Simuliranje spektara kvazara iz Sloan Digital Sky Survey

Milaković, Dinko

Master's thesis / Diplomski rad

2015

Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj: **University of Zagreb, Faculty of Science / Sveučilište u Zagrebu, Prirodoslovno-matematički fakultet**

Permanent link / Trajna poveznica: https://urn.nsk.hr/urn:nbn:hr:217:636197

Rights / Prava: In copyright / Zaštićeno autorskim pravom.

Download date / Datum preuzimanja: **2024-03-29**

Repository / Repozitorij:

Repository of the Faculty of Science - University of Zagreb

SVEUČILIŠTE U ZAGREBU
PRIRODOSLOVNO-MATEMATIČKI FAKULTET
FIZIČKI ODSJEK

Dinko Milaković

SIMULIRANJE SPEKTARA KVAZARA IZ SLOAN DIGITAL SKY SURVEY

Diplomski rad

Zagreb, 2015.

SVEUČILIŠTE U ZAGREBU
PRIRODOSLOVNO-MATEMATIČKI FAKULTET
FIZIČKI ODSJEK

SMJER: ISTRAŽIVAČKI

Dinko Milaković

Diplomski rad

Simuliranje spektara kvazara iz Sloan Digital Sky Surveya

Voditelj diplomskog rada: doc. dr. sc. Vernesa Smolčić

Ocjena diplomskog rada: _____

Povjerenstvo: 1. _____

2. _____

3. _____

Datum polaganja: _____

Zagreb, 2015.

Acknowledging everyone who contributed to the production of this thesis would be a daunting task, filling several pages.

I thank all of my classmates and friends – Dean, Branimir, Nikolina, Matej, Davor, Tamara, Petra, Benjamin, and Petar, along with many others whom I spent uncountable productive days and unproductive nights with – it would not have been the same. To the Astro team in Zagreb, thanks for valuable insights into the Astrophysical community and all of the fun conversations over the shared lunches and coffees. Special thanks to Jacinta for her invaluable comments on this thesis.

I thank my 'adult' friends with paying jobs, Mislav, Dinko, Filip, and Kristijan, for showing me that there are both better and worse things than being a student.

I thank my Mother for all her love, and for fulfilling all of my wishes without hesitation. My Australian adventure would not be possible without the support of my family, both domestic and international. I thank all of them for it. Constant supervision of my work by my Aunt in Sydney sped things up significantly.

Finally, the greatest thanks are extended to my supervisors: Assistant Professor Vernesa Smolčić and Professor John Kelvin Webb. I thank Vernesa for introducing me into the world of Astrophysics research, many friendly conversations and advice, and most importantly – paying for my ticket. I thank John for being a great supervisor in every sense of the word – not everyone would accept to supervise a student who contacted them from the other side of the world, and make them feel welcome. Thanks for providing the opportunity to work on such an interesting (and demanding) project.

Simulating Quasar Spectra from the Sloan Digital Sky Survey

Abstract

Quasars are the most distant astrophysical objects in the observable Universe that can be studied at high precision. The features seen in their spectra shed light onto the evolution of the Universe in the last 12 billion years. The quasar's emission line properties and the absorption features created by intervening gas clouds, known as the Lyman α forest, contain a wealth of data useful in cosmological studies. They can be used to directly probe the distribution of baryonic and dark matter, to study baryonic acoustic oscillations, the expansion history of the Universe, locate anisotropies, constrain cosmological models, and many more.

The purpose of this thesis is to generate a comprehensive set of simulated quasar spectra, providing an important tool for exploring a wide range of scientific topics derivable from observed quasar spectra. The main focus will be on Sloan Digital Sky Survey data, motivated by a project seeking to constrain any large-scale anisotropies in the Universe. The project uses quasar spectra from Data Release 10 to look for variations in the observed optical depth of neutral hydrogen across the sky, at any given redshift. This thesis will not discuss the actual spectral analysis performed to study the opacity of neutral hydrogen, but will focus on the necessity and production of mock data.

Mock catalogues will allow us to identify and remove systematic errors, provide quantitative checks on other statistical methods used, and help to interpret scientific results. To produce the artificial quasar spectra with the observed properties of the Lyman α forest, I continued on previous work of John K. Webb and his program QSOSIM9. My new version of the program, QSOSIM10, generates the Lyman α forest under the assumption that the forest is represented fairly well by a set of discrete absorption lines. The program features an far more extensive set of the emission line features and new models of the column density and redshift distributions of neutral hydrogen, derived from high-resolution quasar spectra. The simulated spectra include physical effects seen in real spectra, such as the clustering of the intergalactic medium along the line-of-sight, the quasar proximity effect, and the heavy element transitions statistically, using known parameter distributions. The simulations employ the use of CLOUDY (a plasma simulation code) and CUBA (a radiative transfer code) calculations to derive heavy element column densities and model the cosmic ultraviolet background. The simulated spectra and the software developed in this thesis will also be suitable for other applications.

Simuliranje spektara kvazara iz Sloan Digital Sky Surveya

Prošireni sažetak

Kvazari su klasa najudaljenijih astrofizičkih objekata čijim proučavanjem možemo detaljno proučavati povijest svemira kroz zadnjih 12 milijardi godina. Niz značajki vidljiv u spektrima kvazara, poput emisijskih linija kvazara i linija nastalih apsorpcijom na oblacima međugalaktičkog plina između nas i kvazara koje nazivamo Lyman α šumom, sadrže mnoštvo korisnih informacija za izučavanje kozmoloških pitanja. Lyman α šuma je korisna u istraživanjima koja proučavaju raspodjelu barionske i tamne materije u svemiru, akustične barionske oscilacije, povijest širenja svemira, određuju kozmološke parametre i mnoge druge.

Cilj ovog rada je proizvesti sveobuhvatan skup simuliranih spektara kvazara i pritom konstruirati alat koristan za izučavanje širokog raspona znanstvenih pitanja koja proizlaze iz spektara kvazara. Fokus rada je na simuliranju spektara iz Sloan Digital Sky Survey (SDSS). Simulirani spektri biti će korišteni u sklopu projekta koji za cilj ima proučavanje anizotropije svemira. Ovaj projekt analizira spektre velikog broja kvazara iz Objave podataka 10[1] u potrazi za varijacijama u optičkoj dubini neutralnog vodika u svim dostupnim smjerovima i na svim udaljenostima. Ovaj rad izlaže motivaciju i postupak kreiranja simuliranih spektara i neće se baviti opisivanjem spektralne analize korištene u projektu.

Simulirani katalog, finalni produkt ovog rada, pomoći će pri otkrivanju i otklanjanju sistematskih efekata prisutnih u stvarnom uzorku, povećati pouzdanost u korištene metode i pomoći u interpretaciji rezultata. Ovaj rad nastao je na temelju prethodnog programa za simulaciju spektra kvazara, QSOSIM9, kojeg je napisao John K. Webb. Nova verzija spomenutog programa, nazvana QSOSIM10, simulira Lyman α šumu karakterizirajući ju kao niz apsorpcijskih linija u spektru. Novi program unaprijeđuje prethodnu verziju uključivanjem većeg broja emisijskih linija, novih modela evolucije međugalaktičkog medija i raspodjele stupčane gustoće[2] oblaka vodika u njemu. Nadalje, program koristi statističke metode kako bi simulirao fizikalne efekte poput zgrudnjavanja oblaka međugalaktičkog plina u jata duž doglednice, inverznog efekta u blizini kvazara[3] i apsorpcije na metalima u oblacima. Program također koristi kod za simulaciju plazme CLOUDY i kod za simulaciju propagacije zračenja CUBA za modeliranje apsorpcije na metalima i kozmičke

[1] engl. *Data Release 10*
[2] engl. *column density*
[3] engl. *proximity effect*

ultraljubičaste pozadine. Osim u gore navedene svrhe, kod i njegovi rezultati biti će podobni za mnoge druge primjene.

Ovaj produljeni sažetak namjenjen je vođenju čitatelja kroz engleski tekst rada. Svaki dio produljenog sažetka ukratko izlaže sadržaj i pojašnjava bitne koncepte pojedinog poglavlja u radu.

Prvo poglavlje daje kratki osvrt na povijest razvoja fizikalne kozmologije, izlaže detalje o Sloan Digital Sky Survey (SDSS) – najvećem modernom pregledu neba, pojašnjava ulogu simulacija u kozmologiji, daje pregled povijesti kvazara i njihove važnosti za kozmologiju, te izlaže probleme čije rješavanje je motiviralo ovaj rad.

Drugo poglavlje daje pregled već postojećeg programa za simulaciju spektra kvazara, QSOSIM9, te izlaže pretpostavke i modele korištene za simulaciju spektra.

Treće poglavlje je ujedno i središnji dio ovog rada. Ovo poglavlje predstavlja unaprijeđenja uvedena u QSOSIM10. Poglavlje izlaže teorijsku pozadinu novih modela emisijskih linija kvazara, raspodjele stupčanih gustoća neutralnog vodika u intergalaktičkom mediju, evolucije broja apsorpcijskih linija u Lyman α šumi, zgrudnjavanje oblaka duž doglednice, distribucije širina apsorpcijskih linija, uporabe programa CLOUDY za modeliranje apsorpcije na metalima i uporabe programa CUBA za modeliranje inverznog efekta. Konačno, poglavlje opisuje modeliranje karakteristika SDSS spektroskopa i njihovo uvođenje u konačni spektar.

Rezultati programa QSOSIM10 predstavljeni su u četvrtom poglavlju. U ovom poglavlju dana je usporedba simuliranih spektara i spektara opaženih u SDSS-u te su raspravljeni testovi za kvantifikaciju slaganja simuliranih i realnih spektara.

Peto poglavlje zaključuje ovaj rad raspravom o mogućim primjenama i planovima za buduća poboljšanja programa QSOSIM10.

Uvod

Kruna fizikalne kozmologije je teorija Velikog praska, koja prepostavlja da je svemir *homogen* i *izotropan* – odnosno, da svemir izgleda jednako u svakom svom dijelu i u svakom smjeru. Ove dve pretpostavke zajednički se nazivaju *kozmološkim načelom*. Istinitost kozmološkog načela potvrđuju rezultati mnogih studija provedenih tijekom zadnja dva desetljeća, poput misija satelita COBE [1], WMAP [2] i Planck [3]. Međutim, nekoliko nedavnih studija dalo je indikacije o narušenosti kozmološkog načela, poput neočekivane usklađenosti smjera gibanja velikog broja galaksija na skalama ~ 100 Mpc [4], varijacije fundamentalne konstante fine strukture [5], [6] i drugih.

Projekt izučavanja barionskih akustičnih oscilacija u sklopu SDSS-a, BOSS[4], priku-

[4]Baryonic Oscillations Spectroscopic Survey

pio je optičke spektre više od 294 000 kvazara raspoređenih preko 10 000 kvadratnih stupnjeva neba. SDSS BOSS koristi specijalizirani teleskop za prikupljanje fotometrijskih i spektroskopskih podataka velikog broja objekata istovremeno. Detalji o fotometrijskim i spektrometrijskim karakteristikama SDSS teleskopa i BOSS spektrografa dani su, redom, u [7] i [8]. Korištenje velikih skupova spektrografskih podataka kvazara zahtjeva detaljno razumjevanje mogućih sistematskih efekata u korištenom uzorku. Kvantificiranje takvih efekata olakšano je uporabom računalnih simulacija. Monte Carlo metode posebna su vrsta metoda korištena u simulacijama koje koriste nasumične brojeve kako bi ostvarili jednu od mogućih raspodjela vjerojatnosti. U ovom radu ćemo koristiti Monte Carlo metode za simulaciju međugalaktičkog medija u Lyman α šumi.

Motivacija ovog rada je višestruka:

1. SDSS algoritam za odabir kvazara bira kandidate za spektroskopsko opažanje na temelju njihovih položaja u boja-boja prostoru, koji se razlikuju od položaja drugih vrsta objekata poput galaksija ili zvijezda. Međutim, kvazari na određenim crvenim pomacima (najviše na crvenim pomacima $2.7 < z < 3.5$) bojama su jako slični zvjezdama, što otežava njihovo identificiranje. U području jakog miješanja zvijezda i kvazara, samo nekoliko postotaka kvazara na nebu je odabrano za spektroskopsko opažanje automatskim SDSS BOSS algoritmom zbog velike sličnosti fotometrijskih boja kvazara u navedenom rasponu crvenih pomaka i zvijezda u našoj galaksiji [9], [10] (vidi sliku 1.8). Osim toga, kriteriji korišteni pri odabiru kandidata za spektroskopiju su mijenjani 19 puta tokom prve dvije godine SDSS BOSS projekta [11] (vidi sliku 1.9). Primjena istih kriterija na simulirane podatke omogućit će kvantifikaciju utjecaja ovih efekata na stvarni uzorak.

2. Rezultati svih studija koje koriste podatke iz Lyman α šume ovise o preciznom određivanju toka zračenja kvazara u slučaju bez apsorpcije u intergalaktičkom mediju. Sveobuhvatni set simuliranih podataka može pomoći unaprijediti metode određivanja istog.

3. Nehomogenosti svemira na malim skalama uzrokuju statistička odstupanja mjerenih vrijednosti kozmoloških parametara. Utjecaj tih odstupanja onemogućava zaključivanje o svojstvima svemira na skalama značajno većim od proučavanog područja i naziva se *kozmičkom varijancom*. Simulacije različitih ostvarenja neba poslužit će za kvantificiranje ovog efekta.

Simuliranje spektra kvazara - QSOSIM9

Ovaj rad nastavlja i nadograđuje prethodno napisani program za simulaciju spektra kvazara proizvoljne rezolucije, QSOSIM9 [12], kojeg je napisao profesor John K. Webb sa Sveučilišta Novog Južnog Walesa u Australiji. Ovo poglavlje daje općeniti pregled konstrukcije tog programa i modela korištenih pri stvaranju spektra. QSOSIM9 simulira najbitnije značajke prisutne u spektru kvazara poput emisijskih linija kvazara i Lyman α šume (pretpostavljajući da je šuma dobro opisana kao skup pojedinačnih apsorpcijskih linija), koristeći se parametrima unijetim od strane korisnika: crvenim pomakom kvazara z_{QSO}, spektralnim ideksom α, V magnitudom Johnson-Cousins sustava i drugih. Za modeliranje šume, program prvo treba znati koliko je linija raspoređeno duž doglednice. Za izračun broja apsorpcijskih linija program se koristi raspodjelom diferencijalnog broja apsorpcijskih linija neutralnog vodika po jedinici crvenog pomaka. Ta je raspodjela empirijski dobro opisana jednadžbom (2.3) te se stoga broj apsorpcijskih linija duž doglednice u Lyman α šumi može izračunati jednadžbom (2.4).

Svaka apsorpcijska linija jedinstveno je opisana trima parametrima: svojim položajem u prostoru brzina (crvenim pomakom), stupčanom gustoćom i širinom linije (Dopplerovim b-parametrom). Za određivanje svakog od tri navedena parametra, QSOSIM9 koristi jednostavne analitičke modele i Monte Carlo algoritme. Vrijednosti crvenih pomaka i stupčanih gustoća program uzorkuje koristeći bijekciju između skupa realnih brojeva u intervalu $x \in [0,1]$ i pripadajućih kumulativnih distribucija. Kumulativna distribucija crvenih pomaka izvedena je iz jednadžbe (2.3) i opisana u potpoglavlju 2.3. Distribucija stupčane gustoće neutralnog vodika (u daljnjem tekstu: N_{HI}) opisana je jednostavnim analitičkim modelom u obliku potencijske funkcije, dane jednadžbom (2.6). Ovim potencijskim zakonom, QSOSIM9 opisuje distribuciju stupčanih gustoća u rasponu od $N_{\mathrm{HI}} = 10^{13}$ cm^{-2} do $N_{\mathrm{HI}} = 10^{17}$ cm^{-2}. Kumulativna distribucija stupčanih gustoća vodika opisana je u potpoglavlju 2.4. Konačno, vrijednosti b-parametara program uzorkuje normalnu distribuciju sa srednjom vrijednosti $\mu = 23$ km s^{-1} i standardnom devijacijom $\sigma = 3$ km s^{-1}. QSOSIM9 također dopušta korisniku unos do 20 apsorpcijskih sustava s proizvoljnim parametrima.

Nakon unosa apsorpcijskih linija, program konvoluira spektar s Gaussovim profilom čija je standarnda devijacija jednaka vrijednosti unešenoj od strane korisnika i svodi spektar na željenu rezoluciju. Konačno, program uvodi šum u generirani spektar. Primjer spektra generiranog pomoću QSOSIM9 prikazan je na slici 2.1. Generirani spektar je spremljen u numeričkom obliku u tekstualnu datoteku '*spec.dat*'. Opis konstrukcije spektra, korak po korak, također je dan u ovom poglavlju.

Simuliranje spektra kvazara - QSOSIM10

Ovo poglavlje predstavlja unaprijeđenja uvedena u novu iteraciju QSOSIM programa, nazvanu QSOSIM10. Cilj nadogradnji uvedenih u QSOSIM10 je simuliranje spektara kvazara sa svojstvima sto sličnijima stvarnim spektrima u SDSS BOSS katalogu [13]. Svaki kvazar simuliran ovim programom odgovara stvarnom kvazaru u pogledu pripadajućeg crvenog pomaka, spektralnog indeksa, magnituda (u SDSS sustavu [14]) i pozicije na nebu. Osim toga, simulirani spektri moraju reproducirati instrumentalne karakteristike SDSS BOSS spektrografa, poput raspona valnih duljina, odnosa signala i šuma u spektru te rezolucije. Nadalje, QSOSIM10 koristi poboljšane algoritme za modeliranje intergalaktičkog medija i Lyman α šume te nove modele distribucija stupčane gustoće, crvenog pomaka i Dopplerovog b-parametra, a u skladu s novijim saznanjima dobivenih iz spektara visoke rezolucije. Detalji o unaprijeđenjima i modelima korištenima u QSOSIM10 programu opisani su u ovom poglavlju, no mogu se sažeti u sljedećih 8 točaka:

1. Program predstavlja gotovo dvostruko brojniji skup širokih i uskih emisijskih linija iz područja u neposrednoj blizini crne rupe. Prisustvo širokih emisijskih linija jedno je od identifikacijskih osobina kvazara, također važno u studijama masa supermasivnih crnih rupa u središtu svakog kvazara. Svojstva linija, poput njihove relativne jakosti i širine, preuzeta su iz istraživanja kompozitnih (uprosječenih) spektara kvazara i tabulirana u Dodatku A.

2. Model stupčanih gustoća neutralnog vodika je unaprijeđen u skladu s najnovijim empirijskim rezultatima te pokriva veći raspon: $N_{\mathrm{HI}} = 10^{12} - 10^{22}$ cm^{-2}. Potencijski zakon korišten u QSOSIM9 zamijenjen je krivuljom bez analitičkog opisa, prikazane na slici 3.5. Nije poznato mijenja li se oblik ove krivulje s porastom crvenog pomaka pa QSOSIM10 pretpostavlja da je oblik stalan. Algoritam kojim se svakoj apsorpcijskoj liniji neutralnog vodika H I dodjeljuje vrijednost stupčane gustoće u osnovi je ostao nepromijenjen u odnosu na prethondu verziju, no kumulativnu distribuciju gustoće QSOSIM10 računa numeričkom integracijom gorespomenute neanalitičke krivulje. Za određivanje stupčane gustoće svake linije, program uzorkuje kumulativnu distribuciju nasumičnim brojem. Ovaj je proces pojašnjen u potpoglavlju 3.3.1.

3. Iz spektara kvazara visoke razlučivosti poznato je da oblaci različitih stupčanih gustoća evoluiraju različitim brzinama [15]–[17]. Model evolucije broja oblaka H I po jedinici crvenog pomaka, $\mathrm{d}n/\mathrm{d}z$ (opisan jednadžbom (3.12)), unaprijeđen je da opisuje evoluciju tri populacije oblaka različitih stupčanih gustoća: $N_{\mathrm{HI}} = 10^{12} - 10^{14}$ cm^{-2}, $N_{\mathrm{HI}} = 10^{14} - 10^{17}$ cm^{-2}, i $N_{\mathrm{HI}} = 10^{17} - 10^{22}$ cm^{-2}.

Obzirom na nedostatak detaljnih informacija o evoluciji oblaka najmanjih stupčanih gustoća, 10^{12} cm^{-2}< $N_{\rm HI}$ < $10^{13.1}$ cm^{-2}, njihov broj je ekstrapoliran pod pretpostavkom da tvore istu populaciju skupa s oblacima nešto veće stupčane gustoće. Potpoglavlje 3.3.2 pojašnjava pretpostavke i postupak renormalizacije broja oblaka populacije najmanjih stupčanih gustoća. Nadalje, isto potpoglavlje objašnjava kako QSOSIM10 predviđa broj apsorpcijskih linija u simuliranom spektru (vidi jednadžbu (3.14)), uzimajući u obzir različite brzine evolucije diferencijalnog broja linija za tri populacije oblaka. Koristeći se Monte Carlo metodama, program određuje ukupni broj linija po jednadžbi (3.15).

4. Diferencijalni broj linija po jedinici crvenog pomaka, $\mathrm{d}n/\mathrm{d}z$, u stvari opisuje generalni trend i ne reflektira *stvarnu* raspodjelu oblaka duž doglednice. Varijacije u gustoći tvari u svemiru (odnosno zgrudnjavanje) uzrokuju odstupanja stvarne raspodjele od ovog trenda. QSOSIM10 unaprijeđuje stari model iz QSOSIM9, te statističkim metodama modelira zgrudnjavanje oblaka međugalaktičkog plina. Statističke studije osobina međugalaktičkog plina su pokazale da se oblaci zgrudnjavaju na skalama $v_0 \sim 100-250$ km s^{-1} (u prostoru brzina) [18]. Program koristi Monte Carlo metode i jednadžbu (3.16) kako bi odredio broj klastera (mjesta zgrudnjavanja oblaka) u spektru, nakon čega svakom klasteru nasumce dodijeli poziciju u prostoru crvenog pomaka, z_{cl}. Radi jednostavnosti, svakom klasteru je dodijeljen jednak profil gustoće $\nu(z_{cl})$ i bogatstvo $h(z_{cl})$. Svaki klaster utječe na vrijednost $\mathrm{d}n/\mathrm{d}z$ te se stoga ta vrijednost za određeni crveni pomak z računa kao zbroj doprinosa svakog klastera u spektru:

$$\frac{\mathrm{d}n}{\mathrm{d}z} = \sum_{\text{klaster}} h(z_{cl})\nu(\Delta z; z_{cl}).$$

Numeričkom integracijom vrijednosti $\mathrm{d}n/\mathrm{d}z$ duž doglednice program računa kumulativnu distribuciju crvenih pomaka linija. Koristeći Monte Carlo algoritam, QSOSIM10 svakoj apsorpcijskoj liniji dodjeljuje nasumični broj između 0 i 1. Nadalje, koristeći se bijektivnom funkcijom koja povezuje kumulativnu distribuciju crvenih pomaka i interval realnih brojeva $x \in [0,1]$, program linijama dodjeljuje vrijednosti crvenog pomaka. Detalji modela zgrudnjavanja dani su u potpoglavlju 3.3.3. Postupak pridruživanja crvenih pomaka linijama ilustriran je na slici 3.7, uz usporedbu modela zgrudnjavanja s modelom korištenim u QSOSIM9. Slika prikazuje primjer kvazara na crvenom pomaku $z_{QSO} = 3.5383$, u čiji spektar je umetnuto 197 klastera između crvenih pomaka $z_1 = 1.9261$ i $z_2 = 3.5383$, a ukupni broj H I apsorpcijskih linija u Lyman α šumi je 1820. Svakoj liniji u spektru je dodjeljen nasumični broj između 0 i 1, nakon čega

je svakoj dodjeljena vrijednost crvenog pomaka pomoću modela iz QSOSIM9 i QSOSIM10. Vrijednosti crvenih pomaka dobivene korištenjem novog modela prikazane su kao niz crvenih linija na vrhu slike 3.7, dok su vrijednosti dobivene korištenjem starog modela iz QSOSIM9 prikazane kao niz plavih linija. Model korišten u QSOSIM10 reproducira zgrudnjavanje linija u klastere, u skladu sa opažanjima.

5. QSOSIM10 koristi isti algoritam za određivanje vrijednosti b-parametra kao i QSOSIM9, odnosno uzorkuje normalnu raspodjelu Monte Carlo algoritmom. Međutim, novi model koristi parametre koji bolje opisuju stvarnu raspodjelu, opisanu u potpoglavlju 3.3.4 i prikazanu na slici 3.8. Analiza ove raspodjele pokazala je da se ona najbolje može opisati kao superpozicija četiri normalne raspodjele. Dvije od četiri raspodjele su asocirane s neutralnim vodikom (raspodjele sa srednjim vrijednostima $b \approx 22$ km s^{-1}), jedna s metalima (raspodjela sa srednjom vrijednosti $b \approx 6$ km s^{-1}), a jedna je rezultat artefakata u analizi. Rezultati najbolje prilagodbe spomenutih četiriju normalnih raspodjela na raspodjelu širina linija dani su u tablici 3.4. Također, analizom širine linija u ovisnosti o crvenom pomaku istih, pokazano je da se srednje vrijednosti i širina raspodjele ne mijenjaju značajno porastom crvenog pomaka (slika 3.9). Na temelju te analize, QSOSIM10 uzorkuje normalnu raspodjelu sa srednjom vrijednosti $b = 22$ km s^{-1} i standardnom devijacijom $\sigma = 5$ km s^{-1} za određivanje širine apsorpcijskih linija neutralnog vodika, te raspodjelu sa srednjom vrijednosti $b = 6$ km s^{-1} i standardnom devijacijom $\sigma = 1.5$ km s^{-1} za određivanje širine apsorpcijskih linija metala.

6. Još jedna novina uvedena u novu inačicu programa je apsorpcija na metalima u međugalaktičkom mediju, opisana u potpoglavlju 3.3.5. U astrofizici, metalima se nazivaju svi elementi teži od helija. Metali u Lyman α šumi bitni su za izučavanje formiranja galaksija i njihove evolucije, evolucije kozmičkog metaliciteta, ultraljubičaste pozadine i mnoge druge studije. Metali su asocirani s oblacima neutralnog vodika velikih stupčanih gustoća – oblaci s $N_{\mathrm{HI}} \geqslant 10^{15}$ cm^{-2} pokazuju asocijaciju s metalima u 73% slučajeva [19]. QSOSIM10 modelira apsorpciju na metalima za sustave iznad praga $N_{\mathrm{HI}} = 10^{17}$ cm^{-2}. Kao i u slučaju neutralnog vodika, za modeliranje apsorpcije na metalima, potrebno je znati vrijednost crvenog pomaka, stupčane gustoće i širine svake linije. Crveni pomak metala, z, isti je kao i crveni pomak neutralnog vodika u oblaku. Za određivanje stupčanih gustoća metala, QSOSIM10 koristi napredni kod za simulaciju plazme CLOUDY, kojeg su razvili znanstvenici sa Sveučilišta Cambridge. CLOUDY simulira radijativni prijenos u plinu van

termodinamičke ravnoteže i predviđa niz svojstava, među kojima i stupčane gustoće. Popis metala s najvećim stupčanim gustoćama, a koji su modelirani u spektru, dan je u ovom potpoglavlju. Kao ulazne parametre, CLOUDY uzima vrijednosti crvenog pomaka oblaka međugalaktičkog plina i njegovog metaliciteta, kao i volumne i stupčane gustoće vodika. Metalicitet svakog sustava međugalaktičkog plina iznad gore definiranog praga, Z, određen je primjenom Monte Carlo algoritma i formulom (3.30). Korištenje CLOUDY-ja za svaki sustav iznad praga produljilo bi proces proizvodnje spektra za red veličine. QSOSIM10 stoga koristi bazu podataka s podacima o stupčanim gustoćama metala, koju poziva za svaki sustav koji ima stupčanu gustoću vodika iznad definiranog praga. Baza podataka je konstruirana diskretno uzorkujući trodimenzionalni parametarski prostor (Z,N_{HI},z) u 20 000 točaka (20 točaka u metalicitetu Z, 50 točaka u stupčanoj gustoći neutralnog vodika N_{HI} i 20 točaka u crvenom pomaku z). Za računanje stupčanih gustoća metala u svakoj točki parametarskog prostora korišten je CLOUDY 13.03 [20]. Rezultati su spremljeni u bazu podataka iz koje QSOSIM10 čita vrijednosti stupčanih gustoća metala točke parametarskog prostora koja je najbliža vrijednostima sustava. Primjer ulazne datoteke korištene za konstrukciju baze dan je u Dodatku B. Vrijednosti b-parametra svake linije određene su nasumičnim uzorkovanjem normalne distribucije sa srednjom vrijednosti $b = 6$ km s^{-1} i standardnom devijacijom $\sigma = 1.5$ km s^{-1} (vidi potpoglavlje 3.3.4).

7. U spektrima visoke rezolucije primjećen je fizikalni efekt smanjenja broja apsorpcijskih linija iznad određenog praga stupčane gustoće [21]. Ovaj efekt, izazvan snažnim tokom ionizirajućeg zračenja u neposrednoj blizini kvazara, zbog prividnog okretanja trenda porasta broja linija porastom crvenog pomaka, nazvan je *inverznim* efektom. QSOSIM10 uključuje ovaj efekt pri simulaciji spektra na način koji je opisan u potpoglavlju 3.3.6. Daleko od izvora ionizirajućeg zračenja, međugalaktički medij je ioniziran ultraljubičastom kozmičkom pozadinom (integriranim ultraljubičastim zračenjem galaksija i aktivnih galaktičkih jezgara). U blizini snažnog izvora zračenja poput kvazara udio neutralnog vodika značajno je smanjen, odnosno izmjerena stupčana gustoća je manja. Smanjenje izmjerene stupčane gustoće dano je izrazom 3.32, gdje je veličina ω proporcionalna omjeru ionizirajućeg zračenja koje dolazi od kvazara i zračenja ultraljubičaste pozadine. Količina zračenja koja dolazi od kvazara obrnuto je proporcionalna kvadratu luminozitetne udaljenosti od oblaka do kvazara, a proporcionalna monokromatskom luminozitetu kvazara na Lymanovoj granici ($\lambda = 912$ Å). S druge strane, za određivanje doprinosa ul-

traljubičaste kozmičke pozadine korišten je program za radijativni prijenos u međugalaktičkom mediju zvan CUBA, koji je opisan u [22]. Rezultati CUBA-e javno su dostupni na web stranicama [23]. Kako bi se izbjegla interpolacija između točaka, ultraljubičasta pozadina prilagođena je kao superpozicija dvije normalne i jedne logaritmirane normalne krivulje, odnosno izrazom (3.40). Rezulat prilagodbe dan je u tablici 3.5 i prikazan na slici 3.11. Konačno, primjer inverznog efekta na mjerene stupčane gustoće dan je na primjeru kvazara $z_{QSO} = 3.5383$, na slici 3.12.

8. QSOSIM10 je konstruiran u cilju simuliranja spektara kvazara opaženih SDSS BOSS spektrografom. Simulirani spektri stoga moraju imati svojstva BOSS spektrografa, poput rezolucije i karakterističnog šuma u signalu. Takav zahtjev traži temeljito poznavanje elektroničkih i mehaničkih komponenti BOSS spektrografa, kao i načina na koji se signal obrađuje. Rezolucijski elementi BOSS spektrografa (u daljnjem tekstu 'pikseli') su konstantne veličine u prostoru brzina, $p^{\,v} = 69$ km s^{-1} [24]. Nakon unosa emisijskih i apsorpcijskih linija u spektar, simulirani se spektar konvoluira s Gaussovim profilom čija je standardna devijacija $\sigma = p^{\,v}$, kako bi simulirani spektri veličinom odgovarali BOSS pikselima. Simulacija razine šuma u signalu zahtjeva poznavanje propusnosti teleskopa za sve fotone koji od kvazara dolaze do Zemlje i načina na koji se broj upadnih fotona pretvara u tok zračenja kvazara. Propusnost teleskopa ovisi o efektima poput atmosferske apsorpcije, kvalitete optičkih i elektroničkih komponenti teleskopa ili učinkovitosti pretvaranja fotona u elektrone u kameri. Grafički prikaz propusnosti BOSS spektrografa prikazana je na slici 3.13, a model korišten u QSOSIM10 je prikazan na slici 3.14. Poznavajući tehničke karakteristike teleskopa, QSOSIM10 koristi formulu (3.41) za pretvaranje toka zračenja u broj detektiranih fotona i obrnuto. Tok zračenja simuliranog kvazara pretvoren je u broj upadnih fotona za svaki piksel spektra nakon čega je, korištenjem formule (3.44), za svaki piksel izračunata recipročna vrijednost varijance, $1/\sigma^2$ (u broju fotona). Za konačno simuliranje šuma, svakom pojedinačnom pikselu u spektru dodana je vrijednost šuma. Ta je vrijednost određena Monte Carlo metodom, odnosno uzorkovanjem normalne distribucije sa srednjom vrijednosti $\mu = 0$ i standardnom devijacijom σ (iz formule (3.44)). Rezultantna vrijednost toka sa pridodanim šumom je na kraju pretvorena iz broja fotona u tok zračenja inverzijom formule (3.41).

Ovdje sažeti opis 3. poglavlja daje osnovni pregled unaprijeđenja i modela fizikalnih efekata korištenih u QSOSIM10, no čitatelja se upućuje na čitanje cijelog poglavlja za potpunu sliku.

Rezultati programa QSOSIM10

Set simuliranih spektara kvazara iz Sloan Digital Sky Survey glavni je rezultat razvoja programa QSOSIM10. Nova verzija QSOSIM programa poboljšava prije korištene modele i simulira fizikalne efekte koje prethodne verzije nisu uključivale. Proces generacije čitavog skupa kvazara iz SDSS BOSS DR10 kataloga [13] opisan je, korak po korak, u ovom poglavlju. Proces simulacije spektra ilustriran je na primjeru kvazara na crvenom pomaku $z_{QSO} = 3.5383$, r magnitude $r = 19.99$ i spektralnog indeksa $\alpha = 0.6$ (vidi slike 4.1 i 4.2). U ovom poglavlju prikazana je i vizualna usporedba simuliranih i pravih SDSS spektara za deset kvazara. Vizualnim pregledom vidi se da su simulirani i pravi spektri naoko vrlo slični: simulirani spektri oponašaju gotovo sve spektralne karakteristike vidljive u pravim spektrima, poput jačine i širine većine emisijskih linija te razine šuma. Očigledna nekonzistentnost jest razina šuma na valnim duljinama $\lambda \gtrsim 7200$ Å, koja ne oponaša apsorpciju fotona na Zemljinoj atmosferi. Također, zbog statističke prirode jakosti emisijskih linija u simuliranim spektrima, oni ponekad ne uspijevaju vjerno reproducirati snažne i uske vrhove Ly α emisijske linije kvazara koje su vidljive u pravim spektrima.

Postoji nekoliko testova kojima bi se kvantificiralo slaganje simulacija i stvarnih spektara – ukoliko simulacije vjerno opisuju svojstva kvazara i međugalaktičkog medija, spektralna analiza kojom proučavamo optičku dubinu neutralnog vodika u svemiru trebala bi dati konzistentne rezultate za simulirane i prave spektre. Osim toga, kompozitni spektri dobiveni uprosječavanjem simuliranih spektara trebali bi imati ista svojstva kao i kompozitni spektri dobiveni iz realnih, te bi ekvivalentne širine emisijskih linija kvazara trebale biti konzistentne. Nijedan od predloženih testova, međutim, nije napravljen na vrijeme da bude uključen u ovaj rad.

Zaključak, primjene i budućnost programa QSOSIM10

Ovaj rad predstavio je proces konstrukcije programa QSOSIM10 i njegovog najbitnijeg rezultata – najopsežnijeg i najsofisticiranijeg skupa simuliranih spektara kvazara iz SDSS-a. Program proizvodi simulirane spektre kvazara koristeći se podacima iz SDSS kataloga kvazara i karakteristika BOSS spektrografa, pritom modelirajući fizikalne efekte pomoću najtočnijih i najnovijih empirijskih rezultata iz spektara visoke rezolucije i kompozitnih spektara. QSOSIM10 tako modelira sljedeće karakteristike vidljive u spektru kvazara: skup od 59 širokih i uskih emisijskih linija kvazara, apsorpciju na međugalaktičkim oblacima u Lyman α šumi, zgrudnjavanje međugalaktičkog medija duž doglednice, inverzni efekt u blizini kvazara i apsorpciju na metalima u oblacima visoke gustoće. Većina ovih karakteristika nije bila prisutna u prijašnjim verzijama QSOSIM programa.

Skup simuliranih spektara koristiti će se kako bi se kvantificirali sistematski efekti prisutni u skupu pravih SDSS spektara, poput slabog uzorkovanja kvazara koji su bojama slični zvijezdama i promjena kriterija za odabir kandidata za spektroskopsko opažanje u prve dvije godine prikupljanja podataka. Svoju prvu primjenu simulirani spektri će dobiti u istraživanju anizotropija svemira na velikim skalama. Ovo istraživanje analizira spektre kvazara iz SDSS Objave podataka 12 u potrazi za odstupanjima optičke dubine svemira od očekivane vrijednosti, τ_{eff}, na svim crvenim pomacima i u svim dostupnim smjerovima. Preliminarni rezultati ovog istrazivanja prikazani su na slici 5.1 i predstavljaju najtočnija mjerenja prosječne optičke dubine svemira do sad. Slika 5.1 prikazuje nekoliko do sad nepoznatih značajki u evoluciji optičke dubine na crvenim pomacima $z > 4$. Ukoliko se utvrde točnima, rezultati će biti značajna kozmološka otkrića. Za utvrđivanje točnosti rezultata, potrebno je detaljno razmatranje korištenih metoda i postojećih sistematskih efekata u korištenim podacima. Svakom simuliranom spektru biti će određene fotometrijske magnitude u SDSS fotometrijskom sustavu, te će biti konstruiran simulirani boja-boja prostor korišten za odabir kandidata za spektroskopsko opažanje. Primjenom istih rezova na simulirani boja-boja prostor koji su primjenjeni na boja-boja prostor konstruiran iz pravih podataka, kvantificirati će se utjecaj sistematskih efekata u pravom uzorku na rezultate optičke dubine svemira.

Skup kvazara u konačnoj objavi podataka u sklopu SDSS-BOSS III projekta, Objavi podataka 12, broji više od 294 000 spektara raspoređenih na preko 10 000 kv. stupnjeva neba. Ovo je najveći dostupan skup spektara srednje rezolucije, prigodan za korištenje u širokom rasponu znanstvenih istraživanja. Skup simuliranih spektara može poslužiti za provjeru učinkovitosti metoda i pomoći u interpretaciji rezultata dobivenih korištenjem pravih spektara.

Usporedba rezultata dobivenih iz simuliranih spektara može biti od pomoći u istraživanjima ustrojstva svemira na velikim skalama, jednadžbe stanja međugalaktičkog medija, praiskonskih magnetskih polja, evoluciji kozmičkog metaliciteta, povijesti stvaranja zvijezda u svemiru, kao i istraživanjima kozmičke ultraljubičaste pozadine i mnogih drugih.

Povećanje rezolucije simuliranih spektara jedan je od prvih koraka u daljnjem razvoju programa QSOSIM10. Ciljana rezolucija simuliranih spektara trebala bi nadmašiti i rezoluciju teleskopa Keck i VLT, najboljih teleskopa današnjice. Ovime će se povećati svestranost primjene programa, omogućivši primjenu na istraživanjima koja koriste spektre kvazara iznimno visoke razlučivosti.

Konačno, skup simuliranih spektara kvazara dan je znanstvenoj zajednici na slobodno korištenje, u nadi da će pomoći iznjedriti dobru znanost.

Contents

1 Introduction **1**

1.1 Modern cosmology . . . 1

1.2 Sky surveys . . . 2

1.2.1 Sloan Digital Sky Survey – SDSS . . . 3

1.3 Simulations . . . 6

1.4 Quasars . . . 6

1.4.1 Historical overview . . . 6

1.4.2 Quasar spectral properties . . . 7

1.4.3 Quasar evolution through cosmic time . . . 9

1.4.4 The Lyman α forest . . . 10

1.5 Thesis motivation . . . 13

2 Simulating QSO spectra – QSOSIM9 **18**

2.1 QSOSIM9 overview . . . 18

2.2 Statistical properties of HI in the Lyman α forest . . . 20

2.3 Redshift distribution of HI absorption systems . . . 23

2.4 Column density distribution of HI absorption systems . . . 23

2.5 Doppler b-parameter . . . 24

2.6 Noise . . . 25

3 Simulating QSO spectra – QSOSIM10 **26**

3.1 Synchrotron mechanism . . . 29

3.2 QSO emission lines . . . 30

3.3 The Lyman α forest . . . 32

3.3.1 Column density distribution of HI absorption systems . . . 32

3.3.2 Redshift evolution of the HI absorption systems . . . 36

3.3.3 Clustering . . . 39

3.3.4 Doppler b-parameter . . . 42

3.3.5 Metal absorption . . . 45

3.3.6 The QSO proximity effect . . . 48

3.4 BOSS characteristics . . . 51

4 Results of QSOSIM10 **56**

5 Conclusions & future prospects **65**

5.1 Summary . . . 65

5.2 Applications & future prospects . . . 65

Appendices **69**

A Table of emission lines **69**

B An example of a CLOUDY input file **70**

1 Introduction

1.1 Modern cosmology

The word "Cosmology" comes from the Ancient Greek words κόσμος ("order" or "arrangement") and λόγος ("reasoning") [25], making it the study of the origin and order of the world and all the things around us – the Universe – as well as its properties. The questions about the nature of the Universe were important to virtually every human civilisation, with each one offering their own cosmological view of the world, often combining observations with philosophy and theology. It was not until the Scientific Revolution in Europe that the scientific method and scientific evidence became the foundations of modern cosmology. For its basis in physical laws it is also known as *physical cosmology*, and as such it endeavours to understand the origin, evolution and ultimate fate of the Universe.

Pioneering work in the field was that of Copernicus, Brache, Kepler, Newton, and many others who paved the way for experiment and theory as the basis for acquiring knowledge. More recently, Einstein's General Theory of relativity and his equations describing space-time changed our understanding of space and time themselves. On his work continued de Sitter, Friedmann, Robertson, Walker, and Lemaître, who produced important solutions to those equations.

The pinnacle of modern cosmology is the Big Bang Theory - a widely accepted scientific theory of the origin and evolution of the Universe, that states that the Universe underwent a highly energetic and extremely quick expansion (inflation) some 13.8 billion years ago [3]. In its inflationary phase, the Universe's volume increased by the factor of 10^{40} in just 10^{-33} s [26]. Evidence supporting the Theory are numerous, including the abundance of light elements and their relative ratios, the cosmic microwave background and its homogeneity (up to 10^{-5} K)[3], [26], large scale structure and baryonic acoustic oscillations (BAO), and the Hubble's law.

Woven into the Big Bang Theory are two important assumptions on the nature of the Universe: *homogeneity* and *isotropy*. The assumption of homogeneity implies that the Universe appears to be the same everywhere. This is valid observationally, given that one looks at a scale large enough ($\sim$ 100 Mpc) [26]. The assumption of isotropy implies that there are no preferred directions in the Universe, i.e. that the Universe appears to be the same in every direction, and that physical laws are the same in the whole Universe. Together, they are known as the *Cosmological principle*. Assumptions of homogeneity and isotropy have been tested and verified numerous times in the last two decades, by missions such as COBE [1] and WMAP [2], and most recently by the ESA's Planck mission [3]. Recently, however, deviations from

the Cosmological principle have been reported. Study of the peculiar velocities of local galaxies reported an unexpected correlation in the movement of galaxies on scales ~ 100 Mpc, a so-called *dark flow* [4]. Furthermore, reports on the space-time variations of the fine-structure constant (α) [5], [6], and the observed anisotropy in the luminosities of standard candles (supernovae type Ia) [27], give further indications about the violation of the Cosmological principle.

Although the methods to determine the validity of the Cosmological principle are numerous and the evidence to support it is not lacking, big sky surveys of the latest generation, with an unprecedented number of objects observed, provide us with a new way to search for inhomogeneity and anisotropy of the Universe.

1.2 Sky surveys

For centuries, telescope design was limited by the optical technology available, resulting with a narrow field of view and suited to observe only a small patch of sky. That limitation made systematic observations of large areas of the sky very tedious, and thus small but detailed surveys were preferred. With advances in optical technology and digital image processing in the last few decades, it became feasible to build wide angle telescopes with which large scale sky surveys were possible. Surveys in which a part of the sky was observed under the same (or similar) conditions, with no particular target, are called photometric surveys. The huge number of objects included in sky surveys made it possible to produce catalogues of various classes of astronomical objects. Also, large populations of different objects observed made statistical studies more accurate and, because of uniformity provided by using the same instrumentation, less biased.

Depending on the part of the electromagnetic spectrum in which the sky is observed, a survey can be categorised as optical (e.g. Sloan Digital Sky Survey - SDSS [28], Palomar Distant Solar System Survey [29]), infrared (e.g. The 2-micron All-Sky Survey - 2-MASS [30]), radio (e.g. NRAO VLA Sky Survey, PALFA Survey [31]) or gamma-ray (e.g. Fermi Gamma-ray Space Telescope [32]). If a survey is covering more than one wavelength range it is called a multi-wavelength survey (e.g. Cosmic Evolution Survey - COSMOS [33]). Besides its wavelength range, each survey can also be characterised by its sensitivity. Sensitivity, sometimes referred to as 'depth' or detection limit, of a survey is determined by technical characteristics of the observational instrument (i.e. telescope) used and the time dedicated to observations. It is usually expressed as the magnitude of faintest object observable or the minimal flux detectable. Surveys with high sensitivity can detect fainter objects than ones with low sensitivity. Because of high resource consumption needed to cover a large part of

the sky with great sensitivity, sky surveys are made with a 'wedding cake' approach - wide and shallow (large sky coverage at low depth) or narrow and deep (small sky coverage at high depth). Photometric data gathered by the survey is often used to select objects with interesting or desired properties, after which they are observed in a different wavelength range or observed spectroscopically. In SDSS, for instance, the photometric data is used to select stars, galaxies, and quasars whose spectra are later observed. Those spectra are collected and analysed to create catalogues, such as the SDSS DR10 QSO catalogue that contains data on over 160 000 quasars, which are open for public use [13]. Many telescopes and surveys, such as the Keck telescope, the Very Large Telescope, and the SDSS, allow open access to their databases. This eliminates the need for repeated observations of the same areas of sky or repeated spectroscopy of certain objects, while also subjecting scientific results to independent verification.

1.2.1 Sloan Digital Sky Survey – SDSS

The Sloan Digital Sky Survey, a photometric and spectroscopic survey covering approximately 35% of the sky is one of the largest sky surveys to date. The survey uses a dedicated 2.5 m wide-angle optical telescope at Apache Point Observatory (APO) in New Mexico [34], shown in Fig. 1.1. It was conceived in the 1980s with the primary objective of providing a better insight into the large-scale structure of the Universe. SDSS has been operational since 2000, and has gone through three phases: SDSS-I (2000-2005), SDSS-II (2005-2008), SDSS-III (2008-2014), and has started its fourth phase SDSS-IV in 2014. During the first three phases, SDSS has imaged approximately 14 555 square degrees of sky at high Galactic latitudes [8].

SDSS-III has four components: The Sloan Extension for Galactic Understanding and Exploration 2 (SEGUE-2) [35], the Baryon Oscillations Spectrographic Survey (BOSS) [36], the Multi-Object APO Radial-Velocity Exoplanet Large-area Survey (MARVELS) [37], and the Apache Point Observatory Galactic Evolution Experiment (APOGEE) [38]. SDSS BOSS is aimed at obtaining the spectra of galaxies and quasars with the primary goal of measuring the oscillation signature in the clustering of matter, in an effort to constrain cosmological models. It is focused on obtaining spectra of those galaxies and quasars whose Lyman α forest is within the BOSS spectral coverage, i.e. those with redshifts $z > 2.15$.

The SDSS telescope was built to obtain both photometric imaging and spectroscopic data, for which it has a wide-field mosaic 120 mega-pixel CCD camera, and two twin multi-object fiber spectrographs. A single twin spectrograph has two arms, each covering a different wavelength range (the 'red' and 'blue' spectrograph arms) [7]. The photometric imaging survey obtains multi-band photometry using the CCD

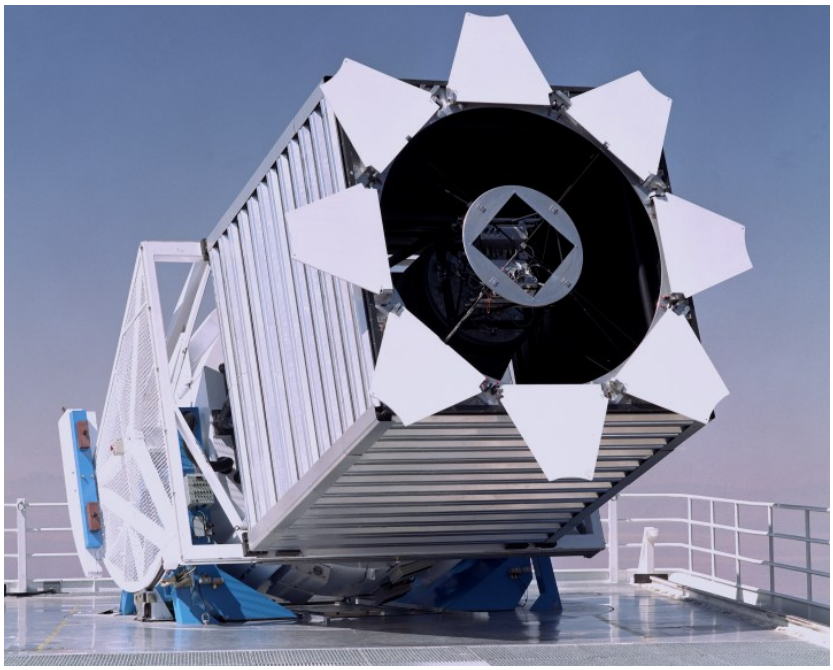

Figure 1.1: Sloan Digital Sky Survey's dedicated optical telescope at Apache Point Observatory in New Mexico. The telescope is equipped with 2.5 m wide-angle optical system, a 120 mega-pixel camera, and two high resolution BOSS spectrographs. Figure sourcedsourced from [39].

camera, while the two fiber-fed BOSS spectrographs are used to obtain the spectra of 1000 objects simultaneously (500 per spectrograph) [8], [14]. The SDSS photometric system comprises five colour bands (u, g, r, i, z) that divide the range from 3000 Å to 11000 Å into non-overlapping pass bands [14]. For object spectroscopy, each spectrograph gathers light through 500 fibers which terminate inside the slithead. Light gets collimated on a spherical mirror and then splits on a dichroic 45° beamsplitter. The beamsplitter reflects the blue portion of the bandpass ($\lambda < 6000$ Å) and transmits the red portion ($\lambda > 6000$ Å). Each of the spectrograph arms has a grism (a grating sandwiched between two prisms) immediately after the beamsplitter. Light diffused on the grisms then enters the cameras, where it's recorded on a CCD. The optical layout of the BOSS spectrograph is shown in Fig. 1.2, while the details of its construction and the spectrograph's properties can be found in [8].

The SDSS project regularly releases gathered data in a publicly accessible Data Release (DR), with DR12 being the latest [40]. Data Releases are cumulative, meaning that each new release contains all of the previous ones. SDSS DR12 contains imaging data of almost 470 million unique objects covering more than 10 000 sq. deg. of the sky, shown on Fig. 1.3. In DR12, BOSS contains spectra of 2.26 million unique objects, of which 1.37 million are galaxies and 294 000 are quasars, raising the total number of spectra in SDSS DR12 to 4.08 million, of which 2.40 million are galaxies and 477 000 are quasars [40]. The BOSS spectra cover the wavelength range between 3560 - 10400 Å[8].

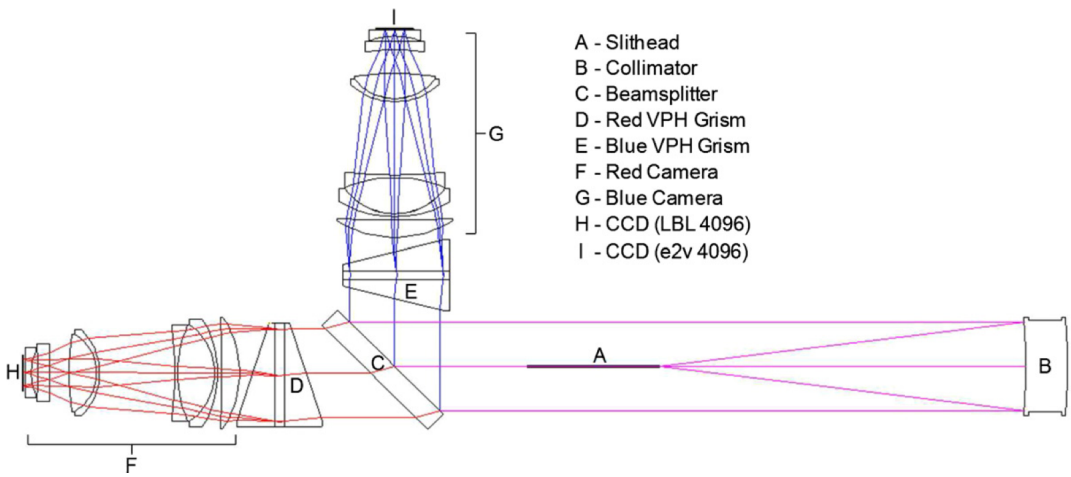

Figure 1.2: Optical layout of the BOSS spectrographs: light enters each of the spectrographs via 500 optical fibers that terminate in the slithead (A). Light gets collimated on the spherical collimating mirror (B), after which it gets split on the beamsplitter (C). The 'red' ($\lambda > 6000$ Å) portion of the light is transmitted, while the 'blue' ($\lambda < 6000$ Å) portion is reflected. After that, light gets diffused on the grisms (D, E) and enters the cameras (F, G), where it is recorded on a CCD (H, I). Figure sourced from [8].

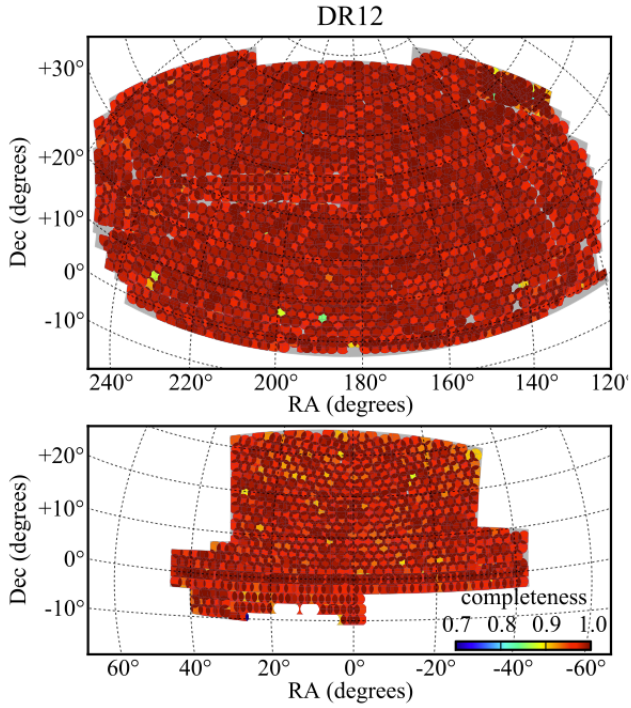

Figure 1.3: SDSS BOSS DR12 spectroscopic sky coverage in the Northern (top) and Southern (bottom) Galactic Caps. The area covered by the spectroscopic survey of DR12 exceeds 10 000 sq. deg. The colour coding indicates the completeness of the surveyed area. Figure sourced from [40].

1.3 Simulations

When dealing with large data sets, such as the SDSS, it is necessary to have a complete understanding of the data. We need to ensure that every step of the methodology is understood, so that the results are correctly interpreted and the systematic errors quantified. One way to do this is through the use of computer simulations. Computer simulations represent a cheap method of testing our understanding of the system's dynamics and properties – be it quantum interactions, material properties, galaxy mergers, or distribution of matter in the Universe. Simulations are computer programs that use step-by-step methods to explore system behaviour using mathematical models. Comparison of simulated, or *mock*, data sets to the data derived from measurement or experiment is used to validate methods and verify results. Mock data sets make systematic errors easier to account for, making results derived from real data more reliable.

One of the methods most widely used in simulations are the Monte Carlo methods; a set of powerful methods that use random numbers. They can be used to generate a particular realisation of a known probability distribution, calculate multidimensional integrals, and analyse behaviours of complex systems. These methods are especially useful when modelling a system in which many different realisations of the probability distribution are possible. We will use Monte Carlo methods to model the intergalactic medium seen in quasar spectra.

1.4 Quasars

1.4.1 Historical overview

The term *quasar* was coined in the 1960s from an acronym QSR, standing for 'Quasi-Stellar Radio source'. Quasars are one of the most luminous objects in the Universe, more than a 1000 times brighter than an average galaxy, and have been observed up to the highest measured redshifts [26]. In 1960s, two radio surveys of the northern sky - 3C and 3CR (at 158 MHz and 178 MHz, respectively) were completed [41], [42]. Low angular resolution of radio telescopes used made finding optical counterparts of radio sources difficult. However, in 1963 Matthews and Sandage [43] identified three radio sources from 3C (3C 48, 3C 196, and 3C 286) with their optical star-like counterparts. These three objects showed peculiar properties – their spectra did not resemble those characteristic of either stars or galaxies (i.e. Planck's black-body spectrum), and had a few very broad emission features. They were compact, point-like sources and therefore classified as 'Quasi-Stellar Objects' or QSOs. In the rest of this thesis, terms QSO and quasar are used interchangeably, with no

implied difference between the two.

Later that year, two astronomers studied another 3C radio source, 3C 273. Using the Parkes radio observatory, Hazard constrained the location of 3C 273 on the sky using the method of Lunar occultation [44]. Data provided by Hazard enabled Schmidt to observe the 3C 273's optical spectrum using the Palomar 200in telescope. In the 3C 273's optical spectrum, Schmidt identified the Balmer series of hydrogen at, for that time extremely high, redshift $z = 0.158$ [45]. Assuming that this redshift is entirely due to cosmological expansion and that Hubble's law is valid, this discovery implied that the source is the most distant object yet discovered, and about 1000 times brighter than normal galaxies. Furthermore, point-like nature of those objects in the optical meant that the emitting region is orders of magnitude smaller than a typical size of a galaxy. A galactic nucleus undergoing a very energetic process that makes the nucleus outshine the rest of the galaxy is called an *Active Galactic Nucleus* (AGN). Quasars are a type of AGN. Other types of AGNs include Seyfert galaxies and radio galaxies. Seyfert galaxies, named after their discoverer Carl Seyfert, have broad emission lines visible in their spectra, luminosities much lower than those of quasars, and are faint at radio wavelengths [26]. Radio galaxies, as their name suggests, have strong radio emission but resemble ordinary galaxies in the visible part of the spectrum. In some cases, radio galaxies have very energetic jets coming out of their centre that emit very strongly in the radio part of the spectrum. Jets are comprised of relativistic matter ejected by the black hole in opposite directions, perpendicular to the plane of the galaxy, and can span a distance that is 5 to 10 times the size of a parent galaxy [46]. Although thought of as separate classes of objects in the past, it is now believed that all AGN are a single population of objects observed at different angles. This is known as the AGN unification scheme [47, and the references therein]. According to the hypothesis, many features seen in the spectrum of the AGN are dependent of the angle at which we observe the AGN. The broad line emitting region of the AGN is invisible when the observer's line-of-sight forms a large angle with the axis of symmetry and vice versa. When the observer's line-of-sight is parallel to the symmetry axis of the AGN (i.e. the observer is 'looking down'), they are effectively looking down the AGN's radio jet. An illustration of the AGN unification scheme is shown in Fig. 1.4.

1.4.2 Quasar spectral properties

For such a luminous object to be seen from a great distance, a quasar must have a very powerful energy source. It is believed that almost every galaxy, with the exception of dwarf and irregular galaxies contains a supermassive black hole (SMBH) in its centre (e.g. [49]–[51]) which is active at certain periods of the galaxy's life [52], [53].

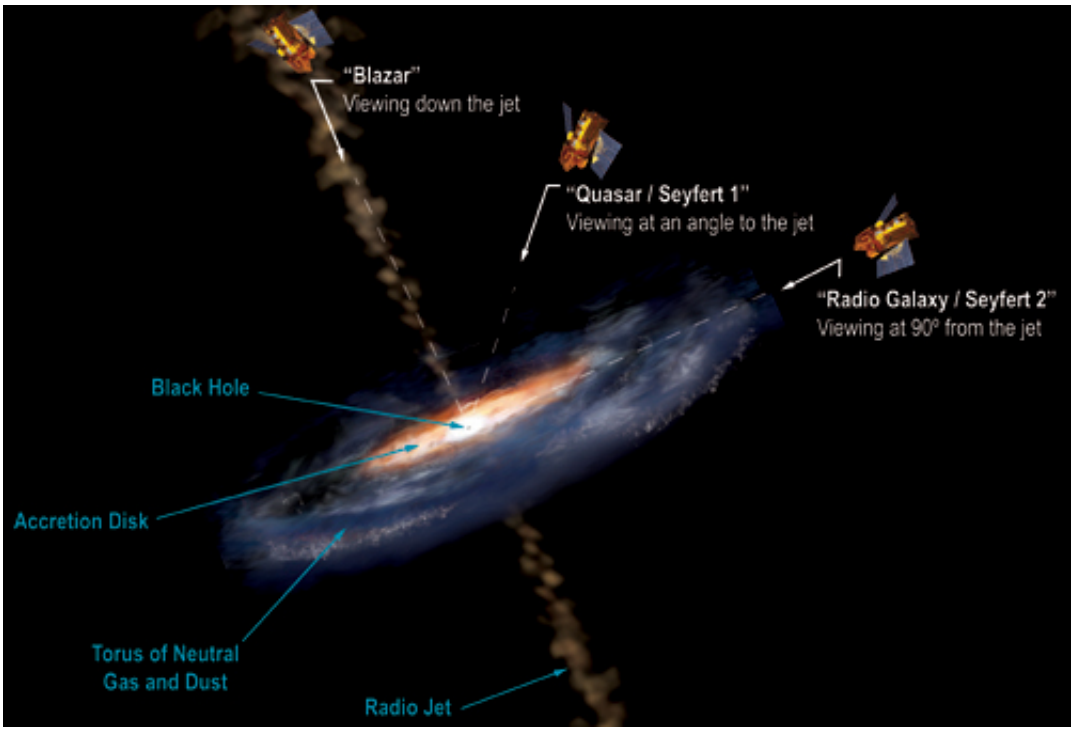

Figure 1.4: Under the unified model of the Active Galactic Nuclei, their observed spectral properties depend on the angle at which they are observed. Previous classifications such as Seyfert galaxies (type 1 & 2), Quasars, Blazars, and Radio Galaxies are a manifestation of this effect. Figure sourced from [48].

A galactic nucleus is active when the SMBH in its centre is actively accreting matter, a process which results in highly energetic emission of particles and electromagnetic radiation from the accretion disk [54]. Quasars emit electromagnetic radiation at all wavelengths, from radio to X-ray. Gravitational infall of matter towards the black hole (BH) and friction within the disk heat up the matter to temperatures as high as 10^5 K, corresponding to thermal emission in X-ray. Numerous non-thermal processes in the hot accretion disk influence the shape of the quasar's spectrum, but over a broad frequency range, its continuum spectrum can often be described with a power law:

$$F_\nu \propto \nu^{-\alpha} \tag{1.1}$$

where F_ν is the flux at frequency ν and α is the spectral index. The spectral index of quasar continuum α differs for various ranges of the electromagnetic spectrum, with mean values of $\alpha = 0.5$ in the radio [55], $\alpha = 0.44$ in the optical [56], and $\alpha = 0.7$ in the X-ray [26]. An example of a QSO spectrum at optical wavelengths is shown on Fig. 1.5.

This spectral form indicates that the bulk of the electromagnetic emission is produced by *synchrotron radiation* of relativistic electrons [26]. Synchrotron radiation is the result of constant acceleration of charged particles, such as electrons, by ex-

tremely strong magnetic fields in the vicinity of the black hole. Quasar luminosities fall in the range from 10^{38} to 10^{42} W. For comparison, large galaxies such as the Milky Way have a luminosity of the order of 10^{37} W, or 10^{10} Solar luminosities [26].

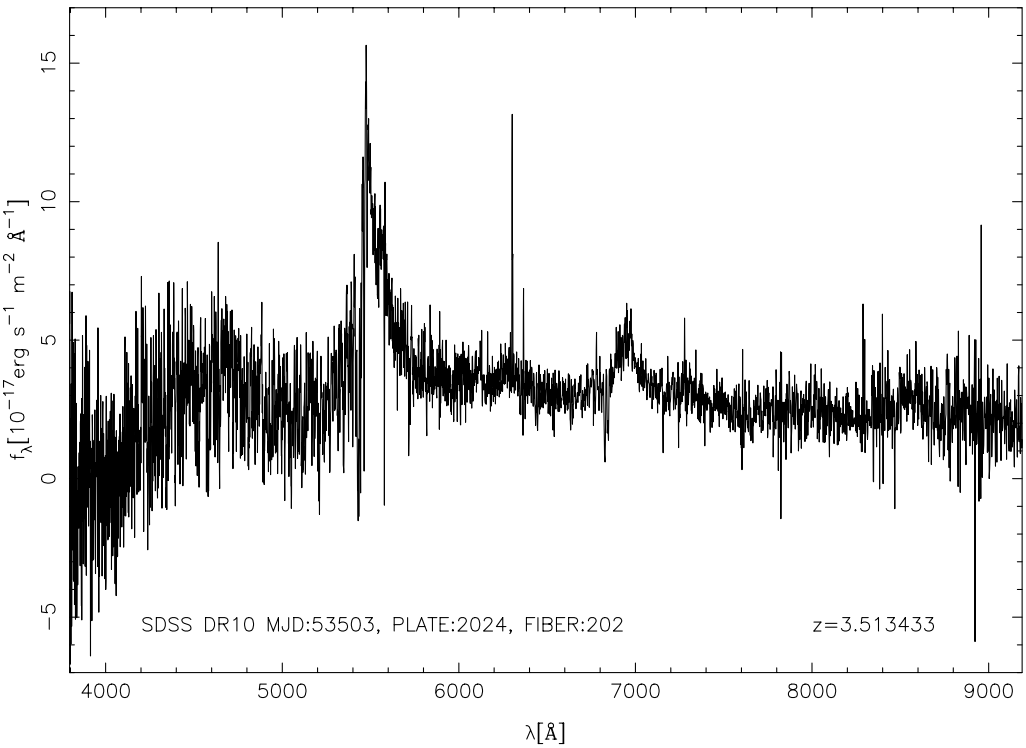

Figure 1.5: A typical example of the QSO spectrum from SDSS DR10 at redshift $z = 3.513433$. The Lyman α forest is visible at wavelengths $\lambda \lesssim 5500$ Å.

1.4.3 Quasar evolution through cosmic time

The fraction of galaxies that have an AGN is not constant in time. While quasars are almost non-existent in today's Universe, they were a common feature in the Universe some 10-12 billion years ago. The nearest (and also the first identified [44], [45]) quasar, 3C 273, is located more than 2 billion light years away. Because there are no other low redshift quasars, it follows that there are no other nearby quasars. However, the number of quasars rises sharply with lookback time and peaks at some 1.3 billion years after the Big Bang (i.e. $z \approx 2.4$), only to drop once again at earlier epochs. The evolution of the number of quasars in the Universe with cosmic time can be seen on Fig. 1.6 using a sample of the SDSS DR10 quasars.

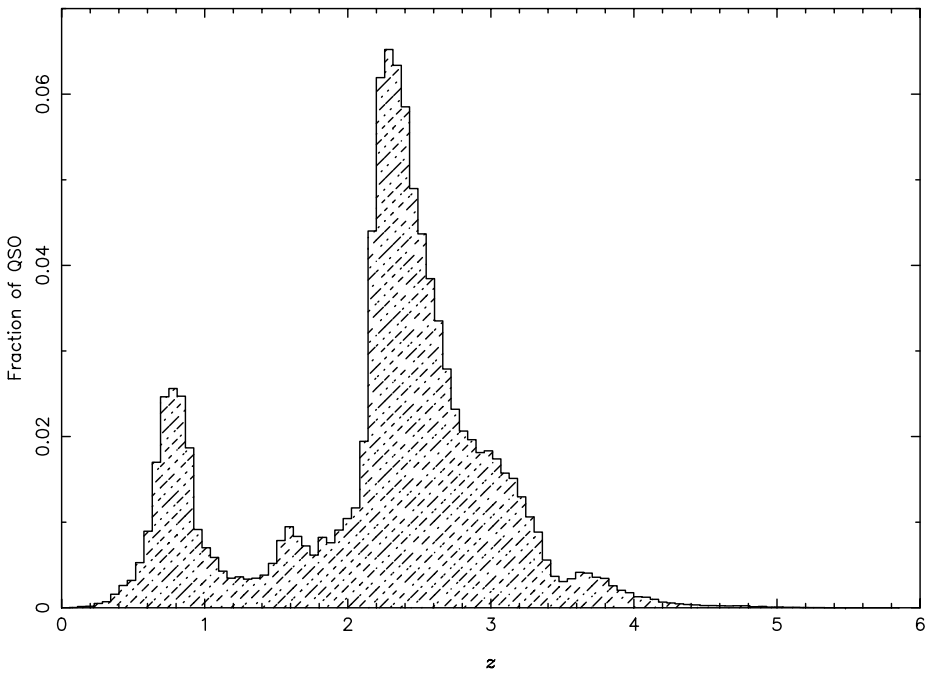

Figure 1.6: Normalised distribution of QSOs observed by SDSS as a function of redshift. Majority of the detected QSOs are located between redshifts $z = 2$ and $z = 3.5$, indicating at an epoch dominated by QSO activity.

1.4.4 The Lyman α forest

The Lyman α forest is a term used to describe a series of absorption lines seen in QSO spectra located bluewards (i.e. at shorter wavelengths – higher frequencies) of the QSO's Ly α emission line of neutral hydrogen. The phenomena was first discovered in 1971 by Lynds [57]. On its way towards us, light from a distant object passes through clouds of intergalactic gas and clouds associated with foreground galaxies – and is absorbed and re-emitted by them. Absorption by the gas clouds leaves an imprint on the spectrum that is unique to the absorbing cloud – providing us with information on the cloud's position and dominant physical processes, as well as its chemical composition and atomic abundances. Because of this property, statistical studies of the Lyman α forest give us a unique and direct insight into the intergalactic medium (IGM) and its evolution through cosmic time, as well as into the small- and large-scale structure of the Universe. An example of absorption seen in the Lyman α forest seen in QSO spectrum is shown on Fig. 1.7.

As the name says, a vast majority of absorption lines in the Lyman α forest comes from the Ly α transition of neutral hydrogen at wavelength $\lambda = 1215.67$ Å (in the absorbing cloud's rest frame), with a smaller number of absorption lines associated

with other atomic species e.g. transitions from metal[5] species. The Lyman α forest was first observed in the optical wavelength band (4000-9000 Å), implying that the clouds are highly redshifted by cosmic expansion.

The IGM clouds have simple observational signatures in the spectra. With only the intrinsic atomic line width and motion of the atoms contributing to line broadening, the line profile is a convolution of a Lorentzian resonance curve and a Maxwell-Boltzmann velocity distribution. The result is a so-called Voigt profile, which is fully specified by three parameters: its position in velocity space (redshift), column density, and line width (Doppler b-parameter). Column density is defined as the number of atoms of a substance per unit area, integrated along a path. It is directly related to the matter density of the cloud, and is expressed in units of cm^{-2}. Line width, or b-parameter, is a measure of the total velocity distribution in the cloud. Thermal motion, turbulence, bulk flows, and Hubble expansion (in extended objects) all contribute, in varying relative amounts, to absorption line broadening.

The IGM gas clouds were at first thought of as discrete patches of intergalactic gas held in hydrostatic, thermal and ionisation equilibrium by the pressure from hot intergalactic medium that prevented it from dissipating. However, that hypothesis could not explain some of the observational evidence, such as the large range of column densities observed and the evolution of the number of absorption systems with redshift. Finally, the COBE mission [1] provided definitive arguments against the existence of the proposed cloud confining mechanism and in favour of the now overwhelmingly accepted *dark energy – cold dark matter* cosmology (ΛCDM). According to the ΛCDM cosmological model, large quantities of warm intergalactic baryonic gas are trapped by the gravitational collapse of dark matter halos, but the thermal pressure of the warm gas at low column densities prevents star formation and galaxy formation. Ly α absorption systems visible in the IGM are divided into three classes in accordance with their H I column density and properties:

1. *Lyman α forest* – clouds with H I column densities between $N \sim 10^{12}$ cm^{-2} and $N \sim 10^{17.2}$ cm^{-2} that make a majority of observed absorption features along the line of sight to the QSO.

2. *Lyman Limit System (LLS)* – clouds with H I column densities larger than $N \sim 10^{17.2}$ cm^{-2} and smaller than $N \sim 10^{19}$ cm^{-2}. They start to exhibit a Lyman series break (at the Lyman limit, 912 Å restframe) and are dense enough to begin shielding themselves against ionising radiation from the outside.

[5] In astrophysics, all elements heavier than helium are referred to as metals.

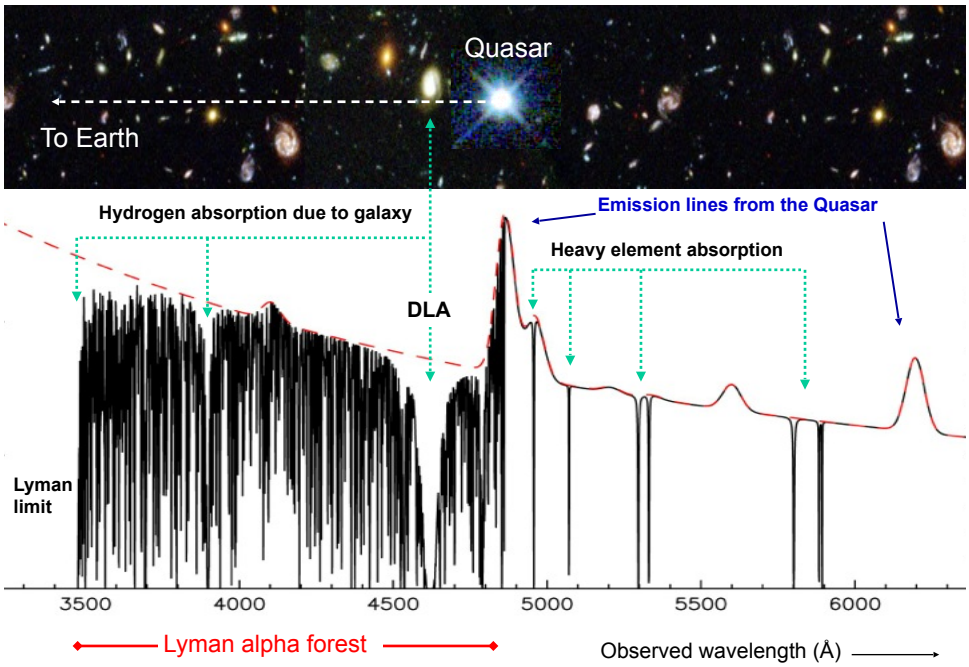

Figure 1.7: A visual representation of absorption of the quasar's light on the intervening gas clouds and the resulting Lyman α forest. Light from the quasar is absorbed and reemitted by clouds of varying column densities and at different redshifts, leaving a trace of the intergalactic medium along the sightline towards the quasar. Light can also be obscured by high density gas clouds associated with nearby foreground galaxies, a feature called Damped Lyman α (DLA), as seen on the figure. Modified from [58].

3. *Damped Lyman α (DLA)* – clouds with H I column densities larger than $N \sim 10^{19}$ cm^{-2}. The clouds are so dense that they almost completely shield themselves from ionising radiation and exhibit dampening wings.

A smaller number of features in the forest come from metal absorption in the IGM, and will be discussed further in Sec. 3.3.5.

The Lyman α forest is an important tracer of both large-scale and small-scale structure [59]. The nature of the present-day Universe, ~13.8 billion years old, can be studied via direct detections of local galaxies and their environments, within the sensitivity limits of available instruments (a few hundred Mpc). Less than a decade before the discovery of the Lyman α forest, the Cosmic Microwave Background (CMB) was discovered by Penzias and Wilson in 1964 [60]. The CMB gave astrophysicists an insight into the early Universe, only ~380 000 years old at the time of radiation decoupling. The discovery of the Lyman α forest provided a way to fill in the gap between the local and distant observations. The ΛCDM cosmological model predicts that large-scale structure in the Universe formed by constant clustering of smaller structures into larger ones. Clustering results in a continuous distribution of filamentary structures known as the *cosmic web*. According to the model, galaxies form in huge haloes of dark matter in which clouds of baryonic matter are trapped. Clustering of baryonic matter in such haloes results in clouds of high matter density that eventually collapse to form stars and galaxies. This is known as the *bottom-up*

process of structure formation. The Lyman α forest is a direct tracer of this process, as it traces the evolution of the IGM, and can be used to study the process of galaxy formation [61]. Because the baryonic clouds are closely correlated with dark matter haloes, it can also be used to study dark matter distribution and constrain cosmological models [62], [63]. Furthermore, the distribution of baryonic matter over the largest scales ($\sim$300 Mpc) can be used to look for baryonic acoustic oscillations and cosmological anisotropies [64]–[66].

1.5 Thesis motivation

Big sky surveys are exquisite tools to study numerous topics in observational cosmology. Statistical studies of galaxy growth, evolution, and clustering are made possible by the huge sample of galaxies (both AGN and non-AGN) surveyed (e.g. COSMOS [33], Two-Degree Field Galaxy Redshift Survey - 2dFGRS [67]). In particular, the SDSS is used to constrain cosmological models [68]–[70], put limits on the neutron masses [71], [72], investigate the baryonic acoustic oscillations that seeded the matter distribution of the Universe [64]–[66], and the evolution of the distribution of non-galactic baryonic matter in the Universe [73], [74].

The purpose of this thesis is to create a catalogue of realistic QSO mock spectra, using data provided by SDSS and Monte Carlo methods to simulate the IGM and intervening gaseous halos intersecting quasar sightlines. These mock spectra will be used as a part of a larger project aiming to test the Cosmological principle using the SDSS QSO data. The project will study the opacity of neutral hydrogen as a function of redshift and position on the sky, in order to examine recent reports on deviations from isotropy [4]–[6], [27] (and thus from the ΛCDM model). This requires a spectral analysis of the Lyman α forest of a large sample of QSO spectra. The latest SDSS data release, DR12 [40], contains medium-resolution spectra of more than 294 000 QSO between redshifts $z \approx 0.5$ and $z \approx 5.8$, covering an area larger than a third of the sky – making it ideal for the study of the IGM seen in the Lyman α forest. This thesis will not discuss the actual spectral analysis performed to study the opacity of neutral hydrogen, but will focus on the necessity and production of mock data.

There are three main issues that can introduce a systematic bias into the results of such a study, the effects of which can be quantified using mock spectra:

1. The systematic effects of the SDSS QSO target selection (see [9], [11]) on the scientific results derived from the sample are still not completely understood. The SDSS uses a colour-colour space derived from the photometric survey to select QSO candidates for the spectroscopic follow-up [9]. The SDSS selection algorithm uses the spectroscopy of known stars to construct stellar loci; regions

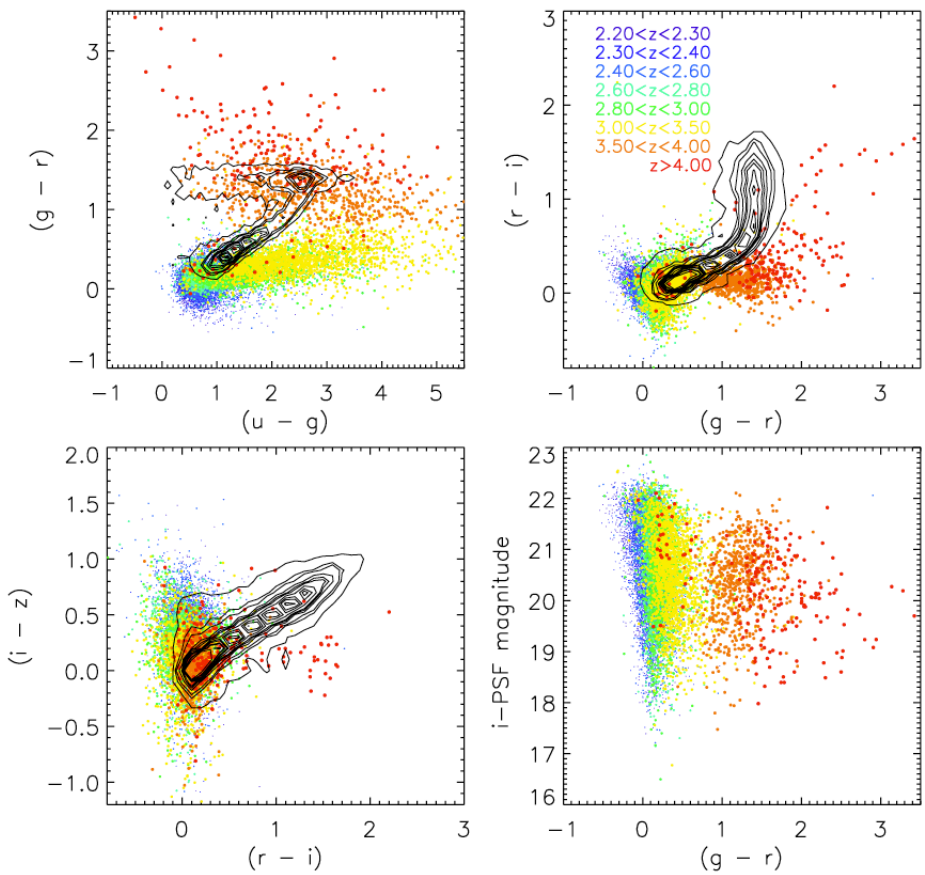

Figure 1.8: Colour-colour diagrams for the First Year data of the SDSS DR9 [75] . Spectroscopically confirmed quasars with redshifts above $z = 2.2$ are shown as coloured dots. The stellar locus is shown as black contours. The areas of the colour-colour space with high stellar contamination are poorly sampled by the SDSS quasar selection algorithm, the effects of which are still unquantified. The lower right panel shows the i magnitude as a function of the $g - r$ colour. Figure sourced from [11].

of the colour-colour space assumed to contain stars. The QSO candidates are selected via their non-stellar colours in the colour-colour space, by selecting objects which are inconsistent with being stars above a 4σ threshold [9]. The QSO candidates are selected from a region of the colour-colour diagram known to be occupied by QSOs, but areas with high stellar contamination exist at redshifts $2.7 < z < 3.5$ [10]. The positions of QSOs in this colour-colour space vary with redshift, as spectral features (e.g. the Lyman α forest, broad emission lines) redshift from one of the SDSS photometric bands into the other. The best QSO candidates to study the Lyman α forest are the most difficult ones to target, as their positions are the closest to the stellar loci [11]. This is illustrated in Fig. 1.8, which shows the positions of the QSOs and the stellar locus in the colour-colour space. The contaminated areas are mostly avoided when selecting QSO candidates and are thus poorly sampled by the quasar target selection algorithm. The poor sampling of these areas can introduce a systematic

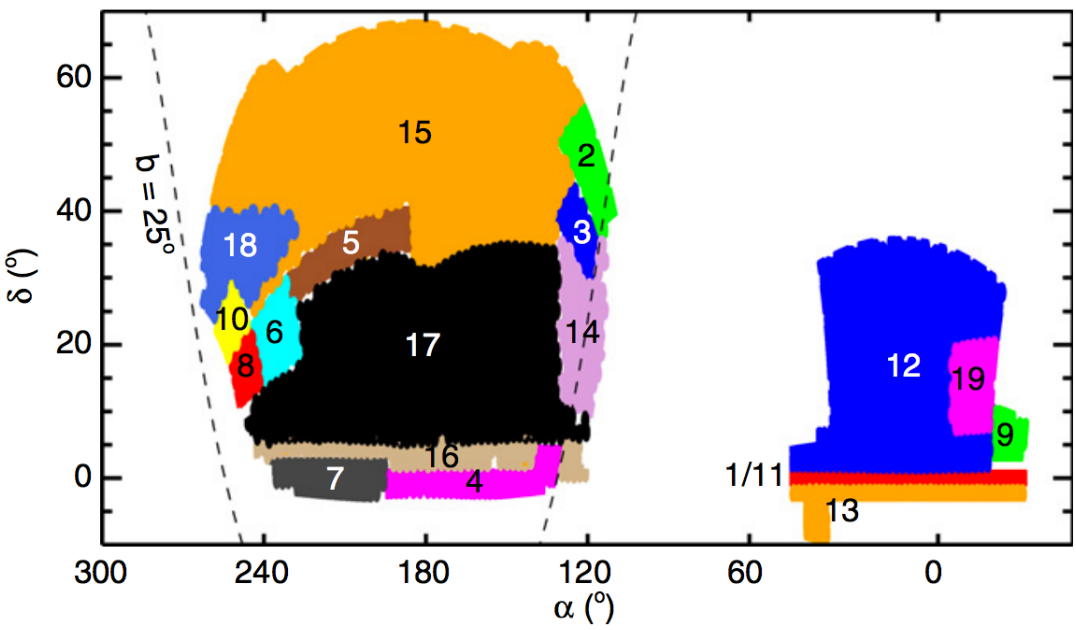

Figure 1.9: Targeting footprint for the first two years of the SDSS BOSS Lyman α forest and QSO survey [11]. The survey was conducted in 19 areas, each with uniform quasar targeting selection criteria. The targeting criteria changed during the first two years of BOSS survey in an effort to maximise the surface density of QSOs. Each of the 19 areas is shown in different colour in the plot. Simulated spectra presented in this thesis will be used to quantify the effect this has on the scientific results derived from the SDSS spectra. Dashed line is at Galactic latitude $b = 25°$. Figure sourced from [11].

effect, as it avoids a particular population of QSOs. The SDSS Collaboration performed simulations of the QSO spectra to quantify the effectiveness of the SDSS QSO target selection algorithm [10], [76]. They report a significant drop in the completeness of the sample in the redshift range $2.35 < z < 3.10$, where the algorithm selects only a few percent of the QSO known to be present [10].

During the commissioning of the SDSS BOSS in 2012, QSO target selection criteria were often redefined. The changes in the target selection were motivated by the effort to improve the completeness of the survey, but complicated by the stellar contamination and the photometric errors of objects near the detection limit [11]. The SDSS BOSS QSO target selection algorithms were improved over the course of the first two years of the survey, resulting in 19 areas with different target selection criteria (see Fig. 1.9). The effect of the variation of the selection criteria with respect to the position on the sky has not yet been quantified. Simulations present the best way to assess the systematics of the two aforementioned effects. The simulated spectra presented here will be the largest and most comprehensive set of simulated QSO spectra, with characteristics of real data obtained by SDSS. The spectra will be used to increase our understanding of the systematic effects introduced by the under-sampling and

varying QSO target selection criteria. The spectra will be used to generate the colour-colour space used by the SDSS to select QSO candidates, in a way similar to the simulations performed by [10], [76]. By making the same cuts on the points in this mock colour-colour space as with the real data, we can quantify the effect of under-sampling and the varying target selection criteria.

2. Correct flux estimation is required by all studies that employ the Lyman α forest in their methods. Examples of this are the studies of the neutrino masses [72], the baryonic acoustic oscillations [77], the epoch of H I and He II reionisation [78]–[80], the thermal properties of the IGM [81], [82], the chemical enrichment and evolution of galaxies and the IGM [17], [74], [83]–[85], and our study of the opacity of neutral hydrogen, among others. All of these studies use datasets that are incomplete (i.e. do not cover the whole sky and all redshifts), contain instrumental limitations (finite instrumental resolution, finite signal-to-noise), and contain systematics (e.g. uncertainties in the continuum estimation). A precise quantification of systematics for one of the biggest databases, i.e. SDSS, creates new possible solutions to these problems and renders the SDSS database more useful in all these contexts.

3. Another uncertainty comes from the fact that the Universe is not homogeneous on small scales. All measurements of cosmological observables are influenced by the local variations in the distribution of matter in the Universe. The value of any cosmological observable X in a particular region of the Universe is different from the mean value, $\widehat{X}$. However, averaged over measurements from many different regions, it is expected that:

$$\langle \widehat{X} - X \rangle = 0, \tag{1.2}$$

and

$$\left\langle (\widehat{X} - X)^2 \right\rangle \neq 0. \tag{1.3}$$

The observables can be the mean number of galaxies in a particular region of the Universe, the matter density, mean opacity, and many others. The restriction on what we can conclusively claim about the properties of the Universe on the largest scales, imposed by using only local data, is known as the *cosmic variance* [26]. Using the extremely large sample of QSO in the SDSS BOSS will minimise the cosmic variance of the opacity of the Universe, as it provides a statistically significant sample covering a large range of redshifts. The opacity cosmic variance is caused by a combination of things, including: the shape of the H I column density distribution, the b-parameter distribution, the clustering

and evolution of the number density of the clouds comprising the intergalactic medium, as well as the spectral resolution of the telescope and uneven sky coverage. Even with its unprecedented sky coverage, the SDSS covers only slightly more than a third of the sky. The SDSS covers a large area of the Northern Galactic Cap, and a smaller area of the Southern Galactic Cap (see Fig. 1.3). Using this newly-developed software, we can create artificial SDSS skies that cover the same, greater, or different areas. This will help us to constrain the opacity cosmic variance as a function of position of the QSO on the sky. Comparing the results from artificial skies to the real one will enable us to study how different realisations of the sky influence the results of our study.

This thesis will present a unique and comprehensive database of simulated QSO spectra with SDSS BOSS characteristics. The simulated spectra will include physical phenomena seen in real QSO spectra, such as the QSO proximity effect [86], clustering of absorption systems [18] (both of H I and heavy elements), metal absorption associated with high H I column density systems, and a large set of quasar emission lines [56]. Chapter 2 presents an overview of the pre-existing spectral generation code, QSOSIM9. Chapter 3 presents the software developed here, QSOSIM10, and the improvements introduced. Chapter 4 discusses the results, compares the mock spectra to real spectra, and discusses the improvements planned in the future.

2 Simulating QSO spectra – QSOSIM9

2.1 QSOSIM9 *overview*

This thesis is concerned with accurately producing simulated SDSS QSO spectra and improves upon the previous work done by John K. Webb from University of New South Wales (UNSW) in Australia. Webb has written a FORTRAN code, named 'QSOSIM9' [12], that generates an artificial QSO spectrum of arbitrary resolution for a given set of input parameters. This section briefly overviews the simulation process of QSOSIM9, with details on the more important steps given in subsequent subsections. It also presents the mathematical models and assumptions used in the generation of the Lyman α forest. The next chapter will describe improvements implemented to create QSOSIM10.

QSOSIM9's input parameters are the wavelength coverage of the spectrum, the QSO's emission redshift z_{QSO} , spectral index α, and the V magnitude in the Johnson-Cousins' magnitude system, along with other user-defined parameters relevant for modelling the spectrum.

QSOSIM9 reproduces most important features of the QSO spectrum – underlying continuum from synchrotron radiation and the broad emission lines of hydrogen, nitrogen, carbon, oxygen, sulphur, and many others. The code also reproduces the Lyman α forest in the spectrum, assuming that it can be represented well by a collection of individual absorption lines. The steps of the program are described below.

1. In the first step, the code reads in a file with the user provided input parameters, which are listed in QSOSIM9's documentation [12]. It uses the spectral index and magnitude provided in the file to plot the underlying QSO continuum. Over the continuum background, the code plots broad emission lines of Ly α, Ly β, N v, C II, C IV, O IV, O VI], Si IV, and many other atomic species (30 in total). Relative strengths of the emission lines were determined by visual inspection of a number of quasar spectra and are hard coded into the program. Emission lines have a Gaussian profile, with the FWHM $= 60$ m s^{-1} and the mean redshifted wavelength (λ) corresponding to $\lambda = \lambda_0(1+z_{QSO})$, where λ_0 is the restframe emission wavelength.

2. In the next step, QSOSIM9 simulates the absorption lines in the Lyman α forest that arise from absorption by the IGM, described in Sec. 1.4.4. The program deals with neutral hydrogen absorption only. Assuming the forest can be represented well by a collection of individual absorption lines, the program calculates the number of absorption features likely to be seen in the spectrum. It then uses

simple statistical models, described in subsections below, to create an artificial line list. Absorption lines are characterised as Voigt profiles. The Voigt profile of every absorption line is fully defined by its position (in redshift space), absorption strength (column density) and line width (Doppler b-parameter). To assign redshifts z and column densities $N_{\rm HI}$ to lines, QSOSIM9 uses simple models to generate cumulative distribution functions (CDFs) of these two parameters. Values are then determined by randomly sampling the two CDFs. The b-parameter distribution is modelled as a normal distribution with a mean $\mu = 23$ km s^{-1} and $\sigma = 3$ km s^{-1}. The value of the b-parameter is determined by using a FORTRAN numerical routine GASDEV taken from [87]. This routine returns a double-precision number that is normally distributed with a mean $\mu = 0$ and deviation $\sigma = 1$. Section 2.2 explains how the total number of lines in a certain redshift range is determined, while the algorithms that generate the line list parameters z, $N_{\rm HI}$, and b are described in Sections 2.3, 2.4, and 2.5, respectively. Voigt profile parameters are assigned independently for each line, which are input into the spectrum one-by-one.

3. The third step allows for the input of up to 20 Lyman limit systems (LLS) in the forest. User defined systems are input by defining their redshift, column density, and b-parameter values in the input file. If the input file contains details on any LLS, QSOSIM9 will automatically include it after it finishes the automatic generation of H I absorption in the forest.

4. In the fourth step, the code includes the limitations imposed by the telescope technology. In modern optical telescopes, complex optical systems are used to decompose the light of the object into a spectrum inside of the spectrograph. The measure of the spectrograph's ability to resolve features of width $\Delta\lambda$ at wavelength λ in the electromagnetic spectrum is called the *spectral resolution* or *resolving power* – R, defined as

$$R = \frac{\lambda}{\Delta\lambda}. \tag{2.1}$$

In real spectrographs $\Delta\lambda$ is a function of wavelength. For SDSS BOSS spectrographs, $\Delta\lambda$ is equal to the FWHM of a Gaussian that is a best-fit solution to the observed spectral profiles at various wavelengths [8]. Spectral resolution then becomes:

$$R = \frac{\lambda}{2.35 \times \sigma_\lambda}, \tag{2.2}$$

where σ_λ is the width of the best-fit Gaussian to the one-dimensional point-spread function at the wavelength λ. The BOSS spectrographs have a resolving

power between $R = 1000$ and $R = 2500$. In high-resolution spectrographs such as the High Resolution Echelle Spectrometer (HIRES) on the Keck Telescope in Hawaii, resolving power varies between $R = 25000$ and $R = 85000$ (depending on the configuration), while the resolving power of the Ultraviolet and Visual Echelle Spectrometer on the European Southern Observatory's Very Large Telescope (VLT/UVES) can reach values upward of $R = 100000$. QSOSIM9 neglects the wavelength dependance of the resolving power and uses a single value of σ_λ to convolve the the spectrum. Resolution of QSOSIM9's spectra is arbitrary and is provided by the user in the input file.

5. In the fifth step, noise is introduced into the spectrum. Even with great leaps in telescope technology, it is impossible to obtain a perfect spectrum of an object. Thermal excitations of electrons in the equipment, atmospheric distortions of the photon path, cosmic rays, and light contamination from other nearby objects (or the seemingly empty sky) introduce an additional component to the signal that is registered by the CCDs. All contributions to the signal that do not come from the targeted object are called *noise*. QSOSIM9 provides two noise models from which a user can choose, described in Sec. 2.6.

6. The spectrum generated by QSOSIM9 is writen in an ASCII file named *spec.dat* in the final step. The output file contains four columns: wavelength, flux, the error on flux, and the values of flux with no noise introduced.

The output of QSOSIM9 is shown in Fig. 2.1, using an example of a QSO at redshift $z_{QSO} = 3.3$, with the spectral index $\alpha = -0.7$, and the magnitude $V = 16$.

2.2 Statistical properties of HI in the Lyman α forest

To correctly model the Lyman α forest, we need to have a mathematical description of the distribution of the H I absorbers. This section gives an overview of the models used by QSOSIM9, and the way the program uses them to generate the artificial list of H I absorption lines.

The number of absorption lines that are seen in the QSO spectrum, n, strongly depends on the wavelength coverage of the spectrum and its resolution. More low column density clouds are seen in high-resolution spectra because they are easier to resolve. Similarly, more clouds of all column densities are seen with a larger spectral coverage.

Let us define $\mathrm{d}n/\mathrm{d}z$ as the number of H I absorption lines that are seen between redshifts z and $z + \mathrm{d}z$ in the spectrum. This number will tell us how likely it is for our

line-of-sight towards the QSO to intersect a cloud at a given redshift z. Observationally, this is well described by a power law of the form

$$\frac{\mathrm{d}n}{\mathrm{d}z} = A(1+z)^{\gamma}, \tag{2.3}$$

where A is a local comoving number density, measured for a specific cloud population (e.g. H I absorbers with $N_{\mathrm{HI}} \geqslant 10^{15}$ cm^{-2}, between redshifts $2.0 < z < 3.0$) and γ is the evolution parameter of the cloud population. A and γ depend on the population of clouds studied, i.e. they are functions of both column density and redshift. Values of $\gamma = 1.85$ for $z > 2.5$ and $\gamma = 0.16$ for $z < 2.5$ are cited by Penton et al. [88]. QSOSIM9 uses the values $A = 10$ and $\gamma = 2.0$ for clouds at all column densities and redshifts. Integrating Eq. 2.3 gives the total number of lines detected in redshift ranges z_1 to z_2

$$n = \frac{A}{\gamma+1}\left[\left(1+z_2\right)^{\gamma+1} - \left(1+z_1\right)^{\gamma+1}\right]. \tag{2.4}$$

The lower limit, z_1, is indirectly defined by the user when specifying the wavelength range of the artificial spectrum, while the upper limit is the quasar's redshift, $z_2 = z_{QSO}$.

The value of z_1 is defined as the lowest redshift at which a Ly α absorption system could be observed in the generated spectrum:

$$z_1 = \frac{\lambda_{\mathrm{low}}}{\lambda_{\mathrm{Ly}\ \alpha}} - 1, \tag{2.5}$$

where λ_{low} is the lowest wavelength observable in the spectrum, as specified by the user in the input file, and $\lambda_{\mathrm{Ly}\ \alpha} = 1215.67$ Å is the wavelength of the Ly α transition.

In a similar way, if we define $\mathrm{d}n/\mathrm{dN_{HI}}$ as the number of absorption lines with H I column density between N_{HI} and $N_{\mathrm{HI}}+\mathrm{dN_{HI}}$, this number tells us how likely it is for our line-of-sight to the background QSO to intersect a cloud with given neutral hydrogen column density N_{HI}. The model used by QSOSIM9 assumes that the distribution of H I column densities is well described by a single power-law for column densities in the range from $N_{\mathrm{HI}} = 10^{13}$ cm^{-2} to $N_{\mathrm{HI}} = 10^{17}$ cm^{-2}:

$$\frac{\mathrm{d}n}{\mathrm{d}N_{\mathrm{HI}}} = {N_{\mathrm{HI}}}^{-\beta}. \tag{2.6}$$

The exact shape of this distribution has been a topic of investigation for many years [15], [17], [86], [88]–[90]. Limitations in telescope design (resolution, spectral coverage) and effects such as line blending (a few low column density systems that are close together can be misidentified as one system with high column density) have led to ambiguous, and sometimes conflicting results.

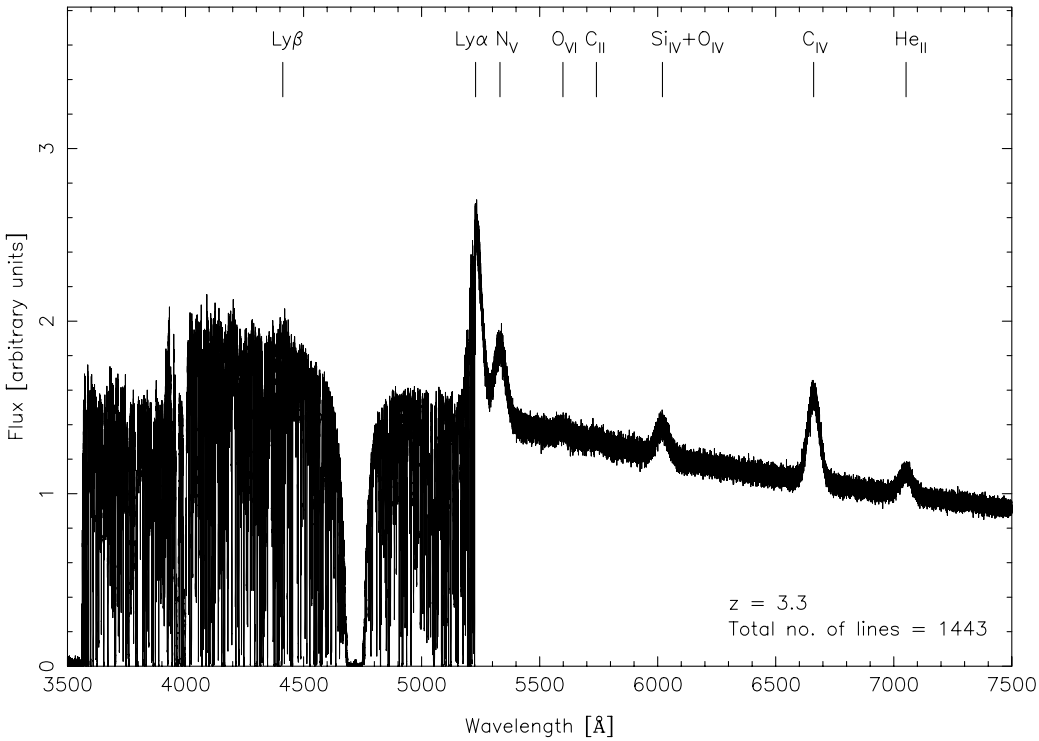

Figure 2.1: An example output of QSOSIM9 for a quasar with redshift $z_{\rm QSO} = 3.3$, $\alpha = 0.7$ and a magnitude $V = 16$. The spectrum has a resolution of $R \approx$100 000, and a SNR=50. QSOSIM9 calculates the total number of lines from redshift $z \approx 1.88$ to $z_{\rm QSO}$ and randomly selects their redshifts, H I column densities and b-parameters, assuming simple evolution models. If the randomly selected column density exceeds a predetermined value, the code inputs a Lyman limit system (such as the one at $\lambda \approx 4750$ Å).

Penton et al. [88] suggest $\beta = 1.65 \pm 0.007$ for $10^{12.5} < N_{\rm HI} < 10^{14.5}$ cm^{-2}, but a flatter slope for $N_{\rm HI} > 10^{14.5}$ cm^{-2}. The value used by QSOSIM9 is $\beta = 1.7$. It should be noted that the shape of the column density distribution in model does not fit the observations recently obtained from high-resolution QSO spectra, and assumes that the shape of the column density distribution does not change with redshift. A new model that fits the new data much better is used in QSOSIM10, and will be discussed in Sec. 3.3.1.

After calculating the number of lines using Eq. 2.4, QSOSIM9 will assign values of column density, redshift, and b-parameter to each line, creating an artificial line list. The algorithms assigning these values are described below.

2.3 Redshift distribution of HI absorption systems

The idea behind the redshift assigning algorithm of QSOSIM9 is as follows. The cumulative distribution function of a line's redshift z is given by

$$\mathrm{CDF}(z) = \frac{\int_{z_1}^{z} \frac{\mathrm{d}n}{\mathrm{d}z'}\,\mathrm{d}z'}{\int_{z_1}^{z_2} \frac{\mathrm{d}n}{\mathrm{d}z'}\,\mathrm{d}z'} = \frac{(1+z)^{\gamma+1} - (1+z_1)^{\gamma+1}}{(1+z_2)^{\gamma+1} - (1+z_1)^{\gamma+1}} \equiv c(z), \tag{2.7}$$

where z_1 and z_2 are the lowest and highest values of H I absorption redshifts in the spectrum (respectively), and z is a variable. For $z_1 < z < z_2$, c is a uniform random variable with values $0 < c < 1$. This property makes the problem of assigning a redshift value to the line trivial, as it is possible to rearrange Eq. 2.7 to yield the function $z(c)$.

Defining:

$$p = (1+z_2)^{\gamma+1}, \tag{2.8}$$

$$q = (1+z_1)^{\gamma+1}, \tag{2.9}$$

and rearranging Eq. 2.7, we get:

$$c(p-q)+q = (1+z)^{\gamma+1}, \tag{2.10}$$

$$\left[\frac{\log_{10}(c\,(p-q)+q)}{\gamma+1}\right] = \log_{10}(1+z) \equiv x(c), \tag{2.11}$$

i.e.

$$z(c) = 10^{x(c)} - 1. \tag{2.12}$$

QSOSIM9 uses a randomly generated number $0 < c < 1$ for each of the n lines, which it passes to equations Eq. 2.11 and Eq. 2.12. The result is a redshift value z that is assigned to the line.

2.4 Column density distribution of HI absorption systems

QSOSIM9's algorithm to assign N_{HI} values to the emission lines is conceptually equivalent to the one that assigns redshift values. After rearranging the expression for the CDF of H I column density, the code generates a random number between 0 and 1 that is used to assign a value of N_{HI} to a line. The CDF of N_{HI} is given by the

expression:

$$\mathrm{CDF}(N_{\mathrm{HI}}) = \frac{\displaystyle\int_{N_{\mathrm{HI}}^{\mathrm{low}}}^{N_{\mathrm{HI}}} \frac{\mathrm{d}n}{\mathrm{d}N_{\mathrm{HI}}'}\,\mathrm{d}N_{\mathrm{HI}}'}{\displaystyle\int_{N_{\mathrm{HI}}^{\mathrm{low}}}^{N_{\mathrm{HI}}^{\mathrm{high}}} \frac{\mathrm{d}n}{\mathrm{d}N_{\mathrm{HI}}'}\,\mathrm{d}N_{\mathrm{HI}}'} = \frac{\left(N_{\mathrm{HI}}^{\mathrm{low}}\right)^{1-\beta} - \left(N_{\mathrm{HI}}\right)^{1-\beta}}{\left(N_{\mathrm{HI}}^{\mathrm{low}}\right)^{1-\beta} - \left(N_{\mathrm{HI}}^{\mathrm{high}}\right)^{1-\beta}} \equiv d\left(N_{\mathrm{HI}}\right), \tag{2.13}$$

where $N_{\mathrm{HI}}^{\mathrm{low}}$ and $N_{\mathrm{HI}}^{\mathrm{high}}$ are lower and upper cut-off values of column density, specified by the user in the input file. Typical values of $N_{\mathrm{HI}}^{\mathrm{low}}$ and $N_{\mathrm{HI}}^{\mathrm{high}}$ in QSOSIM9 are at the order of 10^{13} cm^{-2} and 10^{16} cm^{-2}, respectively.

For $N_{\mathrm{HI}}^{\mathrm{low}} < N_{\mathrm{HI}} < N_{\mathrm{HI}}^{\mathrm{high}}$, d takes the values between 0 and 1. Taking into account that $N_{\mathrm{HI}}^{\mathrm{high}} \gg N_{\mathrm{HI}}^{\mathrm{low}}$ and $1-\beta < 0$, we get:

$$d(N_{\mathrm{HI}}) = \frac{\left(N_{\mathrm{HI}}^{\mathrm{low}}\right)^{1-\beta} - \left(N_{\mathrm{HI}}\right)^{1-\beta}}{\left(N_{\mathrm{HI}}^{\mathrm{low}}\right)^{1-\beta}}. \tag{2.14}$$

Finally, rearranging Eq. 2.14 yields:

$$\log_{10}\left(N_{\mathrm{HI}}\right) = \frac{\log_{10}(1-d)}{1-\beta} + \log_{10}\left(N_{\mathrm{HI}}^{\mathrm{low}}\right). \tag{2.15}$$

The program generates a random number $0 < d < 1$ for each of the n lines, which it feeds into Eq. 2.15 to assign a value of N_{HI} column density to the line.

2.5 Doppler b-parameter

The b-parameter of an emission line is a measure of its width. The shape of the line profile and the physical processes that influence it are discussed in more detail in Sec. 3.3.4. In the model used by QSOSIM9, the b-parameter is a normally distributed with a mean $\mu = 23$ km s^{-1} and standard deviation $\sigma = 3$ km s^{-1}. QSOSIM9 uses a *Numerical recipes* function named GASDEV to randomly sample a normal distribution. It is an algorithm taken from and described in [87]. When called, GASDEV returns a double-precision number that is normally distributed with a mean $\mu_0 = 0$ and standard deviation $\sigma_0 = 1$. A single absorption line's b-parameter is determined by the algorithm:

$$b = \mu + \sigma \times t, \tag{2.16}$$

where t is the number returned by GASDEV.

2.6 Noise

QSOSIM9 has two noise models. One of them produces a constant signal-to-noise ration (SNR) per pixel in the continuum, has no wavelength dependance, and uses a base-value at zero-flux levels [12]. It produces a 1-σ error array:

$$\sigma = \frac{\mathrm{I_{obs}}}{\mathrm{SNR}} + 0.2\frac{\mathrm{I_0}}{\mathrm{SNR}}, \tag{2.17}$$

where the value of SNR is provided by the user in the input file, $\mathrm{I_{obs}}$ the observed intensity, and $\mathrm{I_0}$ is the unabsorbed intensity. The reason for this, as stated in [12], is that saturated lines and Lyman limits would otherwise remain noise-free. The other noise model of QSOSIM9 degrades the noise with decreasing wavelengths (towards the blue), by:

$$\sigma \propto (1 + e^{-x}), \tag{2.18}$$

where

$$x = \frac{\lambda - c_1}{c_2} \tag{2.19}$$

and where default and hard-coded values are $c_1 = 3532$ and $c_2 = 117$ [12].

3 Simulating QSO spectra – QSOSIM10

It is the aim of this thesis to create a new iteration of the QSOSIM program – QSOSIM10, taking QSOSIM9 (the version dated 11/06/2014, described in Sec. 2.1) as a starting point. While QSOSIM9 produced arbitrary resolution spectra, QSOSIM10 is specifically designed to create artificial QSO spectra as similar as possible to real SDSS BOSS QSO spectra. Each simulated spectrum corresponds to a real QSO in the SDSS DR10 quasar catalogue [13] with respect to the redshift, spectral index, magnitudes, and position in the sky. The simulated spectra also need to reproduce instrument specific characteristics of introduced by the use of the BOSS spectrographs, such as the wavelength range, signal-to-noise ratio, and instrumental beam properties. QSOSIM10 improves upon algorithms used by QSOSIM9 to model the IGM and the Lyman α forest, and uses new models of column density, redshift, and b-parameter distributions, in closer agreement with the recent data derived from high-resolution QSO spectra (e.g. [17]–[19], [82], [89]–[94]).

The various modifications introduced in QSOSIM10 are described in detail in this Chapter, and can be summarised as follows:

1. Improved modelling of the broad and weak emission lines coming from the gas in the proximity of the SMBH (Sec. 3.2)

2. An improved model of H I column density distribution, covering the H I column density range $N_{\rm HI} = 10^{12} - 10^{22}$ cm^{-2}(Sec. 3.3.1)

3. An improved model for the evolution of H I clouds' number density with redshift, for three populations of H I absorbers (Sec. 3.3.2)

4. The incorporation of clustering of H I clouds along the line-of-sight (Sec. 3.3.3)

5. An improved model of the b-parameter distribution of H I and metal absorption lines (Sec. 3.3.4)

6. The incorporation of metal absorption in clouds associated with high H I column density ($N_{\rm HI} \geqslant 10^{17}$ cm^{-2}). This is achieved by implementing CLOUDY, a radiative transfer code (Sec. 3.3.5).

7. The incorporation of the proximity effect – a decrease in the number of high column density systems in the proximity of the QSO, due to the strong ionising radiation of the QSO (Sec. 3.3.6)

8. The introduction of SDSS BOSS specific noise and resolution models. (Sec. 3.4)

QSOSIM10 was coded in FORTRAN. The reason for using an early-generation computer language is twofold: (1) previous iterations of QSOSIM and all of their subroutines have also been written using FORTRAN, and (2) FORTRAN's performing speed. Speed is an absolutely crucial factor when running a program hundreds of thousands of times in a loop. Therefore, benefits of using FORTRAN outweigh the use of some other computer languages popular in the astrophysical community like C, C++, Python, or Interactive Data Language (IDL).

Considering that the resulting spectra will be used on-par with SDSS data, the most convenient format for the output is a Flexible Image Transport System (FITS) format. FITS is the most commonly used format for digital files in astronomy, useful for storage, transmission, and processing of scientific images. FITS file consists of one or more Header and Data Units (HDUs) that can contain an image or a table with the information on one or many objects.

The reading and writing of FITS files is done using CFITSIO for FORTRAN. CFITSIO is a popular machine-independent library of routines for reading and writing FITS files. It provides a powerful interface for accessing and creating FITS files and was initially developed by the High Energy Astrophysics Science Archive Research Center (HEASARC) at the National Aeronautics and Space Administration's Goddard Space Flight Center. QSOSIM10 uses an input FITS file containing the simulation parameters used to generate the spectra. The input file was created from the DR10 quasar catalogue found on the SDSS website [95], and the quasar spectra downloaded from the SDSS Science Archive Server [96]. A full list of the simulation parameters required and used by QSOSIM10 is in Tab. 3.1. QSOSIM10 creates a FITS file for every spectrum it creates.

Each of the output FITS files created by QSOSIM10 consists of four HDUs:

1. The 'Primary' HDU that contains no information on QSO spectrum but is required in every FITS file.

2. The 'QSO_INFO' HDU that contains the SDSS designation of the QSO along with the QSO's unique identifier, the information on the date of observation, the plate and fiber used for spectroscopy, the QSO's right-ascension and declination coordinates (in J2000 system), redshift, spectral index (two values: one derived from the principal component analysis by Paris et al. [13], and the other derived from the spectral analysis performed by our group), the PSF magnitudes, the lowest wavelength observed in the spectrum and the number of pixels comprising the spectrum.

3. The 'QSO_SPECTRUM' HDU that contains the artificial QSO spectrum in seven columns: the wavelengths (in $\log_{10}$ scale) – '*loglam*', the flux – '*flux*', the noise

– '*noise*', the inverse variance of the flux – '*ivar*', the unconvolved flux with no noise – '*ncflux*', and the convolved flux with no noise – '*nnflux*'.

4. The 'LINE_LIST' HDU that lists the properties of absorption lines that were input into the spectrum: the atomic species, ionisation state, column density, redshift, and the b-parameter.

Each output file is named in accordance with the SDSS naming system, and follows the template '*mockspec-[PLATE]-[MJD]-[FIBER].FITS*', where *PLATE*, *MJD*, and *FIBER* are the same as in the real SDSS QSO catalogue.

In the first step, QSOSIM9 was modified from a stand-alone program into a subroutine. This subroutine is called from within a loop in the main program, each time with different input parameters for a different QSO. Sec. 3.1 describes how QSOSIM10 uses the QSO's spectral index α and the measured r magnitude to plot the underlying QSO continuum. Sec. 3.2 explains how emission lines are introduced into the spectrum, while Sec. 3.3 describes how QSOSIM10 uses statistics of the H I absorption systems and the plasma simulation code CLOUDY to model the Lyman α forest using Monte Carlo methods. Finally, Sec. 3.4 describes how SDSS characteristics have been introduced in the resulting QSO spectrum.

Table 3.1: Simulation parameters used by QSOSIM10 to generate artificial SDSS quasar catalogue.

Parameter	Description
SDSS_NAME	SDSS DR10 designation
RA	Right Ascension (J2000)
DEC	Declination (J2000)
THING_ID	Unique identifier
PLATE	Plate number
MJD	Spectroscopic MJD
FIBER	Fiber number
Z_VI	Visual inspection QSO emission redshift, z_{QSO}
ALPHA_NU	Spectral index α (PCA[a])
ALPHA_FIT	Spectral index α (independent analysis)
NPIX	Number of pixels comprising the spectrum
BEGIN_WAVE	Starting wavelength, $\log_{10}$ value
PSF_MAG	PSF magnitudes in the u, g, r, i, z bands

[a] Principal Component Analysis, [13]

3.1 Synchrotron mechanism

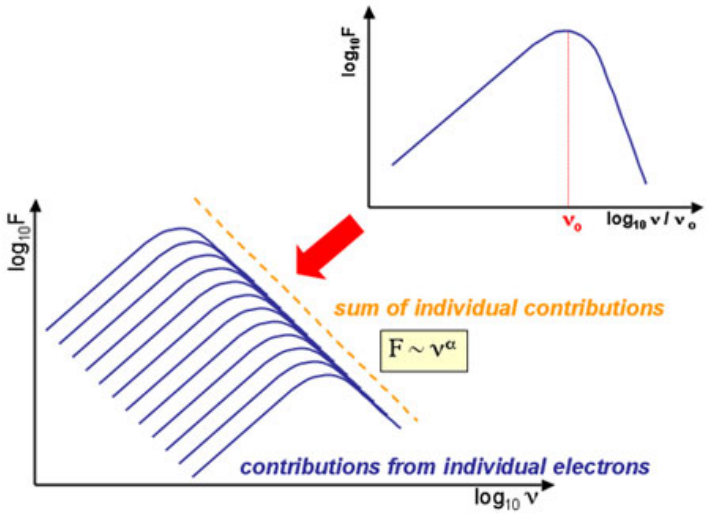

Figure 3.1: Synchrotron emission of relativistic electrons is thought to be the main contribution to the observed power-law shape of the underlying QSO spectrum. Image sourced from [97].

The underlying continuum emission of QSOs, is consistent with the assumption of synchrotron emission of relativistic electrons. The electrons are thought to be ejected from a confined area in the vicinity of the accreting black hole [54], [98]. When summed over the individual emissions of electrons (see Fig. 3.1), the QSO spectrum in the optical wavelength range can be described fairly well by a power law

$$F_\nu \propto C_\nu \; \nu^{-\alpha}. \tag{3.1}$$

F_ν is the value of flux at frequency ν, C_ν the normalisation constant, and α the spectral index. In QSOSIM10, the normalisation constant is determined from the information on the QSO's magnitude and spectral index. Magnitude is a measure of brightness of an object, measured at a certain frequency or a range of frequencies. There are many magnitude systems used in astrophysics, covering a wide range of bands (ranges in frequency/wavelength) from X-ray to radio. The SDSS uses modified Gunn u, g, r, i, z magnitude system, described by Fukugita et al. [14]. The relation between the physical flux of the QSO at wavelength $\lambda = 6182$ Å and the r magnitude[6] of the SDSS magnitude system is given by

$$r = -2.5 \log \left(\frac{F_{\lambda 6182}}{\mathrm{erg/s/cm^2/Å}} \right) - 21.364, \tag{3.2}$$

when all of the appropriate conversions have been taken into account. Inverting this equation gives us the value of $F_{\lambda 6281}$. This information can be used in conjunc-

[6]The effective wavelength of SDSS magnitude r is at $\lambda = 6182$Å. [14]

tion with the wavelength form of the power-law, given by Eq. 3.1, to plot the QSO continuum emission:

$$F_\lambda = C_\lambda \lambda^{-\alpha-2}, \tag{3.3}$$

where λ is the wavelength. The transformation between constants C_ν and C_λ is given by $C_\lambda = C_\nu\, c^{1-\alpha}$, where c is the speed of light in vacuum. This transformation is cosmetic only, as we are free to choose whether to consider the flux to be a function of wavelength or frequency. In optical astronomy, it is customary to express flux as a function of wavelength. The constant C_λ at $\lambda = 6182$ Å is then connected to the r magnitude via the physical flux at the same wavelength by:

$$C_{\lambda 6182} = \frac{F_{\lambda 6182}}{\left(6182\ \text{Å}\right)^{-(\alpha+2)}} = 10^{-\frac{r+21.364}{2.5}} \left(6182\ \text{Å}\right)^{\alpha+2}, \tag{3.4}$$

QSOSIM10 evaluates Eq. 3.4 using the α and r values read from the input file for every QSO, and uses it in conjunction with Eq. 3.3 to plot the underlying continuum emission of the QSO. The r magnitude was chosen over the others for two reasons – (1) the SDSS r magnitude is equivalent to the r magnitude in the AB photometric system and therefore requires no conversion, and (2) the CCD response of SDSS photometric system is highest at the wavelength range corresponding to the r magnitude [14].

3.2 QSO emission lines

One of the defining features of a QSO are the broad emission lines visible in its spectrum. The broad emission lines are produced by virialised photoionised gas in the vicinity of the central supermassive black hole, through a mechanism whose details are still not completely understood. We know that the line broadening does not come from thermal mechanisms, as temperatures in excess of $T \sim 10^{10}$ K would be necessary. At that temperature, all of the gas would be ionised and no emission would be possible. Furthermore, there would be a 511 keV signal from $e^+ - e^-$ pair annihilation visible in the spectrum. Therefore, broadness of emission lines is interpreted as resulting from Doppler broadening – the extreme environment in the proximity of the black hole results in extremely strong magnetic fields, accelerating particles and broadening emission lines in that region of space. We call this region the *broad line region* (BLR). The typical broad emission lines are Ly α, Mg II, C IV, C III], with typical widths of several thousand km s^{-1} [26].

A great majority of AGNs also exhibit a set of much narrower emission lines. These lines originate in an area further away from the accreting black hole than the BLR, where gas density is significantly lower. Because of this, it is called the *narrow*

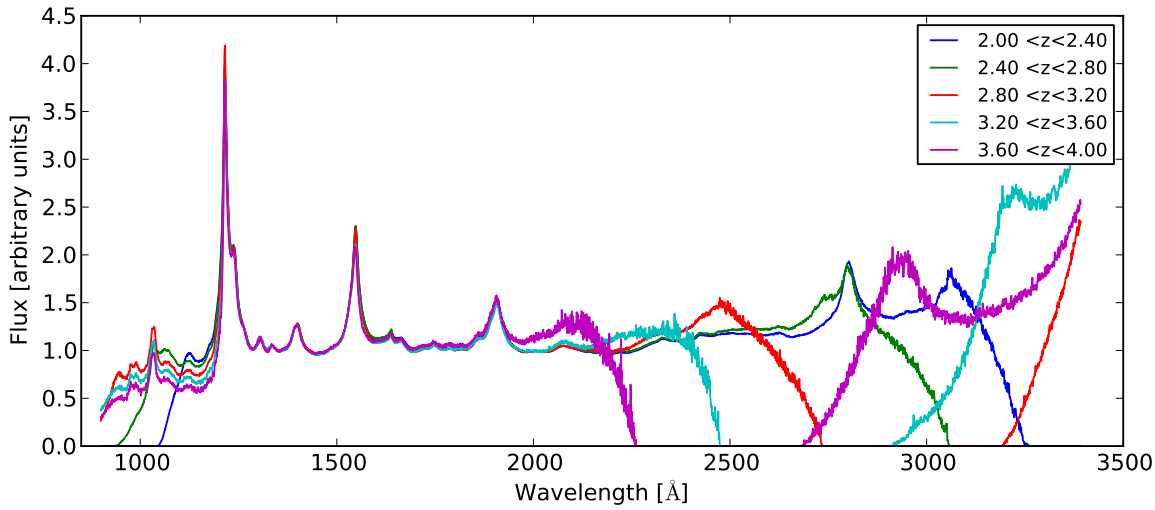

Figure 3.2: The normalised unweighted composite SDSS QSO spectra for 5 bins in redshift, $2.0 < z < 4.0$, $\Delta z = 0.4$. Relative strengths and widths of the emission lines between 1200 Å $\leqslant \lambda \leqslant$ 2000 Å (restframe) do not vary significantly with redshift. The most prominent emission line seen in the composites is the Ly α emission line at $\lambda \approx 1215$ Å. The continuum level bluewards of the Ly α line is seen to vary with redshift, indicating the change in the average opacity of the IGM, a characteristic called the Gunn-Peterson effect [59]. SDSS composite spectra provided by Sebastian Schmidt (private communication).

line region (NLR). Lower velocities of the gas result in emission lines with typical widths on the order of few hundred km s^{-1}. The strongest narrow lines are Ly α, C IV, and O III].

Naturally, no two QSOs have the same spectrum, but there are a number of broad and narrow emission lines that are common to the vast majority of QSOs. A composite spectrum, created by making an average of many individual ones, is used to study the statistical properties of the emission lines seen in QSO spectra. An example of five composite spectra at different redshifts made from SDSS data is shown in Fig. 3.2. The figure shows that many emission lines have the same relative strengths and widths across a large scale of redshifts. A list of emission lines from composite high-resolution QSO spectra was tabulated by several authors (e.g. [99]–[101]).

QSOSIM10 models the emission of 59 lines, almost double the number used by QSOSIM9. The restframe wavelength, the profile width, and the relative strength (compared to the Ly α) of each line is modelled using data from composite spectra [99]–[101]. Each emission line in QSOSIM10 is modelled as a Gaussian with mean λ redshifted from its restframe value λ_0 by an amount corresponding to the the QSO's redshift z_{QSO}, as below:

$$\lambda = \lambda_0(1 + z_{QSO}). \tag{3.5}$$

The entire list of emission lines that are used in QSOSIM10 can be found in Tab. A.1 of Appendix A.

3.3 *The Lyman α forest*

As in QSOSIM9, the underlying assumption in generating the Lyman α forest is that it is represented well by a collection of individual absorption lines. Starting with the statistical distribution of redshifts of H I absorbers, QSOSIM10 first calculates the number of H I absorption lines that are expected to be seen in the spectrum of a given QSO, n_0. Assuming that the distribution of the IGM clouds is random and independent, the predicted number of lines is susceptible to a counting error of the order $\sqrt{n_0}$. The total number of H I absorption lines that are used for generating the QSO spectrum – n, is hence determined by randomly sampling a normal distribution with a mean $\mu = n_0$ and standard deviation $\sigma = \sqrt{n_0}$. As before, the Monte Carlo routine GASDEV is used to generate a normally distributed double-precision number with $\mu = 0$ and $\sigma = 1$. Total number of lines, n, is then the nearest integer of a random number selected from this distribution:

$$n = \|n_0 + \sqrt{n_0} \times t\| \tag{3.6}$$

where t is a random number returned by GASDEV. QSOSIM10 generates a list of n H I absorption lines using the clustering model (described in Sec. 3.3.3), and associates metal absorption with lines above a certain H I column density threshold (described in Sec. 3.3.5). Algorithms that are used to generate the H I line list, i.e. the column densities, redshifts, and b-parameters of $N_{\rm HI}$ absorbers are described in Sec. 3.3.1, Sec. 3.3.2, and Sec. 3.3.4, respectively.

3.3.1 Column density distribution of HI absorption systems

While QSOSIM9 assumes that the number of absorption systems per unit H I column density is well described by a single power law, recent studies [15], [17], [73], [74], [82], [89]–[91], [102], [103] showed that is not the case. Some authors suggest that a broken power law fits the empirical data better (e.g. [15], [73], [91]), while others suggest a Γ-function fit on the high-end of the H I column density range ($N_{\rm HI} \geqslant 10^{17.2}$ cm^{-2}) [102]. The distribution of H I absorbers is often described by the H I column density distribution function (CDDF),

$$f(N_{\rm HI}, z) = \frac{{\rm d}^2 n}{{\rm d}N_{\rm HI}\,{\rm d}X} = \frac{{\rm d}^2 n}{{\rm d}N_{\rm HI}\,{\rm d}z}\frac{{\rm d}z}{{\rm d}X}, \tag{3.7}$$

defined as the number of absorption lines, ${\rm d}^2 n$, per unit column density ${\rm d}N_{\rm HI}$ per unit absorption distance ${\rm d}X$. The CDDF contains information on the primary statistical characteristics of the IGM. *Absorption distance X* was first defined by Bachall &

Peebles [104] as

$$X(z) = \int_0^z (1+z')^2 \frac{H_0}{H(z')} \, \mathrm{d}z', \tag{3.8}$$

where $H(z)$ is the Hubble constant at redshift z and H_0 its value at $z = 0$. In ΛCDM cosmology, evolution of the Hubble constant is described by:

$$H(z) = H_0 \left[(1+z)^3 \Omega_m - (1+z)^2 \Omega_r + \Omega_\Lambda \right]^{1/2}. \tag{3.9}$$

From Eq. 3.8, $\mathrm{d}X$ is connected with $\mathrm{d}z$ via the relation:

$$\frac{\mathrm{d}z}{\mathrm{d}X} = \frac{(1+z)^2}{\left[(1+z^3)\Omega_m + \Omega_\Lambda \right]^{1/2}}, \tag{3.10}$$

where Ω_m, Ω_r, and Ω_Λ are the matter, radiation, and dark energy densities of the Universe, respectively. In Eq. 3.10, Ω_r is taken to be ≈ 0 [3]. The usefulness of $X(z)$ becomes evident when considering non-evolving objects that are fixed in the co-moving coordinate system, as they will have a constant number density per unit absorption distance interval. Absorption distance is however not useful to us, for in order to calculate the total number of absorption lines in the QSO spectrum we first have to decide on a cosmological model. More importantly, H I absorption clouds are known to be evolving faster than predicted by any cosmological model, suggesting an intrinsic evolution of the gas [105]. It can be seen from Eq. 3.7 and Eq. 3.10 that the number of lines per unit redshift and the number of lines per unit absorption distance differ only by a constant. We have therefore decided not to make any assumptions on cosmology, and to rely on empirical distribution of H I absorbers' column density only.

In their recent paper, Prochaska et al. [90] made a meta-analysis of the data available on the column density distribution of H I absorption systems and tried to provide a unified description of the intergalactic medium at $z \approx 2.5$. They analysed the work of Prochaska & Wolfe [91], Noterdaeme et al. [74], Kim et al. [89], O'Meara et al. [103], Rudie et al. (henceafter R13, [82]), and others who have contributed in this area. R13 favour a broken power-law model with four parameters across the N_{HI} range $10^{14} < N_{\mathrm{HI}} < 10^{17}$ cm^{-2}, split at $N_{\mathrm{HI}} = 10^{15.4}$ cm^{-2}. This model yields $\chi^2_\nu = 1.44$ for their measurements. When applied to constraints provided by O'Meara et al., the R13 model yields quite a high value of $\chi^2_\nu = 22.3$ [90]. The reason for such an inconsistency with O'Meara et al. constraints lies in the fact that R13 did not include the absorption systems with the largest N_{HI} values in their analysis. Because of this, we can rule out the R13 model at a high confidence level [90]. O'Meara et al. suggest a six-parameter continuous set of power-laws which cover the range

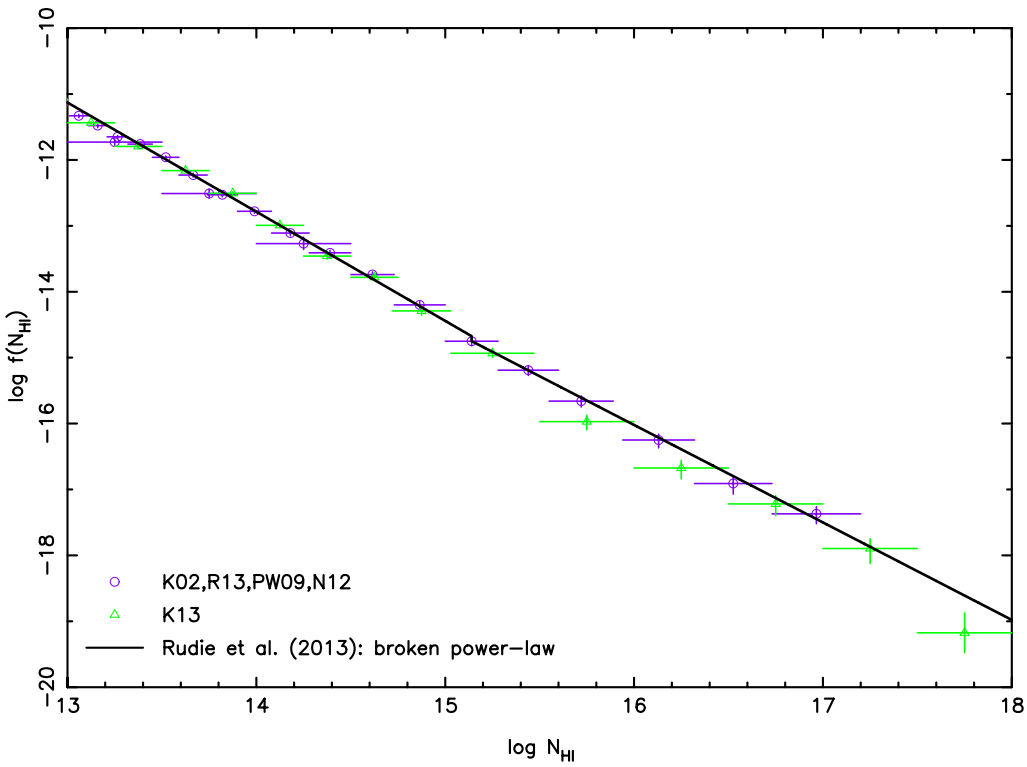

Figure 3.3: Column density distribution broken power-law best-fit model reported by Rudie et al. (R13) [82] shows very good agreement with measurements made by Kim et al. (2002) (K02), Prochaska & Wolfe (PW09) and Noterdaeme et al. (N12), shown in purple, but is inconsistent with newer data provided by Kim et al. (2013) (K13) [17], shown in green. The power-law breaks at $\log N_{\rm HI} = 15.4$.

10^{12} cm^{-2} $< \mathrm{N_{HI}} < 10^{20}$ cm^{-2}. Like the previous model, it too yields a high value of χ^2 when applied to the measurements from different authors, i.e. $\chi^2_\nu = 11.1$ when applied to the R13 data. In Fig. 3.3, it can be seen that the R13 broken power-law also does not fit well with independent measurements published recently by Kim et al. (2013) (henceafter K13, [17]). K13 did a double Voigt line profile fitting analysis of 18 high S/N, echelle quasar spectra obtained from the ESO[7] VLT/UVES archive, in an effort to describe the evolution of H I and C IV quasar absorption line systems at $1.9 < z < 3.2$ [17]. They fit a single and a triple power-law to their measured column density distribution function in three redshift bins, and report their best-fit values to the data. The two power-laws cover the range $10^{12.75}$ cm^{-2} $< \mathrm{N_{HI}} < 10^{18.0}$ cm^{-2}. The triple power-law has breaking points at $N_{\rm HI} = 10^{14.0}$ cm^{-2} and $N_{\rm HI} = 10^{15.0}$ cm^{-2}.

Prochaska et al. tried fitting a 7-parameter continuous power-law, using a Markov Chain Monte Carlo (MCMC) algorithm on constrains given by R13, O'Meara et al. and Prochaska & Wolfe . This model, as seen in Fig. 3.4, was in good agreement with the available data and yielded an acceptable value of $\chi^2_\nu = 1.36$ [90], but was in conflict with recent results published by K13. Prochaska et al. argue that the discrepancy arises from a systematic uncertainty in assessing $f(N_{\rm HI})$ at $N_{\rm HI} > 10^{15}$ cm^{-2} that the authors did not take into account.

[7]European Southern Observatory

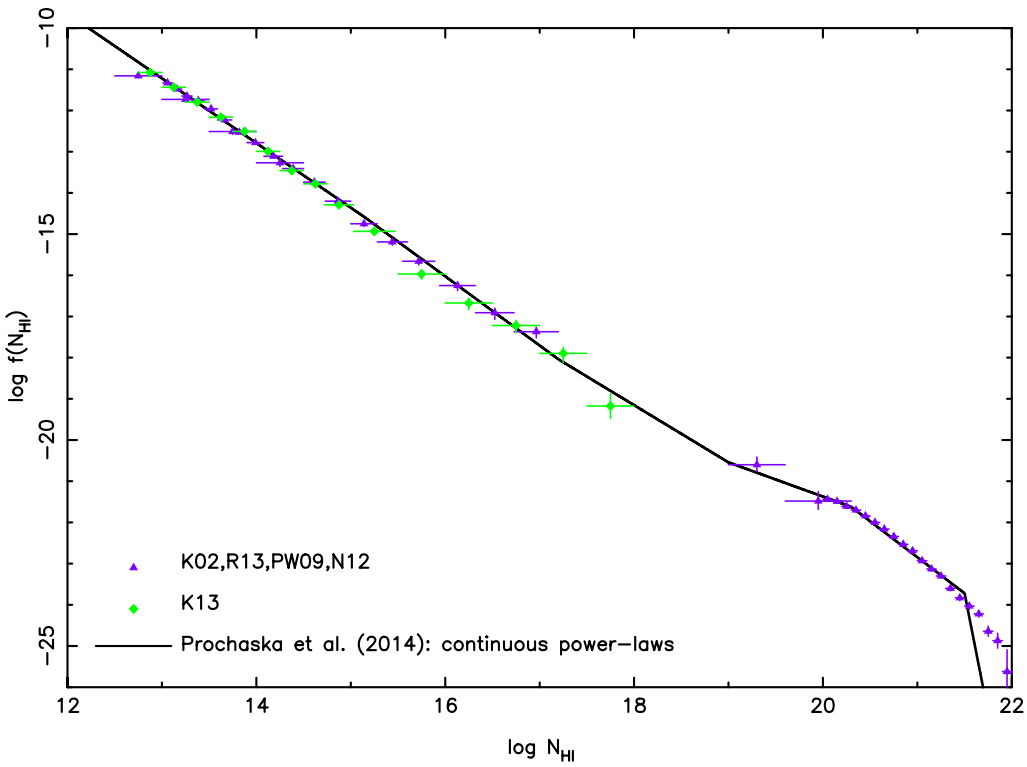

Figure 3.4: Prochaska et al. [90] attempted to reconcile the constraints reported by Rudie et al. and O'Meara et al. with a 7-parameter set of continuous power-laws. This model has a very good agreement with the measurements of Kim et al. (2002) (K02), Rudie et al. (R13), Prochaska & Wolfe (PW09), and Noterdaeme et al. (N12) at $N_{\rm HI}$ values up to 10^{17} cm^{-2}, but is inconsistent with data from Kim et al. (2013).

In an effort to provide a unified model of the IGM taking into account all constrains described above, Prochaska et al. suggest using a new functional form for $f(N_{\rm HI})$ in the form of a monotonically declining spline using the Hermite spline algorithm of Fritsch & Carlson [106]. The spline is parametrised by eight points which are only allowed to vary in amplitude. This model describes the IGM across a large range of column densities, from $N_{\rm HI} = 10^{12}$ cm^{-2} to $N_{\rm HI} = 10^{22}$ cm^{-2}, with a $\chi^2_\nu = 0.86$. This is shown in Fig. 3.5. We adopt this spline model in QSOSIM10 when assigning H I column densities. It is, however, unknown whether the shape of the spline changes with redshift, as this analysis hasn't yet been done. Therefore, when modelling of the Lyman α forest, we assume that the shape of the CDDF is constant and does not evolve with redshift.

Random selection of $N_{\rm HI}$ values for the artificial spectrum is now done in a similar way to QSOSIM9: we integrate the CDDF numerically in 10 000 points across the whole range of column densities to construct a cumulative distribution function (CDF). This CDF is sampled by a set of computer-generated random numbers when generating the H I absorption in the forest.

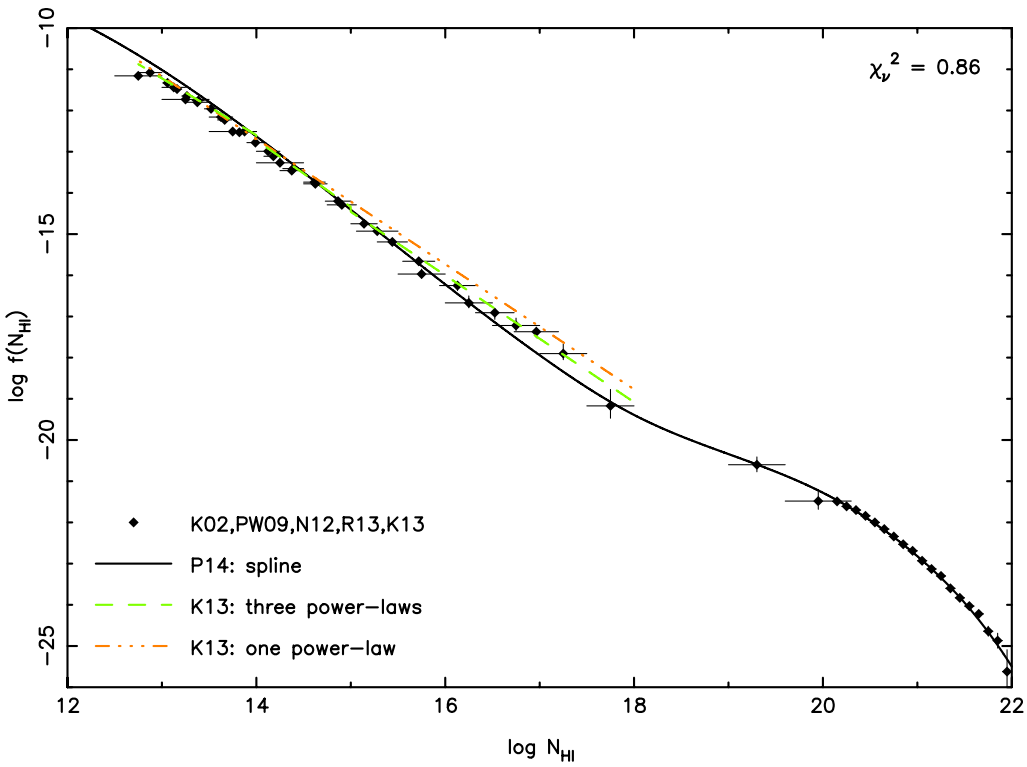

Figure 3.5: The new functional form of the column density distribution function, $f(N_{\rm HI}, z)$, proposed by Prochaska et al. [90], fits the measurements of Kim et al. (2002) (K02), Prochaska & Wolfe (PW09), Noterdaeme et al. (N12), Rudie et al. (R13), and Kim et al. (2013) (K13) reasonably well with a $\chi_\nu^2 = 0.86$. Also shown are the single and triple power-laws from K13, which do not fit the data as well.

3.3.2 Redshift evolution of the HI absorption systems

Assuming that the absorption systems have a constant size and a constant comoving number density, the number of absorption systems seen in quasar spectra per unit redshift is:

$$\frac{\mathrm{d}n}{\mathrm{d}z} = \pi R^2 N_0 c H^{-1}(z)(1+z)^2, \tag{3.11}$$

with R being the cloud's radius, N_0 local comoving number density, c the speed of light and $H(z)$ the Hubble parameter at a redshift z [104]. However, intergalactic clouds of hydrogen are known to evolve with time (see [15], [82], [86], [90] for example). Whether there is significant evolution of the shape of the CDDF with redshift is still unknown, although there are some indications that is the case [17]. The general trend of the number density evolution with redshift is best described by a power law

$$\frac{\mathrm{d}n}{\mathrm{d}z} = A(1+z)^\gamma, \tag{3.12}$$

where A is the local comoving number density for a specific cloud population, and γ the evolution parameter. Both A and γ are determined from high-resolution QSO spectra. Many authors have investigated the evolution of Ly α lines for different populations (low, medium and high H I column densities at low and high redshift)

[15], [86], [88], [89], [91], [107]. Recently, the evolution parameter of systems with column densities $N_{\rm HI} = 10^{14} - 10^{17}$ cm^{-2} and $N_{\rm HI} = 10^{13.1} - 10^{14}$ cm^{-2} at redshifts $1.9 < z < 3.2$ was determined by Kim et al. [17] to be $\gamma = 2.16 \pm 0.14$, and $\gamma = 1.51 \pm 0.09$, respectively. It should be noted that Eq. 3.12 describes a general trend of the line distribution, and that it does not follow the power-law *exactly*. Two-point correlation functions of H I clouds have shown that clouds are correlated on scales of ~ 10 Mpc [18]. Cloud clustering will be discussed in more detail in Sec. 3.3.3.

Like in QSOSIM9, the first step in Lyman α forest generation is to determine the total number of H I absorption systems most likely to be seen in the spectrum, n_0. Unlike in QSOSIM9, where we treated clouds across a wide range of column densities as a single population, the model used in QSOSIM10 divides the range $N_{\rm HI} = 10^{12} - 10^{22}$ cm^{-2} into three distinct populations: $N_{\rm HI} = 10^{12} - 10^{14}$ cm^{-2}, $N_{\rm HI} = 10^{14} - 10^{17}$ cm^{-2}, and $N_{\rm HI} = 10^{17} - 10^{22}$ cm^{-2}.

In order to model the evolution of absorption systems down to column densities $N_{\rm HI} = 10^{12}$ cm^{-2}, for which we have no reliable data, we make the assumption that they form the same population of gas clouds as $N_{\rm HI} = 10^{13.1} - 10^{14}$ cm^{-2} systems. Consequently, we assume that systems with column densities $N_{\rm HI} = 10^{12} - 10^{13.1}$ cm^{-2} evolve in the same way as the systems with $N_{\rm HI} = 10^{13.1} - 10^{14}$ cm^{-2}, and are attributed the same value of γ. Also, we need to correct the normalisation constant A to reflect the increase in population.

We correct A by determining the relative increase in the number of H I absorbers. This is done by integrating the CDDF, f, in ranges $N_{\rm HI} = 10^{12} - 10^{14}$ cm^{-2} and $N_{\rm HI} = 10^{13.1} - 10^{14}$ cm^{-2}. We multiply the value of A that corresponds to the old population, A_{old}, by the ratio of the two integrals to get the value that corresponds to the new population, A_{new}:

$$A_{new} = A_{old} \frac{\int_{10^{12}}^{10^{14}} f(N_{\rm HI})\, {\rm d}N_{\rm HI}}{\int_{10^{13.1}}^{10^{14}} f(N_{\rm HI})\, {\rm d}N_{\rm HI}}. \tag{3.13}$$

In other words, extending the column density range results in a larger cloud popula-

Table 3.2: Values of A and γ for different $N_{\rm HI}$ column density bins from literature.

$\Delta \log N_{\rm HI}$	$\log A$	γ	Source
13.1 – 14.0	0.72 ± 0.08	1.51 ± 0.09	[17]
14.0 – 17.0	1.52 ± 0.05	2.16 ± 0.14	[17]
17.0 – 22.0	-0.82 ± 0.18	1.94 ± 0.34	[16]

tion. The normalisation constant needs to be corrected to reflect this by multiplying the old constant with the relative increase in the population number. The relative increase is determined by integrating the distribution of column densities in the limits that define the old and new populations.

The number of absorption systems is the sum of the number of systems in each of the three populations. To calculate the total number of absorption lines in three populations, QSOSIM10 integrates Eq. 3.12 using values of A and γ corresponding to each population. The upper and lower integration limits are z_1 and z_2, with the latter being equal to the QSO's emission redshift z_{QSO}. As in QSOSIM9, z_1 is defined as the lowest possible redshift of Ly α that can be observed in the spectrum (see Eq. 2.5). The number of H I absorption systems, i.e. H I absorption lines, most likely to be seen in the spectrum is:

$$n_0 \Big|_{z_1}^{z_2} = \sum_{i=1}^{3} \int_{z_1}^{z_2} A_i (1+z)^{\gamma_i} \, \mathrm{d}z = \sum_{i=1}^{3} \frac{A_i}{1+\gamma_i} \left[(1+z_2)^{1+\gamma_i} - (1+z_1)^{1+\gamma_i} \right], \tag{3.14}$$

where i denotes one of the three populations.

Assuming that H I absorption systems are distributed randomly and independently from each other, we can estimate the error of n_0 to be $\sqrt{n_0}$, on the basis of counting statistics. When generating the Lyman α forest in a given QSO spectrum, QSOSIM10 evaluates Eq. 3.14, and randomly samples a normal distribution with a mean $\mu = n_0$ and standard deviation $\sigma = \sqrt{n_0}$ to determine the number of H I absorption lines that will be inputted in the Lyman α forest. The total number of H I absorption lines in each QSO spectrum is defined as the nearest integer of the expression:

$$n = \| n_0 + \sqrt{n_0} \times t \|, \tag{3.15}$$

where t is a random, computer-generated, normally distributed number with a mean $\mu = 0$ and $\sigma = 1$.

To verify the validity of this method and the way we correct the normalisation constant A, we calculate $\mathrm{d}n/\mathrm{d}z$ for a set of QSOs for which this number is known. Kim et al. [17] provide $\mathrm{d}n/\mathrm{d}z$ of two populations of H I absorbers seen in the high-

Table 3.3: Values of A and γ for the three populations of H I absorption systems used by QSOSIM10, after corrections.

i	$\Delta \log N_{\mathrm{HI}}$	$\log A_i$	γ_i
1	12.0 – 14.0	2.17 ± 0.11	1.51 ± 0.09
2	14.0 – 17.0	0.72 ± 0.08	2.16 ± 0.14
3	17.0 – 22.0	-0.82 ± 0.18	1.94 ± 0.34

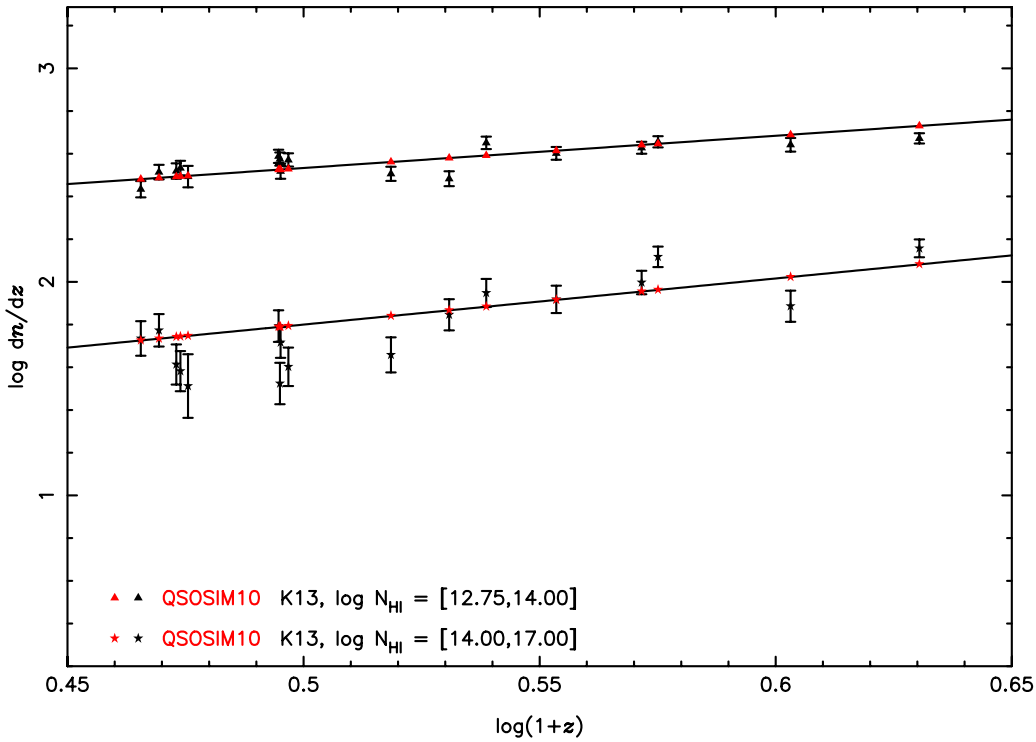

Figure 3.6: The measured and modelled values of $\mathrm{d}n/\mathrm{d}z$ for two cloud populations, as seen in the spectra of 18 QSOs. The black stars are values from a study of high-resolution spectra by Kim et al. (K13) [17]. Red stars are the values predicted by the QSOSIM10 model, using the corrections described in the text. The best fits to the data, as reported by [17], are shown as black lines, and are used by QSOSIM10.

resolution spectra of 18 QSOs. We calculate $\mathrm{d}n/\mathrm{d}z$ for the same two populations, $N_{\mathrm{HI}} = 10^{12.75} - 10^{14}$ cm^{-2} and $N_{\mathrm{HI}} = 10^{14} - 10^{17}$ cm^{-2}. We correct the normalisation constant to encompass systems down to column density $N_{\mathrm{HI}} = 10^{12.75}$ cm^{-2} in the same way as before. The fit of our model to the observed data is given in Fig. 3.6. We can see that our model reproduces the data well and that normalisation correction works, providing further evidence that the shape of the CDDF at low column densities is accurate.

3.3.3 Clustering

The matter density along the line-of-sight towards a quasar is not constant. Variations in the matter density produce areas of increased and decreased density, compared to the average value. This effect is reflected in the redshift distribution of the observed number of hydrogen clouds (i.e. H I absorption lines) in the spectrum. Areas of increased matter density will have a higher-than-average number of absorption lines, while the opposite is true for areas of decreased matter density. This effect, known as *clustering* of the Lyman α forest, was first reported by Webb [108], [109], and further investigated by many others [18], [88], [89], [110]. The studies conclude that the IGM clusters on a scale of $150 - 250$ km s^{-1} [18], [110].

A major improvement to QSOSIM9 is the incorporation of the clustering effect

in QSOSIM10. The mathematical model of $\mathrm{d}n/\mathrm{d}z$ distribution (see Eq. 3.12) is perturbed to reflect the fluctuations in the number density of H I absorption lines. QSOSIM10 simulates line clustering using the model proposed by Barcons and Webb [18].

The clustering algorithm works in five steps:

1. Let the average number of (randomly distributed) clusters per unit redshift be $\mathrm{d}n_{cl}/\mathrm{d}z$. In the framework of the Barcons & Webb model, this number is:

$$\frac{\mathrm{d}n_{cl}}{\mathrm{d}z} = \frac{1}{4(1+z)\,[v_0(z)/c]\,\xi(0;z)}, \tag{3.16}$$

   where v_0 is the clustering scale, c the speed of light in vacuum, and $\xi(0;z)$ the amplitude of the two-point correlation function (TPCF) at zero velocity separation. The values used by QSOSIM10 are $v_0 = 150$ km s^{-1}, and $\xi(0;z) = 0.5$. The algorithm first calculates the expected number of clusters, n_{cl}^0, integrating Eq. 3.16 in the redshift range from z_1 to z_2 (see Sec. 3.3.2 for the definitions of z_1 and z_2). Using a Monte Carlo algorithm, the program samples a normal distribution with the mean n_{cl}^0 and the deviation equal to the square root of this number, to return the number of clusters to be input into the spectrum, n_{cl}.

2. A redshift value is assigned to each of the cluster centres, using random sampling of a uniform distribution between z_1 and z_2. All clusters are assumed to have a simple (exponential) density profile:

$$\nu(\Delta z; z_{cl}) = \exp\left(-\frac{|\Delta z|}{R(z_{cl})}\right), \tag{3.17}$$

   where z_{cl} is the redshift of the cluster centre, and Δz the distance from the centre. The cluster radius, R, is defined as:

$$R(z) = (1+z)\,\frac{v_0}{c}. \tag{3.18}$$

3. The effect of clustering on the average number density of H I lines, $\mathrm{d}n/\mathrm{d}z$, at redshift z is quantified by summing the contributions of all clusters :

$$\frac{\mathrm{d}n}{\mathrm{d}z} = \sum_{\text{clusters}} h(z_{cl})\,\nu(\Delta z; z_{cl}), \tag{3.19}$$

   where

$$h(z_{cl}) = 2\xi(0;z)A(1+z)^{\gamma} \tag{3.20}$$

   is the cluster richness. The values of A and γ are the same as in Tab. 3.3, and

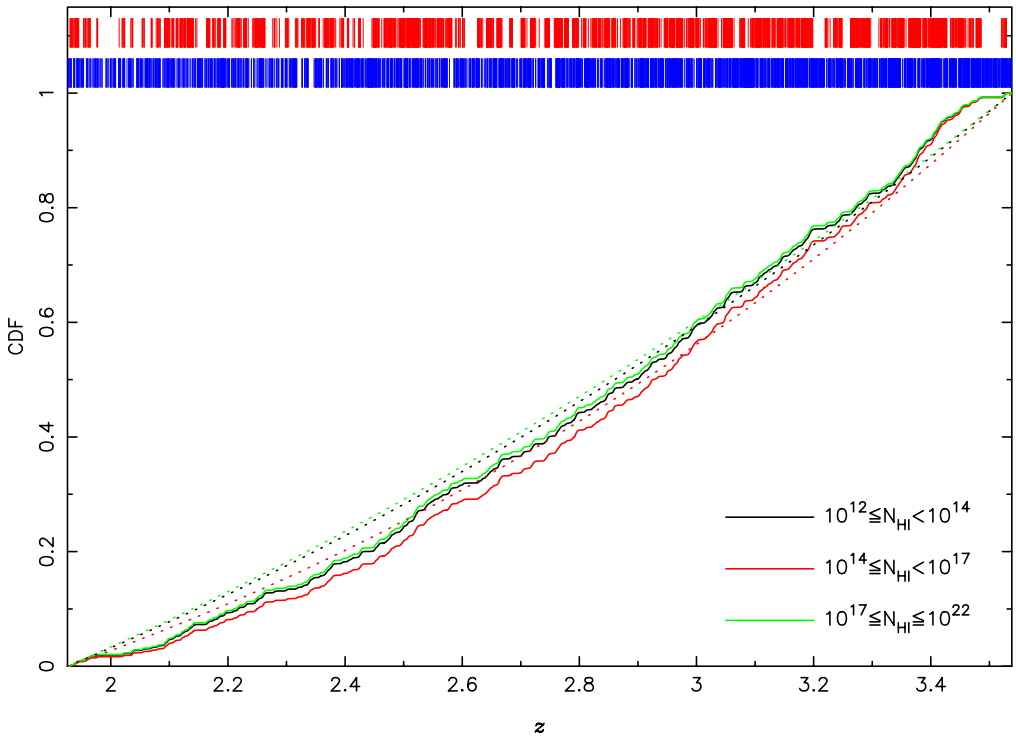

Figure 3.7: The cumulative distribution function (CDF) of the number of H I absorption lines per unit redshift (for the three cloud populations) for a particular realisation of a sightline towards a QSO at redshift $z_{QSO} = 3.5383$. Full lines show the CDF of the clustering model (as used by QSOSIM10); while the dotted lines show the CDF excluding clustering model (as used in QSOSIM9). The redshifts of absorption lines retrieved using the clustering model is shown as a series of red lines at the top of the figure, while the ones retrieved using the non-clustering model is shown as a series of blue lines.

are different for different H I cloud populations.

4. Eq. 3.19 is integrated numerically in the redshift range from z_1 to z_2 to generate the cumulative distribution function of redshifts for H I absorption lines. Every real number $x \in [0,1]$ is assigned a redshift value $z \in [z_1, z_2]$, by a bijective map defined by the CDF:

$$f : x \mapsto z. \tag{3.21}$$

5. The total number of H I absorption lines in the spectrum is derived using Eq. 3.15. To each line, a randomly generated number $x \in [0,1]$, is assigned. Using the previously defined map f between the CDF and redshift values, a redshift is assigned to the line:

$$z = f(x). \tag{3.22}$$

A particular realisation of a sightline towards a QSO at redshift $z_{QSO} = 3.5383$ is shown on Fig. 3.7. QSOSIM10's clustering algorithm has distributed 197 clusters between redshifts $z_1 = 1.9261$ and $z_2 = 3.5383$ (see point 1., above), and generated the three CDFs shown. The total number of H I absorption lines in the Lyman α forest was determined to be 1820, using Eq. 3.15. The redshifts of the lines were determ-

ined from a series of 1820 random numbers between 0 and 1. Each line was assigned a redshift corresponding to the assigned random number (see points 4. and 5., above). For a visual comparison of the models used by QSOSIM9 and QSOSIM10, in Fig. 3.7 we show the redshifts of lines drawn from the same sample of random numbers using both models. We see that the QSOSIM10's model reproduces the clustering of H I lines, as seen in real QSO spectra.

3.3.4 Doppler b-parameter

Correct modelling of H I absorption in the Lyman α forest requires knowledge of the absorption line widths distribution. Physical processes in the intergalactic medium cause the otherwise narrow emission and absorption lines to broaden. In addition to the line broadening caused by the quantum uncertainty, broadening can be caused by thermal and/or turbulent motion of the intergalactic gas. Natural line broadening caused by the Uncertainty principle reflects the finite lifetime of atomic levels. A basic atomic transition at frequency ν has a Lorentzian profile,

$$\phi(\nu) = \frac{1}{\pi} \frac{\Gamma/4\pi}{(\nu - \nu_0)^2 + (\Gamma/4\pi)^2}. \tag{3.23}$$

The quantum-mechanical damping constant Γ is a measure of probability for the decay of the quantum state of the atom, and is expressed in units s^{-1}.

In gas with non-zero temperature, thermal motion of gas particles causes the wavelength of the transition to be red- or blue-shifted with respect to the zero-temperature gas. The observed frequency is dependent on the velocity $\mathbf{v}$ of the particle and its direction with respect to the observer $\hat{\mathbf{r}}$. The Doppler frequency, ν_D, is defined as:

$$\nu_D = \nu \left(1 - \frac{\mathbf{v} \cdot \hat{\mathbf{r}}}{c}\right), \tag{3.24}$$

where c is the speed of light. The velocity distribution of a gas at a kinetic temperature T_k is well described by the Maxwell-Boltzman distribution,

$$f(V)\,\mathrm{d}V = 4\pi \left(\frac{m}{2\pi k_B T_k}\right)^{3/2} V^2 \exp\left(-\frac{1}{2}\frac{mV^2}{k_B T_k}\right) \mathrm{d}V, \tag{3.25}$$

where $V^2 = v_x^2 + v_y^2 + v_z^2$ is the total velocity of a particle, m is the particle's mass, and k_B is the Boltzmann constant. When considering the Doppler broadening of a line, only the velocity component along the line-of-sight towards the observer is relevant. For a particle of mass m at temperature T_k, thermal broadening of a line with a restframe

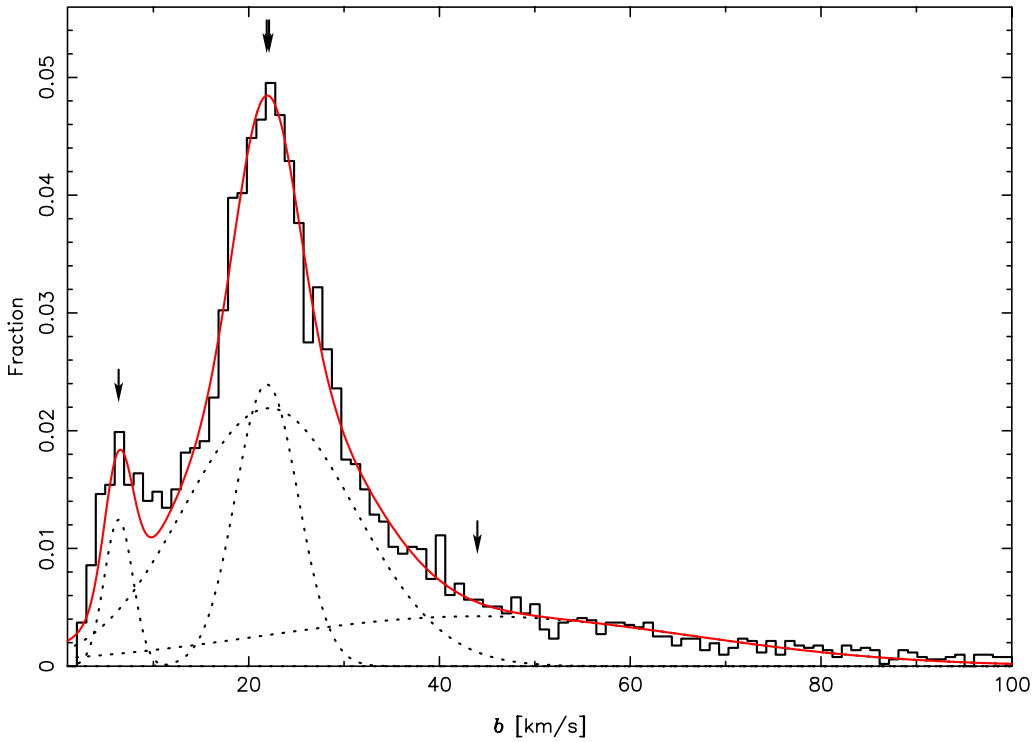

Figure 3.8: Normalised probability distribution of the Doppler b-parameter of absorption lines in the Lyman α forest. The observed distribution (black filled line) is best parametrised as a superposition of four normal distributions, shown in red. The four component distributions are shown as dotted lines, with arrows indicating the position of the peaks. The component peaking at $b \approx 6.3$ km s^{-1} is due to absorption by metals, while the two components that peak at $b \approx 22$ km s^{-1} are due to H I absorption. The fourth, widest, component is most likely caused by artefacts in the Lyman α forest. Data provided by M. Bainbridge (private communication).

frequency ν_0 is defined as

$$\Delta\nu_{th} = \frac{\nu_0}{c}\sqrt{\frac{2k_B T_k}{m}}. \tag{3.26}$$

Therefore, thermal broadening of a line results in a Gaussian profile:

$$\theta(\nu) = \frac{1}{\pi}\frac{1}{\Delta\nu_{th}}\exp\left[-\left(\frac{\nu-\nu_0}{\Delta\nu_{th}}\right)^2\right]. \tag{3.27}$$

Physical processes that contribute to line broadening are not necessarily thermal in origin, and are collectively called *turbulent* motions. They include chaotic internal motions inside the gas cloud and other non-physical processes that are well described by a normal distribution of velocities along the line-of-sight. This makes the total Doppler line broadening easy to obtain by simply adding the two velocity components together

$$V_D^2 = V_{th}^2 + V_{turb}^2, \tag{3.28}$$

while keeping the gaussian profile of the line.

The resulting convolution of the two different line profiles (Lorentzian and Gaussian) is called a Voigt profile, whose width is referred to as *Doppler b-parameter*. The

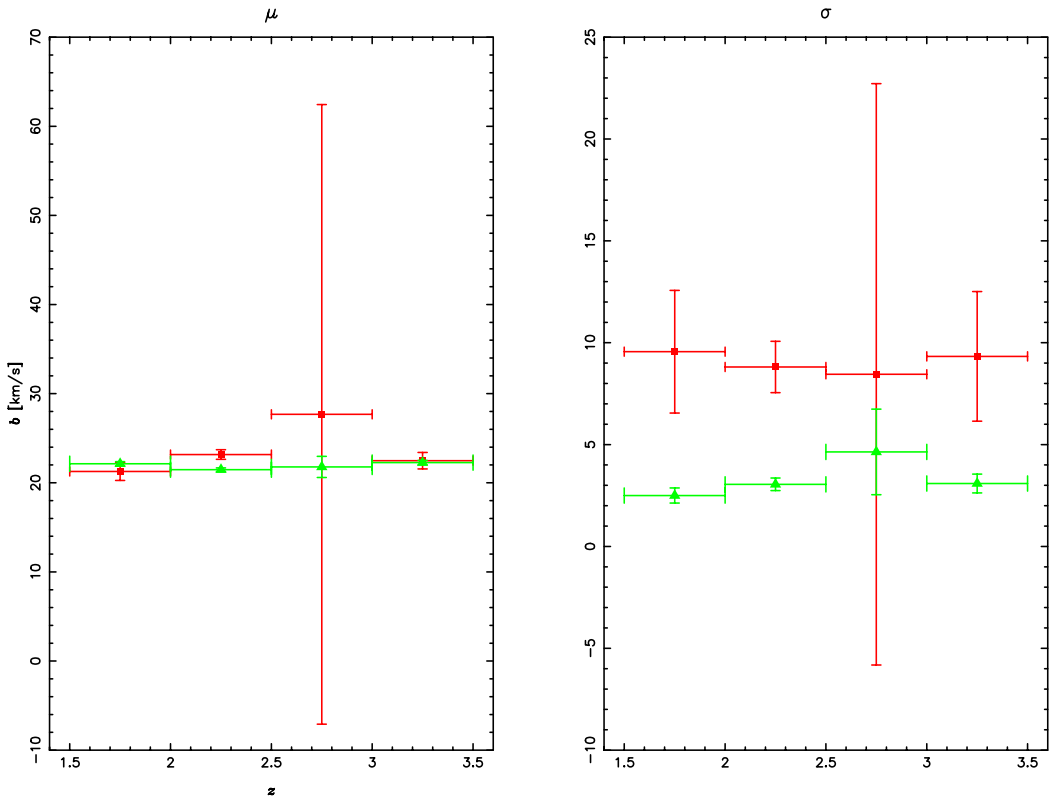

Figure 3.9: The means μ (left) and standard deviations σ (right) of the two Gaussian components of the H I absorption line distribution show no significant evolution with redshift. Green points correspond to the $i = 2$ component, while the red ones correspond to the $i = 3$ component (see Eq. 3.29 and Tab. 3.4). Data provided by M. Bainbridge (private communication).

distribution of the b-parameter in H I absorbers has been studied by numerous authors (e.g. [19], [111]–[113]). The observed distribution is in agreement with the assumption that thermal broadening provides a substantial component to line widths [111].

For QSOSIM10 the algorithm that assigns the value of the b-parameter to the H I absorption line is the same as in QSOSIM9 (see Sec. 2.5), but uses a mean and standard deviation that better reflect the observed distribution. This observed distribution, shown in Fig. 3.8, was provided by Matthew Bainbridge (private communication) and was derived from absorption lines in 58 high-resolution VLT/UVES+Keck QSO spectra (Bainbridge et al., in prep). The probability distribution of line widths is best parametrised by as a superposition of four normal distributions:

$$\mathscr{P}(b = x) = \sum_{i=1}^{4} C_i \, \exp\left(\frac{x - \mu_i}{\sqrt{2}\,\sigma_i}\right)^2, \tag{3.29}$$

where $\mathscr{P}(b)$ is the probability of b being equal to x, C_i is the normalisation constant, and μ_i and σ_i are the mean and standard deviation of the ith component, respectively. Using least-squares regression to fit these superimposed Gaussians to the data returns the parameters tabulated in Tab. 3.4, with a $\chi^2 = 1.13$. The two strongest components, peaking at $b \approx 22$ km s^{-1} are attributed to H I absorption, while the

component peaking at $b \approx 6.3$ km s^{-1} is most likely due to metal absorption in the Lyman α forest. The weakest and broadest component is most likely a systematic effect in the analysis, and comes from partially saturated lines that were not removed from the sample, or erroneous modelling of noise spikes in the spectrum. The two H I associated peaks and their standard deviations do not vary with redshift (see Fig. 3.9).

Based on these fitted parameters for components 2 and 3 in the Tab. 3.4, we choose to sample a normal distribution with mean $\mu_{\rm HI} = 22$ km s^{-1} and standard deviation $\sigma_{\rm HI} = 5$ km s^{-1} to model the H I absorption line widths in QSOSIM10. The b-parameter for metal absorbers is required for the work presented in the next section. Their line widths of metal absorbers are sampled from the component 1 – the normal distribution with the mean $\mu_{metals} = 6$ km s^{-1} and standard deviation $\sigma_{metals} = 1.5$ km s^{-1}.

Table 3.4: Best-fit parameters for the four superimposed normal distributions defined in Eq. 3.29 and fitted to the observed data shown in Fig. 3.8.

i	C_i [$\times$100]	μ_i^a	σ_i^a
1	1.2 ± 0.1	6.4 ± 0.2	1.5 ± 0.2
2	2.4 ± 0.3	21.9 ± 0.2	3.3 ± 0.3
3	2.2 ± 0.3	22.1 ± 0.4	9.1 ± 1.3
4	0.4 ± 0.1	44 ± 15	22.6 ± 8.1

^{a} in units: km s^{-1}

3.3.5 Metal absorption

Another major improvement introduced in QSOSIM10 is the incorporation of metal absorption lines in the simulated QSO spectra. Metals produced in stars are expelled into the intergalactic medium by supernova explosions, galactic winds, and galaxy interactions [114]. This is an important new feature, as metals in QSO spectra are used to study the evolution of stellar populations in the early Universe (e.g. [115], [116]), galaxy formation and evolution (e.g. [117]), and the evolution of cosmic metallicity and the ultraviolet background (e.g [94], [118]). Metals in the IGM also shed light on the influence of the galactic environment on the star-formation and its evolution (e.g. [83], [119]). The evolution and distribution of metals and their abundances in the IGM can be determined by studying high-resolution (R $\gtrsim$ 40 000) quasar spectra [84]. The majority of extragalactic absorption lines seen in quasar spectra come from the Lyman series of hydrogen, the greatest number of which come from Ly α absorption. Metal absorption is associated with higher H I column density systems; the fraction of systems with associated metal absorption increases with increasing H I column density. Absorbers with $N_{\rm HI} \geqslant 10^{15}$ cm^{-2} have associated metal

absorption approximately 73% of the time [19].

Like the H I absorption, we assume that the metal absorption seen along the lines-of-sight toward distant QSOs is well represented by a series of individual absorption lines. Therefore, it is necessary to know the distribution of metal column densities, Doppler b-parameters and redshifts. Compared to H I absorbers, metals have a smaller range of column densities (eg. 10^{13} cm^{-2}< $N_{\rm OVI} < 10^{15}$ cm^{-2} compared to 10^{12} cm^{-2}< $N_{\rm HI} < 10^{22}$ cm^{-2} of neutral hydrogen) and widths ($b \approx 3-6$ km s^{-1} of metal absorbers compared to $b \approx 15-35$ kms^{-1} of neutral hydrogen absorbers, see Sec. 3.3.4).

The column densities of metals in the intergalactic medium clouds depend on the physical conditions present in the cloud and the cloud's environment, such as the density of hydrogen gas in the cloud, its metallicity, and its proximity to background sources of radiation (i.e. galaxies and quasars). QSOSIM10 uses a highly advanced simulation program CLOUDY, developed by the researchers at the Cambridge University, to model the metal absorption. CLOUDY is designed to simulate conditions in a non-equilibrium gas and predicts a number of its properties [20]. Given a set of input parameters, CLOUDY determines column densities of 30 atomic species. For each H I absorption system above a threshold value of $N_{\rm HI} = 10^{17}$ cm^{-2} in the simulated QSO's spectrum, QSOSIM10 uses CLOUDY calculations to determine column densities of metals that will be input into the spectrum. The calculations were performed using version 13.03 of CLOUDY, described in [20]. CLOUDY takes input values for four parameters: redshift of the absorber system z, its metallicity Z, hydrogen volumetric number density, and H I column density. We set the hydrogen volumetric number density to be 1/100 of Solar, and equal for all systems. An example of the CLOUDY input file used by QSOSIM10 to model metal absorption in the Lyman α forest is given in Appendix B.

The evolution of the DLA metallicities through cosmic time (up to $z \approx 5$) was a topic of study by Rafelski et al. [85], who measured metal abundances [M/H] in 47 DLAs. They found that the cosmic metallicity $\langle Z \rangle$ decreases with increasing redshift, but has a large intrinsic scatter. Their bootstrap analysis on the Echellette Spectrograph and Imager (ESI), HIRES and literature data yields a simple parametrisation of cosmic metallicity:

$$\langle Z \rangle = (-0.22 \pm 0.03)\, z - (0.65 \pm 0.09), \tag{3.30}$$

The data points and the fit are shown in Fig. 3.10.

Calling CLOUDY for every absorption system above the threshold in H I column density would be very computationally intensive, increasing the time needed for

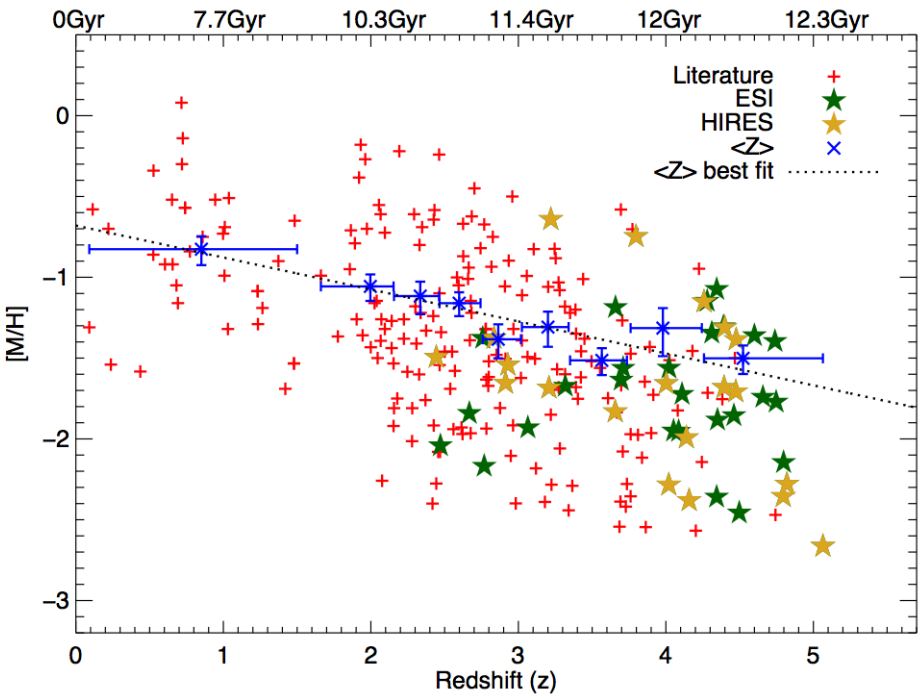

Figure 3.10: Measures of metallicity in the Damped Lyman α (DLA) systems as a function of redshift. Red plus signs are the literature values, while green and gold stars represent ESI and HIRES metallicity measurements. The blue crosses with error bars show measured cosmic metallicity $\langle Z \rangle$. Binwidths are determined such that there are equal numbers of data points in every redshift bin. The vertical error bars represent 1-σ confidence levels. The dashed line is the linear fit to the $\langle Z \rangle$ data points. Figure sourced from [85].

spectrum generation by an order of magnitude. Instead, we make a database of CLOUDY output that is read by QSOSIM10 when needed. We discretely sample a three-dimensional parametric space $(Z, N_{\rm HI}, z)$ with 20 000 points (20 points in the metallicity range $-3 \leqslant Z \leqslant 0$, 50 points in the $N_{\rm HI}$ range 10^{12} cm$^{-2} \leqslant N_{\rm HI} \leqslant 10^{22}$ cm^{-2}, and 20 points in the redshift range $1.9 \leqslant z \leqslant 5.2$). We use CLOUDY to calculate the column densities of metals for all points in this parametric space and save them in an output file. The database contains column densities of 30 atoms and almost 500 of their ionisation states. The strongest ones of these, which we incorporate into our modelled spectra, are: C I, C II, C III, C IV, N I, N II, N III, N V, O I, O VI, Mg I, Mg II, Al II, Al III, Si II, Si III, Si IV, Cr II, Mn II, Fe II, Fe III, Ni II, Zn II. This database is read by QSOSIM10, and used to input metal absorption in the QSO spectrum.

For a system above the threshold value, QSOSIM10 calculates the metallicity of the system, Z_{sys}, using Eq. 3.30 and a Monte Carlo algorithm. Using the known values of the neutral hydrogen column density $N_{{\rm HI}sys}$, and redshift z_{sys}, QSOSIM10 reads the metal column density values of the point closest in $(Z, N_{\rm HI}, z)$ parametric space. QSOSIM10 then inputs metal absorption lines into the QSO spectrum using the column densities at that point in the parametric space. The line widths are sampled from a normal distribution with a mean $\mu = 6$ km s^{-1} and standard deviation $\sigma = 1.5$ km s^{-1}, as described in Sec. 3.3.4.

3.3.6 The QSO proximity effect

Carswell et al. [21] were the first to notice this peculiar effect seen in QSO spectra – the number of H I absorption lines drops significantly in the close proximity of the QSO. This effect, caused by the strong radiation emitted by the QSO that ionises the neutral hydrogen gas in its proximity, is called the *proximity effect*. QSOSIM10 incorporates the proximity effect when generating QSO spectra, further increasing their sophistication.

The degree of ionisation of hydrogen in the IGM depends on the rate at which the hydrogen gas is ionised, and the rate at which it can recombine into neutral hydrogen. The density of neutral hydrogen in the system, n_{HI} is inversely proportional to the photoionisation rate Γ_{HI}:

$$n_{\mathrm{HI}} \propto \frac{1}{\Gamma_{\mathrm{HI}}}, \tag{3.31}$$

which is in turn proportional to the total number of ionising photons.

Far from the QSO, the IGM is ionised by the integrated ultraviolet emission from the AGN and star-forming galaxies, known as the ultraviolet background (UVB). The spectral energy distribution of the UVB and its evolution is studied from the metal absorption in QSO spectra [120], galaxy number and luminosity function evolutions [121], and the proximity effect [92], [93]. In the proximity of the QSO, the QSO itself is the dominating source of ionising radiation. This radiation will cause the local gas to be highly ionised, resulting in the decrease of the detected H I column density of clouds. The number of H I absorption systems in the QSO proximity, above a certain column density threshold, will therefore also decrease.

The decrease in the H I column density of a system is dependent on the contribution of the QSO to the total ionising radiation, relative to the UVB. The detected H I column density of a cloud near to the QSO is:

$$N = N_0 \left(1 + \omega\right)^{-1}, \tag{3.32}$$

where N_0 is the H I column density in the absence of any QSOs, and

$$\omega = \frac{{F_\nu}^Q}{4\pi J_\nu}. \tag{3.33}$$

In this equation, ${F_\nu}^Q$ is the local Lyman limit (LL, $\lambda_{LL} = 912$ Å) flux density (in units erg/s/cm^2/Hz) due to the QSO, and J_ν is the ultraviolet background intensity at the Lyman limit (in the same units). The QSO will have a noticeable impact on the ionisation rate of hydrogen (and by extension, on the detected H I column density) when its contribution to the ionising radiation incident upon on the cloud is comparable to

the contribution from the UVB.

The local LL flux density is proportional to the intrinsic QSO monochromatic luminosity at the LL, L_ν, and inversely proportional to the QSO's luminosity distance to the cloud, r_L:

$$F_\nu{}^Q = \frac{L_\nu}{4\pi r_L^2}. \tag{3.34}$$

The monochromatic luminosity, or the energy output at a certain wavelength (frequency), of a QSO at a redshift z_{QSO} is related to the flux at the same wavelength (frequency) measured by an observer on Earth, f_ν, via:

$$L_\nu = 4\pi\, d_L^2\, f_\nu\, (1+z_{QSO})^{-1}, \tag{3.35}$$

where d_L is the QSO's luminosity distance to Earth, and the factor $(1+z_{QSO})^{-1}$ is the bandwidth correction factor [93]. Combining Eq. 3.34 and Eq. 3.35, the local LL flux density impingent on the cloud becomes:

$$F_\nu{}^Q = f_\nu \left(\frac{d_L}{r_L}\right)^2 (1+z_{QSO})^{-1}. \tag{3.36}$$

QSOSIM10 calculates $F_\nu{}^Q$ from the information on the r magnitude and the redshift z_{QSO} of a QSO, found in the SDSS QSO catalogue (see Sec. 3.1 for details). The luminosity distance to an object at redshift z is defined as:

$$d_L = \int_0^z \frac{\mathrm{d}z}{\sqrt{(1+z)^3\,\Omega_m + (1+z)^2\,\Omega_r + \Omega_\Lambda}}. \tag{3.37}$$

For an observer on Earth, $z = z_{QSO}$. However, the redshift of the QSO is different for a (theoretical) observer (in a theoretical starship) located at the position of the absorbing cloud, i.e. at redshift z_{cl}. This observer will measure the QSO to be at redshift z':

$$z' = \frac{z_{QSO}+1}{z_{cl}+1} - 1. \tag{3.38}$$

Therefore, the luminosity distance from the absorbing cloud to the QSO is:

$$r_L = \int_0^{z'} \frac{\mathrm{d}z}{\sqrt{(1+z)^3\,\Omega_m + (1+z)^2\,\Omega_r + \Omega_\Lambda}}. \tag{3.39}$$

To calculate the luminosity distance, QSOSIM10 uses the cosmological parameters published by the Planck Collaboration [3]: $\Omega_m = 0.307$, $\Omega_r = 5.38 \times 10^{-5}$, $\Omega_\Lambda = 0.691$, and $H_0 = 67.8$ km/s/Mpc.

To quantify the contribution of the UVB to the ionising radiation across a large

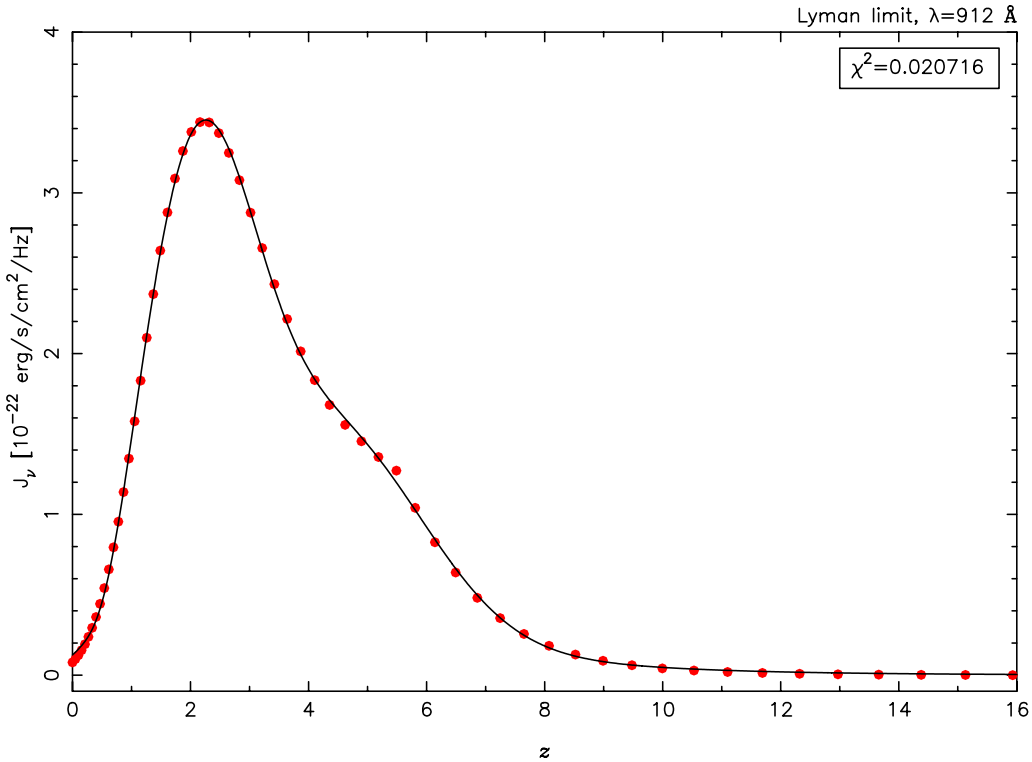

Figure 3.11: The Haardt & Madau [22] model of the redshift evolution of the UVB, J_ν, at the Lyman limit ($\lambda = 912$ Å), as derived by CUBA. The UVB is the integrated ultraviolet emission from the cosmic galaxy and AGN populations. The red points are derived by the radiative transfer code CUBA, and are downloaded from [23]. In QSOSIM10, we model the background by a superposition of a single lognormal and two Gaussian distributions. The parameters of the fit are tabulated in Tab. 3.5.

range in redshifts (J_ν in Eq. 3.33), we use the a radiative transfer code, CUBA, developed by Haardt & Madau [22]. CUBA simulates the propagation of the Lyman-continuum photons through a partially ionised inhomogeneous IGM. This code calculates the photoionisation rate $\Gamma_{\rm HI}$ and the UVB's spectral energy distribution as functions of redshift, taking into account the number and emissivity of star-forming galaxies and QSOs at different epochs. We use the publicly available CUBA output to model the UVB at the LL across a large range of redshifts [23]. In order to avoid interpolating between the points, we look for an analytical expression that best fits the results from CUBA. We found that the best fit to the intensity of the UVB at the LL, J_ν is a superposition of two Gaussians and a single lognormal distribution:

$$J_\nu(z) = C_1 \exp\left(\frac{z-\mu_1}{\sqrt{2}\sigma_1}\right)^2 + C_2 \exp\left(\frac{z-\mu_2}{\sqrt{2}\sigma_2}\right)^2 + C_3 \exp\left(\frac{\log z-\mu_3}{\sqrt{2}\sigma_3}\right)^2, \tag{3.40}$$

where indices 1 and 2 refer to the two Gaussian distribution parameters, and the index 3 refers to the lognormal distribution parameters. Our best fit to the parameters is given in Tab. 3.5.

Fig. 3.11 shows the redshift evolution of the UVB modelled by CUBA in 60 points. The best-fit model used by QSOSIM10 is also shown, with a $\chi^2 = 0.02$.

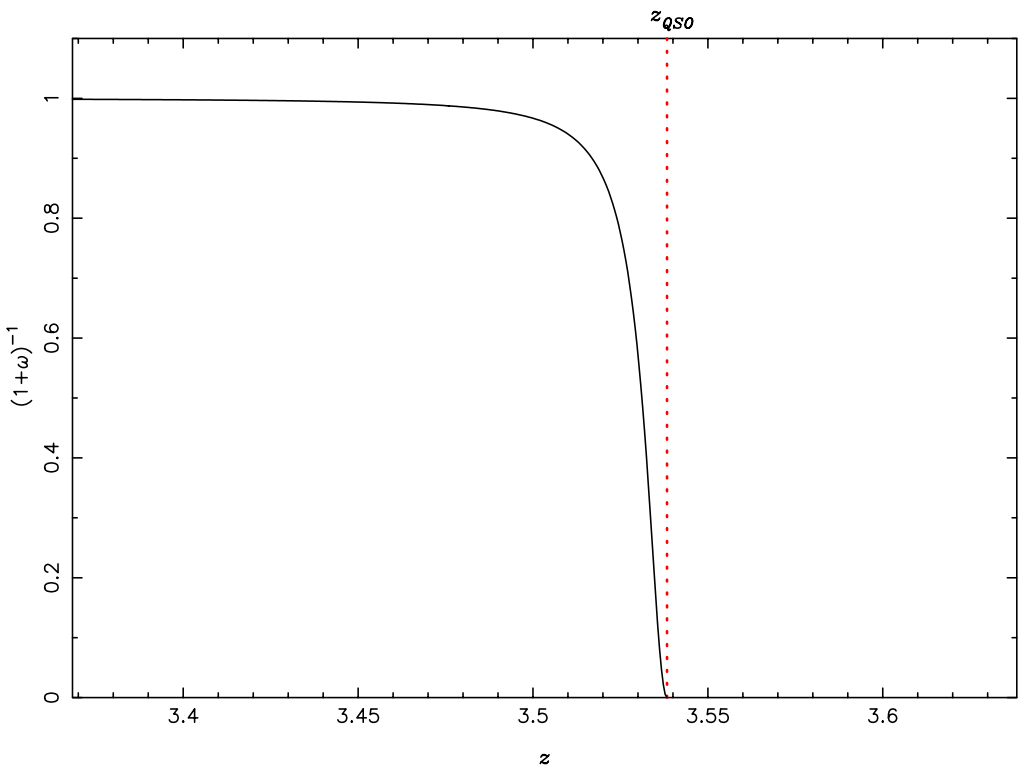

Figure 3.12: The scaling of H I column densities as a function of QSO proximity. The column density of an distant absorber is unchanged by the influx of ionising radiation from the QSO, as it is negligible compared to the UVB. An absorber that is close to the QSO, however, is highly ionised by the QSO's radiation and its H I column density drops steeply with proximity.

Table 3.5: Parameters of the fit to the UV background model of Haardt & Madau [22]. We fit the UV background as a superposition of a single log-normal and two Gaussian distributions (see Eq. 3.40).

	C^a	μ	σ
1st Gaussian	1.90 ± 0.04	2.17 ± 0.02	0.92 ± 0.02
2nd Gaussian	0.96 ± 0.08	4.6 ± 0.1	1.44 ± 0.01
Lognormal	1.3 ± 0.1	0.76 ± 0.02	0.422 ± 0.008

[a] in units: 10^{-22} erg/s/cm^2/Hz

QSOSIM10 uses the values of ${F_\nu}^Q$ and J_ν to calculate the scaling factor of H I column density, $(1+\omega)^{-1}$ as per Eq. 3.32. The dependance of the scaling factor on the distance from the QSO is shown on Fig. 3.12 for an example of a $z_{QSO} = 3.5383$ QSO.

3.4 BOSS characteristics

QSOSIM10 was specifically constructed to emulate the QSO spectra obtained by the SDSS BOSS spectrographs. The models used in the simulations therefore need to reproduce the resolving power and noise characteristics of BOSS. This requires deep understanding of the BOSS hardware and the way it interprets and reconstructs the QSO signal.

The resolution elements of BOSS (hereafter referred to as 'pixels') have a constant size in velocity space, equal to $p^{v} = 69$ km s^{-1} [24]. To make the size of the resolution elements in the simulated spectrum equal to their real BOSS counterparts, we convolve the simulated spectrum with a Gaussian profile with a standard deviation $\sigma = p^{v}$. The convolution smears out the spectrum and blends the spectral absorption lines. Convolving the spectrum removes the information on the absorption lines with the lowest column densities from the spectrum, as they are no longer resolvable.

The signal-to-noise ratio of the observed object's spectrum depends on several factors, such as: the observing time, the collecting area of the telescope, the luminosity of the background sky, the dark current and the read-out-noise of the CCD. Larger telescopes can gather more photons per unit time coming from the object, thus reducing the noise. The same effect is accomplished by collecting more photons by observing for a longer period of time. Generally, objects will have significantly lower SNR if they are observed during periods of increased background sky brightness, such as the period of the full Moon. This is especially true for very faint objects, with brightness comparable to that of the night sky, whose spectra become heavily contaminated by background sky photons. Telescope electronics are another source of noise in the observed spectrum. Thermal excitations of electrons in various electronic components of the telescope cause false-positive detections of photons, decreasing the SNR ratio. This effect can be diminished by keeping the integral electronic components sufficiently cooled.

The quantity that influences the noise seen in the spectrum the most is the throughput. The throughput of a telescope is defined as the ratio of the number of photons that are detected by the telescope's CCD and the number of photons from the target object incident on the atmosphere. The total throughput is influenced by all effects that diminish the total number of photons detected by the CCD. They include, but are not limited to, atmospheric absorption, sensitivity of the telescope electronics and telescope optics, as well as the CCD's effectiveness to convert photons into electrons. Both of the twin BOSS spectrometers have a 'red' and a 'blue' arm, each covering a different range of wavelengths, and generally have slightly different properties. The CCDs in each of the arms are optimised for performance at different wavelengths, which results in a variation of throughput with respect to incident photon's energy. QSOSIM10 omits this, and uses a single model for both arms. The major effects contributing to the lowering of the total throughput of SDSS BOSS are shown in Fig. 3.13.

To correctly model the noise characteristics of BOSS, QSOSIM10 uses the information on the workings of the BOSS spectrographs to convert the values of flux at every pixel into photon counts and vice-versa. Every resolution element in the fi-

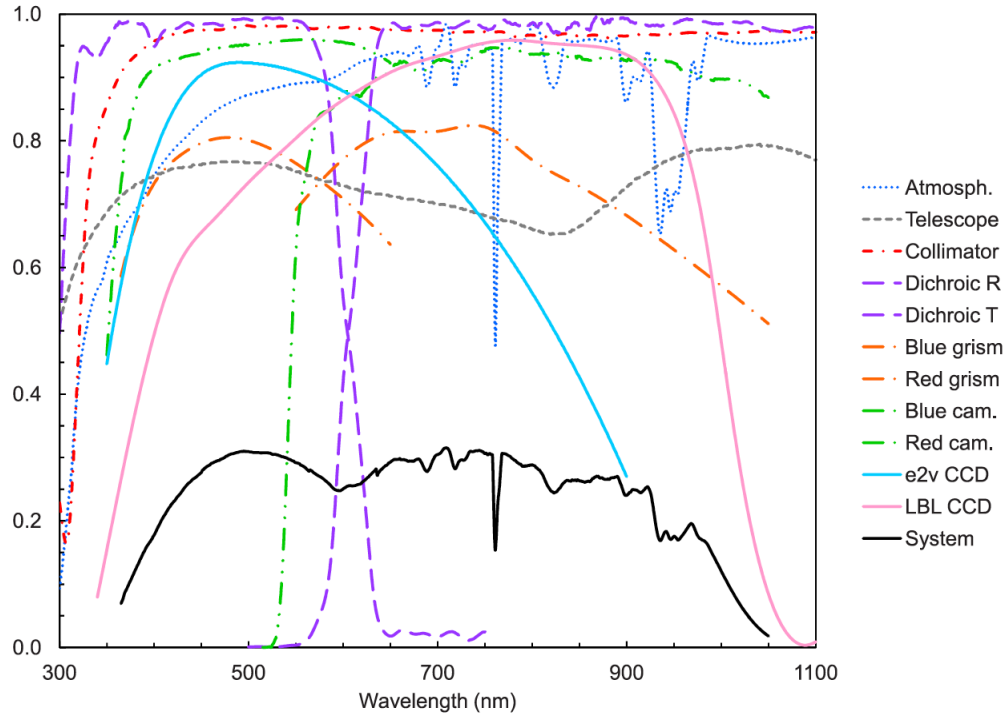

Figure 3.13: The total throughput of SDSS BOSS spectrographs (black line) has contributions from atmospheric absorptions, telescope optics and electronics sensitivity and efficiency, as indicated in the legend. Figure sourced from [8].

nal QSO spectrum represents the number of photons detected in a small energy (i.e. wavelength[8]) range. The number of photons counted at the position of the ith pixel, PC_i, is connected to the value of flux at the same pixel by the following equation:

$$PC_i = \frac{F_i \; p_i^{\lambda}}{E_i \; T_i} \; A \; t. \tag{3.41}$$

Here, E_i is the energy of photons counted by the ith pixel (in erg), and F_i the physical flux of the source (in units erg/s/cm^2/Hz). A is the collecting area of the BOSS telescope, and is $\approx (2.5 \text{ m})^2\pi$ [7]. Time spent observing the source is t. p_i^{λ} is the wavelength range covered by the ith pixel. BOSS spectra have pixels of constant size in velocity space. The conversion between the pixel size in the velocity space p^{v} and the pixel size in the wavelength space, p^{λ}, is given by:

$$p^{\lambda} = \frac{p^{v}}{c} \; \lambda. \tag{3.42}$$

The throughput of SDSS BOSS shown in Fig. 3.13 was digitised into 856 points using a web-tool DEXTER [122]. The throughput was found to be well fitted by a superposition of four Gaussian profiles, shown on Fig. 3.14. QSOSIM10 calculates

[8]The energy of the photon is related to its wavelength by $E = \hbar \frac{c}{\lambda}$

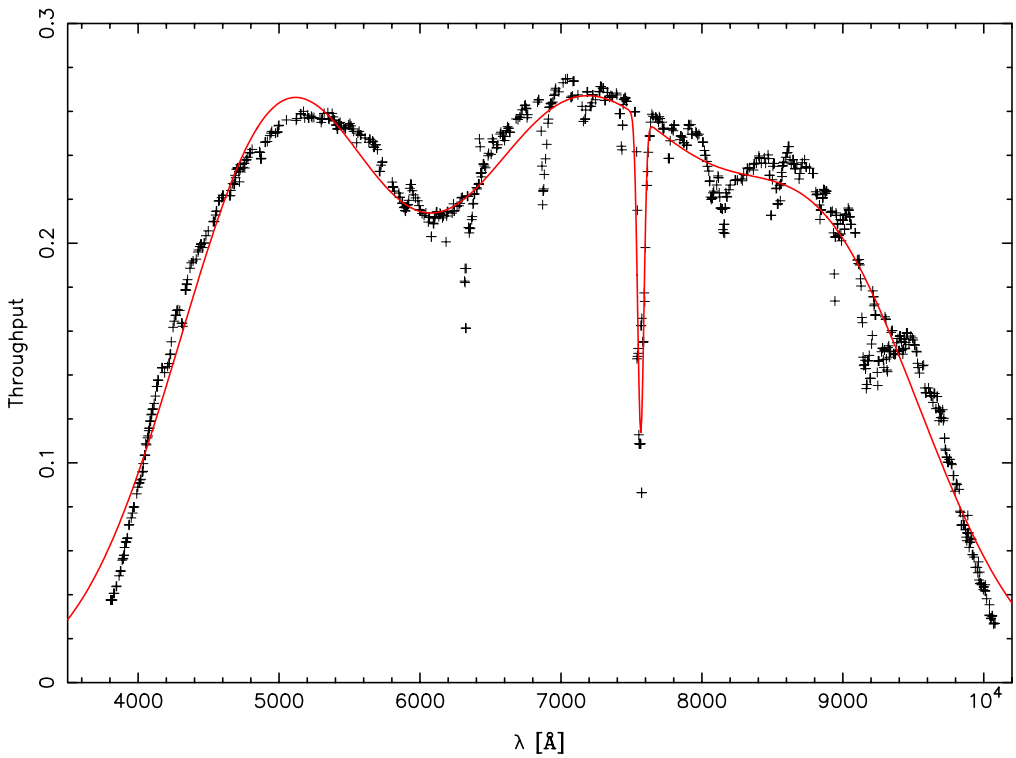

Figure 3.14: The values of the throughput of the SDSS BOSS spectrographs was digitised from Fig. 38. of [8] using the web-tool DEXTER [122]. QSOSIM10 crudely models the throughput of SDSS BOSS spectrographs as a superposition of four Gaussian profiles (red line).

the throughput T_i of the pixel i using the analytical expression:

$$T_i = \sum_{j=1}^{4} C_j \, \exp\left(\frac{\lambda_i - \mu_j}{\sqrt{2}\,\sigma_j}\right)^2, \tag{3.43}$$

where C_j, μ_j and σ_j are the normalisation constant, mean, and standard deviation of the jth Gaussian. The parameters of the fit are given in Tab. 3.6.

QSOSIM10 converts the flux of the simulated QSO into photon counts using Eq. 3.41 and the throughput model described above. We refer to this as the 'signal', or S. The information on the signal is used to calculate the inverse of the variance for every pixel. For the ith pixel, the inverse of the variance σ_i^2 is:

$$\frac{1}{\sigma_i^2} = S_i + \left[S_i^{sky} + \text{DC}\, t\right] n_{pix} + \text{RON}^2 \frac{t}{t_{RO}}. \tag{3.44}$$

In Eq. 3.44, S denotes the QSO photon count signal 'recorded' by the ith pixel, S^{sky} the signal from the background sky, DC the dark current per unit time, n_{pix} the number of pixels that make up the spectrum, RON and t_{RO} the read-out-noise and read-out period of the CCD, and t time of exposure. Values of these parameters were

taken from [8] and hardcoded into the program. Finally, a Monte Carlo method is used to assign the value of noise (in photon counts) to every pixel by sampling a normal distribution with the mean $\mu = 0$ and the standard deviation equal to the value of σ_i (obtained from Eq. 3.44). The inverse of the variance and the noise photon counts are converted into physical fluxes by inverting Eq. 3.41. The resulting noise is then added to the simulated flux coming from the QSO.

The inverse of the variance and the noise of every pixel are saved in the 'QSO_SPECTRUM' HDU of the FITS file.

Table 3.6: Parameters of the fit to the BOSS throughput consisting of four superimposed Gaussian, as in Eq. 3.43.

j	C	μ [Å]	σ [Å]
1	0.248 ± 0.005	5002 ± 25	723 ± 15
2	0.254 ± 0.004	7108 ± 36	898 ± 62
3	0.18 ± 0.01	8897 ± 55	728 ± 27
4	-0.143 ± 0.004	7566.1 ± 0.9	22 ± 1

4 Results of QSOSIM10

This chapter presents and discusses the main results of QSOSIM10 – the simulated quasar spectra. We generate the entire SDSS BOSS DR10 QSO data set, using the data provided in the SDSS DR10 QSO catalogue [13]. The spectrum generation process is similar to the one of QSOSIM9, but includes the more sophisticated models and methods described in Chapter 3. The process of generating the artificial SDSS QSO data set can be summarised as follows.

1. In the first step, QSOSIM10 opens the SDSS DR10 QSO catalogue and reads the relevant information on every QSO, listed in Tab. 3.1. For each QSO, the parameters read from the QSO catalogue are forwarded to the part of the code concerned with generating the spectrum.

2. Before generating the spectrum, QSOSIM10 uses the wavelength information on the first resolution element of the real SDSS spectrum that contains data and the QSO redshift to determine the number of H I absorption lines in the Lyman α forest, n, as in Sec. 3.3.2.

3. QSOSIM10 then generates the cumulative distribution functions of lines in column density and in redshift (described in Sec. 3.3.1 and Sec. 3.3.2, respectively), models the clustering of lines along the line-of-sight (described in Sec. 3.3.3), and computes the column density scaling factor for lines in the proximity of the QSO (described in Sec. 3.3.6).

4. Using the results of the previous three steps, QSOSIM10 uses Monte Carlo methods to sample the two CDFs generated and the b-parameter distribution, returning the redshift, column density, and b-parameter values for each of the n lines. In other words, this step uses Monte Carlo methods to create the line list of H I absorbers.

5. This step deals with the generation the spectrum of a particular QSO. A modified version of the QSOSIM9 program takes the information on the properties of the QSO and the H I line list to generate the spectrum. Following the QSOSIM9 template, this subroutine first calculates the flux at $\lambda = 6182$ Å by knowing the r magnitude of the QSO, and models the synchrotron emission (see Sec. 3.1). The subroutine then adds the broad and narrow QSO emission lines to the spectrum (see Sec. 3.2). Completing this, the subroutine uses the line list generated in the previous step to input H I absorption. If the column density of an absorber is larger than the metal threshold, the subroutine assigns it a metallicity value according to Eq. 3.30 and searches for the point in

the metallicity database that is closest in the (Z, $N_{\rm HI}$,z) parameter space (for a more detailed description, see Sec. 3.3.5). Metal absorption is added for every such system, and the parameters of all metal absorption lines introduced are saved in the line list. Using the information on the characteristics of the BOSS spectrographs provided by the SDSS Collaboration, the spectrum is convolved to the constant pixel size in velocity space, 69 km s^{-1} [24]. Finally, the subroutine introduces the BOSS specific noise into the spectrum, as described in Sec. 3.4. The generated QSO spectrum is returned into the main program, together with the absorption line list, and every intermediate step of the spectrum generation (i.e the flux with no noise introduced, the unabsorbed flux, the unconvolved flux, the inverse of the flux variance, and the noise).

6. In the final step, a FITS file containing all relevant data on the QSO spectrum is created. The FITS file contains the information provided in the SDSS QSO catalogue, the QSO spectrum, and the line list. The contents of the FITS file are described in more detail in the introductory section of Chapter 3.

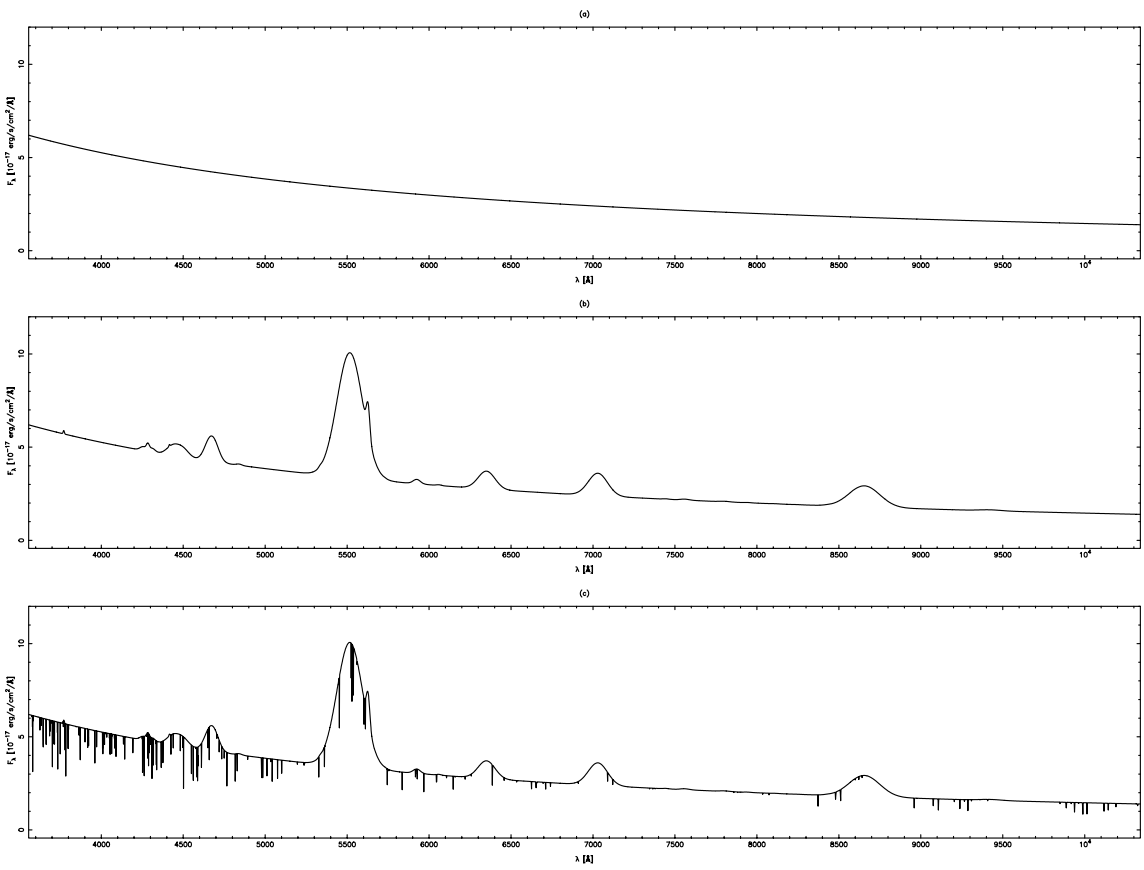

Figure 4.1: An example of the spectrum generation for the QSO SDSS000021.47+062215.7. The QSO is at redshift $z_{QSO} = 3.5383$, with a magnitude $r = 19.99$, and spectral index $\alpha = 0.6$. The three panels shown illustrate the modelling of the underlying continuum emission from synchrotron radiation (top), the input of the QSO broad and narrow emission lines (middle), and the absorption by metals in the IGM (bottom).

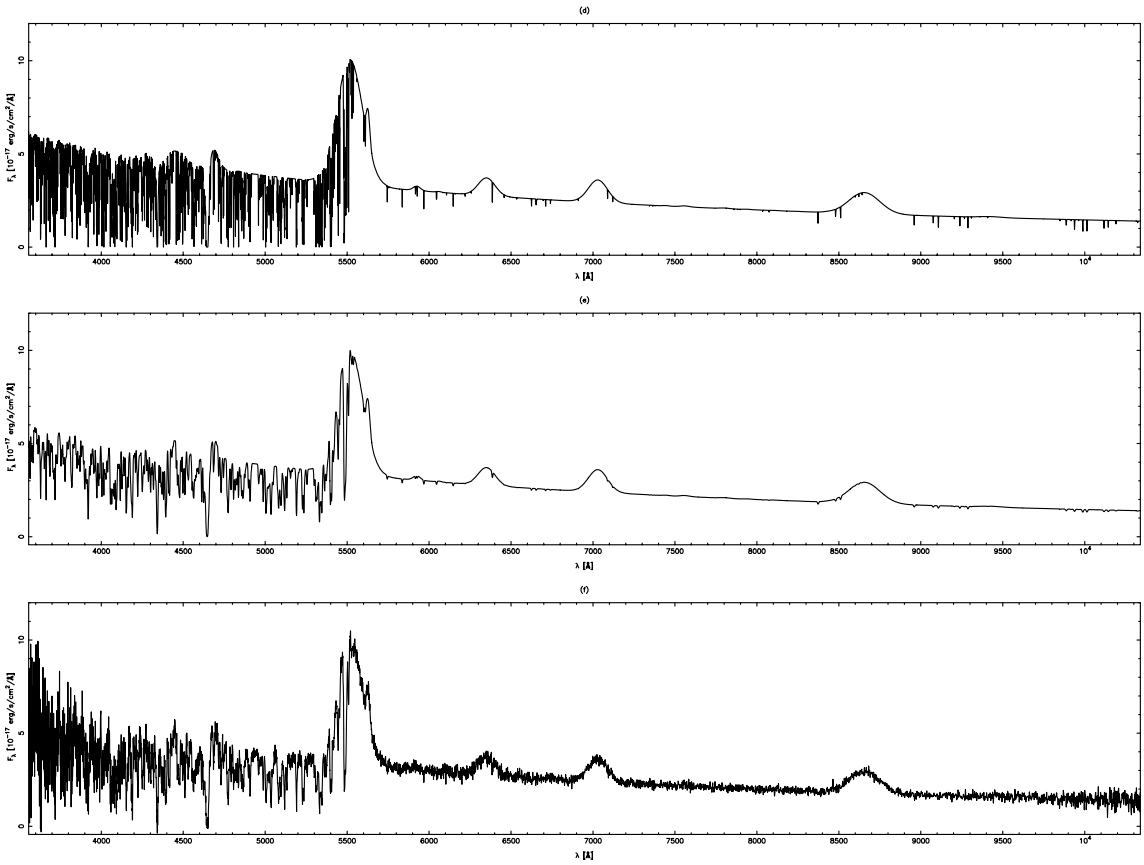

Figure 4.2: An example of the spectrum generation for the QSO SDSS000021.47+062215.7. The QSO is at redshift $z_{QSO} = 3.5383$, with a magnitude $r = 19.99$, and spectral index $\alpha = 0.6$. The three panels shown illustrate the modelling of H I absorption in the Lyman α forest (top), convolution of the spectrum to the BOSS resolution (middle), and introduction of the noise characteristics of BOSS spectrographs (bottom).

Fig. 4.1 and Fig. 4.2 show the six most important steps in the process of spectrum generation. The QSO in the figure has a redshift of $z_{QSO} = 3.5383$, is relatively faint with an r magnitude $r = 19.99$, and has the spectral index $\alpha = 0.6$. The steps in the spectrum generation process of a single QSO are described in the fifth point, above. Panel (a) shows the underlying synchrotron emission, while panel (b) shows the spectrum after broad and narrow QSO emission lines were included. Panel (c) shows only the absorption by metals associated with systems with high H I column density. Absorption by both metals and the H I is shown in panel (d). Although metals are input in the spectrum in parallel with H I, absorption by metals was plotted in panel (c) to clearly illustrate the significance of metals in the Lyman α forest. The convolved QSO spectrum is shown in panel (e). Convolution irreversibly changes the spectrum, blends together absorption lines in the forest and smears out metal absorption redwards of the Ly α transition, making information contained in the spectrum much more difficult to extract. The final panel, (f), shows the QSO spectrum after the introduction of BOSS specific noise.

We show a visual comparison of ten pairs of simulated and real QSO spectra below. Visual inspection of the spectra shows a very good agreement between the two. The simulated spectra follow the features seen redwards of the Ly α transition in real spectra, such as the strengths of the emission lines, their widths and strengths. The noise in the simulated spectra resembles that seen in the real spectra across a large range in wavelengths, with one noticeable exception: the simulated noise at $\lambda \gtrsim 7200$ Å does not contain telluric lines seen in real spectra, as QSOSIM10 does not model telluric absorption. As QSOSIM10 uses the average values of the line strengths and widths from composite spectra to model the emission lines, it sometimes fails in reproducing the strong narrow peak of Ly α present in real spectra. An example of this is shown in Fig. 4.3, below.

Several tests can be made to quantify how well the simulated spectra emulate the spectra of real QSOs as seen by SDSS. One way to do this would be to use simulated spectra to perform the same spectral analysis that is used to study the optical depth of the neutral hydrogen in real spectra. Eventual consistency between the two results would suggest that the IGM in the simulated spectra has the same properties as the real IGM. Other tests include comparing composite spectra made from simulated and real QSO spectra, or comparing equivalent widths of the broad emission lines of the simulated and real spectra. However, none of the proposed tests for the quantitative comparison of the simulated and real QSO spectra proposed have been done in time to be included into this thesis.

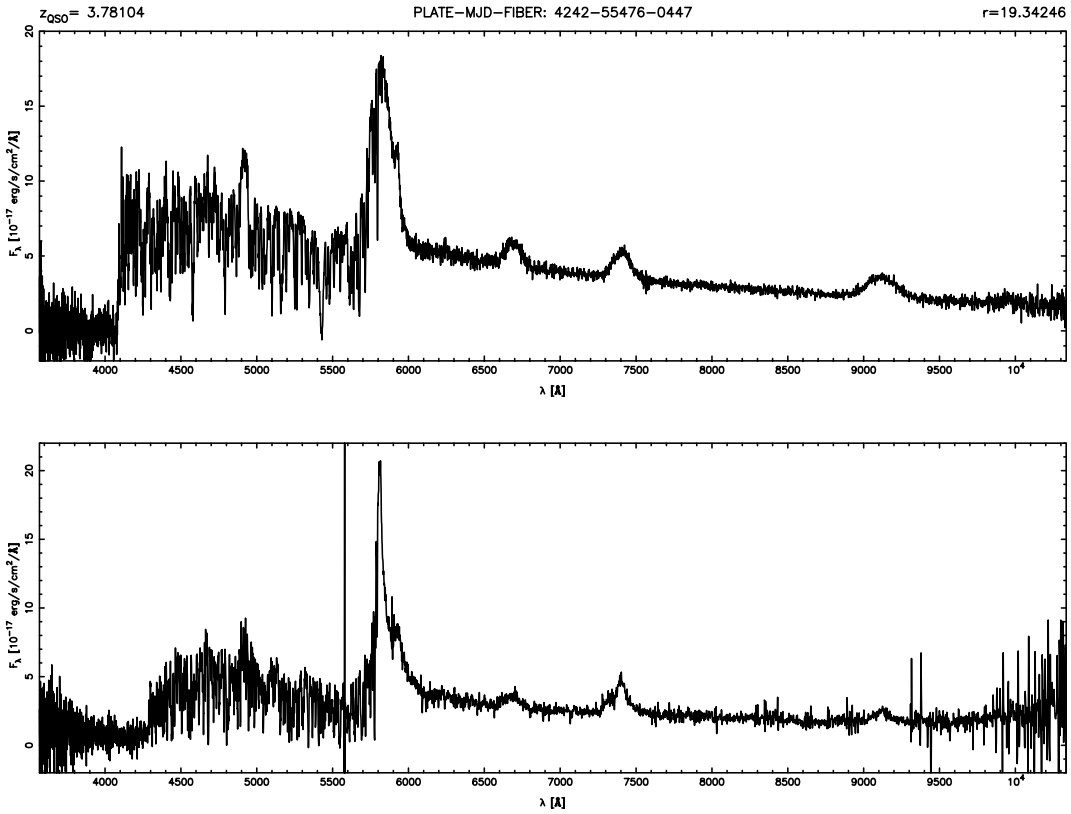

Figure 4.3: Comparison of simulated (top) and real (bottom) spectra for the QSO SDSS025112.44-005208.1, at a redshift $z = 3.78104$ and with a magnitude $r = 19.34246$.

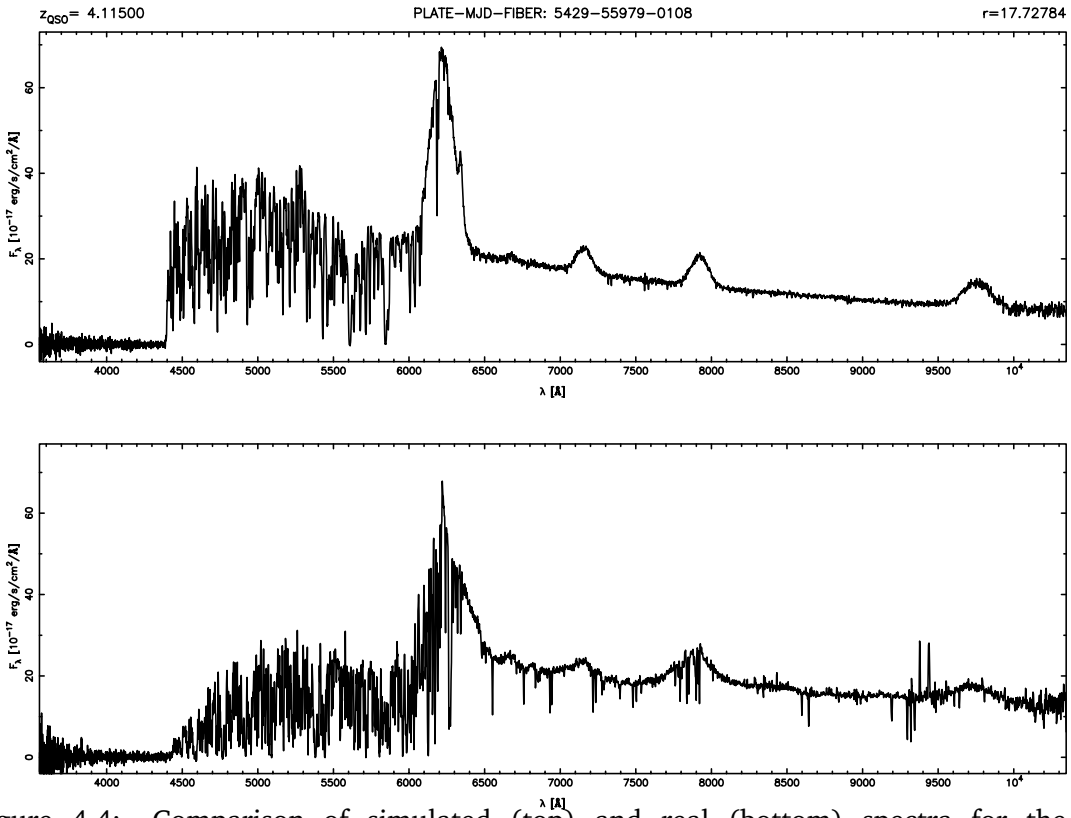

Figure 4.4: Comparison of simulated (top) and real (bottom) spectra for the QSO SDSS132611.85+074358.3, at a redshift $z = 4.11500$ and with a magnitude $r = 17.72784$.

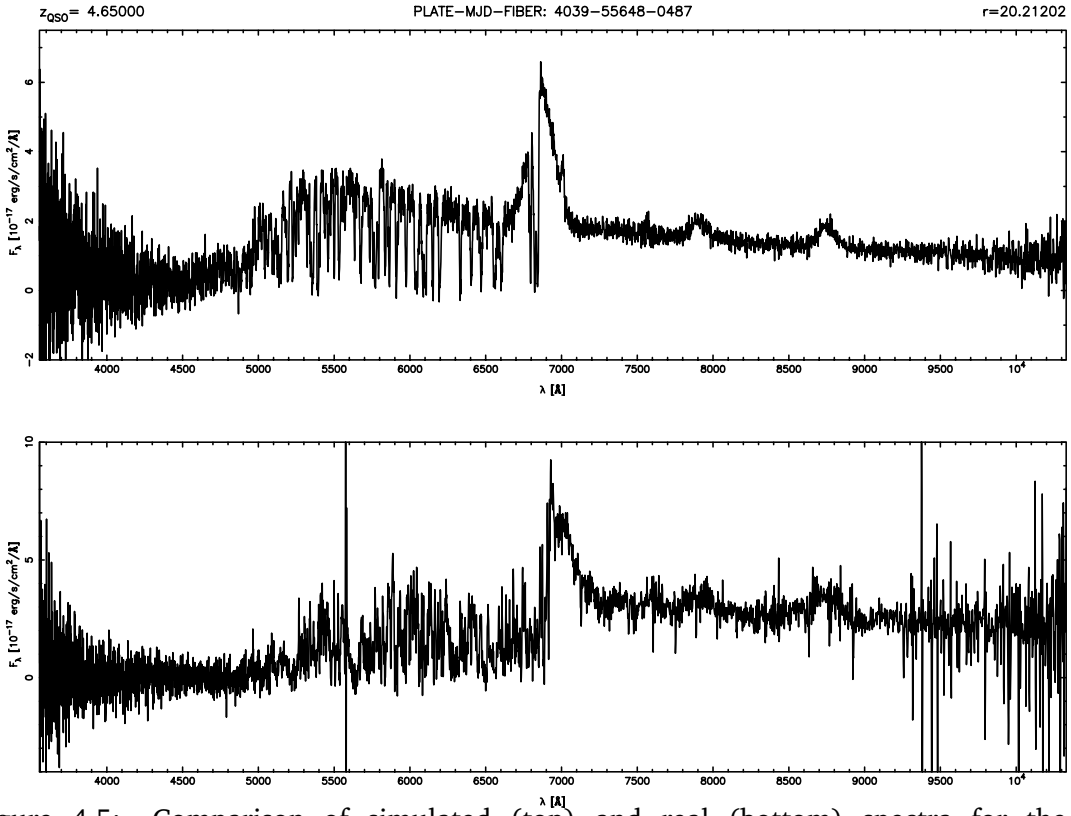

Figure 4.5: Comparison of simulated (top) and real (bottom) spectra for the QSO SDSS134134.19+014157.6, at a redshift $z = 4.56000$ and with a magnitude $r = 20.21202$.

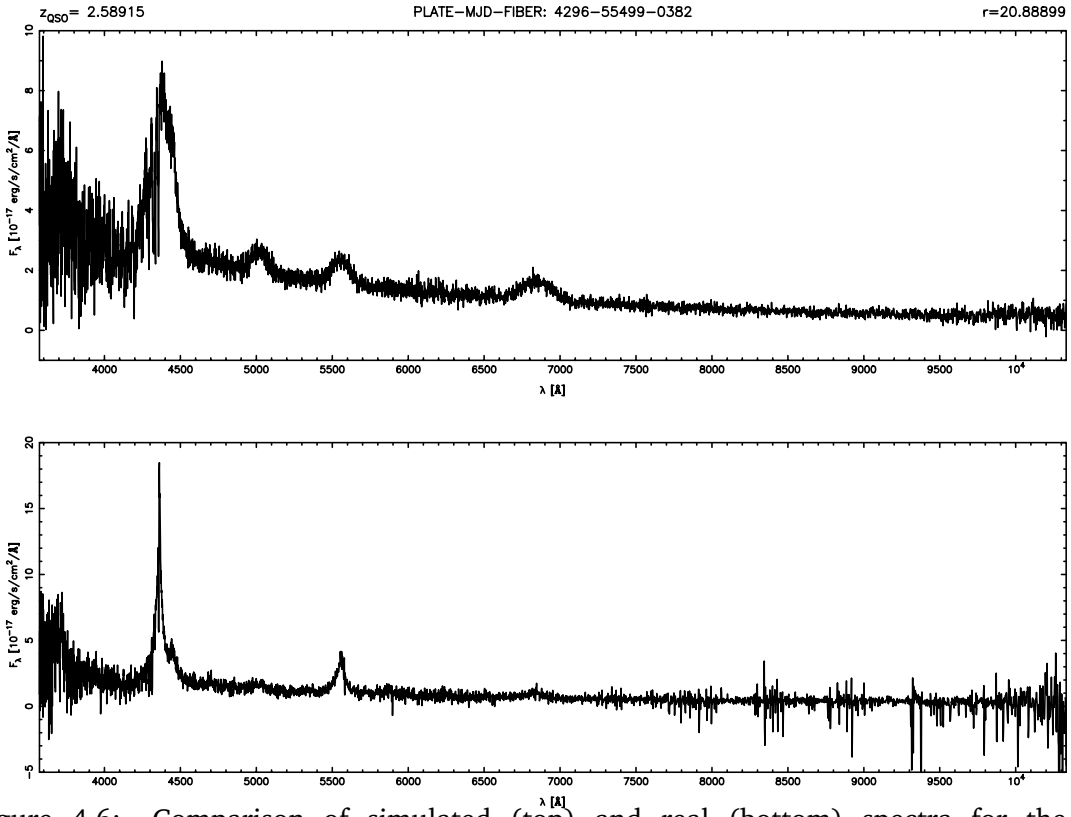

Figure 4.6: Comparison of simulated (top) and real (bottom) spectra for the QSO SDSS000027.31+013126.1, at a redshift $z = 2.58915$ and with a magnitude $r = 20.88899$.

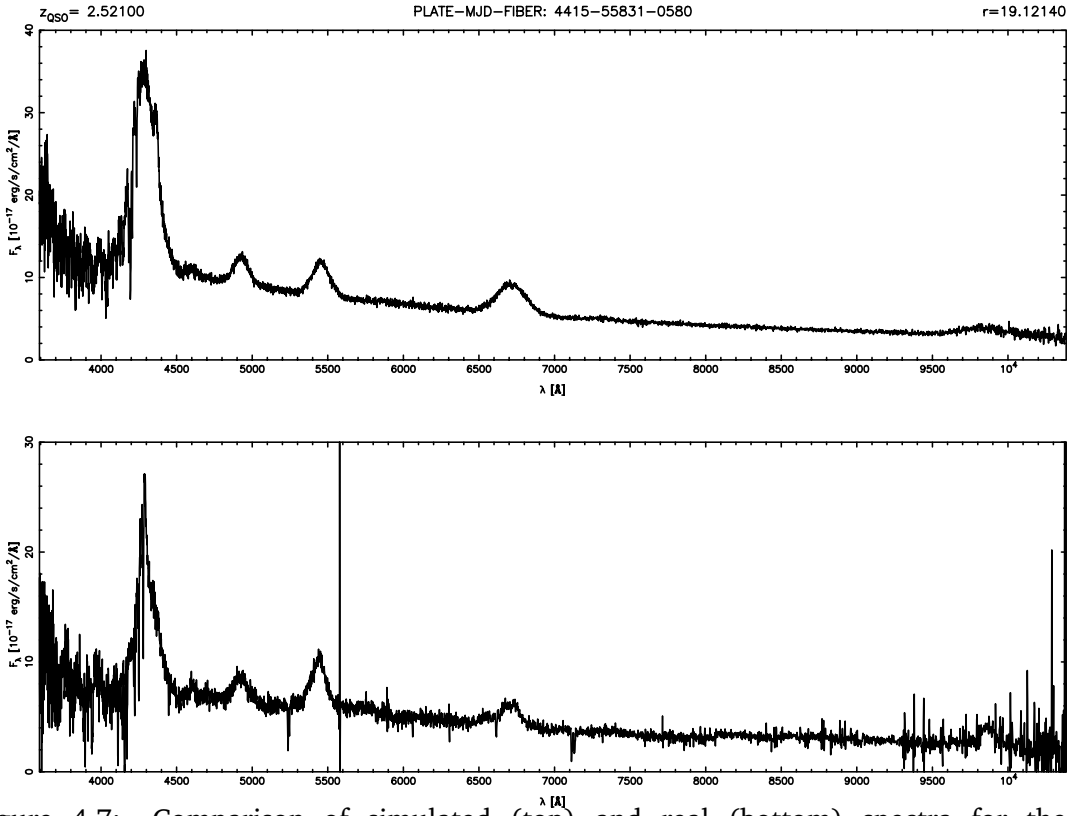

Figure 4.7: Comparison of simulated (top) and real (bottom) spectra for the QSO SDSS000021.29+052914.7, at a redshift $z = 2.52100$ and with a magnitude $r = 19.12140$.

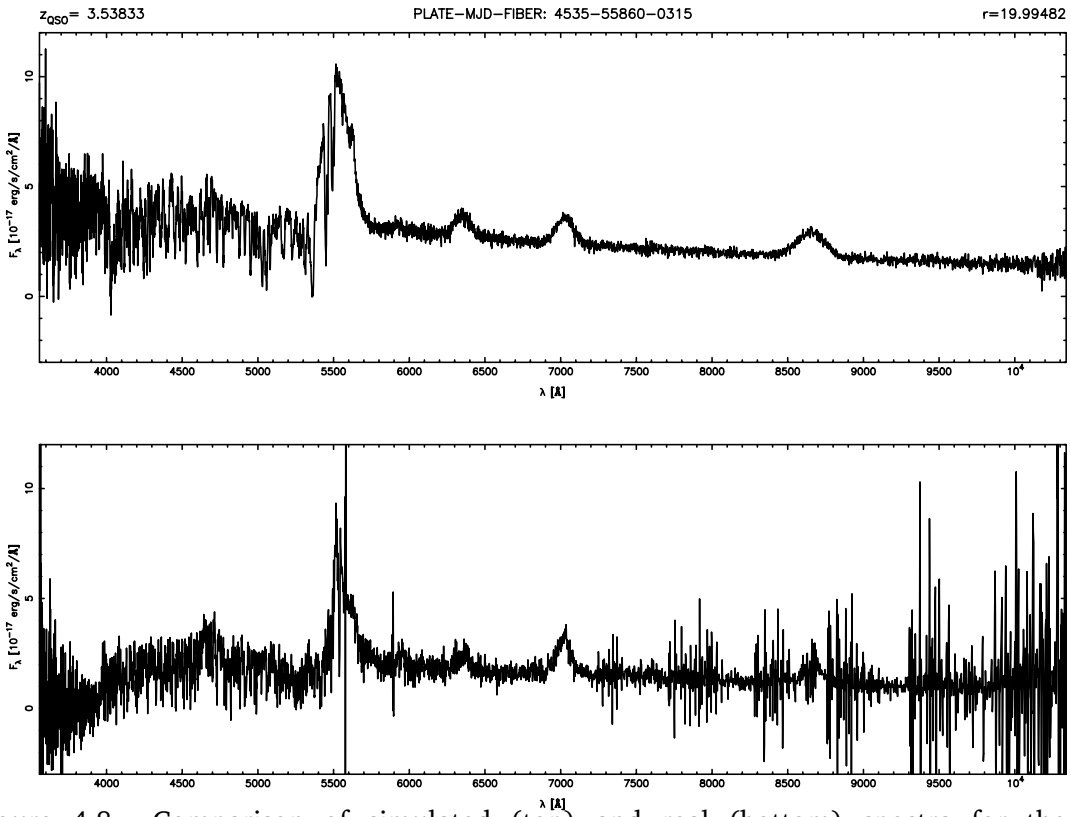

Figure 4.8: Comparison of simulated (top) and real (bottom) spectra for the QSO SDSS000021.47+062215.7, at a redshift $z = 3.53833$ and with a magnitude $r = 19.99482$.

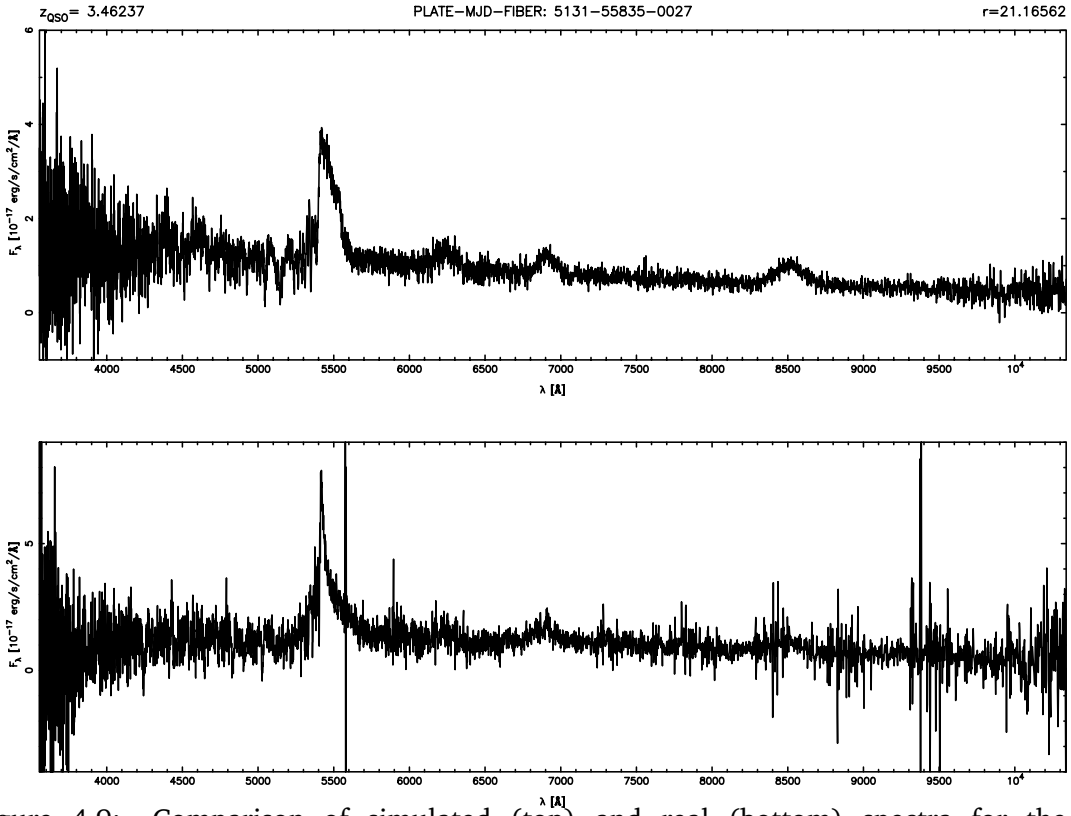

Figure 4.9: Comparison of simulated (top) and real (bottom) spectra for the QSO SDSS011208.38+150122.8, at a redshift $z = 3.46237$ and with a magnitude $r = 21.16562$.

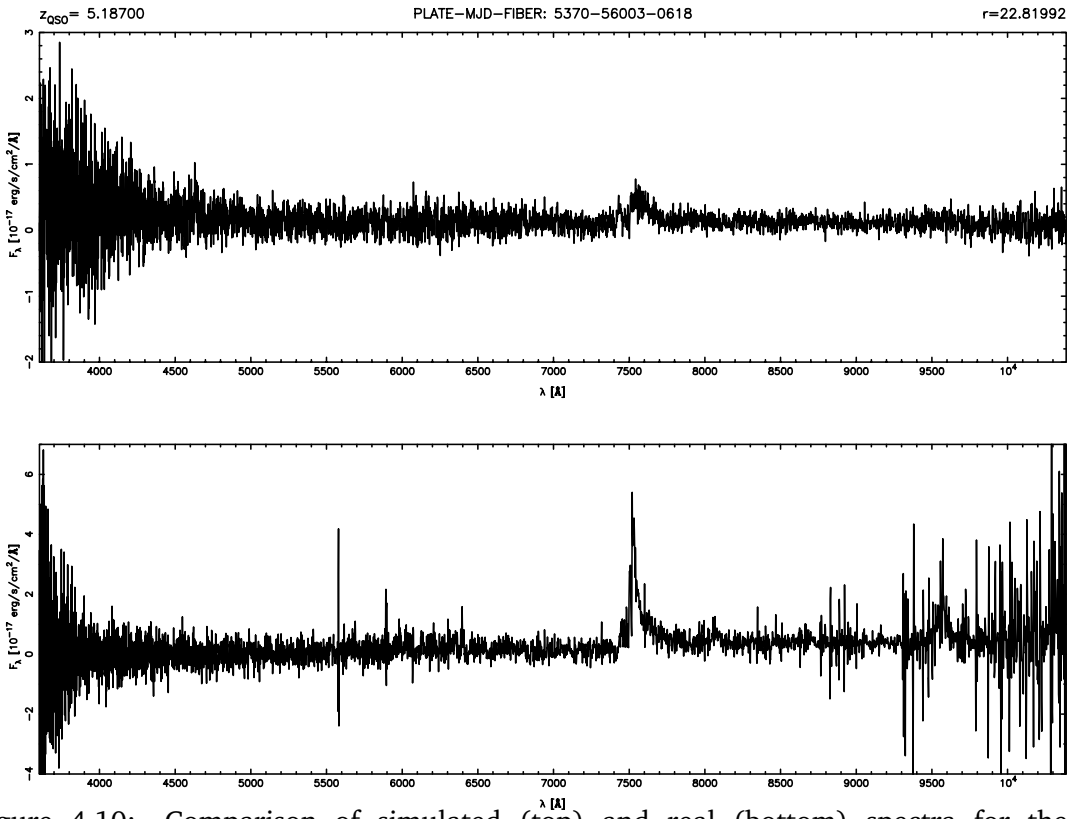

Figure 4.10: Comparison of simulated (top) and real (bottom) spectra for the QSO SDSS112250.04+134236.4, at a redshift $z = 5.18700$ and with a magnitude $r = 22.81992$. This is the highest redshift QSO in the simulated sample.

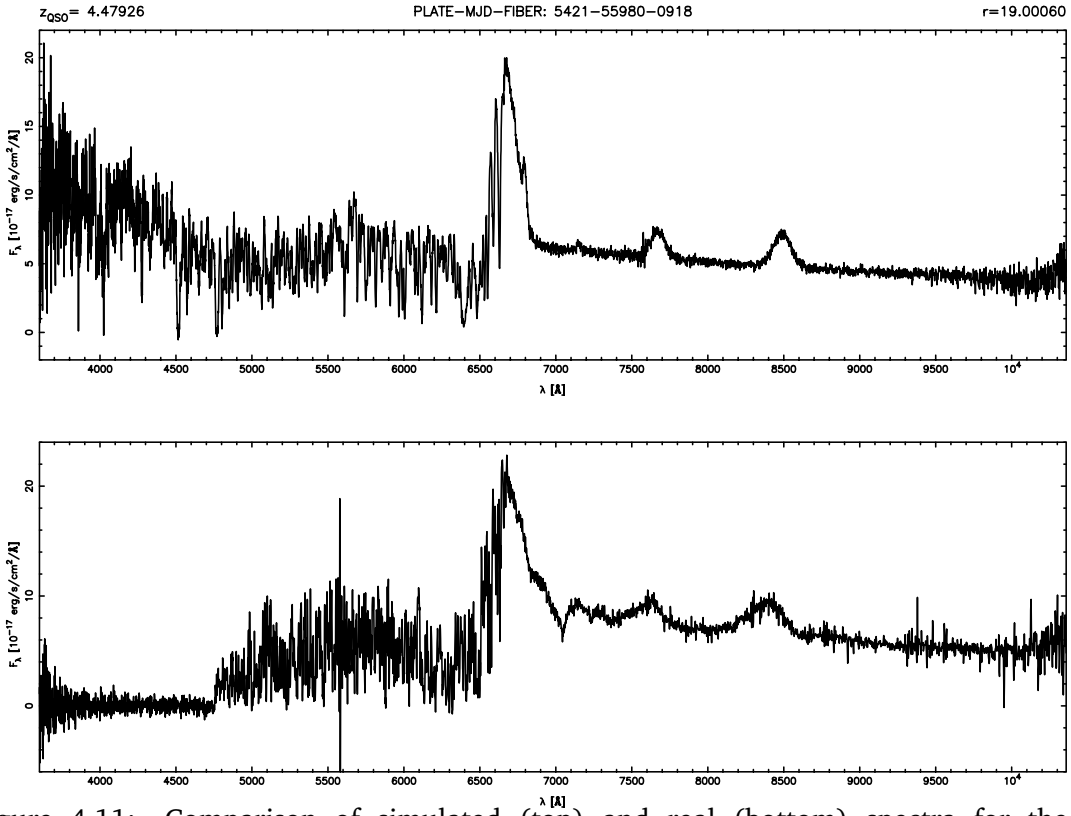

Figure 4.11: Comparison of simulated (top) and real (bottom) spectra for the QSO SDSS131056.86+105530.1, at a redshift $z = 4.47926$ and with a magnitude $r = 19.00060$. The real spectrum exhibits a DLA system that is not present in its simulated counterpart.

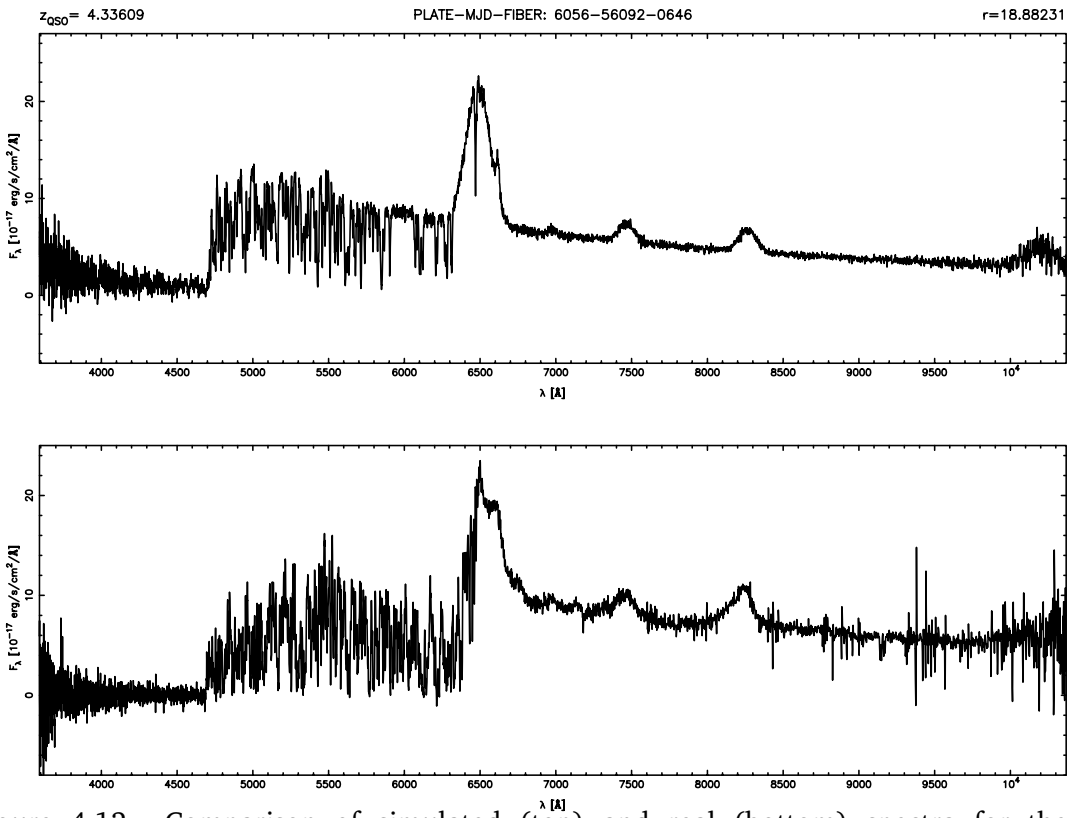

Figure 4.12: Comparison of simulated (top) and real (bottom) spectra for the QSO SDSS141831.70+444937.6, at a redshift $z = 4.33609$ and with a magnitude $r = 18.88231$.

5 Conclusions & future prospects

5.1 Summary

This thesis has presented the construction of QSOSIM10 and its output – the most comprehensive and sophisticated database of simulated SDSS QSO spectra. QSOSIM10 improves upon the previously available code by incorporating physics based on the most accurate and recent observational results available, such as the statistical properties of QSO spectra derived from composite spectra, the properties of the IGM derived from modern high-resolution echelle spectrographs, data on the real QSOs from the SDSS BOSS DR10 QSO catalogue and the technical characteristics of SDSS BOSS. This information in used to model the spectral characteristics seen in the real QSO spectra, including: a set of 59 broad and narrow QSO emission lines, absorption by the intergalactic clouds in the Lyman α forest, the clustering of the IGM along the sightline, the QSO proximity effect, and absorption by metals associated with intergalactic clouds of high densities. Most of these effects were not included in the previous versions of the QSOSIM program.

5.2 Applications & future prospects

The simulated spectra will have their first application in a study looking to constrain any large-scale anisotropies in the Universe. The project uses SDSS DR12 QSO spectra to look for variations in the mean optical depth of the Universe, $\tau_{\rm eff}$, across all redshifts and all available sightlines. The optical depth of Lyman α absorption in the IGM at redshift z is given by:

$$\tau(z) = \frac{\sigma_0 \; c \; x \; n_{\rm H}}{H(z)}, \tag{5.1}$$

where σ_0 is the scattering cross-section of the Ly α transition, $x = n_{\rm HI}/n_{\rm H}$ the neutral fraction, and $n_{\rm H}$ the hydrogen number density [79]. The optical depth of the intergalactic gas consisting primarily of ionised hydrogen and singly ionised helium, which is at density $(1+\delta) \equiv \rho/\langle\rho\rangle$ relative to the average background density $\langle\rho\rangle$ and is in photoionisation equilibrium at redshift z, is:

$$\tau(z) \approx 0.7 \left(\frac{\Omega_b \; h^2}{0.019}\right)^2 \left(\frac{\Omega_m \; h^2}{0.3 \times 0.65^2}\right)^{-1/2} \left(\frac{1+z}{4}\right)^{4.5} \left(\frac{T}{10^4 \text{ K}}\right)^{-0.7} \times \left(\frac{\Gamma}{10^{-12} \text{ s}^{-1}}\right)^{-1} \frac{(1-Y)}{0.76} \frac{(1-Y/4)}{0.94} (1+\delta)^2, \tag{5.2}$$

where Ω_b is the baryon density, $H_0 = 100h$ km s^{-1} Mpc^{-1} is the Hubble's constant, Ω_m is the matter density, T the local gas temperature, Γ is the photoionisation rate, and Y is the helium abundance by mass ([123], but see [124] for the derivation). The optical depth measurements can therefore be used to study the photoionisation rate evolution (with the temperature of the IGM constrained by the line width distribution), the ratio of the baryon-to-matter density, and other parameters in the equation.

The mean transmittance of the IGM at an absorption redshift z_{abs}, $\langle F \rangle(z_{abs})$, is defined as the average fractional loss of the QSO flux at the corresponding absorption wavelength[9] λ, caused by the absorption by the IGM:

$$\langle F \rangle(z_{abs}) = \left\langle \frac{f_\lambda^{observed}}{f_\lambda^{continuum}} \right\rangle. \tag{5.3}$$

This number depends on the average number of intergalactic clouds at redshift z_{abs}, as well as their column densities and line widths. The mean transmittance is used to define the effective optical depth at the absorption redshift, $\tau_{\rm eff}(z_{abs})$, via:

$$\exp\left[-\tau_{\rm eff}(z_{abs})\right] \equiv \langle F \rangle = \langle \exp\left[-\tau(z_{abs})\right] \rangle. \tag{5.4}$$

The relation that connects $\tau_{\rm eff}$ at redshift z_{abs} and the observable statistical properties of the IGM is:

$$\tau_{\rm eff}(z_{abs}) = B(1+z_{abs})^{\gamma+1}, \tag{5.5}$$

where γ is the evolution parameter, and

$$B = \frac{1}{\lambda_{\rm Ly\alpha}} \int_0^\infty \int_0^\infty F(N,b) W(N,b)\, {\rm d}N\, {\rm d}b. \tag{5.6}$$

B is dependant on the column density and line width distributions of absorption lines in the Lyman α forest at redshift z_{abs} ($F(N,b)$) and the absorption properties of the IGM ($W(N,b)$) [125]. These distributions are derived from high-resolution QSO spectra (see [125], Chapter 3 and the references cited). More details on this method can be found in [123], [125]–[127] and the references therein. Using of a large number of QSO sightlines, we can study the evolution of the optical depth of the Universe.

The preliminary results of our study are shown in Fig. 5.1. This is the most accurate measurement of the optical depth produced so far, utilising more than 155 000 QSO spectra from the SDSS DR12. The size of the sample ensures that a large num-

[9]The absorption wavelength is the wavelength of the Ly α absorption at the absorption redshift, $\lambda = (1+z_{abs})\lambda_{\rm Ly\alpha}$.

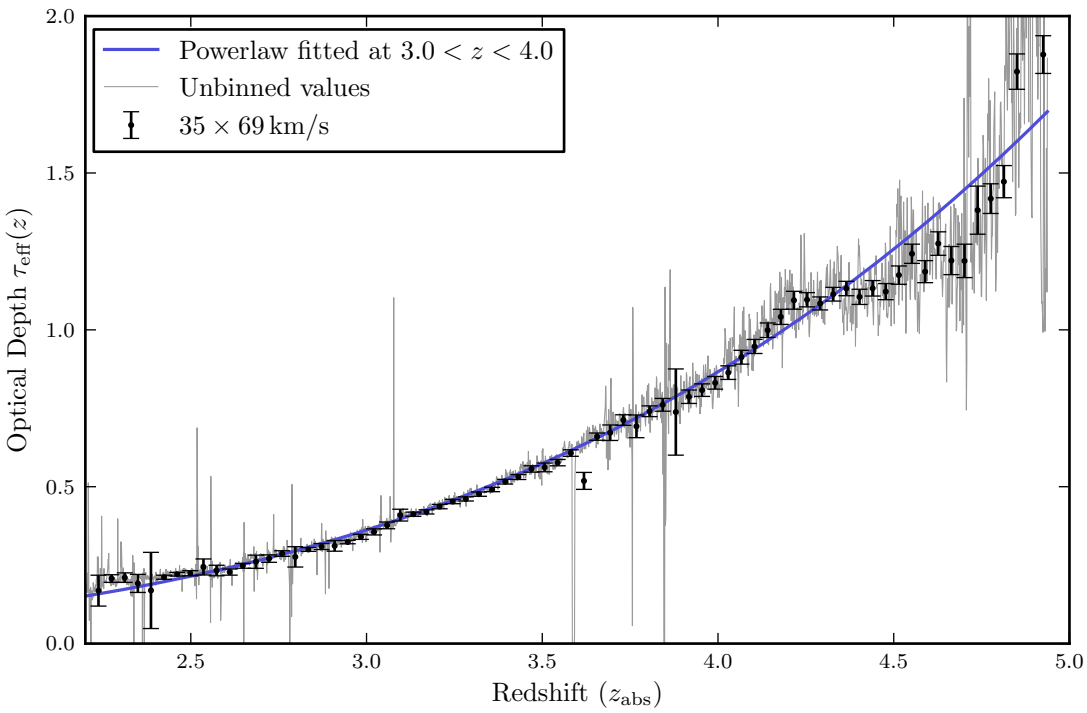

Figure 5.1: Preliminary results of the optical depth evolution of the Lyman α forest. The grey line connects the τ_{eff} values for every SDSS resolution element ('pixel'), while the black dots represent the weighted mean τ_{eff} in redshift bins covered by 35 pixels and the corresponding errors. The power law of Eq. 5.5 was fitted to the unbinned data in the redshift range $3.0 < z < 4.0$. The mean optical depth was calculated using more than 155 000 QSOs from the SDSS DR12. Figure provided by Sebastian Schmidt (private correspondence).

ber of QSOs contribute to every redshift bin, resulting in high result confidence. There are several previously unreported features seen above redshift $z > 4$, which – if real – reflect important cosmological discoveries and will ultimately yield new and fundamental information about ionisation processes in the early Universe. If the departures from the predictions of Eq. 5.5 are caused by the change in the photoionisation rate of hydrogen, this would give stringent constraints on the evolution of the cosmic ultraviolet background and the evolution of structure formation. A careful consideration of the possible artefacts introduced by the methods used in the analysis, and the systematic effects present in the observed data is required to verify these results. QSOSIM10 will permit an assessment of the systematics and hopefully shed light on the nature of the observed features.

The spectra will be used to resolve the effect of under-sampling caused by an overlap of the QSO and stellar populations in the colour-colour space used to select QSOs for the spectroscopic survey, as well as to quantify the effect of the varying selection criteria during the first two years of SDSS BOSS, and improve the flux estimation in the Lyman α forest. The values of SDSS photometric magnitudes will be derived from the simulated spectra, and used to generate a mock colour-colour

space used by SDSS to select QSO candidates. Applying the same criteria to the simulated data as were applied to the real data, we will be able to quantify the effects the selection criteria have on the study of the mean optical depth of the Universe.

With the release of the DR12, the final data release of SDSS-III, the SDSS QSO sample counts more than 294 000 QSO spectra distributed over an area covering more than 10 000 sq. deg. of the sky. This is the largest set of medium-resolution QSO spectra available to date and is suitable for a wide range of scientific pursuits. The simulated database and the QSOSIM10 itself can be used in conjunction with the observed data to provide quantitative checks on the methods used and help interpret results in all studies utilising SDSS data.

Comparison with simulations can also be used to improve a broader range of studies, including those of the large-scale structure, the equation of state of the intergalactic medium, primordial magnetic fields, evolution of the cosmic metallicity and star-formation, the ultraviolet background, and many more. The European Space Agency's satellite Gaia [128] recently started its mission to obtain over 500 000 low- and high-resolution optical QSO spectra. QSOSIM10 can be modified and used to quantify systematic effects of the Gaia sample and maximise its scientific output.

One of the next steps in the development of QSOSIM10 will be to increase its versatility by increasing the resolution of the simulated spectra, surpassing the resolution obtained by both the Keck and the VLT telescopes. This improvement will make QSOSIM10 usable in many studies utilising high-resolution QSO spectra.

We make the database of simulated QSO spectra publicly available and free to use in hope that it will help yield some good science.

Appendices

Appendix A Table of emission lines

Table A.1: Table of emission lines included in QSOSIM10.

Line ID	λ [Å]	σ [Å]	Rel. str.	Line ID	λ [Å]	σ [Å]	Rel. str.
O VI	722.00	3.351	6.3	Fe II	1787.00	6.060	0.229
O III	831.00	0.957	1.6	Si II	1816.98	5.72	0.13
S VI	944.00	2.234	3.2	Al III	1857.40	4.950	0.333
Ly γ	973.00	0.798	1.1	Si III]	1892.03	3.09	0.158
Ly η	937.80	4.730	2.11	Fe III	1892.64	3.09	0.158
Ly δ	949.74	4.730	2.11	Fe III	1905.97	23.58	15.943
C III	977.02	8.950	5.195	C III]	1908.73	23.58	15.943
N III	990.69	8.950	5.195	Fe III	2076.62	16.99	1.58
Ly β	1025.72	7.760	9.615	C II]	2326.44	4.950	0.183
O VI	1033.83	7.760	9.615	Ne [IV]	2423.93	8.420	0.437
Ar I	1066.66	4.179	0.816	O [II]	2471.03	4.54	0.092
C III	1175.70	2.280	0.870	Al II]	2669.95	5.1	0.067
Ly α	1215.67	19.46	100.0	Mg II	2798.75	34.95	14.725
N V	1240.14	2.71	24.61	Fe II	2964.28	22.92	2.017
Si II	1262.59	2.74	0.306	O III	3133.70	9.38	0.326
O I	1304.34	5.42	1.992	He I	3188.67	12.77	0.445
S II	1306.82	5.42	1.992	Fe I	3261.40	3.270	0.032
C II	1335.30	3.86	0.688	Ne [V]	3346.82	5.50	0.118
Si IV	1396.76	12.5	8.916	Ne [V]	3426.84	9.09	0.405
O IV]	1402.06	12.5	8.916	Fe II	3498.92	16.79	0.432
C IV	1549.06	14.33	25.291	Fe VII	3581.70	7.98	0.10
He II	1640.42	4.430	0.521	He I	3588.30	7.98	0.10
O III]	1663.48	5.5	0.48	O [II]	3728.48	3.32	0.424
Fe II	1664.74	5.5	0.48	Ne [III]	3869.85	5.32	0.346
Al II	1670.79	5.5	0.48	Fe [V]	4072.39	3.20	0.039
Fe II	1716.88	7.36	0.258	Fe II	4564.71	61.69	3.757
N IV	1718.55	7.36	0.258	O [III]	4960.30	3.85	0.686
Al II	1721.89	7.36	0.258	O [III]	5008.24	6.04	2.49
N III]	1750.26	5.120	0.382	Fe II	5305.97	74.83	3.522

Appendix B An example of a CLOUDY input file

We used CLOUDY to model the metal absorption in the IGM. The metal absorption input algorithms and database construction are explained in detail in Sec. 3.3.5. Here we present an example of an input file for an absorption system at redshift $z = 1.9$, with H I column density $N_{\rm HI} = 10^{22}$ cm^{-2}, and metallicity $Z = -2.84$, that was used by CLOUDY to generate the database containing metal column densities.

```
c Lyman alpha system for QSOSIM10
cmb z=1.900
table hm05 z=1.900
hden -2
metals -2.84
stop neutral column density 22.00
double
iterate to convergence
print line faint -1
save overview last "s_Z_-2.84_NHI_22_z_1.90.ovr"
save results last "s_Z_-2.84_NHI_22_z_1.90.rlt"
```

The first two lines specify the sources of the cosmic background radiation. The '`cmb`' command specifies to include the contribution from the cosmic microwave background at redshift z, while the '`table hm05`' tells CLOUDY to include the Haardt & Madau model of the UVB, for the same redshift. The next three lines are related to the physical properties of the H I cloud. The '`hden`' command specifies the total hydrogen volumetric density of the system, and the '`metals`' command specifies its metallicity. Both of those quantities are expressed in comparison to Solar, on a logarithmic scale. The '`stop neutral column density`' tells CLOUDY to stop with the execution of the program once the logarithm (base 10) of H I column density reaches the specified number. The last five commands are not related to any physical properties in the cloud, but are related to the execution of the program. Further details on the workings of CLOUDY can be found in the CLOUDY documentation on the official webpage [129].

References

[1] J. C. Mather, 'The Cosmic Background Explorer (COBE)', *Optical Engineering*, vol. 21, no. 4, p. 214 769, 1982.

[2] C. L. Bennett *et al.*, 'The Microwave Anisotropy Probe Mission', *The Astrophysical Journal*, vol. 583, pp. 1–23, 2003. eprint: `astro-ph/0301158`.

[3] Planck Collaboration, 'Planck 2013 results. XVI. Cosmological parameters', *Astronomy & Astrophysics*, vol. 571, A16, A16, 2014.

[4] H. A. Feldman, R. Watkins and M. J. Hudson, 'Cosmic flows on 100 h^{-1} Mpc scales: standardized minimum variance bulk flow, shear and octupole moments', *Monthly Notices of the Royal Astronomical Society*, vol. 407, no. 4, pp. 2328–2338, 2010.

[5] J. K. Webb *et al.*, 'Further Evidence for Cosmological Evolution of the Fine Structure Constant', *Physical Review Letters*, vol. 87, no. 9, 091301, p. 091 301, 2001. eprint: `astro-ph/0012539`.

[6] J. A. King *et al.*, 'Spatial variation in the fine-structure constant - new results from VLT/UVES', *Monthly Notices of the Royal Astronomical Society*, vol. 422, pp. 3370–3414, 2012. arXiv: `1202.4758 [astro-ph.CO]`.

[7] J. E. Gunn *et al.*, 'The 2.5 m Telescope of the Sloan Digital Sky Survey', *The Astronomical Journal*, vol. 131, no. 4, pp. 2332–2359, 2006.

[8] S. A. Smee *et al.*, 'The Multi-object, Fiber-fed Spectrographs for the Sloan Digital Sky Survey and the Baryon Oscillation Spectroscopic Survey', *The Astrophysical Journal*, vol. 146, p. 32, 2013.

[9] G. T. Richards *et al.*, 'Spectroscopic Target Selection in the Sloan Digital Sky Survey: The Quasar Sample', *arXiv.org*, no. 6, pp. 2945–2975, 2002. arXiv: `astro-ph/0202251v1 [astro-ph]`.

[10] G. T. Richards *et al.*, 'The Sloan Digital Sky Survey Quasar Survey: Quasar Luminosity Function from Data Release 3', *The Astronomical Journal*, vol. 131, pp. 2766–2787, 2006. eprint: `astro-ph/0601434`.

[11] N. P. Ross *et al.*, 'The SDSS-III Baryon Oscillation Spectroscopic Survey: Quasar Target Selection for Data Release Nine', *The Astrophysical Journal Supplement Series*, vol. 199, no. 1, p. 3, 2012.

[12] J. K. Webb, *Notes on the program QSOSIM9*, 2014.

[13] I. Paris *et al.*, 'The Sloan Digital Sky Survey quasar catalog: tenth data release', *Astronomy and Astrophysics*, vol. 563, p. 54, 2014.

[14] M. Fukugita *et al.*, 'The Sloan Digital Sky Survey Photometric System', *The Astronomical Journal*, vol. 111, p. 30, 1996.

[15] T.-S. Kim, S. Cristiani and S. D'Odorico, 'The Lyman-Alpha Forest at $1.5 < z < 4$', in *The Physics of Galaxy Formation, ASP Conference Proceedings*, Masayuki, Umemura and Hajime, Susa, Ed., vol. 222, 2001.

[16] A. Songaila and L. L. Cowie, 'The Evolution of Lyman Limit Absorption Systems to Redshift Six', *The Astrophysical Journal*, vol. 721, pp. 1448–1466, 2010. arXiv: `1007.3262 [astro-ph.CO]`.

[17] T.-S. Kim *et al.*, 'The evolution of H I and C IV quasar absorption line systems at $1.9<z<3.2$', *Astronomy & Astrophysics*, vol. 552, A77, 2013.

[18] X. Barcons and J. K. Webb, 'The effect of clustering on the equivalent width distribution of QSO Lyman-alpha clouds', *Monthly Notices of the Royal Astronomical Society*, vol. 253, pp. 207–211, 1991.

[19] C. W. Danforth *et al.*, 'An HST/COS Survey of the Low-Redshift IGM. I. Survey, Methodology, & Overall Results', *arXiv.org*, p. 2655, 2014. arXiv: `1402.2655 [astro-ph.CO]`.

[20] G. J. Ferland *et al.*, 'The 2013 Release of Cloudy', *Revista Mexicana de Astronomia y Astrofisica*, vol. 49, pp. 137–163, 2013.

[21] R. F. Carswell *et al.*, 'Observations of the spectra of Q0122-380 and Q1101-264', *Monthly Notices of the Royal Astronomical Society*, vol. 198, pp. 91–110, 1982.

[22] F. Haardt and P. Madau, 'Radiative Transfer in a Clumpy Universe. IV. New Synthesis Models of the Cosmic UV/X-Ray Background', *The Astrophysical Journal*, vol. 746, no. 2, p. 125, 2012.

[23] The UV background model from CUBA, `http://www.ucolick.org/~pmadau/CUBA/HOME.html` (visited on 21/01/2015).

[24] A. S. Bolton *et al.*, 'Spectral Classification and Redshift Measurement for the SDSS-III Baryon Oscillation Spectroscopic Survey', *The Astronomical Journal*, vol. 144, 144, p. 144, 2012. arXiv: `1207.7326 [astro-ph.CO]`.

[25] Perseus digital library, `http://www.perseus.tufts.edu/hopper/` (visited on 08/09/2014).

[26] P. Schneider, *Extragalactic Astronomy & Cosmology*. Springer Berlin Heidelberg New York, 2006.

[27] A. Mariano and L. Perivolaropoulos, 'Is there Correlation between Fine Structure and Dark Energy Cosmic Dipoles?', *arXiv.org*, no. 8, p. 083 517, 2012. arXiv: 1211.5915 [astro-ph.CO].

[28] H. Aihara *et al.*, 'The Eighth Data Release of the Sloan Digital Sky Survey: First Data from SDSS-III', *The Astrophysical Journal Supplement*, vol. 193, no. 2, p. 29, 2011.

[29] I. N. Reid *et al.*, 'The second Palomar Sky Survey', *Astronomical Society of the Pacific*, vol. 103, pp. 661–674, 1991.

[30] R. M. Cutri *et al.*, *2MASS All Sky Catalog of point sources.* 2003.

[31] J. J. Condon *et al.*, 'The NRAO VLA Sky Survey', *The Astronomical Journal*, vol. 115, no. 5, pp. 1693–1716, 1998.

[32] A. A. Abdo *et al.*, 'Fermi/Large Area Telescope Bright Gamma-Ray Source List', *The Astrophysical Journal Supplement Series*, vol. 183, no. 1, p. 46, 2009.

[33] N. Scoville *et al.*, 'The Cosmic Evolution Survey (COSMOS): Overview', *The Astrophysical Journal Supplements*, vol. 172, pp. 1–8, 2007. eprint: astro-ph/0612305.

[34] C. P. Ahn *et al.*, 'The Tenth Data Release of the Sloan Digital Sky Survey: First Spectroscopic Data from the SDSS-III Apache Point Observatory Galactic Evolution Experiment', 2013. arXiv: 1307.7735.

[35] D. J. Eisenstein *et al.*, 'SDSS-III: Massive Spectroscopic Surveys of the Distant Universe, the Milky Way, and Extra-Solar Planetary Systems', *The Astronomical Journal*, vol. 142, no. 3, p. 72, 2011.

[36] K. S. Dawson *et al.*, 'The Baryon Oscillation Spectroscopic Survey of SDSS-III', *The Astronomical Journal*, vol. 145, 2013.

[37] J. Ge *et al.*, 'The Multi-object APO Radial-Velocity Exoplanet Large-area Survey (MARVELS)', in *Extreme Solar Systems*, D. Fischer *et al.*, Eds., ser. Astronomical Society of the Pacific Conference Series, vol. 398, 2008, p. 449.

[38] C. A. Prieto, S. R. Majewski and R. Schiavon, 'APOGEE: The Apache Point Observatory Galactic Evolution Experiment', *Astronomische Nachrichten*, vol. 329, p. 1018, 2008.

[39] Sloan Digital Sky Survey III, http://www.sdss3.org/instruments/telescope.php (visited on 17/10/2014).

[40] S. Alam, F. D. Albareti and C. Allende Prieto, 'The Eleventh and Twelfth Data Releases of the Sloan Digital Sky Survey: Final Data from SDSS-III', *arXiv.org*, p. 963, 2015. arXiv: 1501.00963 [1501].

[41] D. O. Edge *et al.*, 'A survey of radio sources at a frequency of 159 Mc/s.', *Memoirs of the Royal Astronomical Society*, vol. 68, pp. 37–60, 1959.

[42] A. S. Bennett, 'The revised 3C catalogue of radio sources.', *Memoirs of the Royal Astronomical Society*, vol. 68, p. 163, 1962.

[43] T. A. Matthews and A. R. Sandage, 'Optical Identification of 3c 48, 3c 196, and 3c 286 with Stellar Objects.', *The Astrophysical Journal*, vol. 138, p. 30, 1963.

[44] C. Hazard, M. B. Mackey and A. J. Shimmins, 'Investigation of the Radio Source 3C 273 By The Method of Lunar Occultations', *Nature*, vol. 197, no. 4, pp. 1037–1039, 1963.

[45] M. Schmidt, '3C 273 : A Star-Like Object with Large Red-Shift', *Nature*, vol. 197, no. 4, pp. 1040–1040, 1963.

[46] G. Miley, 'The structure of extended extragalactic radio sources', *Annual Review of Astronomy and Astrophysics*, 1980.

[47] R. Antonucci, 'Unified models for active galactic nuclei and quasars', *Annual review of astronomy and astrophysics*, vol. 31, pp. 473–521, 1993.

[48] Schematic of the AGN unification scheme, `http://www.newscientist.com/data/images/ns/cms/dn12386/dn12386-2_600.jpg` (visited on 04/10/2014).

[49] J. Kormendy and D. Richstone, 'Inward Bound—The Search For Supermassive Black Holes In Galactic Nuclei', *Annual Review of Astronomy and Astrophysics*, vol. 33, no. 1, pp. 581–624, 1995.

[50] D. Richstone *et al.*, 'Supermassive black holes and the evolution of galaxies.', *Nature*, vol. 395, no. 6, p. 14, 1998.

[51] J. Magorrian *et al.*, 'The Demography of Massive Dark Objects in Galaxy Centers', *The Astronomical Journal*, vol. 115, no. 6, pp. 2285–2305, 1998.

[52] P. Martini, 'QSO Lifetimes', *Coevolution of Black Holes and Galaxies*, p. 169, 2004. eprint: `astro-ph/0304009`.

[53] P. F. Hopkins *et al.*, 'Black Holes in Galaxy Mergers: Evolution of Quasars', *The Astrophysical Journal*, vol. 630, no. 2, pp. 705–715, 2005.

[54] G. A. Shields, 'Thermal continuum from acretion disks in quasars', *Nature*, vol. 272, pp. 706–708, 1978.

[55] P.-C. Zinn *et al.*, 'The Australia Telescope Large Area Survey: 2.3 GHz observations of ELAIS-S1 and CDF-S. Spectral index properties of the faint radio sky', *Astronomy & Astrophysics*, vol. 544, p. 38, 2012.

[56] D. E. V. Berk *et al.*, 'Composite Quasar Spectra from Sloan Digital Sky Survey', *The Astrophysical Journal*, vol. 122, p. 549, 2001.

[57] R. Lynds, 'The absorption-line spectrum of 4c 05.34', *The Astrophysical Journal*, 1971.

[58] Homepage of John K. Webb, `http://newt.phys.unsw.edu.au/~jkw/qsoabs_cartoon/` (visited on 01/03/2015).

[59] J. E. Gunn and B. A. Peterson, 'On the Density of Neutral Hydrogen in Intergalactic Space.', *The Astrophysical Journal*, vol. 142, pp. 1633–1641, 1965.

[60] A. A. Penzias and R. W. Wilson, 'A Measurement of Excess Antenna Temperature at 4080 Mc/s.', *The Astrophysical Journal*, vol. 142, pp. 419–421, 1965.

[61] J. E. Rhoads and S. Malhotra, 'Probing Reionization and Galaxy Formation with Lyman Alpha Galaxies', *American Astronomical Society*, vol. 199, p. 1354, 2001.

[62] J. Miralda-Escude *et al.*, 'The Lyman alpha forest from gravitational collapse in the CDM + Lambda Model', *The Astrophysical Journal*, vol. 471, p. 582, 1996.

[63] R.-y. Cen, 'Testing cosmological models with a Lyman alpha cloud statistic: The Fraction of a quasar spectrum with high Lyman alpha optical depth', *The Astrophysical Journal*, vol. 479, p. L85, 1997.

[64] S. E. Nuza, A. G. Sánchez and F. Prada, 'The clustering of galaxies at z$\approx$ 0.5 in the SDSS-III Data Release 9 BOSS-CMASS sample: a test for the ΛCDM cosmology', *Monthly Notices of the Royal Astronomical Society*, 2013.

[65] L. Anderson, E. Aubourg and S. Bailey, 'The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: baryon acoustic oscillations in the Data Release 9 spectroscopic galaxy sample', *Monthly Notices of the Royal Astronomical Society*, 2012.

[66] L. Anderson, É. Aubourg and S. Bailey, 'The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: baryon acoustic oscillations in the Data Releases 10 and 11 Galaxy samples', *Monthly Notices of the Royal Astronomical Society*, 2014.

[67] M. Colless *et al.*, 'The 2dFGRS Galaxy Redshift Survey: Spectra and redshifts', *Monthly Notices of the Royal Astronomical Society*, vol. 328, no. astro-ph/0106498, p. 1039, 2001.

[68] M. Tegmark *et al.*, 'Cosmological constraints from the SDSS luminous red galaxies', *Physical Review D*, 2006.

[69] A. G. Sánchez, M. Crocce and A. Cabré, 'Cosmological parameter constraints from SDSS luminous red galaxies: a new treatment of large-scale clustering', *Monthly Notices of the Royal Astronomical Society*, 2009.

[70] A. G. Sánchez, C. G. Scóccola and A. J. Ross, 'The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: cosmological implications of the large-scale two-point correlation function', *Monthly Notices of the Royal Astronomical Society*, 2012.

[71] G. B. Zhao, S. Saito and W. J. Percival, 'The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: weighing the neutrino mass using the galaxy power spectrum of the CMASS sample', *Monthly Notices of the Royal Astronomical Society*, 2013.

[72] G. Rossi, 'Neutrino mass from the Lyman-Alpha forest', *arXiv.org*, vol. astro-ph.CO, pp. 221–224, 2014.

[73] P. Noterdaeme *et al.*, 'Evolution of the cosmological mass density of neutral gas from Sloan Digital Sky Survey II - Data Release 7', *arXiv*, 2009. arXiv: `0908.1574`.

[74] P. Noterdaeme *et al.*, 'Column density distribution and cosmological mass density of neutral gas: Sloan Digital Sky Survey-III Data Release 9', *Astronomy & Astrophysics*, vol. 547, p. L1, 2012.

[75] C. P. Ahn, R. Alexandroff and C. A. Prieto, 'The ninth data release of the Sloan Digital Sky Survey: first spectroscopic data from the SDSS-III Baryon Oscillation Spectroscopic Survey', *The Astrophysical Journal*, vol. 203, p. 21, 2012.

[76] X. Fan, 'Simulation of Stellar Objects in SDSS Color Space', *The Astronomical Journal*, vol. 117, pp. 2528–2551, 1999. eprint: `astro-ph/9902063`.

[77] A. Font-Ribera *et al.*, 'Quasar-Lyman α Forest Cross-Correlation from BOSS DR11 : Baryon Acoustic Oscillations', *The Astronomical Journal*, vol. 1405, p. 027, 2014.

[78] A. Songaila and L. L. Cowie, 'Approaching Reionization: The Evolution of the Ly α Forest from z=4 to z=6', *The Astronomical Journal*, vol. 123, pp. 2183–2196, 2002. eprint: `astro-ph/0202165`.

[79] T. Theuns *et al.*, 'Detection of He II Reionization in the Sloan Digital Sky Survey Quasar Sample', *The Astrophysical Journal Letters*, vol. 574, pp. L111–L114, 2002. eprint: `astro-ph/0206319`.

[80] J. S. Bolton, S. P. Oh and S. R. Furlanetto, 'The evolution of the Lyman-alpha forest effective optical depth following HeII reionisation', *Monthly Notices of the Royal Astronomical Society*, vol. 396, p. 2405, 2009.

[81] T. Theuns *et al.*, 'Temperature fluctuations in the intergalactic medium', *Monthly Notices of the Royal Astronomical Society*, vol. 332, pp. 367–382, 2002. eprint: `astro-ph/0110600`.

[82] G. C. Rudie *et al.*, 'The Column Density Distribution and Continuum Opacity of the Intergalactic and Circumgalactic Medium at Redshift $< z > = 2.4$', *The Astrophysical Journal*, vol. 769, no. 2, p. 146, 2013.

[83] C. G. Díaz *et al.*, 'Large-scale environment of z$\sim$5.7 C IV absorption systems - I. Projected distribution of galaxies', *Monthly Notices of the Royal Astronomical Society*, vol. 442, pp. 946–978, 2014. arXiv: `1404.7656`.

[84] S. L. Ellison, J. X. Prochaska and J. T. Mendel, 'Metallicities and dust content of proximate damped Lyman α systems in the Sloan Digital Sky Survey', *Monthly Notices of the Royal Astronomical Society*, vol. 412, pp. 448–468, 2011. arXiv: `1010.5244 [astro-ph.CO]`.

[85] M. Rafelski *et al.*, 'Metallicity Evolution of Damped Lyman-alpha Systems out to $z \sim 5$', *The Astrophysical Journal*, vol. 755, no. 1, p. 21, 2012. arXiv: `1209.1772 [203]`.

[86] R. F. Carswell *et al.*, 'High-redshift QSO absorbing clouds and the background ionizing source', *The Astrophysical Journal*, vol. 319, pp. 709–722, 1987.

[87] W. H. Press *et al.*, *Numerical recipes in Fortran 77*. Cambridge University, 1992.

[88] S. V. Penton, J. T. Stocke and J. M. Shull, 'The Local Lyα Forest. IV. Space Telescope Imaging Spectrograph G140M Spectra and Results on the Distribution and Baryon Content of H I Absorbers', *The Astrophysical Journal Supplement Series*, vol. 152, no. 1, pp. 29–62, 2004.

[89] T. S. Kim *et al.*, 'The physical properties of the Lyα forest at $z > 1.5$', *Monthly Notices of the Royal Astronomical Society*, vol. 335, no. 3, pp. 555–573, 2002.

[90] J. X. Prochaska *et al.*, 'Towards a unified description of the intergalactic medium at redshift z $\approx$ 2.5', *Monthly Notices of the Royal Astronomical Society*, vol. 438, no. 1, pp. 476–486, 2014.

[91] J. X. Prochaska and A. M. Wolfe, 'On the (Non)Evolution of H I Gas in Galaxies Over Cosmic Time', *The Astrophysical Journal*, vol. 696, no. 2, pp. 1543–1547, 2009.

[92] A. P. Calverley *et al.*, 'Measurements of the ultraviolet background at $4.6 < z < 6.4$ using the quasar proximity effect', *Monthly Notices of the Royal Astronomical Society*, vol. 412, pp. 2543–2562, 2011. arXiv: 1011.5850 [astro-ph.CO].

[93] S. Bajtlik, R. C. Duncan and J. P. Ostriker, 'Quasar ionization of Lyman-alpha clouds - The proximity effect, a probe of the ultraviolet background at high redshift', *The Astrophysical Journal*, vol. 327, pp. 570–583, 1988.

[94] J. M. Shull, C. W. Danforth and E. M. Tilton, 'Tracing the Cosmic Metal Evolution in the Low-redshift Intergalactic Medium', *The Astrophysical Journal*, vol. 796, 49, p. 49, 2014. arXiv: 1409.6720.

[95] Sloan Digital Sky Survey Data Release 10 Quasar Catalog, http://data.sdss3.org/datamodel/files/BOSS_QSO/DR10Q/DR10Q.html (visited on 04/11/2014).

[96] Data Access for SDSS DR10, http://www.sdss3.org/dr10/data_access/ (visited on 12/09/2014).

[97] Synchrotron emission of relativistic electrons, http://astronomy.swin.edu.au/cms/cpg15x/albums/userpics/synchrotron2.jpg (visited on 13/12/2014).

[98] K. Takarada, 'Radiation Spectra of Quasars', *Progress of Theoretical Physics*, vol. 40, pp. 770–780, 1968.

[99] J. M. Shull, M. Stevans and C. W. Danforth, 'HST-COS Observations of AGNs. I. Ultraviolet Composite Spectra of the Ionizing Continuum and Emission Lines', *The Astrophysical Journal*, vol. 752, no. 2, p. 162, 2012.

[100] R. C. Telfer *et al.*, 'The Rest-Frame Extreme-Ultraviolet Spectral Properties of Quasi-stellar Objects', *The Astrophysical Journal*, vol. 565, pp. 773–785, 2002. eprint: astro-ph/0109531.

[101] D. E. V. Berk *et al.*, 'Composite Quasar Spectra from the Sloan Digital Sky Survey', *The Astronomical Journal*, vol. 122, no. 2, p. 549, 2001.

[102] M. A. Zwaan *et al.*, 'Reconciling the local galaxy population with damped Lyman α cross-sections and metal abundances', *Monthly Notices of the Royal Astronomical Society*, vol. 364, no. 4, pp. 1467–1487, 2005.

[103] J. M. O'Meara *et al.*, 'The Keck+Magellan Survey for Lyman Limit Absorption. I. The Frequency Distribution of Super Lyman Limit Systems', *The Astrophysical Journal*, vol. 656, no. 2, pp. 666–679, 2007.

[104] J. N. Bahcall and P. J. E. Peebles, 'Statistical Tests for the Origin of Absorption Lines Observed in Quasi-Stellar Sources', *The Astrophysical Journal*, vol. 156, p. L7, 1969.

[105] H. S. Murdoch, R. W. Hunstead and M. Pettini, 'Absorption spectrum of the Z= 3.78 QSO 2000-330. II-The redshift and equivalent width distributions of primordial hydrogen clouds', *The Astrophysical Journal*, vol. 309, pp. 19–32, 1986.

[106] F. N. Fritsch and R. E. Carlson, 'Monotone Piecewise Cubic Interpolation', *SIAM Journal on Numerical Analysis*, vol. 17, pp. 238–246, 1980.

[107] J. X. Prochaska, J. M. O'Meara and G. Worseck, 'A Definitive Survey for Lyman Limit Systems at z $\sim$ 3.5 with the Sloan Digital Sky Survey', *The Astrophysical Journal*, vol. 718, pp. 392–416, 2010.

[108] J. K. Webb, R. F. Carswell and M. J. Irwin, 'High Resolution Spectroscopy in the Lyman Alpha Forest; Possible Evidence for Absorption Line Clustering', in *Bulletin of the American Astronomical Society*, ser. Bulletin of the American Astronomical Society, vol. 16, 1984, p. 733.

[109] J. K. Webb, 'Clustering of Lyα Absorbing Clouds at High Redshift', in *Observational Cosmology*, A. Hewitt, G. Burbidge and L. Z. Fang, Eds., ser. IAU Symposium, vol. 124, 1987, p. 803.

[110] M. M. Ćirković and K. M. Lanzetta, 'On the small-scale clustering of the Lyα forest clouds', *Monthly Notices of the Royal Astronomical Society*, vol. 315, pp. 473–478, 2000. eprint: `astro-ph/0010209`.

[111] R. Davé, 'The Evolution of the Lyman Alpha Forest From Redshift 3 to 0', *arXiv.org*, 2001. arXiv: `astro-ph/0105085v1 [astro-ph]`.

[112] C. W. Danforth and J. M. Shull, 'The Low-z Intergalactic Medium. III. H I and Metal Absorbers at z $<$ 0.4', *The Astrophysical Journal*, vol. 679, no. 1, pp. 194–219, 2008.

[113] G. C. Rudie, C. C. Steidel and M. Pettini, 'The Temperature-Density Relation in the Intergalactic Medium at Redshift z=2.4', *arXiv.org*, 2012. arXiv: `1209.0005v1 [astro-ph.CO]`.

[114] B. Aracil *et al.*, 'Metals in the intergalactic medium', *Astronomy and Astrophysics*, vol. 419, no. 3, pp. 811–819, 2004.

[115] G. D. Becker *et al.*, 'Iron and α-element Production in the First One Billion Years after the Big Bang', *The Astrophysical Journal*, vol. 744, 91, p. 91, 2012. arXiv: `1111.4843 [astro-ph.CO]`.

[116] R. Cooke, M. Pettini and M. T. Murphy, 'A new candidate for probing Population III nucleosynthesis with carbon-enhanced damped Lyα systems', *Monthly Notices of the Royal Astronomical Society*, vol. 425, pp. 347–354, 2012. arXiv: 1201.1004 [astro-ph.CO].

[117] R. J. Cooke, M. Pettini and R. A. Jorgenson, 'The Most Metal-poor Damped Lyα Systems: An Insight into Dwarf Galaxies at High-redshift', *The Astrophysical Journal*, vol. 800, 12, p. 12, 2015. arXiv: 1406.7003.

[118] P. Madau and J. M. Shull, 'Cosmic Metal Production and the Contribution of QSO Absorption Systems to the Ionizing Background', *The Astrophysical Journal*, vol. 457, p. 551, 1996. eprint: astro-ph/9509136.

[119] L. A. Barnes, T. Garel and G. G. Kacprzak, 'Ly-α and Mg II as Probes of Galaxies and Their Environment', *Publications of the Astronomical Society of the Pacific*, vol. 126, pp. 969–1009, 2014. arXiv: 1409.6320.

[120] C. Fechner, 'Reconstructing the intergalactic UV background with QSO absorption lines', *Astronomy and Astrophysics*, vol. 532, A62, A62, 2011. arXiv: 1106.2821 [astro-ph.CO].

[121] F. Fontanot *et al.*, 'On the evolution of the cosmic ionizing background', *Monthly Notices of the Royal Astronomical Society*, vol. 438, pp. 2097–2104, 2014. arXiv: 1312.0615 [astro-ph.CO].

[122] Dexter, http://dc.zah.uni-heidelberg.de/sdexter (visited on 08/01/2015).

[123] M. Bernardi *et al.*, 'A Feature at z$\sim$3.2 in the Evolution of the Lyα Forest Optical Depth', *The Astronomical Journal*, vol. 125, pp. 32–52, 2003. eprint: astro-ph/0206293.

[124] P. J. Peebles, *Principles of Physical Cosmology*. Princeton University Press, 1993.

[125] J. Liske, J. K. Webb and R. F. Carswell, 'Large-scale structure in the Lyman-alpha forest: a new technique', *The Monthly Notices of the Royal Astronomical Society*, vol. 301, pp. 787–796, 1998. eprint: astro-ph/9808082.

[126] J. B. Oke and D. G. Korycansky, 'Absolute spectrophotometry of very large redshift quasars', *The Astrophysical Journal*, vol. 255, pp. 11–19, 1982.

[127] J. N. Bahcall *et al.*, 'The Hubble Space Telescope quasar absorption line key project. I - First observational results, including Lyman-alpha and Lyman-limit systems', *The Astrophysical Journal*, vol. 87, pp. 1–43, 1993.

[128] M. A. Perryman, 'Overview of the Gaia Mission', in *Astrometry in the Age of the Next Generation of Large Telescopes*, P. K. Seidelmann and A. K. Monet, Eds., ser. Astronomical Society of the Pacific Conference Series, vol. 338, 2005, p. 3.

[129] Cloudy & Associates, official CLOUDY webpage, `http://nublado.org/` (visited on 10/11/2014).