UNIVERSITAT DE
BARCELONA

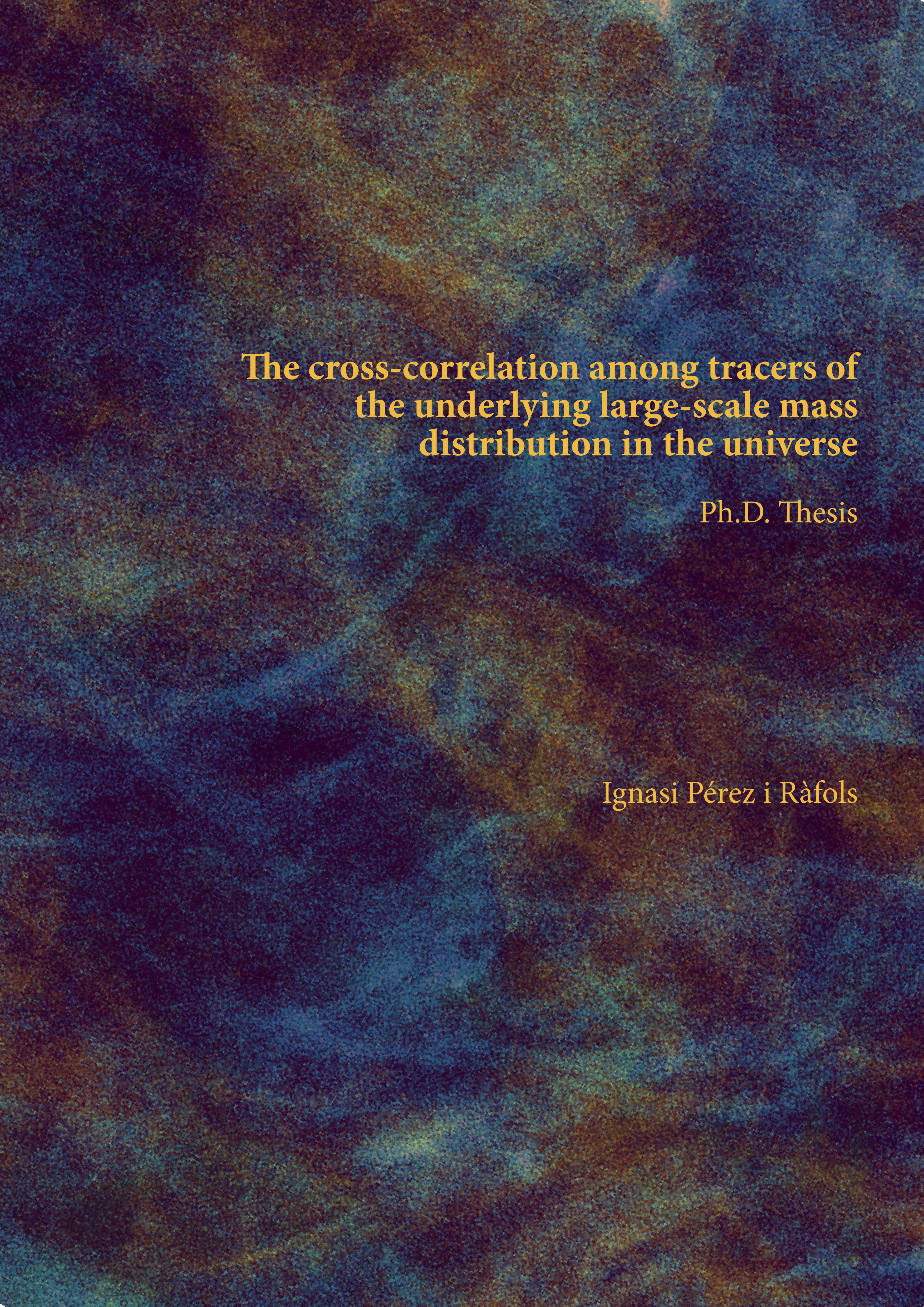
# The cross-correlation among tracers of the underlying large-scale mass distribution in the universe

Ignasi Pérez i Ràfols

Aquesta tesi doctoral està subjecta a la llicència ***Reconeixement- NoComercial – CompartirIgual 4.0. Espanya de Creative Commons***.

Esta tesis doctoral está sujeta a la licencia ***Reconocimiento - NoComercial – CompartirIgual 4.0. España de Creative*** **Commons.**

This doctoral thesis is licensed under the ***Creative Commons Attribution-NonCommercial-ShareAlike 4.0. Spain License***.



# The cross-correlation among tracers of the underlying large-scale mass distribution in the universe

Ph.D. Thesis

Ignasi Pérez i Ràfols

UNIVERSITAT DE
BARCELONA

# The cross-correlation among tracers of the underlying large-scale mass distribution in the universe

Ignasi Pérez i Ràfols

PH.D. THESIS IN PHYSICS

BARCELONA (CATALONIA), MAY 2016

PROGRAMA DE DOCTORAT EN FÍSICA
LÍNIA DE RECERCA EN ASTRONOMIA I ASTROFÍSICA
2013-2016

Tesi doctoral presentada per **Ignasi Pérez i Ràfols**
per optar al grau de Doctor per la Universitat de Barcelona

Director

Prof. Jordi Miralda Escudé

Tutor

Prof. Alberto Manrique Oliva

cc BY NC SA

Ignasi Pérez i Ràfols: *The cross-correlation among tracers of the underlying large-scale mass distribution in the universe*, Ph.D. Thesis, May 2016.

Front cover: Artistic impression of the IGM. The image is composited from three black-and-white firewood smoke images converted to red, blue, and green.

Back cover: Superimposed to the front cover, a word frequency cloud using this thesis text generated by `https://tagul.com/`.

# Agraïments

En primer lloc voldria agraïr al meu director de tesi, el doctor Jordi Miralda Escudé, per la seva dedicació, esforç, paciència i tracte personal al llarg de la realització d'aquesta tesi. El seu àmpli coneixement i el seu rigor científic han constituït sens dubte un dels pilars sobre els que s'ha construït aquesta tesi doctoral.

També voldria agraïr als membres de la col·laboració de BOSS la seva acollida i ensenyament, tot i que no he participat en tantes *telecons* com voldria, n'he après un munt. Voldria agraïr en especial a la doctora Britt Lundgren, al doctor Andreu Font Ribera, al doctor Nicolás Busca, al doctor Michael Blomqvist i al doctor David Kirkby per la seva acollida i ajuda en les estades de recerca.

Igualment voldria agraïr la feina feta per la secretaria del DAM i per l'equip d'informàtics, en especial a el JR i al Gaby, que m'han facilitat enormement la feina.

També voldria agraïr als meus companys del DAM (als nous i als que ja han marxat), que m'ha sabut aguantar dins i fora del deparament durant aquests anys i a tota la gent que m'ha donat suport al llarg d'aquest camí. Finalment, voldria agraïr a la meva família per l'estima, el suport, la paciència i també els tocs d'atenció quan eren necessaris, tot això m'ha permès arribar fins aquí.

A tots i totes, gràcies.

# Contents

**Resum de la tesi** **vii**

**I Introduction** **1**

**1 Introducció** **3**

1.1 L'univers homogeni . . . . . 3

1.2 La matèria fosca i la formació d'estructura . . . . . 6

1.3 L'expansió accelerada de l'univers: el model ΛCDM . . . . . 7

1.4 Els traçadors de la distribució subjacent de matèria . . . . . 8

1.4.1 Quàsars i galàxies . . . . . 11

1.4.2 El bosc de Lyα . . . . . 12

1.4.3 Sistemes Lyα esmorteïts . . . . . 12

1.4.4 Sistemes absorbidors metàl·lics . . . . . 12

1.5 Descripció de la tesi . . . . . 13

**2 Introduction** **15**

2.1 The homogeneous universe . . . . . 15

2.2 Dark matter and formation of structure . . . . . 18

2.3 Accelerated expansion of the universe: ΛCDM . . . . . 19

2.4 Tracers of the underlying dark matter distribution . . . . . 20

2.4.1 Quasars and galaxies . . . . . 22

2.4.2 Lyα forest . . . . . 23

2.4.3 Damped Lyα Systems . . . . . 24

2.4.4 Metal absorbers . . . . . 24

2.5 Description of the thesis . . . . . 24

**II Results** **27**

**3 Mg II absorbers - CMASS galaxies cross-correlations** **29**
- 3.1 Introduction . . . . . 30
- 3.2 Data Sample . . . . . 32
- 3.3 Stacking procedure . . . . . 34
  - 3.3.1 Method 1: Mean subtraction . . . . . 35
  - 3.3.2 Method 2: Variable smoothing . . . . . 41
  - 3.3.3 Unbiasing the composite spectra . . . . . 41
  - 3.3.4 Bootstrap errors . . . . . 41
- 3.4 Model for the Absorption Profile . . . . . 42
- 3.5 Results . . . . . 44
- 3.6 Discussion . . . . . 49
  - 3.6.1 Relation of the mean equivalent width to the bias factor of Mg II absorption systems . . . . . 49
  - 3.6.2 Mean absorption from Mg II systems . . . . . 52
  - 3.6.3 Derivation of the bias factor of Mg II systems . . . . . 54
  - 3.6.4 The ratio of Mg II-absorbing gas to the total mass . . . . . 58
- 3.7 Conclusions . . . . . 59
- 3.A Tests of the continuum fitting methods . . . . . 60
  - 3.A.1 Systematic errors in the absence of correlations . . . . . 60
  - 3.A.2 Tests of the equivalent width measurement for individually detected systems 61
  - 3.A.3 Impact of individually undetected systems on the mean equivalent width 62

**4 DLA-Ly$\alpha$ forest cross-correlations** **67**
- 4.1 Introduction . . . . . 68
- 4.2 Data Sample . . . . . 70
  - 4.2.1 DLA Sample . . . . . 70
  - 4.2.2 Ly$\alpha$ Sample . . . . . 72
- 4.3 Cross-correlation . . . . . 72
  - 4.3.1 Projector . . . . . 74
  - 4.3.2 Estimator for the cross-correlation . . . . . 77
  - 4.3.3 Covariance matrix . . . . . 77
- 4.4 Fitting the cross-correlation . . . . . 78
  - 4.4.1 Distortion matrix . . . . . 78
  - 4.4.2 Model . . . . . 79
- 4.5 Results . . . . . 80
  - 4.5.1 Measured bias . . . . . 82
  - 4.5.2 Bias dependence on redshift . . . . . 85
  - 4.5.3 Bias dependence on neutral hidrogen column density . . . . . 86

4.5.4 Scale dependence of the bias factor . . . . . . . . . . . . . . . . . . . . . 86
4.5.5 Dependence on the Lyα bias . . . . . . . . . . . . . . . . . . . . . . . . . 92
4.6 Summary and Conclusions . . . . . . . . . . . . . . . . . . . . . . . . . . . 93
4.A Lyα autocorrelation contribution to the covariance matrix . . . . . . . . . . . 94
4.B Limits on the linear theory . . . . . . . . . . . . . . . . . . . . . . . . . . . 95

**5 Quasar-Lyα forest cross-correlations** **99**
5.1 Introduction . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 100
5.2 Data Sample . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 101
5.2.1 Quasar Sample . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 101
5.2.2 Lyα Sample . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 102
5.3 Quasar-Lyα cross-correlations . . . . . . . . . . . . . . . . . . . . . . . . . . 102
5.4 Radiation model . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 106
5.4.1 Radiation contribtuion to the cross correlation . . . . . . . . . . . . . . 109
5.4.2 Amplitude of the effect . . . . . . . . . . . . . . . . . . . . . . . . . . . 110
5.4.3 Feasibility of the model to measure the quasar average lifetime . . . . . . 111
5.5 Radiation model in Fourier space . . . . . . . . . . . . . . . . . . . . . . . . 111
5.6 Summary and conclusions . . . . . . . . . . . . . . . . . . . . . . . . . . . . 114

**III Summary and conclusions** **117**

**6 Conclusions and future perspectives** **119**
6.1 Conclusions . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 119
6.2 Future perspectives . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 120

**Bibliography** **121**

**List of Figures** **127**

**List of Tables** **133**

**List of Acronyms** **135**

# Resum de la tesi

El treball fet en aquesta tesi doctoral es basa principalment en la mesura i interpretació de la correlació creuada entre diferents traçadors. Els traçadors emprats en aquesta tesi són les galàxies del catàleg CMASS, els quàsars del catàleg de Baryon Oscillations Spectroscopic Survey (BOSS), els sistemes absorbidors de Mg II, el bosc de Ly$\alpha$ i els sistemes Ly$\alpha$ esmorteïts o DLAs (els tres darrers són sistemes absorbidors detectats als espectres dels quàsars).

El projecte BOSS és un dels quatre projectes que han composat la tercera fase de la col·laboració de l'Sloan Digital Sky Survey (SDSS) el principal objectiu del qual ha estat la detecció del senyal de les oscil·lacions acústiques dels barions (BAO, de l'anglès *Baryon Acoustic Oscillations*) i que ha estat recollint espectres electromagnètics de quàsars i galàxies des de la tardor de 2009 fins a la tardor de 2014.

L'estructura d'aquesta tesi es divideix en tres capítols que fan referència a l'estudi de diferents traçadors. En el capítol 3 els traçadors emprats són les galàxies de CMASS i els sistemes absorbidors de Mg II; en el capítol 4, els DLA i el bosc de Ly$\alpha$, i en el capítol 5, els quàsars i el bosc de Ly$\alpha$. A més la tesi compta amb una introducció en català, una altra en anglès i un compendi de les conclusions obtingudes. A continuació es resumeix el contingut dels capítols principals.

## Correlació encreuada entre absorbidors de Mg II i les galàxies CMASS

En el capítol 3 presento un estudi de la correlació encreuada dels absorbidors de Mg II amb les galàxies CMASS, centrant-me en els efectes a gran escala. La gran quantitat d'espectres de quàsars (d'on mesurem l'absorció de Mg II) i galàxies de la nostra mostra ens permeten avaluar el factor de biaix dels absorbidors de Mg, atès que el biaix de les galàxies CMASS és conegut prèviament.

En aquest capítol he desenvolupat un nou mètode, amb la col·laboració de la Dra. Britt Lundgren i del Dr. Jordi Miralda Escudé, per a poder incloure a l'anàlisi els sistemes absorbidors més febles, impossibles de detectar-se individualment. El mètode proposat es basa en no emprar la zona de l'espectre on es sospita que hi pugui haver absorció per tal de determinar el continu del quàsar. El mètode desenvolupat ve acompanyat per diversos tests que en verifiquen el funcionament. Aquests test també mostren que aquests sistemes febles havien estat obviats en estudis anteriors i implica que les mesures obtingudes en aquells estudis podrien estar esbiaixades per aquesta raó.

Si bé aquest mètode permet incloure absorbidors molt febles, el fet que no serveixi per a la detecció individualitzada d'aquests absorbidors limita la seva versatilitat a estudis de caire estadístic, com és el cas de la correlació encreuada. Tot i aquesta limitació, el mètode presentat permetrà mesurar la correlació encreuada per a altres sistemes absorbidors que presenten una absorció més lleu. Aquests absorbidors són massa febles per a tenir un catàleg d'absorbidors amb un nombre de deteccions suficientment gran com per a que els resultats de la correlació encreuada siguin significatius.

## Correlació encreuada entre DLA i el bosc de Ly$\alpha$

En el capítol 4 presento un estudi de la correlació encreuada dels DLAs amb el bosc de Ly$\alpha$, centrant-me, també en aquest cas, en els efectes a gran escala. Tot i que el biaix dels DLAs ja havia estat calculat amb anterioritat a Font-Ribera et al. (2012), en aquest capítol hi trobem una mesura actualitzada i més acurada.

A més, el fet de tenir una mostra amb entre dues i tres vegades més quantitat d'espectres que les mostres anteriors ens ha permès de subdividir-la primer en funció de la densitat de columna i després en funció del desplaçament cap al roig observat en els DLAs i poder realitzar mesures del biaix dels DLAs presents en aquestes submostres. Així doncs, hem analitzat un total de 9 mostres: la mostra complerta que conté tots els DLAs, dues submostres on hem limitat els talls per incrementar la puresa del catàleg, tres submostres que contenen els DLAs en funció de la seva densitat de columna i tres submostres que contenen els DLAs en funció del seu desplaçament cap al roig.

D'aquesta manera he pogut estudiar, amb la col·laboració de Miralda Escudé i del Dr. Andreu Font Ribera, l'evolució del biaix dels DLAs en funció tant de la densitat de columna com del seu desplaçament cap al roig. Les dades actuals apunten a una discrepància del biaix dels DLAs respecte de les mesures prèvies (Font-Ribera et al. 2012) a un nivell de $\sim 1.5\sigma$. Aquesta discrepància pot ser deguda a diferents efectes.

Una opció és que el formalisme emprat per a calcular la correlació encreuada no acabi de corregir les distorsions introduïdes per l'estimació de l'espectre d'emissió dels quàsars, o bé que la correcció emprada a Font-Ribera et al. (2012) en sobreestima la correcció. Una segona possible explicació rau en el fet que trobem una dependència del biaix dels DLAs que és diferent a la trobada per Font-Ribera et al. (2012).

En l'estudi de les submostres observem una lleugera disminució del biaix dels DLAs tant quan augmenta el desplaçament cap al roig com quan augmenta la densitat de columna. Tanmateix, aquestes tendències són molt poc estadísticament significatives i és força probable que siguin degudes a una fluctuació estadística de la mesura més que no pas una disminució real.

Finalment, la gran quantitat d'espectres presents en la mostra complerta ens ha permès realitzar un estudi de la dependència del biaix dels DLAs amb l'escala a la qual es realitza la mesura. Comparem aquesta mesura amb la predicció teòrica que diu que el biaix dels DLAs hauria de prendre el mateix valor independentment de l'escala a la qual es realitzi la mesura.

## Correlació encreuada entre quàsars i el bosc de Ly$\alpha$

En el capítol 5 presento un estudi de la correlació encreuada dels quàsars amb el bosc de Ly$\alpha$. En aquest cas, contràriament als estudis presentats en els altres capítols, m'he centrat en els efectes a petita escala. Hi trobem una mesura d'aquesta correlació encreuada i una descripció dels efectes més importants que s'observen a petita escala.

Aquests efectes inclouen la contaminació del bosc de Ly$\alpha$ per diverses línies metàl·liques que afegeixen alguns bonys en la correlació encreuada quan la distància perpendicular dels quàsars i el bosc de Ly$\alpha$ és petita. Modelar aquest efecte ens pot permetre millorar els ajustos a models teòrics i també ens aporta informació sobre l'absorció de metalls.

He observat també una asimetria en la mesura de la correlació encreuada en respecte el cas en què el quàsar es trobi davant o darrere del corresponent núvol d'hidrogen amb qui se l'estigui aparellant. Aquesta asimetria podria ser explicada si es té en compte que l'emissió dels quàsars presenta una alta variabilitat temporal.

Per provar d'entendre més a fons l'origen d'aquesta asimetria, en aquest capítol he adoptat una vessant més teòrica al desenvolupar un model sobre com els efectes de la radiació dels quàsars sobre els núvols d'hidrogen més propers es manifesten a la mesura de la correlació encreuada. Aquest model, desenvolupat amb la col·laboració de Miralda Escudé, està calculat tant a l'espai real com a l'espai de Fourier.

# Part I

# Introduction

# 1

# Introducció

## 1.1 L'univers homogeni

Els models cosmològics actuals es basen en l'anomenat principi cosmològic, que diu que l'univers és isòtrop i homogeni. Atès que s'observen estructures a petita escala és clar que aquest principi no és estrictament vàlid. Tot i això el principi cosmològic és una bona aproximació quan estudiem l'univers a gran escala. Així doncs, un enunciat més acurat per al principi cosmològic és que l'univers esdevé isòtrop i homogeni en el límit de grans escales.

Les simetries presents en el principi cosmològic van permetre a Alexander Friedmann, a Georges Lemaître, a Howard Percy Robertson i a Arthur Geoffrey Walker desenvolupar de forma independent una solució general per a la cosmologia en el marc de la relativitat general (Friedmann 1922, 1924; Lemaître 1927; Robertson 1935, 1936a,b; Walker 1937). Aquesta solució, implica que l'evolució de l'univers en el temps serà o bé una expansió homogènia o bé una contracció, també homogènia, i sovint se l'anomena mètrica de Friedmann–Lemaître–Robertson–Walker (FLRW) o amb altres combinacions d'aquests noms. El desenvolupament en detall d'aquesta solució es troba en molts llibres de text, com ara al capítol 3 de Peter & Uzan (2009).

L'any 1929, l'astrònom Edwin Hubble va publicar un estudi comparant les velocitats i les distàncies de les 46 galàxies de les quals s'havia pogut mesurar tant la velocitat com la distància (Hubble 1929). En aquest estudi Hubble va observar que totes les galàxies s'allunyaven de nosaltres (excepte un petit grapat que es troben molt a prop i estan gravitacionalment lligades al pou de potencial local) i que, a més, aquestes ho feien més ràpidament com més lluny es trobessin:

$$v = Hd\,. \tag{1.1}$$

Aquesta relació és la coneguda llei de Hubble que ens indica que la velocitat a la qual s'allunya una galàxia és proporcional a la distància a la qual es troba i és una conseqüència natural del principi cosmològic.

La llei de Hubble, a més, ens permet definir una forma alternativa de mesurar les distàncies a objectes astronòmics que serà emprada constantment en aquesta tesi: el desplaçament cap al roig, $z$. El desplaçament cap al roig es correspon al desplaçament Doppler de la llum provinent dels diferents objectes astronòmics pel simple fet que s'estan allunyant de nosaltres. Atès que les galàxies i altres objectes astronòmics s'estan allunyant de nosaltres, el desplaçament Doppler es produeix sempre cap a longituds d'ona més llargues. En l'espectre òptic, el vermell és el color que presenta una longitud d'ona major i sovint es fa referència a longituds d'ona més vermelles quan es vol parlar de longitud d'ona majors. D'aquí el nom desplaçament cap al *roig*. Aquest concepte ens permet mesurar distàncies gràcies a la llei de Hubble. A més a més, atesa la velocitat constant i finita de la llum, el desplaçament cap al roig també pot ser interpretat com una mesura del temps.

En les dècades que van seguir el descobriment de la llei de Hubble, hi va haver dos models d'un univers en expansió alternatius. D'una banda hi havia el model del Gran Esclat o *Big Bang*, que va ser proposat per Lemaître (Lemaitre 1931) i desenvolupat per George Gamow entre d'altres (Gamow 1948a,b). Aquest model proposa que l'univers inicialment es trobava en un estat molt dens i molt calent i que s'ha anat refredant a mida que s'ha anat expandint. L'origen de l'expansió vindria donat per una *explosió* primordial (i d'aquí el nom del model: en anglès l'onomatopeia per a una explosió és *bang*). L'alternativa a aquest model va ser proposada per Fred Hoyle, Thomas Gold i Hermann Bondi entre d'altres (Bondi & Gold 1948; Hoyle 1948) i afirma que l'univers s'ha mantingut en un estat on la temperatura i la densitat són constants en el temps. Per compensar els efectes de l'expansió de l'univers cal que es vagi creant matèria de forma continuada a mida que l'univers es va expandint, de manera que les seves propietats es poden mantenir constants. Aquest model es va anomenar *Steady State* (estat constant, en anglès).

Gràcies a la seva capacitat predictiva, la teoria del *Big Bang* va anar guanyant acceptació a mida que es confirmaven algunes de les seves prediccions. Una d'aquestes prediccions fou la de les abundàncies dels elements lleugers que van ser creats durant la nucleosíntesi primordial (BBN, de l'anglès *Big Bang Nucleosynthesis*). Per a entendre la BBN primer ens cal repassar l'evolució de l'univers primigeni. D'acord amb les prediccions del model del *Big Bang*, la combinació d'una alta densitat i d'una alta temperatura a l'univers primigeni fa que els ritmes de col·lisió entre les partícules i els fotons siguin prou elevats per a garantir l'equilibri cinètic i tèrmic (amb l'excepció de partícules que només interactuen gravitacionalment, o amb altres interaccions extremament febles). A mida que l'univers es va expandint, aquests ritmes de col·lisió van disminuint fins que no són prou elevats com per a mantenir l'estat d'equilibri. Depenent de la magnitud de les interaccions, cada partícula pot abandonar l'equilibri a temps diferents. Aquest procés, que s'anomena desacoblament, no és un procés abrupte, sinó que passa gradualment a mida que les interaccions esdevenen menys freqüents.

En els primers minuts de l'univers, aquest era prou calent i prou dens com perquè les reaccions nuclears fossin eficients, de manera que les reaccions entre els protons, $p$, i els neutrons, $n$, (que formen deuteró, $D$, i fotons, $\gamma$) estaven en equilibri: $p + n \leftrightarrow D + \gamma$. Això implica que el deuteró format en aquestes reaccions era fotodissociat abans de poder reaccionar per a formar nuclis més complexos. Quan la temperatura de l'univers va baixar per sota de l'energia de lligam del deuteró, l'abundància d'aquests nuclis era prou elevada com per a reaccionar i formar nuclis més complexos, iniciant així la BBN. Durant la BBN els protons i neutrons es van combinar per a formar $D$, $^3H$, $^3He$ i $^4He$. D'elements més massius quasi no se'n van formar doncs l'$^4He$ és el nucli lleuger més lligat i no n'hi ha cap d'estable que tingui 5 nucleons. Tot i això, es van formar petites quantitats de $^7Li$ i $^7Be$ abans l'univers no es refredés fins al punt en què les reaccions nuclears ja no eren possibles. La BBN va durar només uns minuts de manera que no hi va haver prou temps com per a formar quantitats significatives de C i d'altres elements més pesats.

Així doncs, les observacions de les abundàncies primordials estan limitades bàsicament a mesurar la quantitat de deuteró present en zones que no hagin estat en contacte amb ejeccions estel·lars. L'$^3H$ decau a $^3He$ i l'abundància de l'heli es veu àmpliament afectada per l'evolució estel·lar: una fracció elevada de l'hidrogen es transforma en heli al llarg de la vida de les estrelles i es barreja amb l'heli primordial després que les primeres supernoves (SN) explotin. Una explicació més detallada de la BBN es pot trobar a Steigman (2007).

Encara més important fou el descobriment del fons còsmic de microones (CMB, de l'anglès *Cosmic Microwave Background*). El CMB és un fons de radiació homogeni que aporta informació sobre les primeres fases de la història de l'univers, quan la matèria i la radiació van desacoblar-se. Fou introduït per primer cop l'any 1948 per Ralph Alpher i Bob Herman (Alpher & Herman 1948) i la seva existència va ser confirmada per les observacions de Penzias i Wilson, que van trobar una font de soroll desconeguda a les freqüències de ràdio, idèntica en totes direccions (Penzias & Wilson 1965). Aquesta font de soroll va ser ràpidament interpretada per Dicke i col·laboradors com el CMB (Dicke et al. 1965). Amb aquesta confirmació, el model del *Big Bang* va esdevenir el model estàndard de cosmologia.

El CMB es va originar quan els fotons de l'univers primigeni, que estaven en equilibri tèrmic amb la matèria, es van desacoblar de la matèria. Després del desacoblament els fotons mantenen l'espectre d'un cos negre, doncs la densitat a l'espai de fases es conserva en l'expansió de l'univers. La temperatura del CMB va baixant a mida que l'univers es va expandint i els fotons es veuen desplaçats cap al roig, cap a la banda de freqüències de les microones, d'on li prové el nom. El CMB és el cos negre més perfecte que es coneix: les desviacions respecte del cos negre ideal són menors que 50 parts d'un milió (Fixsen et al. 1996). Per ampliar la informació sobre aquest punt es refereix al lector a, per exemple, Samtleben et al. (2007); Hu & Dodelson (2002).

Resumint, hi ha algunes conclusions de la teoria del *Big Bang* que són completament genèriques, independentment del valor que prenguin els diferents paràmetres del model:

- L'estat inicial de l'univers és un estat molt dens i molt calent, on la matèria està ionitzada i es troba en equilibri termodinàmic.

- Hi ha diverses fases en la història de l'univers que es diferencien pel component dominant de l'energia de l'univers. Inicialment hi hagué una era dominada per la radiació, seguida d'una altra dominada per la matèria.
- A mida que s'expandeix, l'univers es va refredant. Presenta, doncs, una història tèrmica. El refredament de l'univers fa que alguns dels processos físics que mantenen acoblades diferents espècies de matèria i radiació esdevinguin menys efectius, provocant un desacoblament de les mateixes. El desacoblament de les espècies es dóna a temps diferents en funció de la magnitud de les interaccions.

## 1.2 La matèria fosca i la formació d'estructura

Un dels grans punts forts de la teoria del *Big Bang* és la seva adaptabilitat: la teoria del *Big Bang* suposa que hi ha matèria i radiació, sense especificar ni de quin tipus de matèria es tracta ni que les distribucions tant de radació com de matèria siguin homogènies; admetent, fins i tot, la presència d'una *constant cosmològica*. Al llarg d'aquest apartat veurem que aquesta adaptabilitat va permetre explicar els problemes observacionals que van anar sorgint. Cal tenir en compte que en el moment en què es va pensar la teoria del *Big Bang*, només es tenia constància de l'existència de l'anomenada matèria bariònica que, segons les darreres observacions del satèl·lit Planck, constitueix només un $\sim 4.9\%$ del total de l'energia de l'univers (Planck Collaboration et al. 2015). La resta de l'univers estaria constituït per l'anomenat sector fosc.

Una petita part d'aquest sector fosc el constitueix la matèria fosca, que representa el $\sim 26\%$ del total de l'energia de l'univers (Planck Collaboration et al. 2015). Les primeres evidències indirectes de l'existència d'un component de la matèria que no interacciona amb la radiació electromagnètica (i d'aquí el nom de matèria fosca) van ser descobertes per Zwicky (1933). En un estudi dinàmic de l'agrupació de galàxies de Coma va concloure que les velocitats mesurades no es podien explicar per l'atracció gravitatòria de la massa observada, aquesta es quedava curta: només és capaç d'explicar entre un 30 i un 50% de l'atracció gravitatòria necessària per explicar la dispersió de velocitats mesurada. Aquestes primeres troballes van ser reforçades per Ostriker et al. (1974), Einasto et al. (1974) i Rubin (1983), que van mesurar que la quantitat de matèria en les galàxies espirals augmentava de forma lineal amb la distància al centre fins a com a mínim 100 kpc.

A mida que les evidències indirectes de l'existència de matèria fosca s'anaven apilant aquesta es va incorporar a la teoria del *Big Bang*. A més, les observacions imposaven que aquesta matèria fosca havia de ser més aviat freda, és a dir, que la seva dispersió de velocitats fos suficientment petita com per a ser negligida per a tots els efectes pràctics observables. Per això el nou paradigma es va anomenar model dematèria fosca freda (CDM, de l'anglès *Cold Dark Matter*), una explicació més detallada del qual es pot trobar a Faber & Gallagher (1979) o Blumenthal et al. (1984). Tot i que altres teories que no necessitaven de l'existència de la matèria fosca van aflorar contemporàniament a la CDM (veure, per exemple Sanders & McGaugh 2002), la capacitat predictiva de la teoria CDM va fer que aquesta fos acceptada per la majoria de científics. Per exemple, la teoria CDM estableix una explicació natural per a les anisotropies presents a l'estructura a gran escala i al CMB.

Segons la teoria CDM, l'estructura a gran escala sorgeix de petites fluctuacions en la densitat a l'univers primigeni que creixen fins a convertir-se en l'estructura a gran escala que mesurem. El fet que les pertorbacions en la densitat creixin s'entén de forma molt intuïtiva: les zones amb una densitat lleugerament superior produeixen una atracció gravitatòria més elevada; això fa que la matèria del voltant tingui tendència a caure dins la pertorbació, fent-la créixer. Si tenim en compte que en els instants inicials d'alta densitat i temperatura la matèria estava acoblada a la radiació, veurem que el mateix mecanisme ens permet explicar les anisotropies del CMB.

Un cop afegides la matèria fosca i les fluctuacions de densitat a la teoria del *Big Bang* original es van poden explicar els nous resultats observacionals. Tanmateix, la teoria del *Big Bang* no pot explicar l'origen d'unes condicions inicials tan especials: l'univers és pràcticament homogeni llevat d'unes fluctuacions de $\delta\rho/\bar{\rho} = 10^{-5}$, on $\delta\rho$ és l'amplitud de les fluctuacions i $\bar{\rho}$ la densitat mitjana de l'univers. A més, aquestes condicions inicials porten a l'anomenat *problema de l'horitzó*: com és possible que regions de l'univers que no estan connectades causalment presentin les mateixes propietats físiques i la mateixa temperatura?

Una solució per al problema de l'horitzó va venir donada com a una extensió de la teoria del *Big Bang* de la mà de Guth (1981) anomenada inflació. A més, la inflació permet derivar les amplituds de les fluctuacions de densitat de forma natural. Aquest model proposa que l'univers es va expandir exponencialment en els primers instants de la seva història, tot disminuint l'horitzó a escala comòbil. Segons aquest escenari les fluctuacions primordials no són altra cosa que fluctuacions quàntiques de l'univers primigeni que van créixer durant el període d'inflació. Aquesta teoria va ser àmpliament acceptada després que les dades preses pel satèl·lit COBE (Smoot et al. 1992), i més endavant per la missió WMAP (Hinshaw et al. 2003), van premetre la mesura de l'espectre de potències primordial. Per a una lectura complementària, llegir els compendis de Narlikar & Padmanabhan (1991) o, per a una argumentació més teòrica, de Brandenberger (1985).

## 1.3 L'expansió accelerada de l'univers: el model ΛCDM

Una de les prediccions principals de qualsevol model cosmològic és la història de l'expansió de l'univers, que es mesura fent servir les anomenades candeles estàndard. Les candeles estàndard són objectes astronòmics dels quals en coneixem molt acuradament la seva lluminositat, fet que ens permet emprar-los per tal de mesurar les distàncies a aquests objectes amb una alta precisió.

Una d'aquestes candeles estàndard són les SN Ia (Phillips 1993). Els grups *SN Search Team* (Riess et al. 1998) i *SN Cosmology Project* (Perlmutter et al. 1999) van determinar de forma independent que l'expansió de l'univers és accelerada, fet que va posar de manifest que encara hi havia un component desconegut que contribuïa de forma significativa a l'energia de l'univers. A més, aquests resultats van ser confirmats per subseqüents observacions de SN (per a més detalls veure Goobar & Leibundgut 2011).

Segons les equacions de la relativitat general, l'expansió accelerada de l'univers no es pot explicar considerant que aquest està format només per matèria i radiació. La manera més simple d'explicar aquesta acceleració és introduir una constant cosmològica $\Lambda$. Aquesta constant cosmològica pot ser pensada com un camp més a l'univers, afegit a la matèria i la radiació, que tindria una pressió negativa i que sovint s'anomena energia fosca. Per a un compendi sobre l'energia fosca i l'expansió accelerada de l'univers veure, per exemple, Frieman et al. (2008).

El nou paradigma, una extensió de la teoria CDM, fou anomenat $\Lambda$CDM, i, tot i que hi ha nombroses alternatives a aquest model, per ara aquest és el model més simple que permet explicar les observacions. Les mesures més precises de la quantitat d'energia fosca (Planck Collaboration et al. 2015) donen un valor del $\sim 69.1\%$ del total d'energia de l'univers (recordem de l'apartat 1.2 a la pàgina 6 que segons Planck Collaboration et al. 2015, la quantitat de matèria bariònica i de matèria fosca correspon al $\sim 5\%$ i al $\sim 26\%$ del total de l'energia de l'univers).

Tot i això, la naturalesa de l'energia fosca ens és encara desconeguda. El que sí que sabem, però, és que juga un paper molt important en l'expansió de l'univers, de manera que afecta indirectament la mesura de la relació entre el desplaçament cap al roig i la distància (el paràmetre de Hubble, $H(z)$). És important, doncs, determinar-ne acuradament l'evolució amb el desplaçament cap al roig.

Tal com hem vist, per a distàncies properes podem mesurar el paràmetre de Hubble fent servir les SN. Tanmateix, com més lluny vulguem observar-les, més febles seran fins que arriba un punt en què la seva detecció no resulta possible. Per a distàncies més llunyanes fem servir una altra candela estàndard: l'escala característica de les oscil·lacions acústiques dels barions (BAO, de l'anglès *Baryon Acoustic Oscillations*).

Per entendre quin és l'origen de l'escala característica de les BAO hem de retrocedir en el temps i entendre com es comportaven els barions en l'univers primitiu. L'univers primigeni era molt calent i molt dens i la radiació estava acoblada amb la matèria bariònica de manera que la seva interacció prevenia la formació dels primers àtoms. La pressió de la radiació produïa ones acústiques en el plasma constituït per la barreja de barions i fotons. Un cop l'univers es va refredat prou perquè els fotons es desacoblessin dels barions, ambdues espècies es van trobar amb una marca produïda per aquestes ones acústiques. Aquesta marca té una escala característica, l'escala de les BAO, determinada per la mida de l'horitzó a l'època de recombinació. Una explicació més detallada al respecte es pot trobar a Eisenstein (2005).

## 1.4 Els traçadors de la distribució subjacent de matèria

En els apartats anteriors hem vist l'evolució del model estàndard de cosmologia per adaptar-se a les noves observacions. També hem vist que la formació d'estructures està íntimament lligada amb la matèria fosca i amb l'energia fosca. Ara bé, no podem observar directament cap de les dues. El que podem mesurar és la matèria bariònica a través de la seva interacció amb els fotons. Per comprendre com podem obtenir informació de la matèria fosca hem d'entendre que els barions tracen la distribució subjacent de matèria fosca i, per això, ens cal desenvolupar la teoria amb una mica més de detall. Aquest és l'objecte d'aquest apartat.

Un dels pilars dels models cosmològics actuals és el principi cosmològic, que diu que l'univers és isotròpic i homogeni. Hem vist que aquest principi l'hem d'entendre com una aproximació: l'univers no és exactament isòtrop i homogeni sinó que, ja des d'un bon principi, hi ha unes fluctuacions primordials de densitat que evolucionaran fins a formar les estructures que observem a dia d'avui.

Assumint que la matèria no presenta cap mena de pressió, fet que és sempre vàlid per a la matèria fosca, les petites pertorbacions de densitat creixen adiabàticament amb el temps fins a formar les galàxies i altres objectes. A un instant $t$, les pertorbacions es poden descriure mitjançant un contrast de densitats, $\delta\left(\vec{x}\right)$:

$$\delta\left(\vec{x}\right) \equiv \frac{\rho\left(\vec{x}\right) - \bar{\rho}}{\bar{\rho}} \,, \tag{1.2}$$

on $\rho\left(\vec{x}\right)$ és la densitat en un temps $t$ i en un punt $\vec{x}$ en particular i $\bar{\rho}$ és la densitat mitjana de l'univers en aquell instant de temps. Per simplicitat en la notació, aquí hem optat per no expressar la dependència temporal de forma explícita. Més detalls sobre l'evolució d'aquestes pertorbacions a, per exemple, Padmanabhan (1993) o Byrd et al. (2007).

Un cop definit el contrast de densitats podem mesurar l'amplitud de les pertorbacions mitjançant la mesura de la funció de correlació. L'amplitud de les pertorbacions a una escala $r$ ve donada per

$$\xi\left(\vec{r}\right) \equiv \left\langle \delta\left(\vec{x}\right) \delta\left(\vec{x} + \vec{r}\right) \right\rangle \,, \tag{1.3}$$

on $\langle \cdots \rangle$ indica que avaluem la mitjana sobre totes les realitzacions possibles de la fluctuació $\delta$ (la qual cosa és, segons el teorema ergòdic, equivalent a calcular la mitjana sobre un volum molt gran de l'espai).

A més de la funció de correlació, sovint es fa servir la seva transformada de Fourier: l'espectre de potències, que ve donat per

$$P\left(\vec{k}\right) = \int d^3k \, \xi\left(\vec{r}\right) e^{-i\vec{k}\vec{r}} \,. \tag{1.4}$$

A la vegada, la funció de correlació és simplement la transformada de Fourier inversa de l'espectre de potències:

$$\xi\left(\vec{r}\right) = \frac{1}{(2\pi)^3} \int d^3r \, P\left(\vec{k}\right) e^{i\vec{k}\vec{r}} \,. \tag{1.5}$$

Una de les raons que fa útil treballar en l'espai de Fourier és que la convolució de funcions pren una expressió molt més simple en l'espai de Fourier. En general, els models teòrics prediuen l'espectre de potències, ja que cada mode de Fourier evoluciona de forma independent en el règim lineal de les fluctuacions, i la funció de correlació s'ha de deduir a partir de l'espectre de potències predit fent la transformada de Fourier. Si les fluctuacions de densitat segueixen una distribució gaussiana, l'espectre de potències proporciona una descripció estadística de les fluctuacions completa.

Un cop les fluctuacions superen la massa de Jeans, aquestes es tornen inestables i col·lapsen tot formant halos de matèria fosca. El temps de col·lapse d'aquests halos depèn de la seva massa, sent més petit com més petit és l'halo, cosa que vol dir que les estructures més petites es formen abans. A més, els halos d'una certa massa tracen la distribució subjacent de matèria de forma esbiaixada. Concretament, la seva funció de correlació serà proporcional a la funció de correlació de la distribució subjacent de massa amb una constant de proporcionalitat que és el quadrat del biaix d'aquests halos Mo & White (1996). Resultats tant de càlculs analítics (per exemple Sheth & Tormen 1999) com de simulacions (per exemple Tinker et al. 2010) indiquen que el biaix dels halos augmenta amb la seva massa.

La matèria bariònica es comporta de manera lleugerament diferent. D'entrada, la matèria bariònica segueix acoblada amb la radiació durant més temps perquè pot interaccionar electromagnèticament. Això vol dir que la formació d'estructures de la matèria bariònica s'endarrereix respecte de la matèria fosca. A més, la interacció electromagnètica permet als barions refredar-se, emetent l'energia sobrant en forma de fotons. Un cop desacoblats de la radiació, els barions es veuen atrets pels pous de potencial dels halos de matèria fosca. A mida que es van refredant, els barions tendeixen a situar-se al centre dels halos de matèria fosca i formar galàxies. Tot i això, la majoria dels barions es troben en el medi intergalàctic. Això vol dir que, a gran escala i de forma aproximada, la funció de correlació de qualsevol població de galàxies serà proporcional a la funció de correlació de la distribució subjacent de matèria amb una constant de proporcionalitat que serà igual al quadrat del biaix dels halos que acullin aquesta població de galàxies.

És interessant mesurar la correlació de diferents traçadors, ja que ens aporten informació independent que permet constrènyer els models cosmològics. Els diferents traçadors sovint es troben en halos de diferent massa i a diferent desplaçament cap al roig, de manera que permeten determinar l'evolució del paràmetre de Hubble, $H(z)$. De fet, algunes de les millors mesures del paràmetre de Hubble provenen de la posició del pic de les BAO a l'autocorrelació de galàxies (Gil-Marín et al. 2015; Cuesta et al. 2016) o del bosc de Ly$\alpha$ (Delubac et al. 2015).

A més de mesurar l'autocorrelació de diferents traçadors, també és possible mesurar la correlació encreuada entre parelles de traçadors. De la mateixa manera que en el cas de l'autocorrelació, la correlació encreuada de dos traçadors és proporcional a la funció de correlació de la distribució subjacent de matèria. La constant de proporcionalitat és, en aquest cas, el producte dels biaixos de les dues poblacions. La mesura de la correlació encreuada permet trencar algunes degeneracions entre diferents paràmetres dels models (Aubourg et al. 2015).

Resumint, els models actuals indiquen que la matèria bariònica traça la distribució subjacent de matèria fosca. La correlació encreuada i l'autocorrelació són proporcionals a l'autocorrelació de la matèria i la constant de proporcionalitat depèn dels biaixos dels traçadors involucrats. La mesura d'aquestes correlacions proporciona límits en els paràmetres dels models que són competitius comparat amb altres mètodes de mesura. En particular, el diferent desplaçament cap al roig dels diversos traçadors permet mesurar l'evolució del paràmetre de Hubble amb molta precisió. Als apartats 1.4.1 a 1.4.4 a la pàgina següent i a la pàgina 12) descrivim els traçadors emprats en aquesta tesi.

### 1.4.1 Quàsars i galàxies

Després del desacoblament de la matèria bariònica i la radiació a l'època de recombinació no hi havia fonts lluminoses a l'univers, que havia entrat a les Edats Fosques. Durant les Edats Fosques les pertorbacions van créixer fins que es van formar els primers halos de matèria fosca. La matèria bariònica va condensar al centre d'aquests halos fins que va ser prou densa i prou calenta com per a començar reaccions termonuclears, formant així les primeres estrelles i donant per acabades les Edats Fosques. El procés de formació d'aquestes primeres estrelles encara no està del tot clar: són molt diferents de la resta d'estrelles a causa de la seva manca de metalls; només els elements més lleugers es formaren durant la nucleosíntesi primordial. Aquí no entrarem en detall sobre les diferents teories respecte de la formació d'aquestes primeres estrelles, referim al lector a Bromm & Larson (2004) i altres referències allà contingudes.

Després de les primeres SN, la radiació emesa per aquestes primeres estrelles i pels discs d'acreció dels forats negres va escalfar i ionitzar el gas del seu voltant, frenant així el ritme de creació de noves estrelles. Tot i que hi ha diversos models per a la formació de les galàxies, la idea principal és que el gas va col·lapsar al centre dels halos de matèria fosca, dissipant l'energia a través del refredament radiatiu, i va formar finalment estrelles en una estructura gravitacionalment lligada que anomenem galàxia. Més detalls sobre la formació i l'evolució de les primeres galàxies i la seva estructura es poden trobar a Bromm & Yoshida (2011); Silk & Mamon (2012); Conselice (2014).

Tot i que la formació i evolució de les galàxies no és l'interès central d'aquesta tesi, hi ha una fase d'aquesta evolució que mereix una menció especial. Aquesta fase es coneix com a galàxia de nucli actiu (AGN, de l'anglès *Active Galactic Nuclei*) i està descrita en detall a Antonucci (1993). Tal com el nom indica, les galàxies en aquesta fase presenten un nivell d'activitat molt elevat situat a les regions centrals, consistent en l'acreció d'enormes quantitats de gas pel forat negre supermassiu situat al centre d'aquestes galàxies. Aquesta acreció comporta l'emissió de poderosos feixos de material, formats de partícules energètiques (no se sap si són principalment de material bariònic accelerat o de parelles electró-positró), que pot assolir fàcilment velocitats relativistes degut a processos altament complexos. Aquesta fase de l'evolució es pot produir diverses vegades, fins que el nucli de la galàxia es queda pràcticament sense gas.

Quan una galàxia primigènia entra en la fase d'AGN, encara té una quantitat de gas molt elevada, fet que fa que hi hagi molt més material disponible per a ser acretat. A la pràctica això implica que els feixos que emetrà aquesta galàxia són tan potents que, si estan orientats cap a nosaltres, els podem observar fàcilment amb telescopis no gaire grans. Aquests objectes van ser descoberts l'any 1963 (Hazard et al. 1963; Schmidt 1963) i van ser anomenats quàsars. Al ser tan brillants, l'espectre d'aquests quàsars permet detectar noves poblacions de traçadors que discutirem als apartats següents.

### 1.4.2 El bosc de Ly$\alpha$

L'any 1965, Maarten Schmidt va prendre espectres del quàsar 3C9 i va mesurar-ne un desplaçament cap al roig de $z = 2.01$ (Schmidt 1965). Va ser la primera vegada que es va trobar un quàsar a un desplaçament cap al roig més gran de 2, però posteriorment se n'han descobert molts més. Els catàlegs actuals en contenen centenars de milers (veure per exemple el catàleg del projecte Baryon Oscillations Spectroscopic Survey (BOSS) DR12Q Pâris & et al. 2016). Tots aquests quàsars tenen una cosa en comú: l'emissió de la línia de Ly$\alpha$ s'ha desplaçat fins a la part visible de l'espectre electromagnètic.

A partir de l'espectre d'aquests quàsars, a unes energies més altes que la línia d'emissió de Ly$\alpha$, podem observar l'absorció produïda per hidrogen neutre al medi intergalàctic (IGM, de l'anglès *intergalactic medium*) (Gunn & Peterson 1965). Les absorcions de Gunn-Peterson, que és com s'anomena aquest efecte, consisteixen en el següent: els fotons emesos a una energia més alta que la de la transició de Ly$\alpha$ poc a poc es van desplaçant cap al roig fins que tenen exactament l'energia necessària per excitar els àtoms d'hidrogen. Si quan això passa només una petita part del medi intergalàctic està ionitzat, llavors l'absorció de Ly$\alpha$ produeix una baixada important de la quantitat de flux transmesa de fotons amb precisament aquesta energia.

Així doncs, normalment es detecten múltiples absorcions a freqüències més blaves que la de la línia de Ly$\alpha$. Al conjunt d'aquestes absorcions se'ls anomena el bosc de Ly$\alpha$; va ser proposat per Bahcall & Salpeter (1965) i més endavant confirmat per Lynds (1971). Aquesta absorció es correspon amb la presència de núvols d'hidrogen a la línia de visual de la qual s'obté l'espectre. La densitat de columna del gas es pot estimar a partir de l'absorció mesurada, si bé el seu valor és força imprecís. Més informació sobre el bosc de Ly$\alpha$ es pot trobar a Rauch (1998).

### 1.4.3 Sistemes Ly$\alpha$ esmorteïts

Dins el bosc de Ly$\alpha$ hi ha alguns absorbidors que mereixen una atenció especial: els sistemes Ly$\alpha$ esmorteïts (DLAs, de l'anglès *Damped Ly$\alpha$ Absorbers*). L'absorció produïda pels DLAs mostra un perfil característic observable fins i tot en els espectres de baixa resolució que permet determinar la densitat de columna amb molta més precisió que no pas els absorbidors habituals. Per als DLAs, les densitats de columna mesurades són $N_{\mathrm{HI}} \geqslant 10^{20.3}$ cm$^{-2}$ (Wolfe et al. 1986).

Els DLAs estan constituïts per núvols densos d'hidrogen que són prou densos per poder-se protegir de la radiació còsmica ionitzant, un efecte que s'anomena autoapantallament. Només les capes més exteriors d'aquets núvols mantenen l'hidrogen ionitzat, la resta del núvol està compost principalment per hidrogen atòmic Vladilo et al. (2001). Veure Wolfe et al. (2005) per a una descripció més detallada.

### 1.4.4 Sistemes absorbidors metàl·lics

En el camp de l'astronomia sovint es parla dels diferents elements emprant un argot característic que pot portar a confusions. Concretament, el mot *metalls* fa referència de manera genèrica a qualsevol element llevat de l'hidrogen i l'heli. Així, pels astrònoms l'oxigen és tan metàl·lic com el ferro. Aquesta notació probablement té origen en el fet que l'abundància relativa de tots els elements junts excepte l'hidrogen i l'heli no arriba al 2% i no té res a veure amb el tipus d'enllaç que els diferents elements fan servir per a formar mol·lècules. En aquesta tesi farem ús d'aquesta mateixa nomenclatura, entenent per metall qualsevol element que no sigui ni l'hidrogen ni l'heli.

De forma semblant al bosc de Ly$\alpha$, l'absorció per altres línies metàl·liques pot ser detectada als espectres. Tanmateix, aquesta absorció és típicament més feble i molt menys habitual, fet que es deu a la menor quantitat de metalls present a l'univers. Els metalls es formen principalment a les explosions de SN en les fases finals de l'evolució estel·lar, on només una petita part de la massa de l'estrella (principalment hidrogen i heli) acaba fusionant-se fins a formar els metalls més pesats.

Tot i que actualment hi ha indicis de boscos d'altres metalls (Pieri 2014), en la seva majoria l'absorció dels metalls es detecta de forma individual en absorbidors aïllats. El principal problema que presenta la detecció d'aquests absorbidors rau en la identificació del metall responsable de l'absorció. Un dels absorbidors més estudiats és el corresponent a la línia d'absorció de Mg II, que presenta un doblet característic i és fàcil d'identificar. A més, aquesta línia té una alta freqüència d'oscil·lació i pot ser observada a baix desplaçament cap al roig emprant telescopis òptics. Això fa que els absorbidors de Mg II siguin un dels millors traçadors metàl·lics de la distribució subjacent de matèria, motiu pel qual l'hem triat com un dels traçadors a estudiar en aquesta tesi.

## 1.5 Descripció de la tesi

El treball presentat en aquesta tesi es basa principalment en la mesura i la interpretació de la correlació encreuada entre diferents traçadors. Per a fer-ho s'han fet servir les dades espectroscòpiques de BOSS, un dels principals cartografiatges realitzats per la col·laboració Sloan Digital Sky Survey III (SDSS-III) (Eisenstein et al. 2011), de la qual en formo part.

Al llarg de la tesi repassaré la correlació encreuada de diferents traçadors, tots ells introduïts als apartats 1.4.1 a 1.4.4 a la pàgina 11 i a la pàgina 12. La primera correlació encreuada que estudio, presentada al capítol 3 a la pàgina 29, és la correlació encreuada entre sistemes absorbidors de Mg II i les galàxies CMASS. En aquest capítol presento un nou mètode per la mesura de l'amplada d'aquests absorbidors desenvolupat amb la col·laboració de Britt Lundgren i de Jordi Miralda Escudé que aplico al doblet de Mg II. Aquest capítol és bàsicament una reproducció de Pérez-Ràfols et al. (2015).

Més endavant, al capítol 4 a la pàgina 67, em centro en els DLAs i el bosc de Ly$\alpha$. Els paràmetres del bosc de Ly$\alpha$ han estat estudiats anteriorment (Delubac et al. 2015; Blomqvist et al. 2015) de manera que són coneguts. Emprant els valors obtinguts en aquests estudis és possible obtenir mesures acurades dels paràmetres dels DLAs a partir de la seva correlació encreuada amb el bosc de Ly$\alpha$. El treball presentat en aquest capítol s'ha fet en col·laboració amb Andreu Font Ribera, Nicolás Busca i Jordi Miralda Escudé. Concretament, l'estimador de la correlació encreuada va ser desenvolupat per Nicolás Busca per a l'estudi de l'autocorrelació i més endavant es va adaptar a la mesura de la correlació encreuada. Per als ajustos del model s'ha emprat el codi *baofit*, un ajustador desenvolupat principalment per David Kirkby i que actualment està mantingut principalment per Michael Blomqvist(Kirkby et al. 2013).

En els capítols 3 i 4 a la pàgina 29 i a la pàgina 67 la mesura de la correlació encreuada i, més important, la seva interpretació estan centrades en els efectes a gran escala, on podem emprar la teoria lineal. A escales més petites entren en joc efectes no lineals i l'aproximació lineal descrita en aquesta introducció deixa de ser vàlida. Al capítol 5 a la pàgina 99 mesuro la correlació encreuada dels quàsars amb el bosc de Ly$\alpha$, centrant-me en els efectes a petita escala. Es poden observar efectes de la contaminació d'absorbidors metàl·lics en el bosc de Ly$\alpha$. A més, discuteixo alguns dels possibles efectes no lineals i proporciono un model teòric simple que explica l'efecte que la radiació dels quàsars té en els núvols d'hidrogen més propers i com aquest efecte es pot observar en la mesura de la correlació encreuada. Aquest estudi s'ha realitzat en col·laboració amb Jordi Miralda Escudé.

Per acabar, en el capítol 6 a la pàgina 119, presento les conclusions globals d'aquesta tesi i comento breument alguns projectes de futur.

# 2

# Introduction

## 2.1 The homogeneous universe

The basis of modern cosmology is the so-called cosmological principle. This principle states that the Universe is isotropic and homogeneous. This principle is not strictly valid, we observe structure at small scales. However, the cosmological principle is a good approximation at large scales. Thus, a better formulation of this principle would be that the Universe becomes isotropic and homogeneous in the limit of large scales.

The symmetries implied in the cosmological principle allowed Alexander Friedmann, Georges Lemaître, Howard Percy Robertson, and Arthur Geoffrey Walker to independently derive a general cosmological solution in the framework of general relativity (GR) (Friedmann 1922, 1924; Lemaître 1927; Robertson 1935, 1936a,b; Walker 1937). This solution, often called Friedmann–Lemaître–Robertson–Walker (FLRW) metric or Friedmann–Robertson–Walker (FRW) metric, implies that the Universe would experience a homogeneous contraction or expansion. For a detailed derivation and further reading, the reader is referred to, e.g., Chapter 3 of Peter & Uzan (2009).

In 1929, Edwin Hubble published a study comparing the velocity and the distance of 46 galaxies (Hubble 1929). Hubble showed that all galaxies were moving away from us (except a few that are vrey close to us and are falling to the center of the local potential well), and that the further away a galaxy was, the greater its velocity of recession:

$$v = \mathrm{H} \cdot \mathrm{d} \,. \tag{2.1}$$

This relation, nowadays known as Hubble's law, is a consequence of the cosmological principle.

Hubble's law allow us to define an alternative way of measuring distances that will be frequently used in this thesis: the redshift, $z$. This quantity corresponds to the Doppler shift observed in an astronomical object due to its recession velocity. Since galaxies are recessing from us in an expanding universe, this shift will always occur to the red, hence *red*shift. As derived from Hubble's law, this quantity is related to the distance. What is more, it can also be related to a measure of time due to the constant and finite speed of light.

During the following decades, two alternative models of an expanding Universe were discussed. On the one hand, there was the Big Bang model. This model, proposed by Lemaître (Lemaitre 1931) and then developed by George Gamov, amongst others (Gamow 1948a,b), proposes that the Universe was once in a very dense, hot state and that it has been expanding ever since as a result of a primordial *explosion* (hence the name). On the other hand, there was the Steady State model, developed by Fred Hoyle, Thomas Gold, and Hermann Bondi among others (Bondi & Gold 1948; Hoyle 1948), which hypothesises that the universe is in a steady state, meaning that its density and temperature are constant in time. To compensate for the effects of the expansion, this model requires matter to be continuously created in the Universe, allowing its properties to remain constant.

One of the key predictions that led to the acceptance of the Big Bang model in favour of the Steady State model is that it naturally explains the abundances of the lightest elements, created during Big Bang Nucleosynthesis (BBN). However, to understand BBN, first we need to review the general evolution of the Universe in its early stages. According to the predictions of the Big Bang model, the combination of high density and high temperature in the early-Universe ensured that the collision rates of all particles and photons were frequent enough to guarantee their in kinetic and thermal equilibrium (the possible exception are those particles that interact only gravitationally). As the Universe expands and cools, the interaction rates drop until they are not high enough to maintain the state of equilibrium. Depending on the strength of their interactions, different particles may depart from equilibrium at different times. This departure from equilibrium, commonly referred to as decoupling, is not abrupt, but it occurs gradually over some time as the interactions become less and less common.

In the first minutes of the Universe, the temperature and density were sufficiently high to enable efficient nuclear reactions. At this point nuclear reactions between protons, $p$, and neutrons, $n$, (which generate deuteron, $D$, and photons, $\gamma$) were still in equilibrium: $p + n \leftrightarrow D + \gamma$. This implies that any deuteron that is produced will be photodissociated by photons before more complex nuclei can be formed. As the temperature drops below the binding energy of the deuteron, the abundance of deuteron increased enough for it to react and form more complex nuclei. It is the starting point of the BBN. During the BBN, protons and neutrons quickly combined to form $D$, $^3H$, $^3He$, and $^4He$. Very few more massive nuclei were formed, the $^4He$ is the most tightly bound of the light nuclei and there is no stable nuclei with 5 nucleons. Even so, some $^7Li$ and $^7Be$ managed to form before the Universe cooled enough to stop the nuclear reactions. BBN only lasted for a few minutes, there is not enough time to form any significant trace of C and other heavier elements.

Observations of the primordial abundances of the relic nuclei are limited mostly to measuring the abundance of deuterium in environments that have not been mixed with stellar ejecta. The $^3$H decays into $^3$He, and the abundance of helium is largely affected by stellar contamination. An important fraction of hydrogen is turned into helium in stars and mixed with the primordial gas after the first supernovae (SNe) start to occur. For an extended review on the BBN see e.g. Steigman (2007).

An even more important prediction is the existence of the Cosmic Microwave Background (CMB). The idea of a CMB was first introduced by Ralph Alpher and Bob Herman (Alpher & Herman 1948). The prediction of the CMB was confirmed by Penzias and Wilsonm who discovered an unexplained noise signal in the radio band, identical in every direction (Penzias & Wilson 1965). Dicke et al. (1965) immediately interpreted this signal as the CMB. This observational discovery validated the Big Bang model. The Steady State model was ruled out shortly after that and the Big Bang model became the standard model of cosmology.

The CMB originates as the photons in the early-Universe that are in thermal equilibrium decouple from matter. Before decoupling, this radiation was the dominant energy field. After decoupling, the photons maintained the thermal spectrum of a black body because the phase-space density of photons is conserved by the expansion of the Universe. The temperature of the CMB goes down as the Universe expands and photons are redshifted to the microwave frequency band that this background radiation is named after. The CMB spectrum is the closest to an ideal black body that has ever been observed. Its deviations from the perfect black body are less than 50 parts per million (Fixsen et al. 1996). For extensive reviews on this subject the reader is referred to, e.g., Samtleben et al. (2007); Hu & Dodelson (2002).

To summarize, some conclusions derived from the Big Bang model are completely generic, independently of the values of any model parameters:

- The initial stage of the Universe is a very hot and dense state, where matter is ionized and in thermodynamical equilibrium.

- Several distinct eras occur in the history of the Universe that are characterized by the dominant contributor to the total energy density of the Universe. Initially there was a radiation-dominated era, followed by a matter-dominated era.

- The Universe has a thermal history. As it expands, the Universe cools down. The temperature decrease and the Universe expansion makes the interactions between particles less effective and, when that happens, those particles decouple from the rest. The decoupling of the different species happens at different times depending on the strength of the interaction.

## 2.2 Dark matter and formation of structure

One of the key points in the Big Bang theory is its adaptability: the theory only assumes that there is matter and radiation. It does not specify neither the type of matter it is dealing with, nor the fact that its distribution, or the radiation's, has to be homogeneous. It even admits the presence of a *cosmological constant*. In this section, we will see that this adaptability of the Big Bang theory allowed it to explain the coming observational problems that arose as the observational data improved. Keep in mind that, when the Big Bang theory was first proposed, there was only evidence of the so-called baryonic matter. Current observations from Planck Collaboration et al. (2015) suggest that this baryonic matter accounts for only $\sim 4.9\%$ of the total energy in the Universe. The rest of the Universe energy would be comprised by what is called dark sector.

Part of this dark sector would be the dark matter, which constitutes $\sim 26\%$ of the total energy of the Universe (Planck Collaboration et al. 2015). The first indirect evidence of the dark matter component came from Zwicky (1933). He performed a dynamical study of the galaxies in the Coma cluster and concluded that the measured velocities could not be explained by the gravitational influence of the visible matter, it fell shy by 30-50%. This early hints were later supported by Ostriker et al. (1974); Einasto et al. (1974); Rubin (1983) who found that matter in galaxies increased linearly with the distance to the center out to at least $100\,\mathrm{kpc}$ for typical spiral galaxies.

As more and more indirect evidence piled up, the dark matter was included in the Big Bang theory. Observations required the dark matter to be cold, i.e. with a velocity dispersion small enough to be neglected in any observable effect. Thus, the new paradigm was named Cold Dark Matter (CDM). For a more detailed description refer to e.g. the reviews Faber & Gallagher (1979); Blumenthal et al. (1984) and references therein. Even thought there were alternative theories to the CDM extension of the Big Bang which didn't include dark matter (see e.g. the review Sanders & McGaugh 2002), the CDM extension of the Big Bang theory was accepted by the majority of the scientific community due to its predictive capability. For example, the anisotropies of the large-scale structure (LSS) and the CMB arise naturally in this model.

In the CDM paradigm, the LSS of the present Universe arises from small density fluctuations in the early Universe. Places with slightly higher density will have slightly higher gravity which will, in turn, attract more matter. Therefore, these fluctuations will grow with time and end up forming the LSS. At very high densities, the radiation was coupled with matter. This means that the same process that created the LSS, was responsible for introducing the anisotropies observed in the CMB.

With the addition of the DM and small density fluctuations, the Big Bang theory was able to accommodate the new observational measurements. However, the Big Bang theory is unable to explain the origin of the initial conditions of the Universe. The fact that the Universe was practically homogeneous but with small fluctuations of $\delta\rho/\bar{\rho} = 10^{-5}$, where $\delta\rho$ is the amplitude of the density fluctuations, and $\bar{\rho}$ the mean value of the density in the Universe, also arise the *horizon problem*: how is it possible that regions which are not causally connected happen to have the same temperature and physical properties?

A solution for this horizon problem was introduced as an extension of the model by Guth (1981). The inflation, as it was called, also provided for a natural explanation for the small density fluctuations. This model proposes that during the very early stages, the Universe underwent an exponential and sudden phase of expansion. Such an expansion would shrink the causal horizon at a comoving scale, thus solving the horizon problem. In this scenario, the primordial fluctuations can be accounted for as quantum fluctuations in the early Universe that grow during the inflationary phase. This inflationary CDM model was extensively accepted by the scientific community when the power spectrum of the primordial fluctuations was first measured by the data taken by the COBE satellite (Smoot et al. 1992) and later confirmed by WMAP (Hinshaw et al. 2003). See e.g. Narlikar & Padmanabhan (1991) for a review on inflation or e.g. Brandenberger (1985) for a more theoretical review.

## 2.3 Accelerated expansion of the universe: $\Lambda$CDM

One of the main predictions of the cosmological models is the expansion history of the Universe. To measure the expansion history of the Universe one needs to make use of standard candles. A standard candle is a population of objects the intrinsic luminosity of which is well understood. This means that we can use them as a means to give accurate estimates of the distances to those objects.

The SNe Ia are one such standard candles (Phillips 1993). These SNe showed that there was still a missing component in the Universe energy pie. The SN Search Team (Riess et al. 1998) and the SN Cosmology Project (Perlmutter et al. 1999) independently detected an accelerated expansion of the Universe. These results were reinforced by subsequent SNe observations (see e.g. Goobar & Leibundgut 2011, for a review).

According to the laws of GR, the accelerated expansion of the Universe cannot be explained by a universe containing only matter and radiations. The simplest way to accommodate the accelerated expansion is to consider a non-zero value for the integral constant in Einstein's equations, the cosmological constant $\Lambda$. This cosmological constant can be thought of as an extra field in the Universe with negative pressure, usually referred to as dark energy (DE). For a review on DE and the acceleratied expansion of the Universe see e.g. Frieman et al. (2008).

This new extension of the CDM paradigm was named $\Lambda$CDM,. While there are plenty of alternatives to $\Lambda$CDM, this model is the simplest that can successfully explain the observations. The most precise measurement of the DE content (from Planck Collaboration et al. 2015) suggests that it accounts for $\sim 69.1\%$ of the total energy in the Universe (recall from section 2.2 on page 18 that according to Planck Collaboration et al. 2015, the baryonic matter and the dark matter account for $\sim 4.9\%$ and $\sim 26\%$ of the total energy bucket).

Nevertheless the nature of DE is still unknown. We know, however, that the primary effect of DE is on the expansion of the Universe. Therefore, it has an indirect effect on the redshift-distance relation (the Hubble parameter, $H(z)$) and on the growth of structure. Thus, it is important to accurately measure the evolution of the Hubble parameter with redshift.

As we have seen, at low redshift the Hubble parameter can be measured using SNe. However, the SNe grow fainter the further away we look, and they are not suitable standard candles at higher values of redshift. We simply cannot observe them at those redshifts. At higher redshift, the Baryon Acoustic Oscillations (BAO) scale is commonly used as standard candles.

To understand the origin of the BAO scale we have to go back to the early Universe and focus on the behaviour of baryons at that epoch. The early Universe was very dense and matter and radiation were tightly coupled, their interaction preventing the formation of the first atoms. The competition of gravitational attraction and baryon pressure produced sound waves in the photon-baryon plasma. Once the Universe was cold enough so that radiation decoupled from baryons, both the photons and the baryons were left with an imprint caused by these sound waves. This imprint has a characteristic scale, the BAO scale, set by the sound horizon at the epoch of recombination. A more detailed explanation can be found in a review from Eisenstein (2005).

## 2.4 Tracers of the underlying dark matter distribution

In the previous sections we have seen the evolution of the standard model for the origin of the Universe to adapt to the new observational results. In particular, we have seen that the formation of structure was tightly coupled with the presence of DM and DE, but that we cannot directly measure neither. We can only acquire information on the baryon component. Hence, we need to develop the theory in more detail, enough to understand that the baryonic matter indeed trace the underlying dark matter distribution. This is the object of this section.

One of the pillars of the current cosmological models is the cosmological principle. According to this principle, the universe is isotropic and homogeneous. We have seen that we have to understand this principle in a statistical sense. The Universe is not exactly homogeneous, already in its very early times there are some primordial fluctuations that will later evolve to form the structures we observe today.

Assuming that the matter is pressureless, which is always true for DM, these small perturbations grow adiabatically with time to eventually form galaxies. At a given time $t$, the perturbations can be described by means of a density contrast, $\delta\left(\vec{x}\right)$:

$$\delta\left(\vec{x}\right) \equiv \frac{\rho\left(\vec{x}\right) - \bar{\rho}}{\bar{\rho}}\,, \tag{2.2}$$

where $\rho\left(\vec{x}\right)$ is the density at time $t$ and in a particular point $\vec{x}$, and $\bar{\rho}$ is the mean density of the universe at such a time. Note that for simplicity in the notation we drop the explicit dependence in time. A more detailed explanation of the evolution of the perturbations can be found in e.g. Padmanabhan (1993) or Byrd et al. (2007).

Having this density contrast field defined, we can measure the amplitude of the perturbations by measuring the two-point statistics of this field. The amplitude of the perturbations at some scale $r$ is given by the correlation function,

$$\xi\left(\vec{r}\right) \equiv \left\langle \delta\left(\vec{x}\right) \delta\left(\vec{x}+\vec{r}\right) \right\rangle \ , \tag{2.3}$$

where $\langle \cdots \rangle$ represents the average over all possible realizations of the fluctuation $\delta$ (according to the ergodic theorem, this is equivalent to computing the average over a large volume).

Another commonly used statistical descriptor is the power spectrum, $P\left(\vec{k}\right)$. The power spectrum is the Fourier transform of the cross-correlation. Namely,

$$P\left(\vec{k}\right) = \int d^3k\, \xi\left(\vec{r}\right) e^{-i\vec{k}\vec{r}} \, . \tag{2.4}$$

Correspondingly, the cross-correlation can be expressed as the inverse Fourier transform of the power spectrum

$$\xi\left(\vec{r}\right) = \frac{1}{(2\pi)^3} \int d^3r\, P\left(\vec{k}\right) e^{-i\vec{k}\vec{r}} \, . \tag{2.5}$$

Theoretical models often make their predictions in the power spectrum since the convolution of functions is easier to model in Fourier space. Plus, in the case that the density fluctuations follow a Gaussian distribution, the power spectrum gives a full statistical description of these perturbations.

When perturbations grow larger than the Jeans mass, they become unstable and grow exponentially, forming DM halos. The collapse time of these halos is shorter for lower mass halos, which means that smaller structures form earlier than more massive ones. Also, halos of a given mass will trace the underlying matter distribution in a somewhat biased way. It is now well known that their autocorrelation is proportional to the matter autocorrelation, their bias factor being the square root of the proportionality constant (Mo & White 1996). It is now well understood that the bias factor of the halos increases with their mass both analytically (see e.g. Sheth & Tormen 1999) and from accurate predictions from simulations (see e.g. Tinker et al. 2010).

Baryonic matter behaves in a slightly different way. First, it stays coupled with radiation way longer than DM, since it can interact electromagnetically. This means that the formation of structure is pushed to later times. Also, the electromagnetic interaction allows the baryons to cool. Once decoupled from the radiation, baryons feel the gravitational pull of DM and are therefore attracted into DM halos. Their cooling capacity reduces the velocity dispersion of baryons, making them cluster preferentially close to the center of DM halos. However, most of the baryons are still found in the intergalactic medium (IGM). This means that, at large scales, the autocorrelation of a population of tracers is approximately proportional to the matter autocorrelation. The constant of proportionality is the square of the bias factor and will depend on the mass of the halos that host the tracers.

It is interesting to measure the correlation of different tracers since different tracers provide independent constraints on the cosmological models. Different populations of baryonic matter are typically hosted in halos of different masses. Also, different tracers are located at different redshifts, which provides for a way of measuring the evolution of the Hubble parameter, $H(z)$. In fact some of the most competitive measurements of the $H(z)$ come from the analysis of BAO peak position in the autocorrelation of galaxies (Gil-Marín et al. 2015; Cuesta et al. 2016) or in the Ly$\alpha$ forest (Delubac et al. 2015).

Besides measuring the autocorrelation of several tracers one can measure the cross-correlations between pairs of tracers. Similarly to the autocorrelation of a given population of tracers, the cross-correlation of two different traces is proportional to the matter autocorrelation. The constant of proportionality in this case is the product of the two bias factors. The cross-correlation has the advantage that can be applied to tracers that don't have enough statistics to provide competitive constraints when measuring the autocorrelation. Also, measuring the cross-correlations can help break degeneracies between the different parameters of the models (Aubourg et al. 2015).

Summing up, current models indicate that baryonic matter trace the underlying dark matter density. The cross- and autocorrelations are proportional to the matter autocorrelation, the proportionality constant depending on the bias of the tracers involved. Measuring the cross-correlation and the autocorrelation we can give competitive constraints on the cosmological models. In particular, the different redshift of different populations allow us to measure the evolution $H(z)$ with high precision. In the next sections (sections 2.4.1 to 2.4.4 on pages 22–24) we will provide the reader with a brief description of the tracers used in this thesis.

### 2.4.1 Quasars and galaxies

After the decoupling of radiation and baryonic matter at the epoch of recombination there were no sources of light in the Universe but the CMB. The Universe had entered the so-called Dark Ages. During the Dark Ages the density perturbations grew until the first DM halos were formed. Baryonic matter collapsed into the center of these halos until eventually they were massive enough and dense enough to trigger thermonuclear reactions and so the first stars were born. It is not yet clear the exact picture of the formation of these first stars: they are very different from current stars since they had virtually no metals; only the very light elements formed during the primordial nucleosynthesis. We are not going to discuss the formation of the first stars here, instead we refer to the review Bromm & Larson (2004) and references therein.

After the first SNe, the radiation emitted from these primordial stars and the accretion disks of the black holes heated and ionized the surrounding gas slowing down star formation. While there are several approaches for understanding how galaxies form, the main picture is that several millions of stars and huge amounts of baryonic gas collapsed inside these DM halos forming a gravitationally bound structure encompassing all of them which we call galaxies. More details in the formation and evolution of the first galaxies and their structure can be found in e.g. Bromm & Yoshida (2011); Silk & Mamon (2012); Conselice (2014).

Even though the formation and evolution of galaxies is not the interest of this thesis, there is a particular phase on galactic evolution that deserves special attention. These would be the Active Galactic Nuclei (AGN) phase (see the review Antonucci 1993). As the name indicates, this phase involves huge amount of activity near the gravitational center of the galaxies. In more detail the activity is produced by the accretion of a huge quantity of gas by the supermassive black holes placed at the center of the galaxies. Through processes not yet fully understood, this accretion results in the ejection of powerful jets made mostly from gas that can achieve relativistic speeds. A galaxy can undergo this AGN phase several times during its evolution history until the galaxy is depleted of gas.

When an early galaxy undergoes the AGN phase, the gas content is still very high. This means that the available material to accrete is much more massive than it would be at later stages. In practice this results to much jets so powerful that, if oriented towards us, can easily be observed with medium sized telescopes. Such an object is called quasar and was first observed back in 1963 (Hazard et al. 1963; Schmidt 1963). Being such bright and distant objects, the spectra of quasars provide means to uncover new populations of tracers, as we shall see in the following sections.

### 2.4.2 Ly$\alpha$ forest

In 1965, Maarten Schmidt spectroscopically observed the quasar 3C9 and reported a redshift of $z = 2.01$ (Schmidt 1965). This was the first of many quasars to be observed at redshifts greater than 2. Current datasets contain hundreds of thousands of such quasars (see e.g. the Baryon Oscillations Spectroscopic Survey (BOSS) twelfth Data Release quasar catalogue (DR12Q), Pâris & et al. 2016). All these quasars have one important thing in common: their Ly$\alpha$ emission line is redshifted to the optical part of the spectrum.

From the spectra of these quasars, and due to the fact that there is non-zero flux at higher energies than the Ly$\alpha$ transition, a strong upper limit can be placed on the abundance of neutral hydrogen in the IGM (Gunn & Peterson 1965). The Gunn-Peterson trough, as the effect is known, consists on the following: photons emitted at higher energies than the Ly$\alpha$ alpha transition are gradually redshifted until they have the exact frequency to excite neutral hydrogen. If a small fraction of the IGM is ionized when this happens, then the Ly$\alpha$ absorption causes a dramatic decrease in the flux transmitted fraction at a the particular energy of these photons.

Bluewards of the Ly$\alpha$ emission line multiple absorption features are usually detected. These hydrogen clouds constitute the so-called Ly$\alpha$ forest. It was first proposed by Bahcall & Salpeter (1965) and later confirmed by Lynds (1971). These absorption features correspond to the line of sight crossing a gas cloud with some neutral hydrogen content. The column density of the neutral hydrogen can be inferred from the flux transmitted fraction, even though its value is quite uncertain. For a review on the Ly$\alpha$ forest see e.g. Rauch (1998).

### 2.4.3 Damped Ly$\alpha$ Systems

Within the Ly$\alpha$ forest there are some absorbers that distinguish themselves from the rest. They show a damped profile of their Ly$\alpha$ absorption which is distinguishable even in low resolution spectra. The hydrogen column density can be determined from this damped profile with much higher precision than the regular absorbers in the Ly$\alpha$ forest. The values for the column density are measured to be $N_{HI} \geqslant 10^{23}$ cm$^{-2}$ (Wolfe et al. 1986).

These damped absorbers or Damped Ly$\alpha$ Absorbers (DLAs) are dense hydrogen clouds, dense enough so that they become self-shielded of external cosmic ionizing background radiation. Thus, except for the outermost part of the clouds, their hydrogen is mostly in atomic form (Vladilo et al. 2001). See Wolfe et al. (2005) for a review on these absorbers.

### 2.4.4 Metal absorbers

In astronomy, the different elements are referred to using a particular jargon, which may sometimes lead to confusions. In particular the word *metals* is used to refer to any element but hydrogen and helium in a generic way. This means that, for an astronomer, oxygen is considered a metal, same as, e.g., iron. This probably comes from the fact that the combined relative abundance of all the elements but hydrogen and helium hardly reach 2%, and has nothing to do on the way the different elements bond to form molecules. Note that in this thesis we have adopted the same nomenclature.

Similarly to the Ly$\alpha$ absorption produced in the Ly$\alpha$ forest, absorption by other metal lines can also be detected in quasar spectra. However, these absorption features are much fainter and rarer than those of the Ly$\alpha$ lines. The reason for this is that the abundances of metals in the Universe are much lower than that of hydrogen. Metals are formed mostly in the SNe resulting from the final stages of stellar evolution; only a small percentage of the stellar mass (mostly hydrogen and helium) is fused into heavier metals.

While there are now hints of an actual forest for other metal lines (see e.g., the C IV forest, Pieri 2014), most metal absorption occurs in isolated, relatively rare absorbers. The main problem is to identify the correct species producing the absorption. One of the most studied lines is the Mg II line. This line presents a characteristic doublet that is easy to identify. Also, it presents a high oscillation frequency and can be observed at low redshift using optical telescopes. This makes Mg II absorbers as one of the best metal tracers of the underlying matter distribution. For this reason we have chosen these absorbers as one of the tracers of study in this thesis.

## 2.5 Description of the thesis

The work presented in this Ph.D. thesis is mostly related to the measurement and interpretation of the cross-correlation of different species. This is achieved using spectroscopic data from the BOSS survey (part of the Sloan Digital Sky Survey III (SDSS-III) Collaboration, Eisenstein et al. 2011), of which I am a member.

Along the thesis I will be visiting the cross-correlation of different tracers, which have been introduced in sections 2.4.1 to 2.4.4 on pages 22–24. The first cross-correlation I address, in chapter 3 on page 29, is that of Mg II absorbers with CMASS galaxies. I present a new method for measuring the equivalent width of metal absorbers, developed in collaboration with Britt Lundgren and Jordi Miralda-Escudé, and I apply it to the Mg II absorption doublet. This chapter is basically a reproduction of Pérez-Ràfols et al. (2015).

Then I shift my attention to the DLAs and the Ly$\alpha$ forest. Since the Ly$\alpha$ forest autocorrelation has already been studied (see e.g. Delubac et al. 2015; Blomqvist et al. 2015), the Ly$\alpha$ parameters are well constrained. I use those constraints to provide more accurate measurements of the DLA bias when analysing the cross-correlation of DLAs with the Ly$\alpha$ forest. The work presented in this chapter has been made in collaboration with Andreu Font-Ribera, Nicolás Busca, and Jordi Miralda-Escudé. In particular, the estimator for the cross-correlation was originally developed by Nicolás Busca for the analysis of the Ly$\alpha$ autocorrelation and later adapted to the measurement of the cross-correlation. To fit the model I make use of *baofit*, a fitter developed mostly by David Kirkby and maintained by Michael Blomqvist (Kirkby et al. 2013).

The cross-correlations studied in chapters 3 and 4 on page 29 and on page 67, and more importantly their interpretation, are mostly focused on large scales, where the linear theory is valid. At smaller scales, non-linear effects start to kick in and the linear approximation described in this introduction is no longer valid. In chapter 5 on page 99 I measure the cross-correlation of quasars with the Ly$\alpha$ forest, and focus on the small scales effects. I show that a contamination from metals is clearly present in the cross-correlation. Also, I discuss some of the possible non-linear effects and give a simple theoretical model to explain the effect that quasar radiation can have in the surrounding hydrogen clouds. This study has been made in collaboration with Jordi Miralda-Escudé.

Finally, in chapter 6 on page 119 I present the overall conclusions of this thesis and I briefly comment on some future perspectives.

# Part II

# Results

# 3

# Mg II absorbers - CMASS galaxies cross-correlations

We present a measurement of the cross-correlation of Mg II absorption and massive galaxies, using the DR11 main galaxy sample of the Baryon Oscillations Spectroscopic Survey of SDSS-III (CMASS galaxies), and the DR7 quasar spectra of SDSS-IISDSS-II. The cross-correlation is measured by stacking quasar absorption spectra shifted to the redshift of galaxies that are within a certain impact parameter bin of the quasar, after dividing by a quasar continuum model. This results in an average Mg II equivalent width as a function of impact parameter from a galaxy, ranging from 50 kpc to more than 10 Mpc in proper units, which includes all Mg II absorbers. We show that special care needs to be taken to use an unbiased quasar continuum estimator, to avoid systematic errors in the measurement of the mean stacked Mg II equivalent width. The measured cross-correlation follows the expected shape of the galaxy correlation function, although measurement errors are large. We use the cross-correlation amplitude to derive the bias factor of Mg II absorbers, finding $b_{\mathrm{MgII}} = 2.33 \pm 0.19$, where the error accounts only for the statistical uncertainty in measuring the mean equivalent width. This bias factor is larger than that obtained in previous studies and may be affected by modelling uncertainties that we discuss, but if correct it suggests that Mg II absorbers at redshift $z \simeq 0.5$ are spatially distributed on large scales similarly to the CMASS galaxies in BOSS.

*This chapter is basically a reproduction of Pérez-Ràfols et al. (2015).*

## 3.1 Introduction

The Mg II doublet absorption line, at rest-frame wavelengths $\lambda = 2796.35$Å and 2803.53Å, is an extremely useful tracer of photoionized gas clouds in galactic halos. Several reasons account for this trait: magnesium is among the most abundant of the heavy elements, the oscillator strength of this Mg II doublet is particularly large, its rest-frame wavelength makes it easily observable from ground-based telescopes at redshift $z > 0.3$, and magnesium is mostly in the form of Mg II in photoionized, self-shielded clouds at the typical pressures of galactic halos. Gas clouds with atomic hydrogen column densities $N_{HI} \lesssim 2 \times 10^{17}$ cm$^{-2}$ are optically thin to photons above the hydrogen ionization potential of 13.6 eV, as well as to the harder photons that ionize Mg II. Most of the magnesium is therefore ionized more than once in thin clouds, greatly reducing the Mg II column density. On the other hand, in optically thick clouds, the shielded magnesium ions are mostly recombined to Mg II, but Mg I has an ionization potential of 7.65 eV (below that of hydrogen) and is ionized by photons that penetrate the region of self-shielded hydrogen. Hence, self-shielded clouds in galactic halos should have most of their magnesium as Mg II (e.g., Bergeron & Stasińska 1986). In fact, most strong Mg II systems are also Lyman limit systems, i.e., their HI column densities are high enough to be self- shielding (Rao et al. 2006).

The rest-frame equivalent width distribution of Mg II absorption systems is approximated by an exponential form, $dN/dz \propto \exp(-W/W^*)$, with a value $W^* \simeq 0.6$Å at $z = 0.5$ that increases gradually with redshift, and an excess of systems over this form at $W < 0.3$Å (Nestor et al. 2005; Narayanan et al. 2007, and references therein). Most of the strong systems have a complex velocity structure, with a velocity dispersion of multiple absorbing components that is characteristic of galaxy velocity dispersions, favouring models of a collection of photoionized clouds randomly moving through a galactic halo (Bahcall 1975; Sargent et al. 1979; Churchill et al. 2000), as had already been proposed by Bahcall & Spitzer (1969).

The association of Mg II absorption systems with galactic halos was firmly established with the work of Lanzetta & Bowen (1990); Bergeron & Boissé (1991); Steidel et al. (1994). The observations of the frequency of occurrence of Mg II absorbers at different impact parameters from luminous galaxies led to a simple model of halos that are close to spherical, in which absorption with rest-frame equivalent width $W > 0.3$Å is nearly always observed within an impact parameter $r_p \lesssim 50(L_K/L_K^*)^{0.15}$ kpc of a galaxy of K-band luminosity $L_K$, and becomes rapidly weaker at larger radii, independently of the type of galaxy being considered (Steidel 1995). Actually, the natural expectation is that there is a smooth profile of declining mean Mg II absorption strength with impact parameter around a galaxy, caused by a decreasing density of clouds with radius in a galactic halo. This is consistent with more recent work, where the mean Mg II equivalent width (which is indicative of the number of intersected individual clumps with saturated absorption) has been shown to steeply decline with the impact parameter $r_p$ roughly as $\overline{W} \propto r_p^{-1.5}$ (Chen et al. 2010a), with a characteristic radius at a fixed $\overline{W}$ that scales proportionally to $R_{\mathrm{Mg\,II}} \propto M_*{}^{0.3}(\mathrm{sSFR})^{0.1}$ (Chen et al. 2010b), where $M_*$ is the stellar mass in the galaxy and sSFR is the star formation rate per unit of stellar mass. We also note that this describes the mean profile of Mg II absorption around galaxies, and that there are rare cases where Mg II absorption is absent even at very small impact parameter from luminous galaxies (Johnson et al. 2014).

The large extent of the gaseous halos traced by metal absorbers and their nearly ubiquitous presence around all massive galaxies, with only a weak dependence on the specific star formation rate, are observational facts providing support to a scenario in which the Mg II absorbers are a signature of the accretion process of new material onto galaxies. Accreting gas at the temperatures of virialization in galactic halos is thermally unstable and should naturally form photoionized clouds whenever the cooling time of hot halo gas is short compared to the age of the galaxy. This behaviour naturally leads to a two-phase model of gaseous galactic halos, where cool clouds can form in approximate pressure equilibrium with a hot medium and are produced in abundance within the cooling radius (Spitzer 1956; Mo & Miralda-Escudé 1996; Maller & Bullock 2004). On the other hand, there is evidence at small radii pointing to the impact of galactic winds on Mg II absorbers: the absorbers are more numerous near the minor axis of their associated galaxies (Bordoloi et al. 2011, 2014; Kacprzak et al. 2011, 2012; Lundgren et al. 2012). The distribution of Mg II absorption systems around galaxies is therefore likely to be sensitive to processes involving both inflow and outflow of material into and from the regions containing the bulk of the stellar mass in galaxies.

The association of Mg II absorbers with galaxies implies a large-scale cross-correlation of these objects. The cross-correlation is in general a result of two effects. First, if every galaxy is surrounded by a gas halo, a Mg II absorber located at an impact parameter $r_p$ from a galaxy may actually be associated with this galaxy, and be part of the gas halo around it. Second, the Mg II absorber may be associated with a different galaxy that may be a satellite of the first, or with an unrelated galaxy that is spatially correlated with the first, following the usual galaxy autocorrelation. These are usually described as the 1-halo and 2-halo terms, although Mg II systems are considered to be associated to galaxies, rather than dark matter halos. In the limit of impact parameters much smaller than the typical size of a galaxy halo, the first term, determined by the gas halo profile around each galaxy (e.g., Tinker & Chen 2008), dominates, whereas in the limit of large impact parameters, the second term is the important one. The total cross-correlation is in general a combination of the two terms over a wide range of intermediate impact parameters, and it is impossible to cleanly separate the two contributions. But in the large-scale limit, the cross-correlation of Mg II absorbers and galaxies should follow the form of the galaxy correlation function, with an amplitude that is proportional to the product of the two bias factors of the two populations. Hence, measuring the large-scale clustering amplitude helps determine the bias factor, and therefore the halo population that the Mg II absorbing clouds are associated with.

The large-scale Mg II-galaxy cross-correlation was first measured using the photometric catalogue of Luminous Red Galaxies in the SDSS-III (see York et al. 2000) and the set of individually detected Mg II absorbers in the spectra of SDSS quasars by Bouché et al. (2004, 2006); Lundgren et al. (2009); Gauthier et al. (2009); Lundgren et al. (2011). In the absence of precise measurements of galaxy redshifts, only the projected cross-correlation function can be measured. The work by Lundgren et al. (2009), based on a set of 2705 Mg II absorbers with rest-frame equivalent width $W > 0.8$Å over the redshift interval $0.36 < z < 0.8$, found that the form of the projected cross-correlation is well matched by the Luminous Red Galaxy autocorrelation in the impact parameter range $0.3\,h^{-1}\mathrm{Mpc} < r_p < 30\,h^{-1}\mathrm{Mpc}$, and measured the bias factor of these Mg II systems to be $b_{Mg} = 1.10 \pm 0.24$. Gauthier et al. (2009) performed a similar analysis for a sample of 1158 Mg II absorbers with $W < 1$Å over the redshift interval $0.4 < z < 0.7$ and found the bias factor of these Mg II systems to be $b_{Mg} = 1.36 \pm 0.38$. Both works found an indication that weak absorbers, with $W < 1.5$Å, cluster more strongly (i.e., they have a larger bias factor) than strong absorbers, and are therefore located in more massive halos on average, although this result was not of high statistical significance. The method used by Bouché et al. (2006), Lundgren et al. (2009) and Gauthier et al. (2009), based on a catalogue of identified Mg II systems, requires careful attention to the selection function of *both* galaxies and Mg II systems through an extensive use of simulations, because the number of absorbers will obviously be enhanced in regions of the survey containing more galaxies and more quasar spectra, or where the spectra are of higher signal-to-noise ratio, owing to variable observing conditions and also intrinsic clustering of the quasar sources.

We present a different approach in this chapter to measure the cross-correlation of massive galaxies and Mg II absorbers, based on the galaxies with spectroscopically-measured redshifts of the new BOSS of the SDSS-III Collaboration (Eisenstein et al. 2011; Dawson et al. 2013). Instead of identifying individual Mg II absorbers, we use a stacking method to measure the average Mg II absorption around a galaxy as a function of the impact parameter and redshift separation. In other words, we measure the redshift-space cross-correlation function of galaxies and Mg II absorption. Our approach is similar to that of Zhu & Ménard (2013b), who have measured the mean Ca II absorption around galaxies, and to Zhu et al. (2014), who have obtained a similar measurement of the large-scale Mg II absorption; we compare their results with ours and discuss the differences near the end of this chapter. The data set we use is described in section 3.2, and the method is presented in detail in sections 3.3 and 3.4 on page 34 and on page 42. We present the results in section 3.5 on page 44, which are applied to infer the mean bias factor of Mg II systems in section 3.6 on page 49. Finally we summarize our conclusions in section 3.7 on page 59. Throughout this chapter we use a flat ΛCDM model with $H_0 = 68\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$ and $\Omega_m = 0.3$.

## 3.2 Data Sample

The first step in the analysis is to identify quasar-galaxy pairs in which the quasar line of sight passes within a specified bin of projected proper radius, or impact parameter $r_p$, from the foreground galaxy. For the background quasar sample we use the quasar catalogue of Schneider et al. (2010) from the DR7 (Abazajian et al. 2009) of the SDSS-II Collaboration (Gunn et al. 1998; York et al. 2000; Gunn et al. 2006; Eisenstein et al. 2011; Bolton et al. 2012; Smee et al. 2013), with 105,783 spectroscopically confirmed quasars. For the galaxies, the CMASS catalogue (Dawson et al. 2013) of the SDSS-III Collaboration that is prepared for the DR11 (extension of the SDSS DR9 (Ahn et al. 2012) and the SDSS DR10 (Ahn et al. 2014)) is used, which contains a total of 938,280 galaxies. The DR11 galaxies represent the majority of the final BOSS sample, also covering most of the sky area included in DR7. We note that this galaxy catalogue is the same that was used by Zhu et al. (2014), even though they refer to it as the Luminous Red Galaxy catalogue. We exclude any galaxies at redshift lower than $0.35$, corresponding to $1 + z = \lambda_{\rm min}/\lambda_{\rm Mg\,II}$, where we set $\lambda_{\rm min} = 3800$Å as the shortest wavelength with sufficient signal-to-noise ratio to provide a useful Mg II absorption signal. This requirement reduces the galaxy sample to 895,472 galaxies. The quasar sample could be increased by including the DR11 sample from BOSS, but most of the new quasars in DR11 are fainter than those in DR7 (having therefore lower signal-to-noise) and are at a high redshift, at which the Mg II lines associated with the CMASS galaxies appear superposed with the Ly$\alpha$forest.

The mean Mg II absorption around galaxies is computed by stacking all the spectra within a certain range of impact parameters from a galaxy. In general, all quasars probing a line-of-sight within a maximum proper impact parameter that was set to $r_{\rm p,max} = 12.8$ Mpc are included in the spectral stacking, provided that the following restrictions are met: first, in order to avoid broad absorption line systems associated with quasar outflows, the redshift of the quasar, $z_{\rm q}$, must be larger than that of the galaxy, $z_{\rm gal}$, by a minimum amount, corresponding to a velocity $v = 0.04c$,

$$1 + z_{\rm q} > R(1 + z_{\rm gal}) \,, \tag{3.1}$$

where $R = \sqrt{(1 + v/c)/(1 - v/c)} \simeq 1.041$. Second, because the presence of the Ly$\alpha$ forest substantially increases the noise in quasar spectra, pairs where the galaxy Mg II line would fall in the Ly$\alpha$ forest region of the quasar are also excluded; in other words, we require

$$(1 + z_{\rm gal})\lambda_{\rm Mg\,II} > (1 + z_{\rm q})\lambda_{\alpha} \,, \tag{3.2}$$

where $\lambda_{\rm Mg\,II} = 2798.74$Å is the mean wavelength of the Mg II doublet, and the hydrogen Ly$\alpha$ wavelength is $\lambda_{\alpha} = 1215.67$Å at the rest-frame. Note that the quasars that have been reported to have broad absorption lines are also included in the sample. We tested the effect of removing them and we found that the measurements are not significantly changed. We believe that this is because we are only using small windows on the quasar spectra, and therefore the broad absorption line systems are typically not included in the analysis.

The luminosity and redshift distributions of all the DR11 CMASS galaxies meeting these two conditions for at least one quasar that is within an impact parameter $b_{\rm max} = 12.8$ Mpc, as computed from the angular separation at the redshift of the galaxy, are shown in figure 3.1 on the next page. Note that the density of quasars in DR7 implies that each galaxy has on average $\sim 3$ quasars within this impact parameter, therefore the majority of galaxies have at least one quasar pair and these distributions are nearly the same as those of the whole DR11 CMASS sample. These distributions represent the characteristics of the galaxies for which we measure the mean Mg II

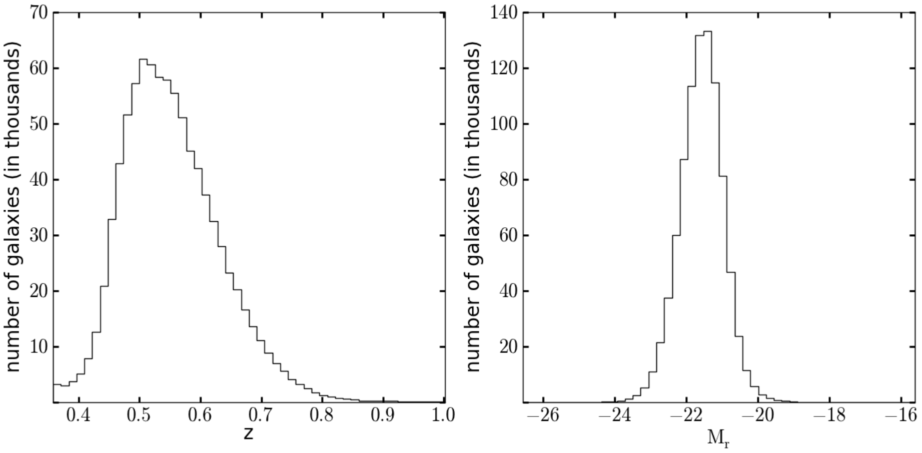


Figure 3.1: Redshift (left panel) and luminosity (right panel) distributions for the selected CMASS galaxies (see text for details) .

absorption equivalent width as a function of impact parameter in this chapter.

## 3.3 Stacking procedure

This section describes the method used to measure the average Mg II absorption equivalent width as a function of impact parameter from the CMASS galaxies. In general, the mean transmission fraction $\bar{F}$ of light from a background quasar due to Mg II absorption line systems can be written as a function of the impact parameter $r_p$ and velocity separation $v$ from a galaxy, as

$$\bar{F}(r_p, v) = \exp\left[-\tau_e(r_p, v)\right] = \exp\left[-\tau_{eo}(1 + \delta_{Mg}(r_p, v))\right] \ . \tag{3.3}$$

Here, $\tau_e$ is the effective optical depth, and $\tau_{eo}$ is its average value at any random position, irrespective of the presence of a galaxy. The perturbation $\delta_{Mg}$ is the relative increase of the effective optical depth of Mg II absorption associated with the presence of a galaxy at impact parameter $r_p$ and velocity separation $v$. The shape of this perturbation as a function of $v$, for a given impact parameter, is rather complicated because it arises from the distribution of relative velocities of a set of doublet lines with different degrees of saturation; besides, the observations are affected by the instrumental resolution and the binning. We will return to these details later.

The mean effective optical depth at a random position can be expressed in terms of the rest-frame equivalent width distribution of Mg II absorbers, as

$$\tau_{eo} = \int_0^\infty dW \, \frac{\partial^2 \mathcal{N}}{\partial W \partial z} \, (1 + z) \frac{W[1 + \bar{q}(W)]}{\lambda_{\mathrm{Mg\,II}}} \ , \tag{3.4}$$

where the equivalent width $W$ is that of the strongest line of the Mg II doublet at $\lambda = 2796.35$Å, and $\bar{q}(W)$ is the mean equivalent width ratio of the two doublet lines.

Our aim in this chapter is to measure the excess in the effective optical depth,

$$\delta\tau_e(r_p, v) = \tau_e(r_p, v) - \tau_{eo} = \tau_{eo}\delta_{Mg}(r_p, v) \, , \tag{3.5}$$

which is induced by the presence of a galaxy at impact parameter $r_p$ and velocity separation $v$. This quantity is directly related to the cross-correlation of Mg II clouds with CMASS galaxies. We will focus in particular on the projected value of $\delta\tau_e$, obtained after integration over velocity, and its relationship to the projected cross-correlation. The method we use to measure this cross-correlation is to average the transmitted fraction over a large number of lines-of-sight within a given range of impact parameter, in order to reduce the photon noise and the noise arising from quasar continuum variability.

A crucial step to measure the Mg II-galaxy cross-correlation, in the form of $\delta\tau_e$, is to estimate the quasar continuum with a method that is, to the best possible degree, free of systematic errors when averaging over a large number of lines-of-sight. In particular, it is important to ensure that the presence of a Mg II line itself, which in most cases may be too weak to be individually detected in the relatively low signal-to-noise ratio SDSS quasar spectra, does not bias the estimate of the continuum. Obviously, if the spectral region where we expect to find the Mg II line associated with a galaxy is used for the continuum determination, the continuum estimate may be systematically biased too low because of the presence of an undetected Mg II line. This systematic bias may be important when stacking large numbers of spectra, even though the Mg II lines causing the bias are not individually detected in any single spectrum.

To illustrate the importance of the quasar continuum determination for the problem of measuring the average Mg II absorption around galaxies at large impact parameters, we explore two different methods and we perform a number of tests for the presence of systematic errors. The first method, designated as *mean subtraction*, is specifically designed for our problem, while the second method, designated as *variable smoothing*, fits a continuum by iteratively smoothing the spectrum with a variable smoothing length to flexibly fit both emission lines and featureless continuum regions. The latter method uses the entire spectrum to determine the continuum, including the region where the associated Mg II line is expected, and is therefore subject to the systematic error described above. We now describe each method in detail. Tests of the methods that show that the mean subtraction method correctly recovers the mean equivalent width and the variable smoothing method is subject to various systematic errors, including the continuum fitting bias mentioned above, are presented in appendix 3.A on page 60.

### 3.3.1 Method 1: Mean subtraction

The first approach uses the mean spectrum of all quasars as a continuum fit model. Each quasar spectrum is divided by the mean quasar spectrum, and then a linear fit to this ratio is obtained around the spectral region of the expected Mg II line for each galaxy-quasar pair, but without using the narrower central interval where the Mg II line should be located. This linear fit is used to further improve the continuum estimate, allowing for intrinsic variations of the quasar continua. The results are then stacked for all quasar-galaxy pairs at each interval of impact parameter, after shifting to the redshift of the galaxy in each pair, to obtain the final composite Mg II absorption spectrum. We now explain each of these steps in detail.

### Generating the mean spectrum

The mean quasar spectrum is generated using the DR7 quasar spectra following a similar approach as the one undertaken by Vanden Berk et al. (2001). The mean spectrum in Vanden Berk et al. (2001) was generated using 1,800 quasars (figure 3 of Vanden Berk et al. 2001), so our mean spectrum is more accurate because of the much larger number of quasars available in DR7. In addition, we normalize the spectra using a spectral window that is particularly suited to obtaining the most accurate continuum model in the region where Mg II absorption lines are found. There is therefore a small difference between our mean spectrum and that of Vanden Berk et al. (2001).

The quasar spectra are first shifted to the rest-frame, using the redshift measurements of Schneider et al. (2010), and rebinned into a common wavelength scale of 1Å per bin, which is close to the resolution of the observed spectra when shifted to the rest-frame. The values of the flux and the error at each pixel in the new binning are determined by the average values of the flux and error in the original pixels that are projected, partly or fully, to the new pixel, weighted by the fraction of the new bin covered by each original bin. Each quasar spectrum, denoted by an index $j$, is normalized with a normalizing factor $n_j$ equal to the mean flux in the interval $2000 - 2600$Å,

$$n_j = \sum_i f_{ij}/N_j \ . \tag{3.6}$$

where $f_{ij}$ is the measured flux value at pixel $i$ of spectrum $j$, and $N_j$ is the number of pixels in the rest-frame wavelength interval $2000\text{Å} < \lambda_{ij}(1+z_j) < 2600\text{Å}$. Any quasar spectra that do not cover this entire range of rest-frame wavelength are discarded. The final number of quasar spectra that are averaged in each 1Å bin is shown in figure 3.2 on the next page. The flat top corresponds to the spectral window used to compute the normalizing factor $n_j$. Note that the total number of quasar spectra used in this method is $70,650$. The maximum number of quasars shown in figure 3.2 on the facing page, roughly $68,600$ quasars, is lower than the total number of spectra used because we remove pixels affected by sky lines. Even though these sky lines are corrected by the BOSS pipeline, the noise in the affected pixels may sometimes not be well characterized, so it is best to remove these pixels. For each sky line, we remove a set of neighbouring pixels following the algorithm summarized in Palanque-Delabrouille et al. (2013) (for a more detailed explanation of the algorithm refer to Lee et al. 2013).

The rest-frame wavelength interval of $2000 - 2600$Å is also used to assign a mean signal-to-

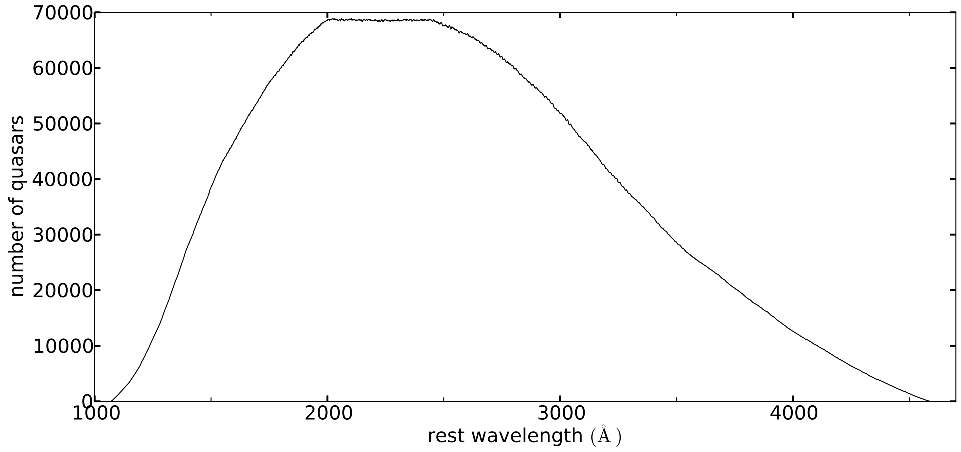


Figure 3.2: Number of quasar spectra contributing to each 1Å bin of the mean spectrum, as a function of rest-frame wavelength. The flat feature corresponds to the spectral range used to compute the normalizing factor $n_j$. Outside this range the number of quasars contributing to the mean quasar spectrum decreases because some of the quasar spectra do not extend to that wavelength.

noise ratio $s_j$ to each spectrum, calculated as

$$s_j = \frac{\sum_i f_{ij}/N_j}{\left(\sum_i e_{ij}^2/N_j\right)^{1/2}}, \tag{3.7}$$

where $e_{ij}$ is the noise of the flux $f_{ij}$. The mean, normalized quasar spectrum is then obtained as a weighted average of all the quasar spectra,

$$\bar{f}_i = \frac{\sum_j w_j \left(f_{ij}/n_j\right)}{\sum_j w_j}, \tag{3.8}$$

where the weights $w_j$ are set equal to

$$w_j = \frac{1}{s_j^{-2} + \sigma_I^2}. \tag{3.9}$$

The constant $\sigma_I$ is introduced to avoid the highest signal-to-noise ratio spectra from excessively contributing to the final average, taking into account the presence of intrinsic quasar spectral variability, while reasonably weighting down the more noisy spectra. We fix this constant to $\sigma_I = 0.05$ (a reasonable estimate for the typical fractional intrinsic variability) throughout this chapter. The resulting mean spectrum is shown in figure 3.3 on the next page.

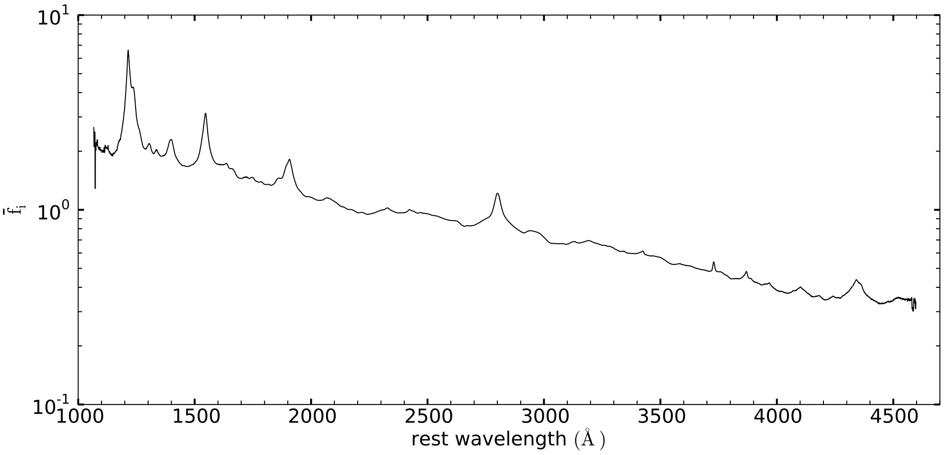


Figure 3.3: Mean spectrum of the weighted-average obtained from the 70, 650 DR7 quasars, normalized in the rest-frame wavelength interval from 2000 to 2600Å. This mean spectrum is used as a continuum model.

### Generating the composite Mg II absorption spectra

The composite Mg II absorption spectra are obtained by stacking the spectra of all the quasars having a galaxy within the corresponding impact parameter bin, after being shifted to the rest-frame of the galaxy in the region around the Mg II doublet wavelength. For each quasar-galaxy pair, the quasar spectrum is first rebinned in a velocity variable $v$, defined in terms of the wavelength separation from the Mg II absorption line in the rest-frame of the galaxy at redshift $z_{\rm gal}$,

$$v = c \cdot \frac{\lambda - \lambda_0}{\lambda_0} \,, \tag{3.10}$$

where $\lambda_0 = \lambda_{\rm Mg\,II}(1 + z_{\rm gal})$, and $\lambda_{\rm Mg\,II} = 2\,798.743$ Å. We use the same linear rebinning method described in section 3.3.1 on page 36, with a bin size $\Delta v = 50$ km s$^{-1}$. The mean spectrum is also rebinned in the same manner, but using $\lambda_0 = \lambda_{\rm Mg\,II}(1 + z_{\rm gal})/(1 + z_q)$, where $z_q$ is the quasar redshift, because the mean quasar spectrum is computed in the quasar rest-frame.

The rebinned spectra $f_{ik}^{(r)}$, where the $i$ index now labels the new bins in $v$, and the $k$ index labels each quasar-galaxy pair in a certain impact parameter bin, are then divided by the continuum to obtain a first estimate of the transmission fraction $F_{ik}^{(0)}$,

$$F_{ik}^{(0)} = \frac{f_{ik}^{(r)}/n_{j(k)}}{\bar{f}_i} \,, \tag{3.11}$$

where the normalization factor $n_{j(k)}$ is that of the $j$ quasar corresponding to each pair $k$. The mean quasar spectrum $\bar{f}$ is understood to be the rebinned one and evaluated at the same bins in $v$

for each quasar-galaxy pair. Hence, if all quasars had identical intrinsic emission spectra, and in the absence of intervening absorption and observational noise, this transmission would be equal to unity for all quasars. The errors are normalized in the same way and computed according to $E_{ij} = e_{ij}/n_j/\bar{f}_i$.

In order to account for intrinsic variations in the spectra of quasars, we allow for a local smooth gradient in the ratio of each quasar spectrum to the mean spectrum in the region around each expected Mg II absorption line. This is modelled by first calculating a weighted average value of $F_{ik}^{(0)}$ on two intervals in $\nu$ on each side of the expected Mg II absorption associated with the galaxy, which are far enough from the center so that any associated absorption can be neglected. The intervals used throughout this chapter are $-5\,000\,\mathrm{km\,s^{-1}} < \nu < -2\,000\,\mathrm{km\,s^{-1}}$ and $2\,000\,\mathrm{km\,s^{-1}} < \nu < 5\,000\,\mathrm{km\,s^{-1}}$, with their weighted averages designated as $F_k^{(-)}$ and $F_k^{(+)}$, respectively. The weights for each bin are set to $w_{ik} = (\sigma_{n,ik}^2 + \sigma_I^2)^{-1}$, where $\sigma_{n,ik} = E_{ik}/F_k^{(+,-)}$ is the inverse signal-to-noise ratio at each pixel (we use the averages $F_k^{(+)}$ and $F_k^{(-)}$ for the signal, instead of the values at each pixel $F_{ik}^{(0)}$, to avoid biasing the result by systematically giving higher weights to pixels with positive noise). We use, as before, $\sigma_I = 0.05$. Weighted averages of the mean velocities $\nu_k^{(-)}$ and $\nu_k^{(+)}$ are computed in the same manner, which are usually close to the central values of the intervals $-3500$ and $+3500\,\mathrm{km\,s^{-1}}$, but not exactly so. A linear function $L_{ik}$ matching these two points is then defined,

$$L_{ik} = F_k^{(-)} + (F_k^{(+)} - F_k^{(-)})(\nu_i - \nu_k^{(-)})/(\nu_k^{(+)} - (\nu_k^{(-)}) \; . \tag{3.12}$$

In order to better adjust the quasar continuum in the presence of unrelated random absorption lines or bad pixels, the calculation of the two weighted averages $F_k^{(-)}$ and $F_k^{(+)}$ is recomputed after eliminating all outlier pixels in which the normalized flux deviates by more than $3\sigma$ from the fitted linear function, i.e., pixels where $\|F_{ik}^{(0)} - L_{ik}\| > 3(\sigma_{n,ik}^2 + \sigma_I^2)^{1/2}$. If more than 20% of the pixels in any of the two intervals are rejected under this criterion, the quasar spectrum is considered anomalous in the region of the expected Mg II line and the quasar-galaxy pair under consideration is rejected and not included in the final processing.

The transmission fraction is then corrected by this linear fit as

$$F_{ik} = F_{ik}^{(0)} + (1 - L_{ik}) \; . \tag{3.13}$$

We note here that although it would be in principle more correct to divide by the linear fit, setting $F_{ik} = F_{ik}^{(0)}/L_{ik}$, we found that this procedure inevitably introduces a systematic feature in the final stacked spectrum owing to the fact that a Gaussian error in the function $L$ results in a non-Gaussian distribution of the final transmission $F$ when $L$ is in the denominator, which is very difficult to correct for. We therefore decided to subtract $L$ following equation 3.13.

Finally, we use the same weights as in equation 3.9 on page 37 to compute the final composite

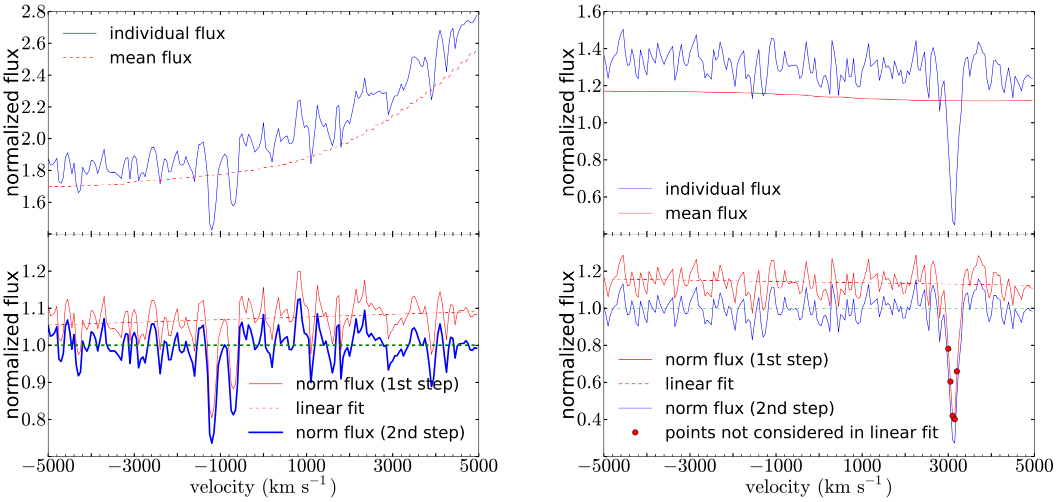


Figure 3.4: Examples illustrating the procedure explained in section 3.3.1 on page 38. The left panels shows a case with a detected individual Mg II absorption system, and the right panels a case with no detectable associated Mg II absorption system, but with an unrelated metal absorption line. The top panels show the normalized flux of the spectral region, $f_{ik}^{(r)}/n_{j(k)}$ (solid blue line), and the normalized mean spectrum $\bar{f}_i$ (dashed red line). The bottom panels show the transmission $F_{ik}^{(0)}$ (thin, solid red line), the computed linear fit $L_{ik}$ (thin, dashed red line), and the corrected transmission $F_{ik}$ (thick, solid blue line). In the bottom right panel, the points indicate pixels that are excluded from the linear fit. A horizontal thick dashed green line at a transmission of 1 is included for visual aid.

spectrum and its errors,

$$\bar{F}_i = \frac{\sum_k F_{ik} w_{j(k)}}{\sum_k w_{j(k)}} \; ; \qquad \bar{E}_i^{-2} = \frac{\sum_k E_{ik}^{-2} w_{j(k)}}{\sum_j w_{j(k)}} \; . \tag{3.14}$$

The index $j(k)$ refers to the quasar index $j$ corresponding to each quasar-galaxy pair labelled by the index $k$. In these sums, any pixels over the interval from $-5000$ to $+5000\,\mathrm{km\,s^{-1}}$ with a normalized flux $F_{ik}$ below $-2$ or above $+3$ are eliminated, to exclude bad pixels or values that may have excessive noise. This eliminates only 0.02% of the pixels.

The whole procedure is illustrated in figure 3.4 with a couple of examples, one with an individually detected Mg II absorption system and one without any individually detected Mg II absorption but with a random metal absorption line. Note that the contribution of these random metal absorption lines is later corrected for (section 3.3.3 on the facing page).

### 3.3.2 Method 2: Variable smoothing

As a second approach to determine the continuum, we use the method described in York et al. (2005), based on an iterative smoothing of the spectrum with a variable smoothing length that is adjusted to decrease in spectral regions of known quasar emission lines. Each quasar spectrum is normalized by dividing both the flux and error by the variable smoothing continuum model, $c_{ij}$:

$$F'_{ij} = f_{ij}/c_{ij} \; ; \qquad E'_{ij} = e_{ij}/c_{ij} \; . \tag{3.15}$$

In the same way as for the first method, the normalized spectra are rebinned into the variable $\nu$ with bin width of $50\,\mathrm{km\,s^{-1}}$. The final composite spectrum is calculated using the same weights as in equation 3.9 on page 37, following equation 3.14 on page 40, but using the primed variables $F'_{ij}$ and $E'_{ij}$.

### 3.3.3 Unbiasing the composite spectra

After the stacking of all the normalized spectra of quasar-galaxy pairs as a function of the velocity variable $\nu$ has been completed by using either continuum fitting method, the mean value of the transmission $\bar{F}_i$ that is obtained far from the expected Mg II line (i.e., at large values of $\|\nu\|$) is close to unity but not exactly so. The main reason for this is the presence of random metal absorption lines (unrelated to the galaxy of the pair) that are detected above a $3\sigma$ fluctuation and therefore excluded when fitting the continuum. Other reasons may be affecting this mean background of $\bar{F}_i$ related to systematic effects in the noise distribution and the continuum fitting method. We eliminate this bias by performing the same linear fit to the stacked spectrum as in section 3.3.1 on page 38, using again the velocity intervals $(-5000, -2000)$ and $(2000, 5000)\,\mathrm{km\,s^{-1}}$ to measure two average values of $\bar{F}_i$, and obtaining a linear fit based on two points at the center of these intervals. The final normalized stacked flux is found by dividing by this linear fit.

Our results will actually be shown, for convenience, in terms of the effective excess optical depth, defined according to $\delta\tau_{e,i} = -\log \bar{F}_i$, in figures 3.5 to 3.7 on pages 45–47.

### 3.3.4 Bootstrap errors

The errors computed from the known observational noise in the observed quasar spectra that are stacked are actually a lower limit to the true errors. In reality, the intrinsic variability of real quasar continua and of the associated Mg II and other random metal absorption lines imply the presence of additional errors that are not taken into account and which are correlated among the pixels of the final stacks. We therefore compute bootstrap errors, which are generally used in our analysis and model fits in this chapter.

To calculate these bootstrap errors, we use the BOSS plates as the regions of the sky in which the sample is divided. Each quasar is tagged with the plate number at which its best spectrum was observed. For DR7 there are $N_{\mathrm{plates}} = 1822$ plates. Pairs are counted as belonging to the plate that contains the quasar, irrespective of the galaxy position.

Bootstrap errors are computed in the standard way, generating $N = 1000$ samples by randomly selecting $N_{\rm plates}$ among all the plates with repetition, and then recalculating the composite spectra using both methods. Bootstrap errors are assigned to the final effective optical depth $\tau_e$, at each velocity bin and each impact parameter interval from the dispersion found among the 1000 random samples. These bootstrap errors are computed for both continuum fitting methods.

We note that, specially at large impact parameter, some galaxies will be paired to quasars in different plates. This implies the presence of residual correlations among the bootstrap samples because of the common galaxies in pairs belonging to different plates, but we believe this effect is negligibly small because the most important error correlations should arise from the quasar spectra.

## 3.4 Model for the Absorption Profile

The usual analysis in the astronomical literature of individual Mg II absorption lines is done by fitting with Voigt profiles, with the equivalent width and the velocity dispersion as free parameters. Whenever the observed absorption profile is not adequately fitted in terms of the two Voigt profiles of the Mg II doublet, one can include the presence of multiple cloud components with blended absorption lines in order to improve the fit. Here, we are analysing a stack of a large number of Mg II absorption systems that may be mostly undetected individually, but for which we can accurately predict the expected mean position from the redshift of the galaxy near the quasar line of sight. The effective optical depth in the stack, $\tau_e = -\log(\bar{F})$, should in this case have a single symmetric component for each line in the doublet, with a profile reflecting the velocity dispersion of Mg II absorbing clouds and galaxies in halos, as well as the large-scale halo correlation in redshift space for large impact parameters. This cross-correlation function can be modelled in terms of the Halo Occupation Distribution formalism (e.g., Gauthier et al. 2008; Zhu & Ménard 2013a), but the density profile of Mg II clouds in halos does not have to follow that of galaxies, and it will generally depend on complex physics of galaxy winds and gas accretion in the circumgalactic medium. For simplicity, we shall assume in this section a model with a Gaussian velocity distribution and a power-law form for the projected correlation function, as an approximation to the generally complex form of the galaxy-absorber cross-correlation function. In the next section, we shall use a more accurate form of the correlation function obtained from halo simulations to determine the amplitude of our measured cross-correlation.

Our model has four free parameters. The first three are the central effective optical depth $\tau_0$ of the strongest line in the doublet at a conventional normalization value of the impact parameter $r_{p0}$ (set here to 1 Mpc), the power-law slope $\alpha$ of the projected cross-correlation, and the mean equivalent width ratio $q$ (where $q = 1/2$ if all the absorption lines were unsaturated, and $q = 1$ in the saturated case). We consider a variation of the velocity dispersion $\sigma$ with radius, taking into account that in the limit of large radius, the Hubble expansion should lead to a linear increase of the effective dispersion. The fitted profile of the excess effective optical depth, $\delta\tau_e(b,\nu)$ (as defined in equation 3.5 on page 35), is

$$\delta\tau_e(r_p,\nu) = \frac{\tau_0}{\sqrt{1+\left(xHr_p/\sigma_0\right)^2}} \left(\frac{r_p}{r_{po}}\right)^{-\alpha} \times$$

$$\times \left[e^{-(\nu-\nu_1)^2/2\sigma^2} + q\, e^{-(\nu-\nu_2)^2/2\sigma^2}\right] , \tag{3.16}$$

where

$$\sigma^2 = \sigma_0^2 + \left(xHr_p\right)^2 . \tag{3.17}$$

The fourth model parameter is therefore the dimensionless constant $x$, which is the scale at which the velocity dispersion starts to increase, compared to $\sigma_0/H$, with $H(z = 0.55) = 91.7$. In principle, the central velocity dispersion $\sigma_0$ should also be left as a free parameter, but in this chapter. we have fixed it to $250\,\mathrm{km\,s^{-1}}$. The reason is that if this condition is relaxed, the obtained fits have a large degeneracy in $\sigma_0$ and $x$ and they are largely dominated by an excess of absorption that is found at the largest impact parameters that we do not fully understand, as we shall see below. The value of $250\,\mathrm{km\,s^{-1}}$ allows us to fit well the width of the doublet line at small impact parameter, where it is resolved. Note that the parameter $q$ is assumed to be independent of impact parameter, even though it may generally depend on it (the mean saturation should decrease with impact parameter if the mean absorber equivalent width also decreases, as suggested by observations; see e.g. Gauthier et al. 2009; Chen et al. 2010a). Note also that the instrumental Point Spread Function (PSF) (Smee et al. 2013) and finite spectral bin size is effectively included in the value of the velocity dispersion in our model. A more detailed modelling of these effects is neglected here for simplicity. The origin of the velocity coordinate $\nu$ is chosen by convention as the central position of the Mg II doublet for an unsaturated line ($q = 0.5$) at the redshift of the galaxy. Under this convention, $\nu_1 = -256.05\,\mathrm{km\,s^{-1}}$ and $\nu_2 = 513.28\,\mathrm{km\,s^{-1}}$.

The fitting is performed by $\chi^2$ minimization, computed including all the pixels of the stacked spectra in the central interval $|\nu| < 2\,000\,\mathrm{km\,s^{-1}}$. Each pixel is weighted according to the bootstrap error measured in that pixel as described in section 3.3.4 on page 41, but without considering the cross-correlations of the errors between pixels. In practice, the bootstrap errorbars in different pixels of the stacked spectra are nearly equal. The spectrum outside the interval $|\nu| < 2\,000\,\mathrm{km\,s^{-1}}$, which is used for deriving the continuum fit in Method 1, is not considered here for the fit. We note also that for a real Gaussian profile for each component of the Mg II doublet, the continuum fitting of Method 1 implies that we formally need to subtract a small constant from the double Gaussian in equation 3.16, equal to the integrated value of the model absorption over the interval $\nu \in (2000, 5000)\,\mathrm{km\,s^{-1}}$ used for determining the continuum, but we neglect this effect here.

We use a Monte Carlo Markov Chain method to perform the $\chi^2$ minimization to the five-parameter model fit. The errors of the parameters are also obtained by repeating the model fit with bootstrap realizations. The average integrated equivalent width as a function of impact parameter can be obtained directly by integrating the effective optical depth over the interval used to fit the model,

$$W_e(r_p) = \frac{\lambda_{\mathrm{Mg\,II}}}{c} \int_{-2000\ \mathrm{km\,s^{-1}}}^{2000\ \mathrm{km\,s^{-1}}} \delta\tau_e(r_p, v)\, dv\ . \tag{3.18}$$

The fitted model also predicts a mean equivalent width,

$$W_e(r_p) = \frac{\lambda_{\mathrm{Mg\,II}}}{c} \sqrt{2\pi}\,\tau_0 \sigma_0 (1+q) \left(\frac{r_p}{r_{p0}}\right)^{-\alpha} = $$

$$= W_{e0} \left(\frac{r_p}{r_{p0}}\right)^{-\alpha}\ , \tag{3.19}$$

although this value is for the absorption over the whole velocity range, not restricted to the interval $-2000\,\mathrm{km\,s^{-1}} < v < 2000\,\mathrm{km\,s^{-1}}$.

## 3.5 Results

The results of the stacked absorption profiles are obtained for a total of 17 impact parameter intervals, measured in proper units at the redshift of the galaxy. The first interval is for $r_p < 50$ kpc, and the other 16 intervals are $2^{(i-1)/2} < (r_p/50\,\mathrm{kpc}) < 2^{i/2}$, for $i = 1$ to 16, up to a maximum impact parameter of 12.8 Mpc. The stacked profiles are shown as the effective optical depth, $\tau_e = -\log(\bar{F})$, in figures 3.5 to 3.7 on pages 45–47. Results are presented for our two continuum fitting methods, the mean subtraction method (thick, solid blue line) and the variable smoothing method (thin, solid red line). The errorbars plotted on the left side are the root mean square (RMS) value of the bootstrap error of $\tau_e$ in one pixel, which has little variation among different pixels in each stacked spectrum. The results of the fitted model parameters and their bootstrap errors computed by repeating the fit for bootstrap realizations of the stacked profiles, are given in table 3.1 on page 48 (the cross-correlation of the parameter errors are omitted).

The stacks in figures 3.5 to 3.7 on pages 45–47 show the mean absorption profile of the Mg II doublet line, clearly seen at the expected positions at small impact parameters. The amplitude of the random pixel-to-pixel variations outside of the central absorption feature is generally consistent with the computed bootstrap errorbars. The expected doublet feature of the Mg II line is generally resolved at $r_p \lesssim 200$ kpc, and is smoothed out at larger impact parameter by the velocity dispersion of the absorbers, which should increase linearly with $r_p$ in the limit of large radius because of Hubble expansion. The fact that our parameter $x$ is close to unity supports this interpretation, although we note that the precise expected theoretical value of $x$ in the linear regime is not one because of redshift distortions. We do not analyse this issue further in this chapter because the maximum impact parameter that we analyse is not yet in the linear regime, and a more detailed model would be necessary for the velocity distribution of absorbers as a function of impact parameter. We mention here that we initially considered a simpler model with $x = 0$ and

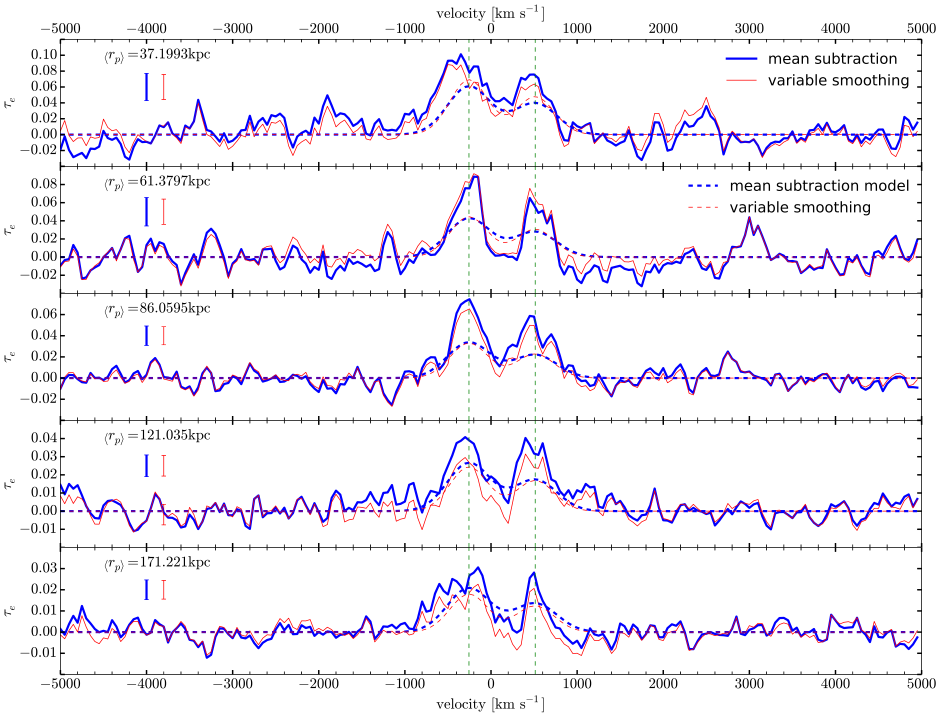


Figure 3.5: From top to bottom, composite spectra for increasing impact parameter intervals (in proper kpc). The effective optical depth is shown against velocity for the mean subtraction method ( thick, solid blue line) and the variable smoothing method (thin, solid red line). The RMS value of the bootstrap error in individual pixels is shown by the errorbars on the left. The thick, dashed blue line and the thin, dashed red line are the best fit model (equation 3.16 on page 43) for the mean subtraction and variable smoothing methods respectively. A single set of parameters are fitted to all the 17 regions. Figures 3.6 and 3.7 on the following page and on page 47 show the spectra for the remaining impact parameter intervals. The stacks show a mean absorption profile for the presence of the Mg II doublet line at the expected position. For visual guidance, vertical, dashed green lines mark the predicted position of the Mg II doublet.

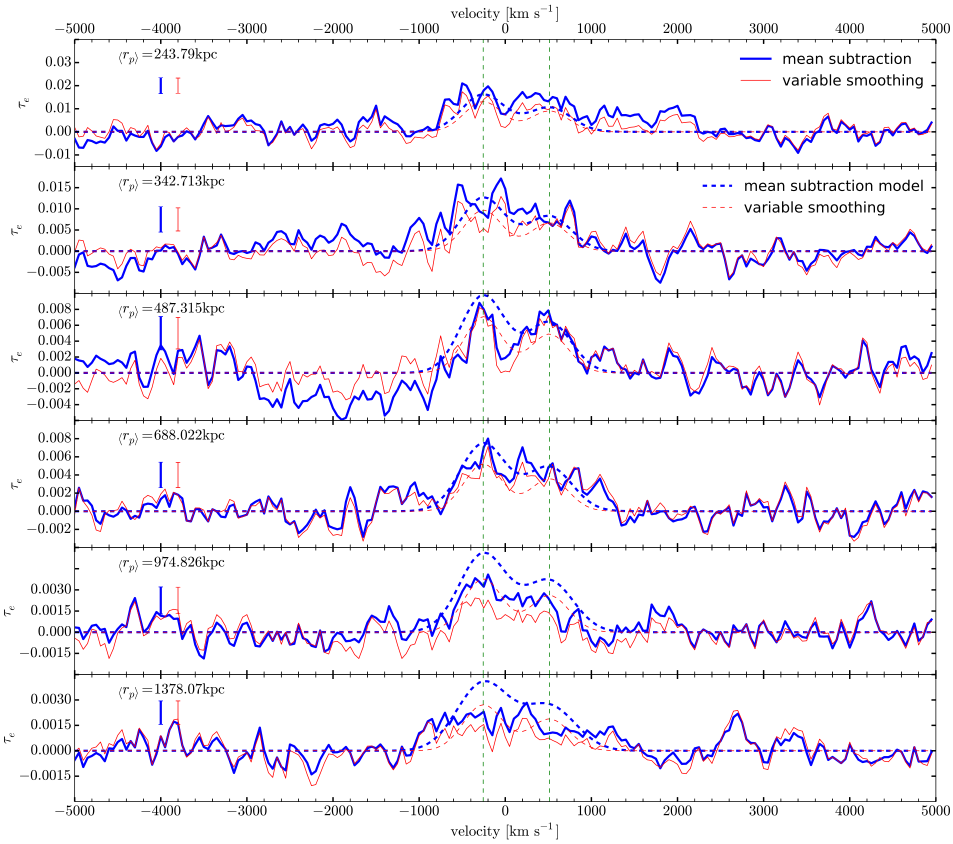


Figure 3.6: Continuation of figure 3.5 on page 45.

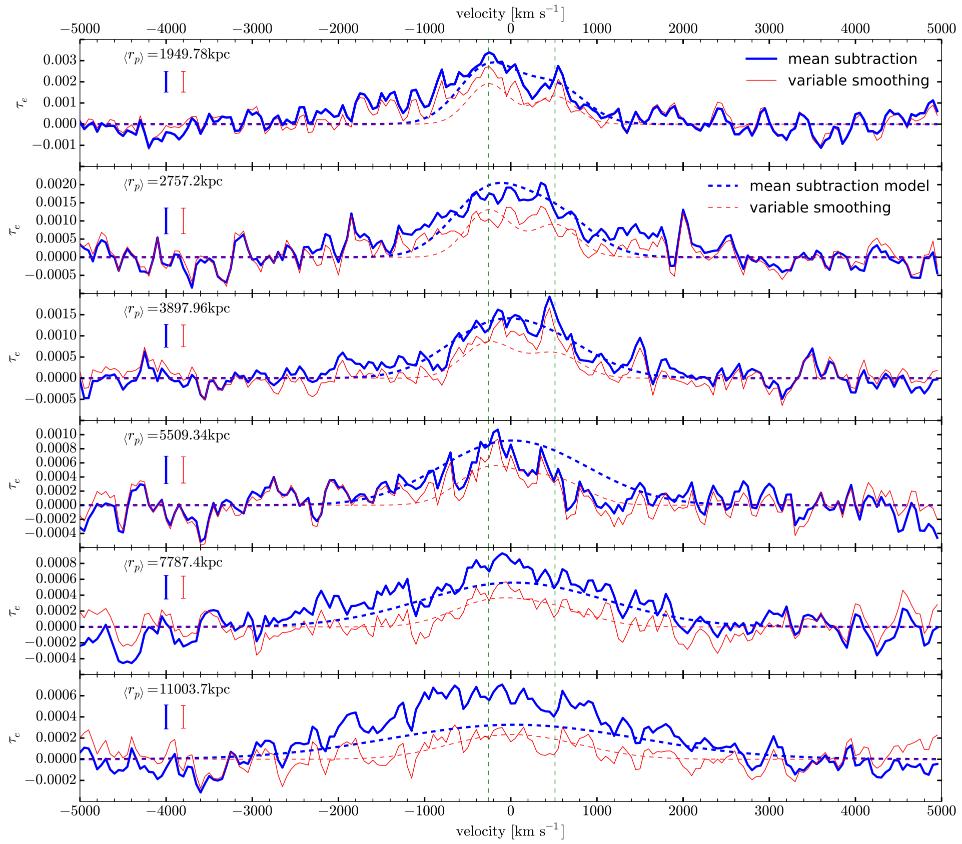


Figure 3.7: Continuation of figure 3.5 on page 45.

| | mean subtraction | variable smoothing |
|---|---|---|
| $\tau_0$ | $0.0060 \pm 0.0001$ | $0.0034 \pm 0.0001$ |
| $\alpha$ | $0.70 \pm 0.01$ | $0.88 \pm 0.01$ |
| $\sigma_0\,[\mathrm{km\,s^{-1}}]$ | 250 | 250 |
| $x$ | $1.35 \pm 0.06$ | $0.46 \pm 0.08$ |
| q | $0.65 \pm 0.03$ | $0.66 \pm 0.03$ |
| $W_{e0}[\mathrm{km\,s^{-1}}]$ | $6.27 \pm 0.10$ | $3.58 \pm 0.08$ |

Table 3.1: Best-fit values for the fitted parameters of the model described in section 3.4 on page 42 and shown as the solid lines in figure 3.8 on page 51. *First column:* Results using the mean subtraction method (see section 3.3.1 on page 35). *Second column:* Results using the variable smoothing method (see section 3.3.2 on page 41). Four independent parameters are fitted, and $W_{eo}$ is related to the other parameters according to equation 3.19 on page 44. As explained in section 3.4 on page 42, $\sigma_0$ is fixed at $250\,\mathrm{km\,s^{-1}}$. Errors are computed from repeating the fits on bootstrap realizations of the stacked profiles.

$\sigma_0$ as free parameter, but this choice gave a substantially worse fit and resulted in a high value of the velocity dispersion because of its increase with radius. As mentioned previously, fits leaving both $\sigma_0$ and $x_0$ as free parameters lead to a substantial degeneracy and to solutions with very large values of $\sigma_0$, driven by the excess of absorption that is seen in the last $r_p$ bin for the mean subtraction method. We have not found an explanation for this excess in the last bin, which is only marginally consistent (at the $\sim 3-\sigma$ level, as shown below in figure 8) with our fit.

A value $q \simeq 0.65$ is obtained for the line ratio of the Mg II doublet, consistent with a mixture of saturated and unsaturated lines. It has been previously reported that the mean equivalent width of the Mg II lines decreases with impact parameter (Gauthier et al. 2009, and references therein). This result may imply that absorbers are less saturated at larger impact parameters, and should therefore have a decreasing value of q, although this interpretation depends on whether the internal velocity dispersion of the absorbing clouds varies with impact parameter (note that this internal velocity dispersion is much smaller than the velocity dispersion of absorbing components around their host galaxies). Our model assumes a constant value of q for simplicity.

The mean equivalent widths obtained with our two methods, by directly integrating the effective optical depth in the stacked spectra as in equation 3.18 on page 44, are shown in figure 3.8 on page 51 as blue triangles for the mean subtraction method, and red squares for the variable smoothing method. These values and their bootstrap errors are also given in table 3.2 on page 50, together with the number of galaxy-quasar pairs that contribute to the stacked spectrum at each impact parameter bin. Note that this mean equivalent width is for the sum of the two lines in the Mg II doublet. There is a systematic difference in the mean equivalent width obtained with the two methods; the variable smoothing method yields a systematically smaller equivalent width compared to the mean subtraction method, and the discrepancy increases with impact parameter, reaching more than a factor 2 at $r_p = 10\,\mathrm{Mpc}$. As mentioned previously in section 3.3 on page 34, the reason for this difference is that the spectral region where Mg II absorption is expected is used to determine the continuum in the variable smoothing method. The presence of weak lines that remain undetected in individual spectra causes the continuum to be underestimated in a way that depends on the signal-to-noise ratio and the equivalent width of the undetected line in a complex manner. This systematic underestimate of the continuum causes the underestimate of the mean equivalent width. Appendix 3.A on page 60 presents quantitative tests demonstrating the presence of this systematic error of the variable smoothing method, and shows also that the result obtained with the mean subtraction method, which does not use the stacked spectrum in the region of the Mg II absorption to determine the continuum, is free of any similar systematic effect to the extent that we are able to discern.

The green circles with errorbars in figure 3.8 on page 51 show the results of Zhu et al. (2014), who have used a sample of galaxies and quasar spectra similar to ours to infer the same mean equivalent width as a function of impact parameter. Their result is systematically below ours, roughly by a factor $\sim 2$ at all impact parameters, and is lower even compared to our variable smoothing method. We believe the reason is again due to the systematic underestimate of the continuum. The continuum fitting method used by Zhu et al. (2014) also uses the observed spectra in the region where Mg II absorption is expected in a rather complex way that is described in Zhu & Ménard (2013b), and the systematic error that this induces is difficult to predict but may in principle explain why it produces a systematically low estimate of the equivalent width. Note that in principle this underestimation will only affect the individually undetected systems. Individually undetected systems cannot be distinguished from the noise and will thus be fitted away by the continuum fitter. Individually detected systems will, in principle, not suffer from this effect. The errorbars of Zhu et al. (2014) are also smaller than ours, since their continuum fitting can better remove any features of the quasar spectrum that are superposed with the Mg II absorption lines, at the cost of introducing a systematic bias in the continuum estimate.

## 3.6 Discussion

### 3.6.1 Relation of the mean equivalent width to the bias factor of Mg II absorption systems

Our measurement of the cross-correlation of Mg II absorption systems and galaxies in the CMASS catalog of BOSS clearly reflects properties of the spatial distribution of these two objects. In the limit of large scales, when the fluctuations are in the linear regime, any population of objects that traces the large-scale mass perturbations is characterized by its bias factor and the autocorrelation in real space is equal to the correlation function of the mass times the square of the bias factor, with redshift distortions added in redshift space (Kaiser 1987; Cole & Kaiser 1989). The cross-correlation of two classes of objects is, in the same large-scale limit, equal to the mass correlation function times the product of the two bias factors. On small, non-linear scales, the correlations are more complex and they depend on other physics that determine the distribution of galaxies and Mg II

| | mean subtraction | | variable smoothing | | |
|---|---|---|---|---|---|
| $r_p$ [kpc] | $< W >$ | $\sigma(< W >)$ [Å] | $< W >$ | $\sigma(< W >)$ [Å] | $N_{pairs}$ |
| $(0, 50]$ | 1.10 | 0.35 | 0.84 | 0.25 | 76 |
| $(50, 70.71]$ | 0.22 | 0.40 | 0.54 | 0.25 | 94 |
| $(70.71, 100]$ | 0.49 | 0.15 | 0.33 | 0.10 | 204 |
| $(100, 141.42]$ | 0.48 | 0.12 | 0.230 | 0.074 | 423 |
| $(141.42, 200]$ | 0.229 | 0.080 | 0.084 | 0.049 | 758 |
| $(200, 282.84]$ | 0.317 | 0.058 | 0.168 | 0.041 | 1396 |
| $(282.84, 400]$ | 0.181 | 0.076 | 0.102 | 0.042 | 2621 |
| $(400, 565.69]$ | 0.042 | 0.049 | 0.069 | 0.028 | 4904 |
| $(565.69, 800]$ | 0.081 | 0.029 | 0.069 | 0.019 | 9576 |
| $(800, 1131.37]$ | 0.040 | 0.018 | 0.006 | 0.012 | 19166 |
| $(1131.37, 1600]$ | 0.032 | 0.015 | 0.015 | 0.010 | 38922 |
| $(1600, 2262.74]$ | 0.0506 | 0.0093 | 0.0354 | 0.0068 | 75567 |
| $(2262.74, 3200]$ | 0.0360 | 0.0072 | 0.0211 | 0.0051 | 148769 |
| $(3200, 4525.48]$ | 0.0258 | 0.0062 | 0.0152 | 0.0038 | 290340 |
| $(4525.48, 6400]$ | 0.0119 | 0.0044 | 0.0073 | 0.0028 | 559840 |
| $(6400, 9050.97]$ | 0.0180 | 0.0036 | 0.0053 | 0.0022 | 1062482 |
| $(9050.97, 12800]$ | 0.0169 | 0.0028 | 0.0041 | 0.0020 | 1961450 |

Table 3.2: Results on the mean equivalent width and errors shown in figure 3.8 on the next page, presented here as a table. From left to right, impact parameter interval, mean rest-frame Mg II equivalent width and its bootstrap error for the mean subtraction and variable smoothing methods, and number of QSO-galaxy pairs used in the interval. The mean rest-frame equivalent widths are the sum of both lines in the doublet.

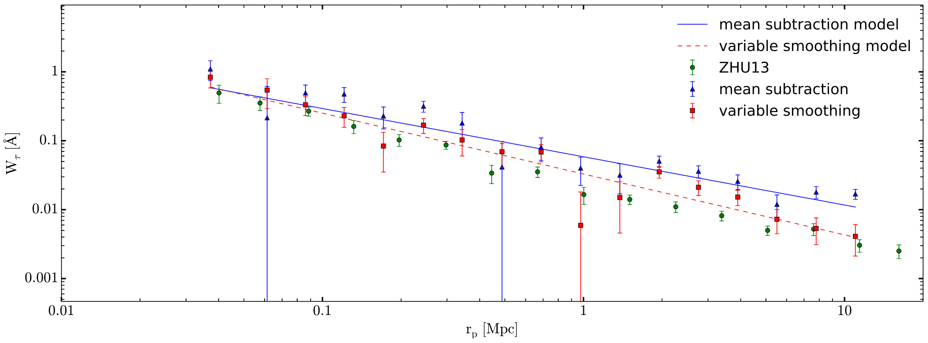


Figure 3.8: Measured rest-frame mean equivalent width $W_e$ of the Mg II doublet, versus proper impact parameter. Blue triangles are obtained from the mean subtraction continuum method, and red squares use the variable smoothing method. Error bars have been obtained by the bootstrap method. Lines are the best-fit power-law model to the data for both methods. Green circles are the results of Zhu et al. (2014). Note that the results for the mean subtraction method are systematically higher than the rest. This is explored in further detail in appendix 3.A on page 60.

absorbers in relation to dark matter halos.

Our stacked spectra measure the mean excess of the effective optical depth as a function of impact parameter $r_p$ and velocity separation $v$ from a galaxy. This quantity is related to the mean Mg II absorption perturbation, $\delta_{Mg}(r_p, v)$, as

$$\tau_e(r_p, v) = \tau_{eo}\left(1 + \delta_{Mg}(r_p, v)\right) \tag{3.20}$$

where $\tau_{eo}$ is the mean Mg II effective optical depth, defined in equation 3.4 on page 34. This perturbation is equal to the cross-correlation function of Mg II absorbers and galaxies, convolved with the mean doublet absorption profile of a Mg II system, and is the function that is measured in our stacking results in figures 3.5 to 3.7 on pages 45–47. In this work, our interest is focused on the projected correlation function, related to the integrated absorption $W_e(r_p)$ in equation 3.18 on page 44. The projected cross-correlation is not affected by redshift distortions and by the complications arising from the convolution with the mean doublet line profile and the spectrograph resolution.

Here, we shall make two approximations to physically interpret our measurement of $W_e(r_p)$: first, we neglect the effect of the finite integrating range $\pm 2000\ \mathrm{km\ s^{-1}}$ that we have used, ignoring the difference from the true projected correlation that is obtained by integrating to infinity. This approximation is likely not very good for the largest impact parameters we use; we discuss this further below. Second, we assume that the cross-correlation of Mg II systems and CMASS galaxies is the same as the autocorrelation of CMASS galaxies times the ratio of bias factors $b_{Mg}/b_g$ of the two types of objects. In other words, we assume the linear relation can be extended into the non-linear regime as far as the ratio of the cross-correlation to the autocorrelation is concerned.

This second assumption can be justified from observations of the correlations of galaxies of different luminosity. Zehavi et al. (2011) measured the projected correlation of galaxies in the DR7 catalogue in different luminosity ranges, and, to a good approximation, in the impact parameter range of our interest, the result is a fixed shape times the variable bias factor, as seen for example in their figure 6. The shape of the galaxy correlation can be interpreted as arising from the correlation among virialized halos and the distribution of galaxies within each halo (e.g. Zheng et al. 2005). This shape does vary slightly with luminosity, but the most important variation is the normalization determined by the bias factor. There is a greater variation of the shape of the projected correlation with galaxy color (see figure 21 in Zehavi et al. 2011). In addition, the projected cross-correlation of galaxies of different color is not exactly equal to the geometric mean of the projected autocorrelations of the two types of galaxies (see their figure 15). Our assumption can only be considered as a first approximation that will need to be tested in the future, but it allows us to obtain a bias factor for Mg II absorption systems assuming that they behave in a similar way as galaxies in the CMASS catalogue.

These assumptions lead to the relation

$$\begin{aligned} W_e(r_p) &= \frac{\tau_{eo}\lambda_{\mathrm{Mg\,II}}}{c}\int dv\,\delta_{\mathrm{Mg}}(r_p, v) = \\ &= \frac{\lambda_{\mathrm{Mg\,II}}}{c}\,\frac{\tau_{eo}\,H(z)}{1+z}\,\frac{b_{\mathrm{Mg}}}{b_g}\,w_{gg}(r_p)\,, \end{aligned} \tag{3.21}$$

where $w_{gg}(r_p)$ is the projected galaxy correlation function, $b_g$ is the galaxy bias factor and $b_{\mathrm{Mg}}$ is the mean bias factor of Mg II absorption systems, weighted in proportion to their equivalent width. We have used $dv = H(z)/(1+z)\,dx$, where $dx$ is the comoving space coordinate that is integrated to obtain the projected galaxy correlation function, and $z$ is the mean redshift of the galaxies and associated Mg II absorption systems. This relation allows us to infer the bias factor of Mg II systems empirically, using only the directly measured projected galaxy correlation. Its validity is strictly valid in the limit of large scales, but, as we shall see below, the ratio $W_e(r_p)/w_{gg}(r_p)$ is roughly constant, making our assumption plausible as a first approximation.

### 3.6.2 Mean absorption from Mg II systems

The value of $\tau_{eo}$, representing the average absorption from the population of Mg II absorbers, must be independently known before we can use the measured mean excess of Mg II absorption around galaxies to infer the bias factor of Mg II systems with equation 3.21. This parameter can be estimated from equation 3.4 on page 34 using models of the equivalent width distribution that fit the observational data.

We use the double exponential model of Nestor et al. (2005),

$$\frac{\partial^2 \mathcal{N}}{\partial W \partial z} = \frac{N^*_{\mathrm{wk}}}{W^*_{\mathrm{wk}}}\exp^{-W/W^*_{\mathrm{wk}}} + \frac{N^*_{\mathrm{str}}}{W^*_{\mathrm{str}}}\exp^{-W/W^*_{\mathrm{str}}}\,, \tag{3.22}$$

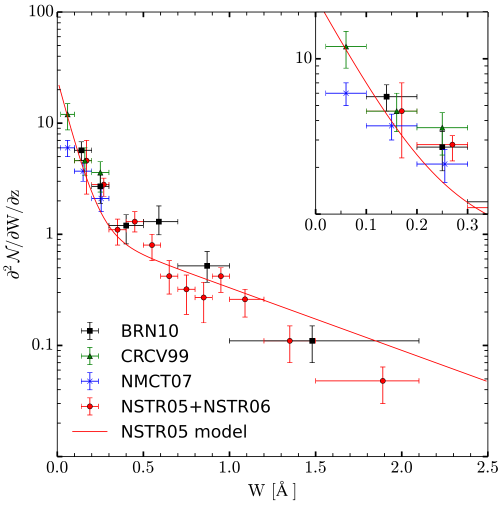


Figure 3.9: Rest-frame equivalent width distribution of Mg II absorption systems. Data points are from Bernet et al. (2010) (black squares), Churchill et al. (1999) (green triangles), Narayanan et al. (2007) (blue crosses) and Nestor et al. (2005, 2006) (red circles). The overplotted solid line is the double exponential fit of Nestor et al. (2005) (see text). Top right panel is a zoomed view of the weakest absorption systems.

where $\mathrm{N}^*_{\mathrm{str}}$ and $\mathrm{N}^*_{\mathrm{wk}}$ are the number of absorbers per unit of redshift in the strong and weak population, and $W^*_{\mathrm{str}}$ and $W^*_{\mathrm{wk}}$ are the characteristic rest-frame equivalent widths of the two exponential distributions. This model was fitted by Nestor et al. (2005) to their data, using mocks to correct for incompleteness at low equivalent widths. More recently, a compilation of high-resolution data was shown by Bernet et al. (2010) in their figure 5, reaching down to lower equivalent widths. We include these observational results in figure 3.9, overplotting the fit that was obtained by Nestor et al. (2005), which has the following parameters: $\mathrm{N}^*_{\mathrm{wk}} = 1.71 \pm 0.02$, $W^*_{\mathrm{wk}} = 0.072 \pm 0.001$Å, $\mathrm{N}^*_{\mathrm{str}} = 0.932 \pm 0.011$ and $W^*_{\mathrm{str}} = 0.771 \pm 0.014$Å. The observations are well reproduced by this fit, which we therefore use to compute $\tau_{\mathrm{eo}}$.

Unfortunately, the sample of absorbers of Nestor et al. (2005) is somewhat heterogeneous, and the main uncertainty we encounter in using it to compute $\tau_{\mathrm{eo}}$ is due to the redshift evolution. A fit to this evolution was determined by Nestor et al. (2005), where the parameters of the exponential model vary as $W^* \propto (1+z)^{0.634\pm0.097}$ and $\mathrm{N}^* \propto (1+z)^{0.226\pm0.170}$, both for the weak and strong populations. We infer from their model fits and the mean value of $\mathrm{N}^*$ that the mean redshift of their sample is $z \simeq 1.1$ and we use these relations to convert the product $\mathrm{N}^*W^*$ to the mean redshift of the CMASS galaxy catalogue, $z \simeq 0.55$. We find $\mathrm{N}^*_{\mathrm{str}}W^*_{\mathrm{str}} = 0.55$Å and $\mathrm{N}^*_{\mathrm{wk}}W^*_{\mathrm{wk}} = 0.095$Å, with an error that is close to 10%, although it is poorly defined because the errors in the redshift evolution of $\mathrm{N}^*_{\mathrm{str,wk}}$ and $W^*_{\mathrm{str,wk}}$ should be correlated, and this information (and the exact redshift distribution of the absorbers) was not provided in Nestor et al. (2005).

From equation 3.4 on page 34, we derive the value of $\tau_{eo}$ in the double exponential model of equation 3.22 on page 52,

$$\tau_{eo} = \frac{1+z}{\lambda_{\mathrm{Mg\,II}}} (1 + \bar{q}) \left( N^*_{wk} W^*_{wk} + N^*_{str} W^*_{str} \right) . \tag{3.23}$$

which yields a value $\tau_{eo}(z = 0.55) = 5.0 \times 10^{-4}$, with an error of about 10% but which is subject to uncertainties owing to the redshift evolution, the accuracy of the fit to the equivalent width distribution, and the value of $\bar{q}$. Our results will be given in terms of $\tau_{eo}$ without including its error, with the understanding that this quantity will need to be better determined in the future from studies of the field population of Mg II absorbers.

### 3.6.3 Derivation of the bias factor of Mg II systems

We now use equation 3.21 on page 52 to infer the bias factor of the Mg II systems, as

$$b_{Mg} = b_g \frac{c\, W_e(r_p)(1+z)}{\tau_{eo} \lambda_{\mathrm{Mg\,II}}\, w_{gg}(r_p)\ H(z)} . \tag{3.24}$$

Note that the factor $(1+z)/H(z)$ appears because of our convention that the effective equivalent width $W_e$ is measured in Å, whereas $w_{gg}$ is assumed to have been transformed to comoving length units. Instead of fitting our Mg II-galaxy cross-correlation measurement with a power-law dependence with impact parameter (as in equation 3.16 on page 43), we can directly fit the functional form that is determined from the observed projected galaxy correlation function, assuming that the shape is the same. To do this, we use the projected galaxy autocorrelation function obtained in Nuza et al. (2013) from the BOSS DR9 catalog of CMASS galaxies, and their prediction for the galaxy correlation function based on assigning galaxies to halos and subhalos in their MultiDark simulation. The measurements of Nuza et al. (2013) are represented as black triangles with errorbars in figure 3.10 on the next page, and their model is shown as the thick black line (given in their figure 6 and table B1; note that we have corrected for the different cosmological model they use, with a present matter density $\Omega_m = 0.27$ instead of our value $\Omega_m = 0.3$). Blue triangles, red squares and green circles are the mean equivalent width $W_e(r_p)$ times the factor $(1+z)/H(z)/\tau_{eo}$ (equal to the cross-correlation of Mg II systems and CMASS galaxies), times $r_p(1+z)$ for the mean subtraction method, the variable smoothing method and measurements from Zhu et al. (2014) respectively.

The galaxy bias factor in the model of Nuza et al. (2013) shown as the dashed black line is $b_g = 2.00 \pm 0.07$. Note that this value is lower than that obtained by Guo et al. (2013), $b_g = 2.16 \pm 0.01$, for the average CMASS galaxy. Using the value given by Guo et al. (2013) would lead to a larger measured value of the bias factor of the Mg II systems. As explained in Guo et al. (2013), the value of the galaxy bias increases with luminosity and redshift, and it can also depend on the range of scales used to fit its value. Here we use the galaxy bias value and the projected

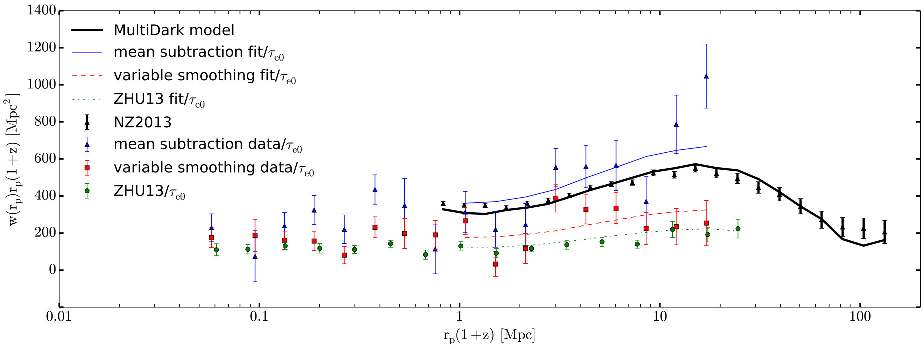


Figure 3.10: Projected correlation functions multiplied by the comoving impact parameter as a function of the comoving impact parameter $r_p(1+z)$. Thick black triangles with errorbars show the autocorrelation of CMASS galaxies from Nuza et al. (2013). Blue triangles, red squares and green circles are the mean equivalent width $W_e(r_p)$ times the factor $(1+z)/H(z)/\tau_{eo}$ (equal to the cross-correlation of Mg II systems and CMASS galaxies), times $r_p(1+z)$ for the mean subtraction method, the variable smoothing method and measurements from Zhu et al. (2014) respectively. The thick solid black line is the MultiDark model prediction described in Nuza et al. (2013). The solid blue, the red dashed and the green dashed-dotted lines are the fit to $W_e$ for the mean subtraction method, the variable smoothing method and measurements from Zhu et al. (2014) respectively. The ratio of the each of these lines with the thick solid black line is the ratio of bias factors, $b_{Mg}/b_g$.

galaxy autocorrelation function of Nuza et al. (2013), but the discrepancy with the higher value obtained by Guo et al. (2013) needs to be resolved to remove this source of uncertainty on the derived bias value of the Mg II absorption systems.

We now redo the fit to the measured $\delta\tau_e(r_p, \nu)$ profiles presented in section 3.5 on page 44, after replacing equation 3.16 on page 43 by

$$\delta\tau_e(r_p,\nu) = \frac{b_{Mg}\,\tau_{eo}}{b_g\sqrt{2\pi}(1+q)}\,\frac{w_{gg}(r_p)H(z)}{(1+z)\sqrt{\sigma_0^2+(xHr_p)^2}} \times$$

$$\left[e^{-(\nu-\nu_1)^2/2\sigma^2} + q\,e^{-(\nu-\nu_2)^2/2\sigma^2}\right] . \qquad (3.25)$$

That is to say, we replace the power-law dependence of $W_e$ on impact parameter by the model to the observed $w_{gg}$ of Nuza et al. (2013). All the parameters except for $b_{Mg}$ are kept fixed to the best fit obtained in section 3.5 on page 44, and a new fit is obtained by matching only $b_{Mg}$. The function $w_{gg}(r_p)$ is computed at our impact parameter bins by linear interpolation of the values of the model of Nuza et al. (2013). For convenience the fit is performed in the equivalent width space, namely,

$$W_e\,(r_p) = \tau_{e0}\frac{\lambda_{Mg\,II}}{c}\,\frac{b_{Mg}}{b_g}\,\frac{w_{gg}\,(r_p)\,H(z)}{1+z} \qquad (3.26)$$

The results we obtain for the Mg II absorption bias factor are

$$b_{\text{Mg mean subtraction}} = 2.33 \pm 0.19\,, \tag{3.27}$$

$$b_{\text{Mg variable smoothing}} = 1.14 \pm 0.36\,, \tag{3.28}$$

and the resulting fit for $W_e(r_p)r_p$ are shown as the solid blue line for the mean subtraction method and the dashed red line for the variable smoothing method in figure 3.10 on page 55 (these lines are simply the rebinned thick solid black line shifted by the factor $b_{Mg}/b_g$). The error on this bias factor includes only the uncertainty of the fit that assumes the bootstrap errors in our stacked spectra, and does not include the error of $\tau_{eo}$, assumed to be $\tau_{eo} = 5.0 \times 10^{-4}$. For completeness we also repeat the fit using Zhu et al. (2014) datapoints. The value we obtain is

$$b_{\text{Mg ZHU13}} = 0.78 \pm 0.05\,, \tag{3.29}$$

Note that the error here is computed not from the bootstrap errors but directly from the $\chi^2$ fitting instead. We neglect the fact that the datapoints are not independent.

We find a huge difference for $b_{Mg}$ for the different methods. Thus, we stress once again the importance of the quasar continuum estimate. Getting the estimate right is crucial for the measurement of the bias and failing to do so may lead to really different results.

Our measurement for the mean subtraction method is discrepant from the previously reported values by Gauthier et al. (2009) of $b_{Mg} = 1.36 \pm 0.38$, and by Lundgren et al. (2009) of $b_{Mg} = 1.10 \pm 0.24$. Our result is closer to the bias factor measured for the galaxies, implying that most of the Mg II systems are associated to massive galaxies like the CMASS ones or even more massive. On the other hand, the measurement for the variable smoothing method is compatible with the previous ones. Finally the measurement using Zhu et al. (2014) datapoints is also not compatible with the reported values but for the opposite reason. It is too low.

However, as we discussed in section 3.5 on page 44 and in the appendix 3.A on page 60, both the variable smoothing method and the method used in Zhu et al. (2014) underestimate the observed mean equivalent width. We have presented proof that it is indeed true for the variable smoothing case. We also argued that this is also likely to be true for the method used in Zhu et al. (2014). This means that the 'correct' value should be the $2.33 \pm 0.19$ obtained for the mean subtraction method even though it is not compatible with the values reported in Gauthier et al. (2009) and Lundgren et al. (2009).

Note that our measurement of the bias factor remains subject to systematic uncertainties that will need to be improved: the determination of $\tau_{e0}$, the use of a wider velocity interval for determining the quasar continuum and the mean Mg II absorption compared to the ones used in this chapter, and the use of a better modelling of the cross-correlation that includes redshift distortions in the regime of large impact parameters, and a more general density profile of Mg II absorbers in halos of different mass.

One possible reason for our high value of the Mg II bias is the degeneracy between $\tau_{eo}$ and the Mg II bias. What we are actually fitting is the product of both. We can only recover the Mg II bias once we fix the value of $\tau_{eo}$. This means that an underestimation of $\tau_{eo}$ will result in an overestimation of the Mg II bias. Thus, a more robust measurement of $\tau_{eo}$ is required. Another possible reason for our high value of the Mg II bias is that the bias may decrease with the equivalent width of the Mg II absorbers, as found both by Lundgren et al. (2009) and Gauthier et al. (2009). Our method includes all the Mg II absorbers and measures their average bias, weighting each absorber by its equivalent width. This average bias would be larger than the one for strong, individually detected Mg II lines if the weak absorbers are associated with massive halo environments, whereas absorbers of high equivalent width occur in galaxies hosted by low mass halos. Yet another possible explanation is our use of a limited velocity range for evaluating the projected cross-correlation of Mg II absorption and galaxies. We note that the high value of the bias we obtain is driven by the last point (at largest impact parameter) in figure 3.10 on page 55. This point might be too high because linear redshift distortions have increased the density of Mg II absorbers in the interval used for integration, and decreased them in the interval used for continuum fitting. The projected cross-correlation should not be affected by these redshift distortions when it is computed by integrating over the whole line of sight, but at the largest impact parameters our integrating intervals are probably not large enough. This systematic effect can only be addressed by improving the continuum fitting method and the model of the cross-correlation in future work.

We now compare our results for the Mg II-galaxy cross-correlation with those of Zhu et al. (2014). The modelling of Zhu et al. (2014) involves a free parameter, which they designate as $f_{\rm Mg\,II}$, that reflects the gas-to-mass ratio in halos (ignoring the degree of saturation of the absorption lines), and they assume a fixed density profile for the absorbers in halos. Zhu et al. (2014) do not relate this parameter to the mean absorption of the field population of Mg II absorbers, so they fit the amplitude of the cross-correlation with this unconstrained, free parameter. They determine a characteristic host halo mass for the Mg II absorbers of $M_h \simeq 10^{13.5}\,M_\odot$ based on the presence of a feature in the shape of the cross-correlation at $r_p \sim 1$ Mpc that reflects a transition from the 1-halo term to the 2-halo term in their modelling. The weakness of this feature implies a poor determination of this characteristic halo mass (see the contours in their figure 6, showing the large degeneracy with the $f_{\rm Mg\,II}$ parameter). Their determination of this halo mass is therefore highly dependent on their model of the halo density profile, and does not generally match the total observed abundance of Mg II absorbers. Moreover, the Mg II absorbers are likely hosted in halos with a very broad mass range, which should cause a smoothing of any feature due to the transition from the 1-halo to the 2-halo term. The specific density profile of absorbers they assume has not been tested and is not theoretically well motivated. Mg II absorbers can be distributed in halos differently from galaxies, depending on the physics of gas cooling and galactic winds in halos. autocorrelations and cross-correlations of galaxies of different types have been found to have widely different shapes (see, e.g., figures 15 and 16 in Zehavi et al. 2011) which do not always possess the clear feature that is predicted for a tracer that follows the dark matter profile in halos of a specific mass. Therefore, we think there is insufficient evidence for the presence of any feature in the Mg II-galaxy cross-correlation that may be used to determine a characteristic host halo mass.

Instead, we propose that the *amplitude* of this cross-correlation should be related to the field population of Mg II absorbers (e.g., through the effective optical depth $\tau_{eo}$ that we have introduced), and can then be used to determine the mean bias factor $b_{\mathrm{Mg\,II}}$, which should be equal to the mean bias factor of the host halos of the absorbers (weighted by their rest-frame equivalent width), and is robustly defined even if the range of host halo masses is very broad.

### 3.6.4 The ratio of Mg II-absorbing gas to the total mass

Measurements of the average Mg II absorption around galaxies can be compared with mass measurements averaged in the same way obtained from weak gravitational lensing. This comparison was done in Zhu et al. (2014) to obtain an estimate for the ratio of gas mass to total mass in the halos around the CMASS galaxies of the BOSS survey. We now examine this question to point out a number of uncertainties in this derivation.

In general, the total column density of Mg II in an individual absorber, $N_{\mathrm{Mg\,II}}$, is related to its integrated optical depth according to

$$N_{\mathrm{Mg\,II}} = \frac{m_e c^2}{\pi e^2} \frac{W_\tau}{f \lambda^2_{\mathrm{Mg\,II}}} = 1.13 \times 10^{20} \frac{W_\tau \text{Å}}{f \lambda^2_{\mathrm{Mg\,II}}} \ \mathrm{cm}^{-2} \, . \tag{3.30}$$

where $e$ and $m_e$ are the electric charge and mass of the electron, and $f = 0.921$ is the total oscillator strength of the Mg II doublet. The integrated optical depth is

$$W_\tau = \int d\lambda \tau(\lambda) \, , \tag{3.31}$$

where the integration is performed over a wavelength range that includes the entire absorption profile. However, the only quantity that is observed is the equivalent width,

$$W = \int d\lambda \left[1 - e^{-\tau(\lambda)}\right] \, . \tag{3.32}$$

When the optical depth of the absorber is much less than unity over the whole wavelength range, the absorber is unsaturated and $W_\tau \simeq W$. Otherwise, the column density is not directly measurable simply from the equivalent width. We now define an average saturation level for the population of absorbers, $\bar{S}$, as

$$\bar{S} = \frac{\int dW \, (\partial^2 \mathcal{N}/\partial W \partial z) \, W_\tau}{\int dW (\partial^2 \mathcal{N}/\partial W \partial z) \, W} \, . \tag{3.33}$$

Defining also $x_{\mathrm{Mg\,II}}$ as the fraction of magnesium atoms in the absorbing gas that are in the Mg II ionized species, $g_{\mathrm{Mg}}$ as the fraction of magnesium in the absorbers that is in the gas phase (as opposed to dust grains), and $Z_{\mathrm{Mg}}$ as the magnesium mass fraction compared to that of the

Sun (we use a solar magnesium abundance by mass of $7.0 \times 10^{-4}$, and a magnesium mass $m_{\mathrm{Mg}} = 4.07 \times 10^{-23}$ g), we find that the total gas mass surface density in the Mg II absorbers is

$$\Sigma_g(r_p) = 9.15 \times 10^{-7} \frac{\bar{S}}{x_{\mathrm{Mg\,II}}\, g_{\mathrm{Mg}}\, Z_{\mathrm{Mg}}} \frac{W_e(r_p)}{\text{Å}}\, \mathrm{g\,cm^{-2}}\,. \tag{3.34}$$

The total mass surface density around a CMASS galaxy in the BOSS sample has also been measured using weak gravitational lensing. We can therefore obtain the ratio of gas in Mg II absorbers to the total mass by combining the two observational measurements. We use the recent measurement by Miyatake et al. (2015) based on weak lensing measurements in the CFHTLenS survey area. As an example, we compute the gas-to-mass ratio at a comoving projected radius of $r_p(1+z) = 3$ Mpc. The differential surface density measured by Miyatake et al. (2015) at this radius is $\Delta\Sigma = \bar{\Sigma} - \Sigma \simeq 2\,\mathrm{M_\odot\,pc^{-2}}$ (see their figure 7). Near this radius, $\Delta\Sigma(r_p)$ is falling with radius roughly as $r_p^{-1}$, so the mean surface density within $r_p$ is $\bar{\Sigma} \simeq 2\Sigma$, and we can therefore use $\Sigma \simeq \Delta\Sigma$. At this same radius, $W_e \simeq 0.03$Å, and substituting this value in equation 3.34, it produces $\Sigma_g/\Sigma \simeq 8 \times 10^{-5}\bar{S}/(x_{\mathrm{Mg\,II}}\, g_{\mathrm{Mg}}\, Z_{\mathrm{Mg}})$. Using the mean ratio of baryons to total matter in the universe of $\Omega_b/\Omega_m = 0.17$, the fraction of baryons in the Mg II absorbing gas would be $\Sigma_g/\Sigma_b \simeq 5 \times 10^{-4}\bar{S}/(x_{\mathrm{Mg\,II}}\, g_{\mathrm{Mg}}\, Z_{\mathrm{Mg}})$.

Therefore, the fraction of baryons in the Mg II clouds can be a small one even if the mean metallicity is relatively low. However, the mean saturation parameter $\bar{S}$ is likely much larger than unity, so it is possible that the Mg II absorbers account for an important fraction of the baryons in galactic halos, and for the accreting material that fuels the star formation rate. We note that any further comparison of the detailed radial profiles of $W_e(r_p)$ and $w_{gg}(r_p)$ cannot be reliably used to infer a profile of the gas-to-mass ratio, because $Z_{\mathrm{Mg}}$ is likely to vary with $r_p$, since the heavy elements must have originated from galactic winds, and the values of $\bar{S}$, $x_{\mathrm{Mg\,II}}$ and $g_{\mathrm{Mg}}$ may also vary substantially with $r_p$. The mean value of the gas-to-mass ratio is still highly uncertain because of the unknown value of $\bar{S}/(x_{\mathrm{Mg\,II}}\, g_{\mathrm{Mg}}\, Z_{\mathrm{Mg}})$.

## 3.7 Summary and conclusions

In this chapter we have used the Mg II line to measure the cross-correlation of Mg II absorption and galaxies in BOSS. The large size of the samples we use (SDSS DR7 quasar catalogue as background sample and DR11 CMASS galaxy catalogue as foreground sample) enables a statistical approach to detect Mg II absorption that is too weak to be detected individually and would otherwise be missed. We present a method to estimate the quasar continuum designed for this type of measurements and compare our results with those obtained by a more typical continuum estimate. Our main results can be summarized as follows:

- The method to fit the quasar continuum is crucial to measure the mean Mg II equivalent width as a function of impact parameter. Methods that use the observed flux in the spectral region near the Mg II line wavelength at the galaxy redshift to determine the continuum suffer from a systematic bias, because the absorption from individually undetected systems inevitably lowers the continuum estimate and causes an underestimate of the mean absorption. The tests presented in appendix 3.A show that our mean subtraction method does not suffer from any systematic effect to the extent that we are able to discern.

- We find that the cross-correlation of Mg II absorption and CMASS galaxies follows the shape of the CMASS galaxies autocorrelation at large scales. We use the CMASS autocorrelation model from Nuza et al. (2013) and the measured galaxy bias factor to derive a bias factor of Mg II absorbers of $b_{Mg} = 2.33 \pm 0.19$. This value is substantially larger than the previous measurements by Gauthier et al. (2009) and Lundgren et al. (2009). This discrepancy may be due to a real difference, because our measurement includes a contribution from weak Mg II absorption systems which may be more strongly clustered than strong absorbers, and may also be affected by our imperfect determination of the projected cross-correlation at large impact parameters owing to our limited integrating range. More accurate measurements and better modelling will be necessary to clarify this question.

## 3.A Tests of the continuum fitting methods

The method to fit the quasar continuum is a crucial part of the measurement of the mean Mg II absorption equivalent width as a function of impact parameter from a galaxy, $W_\tau(b)$, presented in this chapter. The two methods we have used produce a different result, which is also different from the result reported by Zhu et al. (2014). It is therefore important to perform tests on these methods that can reveal the presence of systematic errors in the $W_\tau(b)$ estimates. This appendix presents the results of three tests. The first one (section 3.A.1) checks for any systematic mean absorption that might be artificially introduced by the continuum fitting method when there is no correlation between Mg II absorbers and galaxies. The second one (section 3.A.2 on the facing page) verifies that the correct equivalent width of an individually detected Mg II absorption system in a spectrum is correctly recovered. Finally, section 3.A.3 on page 62 reveals the effect on the fitted continuum of the presence of weak absorbers that are individually undetected, and the way these absorbers can bias the estimate of the mean equivalent width.

### 3.A.1 Systematic errors in the absence of correlations

One might suspect that a small average absorption (either positive or negative) is artificially introduced by the method to fit the continuum, even when the regions selected to search for absorbers are completely random and should therefore have no average absorption. This might happen if the quasar continuum is systematically overestimated or underestimated, depending in a complex manner on the varying noise properties and the shape of the true quasar continuum. To test for this possibility, we have remeasured the mean Mg II equivalent widths after rotating the right ascension coordinate of all the quasar by 5 and 10 degrees, in the two possible directions, and after increasing and decreasing the redshift of the galaxies by 0.05. These separations are large enough to make any residual cross-correlation of galaxies and Mg II absorbers completely negligible, so the measured correlation should be consistent with zero. Note that this procedure ensures that the autocorrelations that are present among the Mg II absorbers, quasars and galaxies are preserved, so their contribution to the measurement errors of the cross-correlation is the same.

The result of this test is shown in figure 3.11 on the following page, in the left panel for the mean subtraction method and the right panel for the variable smoothing method. The real data set is shown as big red circles with errorbars, after dividing by the best-fit power-law model that is described in section 3.4 on page 42 and plotted in figure 3.8 on page 51. The average of 6 mock data sets are shown also after dividing by the same model as blue stars. In the absence of any systematics, the mean absorption in the mock data sets is expected to be zero, whereas the real data should produce a ratio to the best-fit model that is consistent with one. The results do not show any systematic errors for the mean subtraction method. There appears to be a small negative systematic absorption that is introduced by the variable smoothing method, as indicated by the negative values of the mock sample in the right panel, at large impact parameters (where the mean equivalent width can be measured with the smallest errorbar). This average negative absorption is approximately equal to the best-fit model prediction at $\sim 10\,\mathrm{Mpc}$, or $\sim 0.3\,\mathrm{km\,s^{-1}}$ (see figure 3.8 on page 51). This result may be due to some subtle effect in the variable smoothing method that introduces a small bias by systematically underestimating the continuum in the absence of real absorption lines, possibly due to occasional false identifications of noise spikes as real absorbers. As we shall see below, the variable smoothing method is actually affected by a more serious systematic error that partially eliminates the contribution of weak absorption systems to the mean equivalent width.

### 3.A.2 Tests of the equivalent width measurement for individually detected systems

We now test that the mean equivalent width measured for absorbing systems that are individually clearly detected above the noise agrees with other well established methods. For this purpose, we use the Mg II absorber catalogue of Zhu & Ménard (2013b), which contains $35,752$ absorption systems from the SDSS DR7 quasar spectra sample. The integrated equivalent widths we obtain for systems in this catalogue, using our two methods of mean subtraction and variable smoothing, are compared with the equivalent widths provided in the catalogue in figure 3.12 on page 63. There is a large scatter in the equivalent widths obtained with different methods. This result is not surprising, because the noise can change the determination of the continuum in random ways in different methods. In particular, in the method of the mean subtraction, the equivalent width is obtained by integrating the absorbed fraction over a wide interval around the absorber, according to equation 3.18 on page 44, adding noise to the estimate. However, the average of the equivalent width estimator in our mean subtraction method, shown by the black points in figure 3.12 on page 63 (with an RMS dispersion indicated by the errorbars), agrees very well with

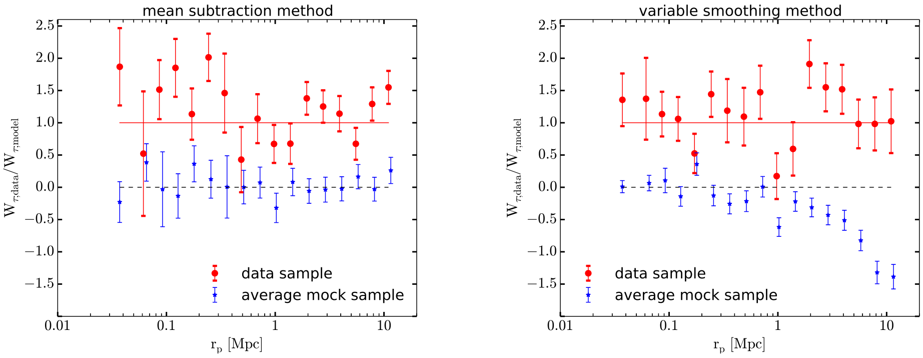


Figure 3.11: Ratio of the mean equivalent width in stacked spectra to the best-fit power-law model prediction, for the real data sample (big red circles), and for the average of 6 mock samples (blue stars). Errorbars are computed with the bootstrap method. This ratio should be consistent with zero for the mock sample in the absence of systematic errors, and with unity for the real data if the power-law model provides a good fit to the data.

the equivalent width provided by the Zhu & Ménard (2013b) catalogue. The variable smoothing method apparently suffers from a bias causing a 10 to 20% increase of the average of the equivalent width (see black points in middle panel of figure 3.12 on the next page), which may be due to a tendency of this method to overestimate the continuum level around detected absorption lines.

### 3.A.3 Impact of individually undetected systems on the mean equivalent width

We now test how the presence of weak absorption systems that cannot be individually detected in a single spectrum but contribute to the mean equivalent width as a function of impact parameter from a galaxy can bias the estimate of the quasar continuum in the different methods we use. For this purpose, artificial absorbers are introduced in a spectrum, and then we refit the continuum and measure the change in the measured equivalent width.

As an illustrative example, we have selected a set of 10 random quasar spectra, and we have introduced absorbers at 10 random redshift values and computed the average values for the recovered width. The absorbers are inserted with a double Gaussian profile in the optical depth, as expected for the Mg II doublet,

$$\tau(\nu) = \frac{W_0}{\sqrt{2\pi}\sigma}\left(\frac{2}{3}\exp\left[-(\nu-\nu_1)^2/2\sigma^2\right] + \right.$$
$$\left. +\frac{1}{3}\exp\left[-(\nu-\nu_2)^2/2\sigma^2\right]\right) , \tag{3.35}$$

where $W_0$ is the total equivalent width of the doublet and $\sigma$ is velocity dispersion. The zero velocity is conventionally chosen to be the central position of the Mg II line for an unsaturated line, so

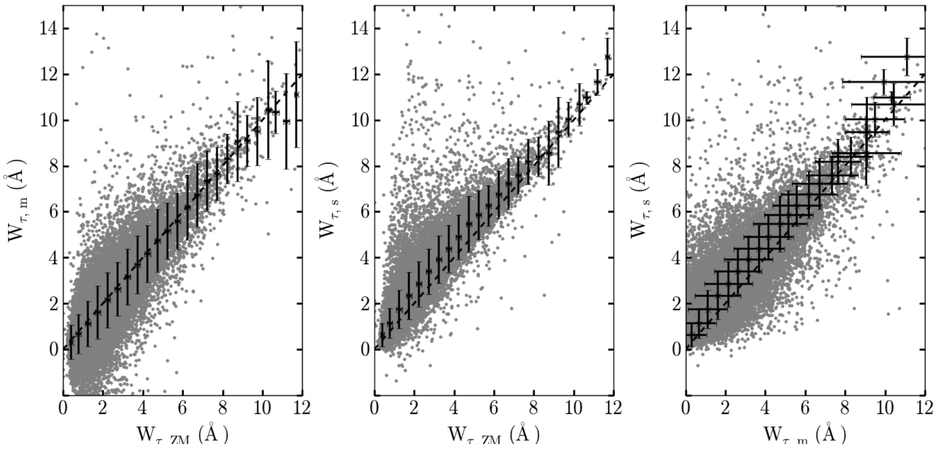


Figure 3.12: Comparison of the equivalent width estimates from different methods (see section 3.3 on page 34). From left to right, equivalent widths measured using the mean subtraction method, $W_{\rm m}$, versus the value of the catalogue from Zhu & Ménard (2013b), $W_{\rm ZM}$; variable smoothing method value, $W_{\rm s}$, versus $W_{\rm ZM}$; and $W_{\rm s}$ versus $W_{\rm m}$. Black points are the average values in bins of $\Delta W = 0.5$Å in the horizontal axis, with the dispersion in each bin indicated by the errorbars. One-to-one correspondence is marked by the black dashed line for visual guidance.

that $v_1 = -256.05\,{\rm km\,s^{-1}}$ and $v_2 = 513.28\,{\rm km\,s^{-1}}$. In the absence of any inserted absorber, the continuum determined in this spectrum is $c(v)$, the flux is $f(v)$, and the transmitted fraction is $F(v) = f(v)/c(v)$. This results in a certain value of the integrated equivalent width $W$ measured over the interval $-2000\,{\rm km\,s^{-1}} < v < 2000\,{\rm km\,s^{-1}}$, by integrating $F(v)$ over this range. To insert the absorber, the spectral flux is modified according to

$$f'(v) = f(v) - c(v)\left(1 - \exp\left[-\tau(v)\right]\right) \ . \tag{3.36}$$

Then, a new continuum $c'(v)$ is determined with the new flux, and a new transmitted fraction $F'(v) = f'(v)/c'(v)$ is derived. Finally, the new equivalent width $W'$ is determined by integrating $F'$ over the same velocity interval.

The change in equivalent width caused by the insertion of an absorber, $\Delta W = W' - W$, is plotted in figure 3.13 on the following page as a function of the equivalent width $W_i$ of the inserted absorber, obtained by integrating $1 - \exp[-\tau(v)]$ over the same velocity interval that is used for determining $W$ and $W'$ ($W_i$ is nearly equal to $W_0$ in equation 3.35 on page 62, except that the integrating interval does not extend to infinity). The different lines correspond to different values of the absorber velocity dispersion, $\sigma$. The solid blue lines represent the mean subtraction method, and they coincide precisely with $\Delta W = W_i$ for all values of $\sigma$. The result, as expected, is that the continuum determined by this method is unaffected by the presence of the absorbers that have

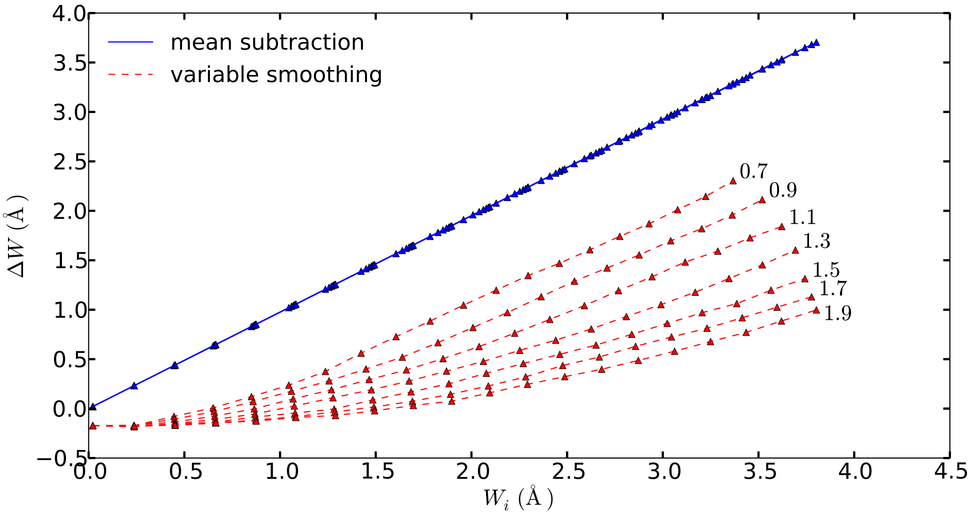


Figure 3.13: Change in the measured equivalent width, $\Delta W$, caused by the insertion of an absorbing system with equivalent width $W_i$, for the mean subtraction method (solid blue lines) and the variable smoothing method (dashed red lines). Results are shown as a function of the velocity dispersion $\sigma$, with values indicated to the right of the lines in Å. The blue lines all nearly coincide at $\Delta W = W_i$. Points show the values of $W_i$ for which $\Delta W$ has been computed; the sudden changes in the red lines indicate discontinuities in the variable smoothing method as the inserted line becomes detected or covers different pixels, which causes a change in the continuum estimate.

narrow widths compared to the chosen integrating interval width of 4000 km s$^{-1}$. The reason is that, in the mean subtraction method, the continuum $c'(\nu)$ is determined using only the measured flux outside of this interval.

On the other hand, the variable smoothing method (dashed red lines in figure 3.13) is strongly biased to lowering the estimated continuum in response to the presence of a weak absorption system. The result is that the change caused by the absorber in the estimated equivalent width, $\Delta W$, can be much less than the true value, and the difference is a complex function of the equivalent width $W_0$, the velocity dispersion and the signal-to-noise ratio of the spectrum. The underestimate of the continuum level is naturally smaller for narrower lines (lower $\sigma$), because the lines are detected and eliminated from the continuum estimation for lower values of $W_0$.

To summarize, the three tests of our continuum fitting methods presented in this appendix demonstrate that the variable smoothing method suffers from several systematic errors. The first test shows that a small negative absorption, of equivalent width $\sim -0.3$ km s$^{-1}$, is induced where there is none. The second test indicates that the equivalent width of strong, detected systems is overestimated by $10 - 20\%$. Finally, the third test shows that for weak systems, the continuum is systematically underestimated, thereby strongly reducing the contribution of these systems to the measured equivalent width. However, the mean subtraction method successfully passes all these tests and should therefore provide a reliable estimate of the stacked equivalent width as a function of impact parameter.

# 4

# DLA-Ly$\alpha$ forest cross-correlations

We present an update on the measurement of the DLA bias from the cross-correlation of DLAs and the Ly$\alpha$ forest (Font-Ribera et al. 2012). In our analysis, we use of the final DR of BOSS. We find the bias of the DLAs to be $b_d = 1.87 \pm 0.05$, which is smaller than previously reported. The lower value of the bias is yet to be fully understood. Several different reasons, or a combination of them, might be responsible for the decrease. With the improved statistics, we confirm that the DLA bias is scale independent. Also, we explore the evolution of the DLA bias with column density and redshift. We find no clear evidence of any evolution in neither of them. This means that the dark matter halos hosting the DLAs are essentially the same independently of the properties of the DLAs.

## 4.1 Introduction

DLAs are defined as absorption systems with neutral hydrogen column density $N_{\rm HI} \geqslant 2 \times 10^{20}$ cm$^{-2}$ Wolfe et al. (1986). The limit on $N_{\rm HI}$ is not only an arbitrary cut, the DLAs have two characteristics that distinguish them as a separate population of absorbers: *(i)* their hydrogen is mostly in atomic form since they are self-shielded of the external cosmic ionizing background radiation (Vladilo et al. 2001), and *(ii)* the damped profile of their hydrogen Ly$\alpha$ line is distinguishable even in low resolution spectra, which. means that the column density can be measured directly from the absorption profile. For a detailed review on the DLAs and their properties see e.g Wolfe et al. (2005); Barnes et al. (2014).

DLAs are, then, a powerful probe of the physical gas that has condensed into high density. Measurements of their column density distribution suggest that the mean density of baryons contained in these systems, which account for a fraction of the critical density, is about $\Omega_{\rm DLA} \simeq 10^{-3}$ at redshifts $2 < z < 3.5$ (Péroux et al. 2003; Prochaska et al. 2005; Zafar et al. 2013; Crighton et al. 2015; Padmanabhan et al. 2016; Sánchez-Ramírez et al. 2016). This quantity (~ 2% of all baryons in the universe) is comparable to the fraction of baryons inside stars at these redshifts (Prochaska et al. 2005; Prochaska & Wolfe 2009; Noterdaeme et al. 2009, 2012). Thus, these absorption systems provide reservoirs of atomic gas clouds which were available for the formation of galaxies.

The study of the DLAs and their properties can be used to provide powerful constraints on the galaxy formation models. As such, detailed studies of the metal absorption lines of DLAs have been explored for a long time. The metallic abundances derived from these studies reveal that these systems have generally very low metallicities, distributed over a range of $10^{-3}$ to $10^{-1} Z_{\odot}$ (Prochaska & Wolfe 2002; Prochaska et al. 2003b; Kulkarni et al. 2005; Ledoux et al. 2006). This distribution is not Gaussian, DLAs at higher redshift present, on average, lower metallicities (Kulkarni & Fall 2002; Vladilo 2002; Prochaska et al. 2003a; Calura et al. 2003; Khare et al. 2004; Akerman et al. 2005; Kulkarni et al. 2005; Rafelski et al. 2012, 2014; Jorgenson et al. 2013; Neeleman et al. 2013; Møller et al. 2013). The mean metal equivalent width for metal absorption tend to increase with column density but the increase is different for low- and high-ionization species (Mas-Ribas et al. 2016).

Kinematic studies of metal absorption lines in high-resolution spectra of DLAs reveal a diversity of velocity structures within these absorbers. In most cases several components are seen over a velocity range of $\sim 100\,{\rm km\,s^{-1}}$ (Prochaska & Wolfe 1997, 1998; Wolfe 1998). In particular, the velocity profiles of low- and high-ionization species are different, which suggests that there is more than one phase in the DLA gas (Wolfe & Prochaska 2000; Prochaska & Wolfe 2002; Fox et al. 2007a,b). What is more, the complex velocity profiles suggest a highly turbulent environment probably arising from collisions with other DLAs. As a result of these collisions, a complex variety of gas phases is believed to arise in the DLAs, leading to a broad range of temperatures and densities. Even so, models of gaseous galactic halos can generally explain these complex velocity profiles (see e.g. Haehnelt et al. 1998; McDonald & Miralda-Escudé 1999; Fumagalli et al. 2011; Cen 2012; Rahmati & Schaye 2014; Bird et al. 2015; Neeleman et al. 2015).

One may think that with all these observational constraints we now have a clear understanding of the precise nature of these DLAs. However, this is not the case, the association of DLAs and galaxies is yet to be understood. In particular, we do not know neither the mass of their host halos, nor the type of galaxies they are associated with. It is also possible that different types of DLA are associated with different types of galaxies. To characterize the population of halos that are hosting the DLAs one can resort to clustering analysis. In the limit of large scales, where linear theory holds, the correlation function of any population of objects that trace the primordial mass perturbations is proportional to the matter autocorrelation (Mo & White 1996), the bias factor of the population being the square root of the proportionality constant. It is now well understood that the bias factor of the halos increases with their mass both analytically (see e.g. Sheth & Tormen 1999) and from accurate predictions from simulations (see e.g. Tinker et al. 2010).

Being large clouds of gas, DLAs are bound to follow the gravitational potential of matter halos. Thus, they are tracers of the underlying dark matter distribution. Hence the autocorrelation of DLA absorbers follows $\xi_d(r) = b_d^2 \xi_m(r)$, where $b_d^2$ is the square of the linear bias factor of the DLAs. The bias factor gives an insight of the clustering of DLAs. The higher the value of the bias, the more clustered the objects are.

This means that we can measure the DLA bias by measuring the DLA autocorrelation. This approach, however, requires a much more uniform and complete sample than those currently available. Alternatively, one can make measure the cross-correlation with another tracer population. Previous attempts to measure the DLA bias gave values of $1.3 < b_d < 4$ (Cooke et al. 2006) and of $b_d = 2.17 \pm 0.20$ (Font-Ribera et al. 2012). The former derived this value from measuring the cross-correlation with luminous Lyman break galaxies, using a very small sample of only 11 DLAs in their spectra. The later used the cross-correlation with the Ly$\alpha$ forest using data from BOSS (Dawson et al. 2013) in the SDSS-III Collaboration (Eisenstein et al. 2011). Their sample contained a total of 7458 DLAs and provided the first measurement of the DLA bias.

While being able to measure for the first time the bias of the DLAs, Font-Ribera et al. (2012) used only part of both the DLAs and the Ly$\alpha$ forest observations from BOSS (the survey hadn't finished collecting data by the time their results were published). Using the entire sample we can decrease those errorbars and we can even explore the dependence of the bias factor with the DLA properties. This is interesting because there are hints that the evolution of DLA properties, in particular the metallicity, are not constant with either redshift or column density. Powerful constraints on galaxy formation models can arise from a deeper understanding on the evolution of DLA properties with redshift and column density. For example, if DLAs tended to cluster more strongly at higher redshift, then it would imply that galaxy formation would be more efficient at high redshift where the metallicities are the lowest.

We start by describing the datasets used to derive the DLA bias in section 4.2 on the next page. We use an improved estimator for the cross-correlation than Font-Ribera et al. (2012), described in section 4.3 on page 72. Section 4.4 on page 78 explain the model used to extract the DLA bias. Then, in section 4.5 on page 80, we present our results and compare them with the previous measurements. Finally, we summarize our conclusions in section 4.6 on page 93. Note that throughout this chapter we use the cosmology from Planck Collaboration et al. (2014), i.e., a flat $\Lambda$CDM cosmology with $\Omega_m = 0.3183$, $\Omega_b = 0.0490$, $h = 0.6704$, $n_s = 0.9619$, and $\sigma_8 = 0.8347$.

## 4.2 Data Sample

The DR12 of the SDSS-III Collaboration (Gunn et al. 1998; York et al. 2000; Gunn et al. 2006; Eisenstein et al. 2011; Bolton et al. 2012; Smee et al. 2013) contains the final sample of BOSS (Dawson et al. 2013). The quasar target selection used in BOSS is summarized in Ross et al. (2012), and combines different targeting methods described in Yèche et al. (2010); Kirkpatrick et al. (2011); Bovy et al. (2011).

In this section we describe the dataset used in this study. We measure the cross-correlation of two tracers of the underlying density field. These are the number density of DLA and the Ly$\alpha$ absorption along a set of lines of sight. The DLA used as tracers are designated here as *DLA sample* and the quasar lines of sight where the Ly$\alpha$ absorption is measured are designated as *Ly$\alpha$ sample*.

### 4.2.1 DLA Sample

For the DLA sample we use the DR12 extension of the DLA catalogue from Noterdaeme et al. (2012). This sample contains a total of 34050 DLA with column density $N_{HI} \geqslant 10^{20}$ cm$^{-2}$. Even though the strict definition of DLA requires its column density to be at least $2 \times 10^{20}$ cm$^{-2}$, we are considering the threshold to be a factor of two lower because systems with column density down to $10^{20}$ cm$^{-2}$ are robustly identified and are not expected to sharply change their nature. Note that when we refer to DLA we also include these systems.

In this chapter we apply several cuts in order to increase the purity of the catalogue. Keep in mind that any object included in the catalogue which is not a DLA will decrease the measured bias of the DLAs. On the other hand, completeness is not as important: eliminating a fraction of the real DLAs will only result in a increase on the errors of the cross-correlation without modifying it systematically. Note that the cuts applied here are similar to those in Font-Ribera et al. (2012) but not exactly so.

The first cut is a cut in redshift. We include in our DLA sample only those DLA in the redshift range $2.0 \leqslant z_d < 3.5$. There are a some DLAs outside this redhsift interval, but they have very few nearby lines of sight with good signal-to-noise ratio to correlate them with. Eliminating them means that we have a well defined redshift interval of our systems. This reduces our sample to 31069 DLAs.

The second cut requires the continuum-to-noise ratio (CNR) of the quasar spectrum to be at least 3. The CNR provides a good estimate of the data quality over the region of interest, and is independent of the presence of DLAs. This second cut is applied to increase the purity of the sample without drastically reducing the number of systems. 23576 DLAs survive this cut.

The third to fifth cuts we apply involve eliminating Broad Absorption Line (BAL) systems, the profiles of which are not easily distinguished from the Voight profiles of DLAs. Our third cut, then, is to exclude all the DLAs with a balnicity index different from 0, leaving 23350 DLAs. Also, BALs are more difficult to detect when they are close to the Ly$\alpha$ emission line. In order to eliminate those BALs close to the Ly$\alpha$ emission line that may be too weak to have been identified, our fourth cut eliminates all the DLAs that are within a velocity separation of 5000 km s$^{-1}$ from the quasar redshift, where

$$\frac{v}{c} = \frac{z_q - z_d}{1 + z_q} = \frac{\lambda_\alpha - \lambda_r}{\lambda_\alpha}\,, \tag{4.1}$$

$z_q$ and $z_d$ are the redshifts of the quasar and the DLA respectively, $\lambda_\alpha$ is the Ly$\alpha$ wavelength, and $\lambda_r$ is the quasar rest-frame wavelength at which the DLA absorption line is centered. This condition is equivalent to require $\lambda_r \geqslant 1195.39$Å in all the DLAs that survive this cut. Application of this constrain reduces our sample to 21415 DLAs. In Font-Ribera et al. (2012), they showed that there was an excess of DLAs with $\lambda_r$ between 1005Å and 1037Å, and that the excess was probably due to BAL contamination. Our fifth cut, then, is to remove all those DLAs with $\lambda_r$ in this interval. 19661 DLAs remain after this cut.

In their analysis of the DLAs, Mas-Ribas et al. (2016) found that a fraction of the absorption produced by DLAs detected bluewards of the Ly$\beta$ line were in fact Ly$\beta$ absorption. For some reason the detection algorithm was unable to found the associated Ly$\alpha$ absorption, and these systems were labelled as DLAs. Our sixth cut gets rid of these systems. We eliminate from our sample all those DLAs detected bluewards of the Ly$\beta$ emission line. This cut drastically reduces our sample to 13740 DLAs.

Finally, from the remaining DLAs there are 6 which have a *ThingID* of $-1$ and a Modified Julian Date (MJD) of 0. This indicates that something went wrong either in the observations or else in the DLA detection procedure. They are also excluded from the final sample. The final sample contains a total of 13734 DLAs. Here, we want to emphasize that the purity of the sample is more important than its completeness. However, we understand that the fourth, fifth, and specially the sixth cut we presented here excludes an important amount of DLAs, most of which will be true DLAs. To analyse the importance of these cuts in the final measurements, different DLA samples are studied in this work. We label the final sample considering all cuts as dataset A and the final sample considering only the cuts that are certain to remove contaminants (i.e. not applying cuts 4 to 6) as C1. Finally we label dataset C2 to the sample resulting from the application of all constraints save the sixth. In this sample, the same cuts as in Font-Ribera et al. (2012) are applied, allowing for a more direct comparison. The dataset properties are summarized in table 4.1 on the following page.

We separate the total dataset in bins according to the redshift value of the DLA and according to its column density. The cuts in column density are applied to DLAs in dataset A in such a way

| Name | Description | Number of DLAs |
|---|---|---|
| A | full DLA sample | 13734 |
| C1 | full DLA sample excluding cuts 4 to 6 | 23342 |
| C2 | full DLA sample excluding cut 6 | 19655 |
| Z1 | DLAs with $z_d < 2.25$ | 3348 |
| Z2 | DLAs with $2.25 \leqslant z_d < 2.5$ | 3455 |
| Z3 | DLAs with $2.5 \leqslant z_d$ | 6931 |
| N1 | DLAs with $\log(N_{HI}) < 20.26$ | 4448 |
| N2 | DLAs with $20.26 \leqslant \log(N_{HI}) < 20.63$ | 4683 |
| N3 | DLAs with $20.63 \leqslant \log(N_{HI})$ | 4603 |

Table 4.1: Summary of the different DLA datasets.

that the resulting sub-sambles have similar number of DLAs. The cuts in redshift are applied to the DLAs in dataset A in such a way that the resulting signal-to-noise in the cross-correlation measured in the different sub-samples are similar. Because the Ly$\alpha$ forest data available for cross-correlation is much more sparse at high redshift in BOSS, the third bin in redshift has to be made wider if the signal-to-noise is to be similar amongst the different subsamples. We label the redshift sub-samples Z1, Z2, and Z3, and the column density sub-samples N1, N2, and N3. The properties of these sub-samples are also summarized in table 4.1. Figure 4.1 on the facing page show the distribution of the total DLA sample according to redshift and column density. The applied cuts to determine the sub-samples Z1 to Z3 and N1 to N3 are shown as red solid lines in figure 4.1 on the next page.

### 4.2.2 Ly$\alpha$ Sample

For the Ly$\alpha$ sample we use the DR12 Ly$\alpha$ spectra computed as in Busca et al. (2013) (N. Busca, private communication). This corresponds to a total of 157922 spectra containing over 27 million Ly$\alpha$ pixels. In particular we use their *analysis pixels* that are the flux average over three adjacent pipeline pixels. Throughout the rest of this chapter, *pixel* refers to analysis pixels unless otherwise stated. The effective width of these pixels is 210 $\mathrm{km\,s^{-1}}$. For the quasar continuum fit their method 1 was used.

## 4.3 Cross-correlation

In this section we describe the method used to measure the cross-correlation and its covariance matrix. The method here differs from the previous measurement (Font-Ribera et al. 2012). There, they used a very simple estimator,

$$\xi^A = \frac{\sum_i \omega_i \delta_i}{\sum_i \omega_i} , \tag{4.2}$$

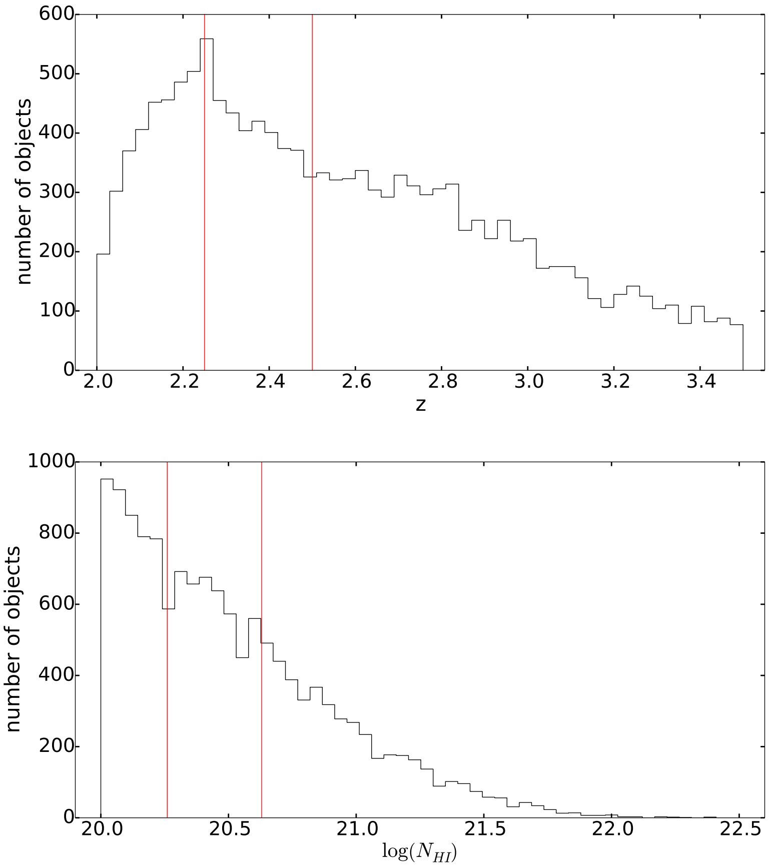


Figure 4.1: Distribution of DLAs in dataset A according to redshift (top panel) and column density (bottom panel). The solid red lines show the applied cuts to construct the different sub-samples (see table 4.1 on page 72)

where the sum is over all pixel-quasar pairs $iq$ that are at separation $r_{iq}$ in bin $A$. Then they performed a Mean Transmission Correction (MTC) to compensate for the effects on the quasar continuum estimation.

Here we present a different approach. The goal is to compute things in such a way that the MTC correction is no longer necessary. The hypothesis is that the measured Ly$\alpha$ transmission fluctuation, $\delta^{(m)}$, differs from the true Ly$\alpha$ transmission fluctuation, $\delta^{(t)}$ according to

$$\delta^{(m)} = \delta^{(t)} + a + b\lambda \,, \tag{4.3}$$

where $a$ and $b$ are small unknown functions that depend on the $\delta$ field in a complicated way, and $\lambda$ is either the wavelength or the logarithm of the wavelength (whichever is used in the computation of the $\delta$ field). Here we assume that $a$, $b$ are constant within a given forest.

This hypothesis is motivated by the definition of the $\delta$ field. The $\delta$ field, is defined as

$$\delta_i = \frac{f_i}{\overline{F}\left(z_i\right) C_i} - 1 \,, \tag{4.4}$$

where $f_i$ is the observed flux, $\overline{F}\left(z_i\right)$ is the mean transmitted fraction at redshift $z_i$, and $C_i$ is our estimation of the quasar continuum. The quasar continuum is believed to have the form $C_i = \overline{C}\left(\lambda_{rf}\right)\left(a + b\lambda_i\right)$, where $\overline{C}$ is the mean flux determined by stacking spectra, estimated at the restframe wavelength, and $a$ and $b$ are fitted constants, different for different forests. We can fit the parameters $a$ and $b$ except for a small error. Namely, $a = a_t - \delta_a$ and $b = b_t - \delta_b$.

If we Taylor expand this expression and keep only the leading order we find

$$\delta_i \approx \delta_i^{(t)} - \frac{\delta_a}{a_t + b_t\lambda_i} - \frac{\delta_b\lambda}{a_t + b_t\lambda_i} \,. \tag{4.5}$$

We can now assume that the average of $b_t$ along the different forests will be zero and that for each individual forest it is a small fluctuation of this average. This assumption is motivated by the fact that the steepness of the flux spectra is accounted for in the estimation of $\overline{C}$. Therefore, we can neglect $b_t\lambda$ over $a$, hence the presented hypothesis (equation 4.3).

### 4.3.1 Projector

Since it is impossible for us to know the values of $a$ and $b$, it is necessary for us to find a projector, $P$, that allow us to *get rid* of these parameters. Namely,

$$P\delta^{(m)} = P\delta^{(\nu)} \,. \tag{4.6}$$

To find an expression for this projector, it is interesting to work in vectorial representation. This way we can treat the forest as a whole. Keep in mind that we are assuming $a$ and $b$ to be constant throughout the forest.

To start with the derivation we are first going to consider the case $b = 0$. In vectorial form, and for a forest of length $N$, we have

$$\begin{pmatrix} \delta_1^{(m)} \\ \vdots \\ \delta_N^{(m)} \end{pmatrix} = \begin{pmatrix} \delta_1^{(\nu)} \\ \vdots \\ \delta_N^{(\nu)} \end{pmatrix} - a v_1 \ , \tag{4.7}$$

where $v_1$ is a vector in the direction of $(1, \cdots, 1)$, properly normalized.

We can now construct a projector $P_0$ which will cancel the second term in the above equation. Let us consider the projector

$$P_0 = \mathbb{I} - v_1 v_1^t \ . \tag{4.8}$$

Such a projector indeed cancels the second term in eq. 4.7:

$$P_0 v_1 = \left(\mathbb{I} - v_1 v_1^t\right) v_1 = \mathbb{I} v_1 - v_1 \underbrace{v_1^t v_1}_{=1} = v_1 - v_1 = 0 \ . \tag{4.9}$$

Now that we have a projector that cancels the contribution of the first correction, let us relax the condition $b = 0$. We now have

$$\begin{pmatrix} \delta_1^{(m)} \\ \vdots \\ \delta_N^{(m)} \end{pmatrix} = \begin{pmatrix} \delta_1^{(\nu)} \\ \vdots \\ \delta_N^{(\nu)} \end{pmatrix} - a v_1 - b v_2 \ , \tag{4.10}$$

where $v_2 = (\lambda_1, \cdots, \lambda_N)$.

We have to expand the projector $P_0$ to a new projector $P$ in such a way that we maintain the condition $P v_1 = 0$ that we imposed in the particular case where $b = 0$ and add the extra condition that $P v_2 = 0$. In order to do that we have to project using a vector which is orthogonal to $v_1$. We can follow the Gram-Schmidt process to determine such a vector: $u_2 = v_2 - \left(v_2^t v_1\right) v_1$. Any vector in this direction will verify $P_1 v_1 = 0$. However, for it to verify $P v_2 = 0$ we need a vector in the direction of $u_2$ which is properly normalized. Therefore the new projector reads

$$P = \mathbb{I} - v_1 v_1^t - \frac{1}{\mathcal{N}^2} u_2 u_2^t \ , \tag{4.11}$$

where $\mathcal{N}^2 = u_2^t u_2 = v_2^t v_2 - v_2^t v_1 v_1^t v_2$.

We can see that the projector verifies both conditions:

$$\begin{aligned} Pv_1 = \left[\mathbb{I} - v_1 v_1^t - \frac{1}{\mathcal{N}^2} u_2 u_2^t\right] v_1 = \underbrace{P_0 v_1}_{=0} - \frac{1}{\mathcal{N}^2} u_2 u_2^t v_1 = \frac{1}{\mathcal{N}^2}\left[v_2 - v_2^t v_1 v_1\right] \times \\ \times \left[v_2 - v_2^t v_1 v_1\right]^t v_1 = \frac{1}{\mathcal{N}^2}\left[v_2 - v_2^t v_1 v_1\right] \underbrace{\left(v_2^t v_1 - v_2^t v_1 \underbrace{v_1^t v_1}_{=1}\right)}_{=0} = 0\,, \end{aligned} \tag{4.12}$$

and

$$\begin{aligned} Pv_2 = \left[\mathbb{I} - v_1 v_1^t - \frac{1}{\mathcal{N}^2} u_2 u_2^t\right] v_2 = \mathbb{I} v_2 - v_1 v_1^t v_2 - \frac{1}{\mathcal{N}^2}\left[v_2 - v_2^t v_1 v_1\right] \times \\ \times \left[v_2 - v_2^t v_1 v_1\right]^t v_2 = v_2 - v_1 v_1^t v_2 - \frac{1}{\mathcal{N}^2}\left[v_2 - v_2^t v_1 v_1\right] \underbrace{\left(v_2^t v_2 - v_2^t v_1 v_1^t v_2\right)}_{=\mathcal{N}^2} = \\ = v_2 - v_1 v_1^t v_2 - \left(v_2 - v_1 v_1^t v_2\right) = 0\,. \end{aligned} \tag{4.13}$$

The projector that we have just derived allow us to compare the real and the measured values without knowing the parameters $a$ and $b$. Note that the derivation has been performed without specifying the scalar product. The expression for the projector (equation 4.11 on page 75) is then valid for any given scalar product. This is interesting because not all the pixels in the Lyα forest are equally noisy. There is a weight associated with each pixel. This weight is now easily introduced into this formalism if one simply defines the scalar product as

$$u^t v = \sum_{i \in f} u_i v_i \omega_i\,, \tag{4.14}$$

where $i$ is an index that runs over pixels in a particular forest $f$

Now that we have specified the scalar product, we can find specific expressions for $v_1$, $u_2$, and $\mathcal{N}^2$.

$$\begin{pmatrix} 1 & \cdots & 1 \end{pmatrix} \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} = \sum_{i \in f} \omega_i \Rightarrow v_1 = \frac{1}{\sqrt{\sum_{i \in f} \omega_i}} (1, ..., 1)\,, \tag{4.15}$$

$$u_2 = v_2 - v_2^t v_1 v_1 = v_2 - \frac{\sum_{i \in f} \lambda_i \omega_i}{\sqrt{\sum_{i \in f} \omega_i}} v_1 = v_2 - \bar{\lambda} \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} = \begin{pmatrix} \lambda_1 - \bar{\lambda} \\ \vdots \\ \lambda_N - \bar{\lambda} \end{pmatrix}\,, \text{ and} \tag{4.16}$$

$$\mathcal{N}^2 = u_2^t u_2 = \sum_{i \in f} \left(\lambda_i - \bar{\lambda}\right)^2 \omega_i\,. \tag{4.17}$$

Note that here $\bar{x} \equiv \sum_{i \in f} x\omega_i / \sum_{i \in f} \omega_i$.

Using this scalar product and the corresponding expressions for $\nu_1$, $u_2$, and $\mathcal{N}^2$ derived above, we can study the behaviour of this projector when it is applied to a vector $\delta$, defined in the forest of interest.

$$P\delta = \left[\mathbb{I} - \nu_1 \nu_1^t - \frac{1}{\mathcal{N}^2} u_2 u_2^t\right] \delta = \delta - \nu_1 \nu_1^t \delta - \frac{1}{\mathcal{N}^2} u_2 u_2^t \delta =$$

$$= \delta - \frac{\sum_{i \in f} \delta_i \omega_i}{\sum_{i \in f} \omega_i} \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} - \frac{\sum_{i \in f} \left(\lambda_i - \bar{\lambda}\right) \delta_i \omega_i}{\sum_{i \in f} \left(\lambda_i - \bar{\lambda}\right)^2 \omega_i} \begin{pmatrix} \lambda_1 - \bar{\lambda} \\ \vdots \\ \lambda_N - \bar{\lambda} \end{pmatrix} . \tag{4.18}$$

As we will see later, it is useful to consider the $i^{\text{th}}$ component of this vector:

$$(P\delta)_i = \sum_{j \in f} P_{ij} \delta_j = \delta_i - \bar{\delta} - \frac{\sum_{j \in f} \delta_j \left(\lambda_j - \bar{\lambda}\right) \omega_j}{\sum_{j \in f} \left(\lambda_j - \bar{\lambda}\right)^2 \omega_j} \left(\lambda_i - \bar{\lambda}\right) . \tag{4.19}$$

### 4.3.2 Estimator for the cross-correlation

While the cross-correlation is a continuous function, in practice we have to discretize it in several bins. The cross-correlation in bin $A$ is computed with a simple estimator:

$$\xi^A = \frac{\sum_{q,f} \sum_{i \in f} \Theta^A_{iq} \omega_i \sum_{j \in f} P_{ij} \delta_j}{\sum_{q,f} \sum_{i \in f} \Theta^A_{iq} \omega_i} , \tag{4.20}$$

where the indexes $q$ and $f$ run over quasars and forests respectively, the indexes $i$, $j$, and $k$, run over pixels in a particular forest, and $\Theta^A_{qi}$ is $1$ if the quasar-pixel pair is in bin A and 0 otherwise.

### 4.3.3 Covariance matrix

The covariance of the cross-correlation for two bins $A$ and $B$ is equal to

$$C^{AB} \equiv \left\langle \xi^A \xi^B \right\rangle - \left\langle \xi^A \right\rangle \left\langle \xi^B \right\rangle = \frac{1}{S^{AB}} \sum_{qf} \sum_{i \in f} \Theta^A_{iq} \omega_i \sum_{q'f'} \sum_{i' \in f'} \Theta^B_{i'q'} \omega_{i'} \xi_{ii'} , \tag{4.21}$$

where $\xi_{ii'}$ is the correlation of the measured values of $\delta$ in pixels $i$ and $i'$, and the normalization factor is

$$S^{AB} = \sum_{qf} \sum_{i \in f} \Theta^{A\omega_i}_{iq} \sum_{q'f'} \sum_{i' \in f'} \Theta^B_{i'q'} \omega_{i'} . \tag{4.22}$$

Note that the correlation $\xi_{ii'}$, with two subindexes for the two correlated pixels, should not be confused with the cross-correlation estimator $\xi^A$, which is a different matrix with index referring to bins in $(\pi, \sigma)$. There are three main contributions to the correlation $\xi_{ii'}$. First, there is a noise component that we assume to be uncorrelated between the different pixels. This contribution arises from the fact the same Ly$\alpha$ pixel contributes to more than one bin in $(\pi, \sigma)$ when paired with different quasars. Second, there is a contribution arising from continuum fitting errors that affect only pixels in the same spectrum. Finally the different Ly$\alpha$ pixels are intrinsically correlated. These three contributions can be reorganized as follows

$$\xi_{ij} = \left(\sigma_N^2 + \sigma_F^2\right)_i \delta_{ij} + \xi_n \left(\frac{z_i + z_j}{2}\right) \delta\left(\sigma_{ij}\right) + \xi_{ij}^{3D} \left(1 - \delta\left(\sigma_{ij}\right)\right) \ . \tag{4.23}$$

Note that here $\delta_{ij}$ and $\delta\left(\sigma_{ij}\right)$ are the Kronecker and the Dirac deltas and should not be confused with $\delta_i$, which is the Ly$\alpha$ transmission fluctuation. The first term contributes only when the two pixels are the same and includes the noise contribution. It is computed as explained in Busca et al. (2013). The second term contributes when the two pixels are found in the same spectrum and includes the contribution from the quasar continuum estimation and the Ly$\alpha$ autocorrelation along the line of sight. It is described in more detail in appendix 4.A on page 94. The third term contributes when the two pixels are in different spectra and includes the contribution of the Ly$\alpha$ autocorrelation.

Due to the large number of quasar and Ly$\alpha$ spectra, the complete sum of pair-of-pairs requires a prohibitively large amount of computer time and is, in practice, not possible. In this work we consider the contribution to the correlation of pixels in different spectra, the third term in equation 4.21 on page 77, to be a second order effect and thus neglect it here.

## 4.4 Fitting the cross-correlation

### 4.4.1 Distortion matrix

The $\chi^2$ statistic for this estimator reads

$$\chi^2 = (\xi - \langle\xi\rangle)^t\, C^{-1}\, (\xi - \langle\xi\rangle) \ , \tag{4.24}$$

where C is the covariance matrix between the different bins of the cross-correlation.

The expected value of the cross-correlation estimator in bin A can be written as

$$\left\langle \xi^A \right\rangle = \frac{\sum_{q,f} \sum_{i \in f} \Theta_{iq}^A \omega_i \sum_{j \in f} P_{ij} \xi_{jq}}{\sum_{q,f} \sum_{i \in f} \Theta_{iq}^A \omega_i} \ , \tag{4.25}$$

where $\xi_{jq}$ is the theoretical prediction of the cross-correlation for the quasar-pixel pair $jq$.

At this point we can discretize the model similarly to the discretization of the data. Then

$$\xi_{jq} = \sum_B \Theta_{jq}^B \xi^B \ . \tag{4.26}$$

Note that this discretization needs not to be the same as the discretization on the data, but it is convenient to do so. The formalism presented here applies to whichever case.

Introducing this discretization into equation 4.25, the expected value of the cross-correlation can be written as

$$\left\langle \xi^{A} \right\rangle = \sum_{B} \frac{\sum_{q,f} \sum_{i \in f} \Theta^{A}_{iq} \omega_i \sum_{j \in f} P_{ij} \Theta^{B}_{jq} \xi^{B}}{\sum_{q,f} \sum_{i \in f} \Theta^{A}_{iq} \omega_i} \equiv \sum_{B} D^{AB} \xi^{B} \,, \tag{4.27}$$

where $D^{AB}$ is the distortion matrix element that relates the cross-correlation measured in bin A and the model for the cross-correlation in bin B. It is defined as

$$D_{AB} = \frac{\sum_{q,f} \sum_{i \in f} \Theta^{A}_{iq} \omega_i \sum_{j \in f} \left( \delta_{ij} - \frac{\omega_j}{\sum_{k \in f} \omega_k} - \frac{(\lambda_j - \bar{\lambda}_f)(\lambda_i - \bar{\lambda}_f)\omega_j}{\sum_{k \in f} (\lambda_k - \bar{\lambda}_f)^2 \omega_k} \right) \Theta^{B}_{jq}}{\sum_{q,f} \sum_{i \in f} \Theta^{A}_{iq} \omega_i} \,, \tag{4.28}$$

where the indexes $q$ and $f$ run over quasars and forests respectively, the indexes $i$, $j$, and $k$, run over pixels in a particular forest, $\delta_{ij}$ is the Kronecker delta and should not be confused with the Ly$\alpha$ transmission fluctuation, and $\Theta^{A}_{qi}$ is 1 if the quasar-pixel pair is in bin A and 0 otherwise.

### 4.4.2 Model

In this chapter we measure the cross-correlation between DLAs ant the Ly$\alpha$ forest for a number of datasets (see section 4.2 on page 70). To interpret and to compare these measurements, we use a single model which is described in this section. To do the actual fitting we use the publicly available fitting code *baofit* (Kirkby et al. 2013). The code has been adapted to include the distortion matrix formalism by Michael Blomqvist and collaborators.

In the limit of large scales the linear theory predicts the shape of the cross-correlations of any two pair of tracers of the LSS. The limit of large scales is broadly defined as the distances at which non-linearities are not important any more. The transition between non-linear and linear scales is not clear and it depends on the type of tracers and the environmental conditions they are subject to. This issue is further discussed in appendix 4.B on page 95.

In Fourier space, linear theory predicts that redshift space distortions of a biased tracer enhance the amplitudes of each Fourier mode by a factor $b\left(1 + \beta\mu_k^2\right)$, where $b$ is the bias factor of the tracer, $\beta$ its redshift space distortion parameter, and $\mu_k$ the cosine of the angle between the Fourier mode and the line of sight. For the DLA-Ly$\alpha$ cross-correlation, the associated power spectrum is then equal to

$$P_{d\alpha}\left(\vec{k}, z\right) = b_d \left(1 + \beta_d \mu_k^2\right) b_\alpha \left(1 + \beta_\alpha \mu_k^2\right) P_L\left(k, z\right) \,, \tag{4.29}$$

where the subscript $d$ stands for DLA, the subscript $\alpha$ for Ly$\alpha$, and $P_L\left(k, z\right)$ is the linear matter power spectrum. Note that in this model we do not include neither a broadband function as done in previous studies with BOSS data (e.g. Font-Ribera et al. 2014; Delubac et al. 2015; Blomqvist et al. 2015) nor the MTC correction used in Font-Ribera et al. (2012).The correction is already implemented in the projection of the $\delta$ field (see section 4.3 on page 72).

As seen in equation 4.29, we are not sensitive to variations on $b_d$ nor $b_\alpha$ but to the product of both. Also, statistics on DLAs are lower than that on the Ly$\alpha$ forest. This makes it difficult to use the DLA-Ly$\alpha$ cross-correlation to provide competitive measurements for the bias and redshift space distortion parameter for both tracers. However, if the parameters of either DLAs or the Ly$\alpha$ forest are independently constrained, then we can give competitive constraints for the other tracer.

Previous analysis of the BOSS Collaboration (see e.g. Blomqvist et al. 2015; Delubac et al. 2015) have studied in detail the Ly$\alpha$ forest autocorrelation and can give such constraints. In particular, for this chapter we will consider both the values for the bias and redshift space distortion parameter for the Ly$\alpha$ forest, and the BAO parameters as those specified in Blomqvist et al. (2015), i.e., we fix $\beta_\alpha = 1.39$ and $b_\alpha (1 + \beta_\alpha) = -0.374$ at reference redshift $z_{\text{ref}} = 2.3$. Also, we fix $\beta_d b_d = f(\Omega) = 0.968897$. To account for the fact that the mean redshift of different bins are not exactly at $z_{\text{ref}}$, we assume the bias factor of the Ly$\alpha$ forest to evolve with redshift as $(1 + z)^{2.9}$, as suggested in McDonald et al. (2006), and the redshift space distortion parameter for the Ly$\alpha$ forest to be constant with redshift. The bias factor of the DLAs is assumed to be constant in redshift as far as the fitting is concerned. The evolution of this parameter will be analysed via the sub-samples Z1, Z2, and Z3.

Summing up, we fix all parameters in equation 4.29 on page 79 but $b_d$. Additionally we allow the symmetry in the cross-correlation to be shifted constant amount $\Delta v$, which is a second parameter of the fit, to account for a possible error in the redshift determination of either the DLAs or the Ly$\alpha$ forest. All fits will be made regarding these specifications unless otherwise noted.

## 4.5 Results

We have measured the cross-correlation for the all the samples listed in table 4.1 on page 72 with $\delta_\pi = \delta_\sigma = 2\,h^{-1}$Mpc and $\pi_{\text{max}} = \sigma_{\text{max}} = 80\,h^{-1}$Mpc. The measured biases are summarized in table 4.2 on the next page. The results from the different datasets are organized throughout this section as follows: in section 4.5.1 on page 82 we present the results for the overall sample (sample A, see section 4.2 on page 70) and compare them with the results from samples C1 and C2, in sections 4.5.2 and 4.5.3 on page 85 and on page 86 we explore the redshift dependence and the column density dependence of the parameters using sub-samples Z1 to Z3 and N1 ro N3 respectively (see table 4.1 on page 72 and section 4.2 on page 70). Finally in sections 4.5.4 and 4.5.5 on page 86 and on page 92 we use dataset A to explore the scale dependence of the bias factor and the dependence on the Ly$\alpha$ bias respectively.

All our fits exclude bins with $r = \left(\pi^2 + \sigma^2\right)^{1/2} < 5\,h^{-1}$Mpc or $r > 90\,h^{-1}$Mpc even though the points are shown in the plots. The lower limit on $r$ is placed to avoid including non-linearities present at small scales into the fit. Even though we may already be entering the

| Dataset | $b_d$ | $\chi^2$(d.o.f.) |
|---|---|---|
| A | $1.87 \pm 0.05$ | 2965.41 (2896-2) |
| C1 | $1.86 \pm 0.04$ | 3115.52 (2896-2) |
| C2 | $1.82 \pm 0.04$ | 2974.40 (2896-2) |
| Z1 | $1.98 \pm 0.09$ | 2918.09 (2896-2) |
| Z2 | $1.80 \pm 0.09$ | 3015.57 (2896-2) |
| Z3 | $1.80 \pm 0.10$ | 3074.20 (2896-2) |
| N1 | $1.91 \pm 0.10$ | 3006.52 (2896-2) |
| N2 | $1.95 \pm 0.09$ | 3092.98 (2896-2) |
| N3 | $1.74 \pm 0.09$ | 3004.70 (2896-2) |

Table 4.2: Summary of the different DLA datasets. The values of the bias are given at the reference redshift $z_{\text{ref}} = 2.3$ and for $\beta_\alpha = 1.39$. See table 4.1 on page 72 for a summary on the datasets' properties.

non-linear regime at 5 $h^{-1}$Mpc, the deviations from the linear theory are small if present at all. The choice of this limit allows for a more direct comparison with the results previously reported by Font-Ribera et al. (2012). This limit is further discussed in appendix 4.B on page 95. The upper limit on $r$ is placed to make sure we stay clear of the BAO peak.

For display purposes alone, in most of the plots the data-points have been averaged in bigger bins. The average is computed as

$$\xi^B = \frac{\sum_{A\in B} \xi^A/\sqrt{C^{AA}}}{\sum_{A\in B} 1/\sqrt{C^{AA}}}, \tag{4.30}$$

$$\pi^B = \frac{\sum_{A\in B} \pi^A/\sqrt{C^{AA}}}{\sum_{A\in B} 1/\sqrt{C^{AA}}}, \tag{4.31}$$

$$\frac{1}{C^{B_1B_2}} = \sum_{A_1\in B_1}\sum_{A_2\in B_2} \frac{1}{C^{A_1A_2}}, \tag{4.32}$$

where the indexes $A$, $A_1$, and $A_2$ stand for the measured bins, the indexes $B$, $B_1$, and $B_2$ stand for the new bins, and the sum $\sum_{A\in B}$ is over all bins $A$ that are included in bin $B$. In this chapter, as far as the figures are concerned, we will use 11 bins in $\sigma$ which will be delimited by (0, 4, 8, 12, 16, 20, 28, 32, 40, 48, 64, 80), and 30 bins in $\pi$, delimited by (-80, -72, -64, -56, -48, -40, -36, -32, -28, -24, -20, -16, -12, -8, -4, 0, 4, 8, 12, 16, 20, 24, 28, 32, 36, 40, 48, 56, 64, 72, 80) with average values computed as described above unless otherwise noted. Note that all delimitations are expressed in $h^{-1}$Mpc.

### 4.5.1 Measured bias

The results of the DLA-Ly$\alpha$cross-correlation as a function of $\pi$ are shown for the different bins of $\sigma$ and for sample A in figures 4.2 and 4.3 on the next page and on page 84. Our fit gives $b_d = 1.87 \pm 0.05$ at $z_{\rm ref} = 2.3$. The solid lines are our best-fit model. Note that only bins with $r = \left(\pi^2 + \sigma^2\right)^{1/2} \geqslant 5\,h^{-1}$Mpc are considered for the fit, and that a single set of parameters was used to fit the data in all bins. The results are also shown in a contour plot in figure 4.4 on page 85. Note that in the contour plot the smoothing of the bins is performed in a slightly different way. For bins with $16\,h^{-1}$Mpc $< r \leqslant 32\,h^{-1}$Mpc we average their value with the bins with $\pi \in (\pi_{\tt bin} - \delta_\pi, \pi_{\tt bin} + \delta_\pi)$ and $\sigma \in (\sigma_{\tt bin} - \delta_\sigma, \sigma_{\tt bin} + \delta_\sigma)$ weighted by $1/\left(e + 0.05\right)$, where e is the squared root of the corresponding diagonal element of the covariance matrix. For bins with $r > 32\,h^{-1}$Mpc we average their value with the bins with $\pi \in (\pi_{\tt bin} - 2\delta_\pi, pi_{\tt bin} + 2\delta_\pi)$ and $\sigma \in (\sigma_{\tt bin} - 2\delta_\sigma, \sigma_{\tt bin} + 2\delta_\sigma)$ weighted in the same way.

The value we obtain for the DLA bias is somewhat lower to that reported by Font-Ribera et al. (2012), $b_d = 2.17 \pm 0.20$ for $\beta_\alpha = 1$. This corresponds to $b_d = 2.33 \pm 0.22$ for $\beta_\alpha = 1.39$ (see their section 4.7). What is more, this value has to be compared with the measurement obtained for sample C2, which has the same cuts as the sample in Font-Ribera et al. (2012), $b_d = 1.82 \pm 0.04$. Part of the discrepancy we see here is not a real discrepancy but it arises because we use different Ly$\alpha$ forest parameters and different fiducial cosmologies. What we measure is the amplitude of the cross-correlation, which is proportional to $b_d b_\alpha \sigma_8^2\left(z\right)$. The bias of the DLAs is then obtained by independently constraining $b_\alpha$ and $\sigma_8$ at our redshift of reference, namely $z_{\rm ref} = 2.3$. Due to the degeneracies present between the different parameters it is useful to represent this proportionally factor as a the combination

$$b_d \frac{b_\alpha\left(1 + \beta_\alpha\right)}{\left(1 + \beta_\alpha\right)} \left(\sigma_8 \frac{D\left(z\right)}{D_0}\right)^2 , \tag{4.33}$$

where $D\left(z\right)$ is the growth factor evaluated at redshift $z$ and $D_0 \equiv D\left(z = 0\right)$.

In their fits, Font-Ribera et al. (2012) use, $b_\alpha\left(1 + \beta_\alpha\right) = -0.336$ at $z = 2.25$ , reported by Slosar et al. (2011), $\sigma_8 = 0.8$, and $D\left(z = 2.25\right)/D_0 = 0.3685$, whereas we use $b_\alpha\left(1 + \beta_\alpha\right) = -0.374$ at $z = 2.3$, reported by Blomqvist et al. (2015), $\sigma_8 = 0.8347$, and $D\left(z = 2.3\right)/D_0 = 0.3787$. The corrected value for the DLA bias from Font-Ribera et al. (2012) is then $b_d = 2.09 \pm 0.19$, where we have assumed the relative error to be conserved.

Once this correction has been applied we observe a discrepancy between our measurement and that of Font-Ribera et al. (2012) at the $\sim 1.5\sigma$ level. The reason for this discrepancy is yet to be fully understood. Interestingly, we find that our measurement is consistent with the NOMTC measurement reported by Font-Ribera et al. (2012): $b_d = 2.03 \pm 0.20$, corrected for the different Ly$\alpha$ forest parameters and different fiducial cosmologies. This may indicate that the correction we are applying is not fully correcting for the effects introduced by the quasar continuum fit, or that the MTC presented by Font-Ribera et al. (2012) overcorrected for these effects. Extensive studies on mock catalogues should be able to clarify this point, but we leave them as future work.

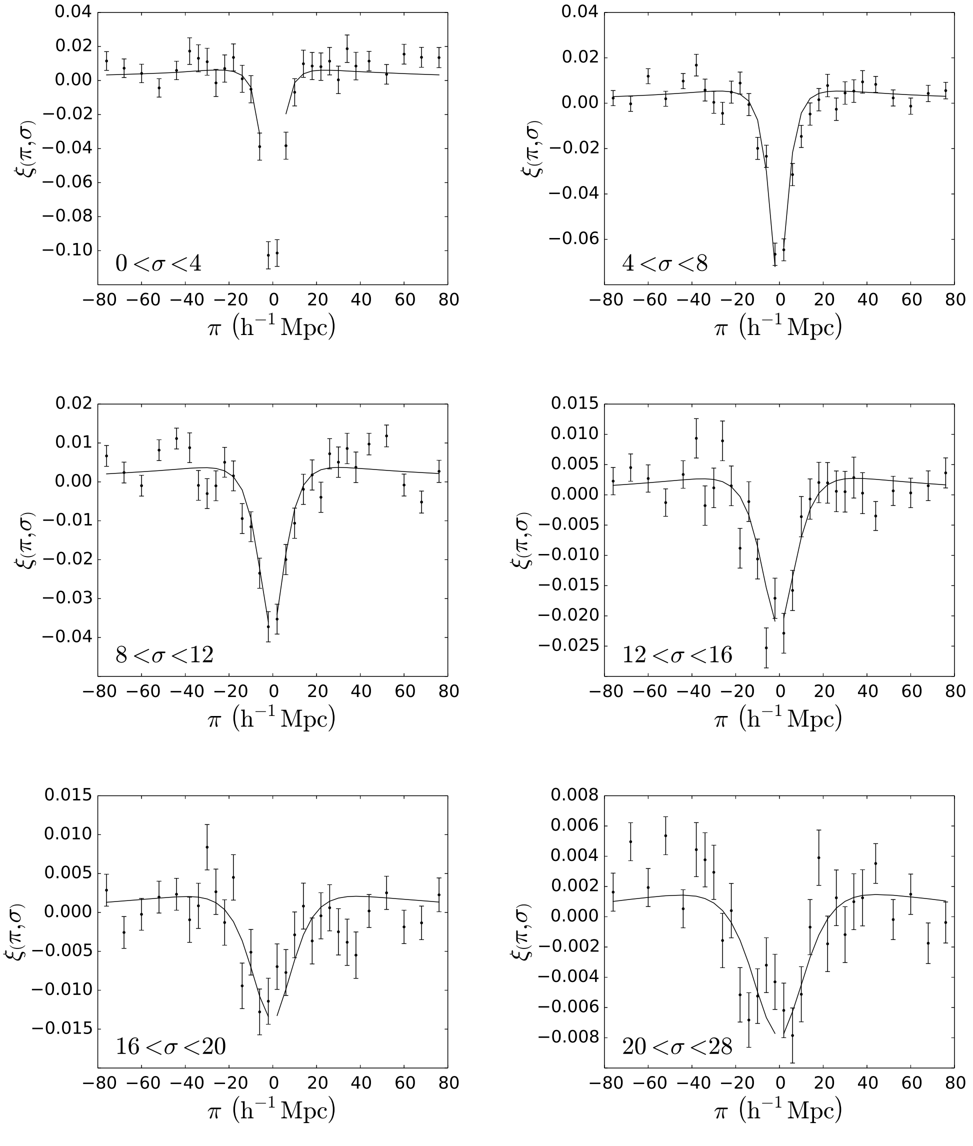


Figure 4.2: Measurement of the cross-correlation for different bins in $\sigma$: from left to right $0\,\mathrm{h}^{-1}\mathrm{Mpc} < \sigma < 4\,\mathrm{h}^{-1}\mathrm{Mpc}$, $4\,\mathrm{h}^{-1}\mathrm{Mpc} \leqslant \sigma/8\,\mathrm{h}^{-1}\mathrm{Mpc}$, $8\,\mathrm{h}^{-1}\mathrm{Mpc} \leqslant \sigma < 12\,\mathrm{h}^{-1}\mathrm{Mpc}$, $12\,\mathrm{h}^{-1}\mathrm{Mpc} \leqslant \sigma < 16\,\mathrm{h}^{-1}\mathrm{Mpc}$, $16\,\mathrm{h}^{-1}\mathrm{Mpc} \leqslant \sigma < 20\,\mathrm{h}^{-1}\mathrm{Mpc}$, and $20\,\mathrm{h}^{-1}\mathrm{Mpc} \leqslant \sigma < 28\,\mathrm{h}^{-1}\mathrm{Mpc}$. The values of $\sigma$ are expressed in $\mathrm{h}^{-1}\mathrm{Mpc}$. Figure 4.3 on the next page contain similar plots for the remaining bins in $\sigma$. In all panels black circles are for sample A. Black solid lines correspond to the best-fit model considering $r_{\min} = 5\,\mathrm{h}^{-1}\mathrm{Mpc}$.

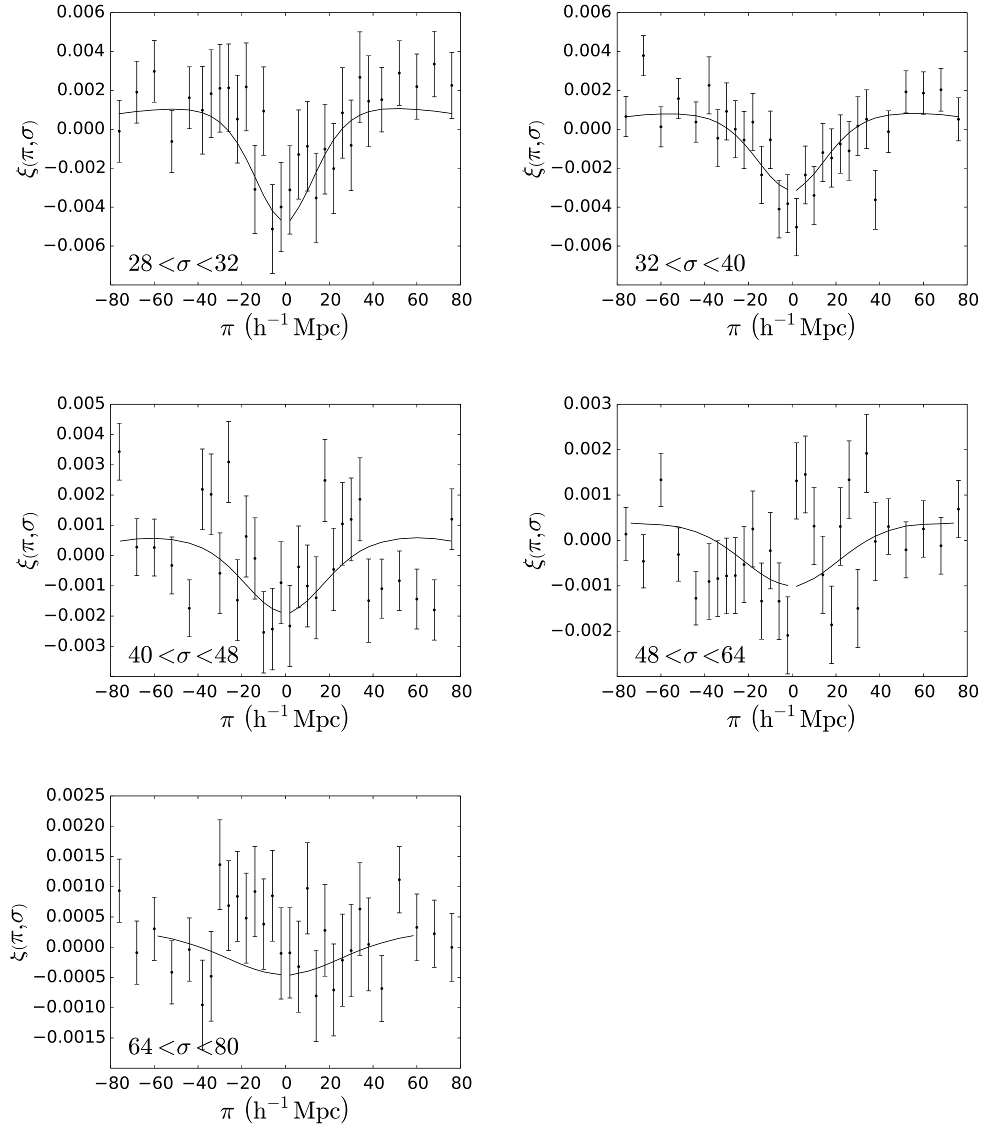


Figure 4.3: From left to right and from top to bottom, same as figure 4.2 on page 83 but for bins with $28\,\mathrm{h}^{-1}\mathrm{Mpc} \leqslant \sigma < 32\,\mathrm{h}^{-1}\mathrm{Mpc}$, $32\,\mathrm{h}^{-1}\mathrm{Mpc} \leqslant \sigma < 40\,\mathrm{h}^{-1}\mathrm{Mpc}$, $40\,\mathrm{h}^{-1}\mathrm{Mpc} \leqslant \sigma < 48\,\mathrm{h}^{-1}\mathrm{Mpc}$, $48\,\mathrm{h}^{-1}\mathrm{Mpc} \leqslant \sigma < 64\,\mathrm{h}^{-1}\mathrm{Mpc}$, and $64\,\mathrm{h}^{-1}\mathrm{Mpc} \leqslant \sigma < 80\,\mathrm{h}^{-1}\mathrm{Mpc}$.

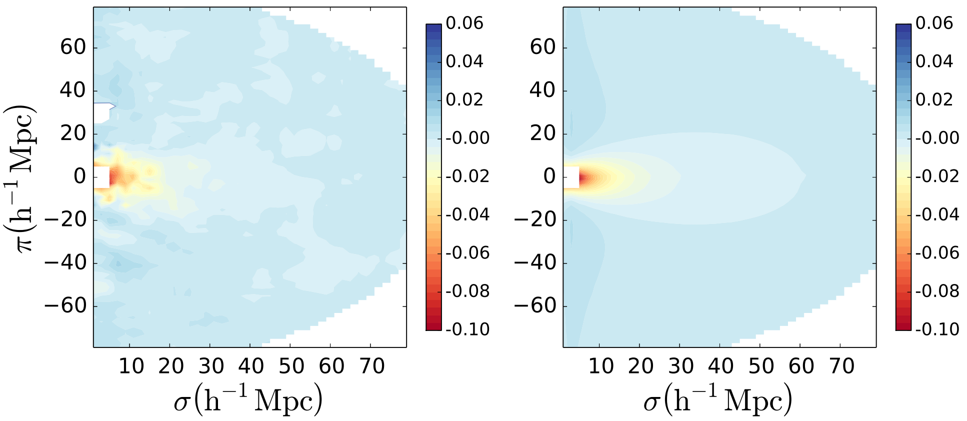


Figure 4.4: Contour plot of the measured DLA-Ly$\alpha$cross-correlation (left) and our best-fit theoretical model considering bins with $5\,h^{-1}\mathrm{Mpc} < r = \left(\pi^2 + \sigma^2\right)^{1/2} < 90\,h^{-1}\mathrm{Mpc}$ (right).

An alternative explanation for this discrepancy can be found the dependence of the DLA bias on the Ly$\alpha$ bias. Font-Ribera et al. (2012) found $b_d \propto \beta_\alpha^{0.22}$ whereas we find $b_d \propto \beta_\alpha^{0.10}$ (see section 4.5.5 on page 92). If we take into account the dependence found here instead of the one reported by Font-Ribera et al. (2012), then the corrected value for the DLA bias from Font-Ribera et al. (2012) is $b_d = 2.01 \pm 0.19$, which is consistent with our measurement at the $1\sigma$ level. Note, however, that these two possible explanations for the discrepancy will have to be further tested and understood before any claim can be made. We have not yet discarded the possibility that the discrepancy arises from a statistical fluctuation.

We now shift our attention to the effects the cuts applied to the DLA samples may be having in the measurement of the bias factor of DLAs. As we have already seen, the bias factor measured in dataset C2 is $\sim$ 3% lower than that measured in dataset A. To this measurements we have to add the $b_d = 1.86 \pm 0.04$ obtained for dataset C1. Naturally, the measurement in C1 is the one presenting a tighter bound on the measurement (C1 is the largest dataset). We find that the discrepancies between these three measurements are consistent with them being statistical noise. However, it is important to note that measurements in datasets C1 and C2 will be systematically biased towards lower bias values due to the contamination of the datasets with wrongly identified DLAs, even if we are unable to measure this bias. These misidentified DLAs will contribute with a 0 to the average value of the DLA bias. Thus, this question will become more relevant as we the DLA sample grows.

### 4.5.2 Bias dependence on redshift

The dependence of the DLA bias with the DLA redshift is analysed by comparing the DLA-Ly$\alpha$ cross-correlation of samples Z1, Z2, and Z3 (see table 4.1 on page 72). Figures 4.5 and 4.6 on the facing page and on page 88 show the measured values of the cross-correlation as a function of $\pi$ for the different bins in $\sigma$. The increase of dispersion for each of the samples is due to the fact that we have reduced the number of DLA in each of the samples. Overplotted are the best-fit theoretical models for each of the samples. In these plots there is no clear trend between the different samples. Left panel in figure 4.7 on page 89 show the measured dependence of the DLA bias with the DLA redshift. The exact measurements are tabulated in table 4.2 on page 81.

We see that the recovered bias at the lowest redshift bin is somewhat higher than the rest but with a low statistical significance. Font-Ribera et al. (2012) reported a similar behaviour. They also found a small decrease in $b_d$ with redshift, but the decrease was measured at higher redshift and was also at low statistical significance. Taking into account both measurements, they seem to point in the direction that theDLA bias is constant with redshift. The difference between the models shown in figures 4.5 and 4.6 on the facing page and on page 88 correspond to them having quite different values of $\Delta\nu$.

### 4.5.3 Bias dependence on neutral hidrogen column density

To estimate the dependence of the DLA bias with the DLA column density, we compute the DLA-Lyα cross-correlations for sub-samples in column density, samples N1, N2, and N3 (see table 4.1 on page 72). Figures 4.8 and 4.9 on page 90 and on page 91 show the measured values of the cross-correlation as a function of $\pi$ for the different bins in $\sigma$. Overplotted are the best-fit theoretical models for each of the samples. We can clearly see that the different samples follow roughly the same pattern and that the models are practically superposed. As in section 4.5.2 on page 85, the increase of dispersion is due to the fact that we have reduced the number of DLA in each of the samples. Right panel in figure 4.7 on page 89 shows the measured dependence of the DLA bias with the DLA column density. The measured biases are tabulated in table 4.2 on page 81. These results show that there is a sudden decrease of the DLA bias factor at high column densities, even though the statistical significance of this is low. This decrease can also be observed directly from the models plotted in figures 4.8 and 4.9 on page 90 and on page 91.

### 4.5.4 Scale dependence of the bias factor

An important test that our measurement agrees with the theoretical expectation is the predicted dependence of the cross-correlation on the separation $r$. In the ΛCDM model, and according to linear theory, we expect there to be no dependence, except maybe at small scales, where non-linear effects may start to be important. Using sample A we measure the evolution of the DLA bias with $r$. For this analysis we assume that the Lyα bias and redshift space parameters are constant with $r$. The actual measurement is performed by fitting the model to sample A but considering only a small interval in $r$. These intervals are selected to form rings of equal areas from. The first bin starts at rmin $= 5\,h^{-1}$Mpc and the bins are computed as $\left[1.35^i r_{\min}, 1.35^{i+1} r_{\min}\right)$, with $i = 0, 1, 2, ..., 9$.

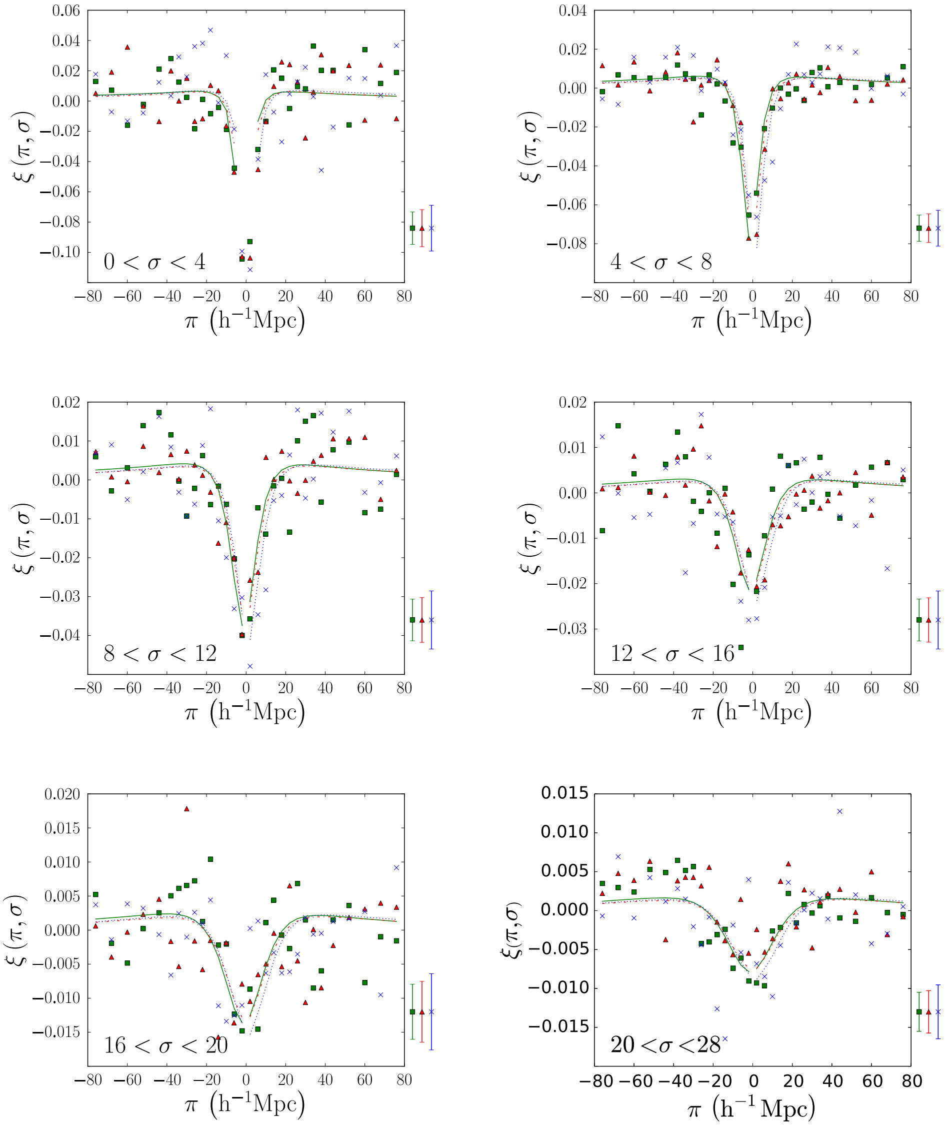


Figure 4.5: Similar to figure 4.2 on page 83 but for the sub-samples in redshift Z1, Z2, and Z3. In this figure from left to right $0\,h^{-1}\mathrm{Mpc} < \sigma < 4\,h^{-1}\mathrm{Mpc}$, $4\,h^{-1}\mathrm{Mpc} \leqslant \sigma/8\,h^{-1}\mathrm{Mpc}$, $8\,h^{-1}\mathrm{Mpc} \leqslant \sigma < 12\,h^{-1}\mathrm{Mpc}$, $12\,h^{-1}\mathrm{Mpc} \leqslant \sigma < 16\,h^{-1}\mathrm{Mpc}$, $16\,h^{-1}\mathrm{Mpc} \leqslant \sigma < 20\,h^{-1}\mathrm{Mpc}$, and $20\,h^{-1}\mathrm{Mpc} \leqslant \sigma < 28\,h^{-1}\mathrm{Mpc}$. Figure 4.6 on the next page contain similar plots for the remaining bins in $\sigma$. In all panels green squares, red triangles, and blue crosses stand for samples Z1, Z2, and Z3 respectively. Green solid lines, red dashed-dotted lines, and blue dotted lines correspond to the best-fit model considering $r_{\mathrm{min}} = 5\,h^{-1}\mathrm{Mpc}$ fitted to samples Z1, Z2, and Z3 respectively. Errorbars at the side of the plot show the typical error of the points in the corresponding $\sigma$ bin.

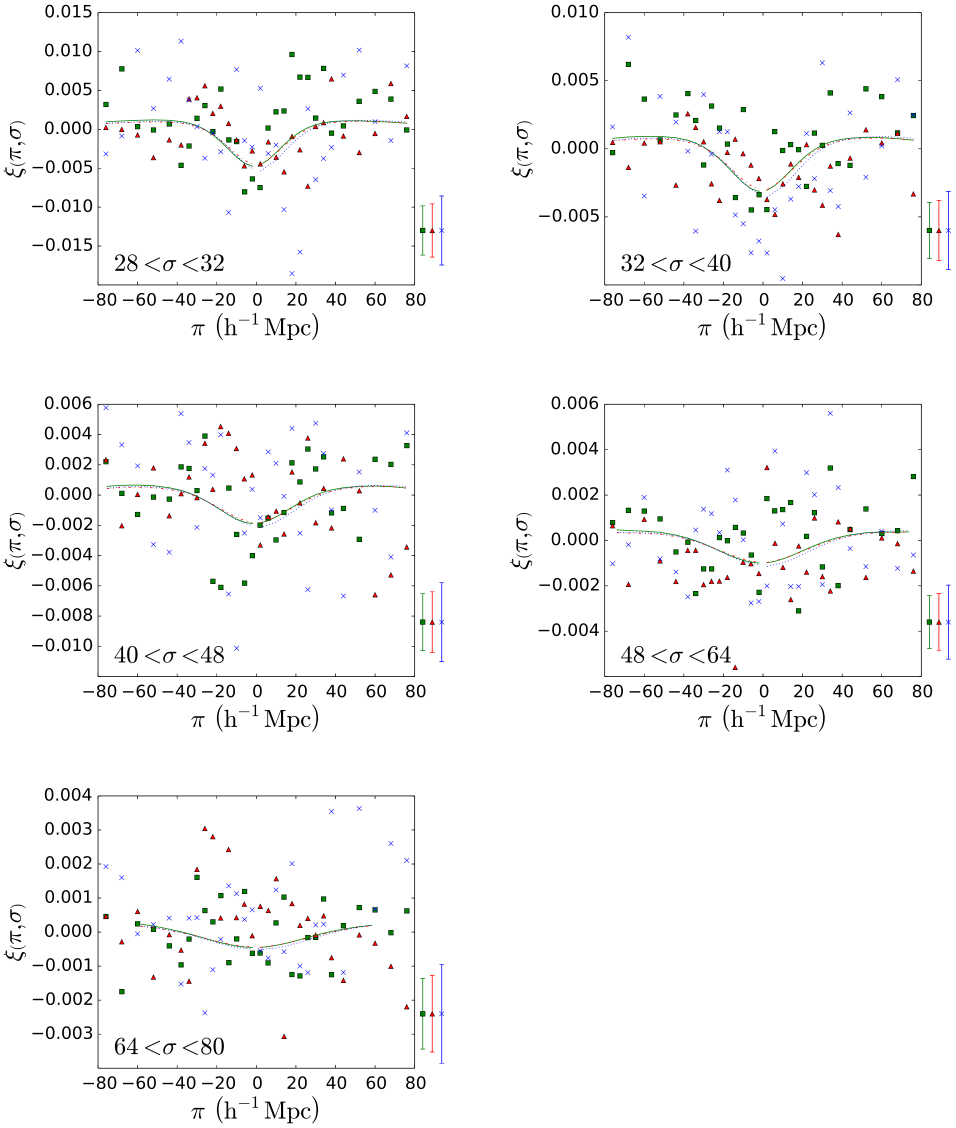


Figure 4.6: From left to right and from top to bottom, same as figure 4.5 on page 87 but for bins with $28\,\mathrm{h}^{-1}\mathrm{Mpc} \leqslant \sigma < 32\,\mathrm{h}^{-1}\mathrm{Mpc}$, $32\,\mathrm{h}^{-1}\mathrm{Mpc} \leqslant \sigma < 40\,\mathrm{h}^{-1}\mathrm{Mpc}$, $40\,\mathrm{h}^{-1}\mathrm{Mpc} \leqslant \sigma < 48\,\mathrm{h}^{-1}\mathrm{Mpc}$, $48\,\mathrm{h}^{-1}\mathrm{Mpc} \leqslant \sigma < 64\,\mathrm{h}^{-1}\mathrm{Mpc}$, and $64\,\mathrm{h}^{-1}\mathrm{Mpc} \leqslant \sigma < 80\,\mathrm{h}^{-1}\mathrm{Mpc}$.

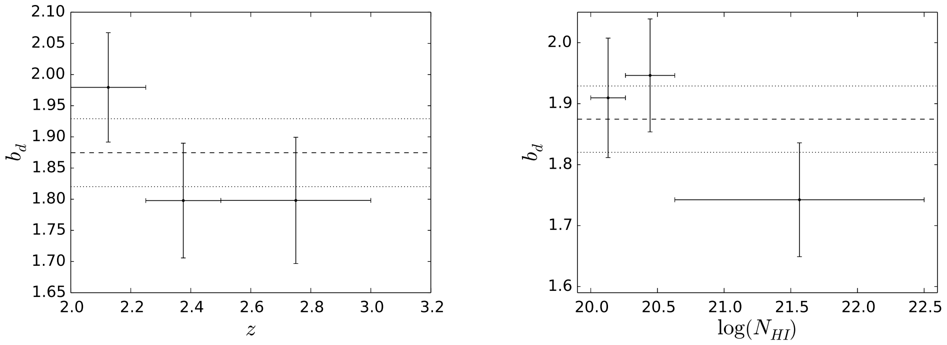


Figure 4.7: Bias of the DLA against $z$ (left) and $\log(N_{\mathrm{HI}})$ (right) obtained by fitting samples Z1, Z2, and Z3, and N1, N2, and N3 respectively (see table 4.1 on page 72). Dashed lines indicate the value obtained by fitting sample A. Dotted lines indicate the $1\sigma$ region.

| $r_{\mathrm{min}}$ | $r_{\mathrm{max}}$ | $b_{\mathrm{d}}$ | $\chi^2$ (d.o.f) |
|---|---|---|---|
| 5.00 | 6.75 | $1.88 \pm 0.12$ | 12.6599 (8-2) |
| 6.75 | 9.11 | $1.83 \pm 0.12$ | 13.4794 (16-2) |
| 9.11 | 12.30 | $2.10 \pm 0.15$ | 34.2189 (28-2) |
| 12.30 | 16.61 | $2.17 \pm 0.10$ | 56.1263 (48-2) |
| 16.61 | 22.42 | $1.44 \pm 0.18$ | 72.0837 (88-2) |
| 22.42 | 30.27 | $1.95 \pm 0.10$ | 188.073 (166-2) |
| 30.27 | 40.86 | $1.79 \pm 0.29$ | 313.483 (290-2) |
| 40.86 | 55.16 | $2.26 \pm 0.36$ | 578.824 (543-2) |
| 55.16 | 74.47 | $1.17 \pm 0.45$ | 1032.11 (977-2) |
| 74.47 | 100.53 | $2.00 \pm 0.72$ | 917.469 (944-2) |

Table 4.3: Bias of the DLA against $r$ obtained by fitting sample A but considering only bins with $r \in [r_{\mathrm{min}}, r_{\mathrm{max}})$.

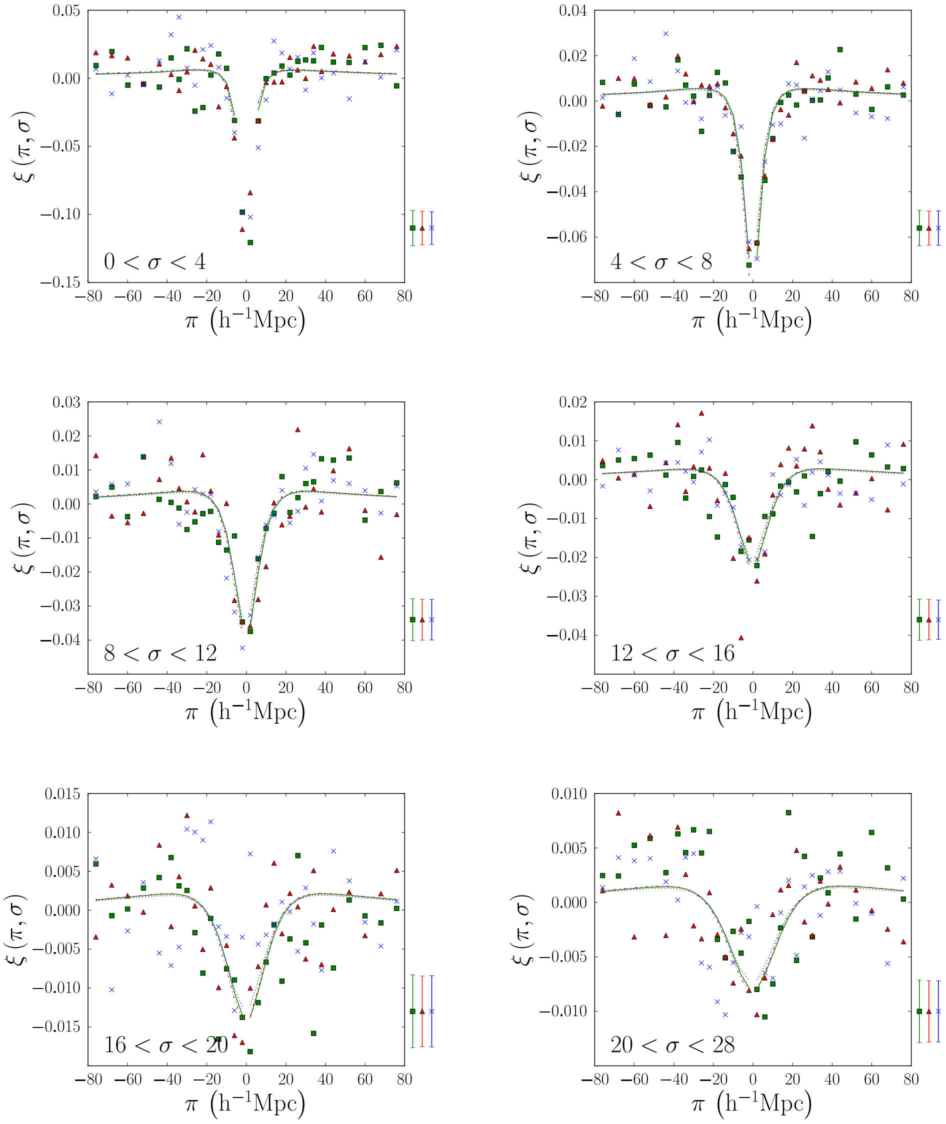


Figure 4.8: Similar to figure 4.2 on page 83 but for the sub-samples in column density N1, N2, and N3. In this figure from left to right $0\,h^{-1}\mathrm{Mpc} < \sigma < 4\,h^{-1}\mathrm{Mpc}$, $4\,h^{-1}\mathrm{Mpc} \leqslant \sigma/8\,h^{-1}\mathrm{Mpc}$, $8\,h^{-1}\mathrm{Mpc} \leqslant \sigma < 12\,h^{-1}\mathrm{Mpc}$, $12\,h^{-1}\mathrm{Mpc} \leqslant \sigma < 16\,h^{-1}\mathrm{Mpc}$, $16\,h^{-1}\mathrm{Mpc} \leqslant \sigma < 20\,h^{-1}\mathrm{Mpc}$, and $20\,h^{-1}\mathrm{Mpc} \leqslant \sigma < 28\,h^{-1}\mathrm{Mpc}$. Figure 4.9 on the next page contain similar plots for the remaining bins in $\sigma$. In all panels green squares, red triangles, and blue crosses stand for samples N1, N2, and N3 respectively. Green solid lines, red dashed-dotted lines, and blue dotted lines correspond to the best-fit model considering $r_{\mathrm{min}} = 5\,h^{-1}\mathrm{Mpc}$ fitted to samples N1, N2, and N3 respectively. Errorbars at the side of the plot show the typical error of the points in the corresponding $\sigma$ bin.

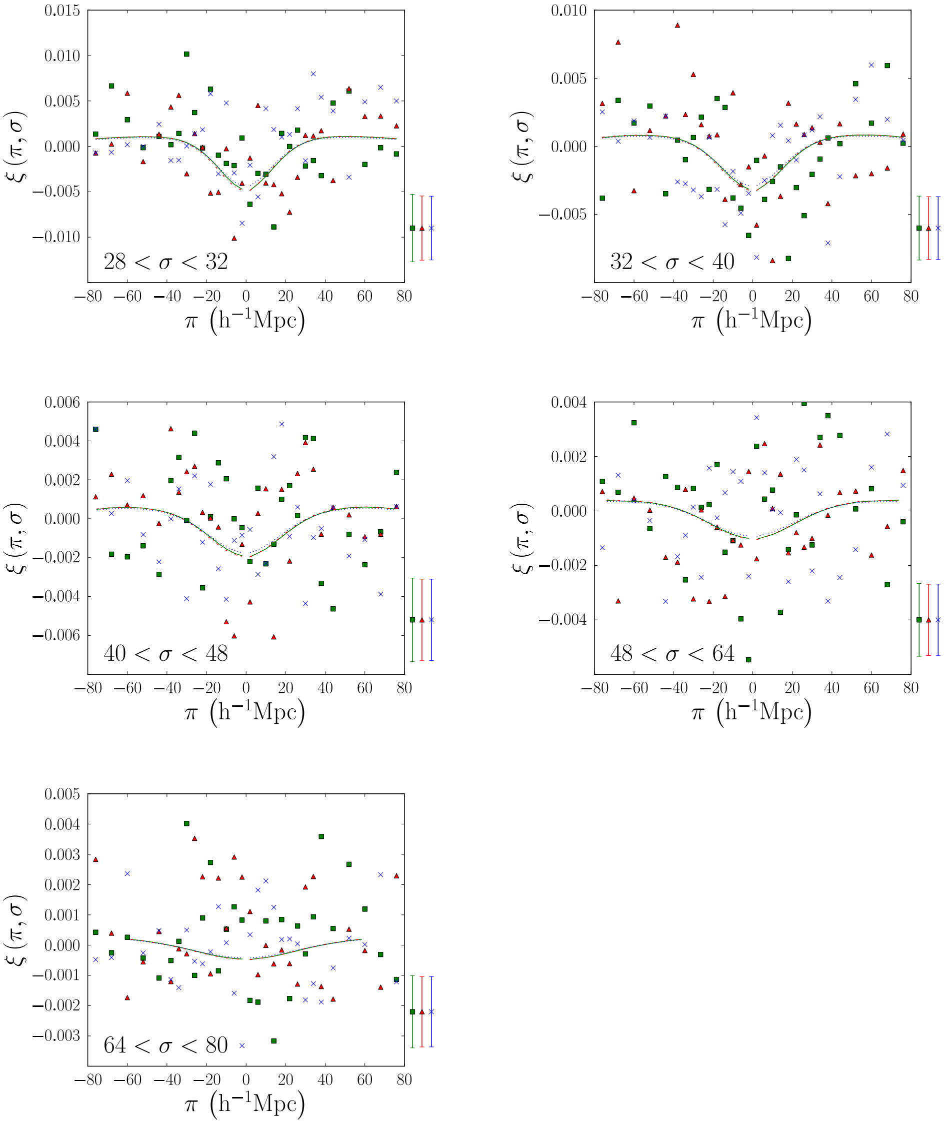


Figure 4.9: From left to right and from top to bottom, same as figure 4.8 on page 90 but for bins with $28\,h^{-1}\mathrm{Mpc} \leqslant \sigma < 32\,h^{-1}\mathrm{Mpc}$, $32\,h^{-1}\mathrm{Mpc} \leqslant \sigma < 40\,h^{-1}\mathrm{Mpc}$, $40\,h^{-1}\mathrm{Mpc} \leqslant \sigma < 48\,h^{-1}\mathrm{Mpc}$, $48\,h^{-1}\mathrm{Mpc} \leqslant \sigma < 64\,h^{-1}\mathrm{Mpc}$, $64\,h^{-1}\mathrm{Mpc} \leqslant \sigma < 80\,h^{-1}\mathrm{Mpc}$.

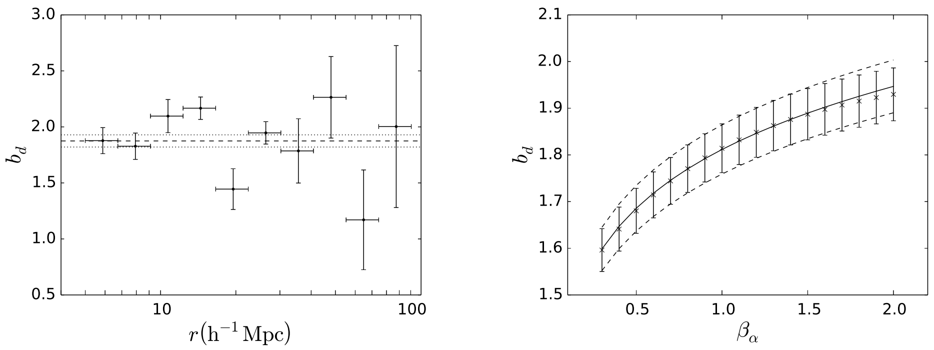


Figure 4.10: *Left*: Bias of the DLA, $b_d$, against $r$ obtained by fitting sample A but considering only certain intervals in $r$, shown by the horizontal errorbars. Dashed lines indicate the value obtained by fitting the entire sample A. Dotted lines indicate the $1\sigma$ interval for this measurement. *Right*: Best fit value of $b_d$ for different assumed values of $\beta_\alpha$. The solid and dashed lines show the fitting formula described in equation 4.34.

The results of these fits are shown in left panel of figure 4.10, and in table 4.3 on page 89. We find no clear dependence of the bias with $r$. This is expected if DLAs are indeed tracers of the underlying dark matter distribution (Mo & White 1996). However, we find a high dispersion around the value measured for the entire range in $r$. This may indicate that the errors given by *baofit* are being underestimated.

### 4.5.5 Dependence on the Lyα bias

Until now, all the fits for the DLA bias have assumed fixed values for the Lyα parameters. In this section we relax this condition. Due to the fact that the constraints on the combination $b_\alpha\,(1+\beta_\alpha)$ are significantly tighter than those on $\beta_\alpha$, we have examined the dependence of our measurement for the DLA bias on the assumed value of $\beta_\alpha$. We repeat the fit assuming different values of the redshift space distortion for the Lyα forest.

In right panel of figure 4.10 and in table 4.4 on the next page we present the results of such a study applied to sample A. We can see that the dependence can indeed be well described using a fitting formula with $b_d \propto \beta_\alpha^n$ (as suggested by Font-Ribera et al. 2012). We find

$$b_d = (1.87 \pm 0.05)\,(\beta_\alpha/1.39)^{0.10}\ , \tag{4.34}$$

which is not the same evolution as that found in Font-Ribera et al. (2012) ($b_d \propto \beta_\alpha^{0.22}$). As we have seen in section 4.5.1 on page 82, this difference may help explain the discrepancies in the measured values of $b_d$. The reason for this difference is still unknown to us. This issue will have to be addressed in the near future.

| $\beta_\alpha$ | $b_d$ | $\chi^2$ (d.o.f.) |
|---|---|---|
| 0.30 | $1.60 \pm 0.05$ | 2939.17 (2896-2) |
| 0.40 | $1.64 \pm 0.05$ | 2937.60 (2896-2) |
| 0.50 | $1.68 \pm 0.05$ | 2937.57 (2896-2) |
| 0.60 | $1.71 \pm 0.05$ | 2938.70 (2896-2) |
| 0.70 | $1.74 \pm 0.05$ | 2940.71 (2896-2) |
| 0.80 | $1.77 \pm 0.05$ | 2943.38 (2896-2) |
| 0.90 | $1.79 \pm 0.05$ | 2946.53 (2896-2) |
| 1.00 | $1.81 \pm 0.05$ | 2950.04 (2896-2) |
| 1.10 | $1.83 \pm 0.05$ | 2953.80 (2896-2) |
| 1.20 | $1.85 \pm 0.05$ | 2957.72 (2896-2) |
| 1.30 | $1.86 \pm 0.05$ | 2961.74 (2896-2) |
| 1.40 | $1.88 \pm 0.05$ | 2965.82 (2896-2) |
| 1.50 | $1.89 \pm 0.05$ | 2969.91 (2896-2) |
| 1.60 | $1.90 \pm 0.06$ | 2973.98 (2896-2) |
| 1.70 | $1.91 \pm 0.06$ | 2978.02 (2896-2) |
| 1.80 | $1.92 \pm 0.06$ | 2982.00 (2896-2) |
| 1.90 | $1.92 \pm 0.06$ | 2985.92 (2896-2) |
| 2.00 | $1.93 \pm 0.06$ | 2989.76 (2896-2) |

Table 4.4: Best fit value of $b_d$ for different assumed values of $\beta_\alpha$.

## 4.6 Summary and Conclusions

We have measured the cross-correlations of DLAs and the Ly$\alpha$ forest for different samples (see table 4.1 on page 72). We make use of the final DR from BOSS. The different samples consist of a set of global samples (using all DLAs, with different cuts) and samples with extra cuts on redshift or column density. We have measured the DLA bias factor for all these samples. Our main conclusions are:

- We measure $b_d = 1.87 \pm 0.05$. This value is a little bit lower than that reported by Font-Ribera et al. (2012). The reason for this discrepancy is still not clear. We propose two possible explanations that will have to be tested in the near future. First, the projection applied to correct for the quasar continuum fitting may not be accurate enough, or the MTC correction proposed by Font-Ribera et al. (2012) overestimate the correction. Second, the observed dependence of the DLA bias on the redshift space distortion parameter for the Ly$\alpha$ forest measured here is different from that measured by Font-Ribera et al. (2012). This dependence plays an important role in the correction of $b_d$ due to the fact that we are using different parameters and cosmologies than Font-Ribera et al. (2012). Any of these reasons (or a combination of the two) may reconcile the measurements of the DLA.
- We observe hints of a redshift dependence for the DLA bias of low statistical significance.

However, the comparison of the evolution measured here and that reported by Font-Ribera et al. (2012) suggests that the bias of the DLAs is constant with redshift. More statistics are required to clarify this point.

- We reject any dependence of the DLA bias on column density at the level of significance provided by our data. We have detected a small decrease of the DLA bias at high column density, but it is of low statistical significance.
- We have analysed the scale dependence of the DLA bias and found that the value of the DLA bias is consistent with being constant in the explored range. This test shows that the errors given by the fitter *baofit* may be underestimated.

## 4.A Ly$\alpha$ autocorrelation contribution to the covariance matrix

To properly estimate the contribution to the covariance matrix of pixels in the same spectrum one has to take several things into account. The most important factor to take into account is that the pixels have a certain width. This means that we are averaging the flux in some wavelength range. One has to take into account other things such as the continuum fitting errors, the PSF of the instrument, and so on and so forth. Also, the contribution will depend on the pixel separation, $r_\parallel$, and the redshift of both pixels. For simplicity, we will consider the redshift dependence to be only through the mean redshift of both pixels. The correlation is, then, the one-dimensional power-spectrum convoluted with a bunch of functions that correspond to the effects we just mentioned.

In practice this means that considering a theoretical model for the Ly$\alpha$ autocorrelation is indeed very complex. Instead, in this work we use the measured 1D power spectrum from Palanque-Delabrouille et al. (2013). In their analysis, they also use SDSS-III data, which means that the instrument used is the same. However, they do not use the analysis pixels as we do but the measured pixels obtained directly after the pipeline reduction. We can therefore assume that all the effects but the averaging of pipeline pixels into analysis pixels are already included in their measurement of the 1D power spectrum. The only exception is the spectral resolution, which was corrected for in the 1D power spectrum measurement. Thus, we have to reintroduce that effect to properly estimate the contribution to the covariance matrix.

To take the averaging of pipeline pixels into account we consider a top-hat function in real space.

$$W(r) = \begin{cases} 1/A & ; \quad |r| < A/2 \\ 0 & ; \quad |r| > A/2 \end{cases} , \tag{4.35}$$

where $A = 201\,\mathrm{km\,s^{-1}}$ is the pixel's width. Note that it is convenient to express this quantity in $\mathrm{km\,s^{-1}}$ since the 1D power spectrum computed in Palanque-Delabrouille et al. (2013) is also expressed in such units. In Fourier space this transforms as

$$\begin{aligned} W(k) &= \int_0^\infty W(r)e^{-ikr}dr = \frac{1}{A}\int_{-A/2}^{A/2} e^{-ikr}dr = \frac{1}{Ak/2}\left.\frac{e^{-ikr}}{-2i}\right|_{-A/2}^{A/2} = \\ &= \frac{1}{Ak/2}\frac{e^{ikA/2}-e^{-ikA/2}}{2i} = \frac{\sin(kA/2)}{kA/2} \,. \end{aligned} \tag{4.36}$$

Thus, the correlation reads

$$\begin{aligned} \xi_n\left(\frac{z_i+z_j}{2}\right) &= \int_{-\infty}^{\infty}\frac{dk}{2\pi}P^{1D}\left(k,\frac{z_i+z_j}{2}\right)\frac{\sin(kA/2)}{kA/2}e^{-iknA}\frac{1}{\sqrt{2\pi}\sigma_{PSF}}\exp\left(-\frac{k^2}{2\sigma_{PSF}^2}\right) = \\ &= 2\int_0^\infty\frac{dk}{2\pi}P^{1D}\left(k,\frac{z_i+z_j}{2}\right)\frac{\sin(kA/2)}{kA/2}\frac{\cos(knA/2)}{\sqrt{2\pi}\sigma_{PSF}}\exp\left(-\frac{k^2}{2\sigma_{PSF}^2}\right) \,, \end{aligned} \tag{4.37}$$

where $\sigma_{PSF} = 70\,\mathrm{km\,s^{-1}}$ is the average value for the PSF. Note that the PSF changes slightly from pixel to pixel but that, in the level of precision we are working with, the variations of the PSF are negligible.

## 4.B Limits on the linear theory

As discussed in section 4.4 on page 78 we are assuming that our measurements (or more precisely the part of the measurements used in the fitting process) are in the linear regime. As previously stated it is not clear which is the smallest distance at which we can safely assume the linear regime. The aim of this appendix is not to find such a limit, but to convince the reader that the ranges considered here are indeed in the linear regime or that, at least, the deviations from the linear regime are small.

In order to do so, we explore the dependence of the model on the minimum distance considered for the analysis. We fit the model considering $r_{min} = 4\,h^{-1}$Mpc, $r_{min} = 5\,h^{-1}$Mpc, $r_{min} = 6\,h^{-1}$Mpc, $r_{min} = 7\,h^{-1}$Mpc, $r_{min} = 8\,h^{-1}$Mpc, $r_{min} = 9\,h^{-1}$Mpc, and $r_{min} = 10\,h^{-1}$Mpc for sample A. Even though the fits are performed using all the bins in $\sigma$, here we only show the first six bins in $\sigma$ as an example (see figure 4.11 on the next page; note that only the cases with $r_{min} = 4\,h^{-1}$Mpc, $r_{min} = 6\,h^{-1}$Mpc, $r_{min} = 8\,h^{-1}$Mpc, and $r_{min} = 10\,h^{-1}$Mpc are plotted). The values of the fitted parameters are listed in table 4.5 on page 97. While there are discrepancies between the models of a few %, the measured bias is consistent between the models even though there seems to be a systematic decrease on the bias with $r_{min}$. The reduced $\chi^2$ is roughly the same for the different models.

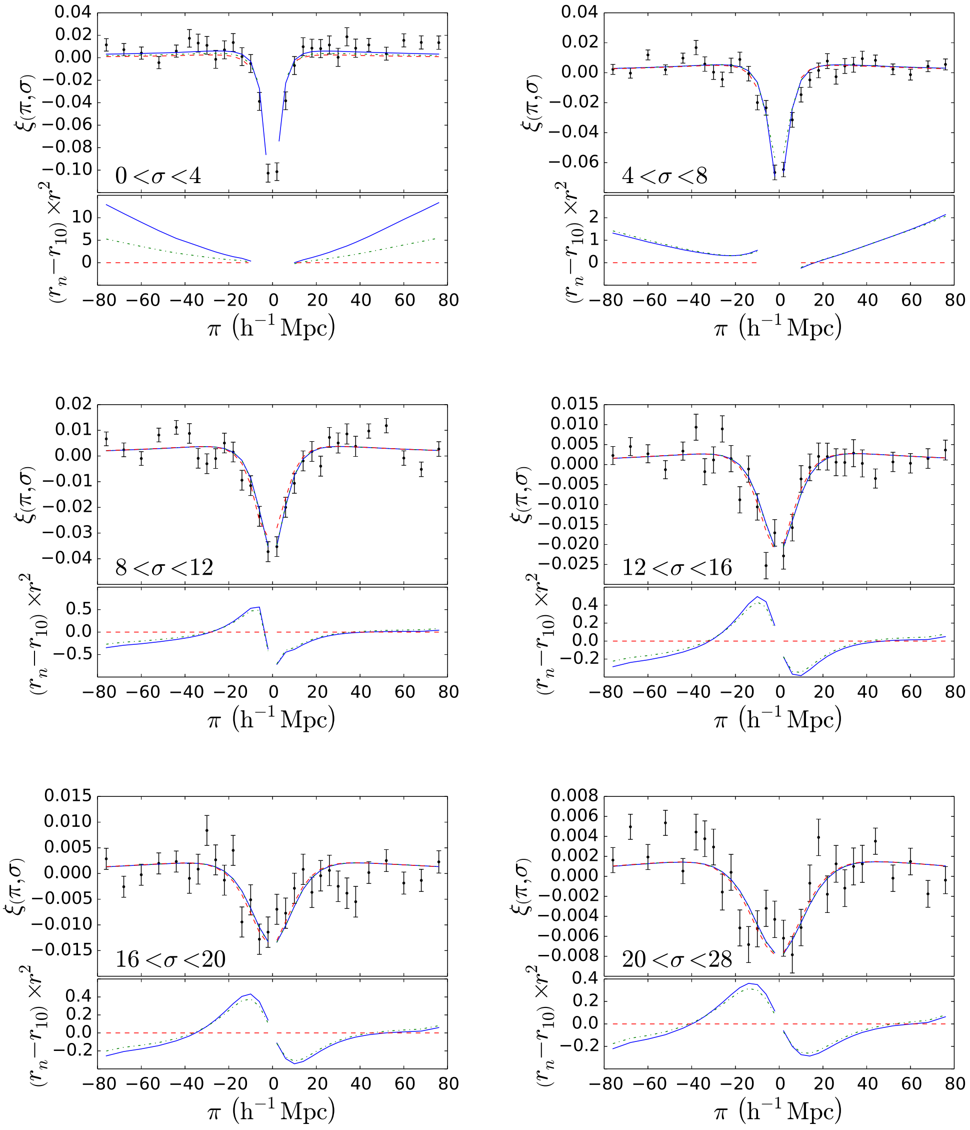


Figure 4.11: Measurement of the cross-correlation for different bins in $\sigma$: from left to right $0\,h^{-1}\mathrm{Mpc} < \sigma < 4\,h^{-1}\mathrm{Mpc}$, $4\,h^{-1}\mathrm{Mpc} \leqslant \sigma < 8\,h^{-1}\mathrm{Mpc}$, $8\,h^{-1}\mathrm{Mpc} \leqslant \sigma < 12\,h^{-1}\mathrm{Mpc}$, $12\,h^{-1}\mathrm{Mpc} \leqslant \sigma < 16\,h^{-1}\mathrm{Mpc}$, $16\,h^{-1}\mathrm{Mpc} \leqslant \sigma < 20\,h^{-1}\mathrm{Mpc}$, $20\,h^{-1}\mathrm{Mpc} \leqslant \sigma < 28\,h^{-1}\mathrm{Mpc}$. The values of $\sigma$ are expressed in $h^{-1}\mathrm{Mpc}$. In all panels blue circles are for sample A. Blue solid line, green dashed-dotted line, black dotted line, and red dashed line correspond to the best-fit model considering $r_{\mathrm{min}} = 4\,h^{-1}\mathrm{Mpc}$, $r_{\mathrm{min}} = 6\,h^{-1}\mathrm{Mpc}$, $r_{\mathrm{min}} = 8\,h^{-1}\mathrm{Mpc}$, and $r_{\mathrm{min}} = 10\,h^{-1}\mathrm{Mpc}$ respectively. Bottom panels show the $r^2$-weighted differences between the different models and the best-fit model considering $r_{\mathrm{min}} = 10\,h^{-1}\mathrm{Mpc}$.

| $r_{\min}\,(h^{-1}\mathrm{Mpc})$ | $b_d$ | $\chi^2(\mathrm{d.o.f.})$ |
|---|---|---|
| 4 | $1.87 \pm 0.05$ | 2970.37 (2898-2) |
| 5 | $1.88 \pm 0.05$ | 2965.22 (2896-2) |
| 6 | $1.88 \pm 0.06$ | 2951.99 (2888-2) |
| 7 | $1.88 \pm 0.06$ | 2958.96 (2888-2) |
| 8 | $1.89 \pm 0.07$ | 2951.60 (2878-2) |
| 9 | $1.89 \pm 0.07$ | 2944.27 (2874-2) |
| 10 | $1.90 \pm 0.08$ | 2939.15 (2864-2) |

Table 4.5: Measured bias of the DLA when fitting the same model but cutting at different $r_{min}$.

# 5

# Quasar-Ly$\alpha$ forest cross-correlations

We present a measurement of the cross-correlation between quasar and theLy$\alpha$ forest using the final DR of BOSS. We study the small scales effects on this cross-correlation. These effects include the metal contamination of the Ly$\alpha$ forest and the effect of the quasar radiation in the nearby hydrogen clouds. We observe contamination from the metal lines corresponding to the Si II doublet at 1190.42Å and 1193.30Å, and the Si III at 1296.50Å. However, we do not observe the expected contamination from other metal lines, such as N V at 1240.15Å. At scales smaller than $\sim 20\,\mathrm{h}^{-1}$ Mpc we detect a small asymmetry with respect to the relative position of the hydrogen cloud responsible for the Ly$\alpha$ absorption and the quasar, of low statistical significance. This asymmetry can arise from the effect the quasar radiation has on the nearby hydrogen clouds. We present a theoretical model that predicts how this effect affects the cross-correlation. The model is presented both in real and Fourier space.

## 5.1 Introduction

Quasars are the most luminous known objects in the universe and are found amongst the most distant as well. As such, these objects are used for many different purposes. Of interest here are two of these applications. First, quasars are used as tracers of the underlying dark matter distribution. Second, the quasar absorption spectra bluewards of the Lyα emission line reveal the Lyα forest.

The quasars were first used as density tracers by the measurements of their large-scale clustering by 2dF Survey (e.g. Porciani et al. 2004; Croom et al. 2005; da Ângela et al. 2005), and by SDSS (Myers et al. 2006; Shen et al. 2007; Ross et al. 2009). In the context of linear theory and at large scales, the quasar correlation function should follow that of the underlying DM distribution times the square of the bias factor of their host halos. Clustering analysis of quasars reported a bias factor of $b_q = 3.8 \pm 0.3$ at redshift $z = 2.4$ and that this bias factor was nearly independent of the luminosity of the quasars (Shen et al. 2009; White et al. 2012).

The Lyα forest is also known to be a tracer of the underlying matter density. Recently, there have been many studies of the Lyα forest autocorrelation. These studies include measurements of the one dimensional Lyα forest power spectrum (Palanque-Delabrouille et al. 2013) and the three-dimensional autocorrelation (Delubac et al. 2015; Blomqvist et al. 2015). The Lyα forest bias factor and its redshift space distortion parameter are determined with errors of a few %.

Provided that the Lyα bias factor is independently estimated elsewhere, another way to estimate the bias factor of quasars is to analyse the cross-correlation with the Lyα. Recent studies from the analysis of the cross-correlation of quasars with the Lyα forest gives a bias factor of $b_q = 3.64^{+0.13}_{-0.15}$ (Font-Ribera et al. 2013), which were consistent with previous estimations from the clustering analysis (White et al. 2012). More recent analysis have even measured the BAO peak position in the quasar-Lyα forest cross-correlation (Font-Ribera et al. 2014).

The quasar-Lyα forest cross-correlation was first measured along the same line of sight of each individual quasar in an attempt to understand the *proximity effect* (Carswell et al. 1982; Murdoch et al. 1986; Bajtlik et al. 1988; Scott et al. 2000). The proximity effect designates the impact of the quasar ionizing radiation on the nearby hydrogen clouds, reducing the Lyα absorption. The ionizing radiation emitted by the quasar is added to the cosmic ionizing background, increasing the ionized fraction of the nearby clouds. Thus, the Lyα absorption is reduced in those clouds.

The proximity effect can not only affect hydrogen clouds in the same line of sight of the quasar, but also in the line of sight of nearby quasars. This was investigated in several papers by examining individual quasar pairs separated by a few arc minutes (Jakobsen et al. 1986; Crotts 1989; Møller & Kjaergaard 1992; Liske & Williger 2001) and by using a large number of pairs at wider separations (Schlegel et al. 1998; Schirber et al. 2004; Croft 2004; Rollinde et al. 2005; Guimarães et al. 2007). The overall conclusion of this studies is that the Lyα absorption is stronger near the quasars, opposed to the weaker absorption predicted by the proximity effect. The reason for this is that near the quasars there is not only more ionizing radiation, but also an overdensity of hydrogen clouds. Also, the anisotropy present in the quasar radiation and the time variability of quasars add into the mix, further antagonizing a robust interpretation of this measurement.

In this chapter we focus on the study of the small scales quasar-Ly$\alpha$ forest cross-correlation and address some of this open issues concerning the effect of the quasar radiation. We make use of the BOSS DR12Q catalogue, which is described in section 5.2. The measurement of the cross-correlation is presented in section 5.3 on the following page, and the observed small-scales effects are discussed also in section 5.3 on the next page. In sections 5.4 and 5.5 on page 106 and on page 111 we present a simple theoretical model to include the effect of the quasar radiation to nearby hydrogen clouds on the quasar-Ly$\alpha$ forest cross-correlation. The model is presented in both redshift space and Fourier space. Finally, we summarize our conclusions in section 5.6 on page 114. Note that throughout this chapter we use the cosmology from Planck Collaboration et al. (2015), i.e., a flat $\Lambda$CDM cosmology with $\Omega_m = 0.3089$, $\Omega_b = 0.0486$, $h = 0.6774$, $n_s = 0.9667$, and $\sigma_8 = 0.8159$.

## 5.2 Data Sample

The DR12 of the SDSS-III Collaboration (Gunn et al. 1998; York et al. 2000; Gunn et al. 2006; Eisenstein et al. 2011; Bolton et al. 2012; Smee et al. 2013) contains the final sample of BOSS (Dawson et al. 2013). The quasar target selection used in BOSS is summarized in Ross et al. (2012), and combines different targeting methods described in Yèche et al. (2010); Kirkpatrick et al. (2011); Bovy et al. (2011).

In this section we describe the data set used in this study. We measure the cross-correlation of two tracers of the underlying density field. These are the number density of quasars and the Ly$\alpha$ absorption along a set of lines of sight. The quasars used as tracers are designated here as *quasar sample* and the quasar lines of sight where the Ly$\alpha$ absorption is measured are designated as *Ly$\alpha$ sample*. For each of the samples the catalogue and cuts as described in sections 5.2.1 and 5.2.2 on the current page and on the following page.

### 5.2.1 Quasar Sample

For the quasar sample we use the DR12Q, an updated version of the DR9Q catalogue (Pâris & et al. 2016). This catalogue contains a total of 297,301 visually confirmed quasars, distributed in an area of 9,376 square degrees in two disconnected parts of the sky: The South Galactic Cap and the North Galactic Cap. In this study, however, we only use 189,715 of these quasars as tracers. In order to use a quasar, we require the redshift of the quasars to be in the range $2.0 \leqslant z_q < 3.5$. The lower limit is set as the performance of the BOSS spectrograph rapidly deteriorates at wavelengths bluer than $\lambda \sim 3650$Å, which corresponds to a Ly$\alpha$ absorption of $z = 2$. The upper limit is set as the number of identified quasars drops rapidly at high redshift. Note that in this chapter we use the visual inspection redshift estimate Z_VI. The average redshift of the selected quasars is 2.54 and its distribution is plotted in figure 5.1 on the next page.

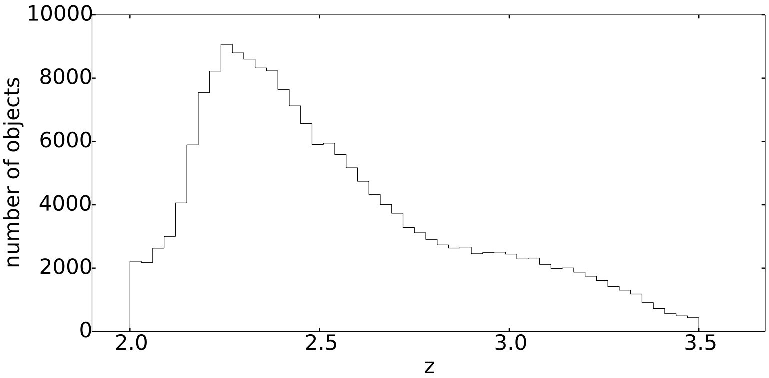


Figure 5.1: Redshift distribution of the 189,715 quasars used as density tracers.

### 5.2.2 Ly$\alpha$ Sample

For the Ly$\alpha$ sample we use the DR12 Ly$\alpha$ spectra computed as in Busca et al. (2013) (N. Busca, private communication). This corresponds to a total of 157922 spectra containing over 27 million Ly$\alpha$ pixels. In particular we use their *analysis pixels* that are the flux average over three adjacent pipeline pixels. Throughout the rest of this chapter, *pixel* refers to analysis pixels unless otherwise stated. The effective width of these pixels is 210 km s$^{-1}$. For the quasar continuum fit their method 1 was used.

## 5.3 Quasar-Ly$\alpha$ cross-correlations

We measure the cross-correlation of quasars and the Ly$\alpha$ forest following the procedure described in section 4.3 on page 72. The cross-correlation is measured in bins of 4 h$^{-1}$Mpc for both parallel and perpendicular separation. The results of the quasar-Ly$\alpha$ cross-correlation as a function of $\pi$ are shown for the different bins of $\sigma$ in figures 5.2 to 5.4 on pages 103–105. For visualization purposes alone, the data-points have been averaged in 12 bins in $\sigma$ delimited by (0, 4, 8, 12, 16, 20, 28, 32, 40, 48, 64, 80, 100), and 32 bins in $\pi$ delimited by (-100, -90, -80, -72, -64, -56, -48, -40, -32, -28, -24, -20, -16, -12, -8, -4, 0, 4, 8, 12, 16, 20, 24, 28, 32, 40, 48, 56, 64, 72, 80, 90, 100). Note that all delimitations are expressed in h$^{-1}$Mpc. The average is computed as

$$\xi^B = \frac{\sum_{A\in B} \xi^A/\sqrt{C^{AA}}}{\sum_{A\in B} 1/\sqrt{C^{AA}}}, \tag{5.1}$$

$$\pi^B = \frac{\sum_{A\in B} \pi^A/\sqrt{C^{AA}}}{\sum_{A\in B} 1/\sqrt{C^{AA}}}, \tag{5.2}$$

$$\frac{1}{C^{B_1B_2}} = \sum_{A_1\in B_1}\sum_{A_2\in B_2} \frac{1}{C^{A_1A_2}}, \tag{5.3}$$

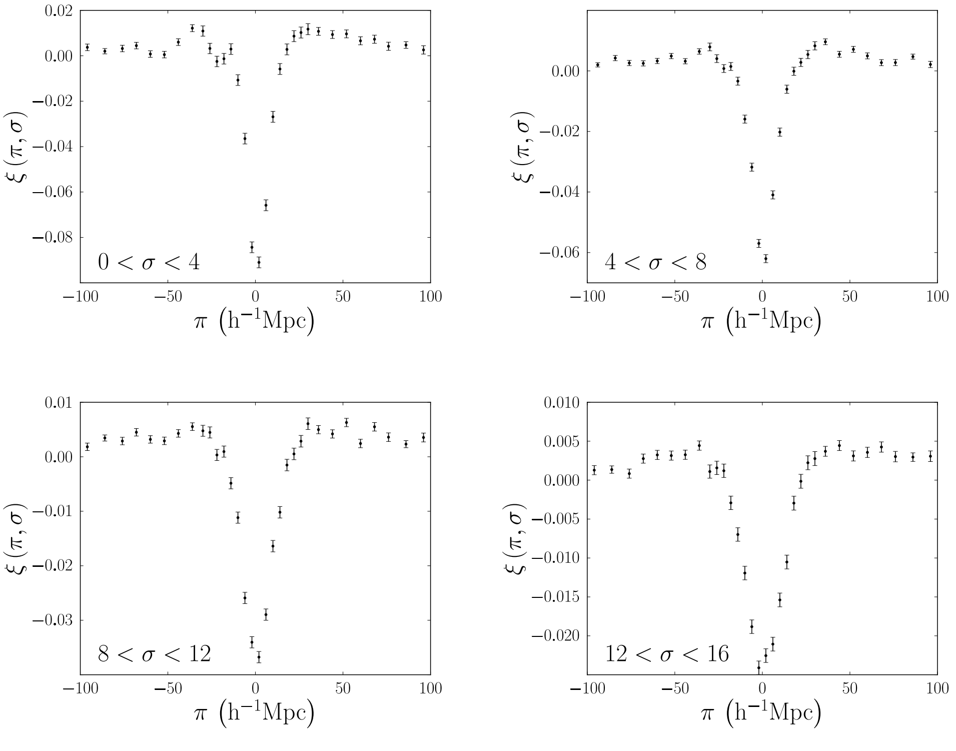


Figure 5.2: Measurement of the cross-correlation for different bins in $\sigma$: from left to right and top to bottom $0\,\mathrm{h}^{-1}\mathrm{Mpc} < \sigma < 4\,\mathrm{h}^{-1}\mathrm{Mpc}$, $4\,\mathrm{h}^{-1}\mathrm{Mpc} \leqslant \sigma/8\,\mathrm{h}^{-1}\mathrm{Mpc}$, $8\,\mathrm{h}^{-1}\mathrm{Mpc} \leqslant \sigma < 12\,\mathrm{h}^{-1}\mathrm{Mpc}$, and $12\,\mathrm{h}^{-1}\mathrm{Mpc} \leqslant \sigma < 16\,\mathrm{h}^{-1}\mathrm{Mpc}$. The values of $\sigma$ are expressed in $\mathrm{h}^{-1}\mathrm{Mpc}$. Figures 5.3 and 5.4 on the following page and on page 105 contain similar plots for the remaining bins in $\sigma$.

where the indexes $A$, $A_1$, and $A_2$ stand for the measured bins, the indexes $B$, $B_1$, and $B_2$ stand for the new bins, and the sum $\sum_{A \in B}$ is over all bins $A$ that are included in bin $B$. The errorbars shown in the plots are the square root of the diagonal elements of the covariance matrix.

The results are also shown in a contour plot in figure 5.5 on page 105. Note that for the contour plot, the smoothing of bins has been done in a slightly different way. For bins with $16\,\mathrm{h}^{-1}\mathrm{Mpc} < r \leqslant 32\,\mathrm{h}^{-1}\mathrm{Mpc}$ we average their value with the bins with $\pi \in (\pi_{\mathrm{bin}} - \delta_\pi, \pi_{\mathrm{bin}} + \delta_\pi)$ and $\sigma \in (\sigma_{\mathrm{bin}} - \delta_\sigma, \sigma_{\mathrm{bin}} + \delta_\sigma)$ weighted by $1/(e + 0.05)$, where $e$ is the squared root of the corresponding diagonal element of the covariance matrix. For bins with $r > 32\,\mathrm{h}^{-1}\mathrm{Mpc}$ we average their value with the bins with $\pi \in (\pi_{\mathrm{bin}} - 2\delta_\pi, pi_{\mathrm{bin}} + 2\delta_\pi)$ and $\sigma \in (\sigma_{\mathrm{bin}} - 2\delta_\sigma, \sigma_{\mathrm{bin}} + 2\delta_\sigma)$ weighted in the same way.

In the uppermost left panel in figure 5.2, corresponding to $0\,\mathrm{h}^{-1}\mathrm{Mpc} < \sigma < 4\,\mathrm{h}^{-1}\mathrm{Mpc}$ we can clearly see two bumps at negative values of $\pi$. Starting from zero and moving to the more negative values of $\pi$ the bumps correspond to contamination from metals. In particular

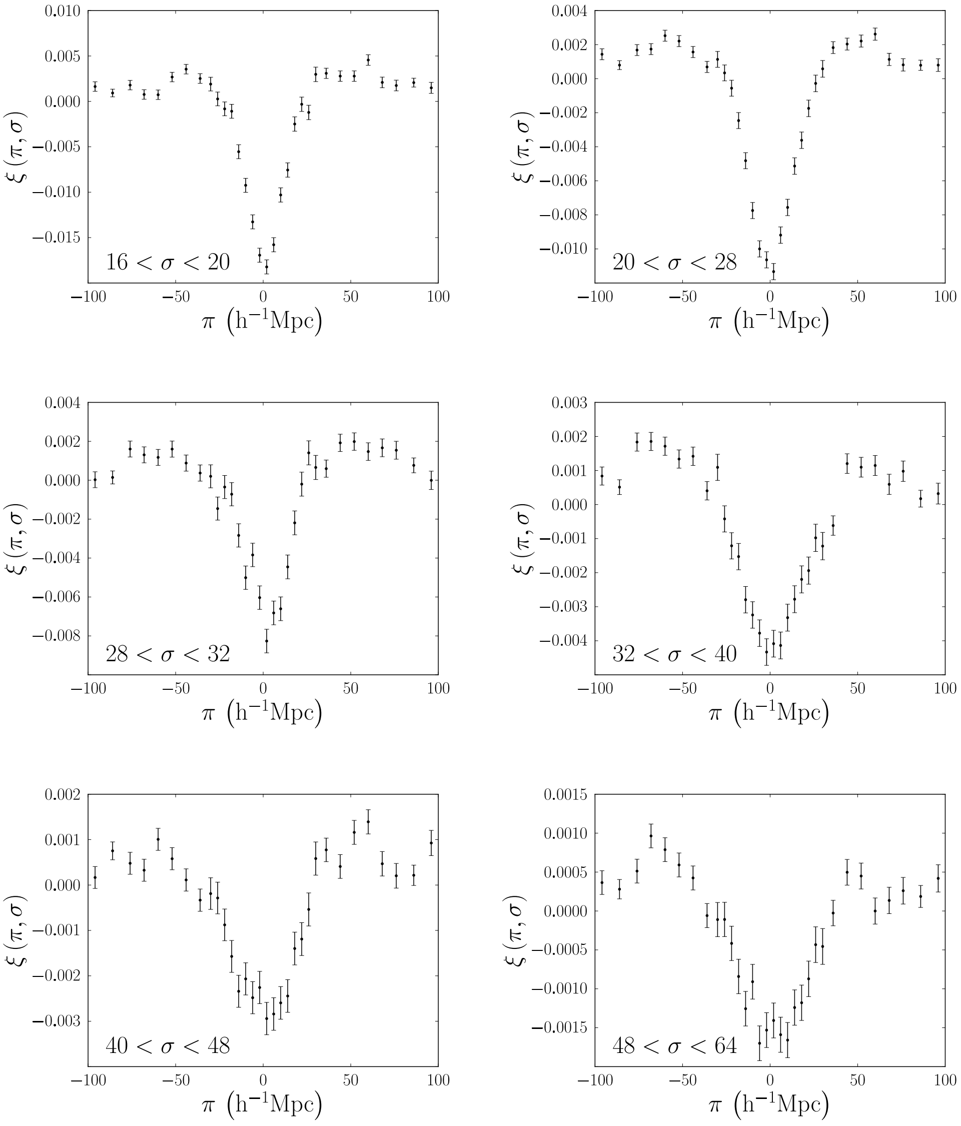


Figure 5.3: From left to right and from top to bottom, same as figure 5.2 on page 103 but for bins with $16\,h^{-1}\mathrm{Mpc} \leqslant \sigma < 20\,h^{-1}\mathrm{Mpc}$, $20\,h^{-1}\mathrm{Mpc} \leqslant \sigma < 28\,h^{-1}\mathrm{Mpc}$, $28\,h^{-1}\mathrm{Mpc} \leqslant \sigma < 32\,h^{-1}\mathrm{Mpc}$, $32\,h^{-1}\mathrm{Mpc} \leqslant \sigma < 40\,h^{-1}\mathrm{Mpc}$, $40\,h^{-1}\mathrm{Mpc} \leqslant \sigma < 48\,h^{-1}\mathrm{Mpc}$, and $48\,h^{-1}\mathrm{Mpc} \leqslant \sigma < 64\,h^{-1}\mathrm{Mpc}$.

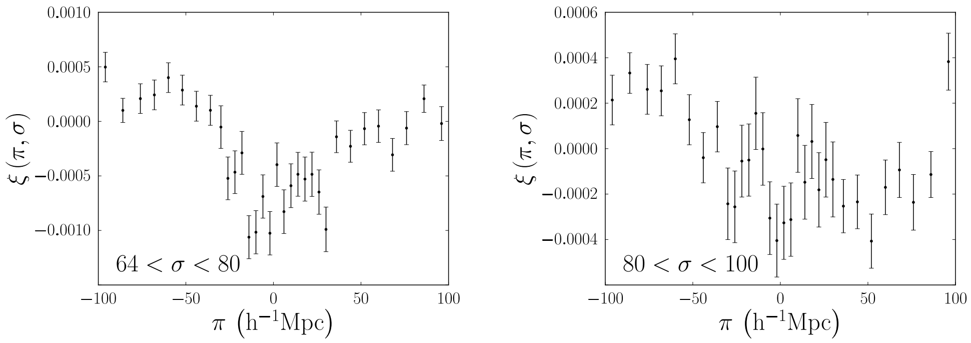


Figure 5.4: From left to right, same as figure 5.2 on page 103 but for bins with $64\,\mathrm{h^{-1}\,Mpc} \leqslant \sigma < 80\,\mathrm{h^{-1}\,Mpc}$ and $80\,\mathrm{h^{-1}\,Mpc} \leqslant \sigma < 100\,\mathrm{h^{-1}\,Mpc}$.

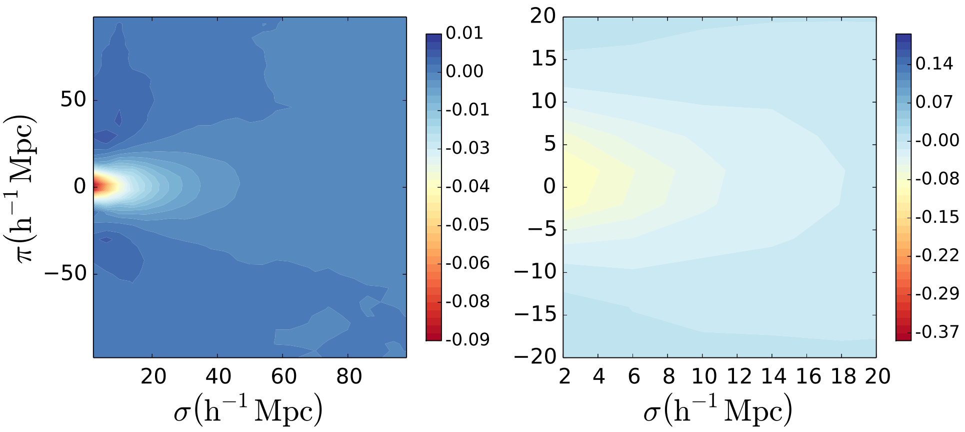


Figure 5.5: Contour plot of the measured quasar-Lyα cross-correlation. The right panel shows a zoom in the central region ($\pi \leqslant 20\,\mathrm{h^{-1}\,Mpc}$, $\sigma \leqslant 20\,\mathrm{h^{-1}\,Mpc}$).

they correspond to the Si II and Si III lines. The Si III absorption line at 1206.50Å[1] is expected at $-21.5\,h^{-1}$Mpc, which corresponds to the position of the first bump. The Si II doublet at 1193.29Å and 1194.10Å[1] is expected at $-52.5\,h^{-1}$Mpc and $-59.3\,h^{-1}$Mpc. The second bump is located also at the expected position. The doublet is responsible for the increase of the width with respect to the first bump. At larger perpendicular distances these bumps are not clearly observed (even though there are hints of their presence up to $\sigma \sim 20\,h^{-1}$Mpc). This suggests that the redshift space distortion parameter for this metals is lower than that of the Ly$\alpha$ forest. Otherwise, the same relative strength would be expected.

If we analyse the right half of this panel, corresponding to positive values of $\pi$, we find that no evidence of contamination from metals or any other strange feature. This means that the contamination from N V line at 1240.146Å[1], expected at $57.4\,h^{-1}$Mpc, is much weaker than that of Si II and Si III.

Finally, to finish our analysis of the cross-correlation at small-scales, we shift our focus to the central shapes of the cross-correlation in the different panels. There are hints of an asymmetry between the positive and negative values of $\pi$. This hints are more clearly seen in the contour plot (figure 5.5 on page 105). However, if present, this asymmetry would be a small effect. To enhance this effect, we compute $\xi\left(\pi,\sigma\right) - \xi\left(-\pi,\sigma\right)$ assuming that the mean value of $\pi$ in each bin is exactly in the middle of the interval considered in that particular bin. The results are shown in figures 5.6 to 5.8 on pages 107–109. These figures show that an asymmetry is detected for $\pi \lesssim 20\,h^{-1}$Mpc in the first five bins in $\sigma$ ($\sigma < 20\,h^{-1}$Mpc) even though the statistical significance of this detection low. Note that the asymmetry in the first bin in $\sigma$ can easily be introduced by the metal contamination described above. However, this asymmetry appears to extend to bins of $\sigma$ where the metal contamination is no longer observed. We discuss a possible interpretation of this asymmetry in section 5.4 and expand it in Fourier space in section 5.5 on page 111.

## 5.4 Radiation model

So far, the analysis of both the cross-correlation and the autocorrelation has focused on measuring the bias factor of the different tracers under the assumption that they follow the underlying matter density, and the position of the BAO peak (see e.g. Font-Ribera et al. 2012; Busca et al. 2013; Font-Ribera et al. 2013, 2014; Blomqvist et al. 2015).

The current model for the quasar-Ly$\alpha$ power spectrum, $P_c\left(k,\mu_k,z\right)$, is defined as

$$P_c\left(k,\mu_k,z\right) = b_\alpha\left(1+\beta_\alpha\mu_k^2\right) b_q\left(1+\beta_q\mu_k^2\right) P_L\left(k,\mu_k,z\right)\ , \tag{5.4}$$

where $b_\alpha$ and $b_q$ are the Ly$\alpha$ forest and quasars bias respectively, $\beta_\alpha$ and $\beta q$ are the redshift-space distortion parameter for the Ly$\alpha$ forest and quasars respectively, and $P_L\left(k,\mu_k,z\right)$ is the linear

---

[1] The values of the wavelength have been taken for excitations from the ground state from Morton (2003).

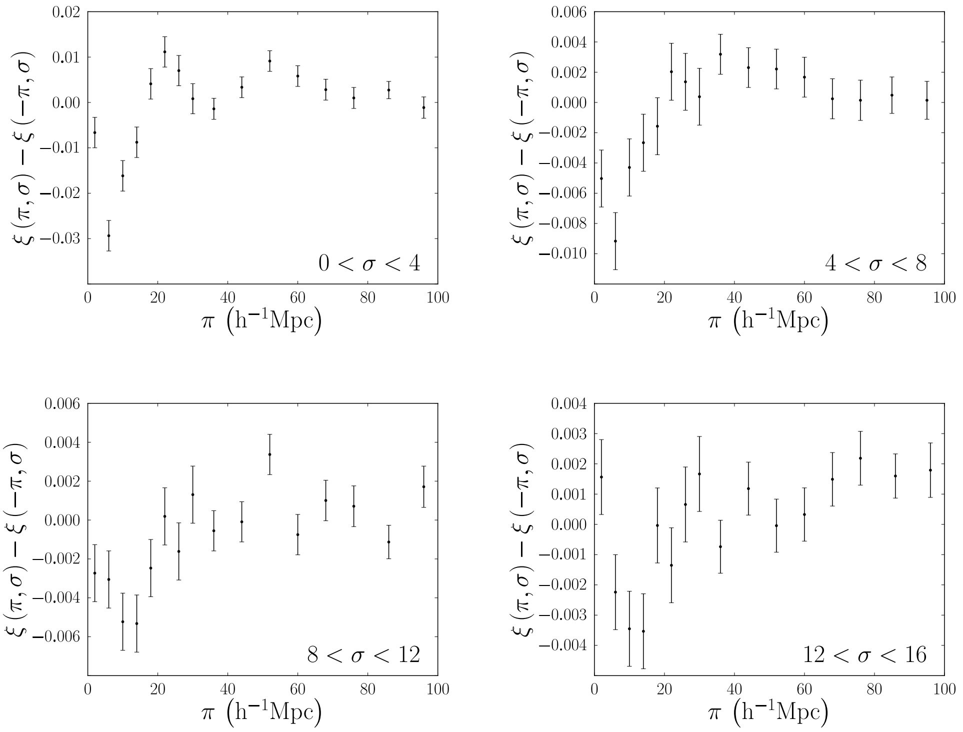


Figure 5.6: $\xi(\pi, \sigma) - \xi(-\pi, \sigma)$ for different bins in $\sigma$: from left to right and top to bottom $0\,h^{-1}\mathrm{Mpc} < \sigma < 4\,h^{-1}\mathrm{Mpc}$, $4\,h^{-1}\mathrm{Mpc} \leqslant \sigma/8\,h^{-1}\mathrm{Mpc}$, $8\,h^{-1}\mathrm{Mpc} \leqslant \sigma < 12\,h^{-1}\mathrm{Mpc}$, and $12\,h^{-1}\mathrm{Mpc} \leqslant \sigma < 16\,h^{-1}\mathrm{Mpc}$. The values of $\pi$ are the averages of the absolute value of $\pi$ in both points. The errors are computed as the quadratic sum of the errors in both points. The values of $\sigma$ are expressed in $h^{-1}\mathrm{Mpc}$. Figures 5.7 and 5.8 on the following page and on page 109 contain similar plots for the remaining bins in $\sigma$. In all panels black circles are for sample A. Black solid lines correspond to the best-fit model considering $r_{\mathrm{min}} = 5\,h^{-1}\mathrm{Mpc}$.

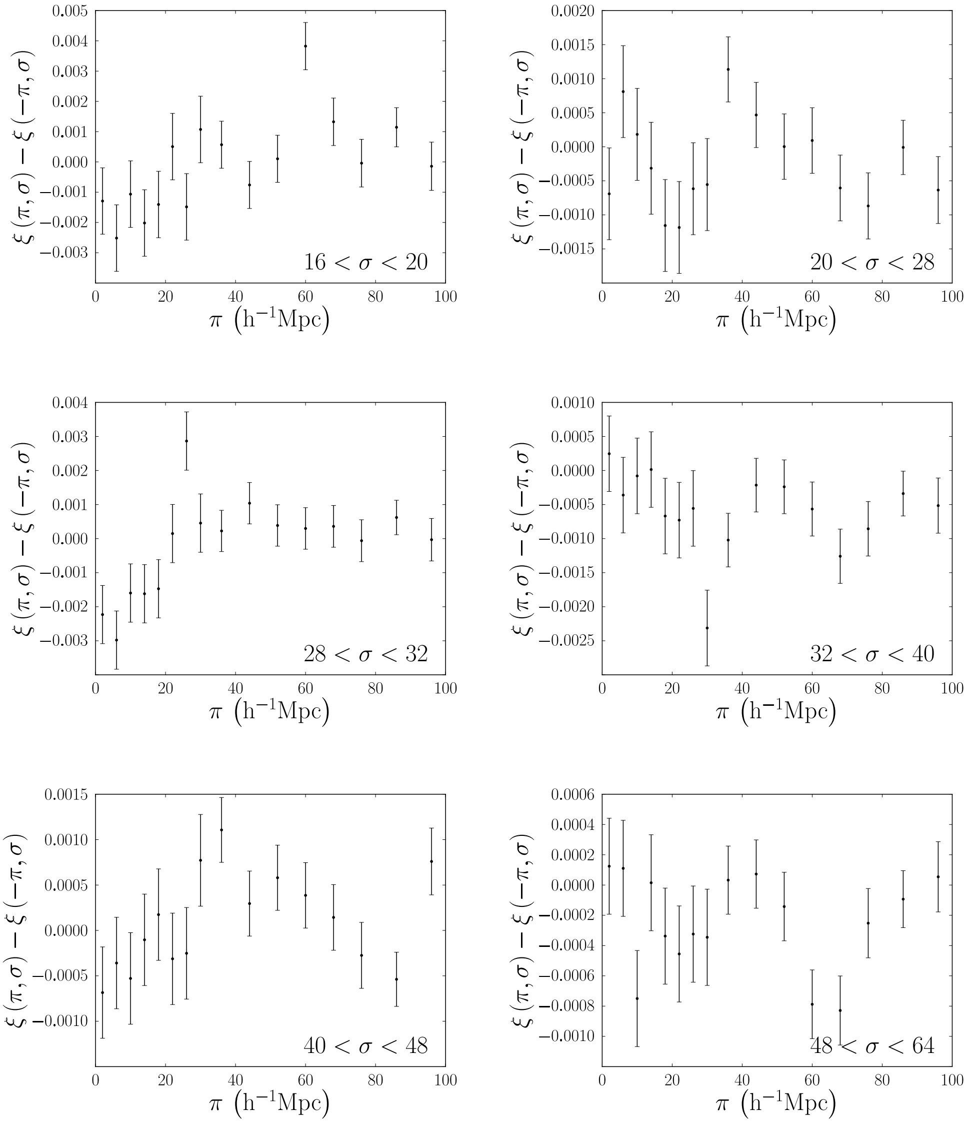


Figure 5.7: From left to right and from top to bottom, same as figure 5.6 on page 107 but for bins with $16\,h^{-1}\mathrm{Mpc} \leqslant \sigma < 20\,h^{-1}\mathrm{Mpc}$, $20\,h^{-1}\mathrm{Mpc} \leqslant \sigma < 28\,h^{-1}\mathrm{Mpc}$, $28\,h^{-1}\mathrm{Mpc} \leqslant \sigma < 32\,h^{-1}\mathrm{Mpc}$, $32\,h^{-1}\mathrm{Mpc} \leqslant \sigma < 40\,h^{-1}\mathrm{Mpc}$, $40\,h^{-1}\mathrm{Mpc} \leqslant \sigma < 48\,h^{-1}\mathrm{Mpc}$, and $48\,h^{-1}\mathrm{Mpc} \leqslant \sigma < 64\,h^{-1}\mathrm{Mpc}$.

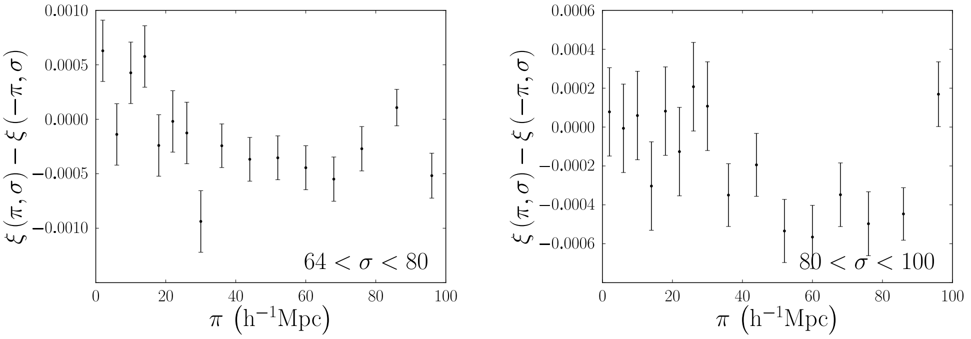


Figure 5.8: From left to right, same as figure 5.6 on page 107 but for bins with $64\,h^{-1}\mathrm{Mpc} \leqslant \sigma < 80\,h^{-1}\mathrm{Mpc}$ and $80\,h^{-1}\mathrm{Mpc} \leqslant \sigma < 100\,h^{-1}\mathrm{Mpc}$.

power spectrum. The quasar-Ly$\alpha$ cross-correlation, $\xi_c\left(r,\mu,z\right)$, is defined as the Fourier transform of $P_c\left(k,\mu_k,z\right)$.

However, on top of this well known effect, the Ly$\alpha$ transmission is also affected by the radiation emitted from the quasars which ionize hydrogen in the surrounding IGM. This results in a small additive contribution to the cross-correlation at small scales. Taking both effects into account, the cross-correlation reads

$$\xi\left(r,\mu,z\right)=\xi_c\left(r,\mu,z\right)+\xi_{\mathrm{rad}}\left(r,\mu,z\right)\ , \tag{5.5}$$

where $\xi_{\mathrm{rad}}$ is the quasar radiation contribution to the cross-correlation. As we will see, this extra term in the cross-correlation might help explain the asymmetries found at small perpendicular separations (see section 5.3 on page 102).

### 5.4.1 Radiation contribtuion to the cross correlation

In order to model this effect we have to take several things into account. In this section we will discuss the different effects considered in this work. For simplicity we shall assume a model without redshift dependence. A more realistic model will have to include this redshift dependence but we leave this to future studies.

First, it is well known that quasar emission is anisotropic. It is not yet clear how does this anisotropic radiation affect the cross-correlation. Here we will assume a toy model that will nevertheless allow us to determine whether or not this contribution is measurable with the current data. This model, $F_a\left(\mu\right)$, reads

$$F_a\left(\mu\right)=1+a\left(1-\mu^2\right)\ . \tag{5.6}$$

Here, $a$ is a free parameter of the model that accounts for the asymmetry of quasar radiation. For

$a = 0$ we would recover a perfectly symmetric radiation, and $a = 1$ corresponds to maximum asymmetry.

The second effect we have to take into account is the time delay between the photons we observe directly from the quasar and the photons that affect the Lyα forest in a parallel line of sight. For example, we see the medium behind the quasar affected by the luminosity of the quasar at an earlier time than for the medium in front of the quasar. This effect should introduce an asymmetry in the quasar-Lyα cross-correlation between positive and negative values of $\pi$ (meaning that the Lyα pixel is located behind or in front of the quasar respectively), which depends on the typical quasar lifetime. Note that because quasars are selected in a flux-limited sample, one expects that the average luminosity of any quasar is lower than the presently observed one as we probe further into the past. Therefore, this asymmetry probes directly the average quasar lifetime, which has not been constrained by many observations so far (except for a similar effect due to He II ionization for which there are very few quasars observed). Here we model this effect as a simple exponential decay

$$F_{\nu} = e^{-\tau/t_q} , \tag{5.7}$$

where $t_q$ is the typical quasar lifetime times the speed of light, and $\tau = r + \pi = r\,(1 + \mu)$.

The contribution to the cross correlation will be proportional to $F_a$ and $F_{\nu}$. Namely,

$$\xi_{rad}\,(r, \mu) \propto F_a\,(\mu)\,F_{\nu}\,(r, \mu) \tag{5.8}$$

### 5.4.2 Amplitude of the effect

In the absence of any absorption, the amplitude of this effect should be $b_{\gamma}\,(r_p/r)^2$, where $b_{\Gamma}$ is the radiation bias factor with respect to the underlying matter density, and $r_p$ is a characteristic distance that regulates the overall amplitude of the effect. The radiation bias factor has been recently estimated in simulations by Arinyo-i-Prats et al. (2015). They report a value of $b_{\Gamma} = 0.093$ for $z = 2.3$.

However, in the presence of hydrogen clouds, this amplitude will rapidly decrease with a characteristic scale length: the mean free path of ionizing photons, $\lambda$. The value of $\lambda$ is somewhat unclear. In the literature there are several measurements using different techniques. Measurements from different techniques are discrepant (see e.g. O'Meara et al. 2013; Rudie et al. 2013). Recently, Prochaska et al. (2014) exoplored this discrepancies and found indications that Rudie et al. (2013) underestimated their value of $\lambda$. Therefore, in this chapter we will consider $\lambda = 244 \pm 42\,h^{-1}$Mpc (from O'Meara et al. 2013, corrected for cosmology).

Taking all this into account, the overall contribution reads

$$\xi_{rad}\,(r, \mu) = b_{\Gamma} \left(\frac{r_p}{r}\right)^2 F_a\,(\mu)\,F_{\nu}\,(\tau)\,e^{-r/\lambda} . \tag{5.9}$$

For an expression for this model in Fourier space, see section 5.5 on the next page.

### 5.4.3 Feasibility of the model to measure the quasar average lifetime

To estimate whether or not the model presented here will be able to estimate the average lifetime given the current data, we pick reasonable values for the model parameters and compare its effect when added to the model for the quasar-Ly$\alpha$ cross-correlation presented in Font-Ribera et al. (2013). In their paper, Font-Ribera et al. (2013), report the values $\beta_\alpha = 1.10 \pm 0.15$, and $b_q = 3.64^{+0.13}_{0.15}$ at $z = 2.38$, while fixing $b_\alpha\,(1 + \beta_\alpha) = -0.336$ at $z = 2.25$ (from Slosar et al. 2011), and $\beta_q = f\,(\Omega)\,b_q$.

The value we choose for $a$ is that of an intermediate case, $a = 0.5$. On the one hand, we know that quasar radiation is not perfectly symmetric, so choosing $a = 0$ would not make much sense. On the other hand, picking the value corresponding to maximum asymmetry does not seem a good choice to give an estimate of the behaviour of this model. We pick $t_q = 29\,h^{-1}\,$Mpc corresponds to a quasar average lifetime of $\sim 5 \times 10^8$ yr. If this value was much longer, it would probably have already been detected. Also, if we pick a value that is too low, then the model will have virtually no effect on the cross-correlation. Finally we take $r_p = 4.5\,h^{-1}\,$Mpc, estimated following the recipe in section 5 of Font-Ribera et al. (2013). In their fit, Font-Ribera et al. (2012) find the cross-correlation to be shifted by a constant amount $\Delta\nu = -115\,\mathrm{km\,s^{-1}}$ (note $\Delta\nu$ is one of the fitted parameters). We apply the same shift to the radiation contribution to the cross-correlation.

Figure 5.9 on the following page shows the difference between the models when the radiation contribution is (right) or is not (left) considered. With the selected parameters the change in the model for quasar-Ly$\alpha$ cross-correlation is visually detectable. What is more, if we visually compare these two models with right panel in figure 5.5 on page 105, we can easily see that the model accommodates better for the data when the radiation model is included. Therefore, this model is promising in the detection of the quasar average lifetime. However, we have to keep in mind that the models shown in figure 5.9 on the following page do not take into account the contamination of metals that is clearly present in the data (see section 5.3 on page 102). Before any fit can be attempted, this radiation model has to be modelled in conjunction with a model for this metal contamination. We leave this issue to future studies.

## 5.5 Radiation model in Fourier space

Theoretical models often compute the power spectrum instead of the cross-correlation. This is mainly because the convolution of functions is much more simpler in Fourier space than it is in real space. Thus, it makes sense to expand our theoretical model to Fourier space. Using the expression of the Fourier transform specialized for cylindrical symmetry, the Fourier-space version for the

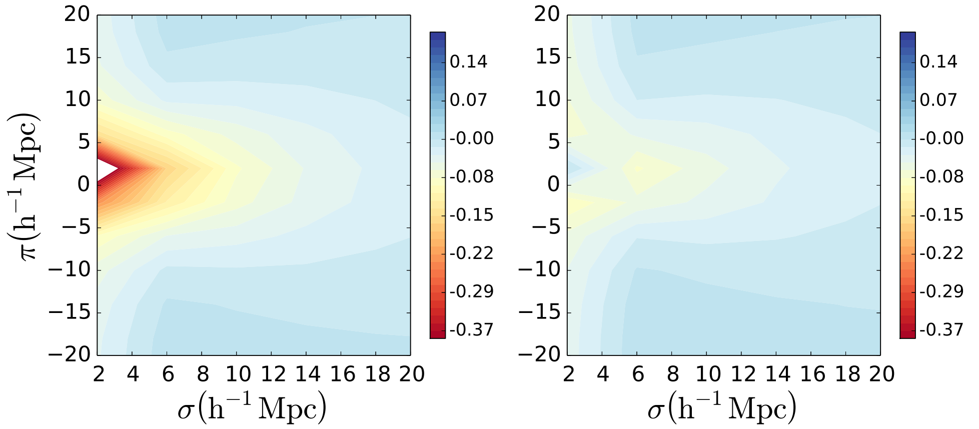


Figure 5.9: Contour plot of the model for the quasar-Lyα cross-correlation with (right) and without (left) the contribution from the quasar radiation. The parameters for the base cosmological model are taken from Font-Ribera et al. (2013). The parameters for the radiation model do not correspond to any fit but are selected to be reasonable.

radiation model presented in section 5.4 on page 106, $P_{rad}(k, \mu_k)$, reads

$$\begin{aligned} P_{rad}(k,\mu_k) = b_\Gamma r_p^2 \int_{-1}^{1} d\mu \left(1 + a\left(1-\mu^2\right)\right) \int_0^\infty dr\, e^{-r/\lambda} e^{-r(1+\mu)/t_q} \times \\ \times e^{-ikr\mu\mu_k} J_0\left(kr\sin\theta\sin k_\theta\right) = b_\Gamma r_p^2 \int_{-1}^{1} d\mu \left(1 + a\left(1-\mu^2\right)\right) \times \\ \times \int_0^\infty dr\, e^{-(1/\lambda+(1+\mu)/t_q+ik\mu\mu_k)r} J_0\left(kr\sin\theta\sin k_\theta\right) \end{aligned} \quad (5.10)$$

Let's identify the integral over $r$, $I_r$, with

$$\int_0^\infty dx\, e^{-\alpha x} J_0(\beta x) = \frac{1}{\sqrt{\alpha^2+\beta^2}} \quad ; \quad \mathrm{Re}(\alpha \pm i\beta) > 0 \quad (5.11)$$

Now if we consider $\sin^2\theta = 1-\mu^2$ and $\sin^2 k_\theta = 1-\mu_k^2$ we can write

$$\begin{aligned} I_r^{-2}(\mu) = \left(\frac{1}{\lambda} + \frac{1+\mu}{t_q} + ik\mu\mu_k\right)^2 + k^2\left(1-\mu^2\right)\left(1-\mu_k^2\right) = \\ = \left(\frac{1}{\lambda} + \frac{1+\mu}{t_q}\right)^2 + k^2\left(1-\mu^2-\mu_k^2\right) + 2ik\mu\mu_k\left(\frac{1}{\lambda}+\frac{1+\mu}{t_q}\right) \end{aligned} \quad (5.12)$$

Thus

$$P_{rad}\left(k, \mu_k\right) = b_\Gamma r_p^2 \int_{-1}^{1} I_r\left(\mu\right)\left(1 + a\left(1 - \mu^2\right)\right) d\mu\,. \tag{5.13}$$

We can separate the integral in two terms, according to the power of $\mu$ in the numerator, as

$$P_{rad}\left(k, \mu_k\right) = b_\Gamma r_p^2 \left(\left(1 + a\right) I_1 - a I_2\right)\ , \tag{5.14}$$

where

$$I_1 = \int_{-1}^{1} I_r\left(\mu\right) d\mu\,, \tag{5.15}$$

and

$$I_2 = \int_{-1}^{1} I_r\left(\mu\right) \mu^2 d\,. \tag{5.16}$$

Let us define

$$\alpha \equiv \left(\frac{1}{\lambda} + \frac{1}{t_q}\right)^2 + k^2\left(1 - \mu_k^2\right)\ , \tag{5.17}$$

$$\beta_1 \equiv \frac{2}{t_q}\left(\frac{1}{\lambda} + \frac{1}{t_q}\right)\ ; \qquad \beta_2 \equiv 2k\mu_k\left(\frac{1}{\lambda} + \frac{1}{t_q}\right)\ , \tag{5.18}$$

$$\gamma_1 \equiv \frac{1}{t_q^2} - k^2\left(1 - \mu_k^2\right)\ ; \qquad \gamma_2 \equiv \frac{2k\mu_k}{t_q}\ , \tag{5.19}$$

Equations 5.15 and 5.16 now become

$$I_1 = \int_{-1}^{1} \frac{d\mu}{\sqrt{\alpha + \left(\beta_1 + i\beta_2\right)\mu + \left(\gamma_1 + i\gamma_2\right)\mu^2}} \tag{5.20}$$

and

$$I_2 = \int_{-1}^{1} \frac{\mu^2 d\mu}{\sqrt{\alpha + \left(\beta_1 + i\beta_2\right)\mu + \left(\gamma_1 + i\gamma_2\right)\mu^2}}\ , \tag{5.21}$$

which can be solved analytically[2]:

$$\begin{aligned} I_1 = &\left[\frac{\ln\left(2\sqrt{\alpha + \left(\beta_1 + i\beta_2\right)\mu + \left(\gamma_1 + i\gamma_2\right)\mu^2} + \frac{\beta_1 + i\beta_2 + 2\left(\gamma_1 + i\gamma_2\right)\mu}{\sqrt{\gamma_1 + i\gamma_2}}\right)}{\sqrt{\gamma_1 + i\gamma_2}}\right]_{-1}^{1} = \\ &= \frac{1}{\sqrt{\gamma_1 + i\gamma_2}} \ln\left(\frac{N_{I1}}{D_{I1}}\right)\ , \end{aligned} \tag{5.22}$$

---

[2] source: Wolfram Alpha: `https://www.wolframalpha.com/`

with

$$N_{I1} = 2\sqrt{\gamma_1 + i\gamma_2}\sqrt{\alpha + \beta_1 + i\beta_2 + \gamma_1 + i\gamma_2} + \beta_1 + i\beta_2 + 2\left(\gamma_1 + i\gamma_2\right) \,, \tag{5.23}$$

and

$$D_{I1} = 2\sqrt{\gamma_1 + i\gamma_2}\sqrt{\alpha - \left(\beta_1 + i\beta_2\right) + \gamma_1 + i\gamma_2} + \beta_1 + i\beta_2 - 2\left(\gamma_1 + i\gamma_2\right) \,, \tag{5.24}$$

and

$$\begin{aligned} I_2 = & \left[\frac{\left(3\left(\beta_1 + i\beta_2\right) - 2\left(\gamma_1 + i\gamma_2\right)\mu\right)\sqrt{\alpha + \left(\beta_1 + i\beta_2\right)\mu + \left(\gamma_1 + i\gamma_2\right)\mu^2}}{4\left(\gamma_1 - i\gamma_2\right)^2}\right]_{-1}^{1} + \\ & + \frac{4\alpha\left(\gamma_1 + i\gamma_2\right) - 3\left(\beta_1 + i\beta_2\right)^2}{8\left(\gamma_1 - i\gamma_2\right)^2} I_1 = \\ = & \frac{\left(3\left(\beta_1 + i\beta_2\right) - 2\left(\gamma_1 + i\gamma_2\right)\right)\sqrt{\alpha + \beta_1 + i\beta_2 + \gamma_1 + i\gamma_2}}{4\left(\gamma_1 - i\gamma_2\right)^2} - \\ & - \frac{\left(3\left(\beta_1 + i\beta_2\right) + 2\left(\gamma_1 + i\gamma_2\right)\right)\sqrt{\alpha - \left(\beta_1 + i\beta_2\right) + \left(\gamma_1 + i\gamma_2\right)}}{4\left(\gamma_1 - i\gamma_2\right)^2} + \\ & + \frac{4\alpha\left(\gamma_1 + i\gamma_2\right) - 3\left(\beta_1 + i\beta_2\right)^2}{8\left(\gamma_1 - i\gamma_2\right)^2} I_1 \,. \end{aligned} \tag{5.25}$$

## 5.6 Summary and conclusions

We have measured the cross-correlation of quasars with the Lyα forest focusing on the small-scales effects. We observe mainly two effects. First, there is contamination from metals, mostly Si II and Si III. The presence of the contamination at small bins in $\sigma$ and the fast recession of the amplitude of this effect suggest that the redhsift space distortion of the metal absorbers is different than that of Lyα forest. Precise measurements of this parameters are required to confirm or reject this hypothesis.

Second, we find that there is a small asymmetry (at the $\sim 1\sigma$ level) in the data at $\sigma \leqslant 20\,h^{-1}$ Mpc and $\pi \lesssim 20\,h^{-1}$ Mpc. The statistical significance of this asymmetry is not explored here. Instead, we present a simple model (both in redhsift space and in Fourier space) that relates the average lifetime of quasars with this asymmetry. Depending on the statistical significance of the asymmetry this may lead to either a measurement of the average lifetime of quasars or else an upper limit on its value. Tests on the feasibility of the model point towards a detection, rather than an upper limit, for the average lifetime of quasars.

We emphasize that the asymmetry detected here is found at roughly the same distance where the contamination from metals is observed. This makes it harder to estimate the statistical significance of the asymmetry detection. Any attempt at modelling this effect should be accompanied by a modelling of the metal contamination as well. Otherwise, the results are likely to be biased. Thus, a more detailed study is needed to better constrain the average lifetime of quasars. Such a study will be performed in the near future.

# Part III

# Summary and conclusions

# 6

# Conclusions and future perspectives

## 6.1 Conclusions

We have measured the cross-correlation of different tracers and have seen that there is still room for improvement in this type of analysis. Apart from using more extensive datasets, we have yet to fully understand the effects of small scales physics have on the cross-correlations.

We have measured the cross-correlation of Mg II absorbers and the CMASS galaxies. We have developed a new method to improve the detection of the Mg II absorption systems. The main improvement of this method is that we add the very weak Mg II absorbers to the analysis, which are commonly systematically not included in the analysis. This happens mainly to absorbers with equivalent widths comparable to the noise in individual spectra or lower. Using this new method we have obtained a bias factor for the Mg II absorbers of $b_{\mathrm{Mg}} = 2.33 \pm 0.19$ which is substantially higher than previous measurements (Gauthier et al. 2009; Lundgren et al. 2009). It is not yet clear whether the discrepancy is real or else a combination of several effects.

We have also measured the cross-correlation of DLAs with the Ly$\alpha$ forest. In this analysis, we reduce the errorbars of the DLA bias factor down to ~ 2.5%. We find $b_d = 1.87 \pm 0.05$, which is a bit lower than the value reported by Font-Ribera et al. (2012). We discuss several reasons that can explain this discrepancy, but further investigation is required to get to the bottom of this. We observe no clear dependence of the DLA bias with neither redshift nor column density, even though we do find hints (at low statistical significance) that suggest the DLA bias may decrease at higher redshift and at higher column densities. More statistics are required to shed light on this.

Finally we show that the cross-correlation analysis has the potential to uncover physics at IGM scales. By analysing the cross-correlation of quasars and the Ly$\alpha$ forest at small scales, we have seen that there are some effects that can give an insight on how astrophysical objects behave. For instance in the asymmetry observed at scales $\lesssim 20\,h^{-1}\,\mathrm{Mpc}$ we see the potential to measure the average lifetime of quasar activity, and we purpose a simple model to constrain it. Also, the contamination of the Ly$\alpha$ forest by metals can potentially lead to the understanding of the behaviour of these metals. By analysing how the contamination of different metals decrease as we go to larger scales, we can uncover differences in the behaviour of the different species.

## 6.2 Future perspectives

As we have been discussing, the study of cross-correlations between different tracers is still far from obsolete. There are still many things to be learned from this type of analysis. For instance, the cross-correlations of metals have not yet being studied. In chapter 5 on page 99 we have seen hints that the behaviour of different metals may not be the same, and with the new methodology presented in chapter 3 on page 29 these hints could be confirmed or denied. Understanding the behaviour of metals on the IGM can provide new insights on the formation and evolution of galaxies.

Also, the new datasets are going to provide much more statistics with which to measure the cross-correlation and the autocorrelations to a much higher precision. The small-scales effects that we are beginning to probe will be fully accessible with these new datasets. This will uncover a new way to probe physics of the IGM that are otherwise difficult to check. An example of this is presented in chapter 5 on page 99 with the model to estimate the average lifetime of quasars.

However, before all this is achievable, we must fully understand the effects the quasar continuum fitting introduce into the cross- and autocorrelations. The new distortion matrix method is promising but not yet complete. Our most recent results show that it may not account for most of the broadband effects, but there are still some alternative explanations that could account for the observed discrepancies. Overall the evolution of this field of research is promising. Many improvements will be made in the near future.

# Bibliography

Abazajian, K. N., Adelman-McCarthy, J. K., Agüeros, M. A., et al. 2009, ApJS, 182, 543

Ahn, C. P., Alexandroff, R., Allende Prieto, C., et al. 2012, ApJS, 203, 21

—. 2014, ApJS, 211, 17

Akerman, C. J., Ellison, S. L., Pettini, M., & Steidel, C. C. 2005, A&A, 440, 499

Alpher, R. A., & Herman, R. 1948, Nature, 162, 774

Antonucci, R. 1993, ARA&A, 31, 473

Arinyo-i-Prats, A., Miralda-Escudé, J., Viel, M., & Cen, R. 2015, JCAP, 12, 017

Aubourg, É., Bailey, S., Bautista, J. E., et al. 2015, Phys. Rev. D, 92, 123516

Bahcall, J. N. 1975, ApJL, 200, L1

Bahcall, J. N., & Salpeter, E. E. 1965, ApJ, 142, 1677

Bahcall, J. N., & Spitzer, Jr., L. 1969, ApJL, 156, L63

Bajtlik, S., Duncan, R. C., & Ostriker, J. P. 1988, ApJ, 327, 570

Barnes, L. A., Garel, T., & Kacprzak, G. G. 2014, PASP, 126, 969

Bergeron, J., & Boissé, P. 1991, A&A, 243, 344

Bergeron, J., & Stasińska, G. 1986, A&A, 169, 1

Bernet, M. L., Miniati, F., & Lilly, S. J. 2010, ApJ, 711, 380

Bird, S., Haehnelt, M., Neeleman, M., et al. 2015, MNRAS, 447, 1834

Blomqvist, M., Kirkby, D., Bautista, J. E., et al. 2015, JCAP, 11, 034

Blumenthal, G. R., Faber, S. M., Primack, J. R., & Rees, M. J. 1984, Nature, 311, 517

Bolton, A. S., Schlegel, D. J., Aubourg, É., et al. 2012, AJ, 144, 144

Bondi, H., & Gold, T. 1948, MNRAS, 108, 252

Bordoloi, R., Lilly, S. J., Kacprzak, G. G., & Churchill, C. W. 2014, ApJ, 784, 108

Bordoloi, R., Lilly, S. J., Knobel, C., et al. 2011, ApJ, 743, 10

Bouché, N., Murphy, M. T., & Péroux, C. 2004, MNRAS, 354, L25

Bouché, N., Murphy, M. T., Péroux, C., Csabai, I., & Wild, V. 2006, MNRAS, 371, 495

Bovy, J., Hennawi, J. F., Hogg, D. W., et al. 2011, ApJ, 729, 141

Brandenberger, R. H. 1985, Reviews of Modern Physics, 57, 1

Bromm, V., & Larson, R. B. 2004, ARA&A, 42, 79

Bromm, V., & Yoshida, N. 2011, ARA&A, 49, 373

Busca, N. G., Delubac, T., Rich, J., et al. 2013, A&A, 552, A96

Byrd, G. G., Chernin, A. D., & J., V. M. 2007, Cosmology: Fundations and Frontiers (URSS)

Calura, F., Matteucci, F., & Vladilo, G. 2003, MNRAS, 340, 59

Carswell, R. F., Whelan, J. A. J., Smith, M. G., Boksenberg, A., & Tytler, D. 1982, MNRAS, 198, 91

Cen, R. 2012, ApJ, 748, 121

Chen, H.-W., Helsby, J. E., Gauthier, J.-R., et al. 2010a, ApJ, 714, 1521

Chen, H.-W., Wild, V., Tinker, J. L., et al. 2010b, ApJL, 724, L176

Churchill, C. W., Mellon, R. R., Charlton, J. C., et al. 2000, ApJ, 543, 577

Churchill, C. W., Rigby, J. R., Charlton, J. C., & Vogt, S. S. 1999, ApJS, 120, 51

Cole, S., & Kaiser, N. 1989, MNRAS, 237, 1127

Conselice, C. J. 2014, ARA&A, 52, 291

Cooke, J., Wolfe, A. M., Gawiser, E., & Prochaska, J. X. 2006, ApJ, 652, 994

Crighton, N. H. M., Murphy, M. T., Prochaska, J. X., et al. 2015, MNRAS, 452, 217

Croft, R. A. C. 2004, ApJ, 610, 642

Croom, S. M., Boyle, B. J., Shanks, T., et al. 2005, MNRAS, 356, 415

Crotts, A. P. S. 1989, ApJ, 336, 550

Cuesta, A. J., Vargas-Magaña, M., Beutler, F., et al. 2016, MNRAS, 457, 1770

da Ângela, J., Outram, P. J., Shanks, T., et al. 2005, MNRAS, 360, 1040

Dawson, K. S., Schlegel, D. J., Ahn, C. P., et al. 2013, AJ, 145, 10

Delubac, T., Bautista, J. E., Busca, N. G., et al. 2015, A&A, 574, A59

Dicke, R. H., Peebles, P. J. E., Roll, P. G., & Wilkinson, D. T. 1965, ApJ, 142, 414

Einasto, J., Kaasik, A., & Saar, E. 1974, Nature, 250, 309

Eisenstein, D. J. 2005, New Astronomy Reviews, 49, 360

Eisenstein, D. J., Weinberg, D. H., Agol, E., et al. 2011, AJ, 142, 72

Faber, S. M., & Gallagher, J. S. 1979, ARA&A, 17, 135

Fixsen, D. J., Cheng, E. S., Cottingham, D. A., et al. 1996, ApJ, 470, 63

Font-Ribera, A., Miralda-Escudé, J., Arnau, E., et al. 2012, JCAP, 11, 59

Font-Ribera, A., Arnau, E., Miralda-Escudé, J., et al. 2013, JCAP, 5, 18

Font-Ribera, A., Kirkby, D., Busca, N., et al. 2014, JCAP, 5, 27

Fox, A. J., Ledoux, C., Petitjean, P., & Srianand, R. 2007a, A&A, 473, 791

Fox, A. J., Petitjean, P., Ledoux, C., & Srianand, R. 2007b, A&A, 465, 171

Friedmann, A. 1922, Zeitschrift fur Physik, 10, 377

—. 1924, Zeitschrift fur Physik, 21, 326

Frieman, J. A., Turner, M. S., & Huterer, D. 2008, ARA&A, 46, 385

Fumagalli, M., Prochaska, J. X., Kasen, D., et al. 2011, MNRAS, 418, 1796

Gamow, G. 1948a, Nature, 162, 680

—. 1948b, Physical Review, 74, 505

Gauthier, J.-R., Chen, H.-W., & Tinker, J. L. 2009, ApJ, 702, 50

Gauthier, J.-R., Tinker, J., & Chen, H.-W. 2008, in COSPAR Meeting, Vol. 37, 37th COSPAR Scientific Assembly, 980

Gil-Marín, H., Percival, W. J., Cuesta, A. J., et al. 2015, ArXiv e-prints, arXiv:1509.06373

Goobar, A., & Leibundgut, B. 2011, Annual Review of Nuclear and Particle Science, 61, 251

Guimarães, R., Petitjean, P., Rollinde, E., et al. 2007, MNRAS, 377, 657

Gunn, J. E., & Peterson, B. A. 1965, ApJ, 142, 1633

Gunn, J. E., Carr, M., Rockosi, C., et al. 1998, AJ, 116, 3040

Gunn, J. E., Siegmund, W. A., Mannery, E. J., et al. 2006, AJ, 131, 2332

Guo, H., Zehavi, I., Zheng, Z., et al. 2013, ApJ, 767, 122

Guth, A. H. 1981, Phys. Rev. D, 23, 347

Haehnelt, M. G., Steinmetz, M., & Rauch, M. 1998, ApJ, 495, 647

Hazard, C., Mackey, M. B., & Shimmins, A. J. 1963, Nature, 197, 1037

Hinshaw, G., Spergel, D. N., Verde, L., et al. 2003, ApJS, 148, 135

Hoyle, F. 1948, MNRAS, 108, 372

Hu, W., & Dodelson, S. 2002, ARA&A, 40, 171

Hubble, E. 1929, Proceedings of the National Academy of Science, 15, 168

Jakobsen, P., Perryman, M. A. C., di Serego Alighieri, S., Ulrich, M. H., & Macchetto, F. 1986, ApJL, 303, L27

Johnson, S. D., Chen, H.-W., Mulchaey, J. S., et al. 2014, MNRAS, 438, 3039

Jorgenson, R. A., Murphy, M. T., & Thompson, R. 2013, MNRAS, 435, 482

Kacprzak, G. G., Churchill, C. W., Evans, J. L., Murphy, M. T., & Steidel, C. C. 2011, MNRAS, 416, 3118

Kacprzak, G. G., Churchill, C. W., & Nielsen, N. M. 2012, ApJL, 760, L7

Kaiser, N. 1987, MNRAS, 227, 1

Khare, P., Kulkarni, V. P., Lauroesch, J. T., et al. 2004, ApJ, 616, 86

Kirkby, D., Margala, D., Slosar, A., et al. 2013, JCAP, 3, 24

Kirkpatrick, J. A., Schlegel, D. J., Ross, N. P., et al. 2011, ApJ, 743, 125

Kulkarni, V. P., & Fall, S. M. 2002, ApJ, 580, 732

Kulkarni, V. P., Fall, S. M., Lauroesch, J. T., et al. 2005, ApJ, 618, 68

Lanzetta, K. M., & Bowen, D. 1990, ApJ, 357, 321

Ledoux, C., Petitjean, P., Fynbo, J. P. U., Møller, P., & Srianand, R. 2006, A&A, 457, 71

Lee, K.-G., Bailey, S., Bartsch, L. E., et al. 2013, AJ, 145, 69

Lemaître, G. 1927, Annales de la Société Scientifique de Bruxelles, 47, 49

Lemaitre, G. 1931, Publications du Laboratoire d'Astronomie et de Geodesie de l'Universite de

Louvain, vol. 8, pp.101-120, 8, 101

Liske, J., & Williger, G. M. 2001, MNRAS, 328, 653

Lundgren, B. F., Wake, D. A., Padmanabhan, N., Coil, A., & York, D. G. 2011, MNRAS, 417, 304

Lundgren, B. F., Brunner, R. J., York, D. G., et al. 2009, ApJ, 698, 819

Lundgren, B. F., Brammer, G., van Dokkum, P., et al. 2012, ApJ, 760, 49

Lynds, R. 1971, ApJL, 164, L73

Maller, A. H., & Bullock, J. S. 2004, MNRAS, 355, 694

Mas-Ribas, L., Miralda-Escudé, J., Arinyo-i-Prats, A., et al. 2016, in Prep

McDonald, P., & Miralda-Escudé, J. 1999, ApJ, 519, 486

McDonald, P., Seljak, U., Burles, S., et al. 2006, ApJS, 163, 80

Miyatake, H., More, S., Mandelbaum, R., et al. 2015, ApJ, 806, 1

Mo, H. J., & Miralda-Escudé, J. 1996, ApJ, 469, 589

Mo, H. J., & White, S. D. M. 1996, MNRAS, 282, 347

Møller, P., Fynbo, J. P. U., Ledoux, C., & Nilsson, K. K. 2013, MNRAS, 430, 2680

Møller, P., & Kjaergaard, P. 1992, A&A, 258, 234

Morton, D. C. 2003, ApJS, 149, 205

Murdoch, H. S., Hunstead, R. W., Pettini, M., & Blades, J. C. 1986, ApJ, 309, 19

Myers, A. D., Brunner, R. J., Richards, G. T., et al. 2006, ApJ, 638, 622

Narayanan, A., Misawa, T., Charlton, J. C., & Kim, T.-S. 2007, ApJ, 660, 1093

Narlikar, J. V., & Padmanabhan, T. 1991, ARA&A, 29, 325

Neeleman, M., Prochaska, J. X., & Wolfe, A. M. 2015, ApJ, 800, 7

Neeleman, M., Wolfe, A. M., Prochaska, J. X., & Rafelski, M. 2013, ApJ, 769, 54

Nestor, D. B., Turnshek, D. A., & Rao, S. M. 2005, ApJ, 628, 637

—. 2006, ApJ, 643, 75

Noterdaeme, P., Petitjean, P., Ledoux, C., & Srianand, R. 2009, A&A, 505, 1087

Noterdaeme, P., Petitjean, P., Carithers, W. C., et al. 2012, A&A, 547, L1

Nuza, S. E., Sánchez, A. G., Prada, F., et al. 2013, MNRAS, 432, 743

O'Meara, J. M., Prochaska, J. X., Worseck, G., Chen, H.-W., & Madau, P. 2013, ApJ, 765, 137

Ostriker, J. P., Peebles, P. J. E., & Yahil, A. 1974, ApJL, 193, L1

Padmanabhan, H., Choudhury, T. R., & Refregier, A. 2016, MNRAS, 458, 781

Padmanabhan, T. 1993, Structure Formation in the Universe (Cambridge University Press), 499

Palanque-Delabrouille, N., Yèche, C., Borde, A., et al. 2013, A&A, 559, A85

Pâris, I., & et al. . 2016, submitted

Penzias, A. A., & Wilson, R. W. 1965, ApJ, 142, 419

Pérez-Ràfols, I., Miralda-Escudé, J., Lundgren, B., et al. 2015, MNRAS, 447, 2784

Perlmutter, S., Aldering, G., Goldhaber, G., et al. 1999, ApJ, 517, 565

Péroux, C., McMahon, R. G., Storrie-Lombardi, L. J., & Irwin, M. J. 2003, MNRAS, 346, 1103

Peter, P., & Uzan, J. 2009, Primordial cosmology, Oxford graduate texts (Oxford University Press)

Phillips, M. M. 1993, ApJL, 413, L105

Pieri, M. M. 2014, MNRAS, 445, L104

Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2014, A&A, 571, A16

—. 2015, ArXiv e-prints, arXiv:1502.01589

Porciani, C., Magliocchetti, M., & Norberg, P. 2004, MNRAS, 355, 1010

Prochaska, J. X., Gawiser, E., Wolfe, A. M., Castro, S., & Djorgovski, S. G. 2003a, ApJL, 595, L9

Prochaska, J. X., Gawiser, E., Wolfe, A. M., Cooke, J., & Gelino, D. 2003b, ApJS, 147, 227

Prochaska, J. X., Herbert-Fort, S., & Wolfe, A. M. 2005, ApJ, 635, 123

Prochaska, J. X., Madau, P., O'Meara, J. M., & Fumagalli, M. 2014, MNRAS, 438, 476

Prochaska, J. X., & Wolfe, A. M. 1997, ApJ, 487, 73

—. 1998, ApJ, 507, 113

—. 2002, ApJ, 566, 68

—. 2009, ApJ, 696, 1543

Rafelski, M., Neeleman, M., Fumagalli, M., Wolfe, A. M., & Prochaska, J. X. 2014, ApJL, 782, L29

Rafelski, M., Wolfe, A. M., Prochaska, J. X., Neeleman, M., & Mendez, A. J. 2012, ApJ, 755, 89

Rahmati, A., & Schaye, J. 2014, MNRAS, 438, 529

Rao, S. M., Turnshek, D. A., & Nestor, D. B. 2006, ApJ, 636, 610

Rauch, M. 1998, ARA&A, 36, 267

Riess, A. G., Filippenko, A. V., Challis, P., et al. 1998, AJ, 116, 1009

Robertson, H. P. 1935, ApJ, 82, 284

—. 1936a, ApJ, 83, 187

—. 1936b, ApJ, 83, 257

Rollinde, E., Srianand, R., Theuns, T., Petitjean, P., & Chand, H. 2005, MNRAS, 361, 1015

Ross, N. P., Shen, Y., Strauss, M. A., et al. 2009, ApJ, 697, 1634

Ross, N. P., Myers, A. D., Sheldon, E. S., et al. 2012, ApJS, 199, 3

Rubin, V. C. 1983, Science, 220, 1339

Rudie, G. C., Steidel, C. C., Shapley, A. E., & Pettini, M. 2013, ApJ, 769, 146

Samtleben, D., Staggs, S., & Winstein, B. 2007, Annual Review of Nuclear and Particle Science, 57, 245

Sánchez-Ramírez, R., Ellison, S. L., Prochaska, J. X., et al. 2016, MNRAS, 456, 4488

Sanders, R. H., & McGaugh, S. S. 2002, ARA&A, 40, 263

Sargent, W. L. W., Young, P. J., Boksenberg, A., Carswell, R. F., & Whelan, J. A. J. 1979, ApJ, 230, 49

Schirber, M., Miralda-Escudé, J., & McDonald, P. 2004, ApJ, 610, 105

Schlegel, D. J., Finkbeiner, D. P., & Davis, M. 1998, ApJ, 500, 525

Schmidt, M. 1963, Nature, 197, 1040

—. 1965, ApJ, 141, 1295

Schneider, D. P., Richards, G. T., Hall, P. B., et al. 2010, AJ, 139, 2360

Scott, J., Bechtold, J., Dobrzycki, A., & Kulkarni, V. P. 2000, ApJS, 130, 67

Shen, Y., Strauss, M. A., Oguri, M., et al. 2007, AJ, 133, 2222

Shen, Y., Strauss, M. A., Ross, N. P., et al. 2009, ApJ, 697, 1656

Sheth, R. K., & Tormen, G. 1999, MNRAS, 308, 119

Silk, J., & Mamon, G. A. 2012, Research in Astronomy and Astrophysics, 12, 917

Slosar, A., Font-Ribera, A., Pieri, M. M., et al. 2011, JCAP, 9, 001

Smee, S. A., Gunn, J. E., Uomoto, A., et al. 2013, AJ, 146, 32

Smoot, G. F., Bennett, C. L., Kogut, A., et al. 1992, ApJL, 396, L1

Spitzer, Jr., L. 1956, ApJ, 124, 20

Steidel, C. C. 1995, in QSO Absorption Lines, ed. G. Meylan, 139

Steidel, C. C., Dickinson, M., & Persson, S. E. 1994, ApJL, 437, L75

Steigman, G. 2007, Annual Review of Nuclear and Particle Science, 57, 463

Tinker, J. L., & Chen, H.-W. 2008, ApJ, 679, 1218

Tinker, J. L., Robertson, B. E., Kravtsov, A. V., et al. 2010, ApJ, 724, 878

Vanden Berk, D. E., Richards, G. T., Bauer, A., et al. 2001, AJ, 122, 549

Vladilo, G. 2002, A&A, 391, 407

Vladilo, G., Centurión, M., Bonifacio, P., & Howk, J. C. 2001, ApJ, 557, 1007

Walker, A. G. 1937, Proceedings of the London Mathematical Society, s2-42, 90

White, M., Myers, A. D., Ross, N. P., et al. 2012, MNRAS, 424, 933

Wolfe, A. M. 1998, in Astronomical Society of the Pacific Conference Series, Vol. 136, Galactic Halos, ed. D. Zaritsky, 159

Wolfe, A. M., Gawiser, E., & Prochaska, J. X. 2005, ARA&A, 43, 861

Wolfe, A. M., & Prochaska, J. X. 2000, ApJ, 545, 591

Wolfe, A. M., Turnshek, D. A., Smith, H. E., & Cohen, R. D. 1986, ApJS, 61, 249

Yèche, C., Petitjean, P., Rich, J., et al. 2010, A&A, 523, A14

York, D. G., Adelman, J., Anderson, Jr., J. E., et al. 2000, AJ, 120, 1579

York, D. G., Vanden Berk, D., Richards, G. T., et al. 2005, in IAU Colloq. 199: Probing Galaxies through Quasar Absorption Lines, ed. P. Williams, C.-G. Shu, & B. Menard, 58–64

Zafar, T., Péroux, C., Popping, A., et al. 2013, A&A, 556, A141

Zehavi, I., Zheng, Z., Weinberg, D. H., et al. 2011, ApJ, 736, 59

Zheng, Z., Berlind, A. A., Weinberg, D. H., et al. 2005, ApJ, 633, 791

Zhu, G., & Ménard, B. 2013a, ApJ, 773, 16

—. 2013b, ApJ, 770, 130

Zhu, G., Ménard, B., Bizyaev, D., et al. 2014, MNRAS, 439, 3139

Zwicky, F. 1933, Helvetica Physica Acta, 6, 110

# List of Figures

3.1 Redshift (left panel) and luminosity (right panel) distributions for the selected CMASS galaxies (see text for details) . . . . . . . . . . . . . . . . . . . . . . . . 34

3.2 Number of quasar spectra contributing to each 1Å bin of the mean spectrum, as a function of rest-frame wavelength. The flat feature corresponds to the spectral range used to compute the normalizing factor $n_j$. Outside this range the number of quasars contributing to the mean quasar spectrum decreases because some of the quasar spectra do not extend to that wavelength. . . . . . . . . . . . . . . . . 37

3.3 Mean spectrum of the weighted-average obtained from the 70,650 DR7 quasars, normalized in the rest-frame wavelength interval from 2000 to 2600Å. This mean spectrum is used as a continuum model. . . . . . . . . . . . . . . . . . . . . . . . 38

3.4 Examples illustrating the procedure explained in section 3.3.1 on page 38. The left panels shows a case with a detected individual Mg II absorption system, and the right panels a case with no detectable associated Mg II absorption system, but with an unrelated metal absorption line. The top panels show the normalized flux of the spectral region, $f_{ik}^{(r)}/n_{j(k)}$ (solid blue line), and the normalized mean spectrum $\bar{f}_i$ (dashed red line). The bottom panels show the transmission $F_{ik}^{(0)}$ (thin, solid red line), the computed linear fit $L_{ik}$ (thin, dashed red line), and the corrected transmission $F_{ik}$ (thick, solid blue line). In the bottom right panel, the points indicate pixels that are excluded from the linear fit. A horizontal thick dashed green line at a transmission of 1 is included for visual aid. . . . . . . . . . . . . . 40

3.5 From top to bottom, composite spectra for increasing impact parameter intervals (in proper kpc). The effective optical depth is shown against velocity for the mean subtraction method ( thick, solid blue line) and the variable smoothing method (thin, solid red line). The RMS value of the bootstrap error in individual pixels is shown by the errorbars on the left. The thick, dashed blue line and the thin, dashed red line are the best fit model (equation 3.16 on page 43) for the mean subtraction and variable smoothing methods respectively. A single set of parameters are fitted to all the 17 regions. Figures 3.6 and 3.7 on page 46 and on page 47 show the spectra for the remaining impact parameter intervals. The stacks show a mean absorption profile for the presence of the Mg II doublet line at the expected position. For visual guidance, vertical, dashed green lines mark the predicted position of the Mg II doublet. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 45

3.6 Continuation of figure 3.5 on page 45. . . . . . . . . . . . . . . . . . . . . . . . . . 46

3.7 Continuation of figure 3.5 on page 45. . . . . . . . . . . . . . . . . . . . . . . . . . 47

3.8 Measured rest-frame mean equivalent width $W_e$ of the Mg II doublet, versus proper impact parameter. Blue triangles are obtained from the mean subtraction continuum method, and red squares use the variable smoothing method. Error bars have been obtained by the bootstrap method. Lines are the best-fit power-law model to the data for both methods. Green circles are the results of Zhu et al. (2014). Note that the results for the mean subtraction method are systematically higher than the rest. This is explored in further detail in appendix 3.A on page 60. 51

3.9 Rest-frame equivalent width distribution of Mg II absorption systems. Data points are from Bernet et al. (2010) (black squares), Churchill et al. (1999) (green triangles), Narayanan et al. (2007) (blue crosses) and Nestor et al. (2005, 2006) (red circles). The overplotted solid line is the double exponential fit of Nestor et al. (2005) (see text). Top right panel is a zoomed view of the weakest absorption systems. . . . . 53

3.10 Projected correlation functions multiplied by the comoving impact parameter as a function of the comoving impact parameter $r_p(1+z)$. Thick black triangles with errorbars show the autocorrelation of CMASS galaxies from Nuza et al. (2013). Blue triangles, red squares and green circles are the mean equivalent width $W_e(r_p)$ times the factor $(1+z)/H(z)/\tau_{eo}$ (equal to the cross-correlation of Mg II systems and CMASS galaxies), times $r_p(1+z)$ for the mean subtraction method, the variable smoothing method and measurements from Zhu et al. (2014) respectively. The thick solid black line is the MultiDark model prediction described in Nuza et al. (2013). The solid blue, the red dashed and the green dashed-dotted lines are the fit to $W_e$ for the mean subtraction method, the variable smoothing method and measurements from Zhu et al. (2014) respectively. The ratio of the each of these lines with the thick solid black line is the ratio of bias factors, $b_{Mg}/b_g$. . . . 55

3.11 Ratio of the mean equivalent width in stacked spectra to the best-fit power-law model prediction, for the real data sample (big red circles), and for the average of 6 mock samples (blue stars). Errorbars are computed with the bootstrap method. This ratio should be consistent with zero for the mock sample in the absence of systematic errors, and with unity for the real data if the power-law model provides a good fit to the data. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 62

3.12 Comparison of the equivalent width estimates from different methods (see section 3.3 on page 34). From left to right, equivalent widths measured using the mean subtraction method, $W_m$, versus the value of the catalogue from Zhu & Ménard (2013b), $W_{ZM}$; variable smoothing method value, $W_s$, versus $W_{ZM}$; and $W_s$ versus $W_m$. Black points are the average values in bins of $\Delta W = 0.5$Å in the horizontal axis, with the dispersion in each bin indicated by the errorbars. One-to-one correspondence is marked by the black dashed line for visual guidance. 63

3.13 Change in the measured equivalent width, $\Delta W$, caused by the insertion of an absorbing system with equivalent width $W_i$, for the mean subtraction method (solid blue lines) and the variable smoothing method (dashed red lines). Results are shown as a function of the velocity dispersion σ, with values indicated to the right of the lines in Å. The blue lines all nearly coincide at $\Delta W = W_i$. Points show the values of $W_i$ for which $\Delta W$ has been computed; the sudden changes in the red lines indicate discontinuities in the variable smoothing method as the inserted line becomes detected or covers different pixels, which causes a change in the continuum estimate. . . . . . . . . . . . . . . . . . . . . . . . . . . . 64

4.1 Distribution of DLAs in dataset A according to redshift (top panel) and column density (bottom panel). The solid red lines show the applied cuts to construct the different sub-samples (see table 4.1 on page 72) . . . . . . . . . . . . . . . . . . 73

4.2 Measurement of the cross-correlation for different bins in σ: from left to right $0\,h^{-1}\text{Mpc} < \sigma < 4\,h^{-1}\text{Mpc}$, $4\,h^{-1}\text{Mpc} \leqslant \sigma / 8\,h^{-1}\text{Mpc}$, $8\,h^{-1}\text{Mpc} \leqslant \sigma < 12\,h^{-1}\text{Mpc}$, $12\,h^{-1}\text{Mpc} \leqslant \sigma < 16\,h^{-1}\text{Mpc}$, $16\,h^{-1}\text{Mpc} \leqslant \sigma < 20\,h^{-1}\text{Mpc}$, and $20\,h^{-1}\text{Mpc} \leqslant \sigma < 28\,h^{-1}\text{Mpc}$. The values of σ are expressed in $h^{-1}$Mpc. Figure 4.3 on page 84 contain similar plots for the remaining bins in σ. In all panels black circles are for sample A. Black solid lines correspond to the best-fit model considering $r_{min} = 5\,h^{-1}$Mpc. . . . . . . . . . . . . . . . . . . . . . . 83

4.3 From left to right and from top to bottom, same as figure 4.2 on page 83 but for bins with $28\,h^{-1}\text{Mpc} \leqslant \sigma < 32\,h^{-1}\text{Mpc}$, $32\,h^{-1}\text{Mpc} \leqslant \sigma < 40\,h^{-1}\text{Mpc}$, $40\,h^{-1}\text{Mpc} \leqslant \sigma < 48\,h^{-1}\text{Mpc}$, $48\,h^{-1}\text{Mpc} \leqslant \sigma < 64\,h^{-1}\text{Mpc}$, and $64\,h^{-1}\text{Mpc} \leqslant \sigma < 80\,h^{-1}\text{Mpc}$. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 84

4.4 Contour plot of the measured DLA-Lyαcross-correlation (left) and our best-fit theoretical model considering bins with $5\,h^{-1}\text{Mpc} < r = \left(\pi^2 + \sigma^2\right)^{1/2} < 90\,h^{-1}\text{Mpc}$ (right). . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 85

4.5 Similar to figure 4.2 on page 83 but for the sub-samples in redshift Z1, Z2, and Z3. In this figure from left to right $0\,h^{-1}\mathrm{Mpc} < \sigma < 4\,h^{-1}\mathrm{Mpc}$, $4\,h^{-1}\mathrm{Mpc} \leqslant \sigma/8\,h^{-1}\mathrm{Mpc}$, $8\,h^{-1}\mathrm{Mpc} \leqslant \sigma < 12\,h^{-1}\mathrm{Mpc}$, $12\,h^{-1}\mathrm{Mpc} \leqslant \sigma < 16\,h^{-1}\mathrm{Mpc}$, $16\,h^{-1}\mathrm{Mpc} \leqslant \sigma < 20\,h^{-1}\mathrm{Mpc}$, and $20\,h^{-1}\mathrm{Mpc} \leqslant \sigma < 28\,h^{-1}\mathrm{Mpc}$. Figure 4.6 on page 88 contain similar plots for the remaining bins in $\sigma$. In all panels green squares, red triangles, and blue crosses stand for samples Z1, Z2, and Z3 respectively. Green solid lines, red dashed-dotted lines, and blue dotted lines correspond to the best-fit model considering $r_{\mathrm{min}} = 5\,h^{-1}\mathrm{Mpc}$ fitted to samples Z1, Z2, and Z3 respectively. Errorbars at the side of the plot show the typical error of the points in the corresponding $\sigma$ bin. . . . . . . . . . . . . . . . . . . . . . . . . . . 87

4.6 From left to right and from top to bottom, same as figure 4.5 on page 87 but for bins with $28\,h^{-1}\mathrm{Mpc} \leqslant \sigma < 32\,h^{-1}\mathrm{Mpc}$, $32\,h^{-1}\mathrm{Mpc} \leqslant \sigma < 40\,h^{-1}\mathrm{Mpc}$, $40\,h^{-1}\mathrm{Mpc} \leqslant \sigma < 48\,h^{-1}\mathrm{Mpc}$, $48\,h^{-1}\mathrm{Mpc} \leqslant \sigma < 64\,h^{-1}\mathrm{Mpc}$, and $64\,h^{-1}\mathrm{Mpc} \leqslant \sigma < 80\,h^{-1}\mathrm{Mpc}$. . . . . . . . . . . . . . . . . . . . . . . . . . . . . 88

4.7 Bias of the DLA against $z$ (left) and $\log\left(N_{\mathrm{HI}}\right)$ (right) obtained by fitting samples Z1, Z2, and Z3, and N1, N2, and N3 respectively (see table 4.1 on page 72). Dashed lines indicate the value obtained by fitting sample A. Dotted lines indicate the $1\sigma$ region. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 89

4.8 Similar to figure 4.2 on page 83 but for the sub-samples in column density N1, N2, and N3. In this figure from left to right $0\,h^{-1}\mathrm{Mpc} < \sigma < 4\,h^{-1}\mathrm{Mpc}$, $4\,h^{-1}\mathrm{Mpc} \leqslant \sigma/8\,h^{-1}\mathrm{Mpc}$, $8\,h^{-1}\mathrm{Mpc} \leqslant \sigma < 12\,h^{-1}\mathrm{Mpc}$, $12\,h^{-1}\mathrm{Mpc} \leqslant \sigma < 16\,h^{-1}\mathrm{Mpc}$, $16\,h^{-1}\mathrm{Mpc} \leqslant \sigma < 20\,h^{-1}\mathrm{Mpc}$, and $20\,h^{-1}\mathrm{Mpc} \leqslant \sigma < 28\,h^{-1}\mathrm{Mpc}$. Figure 4.9 on page 91 contain similar plots for the remaining bins in $\sigma$. In all panels green squares, red triangles, and blue crosses stand for samples N1, N2, and N3 respectively. Green solid lines, red dashed-dotted lines, and blue dotted lines correspond to the best-fit model considering $r_{\mathrm{min}} = 5\,h^{-1}\mathrm{Mpc}$ fitted to samples N1, N2, and N3 respectively. Errorbars at the side of the plot show the typical error of the points in the corresponding $\sigma$ bin. . . . . . . . . . . . . . . 90

4.9 From left to right and from top to bottom, same as figure 4.8 on page 90 but for bins with $28\,h^{-1}\mathrm{Mpc} \leqslant \sigma < 32\,h^{-1}\mathrm{Mpc}$, $32\,h^{-1}\mathrm{Mpc} \leqslant \sigma < 40\,h^{-1}\mathrm{Mpc}$, $40\,h^{-1}\mathrm{Mpc} \leqslant \sigma < 48\,h^{-1}\mathrm{Mpc}$, $48\,h^{-1}\mathrm{Mpc} \leqslant \sigma < 64\,h^{-1}\mathrm{Mpc}$, $64\,h^{-1}\mathrm{Mpc} \leqslant \sigma < 80\,h^{-1}\mathrm{Mpc}$. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 91

4.10 *Left*: Bias of the DLA, $b_d$, against $r$ obtained by fitting sample A but considering only certain intervals in $r$, shown by the horizontal errorbars. Dashed lines indicate the value obtained by fitting the entire sample A. Dotted lines indicate the $1\sigma$ interval for this measurement. *Right*: Best fit value of $b_d$ for different assumed values of $\beta_\alpha$. The solid and dashed lines show the fitting formula described in equation 4.34 on page 92. . . . . . . . . . . . . . . . . . . . . . . . . . . 92

4.11 Measurement of the cross-correlation for different bins in $\sigma$: from left to right $0\,h^{-1}\text{Mpc} < \sigma < 4\,h^{-1}\text{Mpc}$, $4\,h^{-1}\text{Mpc} \leqslant \sigma < 8\,h^{-1}\text{Mpc}$, $8\,h^{-1}\text{Mpc} \leqslant \sigma < 12\,h^{-1}\text{Mpc}$, $12\,h^{-1}\text{Mpc} \leqslant \sigma < 16\,h^{-1}\text{Mpc}$, $16\,h^{-1}\text{Mpc} \leqslant \sigma < 20\,h^{-1}\text{Mpc}$, $20\,h^{-1}\text{Mpc} \leqslant \sigma < 28\,h^{-1}\text{Mpc}$. The values of $\sigma$ are expressed in $h^{-1}\text{Mpc}$. In all panels blue circles are for sample A. Blue solid line, green dashed-dotted line, black dotted line, and red dashed line correspond to the best-fit model considering $r_{\text{min}} = 4\,h^{-1}\text{Mpc}$, $r_{\text{min}} = 6\,h^{-1}\text{Mpc}$, $r_{\text{min}} = 8\,h^{-1}\text{Mpc}$, and $r_{\text{min}} = 10\,h^{-1}\text{Mpc}$ respectively. Bottom panels show the $r^2$-weighted differences between the different models and the best-fit model considering $r_{\text{min}} = 10\,h^{-1}\text{Mpc}$. 96

5.1 Redshift distribution of the 189,715 quasars used as density tracers. . . . . . . . . 102

5.2 Measurement of the cross-correlation for different bins in $\sigma$: from left to right and top to bottom $0\,h^{-1}\text{Mpc} < \sigma < 4\,h^{-1}\text{Mpc}$, $4\,h^{-1}\text{Mpc} \leqslant \sigma/8\,h^{-1}\text{Mpc}$, $8\,h^{-1}\text{Mpc} \leqslant \sigma < 12\,h^{-1}\text{Mpc}$, and $12\,h^{-1}\text{Mpc} \leqslant \sigma < 16\,h^{-1}\text{Mpc}$. The values of $\sigma$ are expressed in $h^{-1}\text{Mpc}$. Figures 5.3 and 5.4 on page 104 and on page 105 contain similar plots for the remaining bins in $\sigma$. . . . . . . . . . . . . . . . . . . . 103

5.3 From left to right and from top to bottom, same as figure 5.2 on page 103 but for bins with $16\,h^{-1}\text{Mpc} \leqslant \sigma < 20\,h^{-1}\text{Mpc}$, $20\,h^{-1}\text{Mpc} \leqslant \sigma < 28\,h^{-1}\text{Mpc}$, $28\,h^{-1}\text{Mpc} \leqslant \sigma < 32\,h^{-1}\text{Mpc}$, $32\,h^{-1}\text{Mpc} \leqslant \sigma < 40\,h^{-1}\text{Mpc}$, $40\,h^{-1}\text{Mpc} \leqslant \sigma < 48\,h^{-1}\text{Mpc}$, and $48\,h^{-1}\text{Mpc} \leqslant \sigma < 64\,h^{-1}\text{Mpc}$. . . . . . . . . . . . . . . 104

5.4 From left to right, same as figure 5.2 on page 103 but for bins with $64\,h^{-1}\text{Mpc} \leqslant \sigma < 80\,h^{-1}\text{Mpc}$ and $80\,h^{-1}\text{Mpc} \leqslant \sigma < 100\,h^{-1}\text{Mpc}$. . . . . . . . . . . . . . . 105

5.5 Contour plot of the measured quasar-Ly$\alpha$ cross-correlation. The right panel shows a zoom in the central region ($\pi \leqslant 20\,h^{-1}\text{Mpc}$, $\sigma \leqslant 20\,h^{-1}\text{Mpc}$). . . . . . . . . 105

5.6 $\xi(\pi, \sigma) - \xi(-\pi, \sigma)$ for different bins in $\sigma$: from left to right and top to bottom $0\,h^{-1}\text{Mpc} < \sigma < 4\,h^{-1}\text{Mpc}$, $4\,h^{-1}\text{Mpc} \leqslant \sigma/8\,h^{-1}\text{Mpc}$, $8\,h^{-1}\text{Mpc} \leqslant \sigma < 12\,h^{-1}\text{Mpc}$, and $12\,h^{-1}\text{Mpc} \leqslant \sigma < 16\,h^{-1}\text{Mpc}$. The values of $\pi$ are the averages of the absolute value of $\pi$ in both points. The errors are computed as the quadratic sum of the errors in both points. The values of $\sigma$ are expressed in $h^{-1}\text{Mpc}$. Figures 5.7 and 5.8 on page 108 and on page 109 contain similar plots for the remaining bins in $\sigma$. In all panels black circles are for sample A. Black solid lines correspond to the best-fit model considering $r_{\text{min}} = 5\,h^{-1}\text{Mpc}$. . . . . . 107

5.7 From left to right and from top to bottom, same as figure 5.6 on page 107 but for bins with $16\,h^{-1}\text{Mpc} \leqslant \sigma < 20\,h^{-1}\text{Mpc}$, $20\,h^{-1}\text{Mpc} \leqslant \sigma < 28\,h^{-1}\text{Mpc}$, $28\,h^{-1}\text{Mpc} \leqslant \sigma < 32\,h^{-1}\text{Mpc}$, $32\,h^{-1}\text{Mpc} \leqslant \sigma < 40\,h^{-1}\text{Mpc}$, $40\,h^{-1}\text{Mpc} \leqslant \sigma < 48\,h^{-1}\text{Mpc}$, and $48\,h^{-1}\text{Mpc} \leqslant \sigma < 64\,h^{-1}\text{Mpc}$. . . . . . . . . . . . . . . 108

5.8 From left to right, same as figure 5.6 on page 107 but for bins with $64\,h^{-1}\text{Mpc} \leqslant \sigma < 80\,h^{-1}\text{Mpc}$ and $80\,h^{-1}\text{Mpc} \leqslant \sigma < 100\,h^{-1}\text{Mpc}$. . . . . . . . . . . . . . . 109

5.9 Contour plot of the model for the quasar-Ly$\alpha$ cross-correlation with (right) and without (left) the contribution from the quasar radiation. The parameters for the base cosmological model are taken from Font-Ribera et al. (2013). The parameters for the radiation model do not correspond to any fit but are selected to be reasonable. 112

# List of Tables

3.1 Best-fit values for the fitted parameters of the model described in section 3.4 on page 42 and shown as the solid lines in figure 3.8 on page 51. *First column:* Results using the mean subtraction method (see section 3.3.1 on page 35). *Second column:* Results using the variable smoothing method (see section 3.3.2 on page 41). Four independent parameters are fitted, and $W_{e0}$ is related to the other parameters according to equation 3.19 on page 44. As explained in section 3.4 on page 42, $\sigma_0$ is fixed at 250 km $s^{-1}$. Errors are computed from repeating the fits on bootstrap realizations of the stacked profiles. . . . . . . . . . . . . . . . . . . . . . . . . 48

3.2 Results on the mean equivalent width and errors shown in figure 3.8 on page 51, presented here as a table. From left to right, impact parameter interval, mean rest-frame Mg II equivalent width and its bootstrap error for the mean subtraction and variable smoothing methods, and number of QSO-galaxy pairs used in the interval. The mean rest-frame equivalent widths are the sum of both lines in the doublet. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 50

4.1 Summary of the different DLA datasets. . . . . . . . . . . . . . . . . . . . . . . . 72

4.2 Summary of the different DLA datasets. The values of the bias are given at the reference redshift $z_{\text{ref}} = 2.3$ and for $\beta_\alpha = 1.39$. See table 4.1 on page 72 for a summary on the datasets' properties. . . . . . . . . . . . . . . . . . . . . . . . 81

4.3 Bias of the DLA against $r$ obtained by fitting sample A but considering only bins with $r \in [r_{\text{min}}, r_{\text{max}})$. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 89

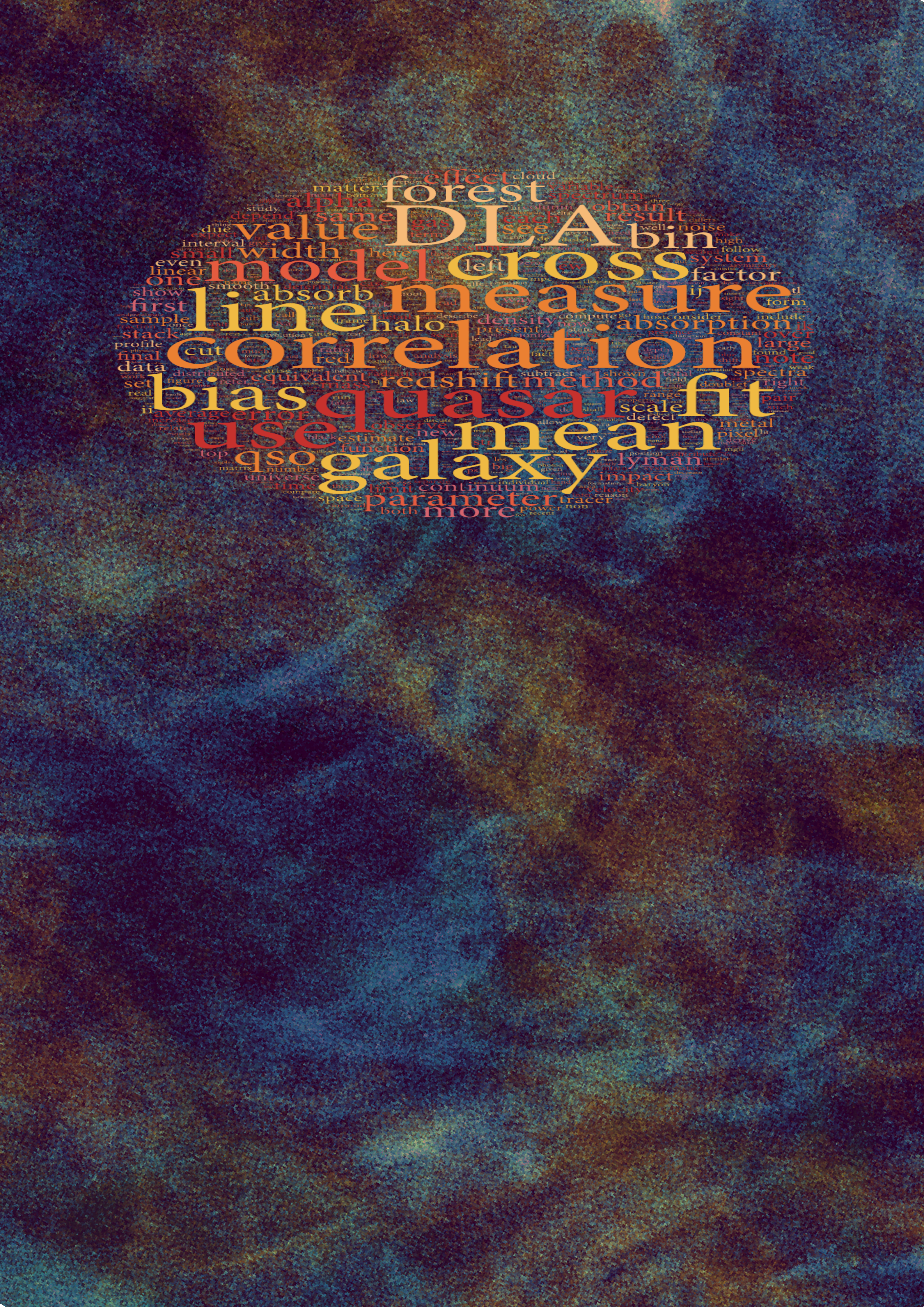
4.4 Best fit value of $b_d$ for different assumed values of $\beta_\alpha$. . . . . . . . . . . . . . . . 93

4.5 Measured bias of the DLA when fitting the same model but cutting at different $r_{min}$. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 97

# List of Acronyms

| | |
|---|---|
| $\Lambda$CDM | $\Lambda$ Cold Dark Matter. 8, 19, 32, 70, 86, 101 |
| 2dF Survey | Two-degree-Field Galaxy Redshift survey. 100 |
| AGN | Active Galactic Nuclei. 11, 23 |
| BAL | Broad Absorption Line. 71 |
| BAO | Baryon Acoustic Oscillations. vii, 8, 10, 20, 22, 80, 81, 100, 106 |
| BBN | Big Bang Nucleosynthesis. 4, 5, 16, 17 |
| BOSS | Baryon Oscillations Spectroscopic Survey. vii, 12, 13, 23, 24, 29, 32, 33, 41, 49, 54, 58, 59, 67, 69, 70, 72, 80, 93, 99, 101, 135 |
| CDM | Cold Dark Matter. 6–8, 18, 19 |
| CFHTLenS | a lensing catalogue. 59 |
| CMASS | a catalogue of “constant mass" luminous infrared galaxies from BOSS. vii, 13, 25, 29, 33–35, 49, 51–56, 58–60, 119, 127, 128 |
| CMB | Cosmic Microwave Background. 5, 7, 17, 18, 22 |
| CNR | continuum-to-noise ratio. 70 |
| COBE | COsmic Background Explorer. 7, 19 |
| DE | dark energy. 19, 20 |
| DLA | Damped Ly$\alpha$ Absorber. vii, viii, 12, 13, 24, 25, 67–73, 79–82, 85, 86, 89, 92–94, 97, 119, 129, 130, 133 |
| DM | dark matter. 18, 20–22, 100 |
| DR | Data Release. 67, 93, 99 |
| DR10 | tenth Data Release. 33 |
| DR11 | eleventh Data Release. 29, 33, 59 |

| | |
|---|---|
| DR12 | twelfth Data Release. 70, 72, 101, 102 |
| DR12Q | DR12 quasar catalogue. 12, 23, 101 |
| DR7 | seventh Data Release. 29, 32, 33, 36, 38, 41, 52, 59, 61, 127 |
| DR9 | ninth Data Release. 33, 54 |
| DR9Q | DR9 quasar catalogue. 101 |
| FLRW | Friedmann–Lemaître–Robertson–Walker. 3, 15 |
| FRW | Friedmann–Robertson–Walker. 15 |
| GR | general relativity. 15, 19 |
| IGM | intergalactic medium. 6, 12, 21, 23, 109, 119, 120 |
| LSS | large-scale structure. 18, 79 |
| MJD | Modified Julian Date. 71 |
| MTC | Mean Transmission Correction. 74, 80, 82, 93 |
| PSF | Point Spread Function. 43, 94, 95 |
| RMS | root mean square. 44, 45, 61, 128 |
| SDSS | Sloan Digital Sky Survey. vii, 31, 33, 35, 59, 61, 100 |
| SDSS-II | Sloan Digital Sky Survey II. 29, 32 |
| SDSS-III | Sloan Digital Sky Survey III. 13, 24, 29, 31–33, 69, 70, 94, 101 |
| SN | supernova. 5, 7, 8, 11, 13, 17, 19, 20, 22, 24 |
| WMAP | Wilkinson Microwave Anisotropy Probe. 7, 19 |


effect
cloud
matter
forest
alpha
obtain
depend
same
DLA
result
value
bin
interval
width
noise
high
follow
even
model
cross
left
system
linear
one
factor
smooth
show
absorb
measure
line
first
form
sample
compute
consider
include
halo
density
absorption
stack
present
over
profile
correlation
large
final
cut
red
note
data
set
bias
quasar
redshift
method
equivalent
mass
spectra
fit
real
scale
average
error
use
mean
detect
observe
metal
estimate
pixel
function
top
qso
galaxy
lyman
matrix
number
universe
impact
continuum
limit
velocity
space
parameter
tracer
both
more
power
non