveaux quasars par deg $^2$ .

### 2.2.3 Pavage du ciel

Une fois que les cibles ont été sélectionnées grâce aux observations photométriques, la phase d'observation spectroscopique peut commencer. Les données sont acquises via 1000 fibres optiques insérées dans une plaque, que l'on dispose au centre du plan focal du télescope. Le processus de « pavage » (M. R. BLANTON et al. 2001) consiste donc à assigner chaque cible ou presque à une fibre optique dans une plaque d'observation, en minimisant le nombre de plaques nécessaires et en maximisant le nombre de cibles qui reçoivent une fibre. Cette opération est réalisée par l'algorithme de sélection des cibles ainsi que du nombre de fibres disponibles.

Afin d'observer les 9000 deg $^2$  constituant le relevé d'eBOSS, celui-ci est divisé en environ 1800 plaques. Chaque plaque couvre une surface de 7 deg $^2$  sur le ciel, et en moyenne, un centre de plaque est assigné tous les 5 deg $^2$  afin d'éviter les trous dans le relevé. Étant donné le nombre d'heures allouées à eBOSS et le nombre de plaques à observer, chaque plaque est observée durant environ 1,5 heure.

Parmi les 1000 fibres disposées sur chaque plaque, 100 sont destinées aux programmes TDSS et SPIDERS et 100 fibres supplémentaires sont réservées à la calibration. Il reste donc 800 fibres par plaque

destinées aux traceurs de eBOSS. Ces fibres sont assignées aux LRG et QSO. Le relevé des ELG se fait sur des plaques indépendantes. Afin de mener ce relevé des ELG, la taille du relevé des LRG et QSO est réduit de 9000 à  $7500\text{deg}^2$ . Ainsi 300 plaques sont rendues disponibles pour l'observation des ELG ( $\sim 1500\text{deg}^2$ ).

Une fois le pavage du ciel effectué, la position des fibres sur le ciel est convertie en coordonnées dans le plan focal du télescope. A cause de la chromaticité de l'instrument, la position dans le plan focal de chaque objet observé dépend de la longueur d'onde. Ainsi chaque fibre dite de science est positionnée de manière à maximiser la lumière en sortie à  $5400\text{\AA}$  pour les galaxies et les quasars destinés aux analyses RSD, et à  $4000\text{\AA}$  pour les quasars Ly $\alpha$ .

Parmi les 100 fibres allouées à la calibration, 80 sont destinées à la soustraction du fond de ciel. Pour chaque plaque, il est requis que chaque spectrographe reçoivent au moins 30 fibres de ciel. Les 20 fibres restantes sont destinées à la calibration du flux. Celle-ci se fait en pointant des étoiles standards de type F. De la même manière, parmi ces 20 fibres, au minimum 6 fibres sont requises sur chaque spectrographe.

#### 2.2.4 Phase d'observation

Une fois le pavage du ciel réalisé et la position de chaque fibre dans le plan focale déterminée, les plaques sont préparées puis percées. Ces plaques sont en aluminium de 3,2mm d'épaisseur et de 80cm de diamètre. La zone contenant les fibres mesure 65,2cm de diamètre. La préparation des plaques est faite à l'université de Washington. Elle est décrite dans Michael R. BLANTON et al. (2017).

Une fois les plaques prêtes, les observations peuvent commencer. La prise de données a débuté en juillet 2014. Pendant les deux premières années, seules les plaques des LRG et des QSO ont été observées. Les deux années suivantes, les plaques des ELG ont été observées en alternance avec les plaques assignées aux LRG et QSO. Les 305 plaques formant le relevé d'ELG ont fini d'être observées en février 2018. En mars 2019, les observations sont arrêtées afin de laisser les autres programmes observer. Contrairement à BOSS, les observations pour eBOSS ont connu un mauvais temps, retardant l'avancement du relevé. Ce retard a été essentiellement répercuté sur l'observation des plaques contenant les LRG et QSO : environ 1000 plaques sur les 1500 initialement prévues ont pu être observées, réduisant le relevé de 7500 à environ  $4700\text{deg}^2$ .

#### 2.2.5 Caractéristiques techniques de l'instrument

L'instrument (J. E. GUNN et al. 2006) utilisé pour eBOSS est celui de SDSS, situé à l'APO. Nous décrivons ici ses composantes importantes.

##### Le télescope

Le télescope est installé à l'APO. Il est commun à tous les programmes SDSS. Afin de mener à bien ces différents programmes, le télescope doit être capable de réaliser un relevé photométrique d'un quart du ciel, puis mesurer le spectre de millions de cibles identifiées via cette photométrie. Ainsi le télescope doit avoir un grand champ de vue, avec très peu de distorsions dans le plan focal. Ces prérequis ont

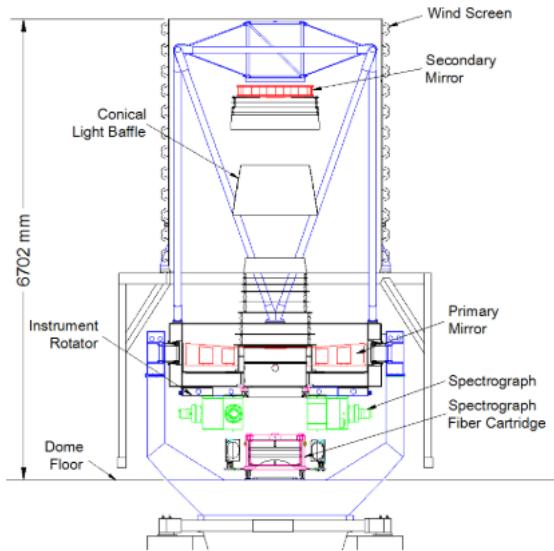


FIGURE 2.2 – Schéma du télescope SDSS. Le miroir principal est représenté en bas en rouge. Le miroir secondaire est visible au sommet du télescope en rouge. Les spectrographes sont représentés en vert sous le télescope. Une cartouche, pas encore disposée au plan focal du télescope, est représentée sous les spectrographes. Crédits : SMEE et al. (2012).

conduit à la construction du télescope SDSS de 2,5m de diamètre de type Ritchey-Chrétien. Il est représenté schématiquement sur la figure 2.2.

Le télescope se compose d'un miroir primaire de 2,5m de diamètre et d'ouverture  $f/2,25$ , et d'un miroir secondaire de 1,08m de diamètre situé à 3,6m du miroir primaire. Avec un plan focal situé à 0,76m derrière le miroir primaire, l'ouverture finale du télescope est  $f/5$ . Le champ de vue qui en résulte est de 3deg de diamètre sur le ciel ( $7\text{deg}^2$ ), correspondant à un diamètre de 0,65m dans le plan focal.

Le télescope inclut aussi 2 correcteurs optiques. Le premier est un correcteur d'astigmatisme de type Gascoigne. Le second est un jeu de correcteurs hautement asphériques et interchangeables situés près du plan focal. L'un, épais, est utilisé pour la photométrie ; l'autre, beaucoup plus fin, est utilisé lors des phases de spectrométrie.

### La caméra

L'instrument nécessite une caméra (J. GUNN et al. 1998) capable de couvrir l'entièreté du plan focal du télescope. Le très grand champ de vue du télescope a imposé l'utilisation des CCD les plus grands disponibles à l'époque : les Tektronix Tk2048E. Ces CCD sont des grilles de  $2048 \times 2048$  pixels de  $24\mu\text{m}$ . Étant donné la longueur focale du télescope, ces  $24\mu\text{m}$  représentent  $0,4''$  sur le ciel. Ainsi, la PSF<sup>1</sup> d'une largeur à mi-hauteur d'environ  $1''$  est bien échantillonnée.

La caméra est constituée de 2 modules, le premier comportant 5 (filtres<sup>2</sup>)  $\times$  6 (colonnes) CCD est destiné à la photométrie. En plus de ces 30 CCD, 22 CCD  $400 \times 2048$  et de même taille de pixel sont

1. La PSF (*Point Spread Function*) désigne la réponse d'un système optique à une source ponctuelle.

2. Il s'agit des 5 filtres u, g, r, i et z. Voir figure 2.4

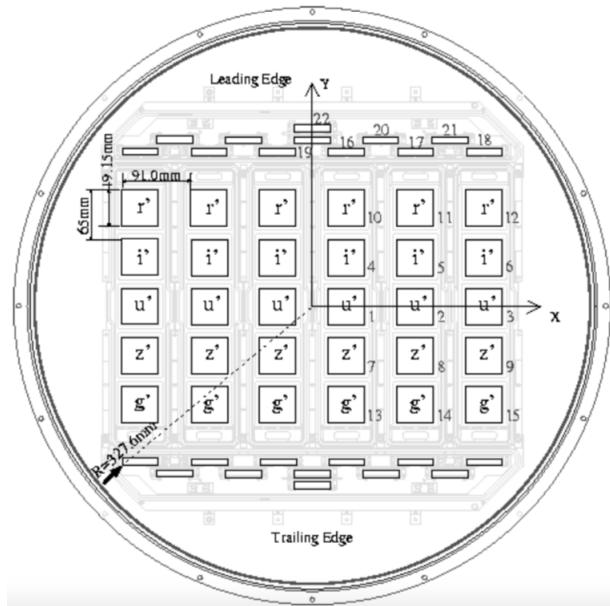


FIGURE 2.3 – Schéma de la disposition des CCD dans le plan focal du télescope SDSS. Les capteurs 1 à 15 sont les CCD dédiés à la photométrie. Le module d'astrométrie se compose des CCD 16 à 21. Enfin le CCD 22 sert au contrôle de la mise au point. Crédits : (J. GUNN et al. 1998)

ajoutés au dessus et en dessous du module dédié à la photométrie. Ces CCD permettent de relier les étoiles standards brillantes aux objets observés lors de la phase photométrique. Ils constituent le module d'astrométrie. Deux CCD supplémentaires sont ajoutés comme dispositif de contrôle pour la mise au point. La figure 2.3 résume leur disposition.

Lors de la phase d'observation photométrique, la caméra est utilisée en mode *time delay integration* (TDI). Ce mode d'observation consiste à laisser le ciel défiler devant la caméra. La lumière de chaque objet est ainsi accumulée durant tout le transit de l'objet. Cette technique permet de gagner en efficacité d'observation, en réduisant le temps de lecture (qu'il y aurait en mode exposition classique) et le temps de pointage du télescope.

### Le spectrographe

Une fois la phase de photométrie effectuée et les cibles sélectionnées, la phase de spectroscopie commence. Les plaques d'observations sont d'abord percées aux positions des cibles dans le plan focal. Ces plaques sont ensuite disposées au plan focal du télescope (à la place de la caméra). Les fibres, insérées dans ces plaques et placées à la position de chaque cible, sont ensuite envoyées vers les spectrographes (SMEE et al. 2012) afin de mesurer le spectre de ces cibles.

Les spectrographes utilisés dans eBOSS sont les mêmes que ceux utilisés dans BOSS. Ces spectrographes, initialement présents dans SDSS, ont été améliorés afin d'atteindre les objectifs de BOSS. Comparé à SDSS, BOSS a augmenté le nombre de spectres mesurés de 35 %, ses objets étant à plus grand redshift, donc de plus faible luminosité. Ainsi le nombre de fibres des spectrographes de BOSS passe de 640 à 1000. Les objets observés étant plus lointains et donc ayant une taille sur le ciel plus petite, le diamètre de ces fibres est réduit d'un tiers, passant à 120 µm, afin d'augmenter le rapport

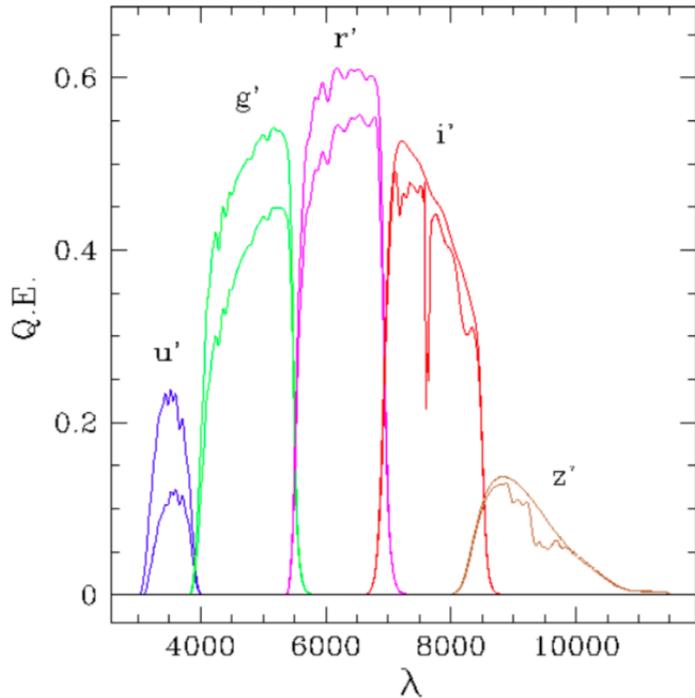


FIGURE 2.4 – Efficacité de chacun des 5 filtres utilisés lors de la phase photométrique de SDSS. Les courbes incluent l’efficacité quantique des CCD, ainsi que l’efficacité du système optique. Les courbes en dessous incluent en plus la transmission de l’atmosphère. Crédits : (J. GUNN et al. 1998)

signal sur bruit des spectres. BOSS inclut aussi un nouveau traceur : la forêt Ly $\alpha$ . Afin de mesurer l’absorption dans la forêt Ly $\alpha$  des quasars à un redshift  $z = 2,2$ , la longueur d’onde d’observation minimale a été diminuée de 3900 à 3560 Å. De la même manière, la longueur d’onde d’observation maximale a été augmenté de 9100 à 10400 Å pour améliorer la détermination des redshifts de l’échantillon de galaxies.

Pour limiter les risques de dommage aux fibres lors du montage et du démontage des différentes plaques, chaque plaque est montée sur un support rigide. Ce support rigide est appelé cartouche, il comporte la plaque d’observation en aluminium, les 1000 fibres insérées dans cette plaque, et deux *slitheads* qui sont ensuite insérés dans chacun des spectrographes. La figure 2.5 donne un aperçu de ces cartouches.

L’instrument dispose de deux spectrographes. Ils sont schématisés sur la figure 2.6. Les 1000 fibres d’une plaque sont séparées en deux. Chaque lot de 500 fibres constitue une *demi-plaque*. Chaque spectrographe reçoit donc les fibres optiques d’une demi-plaque, via les slitheads. La lumière issue de ces fibres est collimatée grâce à un miroir sphérique. Le faisceau parallèle ainsi créé passe à travers un miroir semi-réfléchissant, permettant de séparer les longueurs d’onde plus petites que 6050 Å des longueurs d’onde plus grandes. Enfin, chaque demi-faisceau passe au travers d’un grisme<sup>1</sup>, et arrive sur la caméra bleue pour les longueurs d’onde plus petites que 6050 Å, ou sur la caméra rouge pour les longueurs d’onde plus grandes. Après avoir traversé toutes les pièces d’optique, la lumière arrive

1. Association d’un prisme et d’un réseau de diffraction. Le grisme permet de décomposer la lumière tout en gardant le faisceau parallèle.

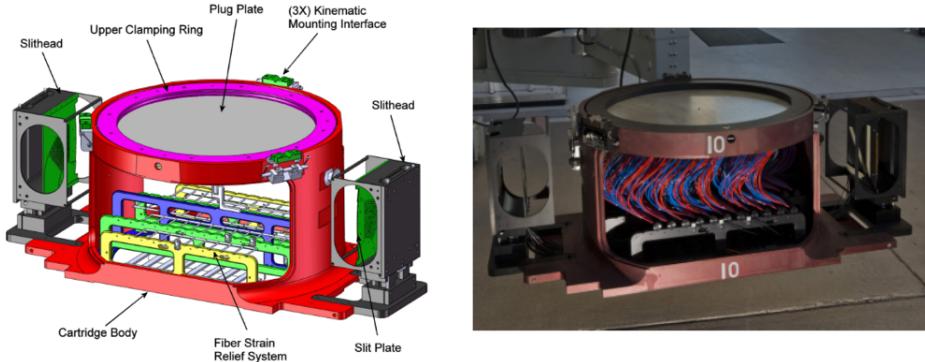


FIGURE 2.5 – Le schéma de gauche présente les différents éléments d'une cartouche. On peut y reconnaître la plaque en aluminium sur le dessus en gris. Sur la droite, une photo d'une cartouche sur laquelle les fibres optiques ont été insérées dans la plaque d'aluminium. L'extrémité des fibres est reliée aux slitheads, visibles à droite et à gauche, qui seront insérés dans les 2 spectrographes. Crédits : SMEE et al. (2012).

sur des CCD de  $4000 \times 4000$  pixels, avec une taille de pixel de  $15\mu\text{m}$ . Ainsi une des dimensions du CCD correspond à la longueur d'onde observée, selon l'axe de dispersion, l'autre dimension parcourt les différentes fibres. Selon cette dimension, chaque spectre possède un profil de 3 pixels de large et est séparé de son voisin par 6 pixels afin d'éviter la contamination d'un spectre à un autre.

### Les performances

Les améliorations apportées à l'instrument pour BOSS ont permis d'augmenter le nombre maximal de spectres observables par nuit, ainsi que la magnitude limite atteignable. Cette dernière est directement liée à l'efficacité optique du système. L'efficacité optique est définie comme le ratio du flux mesuré d'une source ponctuelle sur le flux de cette même source situé en dehors de l'atmosphère. La figure 2.7 présente les estimations des différentes composantes participant à l'efficacité optique globale de l'instrument.

Le pouvoir de résolution traduit la capacité de l'instrument à identifier et mesurer des raies spectrales. Pour mesurer ce pouvoir de résolution, le spectre de lampes à arc dédiées à la calibration est mesuré, puis chaque raie d'émission est ajustée par une gaussienne de largeur  $\sigma_\lambda$ . Ce  $\sigma_\lambda$  est ensuite ajusté par un polynôme en fonction de  $\lambda$ , ce qui donne une estimation de la largeur d'une raie spectrale en fonction de la longueur d'onde observée. Enfin, le pouvoir de résolution est défini comme  $R = \frac{\lambda}{FWHM}$ , où la largeur à mi-hauteur FWHM (pour *Full Width at Half Maximum*) vaut  $2,36\sigma_\lambda$  et donne la résolution de l'instrument. Le pouvoir de résolution a été mesuré sur 100 plaques SDSS et 100 plaques BOSS. La comparaison est présentée sur la figure 2.8. Le pouvoir de résolution est légèrement moins grand pour BOSS que pour SDSS, mais reste au dessus des prérequis.

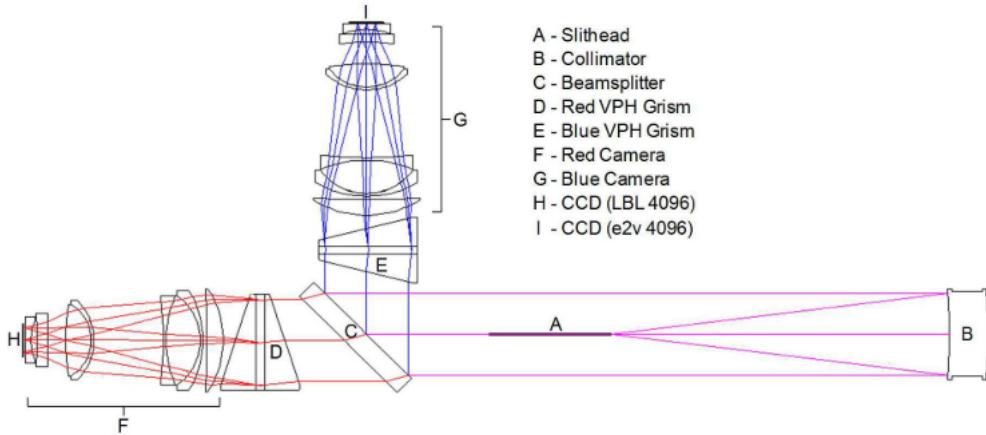


FIGURE 2.6 – Schéma des spectrographes de BOSS. La lumière arrive via les fibres optiques (A). Elle est ensuite collimatée (B) en un faisceau parallèle, puis séparée par le miroir semi-réfléchissant (C). Les longueurs d'ondes  $\lambda < 6050$  sont réfléchies vers la caméra bleue (G), les autres entrent dans le bras rouge (F) du spectrographe. Chaque bras comporte un grisme (D et E), une série de lentille puis le CCD (H et I). Crédits : SMEE et al. (2012).

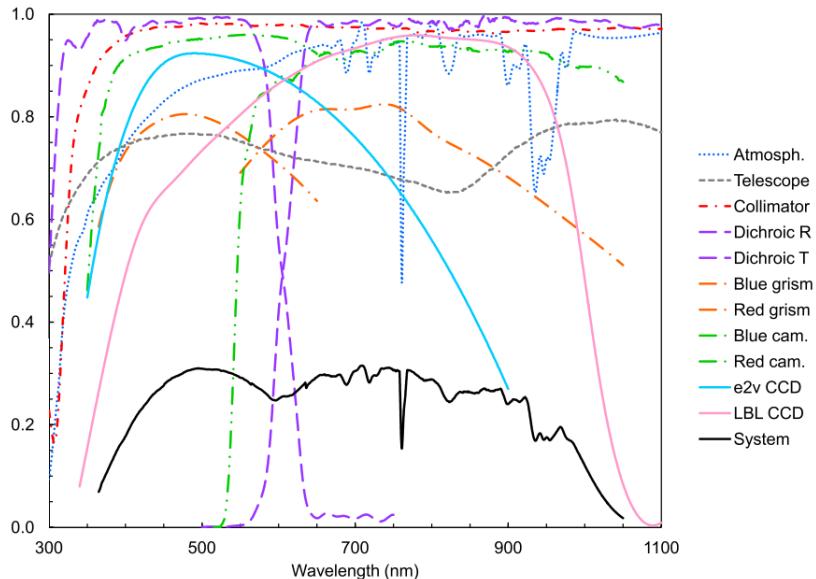


FIGURE 2.7 – Efficacité optique en fonction de la longueur d'onde. Le graphique présente les prévisions de l'efficacité optique de l'instrument avec toutes ses composantes. Crédits : SMEE et al. (2012).

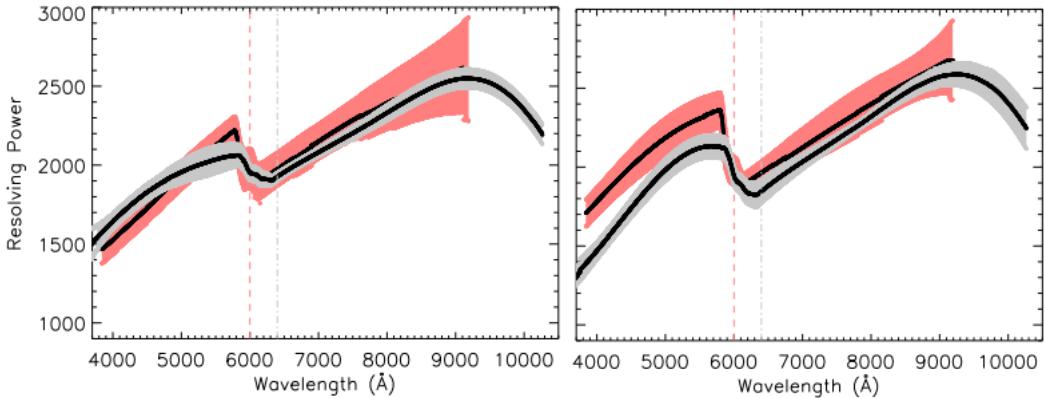


FIGURE 2.8 – Pouvoir de résolution en fonction de la longueur d’onde pour les spectrographes de SDSS (rouge) et BOSS (gris). La courbe de gauche correspond à la mesure sur le spectrographe 1, et celle de droite sur le spectrographe 2. Les régions colorées représentent les régions contenant 68 % des mesures. Crédits : SMEE et al. (2012).

### 2.2.6 Résultats

Le relevé eBOSS fournit les mesures de l’échelle BAO et du taux de croissance  $f\sigma_8$  les plus précises à ce jour. Ces mesures, grâce à la diversité de traceurs utilisés, sont réparties dans une large gamme en redshift, allant de  $z \sim 0.15$  avec les galaxies locales à  $z \sim 2.4$  avec le Ly $\alpha$ . La figure 2.9 présente l’ensemble de ces mesures. Parmi les nombreux résultats produits par eBOSS, nous pouvons citer la mesure de l’échelle BAO qui favorise une densité d’énergie noire non nulle à plus de 8 sigmas. La figure 2.10 montre les contraintes sur les paramètres  $\Omega_\Lambda$  et  $\Omega_m$  apportées uniquement par la mesure de l’échelle BAO. En ajoutant les mesures de la BBN (*Big Bang Nucleosynthesis*), la mesure de l’échelle BAO faite par eBOSS produit une mesure de  $H_0$  en accord avec les mesures de Planck, et en tension avec les mesures directes. La figure 2.11 montre les contraintes sur les paramètres  $H_0$  et  $\Omega_m$  apportées par la mesure de l’échelle BAO. La combinaison des mesures à  $z > 1$  et  $z < 1$  permet de mesurer précisément à la fois  $H_0$  et  $\Omega_m$ . La zone horizontale grise montre la mesure directe de  $H_0$  faite par Riess et al. (2019).

Par ailleurs, les analyses RSD menées grâce aux données d’eBOSS produisent des mesures de  $f\sigma_8$  qui sont compatibles avec les mesures produites par Planck, et compatibles avec les prévisions de la relativité générale.

Enfin, eBOSS permet de réduire la limite supérieure de la somme des masses des neutrinos à 0,111 eV (à 2  $\sigma$ ). Tous les résultats et contraintes cosmologiques produits par eBOSS sont présentés et discutés dans COLLABORATION et al. (2020).

## 2.3 DESI

Le *Dark Energy Spectroscopic Instrument* (DESI, DESI COLLABORATION et al. (2016)) est un projet américain de mesure d’énergie noire de génération 4. Il a vu sa première lumière en octobre 2019. Le début de la prise de données devait avoir lieu en juillet 2020. Cependant, à cause de la pandémie de 2020, les phases de *commissioning* et *survey validation* durant laquelle l’instrument et

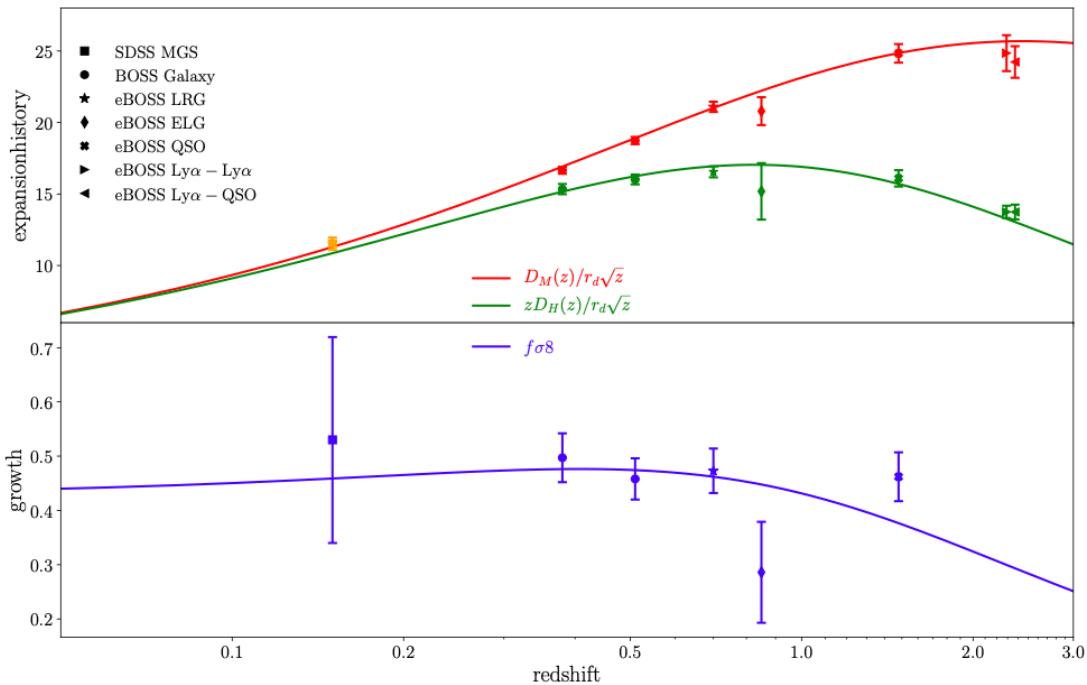


FIGURE 2.9 – L’ensemble des mesures de l’échelle BAO (graphique du haut) et du taux de croissance  $f\sigma_8$  (graphique du bas) produites par tous les relevés SDSS (COLLABORATION et al. 2020). Les lignes continues indiquent les prédictions produites par Planck en supposant un modèle  $\Lambda$ CDM plat.

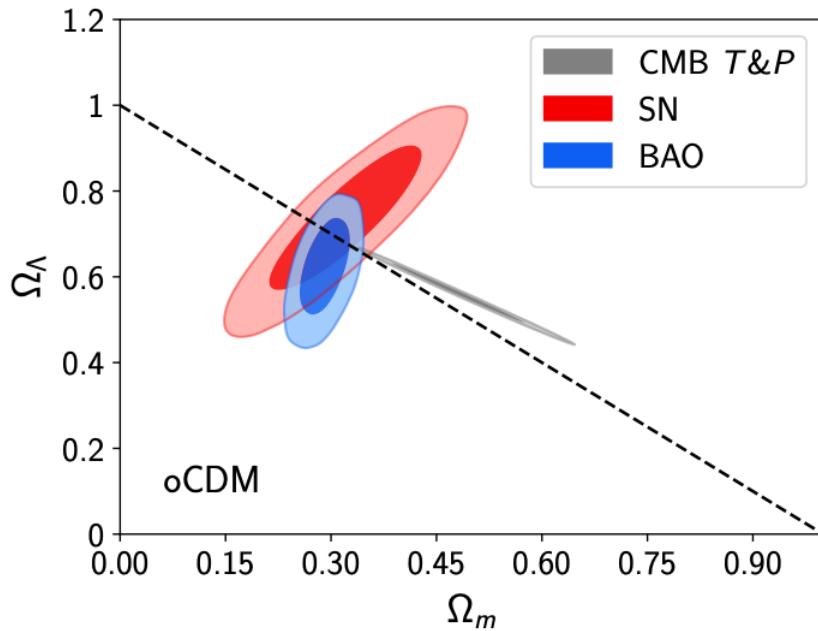


FIGURE 2.10 – Contraintes cosmologiques sur les paramètres  $\Omega_\Lambda$  et  $\Omega_m$ , avec l’hypothèse d’un modèle  $\Lambda$ CDM de courbure libre. Les contours rouges indiquent les contraintes à 1 et 2 sigmas produites à partir des données du relevé de supernovae Pantheon (SCOLNICK et al. 2017), les gris indiquent les contraintes produites par Planck (température et polarisation), et les bleus donnent les contraintes produites par la mesure de l’échelle BAO avec l’ensemble des relevés SDSS (COLLABORATION et al. 2020).

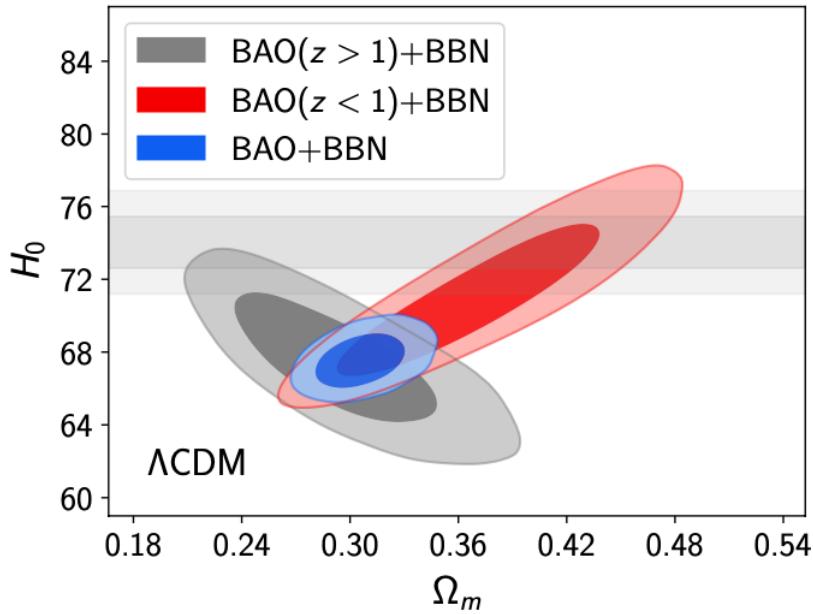


FIGURE 2.11 – Contraintes cosmologiques sur les paramètres  $H_0$  et  $\Omega_m$  produites par la mesure de l'échelle BAO avec l'ensemble des relevés SDSS (COLLABORATION et al. 2020), en supposant un modèle  $\Lambda$ CDM plat. Les contours rouges et gris indiquent les contraintes à 1 et 2 sigmas produites par les mesures à  $z > 1$  et  $z < 1$  respectivement. Les contours bleus indiquent la combinaison de tous les mesures. Enfin, la zone horizontale grise montre la mesure de  $H_0$  faite par RIESS et al. (2019).

la qualité des données sont testées ont été interrompues. Actuellement, le début de la prise de données est estimé à janvier 2021.

Comme eBOSS, DESI étudie les BAO et la croissance des structures à l'aide d'un très grand relevé de galaxies et de quasars. À l'issue des 5 ans d'observation prévus, DESI aura mesuré plus de 30 millions de spectres, distribués sur un relevé de plus de  $14000 \text{ deg}^2$ . Pour atteindre ses objectifs, DESI utilise le télescope Mayall, mesurant 4m de diamètre et situé au Kitt Peak en Arizona. Le champ de vue du télescope est le même que celui de SDSS :  $3 \text{ deg}$  de diamètre sur le ciel. L'instrument inclut un système de fibres optiques, au nombre de 5000, mais celles-ci sont placées au plan focal à l'aide de robots qui ajustent la position de chaque fibre avant chaque exposition. Dix spectrographes reçoivent ces fibres, chacun comportant 3 caméras et couvrant les longueurs d'onde de 3600 à 9800 Å. DESI observe les 4 mêmes traceurs qu'eBOSS : les LRG jusqu'à  $z = 1,0$ , les ELG jusqu'à  $z = 1,7$ , ainsi que les quasars en tant que traceurs directs de la matière et les quasars Ly $\alpha$  sur la gamme  $2,1 < z < 3,5$ . La densité visée de quasars Ly $\alpha$  est de 50 par  $\text{deg}^2$ . En plus de ces traceurs, DESI observera des galaxies brillantes (BG : *Bright Galaxies*) pendant le grey time<sup>1</sup>. Le relevé de ces galaxies contiendra 10 millions d'objets, avec un redshift moyen  $z = 0,2$ .

A la fin des 5 ans d'observations, DESI fournira plus de 30 points de mesure des distances cosmologiques, chacun avec une précision meilleure que le pourcent, et couvrant la gamme  $0 < z < 3,5$ . La figure 2.12 illustre la différence entre BOSS et DESI pour la mesure du taux d'expansion  $H(z)$ . De plus, DESI donnera une mesure de la somme des masses des neutrinos, avec une incertitude de

1. Le grey time, par opposition au dark time, correspond au moment où le ciel n'est pas totalement sombre, lorsqu'il est éclairé par la lune par exemple.

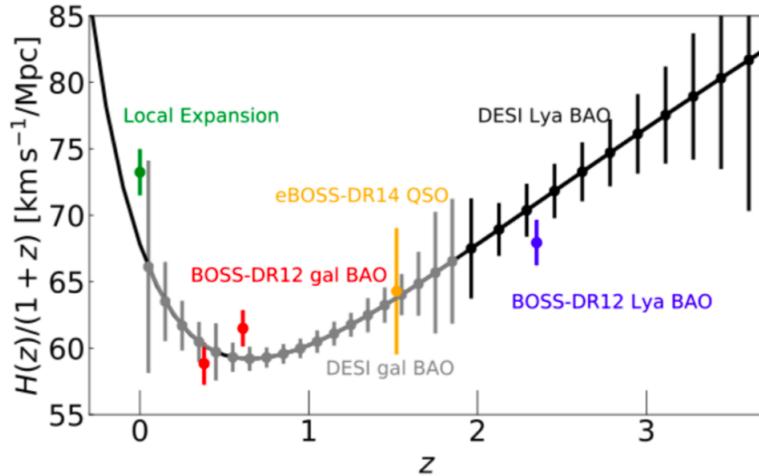


FIGURE 2.12 – Mesure de la distance de Hubble en fonction du redshift. Les points vert (mesure local à l'aide des SN1a), rouge, jaune et bleu donnent les mesures existantes. Les points gris donnent la prédiction pour les galaxies et quasars de DESI, les points noirs donnent la prédiction pour le Ly $\alpha$  de DESI.

0,020 eV. Cette précision est suffisante pour exclure la hiérarchie de masse inversée à  $3\sigma$ .

# 3

## Présentation des données

---

### Sommaire

---

<b>3.1</b>	<b>Réduction des données</b>	60
<b>3.2</b>	<b>Le catalogue de quasar</b>	60
<b>3.3</b>	<b>La sélection des forêts</b>	62
<b>3.4</b>	<b>Définition du champ d'absorption</b>	64
3.4.1	Calcul des $\delta$	64
3.4.2	Prise en compte de l'ajustement du continuum	65

---

Dans le chapitre précédent, nous présentions l'instrument SDSS et le relevé eBOSS. Nous présentons ici comment les spectres sont déduits des photons acquis par les CCD, puis comment le spectre d'absorption est reconstruit à partir de chaque spectre.

### 3.1 Réduction des données

L'observation de chaque plaque fournit des données en deux dimensions : la première correspond aux différentes longueurs d'onde, la seconde aux différentes fibres (figure 3.1). La chaîne de réduction des données transforme ces informations en une liste de spectres. Premièrement, chaque spectre est extrait de l'image acquise par le CCD. Le spectre est étalonné en longueur d'onde à l'aide de lampes à arc. Le flux est calibré en utilisant les spectres des fibres dédiées à des étoiles standards. Puis le fond du ciel, estimé dans chaque demi-plaque grâce aux fibres dédiées, est soustrait à chaque spectre. La variance dans chaque pixel est ensuite estimée. Elle prend en compte le bruit de photon et le bruit de l'électronique. Les pixels affectés par des rayons cosmiques sont rejetés. Enfin, pour chaque objet, toutes les expositions sont ajoutées pour former un seul et même spectre, avec un plus grand rapport signal sur bruit. Le nombre typique d'expositions par objet varie entre 4 et 6.

Les données que nous utilisons dans ce manuscrit ont été rendues publiques lors de la sixième publication de données SDSS (DR16<sup>1</sup>). Elles sont décrites par AHUMADA et al. (2019).

### 3.2 Le catalogue de quasar

Une fois les spectres extraits, il est important de les classifier afin de construire des catalogues utilisables par les différentes analyses. Pour ce faire, chaque spectre est traité par le pipeline de SDSS, décrit par BOLTON et al. (2012). Plusieurs modèles d'étoiles, de galaxies et de quasars sont ajustés à chaque spectre. Chaque modèle est construit à l'aide d'une analyse en composantes principales (PCA),

---

1. <https://www.sdss.org/dr16/>

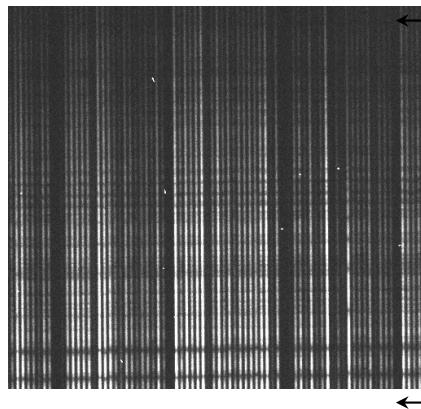


FIGURE 3.1 – Photographie d'un capteur CCD. L'axe des abscisses indique le numéro de la fibre optique. L'axe des ordonnées donne la longueur d'onde observée.

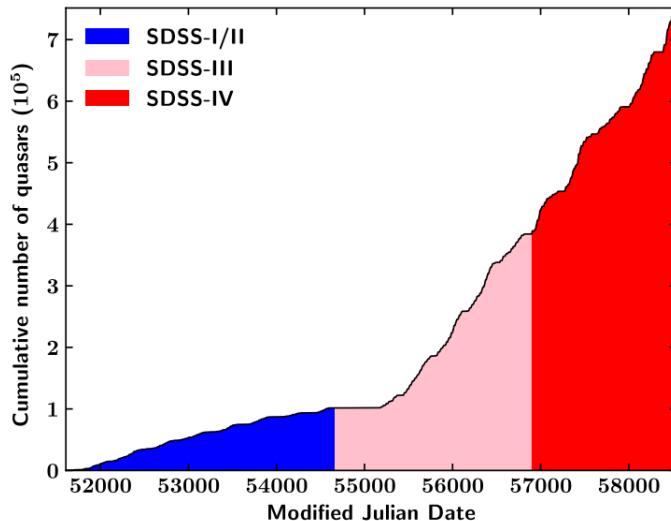


FIGURE 3.2 – Évolution du nombre de quasars observés par les différentes générations de SDSS en fonction du temps. Les générations SDSS I et II correspondent à  $MJD < 54663$  (bleu), SDSS III à  $54663 \leq MJD < 56898$  (rose) et SDSS IV à  $56898 < MJD < 58543$  (rouge). Crédits : LYKE et al. (2020)

propre à chaque type d’objet. Pour chaque type d’objet, l’analyse fournit des spectres typiques. Ces derniers sont de complexités croissantes : le premier donne le plus d’information comme par exemple le spectre moyen, les spectres plus complexes donnent des détails comme certaines raies d’émission ou profil d’absorption. Ces spectres typiques sont pondérés et combinés afin de produire un spectre simulé qui ajuste aux mieux le spectre à classifier. Une fois tous les modèles ajustés, ils sont triés selon leur  $\chi^2$  réduit. Le spectre est alors classifié selon le modèle possédant le plus faible  $\chi^2$  réduit. Tous les spectres mesurés par SDSS et classifiés comme quasar constituent l’échantillon *superset*. Il contient 1 440 627 objets.

Certains quasars peuvent être difficiles à classifier à cause d’absorptions intenses comme les DLA ou les BAL<sup>1</sup>. Tous les quasars du relevé BOSS ont été inspectés visuellement (PÂRIS et al. 2016) afin de confirmer la classification du pipeline de SDSS, et d’estimer leur redshifts. Cet échantillon représente 297 301 objets. Cependant, à cause du grand nombre de quasars observés par eBOSS (figure 3.2), l’observation visuelle n’a pas pu être effectuée pour tous ces objets. Afin de vérifier la première identification faite par le pipeline de SDSS, une seconde classification est alors effectuée (LYKE et al. 2020). Elle est complétée par l’algorithme QuasarNET (BUSCA et BALLAND 2018). Ceci réduit le nombre d’identifications douteuses et donc le nombre d’inspections visuelles requises. A la fin, 0,6 % des spectres, soit 8581 spectres, requièrent une inspection visuelle.

Une fois les spectres classifiés, un sous-échantillon du superset est construit. Il contient 750 426 objets, confirmés comme quasar par la chaîne de traitement précédente. Pour chaque objet, le catalogue fournit plusieurs estimations de redshift. Le pipeline de SDSS produit une première estimation. L’algorithme QuasarNET en fournit une seconde. Les spectres inspectés visuellement possèdent une autre

1. Les BAL (Broad Absorption Line) sont des quasars qui présentent des absorptions intenses au voisinage de leurs raies d’émission. Cette absorption est interprétée comme étant due à un absorbeur dense situé juste devant le quasar. La figure 3.3 présente un spectre de quasar avec un BAL.

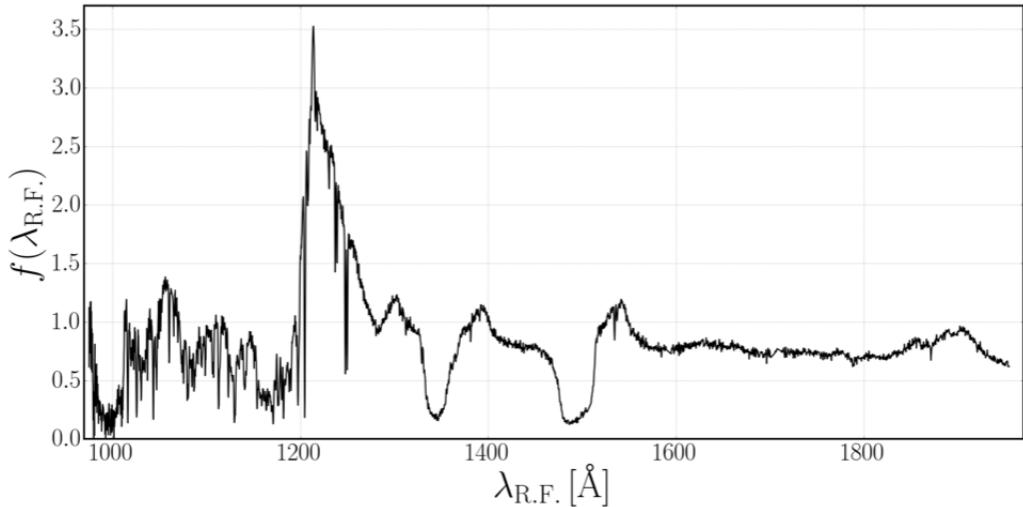


FIGURE 3.3 – Spectre d'un quasar pris par SDSS présentant un BAL. Le spectre est très fortement absorbé pour les longueurs d'onde légèrement inférieures à chaque raie d'émission.

estimation. Enfin, l'algorithme `redvsblue`<sup>1</sup> produit plusieurs estimations de redshift, parmi lesquelles figurent `Z_PCA` et `Z_LYAWG`.

Différents algorithmes sont alors appliqués au catalogue, afin d'identifier les DLA et les BAL présents. L'identification utilise l'addition des différentes expositions pour chaque objet. Concernant les DLA, l'algorithme de détection est décrit dans PARKS et al. (2017). Il est appliqué sur les quasars à un redshift  $2 \leq Z_{\text{PCA}} \leq 6$ , afin d'avoir suffisamment de pixels dans la zone  $900 < \lambda_{\text{RF}} < 1346 \text{ \AA}$ . Parmi les 270 315 spectres inspectés, 39 514 DLA ont été identifiés, distribués dans 35 686 spectres. Concernant les BAL, l'algorithme utilisé est très similaire à celui décrit dans GUO et MARTINI (2019). Les BAL sont recherchés dans les spectres ayant un redshift entre 1,57 et 5,6. L'algorithme fournit la probabilité qu'un spectre possède un BAL. Le champ `BAL_PROB` du catalogue indique cette probabilité.

A la fin, le catalogue ainsi construit (DR16Q dans la suite de ce manuscrit) contient 750 426 quasars confirmés. Nous référons le lecteur à l'article LYKE et al. (2020) pour davantage d'informations.

L'analyse Ly $\alpha$  des données complètes d'eBOSS (DU MAS DES BOURBOUX et al. 2020) utilise le catalogue DR16Q. Le redshift des quasars est choisi comme étant `Z_LYAWG`. Les quasars sont sélectionnés avec un redshift  $1,77 < z \leq 4$ . L'échantillon correspondant représente alors 341 468 objets. La distribution en redshift de ces quasars traceurs (ceux utilisés comme traceurs pour la fonction de corrélation croisée Ly $\alpha$ -QSO) est présentée dans le graphique de gauche de la figure 3.4.

### 3.3 La sélection des forêts

L'analyse Ly $\alpha$  des données complètes d'eBOSS, dont nous nous servons dans ce manuscrit, est présentée dans l'article DU MAS DES BOURBOUX et al. (2020). Les informations que nous donnons dans la suite de ce chapitre sont tirés de cette article. Nous y référons le lecteur pour davantage d'informations.

1. <https://github.com/londumas/redvsblue>

### 3.3. LA SÉLECTION DES FORÊTS

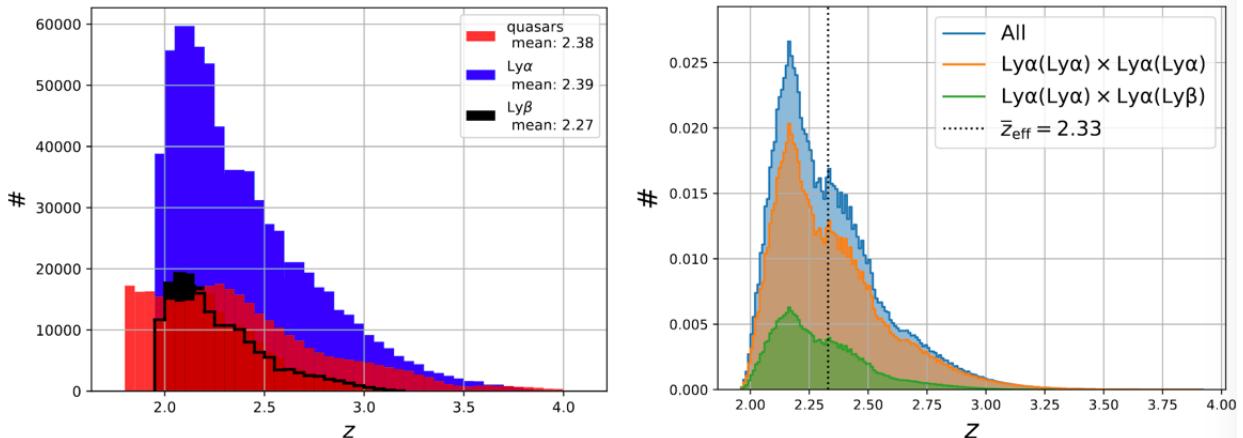


FIGURE 3.4 – Gauche : distribution en redshift des pixels Ly $\alpha$ (Ly $\alpha$ ) (bleu) et des pixels Ly $\alpha$ (Ly $\beta$ ) (noir) (nombres divisés par 50), ainsi que celle des quasars traceurs (rouge). Droite : distribution pondérée du redshift efficace de chaque paire de pixels utilisée pour le calcul de la fonction corrélation du Ly $\alpha$ . La distribution orange présente les paires de pixels Ly $\alpha$ (Ly $\alpha$ )-Ly $\alpha$ (Ly $\alpha$ ), la verte les paires de pixels Ly $\alpha$ (Ly $\alpha$ )-Ly $\alpha$ (Ly $\beta$ ), et la bleue la somme des deux. Les paires représentées sont celles nécessaires pour la mesure de la position du pic BAO, c'est à dire avec une séparation  $r \in [80; 120] h^{-1}$  Mpc. La ligne en pointillés indique le redshift effectif de la mesure, c'est à dire la moyenne de la distribution bleue. Il vaut  $z_{\text{eff}} = 2,334$ .

Les spectres produits par la chaîne de réduction des données SDSS sont *rebinnés* : 3 pixels du spectre original, d'une taille  $\Delta \log_{10}(\lambda) \sim 10^{-4}$ , sont combinés en 1 seul pixel d'analyse, d'une taille  $\Delta \log_{10}(\lambda) \sim 3 \times 10^{-4}$ . Ceci est fait afin de réduire le temps de calcul nécessaire pour estimer les fonctions de corrélation. Dans la suite, l'utilisation de « pixel » réfère à ces pixels d'analyse.

L'absorption Ly $\alpha$  est mesurée dans deux régions distinctes du spectre. La première, dénommée région Ly $\alpha$ , correspond aux longueurs d'onde  $1040 \leq \lambda_{\text{RF}} \leq 1200 \text{ \AA}$ , c'est à dire entre les raies d'émission Ly $\beta$  et Ly $\alpha$ . La seconde, dénommée région Ly $\beta$ , correspond aux longueurs d'onde  $920 \leq \lambda_{\text{RF}} \leq 1020 \text{ \AA}$ , c'est à dire entre la limite de la série de Lyman et la raie d'émission Ly $\beta$ . Les pixels d'absorption Ly $\alpha$  dans la région Ly $\alpha$  sont dénommés pixels Ly $\alpha$ (Ly $\alpha$ ), et ceux dans la région Ly $\beta$  sont dénommés pixels Ly $\alpha$ (Ly $\beta$ ). De plus, l'analyse se limite aux pixels dont la longueur d'onde observée est comprise entre  $3600 \leq \lambda_{\text{obs}} \leq 6000 \text{ \AA}$ . La limite inférieure provient de la sensibilité des CCD et de l'augmentation des absorptions atmosphériques intenses dans l'UV. La limite supérieure provient des raies d'émission du ciel dans le proche infrarouge qui bruent le signal. Ces limites sur  $\lambda_{\text{obs}}$  correspondent à un redshift minimal  $z_{\text{QSO}} = 2$  pour les quasars utilisés pour leurs pixels dans la région Ly $\alpha$  (quasars Ly $\alpha$ ), et  $z_{\text{QSO}} = 2,53$  pour les quasars utilisés pour leurs pixels dans la région Ly $\beta$  (quasars Ly $\beta$ ). Parmi les 341 468 quasars traceurs, 256 328 sont des quasars Ly $\alpha$ , et 103 080 sont des quasars Ly $\beta$ .

D'autres sélections sont aussi appliquées. Les quasars pour lesquels la probabilité d'avoir un BAL est supérieure à 0.9 sont écartés. Les mauvaises observations durant BOSS ou eBOSS sont mises de côté. De plus, chaque région nécessite au moins 50 pixels. Cette sélection décale le redshift minimal à  $z_{\text{QSO}} = 2,10$  pour les quasars Ly $\alpha$ , et à  $z = 2,65$  pour les quasars Ly $\beta$ . Le pixel d'absorption Ly $\alpha$  de plus bas redshift se trouve à  $z = 1,96$ . Enfin, l'ajustement du continuum (voir section suivante) échoue pour environ 2 % des spectres). Ces spectres sont aussi écartés. Après toutes ces sélections,

l'échantillon final contient 210 005 quasars Ly $\alpha$  et 69 656 quasars Ly $\beta$ .

Enfin, des corrections sur le flux et la variance du flux sont appliquées afin de corriger les imperfections de la chaîne de réduction. De plus, certaines régions spectrales en longueur d'onde observée sont masquées, à cause de l'augmentation de la variance du flux causée par les raies du ciel. Tout ceci est décrit en détail dans DU MAS DES BOURBOUX et al. (2020).

## 3.4 Définition du champ d'absorption

Dans cette section, nous décrivons comment, du flux mesuré  $f_q(\lambda)$  du quasar  $q$  à la longueur d'onde observée  $\lambda$ , nous pouvons déduire les fluctuations du champ d'absorption Ly $\alpha$ . La démarche que nous décrivons ici est très bien expliquée dans la thèse DE SAINTE AGATHE (p.d.). Nous y référerons le lecteur pour plus de détails.

### 3.4.1 Calcul des $\delta$

Comme présenté dans la section 1.5.4, le contraste de l'absorption Ly $\alpha$   $\delta_q(\lambda)$  est défini comme

$$\delta_q(\lambda) = \frac{f_q(\lambda)}{\bar{F}(\lambda)C_q(\lambda)} - 1, \quad (3.1)$$

où  $\bar{F}(\lambda)$  est la transmission moyenne au redshift  $z = \lambda/\lambda_{\text{Ly}\alpha} - 1$ , et  $C_q(\lambda)$  donne le flux du spectre sans absorption. Il est dénommé *continuum* et est différent pour chaque quasar. Le produit  $\bar{F}(\lambda)C_q(\lambda)$  représente donc le flux moyen attendu du quasar  $q$ . La figure 1.19 représente ces quantités pour les régions Ly $\alpha$  et Ly $\beta$ . Du fait de la faible résolution et du faible rapport signal sur bruit des données,  $\bar{F}(\lambda)$  n'est pas mesurable. Pour ne pas dépendre d'analyses externes, on détermine directement le produit  $\bar{F}(\lambda)C_q(\lambda)$  pour chaque spectre. Afin de prendre en compte la variabilité d'un spectre à un autre, le produit  $\bar{F}(\lambda)C_q(\lambda)$  est modélisé par une relation linéaire en  $\log_{10}\lambda$  :

$$\bar{F}(\lambda)C_q(\lambda) = \bar{C}(\lambda_{\text{RF}})(a_q + b_q \log_{10}(\lambda)), \quad (3.2)$$

où  $\bar{C}(\lambda_{\text{RF}})$  donne le continuum moyenné sur tous les spectres, en fonction de la longueur d'onde dans le référentiel du quasar.  $a_q$  et  $b_q$  sont propres à chaque quasar, ils modélisent la diversité des quasars. Les termes  $a_q$ ,  $b_q$  et  $\bar{C}(\lambda_{\text{RF}})$  sont déterminés en minimisant la fonction de vraisemblance  $L$ , qui s'exprime comme

$$-2 \ln L = \sum_i \frac{[f_i - \bar{F}C_q(\lambda_i, a_q, b_q)]^2}{\sigma_q^2(\lambda_i)} + \ln[\sigma_q^2(\lambda_i)]. \quad (3.3)$$

La variance  $\sigma_q^2(\lambda)$  dans chaque pixel est donnée par

$$\sigma_q^2(\lambda) = \eta(\lambda)\sigma_{\text{instru},q}^2(\lambda) + \sigma_{\text{cosmo}}^2(\lambda)(\bar{F}C_q(\lambda))^2 + \frac{\epsilon(\lambda)(\bar{F}C(\lambda))^4}{\sigma_{\text{instru},q}^2(\lambda)}. \quad (3.4)$$

Le terme  $\eta(\lambda)\sigma_{\text{instru},q}^2(\lambda)$  rend compte de la variance qui est provoquée par l'instrument et liée à l'incertitude sur la mesure des flux. Le terme  $\sigma_{\text{cosmo}}^2(\lambda)$  est purement cosmologique et donne la variance

propre au champ d'absorption du Ly $\alpha$ . Il traduit le fait que certaines régions de l'univers peuvent absorber plus ou moins que la moyenne. Du fait de l'isotropie,  $\sigma_{\text{cosmo}}$  ne dépend que de  $z$ . Le terme ad hoc  $\epsilon(\lambda)(\bar{F}C(\lambda))^4/\sigma_{\text{instru},q}^2(\lambda)$  rend compte du fait que la variance augmente avec le rapport signal sur bruit, probablement à cause de la diversité des quasars. Enfin, le terme  $\ln[\sigma_q^2]$  dans l'expression de la vraisemblance provient de la normalisation  $\sqrt{2\pi\sigma^2}$  au dénominateur des distributions gaussiennes. Ce terme est généralement omis car constant. Mais dans notre cas,  $\sigma_q$  dépend de  $a_q$ ,  $b_q$  et  $\bar{C}(\lambda_{\text{RF}})$ .

Afin d'obtenir le produit  $\bar{F}C_q$  pour chaque quasar, il faut maximiser la fonction de vraisemblance  $L$  (équation 3.3), et donc ajuster les paramètres  $a_q$ ,  $b_q$ ,  $\eta(\lambda)$ ,  $\sigma_{\text{cosmo}}(\lambda)$  et  $\epsilon(\lambda)$ . Cependant, celle fonction dépend elle même du produit  $\bar{F}C_q$ . Pour résoudre ce problème, l'ajustement est fait de manière itérative. Premièrement, les 5 paramètres à ajuster, ainsi que  $\bar{C}(\lambda_{\text{RF}})$ , sont initialisés. Ceci permet d'ajuster  $\bar{F}C_q$  sur  $f_q(\lambda)$  en maximisant la fonction de vraisemblance  $L$ . Dans cet ajustement, chaque pixel observé à une longueur d'onde  $\lambda$  est pondéré par un poids

$$w_q(\lambda) = \frac{1}{\sigma_q^2(\lambda)} . \quad (3.5)$$

Une fois cette ajustement fait,  $\bar{C}(\lambda_{\text{RF}})$  est de nouveau calculé. Puis, les paramètres  $\eta(\lambda)$ ,  $\sigma_{\text{cosmo}}(\lambda)$  et  $\epsilon(\lambda)$  sont ajustés de manière à ce que, à lambda fixé,  $\sigma_q(\lambda)$  reproduise la variance estimée à partir des flux  $f_q(\lambda)$  de tous les quasars. Ce nouvel ajustement permet une nouvelle fois d'ajuster  $\bar{F}C_q$ , etc. Cette procédure est répétée jusqu'à obtenir des valeurs stables pour les paramètres ajustés. En pratique, 5 itérations sont nécessaires. Cette procédure permet donc d'estimer, pour chaque quasar, le produit  $\bar{F}(\lambda)C_q(\lambda)$ , et ainsi d'estimer les  $\delta_q(\lambda)$  (équation 3.1) indépendamment dans chaque région.

Lors du calcul des  $\delta_q$ , les DLA identifiés dans chaque forêt sont masqués. Pour ce faire, un profil de Voigt, dépendant de la densité de colonne mesurée par l'algorithme de détection, est ajusté à l'endroit du spectre où le DLA a été identifié. Les pixels pour lesquels l'absorption est plus grande que 20 % ne sont pas utilisés. Les autres sont corrigés en utilisant le profil de Voigt ajusté. La figure 1.20 montre deux spectres présentant un DLA. Le profil de Voigt est montré en dessous de la forêt, en rouge. Les bandes rouges indiquent les pixels non utilisés.

### 3.4.2 Prise en compte de l'ajustement du continuum

La procédure décrite dans la section précédente permet d'ajuster le continuum de chaque quasar afin d'estimer le champ  $\delta$  dans chaque région d'absorption Ly $\alpha$ . Cependant, cet ajustement biaise le champ  $\delta$  mesuré, et introduit de fausses corrélations lors du calcul des fonctions de corrélation. Il est donc important de prendre en compte cet effet.

En redéfinissant les paramètres  $a_q$  et  $b_q$ , l'équation 3.2 peut être mise sous la forme

$$\bar{F}(\lambda)C_q(\lambda) = \bar{C}(\lambda_{\text{RF}})(a_q + b_q(\Lambda - \bar{\Lambda}_q)) , \quad (3.6)$$

où  $\Lambda = \log_{10} \lambda$ , et  $\bar{\Lambda}_q$  est la moyenne de  $\Lambda$  dans chaque forêt :

$$\bar{\Lambda}_q = \frac{\sum_{\lambda} w_q(\lambda) \Lambda}{\sum_{\lambda} w_q(\lambda)} . \quad (3.7)$$

Le champ  $\delta_q$  mesuré est donc donné par

$$\delta_q(\lambda) = \frac{f_q(\lambda)}{\overline{C}(\lambda_{\text{RF}})(a_q + b_q(\Lambda - \bar{\Lambda}_q))} - 1. \quad (3.8)$$

Du fait que  $a_q$  et  $b_q$  sont ajustés sur les  $f_q(\lambda)$ , l'utilisation de l'équation 3.6 pour déterminer le continuum force la moyenne et la pente de chaque région spectrale à être nulle. Ainsi le champ  $\delta$  mesuré est biaisé, et de fausses corrélations sont induites dans le calcul de la fonction de corrélation. Afin de prendre en compte cet effet, le champ  $\delta$  est transformé selon la relation

$$\tilde{\delta}_q(\lambda_i) = \sum_j \eta_{ij}^q \delta_q(\lambda_j), \quad (3.9)$$

avec

$$\eta_{ij}^q = \delta_{ij}^K - \frac{w_j}{\sum_k w_k} - \frac{w_j (\Lambda_i - \bar{\Lambda}_q) (\Lambda_j - \bar{\Lambda}_q)}{\sum_k w_k (\Lambda_k - \bar{\Lambda}_q)^2}, \quad (3.10)$$

où  $\delta_{ij}^K$  est le symbole de Kronecker. Dans ce calcul, les poids  $w_q(\lambda)$  sont corrigés de la dépendance en redshift du biais du Ly $\alpha$  et du facteur de croissance :

$$w_i = \sigma_q^{-2}(\lambda_i) \left( \frac{1+z_i}{1+2.25} \right)^{\gamma_{\text{Ly}\alpha}-1}, \quad (3.11)$$

avec  $\gamma_{\text{Ly}\alpha} = 2.9$  (McDONALD et al. 2004). Grâce à cette transformation, l'effet sur la fonction de corrélation peut ainsi être modélisé et pris en compte. Sa modélisation passe par le calcul de la *matrice de distorsion*, détaillée dans la section 5.2.

Enfin, l'utilisation de  $\overline{C}(\lambda_{\text{RF}})$  dans l'équation 3.2 force la moyenne des  $\delta_q$  dans chaque bin de longueur d'onde observée à zéro. Cependant, la transformation 3.9 modifie légèrement la moyenne des  $\delta_q$  dans chaque bin. Afin que la fonction de corrélation croisée Ly $\alpha$ -QSO tende vers zéro à grand séparation, l'effet est corrigé en soustrayant explicitement la moyenne dans chaque bin :

$$\hat{\delta}_q(\lambda_i) = \tilde{\delta}_q(\lambda_i) - \bar{\tilde{\delta}}(\lambda). \quad (3.12)$$

Dans la suite de ce manuscrit, lorsque nous parlons du champ  $\delta$ , nous référons à ces  $\hat{\delta}_q(\lambda_i)$ . Le graphique de gauche de la figure 3.4 montre la distribution en redshift des  $\delta$  dans chaque région d'absorption. Le graphique de droite présente la distribution pondérée par les poids  $w_i$  (équation 3.11) du redshift effectif de chaque paire, utilisé dans le calcul de la fonction de corrélation.

# 4

## Développement des mocks

---

### Sommaire

---

<b>4.1</b>	<b>Objectifs des mocks</b>	<b>68</b>
<b>4.2</b>	<b>Construction des mocks</b>	<b>69</b>
4.2.1	Les champs de densité	69
	Les quasars	70
	Le champ Ly $\alpha$	71
	Les champs des vitesses	72
4.2.2	Le relevé de quasars	73
4.2.3	Création des lignes de visée	77
4.2.4	De la densité à l'absorption	79
	Les petites échelles	79
	Les RSD	83
	La prédiction	86
4.2.5	Ajout des HCD	88
<b>4.3</b>	<b>Production des mocks</b>	<b>90</b>
<b>4.4</b>	<b>Ajustement des paramètres</b>	<b>95</b>
<b>4.5</b>	<b>Expansion des mocks</b>	<b>100</b>

---

Dans ce chapitre, nous présentons la construction des *mocks* : des spectres de quasars simulés, dont les forêts Ly $\alpha$  sont corrélées entre elles et avec le champ de quasars sous-jacent. Ces mocks visent à reproduire les données d'eBOSS et de DESI. Ils sont nommés **SaclayMocks** et présentés dans ETOURNEAU et al. (2020). Le code est écrit en Python<sup>1</sup> et se trouve en accès libre sur GitHub<sup>2</sup>. L'utilisation de ces mocks et leur validation seront présentés dans les chapitres suivants.

## 4.1 Objectifs des mocks

Contrairement à ce qu'on appelle les simulations, les mocks ne contiennent pas de physique à proprement parler : ils ne sont pas utilisés afin de déduire des paramètres astrophysiques. Certaines simulations, les simulations hydrodynamiques, permettent de mesurer des effets astrophysiques, comme par exemple le biais de l'hydrogène ou du Ly $\alpha$ . Mais ces simulations sont très coûteuses car elles nécessitent de modéliser les effets physiques qui affectent les paramètres mesurés. Les mocks, quant à eux, sont conçus afin de répliquer rapidement un jeu de données, dans le but de tester l'analyse qui sera appliquée sur ces données. Les mocks sont donc utilisés afin

- de vérifier la mesure des paramètres  $\alpha_{\parallel}$ ,  $\alpha_{\perp}$  : cette mesure est-elle non biaisée ?
- d'identifier les potentielles systématiques : la présence de métaux et d'HCD dans les données est-elle bien modélisée ? Affecte-t-elle la mesure de  $\alpha_{\parallel}$ ,  $\alpha_{\perp}$  ? Des autres paramètres ?
- de tester la matrice de distorsion : la distorsion de la fonction de corrélation due à l'ajustement du continuum du quasar est-elle correctement prise en compte par la matrice de distorsion ?
- de vérifier l'estimation de la matrice de covariance : la matrice de covariance, calculée à partir des données, est-elle bien estimée ?

Il est important de produire un grand nombre de réalisations des mocks afin de disposer de suffisamment de statistique pour répondre précisément à ces questions. La production et l'analyse de ces différentes réalisations des mocks sont donc nécessaires pour pouvoir valider les analyses menées sur les données.

Les mocks décrits dans ce manuscrit s'inscrivent dans les projets eBOSS et DESI. Ils sont utilisés dans l'analyse Ly $\alpha$  des données complète d'eBOSS (DU MAS DES BOURBOUX et al. 2020), et seront utilisés dans l'analyse Ly $\alpha$  de DESI. L'objectif de ces mocks est donc de répliquer au mieux les données Ly $\alpha$  d'eBOSS et de DESI. Ces relevés couvrent un volume de plusieurs dizaines de Gpc<sup>3</sup>, et les échelles sondées grâce au Ly $\alpha$  descendant jusqu'à la centaine de kpc. Les mocks nécessitent donc de reproduire un volume immense, avec une bonne résolution. Les simulations dites *N-corps* sont des simulations qui traitent le problème à N corps. Elles sont initialisées à un redshift élevé ( $z > 100$ ), avec une distribution de matière noire représentée par des macro-particules. Actuellement, les simulations les plus performantes utilisent environ  $(10000)^3$  macro-particules. La masse de ces particules dépend du volume simulé. Par exemple, la simulation Outer Rim (HEITMANN et al. 2019) simule un volume d'environ  $(4 \text{Gpc})^3$ , et les particules possèdent une masse  $\sim 10^9 M_{\odot}$ . Puis, à chaque pas de temps,

---

1. <https://www.python.org/>

2. <https://github.com/igmhub/SaclayMocks>

ces macro-particules sont déplacées en considérant uniquement les interactions gravitationnelles. Le champ de matière initial évolue ainsi jusqu'à  $z = 0$ . Ces simulations sont très utiles pour étudier les effets de la gravité à grande échelle. Cependant elles ne sont pas adaptées à notre utilisation : afin d'avoir la résolution et le volume requis, la simulation nécessiterait beaucoup trop de macro-particules pour être réalisable dans un temps raisonnable.

Les simulations hydrodynamiques fonctionnent de la même manière que les simulations N-corps. Elles incluent, en plus des macro-particules de matière noire, la physique baryonique présente dans le milieu galactique. Les baryons sont aussi représentés par des macro-particules. Afin de résoudre l'intérieur des galaxies, les macro-particules utilisées possèdent une masse plus faible que dans le cas des simulations N-corps. En contrepartie, le volume simulé est plus petit. Dans le cas des simulations hydrodynamiques, la densité, la pression et la température sont tracées dans chaque cellule. Certains effets astrophysiques, comme les supernovae ou les AGN peuvent aussi être présents. Ces simulations nécessitent encore plus de temps de calcul que les simulations à N-corps. Elles ne sont donc pas non plus adaptées à notre utilisation.

Ainsi, seuls les *champs aléatoires gaussiens* (GRF pour Gaussian Random Field) permettent de générer un grand volume avec une bonne résolution dans un temps raisonnable, afin de produire un grand nombre de mocks. Les champs aléatoires gaussiens sont des champs qui en chaque point prennent une valeur aléatoire selon une statistique gaussienne. Une fois générés, il est possible de donner à ces champs n'importe quelle fonction de corrélation, en utilisant la transformation de Fourier. Ces champs sont très rapides à générer. Ils nous permettent donc de produire un grand nombre de simulations, afin de disposer d'une puissance statistique suffisante. Ces champs sont utilisés notamment dans les simulations à N-corps, afin de fournir les conditions initiales. Cependant, l'utilisation des GRF ne donne pas accès aux non-linéarités qui peuvent émerger dans l'évolution des simulations N-corps et hydrodynamiques. La seule information provient de la fonction de corrélation que l'on applique au champ aléatoire initial. Mais cela est entièrement suffisant pour l'utilisation que nous en avons dans ce manuscrit : nous générerons un champ gaussien destiné à simuler le champ d'absorption Ly $\alpha$  et le champ de quasars, dont les fonctions d'auto corrélation et de corrélation croisée sont choisies afin de correspondre à ce qui est observé dans les données.

## 4.2 Construction des mocks

Dans cette section, nous détaillons comment les mocks sont générés. Nous présentons d'abord la génération des champs de densité, puis comment les quasars sont tirés à partir de ces champs de densité. Nous expliquons ensuite comment nous calculons la densité le long de la ligne de visée de chaque quasar. Enfin, nous présentons comment cette densité est transformée en fraction de flux transmis, et comment nous tirons les HCD.

### 4.2.1 Les champs de densité

La première étape dans la création des mocks est de générer les boîtes qui contiennent le champ de densité  $\delta$ . D'abord, un GRF est généré dans une boîte de  $2560 \times 2560 \times 1536$  voxels, chaque voxel

faisant  $d_{\text{cell}}^3 = (2,19 h^{-1} \text{Mpc})^3$ . Afin que le champ  $\delta$  possède la bonne fonction de corrélation, une transformation de Fourier 3D<sup>1</sup> est appliquée sur la boîte, puis la boîte  $\delta_k$  ainsi obtenue est multipliée par

$$\sqrt{\frac{P(k)}{d_{\text{cell}}^3}}, \quad (4.1)$$

où  $P(k)$  est le spectre de puissance désiré. Il est ensuite possible d'obtenir la boîte  $\delta$  dans l'espace réel grâce à une transformation de Fourier inverse de la boîte  $\delta_k$ . Ce procédé garanti que le champ  $\delta$  suive le spectre de puissance  $P(k)$ . Le GRF pourrait être tiré directement dans l'espace  $k$ . Dans ce cas, il nous faut tirer deux champs gaussiens : un pour la partie réelle, et un autre pour la partie imaginaire. La transformation de Fourier prenant moins de temps que la génération du champ aléatoire, nous préférons générer le champ dans l'espace réel plutôt que dans l'espace  $k$ . Dans la suite nous décrivons les différentes boîtes nécessaires à la construction des mocks : les boîtes champs utilisées pour tirer les quasars, la boîte utilisée pour créer l'absorption Ly $\alpha$ , ainsi que les boîtes de vitesse et de gradient de vitesse. Afin de garantir leur corrélation, toutes ces boîtes sont construites à partir de la même boîte initial  $\delta_k$ .

## Les quasars

Afin de construire un relevé de quasars corrélés, nous tirons les quasars selon le champ dans l'espace réel  $\delta_{\text{QSO}}$ , construit à partir de  $\delta_k$ . Une première solution serait de tirer les quasars dans les voxels dont le champ  $\delta_{\text{QSO}}$  est supérieur à un certain seuil. Cette solution produit une fonction de corrélation correcte aux grandes échelles, mais pas aux petites. En effet, comme montré dans FONT-RIBERA et MIRALDA-ESCUDÉ (2012), les objets tirés aux endroits où la densité est supérieure à un certain seuil suivent la fonction de corrélation de la densité sous-jacente, avec un biais qui dépend du seuil choisi, si et seulement si la fonction de corrélation est petite devant 1. La fonction de corrélation ainsi obtenue est correcte, sauf pour les petites échelles pour lesquelles la fonction de corrélation est importante. Une solution alternative consiste à considérer que les quasars suivent une distribution log-normale. Ceci permet d'obtenir une meilleure corrélation aux petites échelles. Ce choix est souvent fait pour simuler des relevés de galaxies (AGRAWAL et al. 2017), et est en accord avec ce qui est observé dans les données (CLERKIN et al. 2016). Nous optons donc pour cette seconde solution. Ainsi, les quasars sont tirés dans chaque voxel avec une probabilité

$$P \propto e^{\delta_q}, \quad (4.2)$$

où  $\delta_q$  est le champ de densité dans le voxel considéré. Comme montré par COLES et JONES (1991), afin que les quasars suivent la fonction de corrélation  $\xi(r)$ , le champ  $\delta_q$  doit suivre la fonction de corrélation

$$\xi_q(r) = \ln(1 + \xi(r)). \quad (4.3)$$

---

1. Nous utilisons la librairie pyFFTW (<https://github.com/pyFFTW/pyFFTW>), une adaptation python de la librairie FFTW (<http://www.fftw.org/>).

Nous verrons section 4.2.2 que, de manière à obtenir un relevé synthétique de quasars dont le biais dépend de  $z$ , nous utilisons trois boîtes qui suivent des distributions log-normales, à des redshifts différents. La probabilité pour tirer les quasars dépend de l'interpolation entre ces 3 boîtes. Pour chacune des boîtes, nous partons du spectre de puissance de la matière  $P_{\text{matière}}(k)$  à  $z = 0$ , fourni par Camb (LEWIS et al. 1999). Nous multiplions ensuite ce spectre de puissance par  $(b_{\text{QSO}}(z_i)G(z_i))^2$ , où  $i \in [1; 2; 3]$ . A l'aide de la transformation de Fourier, nous calculons la fonction de corrélation  $\xi_i(r)$ . Puis, nous déterminons le spectre de puissance  $P_{\text{QSO},i}(k)$ , à appliquer à la boîte  $\delta_k$ , comme la transformée de Fourier de  $\xi_{\text{QSO},i}(r) = \ln(1 + \xi_i(r))$  (équation 1.45). Une fois les trois spectres de puissances  $P_{\text{QSO},i}(k)$  obtenus, nous construisons 3 boîtes

$$\delta_{k,i}(k) = \delta_k(k) \sqrt{\frac{P_{\text{QSO},i}(k)}{V_{\text{cell}}}}, \quad (4.4)$$

où  $\delta_k$  est le GRF dans l'espace de Fourier. Une fois ces 3 boîtes construites, nous appliquons à chacune d'entre elle une transformation de Fourier inverse afin d'obtenir les boîtes  $\delta_{\text{QSO},i}$ . Ces boîtes seront interpolées en  $z$ , puis les quasars seront ensuite tirés avec une probabilité  $\propto \exp(\delta_{\text{QSO}}(z))$ , où  $\delta_{\text{QSO}}$  est la boîte interpolée. Nous expliquons cette étape dans la section 4.2.2.

### Le champ Ly $\alpha$

Afin de construire le champ d'absorption Ly $\alpha$ , nous avons besoin du champ de densité de l'hydrogène neutre. Comme expliqué dans la section 1.5.4, la fraction de flux transmis  $F$  est reliée à la profondeur optique  $\tau$  par

$$F = \exp(-\tau). \quad (4.5)$$

De plus, l'approximation FGPA (Fluctuating Gunn Peterson Approximation) permet de relier la profondeur optique  $\tau$  au contraste de densité  $\delta$  à  $z = 0$  :

$$\tau(z) = a(z) \exp(b(z)G(z)\delta). \quad (4.6)$$

Elle est décrite dans l'appendice A. Dans cette approximation, l'élargissement des raies d'absorption causé par la température est négligé. Étant donné que l'élargissement thermique est la source dominante d'élargissement des raies d'absorption Ly $\alpha$ , et que cet effet est environ cinq fois inférieur à la résolution des données que nous cherchons à modéliser, l'utilisation de la formule FGPA est approprié.

Dans cette approximation (équation 4.6), les paramètres  $a$  et  $b$  sont des paramètres à ajuster afin d'obtenir le bon biais du Ly $\alpha$  et la bonne transmission moyenne  $\bar{F}$ . Leur détermination est décrite dans la section 4.4. Le facteur de croissance  $G$  prend en compte l'évolution avec le redshift du champ de densité  $\delta$ . Ainsi il nous suffit de construire un GRF qui suit la fonction de corrélation à  $z = 0$  pour simuler le champ d'absorption du Ly $\alpha$ . Pour ce faire, nous partons de la même boîte  $\delta_k$  utilisée pour construire les 3 boîtes log-normales des quasars. Ceci garanti la corrélation croisée entre les quasars et le champ d'absorption Ly $\alpha$ . Le spectre de puissance de la matière  $P_{\text{matière}}(k)$  à  $z = 0$  est ensuite appliqué à la boîte  $\delta_k$ . Enfin, nous obtenons la boîte de densité  $\delta_{\text{matière}}$  à  $z = 0$  qui servira au calcul

du champ d'absorption du Ly $\alpha$  en effectuant la transformée de Fourier de la boîte

$$\delta_{k,\text{matière}}(\vec{k}) = \delta_k(\vec{k}) \sqrt{\frac{P_{\text{matière}}(\vec{k})}{V_{\text{cell}}}}. \quad (4.7)$$

### Les champs des vitesses

Afin d'inclure les RSD dans nos mocks, nous simulons aussi le champ des vitesses. A l'ordre linéaire, le champ des vitesses  $v_{k,n}$  dans l'espace  $k$  selon la direction  $\vec{u}_n$ , avec  $n \in [\text{x}; \text{y}; \text{z}]$ , est relié au champ de densité  $\delta_k$  par la relation (voir équation 9.18 et 9.20 de DODELSON (2003))

$$v_{k,n}(\vec{k}) = \frac{ik_n}{k^2} aHf \delta_k(\vec{k}). \quad (4.8)$$

En théorie linéaire, les traceurs sont considérés avoir la même vitesse que le champ de matière sous-jacent. Autrement dit, le champ de vitesse des traceurs est non biaisé. Nous adoptons ici cette hypothèse. En ce qui concerne les quasars, nous simulons les RSD en déplaçant chaque quasar proportionnellement à sa vitesse le long de la ligne de visée (équation 1.60). Dans ce but, nous calculons les trois boîtes de vitesses  $v_{k,x}$ ,  $v_{k,y}$  et  $v_{k,z}$ , comme

$$v_{k,n}(\vec{k}, z) = \frac{-ik_n}{k^2} H(z) \frac{dG}{dz}(z) \delta_{k,\text{matière}}(\vec{k}); \quad n \in [\text{x}; \text{y}; \text{z}]. \quad (4.9)$$

Ces boîtes sont construites pour  $z = 0$ . Cette équation est équivalente à l'équation 4.8, car

$$f = \frac{d \ln G}{d \ln a} = \frac{a}{G} \frac{dG}{da} = -\frac{1}{aG} \frac{dG}{dz} \quad (4.10)$$

Comme précédemment, la boîte  $\delta_{k,\text{matière}}$  est la même que celle utilisée pour construire les boîtes  $\delta_{\text{QSO},i}$  des quasars et la boîte  $\delta_{\text{matière}}$  utilisée pour le Ly $\alpha$ , ceci afin de garantir la correspondance entre la densité des traceurs et leurs vitesses particulières. A l'aide d'une transformation de Fourier inverse, nous obtenons les trois boîtes de vitesse à  $z = 0$  dans l'espace réel  $v_x$ ,  $v_y$  et  $v_z$ . Le calcul pour obtenir la vitesse parallèle  $v_{\parallel}$  à un redshift  $z$  est décrit dans la section 4.2.2.

**Concernant le champ d'absorption Ly $\alpha$ , les RSD sont prises en compte par une modification de la formule FGPA (voir la fin de l'appendice A).** Pour ce faire, nous avons besoin du gradient de vitesse  $\eta_{\parallel}$  à  $z = 0$ . Le gradient de vitesse est donné par

$$\eta = \frac{1}{aH} \frac{\partial v_p}{\partial x_p}, \quad (4.11)$$

où  $\eta$  est le gradient sans dimension de la vitesse  $v_p$  le long de la ligne de visée, et  $x_p$  est la coordonnées comobile le long de cette ligne de visée. En utilisant la définition précédente, et l'équation 4.8, nous obtenons le gradient  $\eta_{nm}$  de la vitesse  $v_n$  selon la direction  $\vec{u}_m$  en fonction du champ  $\delta_k$  :

$$\eta_{nm}(\vec{k}) = -f \frac{k_n k_m}{k^2} \delta_k(\vec{k}). \quad (4.12)$$

Ceci nous permet, lorsque nous soustrayons  $\eta$  au contraste  $\delta$  (appendice A), de retrouver

la formule de Kaiser :

$$\begin{aligned}\delta_k^s(\vec{k}) &= \delta_k(\vec{k}) - \eta_{\parallel}(\vec{k}), \\ &= (1 + f\mu_k^2)\delta_k(\vec{k}).\end{aligned}\quad (4.13)$$

La boîte  $\delta_k$  utilisée est le GRF initial, afin de garantir les corrélations entre les différents champs. A l'aide d'une transformation de Fourier inverse, nous obtenons<sup>1</sup> les 6 boîtes de gradients de vitesses à  $z = 0$  dans l'espace réel  $\eta_{XX}$ ,  $\eta_{YY}$ ,  $\eta_{ZZ}$ ,  $\eta_{XY}$ ,  $\eta_{YZ}$  et  $\eta_{XZ}$ .

### 4.2.2 Le relevé de quasars

Une fois toutes ces boîtes construites, nous définissons la géométrie du relevé. Les boîtes, d'une taille  $2560 \times 2560 \times 1536$  selon les axes x, Y et z respectivement, sont placées à un redshift central  $z_0 = 1,71$ , et à une ascension droite  $\alpha_0$  et une déclinaison  $\delta_0$  (équation 4.20). Leurs dimensions permettent de couvrir les redshifts  $1,3 < z < 3,6$ . Cette limite basse est choisie de manière à pouvoir inclure les absorptions du CIV. Comme indiqué dans la section 3.3, le pixel d'absorption de plus bas redshift dans la région Ly $\alpha$  se trouve à  $z = 1,96$ . Si cette absorption est causée par du CIV, l'absorbeur se situe à un redshift  $z = (1 + 1,96)\lambda_{Ly\alpha}/\lambda_{CIV} - 1 \sim 1,32$ . Enfin, l'observateur est considéré être à  $z = 0$ , au centre du plan (x, y).

Afin d'obtenir un biais des quasars qui dépend du redshift, nous pouvons modifier l'équation 4.2 et inclure une facteur  $\alpha(z)$  qui prend en compte l'évolution avec le redshift du champ  $\delta_q$  associé aux quasars :

$$P \propto \exp(\alpha(z)\delta_q), \quad (4.14)$$

où  $\alpha(z)$  est donné par

$$\alpha(z) = \frac{b_{QSO}(z)(1+z_0)}{b_{QSO}(z_0)(1+z)}, \quad (4.15)$$

et  $z_0$  est le redshift auquel est construit le champ log-normal  $\delta_q$ , avec  $b_{QSO}$  son biais. La paramétrisation que nous utilisons pour le biais des quasars est donnée dans l'équation 1.64. La fonction de corrélation  $\xi_i$  à un redshift  $z_i$  est obtenue à partir de l'équation 4.3, et s'exprime comme

$$\xi_i = \exp(\alpha^2(z_i)\ln(1+\xi_0)) - 1, \quad (4.16)$$

où  $\xi_0$  est la fonction de corrélation à  $z_0 = 2,33$ , obtenue comme la transformation de Fourier du spectre de puissance  $P_{\text{matière}}(k)$  à  $z = 0$  fourni par Camb, et multipliée par le facteur de croissance  $G$  et le biais des quasars  $b_{QSO}$  au redshift  $z_0$ . Dans la limite où  $\xi_0 \ll 1$  ou  $|1 - \alpha| \ll 1$ , nous obtenons comme souhaité  $\xi_i = \alpha^2(z_i)\xi_0$ . La figure 4.1 compare les rapports  $\xi_i/\alpha^2(z_i)$  avec la fonction de corrélation de référence  $\xi_0$ . Nous pouvons remarquer que plus le redshift  $z_i$  est éloigné de  $z_0$ , et plus l'écart entre  $\xi_i/\alpha^2(z_i)$  et  $\xi_0$  est grand. Ceci est d'autant plus vrai pour les petites séparations  $r$ , pour lesquelles  $\xi_0$  n'est pas négligeable devant 1. Nous pouvons aussi remarquer sur le graphique de droite de la figure 4.1 que les formes des rapports  $\xi_i/\alpha^2(z_i)\xi_0$  pour  $z_i > z_0$  et  $z_i < z_0$  sont semblables et de signe opposé. Nous

1. Lors de la construction des 6 boîtes de gradients de vitesses, nous omettons volontairement le facteur  $f(z = 0)$  de l'équation 4.12. Ce facteur  $f$  manquant sera pris en compte lors de l'ajout de la dépendance en redshift (voir section 4.2.4)

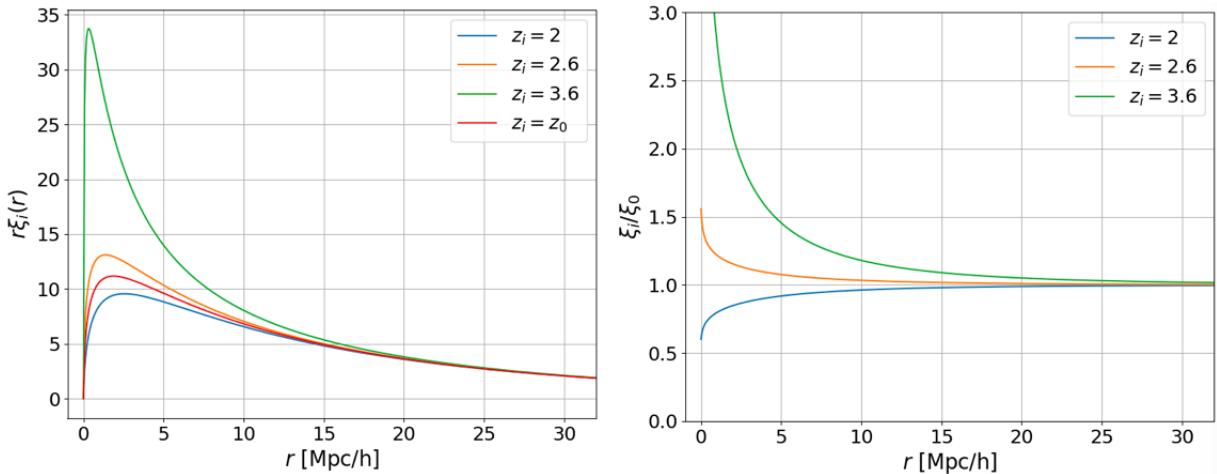


FIGURE 4.1 – Gauche : fonctions de corrélation  $\xi_i$  obtenues à partir de l'équation 4.16 et normalisées par un facteur  $1/\alpha^2(z_i)$ . La fonction de référence  $\xi_0$  est calculée à un redshift  $z_0 = 2,33$ . Droite : rapports  $\xi_i/\alpha^2(z_i)\xi_0$ .

pouvons donc combiner ces fonctions de corrélation afin d'avoir un résultat plus satisfaisant.

Nous considérons donc maintenant deux champs log-normaux  $\delta_{\text{QSO},1}$  et  $\delta_{\text{QSO},2}$  à  $z_1 = 2,1$  et  $z_2 = 3,5$ . Pour chaque champ, comme précédemment, nous construisons le champ

$$\hat{\delta}_{\text{QSO},i}(z) = \exp(\alpha(z)\delta_{\text{QSO},i}), \quad i \in [1; 2], \quad (4.17)$$

puis, nous définissons le champ  $\hat{\delta}_{\text{QSO},12}$  comme une interpolation linéaire des deux champs définis dans l'équation précédente :

$$\hat{\delta}_{\text{QSO},12}(z) = \hat{\delta}_{\text{QSO},1}(z) \frac{z_2 - z}{z_2 - z_1} + \hat{\delta}_{\text{QSO},2}(z) \frac{z - z_1}{z_2 - z_1}. \quad (4.18)$$

La figure 4.2 présente les fonctions de corrélation normalisées  $\xi_i$  qui correspondent au champ  $\hat{\delta}_{\text{QSO},12}$  à différents redshifts  $z_i$ . Elles sont obtenues en interpolant linéairement (équation 4.18) les fonctions de corrélation  $\xi_1$  et  $\xi_2$ , elles-mêmes obtenues aux redshifts  $z_1$  et  $z_2$  grâce aux transformations log-normales (équation 4.16) de  $\alpha^2(z_1)\xi_0$  et  $\alpha^2(z_2)\xi_0$ . Nous pouvons remarquer sur cette figure 4.2 que l'écart entre les fonctions de corrélation  $\xi_1$  et  $\xi_2$  (en rouge) et les fonctions de corrélation interpolées est moins important que dans le cas avec un seul champ log-normal. La courbe orange présente la fonction de corrélation  $\xi_i$  au redshift  $z_i = 3,0$  pour lequel l'écart avec  $\xi_1$  et  $\xi_2$  est maximal. Comme précédemment, nous pouvons constater que l'effet entre une interpolation (orange) et une extrapolation (vert, par exemple) et similaire et de signe opposé. Afin d'affiner encore l'évolution avec le redshift du champ utilisé pour tirer les quasars, nous pouvons considérer trois champs log-normaux, et combiner une interpolation et une extrapolation obtenues à partir de ces trois champs.

Nous considérons donc finalement trois champs log-normaux  $\delta_{\text{QSO},1}$ ,  $\delta_{\text{QSO},2}$  et  $\delta_{\text{QSO},3}$  à  $z_1 = 1,9$ ,  $z_2 = 2,75$  et  $z_3 = 3,6$ . C'est la solution qui a été choisie et implémentée dans les mocks. Ces trois champs sont stockés dans les boîtes construites dans la section 4.2.1. A partir de ces trois boîtes, nous construisons les 3 boîtes  $\hat{\delta}_{\text{QSO},1}$ ,  $\hat{\delta}_{\text{QSO},2}$  et  $\hat{\delta}_{\text{QSO},3}$  (équation 4.17), puis les deux boîtes interpolées

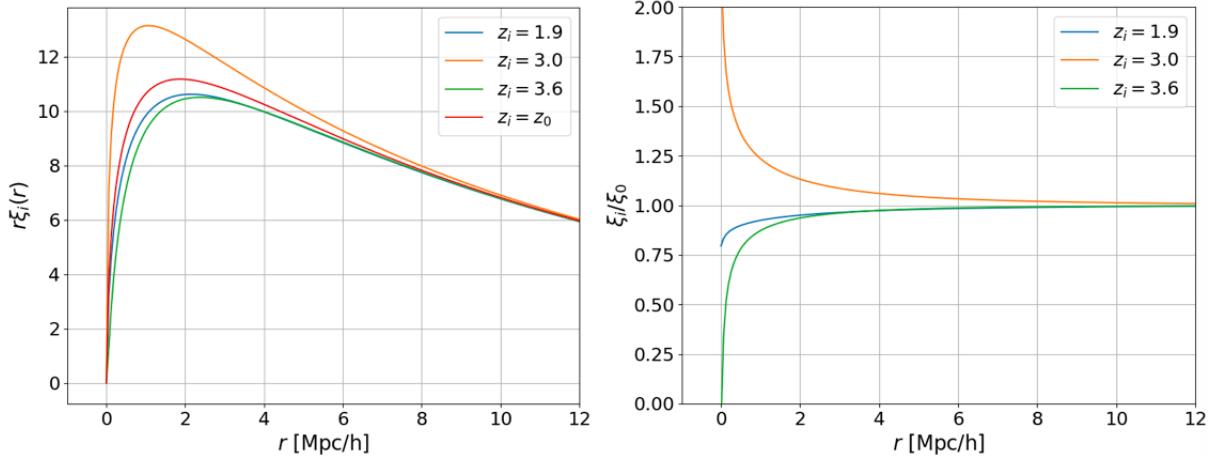


FIGURE 4.2 – Gauche : fonctions de corrélation  $\xi_i(z_i)$ , normalisées par un facteur  $1/\alpha^2(z_i)$  et obtenues par interpolation (orange) et extrapolation (bleu et vert) des fonctions de corrélation  $\xi_1$  et  $\xi_2$  à  $z_1 = 2,1$  et  $z_2 = 3,5$ . Leur obtention est détaillée dans le texte. Droite : rapports  $\xi_i/\alpha^2(z_i)\xi_0$ . Le redshift  $z = 3,0$  est le redshift pour lequel la déviation par rapport à  $\xi_0$  est la plus importante.

$\hat{\delta}_{\text{QSO},12}$  et  $\hat{\delta}_{\text{QSO},23}$  (équation 4.18). Enfin, nous construisons la boîte

$$\hat{\delta}_{\text{QSO}}(z) = K(z)\hat{\delta}_{\text{QSO},12}(z) + (1 - K(z))\hat{\delta}_{\text{QSO},23}(z), \quad (4.19)$$

comme la combinaison linéaire de deux boîtes interpolées  $\hat{\delta}_{\text{QSO},12}$  et  $\hat{\delta}_{\text{QSO},23}$ . Le paramètre  $K(z)$  est déterminé tel que, pour chaque  $z$ , le rapport de la fonction de corrélation  $\xi_{\text{QSO}}$ , normalisée par  $1/\alpha^2(z)$  et obtenue avec l'équation 4.19, par la fonction de corrélation  $\xi_0$ , vaille 1 à  $r = 5 h^{-1} \text{ Mpc}$ . La figure 4.3 présente les rapports normalisés  $\xi_{12}(z)/\xi_0$  (correspondant à l'interpolation, en bleu) et  $\xi_{23}(z)/\xi_0$  (correspondant à l'extrapolation, en orange) au redshift  $z = 2,38$ . La combinaison linéaire des deux (équation 4.19) est montrée en vert. Le redshift  $z = 2,38$  correspond au redshift pour lequel la différence entre  $\xi_{\text{QSO}}$  et  $\xi_0$  est la plus grande. Le graphique de droite de cette figure 4.3 présente un zoom de la courbe verte. Sur ce zoom, nous pouvons voir que  $\xi_{\text{QSO}}$  ne dévie pas de  $\xi_0$  de plus de  $5 \times 10^{-4}$  pour  $r > 5 h^{-1} \text{ Mpc}$ .

Les quasars sont ensuite tirés dans chaque voxel, avec une probabilité  $P \propto \hat{\delta}_{\text{QSO}}$ . Pour ce faire, nous générerons une variable  $\phi$  aléatoire uniforme entre 0 et 1 dans chaque voxel. Les voxels pour lesquelles  $\phi < N(z)\hat{\delta}_{\text{QSO}}$  hébergent un quasar. Le facteur  $N(z)$  contient l'évolution avec le redshift du nombre de quasars par degré carré. Il contient aussi un facteur de normalisation. Une fois les quasars tirés, nous les plaçons aléatoirement dans leur voxel. Puis, les quasars dont le redshift est en dehors de l'intervalle  $[1,8; 3,6]$  sont écartés. Les quasars dont l'ascension droite et la déclinaison sont en dehors des intervalles  $[-\Delta\alpha; \Delta\alpha]$  et  $[-\Delta\delta; \Delta\delta]$  sont aussi écartés. L'ascension droite  $\alpha$  et la déclinaison  $\delta$  du point  $(x, y, z)$  sont données par

$$\alpha = \arctan\left(\frac{\cos(\alpha_0)x - \sin(\delta_0)\sin(\alpha_0)y + \cos(\delta_0)\sin(\alpha_0)z}{-\sin(\alpha_0)x - \sin(\delta_0)\cos(\alpha_0)y + \cos(\delta_0)\cos(\alpha_0)z}\right), \quad (4.20)$$

$$\delta = \arcsin\left(\frac{\cos(\delta_0)y + \sin(\delta_0)z}{\sqrt{x^2 + y^2 + z^2}}\right). \quad (4.21)$$

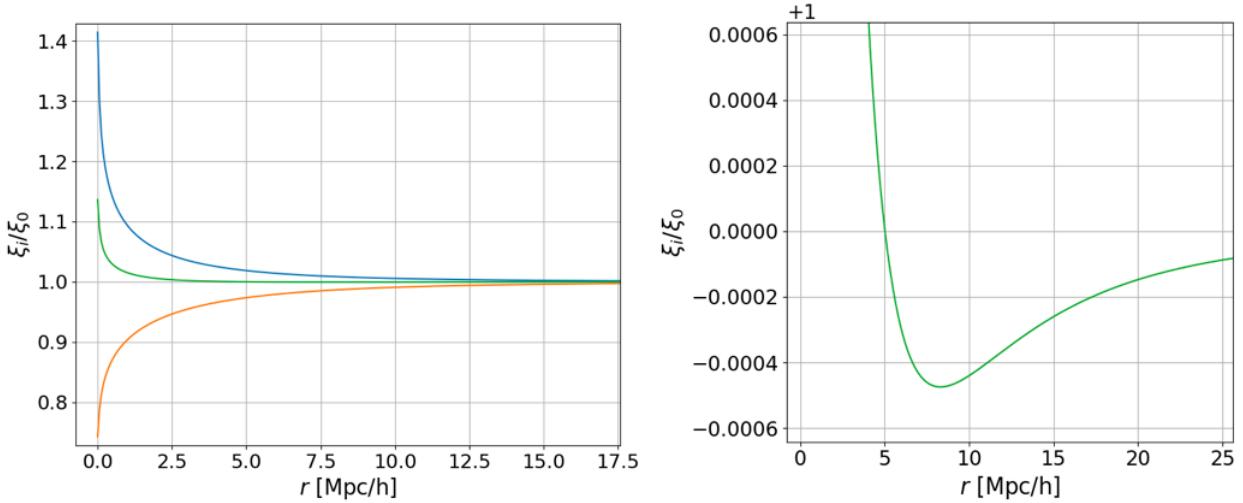


FIGURE 4.3 – Rapports normalisés  $\xi_{12}(z)/\xi_0$  (bleu) et  $\xi_{23}(z)/\xi_0$  (orange) à  $z = 2,38$ . Pour ce redshift,  $\xi_{12}$  est obtenue comme une interpolation de  $\xi_1$  et  $\xi_2$ , et  $\xi_{23}$  comme une extrapolation de  $\xi_2$  et  $\xi_3$ . La courbe verte est obtenue comme une combinaison linéaire de  $\xi_{12}$  et  $\xi_{23}$ . Le redshift  $z = 2,38$  est le redshift pour lequel la courbe verte dévie le plus de  $\xi_0$ . Le détail est donné dans le texte. Le graphique de droite présente un zoom de celui de gauche.

Enfin, grâce au facteur  $N(z)$ , les quasars sont tirés selon la distribution en  $z$  prédictive pour DESI. Cette distribution est présentée sur la figure 4.4. Cependant, nous tirons environ deux fois plus de quasars, afin de pouvoir simuler, entre autre, la sélection des cibles à l'aide du code `quickquasars` (présenté dans la section 4.5). De plus, cela permet d'utiliser les mocks pour d'autres projets, comme WEAVE<sup>1</sup>, qui possèdent une densité de cible plus élevée que celle de DESI. A la fin, nous obtenons environ 100 quasars à  $z > 2,1$  par degré carré.

Une fois les quasars tirés, nous les déplaçons proportionnellement à leur vitesse  $v_{\parallel}(z)$  le long de la ligne de visée. La vitesse, à  $z = 0$ , le long de la ligne de vitesse est obtenue comme

$$v_{\parallel} = \frac{v_X X + v_Y Y + v_Z Z}{\sqrt{X^2 + Y^2 + Z^2}}, \quad (4.22)$$

où les vitesses  $v_X$ ,  $v_Y$  et  $v_Z$  sont les vitesses obtenues par transformation de Fourier inverse des vitesses  $v_{k,X}$ ,  $v_{k,Y}$  et  $v_{k,Z}$ , définies dans l'équation 4.9. Ainsi, un quasar situé à une distance  $R$  sera replacé le long de la ligne de visée à une distance

$$R \rightarrow R + \frac{(1+z)}{H(z)} v_{\parallel}(z), \quad (4.23)$$

où  $v_{\parallel}(z)$  est relié à  $v_{\parallel}$  à  $z = 0$  par (voir équation 4.9)

$$v_{\parallel}(z) = \frac{H(z)dG/dz}{[H(z)dG/dz]_{z=0}} v_{\parallel}. \quad (4.24)$$

1. WEAVE est un spectrographe à fibre multi-objets. Il possède 1000 fibres avec un champ de vue de  $3,1 \text{ deg}^2$ . Le relevé WEAVE-QSO prévoit d'observer 350 000 quasars à grand redshift, sur un relevé de  $6000 \text{ deg}^2$  (PIERI et al. 2016).

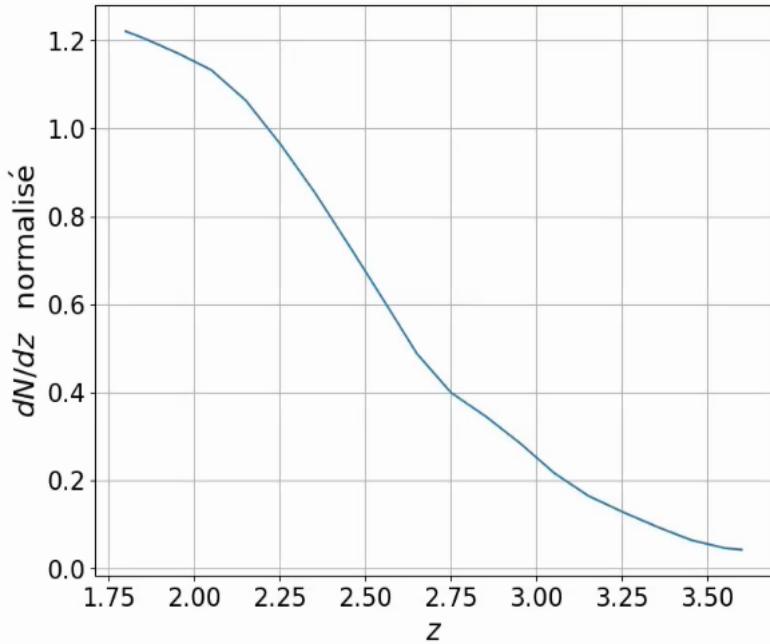


FIGURE 4.4 – Distribution normalisée en redshift des quasars tirés dans les mocks.

Le facteur  $(1 + z)$  dans l'équation 4.23 vient de la conversion des distances en distances comobiles. Une fois tous les quasars déplacés, leur redshift est recalculé, puis ils sont stockés dans un catalogue. Pour chaque quasar, le catalogue contient leur position dans le ciel  $(\alpha, \delta)$ , leur redshift avec et sans RSD, ainsi qu'un identifiant unique.

#### 4.2.3 Crédation des lignes de visée

A cette étape, nous disposons d'un catalogue de quasars corrélés avec le champ de densité  $\delta_{\text{matière}}$  qui sera utilisé pour construire l'absorption Ly $\alpha$ . Nous pouvons donc créer les lignes de visées à partir de chaque quasar et jusqu'à l'observateur, puis interpoler la boîte contenant le champ de densité le long de ces lignes de visée. Dans un premier temps, nous commençons par définir le tableau en longueurs d'onde observées, sur lequel sera interpolé la boîte de densité. Nous choisissons une taille de pixel  $d_{\text{pix}} = 0,2 h^{-1} \text{ Mpc}$ . Les limites  $1,8 < z < 3,6$  en redshift se traduisent par des limites  $3403,876 < \lambda < 5592,082 \text{ \AA}$  sur la longueur d'onde observée pour le Ly $\alpha$ . Nous ajoutons la limite basse des spectrographes de DESI :  $\lambda_{\min} = 3530 \text{ \AA}$ , que nous réduisons afin d'inclure certains métaux dans les forêts (cela sera expliqué dans la section 5.4.1). Les longueurs d'onde observées couvrent donc  $3476,1877 < \lambda < 5591,566 \text{ \AA}$  à l'aide de 6524 pixels<sup>1</sup>.

Une fois ce tableau en longueur d'onde défini, nous faisons correspondre le pixel  $\lambda_{\text{QSO}} = (1 + z_{\text{QSO}})\lambda_{\text{Ly}\alpha}$  à la position  $(x_{\text{QSO}}, y_{\text{QSO}}, z_{\text{QSO}})$  du quasar. Puis, pour chaque pixel  $i$  entre  $\lambda_{\min} = 3476,1877$  et  $\lambda_{\text{QSO}}$ , la position  $(x_i, y_i, z_i)$  du pixel est déterminée. Pour chaque pixel, nous calculons une moyenne pondérée des voxels voisins avec un lissage gaussien. Ce lissage est nécessaire afin d'éviter le crénelage (*aliasing*) aux petites échelles. Sans lissage, les spectres interpolés possèdent des discontinuités (figure 4.5). Ces discontinuités rajoutent de la puissance parasite aux petites échelles lors du calcul

1. Ces limites sont aussi choisies afin de garantir un nombre entier de pixels.

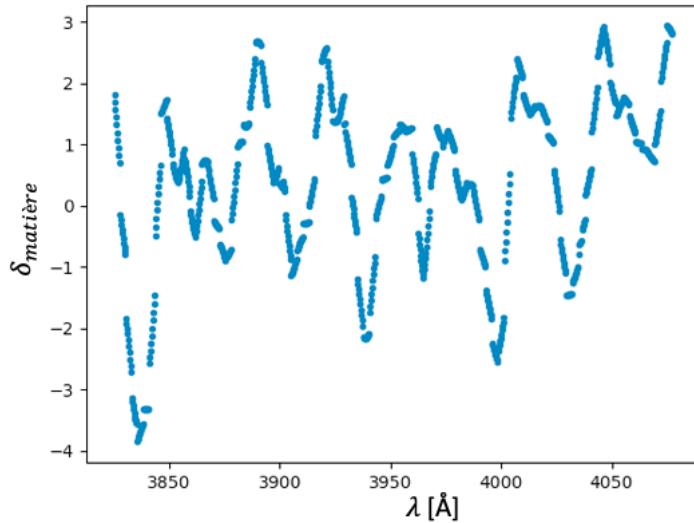


FIGURE 4.5 – Exemple de champ interpolé en utilisant les voxels à  $\pm 1\sigma$  du voxel central. Le champ ainsi interpolé possède de nombreuses discontinuités. Ces discontinuités produisent de la puissance supplémentaire aux petites échelles (effet de crénage ou d'*aliasing*).

du spectre de puissance. Nous appliquons donc un lissage gaussien, de largeur  $\sigma = d_{\text{cell}}$ . La figure 4.5 présente le champ interpolé en considérant uniquement les voxels à  $\pm 1\sigma$ . Trop peu de voxels sont considérés pour le calcul de chaque pixel, et le champ obtenu contient des discontinuités. La figure 4.6 présente le champ interpolé en considérant les voxels à  $\pm 2\sigma$  (orange) et  $\pm 3\sigma$  (bleu). La différence entre les deux est faible, cependant certaines discontinuités subsistent pour le champ calculé avec  $2\sigma$ , comme le montre le zoom sur le graphique de droite. Enfin, nous avons vérifié que les différences entre les champs obtenus avec 3 et  $4\sigma$  sont plus faibles que  $10^{-3}$ , et donc négligeables par rapport au bruit dans chaque pixel. Nous nous limitons donc aux voxels compris à  $\pm 3\sigma$  du voxel central pour limiter le temps CPU. Pour chaque pixel, nous considérons les voxels appartenant au cube de 7 voxels de côté, centré sur le voxel dans lequel se trouve le pixel. Ceci représente donc, pour chaque pixel, une moyenne sur 343 voxels. Le champ dans le pixel  $i$  est alors donné par

$$\delta_i = \frac{\sum_{j=0}^{342} \delta_j w_{ij}}{\sum_{j=0}^{342} w_{ij}}, \quad (4.25)$$

avec

$$w_{ij} = \exp\left(\frac{-(\vec{r}_j - \vec{r}_i)^2}{2\sigma^2}\right), \quad (4.26)$$

où  $\vec{r}_j$  est la position du centre du voxel  $j$ ,  $\vec{r}_i$  celle du pixel  $i$ , et  $\delta_j$  est la valeur de la boîte au pixel  $j$ .

Ce calcul est effectué pour tous les pixels qui vérifient  $\lambda_i < \lambda_{\text{QSO}}$ , pour chaque quasar. Les boîtes interpolées sont la boîte  $\delta_{\text{matière}}$  utilisée pour construire l'absorption Ly $\alpha$ , les trois boîtes de vitesse utilisées pour ajouter les RSD aux HCD tirés dans chaque ligne de visée (section 4.2.5), et les six boîtes de gradient de vitesse afin d'ajouter les RSD au champ Ly $\alpha$ .

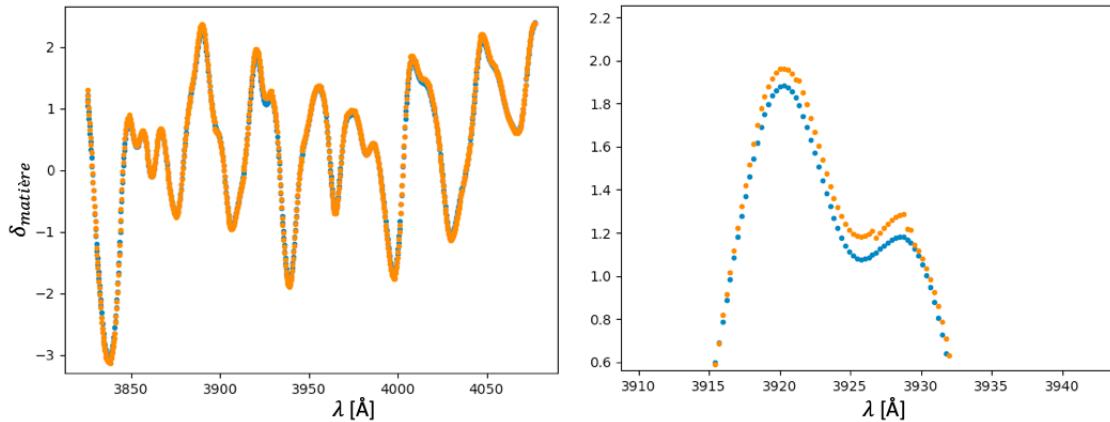


FIGURE 4.6 – Exemple de champ interpolé en utilisant les voxels à  $\pm 2\sigma$  (orange) et à  $\pm 3\sigma$  (bleu) du voxel central. Les discontinuités sont beaucoup moins nombreuses que pour le champ interpolé avec  $\pm 1\sigma$  (figure 4.5). Cependant, certaines sont encore visibles sur le zoom, présenté sur le graphique de droite.

#### 4.2.4 De la densité à l'absorption

Une fois les lignes de visés interpolées, nous pouvons transformer le champ de densité en absorption Ly $\alpha$ . Ceci est fait via la formule FGPA :

$$F = \exp[-a(z) \exp(b(z)G(z)\delta_{\text{matière}})] . \quad (4.27)$$

#### Les petites échelles

La boîte  $\delta_{\text{matière}}$ , utilisée dans l'équation 4.27, contient le champ de matière à grande échelle. Elle est construite grâce à la transformation de Fourier de la boîte  $\delta_k$ . Ce champ est construit sur une grille d'intervalle  $d_{\text{cell}} = 2,19 h^{-1} \text{ Mpc}$ . Par conséquent, la plus petite échelle accessible est

$$k_N = \frac{\pi}{d_{\text{cell}}} \sim 1,43 h \text{ Mpc}^{-1} . \quad (4.28)$$

Nous manquons donc toutes les fluctuations pour lesquelles  $k > k_N$ . Ces fluctuations ne sont pas importantes pour la corrélation à trois dimensions. La figure 4.7 présente le spectre de puissance de la matière  $P_{\text{matière}}(k)$  à  $z = 0$  (noir), obtenu avec Camb, ainsi que l'effet du lissage gaussien et l'effet de la taille non nulle de ces voxels. La courbe bleu donne le spectre de puissance qui inclut l'effet du lissage gaussien. Il s'exprime comme

$$P(k) = W^2(k)P_{\text{matière}}(k) , \quad (4.29)$$

où  $W(k)$  est le terme représentant le lissage gaussien appliqué à l'interpolation de la boîte  $\delta_l$  :

$$W(k) = \exp\left(-\frac{k^2 d_{\text{cell}}^2}{2}\right) . \quad (4.30)$$

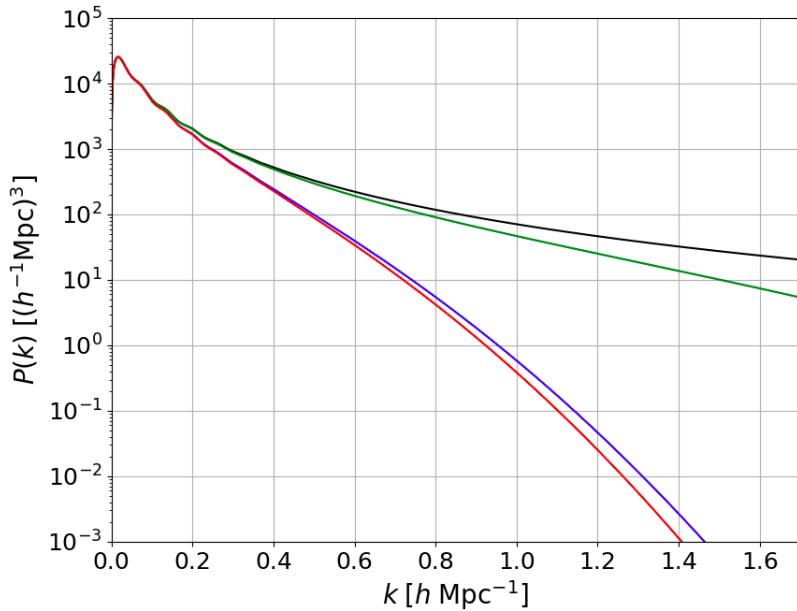


FIGURE 4.7 – Effet du lissage gaussien et de la taille non nulle des pixels sur le spectre de puissance. La courbe noire donne le spectre de puissance de la matière, obtenu avec Camb. La courbe bleu donne le spectre de puissance de la matière multiplié par  $W^2(k)$  (équation 4.30). La courbe verte donne le spectre de puissance multiplié par  $G^2(k)$  (équation 4.30). Enfin la courbe rouge donne le spectre de puissance multiplié par  $W^2(k)G^2(k)$ .

La courbe verte donne le spectre de puissance qui inclut l'effet de la taille des voxels. Celui-ci s'exprime comme

$$P(k) = G^2(k)P_{\text{matière}}(k), \quad (4.31)$$

où  $G(k)$  est le terme qui prend en compte la taille des voxels :

$$G(k) = \text{sinc}\left(\frac{kd_{\text{cell}}}{2}\right). \quad (4.32)$$

La courbe rouge donne le spectre de puissance qui inclut ces deux effets, c'est à dire multiplié par  $W^2(k)G^2(k)$ . La figure 4.8 présente les fonctions de corrélation, qui sont les transformée de Fourier de ces spectres de puissance. Les effets liés au lissage gaussien et à la taille des voxels sont principalement localisés à petit  $r$  et autour du pic BAO. Ces effets sont faibles, en particulier celui produit par la taille non nulle des voxels. Ainsi, ces fluctuations à petite échelle ne sont pas importantes pour la fonction de corrélation à trois dimension.

Cependant, ces fluctuations sont importantes pour le spectre de puissance à une dimension. Celui-ci est calculé en corrélant les pixels d'une même forêt. Il donne la puissance mesurée le long de la ligne de visée. Pour un mode  $k_{\parallel}$  donné, on peut montrer que le spectre de puissance à une dimension est donné par l'intégrale du spectre de puissance à trois dimensions sur la direction transverse à la ligne de visée (LUMSDEN et al. 1989) :

$$P^{1\text{D}}(k_{\parallel}) = \frac{1}{2\pi} \int_0^{\infty} k_{\perp} P(k_{\parallel}, k_{\perp}) dk_{\perp}. \quad (4.33)$$

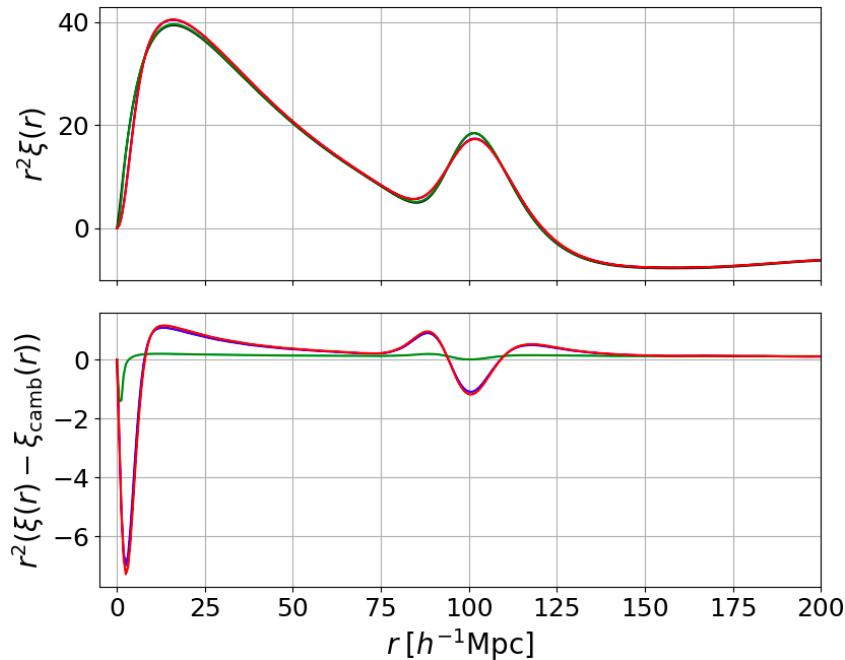


FIGURE 4.8 – Effet du lissage gaussien et de la taille non nulle des pixels sur la fonction de corrélation des mocks. Le graphique du haut présente les fonctions de corrélation obtenues à partir du spectre de puissance de Camb (noir), du spectre de puissance de Camb multiplié par  $W^2(k)$  en bleu, du spectre de puissance de Camb multiplié par  $G^2(k)$  en vert, et du spectre de puissance de Camb multiplié par  $W^2(k)G^2(k)$  en rouge (équations 4.30 et 4.32). Le graphique du bas présente la différence entre les courbes bleue et noire, en bleu, entre les courbes vert et noire, en vert, ainsi qu'entre les courbes rouge et noire, en rouge.

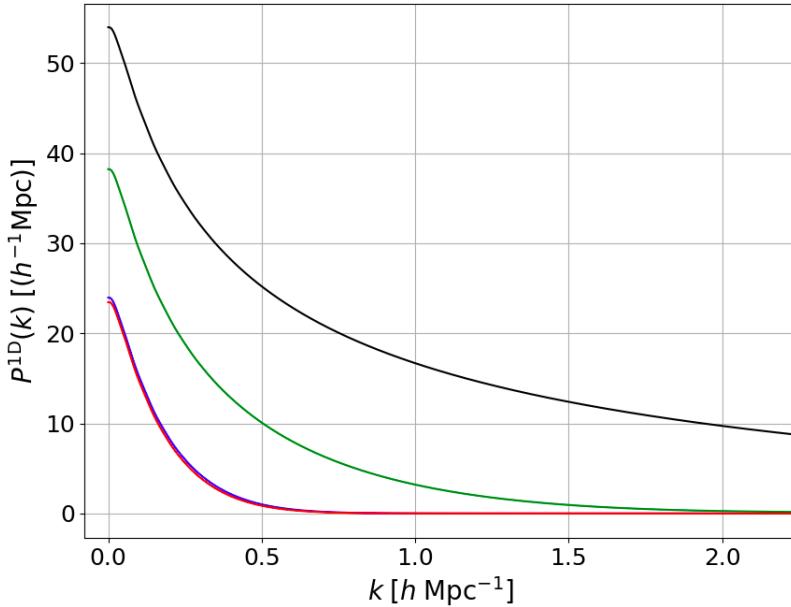


FIGURE 4.9 – Effet du lissage gaussien et de la taille non nulle des pixels sur le spectre de puissance à une dimension. La courbe noire donne le spectre de puissance à une dimension obtenu à partir du spectre de puissance de la matière (équation 4.33). La courbe bleu donne  $P^{1D}$  obtenu à partir du spectre de puissance de la matière multiplié par  $W^2(k)$  (équation 4.30). La courbe verte donne  $P^{1D}$  obtenu à partir du spectre de puissance de la matière multiplié par  $G^2(k)$  (équation 4.30). Enfin la courbe rouge donne  $P^{1D}$  obtenu à partir du spectre de puissance de la matière multiplié par  $W^2(k)G^2(k)$ . L’effet dominant sur le spectre de puissance à une dimension est l’effet du lissage gaussien.

Un spectre de quasar sonde la densité de matière le long de la ligne de visée. Il procure ainsi un échantillonnage, transverse à la ligne de visée, de la densité à une échelle très faible, de l’ordre du parsec. Le calcul du spectre de puissance sur de tels spectres implique donc une intégrale jusqu’à  $k_\perp = \infty$ . Dans notre cas, le spectre de puissance à trois dimensions  $P(k_\parallel, k_\perp)$  à intégrer est le spectre de puissance de la matière  $P_{\text{matière}}(k_\parallel, k_\perp)$  multiplié par  $G^2(k_\parallel, k_\perp)W^2(k_\parallel, k_\perp)$ . L’effet de  $G^2$  et  $W^2$  sur  $P^{1D}$  est présenté sur la figure 4.9. Contrairement au cas de la fonction de corrélation à trois dimensions, l’effet de la taille des voxels et du lissage gaussien est important pour le spectre de puissance à une dimension : les termes  $G^2$  et  $W^2$  réduisent l’amplitude des modes aux petites échelles, qui ont une contribution importante pour tous les  $k_\parallel$  lors du calcul de l’intégrale de l’équation 4.33. Pour palier ce problème, nous rajoutons indépendamment sur chaque ligne de visée un champ  $\delta_s$  (*small scales* : petites échelles) au champ  $\delta_{\text{matière}}$  que nous appelons désormais  $\delta_l$  (*large scales* : grandes échelles) :

$$F = \exp[-a(z)\exp(b(z)G(z)(\delta_l + \delta_s))];. \quad (4.34)$$

Comme ce champ n’est pas corrélé d’une ligne de visée à une autre, il ne participe pas à la fonction de corrélation à trois dimensions<sup>1</sup>. Afin d’ajouter la bonne quantité de fluctuations aux petites échelles, pour chaque ligne de visée nous générerons un GRF à une dimension  $\delta_{k,s}$ . Puis, nous multiplions ce

1. Lors du calcul de la fonction de corrélation à trois dimensions, nous ne considérons pas les paires de pixels issues de la même forêt (voir section 5.1.1)

champ par

$$\sqrt{\frac{P_s(k, z_{\text{eff}})}{d_{\text{pix}}}}, \quad (4.35)$$

où  $z_{\text{eff}}$  est le redshift moyen de la forêt, et  $P_s$  est le spectre de puissance qu'il faut appliquer à  $\delta_{k,s}$  afin que  $\delta_F$  possède le bon  $P^{1D}$ .  $\delta_{k,s}$  est créé avec une taille supérieure à celle de la forêt, pour éviter les corrélations d'une extrémité à l'autre. La détermination de  $P_s$  est détaillée dans la section 4.4. Puis, nous obtenons  $\delta_s$  à l'aide de la transformation de Fourier inverse de  $\delta_{k,s}$ . Avant d'ajouter  $\delta_s$  à  $\delta_l$ , nous corrigons la dépendance en  $z$  de chacun des  $\delta_s$  de la forêt. En effet,  $\delta_s$  est construit de façon à obtenir le bon spectre de puissance à une dimension au redshift moyen de la forêt. Les pixels situés en bord de forêt n'auront donc pas le bon  $P^{1D}$ . Ainsi, nous corrigons chaque  $\delta_s$  comme :

$$\delta_s(z) \rightarrow \delta_s(z) \frac{\sigma_s(z)}{\sigma_s(z_{\text{eff}})}, \quad (4.36)$$

où  $\sigma_s$  est l'écart type du champ  $\delta_s$ . Celui-ci est relié au spectre de puissance  $P_s$  par :

$$\sigma_s = \frac{1}{d_{\text{pix}}N} \left( P_s(0) + P_s(k_{\text{Ny}}) + 2 \sum_{j=1}^{\frac{N}{2}-1} P_s \left( \frac{2\pi}{d_{\text{pix}}N} j \right) \right), \quad (4.37)$$

où  $k_{\text{Ny}} = \pi/d_{\text{pix}} \sim 15,7 h^{-1} \text{ Mpc}$  est le mode de Nyquist : c'est le mode maximal accessible pour une taille de pixel donnée.

## Les RSD

Une fois les petites échelles ajoutées, nous obtenons un champ d'absorption  $F$  qui possède le bon spectre de puissance à trois dimensions, ainsi que le bon spectre de puissance à une dimension. Cependant, le champ d'absorption  $F$  ne possède pas de RSD pour l'instant, car il est construit à partir du spectre de puissance  $P_{\text{matière}}(k)$  qui est isotrope. Initialement, nous pensions ajouter les RSD au niveau du spectre de puissance : multiplier le GRF initial  $\delta_k$  par  $(1+\beta\mu^2)P_{\text{matière}}(k,\mu)$ , avec  $\mu = k_z/k$ . Mais cette solution n'est pas envisageable car les lignes de visées ne sont pas parallèles<sup>1</sup>, et nous ne pouvons donc pas confondre l'axe  $k_z$  avec l'axe de la ligne de visée  $k_{\parallel}$ . Une autre solution est d'utiliser le champ de vitesse défini dans l'équation 4.8. Chaque pixel d'absorption est alors déplacé proportionnellement à la vitesse particulière du gaz dans la cellule considérée. Puis l'absorption est modifiée en fonction de la différence des vitesses des pixels voisins. En effet, si cette différence est non nulle, le gaz se retrouve comprimé par endroit, et détendu dans d'autres, modifiant l'absorption dans chaque pixel. Cette méthode pour ajouter les RSD dans les mocks Ly $\alpha$  est la méthode choisie par LE GOFF et al. (2011) et FARR et al. (2019). Ce n'est pas la solution que nous utilisons ici. Afin d'inclure les RSD dans le champ d'absorption  $F$ , nous utilisons le champ de gradient de vitesse  $\eta_{\parallel}$ , présenté dans la section 4.2.1. Contrairement à la méthode qui utilise le champ de vitesse, la solution qui ajoute le champ  $\eta_{\parallel}$  au champ  $\delta_l$  permet de prédire les fonctions de corrélation des mocks. Ceci est un atout majeur pour nos simulations, car cela nous permet de prédire la valeur des paramètres que

1. Les lignes de visées étaient parallèles dans les mocks développés pour BOSS. Nous avons choisi d'implémenter des lignes de visée non parallèles afin que nos mocks, développés pour eBOSS et DESI, soient plus fidèles aux données.

nous devons obtenir lorsque nous ajustons les fonctions de corrélation des mocks. Cela est notamment très utile pour vérifier que la mesure du biais  $b_{\text{Ly}\alpha}$  et du paramètre RSD  $\beta_{\text{Ly}\alpha}$  faite lors de l'ajustement des fonctions de corrélation des mocks, dans l'optique d'effectuer une analyse RSD, est une mesure non biaisée. Nous décrivons la prédiction des fonctions de corrélation des mocks dans la section suivante.

Plusieurs essais ont été menés afin de savoir comment inclure correctement le champ  $\eta_{\parallel}$  dans FGPA. La première solution envisagée était de modifier la profondeur optique  $\tau$ . Comme  $\delta_{\tau} = \tau/\bar{\tau} - 1$  est conservée lors du passage dans l'espace des redshifts, nous modifions  $\tau$  (équation 4.6) comme

$$\tau \rightarrow \tau + \bar{\tau}\eta_{\parallel}. \quad (4.38)$$

Cette solution produisait une trop faible quantité de RSD. Nous avons alors décidé d'inclure les RSD causées par les petites échelles. Similairement à  $\delta_l$  et  $\delta_s$ , nous ajoutons à la profondeur optique la contribution aux grandes et petites échelles  $\eta_l$  et  $\eta_s$ . En faisant ceci, nous étions plus proche de ce que nous aurions obtenu avec une taille de voxel plus petite. La profondeur optique était alors donnée par

$$\tau = a \exp[bG(\delta_l + \delta_s)] + \bar{\tau}(\eta_l + \eta_s). \quad (4.39)$$

Cette solution produisait la bonne quantité de RSD. Cependant, la profondeur optique était négative pour un nombre non négligeable de pixels, résultant en une absorption  $F > 1$  non physique pour ces pixels. Finalement, nous avons décidé d'inclure le gradient de vitesse directement au niveau du champ  $\delta$ . Le champ  $\eta_{\parallel}$  est donc ajouté, en plus du champ  $\delta_s$ , au champ  $\delta_l$ . Ceci nous permet de retrouver la formule de Kaiser (équation 4.13). De plus, cette solution ne produit pas de pixel avec  $F > 1$ . Afin de gérer la bonne quantité de RSD, nous ajoutons un coefficient  $c$ , qui dépend de  $z$ . L'ajustement de ce paramètre nous permet d'obtenir la bonne dépendance en  $z$  pour  $\beta_{\text{Ly}\alpha}$ . La formule FGPA devient donc

$$F = \exp[-a(z) \exp(b(z)G(z)(\delta_l + \delta_s + c(z)\eta_{\parallel}))]. \quad (4.40)$$

Les champs  $\delta_l$ ,  $\delta_s$  et  $\eta_{\parallel}$  sont calculés à  $z = 0$ . De plus, le facteur  $f$  que nous avons laissé de côté dans la section 4.2.1 n'est pas explicité ici : pour les redshifts  $z > 2$ , l'univers est dominé par la matière et donc, en bonne approximation, nous avons  $f(z) \sim 1$ . Les paramètres  $a(z)$ ,  $b(z)$ ,  $c(z)$ , ainsi que  $P_s(k, z)$  sont ajustés afin d'obtenir les bons  $b_{\text{Ly}\alpha}(z)$ ,  $\beta_{\text{Ly}\alpha}(z)$ ,  $\bar{F}(z)$  et  $P_{\text{Ly}\alpha}^{\text{1D}}(k, z)$ . L'ajustement est décrit dans la section 4.4.

Les différentes étapes relatives à la création des forêts de chaque quasar sont résumées sur la figure 4.10. De haut en bas, le premier graphique de cette figure présente les champs à grande échelle  $\delta_l$ ,  $c(z)\eta_{\parallel}$  et la somme des deux. Le deuxième présente la somme  $\delta_l + \delta_s$  et le champ  $\delta_g$  défini comme  $\delta_g = \delta_l + \delta_s + c(z)\eta_{\parallel}$ . Le troisième présente la fraction de flux transmise  $F$ , et enfin le dernier graphique présente le spectre du quasar. Cette dernière étape est produite par le code `quickquasars` (voir section 4.5). Un DLA avec une grande densité de colonne est visible à  $\lambda_{\text{RF}} \sim 1165 \text{ \AA}$ . La remontée observée pour les longueurs d'onde proches de  $1200 \text{ \AA}$  est causée par la proximité de la raie Ly $\alpha$ .

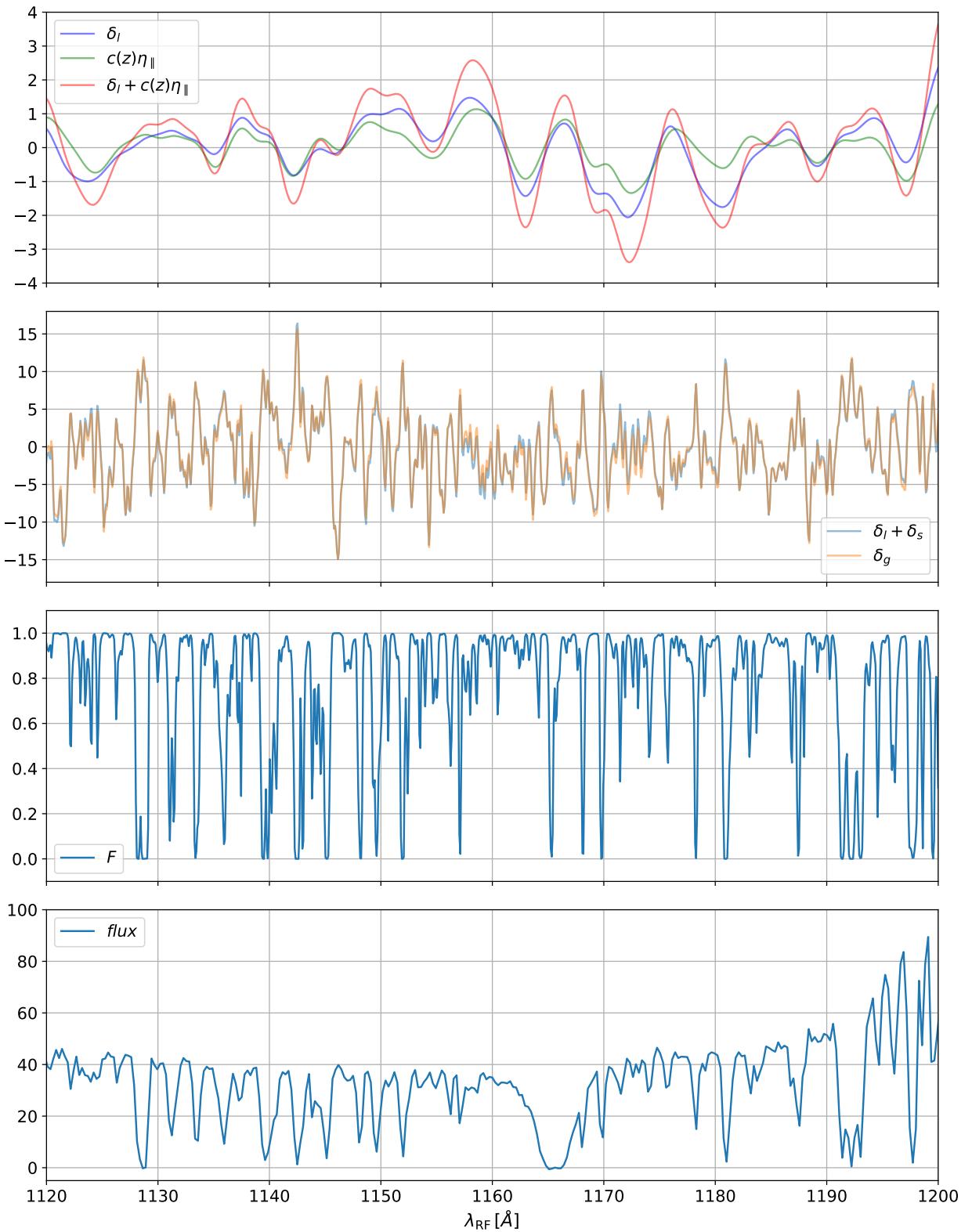


FIGURE 4.10 – De haut en bas, le premier graphique présente les champs à grande échelle  $\delta_l$ ,  $c(z)\eta_{\parallel}$  et la somme des deux. Le deuxième présente la somme  $\delta_l + \delta_s$  et le champ  $\delta_g$  défini comme  $\delta_g = \delta_l + \delta_s + c(z)\eta_{\parallel}$ . Le troisième présente la fraction de flux transmise  $F$ , et enfin le dernier graphique présente le spectre du quasar. Un DLA avec une grande densité de colonne est visible à  $\lambda_{\text{RF}} \sim 1165 \text{ \AA}$ .

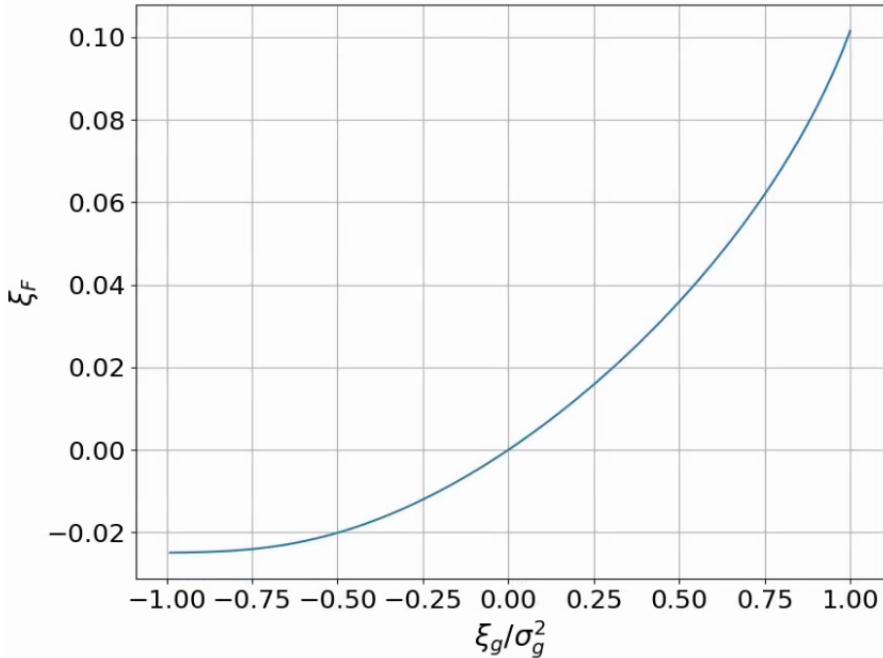


FIGURE 4.11 – La fonction de corrélation  $\xi_F$  obtenue à l'aide de l'équation 4.41 pour chaque valeur de  $\xi_g/\sigma_g^2 \in [-1; 1]$ .

### La prédition

La méthode que nous utilisons pour ajouter les RSD, décrite dans la section précédente, a l'avantage d'avoir une fonction de corrélation prédictible. Dans les lignes qui suivent, nous expliquons comment calculer la prédition de la fonction de corrélation. Comme décrit par FONT-RIBERA et al. (2012), dans le cas où le champ  $\delta_F$  est une fonction d'un champ gaussien  $\delta_g$ , il est possible de relier la fonction de corrélation  $\xi_F$  du champ  $\delta_F$  à la fonction de corrélation  $\xi_g$  du champ  $\delta_g$ . La fonction de corrélation  $\xi_F(r_{12})$  pour la séparation  $r_{12}$  est reliée à  $\xi_g(r_{12})$  par

$$\xi_F(r_{12}) = \int_{-\infty}^{\infty} d\delta_{g1} \int_{-\infty}^{\infty} d\delta_{g2} \frac{\exp\left[-\frac{\delta_{g1}^2 + \delta_{g2}^2 - 2\delta_{g1}\delta_{g2}\xi_g(r_{12})}{2(1 - \xi_g^2(r_{12}))}\right]}{2\pi\sqrt{1 - \xi_g^2(r_{12})}} \delta_F(\delta_{g1})\delta_F(\delta_{g2}), \quad (4.41)$$

où  $\delta_g$  est un GRF de variance 1 et  $\delta_F$  est le champ d'absorption calculé à partir du champ gaussien. Dans notre cas, le champ  $\delta_g$  représente le champ  $\delta_l + \delta_s + c(z)\eta_{\parallel}$ . Ce champ est un champ gaussien, de valeur moyenne nulle et de variance  $\sigma_g^2$ . Cette variance est dominé par la variance de  $\delta_s$ . Nous compensons le fait que  $\sigma_g^2 \neq 1$  en remplaçant  $\xi_g$  par  $\xi_g/\sigma_g^2$  dans l'équation 4.41. Dans cette équation,  $\xi_F(r_{12})$  ne dépend que de la valeur de  $\xi_g(r_{12})/\sigma_g^2$ . Nous construisons donc une table qui permet de relier chaque valeur de  $\xi_g/\sigma_g^2 \in [-1; 1]$  à la valeur  $\xi_F$  correspondante. La figure 4.11 présente la fonction de corrélation  $\xi_F$  obtenue, en fonction de  $\xi_g/\sigma_g^2$ . Par ailleurs, puisque  $\delta_s$  n'est pas corrélé d'une forêt à une autre, il ne participe pas au spectre de puissance à trois dimensions. D'après la formule de Kaiser (équation 4.13), le champ  $\delta_g$  suit le spectre de puissance

$$P_g(k) = (1 + c\mu_k^2)^2 P_l(k), \quad (4.42)$$

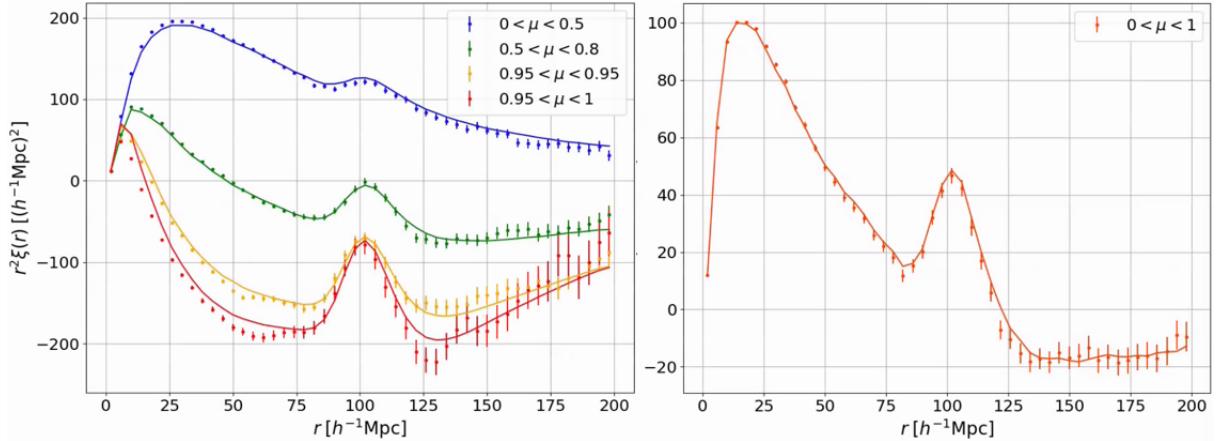


FIGURE 4.12 – Fonctions de corrélation du champ  $\delta_g$  (points) et prédition de cette fonction de corrélation (équations 4.42 et 4.43). Le graphique de gauche présente les corrélations dans quatre bins en  $\mu$ . Celui de droite présente les corrélations dans la gamme  $0 < \mu < 1$ . La corrélation mesurée sur le champ  $\delta_g$  est en bon accord avec la corrélation prédictive, même s'il existe un écart le long de la ligne de visée.

où  $P_l$  est le spectre de puissance que suit le champ  $\delta_l$ . La figure 4.12 présente la fonction de corrélation du champ  $\delta_g$  mesuré dans les mocks ainsi que la fonction de corrélation obtenue à partir du spectre de puissance de l'équation 4.42 (voir le passage de  $P_g(k)$  à  $\xi_g(r, \mu)$  dans les lignes qui suivent). Nous vérifions que ces deux fonctions de corrélation sont en accord. Nous observons cependant de petites différences essentiellement le long de la ligne de visée. Cela pourrait être dû à l'effet de la taille des voxels utilisés pour construire les boîtes de gradient de vitesse  $\eta$ , ou possiblement au lissage gaussien appliqué à  $\eta_{\parallel}$ .

Afin de relier la fonction de corrélation dans l'espace des redshifts  $\xi_g(r, \mu)$  que suit le champ  $\delta_g$  à la fonction de corrélation  $\xi(r)$  que suit le champ  $\delta_l$ , nous utilisons les formules données dans HAMILTON (1992) :

$$\xi_g(r, \mu) = \xi_0(r) + P_2(\mu)\xi_2(r) + P_4(\mu)\xi_4(r), \quad (4.43)$$

avec

$$\xi_0(r) = \left(1 + \frac{2}{3}c + \frac{1}{5}c^2\right) \xi(r), \quad (4.44)$$

$$\xi_2(r) = \left(\frac{4}{3}c + \frac{4}{7}c^2\right) [\xi(r) - \bar{\xi}(r)], \quad (4.45)$$

$$\xi_4(r) = \frac{8}{35}c^2 \left[\xi(r) + \frac{5}{2}\bar{\xi}(r) - \frac{7}{2}\bar{\bar{\xi}}(r)\right], \quad (4.46)$$

et

$$\bar{\xi}(r) = 3r^{-3} \int_0^r \xi(s)s^2 ds, \quad (4.47)$$

$$\bar{\bar{\xi}}(r) = 5r^{-5} \int_0^r \xi(s)s^4 ds. \quad (4.48)$$

Les  $P_l(\mu)$  sont les polynômes de Legendre :

$$P_0(\mu) = 1 ; \quad P_2(\mu) = \frac{1}{2}(3\mu^2 - 1) ; \quad P_4(\mu) = \frac{1}{8}(35\mu^4 - 30\mu^2 + 3). \quad (4.49)$$

Afin d'obtenir la prédiction  $\xi_F^{pred}(r, \mu)$ , nous commençons par calculer le spectre de puissance que suit la boîte  $\delta_l$  :

$$P(k) = W^2(k)P_{\text{matière}}(z=0), \quad (4.50)$$

où  $W(k)$  est le terme représentant le lissage gaussien appliqué à l'interpolation de la boîte  $\delta_l$  (équation 4.30). Nous négligeons ici l'effet de la taille non nulle des voxels sur le spectre de puissance (équation 4.32). La figure 4.8 présente l'effet de  $W(k)$  et  $G(k)$  sur la fonction de corrélation. L'effet lié à  $G(k)$  est nettement moins important que celui lié à  $W(k)$ . Puis, à l'aide d'une transformation de Fourier du spectre de puissance, nous obtenons la fonction de corrélation  $\xi(r)$  que suit le champ  $\delta_l$  obtenu par interpolation avec lissage des boîtes. Nous calculons ensuite la fonction de corrélation dans l'espace des redshifts  $\xi_g(r, \mu)$  que suit le champ  $\delta_g$  grâce à l'équation 4.43. La figure 4.12 présente cette fonction de corrélation mesurée sur une réalisation. Enfin, pour tous les couples  $(r, \mu)$  nécessaires, nous obtenons la fonction de corrélation  $\xi_F(r, \mu)$  du champ  $\delta_F$  comme la valeur correspondante à la valeur tabulée  $\xi_g(r, \mu)/\sigma_g^2$  pour  $\xi_g$ .

#### 4.2.5 Ajout des HCD

Les HCD ont un effet important sur les fonctions de corrélation, nous simulons donc aussi leur présence. De manière à avoir une corrélation entre les HCD et les autres traceurs des mocks, nous utilisons la boîte de densité  $\delta_l$  pour tirer les HCD. Nous ne considérons pas la somme  $\delta_l + \delta_s$  car les HCD sont des surdensités qui se situent dans les structures à grandes échelles : une résolution de  $2,19 h^{-1}$  Mpc est suffisante. De plus, l'ajout de  $\delta_s$  bruitterait la corrélation entre les HCD et les autres traceurs. En effet, l'écart type  $\sigma_s$  du champ  $\delta_s$  est entre 4 et 6 fois plus important que celui du champ  $\delta_l$ , et l'écart type  $\sigma_g$  est dominé à plus de 95 % par  $\sigma_s$ . Du fait que  $\delta_s$  ne possède pas de corrélation à 3 dimensions, les HCD se situeraient principalement dans des pics de bruit non corrélés.

Contrairement aux quasars, les HCD sont tirés selon les pics du champ  $\delta_l$  : nous identifions les pixels dans lesquelles  $\delta_l$  est au dessus d'un certain seuil, puis les HCD sont tirés dans ces pixels selon une loi de Poisson. Le seuil  $\nu$  est défini en fonction du biais souhaité pour les HCD. L'appendice A de FONT-RIBERA et MIRALDA-ESCUDÉ (2012) relie le rapport  $b_\nu/b_g$  au seuil  $\nu$  :

$$\left(\frac{b_\nu}{b_g}\right)^2 = \frac{p_g(\nu)}{\left(\int_\nu^\infty d\delta_g p_g(\delta_g)\right)^2} \int_\nu^\infty dy_1 p_g(y_1) y_1, \quad (4.51)$$

où  $b_\nu$  est le biais obtenu avec le seuil  $\nu$ ,  $b_g$  est le biais du champ gaussien, et  $p_g$  donne la densité de probabilité du champ gaussien de variance 1. Dans notre cas,  $b_g = 1$ . En utilisant le changement de variable  $u = y_1^2/2$ , l'intégrale au dénominateur de l'équation 4.51 vaut simplement  $p_g(\nu)$  :

$$\int_\nu^\infty dy_1 p_g(y_1) y_1 = \int_\nu^\infty dy_1 \frac{\exp(-y_1^2/2)}{\sqrt{2\pi}} y_1 = \int_{\nu^2/2}^\infty du \frac{\exp(-u)}{\sqrt{2\pi}} = \frac{\exp(-\nu^2/2)}{\sqrt{2\pi}} = p_g(\nu). \quad (4.52)$$

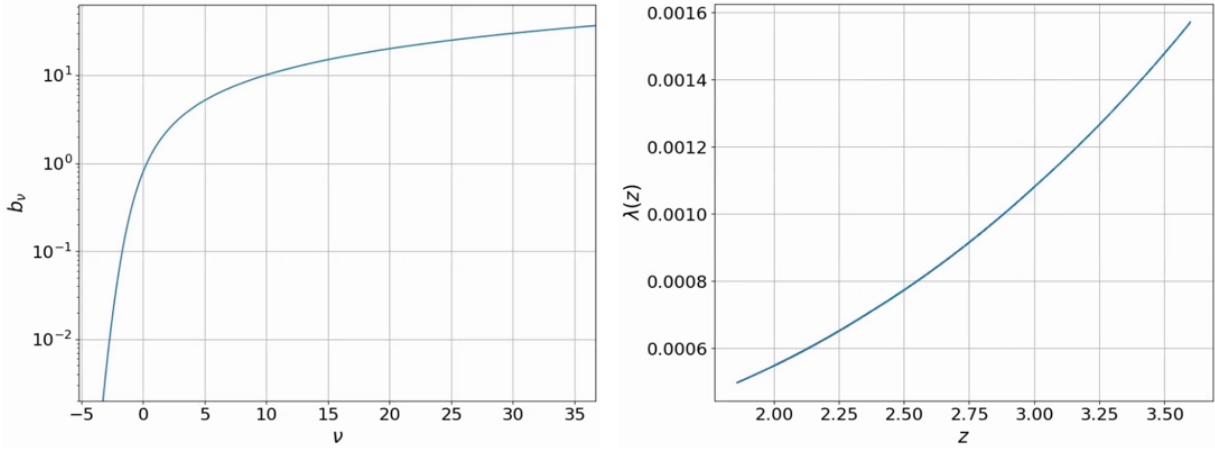


FIGURE 4.13 – Le graphique de gauche présente le biais  $b_\nu$  obtenu pour un seuil  $\nu$ . Le graphique de droite présente le paramètre  $\lambda(z)$ , utilisé pour tirer les HCD.

L'équation 4.51 est donc équivalente à

$$b_\nu = \frac{p_g(\nu)}{\int_\nu^\infty d\delta_g p_g(\delta_g)} . \quad (4.53)$$

Afin d'obtenir le seuil pour un biais donné, nous tabulons  $b_\nu$  pour une large gamme de seuils  $\nu$ , puis nous interpolons  $\nu$  en fonction de  $b_\nu$ . Le graphique de gauche de la figure 4.13 présente le biais  $b_\nu$  en fonction du seuil  $\nu$ . Dans notre cas, le champ  $\delta_l$  suit une distribution de probabilité gaussienne. Cependant, sa variance n'est pas égale à 1. De plus, le champ  $\delta_l$  interpolé le long des lignes de visée correspond au champ de matière à  $z = 0$ . Ainsi, pour obtenir un biais  $b_{\text{HCD}}$ , pour chaque redshift nous calculons le seuil  $\nu$  comme si nous visions un biais  $b = b_{\text{HCD}} \sigma_l G(z)$ . Le terme  $\sigma_l$  prend en compte la variance du champ  $\delta_l$ , et  $G(z)$  le fait que  $\delta_l$  soit construit à  $z = 0$ . Nous choisissons de tirer les HCD avec un biais  $b_{\text{HCD}} = 2$ . Une fois les pixels pouvant héberger un HCD identifiées, nous tirons dans chacune d'entre elles les HCD avec une loi de poisson de paramètre

$$\lambda(z) = \frac{N(z)}{\text{cdf}(-\nu(z))} , \quad (4.54)$$

où  $N(z)$  donne le nombre moyen de HCD attendu par pixel et  $\nu(z)$  le seuil au redshift  $z$ . Le nombre de HCD attendu est donné par la librairie `pyigm`<sup>1</sup>. La distribution en redshift des HCD est présentée sur le graphique de gauche de la figure 4.14, et le paramètre  $\lambda(z)$  sur le graphique de droite de la figure 4.13. En pratique, du fait que  $\lambda(z) \ll 1$ , il est très rare d'avoir plus d'un HCD par pixel<sup>2</sup>. Une fois tous les HCD tirés, nous leur assignons une densité de colonne dans la gamme  $17,2 < \log(n_{\text{HI}}) < 22,5$ , selon la distribution donnée par `pyigm`. Cette distribution est présentée sur le graphique de droite de la figure 4.14. Enfin, nous ajoutons les RSD aux HCD tirés. Chaque HCD tiré est déplacé le long de la ligne de visée proportionnellement à la vitesse  $v_{\parallel}(z)$ . Comme pour les quasars, la vitesse le long de la

1. <https://github.com/pyigm>

2. Lorsqu'un pixel contient plusieurs HCD, ces HCD sont gardés et une densité de colonne leur est assignnée. Ces deux HCD sont alors ajoutés au spectre synthétique par `quickquasars` (voir section 4.5), ce qui résulte en un seul HCD avec une densité de colonne valant la somme des densités de colonne des deux HCD tirés dans le pixel.

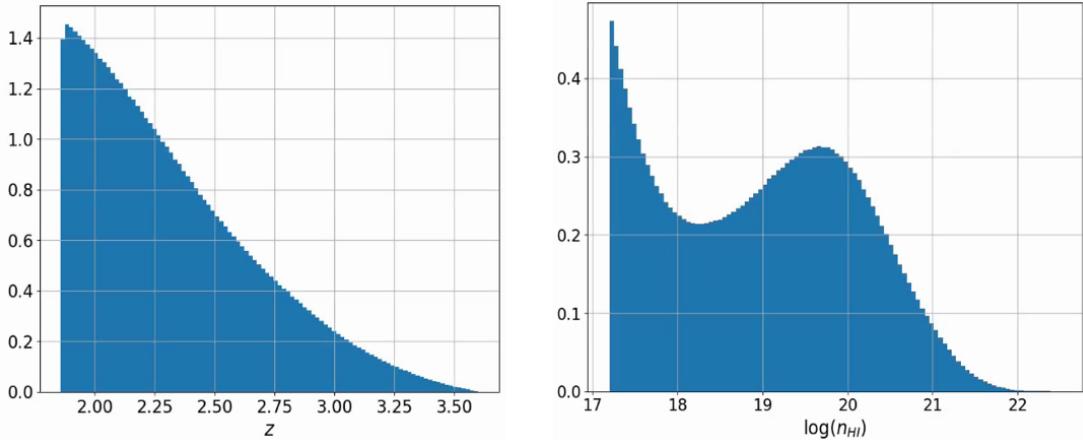


FIGURE 4.14 – Gauche : distribution normalisée en redshift des HCD. Droite : distribution normalisée de  $\log(n_{\text{HI}})$  des HCD. Ces distributions proviennent de la librairie pyigm.

ligne de visée à  $z = 0$  est donnée par

$$v_{\parallel} = \frac{v_x X + v_y Y + v_z Z}{\sqrt{x^2 + y^2 + z^2}}. \quad (4.55)$$

où, cette fois ci,  $v_x$ ,  $v_y$  et  $v_z$  sont les vitesses le long de la ligne de visée provenant de l'interpolation des boîtes de vitesse (section 4.2.3). Ainsi, un HCD à un redshift  $z$  sera déplacé à un redshift

$$z \rightarrow z + \frac{(1+z)}{c} \frac{H(z)dG/dz}{[H(z)dG/dz]_{z=0}} v_{\parallel}. \quad (4.56)$$

Dans les mocks que nous décrivons ici, le profil d'absorption des HCD n'est pas ajouté dans les forêts. Nous produisons uniquement un catalogue qui regroupe tous les HCD tirés. Le profil d'absorption est ajouté au spectre de chaque quasar par le code `quickquasars`, qui utilise le catalogue de HCD que nous produisons.

### 4.3 Production des mocks

Comme expliqué au début de ce chapitre, l'objectif des mocks est de reproduire les données d'eBOSS et de DESI. Étant donné que le relevé d'eBOSS est contenu dans le relevé de DESI, nous simulons directement le relevé DESI. Ainsi, lorsque nous avons besoin de simuler le relevé d'eBOSS, nous retirons les quasars qui ne sont pas contenu dans ce relevé. Les mocks ont été produits au centre de calcul NERSC<sup>1</sup>, avec la machine Cori. Sur cette machine, nous avons utilisé les nœuds « Haswell ». Ces nœuds possèdent 128Go de mémoire vive. De manière à créer nos boîtes de densité, via la transformation de Fourier à trois dimensions, il faut, pour chaque boîte, que l'intégralité de son contenu soit accessible depuis un même endroit. Nous pourrions distribuer la mémoire et effectuer la transformation de Fourier sur plusieurs nœuds à l'aide de la librairie MPI. Cependant nous n'avions pas l'expertise nécessaire. Nous effectuons donc les transformations de Fourier sur un seul nœud.

1. National Energy Research Scientific Computing Center (NERSC), a U.S. Department of Energy Office of Science User Facility operated under Contract No. DE-AC02-05CH11231.

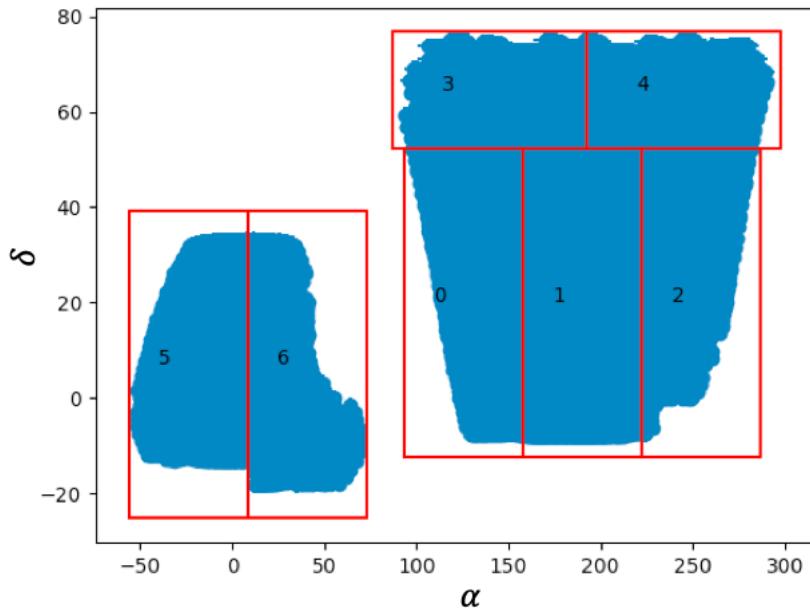


FIGURE 4.15 – Découpage du relevé de DESI en 7 chunks.

Nous profitons néanmoins des 32 coeurs par noeud pour paralléliser notre code. Chaque cœur possède 2 hyper-threads, ce qui permet de gérer 64 tâches simultanément sur un même noeud. Lors de la construction des boîtes (section 4.2.1), nous avons besoin de stocker en mémoire la boîte dans l'espace réel, la boîte dans l'espace  $k$ , et la boîte contenant la norme de  $k$ , cette dernière étant utilisée pour construire les boîtes de vitesses et de gradients de vitesse (équations 4.9 et 4.12). Étant donné la mémoire disponible sur chaque noeud, les boîtes ne doivent pas dépasser 42 Go. Cette mémoire disponible et la taille des voxels choisie ( $2,19 h^{-1}$  Mpc) ne permettent pas de générer l'entièreté du relevé DESI avec une seule boîte. Nous choisissons donc de découper le relevé en sept chunks indépendants, chacun étant généré par une boîte de  $2560 \times 2560 \times 1536$  voxels. Chacune des boîtes représente un volume de 37,5 Go. A la fin de la production, ces sept chunks sont assemblés pour reconstruire le relevé de DESI. Le découpage du relevé en sept chunks est montré sur la figure 4.15. Dans les lignes qui suivent, nous détaillons les différents éléments du code. Le schéma 4.16 résume la situation. Le première module, `interpolate_pk.py` permet de calculer puis d'interpoler les quatre spectres de puissance  $P_{\text{QSO},i}$  et  $P_{\text{matière}}$  sur la grille  $2560 \times 2560 \times 1536$  dans l'espace  $k$ . Le code est lancé séparément sur 16 morceaux de la boîte, puis le code `merge_pk.py` permet de rassembler des 16 morceaux des spectres de puissance interpolés et de les sauver au format FITS (Flexible Image Transport System). Cette étape est effectuée une seule fois, car les spectres de puissance sont communs à toutes les réalisations. Le module suivant est `make_boxes.py`. Ce code lit les spectres de puissance interpolés puis construit les trois boîtes de densité relatives aux quasars, la boîte de densité relative au Ly $\alpha$ , les trois boîtes de vitesse et les six boîtes de gradient de vitesse, décrites dans la section 4.2.1. Au total, 13 transformations de Fourier inverses sont effectuées. Ce code est lancé sept fois, afin de produire les boîtes pour les sept chunks. Une fois toutes les boîtes produites, les quasars sont tirés (section 4.2.2) grâce au code `draw_qso.py`. Afin d'accélérer la production des mocks, les boîtes sont partagées en 512 tranches selon l'axe x. Ces tranches, de taille  $5 \times 2560 \times 1536$ , sont traitées séparément. Ce code est donc tourné

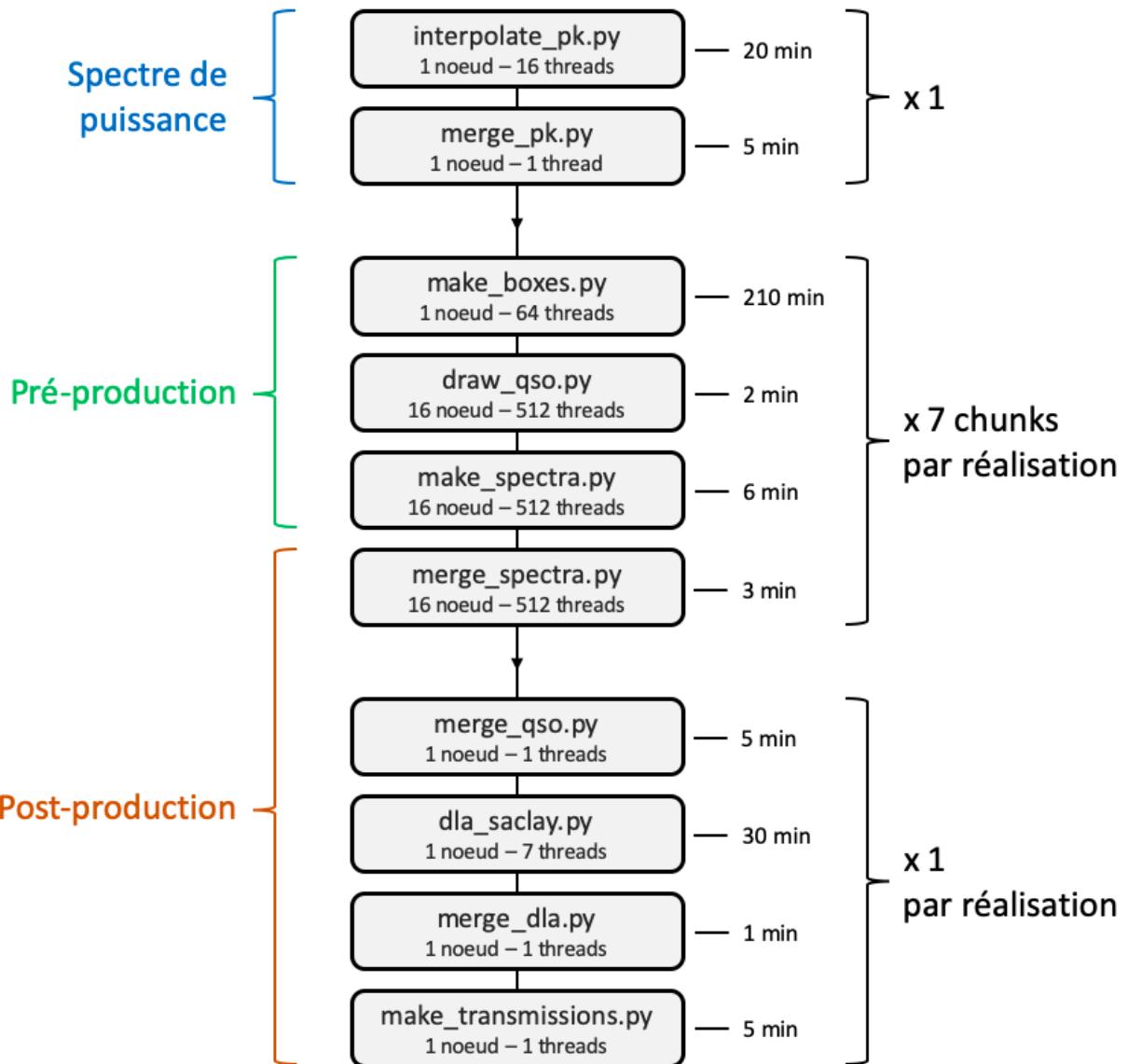


FIGURE 4.16 – Schéma illustratif du fonctionnement des mocks. Le premier bloc (bleu) génère le spectre de puissance nécessaire à la création des boîtes. Puis, pour chaque chunk de chaque réalisation, la pré-production (vert) génère les boîtes, tire les quasars, et interpole les lignes de visée. Enfin, la post-production (orange) rassemble les morceaux de lignes de visée et construit les forêts Ly $\alpha$  pour chaque chunk. Puis, elle regroupe le résultat de chaque chunk, construit les catalogues finaux de quasars et HCD et produit les fichiers transmissions pour chaque réalisation. Chaque case indique le nom du code, ainsi que les conditions dans lesquelles il est tourné. Les temps d'exécution indiqués sont approximatifs.

en parallèle 512 fois, sur 16 noeuds  $\times$  32 threads. Une fois les quasars tirés, les lignes de visée sont interpolées (section 4.2.3) avec le code `make_spectra.py`. De la même manière que le code précédent, il tourne en parallèle 512 fois, sur 16 noeuds  $\times$  32 threads. Chaque instance du code interpole les densités, vitesses et gradients de vitesse le long de chaque morceau de ligne de visée présente dans la tranche traitée. Une fois toutes les tranches traités, les morceaux de ligne de visée sont mis bout à bout grâce au code `merge_spectra.py`. Encore une fois, le code tourne en parallèle 512 fois : chaque instance du code traite toutes les lignes de visée correspondant aux quasars situés dans une même tranche. Le code lit alors tous les morceaux de spectre relatifs à ces quasars, puis les assemble. Lorsque les lignes de visée sont toutes reconstruites, la formule FGPA est appliquée afin d'obtenir le champ de transmission  $F$  pour chaque ligne de visée. La dernière étape consiste à regrouper le résultat de tous les chunks. Le module `merge_qso.py` permet de lire tous les quasars tirés dans chaque tranche de chaque chunk et de créer un catalogue global appelé `master.fits`. Une fois le catalogue construit, le code `dla_saclay.py` tire les HCD le long de chaque ligne de visée. Le code est tourné sur les sept chunks en parallèle. Puis le module `merge_dla.py`, comme pour les quasars, permet de regrouper tous les HCD tirés et de construire le catalogue global `master_DLA.fits`. Enfin le code `make_transmissions.py` permet de mettre les fichiers contenant les forêts au bon format : Les forêts sont regroupées par pixel HEALPix (GORSKI et al. 2004) dans des fichiers FITS, puis ces fichiers FITS sont regroupés par 100 selon leur pixel HEALPix :

`n/pix/transmission-nside-pix.fits.gz`,

où `nside` = 16 est la résolution du schéma HEALPix utilisé, `pix` donne le numéro du pixel HEALPix, et `n` est le résultat de la division euclidienne de `pix` par 100. Cette dernière étape est effectuée sur un noeud, `make_transmissions.py` étant lancé en parallèle sur 64 sous-échantillons des pixels HEALPix.

Comme expliqué à la fin de la section 1.5.4, le calcul de l'auto corrélation des traceurs objets, tels les quasars et les HCD, nécessite l'estimateur de Landy-Szalay (décrit dans la section 5.1.5). Afin d'utiliser cette estimateur, nous avons besoin de catalogues de quasars et de HCD qui suivent une distribution aléatoire. Les distributions issues des vrais catalogues sont ensuite comparées aux distributions issues des catalogues aléatoires. Pour construire ces catalogues aléatoires de quasars et HCD, nous utilisons les codes `draw_qso.py` et `dla_saclay.py`. Une option nous permet de tirer les objets sans tenir compte de la densité dans chaque voxel ou pixel. Les quasars sont donc tirés uniformément dans la boîte. Les HCD sont tirés uniformément le long de chaque ligne de visée, avec  $z_{\text{HCD}} < z_{\text{QSO}}$  où  $z_{\text{QSO}}$  est le redshift du quasar hôte. Dans les deux cas, nous tirons plus de quasars et de HCD pour les distributions aléatoires que pour les vrais catalogues, afin que la statistique des fonctions de corrélation soit limitée uniquement par les vrais catalogues. A la fin, les catalogues aléatoires de quasars et de HCD contiennent respectivement environ 10 et 3 fois plus d'objets que les vrais catalogues.

Pour les analyses Ly $\alpha$  d'eBOSS et de DESI, nous avons décidé de produire 100 réalisations des mocks, afin d'avoir suffisamment de statistique pour étudier finement les potentielles systématiques. Nous avons organisé la production de ces 100 réalisations en deux étapes. Premièrement, nous avons effectué la *pré-production*. Cette étape consiste à créer les boîtes contenant les différents GRF, puis à

tirer les quasars et enfin interpoler la densité, les vitesses et gradients de vitesse, le long de chaque ligne de visée. Ceci correspond aux codes `make_boxes.py`, `draw_qso.py` et `make_spectra.py`. Cette étape est la plus coûteuse en temps de calcul : environ 43 heures CPU sur un noeud de Cori pour produire les sept chunks d'une réalisation. L'ensemble des fichiers temporaires propres à une réalisation pré-produite représente  $\sim 7 \times 550$  Go sur disque. La majorité de cet espace disque est pris par les différentes boîtes. Afin de réduire cet espace disque, la seule boîte que nous sauvons est la boîte contenant  $\delta_k$ . A partir de cette boîte  $\delta_k$ , nous pouvons reconstruire toutes les autres boîtes, puis reconstruire le relevé de quasars et les lignes de visée. Ceci nous permet de vérifier la production des réalisations. Ainsi, pour chaque réalisation pré-produite, nous stockons les lignes de visée interpolées et la boîte  $\delta_k$ , ce qui représente  $\sim 340$  Go sur disque.

Une certain nombre de problèmes informatiques, liés au centre de calcul NERSC, ont ralenti la phase de pré-production. Le principal problème venait du temps de lecture et d'écriture, qui par moment pouvait être multiplié par un facteur cent. Plus de trente secondes étaient parfois nécessaires pour accéder à un simple fichier. Ce problème a été identifié comme venant du transfert des *meta data* sur le centre de calcul. Le code le plus affecté est `make_boxes.py`, car il écrit énormément de fichiers différents, correspondant aux différents slices des boîtes, et destinés à être lus par les codes `draw_qso.py` et `make_spectra.py`. Pour palier ce problème, nous avons décidé d'essayer de faire tourner les codes sur les nœuds de Cori appelés *Burst Buffer*. Ces nœuds possèdent des disques SSD (Solid-State drive), ce qui permet une lecture et une écriture très rapide. Une fois les codes exécutés, les données produites sont déplacées sur les disques durs habituels. Les nœuds Burst Buffer ont permis d'accélérer l'exécution du code `make_boxes.py` par un facteur  $\sim 2$ , passant d'environ deux heures à une heure seulement. Une vingtaine de réalisation ont été produites en utilisant ces nœuds. Cependant, ils sont devenus instables au cours de la production. D'autre part, les problèmes liés aux transferts des meta data avaient été stabilisés entre temps. Nous sommes donc retournés à l'utilisation des nœuds classiques de Cori pour finir la production. Pour les 40 dernières réalisations, le temps d'exécution de `make_boxes.py` variait entre trois et quatre heures.

Le temps pris par les six mois nécessaires à la production<sup>1</sup> nous a permis de choisir sur quelle modélisation des données nous voulions ajuster les paramètres Ly $\alpha$ . En effet, la modélisation des HCD dans les données est complexe et mal comprise. Les paramètres  $b_{\text{HCD}}$  et  $\beta_{\text{HCD}}$  sont très corrélés avec ceux du Ly $\alpha$ . Selon les modélisations, les paramètres Ly $\alpha$  obtenus grâce à l'ajustement des données ne sont pas les mêmes. L'étude de ces modélisations est présentée dans la section 6.2. Après avoir bien défini les paramètres Ly $\alpha$  que nous visons, nous avons ajusté les paramètres de la formule FGPA afin d'obtenir les paramètres Ly $\alpha$  souhaités dans les mocks. Ceci est expliqué dans la section suivante. Une fois ces paramètres ajustés et la phase de pré-production terminée, nous avons mené la phase de *post-production*. Cette phase consiste à créer les spectres d'absorption à partir des densités interpolées le long des lignes de visée, puis à regrouper tous les fichiers de sortie afin de les mettre au format décrit précédemment. Cette étape est beaucoup plus rapide, elle prend l'équivalent d'environ 6 heures CPU sur un nœud de Cori par réalisation. Une fois la production complète effectuée, la place sur disque d'une réalisation, sans compter les fichiers temporaires, correspond à environ 15 Go.

---

1. Dans de bonnes conditions, le temps de production de 100 réalisations est estimé à 4 à 6 semaines.

## 4.4 Ajustement des paramètres

Comme expliqué dans la section 4.2.4, le champ d'absorption Ly $\alpha$  est construit à partir des boîtes  $\delta_l$ ,  $\delta_s$  et  $\eta_{\parallel}$  grâce à la formule FGPA. Contrairement aux quasars pour lesquels le biais est choisi, les paramètres physiques du champ d'absorption Ly $\alpha$  simulé dépendent des quatre paramètres  $a(z)$ ,  $b(z)$ ,  $c(z)$  et  $P_s(z)$  utilisés dans l'équation 4.40. Nous décrivons dans cette section comment ajuster ces paramètres (paramètres FGPA dans la suite) afin d'obtenir les bons  $b_{\text{Ly}\alpha}(z)$ ,  $\beta_{\text{Ly}\alpha}(z)$ ,  $\bar{F}(z)$  et  $P^{\text{1D}}(k, z)$  (paramètres Ly $\alpha$  dans la suite). Nous nous servons de la modélisation des données décrite dans le chapitre 6.1.1 comme référence pour les paramètres Ly $\alpha$ .

Afin d'ajuster ces paramètres FGPA, la méthode standard est de générer un mock avec un jeu de paramètres FGPA, calculer la fonction de corrélation du Ly $\alpha$ , ajuster cette fonction de corrélation et mesurer les paramètres Ly $\alpha$ . Puis une fois ces paramètres mesurés, itérer sur les paramètres FGPA afin de nous rapprocher des paramètres Ly $\alpha$  visés. Cependant cette méthode est très coûteuse, car elle nécessite de générer des spectres et de produire la fonction de corrélation à chaque itération. Pour accélérer la procédure d'ajustement, nous tirons profit du fait que nos mocks possèdent une fonction de corrélation prédictible. Ainsi, au lieu de générer des spectres et calculer la fonction de corrélation sur ces derniers à chaque itération, nous calculons la prédiction (détaillée en section 4.2.4) puis nous générerons une pseudo-fonction de corrélation qui suit cette prédiction. Nous ajustons directement cette pseudo-fonction de corrélation afin de mesurer les paramètres Ly $\alpha$  correspondant au jeu de paramètres FGPA utilisés. Cet ajustement est fait pour les cinq valeurs du redshift  $z_1 = 1,8$ ;  $z_2 = 2,2$ ;  $z_3 = 2,6$ ;  $z_4 = 3,0$  et  $z_5 = 3,6$ .

Dans les lignes qui suivent, nous expliquons comment, pour chaque valeur du redshift, nous choisissons le jeu de paramètres FGPA initial pour générer la prédiction à la première itération. Premièrement, nous choisissons le paramètre  $c$ . Comme expliqué dans la section 4.2.4, le champ  $\delta_s$  ne participe pas à la corrélation à trois dimension et le champ  $g$  suit le spectre de puissance de l'équation 4.42. Le paramètre RSD du champ  $g$  vaut donc  $\beta_g = c$ , et donc, en première approximation, le paramètre RSD du champ  $F$  vaut  $c$  aussi. Pour chaque valeur du redshift, nous choisissons alors  $c(z) = \beta_{\text{Ly}\alpha}(z)$ .

En ce qui concerne les paramètres  $a$  et  $b$ , ils sont ajustés afin de retrouver les bons  $\bar{F}(z)$  et  $P^{\text{1D}}(k, z)$ . Pour ce faire, nous calculons la densité de probabilité de  $F$  en fonction de  $a$  et  $b$  (nous nous plaçons à  $\mu = 0$ , ce qui revient à prendre  $c = 0$ ). Cette densité de probabilité nous donne donc accès à  $\bar{F}$  et  $\bar{F}^2$  en fonction de  $a$  et  $b$ . Nous ajustons alors  $a$  et  $b$  afin d'avoir à la fois  $\bar{F}$  et  $\bar{F}^2$  en accord avec les données. La fraction de flux transmis moyenne  $\bar{F}$  que nous utilisons est paramétrisée par

$$\bar{F}(z) = \exp(-0.00211(1+z)^{3.7}). \quad (4.57)$$

Elle est représentée sur la figure 4.17 par une ligne tiretée bleue. Les points noirs sur cette figure représentent la mesure de  $\bar{F}(z)$  faite par FAUCHER-GIGUERE et al. (2008). La paramétrisation que nous utilisons (équation 4.57) est en bon accord avec les données dans la gamme  $1,8 < z < 3,6$ . En ce qui concerne le terme  $\bar{F}^2$ , nous le calculons pour chaque valeur du redshift. Il est donné par l'intégrale du  $P^{\text{1D}}$ . Pour calculer cette intégrale, nous utilisons  $P_{\text{modèle}}^{\text{1D}}$ , décrit dans les paragraphes suivants.

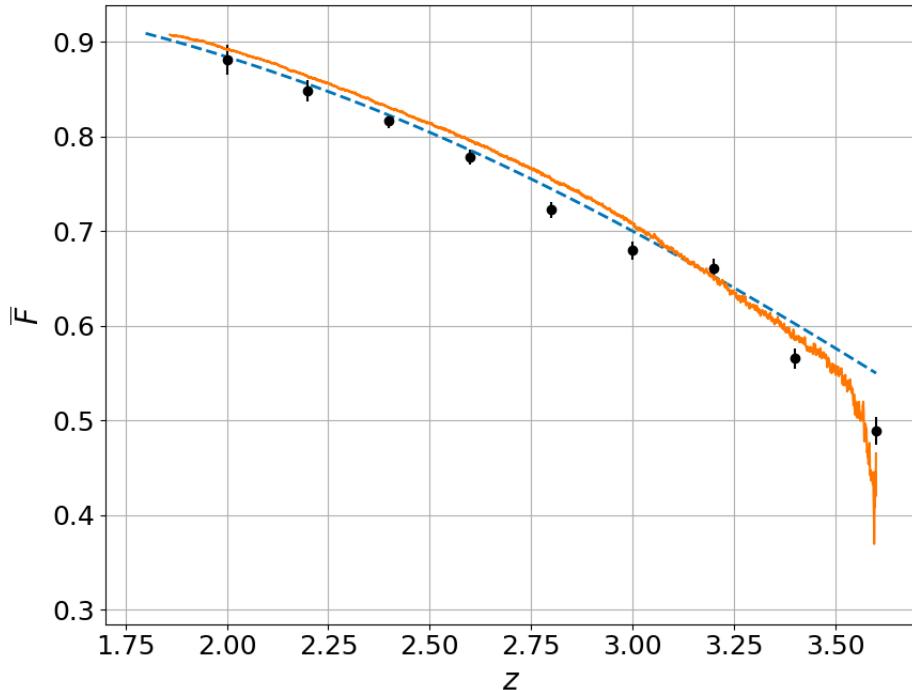


FIGURE 4.17 – Fraction de flux transmis moyenne en fonction du redshift. Les points noirs donnent la mesure faite par FAUCHER-GIGUERE et al. (2008). La ligne tiretée bleue donne la paramétrisation de  $\bar{F}(z)$  que nous utilisons dans les mocks (équation 4.57). La courbe orange donne la mesure faite sur les mocks.

Une fois les paramètres  $a$ ,  $b$  et  $c$  choisis, nous déterminons le spectre de puissance  $P_s$  à appliquer à  $\delta_{k,s}$  afin d'obtenir le bon  $P_{\text{mock}}^{\text{1D}}$ . Nous avons besoin de  $P_s$  car il nous faut connaître  $\sigma_g$ , qui lui-même requiert  $\sigma_s$ , pour chaque valeur du redshift afin de calculer la prédiction. La variance du champ  $g = \delta_l + \delta_s + c(z)\eta_{\parallel}$  est donnée par

$$\sigma_g^2(z) = \langle \delta_l^2 \rangle + \langle \delta_s^2 \rangle + c^2(z) \langle \eta_{\parallel}^2 \rangle + c(z) \left( \langle \delta_l \eta_{\parallel} \rangle - \langle \delta_l \rangle \langle \eta_{\parallel} \rangle \right). \quad (4.58)$$

Le terme entre parenthèses est la covariance des champs  $\delta_l$  et  $\eta_{\parallel}$ . La variance  $\sigma_s$  du champ  $\delta_s$  est reliée à son spectre de puissance  $P_s$  par l'équation 4.37. Dans le but de calculer  $\sigma_s$  puis la prédiction, nous commençons donc par construire le spectre de puissance  $P_{\text{modèle}}^{\text{1D}}$  sur lequel le  $P_{\text{mock}}^{\text{1D}}$  des mocks sera ajusté. Tout d'abord, le spectre de puissance à une dimension provenant des données DR12 dans lequel les oscillations dues au silicium ont été retirées (Nathalie Palanque-Delabrouille, communication privée) est ajusté dans la gamme  $0,2 < k < 2.0 h \text{ Mpc}^{-1}$ . La fonction ajustée aux données est définie comme

$$f(k) = \exp\left(-ak + b + \frac{c}{k+d}\right), \quad (4.59)$$

où  $a$ ,  $b$ ,  $c$  et  $d$  sont quatre paramètres à ajuster. La forme choisie pour  $f$  permet d'avoir  $\log(P_{\text{modèle}}^{\text{1D}})$  linéaire à grand  $k$ , en accord avec ce qui est mesuré dans les simulations hydrodynamiques (ARINYO-I-PRATS et al. 2015). Afin de construire  $\delta_s$ , nous avons besoin de calculer  $P_s$  jusqu'à  $k_{Ny}$ . Le résultat de l'ajustement sur les données est donc extrapolé de  $k = 2,0$  jusqu'à  $k = 20 h \text{ Mpc}^{-1}$ . Pour les  $k$  plus petits que  $0,2 h \text{ Mpc}^{-1}$ , nous calculons le  $P^{\text{1D}}$  correspondant au modèle de  $P^{\text{3D}}$  décrit dans

ARINYO-I-PRATS et al. (2015) et ajusté sur des simulations hydrodynamiques. Ce modèle est défini comme

$$P(k) = b_{\text{Ly}\alpha}^2 (1 + \beta_{\text{Ly}\alpha} \mu_k^2)^2 P_{\text{L}}(k) D(k, \mu), \quad (4.60)$$

où  $P_{\text{L}}$  est le spectre de puissance linéaire, donné par Camb, et  $D(k, \mu)$  représente les déviations par rapport à la théorie linéaire. Le terme  $D(k, \mu)$  tend donc vers 1 à petit  $k$ . La forme choisie dans ARINYO-I-PRATS et al. (2015) diffère de celle utilisée dans McDONALD (2003). Cette nouvelle forme permet d'obtenir le bon comportement à petit  $k$ . Aussi, elle nécessite l'ajustement de moins de paramètres.  $D(k, \mu)$  est donc défini comme

$$D(k, \mu) = \exp \left[ \left( q_1 \Delta^2(k) + q_2 \Delta^4(k) \right) \left( 1 - \left( \frac{k}{k_v} \right)^{a_v} \mu^{b_v} \right) - \left( \frac{k}{k_p} \right)^2 \right], \quad (4.61)$$

avec

$$\Delta^2(k) = \frac{1}{2\pi^2} k^3 P_{\text{L}}(k). \quad (4.62)$$

Les termes  $q_1 \Delta^2(k)$  et  $q_2 \Delta^4(k)$  représentent l'augmentation de puissance aux petites échelles due aux non-linéarités. L'ajustement que nous utilisons force  $q_2 = 0$ . Les autres paramètres ajustés sont donnés dans la section « Planck » de la table 7 de ARINYO-I-PRATS et al. (2015). Puis, pour chaque valeur du redshift, la forme ajustée aux données et définie dans l'équation 4.59 est prolongée de  $k = 0,2 h \text{ Mpc}^{-1}$  jusqu'à  $k = 0$  par le modèle donné dans l'équation 4.61. Pour ce faire, nous calculons le modèle à  $z = 2,4$ , puis nous le multiplions par une constante, dépendant de la valeur du redshift, de façon à ce que le prolongement en  $k = 0,2 h \text{ Mpc}^{-1}$  soit continu. La figure 4.18 montre le modèle ainsi construit. Les données y sont superposées (Nathalie Palanque-Delabrouille, communication privée). Pour le bin  $z = 1,8$ , aucune donnée Ly $\alpha$  n'est disponible. Nous extrapolons donc le modèle à  $z = 2,2$ . Nous considérons que la forme du modèle reste la même. Nous considérons aussi que l'évolution en redshift de  $z = 2,4$  à  $z = 2,2$  est la même jusqu'à  $z = 1,8$ . Le modèle à  $z = 1,8$  est donc donné par

$$P_{\text{modèle}}^{\text{1D}}(k, z = 1,8) = P_{\text{modèle}}^{\text{1D}}(k, z = 2,2) \left( \frac{P_{\text{modèle}}^{\text{1D}}(k, z = 2,2)}{P_{\text{modèle}}^{\text{1D}}(k, z = 2,4)} \right)^2. \quad (4.63)$$

Une fois le spectre de puissance modèle défini, nous estimons le spectre de puissance à une dimension  $P_l^{\text{1D}}$  du champ interpolé  $\delta_l + c\eta_{\parallel}$ . Nous partons du spectre de puissance fourni par Camb. Nous multiplions ce spectre de puissance par le terme de kaiser  $(1 + c\mu_k^2)^2$  puis par le terme  $W^2$  représentant l'effet du lissage gaussien (équation 4.30). Enfin, nous calculons  $P_l^{\text{1D}}$  grâce à l'équation 4.33, en nous restreignant aux  $k < k_{\text{max}}$ . Une première estimation du spectre de puissance à une dimension  $P_s^0$  à appliquer à  $\delta_{k,s}$  afin d'obtenir le bon  $P_{\text{mock}}^{\text{1D}}$  est donnée par

$$P_s^0(k) = P_{\text{modèle}}^{\text{1D}}(k) - P_l^{\text{1D}}(k). \quad (4.64)$$

Le spectre de puissance  $P_s^0$  ainsi construit nous permet d'obtenir un  $P_{\text{mock}}^{\text{1D}}$  convenable. Cependant, il reste des différences avec le spectre de puissance modèle. Afin de corriger ces différences, nous ajustons itérativement la forme de  $P_s$ . A chaque itération  $n$ , nous commençons par générer  $\delta_s$  selon  $P_s^n$ . Puis

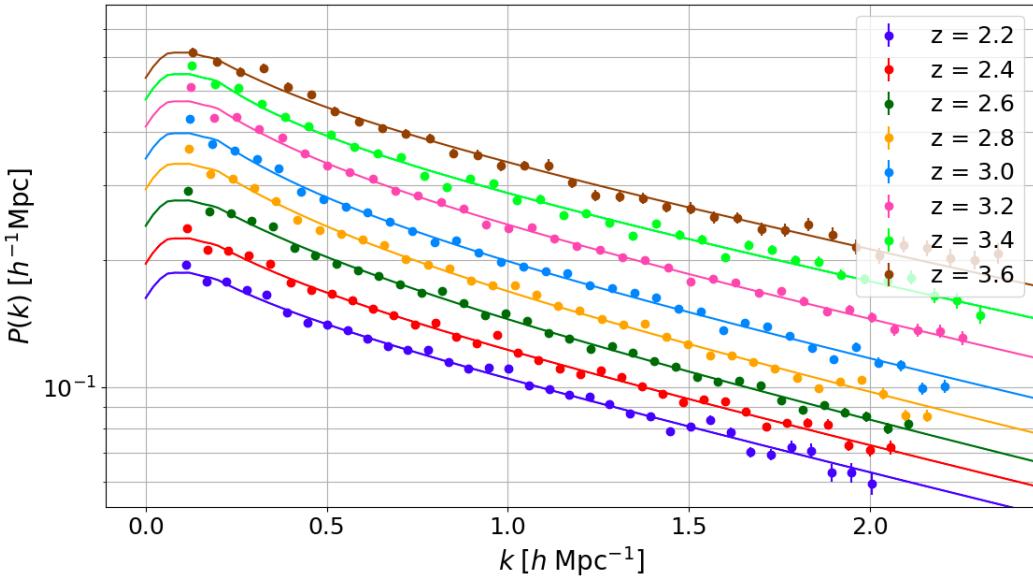


FIGURE 4.18 – Modèle du spectre de puissance à une dimension, pour différentes valeurs du redshift  $z$ , utilisé dans les mocks. Le modèle, donné par les lignes continues, est ajusté sur les données (Nathalie Palanque-Delabrouille, communication privée), représentées par les points. Ce modèle est décrit dans le texte.

nous calculons le  $P_{\text{mock}}^{1\text{D}}$ , correspondant à ce  $P_s^n$ . Le  $P_s^{n+1}$  de l’itération suivante est alors donné par

$$P_s^{n+1}(k) = P_s^n(k) \frac{P_{\text{modèle}}^{1\text{D}}(k)}{P_{\text{mock}}^{1\text{D}}(k)}. \quad (4.65)$$

Quelques itérations sont suffisantes pour obtenir un  $P_s$  qui donne un  $P_{\text{mock}}^{1\text{D}}$  en accord avec les données. Nous en effectuons dix pour chaque valeur du redshift.

A ce stade, nous disposons des informations nécessaires pour calculer la prédiction. Afin de mesurer les paramètres Ly $\alpha$  prédictifs, nous générerons une pseudo-fonction de corrélation qui suit la prédiction calculée avec le jeu de paramètres. Nous ajustons ensuite cette pseudo-fonction de corrélation avec le code **picca**, comme nous le ferions avec les mocks. Le code **picca** ajuste les paramètres  $b_{\eta, \text{Ly}\alpha}$  et  $\beta_{\text{Ly}\alpha}$ . Le biais  $b_{\text{Ly}\alpha}$  est relié à ces deux paramètres par la relation

$$b_{\text{Ly}\alpha} = \frac{b_{\eta, \text{Ly}\alpha} f}{\beta_{\text{Ly}\alpha}}, \quad (4.66)$$

où  $f$  est le taux de croissance des structures. Dans la section 6.1.2, nous verrons que les paramètres Ly $\alpha$  sont corrélés entre eux. Ainsi, afin de comparer au mieux la prédiction avec les données, plutôt que de mesurer les paramètres  $b_{\eta, \text{Ly}\alpha}$  ou  $b_{\text{Ly}\alpha}$  et  $\beta_{\text{Ly}\alpha}$ , nous mesurons les paramètres  $b_{\text{eff}, \text{Ly}\alpha}$  et  $\beta_{\text{Ly}\alpha}$ . Le biais effectif  $b_{\text{eff}, \text{Ly}\alpha}$  est relié au biais et au paramètre RSD du Ly $\alpha$  par

$$b_{\text{eff}, \text{Ly}\alpha} = b_{\text{Ly}\alpha} \sqrt{1 + \frac{2}{3} \beta_{\text{Ly}\alpha} + \frac{1}{5} \beta_{\text{Ly}\alpha}^2}. \quad (4.67)$$

Il est sensible à l’amplitude de la fonction de corrélation, et est moins corrélé avec  $\beta_{\text{Ly}\alpha}$  que l’est

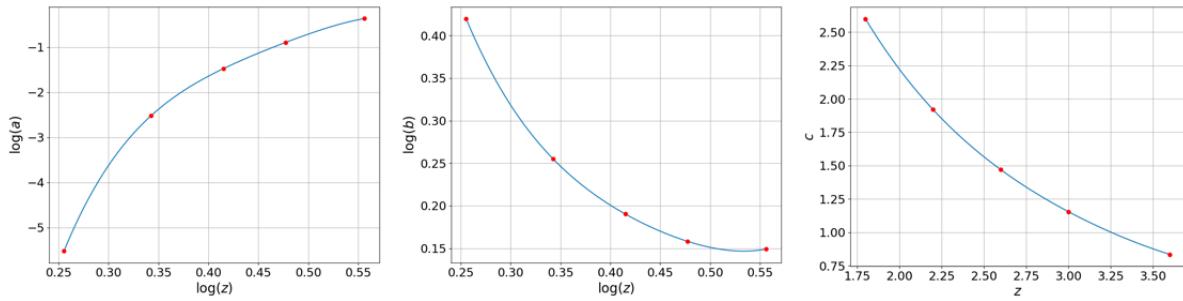


FIGURE 4.19 – Résultat de la procédure d’ajustement des paramètres FGPA.

$b_{\eta,\text{Ly}\alpha}$  ou  $b_{\text{Ly}\alpha}$ . Enfin, nous mesurons aussi  $\bar{F}$  dans la prédition, et nous comparons sa valeur à la paramétrisation définie dans l’équation 4.57.

L’estimation initiale que nous faisons des paramètres  $b$  et  $c$  donne des valeurs de  $\beta_{\text{Ly}\alpha}$  et  $\bar{F}$  proches de ce qui est mesuré dans les données. Cependant, la valeur de  $a$  obtenue est trop grande : les paramètres  $a$  et  $b$  sont estimés à partir des mesures de  $P^{1D}$  et  $\bar{F}$  dans les données. Comme nous le verrons dans la section 5.4.1, la présence de HCD dans les données a pour effet d’augmenter le biais effectif mesuré. Ainsi, l’estimation du paramètre  $a$  est faite avec un  $P^{1D}$  qui possède un biais effectif surestimé. Ceci résulte en une valeur de  $a$  trop grande qui produit un biais effectif trop grand dans nos mocks. Nous itérons alors sur les paramètres  $a$ ,  $b$  et  $c$  : nous diminuons  $a$ , et si besoin modifions légèrement  $b$  et  $c$  afin d’affiner les valeurs de  $\beta_{\text{Ly}\alpha}$  et  $\bar{F}$  des mocks. Puis, nous recalculons le nouveau  $P_s^{10}$  afin de générer la nouvelle prédition. Nous ajustons de nouveau la prédition avec `picca` et comparons les valeurs de  $b_{\text{eff,Ly}\alpha}$ ,  $\beta_{\text{Ly}\alpha}$  et  $\bar{F}$  mesurées aux données. Ces itérations sont faites jusqu’à obtenir des valeurs en accord avec les données.

Une fois que cette procédure itérative a convergé pour les cinq valeurs du redshift, nous interpolons pour avoir des fonctions continues des paramètres  $a(z)$ ,  $b(z)$  et  $c(z)$  sur  $z \in [1,8;3,6]$ . Pour  $a(z)$  et  $b(z)$ , nous calculons les polynômes de degré quatre de  $\log z$  qui passent par les cinq points  $\log a(z)$  et  $\log b(z)$ . Ce choix de fonction est fait dans le but d’éviter des valeurs négatives pour  $a(z)$ , ainsi que des points d’infexion. Pour le paramètre  $c(z)$ , nous calculons le polynôme de  $z$  qui passe par les cinq points  $c(z)$ . La figure 4.19 présente les interpolations ainsi obtenues. Les cinq spectres de puissance  $P_s(k)$  construits précédemment pour chaque valeur du redshift produisent des spectres de puissance  $P_{\text{mock}}^{1D}$  en très bon accord avec le modèle. Les cinq  $P_{\text{mock}}^{1D}$  obtenus pour chaque valeur du redshift sont présentés sur la figure 4.21. Les lignes continues donnent le modèle pour chaque valeur du redshift. Afin d’obtenir  $P_s(k)$  pour n’importe quel redshift, nous créons une grille  $(k, z)$  couvrant  $z \in [1,8;3,6]$  et  $k \in [0;20]h \text{ Mpc}^{-1}$ . Puis,  $P_s(k, z)$  est obtenu via une interpolation cubique sur cette grille. Du fait des très faibles valeurs de  $P_s$  à grand  $k$  et de l’interpolation cubique utilisée, certaines valeurs de l’interpolation sont négatives. Pour chaque redshift de la grille, nous forçons tous les pixels de l’interpolation de  $P_s$  à zéro pour tous les  $k > k_0$ , où  $k_0$  est le premier  $k$  pour lequel l’interpolation est nulle. La figure 4.20 montre l’interpolation de  $P_s(k)$  pour différentes valeurs de  $z$ . Enfin, nous mesurons la fraction de flux transmise moyenne  $\bar{F}$  en fonction du redshift sur une réalisation complète des mocks. Cette mesure est présentée en orange sur la figure 4.17. La ligne en pointillés bleus montre la paramétrisation définie dans l’équation 4.57. La transmission moyenne mesurée dans les mocks (orange) est en accord avec cette paramétrisation.

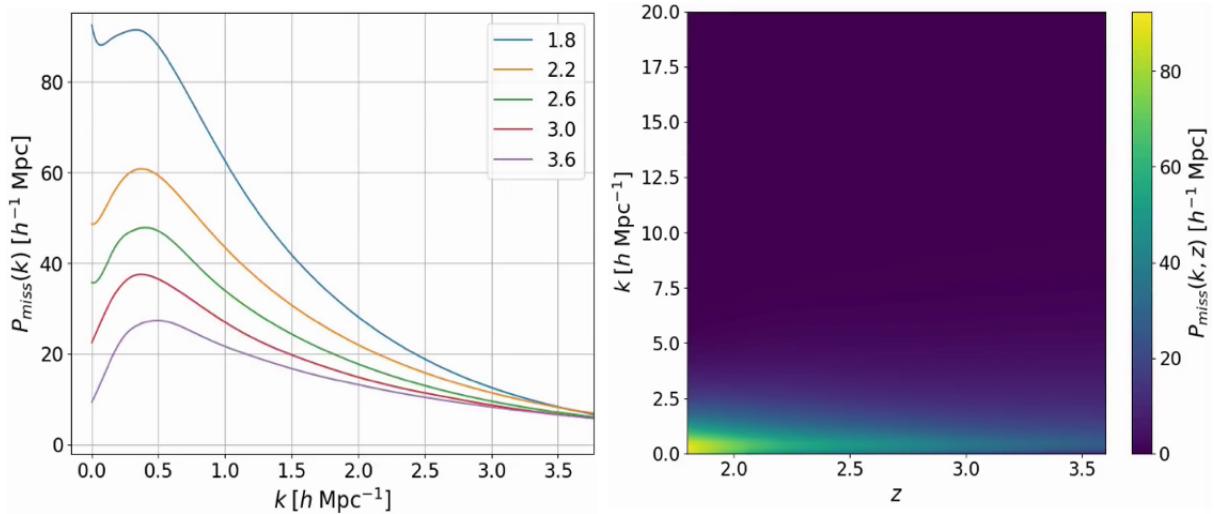


FIGURE 4.20 – Gauche : les cinq spectres de puissance  $P_s(k)$  ajustés pour chaque valeur du redshift. Droite : interpolation cubique de ces cinq spectres de puissance sur une grille  $(k, z)$ .

Nous pouvons noter que la mesure faite pour les redshifts proches de 3,6 est biaisée par le fait que seuls les pixels proches des quasars participent à cette mesure : la corrélation croisée Ly $\alpha$  $\times$ QSO a pour effet de réduire  $F$ .

Le résultat de la procédure d'ajustement décrite précédemment a été utilisé pour produire 50 réalisations indépendantes des mocks. Nous prévoyons de terminer la post-production des 50 autres réalisations pré-produites dans les prochains mois. Parmi les 50 réalisations produites, 30 sont analysées. Le chapitre 5 présente cette analyse.

## 4.5 Expansion des mocks

Les mocks décrits précédemment produisent un relevé de quasars avec, pour chaque quasar, une forêt d'absorption  $F$  variant entre 0 et 1. Afin de reproduire complètement les données et pouvoir simuler la chaîne d'analyse, nous devons ajouter un continuum à ces forêts, puis ajouter du bruit de mesure. Le code utilisé pour créer ces spectres synthétiques est le code `quickquasars`. Il fait parti du package `desisim`<sup>1</sup> et est décrit dans GONZÁLEZ-MORALES et AL (p.d.).

Le continuum des quasars est généré à l'aide du module `SIMQSO`, du package `desisim`. Le modèle utilisé est composé de plusieurs lois de puissance. Ces lois de puissance sont mises bout à bout pour créer le continuum. Les raies d'émission sont ajoutées en utilisant un modèle fondé sur les spectres des quasars observés par BOSS. Une option permet d'ajouter des BAL aux spectres de quasars simulés. Nous n'utilisons pas cette option car les spectres présentant un BAL sont retirés des données lors de l'analyse. Le code permet aussi d'ajouter des HCD et des métaux. Les HCD sont ajoutés dans les spectres aux positions indiquées par le catalogue de HCD que nous générerons (section 4.2.5). Pour chaque HCD, un profil de Voigt, dont la profondeur dépend de la densité de colonne, est créé puis ajouté au spectre. En ce qui concerne les métaux, ils sont ajoutés en utilisant la transmission du Ly $\alpha$  :

1. <https://github.com/desihub/desisim>

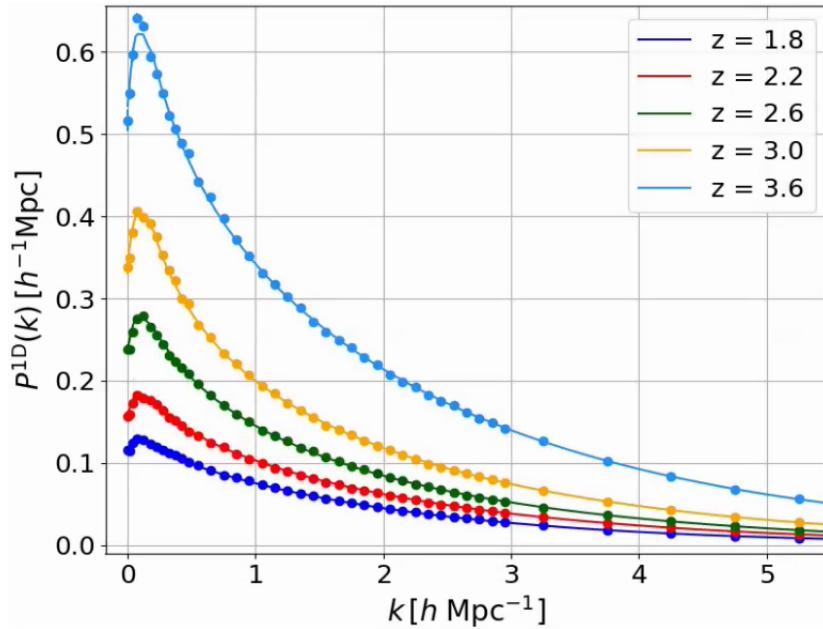


FIGURE 4.21 – Les spectres de puissance  $P_{\text{mock}}^{\text{1D}}$  obtenus pour chaque valeur du redshift à la fin de la procédure d'ajustement. Les lignes continues donnent le modèle pour chaque valeur du redshift. Les mocks sont en train bon accord avec le modèle.

pour chaque métal, la forêt Ly $\alpha$  est décalée en longueurs d'onde observées par un facteur  $\lambda_m/\lambda_{\text{Ly}\alpha}$ , où  $\lambda_m$  est la longueur d'onde du métal considéré. Puis, l'absorption est modifiée : le flux de chaque absorption est corrigé par un facteur  $A_m$ , où  $A_m$  relie la profondeur optique de la raie d'absorption du métal à la profondeur optique du Ly $\alpha$  :  $\tau_m = A_m \tau_{\text{Ly}\alpha}$ . Le paramètre  $A_m$  est ajusté afin d'obtenir le bon biais pour chaque métal. Enfin, le code ajoute les spécificités propres à l'instrument ou au relevé (eBOSS ou DESI), comme par exemple la réponse des spectrographes, le temps d'exposition pour chaque spectre, ou encore la distribution en magnitude des quasars. La figure 4.22 présente le spectre d'un quasar simulé. Un agrandissement de la forêt Ly $\alpha$  de ce spectre est visible sur la figure 4.10.

Nous produisons plusieurs versions différentes des mocks :

- les *raw mocks*, ou mocks *eboss-raw*,
- les mocks *eboss-0.0*,
- les mocks *eboss-0.2*,
- les mocks *eboss-0.3*.

Les raw mocks sont les mocks bruts, sans l'utilisation de `quickquasars`. Ces mocks sont analysés en estimant les fonctions de corrélation directement à partir des fichiers de transmission (voir la section 5.5). Ceci nous permet de vérifier la construction des mocks et les différentes fonctions de corrélation, sans que celles-ci soient affectées par l'ajustement du continuum ou les effets astrophysiques et instrumentaux. Les versions *eboss-0.0*, *eboss-0.2* et *eboss-0.3* sont les versions des mocks après avoir utilisé `quickquasars` pour produire les spectres synthétiques. La version *eboss-0.0* correspond à l'ajout du continuum et du bruit. La version *eboss-0.2* est la version *eboss-0.0* à laquelle les HCD sont ajoutés.

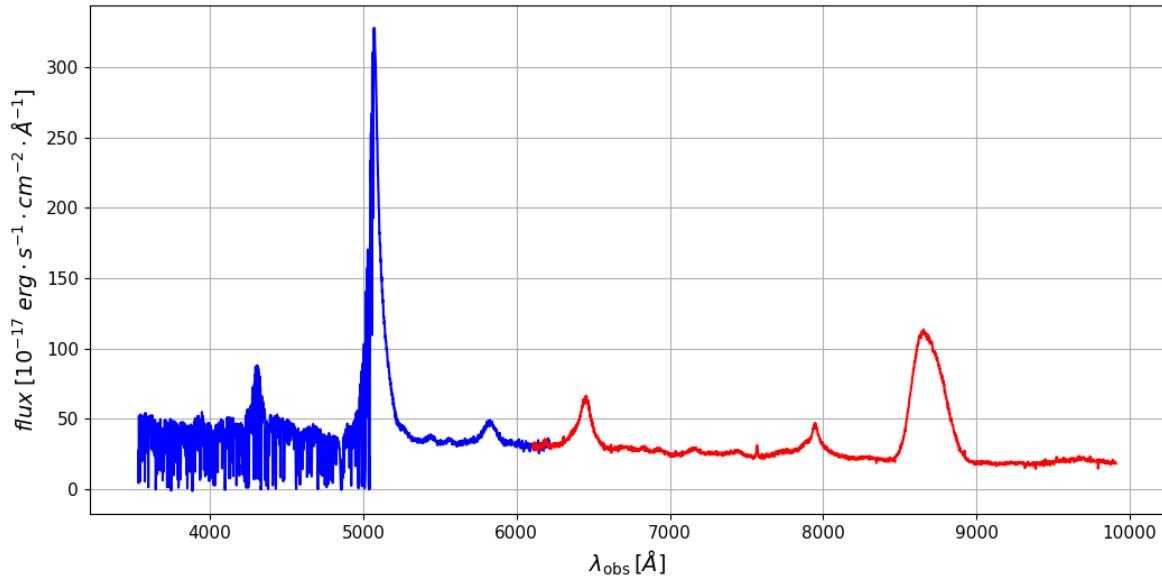


FIGURE 4.22 – Spectre d’un quasar synthétique généré par `quickquasars`. Le quasar se trouve à un redshift  $z = 3,167$ . La partie bleue du spectre correspond au spectrographe bleu, la partie rouge correspond au spectrographe rouge. Pour des questions de visibilité, nous avons réduit la zone de recouvrement des deux spectrographes dans cette figure.

La version eboss-0.3 ajoute le continuum, le bruit, les HCD ainsi que les métaux. Nous présentons l’analyse de ces différentes versions des mocks dans le chapitre suivant.

Un point important à noter est que, pour chaque réalisation, les quatre versions présentées ci-dessus sont produites avec le même sous-échantillon de quasars. Ceci facilite la comparaison des différentes versions.

# 5

## Validation des mocks

---

### Sommaire

<b>5.1</b>	<b>Les estimateurs</b>	<b>104</b>
5.1.1	L'auto-corrélation $\text{Ly}\alpha \times \text{Ly}\alpha$	105
5.1.2	La corrélation croisée $\text{Ly}\alpha \times \text{QSO}$	106
5.1.3	Le spectre de puissance à une dimension	106
5.1.4	La fonction de corrélation à une dimension	107
5.1.5	L'auto-corrélation $\text{QSO} \times \text{QSO}$	107
<b>5.2</b>	<b>Les matrices de distorsion</b>	<b>108</b>
<b>5.3</b>	<b>Les matrices de covariance</b>	<b>109</b>
<b>5.4</b>	<b>Modélisation des fonctions de corrélation</b>	<b>110</b>
5.4.1	Modélisation des données	110
5.4.2	Modélisation des mocks	116
<b>5.5</b>	<b>Analyse des mocks</b>	<b>119</b>
5.5.1	Analyse des raw mocks	119
L'auto-corrélation $\text{Ly}\alpha \times \text{Ly}\alpha$	119	
La corrélation croisée $\text{Ly}\alpha \times \text{QSO}$	121	
Le spectre de puissance à une dimension	122	
La corrélation à une dimension $\text{Ly}\alpha \times \text{Ly}\alpha$	123	
L'auto-corrélation $\text{QSO} \times \text{QSO}$	125	
L'auto-corrélation $\text{HCD} \times \text{HCD}$	126	
5.5.2	Analyse des mocks eboss-0.0	128
L'auto-corrélation $\text{Ly}\alpha \times \text{Ly}\alpha$	129	
La corrélation croisée $\text{Ly}\alpha \times \text{QSO}$	129	
5.5.3	Analyse des mocks eboss-0.2	132
L'auto-corrélation $\text{Ly}\alpha \times \text{Ly}\alpha$	132	
La corrélation croisée $\text{Ly}\alpha \times \text{QSO}$	134	
5.5.4	Analyse des mocks eboss-0.3	136
La fonction de corrélation à une dimension	136	

---

Ce chapitre a pour vocation de présenter l'analyse menée sur les mocks, afin de valider leur construction et le choix des différents paramètres. Nous présentons d'abord les estimateurs des diverses fonctions de corrélation, puis les modèles ajustés sur celles-ci. Nous donnons premièrement le modèle ajusté sur les données et présenté dans DU MAS DES BOURBOUX et al. (2020), puis le modèle modifié ajusté sur les mocks. Enfin, nous présentons l'analyse des 30 réalisations de mocks produites. Le résultat de l'analyse des données, utilisée pour choisir les paramètres Ly $\alpha$  des mocks, est présentée dans le chapitre suivant. Toutes ces analyses sont produites avec le code `picca`<sup>1</sup>. Ce code permet de calculer les champs  $\delta$ , les fonctions de corrélation, les matrices de distorsion, les matrices de covariance ainsi que de réaliser l'ajustement du modèle.

## 5.1 Les estimateurs

Nous présentons ici les estimateurs utilisés pour calculer la fonction d'auto-corrélation du Ly $\alpha$ , la fonction de corrélation croisée Ly $\alpha$ -QSO, ainsi que la fonction d'auto-corrélation d'objets ponctuels tels les quasars. Dans le cas des objets ponctuels, il est nécessaire de produire un catalogue aléatoire d'objets qui prend en compte la complétude<sup>2</sup> du relevé spectrométrique. Il est aussi nécessaire d'estimer l'inhomogénéité de la sélection des cibles, afin de pondérer chaque quasar par un poids qui vise à la corriger. Pour le calcul des fonctions de corrélation impliquant le Ly $\alpha$ , il n'est pas utile d'avoir un catalogue aléatoire. En effet, le champ  $\delta_F$  nous donne directement accès, à un biais près, au contraste de densité. Nous pouvons donc calculer la fonction de corrélation avec une formule du type de l'équation 1.42, ce qui n'est pas possible dans le cas des objets ponctuels, et nous n'avons pas besoin de tenir compte de l'inhomogénéité de la sélection des cibles. Les poids servent donc uniquement à maximiser le rapport signal sur bruit.

Pour le calcul du champ  $\delta_F$ , nous distinguons deux cas : le cas où nous analysons les *raw mocks* et le cas où nous analysons les *cooked mock* ou les données. Les raw mocks (les mocks bruts, non préparés) désignent les mocks avant d'avoir tourné le code `quickquasars`. Les fonctions de corrélation sont alors calculées en utilisant directement la fraction de flux transmise  $F$  donnée dans les fichiers de transmissions. Étant donné que pour chaque forêt, nous avons directement accès à  $F$ , le champ  $\delta_F$  est donné par

$$\delta_F = \frac{F}{\bar{F}(z)} - 1. \quad (5.1)$$

Les cooked mocks quant à eux désignent les mocks après avoir appliqué `quickquasars` à chaque forêt. Dans ce cas, comme pour les données, la fraction de flux transmise  $F$  n'est pas accessible. Il faut alors utiliser la procédure d'ajustement du continuum décrite dans la section 3.4.1. Une fois le champ  $\delta_F$  calculé, nous pouvons calculer les différentes fonctions de corrélation.

---

1. Le code est en accès libre sur GitHub : <https://github.com/igmhub/picca>.

2. La complétude donne la proportion de cibles observées en spectroscopie par rapport au nombre de cibles détectées avec le relevé photométrique.

### 5.1.1 L'auto-corrélation Ly $\alpha$ $\times$ Ly $\alpha$

Comme expliqué dans le chapitre d'introduction, la fonction d'auto-corrélation du champ d'absorption  $F$  du Ly $\alpha$  est définie comme

$$\xi_{\text{Ly}\alpha}(\vec{r}) = \langle \delta_F(\vec{r}') \delta_F(\vec{r}' + \vec{r}') \rangle . \quad (5.2)$$

Nous pouvons utiliser comme estimateur

$$\hat{\xi}(\vec{r}) = \langle \delta_i \delta_j \rangle , \quad (5.3)$$

où  $\langle \cdot \rangle$  désigne la moyenne sur tous les pixels  $i$  et  $j$  qui vérifient  $\vec{r}_{ij} = \vec{r}$ . Afin d'estimer  $\xi_{\text{Ly}\alpha}$ , une grille en  $(r_{\parallel}, r_{\perp})$  est créée. Les bins mesurent  $4 h^{-1}$  Mpc dans chaque direction. Cette taille de bin est choisie de façon à avoir suffisamment de bins pour résoudre correctement la région du pic BAO, mais aussi de façon à ne pas avoir trop de bins pour le calcul de la matrice de covariance. La taille de bin choisie et la gamme en  $r$  utilisée produisent déjà une matrice de covariance de  $2500 \times 2500$  bins. Une fois cette grille construite, la séparation de chaque paires  $(\Delta\theta, \Delta z)$  est transformée en distance comobile afin d'obtenir la séparation  $(r_{\parallel}, r_{\perp})$  :

$$\begin{cases} r_{\parallel} = [D_C(z_i) - D_C(z_j)] \cos\left(\frac{\Delta\theta}{2}\right) , \\ r_{\perp} = [D_M(z_i) - D_M(z_j)] \sin\left(\frac{\Delta\theta}{2}\right) , \end{cases} \quad (5.4)$$

où  $D_C$  est la distance comobile le long de la ligne de visée, et  $D_M$  la distance comobile transverse (voir section 1.2.2, paragraphe *Les distances*).  $z_i$  et  $z_j$  sont les redshifts des pixels  $i$  et  $j$ . Puis, pour chaque bin  $A$  de la grille en  $(r_{\parallel}, r_{\perp})$ , toutes les paires de pixels dont la distance de séparation se trouve dans ce bin  $A$  sont considérées. La fonction de corrélation dans ce bin est alors donnée par

$$\hat{\xi}_A = \frac{\sum_{(i,j) \in A} w_i w_j \delta_i \delta_j}{\sum_{(i,j) \in A} w_i w_j} , \quad (5.5)$$

où  $w_i$  est le poids associé au pixel  $i$ . Ces poids sont définis dans l'équation 3.11. La correction de la dépendance en redshift du biais du Ly $\alpha$  permet de donner plus de poids aux pixels à grand redshift, où l'amplitude de la fonction de corrélation est plus importante. Le calcul de la fonction de corrélation est donc effectué à l'aide d'une double boucle sur les pixels. Afin de réduire le temps de calcul, la fonction de corrélation est calculée uniquement pour les paires de pixels pour lesquelles  $r_{\parallel}$  et  $r_{\perp}$  sont inclus dans  $[0; 200] h^{-1}$  Mpc. La fonction de corrélation est donc calculée dans  $50 \times 50 = 2500$  bins. Les paires formées par des pixels provenant de la même forêt sont exclues du calcul afin d'éviter que les erreurs sur l'ajustement du continuum biaissent la mesure de la fonction de corrélation.

L'analyse Ly $\alpha$  des données complète d'eBOSS (DU MAS DES BOURBOUX et al. 2020) utilise deux fonctions de corrélation distinctes : les fonctions de corrélation Ly $\alpha$ (Ly $\alpha$ ) $\times$ Ly $\alpha$ (Ly $\alpha$ ) et Ly $\alpha$ (Ly $\alpha$ ) $\times$ Ly $\alpha$ (Ly $\beta$ ). La région Ly $\beta$  ne fournit pas suffisamment de pixels pour pouvoir calculer la fonction de corrélation Ly $\alpha$ (Ly $\beta$ ) $\times$ Ly $\alpha$ (Ly $\beta$ ) et y détecter le pic BAO. De plus, le gain de statistique que représente cette

fonction de corrélation est négligeable. L'utilisation de ces deux fonctions de corrélation permet de mesurer la position du pic BAO avec une plus grande précision. Dans ce manuscrit, nous nous intéressons à la mesure de  $b_{\text{Ly}\alpha}$  et  $\beta_{\text{Ly}\alpha}$  afin de construire correctement nos mocks. Pour limiter les potentielles systématiques, nous considérons donc uniquement la fonction de corrélation  $\text{Ly}\alpha(\text{Ly}\alpha) \times \text{Ly}\alpha(\text{Ly}\alpha)$ . Le graphique de droite de la figure 3.4 présente les distributions pondérées en redshift des paires  $\text{Ly}\alpha(\text{Ly}\alpha) \times \text{Ly}\alpha(\text{Ly}\alpha)$  et  $\text{Ly}\alpha(\text{Ly}\alpha) \times \text{Ly}\alpha(\text{Ly}\beta)$ . L'analyse des données DR16 présentée dans ce manuscrit considère donc uniquement la distribution indiquée en orange.

### 5.1.2 La corrélation croisée $\text{Ly}\alpha \times \text{QSO}$

Nous donnons dans cette section l'estimateur de la fonction de corrélation croisée  $\text{Ly}\alpha \times \text{QSO}$ . De la même manière que précédemment, le calcul s'effectue dans des bins en  $(r_{\parallel}, r_{\perp})$ . L'estimateur de la fonction de corrélation dans le bin  $A$  est donnée par

$$\hat{\xi}_A = \frac{\sum_{(i,j) \in A} w_i w_j \delta_i}{\sum_{(i,j) \in A} w_i w_j}, \quad (5.6)$$

où l'indice  $i$  court sur les pixels des forêts et  $j$  sur les quasars pour lesquels la distance de séparation est comprise dans le bin  $A$ . Les paires  $ij$  où le pixel  $i$  appartient à la forêt du quasar  $j$  sont rejetées. Similairement au Ly $\alpha$ , les poids  $w_j$  associés aux quasars favorisent les quasars à plus grand redshift. Ces poids sont définis comme

$$w_j = \left( \frac{1+z_j}{1+2,25} \right)^{\gamma_{\text{QSO}}}, \quad (5.7)$$

où  $\gamma_{\text{QSO}} = 1,44 \pm 0,08$  (DU MAS DES BOURBOUX et al. 2019). Dans le calcul de la fonction de corrélation croisée  $\text{Ly}\alpha \times \text{QSO}$ , nous nous restreignons aux paires pour lesquelles  $r_{\perp} \in [0; 200] h^{-1} \text{ Mpc}$ . Contrairement à l'auto-corrélation du Ly $\alpha$ , la fonction de corrélation croisée n'est pas symétrique en  $r_{\parallel}$ . Cette asymétrie est produite par l'erreur systématique sur la mesure du redshift des quasars, et aussi par le rayonnement des quasars qui n'est pas isotrope. La fonction de corrélation croisée est donc calculée dans les bins pour lesquels  $r_{\parallel} \in [-200; 200] h^{-1} \text{ Mpc}$ . Ceci représente  $100 \times 50 = 5000$  bins.

### 5.1.3 Le spectre de puissance à une dimension

Comme expliqué dans le chapitre précédent, nous ajustons le spectre de puissance  $P_s$  appliqué à  $\delta_s$  de façon à obtenir un spectre de puissance à une dimension  $P^{1D}$  en accord avec les données. Nous présentons donc maintenant la mesure de ce spectre de puissance sur les raw mocks. Pour chaque forêt, nous appliquons une transformation de Fourier au champ  $\delta_F$  afin d'obtenir le champ  $\delta_k$ . Puis, le spectre de puissance de cette forêt est obtenu comme

$$\hat{P}^{1D}(k) = \langle \delta_k^2 \rangle. \quad (5.8)$$

Nous répétons cette procédure pour toutes les forêts, puis le spectre de puissance total est obtenu comme la moyenne du spectre de puissance de chaque forêt.

En ce qui concerne les données et les cooked mocks, la mesure du spectre de puissance à une dimension est plus complexe. Un certain nombre d'effets liés à la mesure doivent être pris en compte. En particulier, le bruit et la résolution doivent être estimés et pris en compte pour ne pas fausser l'estimation du spectre de puissance. Cette analyse est détaillée dans CHABANIER et al. (2018).

### 5.1.4 La fonction de corrélation à une dimension

Nous présentons maintenant l'estimateur de la fonction de corrélation à une dimension. Similairement au  $P^{1D}$ , cette fonction de corrélation est calculée en considérant uniquement les paires de pixels appartenant à la même forêt. Nous l'estimons comme

$$\hat{\xi}_A^{1D} = \frac{\sum_{(i,j) \in A} w_i w_j \delta_i \delta_j}{\sum_{(i,j) \in A} w_i w_j}, \quad (5.9)$$

où  $w_i$  est le poids associé au pixel  $i$ . Ces poids sont définis dans l'équation 3.11. La fonction de corrélation à une dimension  $\xi^{1D}$  correspond à la fonction de corrélation à trois dimensions pour  $\mu = 1$ . En considérant uniquement les paires le long de la ligne de visée, elle permet de mettre en évidence les autres absorbeurs présents dans le milieu intergalactique, tels que les métaux. La longueur d'onde d'absorption au repos de ces espèces étant différente de celle du Ly $\alpha$ , les séparations physiques  $r_{\parallel} = 0$  sont reconstruites à une séparation  $r_{\parallel}$  non nulle (voir l'explication de la matrice des métaux, section 5.4.1). Ceci produit des pics dans la fonction de corrélation à une dimension. Ces pics sont d'autant plus visibles quand la fonction de corrélation est représentée en fonction de  $\lambda_i / \lambda_j$ , où  $\lambda_i$  et  $\lambda_j$  sont les longueurs des deux pixels formant chaque paire.

### 5.1.5 L'auto-corrélation QSO×QSO

Afin de tester les mocks sous tous leurs aspects, nous mesurons aussi l'auto-corrélation QSO×QSO. L'estimateur utilisé pour cette corrélation est l'estimateur de Landy-Szalay (LANDY et SZALAY 1993). C'est l'estimateur qui minimise la variance. Il est défini comme

$$\hat{\xi} = \frac{DD - 2DR + RR}{RR}. \quad (5.10)$$

Les termes  $DD$ ,  $RR$  et  $DR$  donnent les nombres de paires normalisés dans chaque bin ( $r_{\parallel}$ ,  $r_{\perp}$ ). Ils sont donnés par

$$DD(r_{\parallel}, r_{\perp}) = \frac{dd(r_{\parallel}, r_{\perp})}{n_d(n_d - 1)/2}, \quad (5.11)$$

$$RR(r_{\parallel}, r_{\perp}) = \frac{rr(r_{\parallel}, r_{\perp})}{n_r(n_r - 1)/2}, \quad (5.12)$$

$$DR(r_{\parallel}, r_{\perp}) = \frac{dr(r_{\parallel}, r_{\perp})}{n_d n_r}, \quad (5.13)$$

où  $dd$  donne le nombre de paires mesuré dans les données,  $rr$  le nombre de paires mesuré dans le catalogue aléatoire, et  $dr$  le nombre de paires constituées d'un objet présent dans les données et

d'un objet présent dans le catalogue aléatoire.  $n_d$  et  $n_r$  donnent respectivement le nombre d'objet constituant le catalogue des données et le catalogue aléatoire. Nous utilisons aussi cet estimateur pour calculer la fonction d'auto-corrélation HCD $\times$ HCD.

## 5.2 Les matrices de distorsion

L'ajustement du continuum nécessaire au calcul du champ  $\delta_F$  dans les données et les cooked mocks biaise le champ mesuré. Cependant, grâce à la transformation (équation 3.9) décrite dans la section 3.4.2), l'effet sur la fonction de corrélation peut-être pris en compte. L'idée est la suivante : modéliser la distorsion induite par l'ajustement du continuum et par la transformation 3.9 sur la fonction de corrélation, appliquer cette distorsion au modèle, puis ajuster le modèle « distordu » aux données. L'effet le plus important de l'ajustement du continuum et de la transformation 3.9 est de forcer la moyenne et la pente de chaque région spectrale à être nulle. Ceci induit des corrélations entre les pixels d'une même région spectrale, et donc le long de la ligne de visée. Ainsi, au premier ordre, nous pouvons considérer que chaque  $\delta_F$  après distorsion d'une forêt est une combinaison linéaire de tous les  $\delta_F$  avant distorsion de cette forêt. La fonction de corrélation distordue dans le bin  $A$  peut alors être reliée à la vraie fonction de corrélation comme

$$\xi_{\text{distorsion}}(A) = \sum_B D_{AB} \xi_{\text{vraie}}(B), \quad (5.14)$$

où  $D_{AB}$  est appelée la *matrice de distorsion*. Celle-ci s'exprime en fonction du terme  $\eta_{ij}^q$ , défini dans l'équation 3.10. Pour l'auto-corrélation, elle s'exprime comme

$$D_{AB} = \frac{1}{W_A} \sum_{(i,j) \in A} w_i w_j \left( \sum_{(i',j') \in B} \eta_{ii'} \eta_{jj'} \right) \quad (5.15)$$

où  $W_A = \sum_{(i,j) \in A} w_i w_j$  est le poids du bin A. Pour la corrélation croisée, la matrice de distorsion est donnée par

$$D_{AB} = \frac{1}{W_A} \sum_{(i,j) \in A} w_i w_j \left( \sum_{(i',j') \in B} \eta_{ii'} \right). \quad (5.16)$$

Comme précédemment, les indices  $i$  correspondent aux pixels des forêts, et  $j$  aux quasars. A cause de la double somme, le calcul de la matrice de distorsion est très long. Afin de rendre possible l'estimation de cette dernière, le calcul est fait sur 1 % des paires, tirées au hasard.

La figure 5.1 présente l'effet de la distorsion sur l'auto-corrélation Ly $\alpha$  $\times$ Ly $\alpha$  dans les mocks. La fonction de corrélation est présentée dans quatre gammes en  $\mu$ . Les points bleus donnent la corrélation estimée à partir des raw mocks, c'est à dire sans les distorsions liées à l'ajustement du continuum. La ligne continue bleue donne l'ajustement du modèle de **picca**. Les points rouges donnent la corrélation estimée à partir des mocks eboss-0.0. Cette corrélation est affectée par les distorsions de l'ajustement du continuum. La courbe rouge donne l'ajustement du modèle sur cette corrélation. La matrice de distorsion permet ainsi de passer de la courbe bleue, à la courbe rouge.

La matrice de distorsion est un objet uniquement géométrique. Son calcul est indépendant du

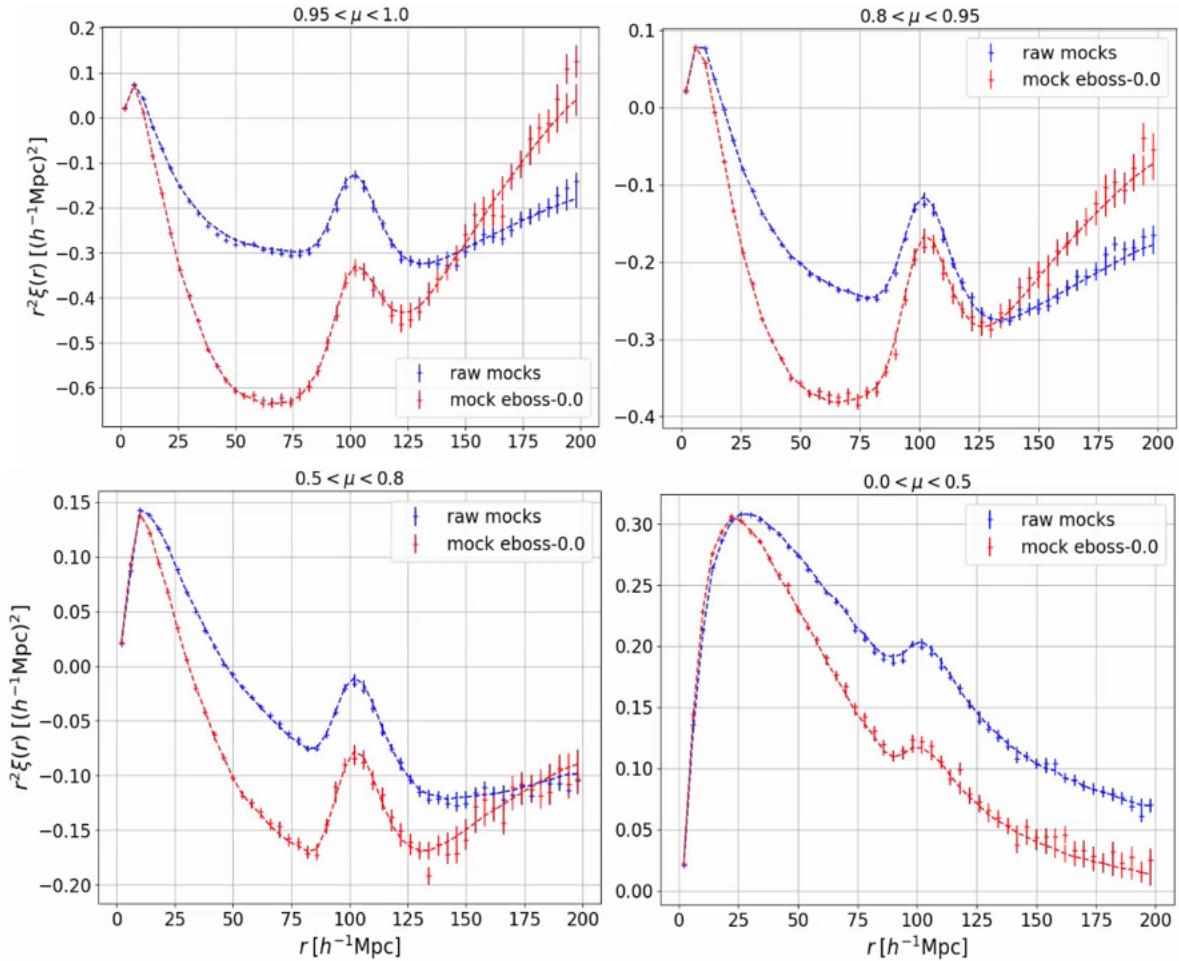


FIGURE 5.1 – L’auto-corrélation Ly $\alpha$  × Ly $\alpha$  estimée à partir de 30 réalisations des raw mocks (bleu) et à partir de 30 réalisations des mocks eboss-0.0 (rouge). Les lignes continues bleues et rouges donnent respectivement l’ajustement du modèle sur les raw mocks et les mocks eboss-0.0.

champ  $\delta_F$ . Elle dépend uniquement de la géométrie du relevé et de la distribution des poids. Il n’est pas nécessaire, en principe, de calculer la matrice de distorsion pour l’ensemble des réalisations des mocks. Nous vérifions cela en comparant le résultat de l’ajustement de la corrélation Ly $\alpha$  × Ly $\alpha$  estimée à partir d’une réalisation et ajustée en utilisant la matrice de distorsion calculée à partir de cette réalisation, au résultat de l’ajustement de la même corrélation, ajustée en utilisant la matrice de distorsion d’une autre réalisation. Les différences observées entre ces deux ajustements sont plus faibles que les erreurs liées à chaque paramètre. Ainsi, dans l’analyse présentée dans la suite de ce chapitre, nous calculons la matrice de distorsion pour l’auto-corrélation et pour la corrélation croisée une seule fois. L’ajustement de chaque fonction de corrélation utilise une de ces deux matrices de distorsion.

### 5.3 Les matrices de covariance

Afin de réaliser l’ajustement de chaque fonction de corrélation, nous avons besoin de calculer les matrices de covariance associées à ces fonctions de corrélation. La covariance de la fonction de

corrélation  $\xi$  dans le bin  $A$  et de  $\xi$  dans le bin  $B$  est définie comme

$$C_{AB} = \langle \xi_A \xi_B \rangle - \langle \xi_A \rangle \langle \xi_B \rangle . \quad (5.17)$$

De cette matrice de covariance, la matrice de corrélation est définie comme

$$Corr_{AB} = \frac{C_{AB}}{\sqrt{C_{AA} C_{BB}}} . \quad (5.18)$$

La matrice de corrélation donne la corrélation, comprise dans  $[-1; 1]$ , d'un bin A avec un bin B. Afin d'estimer la matrice de covariance, le relevé est divisé en pixels HEALPix, en utilisant `nside` = 16. Cette résolution produit des pixels d'une taille sur le ciel de  $3,7 \times 3,7 = 13,4 \text{deg}^2$ , correspondant à  $250 \times 250 (h^{-1} \text{Mpc})^2$  à un redshift  $z = 2,33$ . Ces sous-échantillons sont suffisamment grands pour pouvoir négliger les corrélations entre différents pixels HEALPix et ainsi estimer la matrice de covariance comme la variance d'un sous-échantillon à un autre. La matrice de covariance est donc calculée comme

$$C_{AB} = \frac{1}{W_A W_B} \sum_s W_A^s W_B^s (\xi_A^s \xi_B^s - \xi_A \xi_B) , \quad (5.19)$$

où  $s$  est un sous-échantillon, et  $W_A^s$  les poids du bin  $A$  de ce sous-échantillon. Les éléments les plus importants de cette matrice sont les éléments diagonaux : la variance dans chaque bin. Les éléments non-diagonaux, les covariances entre deux bins distincts, sont faibles. Leur estimation est bruitée. La matrice de covariance est donc lissée après avoir été estimée. En ce qui concerne l'auto-corrélation, la matrice de covariance possède  $2500 \times 2500$  bins. Pour la corrélation croisée, elle en possède  $5000 \times 5000$ .

## 5.4 Modélisation des fonctions de corrélation

Dans cette section, nous présentons les modèles utilisés pour ajuster les fonctions de corrélation Ly $\alpha$  × Ly $\alpha$  et Ly $\alpha$  × QSO. Nous présentons d'abord les modèles utilisés dans l'analyse des données DR16, puis nous donnons les modèles utilisés pour analyser les mocks.

### 5.4.1 Modélisation des données

Pour l'analyse des données DR16, dont les résultats sont présentés dans le chapitre suivant, nous utilisons le modèle décrit dans DU MAS DES BOURBOUX et al. (2020). L'analyse décrite dans cette étude est une analyse BAO : l'auto-corrélation et la corrélation croisée sont modélisées de façon à mesurer au mieux les paramètres BAO  $\alpha_{\parallel}$  et  $\alpha_{\perp}$ . Pour ce faire, le modèle est séparé en deux composantes. La première,  $\xi_{\text{smooth}}$ , correspond à la forme globale de la fonction de corrélation. La seconde,  $\xi_{\text{peak}}$ , correspond au pic BAO. C'est cette seconde composante qui dépend des paramètres BAO :

$$\xi(r_{\parallel}, r_{\perp}, \alpha_{\parallel}, \alpha_{\perp}) = \xi_{\text{smooth}}(r_{\parallel}, r_{\perp}) + \xi_{\text{peak}}(\alpha_{\parallel} r_{\parallel}, \alpha_{\perp} r_{\perp}) . \quad (5.20)$$

Cette distinction entre les composantes *smooth* et *peak* est faite lors du calcul du spectre de puissance modèle (voir paragraphe suivant). Le modèle de la fonction de corrélation est ensuite obtenue à l'aide

d'une transformation de Fourier inverse de ce spectre de puissance.

Le spectre de puissance utilisé dans la modélisation de la fonction de corrélation des traceurs  $i$  et  $j$  s'exprime comme

$$P(\vec{k}) = b_i b_j (1 + \beta_i \mu_k^2) (1 + \beta_j \mu_k^2) P_{\text{QL}}(\vec{k}) F_{\text{NL}}(\vec{k}) G(\vec{k}). \quad (5.21)$$

Les termes  $b_i(1 + \beta_i \mu_k^2)$  et  $b_j(1 + \beta_j \mu_k^2)$  sont les facteurs de Kaiser (équation 1.63) relatifs aux traceurs  $i$  et  $j$ .  $P_{\text{QL}}$  est le spectre de puissance *quasi-linéaire*. Il est découpé en deux composantes  $P_{\text{smooth}}$  et  $P_{\text{peak}}$  comme

$$P_{\text{QL}}(\vec{k}, z) = P_{\text{smooth}}(k, z) + \exp\left(-\frac{k_{\parallel}^2 \Sigma_{\parallel}^2 + k_{\perp}^2 \Sigma_{\perp}^2}{2}\right) P_{\text{peak}}(k, z). \quad (5.22)$$

$P_{\text{smooth}}$  est le spectre de puissance linéaire, sans les BAO. Il est construit à partir du spectre de puissance linéaire  $P_L$  donné par Camb, puis les BAO sont retirées en utilisant la technique *side-band* décrite dans KIRKBY et al. (2013). Le spectre de puissance  $P_{\text{peak}}$  est alors obtenu comme la différence  $P_L - P_{\text{smooth}}$  : il contient uniquement les oscillations dues aux BAO présentes dans le spectre de puissance linéaire. Le terme exponentiel devant  $P_{\text{peak}}$ , paramétrisé par  $\Sigma_{\parallel}$  et  $\Sigma_{\perp}$ , prend en compte l'élargissement non linéaire du pic BAO. Nous utilisons  $\Sigma_{\parallel} = 6,42 h^{-1} \text{ Mpc}$  et  $\Sigma_{\perp} = 3,26 h^{-1} \text{ Mpc}$  (D J EISENSTEIN et al. 2007). Le terme  $F_{\text{NL}}$  prend en compte les non-linéarités aux petites échelles. Nous distinguons  $F_{\text{NL}}^{\text{auto}}$  et  $F_{\text{NL}}^{\text{cross}}$ . Pour l'auto-corrélation, les effets non linéaires proviennent de l'élargissement thermique, des vitesses particulières et de la croissance des structures non linéaire. Comme lors de la modélisation du  $P^{1D}$ , nous utilisons le modèle décrit dans ARINYO-I-PRATS et al. (2015). Nous avons donc  $F_{\text{NL}}^{\text{auto}}(k, \mu) = D(k, \mu)$ , où  $D$  est défini dans l'équation 4.61. Les paramètres utilisés sont une interpolation à  $z = 2,334$  de ceux donnés dans la section « Planck » de la table 7 de ARINYO-I-PRATS et al. (2015). Pour la corrélation croisée, l'effet dominant est dû aux vitesses non linéaires des quasars. Cet effet est modélisé par une lorentzienne :

$$F_{\text{NL}}^{\text{cross}}(k_{\parallel}) = \frac{1}{1 + (k_{\parallel} \sigma_v)^2}, \quad (5.23)$$

où l'inverse de la demi-largeur à mi-hauteur  $\sigma_v$  est un paramètre libre. L'effet dû aux erreurs statistiques sur la mesure du redshift des quasars étant confondu avec l'effet des vitesses non linéaires des quasars, il est aussi pris en compte par le terme  $F_{\text{NL}}^{\text{cross}}$ . Enfin, le terme  $G(\vec{k})$  prend en compte l'effet du binning utilisé lors du calcul de la fonction de corrélation. Il est défini comme le produit des transformés de Fourier de la fonction porte :

$$G(\vec{k}) = \text{sinc}\left(\frac{k_{\parallel} R_{\parallel}}{2}\right) \text{sinc}\left(\frac{k_{\perp} R_{\perp}}{2}\right), \quad (5.24)$$

avec  $R_{\parallel}$  et  $R_{\perp}$  la largeur des bins, soit  $4 h^{-1} \text{ Mpc}$ .

Afin de modéliser la corrélation croisée Ly $\alpha$ ×QSO, nous ajoutons le paramètre  $\Delta r_{\parallel, \text{QSO}}$ , inclus comme

$$r_{\parallel} = r_{\parallel, \text{mesure}} + \Delta r_{\parallel, \text{QSO}}, \quad (5.25)$$

où  $r_{\parallel, \text{mesure}}$  est la séparation mesurée des paires  $(i, j)$ , et  $r_{\parallel}$  est la séparation utilisée dans le modèle de

la corrélation croisée. L'ajout de ce paramètre permet de prendre en compte les erreurs systématiques sur la mesure du redshift des quasars, qui rendent asymétrique la fonction de corrélation Ly $\alpha$ ×QSO.

La présence des facteurs de Kaiser dans l'équation 5.21 permet de mesurer le biais et le paramètre RSD de nos traceurs. En ce qui concerne l'auto-corrélation Ly $\alpha$ ×Ly $\alpha$ , la fonction de corrélation est proportionnelle à  $b_{\text{Ly}\alpha}^2(1+\beta_{\text{Ly}\alpha}\mu^2)^2$ . Cependant, la présence de HCD dans les données modifie le biais et le paramètre RSD du Ly $\alpha$ . En effet, l'efficacité de l'algorithme de détection n'étant pas de 100 %, il subsiste des DLA non identifiés dans les forêts. De plus, les HCD avec  $\log n_{\text{HI}} < 20,3$  ne sont pas identifiés. Ces deux effets participent à augmenter le biais mesuré du Ly $\alpha$  significativement. Nous utilisons les paramètres effectifs

$$b'_{\text{Ly}\alpha} = b_{\text{Ly}\alpha} + b_{\text{HCD}}F_{\text{HCD}}(k_{\parallel}), \quad (5.26)$$

$$b'_{\text{Ly}\alpha}\beta'_{\text{Ly}\alpha} = b_{\text{Ly}\alpha}\beta_{\text{Ly}\alpha} + b_{\text{HCD}}\beta_{\text{HCD}}F_{\text{HCD}}(k_{\parallel}), \quad (5.27)$$

où  $F_{\text{HCD}}$  est une fonction qui dépend de la distribution en  $z$  et en  $\log n_{\text{HI}}$  des HCD (FONT-RIBERA et MIRALDA-ESCUDÉ 2012). Cette fonction est estimée sur des simulations hydrodynamiques (ROGERS et al. 2017). Elle est modélisée comme

$$F_{\text{HCD}}(k_{\parallel}) = \exp(-L_{\text{HCD}}k_{\parallel}), \quad (5.28)$$

où  $L_{\text{HCD}}$  est la taille typique des HCD non masqués. La résolution du spectrographe d'eBOSS rend possible l'identification des DLA dont la largeur est supérieure à 2 nm, correspondant à une taille d'environ  $14 h^{-1}$  Mpc au redshift effectif de la mesure. Par ailleurs,  $L_{\text{HCD}}$  est très dégénéré avec les autres paramètres du modèle, comme le biais des HCD ou les paramètres du Ly $\alpha$ . Nous fixons donc sa valeur à  $10 h^{-1}$  Mpc dans l'ajustement du modèle. Le paramètre  $\beta_{\text{HCD}}$  est aussi très dégénéré avec les autres paramètres du modèle, nous utilisons un prior gaussien centré sur 0,5 et avec une largeur 0,09 lors des ajustements.

Afin de pouvoir ajuster le même modèle sur tous les bins  $(r_{\parallel}, r_{\perp})$ , la dépendance en redshift de  $\delta_F$  est prise en compte. En considérant que  $\beta_{\text{Ly}\alpha}$  est constant avec le redshift, nous avons  $\delta_F(z) \propto G(z)b_{\text{Ly}\alpha}(z)$ , avec  $G(z) \propto (1+z)^{-1}$ . Concernant le biais du Ly $\alpha$ , nous utilisons la même dépendance que celle choisie lors du calcul des poids (équation 3.11), c'est à dire  $b_{\text{Ly}\alpha} \propto (1+z)^{\gamma_{\text{Ly}\alpha}}$ , avec  $\gamma_{\text{Ly}\alpha} = 2.9$  (McDONALD et al. 2004). En ce qui concerne  $\beta_{\text{Ly}\alpha}$ , nous considérons lors de l'ajustement du modèle qu'il est indépendant du redshift. Comme montré dans l'analyse présentée dans le chapitre 6,  $\beta_{\text{Ly}\alpha}$  n'est pas indépendant du redshift. Cependant, le redshift moyen dans chaque bin  $(r_{\parallel}, r_{\perp})$  varie peu. Lors de l'analyse de l'ensemble des données DR16, il varie dans la gamme  $2,31 < z < 2,39$ . Cette variation correspond à une variation de  $\beta_{\text{Ly}\alpha}$  de moins de 5 %. De plus, elle est d'autant plus faible lorsque l'analyse est faite dans différents bins en redshift.

La fonction de corrélation croisée Ly $\alpha$ ×QSO est sensible au produit  $b_{\text{Ly}\alpha}(1+\beta_{\text{Ly}\alpha}\mu^2)b_{\text{QSO}}(1+\beta_{\text{QSO}}\mu^2)$ . L'ajustement de cette seule fonction de corrélation ne permet donc pas de lever la dégénérence des paramètres Ly $\alpha$  et des paramètres des QSO. Nous fixons donc  $b_{\text{QSO}}$  et  $\beta_{\text{QSO}}$ . Concernant le biais des quasars, comme pour le Ly $\alpha$  nous prenons en compte l'évolution avec le redshift. Il est

paramétrisé comme

$$b_{\text{QSO}}(z) = 3,77 \left( \frac{1+z}{1+2,334} \right)^{1,44}. \quad (5.29)$$

Cette paramétrisation est celle choisie dans DU MAS DES BOURBOUX et al. (2020). Nous gardons cette paramétrisation pour notre analyse des données, présentée dans le chapitre suivant. Cependant, lorsque nous analysons les mocks, nous utilisons la paramétrisation adoptée pour construire les mocks. Elle est donnée dans l'équation 1.64.  $\beta_{\text{QSO}}$  est choisi constant et vaut  $\beta_{\text{QSO}} = f/b_{\text{QSO}}(z_{\text{eff}})$ , où  $z_{\text{eff}}$  est le redshift effectif de la mesure. Le taux de croissance  $f$  est lui aussi évalué au redshift effectif de la mesure.

Une fois les composantes multiplicatives incluses au modèle, nous pouvons transformer le  $P(\vec{k})$  modèle définit dans l'équation 5.21 en fonction de corrélation  $\xi(r, \mu)$ . Pour ce faire, la fonction de corrélation est décomposée en polynômes de Legendre  $P_l$  jusqu'à  $l_{\max} = 6$ . Pour chaque  $l \in [0; 2; 4; 6]$ , le spectre de puissance est multiplié par la fonction de Bessel  $j_l$  puis intégré sur  $\mu_k$  (voir l'équation 2.7 de KIRKBY et al. (2013)). Étant donné que la transformation  $P(\vec{k}) \rightarrow \xi(r, \mu)$  est faite à chaque étape de la minimisation lors de l'ajustement du modèle, il est important que cette transformation se fasse très rapidement. Nous utilisons donc l'algorithme FFTLog (HAMILTON 1999), qui apporte à la fois rapidité et précision. Ainsi, en suivant cette procédure, nous obtenons les fonctions de corrélation  $\xi_{\text{Ly}\alpha \times \text{Ly}\alpha}$  en choisissant  $i = j = \text{Ly}\alpha$ , et  $\xi_{\text{Ly}\alpha \times \text{QSO}}$  en choisissant  $i = \text{Ly}\alpha$  et  $j = \text{QSO}$ . Afin de pouvoir ajuster les fonctions de corrélation calculées avec les données, nous devons prendre en compte dans nos modèles les corrélations parasites. Ces corrélations s'ajoutent au  $\xi$  modèle calculé précédemment. À ce stade, nous distinguons le modèle utilisé pour l'auto-corrélation et la corrélation croisée. Le modèle de l'auto-corrélation  $\text{Ly}\alpha \times \text{Ly}\alpha$  est défini comme

$$\xi = \xi_{\text{Ly}\alpha \times \text{Ly}\alpha} + \sum_{m,n} \tilde{\xi}_{m \times n} + \xi_{\text{ciel}}, \quad (5.30)$$

où  $\tilde{\xi}_{m \times n}$  est la corrélation de l'absorbeur  $m$  avec l'absorbeur  $n$ , interprétés comme des absorptions Ly $\alpha$ . Nous verrons plus tard comment relier la corrélation  $\tilde{\xi}_{m \times n}$  à la corrélation physique  $\xi_{m \times n}$ . Ces absorbeurs peuvent être du Ly $\alpha$  ou des métaux ( $m$  et  $n$  ne peuvent pas être tous les deux du Ly $\alpha$ ). Le tableau 5.1 liste les métaux ajustés dans les données.  $\xi_{\text{ciel}}$  est la corrélation induite par la soustraction du fond de ciel. Ces termes sont décrits dans les prochains paragraphes. Le modèle de la corrélation croisée Ly $\alpha \times \text{QSO}$  est défini comme

$$\xi = \xi_{\text{Ly}\alpha \times \text{QSO}} + \sum_m \tilde{\xi}_{m \times \text{QSO}} + \xi_{\text{prox}}, \quad (5.31)$$

où  $\tilde{\xi}_{m \times \text{QSO}}$  est la corrélation du métal  $m$  (interprété comme du Ly $\alpha$ ) avec les quasars, et  $\xi_{\text{prox}}$  donne la corrélation induite par l'effet de proximité des quasars. Ces termes sont décrits dans les prochains paragraphes.

Afin de modéliser la corrélation des métaux, nous utilisons le modèle défini précédemment, utilisé pour décrire les corrélations  $\xi_{\text{Ly}\alpha \times \text{Ly}\alpha}$  et  $\xi_{\text{Ly}\alpha \times \text{QSO}}$ . Comme expliqué dans la section 1.5.5, toutes les

absorptions sont supposées être des absorptions Ly $\alpha$ . Les absorptions causées par les métaux sont donc reconstruites à un mauvais redshift. Ceci résulte dans un décalage de la fonction de corrélation le long de la ligne de visée. Considérons deux absorbeurs  $m$  et  $n$  ( $m \neq n$ ). La fonction de corrélation  $\xi_{m \times n}$  de ces deux absorbeurs est maximale pour les séparations  $r = 0$ . Cependant, puisque l'analyse suppose que ces deux absorptions sont causées par le Ly $\alpha$ , cette séparation  $r = 0$  est reconstruite à  $r_{\perp} = 0$  et  $r_{\parallel} \sim (1+z)D_{\text{H}}(z)(\lambda_m - \lambda_n)/\lambda_{\text{Ly}\alpha}$ , où  $z$  est le redshift moyen des deux absorbeurs. Ainsi, l'excès de corrélation observé n'est pas situé à  $r = 0$ , mais se trouve décalé le long de la ligne de visée. Le tableau 5.1 donne les séparations associées aux corrélations entre le Ly $\alpha$  et les métaux ajustés sur les données. Les métaux étant beaucoup moins présents que l'hydrogène dans le milieu intergalactique, les corrélations mettant en jeu deux métaux sont beaucoup moins importantes que les corrélations mettant en jeu un métal et le Ly $\alpha$ . L'effet principal vient donc des corrélations  $\xi_{\text{Ly}\alpha \times m}$ . En ce qui concerne les corrélations  $\xi_{m \times m}$ , le décalage est d'origine différente. Une séparation physique  $r = 0$  correspond bien à une reconstruction  $r_{\parallel} = r_{\perp} = 0$ . Cependant, le redshift de la paire est mal estimé. Chaque séparation physique  $(r_{\parallel}, r_{\perp})$  est donc reconstruite à  $(D_{\text{H}}(z_m)/D_{\text{H}}(z))r_{\parallel}$  et  $(D_{\text{M}}(z_m)/D_{\text{M}}(z))r_{\perp}$ . Pour la corrélation  $\xi_{m \times \text{QSO}}$ , le décalage est le même que dans le cas  $\xi_{m \times n}$ , en prenant  $z_n = z_{\text{QSO}}$ .

Pour chaque couple  $(m, n)$ , le décalage de la fonction de corrélation  $\xi_{m \times n}$  est pris en compte par la matrice des métaux  $M_{AB}^{m \times n}$ . Nous ne détaillons pas son calcul ici, mais il est donné dans la thèse DE SAINTE AGATHE (p.d.). La matrice des métaux permet donc de relier la fonction de corrélation  $\xi_{m \times n}$  à la fonction de corrélation  $\tilde{\xi}_{m \times n}$ , où les absorbeurs  $m$  et  $n$  sont interprétés comme des absorptions Ly $\alpha$ , utilisée comme modèle dans l'équation 5.30 et 5.31 :

$$\tilde{\xi}_{m \times n}(A) = \sum_B M_{AB}^{m \times n} \xi_{m \times n}(B), \quad (5.32)$$

Les bins  $A$  correspondent aux séparations calculées en supposant une absorption Ly $\alpha$ . Les bins  $B$  correspondent aux séparations physiques, calculées en utilisant les redshifts  $z_m$  et  $z_n$  des absorbeurs. Ainsi, pour chaque couple  $(m, n)$ , la matrice des métaux  $M_{AB}^{m \times n}$  est calculée, puis la fonction de corrélation  $\tilde{\xi}_{m \times n}$  est estimée et ajoutée à la fonction de corrélation  $\xi_{\text{Ly}\alpha \times \text{Ly}\alpha}$  (équation 5.30). Dans le cas de la corrélation croisée, la matrice est calculée pour tous les couples  $(m, \text{QSO})$ , puis la fonction de corrélation  $\tilde{\xi}_{m \times \text{QSO}}$  et estimée et ajoutée à la fonction de corrélation  $\xi_{\text{Ly}\alpha \times \text{QSO}}$  (équation 5.31).

Le modèle utilisé pour construire les fonctions de corrélation  $\xi_{m \times n}$  et  $\xi_{m \times \text{QSO}}$  est le même que celui utilisé pour construire les fonctions de corrélation  $\xi_{\text{Ly}\alpha \times \text{Ly}\alpha}$  et  $\xi_{\text{Ly}\alpha \times \text{QSO}}$  (équation 5.21) mais avec les paramètres  $b_i$ ,  $b_j$ ,  $\beta_i$  et  $\beta_j$  qui sont ceux des métaux. Du fait que les métaux sont mesurables principalement le long de la ligne de visée, il est difficile d'ajuster à la fois le biais et le paramètre RSD de chaque métal. Nous ajustons donc uniquement le biais de chaque métal, le paramètre RSD étant fixé à  $\beta_m = 0,5$ . De plus, comme la quantité de métaux dans le milieu intergalactique est bien moins importante que celle de l'hydrogène, il n'existe pas ou peu de structures denses constituées de métaux. Nous n'incluons donc pas la modélisation des HCD dans les fonctions de corrélation  $\xi_{m \times n}$  et  $\xi_{m \times \text{QSO}}$ .

Le terme additionnel suivant est le terme  $\xi_{\text{ciel}}$ . Ce terme prend en compte les corrélations induites par la soustraction du fond de ciel. Lors de la réduction des données, décrite dans la section 3.1, le

TABLE 5.1 – Liste des métaux inclus dans le modèle ajusté aux données. La 3<sup>e</sup> colonne donne la séparation reconstruire pour une séparation réelle  $r = 0$ . CIV(eff) indique la raie effective du carbone IV : la résolution du spectrographe d'eBOSS étant trop faible pour distinguer le doublet du CIV, nous ajustons la combinaison des deux raies. La séparation liée au CIV est bien supérieure à  $200 h^{-1}$  Mpc, la corrélation Ly $\alpha$  × CIV n'a donc pas d'effet sur nos mesures. Nous modélisons cependant l'effet lié à l'auto-corrélation CIV × CIV.

Raie	$\lambda_m [\text{\AA}]$	$r_{\parallel} [h^{-1} \text{ Mpc}]$
SiII(1190)	1190,4158	-64
SiII(1193)	1193,2897	-56
SiIII(1207)	1206,500	-21
SiIII(1260)	1260,4221	+111
CIV(eff)	1549,06	> 200

spectre du fond de ciel est soustrait à tous les spectres d'une même demi-plaque. Ceci induit alors des corrélations entre tous ces spectres pour  $r_{\parallel} = 0$ . A cause de la distorsion induite par l'ajustement du continuum, cette effet ne se limite pas à  $r_{\parallel} = 0$ . L'effet est modélisé par une fonction gaussienne de  $r_{\perp}$  :

$$\xi_{ciel}(r_{\parallel}, r_{\perp}) = \begin{cases} \frac{A_{sky}}{\sqrt{2\pi\sigma_{sky}^2}} \exp\left(-\frac{r_{\perp}^2}{2\sigma_{sky}^2}\right) & , \text{ si } r_{\parallel} = 0 \\ 0 & , \text{ si } r_{\parallel} \neq 0 \end{cases}. \quad (5.33)$$

Les paramètres  $A_{sky}$  et  $\sigma_{sky}$  sont laissés libres lors de l'ajustement des données. Ils donnent l'amplitude et la largeur de la gaussienne. Le terme  $\xi_{ciel}$  n'est présent que dans l'ajustement de l'auto-corrélation, car ces corrélations parasites ne sont induites que lorsqu'on corrèle des pixels d'absorption issus de deux spectres présents sur la même demi-plaque. Cet effet n'a donc pas lieu d'être pour la corrélation croisée.

Enfin, le dernier terme additionnel est le terme  $\xi_{prox}$ . Ce terme n'est présent que dans la fonction de corrélation croisée  $\xi_{\text{Ly}\alpha \times \text{QSO}}$ . Il prend en compte l'effet du rayonnement produit par les quasars sur l'hydrogène environnant. En effet, à cause de leur grande luminosité, et en particulier dans la direction de leur jet, les quasars ionisent le gaz qui les entoure. Ceci réduit donc la fraction d'hydrogène neutre et donc la profondeur optique au voisinage de chaque quasar, ce qui induit des corrélations supplémentaire entre le champ d'absorption Ly $\alpha$  et la position des quasars. Cet effet est modélisé comme (FONT-RIBERA et al. 2013) :

$$\xi_{prox} = \xi_{0,prox} \left( \frac{1 h^{-1} \text{ Mpc}}{r} \right)^2 \exp\left(-\frac{r}{\lambda_{UV}}\right), \quad (5.34)$$

où  $\xi_{0,prox}$  donne l'amplitude de l'effet. L'émission est supposée isotrope, et le paramètre  $\lambda_{UV}$  est fixé à  $300 h^{-1}$  Mpc.

Une fois tous ces termes inclus, nous obtenons un modèle qui décrit la fonction de corrélation de la matière dans l'espace des redshifts, à un biais près, et qui prend en compte les différents effets astrophysiques ou instrumentaux qui affectent les données. Afin de pourvoir correctement comparer

notre modèle aux fonctions de corrélation calculées avec les données, nous devons prendre en compte la distorsion due à l'ajustement du continuum. Ceci est fait, comme décrit dans la section 5.2, grâce à la matrice de distorsion. Le modèle distordu est alors donné par

$$\xi_{\text{distorsion}}(A) = \sum_B D_{AB} \xi(B), \quad (5.35)$$

où  $D_{AB}$  est la matrice de distorsion, et  $\xi$  est le modèle construit précédemment. Ainsi, le modèle qui est ajusté aux données est  $\xi_{\text{distorsion}}(A)$ .

### 5.4.2 Modélisation des mocks

Afin d'analyser les fonctions de corrélation Ly $\alpha$  × Ly $\alpha$  et Ly $\alpha$  × QSO des mocks, nous utilisons les modèles ajustés sur les données et décrits dans la section précédente. Cependant, un certain nombre d'effets modélisés dans les données ne sont pas présents dans les mocks. Nous modifions donc légèrement les modèles décrits précédemment.

Premièrement, nous n'incluons pas l'élargissement non linéaire du pic BAO dans les mocks. Les non-linéarités prises en compte par le terme  $F_{NL}$  ne sont pas non plus présentes dans nos mocks. Ainsi, le spectre de puissance défini dans l'équation 5.21 et utilisé comme modèle pour l'auto-corrélation ne contient pas le terme  $F_{NL}^{\text{auto}}$ . Cependant, le code **quickquasars** ajoute une vitesse particulière à chaque quasar. Ceci a pour effet d'ajouter une erreur statistique sur la mesure du redshift des quasars. Nous gardons donc le terme  $F_{NL}^{\text{cross}}$  lorsque nous ajustons la corrélation Ly $\alpha$  × QSO issues des mocks avec **quickquasars**. Ce terme n'est pas présent dans le modèle utilisé pour ajuster les raw mocks.

L'effet instrumental causé par la soustraction du fond de ciel sur l'auto-corrélation n'est pas modélisé par **quickquasars**. Nous n'ajoutons donc pas le terme  $\xi_{\text{ciel}}$  à  $\xi_{\text{Ly}\alpha \times \text{Ly}\alpha}$  dans l'ajustement de l'auto-corrélation. L'effet de proximité des quasars sur le champ Ly $\alpha$  environnant n'est ajouté ni dans les mocks, ni dans **quickquasars**. Le terme  $\xi_{\text{prox}}$  n'est donc pas inclus dans le modèle de la corrélation croisée.

En ce qui concerne l'ajustement des HCD et des métaux, cela dépend de la version des mocks analysée. Pour les raw mocks comme pour les mocks eboss-0.0, ni les HCD ni les métaux ne sont présents. Nous n'incluons donc pas leur modélisation dans ces versions des mocks. Les versions eboss-0.2 incluent les HCD, nous modélisons donc leur présence comme décrit dans la section précédente. Les versions eboss-0.3 incluent à la fois les HCD et les métaux. L'ajustement des fonctions de corrélation issues de ces mocks contient donc les paramètres  $b_{\text{HCD}}$  et  $\beta_{\text{HCD}}$ . De plus, nous calculons la matrice des métaux pour ces versions et ajoutons au modèle les termes  $\xi_{m \times n}$  et  $\xi_{m \times \text{QSO}}$ .

Pour l'ajustement des corrélations croisées Ly $\alpha$  × QSO, nous gardons le paramètre  $\Delta_{r_{\parallel}, \text{QSO}}$ . Même si **quickquasars** ajoute une erreur statistique sur les redshifts des quasars, cette erreur est nulle en moyenne. Nous nous attendons donc à obtenir  $\Delta_{r_{\parallel}, \text{QSO}} = 0$  dans l'ajustement des différentes versions des mocks. Ainsi,  $\Delta_{r_{\parallel}, \text{QSO}}$  sert de test de la construction des mocks, pour vérifier par exemple que les lignes de visées sont placées correctement à partir de chaque quasar.

Comme pour les données, les fonctions de corrélation sont évaluées sur une grille de séparation d'intervalle  $4 h^{-1}$  Mpc. Nous gardons donc le terme  $G(\vec{k})$  dans la modélisation des mocks. Dans le cas

des mocks, nous incluons un terme supplémentaire, qui prend en compte le lissage gaussien appliqué au champ  $\delta_l$  interpolé. Ce terme est donné par

$$W(k) = \exp\left(-\frac{k^2 \sigma_{\text{smooth}}^2}{2}\right). \quad (5.36)$$

Comme montré sur la figure 4.8, l'effet du lissage gaussien sur la fonction de corrélation est important surtout à petit  $r$ . Les mocks n'étant pas en très bon accord avec le modèle à petit  $r$ , la modélisation du lissage gaussien a tendance à corriger ces désaccords lorsque le paramètre  $\sigma_{\text{smooth}}$  est laissé libre. De plus,  $\sigma_{\text{smooth}}$  est dégénéré avec d'autres paramètres du modèle. Nous fixons donc  $\sigma_{\text{smooth}}$  à la taille des voxels, soit  $2,19 h^{-1}$  Mpc lors des ajustements. Le spectre de puissance modèle  $P(\vec{k})$  est ainsi multiplié par  $W^2(k)$ .

Enfin, similairement à la modélisation des données, le modèle ajusté sur les mocks est multiplié par la matrice de distorsion  $D_{AB}$  (équation 5.35). Cependant, dans le cas des raw mocks, les fonctions corrélation ne sont pas affectées par la distorsion due à l'ajustement du continuum puisque nous avons accès directement au champ  $\delta_F$ . Dans ce cas, la matrice de distorsion  $D_{AB}$  vaut la matrice identité.

L'accord entre les mocks et le modèle n'est pas excellent à petit  $r$ , malgré le terme  $W(k)$  qui prend en compte en principe l'effet dû au lissage gaussien. La mesure de  $b_{\text{eff},\text{Ly}\alpha}(z)$  et  $\beta_{\text{Ly}\alpha}(z)$  est très affectée par ces petites séparations. Comme nous le montrons ici, celle-ci dépend de la séparation minimale  $r_{\min}$  utilisée pour ajuster le modèle sur les mocks. Toujours dans le but d'avoir une mesure stable de  $b_{\text{eff},\text{Ly}\alpha}(z)$  et  $\beta_{\text{Ly}\alpha}(z)$ , nous produisons plusieurs ajustements de la corrélation Ly $\alpha$  × Ly $\alpha$  estimée à partir des raw mocks et des mocks eboss-0.0 pour différents  $r_{\min}$ . Pour chacun des ajustements, nous mesurons  $b_{\text{eff},\text{Ly}\alpha}$  et  $\beta_{\text{Ly}\alpha}$  et  $\chi^2_{\text{red}}$ . Ce dernier est défini comme

$$\chi^2_{\text{red}} = \frac{\chi^2}{n_{dof}}, \quad (5.37)$$

où  $n_{d.o.f.}$  donne le nombre de degrés de liberté de l'ajustement : c'est le nombre de bins dans laquelle la fonction de corrélation est ajustée moins le nombre de paramètres libres du modèle.  $n_{d.o.f.}$  varie de 1590 pour  $r_{\min} = 10 h^{-1}$  Mpc à 1278 pour  $r_{\min} = 80 h^{-1}$  Mpc. La figure 5.2 montre l'évolution de  $\chi^2_{\text{red}}$ ,  $b_{\text{eff},\text{Ly}\alpha}$  et  $\beta_{\text{Ly}\alpha}$  en fonction de  $r_{\min}$ . La colonne de gauche donne les mesures faites sur les raw mocks, et celle de droite les mesures faites sur les mocks eboss-0.0. Sur cette figure, nous pouvons premièrement remarquer qu'il est difficile de mesurer  $b_{\text{eff},\text{Ly}\alpha}$  et  $\beta_{\text{Ly}\alpha}$  lorsque  $r_{\min}$  est trop grand : l'erreur statistique augmente énormément avec  $r_{\min}$ . Ensuite,  $\chi^2_{\text{red}}$  augmente drastiquement (en particulier pour les raw mocks<sup>1</sup>) pour  $r_{\min} = 10 h^{-1}$  Mpc. De plus,  $b_{\text{eff},\text{Ly}\alpha}$  et  $\beta_{\text{Ly}\alpha}$  ont un comportement erratique pour  $r_{\min} = 10 h^{-1}$  Mpc, et semblent être stables pour  $r_{\min} \geq 10 h^{-1}$  Mpc. Pour toutes ces raisons, et aussi pour garder un maximum de statistique, nous choisissons dans tous les ajustements des mocks (sauf autrement spécifié)  $r_{\min} = 20 h^{-1}$  Mpc. En ce qui concerne la borne supérieure, le modèle est ajusté jusqu'à  $r_{\max} = 180 h^{-1}$  Mpc.

---

1. Les raw mocks possèdent un  $\chi^2$  plus important car la statistique au sein d'une même réalisation est plus importante (le bruit ajouté par `quickquasars` n'est pas présent) : les déviations, s'il y en a, sont moins marquées du fait des barres d'erreur plus grandes dans les mocks avec `quickquasars`.

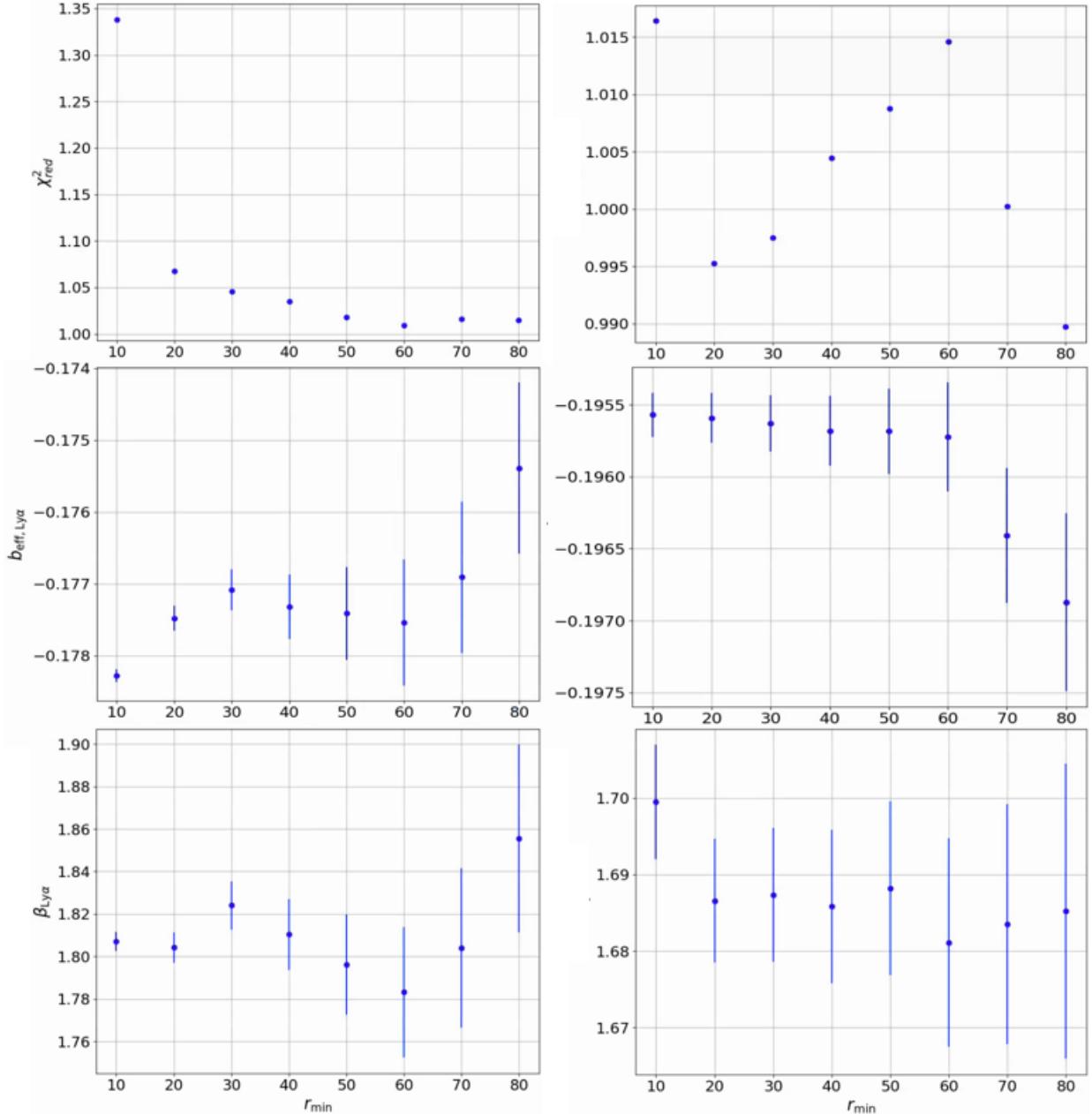


FIGURE 5.2 – Évolution de  $\chi^2_{red}$  (graphiques du haut),  $b_{\text{eff}, \text{Ly}\alpha}$  (graphiques du milieu) et  $\beta_{\text{Ly}\alpha}$  (graphiques du bas) en fonction de  $r_{min}$  (donné en  $h^{-1}$  Mpc). La colonne de gauche donne les mesures avec l'ajustement de la corrélation Ly $\alpha$ ×Ly $\alpha$  estimée à partir des raw mocks. Celle de droite donne la mesure faite sur les mocks eboss-0.0.

## 5.5 Analyse des mocks

Comme expliqué dans la section 4.5, nous produisons différentes versions de mocks : les raw mocks, pour lesquels le champ  $\delta_F$  est obtenu directement à partir des vraies transmissions, et les mocks après l'utilisation de `quickquasars` : eboss-0.0, eboss-0.2 et eboss-0.3. Nous présentons dans cette section l'analyse de ces différentes versions de mocks.

### 5.5.1 Analyse des raw mocks

Pour chacune des 30 réalisations produites, nous avons estimé les fonctions de corrélation  $\text{Ly}\alpha \times \text{Ly}\alpha$  et  $\text{Ly}\alpha \times \text{QSO}$  des raw mocks. L'analyse de ces fonctions de corrélation permet d'identifier plus facilement les problèmes qui peuvent exister au niveau de la construction des mocks, car les effets dus à l'ajustement du continuum et les effets astrophysiques et instrumentaux introduits par `quickquasars` ne sont pas présents. Nous commençons donc par valider la construction des mocks, via l'étude des raw mocks, puis nous présentons l'analyse des versions des mocks avec `quickquasars`.

Les fonctions de corrélation présentées ici sont estimées dans quatre bins en redshift. Ces bins sont les mêmes que ceux choisis pour analyser les données et déterminer  $b_{\text{Ly}\alpha}(z)$  et  $\beta_{\text{Ly}\alpha}(z)$  à utiliser pour la construction des mocks (voir section 6.1.1). Ces bins sont : [0;2,35], [2,35;2,65], [2,65;3,05] et [3,05;10]. Nous utilisons les mêmes bins afin de faciliter la comparaison entre les mocks et les données. Une fois les fonctions de corrélation estimées dans chaque bin et pour chaque réalisation, nous calculons, dans chaque bin en redshift, la moyenne de ces fonctions de corrélation, puis ajustons le résultat de cette moyenne. L'ajustement est fait avec le code `picca`.

#### L'auto-corrélation $\text{Ly}\alpha \times \text{Ly}\alpha$

La figure 5.3 donne la moyenne des 30 fonctions de corrélation  $\text{Ly}\alpha \times \text{Ly}\alpha$  des raw mocks dans chaque bin en redshift. Pour chaque bin en redshift, la fonction de corrélation est présentée dans quatre bins en  $\mu$  différents. Le bin  $0,95 < \mu < 1$  correspond aux paires avec une séparation le long de la ligne de visée. Le bin  $0 < \mu < 0,5$  correspond aux paires perpendiculaires à la ligne de visée. Les lignes continues donnent l'ajustement du modèle produit par `picca`. Ce modèle décrit très bien les raw mocks. Les lignes en pointillés donnent la prédiction des mocks. L'écart visible entre la corrélation estimée à partir des raw mocks et la prédiction provient de l'écart entre la corrélation du champ  $\delta_g$  et la prédiction que nous en faisons (figure 4.12). Cet écart se propage lorsque nous passons de la prédiction de  $\xi_g$  à la prédiction de  $\xi_F$  grâce à l'équation 4.41.

Le tableau 5.2 donne le résultat de l'ajustement dans chaque bin en redshift. Nous pouvons remarquer que les paramètres BAO  $\alpha_{\parallel}$  et  $\alpha_{\perp}$  sont compatibles avec 1. La figure 5.4 présente le biais et le paramètre RSD du  $\text{Ly}\alpha$  obtenus dans l'ajustement des fonctions de corrélation  $\text{Ly}\alpha \times \text{Ly}\alpha$  des raw mocks (vert), ainsi que ceux obtenus dans l'ajustement des données (noir). L'analyse en quatre bins en redshift des données est décrite dans la section 6.1.1. Les valeurs obtenues avec la prédiction lors de l'ajustement des paramètres des mocks sont indiquées en magenta. Enfin, pour chacun des jeux des données présentés sur la figure 5.4, nous ajustons une loi de puissance du type  $a(1+z)^\gamma$ . Les paramètres  $b_{\text{Ly}\alpha}$  et  $\beta_{\text{Ly}\alpha}$  mesurés sur les raw mocks sont en très bon accord avec ceux prédis, et donc,

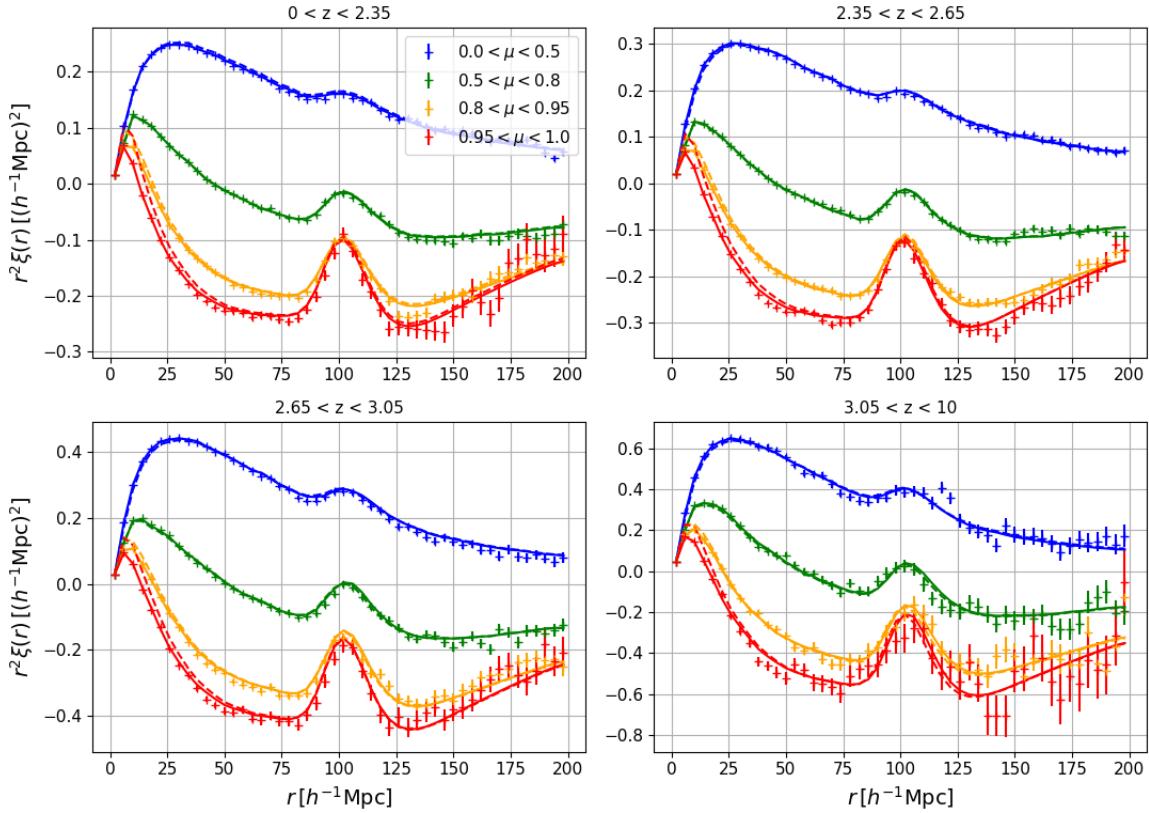


FIGURE 5.3 – L’auto-corrélation Ly $\alpha$  × Ly $\alpha$  calculée sur la moyenne des 30 raw mocks. Chaque graphique donne la moyenne des fonctions de corrélation calculées dans chaque bins en redshift pour chaque réalisation. Les fonctions de corrélation sont présentées dans quatre bins en  $\mu$ . Les lignes continues donnent l’ajustement du modèle. Les lignes en pointillés donnent la prédiction des mocks.

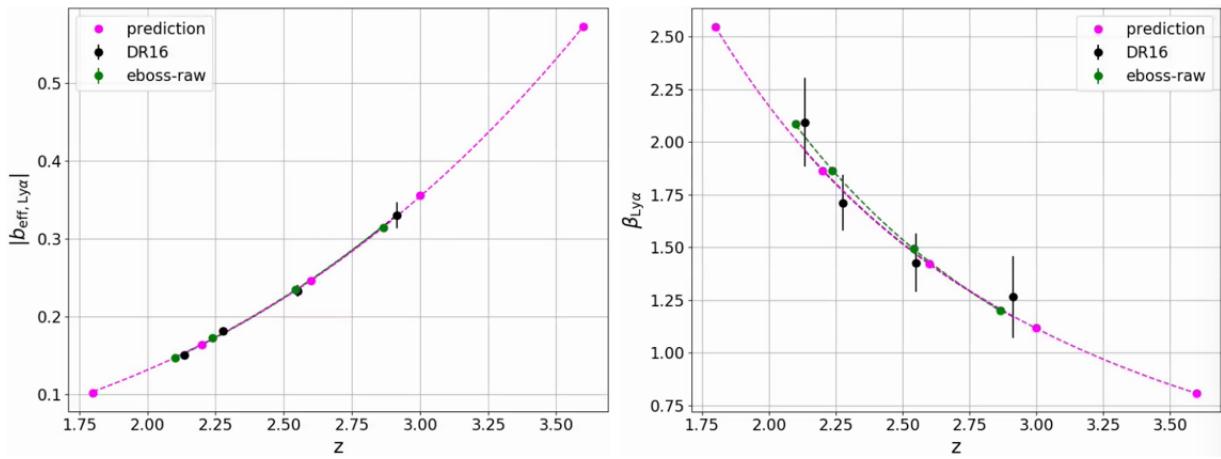


FIGURE 5.4 – Mesures des paramètres  $b_{\text{Ly}\alpha}$  et  $\beta_{\text{Ly}\alpha}$  faites avec les auto-corrélations Ly $\alpha$  × Ly $\alpha$  estimées dans chaque bin en redshift à partir des données DR16 (noir) et de la moyenne des 30 raw mocks (vert). Les barres d’erreur sont affichées pour ces deux jeux de données. Dans le cas des raw mocks, elles sont plus petites que la taille des points. Les points magenta donnent les mesures faites avec la prédiction des mocks. Les lignes en pointillés représentent l’ajustement sur chaque jeu de données d’une loi de puissance  $(1+z)^\gamma$ .

TABLE 5.2 – Résultats de l’ajustement de l’auto-corrélation Ly $\alpha$ ×Ly $\alpha$  calculée sur la moyenne des 30 raw mocks. Chaque colonne donne le résultat de l’ajustement d’un bin en redshift. La dernière colonne donne le résultat de l’ajustement de la corrélation moyennée sur ces quatre bins en redshift. La première section du tableau donne les paramètres du modèle qui sont ajustés. La seconde donne le redshift effectif  $z_{\text{eff}}$  et le  $\chi^2$ . Le nombre de bins sur lesquels le modèle est ajusté est  $N_{\text{bin}} = 1574$ , ce qui donne un nombre de degrés de liberté  $n_{\text{d.o.f.}} = 1570$ . La dernière section donne le biais et le biais effectif du Ly $\alpha$ .

Paramètre	$0 < z < 2,35$	$2,35 < z < 2,65$	$2,65 < z < 3,05$	$3,05 < z < 10$	$0 < z < 10$
$\alpha_{  }$	$1.001 \pm 0.005$	$1.004 \pm 0.004$	$0.998 \pm 0.005$	$0.986 \pm 0.014$	$1.001 \pm 0.003$
$\alpha_{\perp}$	$0.999 \pm 0.007$	$1.002 \pm 0.006$	$0.984 \pm 0.008$	$0.994 \pm 0.018$	$0.995 \pm 0.004$
$b_{\eta, \text{Ly}\alpha}$	$-0.1751 \pm 0.0004$	$-0.1936 \pm 0.0004$	$-0.2311 \pm 0.0008$	$-0.2692 \pm 0.0019$	$-0.1960 \pm 0.0003$
$\beta_{\text{Ly}\alpha}$	$2.085 \pm 0.014$	$1.863 \pm 0.011$	$1.494 \pm 0.011$	$1.200 \pm 0.016$	$1.804 \pm 0.007$
$\chi^2$	1557	1608	1628	1534	1676
$z_{\text{eff}}$	2.101	2.237	2.542	2.866	2.261
$b_{\text{Ly}\alpha}$	$-0.0808 \pm 0.0004$	$-0.1004 \pm 0.0004$	$-0.1507 \pm 0.0007$	$-0.2198 \pm 0.0017$	$-0.1051 \pm 0.0003$
$b_{\text{eff,Ly}\alpha}$	$-0.1458 \pm 0.0002$	$-0.1721 \pm 0.0002$	$-0.2355 \pm 0.0004$	$-0.3176 \pm 0.0011$	$-0.1775 \pm 0.0002$

par construction, avec les données DR16. Nous pouvons cependant noter une légère déviation entre  $\beta_{\text{Ly}\alpha}$  mesuré dans les raw mocks et la valeur prédictive pour les faibles redshifts. Comme nous l’avons déjà fait remarquer, la prédiction ne décrit pas parfaitement la fonction de corrélation du champ  $\delta_g$ , en particulier le long de la ligne de visée (figure 4.12). Ceci peut expliquer les déviations observées dans la mesure de  $\beta_{\text{Ly}\alpha}$ .

### La corrélation croisée Ly $\alpha$ ×QSO

Afin d’estimer la corrélation croisée Ly $\alpha$ ×QSO, nous avons recours à l’estimateur défini dans l’équation 5.6. Cependant, dans le cas des raw mocks, la moyenne des  $\delta_F$  est non nulle dans certaines régions du ciel, et ainsi  $r^2\xi(r)$  ne tend pas vers 0 à grand  $r$ . Pour palier ce problème, la fonction de corrélation « mélangée » est calculée : les forêts sont attribuées aléatoirement à un autre quasar, puis la fonction de corrélation est estimée en utilisant l’équation 5.6. Cette fonction de corrélation estime donc les corrélations parasites, comme celles qui peuvent être créées lorsque la moyenne des  $\delta_F$  est non nulle. Une fois cette corrélation estimée, nous obtenons la corrélation croisée Ly $\alpha$ ×QSO comme la différence de la corrélation croisée « normale » et de la corrélation « mélangée ». Cependant, dans le cas des mocks présentés dans ce manuscrit, cette méthode pour estimer la corrélation croisée Ly $\alpha$ ×QSO à partir des raw mocks ne suffit pas à retirer toutes les corrélations parasites. La figure 5.5 présente la moyenne des 30 corrélations Ly $\alpha$ ×QSO estimées à partir des raw mocks dans chaque bin en redshift. Les lignes continues donnent le modèle ajusté par `picca`. Nous remarquons que  $r^2\xi(r)$  ne tend pas vers 0 à grand  $r$ . Les déviations par rapport au modèle se compensent en partie lorsque nous regardons la corrélation Ly $\alpha$ ×QSO moyennée dans les quatre bins en redshift. Cette corrélation est présentée sur la figure 5.6. L’effet est quand même visible, en particulier sur le graphique de droite, qui présente la corrélation moyennée dans la gamme  $0 < \mu < 1$ .

Nous ne comprenons pas vraiment d’où proviennent ces corrélations parasites à grande échelle. Nous suspectons le calcul du champ  $\delta_F$  ou des corrélations : comme nous le verrons dans les sections 5.5.2

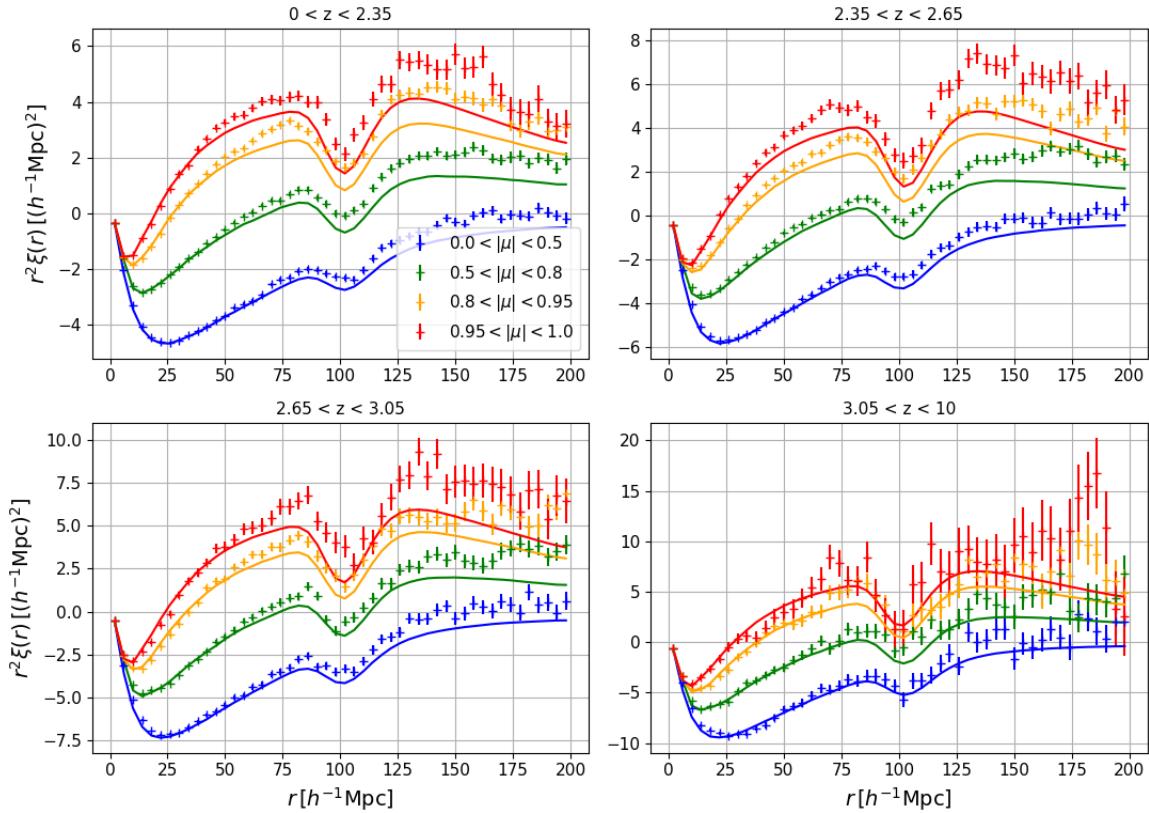


FIGURE 5.5 – La corrélation croisée Ly $\alpha$ ×QSO calculée sur la moyenne des 30 raw mocks. Chaque graphique donne la moyenne des fonctions de corrélation calculées dans chaque bins en redshift pour chaque réalisation. Les fonctions de corrélation sont présentées dans quatre bins en  $\mu$ . Les lignes continues donnent l’ajustement du modèle.

et 5.5.3, nous n’observons aucune déviation de la sorte lorsque nous estimons les corrélations Ly $\alpha$ ×QSO à partir des mocks eboss-0.0 ou eboss-0.2. De plus, nous suspectons aussi le fait que nous calculons les fonctions de corrélation dans différents bins en redshift. Ceci n’était pas fait jusqu’à présent, et est très peu testé. Par exemple, le calcul de la corrélation croisée Ly $\alpha$ ×QSO à partir des mocks eboss-0.0 ou eboss-0.2 ne requiert pas, en principe, de retirer la corrélation mélangée. Cependant, lorsque nous avons estimé ces corrélations dans les quatre bins en redshift à partir des mocks eboss-0.0 et eboss-0.2, nous avons observé le même genre de déviations que celles observées sur la figure 5.5. Ces déviations se compensaient lorsque nous moyennions les corrélations des différents bins en redshift. Cependant, dans le but de mesurer la corrélation Ly $\alpha$ ×QSO dans ces différents bins en redshift, nous avons décidé de retirer la corrélation mélangée dans chacun des bins en redshift. Ceci a résolu le problème et retiré les corrélations parasites. Paradoxalement, cette solution ne fonctionne pas complètement pour les raw mocks. Il serait intéressant de calculer les fonctions de corrélation Ly $\alpha$ ×QSO directement sur l’ensemble des forêts, afin de vérifier que ces corrélations parasites disparaissent.

### Le spectre de puissance à une dimension

Comme expliqué dans la section 4.4, le procédure d’ajustement des mocks nous permet d’obtenir un spectre de puissance à une dimension  $P^{1D}$  correcte dans la gamme  $1.8 < z < 3.6$ . Cependant, lorsque

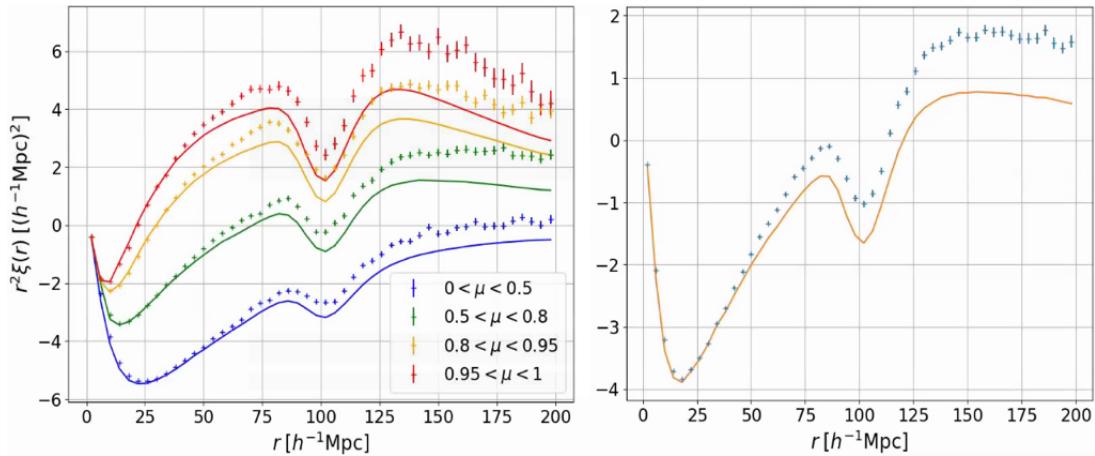


FIGURE 5.6 – La corrélation croisée Ly $\alpha$  × QSO estimée à partir des raw mocks et moyennée dans les quatre bins en redshift. Le graphique de gauche présente les fonctions de corrélation dans quatre bins en  $\mu$ . Le graphique de droite présente la fonction de corrélation moyennée dans la gamme  $0 < \mu < 1$ . Les lignes continues donnent l’ajustement du modèle.

nous calculons le spectre de puissance à une dimension à partir des raw mocks, nous observons des déviations par rapport au modèle sur lequel le  $P^{1D}$  des mocks est ajusté. La figure 5.7 présente le spectre de puissance à une dimension estimé à partir des raw mocks dans différents bins en redshift. Les lignes continues donnent le modèle utilisé pour ajuster les mocks. Pour chacune des valeurs  $z$  du redshift indiquées sur cette figure, le spectre de puissance à une dimension est estimée à l'aide de l'équation 5.8, en considérant uniquement les pixels dont le redshift  $z_i$  vérifie  $z - 0,1 < z_i < z + 0,1$ . Sur la figure 5.7, nous pouvons remarquer une déviation par rapport au modèle pour les redshifts les plus à l'extérieur de l'intervalle [1,8; 3,6]. Ceci est étonnant car, comme montré sur la figure 4.21, les  $P^{1D}$  obtenus à la fin de la procédure d'ajustement sont en parfait accord avec le modèle. Nous suspectons que les déviations observées sont liées d'une certaine façon à la dépendance avec le redshift des pixels d'absorption Ly $\alpha$ . En effet, lorsque nous produisons des forêts où tous les pixels d'absorption sont construits au même redshift, nous n'observons pas ces déviations et le spectre de puissance calculé sur ces forêts est en parfait accord avec le modèle. Heureusement, les déviations que nous observons concernent uniquement les redshifts les plus à l'extérieur de l'intervalle [1,8; 3,6]. Ces déviations ne concernent donc qu'un faible nombre de pixels.

### La corrélation à une dimension Ly $\alpha$ × Ly $\alpha$

Contrairement aux autres fonctions de corrélation, la corrélation à une dimension Ly $\alpha$  × Ly $\alpha$  n'est pas estimée dans différents bins en redshift. Nous l'estimons sur une seule réalisation des raw mocks, en considérant les forêts de l'ensemble des quasars. La figure 5.8 présente cette corrélation. Du fait qu'elle est estimée à partir des raw mocks, elle ne possède ni les distorsions produites par l'ajustement du continuum, ni les HCD et les métaux. Nous ne comparons donc pas cette corrélation à celle des données. Cette comparaison est faite en utilisant la corrélation à une dimension estimée à partir des mocks eboss-0.3, présentée sur la figure 5.5.4.

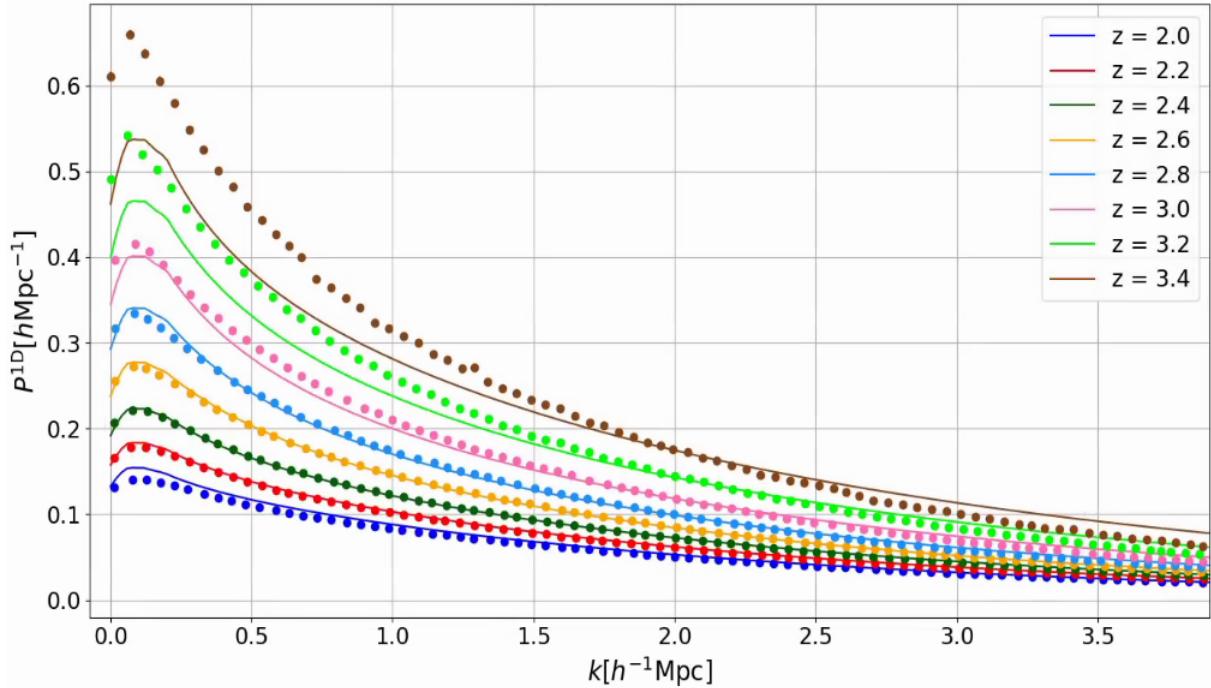


FIGURE 5.7 – Spectres de puissance à une dimension mesurés dans les raw mocks, dans différents bins en redshift. Les lignes continues donnent le modèle utilisé pour ajuster le spectre de puissance à une dimension des mocks. Les barres d'erreur représentées sont inférieures à la taille des points.

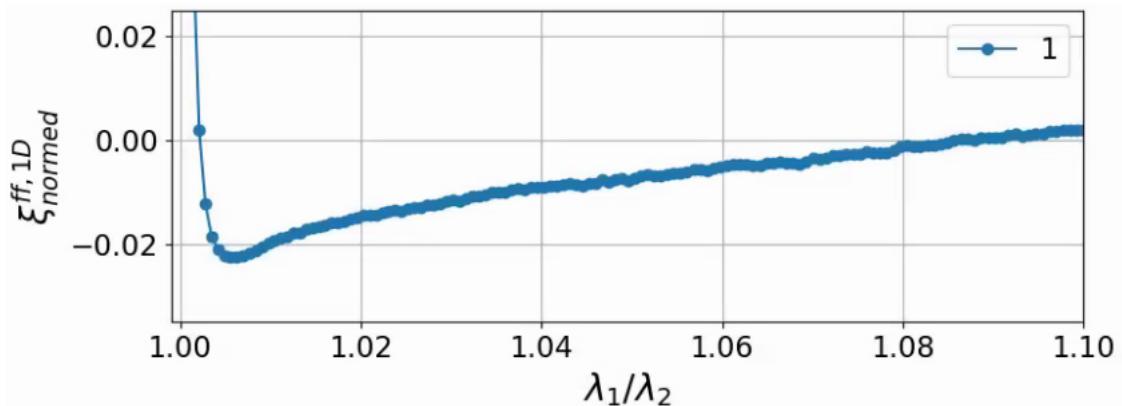


FIGURE 5.8 – La fonction de corrélation à une dimension Ly $\alpha$  × Ly $\alpha$  estimée à partir d'une réalisation de raw mocks (bleu).

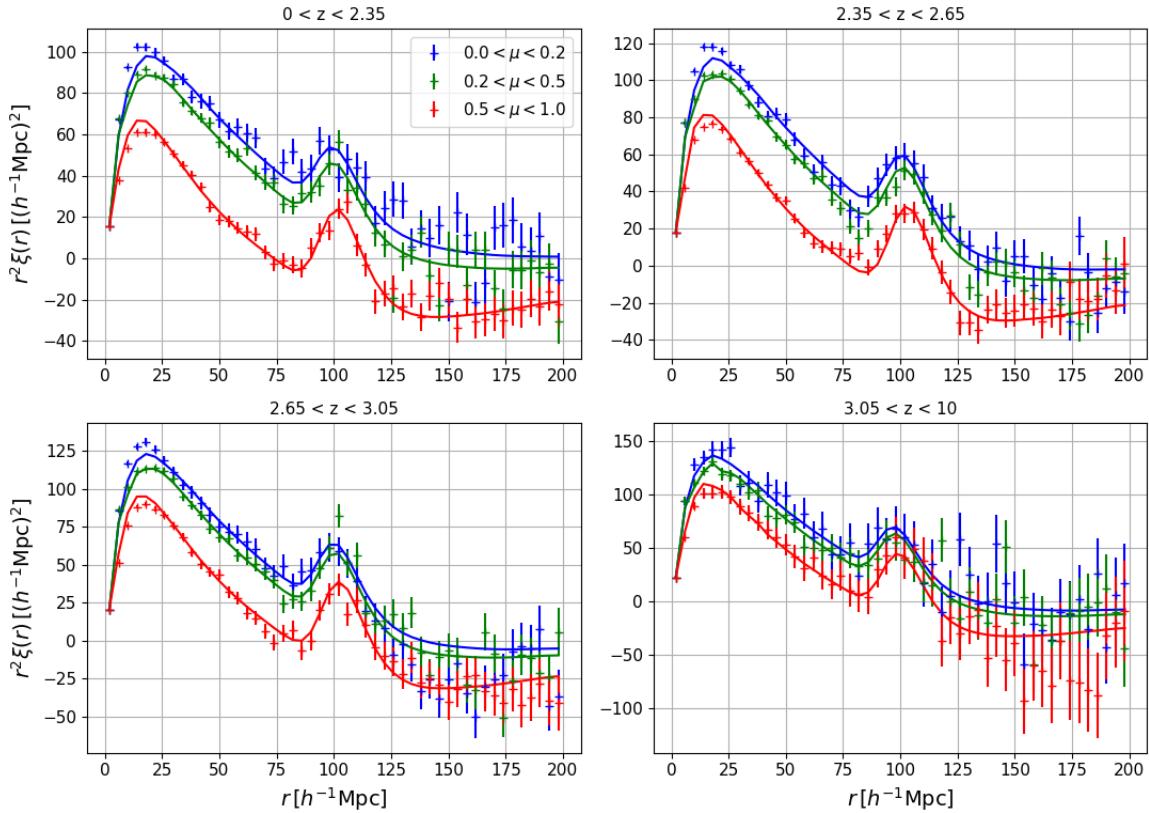


FIGURE 5.9 – L’auto-corrélation QSO×QSO calculée sur 10 réalisations des mocks. Chaque graphique donne la moyenne des dix fonctions de corrélation calculées dans chaque bin en redshift. Les fonctions de corrélation sont présentées dans trois bins en  $\mu$ . Les lignes continues donnent l’ajustement du modèle.

### L’auto-corrélation QSO×QSO

Nous présentons ici la mesure de l’auto-corrélation QSO×QSO, estimée grâce à l’équation 5.10. Comme pour les corrélations Ly $\alpha$ ×Ly $\alpha$  et Ly $\alpha$ ×QSO, la fonction de corrélation QSO×QSO est estimée puis ajustée dans les quatre bins en redshift [0 ; 2,35], [2,35 ; 2,65], [2,65 ; 3,05] et [3,05 ; 10]. La corrélation est estimée à partir de 10 réalisations. La figure 5.9 présente la moyenne des dix fonctions de corrélation et son ajustement dans chaque bin en redshift. Pour chacun des bins, la fonction de corrélation est présentée dans trois bins en  $\mu$ .

Le modèle utilisé pour ajuster la corrélation QSO×QSO est le même que celui utilisé pour ajuster la corrélation Ly $\alpha$ ×Ly $\alpha$ , à la différence que nous n’incluons pas le terme représentant le lissage gaussien. Nous n’incluons pas non plus le terme représentant l’effet produit par la taille non nulle des voxels, mais il serait justifié de le faire. Dans le cas de la corrélation QSO×QSO, contrairement au cas du Ly $\alpha$ , nous avons la relation  $b_{\text{QSO}}\beta_{\text{QSO}} = f$ , où  $f$  est le taux de croissance des structures. Ainsi, plutôt que d’ajuster le paramètre  $b_\eta = b\beta/f$  comme dans le cas du Ly $\alpha$  (équation 4.66), nous fixons le paramètre  $b_{\eta,\text{QSO}} = 1$  et nous ajustons les paramètres  $\beta_{\text{QSO}}$  et  $f$ . Le résultat de l’ajustement dans chaque bin en redshift est donné dans le tableau 5.3. Nous pouvons voir dans ce tableau que les paramètres  $\alpha_{\parallel}$  et  $\alpha_{\perp}$  sont compatibles avec 1 à moins de  $2\sigma$ . Le facteur de croissance  $f$  est lui aussi en accord avec la cosmologie utilisée dans les mocks. Celui-ci varie de 0,960 à  $z = 2,050$  à 0,985 à  $z = 3,257$ .

TABLE 5.3 – Résultats de l’ajustement de l’auto-corrélation QSO×QSO estimée à partir de 10 réalisations des mocks. Chaque colonne donne le résultat de l’ajustement d’un bin en redshift. La dernière colonne donne le résultat de l’ajustement de la corrélation moyennée sur ces quatre bins en redshift. La première section du tableau donne les paramètres du modèle qui sont ajustés. La seconde donne le redshift effectif  $z_{\text{eff}}$  et le  $\chi^2$ . Le nombre de bins sur lesquels le modèle est ajusté est  $N_{\text{bin}} = 1574$ , ce qui donne un nombre de degrés de liberté  $n_{d.o.f.} = 1570$ . La dernière section donne le biais des quasars.

Paramètre	$0 < z < 2,35$	$2,35 < z < 2,65$	$2,65 < z < 3,05$	$3,05 < z < 10$	$0 < z < 10$
$\alpha_{\parallel}$	$0.986 \pm 0.019$	$0.970 \pm 0.016$	$0.978 \pm 0.018$	$1.002 \pm 0.045$	$0.980 \pm 0.010$
$\alpha_{\perp}$	$1.022 \pm 0.012$	$1.010 \pm 0.009$	$1.015 \pm 0.010$	$1.045 \pm 0.024$	$1.016 \pm 0.006$
$\beta_{\text{QSO}}$	$0.314 \pm 0.009$	$0.254 \pm 0.007$	$0.204 \pm 0.008$	$0.180 \pm 0.020$	$0.257 \pm 0.005$
$f$	$0.985 \pm 0.023$	$1.007 \pm 0.023$	$0.962 \pm 0.035$	$1.013 \pm 0.102$	$0.919 \pm 0.015$
$\chi^2$	1739	1612	1561	1021	1546
$z_{\text{eff}}$	2.050	2.488	2.826	3.257	2.273
$b_{\text{QSO}}$	$3.134 \pm 0.020$	$3.968 \pm 0.020$	$4.710 \pm 0.030$	$5.617 \pm 0.094$	$3.570 \pm 0.013$

Sur la figure 5.9, nous pouvons remarquer que le modèle ajusté par `picca` décrit très bien les mocks. Une légère différence est visible à petit  $r$ . Cette différence dépend de  $\mu$  : l’amplitude est trop faible le long de la ligne de visée, et trop importante perpendiculairement à cette dernière. Ceci semble provenir des RSD aux petites échelles. Le graphique de droite de la figure 5.10 confirme cette hypothèse : la corrélation moyennée sur toute la gamme en  $\mu$  est très bien décrite par le modèle de `picca` (ligne continue) ainsi que par la prédiction de cette corrélation (ligne en pointillés).

Enfin, la figure 5.11 présente  $b_{\text{QSO}}$  et  $\beta_{\text{QSO}}$  obtenus avec l’ajustement de la moyenne des dix fonctions de corrélation. La ligne bleue donne la paramétrisation utilisée dans les mocks (équation 1.64). les valeurs ajustées de  $\beta_{\text{QSO}}$  et  $f$  sont corrélées à plus de 99 %. Ceci vient du fait que, pour les quasars, le paramètre RSD est faible. Il y a donc peu de différences entre la corrélation le long de la ligne de visée et perpendiculairement à cette dernière. Malgré cette corrélation, les valeurs obtenues pour  $b_{\text{QSO}}$  et  $\beta_{\text{QSO}}$  sont en très bon accord avec la paramétrisation utilisée pour construire les mocks.

## L’auto-corrélation HCD×HCD

Similairement à l’auto-corrélation QSO×QSO, nous estimons la corrélation HCD×HCD. Cependant, celle-ci est produite dans un unique bin en redshift. Nous estimons la corrélation HCD×HCD sur dix réalisations, puis nous calculons et ajustons la moyenne de ces dix corrélations. La figure 5.12 présente cette fonction de corrélation. Les lignes continues donnent le modèle ajusté par `picca`. La modélisation utilisée pour cette corrélation est la même que celle utilisée pour les quasars. Les deux paramètres ajustés sont donc  $f$  et  $\beta_{\text{HCD}}$ . Le tableau 5.4 donne le résultat de l’ajustement. Le biais des HCD que nous mesurons est  $b_{\text{HCD}} = 2,045 \pm 0,016$ . Cette mesure n’est pas compatible ( $2,8\sigma$ ) avec le biais  $b_{\text{HCD}} = 2$  que nous utilisons pour tirer les HCD. Ce petit écart (2 %) provient probablement d’une mesure pas assez précise de  $\sigma_l$ , utilisé dans le calcul du seuil  $\nu$  auquel nous avons recours pour tirer les HCD (voir section 4.2.5).

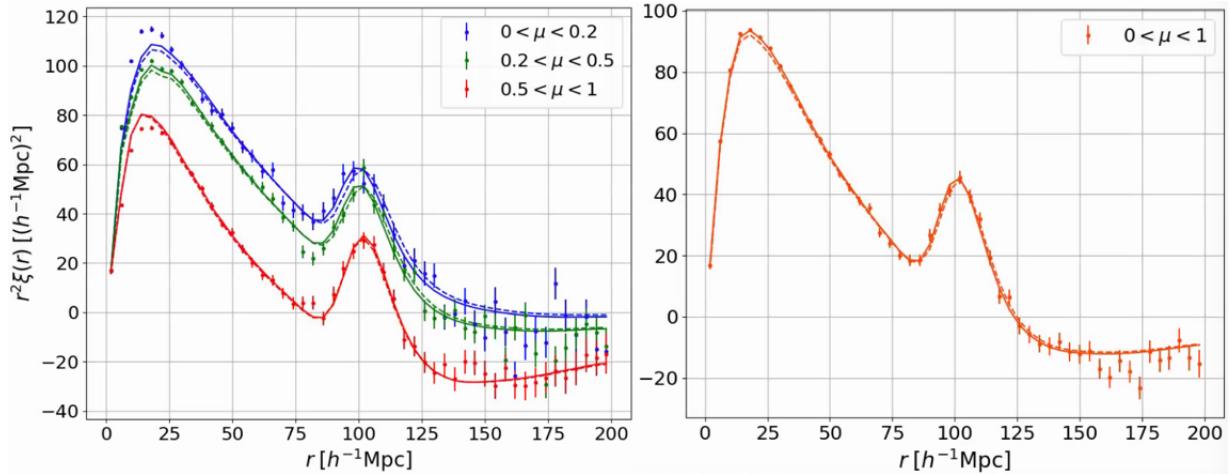


FIGURE 5.10 – L’auto-corrélation QSO×QSO calculée sur 10 réalisations des mocks et moyennée sur les quatre bins en redshift. Le graphique de gauche présente la corrélation dans trois bins en  $\mu$ . Celui de droite présente la corrélation moyennée sur toute la gamme en  $\mu$ . Les lignes continues donnent l’ajustement du modèle. Les lignes en pointillés donnent la prédition.

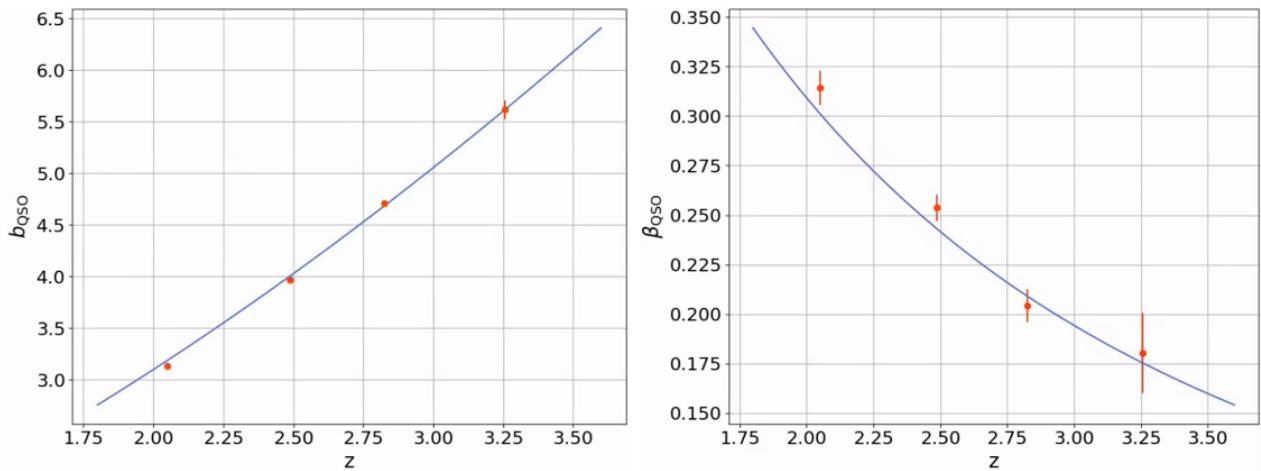


FIGURE 5.11 – Le biais  $b_{\text{QSO}}$  et le paramètre RSD  $\beta_{\text{QSO}}$  mesurés sur l’auto-corrélation QSO×QSO dans les mocks. La mesure est faite dans 4 bins en redshift. La ligne bleue correspond à la paramétrisation utilisée pour construire le relevé de quasars dans les mocks.

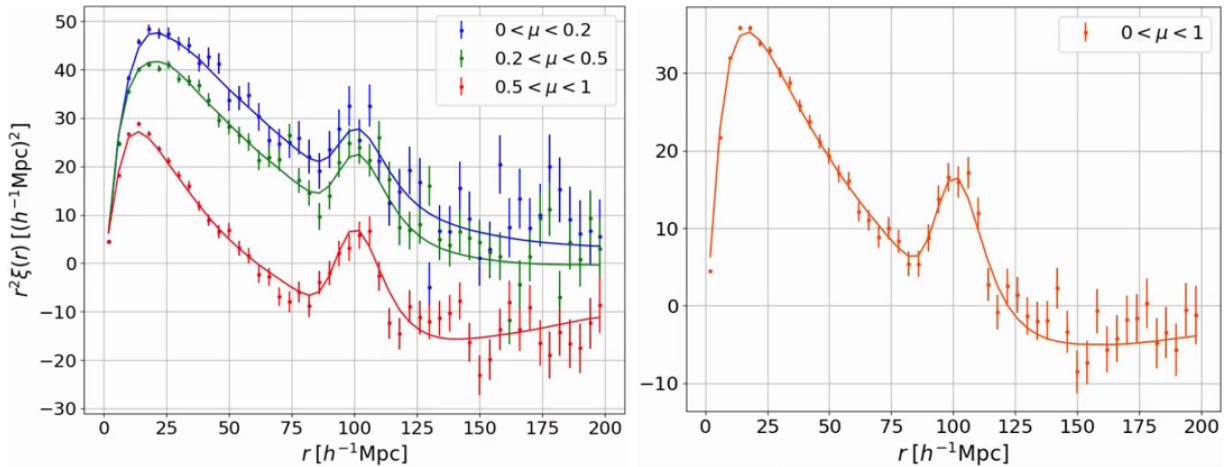


FIGURE 5.12 – L’auto-corrélation HCD $\times$ HCD calculée sur les mocks. La fonction de corrélation sont présentée dans trois bins en  $\mu$ . Les lignes continues donnent l’ajustement du modèle dans chaque bin en  $\mu$ .

TABLE 5.4 – Résultats de l’ajustement de l’auto-corrélation HCD $\times$ HCD calculée sur les mocks. La première section du tableau donne les paramètres du modèle qui sont ajustés. La seconde donne le redshift effectif  $z_{\text{eff}}$  et le  $\chi^2$ . Le nombre de bins sur lesquels le modèle est ajusté est  $N_{\text{bin}} = 1574$ , ce qui donne un nombre de degrés de liberté  $n_{d.o.f.} = 1570$ . La dernière section donne le biais des HCD.

Paramètre	$0 < z < 10$
$\alpha_{\parallel}$	$1.019 \pm 0.020$
$\alpha_{\perp}$	$1.005 \pm 0.014$
$\beta_{\text{HCD}}$	$0.491 \pm 0.012$
$f$	$1.003 \pm 0.018$
$\chi^2$	1802
$z_{\text{eff}}$	2.22
$b_{\text{HCD}}$	$2.045 \pm 0.016$

### 5.5.2 Analyse des mocks eboss-0.0

Maintenant que nous avons vérifié que les raw mocks possèdent les bonnes fonctions de corrélation, nous pouvons analyser les mocks après avoir appliqué `quickquasars`. Nous commençons par présenter l’analyse des mocks eboss-0.0. Contrairement aux raw mocks, nous avons recours à l’ajustement du continuum pour calculer le champ  $\delta_F$ . Ainsi, les fonctions de corrélation possèdent les distorsions liées à cet ajustement, et les modèles sont multipliés par les matrices de distorsions (équation 5.35). De plus, les mocks issus de `quickquasars` contiennent du bruit instrumental. Les fonctions de corrélation sont donc plus bruitées que celles calculées sur les raw mocks. De la même manière que pour les raw mocks, les fonctions de corrélation sont estimées dans les bins en redshift : [0;2,35], [2,35;2,65], [2,65;3,05] et [3,05;10]. Ces fonctions de corrélation sont calculées sur 30 réalisations.

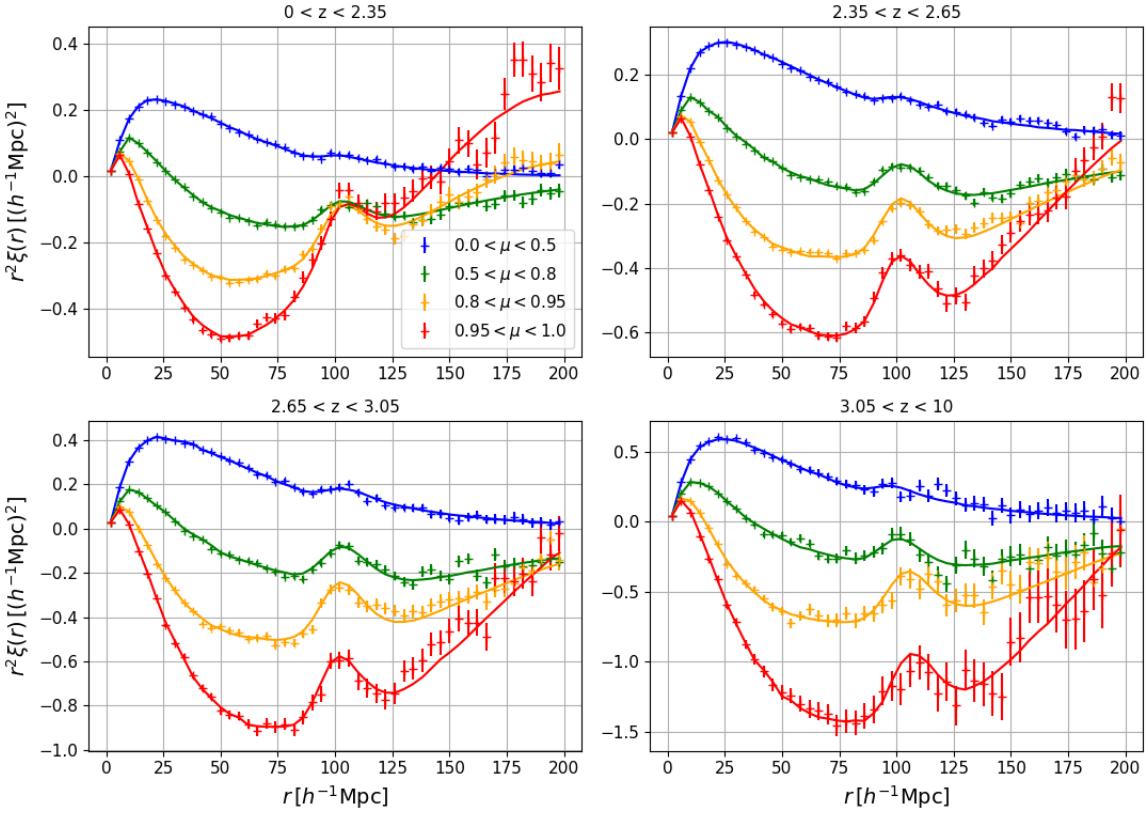


FIGURE 5.13 – L’auto-corrélation  $\text{Ly}\alpha \times \text{Ly}\alpha$  calculée sur les mocks eboss-0.0. Chaque graphique donne la moyenne des fonctions de corrélation calculées dans chaque bins en redshift. Les fonctions de corrélation sont présentées dans quatre bins en  $\mu$ . Les lignes continues donnent l’ajustement du modèle.

### L’auto-corrélation $\text{Ly}\alpha \times \text{Ly}\alpha$

La figure 5.13 présente l’auto-corrélation  $\text{Ly}\alpha \times \text{Ly}\alpha$  calculée sur les mocks eboss-0.0. Chacun des graphiques donne la corrélation dans un des bins en redshift. Pour chaque bins en redshift, la corrélation est présentée dans quatre bins en  $\mu$ . Le modèle ajusté par `picca` décrit très bien l’auto-corrélation  $\text{Ly}\alpha \times \text{Ly}\alpha$  estimée sur les mocks eboss-0.0. l’ajustement du modèle est représenté par des lignes continues. Le résultat des ajustements est donné dans le tableau 5.5. Comme dans le cas des raw mocks, les paramètres  $\alpha_{\parallel}$  et  $\alpha_{\perp}$  sont compatibles avec 1.

La figure 5.16 présente les paramètres  $b_{\text{Ly}\alpha}$  et  $\beta_{\text{Ly}\alpha}$  mesurés dans l’auto-corrélation  $\text{Ly}\alpha \times \text{Ly}\alpha$  estimée à partir des mocks eboss-0.0. Nous pouvons noter un écart statistiquement significatif entre les paramètres  $\text{Ly}\alpha$  mesurés sur les raw mocks (vert) et sur les mocks eboss-0.0 (bleu). Cet écart est probablement dû à la matrice de distorsion qui ne capture pas l’intégralité des effets produits par l’ajustement du continuum.

### La corrélation croisée $\text{Ly}\alpha \times \text{QSO}$

La figure 5.14 présente la corrélation croisée  $\text{Ly}\alpha \times \text{QSO}$  dans chaque bin de redshift. Contrairement à l’auto-corrélation  $\text{Ly}\alpha \times \text{Ly}\alpha$ , la corrélation  $\text{Ly}\alpha \times \text{QSO}$  est estimée pour  $r_{\parallel} \in [-200; 200] h^{-1} \text{Mpc}$ . Ainsi, pour obtenir chaque bin en  $\mu$ , nous moyennons la fonction de corrélation pour les valeurs

TABLE 5.5 – Résultats de l’ajustement de l’auto-corrélation  $\text{Ly}\alpha \times \text{Ly}\alpha$  calculée sur les mocks eboss-0.0. Chaque colonne donne le résultat de l’ajustement d’un bin en redshift. La dernière colonne donne le résultat de l’ajustement de la corrélation moyennée sur ces quatre bins en redshift. La première section du tableau donne les paramètres du modèle qui sont ajustés. La seconde donne le  $\chi^2$  et le redshift effectif  $z_{\text{eff}}$ . Le nombre de bins sur lesquels le modèle est ajusté est  $N_{\text{bin}} = 1574$ , ce qui donne un nombre de degrés de liberté  $n_{d.o.f.} = 1570$ . La dernière section donne le biais et le biais effectif du  $\text{Ly}\alpha$ .

Paramètre	$0 < z < 2,35$	$2,35 < z < 2,65$	$2,65 < z < 3,05$	$3,05 < z < 10$	$0 < z < 10$
$\alpha_{\parallel}$	$0.992 \pm 0.009$	$1.006 \pm 0.007$	$1.005 \pm 0.009$	$0.954 \pm 0.023$	$1.003 \pm 0.005$
$\alpha_{\perp}$	$1.002 \pm 0.016$	$1.000 \pm 0.012$	$0.981 \pm 0.013$	$1.057 \pm 0.033$	$0.995 \pm 0.008$
$b_{\eta, \text{Ly}\alpha}$	$-0.1868 \pm 0.0007$	$-0.2045 \pm 0.0007$	$-0.2381 \pm 0.0012$	$-0.2752 \pm 0.0029$	$-0.2077 \pm 0.0005$
$\beta_{\text{Ly}\alpha}$	$1.962 \pm 0.016$	$1.768 \pm 0.012$	$1.454 \pm 0.013$	$1.172 \pm 0.020$	$1.687 \pm 0.008$
$\chi^2$	1498	1622	1598	1629	1563
$z_{\text{eff}}$	2.118	2.254	2.540	2.867	2.288
$b_{\text{Ly}\alpha}$	$-0.0916 \pm 0.0004$	$-0.1118 \pm 0.0004$	$-0.1595 \pm 0.0007$	$-0.2300 \pm 0.0018$	$-0.1192 \pm 0.0003$
$b_{\text{eff}, \text{Ly}\alpha}$	$-0.1607 \pm 0.0002$	$-0.1873 \pm 0.0002$	$-0.2467 \pm 0.0004$	$-0.3298 \pm 0.0011$	$-0.1956 \pm 0.0002$

positives et négatives de  $\mu$ . Comme pour les autres corrélations, la corrélation  $\text{Ly}\alpha \times \text{QSO}$  est ajustée dans la gamme  $r \in [20; 180] h^{-1} \text{ Mpc}$ , ce qui correspond à un nombre de bins en  $(r_{\parallel}, r_{\perp})$   $N_{\text{bin}} = 3148$ , soit le double du nombre de bins utilisés pour ajuster l’auto-corrélation  $\text{Ly}\alpha \times \text{Ly}\alpha$ .

Comme expliqué dans la section 5.4.1, à cause de la dégénérescence entre les paramètres du  $\text{Ly}\alpha$  et ceux des quasars, nous fixons les paramètres  $b_{\text{QSO}}$  et  $\beta_{\text{QSO}}$ . Ces paramètres sont fixés à la valeur donnée par la paramétrisation (équation 1.64) évaluée au redshift effectif de chaque ajustement. Ceci est validé par la mesure de  $b_{\text{QSO}}$  et  $\beta_{\text{QSO}}$  avec l’auto-corrélation  $\text{QSO} \times \text{QSO}$ , qui est en accord avec la paramétrisation utilisée (voir section 5.5.1). Nous pouvons remarquer sur la figure 5.14 que, comme pour l’auto-corrélation  $\text{Ly}\alpha \times \text{Ly}\alpha$ , le modèle ajusté par `picca` décrit très bien la corrélation croisée  $\text{Ly}\alpha \times \text{QSO}$ .

Le tableau 5.6 présente le résultat de l’ajustement dans chaque bin de redshift. Nous pouvons noter que le paramètre  $\Delta_{r_{\parallel}, \text{QSO}}$  est compatible avec 0, ce à quoi nous nous attendons. Une légère tension ( $2,3\sigma$ ) est visible pour le bin  $2,65 < z < 3,05$ , nous attribuons cela à une fluctuation statistique. Aussi, comme dans le cas de l’auto-corrélation, les paramètres  $\alpha_{\parallel}$  et  $\alpha_{\perp}$  sont compatibles avec 1. Cependant, les paramètres du  $\text{Ly}\alpha$  mesurés dans la corrélation croisée sont en tension avec ceux mesurés dans l’auto-corrélation  $\text{Ly}\alpha \times \text{Ly}\alpha$ . Pourtant, du fait que les paramètres des quasars mesurés sur la corrélation  $\text{QSO} \times \text{QSO}$  sont en accord avec la paramétrisation du biais des quasars utilisée dans les mocks, nous nous attendons à mesurer les mêmes  $b_{\text{Ly}\alpha}$  et  $\beta_{\text{Ly}\alpha}$  dans l’auto-corrélation  $\text{Ly}\alpha \times \text{Ly}\alpha$  et dans la corrélation croisée  $\text{Ly}\alpha \times \text{QSO}$ . Nous avons essayé d’augmenter  $r_{\min}$  à  $30 h^{-1} \text{ Mpc}$  dans l’ajustement des corrélations  $\text{Ly}\alpha \times \text{Ly}\alpha$  et  $\text{Ly}\alpha \times \text{QSO}$ , mais ceci ne réduit pas les tensions observées.

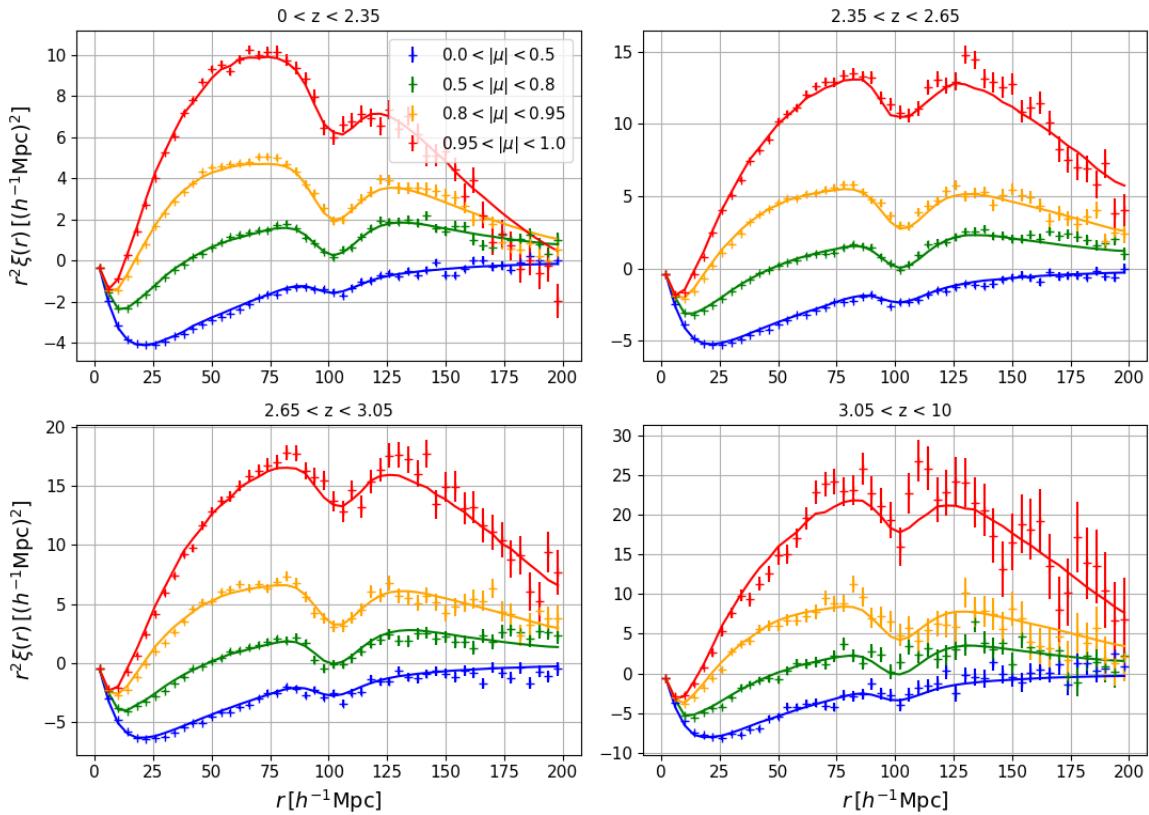


FIGURE 5.14 – La corrélation croisée Ly $\alpha$ ×QSO calculée sur les mocks eboss-0.0. Chaque graphique donne la moyenne des fonctions de corrélation calculées dans chaque bins en redshift. Les fonctions de corrélation sont présentées dans quatre bins en  $\mu$ . Les lignes continues donnent l'ajustement du modèle.

TABLE 5.6 – Résultats de l’ajustement de la corrélation croisée Ly $\alpha$ ×QSO calculée sur les mocks eboss-0.0. Chaque colonne donne le résultat de l’ajustement d’un bin en redshift. La dernière colonne donne le résultat de l’ajustement de la corrélation moyennée sur ces quatre bins en redshift. La première section du tableau donne les paramètres du modèle qui sont ajustés. La seconde donne le  $\chi^2$  et le redshift effectif  $z_{\text{eff}}$ . Le nombre de bins sur lesquels le modèle est ajusté est  $N_{\text{bin}} = 3148$ , ce qui donne un nombre de degrés de liberté  $n_{d.o.f.} = 3143$ . La dernière section donne le biais et le biais effectif du Ly $\alpha$ . Le paramètre  $\Delta r_{\parallel, \text{QSO}}$  est mesuré en  $h^{-1}$  Mpc.

Paramètre	$0 < z < 2,35$	$2,35 < z < 2,65$	$2,65 < z < 3,05$	$3,05 < z < 10$	$0 < z < 10$
$\alpha_{\parallel}$	$1.002 \pm 0.008$	$0.990 \pm 0.008$	$0.986 \pm 0.012$	$1.005 \pm 0.026$	$0.999 \pm 0.005$
$\alpha_{\perp}$	$1.000 \pm 0.010$	$1.001 \pm 0.010$	$1.001 \pm 0.013$	$1.013 \pm 0.031$	$1.000 \pm 0.006$
$b_{\eta, \text{Ly}\alpha}$	$-0.1717 \pm 0.0013$	$-0.1985 \pm 0.0017$	$-0.2333 \pm 0.0030$	$-0.2940 \pm 0.0088$	$-0.1902 \pm 0.0010$
$\beta_{\text{Ly}\alpha}$	$1.787 \pm 0.027$	$1.564 \pm 0.024$	$1.344 \pm 0.029$	$1.208 \pm 0.058$	$1.564 \pm 0.015$
$\Delta r_{\parallel, \text{QSO}}$	$-0.051 \pm 0.053$	$0.039 \pm 0.058$	$-0.182 \pm 0.085$	$0.213 \pm 0.192$	$-0.033 \pm 0.035$
$\chi^2$	3233	3410	3238	3410	3502
$z_{\text{eff}}$	2.128	2.362	2.663	3.043	2.301
$b_{\text{Ly}\alpha}$	$-0.0925 \pm 0.0008$	$-0.1231 \pm 0.001$	$-0.1696 \pm 0.0018$	$-0.2390 \pm 0.0054$	$-0.1177 \pm 0.0006$
$b_{\text{eff}, \text{Ly}\alpha}$	$-0.1556 \pm 0.0005$	$-0.1959 \pm 0.0007$	$-0.2548 \pm 0.0012$	$-0.3462 \pm 0.0036$	$-0.1874 \pm 0.0004$

### 5.5.3 Analyse des mocks eboss-0.2

Nous analysons à présent les mocks eboss-0.2. Ces mocks sont obtenus comme les mocks eboss-0.0, analysés précédemment, à la différence que le code `quickquasars` inclue les HCD dans les spectres synthétiques. Comme pour les données, nous masquons les HCD pour lesquels  $\log n_{\text{HI}} > 20,3$ . Cependant, dans le cas des mocks, le masquage s’effectue à partir du vrai catalogue de HCD. Lors de l’ajustement des fonctions de corrélation, nous modélisons l’effet des HCD non masqués sur ces corrélations. Nous présentons dans la section 6.2.2 l’analyse d’une réalisation eboss-0.2 où les DLA ont été masqués en utilisant l’algorithme d’identification utilisé pour les données DR16.

#### L’auto-corrélation Ly $\alpha$ ×Ly $\alpha$

La figure 5.15 présente les fonctions d’auto-corrélation Ly $\alpha$ ×Ly $\alpha$  dans chaque bin en redshift estimées à partir des mocks eboss-0.2. Comme précédemment, les corrélations sont présentées dans quatre bins en  $\mu$ . Les lignes continues donnent l’ajustement du modèle produit par `picca`. Le tableau 5.7 donne le résultat de l’ajustement dans chaque bin en redshift. A cause du masquage des DLA, le nombre de paires de pixels utilisées pour estimer la corrélation Ly $\alpha$ ×Ly $\alpha$  est réduit, ce qui résulte dans une augmentation légère des barres d’erreurs. Le biais des HCD mesuré dans chaque bin en redshift est compatible avec 0. Mais lorsque nous ajustons la moyenne pondérée des quatre bins en redshift, nous mesurons  $b_{\text{HCD}} = -0,0082 \pm 0,0008$ . Par ailleurs, nous n’observons pas de changement significatif lorsque nous utilisons  $r_{\min} = 10 h^{-1}$  Mpc dans cet ajustement.

La figure 5.16 présente les paramètres  $b_{\text{Ly}\alpha}$  et  $\beta_{\text{Ly}\alpha}$  mesurés dans l’auto-corrélation Ly $\alpha$ ×Ly $\alpha$  estimée à partir des mocks eboss-0.2. Nous pouvons remarquer que la mesure de ces paramètres est très affectée par la présence des HCD. Ceci nous laisse croire que la mesure de  $b_{\text{Ly}\alpha}$  et  $\beta_{\text{Ly}\alpha}$  dans les données DR16, sur laquelle nous avons fondé l’ajustement de nos mocks, n’est pas robuste. Ceci est étudié en détail dans le chapitre suivant.

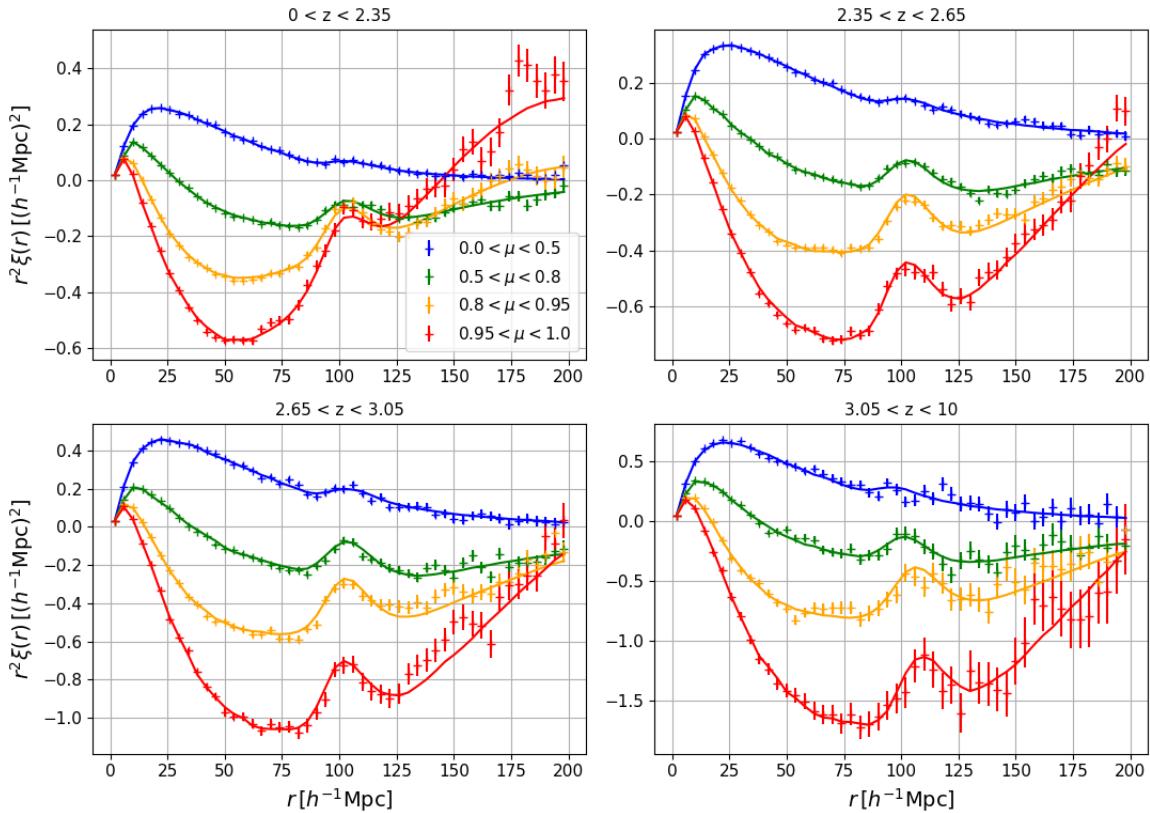


FIGURE 5.15 – L’auto-corrélation Ly $\alpha$  × Ly $\alpha$  calculée sur les mocks eboss-0.2. Chaque graphique donne la moyenne des fonctions de corrélation calculées dans chaque bins en redshift. Les fonctions de corrélation sont présentées dans quatre bins en  $\mu$ . Les lignes continues donnent l’ajustement du modèle.

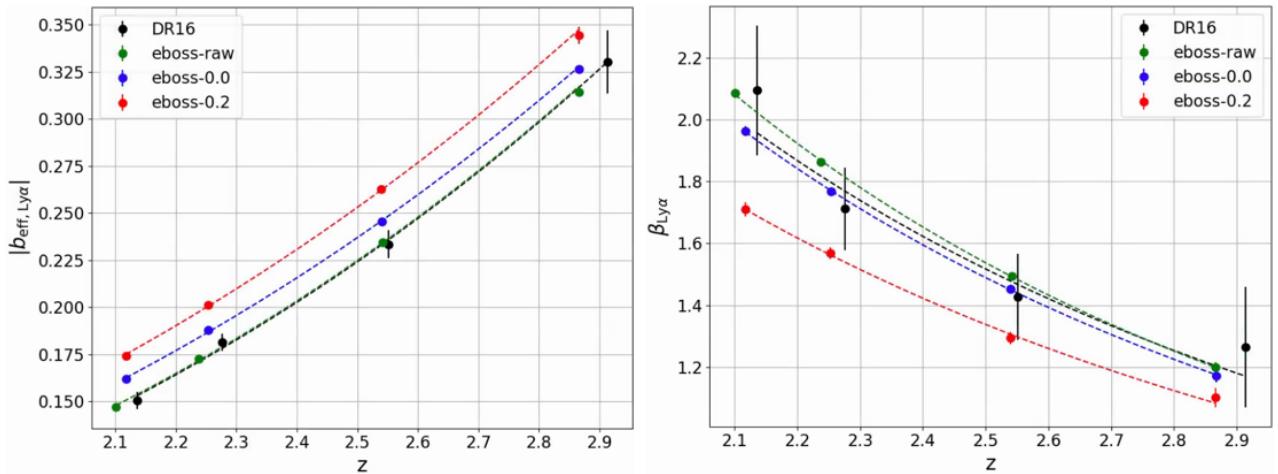


FIGURE 5.16 – Mesures des paramètres  $b_{\text{Ly}\alpha}$  et  $\beta_{\text{Ly}\alpha}$  faites avec les auto-corrélations Ly $\alpha$  × Ly $\alpha$  estimées dans chaque bin en redshift à partir des données DR16 (noir), des raw mocks (vert), des mocks eboss-0.0 (bleu) et des mocks eboss-0.2 (rouge). Pour chacun de ces jeux de données, les barres d’erreur sont représentées, cependant, dans le cas des mocks, elles sont souvent plus petites que la taille des points.

TABLE 5.7 – Résultats de l’ajustement de l’auto-corrélation Ly $\alpha$ ×Ly $\alpha$  calculée sur les mocks eboss-0.2. Chaque colonne donne le résultat de l’ajustement d’un bin en redshift. La dernière colonne donne le résultat de l’ajustement de la corrélation moyennée sur ces quatre bins en redshift. La première section du tableau donne les paramètres du modèle qui sont ajustés. La seconde donne le  $\chi^2$  et le redshift effectif  $z_{\text{eff}}$ . Le nombre de bins sur lesquels le modèle est ajusté est  $N_{\text{bin}} = 1574$ , ce qui donne un nombre de degrés de liberté  $n_{d.o.f.} = 1568$ . La dernière section donne le biais et le biais effectif du Ly $\alpha$ .

Paramètre	$0 < z < 2,35$	$2,35 < z < 2,65$	$2,65 < z < 3,05$	$3,05 < z < 10$	$0 < z < 10$
$\alpha_{  }$	$1.005 \pm 0.009$	$0.996 \pm 0.007$	$1.004 \pm 0.009$	$0.936 \pm 0.021$	$1.000 \pm 0.005$
$\alpha_{\perp}$	$0.985 \pm 0.016$	$1.000 \pm 0.012$	$0.978 \pm 0.013$	$1.066 \pm 0.039$	$0.989 \pm 0.008$
$b_{\eta, \text{Ly}\alpha}$	$-0.1860 \pm 0.0008$	$-0.2042 \pm 0.0008$	$-0.2365 \pm 0.0014$	$-0.2782 \pm 0.0035$	$-0.2063 \pm 0.0007$
$\beta_{\text{Ly}\alpha}$	$1.709 \pm 0.023$	$1.568 \pm 0.019$	$1.295 \pm 0.020$	$1.102 \pm 0.032$	$1.548 \pm 0.014$
$b_{\text{HCD}}$	$-0.0015 \pm 0.0015$	$-0.0017 \pm 0.0014$	$-0.0050 \pm 0.0024$	$-0.0085 \pm 0.0060$	$-0.0080 \pm 0.0011$
$\beta_{\text{HCD}}$	$0.502 \pm 0.090$	$0.500 \pm 0.090$	$0.501 \pm 0.090$	$0.500 \pm 0.090$	$0.487 \pm 0.089$
$\chi^2$	1535	1579	1627	1650	1573
$z_{\text{eff}}$	2.117	2.253	2.539	2.866	2.281
$b_{\text{Ly}\alpha}$	$-0.1048 \pm 0.0012$	$-0.1259 \pm 0.0012$	$-0.1779 \pm 0.0021$	$-0.2474 \pm 0.0053$	$-0.1290 \pm 0.0009$
$b_{\text{eff,Ly}\alpha}$	$-0.1729 \pm 0.0011$	$-0.2006 \pm 0.0010$	$-0.2638 \pm 0.0018$	$-0.3479 \pm 0.0047$	$-0.2044 \pm 0.0007$

### La corrélation croisée Ly $\alpha$ ×QSO

La figure 5.17 présente la mesure de la corrélation croisée Ly $\alpha$ ×QSO. Celle-ci est estimée sur 30 réalisations eboss-0.2 et dans quatre bins en redshift. Pour chacun des bins en redshift, la fonction de corrélation est présentée dans quatre bins en  $\mu$ . Les lignes continues donnent l’ajustement du modèle produit par `picca`. Il décrit correctement la corrélation estimée dans les mocks. Le tableau 5.8 donne le résultat des ajustements dans chaque bin en redshift. Comme pour les mocks eboss-0.0, la mesure de  $\Delta r_{||, \text{QSO}}$  est compatible avec 0. Nous pouvons remarquer dans ce tableau que, comme pour les mocks eboss-0.0, la mesure des paramètres Ly $\alpha$  dans la corrélation croisée Ly $\alpha$ ×QSO est en tension avec la mesure faite sur l’auto-corrélation Ly $\alpha$ ×Ly $\alpha$ . Aussi, les mesures de  $b_{\text{HCD}}$  faites sur les corrélations Ly $\alpha$ ×Ly $\alpha$  et Ly $\alpha$ ×QSO ne sont pas compatibles. Le biais des HCD mesuré dans la corrélation croisée est beaucoup plus important. L’ajustement de la moyenne pondérée des quatre bins en redshift donne  $b_{\text{HCD}} = -0,0377 \pm 0,0020$ , comparé à  $b_{\text{HCD}} = -0,0082 \pm 0,0008$  mesuré dans l’auto-corrélation. Nous pensons que cet écart provient du fait qu’il est difficile lors de l’ajustement du modèle de distinguer l’effet du Ly $\alpha$  de celui des HCD. Une partie des HCD est interprétée comme du Ly $\alpha$ , et vice versa. La proportion de HCD et de Ly $\alpha$  identifiés dans les corrélations Ly $\alpha$ ×Ly $\alpha$  et Ly $\alpha$ ×QSO est différente. Ceci est investigué plus en détails dans le chapitre suivant. Nous pouvons cependant remarquer que les quantités  $b_{\text{Ly}\alpha} + b_{\text{HCD}}$  et  $b_{\text{Ly}\alpha}\beta_{\text{Ly}\alpha} + b_{\text{HCD}}\beta_{\text{HCD}}$  (équation 5.26) mesurés dans les corrélations Ly $\alpha$ ×Ly $\alpha$  et Ly $\alpha$ ×QSO sont comparables aux données DR16. Ces mesures sont présentées sur la figure 5.18.

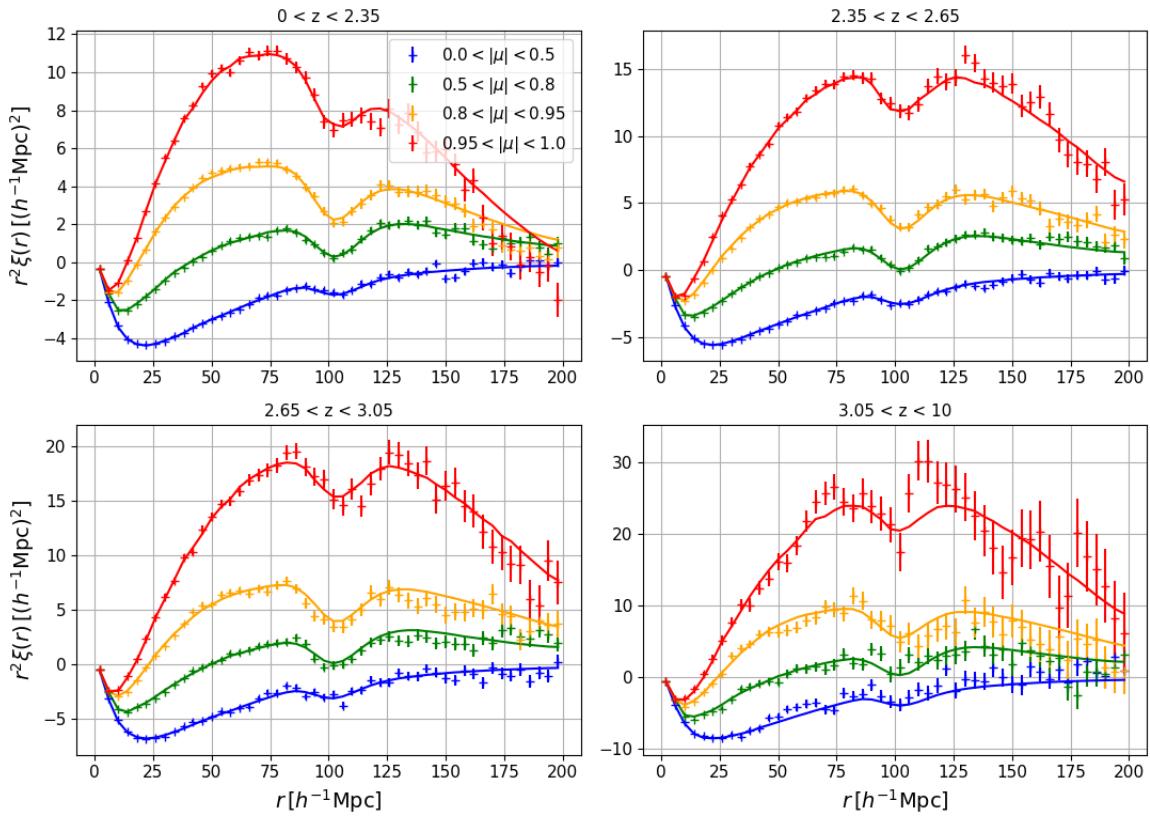


FIGURE 5.17 – La corrélation croisée Ly $\alpha$  $\times$ QSO calculée sur les mocks eboss-0.2. Chaque graphique donne la moyenne des fonctions de corrélation calculées dans chaque bins en redshift. Les fonctions de corrélation sont présentées dans quatre bins en  $\mu$ . Les lignes continues donnent l'ajustement du modèle.

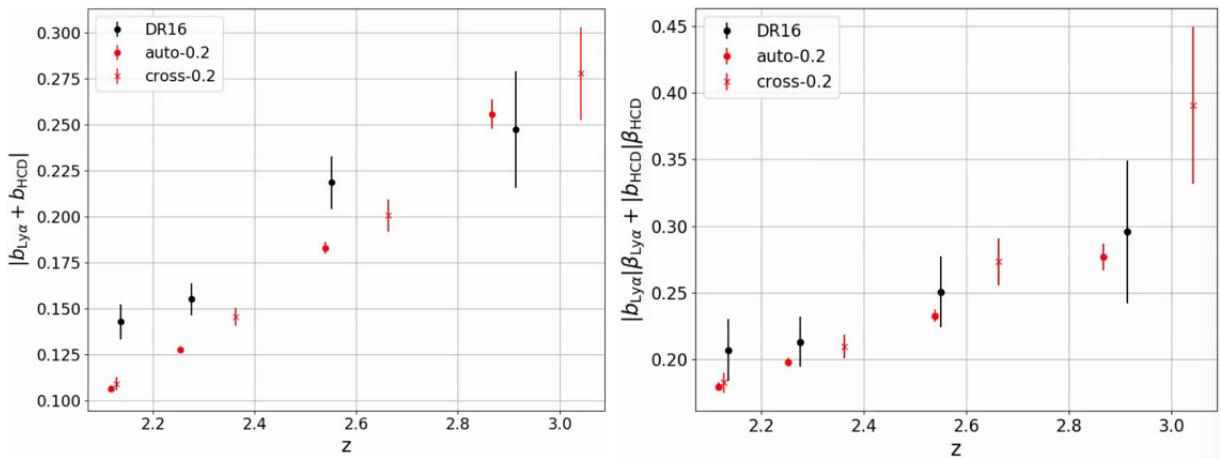


FIGURE 5.18 – Mesures des paramètres  $b_{\text{Ly}\alpha} + b_{\text{HCD}}$  et  $|b_{\text{Ly}\alpha}| \beta_{\text{Ly}\alpha} + |b_{\text{HCD}}| \beta_{\text{HCD}}$  dans la corrélation Ly $\alpha$  $\times$ Ly $\alpha$  estimée à partir des données DR16 (noir), dans la corrélation Ly $\alpha$  $\times$ Ly $\alpha$  estimée à partir des mocks eboss-0.2 (points rouges) et dans la corrélation Ly $\alpha$  $\times$ QSO estimée à partir des mocks eboss-0.2 (croix rouges).

TABLE 5.8 – Résultats de l’ajustement de la corrélation croisée Ly $\alpha$ ×QSO calculée sur les mocks eboss-0.2. Chaque colonne donne le résultat de l’ajustement d’un bin en redshift. La dernière colonne donne le résultat de l’ajustement de la corrélation moyennée sur ces quatre bins en redshift. La première section du tableau donne les paramètres du modèle qui sont ajustés. La seconde donne le  $\chi^2$  et le redshift effectif  $z_{\text{eff}}$ . Le nombre de bins sur lesquels le modèle est ajusté est  $N_{\text{bin}} = 3148$ , ce qui donne un nombre de degrés de liberté  $n_{d.o.f.} = 3141$ . La dernière section donne le biais et le biais effectif du Ly $\alpha$ .

Paramètre	$0 < z < 2,35$	$2,35 < z < 2,65$	$2,65 < z < 3,05$	$3,05 < z < 10$	$0 < z < 10$
$\alpha_{  }$	$1.004 \pm 0.008$	$0.996 \pm 0.009$	$0.987 \pm 0.013$	$1.014 \pm 0.027$	$1.000 \pm 0.005$
$\alpha_{\perp}$	$0.991 \pm 0.011$	$0.992 \pm 0.010$	$1.005 \pm 0.014$	$0.991 \pm 0.029$	$0.994 \pm 0.006$
$b_{\eta, \text{Ly}\alpha}$	$-0.1751 \pm 0.0017$	$-0.2008 \pm 0.0022$	$-0.2427 \pm 0.0044$	$-0.3276 \pm 0.0116$	$-0.1929 \pm 0.0017$
$\beta_{\text{Ly}\alpha}$	$1.991 \pm 0.062$	$1.656 \pm 0.055$	$1.775 \pm 0.093$	$2.277 \pm 0.304$	$1.843 \pm 0.041$
$b_{\text{HCD}}$	$-0.0244 \pm 0.0027$	$-0.0277 \pm 0.0037$	$-0.0672 \pm 0.0066$	$-0.1368 \pm 0.0192$	$-0.0377 \pm 0.0020$
$\beta_{\text{HCD}}$	$0.583 \pm 0.085$	$0.545 \pm 0.087$	$0.540 \pm 0.085$	$0.502 \pm 0.088$	$0.638 \pm 0.078$
$\Delta r_{  , \text{QSO}}$	$-0.053 \pm 0.054$	$-0.037 \pm 0.060$	$-0.178 \pm 0.087$	$0.193 \pm 0.193$	$-0.063 \pm 0.036$
$\chi^2$	3223	3283	3157	3316	3264
$z_{\text{eff}}$	2.127	2.362	2.663	3.043	2.300
$b_{\text{Ly}\alpha}$	$-0.0846 \pm 0.0023$	$-0.1176 \pm 0.0032$	$-0.1335 \pm 0.0057$	$-0.1413 \pm 0.0164$	$-0.1013 \pm 0.0017$
$b_{\text{eff}, \text{Ly}\alpha}$	$-0.1495 \pm 0.0019$	$-0.1915 \pm 0.0027$	$-0.2239 \pm 0.0048$	$-0.2665 \pm 0.0136$	$-0.1728 \pm 0.0014$

### 5.5.4 Analyse des mocks eboss-0.3

Cette section présente l’analyse des mocks eboss-0.3. Ces mocks sont obtenus comme les mocks eboss-0.2, analysés précédemment, à la différence que le code `quickquasars` ajoute les métaux dans les spectres synthétiques. Nous incluons donc les termes  $\xi_{m \times n}$  et  $\xi_{m \times \text{QSO}}$  dans la modélisation des fonctions de corrélation. Pour l’instant, ces mocks n’ont pas été analysés car l’ajustement des paramètres  $A_m$  liés à chaque métal (voir section 4.5) n’est pas terminé. Nous présentons uniquement la fonction de corrélation à une dimension au stade actuel de cet ajustement.

#### La fonction de corrélation à une dimension

La figure 5.19 présente la fonction de corrélation à une dimension obtenue à partir d’une réalisation des mocks eboss-0.3. Les points bleus donnent la corrélation calculée dans les données DR16. Les pics visibles dans cette corrélation sont dus aux métaux. Ces pics dans les mocks n’ont pas exactement la même amplitude que ceux dans les données. Ceci est dû au fait que les paramètres  $A_m$  liés à chaque métal ne sont pas encore parfaitement ajustés. Cependant, nous pouvons noter que la forme global de la corrélation à une dimension estimée à partir des mocks eboss-0.3 est en relativement bon accord avec celle estimée à partir des données DR16.

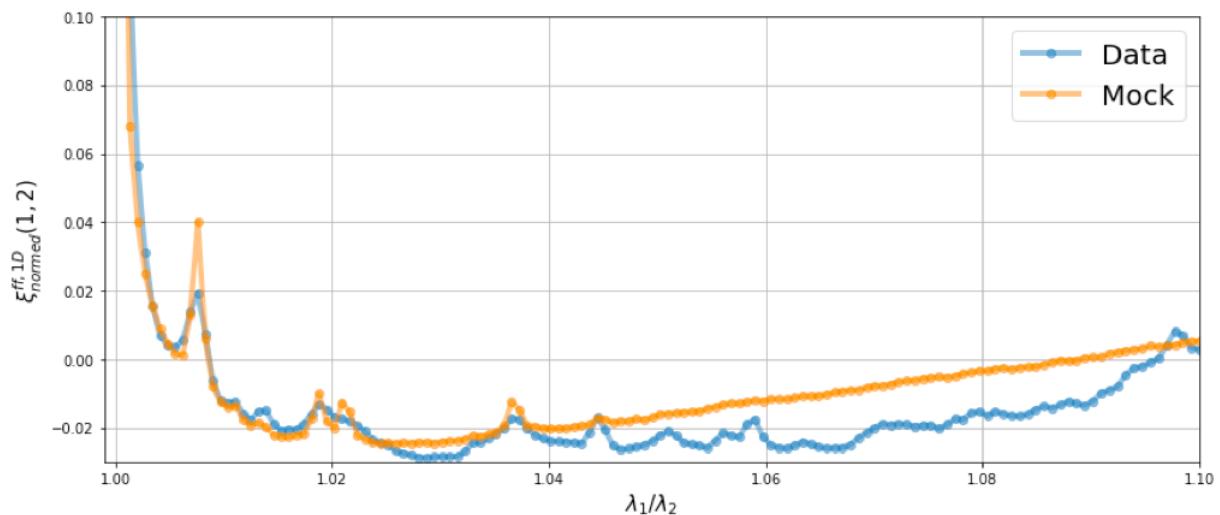


FIGURE 5.19 – Fonctions de corrélation à une dimension obtenues à partir d'une réalisation des mocks eboss-0.3 (orange) et à partir des données DR16 (bleu).



# 6

## Analyse des données et étude des systématiques

---

### Sommaire

---

<b>6.1</b>	<b>L'analyse des données DR16</b>	<b>140</b>
6.1.1	Résultats	140
6.1.2	Stabilité des paramètres Ly $\alpha$	143
<b>6.2</b>	<b>Étude de la modélisation des HCD</b>	<b>144</b>
6.2.1	Comparaison des mocks	145
6.2.2	Effet du masquage des HCD	146
6.2.3	Une meilleure modélisation des HCD	149
6.2.4	Mieux comprendre les HCD	153
	L'effet des HCD sur la fonction de corrélation	153
	Des HCD avec une même densité de colonne	156

---

Dans ce chapitre, nous présentons les diverses analyses que nous avons menées sur les données, avec les mocks comme support de référence. Un élément clé pour la construction des mocks a été de déterminer quels paramètres Ly $\alpha$  nous souhaitions avoir dans nos mocks. Ceci nous a conduit à mener une analyse des données DR16 dans quatre bins en redshift. En produisant cette analyse, nous nous sommes rendus compte que les paramètres Ly $\alpha$  obtenus dépendaient fortement de la modélisation des HCD. Nous avons dû faire un choix quant à cette modélisation. Nous présentons donc d'abord l'analyse des données qui a servi de référence pour l'ajustement des paramètres des mocks. Puis, nous discutons la modélisation des HCD et présentons une meilleure modélisation.

## 6.1 L'analyse des données DR16

### 6.1.1 Résultats

L'analyse des données finale d'eBOSS (DR16), dont nous avons déjà parlé et qui est présentée dans DU MAS DES BOURBOUX et al. (2020), utilise les fonctions de corrélation Ly $\alpha$ (Ly $\alpha$ ) $\times$ Ly $\alpha$ (Ly $\alpha$ ), Ly $\alpha$ (Ly $\alpha$ ) $\times$ Ly $\alpha$ (Ly $\beta$ ), Ly $\alpha$ (Ly $\alpha$ ) $\times$ QSO et Ly $\alpha$ (Ly $\beta$ ) $\times$ QSO. Ces fonctions de corrélation sont estimées sur la gamme en redshift complète, les paramètres ajustés sont donc donnés uniquement pour le redshift effectif  $z_{\text{eff}} = 2,334$  de la mesure. L'appendice F de DU MAS DES BOURBOUX et al. (2020) présente cependant l'analyse des données DR16 dans deux bins en redshift. Mais ces deux bins ne sont pas suffisants pour estimer l'évolution des paramètres Ly $\alpha$  dans la gamme en redshift  $1,9 < z < 3,6$ . Afin d'estimer  $b_{\text{eff}, \text{Ly}\alpha}(z)$  et  $\beta_{\text{Ly}\alpha}(z)$  dans cette gamme, nous avons analysé les données DR16 dans quatre bins en redshift. De manière à limiter les potentielles systématiques, nous nous limitons à l'analyse de la fonction de corrélation Ly $\alpha$ (Ly $\alpha$ ) $\times$ Ly $\alpha$ (Ly $\alpha$ ) (abrégée en Ly $\alpha$  $\times$ Ly $\alpha$  dans la suite de ce chapitre). Pour constituer chacun des bins en redshift, nous pourrions séparer les paires de pixels selon leur redshift effectif. Cependant, à cause de l'ajustement du continuum, cette stratégie induit des corrélations parasites entre bin en redshift lorsqu'une forêt se trouve dans deux bins en redshift à la fois. Pour palier ce problème, nous divisons l'échantillon de forêts selon le redshift des quasars (voir Appendice B de DE SAINTE AGATHE et al. (2019)). Les quatre intervalles choisis pour construire les bins en redshift des quasars sont  $[0; 2,35]$ ,  $[2,35; 2,65]$ ,  $[2,65; 3,05]$  et  $[3,05; 10]$ . Dans chacun des bins, nous calculons la fonction de corrélation Ly $\alpha$  $\times$ Ly $\alpha$ , ainsi que la matrice de distorsion et la matrice des métaux. Enfin, nous procédons à l'ajustement des quatre fonctions de corrélation. Le modèle utilisé pour cet ajustement est le même que celui utilisé pour l'analyse des données finale d'eBOSS, il est présenté dans la section 5.4.1. Le modèle est ajusté pour  $10 \leq r \leq 180 h^{-1} \text{ Mpc}$ . Chacune des fonctions de corrélation est ajustée au redshift effectif de la mesure. Ces redshifts sont  $z_1 = 2,136$ ,  $z_2 = 2,276$ ,  $z_3 = 2,551$  et  $z_4 = 2,914$ .

La figure 6.1 présente la fonction de corrélation et le meilleur ajustement du modèle dans chacun des bins en redshift. Les différents graphiques montrent les différents bins en redshift. Dans chaque graphique, la fonction de corrélation est affichée dans quatre bins en  $\mu$ . Le tableau 6.1 donne le résultat de l'ajustement dans chacun des bins en redshift, ainsi que l'ajustement fait sur la corrélation moyennée sur ces bins. Une fois cette analyse produite, et toujours dans le but d'obtenir  $b_{\text{eff}, \text{Ly}\alpha}(z)$  et  $\beta_{\text{Ly}\alpha}(z)$  pour  $1,9 < z < 3,6$ , nous ajustons les paramètres Ly $\alpha$  mesurés dans les données par une loi de

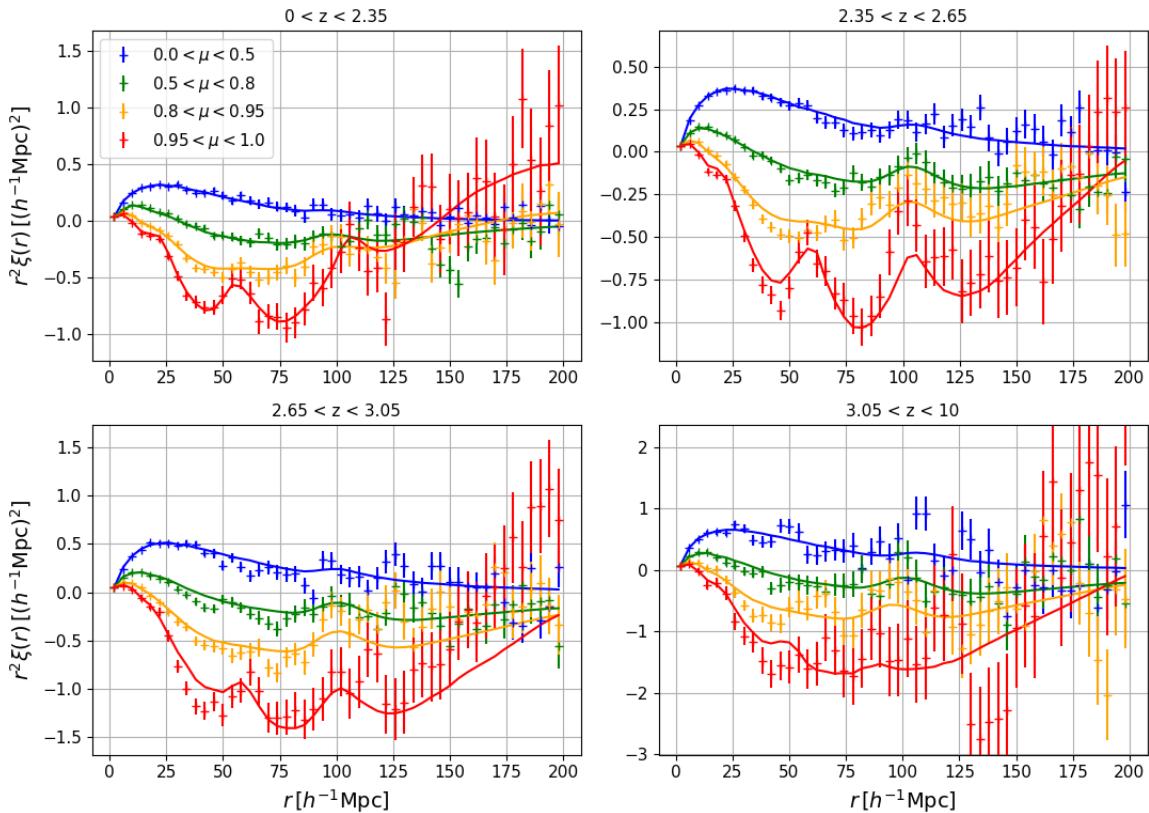


FIGURE 6.1 – Fonctions de corrélation Ly $\alpha$  × Ly $\alpha$  estimées à partir des données DR16. Les courbes en trait plein donne le meilleur ajustement du modèle obtenu avec `picca`. Chaque graphique correspond à un bin en redshift. Pour chacun des bins, la fonction de corrélation et l'ajustement sont montrés dans quatre bins en  $\mu$ .

TABLE 6.1 – Résultats de l’ajustement fait avec **picca** des fonctions de corrélation Ly $\alpha$  × Ly $\alpha$  estimées à partir des données DR16. Chaque colonne donne le résultat de l’ajustement d’un bin en redshift. La dernière colonne donne le résultat de l’ajustement de la corrélation moyennée sur ces quatre bins en redshift. La première section du tableau donne les paramètres du modèle qui sont ajustés. La seconde donne le  $\chi^2$  et le redshift effectif  $z_{\text{eff}}$ . Le modèle comporte 13 paramètres libres. Le nombre de bins sur lesquels le modèle est ajusté est  $N_{\text{bin}} = 1590$ , ce qui donne un nombre de degrés de liberté  $n_{d.o.f.} = 1577$ . Enfin, la dernière section donne le biais et le biais effectif du Ly $\alpha$ . Ils sont reliés aux paramètres  $b_{\eta, \text{Ly}\alpha}$  et  $\beta_{\text{Ly}\alpha}$  par les équations 4.66 et 4.67.

Paramètre	$0 < z < 2,35$	$2,35 < z < 2,65$	$2,65 < z < 3,05$	$3,05 < z < 10$	$0 < z < 10$
$\alpha_{\parallel}$	$1.063 \pm 0.066$	$1.019 \pm 0.041$	$1.029 \pm 0.072$	$1.120 \pm 0.081$	$1.047 \pm 0.034$
$\alpha_{\perp}$	$1.063 \pm 0.108$	$0.965 \pm 0.057$	$1.016 \pm 0.058$	$0.926 \pm 0.072$	$0.980 \pm 0.042$
$b_{\eta, \text{Ly}\alpha}$	$-0.1796 \pm 0.0058$	$-0.1938 \pm 0.0053$	$-0.2237 \pm 0.0084$	$-0.2929 \pm 0.0187$	$-0.1998 \pm 0.0039$
$\beta_{\text{Ly}\alpha}$	$2.094 \pm 0.210$	$1.711 \pm 0.133$	$1.427 \pm 0.138$	$1.265 \pm 0.194$	$1.633 \pm 0.087$
$10^3 b_{\eta, \text{SiII}(1190)}$	$-1.83 \pm 1.10$	$-3.66 \pm 0.68$	$-2.80 \pm 1.01$	$0.36 \pm 1.64$	$-3.02 \pm 0.51$
$10^3 b_{\eta, \text{SiII}(1193)}$	$-4.83 \pm 1.10$	$-1.94 \pm 0.69$	$-0.79 \pm 0.97$	$-2.13 \pm 1.72$	$-2.07 \pm 0.50$
$10^3 b_{\eta, \text{SiII}(1260)}$	$-3.38 \pm 1.33$	$-1.97 \pm 0.80$	$-1.32 \pm 1.05$	$0.90 \pm 1.79$	$-2.16 \pm 0.63$
$10^3 b_{\eta, \text{SiIII}(1207)}$	$-7.87 \pm 1.10$	$-4.52 \pm 0.75$	$-2.11 \pm 1.05$	$-2.89 \pm 1.74$	$-4.59 \pm 0.52$
$10^3 b_{\eta, \text{CIV}(\text{eff})}$	$-4.77 \pm 2.54$	$-5.15 \pm 2.64$	$-5.06 \pm 2.62$	$-5.02 \pm 2.61$	$-5.12 \pm 2.63$
$b_{\text{HCD}}$	$-0.0596 \pm 0.0070$	$-0.0452 \pm 0.0060$	$-0.0665 \pm 0.0100$	$-0.0228 \pm 0.0218$	$-0.0521 \pm 0.0045$
$\beta_{\text{HCD}}$	$0.551 \pm 0.086$	$0.560 \pm 0.086$	$0.508 \pm 0.088$	$0.502 \pm 0.090$	$0.603 \pm 0.083$
$10^2 A_{\text{sky}}$	$1.585 \pm 0.098$	$0.870 \pm 0.082$	$0.729 \pm 0.133$	$0.646 \pm 0.338$	$0.947 \pm 0.060$
$\sigma_{\text{sky}}$	$32.5 \pm 1.8$	$31.6 \pm 2.6$	$31.9 \pm 4.3$	$34.1 \pm 16.0$	$31.4 \pm 1.7$
$\chi^2$	1568	1512	1681	1675	1602
$z_{\text{eff}}$	2.136	2.276	2.551	2.914	2.334
$b_{\text{Ly}\alpha}$	$-0.0832 \pm 0.0065$	$-0.1099 \pm 0.0063$	$-0.1521 \pm 0.0103$	$-0.2247 \pm 0.0230$	$-0.1187 \pm 0.0046$
$b_{\text{eff}, \text{Ly}\alpha}$	$-0.1506 \pm 0.0046$	$-0.1814 \pm 0.0045$	$-0.2336 \pm 0.0074$	$-0.3305 \pm 0.0168$	$-0.1922 \pm 0.0033$

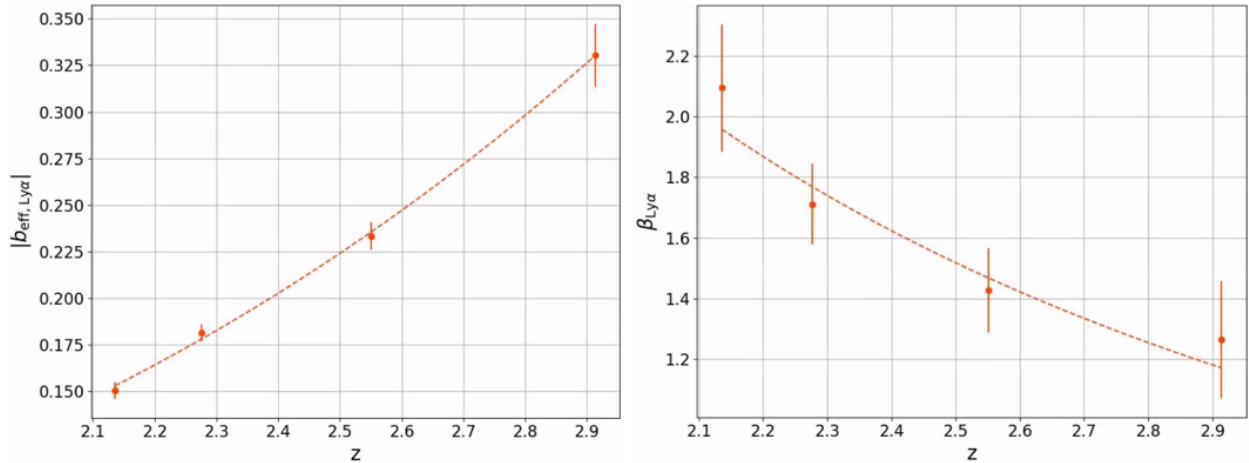


FIGURE 6.2 – Mesure des paramètres  $b_{\text{eff}, \text{Ly}\alpha}$  et  $\beta_{\text{Ly}\alpha}$  dans les données DR16. Les mesures sont faites dans quatre bins en redshift, indiquées par les points. La ligne en pointillés donne le meilleur ajustement par une loi de puissance.

puissance. La figure 6.2 présente les mesures  $b_{\text{eff}, \text{Ly}\alpha}$  et  $\beta_{\text{Ly}\alpha}$  dans les quatre bins en redshift, ainsi que l'ajustement fait sur ces quatre points. Pour le biais effectif, nous obtenons  $b_{\text{eff}, \text{Ly}\alpha}(z) = A(1+z)^\gamma$  avec  $A = 0,0029 \pm 0,0009$  et  $\gamma = 3,47 \pm 0,26$ . Pour le paramètre RSD, nous obtenons  $\beta_{\text{Ly}\alpha}(z) = A(1+z)^\gamma$  avec  $A = 27,7 \pm 47,1$  et  $\gamma = -2,32 \pm 1,40$ . Ces deux ajustements sont utilisés comme référence pour l'ajustement des paramètres des mocks (section 4.4). Ils sont extrapolés de  $z = 1,9$  jusqu'à  $z = 3,6$ .

### 6.1.2 Stabilité des paramètres Ly $\alpha$

Après avoir produit les ajustements présentés dans la section précédente, nous avons cherché à savoir si la mesure des paramètres Ly $\alpha$  était fiable. Nous avons d'abord regardé la corrélation des paramètres  $b_{\eta, \text{Ly}\alpha}$  et  $\beta_{\text{Ly}\alpha}$  avec les autres paramètres du modèle. La table 6.2 présente ces corrélations. Premièrement, nous pouvons remarquer que les deux paramètres Ly $\alpha$  ajustés sont très anticorrelés :  $-87\%$ . Deuxièmement, les paramètres du Ly $\alpha$  sont très corrélés avec ceux des HCD, notamment  $\beta_{\text{Ly}\alpha}$  qui est corrélé à  $-75\%$  avec  $b_{\text{HCD}}$ . Ceci pose plusieurs problèmes : d'abord, la modélisation des HCD choisie dans DU MAS DES BOURBOUX et al. (2020) et utilisée ici consiste à masquer les HCD identifiés par un algorithme. En pratique cela revient à masquer les HCD avec une densité de colonne  $\log n_{\text{HI}} > 20,3$ . Les HCD non masqués sont pris en compte par le terme  $F_{\text{HCD}}$  (voir section 5.4.1). Cependant, l'algorithme utilisé ne possède pas une efficacité de 100 % (Chabanier et al. (in prep)). Des HCD avec une grande densité de colonne ne sont donc pas masqués. Ces HCD produisent des absorptions intenses, non prises en compte par le terme  $F_{\text{HCD}}$ , ce qui a pour effet d'augmenter le biais du Ly $\alpha$ . Ensuite, le paramètre effectif  $L_{\text{HCD}}$  doit être fixé car il est très corrélé avec les autres paramètres et peut empêcher l'ajustement de converger. Sa valeur, qui dépend de la distribution des HCD non masqués, est difficile à déterminer. Une valeur de  $10 h^{-1} \text{ Mpc}$  a été choisie dans l'analyse DR16, de façon un peu arbitraire dans la mesure où cela n'affecte pas la mesure de l'échelle BAO.

Nous remarquons que les paramètres des HCD obtenus dépendent de la valeur de  $L_{\text{HCD}}$  choisie, comme illustré sur la figure 6.3. A cause des corrélations entre les paramètres liés aux HCD et ceux liés au Ly $\alpha$ , le fait de changer  $L_{\text{HCD}}$  change aussi les paramètres Ly $\alpha$  obtenus. La figure 6.4 montre la

TABLE 6.2 – Corrélations des paramètres du modèle avec les paramètres  $b_{\eta,\text{Ly}\alpha}$  et  $\beta_{\text{Ly}\alpha}$  lors de l’ajustement de la fonction de corrélation  $\text{Ly}\alpha \times \text{Ly}\alpha$  estimée à partir des données DR16. L’ajustement est fait sur la moyenne des fonctions de corrélation calculées dans chaque bin en redshift, soit l’ensemble des données DR16.

Paramètre	$b_{\eta,\text{Ly}\alpha}$	$\beta_{\text{Ly}\alpha}$
$\alpha_{\parallel}$	0 %	0 %
$\alpha_{\perp}$	1 %	-2 %
$b_{\eta,\text{Ly}\alpha}$	100 %	-87 %
$\beta_{\text{Ly}\alpha}$	-87 %	100 %
$b_{\eta,\text{SiII}(1190)}$	2 %	-8 %
$b_{\eta,\text{SiII}(1193)}$	3 %	-6 %
$b_{\eta,\text{SiII}(1260)}$	-1 %	-3 %
$b_{\eta,\text{SiIII}(1207)}$	6 %	5 %
$b_{\eta,\text{CIV}(\text{eff})}$	-7 %	-10 %
$b_{\text{HCD}}$	48 %	-75 %
$\beta_{\text{HCD}}$	35 %	-23 %
$A_{\text{sky}}$	34 %	-19 %
$\sigma_{\text{sky}}$	-10 %	-2 %

dépendance de  $b_{\text{eff},\text{Ly}\alpha}$  et  $\beta_{\text{Ly}\alpha}$  avec  $L_{\text{HCD}}$ . Ainsi, le paramètre RSD  $\beta_{\text{Ly}\alpha}$  est très corrélé avec  $L_{\text{HCD}}$ . Lorsque nous laissons libre  $L_{\text{HCD}}$ , en utilisant un prior gaussien centré sur  $10 h^{-1} \text{ Mpc}$  et avec une largeur  $\sigma = 1 h^{-1} \text{ Mpc}$ , nous mesurons une corrélation entre  $L_{\text{HCD}}$  et  $\beta_{\text{Ly}\alpha}$  de -38 %. Lorsque  $L_{\text{HCD}}$  est laissé totalement libre, il est corrélé, en valeur absolue, à plus de 85 % avec les paramètres  $b_{\text{eff},\text{Ly}\alpha}$ ,  $\beta_{\text{Ly}\alpha}$ ,  $b_{\text{HCD}}$  et  $b_{\eta,\text{SiIII}(1207)}$ . Dans ce cas, nous mesurons  $L_{\text{HCD}} = 2,54 \pm 0,73 h^{-1} \text{ Mpc}$ , très éloignée de la valeur à laquelle il est fixé dans l’analyse DR16. Nous discutons cette valeur dans la section 6.2.4).

Enfin, le modèle des HCD choisi influence la mesure des paramètres Ly $\alpha$ . Toujours dans le but d’avoir une mesure robuste des paramètres Ly $\alpha$ , nous avons essayé d’utiliser un autre modèle pour les HCD, développé par Edmond Chaussidon et Julien Guy (modèle C-G). Nous détaillons l’analyse en utilisant ce modèle dans la section 6.2.3.

## 6.2 Étude de la modélisation des HCD

Suite aux différents points énoncés dans la section précédente, nous avons étudié l’effet qu’ont les HCD sur les paramètres Ly $\alpha$  mesurés dans les mocks. En effet, les mocks sont l’outil parfait pour ce genre d’analyse : ils permettent, contrairement aux données, de connaître la quantité de Ly $\alpha$  présente, et de comparer cette quantité à ce qui est mesuré par l’ajustement. De plus, nous connaissons le nombre et les distributions en redshift ( $z$ ) et en densité de colonne ( $\log n_{\text{HI}}$ ) des HCD ajoutés dans les mocks, ce qui n’est pas le cas des données. Dans cette section, nous comparons les paramètres Ly $\alpha$  mesurés dans les mocks sans HCD (raw mocks, eboss-0.0) et avec HCD (eboss-0.2). Nous comparons aussi les paramètres Ly $\alpha$  mesurés en utilisant différentes modélisations des HCD.

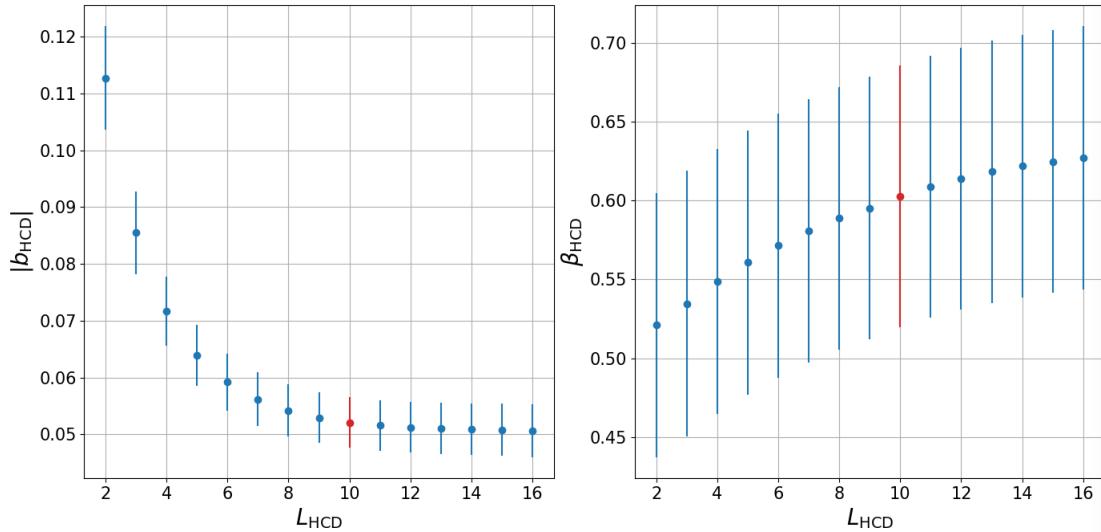


FIGURE 6.3 – Évolution des mesures des paramètres  $b_{\text{HCD}}$  et  $\beta_{\text{HCD}}$  en fonction de la valeur  $L_{\text{HCD}}$  choisie pour l’ajustement. Les mesures sont faites sur l’auto-corrélation  $\text{Ly}\alpha \times \text{Ly}\alpha$  estimée à partir des données DR16 et moyennée sur les quatre bins en redshift.

### 6.2.1 Comparaison des mocks

Comme expliqué dans le chapitre 5, nous avons analysé 30 réalisations des raw mocks, des mocks eboss-0.0 et des mocks eboss-0.2. Dans chacun des cas, nous ajustons le modèle sur  $20 < r < 180 h^{-1} \text{ Mpc}$  et mesurons les paramètres  $\text{Ly}\alpha$ . La figure 5.16 présente les mesures de ces paramètres dans chaque bin en redshift pour chacune des versions des mocks. Nous pouvons remarquer que les valeurs des paramètres  $\text{Ly}\alpha$  mesurés changent selon la version des mocks. Les paramètres mesurés dans les raw mocks sont très proches des paramètres visés, mesurés dans les données DR16. Ceci montre que la procédure d’ajustement des paramètres des mocks que nous avons mise en place est efficace pour obtenir les bons  $b_{\text{eff},\text{Ly}\alpha}$  et  $\beta_{\text{Ly}\alpha}$ .

Lorsque nous comparons maintenant les valeurs de  $b_{\text{eff},\text{Ly}\alpha}$  et  $\beta_{\text{Ly}\alpha}$  mesurées dans les raw mocks à celles mesurées dans les mocks eboss-0.0, nous observons un écart statistiquement significatif. L’effet sur  $\beta_{\text{Ly}\alpha}$  est faible, et les valeurs de  $\beta_{\text{Ly}\alpha}$  mesurées dans les mocks eboss-0.0 restent compatibles avec les données DR16. Cependant, l’effet sur le biais effectif  $b_{\text{eff},\text{Ly}\alpha}$  est important (de l’ordre de 10 %) et statistiquement significatif. Les écarts observés sont attribués à la matrice de distorsion, qui ne capture pas l’intégralité des distorsions produites par l’ajustement du continuum. La figure 6.5 présente la comparaison des mocks eboss-0.0, en rouge, avec les mocks eboss-raw multipliés par la matrice de distorsion, en vert. Si la matrice de distorsion était parfaite, les courbes rouge et verte seraient confondues, ce qui n’est pas le cas. Une étude approfondie de la matrice de distorsion et de l’ajout du continuum par `quickquasars` est nécessaire pour comprendre ces différences.

Enfin, nous observons un écart entre les valeurs de  $b_{\text{eff},\text{Ly}\alpha}$  et  $\beta_{\text{Ly}\alpha}$  mesurées dans les mocks eboss-0.0 et eboss-0.2. L’écart mesuré pour  $b_{\text{eff},\text{Ly}\alpha}$  est comparable à celui mesuré entre les mocks eboss-0.0 et les raw mocks. L’écart mesuré sur  $\beta_{\text{Ly}\alpha}$  est plus important. Les valeurs obtenues dans l’ajustement des mocks eboss-0.2 ne sont pas compatibles avec celles mesurées dans les données DR16. Comme expliqué dans la section 6.1.2, les paramètres  $\text{Ly}\alpha$  sont très corrélés avec ceux des HCD, en particulier avec  $L_{\text{HCD}}$ .

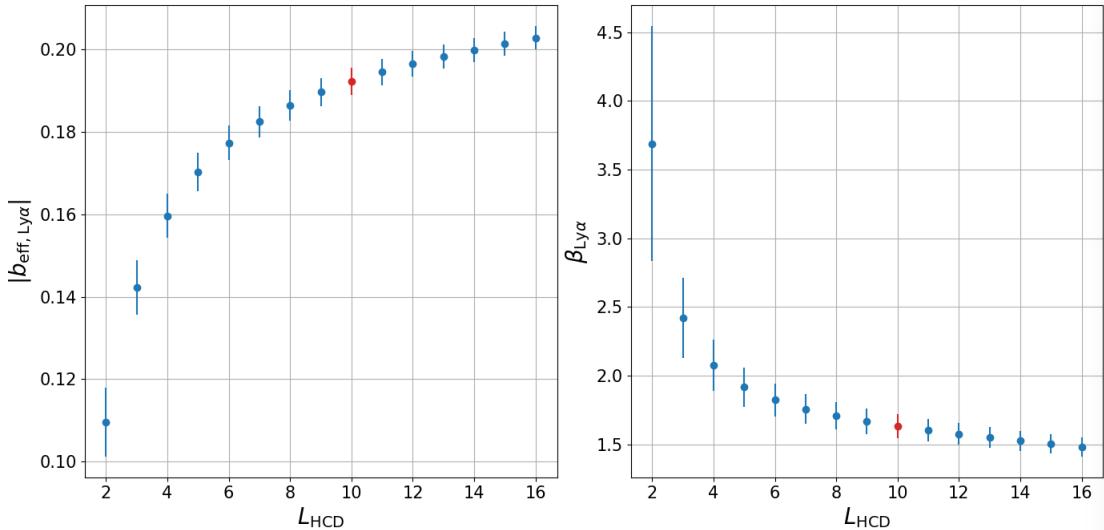


FIGURE 6.4 – Évolution des mesures des paramètres  $b_{\text{eff}, \text{Ly}\alpha}$  et  $\beta_{\text{Ly}\alpha}$  en fonction de la valeur  $L_{\text{HCD}}$  choisie pour l’ajustement. Les mesures sont faites sur l’auto-corrélation  $\text{Ly}\alpha \times \text{Ly}\alpha$  estimée à partir des données DR16 et moyennée sur les quatre bins en redshift.

TABLE 6.3 – Résultats des ajustements de l’auto-corrélation  $\text{Ly}\alpha \times \text{Ly}\alpha$  estimée sur 30 réalisations des mocks et moyennée sur les quatre bins en redshift. La première ligne donne l’ajustement des mocks eboss-0.0. La deuxième donne l’ajustement des mocks eboss-0.2 comme décrit dans la section 5.4.2. La troisième ligne donne le même ajustement que la deuxième mais en utilisant  $L_{\text{HCD}} = 2,8 h^{-1} \text{ Mpc}$ .

version	$b_{\text{eff}, \text{Ly}\alpha}$	$\beta_{\text{Ly}\alpha}$	$b_{\text{HCD}}$	$\beta_{\text{HCD}}$	$L_{\text{HCD}} [h^{-1} \text{ Mpc}]$	$\chi^2 (n_{\text{dof}})$
eboss-0.0	$-0.1956 \pm 0.0002$	$1.687 \pm 0.008$				1562 (1570)
eboss-0.2	$-0.2044 \pm 0.0007$	$1.548 \pm 0.014$	$-0.0080 \pm 0.0010$	$0.487 \pm 0.089$	10	1573 (1568)
eboss-0.2	$-0.1925 \pm 0.0023$	$1.647 \pm 0.034$	$-0.0177 \pm 0.0025$	$0.455 \pm 0.090$	2.8	1578 (1568)

qui est fixé à  $10 h^{-1} \text{ Mpc}$ . Nous regardons dans un premier temps la corrélation des paramètres  $\text{Ly}\alpha$  avec  $L_{\text{HCD}}$  dans les mocks. La figure 6.6 présente la mesure de  $b_{\text{eff}, \text{Ly}\alpha}$  et  $\beta_{\text{Ly}\alpha}$  dans les mocks-0.2 pour différentes valeurs de  $L_{\text{HCD}}$ . Le corrélation avec  $L_{\text{HCD}}$  est similaire à celle observée dans les données (figure 6.4). Cette figure suggère qu’il faut utiliser une valeur de  $L_{\text{HCD}}$  entre 2 et  $4 h^{-1} \text{ Mpc}$  afin d’obtenir une mesure des paramètres  $\text{Ly}\alpha$  en accord avec les mocks eboss-0.0. Ainsi nous produisons un ajustement des mocks eboss-0.2 avec  $L_{\text{HCD}} = 2,8 h^{-1} \text{ Mpc}$  (le choix de cette valeur particulière est expliquée dans la section 6.2.4). Le résultat de cet ajustement est présenté dans le tableau 6.3. Les paramètres  $\text{Ly}\alpha$  obtenus sont maintenant compatibles à  $1,2\sigma$  et  $1,3\sigma$  avec les mocks eboss-0.0. Ceci suggère que la valeur de  $L_{\text{HCD}}$  utilisée dans l’ajustement des données DR16 est mal choisie. Nous investiguons le choix de  $L_{\text{HCD}}$  dans la section 6.2.4.

### 6.2.2 Effet du masquage des HCD

Dans l’analyse des mocks eboss-0.2, afin d’imiter l’analyse des données DR16, les HCD pour lesquels  $\log n_{\text{HI}} > 20,3$  sont masqués lors du calcul des  $\delta_F$ . Comme expliqué dans la section 6.1.2, le masquage des HCD dans les données s’effectue selon le résultat de l’algorithme d’identification. Les HCD identifiés puis reconstruits avec  $\log n_{\text{HI}} > 20,3$  sont masqués. Dans le cas des mocks, les HCD avec  $\log n_{\text{HI}} > 20,3$

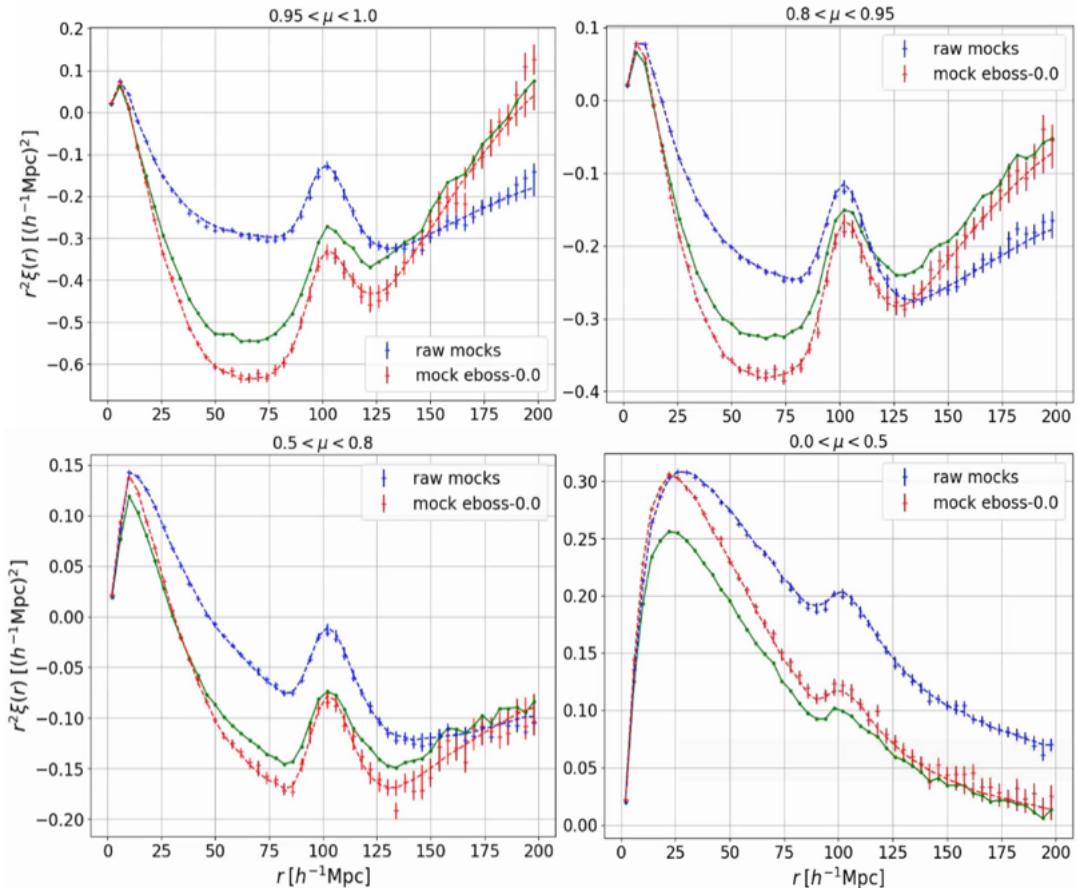


FIGURE 6.5 – Fonction de corrélation Ly $\alpha$ ×Ly $\alpha$  estimée à partir des mocks eboss-raw, en bleu, et des mocks eboss-0.0, en rouge. Les fonctions de corrélation sont estimées à partir de 30 réalisations des mocks. La courbe verte correspond à la corrélation Ly $\alpha$ ×Ly $\alpha$  estimée à partir des mocks eboss-raw et multipliée par la matrice de distorsion.

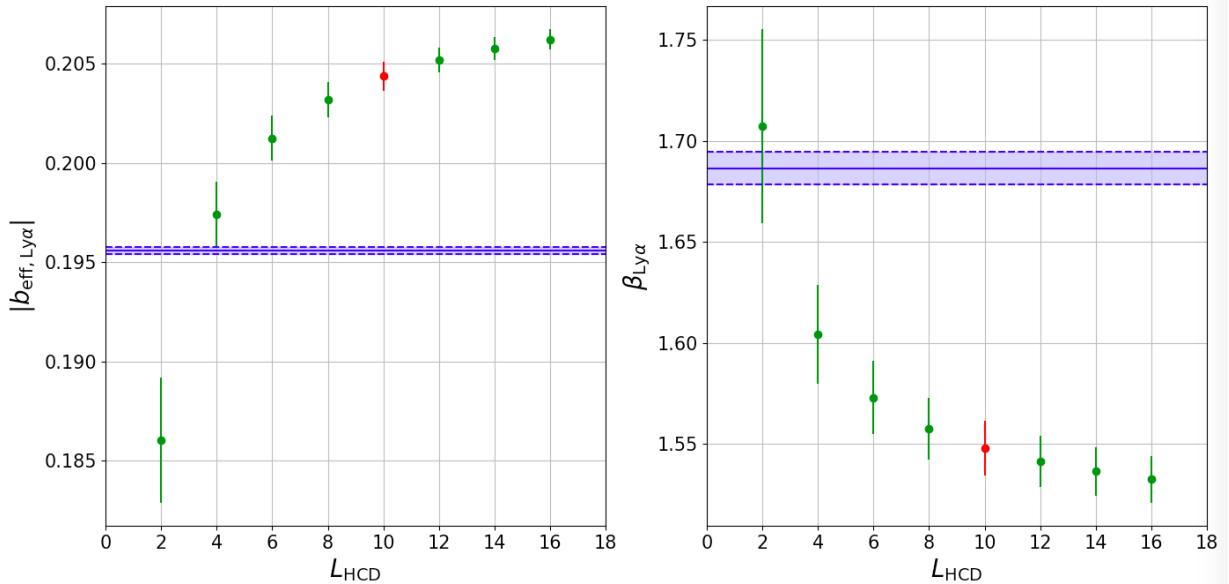


FIGURE 6.6 – Évolution des mesures des paramètres  $b_{\text{eff}, \text{Ly}\alpha}$  et  $\beta_{\text{Ly}\alpha}$  en fonction de la valeur  $L_{\text{HCD}}$  choisie pour l'ajustement. Les mesures sont faites sur l'auto-corrélation  $\text{Ly}\alpha \times \text{Ly}\alpha$  estimée à partir de 30 réalisations des mocks eboss-0.2 et moyennée sur les 4 bins en redshift. La ligne horizontale bleue donne la mesure faite sur les 30 réalisations des mocks eboss-0.0. Les lignes tiretées bleues donnent les erreurs à  $1\sigma$ .

sont masqués à partir du « vrai » catalogue. Nous étudions ici l’effet du masquage à partir du catalogue produit par l’algorithme d’identification. Pour ce faire, nous produisons l’analyse d’une réalisation de mock eboss-0.2, pour laquelle nous utilisons l’algorithme d’identification pour créer un catalogue de HCD. Le champ  $\delta_F$  est calculé en masquant les HCD identifiés par l’algorithme, puis la fonction de corrélation  $\text{Ly}\alpha \times \text{Ly}\alpha$  est estimée dans les quatre bins en redshift utilisés jusqu’ici. Nous nommons cette analyse *eboss-0.2\_finder*. La figure 6.7 présente la fonction de corrélation  $\text{Ly}\alpha \times \text{Ly}\alpha$  estimée à partir de la même réalisation des mocks, en version eboss-0.0, eboss-0.2 et eboss-0.2\_finder. Les fonctions de corrélation affichées sont la moyenne des fonctions de corrélation estimées dans les quatre bins en redshift. Dans les trois versions, le code `quickquasars` utilise les mêmes quasars pour produire les spectres synthétiques, et ajoute le même bruit à ces spectres dans les trois cas. Ceci nous permet d’avoir les mêmes fluctuations statistiques dans le calcul de la corrélation  $\text{Ly}\alpha \times \text{Ly}\alpha$ , et ainsi d’avoir des mesures de biais comparables. L’effet des HCD (avec une densité de colonne  $17,2 < \log n_{\text{HI}} < 20,3$ ) est visible en comparant la corrélation montrée en rouge à celle montrée en bleu. Comme expliqué dans la section 5.4.1, l’effet principal des HCD est d’augmenter le biais effectif. Par ailleurs, le fait que l’effet des HCD soit légèrement plus important sur la corrélation de la version eboss-0.2 que sur celle de la version eboss-0.2\_finder suggère que l’algorithme identifie correctement les HCD pour lesquels  $\log n_{\text{HI}} > 20,3$ , et identifie une petite partie des HCD avec  $\log n_{\text{HI}} < 20,3$  et les reconstruit avec  $\log n_{\text{HI}} > 20,3$ . Ceci a pour effet de masquer des HCD qui ne possèdent pas une densité de colonne supérieure à 20,3. Ceci est confirmé par la figure 6.8, qui compare la densité de colonne  $\log n_{\text{HI}}$  trouvée par le finder (output) à la « vrai » densité de colonne des HCD des mocks (input). Nous voyons sur cette figure que l’algorithme a tendance à surestimer la densité de colonne  $\log n_{\text{HI}}$ , ce qui est en accord

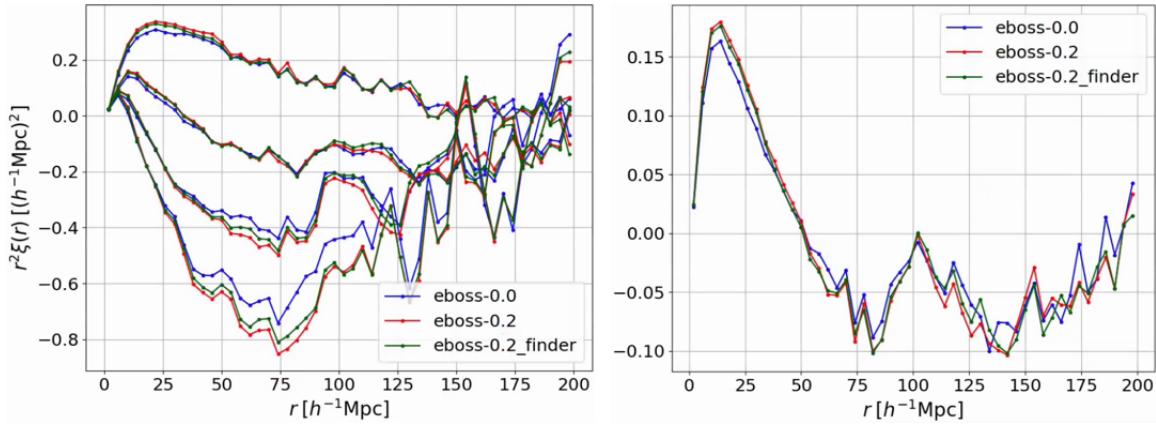


FIGURE 6.7 – Fonctions de corrélation Ly $\alpha$  × Ly $\alpha$  estimées à partir d’une réalisation des mocks eboss-0.0 (bleu), eboss-0.2 (rouge) et eboss-0.2\_finder (vert). La version eboss-0.2\_finder correspond aux mocks eboss-0.2, dans lesquels les HCD ont été masqués en utilisant le catalogue de HCD produit par l’algorithme d’identification. Le graphique de gauche montre les corrélations dans quatre gammes en  $\mu$ . Ces gammes sont, de haut en bas :  $0 < \mu < 0,5$ ,  $0,5 < \mu < 0,8$ ,  $0,8 < \mu < 0,95$  et  $0,95 < \mu < 1$ . Le graphique de droite montre les corrélations moyennées sur  $0 < \mu < 1$ .

TABLE 6.4 – Résultat de l’ajustement de la corrélation Ly $\alpha$  × Ly $\alpha$  estimée à partir d’une réalisation des mocks eboss-0.0, eboss-0.2 et eboss-0.2\_finder. La version eboss-0.2\_finder correspond aux mocks eboss-0.2, dans lesquels les HCD ont été masqués en utilisant le catalogue de HCD produit par l’algorithme d’identification.

version	$b_{\text{eff}, \text{Ly}\alpha}$	$\beta_{\text{Ly}\alpha}$	$b_{\text{HCD}}$	$\beta_{\text{HCD}}$
eboss-0.0	$-0.1970 \pm 0.0009$	$1.641 \pm 0.039$		
eboss-0.2	$-0.1979 \pm 0.0038$	$1.578 \pm 0.069$	$-0.0201 \pm 0.0053$	$0.499 \pm 0.090$
eboss-0.2_finder	$-0.1951 \pm 0.0039$	$1.592 \pm 0.074$	$-0.0186 \pm 0.0054$	$0.494 \pm 0.091$

avec nos observations. Notons par ailleurs que l’algorithme d’identification possède une efficacité de 90 % pour  $\log n_{\text{HI}} > 20,3$ . L’effet de ces HCD non identifiés avec  $\log n_{\text{HI}} > 20,3$  est donc plus faible que l’effet des HCD dont le  $\log n_{\text{HI}}$  est surestimé. Ceci s’explique par le faible nombre de HCD avec  $\log n_{\text{HI}} > 20,3$ .

Le tableau 6.4 donne les résultats des ajustements des trois corrélations présentées sur la figure 6.7. La statistique d’une seule réalisation n’est pas suffisante pour identifier des potentielles systématiques. Cependant, la précision de la mesure des paramètres Ly $\alpha$  dans les données DR16 étant comparable à celle des mocks eboss-0.2, les potentielles systématiques sont inférieures à l’erreur statistique sur cette mesure. Il serait tout de même intéressant de mener cette étude sur un plus grand nombre de réalisations.

### 6.2.3 Une meilleure modélisation des HCD

Toujours dans l’optique de tester la robustesse de la mesure des paramètres Ly $\alpha$ , nous avons utilisé une modélisation des HCD différente de celle décrite dans la section 5.4.1 et utilisée jusqu’ici pour modéliser les mocks et les données. Ce modèle est développé par Edmond Chaussidon et Julien Guy,

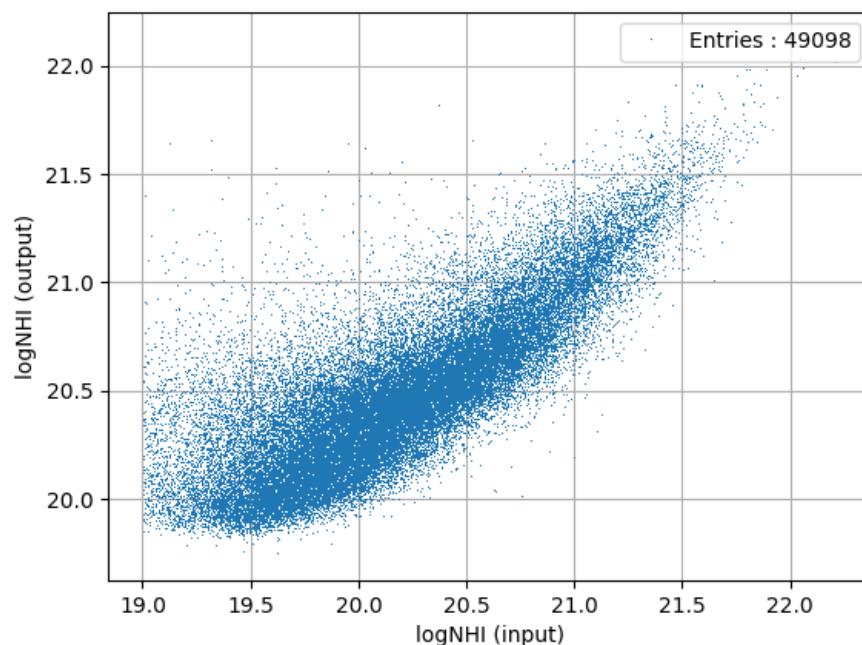


FIGURE 6.8 – Densité de colonne  $\log n_{\text{HI}}$  trouvée par le finder ( $\log \text{NHI}$  output) en fonction de la « vrai » densité de colonne des HCD des mocks ( $\log \text{NHI}$  input). Cette comparaison est produite par Jim Rich et Solène Chabanier, du groupe cosmologie du CEA, à l'aide des mocks (Chabanier et al. (in prep)).

au sein du groupe Ly $\alpha$  de la collaboration DESI. Nous faisons référence à ce modèle via le nom *modèle C-G*. Le modèle des HCD décrit dans la section 5.4.1 est dénommé *modèle de Rogers*. Le modèle C-G, contrairement au modèle de Rogers, n'a pas besoin de masquer les HCD identifiés par l'algorithme. Il prend en compte les effets sur les corrélations Ly $\alpha$  $\times$ Ly $\alpha$  et Ly $\alpha$  $\times$ QSO produits par l'ensemble des HCD dans les données. Les deux modèles utilisent la modélisation définie dans l'équation 5.26. Cependant, dans le cas du modèle C-G, plutôt que de définir  $F_{\text{HCD}}$  comme une fonction exponentielle avec un paramètre effectif  $L_{\text{HCD}}$  qui reflète la taille caractéristique des HCD non masqués (équation 5.28), la fonction  $F_{\text{HCD}}$  est calculée en prenant en compte la distribution en  $\log n_{\text{HI}}$  des HCD présents dans les données. L'avantage de cette méthode est qu'elle n'utilise pas le paramètre effectif  $L_{\text{HCD}}$ . Elle permet donc de modéliser des distributions de HCD avec une plus grande gamme en  $\log n_{\text{HI}}$ , là où le modèle de Rogers, utilisé pour analyser les données DR16, ne fonctionne plus très bien. De plus, le modèle C-G permet de s'affranchir des potentiels systématiques induites par l'utilisation de l'algorithme de détection. Cependant, afin de calculer le terme  $F_{\text{HCD}}$  correspondant à la distribution de HCD présents, nous devons justement connaître cette distribution. Si cela est possible pour les mocks, cela ne l'est pas pour les données, car nous ignorons la distribution en  $\log n_{\text{HI}}$  des HCD non identifiés, c'est à dire pour lesquels  $\log n_{\text{HI}} < 20,3$ . Ainsi, lorsque nous analysons les données DR16 avec le modèle C-G, nous supposons que la distribution de HCD dans les données est celle du modèle `pyigm`, utilisée dans les mocks.

Du fait que, pour le modèle C-G, le terme  $F_{\text{HCD}}$  soit calculé à partir de la distribution en  $\log n_{\text{HI}}$  des HCD présents, nous pouvons envisager d'ajuster avec ce modèle, des mocks avec différentes distributions de HCD : les mocks eboss-0.2 où les HCD avec  $\log n_{\text{HI}} < 20,3$  sont masqués, les mocks eboss-0.2 où tous les HCD sont présents ( $17,2 < \log n_{\text{HI}} < 22,5$ ), ou n'importe quelle autre distribution.

Afin de tester le modèle C-G, et de le comparer au modèle de Rogers, nous vérifions premièrement que ces deux modèles, lorsqu'ils sont ajustés sur les mocks eboss-0.0, mesurent une quantité de HCD compatible avec 0.

Par ailleurs, dans une étude préliminaire au choix des paramètres Ly $\alpha$  à utiliser pour construire les mocks, nous étudions comment ces deux modèles se comportent lorsque nous triplons la quantité de HCD dans les mocks. Pour ce faire, à partir d'une même réalisation des mocks, nous produisons deux versions de quickquasars eboss-0.2 : l'une avec le catalogue de HCD standard, l'autre avec un catalogue contenant 3 fois plus de HCD. Dans ces deux versions, les HCD qui vérifient  $\log n_{\text{HI}} > 20,3$  sont masqués. Afin de faciliter la comparaison des résultats, nous utilisons le même sous-échantillon de quasars et ajoutons le même bruit pour produire les spectres synthétiques puis calculer la corrélation Ly $\alpha$  $\times$ Ly $\alpha$  à partir de ces deux versions. Puis, nous ajustons la fonction de corrélation Ly $\alpha$  $\times$ Ly $\alpha$ , estimées à partir de ces deux versions, avec le modèle de Rogers et le modèle C-G. Le paramètre  $b_{\text{HCD}}$  étant proportionnel au nombre de HCD, nous nous attendons à mesurer un biais trois fois plus grand dans la réalisation possédant trois fois plus de HCD. Cependant, ce n'est pas ce que nous observons. Le tableau 6.5 résume ces mesures. Pour une même version, les mesures de  $b_{\text{HCD}}$  faites avec le modèle de Rogers ou le modèle C-G ne sont pas compatibles. Comme nous le verrons dans la section 6.2.4, cela est dû à la valeur de  $L_{\text{HCD}}$  mal choisie pour le modèle de Rogers. De plus, pour un même modèle, nous remarquons que  $b_{\text{HCD}}$  mesuré dans la version contenant 3 fois plus de HCD n'est pas compatible

TABLE 6.5 – Mesures des paramètres  $b_{\text{eff},\text{Ly}\alpha}$ ,  $\beta_{\text{Ly}\alpha}$ ,  $b_{\text{HCD}}$  et  $\beta_{\text{HCD}}$  faites à partir de l’auto-corrélation  $\text{Ly}\alpha \times \text{Ly}\alpha$  estimée sur une réalisation des mocks eboss-0.2, où les HCD avec  $\log n_{\text{HI}} > 20,3$  ont été masqués. Le détail de chaque ajustement est donné dans le texte.

Version	$b_{\text{eff},\text{Ly}\alpha}$	$\beta_{\text{Ly}\alpha}$	$b_{\text{HCD}}$	$\beta_{\text{HCD}}$
Rogers	$-0.1910 \pm 0.0020$	$1.561 \pm 0.039$	$-0.0167 \pm 0.0030$	$0.509 \pm 0.089$
Rogers & $3 \times \text{HCD}$	$-0.2095 \pm 0.0025$	$1.295 \pm 0.039$	$-0.0296 \pm 0.0037$	$0.514 \pm 0.087$
Rogers & $3 \times \text{HCD}$ & $\text{Ly}\alpha$ fixé	$-0.1910$	$1.561$	$-0.0555 \pm 0.0017$	$0.465 \pm 0.061$
C-G	$-0.1704 \pm 0.0058$	$1.752 \pm 0.096$	$-0.0327 \pm 0.0062$	$0.487 \pm 0.089$
C-G & $3 \times \text{HCD}$	$-0.1713 \pm 0.0071$	$1.561 \pm 0.110$	$-0.0598 \pm 0.0077$	$0.477 \pm 0.087$
C-G & $3 \times \text{HCD}$ & $\text{Ly}\alpha$ fixé	$-0.1704$	$1.752$	$-0.0651 \pm 0.0014$	$0.277 \pm 0.040$

avec 3 fois le biais des HCD mesuré dans l’autre version. Le biais des HCD mesuré dans la version contenant 3 fois plus de HCD est sous-estimé. Ceci a pour effet de réduire la mesure de  $\beta_{\text{Ly}\alpha}$ . Nous pouvons toute fois noter que le modèle C-G produit des mesures de  $b_{\text{eff},\text{Ly}\alpha}$  compatibles dans ces deux versions, ce qui n’est pas le cas du modèle de Rogers.

Enfin, nous regardons si en fixant les paramètres  $\text{Ly}\alpha$ , nous obtenons un biais des HCD compatible avec 3 fois le biais des HCD mesuré dans la version standard. Nous faisons ce test pour le modèle de Rogers et pour le modèle C-G. Le résultat de ces ajustements est donné dans le tableau 6.5 (voir les lignes «  $3 \times \text{HCD} \& \text{Ly}\alpha$  fixé »). Pour le modèle de Rogers, le fait de fixer les paramètres  $\text{Ly}\alpha$  résout bien les tensions observées : nous attendons  $b_{\text{HCD}} = 0,0501 \pm 0,0090$  et mesurons  $b_{\text{HCD}} = -0,0555 \pm 0,0017$ . Cependant, dans le cas du modèle C-G, le paramètre  $\beta_{\text{HCD}}$  compense les paramètres  $\text{Ly}\alpha$  fixés, et la valeur de  $b_{\text{HCD}}$  obtenue n’est pas compatible avec 3 fois celle mesurée dans la version standard. Il serait intéressant de vérifier si en fixant aussi la valeur de  $\beta_{\text{HCD}}$ , nous retrouvons le bon  $b_{\text{HCD}}$ .

Afin d’étudier davantage le modèle C-G, nous comparons maintenant la mesure des paramètres  $\text{Ly}\alpha$  faite avec ce modèle sur les mocks eboss-0.2 où les HCD avec une densité de colonne  $\log n_{\text{HI}} > 20,3$  sont masqués, à la mesure faite sur les mocks eboss-0.2 où aucun HCD n’est masqué. Dans chacun des cas, nous utilisons le terme  $F_{\text{HCD}}$  correspondant à la distribution de HCD considérée. La comparaison est présentée sur la figure 6.9. Nous remarquons que la mesure des paramètres  $\text{Ly}\alpha$  faite avec le modèle C-G sur les mocks eboss-0.2 où les HCD sont masqués (orange) est compatible avec la mesure faite sur les mocks eboss-0.2 où les HCD ne sont pas masqués (rouge). Ceci est encourageant car cela signifie que le modèle C-G est cohérent avec lui-même. Lorsque nous comparons les mesures faites avec le modèle C-G aux mesures faites avec le modèle de Rogers, nous pouvons identifier une différence statistiquement significative, en particulier pour  $b_{\text{eff},\text{Ly}\alpha}$ . Nous étudions la cause de ces différences dans la section suivante. Enfin, les mesures de  $b_{\text{eff},\text{Ly}\alpha}$  faites avec le modèle C-G sur les mocks eboss-0.2 sont en meilleur accord avec la mesure faite sur les mocks eboss-0.0 que ne l’est la mesure faite avec le modèle de Rogers sur les mocks eboss-0.2.

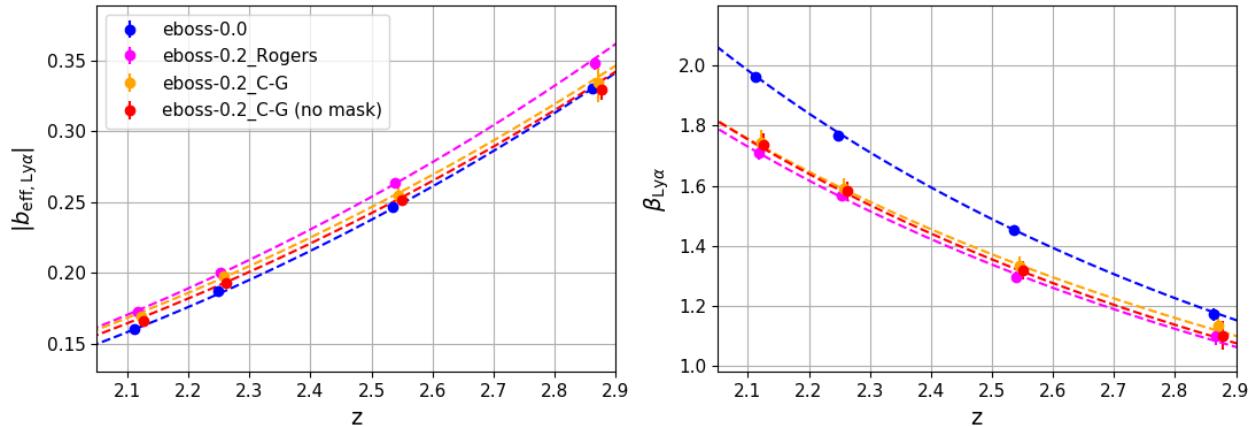


FIGURE 6.9 – Mesure des paramètres  $b_{\text{eff}, \text{Ly}\alpha}$  (gauche) et  $\beta_{\text{Ly}\alpha}$  (droite) sur l’auto-corrélation  $\text{Ly}\alpha \times \text{Ly}\alpha$  estimée à partir des mocks (30 réalisations utilisées pour chaque mesure). Les points bleus présentent la mesure faite sur les mocks eboss-0.0. Les points magenta la mesure faite sur les mocks eboss-0.2, où les HCD avec  $\log n_{\text{HI}} > 20,3$  sont masqués dans le calcul des fonctions de corrélation, et ajustés avec le modèle de Rogers. Les points orange la mesure faite sur ces mêmes fonctions de corrélation, mais ajustées avec le modèle C-G. Enfin, les points rouges présentent la mesure faite sur les mocks eboss-0.2 où aucun HCD n’a été masqué, ajustés avec le modèle C-G. Les courbes tiretées correspondent à l’ajustement de chaque jeu de données par une loi de puissance. Les points bleus sont décalés horizontalement de  $\Delta z = -1 \times 10^{-2}$ , les magenta de  $\Delta z = -5 \times 10^{-3}$  et les rouges de  $\Delta z = 5 \times 10^{-3}$  pour des raisons de visibilité.

### 6.2.4 Mieux comprendre les HCD

#### L’effet des HCD sur la fonction de corrélation

Afin de comprendre pourquoi les deux modèles discutés dans la section précédente ont du mal à distinguer la contribution des HCD de celle du  $\text{Ly}\alpha$ , nous avons essayé de comprendre l’effet des HCD sur les fonctions de corrélation. La figure 6.10 présente les différentes composantes du modèle ajusté sur l’auto-corrélation  $\text{Ly}\alpha \times \text{Ly}\alpha$  estimée à partir des données DR16. Ce modèle est présenté dans la section 5.4.1, il utilise la modélisation de Rogers avec  $L_{\text{HCD}} = 10 h^{-1} \text{ Mpc}$ . La figure 6.11 donnent ces mêmes composantes non multipliées par la matrice de distorsion. En comparant les courbes orange aux courbes vertes, nous pouvons remarquer sur ces figures que l’effet des HCD est assez semblable à celui d’une augmentation du biais. Nous pouvons cependant distinguer l’effet des HCD de celui d’une augmentation du biais en regardant le long de la ligne de visée : les HCD ajoutent une corrélation positive pour les petites échelles, et les courbes vertes et jaune ne se coupent pas au niveau de l’axe  $y = 0$ . Ceci est davantage visible lorsque les modèles ne sont pas multipliés par la matrice de distorsion (figure 6.11). Si nous regardons maintenant la gamme  $0 < \mu < 0,5$ , l’effet des HCD est très semblable à l’effet d’un facteur multiplicatif appliqué à la courbe verte. La distinction entre une augmentation du biais et l’effet des HCD est d’autant plus facile que les modèles utilisent une valeur de  $L_{\text{HCD}}$  importante ( $10 h^{-1} \text{ Mpc}$ ). Lorsque  $L_{\text{HCD}}$  est plus faible, le terme  $F_{\text{HCD}}$  est plus proche de 1, et donc l’effet des HCD ressemble davantage à celui d’une augmentation de biais (équation 5.26). Dans un tel cas, les paramètres  $\text{Ly}\alpha$  et ceux des HCD sont très dégénérés.

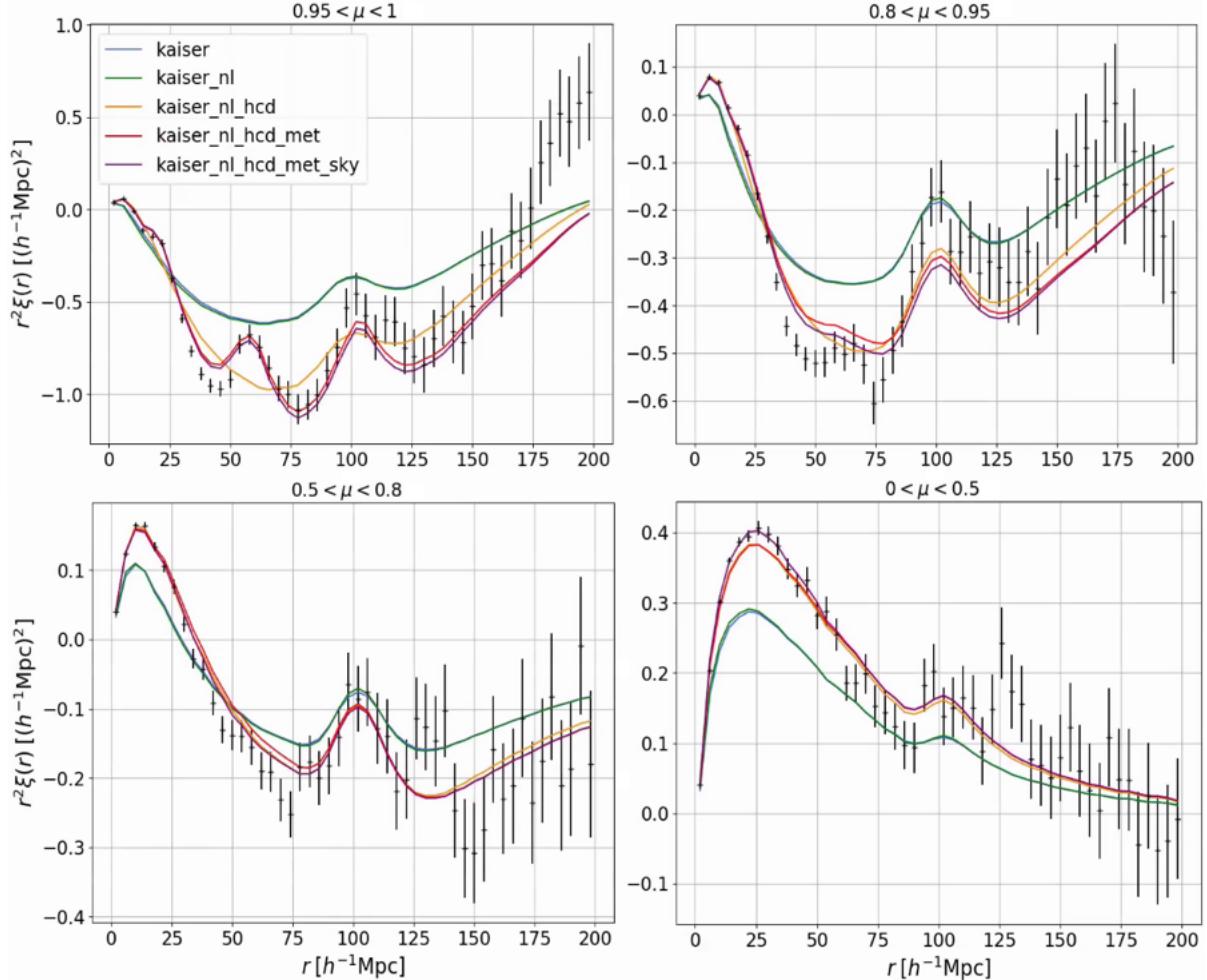


FIGURE 6.10 – Les différentes composantes du modèle ajusté sur l’auto-corrélation Ly $\alpha$ ×Ly $\alpha$  des données DR16. La ligne bleue donne le modèle de Kaiser, c’est à dire la fonction de corrélation obtenue à partir du spectre de puissance  $P_{\text{QL}}$  (équation 5.22) multiplié par le facteur de Kaiser  $b_{\text{eff},\text{Ly}\alpha}^2(1 + \beta_{\text{Ly}\alpha}\mu_k^2)$ . La ligne verte donne le modèle de Kaiser multiplié par le terme  $F_{\text{NL}}^{\text{auto}}$  qui prend en compte les non-linéarités aux petites échelles. La ligne orange donne le modèle précédent auquel la contribution des HCD a été ajoutée (équation 5.26). La ligne rouge donne le modèle précédent plus la contribution des métaux et la ligne violette donne le modèle complet. L’auto-corrélation Ly $\alpha$ ×Ly $\alpha$  estimée à partir des données DR16 et moyennée dans les quatre bins en redshift est représentée par les points noirs. Les quatre graphiques présentent 4 gammes en  $\mu$  différentes.

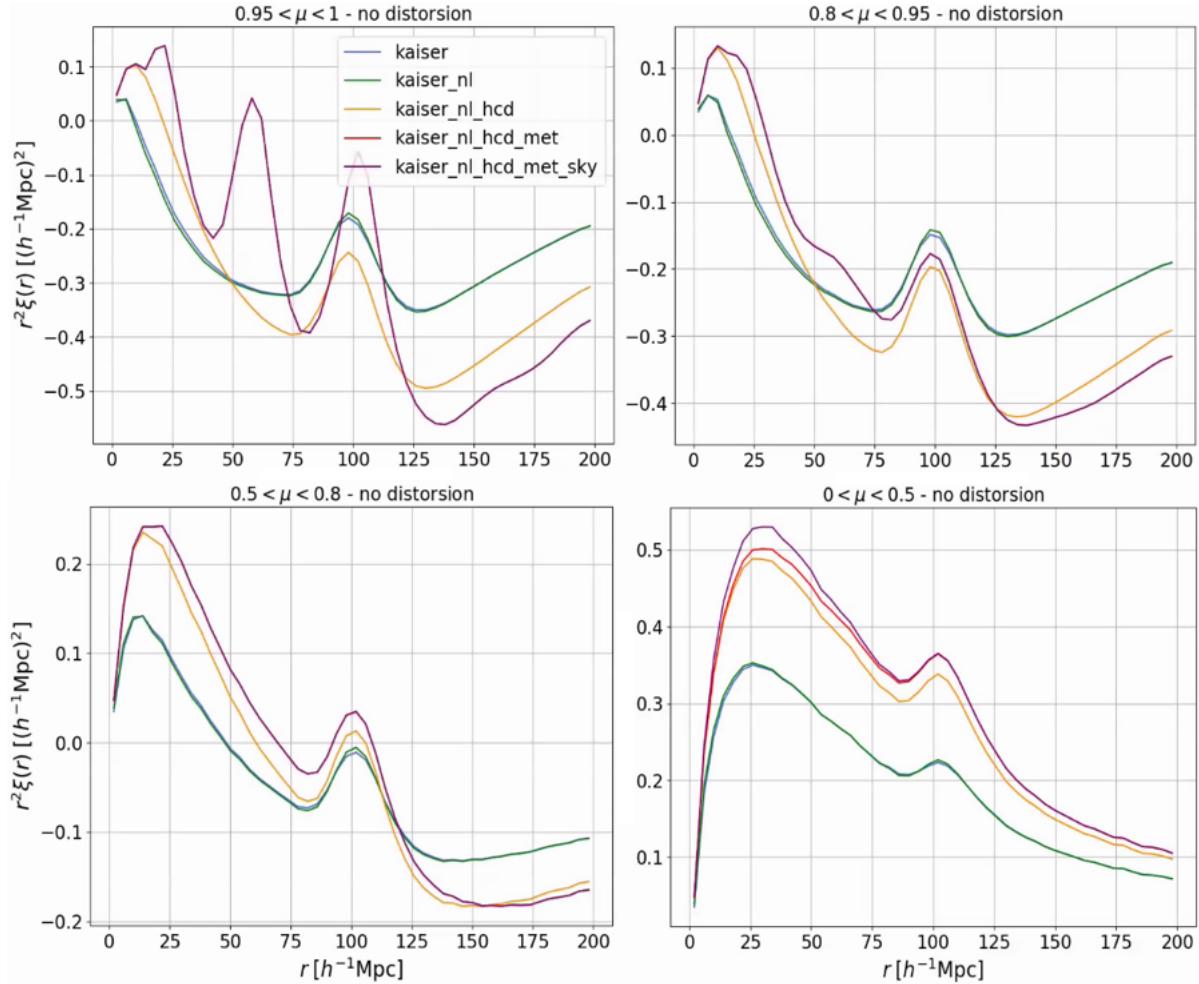


FIGURE 6.11 – Les différentes composantes du modèle ajusté sur l’auto-corrélation Ly $\alpha$  × Ly $\alpha$  des données DR16. Elles sont décrites dans la légende de la figure 6.10. Les composantes présentées ici n’incluent pas les distorsions produites par l’ajustement du continuum. Les quatre graphiques présentent ces composantes dans 4 gammes en  $\mu$ .

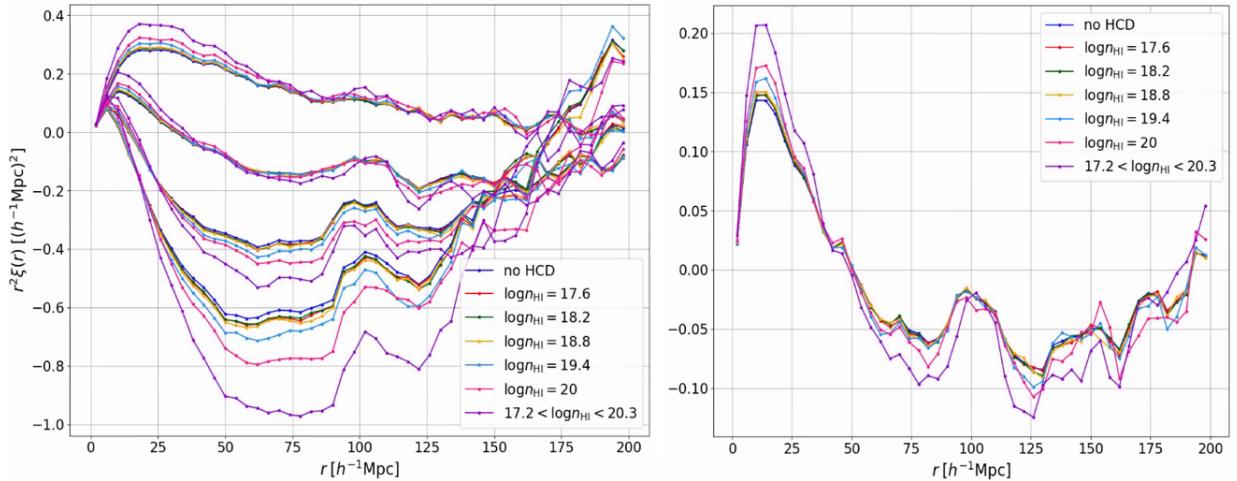


FIGURE 6.12 – Fonctions de corrélation Ly $\alpha$ ×Ly $\alpha$  estimées à partir d’une réalisation des mocks. La ligne bleue foncée donne la corrélation pour les mocks eboss-0.0 (sans HCD). La ligne violette donne la corrélation pour les mocks eboss-0.2 incluant les HCD avec  $17,2 < \log n_{\text{HI}} < 20,3$ . Les autres couleurs donnent les corrélations pour les mocks eboss-0.2 incluant des HCD avec une valeur fixe en  $\log n_{\text{HI}}$ . Le graphique de gauche montre les corrélations dans quatre gammes en  $\mu$ . Ces gammes sont, de haut en bas :  $0 < \mu < 0,5$ ,  $0,5 < \mu < 0,8$ ,  $0,8 < \mu < 0,95$  et  $0,95 < \mu < 1$ . Le graphique de droite montre les corrélations moyennées sur  $0 < \mu < 1$ .

### Des HCD avec une même densité de colonne

Nous avons ensuite regardé l’effet des différentes gammes de densité de colonne des HCD sur la fonction de corrélation Ly $\alpha$ ×Ly $\alpha$ . Pour ce faire, à partir de la même réalisation des mocks, nous avons produit cinq versions eboss-0.2. Dans chacune de ces versions, nous ajoutons les HCD avec une densité de colonne fixe. Ces valeurs pour les cinq versions sont  $\log n_{\text{HI}} \in [17,6; 18,2; 18,8; 19,4; 20]$ . Les nombres relatifs de HCD entre ces versions suivent la distribution en  $\log n_{\text{HI}}$  utilisée pour construire le catalogue standard de HCD (présentée sur la figure 4.14). La distribution en  $z$  utilisée est trois fois plus importante que dans le catalogue standard afin d’avoir suffisamment de HCD dans chaque gamme.

La figure 6.12 présente les fonctions de corrélation Ly $\alpha$ ×Ly $\alpha$  pour les différentes versions des mocks décrites précédemment. Pour simplifier les comparaisons, nous n’utilisons pas de bins en redshift, les fonctions de corrélation sont donc estimées à partir de l’ensemble des forêts. Nous pouvons voir sur la figure 6.12 que les HCD qui ont le plus grand effet sur la fonction de corrélation sont les HCD avec une grande densité de colonne, malgré leur nombre plus restreint. L’effet causé par les HCD avec une densité de colonne de 17,6, 18,2 et 18,8 est similaire : la faible densité de colonne est compensée par le nombre plus important de HCD.

Afin de comprendre les différences entre le modèle de Rogers et le modèle C-G observées sur la figure 6.9, nous étudions maintenant comment ces deux modèles se comportent lorsque nous ajustons les mocks contenant des HCD avec une densité de colonne fixe. Dans chacun des ajustements, nous devons utiliser pour chaque modèle un terme  $F_{\text{HCD}}$  adéquat aux HCD présents. En ce qui concerne le modèle de C-G, le code `picca` permet de calculer le terme  $F_{\text{HCD}}$  pour une distribution en  $\log n_{\text{HI}}$

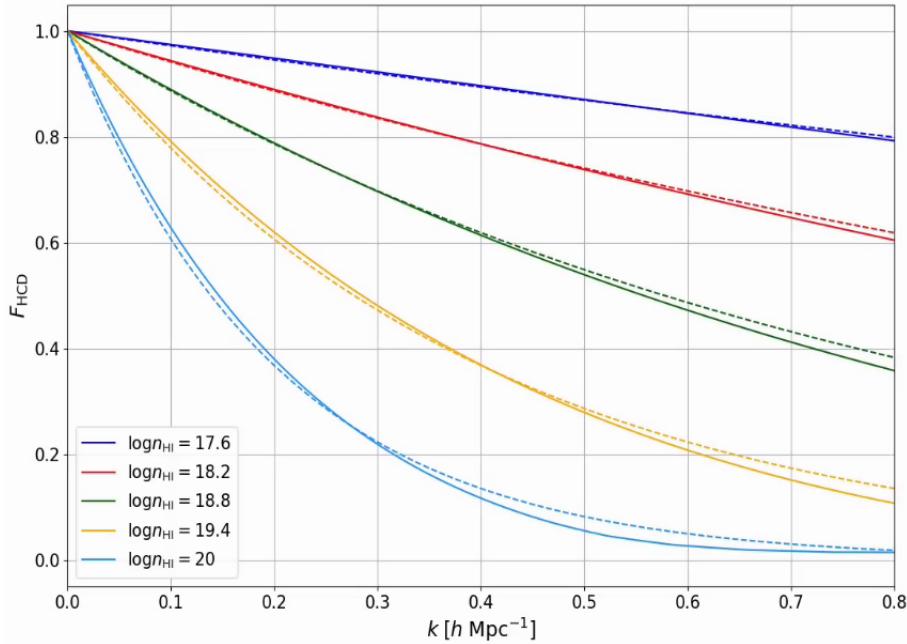


FIGURE 6.13 – Les fonctions  $F_{\text{HCD}}$  utilisées dans le cadre du modèle C-G (lignes continues) ou du modèle de Rogers (lignes tiretées) pour différentes valeurs de  $\log n_{\text{HI}}$ . Les valeurs de  $L_{\text{HCD}}$  utilisées pour le modèle de Rogers sont données dans le texte.

données. Nous calculons donc  $F_{\text{HCD}}$  pour chacune des valeurs de  $\log n_{\text{HI}}$ . Pour le modèle de Rogers, la forme de  $F_{\text{HCD}}$  est fixée (équation 5.28). Il nous suffit, pour chaque valeur de  $\log n_{\text{HI}}$ , de trouver la valeur de  $L_{\text{HCD}}$  à utiliser. Mais ceci n'est pas chose aisée. Dans l'analyse DR16,  $L_{\text{HCD}}$  est choisie de manière un peu arbitraire à  $10 h^{-1}$  Mpc, car il influence très peu la mesure des paramètres BAO. Mais comme nous l'avons vu à plusieurs reprises,  $L_{\text{HCD}}$  influence grandement la mesure des paramètres Ly $\alpha$ . Nous devons donc le choisir avec précaution.

Nous expliquons ici la méthode utilisée pour déterminer  $L_{\text{HCD}}$ . Pour chaque valeur de  $\log n_{\text{HI}}$ , nous choisissons une valeur de  $L_{\text{HCD}}$  qui reproduise au mieux le comportement de  $F_{\text{HCD}}(k)$  calculé dans le modèle C-G dans la gamme  $k \lesssim 0.8 h \text{ Mpc}^{-1}$ . Nous ne nous intéressons pas aux  $k$  plus grands que  $0.8 h \text{ Mpc}^{-1}$  car la fonction de corrélation est estimée dans des bins de  $4 h^{-1}$  Mpc, les échelles plus petites que  $k = \pi/4 \sim 0.8 h \text{ Mpc}^{-1}$  ne sont donc pas accessibles. Ainsi, les valeurs de  $L_{\text{HCD}}$  que nous obtenons pour  $\log n_{\text{HI}} \in [17.6; 18.2; 18.8; 19.4; 20]$  sont  $L_{\text{HCD}} \in [0.28; 0.60; 1.2; 2.5; 5]$ . La figure 6.13 présente les fonctions  $F_{\text{HCD}}(k)$  pour le modèle C-G (lignes continues) et pour le modèle de Rogers (lignes tiretées) en utilisant les valeurs de  $L_{\text{HCD}}$  obtenues précédemment. Pour chaque valeur de  $\log n_{\text{HI}}$ , les fonctions  $F_{\text{HCD}}$  sont très similaires entre les deux modèles.

Nous ajustons maintenant chacune des versions des mocks avec les deux modèles. Pour le modèle de Rogers, nous utilisons les valeurs de  $L_{\text{HCD}}$  déterminées précédemment. Pour chacune des versions des mocks, les ajustements faits avec le modèle de Rogers et le modèle C-G produisent des résultats très similaires. Les différences sont inférieures au dixième de  $\sigma$ . Ceci est rassurant car les fonctions  $F_{\text{HCD}}$  utilisées dans ces deux modèles sont très similaires. Par ailleurs, pour les faibles valeurs de  $\log n_{\text{HI}}$ , les fonctions  $F_{\text{HCD}}$  sont très proches de 1 (figure 6.13). Les paramètres Ly $\alpha$  et HCD sont alors quasiment complètement dégénérés. Le paramètre  $\beta_{\text{Ly}\alpha}$ , en particulier, n'est pas contraint. Lorsque

$\log n_{\text{HI}}$  est plus grand (supérieur à 18,8), les ajustements parviennent à mesurer les paramètres du Ly $\alpha$ . Cependant ces mesures ne sont pas compatibles entre elles et ne sont pas non plus compatibles avec les mesures faites sur les mocks eboss-0.0. Si ces modèles ne parviennent pas à distinguer l'effet du Ly $\alpha$  de l'effet des HCD lorsque ces derniers possèdent une densité de colonne fixe, alors il y a peu d'espoirs qu'ils parviennent à le faire lorsque les HCD sont distribués dans une large gamme en densité de colonne.

Nous comparons maintenant les fonctions  $F_{\text{HCD}}$  correspondant à une large gamme en densité de colonne. Nous considérons premièrement la gamme en densité de colonne des HCD présents dans les mocks eboss-0.2 lorsque les HCD avec  $\log n_{\text{HI}} > 20,3$  sont masqués, c'est à dire la gamme  $17,2 < \log n_{\text{HI}} < 20,3$ . La fonction  $F_{\text{HCD}}$  utilisée dans le modèle de C-G pour cette gamme en  $\log n_{\text{HI}}$  est calculée par `picca`. Comme précédemment, afin de déterminer le  $L_{\text{HCD}}$  à utiliser dans le modèle de Rogers, nous cherchons à reproduire le comportement de  $F_{\text{HCD}}(k)$  utilisé dans le modèle de C-G dans la gamme  $k \lesssim 0,8 h \text{ Mpc}^{-1}$ . Pour ce faire, nous devons choisir  $L_{\text{HCD}} = 2,8 h^{-1} \text{ Mpc}$ . La figure 6.14 présente les fonctions  $F_{\text{HCD}}$  ainsi obtenues (en bleu). Nous pouvons remarquer que pour une large gamme en  $\log n_{\text{HI}}$ , il est plus difficile d'obtenir un accord entre le modèle de Rogers et le modèle C-G. La figure 6.14 présente aussi, à titre de comparaison, la fonction  $F_{\text{HCD}}$  obtenue avec  $L_{\text{HCD}} = 10 h^{-1} \text{ Mpc}$  (en vert). La différence entre la courbe continue bleue et la courbe tiretée verte est très importante. Nous sommes étonnés<sup>1</sup> que les ajustements faits sur les mocks eboss-0.2 avec le modèle de Rogers (avec  $L_{\text{HCD}} = 10 h^{-1} \text{ Mpc}$ ) et le modèle C-G produisent des résultats si similaires (respectivement en rose et en jaune sur la figure 6.9) compte tenu de ces différences sur  $F_{\text{HCD}}$ . Dans tous les cas, nous pensons que la mesure faite avec  $L_{\text{HCD}} = 10 h^{-1} \text{ Mpc}$  n'est pas fiable. La figure 6.14 présente aussi, en rouge, les fonctions  $F_{\text{HCD}}$  obtenues pour la gamme  $17,2 < \log n_{\text{HI}} < 22,5$ . La valeur de  $L_{\text{HCD}}$  choisie est  $6,5 h^{-1} \text{ Mpc}$ . Les fonctions  $F_{\text{HCD}}$  obtenues pour cette gamme dans le cadre du modèle C-G et du modèle de Rogers (courbes rouges) sont très différentes. Le modèle de Rogers atteint ici ses limites.

Suite à ces observations, nous reproduisons l'analyse des mocks eboss-0.2 où les HCD avec une densité de colonne  $\log n_{\text{HI}} > 20,3$  sont masqués. Nous analysons les fonctions de corrélation Ly $\alpha$  × Ly $\alpha$  en utilisant le modèle de Rogers, avec  $L_{\text{HCD}} = 2,8 h^{-1} \text{ Mpc}$ . Ainsi, nous vérifions que les résultats de l'ajustement sont très similaires à ceux obtenus avec le modèle C-G (les résultats varient d'un dixième de  $\sigma$ ). Ceci est attendu car les fonctions  $F_{\text{HCD}}$  sont très similaires. De plus, comme nous l'avons vu dans la section 6.2.3, la mesure des paramètres Ly $\alpha$  faite sur les mocks eboss-0.2 où les HCD avec une densité de colonne  $\log n_{\text{HI}} > 20,3$  sont masqués et ajustés avec le modèle C-G, est compatible avec la mesure faite sur les mocks eboss-0.2 où aucun HCD n'est masqué et ajustés avec le modèle C-G. Ainsi les trois mesures des paramètres Ly $\alpha$  que nous faisons (modèle C-G avec ou sans HCD masqués et modèle de Rogers avec  $L_{\text{HCD}} = 2,8 h^{-1} \text{ Mpc}$ ) sur les mocks eboss-0.2 sont compatibles. Cependant il reste toujours un léger écart entre la mesure faite sur les mocks eboss-0.0 et celle faite sur les mocks eboss-0.2 (environ 5 %, voir figure 6.9).

1. Nous pensons que cela vient du biais des HCD qui est environ 2,2 fois supérieur dans la mesure faite avec le modèle C-G. Nous pensons que le biais des HCD plus faible obtenu avec le modèle de Rogers compense les mesures des paramètres Ly $\alpha$ .

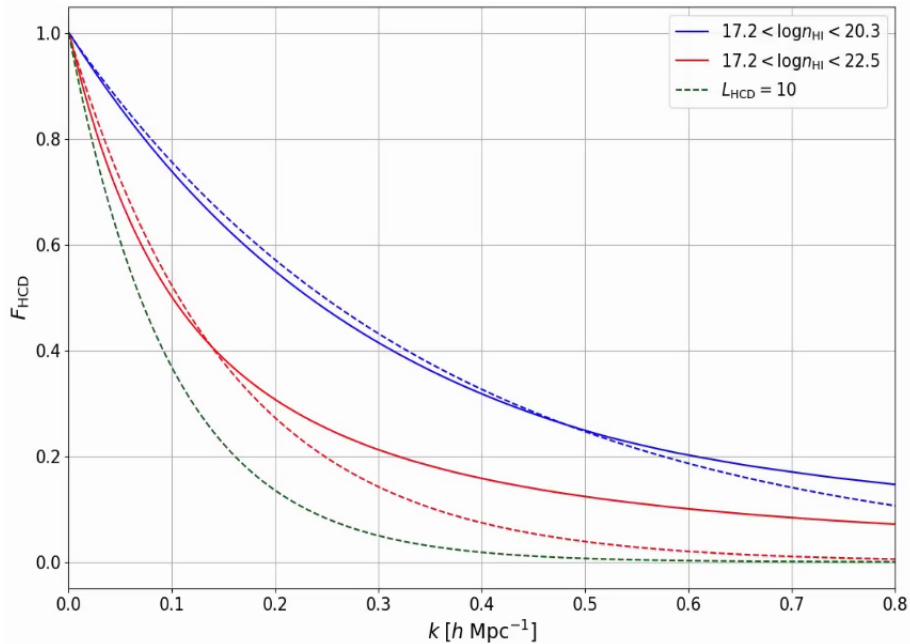


FIGURE 6.14 – Les fonctions  $F_{\text{HCD}}$  utilisées dans le cadre du modèle C-G (lignes continues) ou du modèle de Rogers (lignes tiretées). Les courbes bleues correspondent à la gamme  $17,2 < \log n_{\text{HI}} < 20,3$ . Les courbes rouges correspondent à la gamme  $17,2 < \log n_{\text{HI}} < 22,5$ . La courbe verte donne  $F_{\text{HCD}}$  utilisé dans le modèle de Rogers avec  $L_{\text{HCD}} = 10 h^{-1} \text{ Mpc}$ .

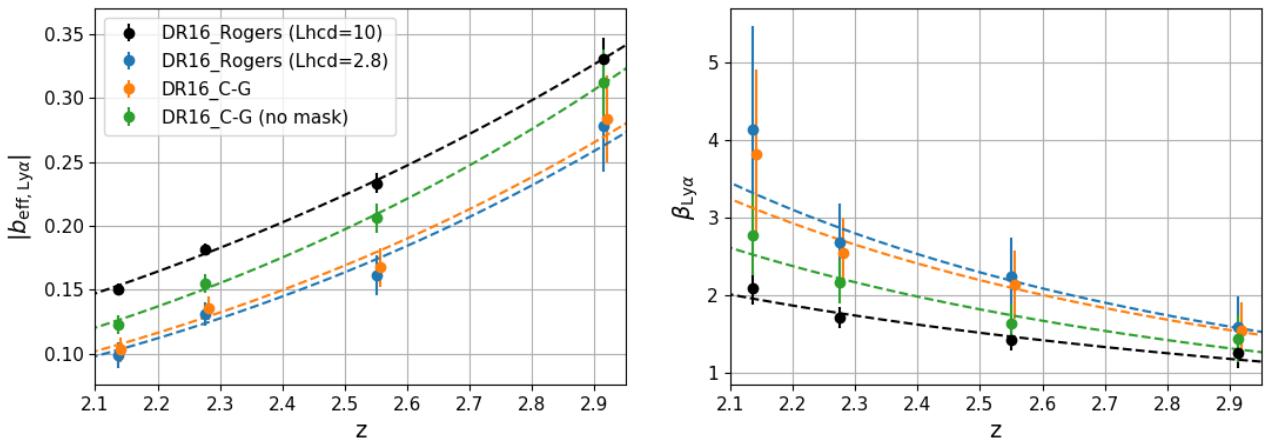


FIGURE 6.15 – Mesure des paramètres  $b_{\text{eff}, \text{Ly}\alpha}$  (gauche) et  $\beta_{\text{Ly}\alpha}$  (droite) sur l'auto-corrélation  $\text{Ly}\alpha \times \text{Ly}\alpha$  estimée à partir des données DR16. Les points bleus présentent la mesure faite avec le modèle de Rogers en utilisant  $L_{\text{HCD}} = 2,8 h^{-1} \text{ Mpc}$ . Les points orange la mesure faite avec le modèle C-G lorsque les HCD avec  $\log n_{\text{HI}} > 20,3$  sont masqués. Les points verts la mesure faite avec le modèle C-G lorsqu'aucun HCD n'est masqué. Enfin, les points noirs présentent la mesure faite avec le modèle de Rogers en utilisant  $L_{\text{HCD}} = 10 h^{-1} \text{ Mpc}$ . Les courbes tiretées correspondent à l'ajustement de chaque jeu de données par une loi de puissance. Les points orange sont décalés horizontalement de  $\Delta z = 5 \times 10^{-3}$  pour des raisons de visibilité.

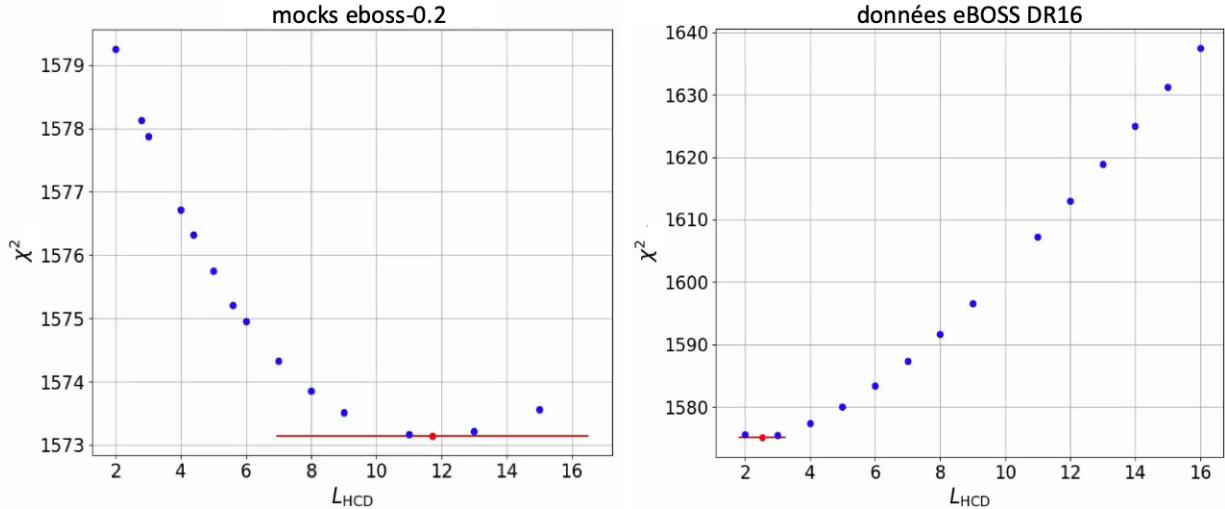


FIGURE 6.16 – Évolution de la valeur de  $\chi^2$  de l'ajustement en fonction du paramètre  $L_{\text{HCD}}$ . La figure de gauche montre l'évolution pour les mocks eboss-0.2 où les HCD avec une densité de colonne  $\log n_{\text{HI}} > 20,3$  sont masqués. Celle de droite l'évolution pour les données DR16 où les HCD sont aussi masqués. Sur chaque graphique, le point rouge présente la valeur, ainsi que la barre d'erreur qui l'accompagne, de  $L_{\text{HCD}}$  qui minimise le  $\chi^2$ .

Après avoir fait ces vérifications sur les mocks, nous analysons les données DR16 avec les modèles de HCD étudiés précédemment. La figure 6.15 présente les mesures des paramètres Ly $\alpha$  faites sur les données DR16 en utilisant ces différents modèles de HCD. Nous pouvons premièrement remarquer que, comme pour les mocks, la mesure faite avec le modèle C-G lorsque les HCD avec une densité de colonne  $\log n_{\text{HI}} > 20,3$  sont masqués (orange) est très similaires à la mesure faite avec le modèle de Rogers lorsque nous utilisons  $L_{\text{HCD}} = 2,8 h^{-1} \text{ Mpc}$  (bleu). Cependant, contrairement aux mocks, ces deux mesures ne sont pas compatibles avec la mesure faite avec le modèle C-G lorsqu'aucun HCD n'est masqué (vert). Le fait que les deux mesures faites avec le modèle C-G (HCD masqués ou non) ne soient pas compatibles, alors que les mêmes mesures faites sur les mocks le sont (voir figure 6.9), pose un sérieux problème. Cela signifie que quelque chose dans les données ne se comporte pas de la même manière que dans les mocks. Il se pourrait qu'un effet inconnu, présent dans les données, soit pris en compte par la modélisation des HCD. Cependant nous n'avons pour l'instant pas d'intuition concernant la nature de cet effet.

La dernière analyse que nous effectuons est d'étudier comment se comporte la valeur de  $\chi^2$  de l'ajustement en fonction du paramètre  $L_{\text{HCD}}$ . Nous effectuons cette étude à la fois pour les mocks et pour les données. La figure 6.16 présente la valeur de  $\chi^2$  de l'ajustement obtenue pour différentes valeurs de  $L_{\text{HCD}}$ . Sur les données DR16 (graphique de droite), nous mesurons  $L_{\text{HCD}} = 2,54 \pm 0,73 h^{-1} \text{ Mpc}$ , ce qui est en accord avec la valeur suggérée par l'analyse présentée dans les paragraphes précédents. Cependant, lorsque nous faisons le même exercice sur les mocks (graphique de gauche), nous mesurons  $L_{\text{HCD}} = 11,73 \pm 4,75 h^{-1} \text{ Mpc}$ . Cette mesure est peu contrainte sur les mocks, et en tension ( $\sim 2\sigma$ ) avec la valeur  $L_{\text{HCD}} = 2,8 h^{-1} \text{ Mpc}$  suggérée par les analyses présentées précédemment.

Ici encore, nous sommes face à une incohérence : lorsque nous étudions la distribution des HCD

présents dans les mocks (avec une densité de colonne comprise dans la gamme  $17,2 < \log n_{\text{HI}} < 20,3$ ), le modèle C-G nous suggère d'utiliser  $L_{\text{HCD}} = 2,8 h^{-1} \text{ Mpc}$ . Pourtant lorsque  $L_{\text{HCD}}$  est laissé libre, la valeur obtenue est en tension avec  $L_{\text{HCD}} = 2,8 h^{-1} \text{ Mpc}$ . Quand nous fixons les paramètres Ly $\alpha$  aux valeurs mesurées sur les mocks eboss-0.0 et laissons libre  $L_{\text{HCD}}$ , nous mesurons  $L_{\text{HCD}} = 4,38 \pm 0,47 h^{-1} \text{ Mpc}$ . Ceci a pour effet de produire une mesure de  $L_{\text{HCD}}$  plus proche de la valeur suggérée par le modèle C-G. Mais cette mesure n'est tout de même pas compatible ( $\sim 3\sigma$ ) avec cette valeur suggérée de  $L_{\text{HCD}} = 2,8 h^{-1} \text{ Mpc}$ .

Après ces nombreuses études, nous ne sommes pas convaincus quant au modèle de HCD à utiliser pour ajuster les données. Le fait que la mesure du paramètre  $L_{\text{HCD}}$ , quand il est laissé libre, soit compatible avec ce que suggère le modèle C-G (grâce à l'étude des fonctions  $F_{\text{HCD}}$ ) encourage à utiliser le modèle de Rogers avec  $L_{\text{HCD}} = 2,8 h^{-1} \text{ Mpc}$ , où de manière équivalente le modèle C-G lorsque les HCD avec une densité de colonne  $\log n_{\text{HI}} > 20,3$  sont masqués. Cependant, le fait que ce résultat ne soit pas également obtenu sur les mocks, c'est à dire que la valeur du paramètre  $L_{\text{HCD}}$  obtenu lorsqu'il est laissé libre dans l'ajustement des mocks eboss-0.2 (avec les HCD masqués) ne soit pas compatible avec la valeur suggérée par l'étude des fonctions  $F_{\text{HCD}}$  avec le modèle C-G, ne nous encourage pas à utiliser le modèle de Rogers avec  $L_{\text{HCD}} = 2,8 h^{-1} \text{ Mpc}$ .

De plus, le fait que les ajustements effectués sur les données DR16 avec le modèle C-G, lorsque les HCD sont masqués ou ne sont pas masqués, ne soient pas compatibles ne nous conforte pas dans l'idée d'utiliser aveuglément ce modèle. Il est probable que des effets systématiques inconnus soient présents dans les données, et que les différents modèles de HCD prennent ces effets en compte.

Étant donné l'état de connaissance actuel de la modélisation des HCD et des effets systématiques qui pourraient affecter les données, nous pensons qu'il est préférable d'utiliser le modèle C-G sans masquer les HCD dans les données. Premièrement, le fait de ne pas masquer les HCD permet d'avoir plus de pixel d'absorptions dans l'analyse, et donc de disposer de plus de statistique pour effectuer la mesure de la position du pic BAO. De plus, lorsque le nombre de HCD présents dans les données est plus important, la fonction  $F_{\text{HCD}}$  est plus importante est donc les paramètres Ly $\alpha$  et ceux des HCD sont moins corrélés. Deuxièmement, le fait de masquer les HCD consiste à retirer une partie non aléatoire des données, qui plus est qui correspond à des régions denses de l'univers. Un tel processus peut introduire des effets systématiques dans l'analyse. Pour ces raisons, nous pensons qu'il est préférable de ne pas masquer les HCD, et donc d'utiliser le modèle C-G.



---

# 7

---

## Conclusion

---

Dans ce manuscrit, j'ai présenté le résultat de mes recherches effectuées au sein du groupe de cosmologie du CEA Saclay, dans le cadre de la préparation de ma thèse de doctorat. Durant ces trois ans, j'ai développé des simulations qui imitent les données Lyman alpha ( $\text{Ly}\alpha$ ) acquises par eBOSS et bientôt par DESI. Ces simulations produisent un relevé de quasars, dont les forêts  $\text{Ly}\alpha$  présentes dans les spectres synthétiques possèdent les bonnes fonctions de corrélation. Pour ce faire, elles utilisent les champs aléatoires gaussiens. D'abord, nous générions un champ aléatoire qui représente le champ de densité. Nous donnons à ce champ aléatoire la bonne fonction de corrélation à l'aide des transformations de Fourier. Puis, grâce à une transformation log-normale, nous tirons un relevé de quasars. Ensuite, à partir de chaque quasar, nous reconstruisons la densité le long de la ligne de visée en interpolant ce champ. Enfin, nous obtenons la fraction de flux transmise en appliquant l'approximation de Gunn-Peterson (FGPA) au champ interpolé le long de la ligne de visée. Ceci nous permet d'obtenir la forêt d'absorption  $\text{Ly}\alpha$  pour chaque quasar de notre relevé. Aussi, nous calculons la prédiction des fonctions de corrélation que nous devons obtenir avec ces simulations. Ceci nous permet de vérifier leur construction en comparant la fonction de corrélation prédictive à la fonction de corrélation mesurée dans nos simulations. Enfin, l'analyse des fonctions de corrélation issues des mocks (chapitre 5) montre que les mocks sont en très bon accord avec la prédiction, et qu'ils sont très bien décrits par les modèles utilisés pour ajuster les données.

Ces simulations ont été utilisées dans l'analyse qui mesure l'échelle des oscillations acoustiques de baryon (BAO) à partir des données finales d'eBOSS DR16 (DU MAS DES BOURBOUX et al. 2020) afin de vérifier que cette analyse produisait des mesures non biaisées de l'échelle BAO. Elles seront aussi utilisées par la collaboration DESI afin de vérifier les analyses  $\text{Ly}\alpha$ . Par ailleurs, ces simulations ont été utilisées par RAVOUX et al. (2020) pour produire une carte tomographique du champ de matière à grande échelle à partir des données d'eBOSS. Elles ont permis de valider l'algorithme utilisé pour reconstruire ce champ de matière.

Nos simulations, avec celles présentées dans FARR et al. (2019), sont à la pointe des programmes qui simulent la distribution d'absorption  $\text{Ly}\alpha$  à grande échelle. Ces simulations sont un atout majeur et indispensable aux analyses de mesure de l'échelle BAO et de mesure des distorsions dans l'espace des redshifts (RSD) avec le  $\text{Ly}\alpha$ . Les simulations que nous présentons dans ce manuscrit ont l'avantage de posséder une fonction de corrélation prédictible, ce qui est très utile pour les analyses RSD. Cette

prédition nous permet d'ajuster efficacement les valeurs du biais  $b_{\text{Ly}\alpha}$ , du paramètre RSD  $\beta_{\text{Ly}\alpha}$ , de la transmission moyenne  $\bar{F}$  et du spectre de puissance  $P^{1D}(k)$  à chaque redshift. De plus, ces simulations bénéficient aussi de distributions de quasars et HCD (High Column Density) très proches de celles observées dans les données, ce qui permet d'obtenir des fonctions de corrélation croisées entre les différents éléments des simulations très correctes. Afin d'améliorer encore ces simulations, il conviendrait d'ajouter les absorptions causées par les HCD et les métaux directement lors de la construction des forêts Ly $\alpha$ , avant l'ajout du continuum de chaque quasar.

Grâce à l'étude approfondie de nos simulations (chapitres 5 et 6), nous montrons que l'analyse des données finales d'eBOSS DR16, présentée dans DU MAS DES BOURBOUX et al. (2020), produit une mesure des paramètres BAO  $\alpha_{\parallel}$  et  $\alpha_{\perp}$  non biaisée à 1 % près. Nous montrons aussi que nous comprenons la forme de la fonction de corrélation Ly $\alpha$  × Ly $\alpha$ , c'est à dire la mesure du biais  $b_{\text{Ly}\alpha}$  et du paramètre RSD  $\beta_{\text{Ly}\alpha}$  du Ly $\alpha$ , à environ 20 % près. Pour affiner cette compréhension, nous devons comprendre pourquoi les distorsions induites sur la fonction de corrélation Ly $\alpha$  × Ly $\alpha$  par l'ajustement du continuum des quasars ne sont pas complètement prises en compte par la matrice de distorsion. Ceci doit passer par une comparaison approfondie des simulations brutes (les raw mocks) et des simulations avec continuum. Afin de simplifier les comparaisons, il est prévu de produire des mocks avec continuum sans ajouter le bruit de mesure et sans inclure la diversité des quasars.

Lors de l'étude des simulations avec HCD (chapitre 6), nous montrons que la modélisation des HCD n'est pas parfaite, et qu'elle mène à des mesures du biais  $b_{\text{Ly}\alpha}$  et du paramètre RSD  $\beta_{\text{Ly}\alpha}$  du Ly $\alpha$  biaisée. Premièrement, la valeur du paramètre  $L_{\text{HCD}}$  dans l'analyse des données d'eBOSS (présentée dans DU MAS DES BOURBOUX et al. (2020)), fixée à  $10 h^{-1} \text{ Mpc}$ , est mal choisie. Nous suggérons d'utiliser  $L_{\text{HCD}} = 2,8 h^{-1} \text{ Mpc}$  si le modèle de Rogers est utilisé. Deuxièmement, nous étudions une modélisation alternative des HCD, que nous appelons le modèle C-G (chapitre 6). Ce modèle produit des mesures compatibles sur les mocks. Cependant, des différences restent à comprendre lorsque nous ajustons les données, notamment lorsque nous ajustons les données avec HCD masqués ou sans masquer les HCD. Suite à ces observations, nous privilégions la piste d'un effet inconnu, présent dans les données (et non dans les mocks) et étant à l'origine de ces différences. Tant que notre compréhension des HCD et de leur modélisation n'aura pas évoluée, nous pensons qu'il est préférable d'utiliser le modèle C-G, sans masquer les HCD. Ceci pour deux raisons : le fait de ne pas masquer les HCD permet de disposer d'une plus grande statistique pour mesurer la position du pic BAO et les paramètres Ly $\alpha$ . De plus, les paramètres Ly $\alpha$  sont moins corrélés dans un tel cas. La seconde raison est que le fait de masquer les HCD peut potentiellement introduire des systématiques dans la fonction de corrélation : les pixels d'absorptions identifiés comme HCD et masqués dans l'analyse ne correspondent pas à une partie aléatoire des données, mais à des régions très denses du ciel. Les masquer peut introduire des systématiques dans l'analyse.

Avec l'arrivée du relevé DESI, une grande quantité de données, et donc une grande puissance statistique, sera mise à la disposition de la communauté scientifique. Cette puissance statistique va permettre d'effectuer des mesures BAO avec le Ly $\alpha$  à différents redshifts, et peut-être de mener une première analyse RSD avec le Ly $\alpha$ . Dans ce but, un effort doit être fait pour comprendre les problèmes liés à la

---

matrice de distorsion et à la modélisation des HCD. Un effort doit aussi être fait pour comprendre les différences qu'il y a entre les mesures faites sur l'auto-corrélation  $\text{Ly}\alpha \times \text{Ly}\alpha$  et les mesures faites sur la corrélation croisée  $\text{Ly}\alpha \times \text{QSO}$  (chapitre 5). Nous montrons dans ce manuscrit que les simulations que nous avons développées permettent d'étudier ces effets et ainsi de comprendre les potentielles systématiques qui affectent ces mesures. Ces simulations sont donc un atout majeur pour préparer les analyses  $\text{Ly}\alpha$  du relevé DESI.



# Appendices

# A

## Fluctuating Gunn Peterson Approximation

---

L'approximation FGPA (Fluctuating Gunn Peterson Approximation), présentée dans la section 4.2.1, relie la profondeur optique  $\tau$  au contraste de densité  $\delta$  :

$$\tau = a \exp(b\delta) , \quad (\text{A.1})$$

où  $a$  et  $b$  sont deux paramètres. Dans cette approximation, l'élargissement thermique des raies d'absorption est négligé.

Afin de relier la profondeur optique  $\tau$  au contraste de densité  $\delta$ , nous débutons avec l'équation qui donne l'équilibre de photo-ionisation de l'hydrogène neutre ( $H_I$ ) :



qui se traduit comme

$$n_\gamma n_{H_I} \langle \sigma_{\text{ioni}} c \rangle = n_e n_p \langle \sigma_{\text{rec}} v \rangle_T , \quad (\text{A.3})$$

où  $n_\gamma$ ,  $n_{H_I}$ ,  $n_e$  et  $n_p$  donnent respectivement les densités de photons, d'hydrogènes neutres, d'électrons et de protons, et  $\sigma_{\text{ioni}}$  et  $\sigma_{\text{rec}}$  donnent respectivement les sections efficaces de l'ionisation d'un atome d'hydrogène neutre et de la recombinaison d'un électron et d'un proton. En négligeant l'hélium et parce que l'hydrogène est quasiment totalement ionisé (entre 1 atome sur 10 000 et 1 atome sur 1000 demeure neutre), nous pouvons approximer les densités d'électron et de protons par la densité de baryons. La densité d'hydrogène neutre s'exprime alors comme

$$n_{H_I} = n_b^2 \frac{\langle \sigma_{\text{rec}} v \rangle_T}{n_\gamma \langle \sigma_{\text{ioni}} c \rangle} . \quad (\text{A.4})$$

De plus, nous avons  $\langle \sigma_{\text{rec}} v \rangle_T \propto T^{-0,7}$ , nous avons donc la relation de proportionnalité

$$n_{H_I} \propto \frac{(1+z)^6 \Omega_b^2 (1+\delta)^2 T^{-0,7}}{n_\gamma \langle \sigma_{\text{ioni}} c \rangle} , \quad (\text{A.5})$$

---

où nous avons explicité la densité de baryon  $n_b^2 = (1+z)^6 \Omega_b^2 (1+\delta)^2$  en fonction du contraste de densité  $\delta$ . Nous pouvons relier de la même manière la température  $T$  au contraste de densité  $\delta$  comme

$$T = \bar{T}(z)(1+\delta)^{\gamma(z)-1}, \quad (\text{A.6})$$

où  $\gamma(z \sim 3) \sim 1,6$  (HUI et GNEDIN 1996). Nous pouvons ainsi relier la profondeur optique  $\tau$  au contraste de densité comme

$$\tau(z) \propto \Omega_b^2 \frac{(1+z)^6 \bar{T}(z)^{-0,7}}{H(z) J_\gamma} \frac{(1+\delta)^\beta}{1+\eta}, \quad (\text{A.7})$$

où  $\eta$  est le gradient de vitesse  $\eta = v_p'(z)/H(z)$ ,  $H(z) J_\gamma$  correspond au flux ionisant de photons, et  $\beta = 2 - 0,7(\gamma(z) - 1) \sim 1,6$  à  $z \sim 3$ .

Lorsque nous nous intéressons uniquement au contraste de densité  $\delta$ , l'équation A.7 et l'approximation log-normale (COLES et JONES 1991) nous donnent

$$\tau(z) \propto \exp(\beta\delta), \quad (\text{A.8})$$

avec  $\beta \sim 1,6$  à  $z \sim 3$ . Cette équation correspond à l'approximation FGPA (Fluctuating Gunn Peterson Approximation). Lorsque nous prenons en compte le gradient de vitesse  $\eta$ , nous obtenons une version légèrement modifiée de l'approximation FGPA

$$\tau(z) \propto \exp(\beta\delta - \eta). \quad (\text{A.9})$$

Cette équation est l'approximation que nous utilisons pour construire les mocks présentés dans ce manuscrit.



# Bibliographie

---

- ABAZAJIAN, K. (2008). « The Seventh Data Release of the Sloan Digital Sky Survey ». In : DOI : [10.1088/0067-0049/182/2/543](https://doi.org/10.1088/0067-0049/182/2/543). arXiv : [0812.0649](https://arxiv.org/abs/0812.0649) (cf. p. 44).
- AGRAWAL, Aniket et al. (2017). « Generating Log-normal Mock Catalog of Galaxies in Redshift Space ». In : DOI : [10.1088/1475-7516/2017/10/003](https://doi.org/10.1088/1475-7516/2017/10/003). arXiv : [1706.09195](https://arxiv.org/abs/1706.09195) (cf. p. 70).
- AHUMADA, Romina et al. (2019). « The Sixteenth Data Release of the Sloan Digital Sky Surveys: First Release from the APOGEE-2 Southern Survey and Full Release of eBOSS Spectra ». In : arXiv : [1912.02905](https://arxiv.org/abs/1912.02905) (cf. p. 60).
- ALAM, Shadab et al. (2015). « The Eleventh and Twelfth Data Releases of the Sloan Digital Sky Survey: Final Data from SDSS-III ». In : DOI : [10.1088/0067-0049/219/1/12](https://doi.org/10.1088/0067-0049/219/1/12). arXiv : [1501.00963](https://arxiv.org/abs/1501.00963) (cf. p. 45).
- ALAM, Shadab et al. (2016). « The clustering of galaxies in the completed SDSS-III Baryon Oscillation Spectroscopic Survey: cosmological analysis of the DR12 galaxy sample ». In : DOI : [10.1093/mnras/stx721](https://doi.org/10.1093/mnras/stx721). arXiv : [1607.03155](https://arxiv.org/abs/1607.03155) (cf. p. 31, 45).
- ARINYO-I-PRATS, Andreu et al. (2015). « The Non-Linear Power Spectrum of the Lyman Alpha Forest ». In : DOI : [10.1088/1475-7516/2015/12/017](https://doi.org/10.1088/1475-7516/2015/12/017). arXiv : [1506.04519](https://arxiv.org/abs/1506.04519) (cf. p. 96, 97, 111).
- BECKER, Robert H., Richard L. WHITE et David J. HELFAND (1995). « The FIRST Survey: Faint Images of the Radio Sky at Twenty Centimeters ». In : *The Astrophysical Journal* 450, p. 559. DOI : [10.1086/176166](https://doi.org/10.1086/176166) (cf. p. 48).
- BENNETT, C. L. et al. (1993). « Scientific results from the Cosmic Background Explorer (COBE) ». In : *Proceedings of the National Academy of Sciences of the United States of America*. T. 90. 11, p. 4766–4773. DOI : [10.1073/pnas.90.11.4766](https://doi.org/10.1073/pnas.90.11.4766) (cf. p. 11).
- BEUTLER, Florian et al. (2011). « The 6dF Galaxy Survey: Baryon acoustic oscillations and the local Hubble constant ». In : *Monthly Notices of the Royal Astronomical Society* 416.4, p. 3017–3032. DOI : [10.1111/j.1365-2966.2011.19250.x](https://doi.org/10.1111/j.1365-2966.2011.19250.x). arXiv : [1106.3366](https://arxiv.org/abs/1106.3366) (cf. p. 31).
- BLANTON, M. R. et al. (2001). « An Efficient Algorithm for Positioning Tiles in the Sloan Digital Sky Survey ». In : DOI : [10.1086/344761](https://doi.org/10.1086/344761). arXiv : [0105535 \[astro-ph\]](https://arxiv.org/abs/0105535) (cf. p. 48).
- BLANTON, Michael R. et al. (2017). « Sloan Digital Sky Survey IV: Mapping the Milky Way, Nearby Galaxies, and the Distant Universe ». In : DOI : [10.3847/1538-3881/aa7567](https://doi.org/10.3847/1538-3881/aa7567). arXiv : [1703.00052](https://arxiv.org/abs/1703.00052) (cf. p. 45, 49).
- BLOMQVIST, Michael et al. (2019). « Baryon acoustic oscillations from the cross-correlation of Lyman-alpha absorption and quasars in eBOSS DR14 ». In : DOI : [10.1051/0004-6361/201935641](https://doi.org/10.1051/0004-6361/201935641). arXiv : [1904.03430](https://arxiv.org/abs/1904.03430) (cf. p. 31).

- BOLTON, Adam S. et al. (2012). « Spectral Classification and Redshift Measurement for the SDSS-III Baryon Oscillation Spectroscopic Survey ». In : DOI : [10.1088/0004-6256/144/5/144](https://doi.org/10.1088/0004-6256/144/5/144). arXiv : [1207.7326](https://arxiv.org/abs/1207.7326) (cf. p. 60).
- BOVY, Jo et al. (2010). « Think Outside the Color Box: Probabilistic Target Selection and the SDSS-XDQSO Quasar Targeting Catalog ». In : DOI : [10.1088/0004-637X/729/2/141](https://doi.org/10.1088/0004-637X/729/2/141). arXiv : [1011.6392](https://arxiv.org/abs/1011.6392) (cf. p. 48).
- BRIDLE, Alan H. et al. (1994). « Deep VLA imaging of twelve extended 3CR quasars ». In : *The Astronomical Journal* 108, p. 766. DOI : [10.1086/117112](https://doi.org/10.1086/117112) (cf. p. 36).
- BUSCA, Nicolas et Christophe BALLAND (2018). « QuasarNET: Human-level spectral classification and redshifting with Deep Neural Networks ». In : arXiv : [1808.09955](https://arxiv.org/abs/1808.09955) (cf. p. 61).
- CALURA, F. et al. (2012). « The Lyman alpha forest flux probability distribution at  $z > 3$  ». In : *Monthly Notices of the Royal Astronomical Society* 422.4, p. 3019–3036. DOI : [10.1111/j.1365-2966.2012.20811.x](https://doi.org/10.1111/j.1365-2966.2012.20811.x). arXiv : [1201.5121](https://arxiv.org/abs/1201.5121) (cf. p. 39).
- CHABANIER, Solène et al. (2018). « The one-dimensional power spectrum from the SDSS DR14 Lyman-alpha forests ». In : DOI : [10.1088/1475-7516/2019/07/017](https://doi.org/10.1088/1475-7516/2019/07/017). arXiv : [1812.03554](https://arxiv.org/abs/1812.03554) (cf. p. 107).
- CHABANIER, Solène et al. (2020). « The impact of AGN feedback on the 1D power spectra from the Lyman-alpha forest using the Horizon-AGN suite of simulations ». In : arXiv : [2002.02822](https://arxiv.org/abs/2002.02822) (cf. p. 35).
- CLERKIN, L. et al. (2016). « Testing the lognormality of the galaxy and weak lensing convergence distributions from Dark Energy Survey maps ». In : DOI : [10.1093/mnras/stw2106](https://doi.org/10.1093/mnras/stw2106). arXiv : [1605.02036](https://arxiv.org/abs/1605.02036) (cf. p. 70).
- COLES, Peter et Bernard JONES (1991). « A lognormal model for the cosmological mass distribution ». In : *Monthly Notices of the Royal Astronomical Society* 248, p. 1–13 (cf. p. 70, 169).
- COLLABORATION, eBOSS et al. (2020). « The Completed SDSS-IV extended Baryon Oscillation Spectroscopic Survey: Cosmological Implications from two Decades of Spectroscopic Surveys at the Apache Point observatory ». In : arXiv 2. arXiv : [2007.08991](https://arxiv.org/abs/2007.08991) (cf. p. 55–57).
- CROOM, Scott M. et al. (2004). « The 2dF QSO Redshift Survey - XIV. Structure and evolution from the two-point correlation function ». In : DOI : [10.1111/j.1365-2966.2004.08379.x/abs/](https://doi.org/10.1111/j.1365-2966.2004.08379.x/abs/). arXiv : [0409314 \[astro-ph\]](https://arxiv.org/abs/0409314) (cf. p. 37).
- DAWSON, Kyle S. et al. (2015). « The SDSS-IV extended Baryon Oscillation Spectroscopic Survey: Overview and Early Data ». In : DOI : [10.3847/0004-6256/151/2/44](https://doi.org/10.3847/0004-6256/151/2/44). arXiv : [1508.04473](https://arxiv.org/abs/1508.04473) (cf. p. 46).
- DE SAINTE AGATHE, Victoria (p.d.). « Mesure de la position du pic d'oscillations acoustiques baryoniques dans les forêts Ly-alpha et Ly-beta des spectres des quasars du relevé eBOSS-SDSS IV ». Thèse de doct. (cf. p. 42, 64, 114).
- DE SAINTE AGATHE, Victoria et al. (2019). « Baryon acoustic oscillations at  $z = 2.34$  from the correlations of Lyman alpha absorption in eBOSS DR14 ». In : DOI : [10.1051/0004-6361/201935638](https://doi.org/10.1051/0004-6361/201935638). arXiv : [1904.03400](https://arxiv.org/abs/1904.03400) (cf. p. 31, 140).
- DESI COLLABORATION et al. (2016). « The DESI Experiment Part I: Science, Targeting, and Survey Design ». In : arXiv : [1611.00036](https://arxiv.org/abs/1611.00036) (cf. p. 55).

- DODELSON, Scott (2003). *Modern Cosmology*. Academic Press. DOI : [10.1023/B:FOOP.0000019699.88151.ed](https://doi.org/10.1023/B:FOOP.0000019699.88151.ed) (cf. p. 6, 14, 26, 72).
- DU MAS DES BOURBOUX, Hélion et al. (2017). « Baryon acoustic oscillations from the complete SDSS-III Ly  $\alpha$  -quasar cross-correlation function at  $z = 2.4$  ». In : *Astronomy & Astrophysics* 608, A130. DOI : [10.1051/0004-6361/201731731](https://doi.org/10.1051/0004-6361/201731731). arXiv : [1708.02225](https://arxiv.org/abs/1708.02225) (cf. p. 45).
- DU MAS DES BOURBOUX, Hélion et al. (2019). « The extended Baryon Oscillation Spectroscopic Survey: measuring the cross-correlation between the MgII flux transmission field and quasars and galaxies at  $z=0.59$  ». In : arXiv : [1901.01950](https://arxiv.org/abs/1901.01950) (cf. p. 106).
- DU MAS DES BOURBOUX, Hélion et al. (2020). « The Completed SDSS-IV extended Baryon Oscillation Spectroscopic Survey: Baryon acoustic oscillations with Lyman-alpha forests ». In : arXiv : [2007.08995](https://arxiv.org/abs/2007.08995) (cf. p. 39, 42, 62, 64, 68, 104, 105, 110, 113, 140, 143, 163, 164).
- EISENSTEIN, D J, H.-J. SEO et M WHITE (2007). « On the Robustness of the Acoustic Scale in the Low-Redshift Clustering of Matter ». In : *ApJ* 664, p. 660–674. DOI : [10.1086/518755](https://doi.org/10.1086/518755) (cf. p. 111).
- EISENSTEIN, D. J. et al. (2005). « Detection of the Baryon Acoustic Peak in the Large-Scale Correlation Function of SDSS Luminous Red Galaxies ». In : DOI : [10.1086/466512](https://doi.org/10.1086/466512). arXiv : [0501171 \[astro-ph\]](https://arxiv.org/abs/0501171) (cf. p. 30, 32, 44).
- EISENSTEIN, Daniel J. et al. (2011). « SDSS-III: Massive Spectroscopic Surveys of the Distant Universe, the Milky Way Galaxy, and Extra-Solar Planetary Systems ». In : DOI : [10.1088/0004-6256/142/3/72](https://doi.org/10.1088/0004-6256/142/3/72). arXiv : [1101.1529](https://arxiv.org/abs/1101.1529) (cf. p. 44).
- ETOURNEAU, Thomas, Jean-Marc LE GOFF et Et AL (2020). « Synthetic quasar spectra with 3D Lyman-alpha correlation ». In : (cf. p. 68).
- FARR, James et al. (2019). « LyCoLoRe: Synthetic Datasets for Current and Future Lyman-alpha Forest BAO Surveys ». In : DOI : [10.1088/1475-7516/2020/03/068](https://doi.org/10.1088/1475-7516/2020/03/068). arXiv : [1912.02763](https://arxiv.org/abs/1912.02763) (cf. p. 83, 163).
- FAUCHER-GIGUERE, C. -A. et al. (2008). « Evolution of the Intergalactic Opacity: Implications for the Ionizing Background, Cosmic Star Formation, and Quasar Activity ». In : DOI : [10.1086/592289](https://doi.org/10.1086/592289). arXiv : [0807.4177](https://arxiv.org/abs/0807.4177) (cf. p. 95, 96).
- FIXSEN, D. J. et al. (1996). « The Cosmic Microwave Background Spectrum from the Full COBE/FIRAS Data Set ». In : DOI : [10.1086/178173](https://doi.org/10.1086/178173). arXiv : [9605054 \[astro-ph\]](https://arxiv.org/abs/9605054) (cf. p. 10).
- FONT-RIBERA, Andreu, Patrick McDONALD et Jordi MIRALDA-ESCUDÉ (2012). « Generating mock data sets for large-scale Lyman-  $\alpha$  forest correlation measurements Generating mock data sets for large-scale Lyman- $\alpha$  forest correlation measurements ». In : *Journal of Cosmology and Astroparticle Physics*. DOI : [10.1088/1475-7516/2012/01/001](https://doi.org/10.1088/1475-7516/2012/01/001) (cf. p. 86).
- FONT-RIBERA, Andreu et Jordi MIRALDA-ESCUDÉ (2012). « The Effect of High Column Density Systems on the Measurement of the Lyman alpha Forest Correlation Function ». In : DOI : [10.1088/1475-7516/2012/07/028](https://doi.org/10.1088/1475-7516/2012/07/028). arXiv : [1205.2018](https://arxiv.org/abs/1205.2018) (cf. p. 70, 88, 112).
- FONT-RIBERA, Andreu et al. (2013). « The large-scale Quasar-Lyman alpha Forest Cross-Correlation from BOSS ». In : DOI : [10.1088/1475-7516/2013/05/018](https://doi.org/10.1088/1475-7516/2013/05/018). arXiv : [1303.1937](https://arxiv.org/abs/1303.1937) (cf. p. 115).
- GONZÁLEZ-MORALES, Alma Xochitl et Et AL (p.d.). « DESI Lyman-alpha synthetic spectra ». In : (cf. p. 100).

- GORSKI, K. M. et al. (2004). « HEALPix – a Framework for High Resolution Discretization, and Fast Analysis of Data Distributed on the Sphere ». In : DOI : [10.1086/427976](https://doi.org/10.1086/427976). arXiv : [0409513 \[astro-ph\]](https://arxiv.org/abs/0409513) (cf. p. 93).
- GUNN, James E., Walter A. SIEGMUND et Edward J. Mannery et AL (2006). « The 2.5 m Telescope of the Sloan Digital Sky Survey ». In : DOI : [10.1086/500975](https://doi.org/10.1086/500975). arXiv : [0602326 \[astro-ph\]](https://arxiv.org/abs/0602326) (cf. p. 49).
- GUNN, J. et al. (1998). « The Sloan Digital Sky Survey Photometric Camera ». In : DOI : [10.1086/300645](https://doi.org/10.1086/300645). arXiv : [9809085 \[astro-ph\]](https://arxiv.org/abs/9809085) (cf. p. 50–52).
- GUO, Zhiyuan et Paul MARTINI (2019). « Classification of Broad Absorption Line Quasars with a Convolutional Neural Network ». In : DOI : [10.3847/1538-4357/ab2590](https://doi.org/10.3847/1538-4357/ab2590). arXiv : [1901.04506](https://arxiv.org/abs/1901.04506) (cf. p. 62).
- HAMILTON, A. J. S. (1992). « Measuring Omega and the real correlation function from the redshift correlation function ». In : *The Astrophysical Journal* 385, p. L5. DOI : [10.1086/186264](https://doi.org/10.1086/186264) (cf. p. 87).
- (1999). « Uncorrelated Modes of the Nonlinear Power Spectrum ». In : DOI : [10.1046/j.1365-8711.2000.03071.x](https://doi.org/10.1046/j.1365-8711.2000.03071.x). arXiv : [9905191 \[astro-ph\]](https://arxiv.org/abs/9905191) (cf. p. 113).
- HANANY, S. et al. (2000). « MAXIMA-1: A Measurement of the Cosmic Microwave Background Anisotropy on angular scales of 10 arcminutes to 5 degrees ». In : *The Astrophysical Journal* 545.1, p. L5–L9. DOI : [10.1086/317322](https://doi.org/10.1086/317322). arXiv : [0005123 \[astro-ph\]](https://arxiv.org/abs/0005123) (cf. p. 30).
- HEITMANN, Katrin et al. (2019). « The Outer Rim Simulation: A Path to Many-Core Supercomputers ». In : DOI : [10.3847/1538-4365/ab4da1](https://doi.org/10.3847/1538-4365/ab4da1). arXiv : [1904.11970](https://arxiv.org/abs/1904.11970) (cf. p. 68).
- HOGG, David W. (1999). « Distance measures in cosmology ». In : arXiv : [9905116 \[astro-ph\]](https://arxiv.org/abs/9905116) (cf. p. 18, 33).
- HUI, Lam et Nickolay Y. GNEDIN (1996). « Equation of State of the Photoionized Intergalactic Medium ». In : DOI : [10.1093/mnras/292.1.27](https://doi.org/10.1093/mnras/292.1.27). arXiv : [9612232 \[astro-ph\]](https://arxiv.org/abs/9612232) (cf. p. 169).
- KAISER, N (1987). « Clustering in real space and in redshift space ». In : *MNRAS* 227, p. 1–21 (cf. p. 33).
- KIRKBY, David et al. (2013). « Fitting Methods for Baryon Acoustic Oscillations in the Lyman-alpha Forest Fluctuations in BOSS Data Release 9 ». In : DOI : [10.1088/1475-7516/2013/03/024](https://doi.org/10.1088/1475-7516/2013/03/024). arXiv : [1301.3456](https://arxiv.org/abs/1301.3456) (cf. p. 111, 113).
- LANDY, Stephen D. et Alexander S. SZALAY (1993). « Bias and variance of angular correlation functions ». In : *The Astrophysical Journal* 412, p. 64. DOI : [10.1086/172900](https://doi.org/10.1086/172900) (cf. p. 107).
- LAURENT, Pierre et al. (2016). « A 14 \$h^{-3}\\$ Gpc\$^3\$ study of cosmic homogeneity using BOSS DR12 quasar sample ». In : DOI : [10.1088/1475-7516/2016/11/060](https://doi.org/10.1088/1475-7516/2016/11/060). arXiv : [1602.09010](https://arxiv.org/abs/1602.09010) (cf. p. 48).
- LAURENT, Pierre et al. (2017). « Clustering of quasars in SDSS-IV eBOSS : study of potential systematics and bias determination ». In : DOI : [10.1088/1475-7516/2017/07/017](https://doi.org/10.1088/1475-7516/2017/07/017). arXiv : [1705.04718](https://arxiv.org/abs/1705.04718) (cf. p. 37).
- LAW, N. M. et al. (2009). « The Palomar Transient Factory: System Overview, Performance and First Results ». In : DOI : [10.1086/648598](https://doi.org/10.1086/648598). arXiv : [0906.5350](https://arxiv.org/abs/0906.5350) (cf. p. 48).
- LAWRENCE, A. et al. (2006). « The UKIRT Infrared Deep Sky Survey (UKIDSS) ». In : DOI : [10.1111/j.1365-2966.2007.12040.x](https://doi.org/10.1111/j.1365-2966.2007.12040.x). arXiv : [0604426 \[astro-ph\]](https://arxiv.org/abs/0604426) (cf. p. 47).

- LE GOFF, J. M. et al. (2011). « Simulations of BAO reconstruction with a quasar Lyman-alpha survey ». In : *Astronomy & Astrophysics* 534, A135. DOI : [10.1051/0004-6361/201117736](https://doi.org/10.1051/0004-6361/201117736). arXiv : [1107.4233](https://arxiv.org/abs/1107.4233) (cf. p. 83).
- LEWIS, Antony, Anthony CHALLINOR et Anthony LASENBY (1999). « Efficient Computation of CMB anisotropies in closed FRW models ». In : DOI : [10.1086/309179](https://doi.org/10.1086/309179). arXiv : [9911177 \[astro-ph\]](https://arxiv.org/abs/9911177) (cf. p. 24, 25, 71).
- LINDER, Eric V. et Robert N. CAHN (2007). « Parameterized Beyond-Einstein Growth ». In : DOI : [10.1016/j.astropartphys.2007.09.003](https://doi.org/10.1016/j.astropartphys.2007.09.003). arXiv : [0701317 \[astro-ph\]](https://arxiv.org/abs/0701317) (cf. p. 34).
- LUMSDEN, S L, A F HEAVENS et J A PEACOCK (1989). « The clustering of peaks in a random Gaussian field ». In : *Monthly Notices of the Royal Astronomical Society* 238, p. 293–318 (cf. p. 80).
- LYKE, Brad W. et al. (2020). « The Sloan Digital Sky Survey Quasar Catalog: Sixteenth Data Release ». In : arXiv : [2007.09001](https://arxiv.org/abs/2007.09001) (cf. p. 61, 62).
- MARTIN, D. Christopher et al. (2004). « The Galaxy Evolution Explorer: A Space Ultraviolet Survey Mission ». In : DOI : [10.1086/426387](https://doi.org/10.1086/426387). arXiv : [0411302 \[astro-ph\]](https://arxiv.org/abs/0411302) (cf. p. 47).
- MASI, S. et al. (2002). « The BOOMERanG experiment and the curvature of the Universe ». In : *Progress in Particle and Nuclear Physics* 48.1, p. 243–261. DOI : [10.1016/S0146-6410\(02\)00131-X](https://doi.org/10.1016/S0146-6410(02)00131-X). arXiv : [0201137 \[astro-ph\]](https://arxiv.org/abs/0201137) (cf. p. 30).
- MCDONALD, Patrick (2003). « Toward a Measurement of the Cosmological Geometry at  $z \sim 2$ : Predicting Ly $\alpha$  Forest Correlation in Three Dimensions and the Potential of Future Data Sets ». In : *The Astrophysical Journal* 585.1, p. 34–51. DOI : [10.1086/345945](https://doi.org/10.1086/345945) (cf. p. 97).
- MCDONALD, Patrick et al. (2004). « The Lyman-alpha Forest Power Spectrum from the Sloan Digital Sky Survey ». In : DOI : [10.1086/444361](https://doi.org/10.1086/444361). arXiv : [0405013 \[astro-ph\]](https://arxiv.org/abs/0405013) (cf. p. 66, 112).
- MILLER, A. et al. (2001). « The QMAP and MAT/TOCO Experiments for Measuring Anisotropy in the Cosmic Microwave Background ». In : *The Astrophysical Journal Supplement Series* 140.2, p. 115–141. DOI : [10.1086/339686](https://doi.org/10.1086/339686). arXiv : [0108030 \[astro-ph\]](https://arxiv.org/abs/0108030) (cf. p. 30).
- PADMANABHAN, T (1993). *Structure Formation in the Universe* (cf. p. 32).
- PÂRIS, Isabelle et al. (2016). « The Sloan Digital Sky Survey Quasar Catalog: twelfth data release ». In : DOI : [10.1051/0004-6361/201527999](https://doi.org/10.1051/0004-6361/201527999). arXiv : [1608.06483](https://arxiv.org/abs/1608.06483) (cf. p. 61).
- PARKS, David et al. (2017). « Deep Learning of Quasar Spectra to Discover and Characterize Damped Ly $\alpha$  Systems ». In : DOI : [10.1093/mnras/sty196](https://doi.org/10.1093/mnras/sty196). arXiv : [1709.04962](https://arxiv.org/abs/1709.04962) (cf. p. 62).
- PIERI, M. M. et al. (2016). « WEAVE-QSO: A Massive Intergalactic Medium Survey for the William Herschel Telescope ». In : arXiv : [1611.09388](https://arxiv.org/abs/1611.09388) (cf. p. 76).
- PLANCK COLLABORATION et al. (2015). « Planck 2015 results. XIII. Cosmological parameters ». In : *ArXiv e-prints* 1502, p. 1589 (cf. p. 31).
- PLANCK COLLABORATION et al. (2018). « Planck 2018 results. VI. Cosmological parameters ». In : arXiv : [1807.06209](https://arxiv.org/abs/1807.06209) (cf. p. 8, 11, 15, 17, 19, 20, 29).
- PLANCK COLLABORATION et al. (2020). « Planck 2018 results: I. Overview and the cosmological legacy of Planck ». In : *Astronomy and Astrophysics* 641. DOI : [10.1051/0004-6361/201833880](https://doi.org/10.1051/0004-6361/201833880). arXiv : [1807.06205](https://arxiv.org/abs/1807.06205) (cf. p. 23).
- RAVOUX, C. et al. (2020). « A tomographic map of the large-scale matter distribution using the eBOSS Stripe 82 Lyman-alpha forest ». In : arXiv : [2004.01448](https://arxiv.org/abs/2004.01448) (cf. p. 163).

- RICH, James (2010). *Fundamentals of Cosmology*. DOI : [10.1063/1.1522222](https://doi.org/10.1063/1.1522222) (cf. p. 6).
- RIESS, Adam G. et al. (2019). « Large Magellanic Cloud Cepheid Standards Provide a 1% Foundation for the Determination of the Hubble Constant and Stronger Evidence for Physics Beyond LambdaCDM ». In : DOI : [10.3847/1538-4357/ab1422](https://doi.org/10.3847/1538-4357/ab1422). arXiv : [1903.07603](https://arxiv.org/abs/1903.07603) (cf. p. 15, 55, 57).
- ROGERS, Keir K. et al. (2017). « Correlations in the three-dimensional Lyman-alpha forest contaminated by high column density absorbers ». In : arXiv : [1711.06275](https://arxiv.org/abs/1711.06275) (cf. p. 42, 112).
- ROSS, Ashley J. et al. (2014). « The Clustering of the SDSS DR7 Main Galaxy Sample I: A 4 per cent Distance Measure at  $z=0.15$  ». In : DOI : [10.1093/mnras/stv154](https://doi.org/10.1093/mnras/stv154). arXiv : [1409.3242](https://arxiv.org/abs/1409.3242) (cf. p. 31).
- SCOLNICK, D. M. et al. (2017). « The Complete Light-curve Sample of Spectroscopically Confirmed Type Ia Supernovae from Pan-STARRS1 and Cosmological Constraints from The Combined Pantheon Sample ». In : DOI : [10.3847/1538-4357/aab9bb](https://doi.org/10.3847/1538-4357/aab9bb). arXiv : [1710.00845](https://arxiv.org/abs/1710.00845) (cf. p. 56).
- SMEE, Stephen et al. (2012). « The Multi-Object, Fiber-Fed Spectrographs for SDSS and the Baryon Oscillation Spectroscopic Survey ». In : DOI : [10.1088/0004-6256/146/2/32](https://doi.org/10.1088/0004-6256/146/2/32). arXiv : [1208.2233](https://arxiv.org/abs/1208.2233) (cf. p. 50, 51, 53–55).
- THE SDSS-II SEGUE COLLABORATION et al. (2009). « SEGUE: A Spectroscopic Survey of 240,000 stars with  $g=14\text{--}20$  ». In : DOI : [10.1088/0004-6256/137/5/4377](https://doi.org/10.1088/0004-6256/137/5/4377). arXiv : [0902.1781](https://arxiv.org/abs/0902.1781) (cf. p. 44).
- WANG, Feige et al. (2018). « Exploring Reionization-Era Quasars III: Discovery of 16 Quasars at  $6.4 < z < 6.9$  with DESI Legacy Imaging Surveys and UKIRT Hemisphere Survey and Quasar Luminosity Function at  $z \sim 6.7$  ». In : DOI : [10.3847/1538-4357/ab2be5](https://doi.org/10.3847/1538-4357/ab2be5). arXiv : [1810.11926](https://arxiv.org/abs/1810.11926) (cf. p. 40).
- WRIGHT, Edward L. et al. (2010). « The Wide-field Infrared Survey Explorer (WISE): Mission Description and Initial On-orbit Performance ». In : DOI : [10.1088/0004-6256/140/6/1868](https://doi.org/10.1088/0004-6256/140/6/1868). arXiv : [1008.0031](https://arxiv.org/abs/1008.0031) (cf. p. 47).
- ZARROUK, Pauline (2018). « Clustering Analysis in Configuration Space and Cosmological Implications of the SDSS-IV eBOSS Quasar Sample ». Thèse de doct. (cf. p. 17).



**Titre :** Les forêts Lyman- $\alpha$  du relevé eBOSS : comprendre les fonctions de corrélation et les systématiques

**Mots clés :** cosmologie, relevé à grand redshift, énergie noire, forêts Lyman- $\alpha$ , mocks, champs aléatoires gaussiens

**Résumé :** Cette thèse s'inscrit dans le cadre des projets eBOSS et DESI. Ces projets utilisent, entre autres, l'absorption Lyman- $\alpha$  ( $\text{Ly}\alpha$ ) afin de sonder la répartition de matière dans l'univers et ainsi mesurer l'échelle des oscillations acoustiques de baryon (BAO). La mesure du rapport de la taille de l'échelle BAO et de la taille de l'horizon acoustique lors du découplage des baryons et des photons permet de contraindre l'expansion de l'univers, et donc les paramètres de l'équation d'état de l'énergie noire.

Cette thèse présente le développement de simulations (ou pseudo-données, ou mocks) à partir de champs aléatoires gaussiens (GRF) utilisées afin de tester les analyses BAO des groupes  $\text{Ly}\alpha$  de eBOSS et DESI. Les GRF permettent de générer un champ de densité  $\delta$ . A partir de ce champ de densité, les positions des quasars (QSO) sont tirées, puis à partir de chaque quasar, les lignes de visées sont constituées. Le champ de densité  $\delta$  est interpolé le long de ces lignes de visées. Enfin, à l'aide de l'approximation FGPA (Fluctuating Gunn Peterson Approximation), la densité interpolée est transformée en profondeur optique  $\tau$ , puis en absorption  $\text{Ly}\alpha$ . Grâce à un programme développé par la communauté de DESI, un continuum est ajouté aux forêts  $\text{Ly}\alpha$  afin de créer des spectres de quasars synthétiques. Les mocks présentées dans ce manuscrit fournissent donc un relevé de quasars dont les forêts  $\text{Ly}\alpha$  présentes dans les spectres possèdent les bonnes

fonctions d'auto-corrélation  $\text{Ly}\alpha \times \text{Ly}\alpha$ , de corrélation croisée  $\text{Ly}\alpha \times \text{QSO}$ , ainsi que d'auto-corrélation  $\text{QSO} \times \text{QSO}$  et  $\text{HCD} \times \text{HCD}$  (High Column Density systems).

L'étude de ces mocks permet de montrer que l'analyse BAO menée sur l'ensemble des données  $\text{Ly}\alpha$  du relevé eBOSS produit une mesure non biaisée des paramètres BAO  $\alpha_{\parallel}$  et  $\alpha_{\perp}$ . Par ailleurs, une étude approfondie du modèle utilisé pour ajuster les fonctions de corrélation montre que la forme de la fonction d'auto-corrélation  $\text{Ly}\alpha \times \text{Ly}\alpha$ , c'est à dire les mesures du biais  $b_{\text{Ly}\alpha}$  et du paramètre RSD (Redshift Space Distortions)  $\beta_{\text{Ly}\alpha}$ , est comprise à environ 20 % près. Les systématiques qui affectent les mesures des paramètres  $\text{Ly}\alpha$  ( $b_{\text{Ly}\alpha}$  et  $\beta_{\text{Ly}\alpha}$ ) sont issues de deux effets. Le premier effet provient de la matrice de distorsion qui ne capture pas l'intégralité des distorsions produites par l'ajustement du continuum des quasars. Le second effet est lié à la modélisation des HCD. La modélisation de ces absorbeurs denses n'est pas parfaite et affecte la mesure des paramètres  $\text{Ly}\alpha$ , en particulier le paramètre RSD  $\beta_{\text{Ly}\alpha}$ .

L'analyse de ces mocks permet donc de valider un bon contrôle des systématiques pour les analyses BAO avec le  $\text{Ly}\alpha$ . Cependant, une meilleure compréhension des mesures des paramètres  $\text{Ly}\alpha$  est nécessaire afin d'envisager une analyse RSD à l'aide de la combinaison de l'auto-corrélation  $\text{Ly}\alpha \times \text{Ly}\alpha$  et de la corrélation croisée  $\text{Ly}\alpha \times \text{QSO}$ .

**Title :** The Lyman- $\alpha$  forests from the eBOSS survey : understanding the correlation functions and the systematics

**Keywords :** cosmology, high-redshift survey, dark energy, Lyman- $\alpha$  forests, mock data sets, gaussian random fields

**Abstract :** This PhD thesis is part of eBOSS and DESI projects. These projects, among other tracers, use the Lyman- $\alpha$  ( $\text{Ly}\alpha$ ) absorption to probe the matter distribution in the universe and measure the baryon acoustic oscillations (BAO) scale. The measurement of the BAO scale to the sound horizon ratio allows to constrain the universe expansion and so the  $\Lambda\text{CDM}$  model, the standard model of cosmology.

This thesis presents the development of mock data sets used in order to check the BAO analyses carried out by the  $\text{Ly}\alpha$  group within the eBOSS and DESI collaborations. These mocks make use of gaussian random fields (GRF). GRF allow to generate a density field  $\delta$ . From this density field, quasar (QSO) positions are drawn. From each quasar, a line of sight is constructed. Then, the density field  $\delta$  is interpolated along each line of sight. Finally, the fluctuating Gunn Peterson approximation (FGPA) is used to convert the interpolated density into the optical depth  $\tau$ , and then into the  $\text{Ly}\alpha$  absorption. Thanks to a program developed by the DESI community, a continuum is added to each  $\text{Ly}\alpha$  forest in order to produce quasar synthetic spectra. The mocks presented in the manuscript provide a survey of quasars whose  $\text{Ly}\alpha$  forests in the quasar spectra have the correct  $\text{Ly}\alpha \times \text{Ly}\alpha$  auto-correlation,  $\text{Ly}\alpha \times \text{QSO}$  cross-correlation, as well as the correct  $\text{QSO} \times \text{QSO}$

and  $\text{HCD} \times \text{HCD}$  (High Column Density systems) auto-correlation functions.

The study of these mocks shows that the BAO analysis run on the whole  $\text{Ly}\alpha$  eBOSS data set produces a non-biased measurement of the BAO parameters  $\alpha_{\parallel}$  et  $\alpha_{\perp}$ . In addition, the analysis of the model used to fit the correlation functions shows that the shape of the  $\text{Ly}\alpha \times \text{Ly}\alpha$  auto-correlation, which is linked to the bias  $b_{\text{Ly}\alpha}$  and redshift space distortions (RSD) parameter  $\beta_{\text{Ly}\alpha}$ , are understood up to 80 %. The systematics affecting the measurement of the  $\text{Ly}\alpha$  parameters ( $b_{\text{Ly}\alpha}$  et  $\beta_{\text{Ly}\alpha}$ ) come from two different effects. The first one originates from the distortion matrix which does not capture all the distortions produced by the quasar continuum fitting procedure. The second one is linked to the HCD modelling. The modelling of these strong absorbers is not perfect and affects the measurement of the  $\text{Ly}\alpha$  parameters, especially the RSD parameter  $\beta_{\text{Ly}\alpha}$ .

Thus, the analysis of these mocks allows to validate the systematic control of the BAO analyses done with the  $\text{Ly}\alpha$ . However, a better understanding of the measurement of the  $\text{Ly}\alpha$  parameters is required in order to consider using the  $\text{Ly}\alpha$ , which means combining the  $\text{Ly}\alpha \times \text{Ly}\alpha$  auto-correlation and  $\text{Ly}\alpha \times \text{QSO}$  cross-correlation, to do a RSD analysis.

Université Paris-Saclay  
Espace Technologique / Immeuble Discovery  
Route de l'Orme aux Merisiers RD 128 / 91190 Saint-Aubin, France