UNIVERSITÀ' DEGLI STUDI DI TRIESTE

XXVII CICLO DEL DOTTORATO DI RICERCA IN

FISICA
(borsa INAF-Trieste)

Probing the IGM with High Resolution Spectroscopy of Quasars

Settore scientifico-disciplinare: FIS/05

DOTTORANDO
Emanuele Pomante

COORDINATORI
Prof. Paolo Camerini
Prof. Marisa Girardi

SUPERVISORI DI TESI
Prof. Stefano Cristiani
Prof. Francesca Matteucci

ANNO ACCADEMICO 2014 / 2015

*Questa tesi e' dedicata con tutto il
cuore alla mia amata moglie la quale
piu' di tutti in questi anni mi ha
sostenuto e al mio fianco ha
combattuto...*

*Thanks to my loving wife who more
than anyone else supported me and
always struggled by my side...*

Contents

Introduction

Quasar spectra have been for a long time one of the main sources of information about the physical conditions of the Inter-Galactic Medium. Indeed, as the light travels from the quasar to the observer, chemical elements encountered along the path leave their imprint on the quasar spectrum, allowing to measure directly useful physical parameters of the intervening gas like the temperature, the column density of the absorbing ions and their velocity shifts.

Since quasars can be observed out to very large cosmological distances (the furthest observed quasar is at redshift $z \sim 7$ (Mortlock et al., 2011)) this allows us to probe different epochs of the evolution of our Universe. In particular neutral hydrogen clouds produce a sequence of absorption lines in the continuum emission as the light from the distant quasars approaches present epoch (z = 0): these typical features represent the so called *Lyman α forest*.

The $Ly\alpha$ forest is characterized by absorption lines with column densities of $log(N_{HI}) < 17cm^{-2}$; lines with higher column densities represent different classes of absorbers called in order of increasing column density: Lyman Limit System, sub-DLA and DLA (Damped Lyman Alpha system).

As a rule of thumb one can imagine that lines with higher column densities (DLA) will trace dense regions of the Universe like galaxies, while going towards the IGM far from the overdensities where galaxies originate one observes lines with lower column densities ($Ly\alpha$ forest).

The work presented in this thesis has been focused on how to study the IGM and in particular on the determination of its metal content pushed to unprecedented low density regimes.

Metals are produced in stars and a large fraction is ejected out to very large distances by Supernovae, galactic winds and fountains. Metal abundances provide some of our most valuable cosmological insight, but it is still contro-

versial how and when the observed IGM enrichment with metals occurred. More stringent constraints to the epoch of enrichment would come from the determination of the fraction of the Universe (or *volume filling factor*) enriched by metals. A large volume filling factor would be a clear indication in favor of early enrichment.

Nevertheless to obtain a significant measure of the volume filling factor it is crucial to probe the gas near the mean density which occupies most of the volume of the Universe and represents about the 80% of the gas mass.

Moreover this kind of measurements allow us to put constraints on feedback models implemented in cosmological simulations and help to figure out a clearer understanding of galactic winds mechanisms.

The most straightforward way to perform this measurement using quasars spectra is to search for metal absorption lines associated, i.e. at the same redshift, with lines of the $Ly\alpha$ forest. Usually the ÅÅ1548, 1550 CIV doublet is used as a metal tracer for two main reasons: Carbon is one of the most abundant elements in the Universe after Hydrogen and Helium; resonance frequencies of this doublet fall redward of the $Ly\alpha$ emission of the QSO (1215, 67Å) and so absorption lines are not crowded in the $Ly\alpha$ forest.

Since this kind of absorption lines are in general narrow and weak, from the observational point of view there are two main limiting factors to be considered: the spectral resolution and the sensitivity of the measured spectrum. Lyman alpha forest associated CIV lines show column densities going from $log(N_{CIV}) \sim 14.5 cm^{-2}$ down to $log(N_{CIV}) \sim 11.5 cm^{-2}$ where the lowest column density lines which can be detected are essentially limited by the SNR of the spectrum.

As an example for a 3σ detection of a CIV 1548 line with column density $log(N_{CIV}) = 13.5 cm^{-2}$ and a typical Doppler parameter of 20 km/s with a spectral resolution $R = 50000$ would be sufficient a spectrum with $SNR > 2$, on the other hand with the same parameters a CIV line with column density $log(N_{CIV}) = 11.8 cm^{-2}$ could be detected only in a spectrum with $SNR \sim 97$.

The signal-to-Noise of measured spectra cannot be easily increased since it mainly depends on the photons collecting area of the telescopes and on the exposure time. Nowadays even with the largest available telescopes, to obtain spectra of $SNR \sim 100$ or higher, tens of hours of observations are required.

Previous studies (Songaila, 2001; Boksenberg et al., 2003; Cooksey et al., 2010; D'Odorico et al., 2010; Simcoe et al., 2011; D'Odorico et al., 2013; Burchett et al., 2015) have successfully accomplished statistical measurement of the CIV lines column density distribution function and the CIV mass density Ω_{CIV} over the range $0 < z < 6$ using large sample of quasars ($\sim$ 30 QSO).

Nevertheless these measurements have been limited by the SNR of the available QSO which rarely reach values larger than $\sim$ 100 per pixel and in most cases does not allow to explore significantly the lowest column density regions.

In most cases the searches for weak associated metal absorption features (based on Voigt profile fitting of CIV lines) have been limited to CIV with $log(N_{CIV}) > 12.0 cm^{-2}$ associated with HI absorption with $log(N_{HI}) > 14.5 cm^{-2}$ (corresponding to overdensities of a few at $z \sim 3$).

Limiting the analysis to $log(N_{CIV}) > 12.0 cm^{-2}$ excludes all the systems which are more likely to represent IGM regions much closer to the mean density of the Universe. This incompleteness becomes evident in the column density distribution function which are not yet clearly defined for $log(N_{CIV}) < 12.0 cm^{-2}$.

Recently an ESO program (*Deep Spectrum* - PI S. Cristiani) has been carried out to deeply explore for the first time the $log(N_{CIV}) < 12.0 cm^{-2}$ region and probe directly the metallicity of the mean density gas.

Using the UVES echelle spectrograph mounted at the ESO-VLT telescope $\sim$ 65 hours of observations of the QSO *HE0940-1050* at $z \sim 3$ have been carried out reaching the impressive $SNR \sim 230$ per pixel in the CIV region. As a co-investigator of the program I have been responsible for the data reduction and for the development of an algorithm for the flux calibration and the co-adding of the individual exposures to produce the final spectrum. The implemented pipeline and procedures, which are presented in this thesis, have been developed with innovative techniques giving particular attention to the propagation of the errors. The final propagated errors have been checked and compared with errors produced by the standard approach provided by the ESO pipeline and procedures, finding a better correspondence of the former with the analysis of the RMS dispersion of the pixels.

Robust error estimation is particularly relevant for the Voigt profile fitting process which determines the goodness of the line fitting from the χ^2 test.

The study presented from the Deep Spectrum has two main limitations: only one sightline is not enough to obtain robust conclusions, from a statistical point of view more sightlines are required to sample appropriately the sky and reduce the cosmic variance; only one QSO will cover a limited redshift range reducing the possibility to study the evolution of the measured physical quantities.

A further step in the exploration of the very low column density regime is to extend the available sample of very high SNR quasars. Obvious difficulties arise from the process required to achieve all the huge amount of observation time required at the most busy telescopes of the world.

Here comes the main point of the thesis: online archives available today represent a precious source of scientific data which in most cases are not yet fully exploited. There are thousands of terabytes of data stored during the operational life of instruments and telescopes which are available for free and without schedules or waiting times, why not to use them to build a very high SNR homogeneous QSO sample?

Focusing our attention on the main high resolution echelle spectrographs available today, the HIRES at Keck telescopes and UVES at VLT telescopes, algorithms have been created to search deeply the archive for multiple recurrences of the same targets. The output of this process has been a list of hundreds of quasars candidates ordered by the total number of available exposures.

QSOs have been selected by redshift, apparent magnitude and available hours of observation, selecting only those with the highest amount of available hours of observation.

Using these selection criteria the final high SNR QSO sample has been built consisting of 12 QSO from $z = 1.71$ to $z = 3.78$ fulfilling the condition $SNR > 100$ for QSO at $z > 3$ and $SNR > 150$ for QSO at $z < 3$, six of them having $SNR > 200$ and one of them $SNR > 250$. The creation of this sample represents one of the main results of this thesis, in particular for the dedicated procedures developed and applied in the data reduction.

The main technical challenge in obtaining such a sample concerns the ability to merge together tens of exposures coming not only from the different setups of the same instruments but also from different spectrographs, when the objects had been observed with both HIRES and UVES.

Since objects from different instruments are usually reduced with different pipelines, differences in the data reduction procedures could in most cases introduce peculiarities in the output spectra.

In order to merge coherently the data from both spectrograph a generalized pipeline has been developed able to reduce in exact the same way the data from HIRES and UVES. This generalized pipeline is an extension of the IDL code developed by Scott Burles and Jason Prochaska for the reduction of MIKE and HIRES data (Bernstein et al., 2015), which uses a novel approach to wavelength calibration, flat fielding and object extraction. This new pipeline has been successfully used for the reduction of all the QSOs used for this study showing its high efficiency and reliability.

All the QSOs of the sample have been analyzed with Voigt profile fitting procedures. Final lists of CIV absorbers have been used to compute the CDDF, the Ω_{CIV} and the two point correlation function. The final results of the analysis are presented in the final chapter of this thesis.

The next pages are organized as follows: a general introduction to the IGM from an observational point of view is presented in the first chapter; in the second chapter are illustrated the technical requirements for a successful observation of the IGM and how it has been possible to build a very high SNR QSO sample; the third chapter is dedicated to an overview of the data reduction and to the presentation of the new HIRES/UVES pipeline specifically developed during this thesis work; the final chapter will present the results of the CIV fitting on the QSO spectra of the sample, the most used statistical estimators like the Column Density Distribution Function and the evolution of the mass density of CIV will be discussed focusing on the lowest column densities regime.

Chapter 1

The intergalactic medium and the formation of structures

The aim of this chapter is to introduce the reader to the theoretical context of the thesis, doing so we will review some of the basic concepts of modern cosmology putting particular emphasis on the Intergalactic Medium and its link with galaxies and the theory of formation of structures. Since this thesis is focused on observations, the reader will find also useful descriptions of the main observables and techniques used in deriving information about the physical conditions and chemical properties of the IGM during its evolution in different epochs.

1.1 Standard cosmological model

The Universe we live in could be described in good approximation through the Robertson-Walker metric and Einstein's general relativity. The observation by Edwin Hubble in the 1929 that distant objects are receding from us at a velocity v proportional to their distance d as

$$v = H_0 d \tag{1.1}$$

where H_0 is the *Hubble constant*, leads to the conclusion that we live in an expanding Universe.

Starting from field equations of general relativity Friedmann in 1922 derived

a solution for an isotropic and homogeneous Universe, from this solution
came two equations known as the *Friedmann equation*

$$H^2 = (\frac{\dot{a}}{a})^2 = \frac{8\pi G}{3}\rho - \frac{kc^2}{a^2} + \frac{\Lambda c^2}{3} \qquad (1.2)$$

and the *Friedmann acceleration equation*

$$\dot{H} + H^2 = \frac{\ddot{a}}{a} = -\frac{4\pi G}{3}(\rho + \frac{3p}{c^2}) + \frac{\Lambda c^2}{3} \qquad (1.3)$$

which contains the cosmological constant Λ, the Hubble parameter $H(t)$, the
scale factor a and its derivative, the density of the Universe ρ, the curvature
k/a^2 and the pressure p (Ryden, 2002).

Results of Hubble observations and expectations from theoretical model were
the first conceptual foundations of the *Big Bang Theory* (Gamow, 1948;
Lemaitre, 1931) in which the present Universe comes from the expansion
of a point-like Universe were all its content was condensed. One of the
predictions of this theory was the existence in the sky of a diffuse radiation
background with a blackbody spectrum with a temperature of $\sim 5K$ and
the discovery of this radiation by Penzias and Wilson in 1965, called after
that the Cosmic Microwave Background or CMB, was a major success of
the Big Bang theory today accepted as the dominant one.

Subsequent observations (Zwicky, 1933) suggested that the matter content
of the Universe is dominated by an unknown *Dark Matter*, in addition to the
baryonic matter, that does not emit any light and that not only Universe
is expanding but this expansion is accelerated[1] requiring the introduction
in the Friedmann equations of a *Dark Energy* that acts like a cosmological
constant.

At present precise measurements of the CMB revealed that the dark matter
represents $\sim 85\%$ of the total matter content and that the cosmological
constant represents the $\sim 73\%$ of the total energy content of the Universe.

From a *Friedmann-Lemaitre-Robertson-Walker* Universe the critical density

$$\rho_c(t) = \frac{3H^2(t)}{8\pi G} \qquad (1.4)$$

can be calculated as the density at which the Universe will have a flat ge-
ometry and so zero value for the curvature k. Usually all other densities are

---

[1]This conclusions was suggested by studies of distant supernovae (Riess, 1998; Perl-
mutter et al., 1999)

expressed in terms of "critical densities" by mean of the *density parameter* defined as

$$\Omega(t) = \frac{\rho(t)}{\rho_c(t)} \qquad (1.5)$$

Using this parametrization the Friedman equation at present epoch $(t = t_0)$ becomes

$$\Omega_{k,0} \equiv -\frac{kc^2}{H_0^2 a_0^2} = 1 - \Omega_0 \qquad (1.6)$$

where $\Omega_{k,0}$ represents the curvature at present time. The term Ω_0 is the total density parameter of the Universe at present time and can be expressed as

$$\Omega_0 = \Omega_{m,0} + \Omega_{\Lambda,0} + \Omega_{r,0} \qquad (1.7)$$

where $\Omega_{m,0}$, $\Omega_{\Lambda,0}$ and $\Omega_{r,0}$ are the matter, cosmological constant and radiation density parameters respectively.

Measurements of the CMB have showed that today the mean density of the Universe equals the critical density, this means $\Omega_0 = \rho_{tot,0}/\rho_c = 1$ and so $\Omega_{k,0} = 0$. Unless future improvements in measurements will show a different behavior the Universe we live in has a flat geometry.

The evolution of these parameters at different epochs can be conveniently derived (Mo and van den Bosch, 2010) from the present values by mean of the function $E(z)$ of the redshift such that for example

$$H(z) = H_0 E(z) \qquad (1.8)$$

where

$$E(z) = [\Omega_{\Lambda,0} + (1 - \Omega_0)(1 + z)^2 + \Omega_{m,0}(1 + z)^3 + \Omega_{r,0}(1 + z)^4]^{1/2} \qquad (1.9)$$

The model in which there is flat geometry $k = 0$, energy dominated by a cosmological constant Λ and matter dominated by dark matter, is called the *Standard Cosmological Model* or ΛCDM, and all measurements presented in this thesis will be interpreted in the framework of this model.

1.1.1 From the hot Big Bang to present day

In the standard cosmological model the Universe began in a very dense and hot point-like state, but what happens in the first instants soon after the Big Bang? How does the Universe evolve from that state to the Universe we see today with stars and galaxies? From theoretical studies and particles
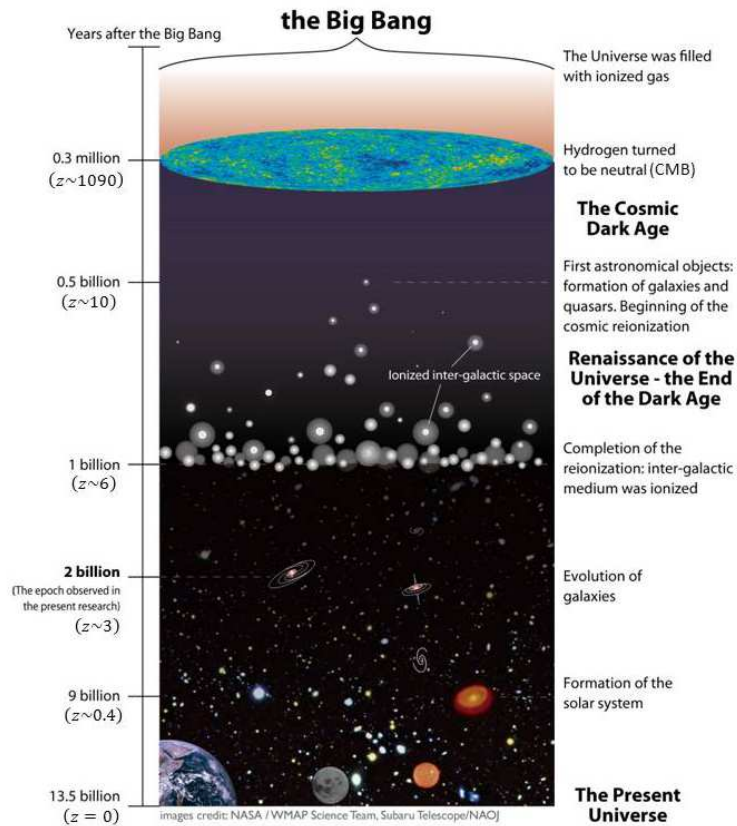
Figure 1.1: The figure shows schematically the main phases the Universe goes trough during its evolution from the Big Bang to the present epoch.

accelerators on earth today scientists are able to describe the different phases the Universe underwent during its evolution, following the schematics in Fig. 1.1

~ 3 **minutes after the Big Bang** stable baryons have formed and for the subsequent ~ 17 minutes temperature is high enough for nuclear fusion and the process called nucleosynthesis can take place to form new atomic nuclei.

Since Universe was rapidly expanding and temperature decrease consequently, nucleosynthesis can occur only for a short time and only lightest nuclei can be produced with a majority of Hydrogen, Helium and Deuterium

Until $z \sim 1090$ Even if nuclei have formed, Universe is still too hot and

dense to form neutral atoms and so it is represented by a fully ionized plasma in which radiation is closely coupled to matter

Recombination When Universe has cooled enough neutral Hydrogen can form, now the majority of free electrons are captured by ions of the plasma and radiation will be no longer coupled with matter and so can be free of travel, this radiation is considered as the "relic" of the Big Bang and is detected today as the CMB

Dark Ages After recombination there was a period called dark ages since no light comes from any sources, distribution of matter is extremely homogeneous and isotropic and no stars are formed yet. Only little anisotropies are present as can be detected from the imprinting left on the CMB, density fluctuations on scale $\lambda \sim 1000h^{-1}Mpc$ are of the order of $\langle(\delta\rho/\rho)^2\rangle \sim 10^{-6}$

Growth of structures Dark matter particles began to accrete via gravitational pull in overdensities originated from the first quantum anisotropies stretched by the inflation, this accretion process gives origin to structure as dark matter haloes, filamentary structures and large voids representing the so called *cosmic web*

Reionization When baryonic matter have cooled enough ($z \sim 15 - 30$), it start to collapse in the potential wells of dark matter haloes forming first stars and galaxy that switch on starting to produce light and metals, anyway the largest fraction of the baryons still reside in the diffuse medium that traces filaments and void in the underlying dark matter structure, this medium is called the *Inter-Galactic Medium*. Light produced from first stars began to *reionize* the neutral medium, this process is not immediate but very slow and proceed in a patchy way starting from little "bubbles" of ionized hydrogen around sources often referred to as a "HII bubble". After the first phase called *pre-overlap* ionizing front of different bubbles continue to expand until the overlap phase began in which different bubbles merge each other forming extended region of ionized medium until $z \sim 6.5$ were Hydrogen is almost all ionized. The Inter-Galactic Medium will carry the imprinting of this process and other processes like the metal enrichment by galactic winds and supernova explosions.

Today All Universe is highly ionized and appears homogeneous and isotropic only on scales larger than $\sim 1700 Mpc$ being characterized by large voids alternated by dense regions where galaxies reside connected by filaments and walls of matter forming together the so called *galaxy clusters*.

1.1.2   Sources of reionization

One of the main difficulties in models of reionization actually concerns the supposed epoch of reionization, which in turn depends on the main sources of ionizing background. To ionize Hydrogen very energetic photons are required with frequency equal or above the Lyman limit 912Å, and sources have to be very efficient producers of this UV radiation in order to reionize all the Universe in the time constraints imposed by observations.

IGM can be reionized by mainly two processes

photo-ionization Ionization of an hydrogen atom by absorption of a photon with frequency larger than the Lyman limit frequency

collisional ionization Ionization of an hydrogen atom by collision with a free electron

The process of collisional ionization requires a temperature of at least $T \gtrsim 10^6 K$ to ionize efficiently the IGM and balance the rate of recombination of ions with free electrons. Since temperatures derived from $Ly\alpha$ forest lines are not high enough, it is generally believed that IGM is predominantly ionized by *photo-ionization* being the derived IGM temperature consistent with this process. Also in this case the flux of ionizing photons has to be high enough to balance the rate of electrons-ions recombination and maintain the IGM at the observed ionization fraction, so the problem of sources able to generate such a flux of ionizing photons is posed.

Main candidates for such sources of ionizing background are:

First population of stars (Pop III) could be the first stars formed in the Universe, soon after the dark ages at redshift $z \sim 10-15$. In principle those stars should have zero metallicity since they have formed from primordial gas (mainly Hydrogen and Helium) falling in the potential well of the first dark matter overdensities evolved in the non-linear regime. Those stars would be characterized by a strong Lyman α emission.

Star forming galaxies are young galaxies whose content is still for the most part in the gas phase, which gets rapidly converted into stars. In general are represented by spiral galaxies characterized by large amount of gas, very young stars and a strong emission of UV-radiation, for this reason those galaxies are generally placed in the blue cloud of the color-magnitude diagram. In this category there are also the so called *Lyman Break Galaxies* or LBG, characterized by strong UV emission and a sudden "break" in the transmitted flux bluer than the Lyman limit wavelength at 912Å indicating the presence of large amount of neutral hydrogen.

Active Galactic Nuclei (AGN) very distant galaxies characterized by the presence of an accretion disk, or torus, feeding a super massive central black hole in the core of the galaxy. As a consequence of the infalling of particles in the deep potential well of the black hole, two powerful jets of particles and radiation are emitted perpendicularly to the torus plane. These objects are characterized by very strong Lyman α emission.

Main distinctive feature in the reionization scenario concerns the epoch of their appearance, while the latter two are favored in cosmological models in agreement with a late reionization scenario at $z < 10$, the first ones are supposed to reionize the IGM prior to the formation of first galaxies at $z \gtrsim 15$ in agreement with models which expect the formation of those Pop III stars soon after the first overdensities begin to enter the non-linear regime.

1.2 The inter-galactic medium

The inter-galactic medium (IGM) is the medium that permeates the space in between galaxies. In order to better understand relation between galaxies and IGM and their interplay during evolution, it will be useful to start from the point of view of the galaxies with reference to the schematics of Fig. 1.2.

Galaxies are made essentially by dark matter, stars and gas. The gas component can be divided in two main contributions

Interstellar Medium (ISM) Gas directly associated with the galaxy
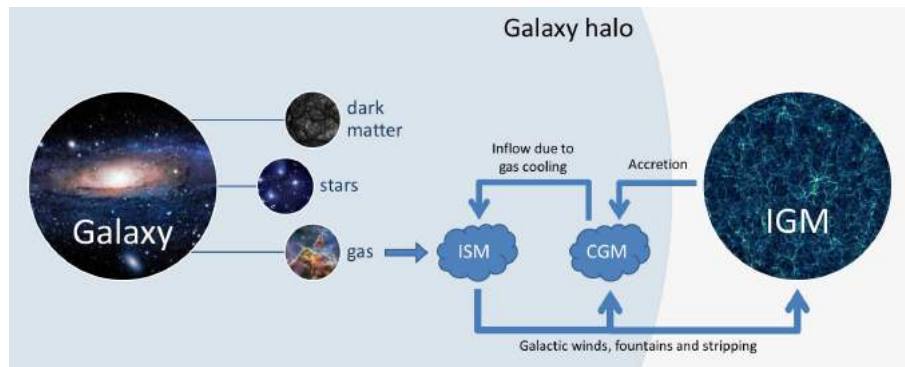
Figure 1.2: In this picture a schematic view of the interplay between gaseous components of the galaxies and the Inter-Galactic Medium is represented. Although there is no net separation between components, in general should be considered IGM the gas outside the virial radius of the dark matter halo.

Circum-galactic Medium (CGM) Halo gas, outside the galaxy, but inside the host dark matter halo of the galaxy

In this representation the IGM can be seen as the gas outside the virial radius of the dark matter halos and so not directly associated with it. It is important to stress that during formation and evolution of galaxies those three components interact actively with each other

- IGM can accrete on dark matter halo becoming parts of the CGM which in turn can cool to form stars in the Galaxy

- ISM can be ejected out very far by galactic winds, fountains (supernova bubbles) and stripping reaching the CGM and IGM

As a display of the deep connections and interplay in the history of formation and evolution of galaxies would be enough to observe that: from one side the IGM reflects the embryonic material out of which galaxies originate, on the other side galaxies can have a significant impact on the IGM by injecting energy in it (UV background - ionization history and temperature of the IGM) and mass (in particular metals that are formed in stars and ejected by supernova explosions).

By investigating the physical properties of the IGM could be possible in principle to understand fundamental aspects of dominant cosmological model.

1.2.1   Quasars as a cosmological probe

In general the IGM has too low density to be detectable in emission, at a variance a powerful method to investigate its properties is the study of absorption produced when light from distant background sources pass through. Quasars represent the ideal candidates as a background sources since they are sufficiently luminous and sufficiently far away, moreover they have well-behaved continua, against which absorption can be analyzed relatively easily. The quantity directly measured when the absorption spectrum is acquired with telescopes and spectrographs is the flux at different wavelengths. Absorbers along the line of sight are responsible for the decrement in flux with respect to the original emission[2] of the quasar, this will be detected on the spectrum as an *absorption line* at a wavelength which depends on the atoms responsible for the absorption. The ratio between the observed flux f_λ and the supposed continuum level f_{cont}, averaged over a specified wavelength range

$$T \equiv \langle f_\lambda / f_{cont} \rangle \tag{1.10}$$

is usually referred to as the *transmission* of the IGM and is directly linked to the opacity of the intervening absorbers.

The flux decrement due to absorption can be related to a quantity called *optical depth* τ by the formula

$$I = I_0 \exp^{-\tau} \tag{1.11}$$

where I_0 is the initial intensity of the source and I is the observed intensity after traveling to the observer.

When a decrement in the observed flux occurs due to an absorption along the light path, the optical depth simply defined in Eq. 1.11 can be related to the number density of absorbing atoms and to the absorption cross section of the atomic transition responsible for the absorption by

$$\tau(\nu_0) \approx \frac{cn_1(t_{\rm a})}{a(t_{\rm a})H_0 E(z_{\rm a})} \left\{ 1 - \exp\left[-\frac{h_P \nu_{12}}{k_B T(t_{\rm a})} \right] \right\} I_{12} \tag{1.12}$$

where

$$I_{12} = \frac{1}{\nu_{12}} \int \sigma_{12}(\nu) d\nu \tag{1.13}$$

---

[2]Flux level that would be observed in the spectrum if no absorption occurred along the sightline is referred as the *continuum level* of the quasar.
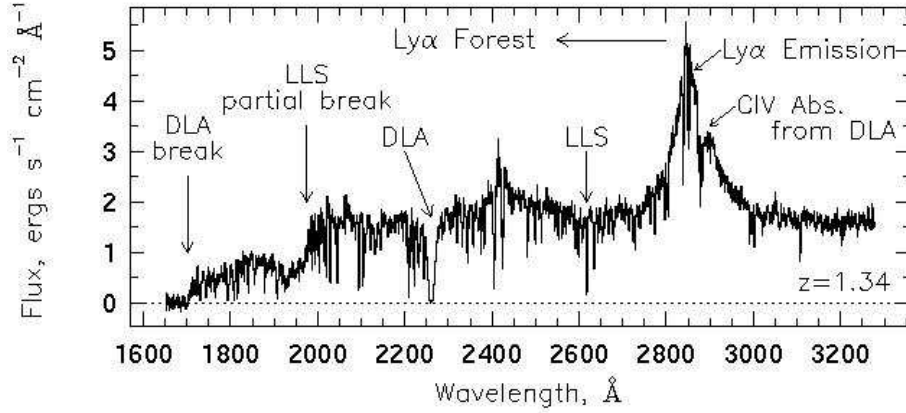
Figure 1.3: Typical quasar spectrum at low redshift. In the image are also indicated the different types of absorption systems which usually characterize quasar spectra (Charlton and Churchill, 2000).

different terms appearing in the equations are: the proper number density $n_1(t_a)$ of the absorbing atoms at the position of the absorption, the Hubble constant $H_0 E(z_a) = H(z_a)$ at the redshift z_a of the absorption (the function $E(z)$ has been defined in Eq. 1.9), the frequency ν_{12} absorbed in the $1 \to 2$ transition and corresponding absorption cross section σ_{12}, the Boltzmann and Planck constants k_B and h_P respectively, and finally the temperature $T(t_a)$ associated with the transition (a detailed derivation can be found in (Mo and van den Bosch, 2010)).

Last equations show clearly that the observed optical depth can be used to infer the physical conditions of the absorbing medium. Since quasars are objects located at cosmological distances and since every absorber along the line of sight leaves its imprinting on the continuum of the quasar spectrum, it is possible to study the physical conditions of astrophysical objects out to very large distances which otherwise would not be accessible. The typical spectrum of a distant quasar is shown in Fig. 1.3.

In the figure are represented various types of absorption systems that could be found in quasars spectrum and the characteristic *Lyα* emission corresponding to the transition from the first excited level of the hydrogen atom to the ground state. The large number of tiny lines blueward of the *Lyα* emission are the signature left on the spectrum by the neutral hydrogen along the line of sight as the light from the quasar approaches us and is known as the *Lyman alpha forest*. It is important to emphasize that the

lines of the Lyman alpha forest represent all the same transition (α transition of the Lyman series exactly) and that the different placement on the spectrum depends on the difference in redshift between subsequent absorptions.

The existence of the Lyman alpha forest was predicted since 1965 (Gunn and Peterson, 1965; Bahcall and Salpeter, 1965; Scheuer, 1965) and few years later discovered (Lynds, 1965) but the meaning and the origin of this absorptions were the center of debate among the scientific community for long years till present day when the advent of large telescopes and better spectrographs allow an improvement in measurements. In the next sections the meaning and interpretations of the QSO[3] absorption systems and related observables will be discussed.

1.2.2 The Gunn-Peterson test

Gunn and Peterson started from the observation that neutral hydrogen HI has a large cross section for $Lyman\alpha$ photons (Gunn and Peterson, 1965), which means that there is an high probability a photon with wavelength (≥ 1215.67Å) being absorbed due to the atomic transition from the ground state to an excited one. If the Universe were filled with a diffuse distribution of HI, a significant absorption of photons bluer than Lyman alpha would be observed, a better understanding of this effect can be achieved from the equations describing the optical depth. If the temperature of the IGM $T \ll T_{Ly\alpha}$, where

$$T_{Ly\alpha} \equiv \frac{h_P \nu_{Ly\alpha}}{k_B} = 1.18 \times 10^5 K \qquad (1.14)$$

the exponential term of Eq. 1.12 can be neglected and the optical depth becomes

$$\tau(\nu_0) \approx \frac{c n_{HI}(t_{\mathrm{a}})}{a(t_{\mathrm{a}}) H_0 E(z_{\mathrm{a}})} I_{Ly\alpha} \qquad (1.15)$$

where

$$I_{Ly\alpha} = \frac{a_{Ly\alpha}}{\nu_{Ly\alpha}} = 4.5 \times 10^{-18} cm^2 \qquad (1.16)$$

which gives an estimate of the Lyman alpha absorption cross section. From the last two equations the relations between the proper number density of

---

[3]This acronym derives from the definition of "Quasi Stellar Objects" which astronomers used for quasar in the first years after their discovery.
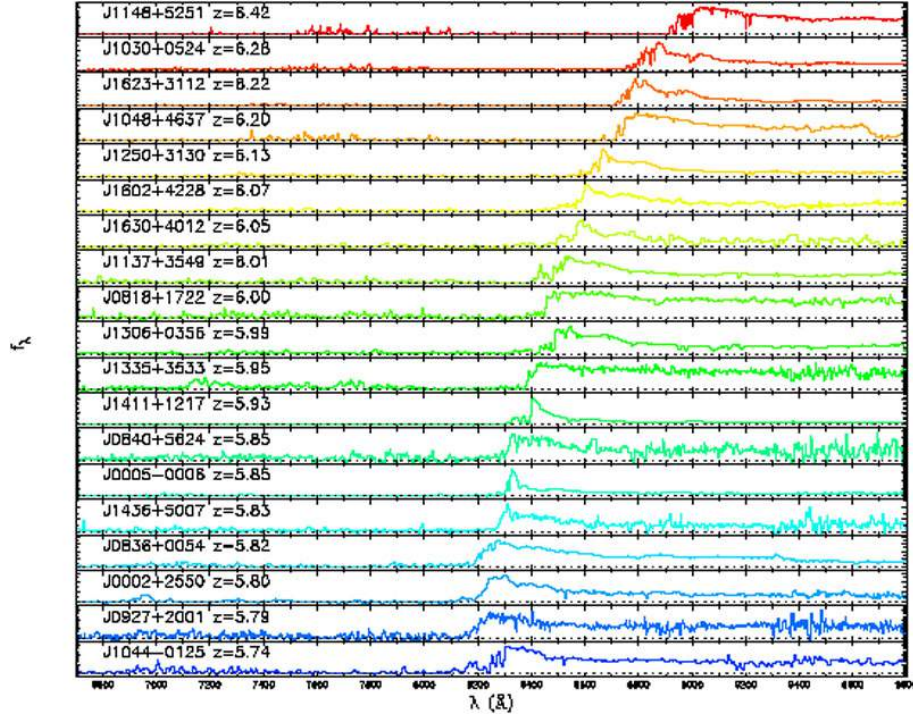
Figure 1.4: A stacked image of 19 quasars spectrum from $z = 5.74$ to $z = 6.42$ shows the rapid increase of the optical depth of the IGM as the redshift increases. For the $z = 6.42$ case almost no flux can be seen blueward the $Ly\alpha$, small flux only in the $Ly\beta$ emission region (Fan et al., 2006).

HI atoms and corresponding optical depth becomes

$$n_{HI}(t_{\mathrm{a}}) = 2.4 \times 10^{-11} hE(z_{\mathrm{a}})\tau \, cm^{-3} \tag{1.17}$$

in an Einstein-deSitter model, which represents a flat and matter dominated Universe, Eq. 1.17 can be rewritten as

$$\tau = 6.6 \times 10^3 h^{-1} \frac{\Omega_b h^2}{0.019} \frac{n_{HI}}{\bar{n}_H} (1+z)^{3/2} \tag{1.18}$$

where Ω_b is the density parameter for the baryonic content of the Universe, and $\bar{n}_H$ is the mean proper number density of Hydrogen atoms, so the ratio $n_{HI}/\bar{n}_H$ represents the fraction of the total Hydrogen in the neutral state. It is clear from Eq. 1.18 that even a small fraction of neutral Hydrogen will result in a large optical depth and so a strong absorption of $Ly\alpha$ photons. This results is confirmed observationally by quasar absorption spectra, measurements of the optical depth of the IGM show $\tau << 1$ out to $z \sim 6$ which means that hydrogen is almost completely ionized out to that distance. Due

to the hydrogen large cross sections to $Ly\alpha$ photons, quasars absorption is not sufficient to probe ionization state of the IGM out to distances greater than $z \sim 6 - 7$ because measurements of the optical depth becomes very difficult. Indeed from Fig. 1.4 becomes clear that as redshift increases the spectrum blueward the $Ly\alpha$ emission line becomes almost completely absorbed, in this case measurement of flux becomes nearly impossible.

A new promising technique to investigate the epoch of reionization, concerns the observation of the $21cm$ emission of neutral hydrogen due to the spin flip transition of the electron with respect to the proton in the nucleus (Ciardi and Ferrara, 2005; Barkana and Loeb, 2007).

1.2.3 Main types of absorption systems

Quasar absorption systems can be distinguished by their characteristic column density which in turn can be related to the opacity of absorbers with respect to photons of frequency equal or greater than Lyman alpha frequency, main systems are shown in Fig. 1.3 and listed in the following:

Lyman alpha Forest These absorption systems show column densities in the range $10^{12} - 10^{17} cm^{-2}$, they appear as narrow lines blueward of the Ly alpha emission and trace the low density regions of the IGM

Lyman Limit Systems (LLS) Column density greater than $10^{17} cm^{-2}$ they become optically thick to radiation with frequency greater than 912Å and show a flux decrement with respect to the QSO continuum bluer than the correspondent wavelength at the redshift of the absorber.

Damped Lyman alpha systems (DLA) Those systems are associated with high density regions and, by definition, show column densities greater than $2 \times 10^{20} cm^{-2}$ and appear in the spectrum as large depression in the mean flux of the QSO continuum as they are optically thick to all photons with frequency greater than Lyman alpha frequency. Probably they are nearby star forming regions and so they show greater metallicity with respect to weaker absorptions.

Metal absorption line systems They are associated with absorbers blueward of the Ly alpha emission but often they appear redward of it making them more easily detectable and measurable. Most commonly observed metals lines are: CIV, SiIV, MgII, NV, FeII, AlIII, OVI...

Their abundances reflect the metallicity and ionization of the different regions they appear in and surprisingly most of them are associated with low density region (IGM) in contrast to what expected if the intergalactic medium reflects the primordial metal content of the Universe (i.e. metals come from elsewhere)

1.3   Observables

Observational progress with traditional $Ly\alpha$ studies can conveniently be charted in terms of two limiting factors: the spectral resolution and the signal-to-noise ratio (SNR). These two factors are in turn related to the technical progress and improvement in instrumentation and dimension of telescopes. Today spectrographs are capable of very high resolution and telescopes reach large diameters as for example the Keck 10 meters telescope in Hawaii or the 8.20 meters of ESO-VLT, but for many years only some features of the QSO spectra were visible and so different ways to obtain information from the spectrum were developed by scientists.

In what follows the main observables that can be obtained from the QSO spectra are listed in order of increasing resolution required to be measured.

1.3.1   Low resolution - Effective optical depth

The most basic observable of the $Ly\alpha$ forest is the flux decrement D, defined as the mean value of the ratio between the observed flux and the expected continuum flux in the absence of absorption. In practice one measures the mean flux decrement D_A between the $Ly\alpha$ and $Ly\beta$ (rest wavelength 1216Å and 1026Å respectively) emission lines

$$D_A = \left\langle 1 - \frac{F_{obs}(\lambda)}{F_{cont}(\lambda)} \right\rangle = \langle 1 - \exp^{-\tau} \rangle = 1 - \exp^{-\tau_{eff}} \tag{1.19}$$

where the average is over the wavelength range covered between $Ly\alpha$ and $Ly\beta$ emission lines. This observable is linked to a quantity called the *effective optical depth* τ_{eff} that can be used in Eq. 1.17 to constrain the number density of neutral hydrogen atoms.

This technique gives the most model-independent measurement possible at the price of getting only one number out of each QSO spectrum, nevertheless with measurements of D_A over a broad redshift range the evolution with redshift of the $Ly\alpha$ forest can be investigated.

The largest uncertainties in the execution of this measurement are probably caused by our ignorance about the precise QSO continuum level and contribution from absorption of metals lines which often cannot be removed in low resolution spectra.

1.3.2 Medium resolution - Line counting

At medium resolution discrete absorbers can be identified in the spectrum and their relative equivalent width W can be measured: a statistic of absorbers in term of their equivalent width and redshift becomes possible. The behavior of $d^2\mathcal{N}/dWdz$, which essentially measures the clumpiness of the $Ly\alpha$ forest gas, can be reasonably expressed by the empirical formula (Sargent et al., 1980)

$$\frac{d^2\mathcal{N}}{dWdz} = \frac{W}{W_*} \exp^{-\frac{W}{W_*}} (1+z)^\gamma \qquad (1.20)$$

with typical value of $W_* \approx 0.27$Å and $1.5 < \gamma < 3$ (see next section for further details on redshift evolution).

Main disadvantage in this case is that without properly deblending the single absorption lines (impossible at low resolution, see 2.1.7 for further details) the curve of growth cannot be used to relate W to more physically meaningful parameters as will be described in next subsection.

1.3.3 High resolution - Voigt profile decomposition

With high resolution spectroscopy it is possible to resolve absorption lines up to $FWHM < 25Km/s$ which is the typical width of Lyman alpha forest lines. Resolved lines of the Lyman alpha forest can be reasonably well approximated by the *Voigt profile.*

The Voigt profile represents a simplification of the profile of an absorption line which generally takes into account different physical aspects which affect the shape of the line and its width. As can be seen in Fig. 1.5 the Voigt profile can be thought as the composition of a Lorentzian function and a Gaussian, the final function being the convolution of these two.

The Lorentzian represents the *natural broadening* of the line due to Heisenberg's uncertainty principle which states that the energy of an excited state could be exactly determined only if the lifetime tends to infinity. Since in
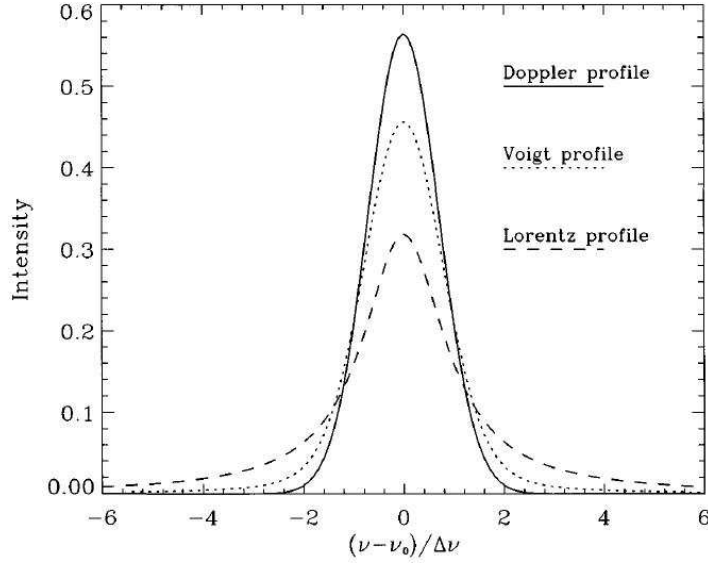
Figure 1.5: The image shows separately the Lorentzian profile due to natural broadening, the Gaussian profile due to the thermal broadening and the Voigt profile resulting from their convolution. Profiles are normalized and x is expressed in term of the Doppler broadening $\Delta\nu$ (Thomas and Stamnes, 1999).

general lifetimes are finite it follows that the energy of an excited state cannot be exactly defined but will spread over a range ΔE. The natural profile can be defined as

$$\mathcal{L}(\nu) \equiv \frac{1}{\pi} \left(\frac{\gamma}{(\nu - \nu_{12})^2 + \gamma^2} \right) \tag{1.21}$$

where ν_{12} is the frequency of the transition, the *full-width-half-maximum* is equal to 2γ and $\gamma \equiv A_{21}/4\pi$, where A_{21} is the spontaneous transition coefficient.

Gaussian represents the *thermal broadening* which can be easily related to the temperature of the absorbing gas by the Doppler parameter b. Assuming that peculiar velocities of the absorbing particles follow a Maxwellian distribution

$$\mathcal{P}(v)dv = \frac{1}{\pi^{1/2}b} \exp\left(-\frac{v^2}{b^2} \right) dv \tag{1.22}$$

this equation describes a Gaussian with a dispersion $\sigma = b/\sqrt{2} = \sqrt{k_B t/m}$ which represents the broadening due to thermal motion of particles with mass m and at temperature T. In the general case where turbulent motion with a Gaussian distribution is included, the Doppler parameter can be

expressed as

$$b^2 = \frac{2k_B T}{m} + b_{turb}^2 \qquad (1.23)$$

where b_{turb} is the Doppler parameter due to turbulent motion.

The general expression for a Voigt profile can be obtained as

$$\phi(\nu) = \int_{-\infty}^{+\infty} \mathcal{L}\left[\nu\left(1 - \frac{v}{c}\right)\right] \mathcal{P}(v)dv \qquad (1.24)$$

which becomes

$$\phi(\nu) = \frac{1}{\sqrt{\pi}} \frac{c}{b} \frac{\mathcal{V}(A, B)}{\nu} \qquad (1.25)$$

where $\mathcal{V}(A, B)$ can be expressed by the *Hjerting function* (or *Voigt function*) as

$$\mathcal{V}(A, B) = \frac{A}{\pi} \int_{-\infty}^{+\infty} \frac{\exp(-y^2)}{(B-y)^2 + A^2} dy \qquad (1.26)$$

with

$$A \equiv \frac{c}{b} \frac{\gamma}{\nu} \qquad B \equiv \frac{c}{b} \frac{\nu - \nu_{12}}{\nu_{12}} \qquad (1.27)$$

A useful approximation to the Voigt function is

$$\mathcal{V}(A, B) \approx \exp(-B^2) + \frac{1}{\sqrt{\pi}} \frac{A}{A^2 + B^2} \qquad (1.28)$$

from this equation can be understood the basilar behavior of the Voigt profile which becomes dominated by thermal broadening near the center of absorption (sometimes called the "Doppler core" of the line) while the edges of the line (where $|\nu - \nu_{12}|/\nu_{12} \gg b/c$ so that $B \gg 1$) are dominated by the Lorentzian profile (those regions are called the "damping wings" of the line profile).

The Voigt profile just defined can be used to derive the optical depth from the equation

$$\tau(\lambda) = N\sigma_{12}(\nu) = \eta_0 \mathcal{V}(A, B) \quad \text{where} \quad \eta_0 \equiv \frac{1}{\sqrt{\pi}} \frac{c}{b} \frac{a_{12}}{\nu_{12}} N \qquad (1.29)$$

which depends from the column density N of the absorbing atoms. In practice the optical depth is related to a quantity usually measured in spectroscopy that is the equivalent width W by the formula

$$W = \int_0^\infty [1 - \exp^{-\tau(\lambda)}]d\lambda \qquad (1.30)$$
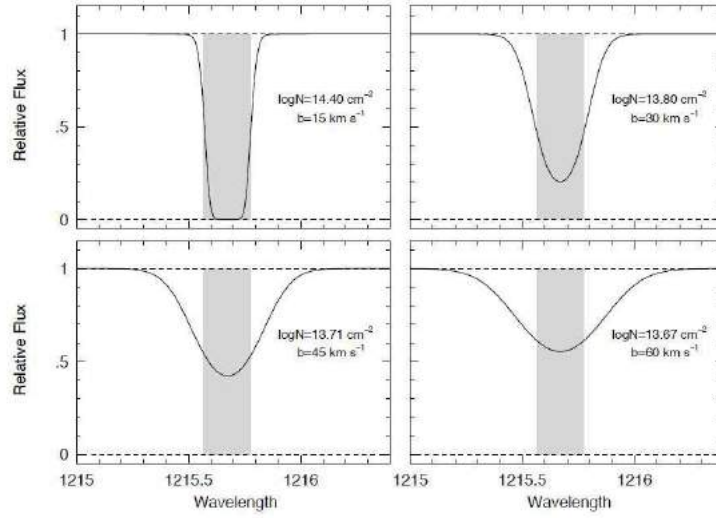
Figure 1.6: Four different representations of absorption lines each with equivalent width $W = 0.21$Å. Even though line profiles are different, the total amounts of flux absorbed, given by Eq. 1.11, are identical. The shaded gray area shows the interpretation of the equivalent width (Churchill, 2016).

The equivalent width is defined as the width that a line should have if all the absorption is represented as a rectangle of height one, as illustrated in Fig. 1.6 different lines could have the same equivalent width depending on their shapes.

Equations 1.29 and 1.30 show that there is a relation between the equivalent width W and the column density N of absorbers. This relation can be expressed graphically by the so called *curve of growth* or COG shown in Fig. 1.7 and approximated at different regimes.

As can be seen from Fig. 1.7, there are three different regimes of the curve of growth depending on the column density:

$\eta_0 \lesssim 1$ - **Linear part** lines are narrow and not saturated, the Lorentzian profile becomes negligible and the line can be well approximated by a Gaussian. The optical depth is small and the equivalent width reduces to

$$W \simeq \int_0^\infty \tau(\lambda)d\lambda = \frac{a_{12}}{\nu_{12}} N\lambda_{12} \qquad (1.31)$$

where a_{12} is the absorption coefficient

$$a_{12} = \frac{\pi e^2}{m_e c} f_{12} \qquad (1.32)$$
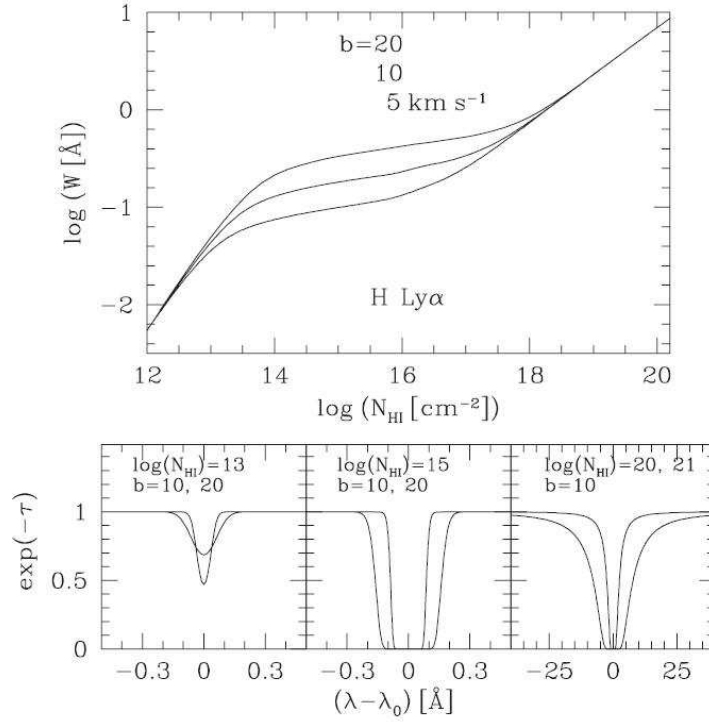
Figure 1.7: In the upper panel the theoretical curve of growth for the *Lyα* transition and three different doppler parameters is shown. The degeneracy between W and N can be clearly observed in the flat part of the curve by noticing that for a small range of equivalent widths corresponds a wide range of column densities. In the lower panel are illustrated three examples showing the shape of a line in the three different regimes of the COG, for each case different values of b and N_{HI} are illustrated (Mo and van den Bosch, 2010).

with e, m_e, c and f_{12} respectively the elementary charge, the electron mass, the speed of light and the oscillator strength for the considered transition. The equivalent width just defined increases linearly with the column density independently of the b parameter, for this reason precise measurements of the column density would be possible

$10 \lesssim \eta_0 \lesssim 1000$ **- Logarithmic part** lines begin to saturate in the Doppler core [i.e. $\tau(\lambda) > 1$ for $\lambda \sim \lambda_{12}$], the equivalent width increases very slowly with N

$$W \approx \frac{2b}{\nu_{12}} \sqrt{\ln \eta_0} \qquad (1.33)$$

this region of the COG is not sensible to column density but increases almost linearly with b allowing for precise temperature estimation.

As can be noticed from Fig. 1.7 the degeneracy between measured equivalent width and relative column density makes very difficult to recover the exact shape of the COG in the central part. One solution to this problem concerns the detection of doublets or multiple transitions of the same atoms or ions like for example the CIV doublet at $\lambda = 1548, 1550$Å. Those transitions should lie on different parts of the same COG making available more points for the definition of the curve.

$\eta_0 > 10^4$ - **Square root part** lines are saturated in the doppler core and are characterized by strong damping wings which begin to dominate the profile, the proportionality between equivalent width and column density can be expressed as

$$W \propto \sqrt{N} \tag{1.34}$$

the curve returns to become sensitive to column density but in practice only the $Ly\alpha$ transition is observed on this part of the COG. Typically systems with $N_{HI} > 10^{20} cm^{-2}$ at $z = 2$ will have $W_{obs} > 22$Å, which correspond to absorption lines so strong that are easily identified even when they are embedded in the $Ly\alpha$ forest.

In summary Voigt profile fitting of Quasar absorption lines, allow to retrieve some of the most meaningful properties of the absorbing medium: the column density, the Doppler parameter and the redshift of the absorber.

1.4 Statistical properties

One of the main advantages of the study of the IGM with quasar absorption spectra is the large portion of space and time that can be directly probed: with Voigt profile fitting of Lyman alpha lines even one QSO spectrum can give column density and doppler parameter of hundreds of absorbers along the line of sight. This wealth of data can be interpreted and useful employed by mean of statistical study on the Lyman alpha forest.
One of the most important and most employed statistic concerns the number of absorbers as a function of redshift and column density described by the so called *Column Density Distribution Function* defined as

$$\mathcal{F}(N,z) \equiv \frac{d^2\mathcal{N}}{dNdz} = n(N,z)\sigma(N,z)(1+z)^3\frac{dl}{dz} = \frac{c}{H(z)}(1+z)^2 n(N,z)\sigma(N,z) \tag{1.35}$$

where $n(N, z)$ represents the number density of absorption line systems with column densities in the interval $(N, N + dN)$ at redshift z, $\sigma(N, z)$ is the proper absorption cross section for column density N at redshift z, $dl = cdt$ is the interval of proper distance corresponding to dz and $H(z)$ is the Hubble constant at z.

In practical cases it would be useful to replace the redshift dz by the so called *redshift absorption path dX* defined as

$$dX = \frac{H_0}{c}(1 + z)^3 dl = \frac{(1 + z)^2}{E(z)} dz \qquad (1.36)$$

where $E(z)$ has been defined in 1.1. In this coordinate the column density becomes

$$\mathcal{F}(N, z) \equiv \frac{d^2 \mathcal{N}}{dN dX} = \frac{c}{H_0} n(N, z)\sigma(N, z) \qquad (1.37)$$

which is independent of redshift for a non-evolving population of absorbers. From the column density distribution function of an element (like HI for example or CIV) it is possible to estimate the mean mass density of that element, moreover if the ionization fraction of the element is known one can estimate the total mass density given by the sum of neutral plus ionized component that will depends on the ionization state of the IGM.

Main drawbacks of this technique concern

- the subjectivity of the decomposition of absorption systems (resulting from the superposition of multiple individual lines) into single Voigt components (a subject that will be discussed in detail in chapter 5)

- blending of lines, mostly in the thick $Ly\alpha$ forest

- incompleteness due to low sensitivity of measurements

- difficulties in precise measurements of particular column density regime like for example the LLS regime (characterized by highly saturated lines)

In the next chapter solutions to some of this problems will be presented and related techniques will be discussed.

1.4.1   Two point correlation function

Once lines of the forest are fitted with Voigt profile and their redshift and column density derived, it is possible to apply the two point correlation

function technique to the entire set of absorbers to study their clustering properties. This technique is basically based on counting of pairs of lines with different velocity separations Δv, in practical applications the two point correlation function can be defined by the formula

$$\delta \mathcal{P}(\bar{z}, \Delta v) = \left(\frac{d\mathcal{N}}{dz}\right)_{\bar{z}} (1 + \bar{z})[1 + \xi(\Delta v)]\frac{\delta v}{c} \qquad (1.38)$$

where $\Delta v = c\Delta z/(1 + \bar{z})$ is the velocity difference in the rest frame, $\bar{z} = (z_1 + z_2)/2$ is the mean redshift of the two lines of a pair, $\xi(\Delta v)$ is the two point correlation function and $\delta \mathcal{P}$ is just the expected number of lines from the mean line density $d\mathcal{N}/dz$. In this case the two point correlation function $\xi(\Delta v)$ is defined as a measure of the excess probability $\delta \mathcal{P}$ of finding two absorption lines at a separation Δv.

Such kind of studies are important to understand the nature and origin of absorbers and their relation with galaxies. In practice the correlation function of absorbers is compared with the correlation function of galaxies in order to see if there is similar clustering properties that could suggest a common origin.

A similar statistic complementary to the this one, is the *void statistics* which study the deviation of the distribution of absorbers from a poissonian one, by measuring the probability of finding a gap Δz in the redshift distribution (Carswell and Rees, 1987).

1.4.2   Pixel optical depth and power spectrum

This method (Croft et al., 1998; Mo and van den Bosch, 2010) considers the $Ly\alpha$ forest as a one-dimensional intensity field sampled on pixels, rather than individual lines, and correlates the pixel intensities.

In practice one observes the transmitted flux $F(\lambda)$ as a function of wavelength which is related to the optical depth $\tau(\lambda)$ by the formula $F = \exp^{-\tau}$. Since optical depth is related to the matter density ρ along the line of sight, in the limit of small perturbations the fluctuations in the transmitted flux $\delta_F \equiv (F - \bar{F})/\bar{F}$ are proportional to fluctuations in the density field $\delta_\rho \equiv (\rho - \bar{\rho})/\bar{\rho}$ and can be used to derive the power spectrum of the gas distribution.

After removing metal lines and highly saturated $Ly\alpha$ lines, and after correcting for possible variations in the continuum, the one-dimensional power spectrum $P_F(k_{1d})$ can be simply recovered from the Fourier transform of the

sampled intensity $F(\lambda)$. The range over which wavenumber k_{1d} can be re-
covered depends on the resolution and the interval of wavelength considered
for the Fourier transformation of the spectrum.

The main advantage of this technique is that the power spectrum calculated
in this way can be directly related to the power spectrum of the matter
distribution by the formula

$$P_F(k_{1d}) = \frac{1}{2\pi} \int_{k_{1d}}^{\infty} W(k_{1d}, k) P_b(k) k dk \qquad (1.39)$$

where a distinction is made between the wavenumber of the one-dimensional
power spectrum k_{1d} and the wavenumber of the three-dimensional power
spectrum k. The terms $W(k_{1d}, k)$ represents a kernel function (Mo and
van den Bosch, 2010) of the form

$$W(k_{1d}, k) = A \exp\left(-\frac{k_{1d}^2 b_0^2}{2H^2}\right) \left[1 + \frac{f(\Omega_m)}{q} \frac{k_{1d}^2}{k^2} - \frac{\Gamma - 1}{4q} \frac{k_{1d}^2 b_0^2}{H^2}\right]^2 \qquad (1.40)$$

where Γ is the polytropic index, $q \equiv 2 + 0.7(1 - \Gamma)$, $b_0 = (2k_B T_0/m)^{1/2}$ with
T_0 the initial temperature of the gas, A is a constant and $f(\Omega_m)$ is related to
the peculiar velocity distortions. The term $P_b(k)$ in Eq. 1.39 represents the
baryon power spectrum which in the linear regime follows the underlying
dark matter distribution and can be expressed as

$$P_b(k) = \frac{P_{DM}}{(1 + k^2/k_J^2)^2} \quad \text{where} \quad k_J = \sqrt{\frac{2}{3}} \frac{a}{c_s t} \qquad (1.41)$$

where P_{DM} represents the dark matter power spectrum, a is the scale factor
and c_s the adiabatic speed of sound.

This kind of approach allow to compare in a relatively straightforward way
observations with results from simulations to obtain constraints on cosmo-
logical parameters and models of structures formation.

1.5   Different column density regime

In the following we are going to give some additional information on the
classes of absorbers previously presented, the aim of this section is to better
characterize the different column density regimes and their properties on
the basis of present understanding derived from observations. In this way
the reader will understand the order of magnitude of the physical quantities
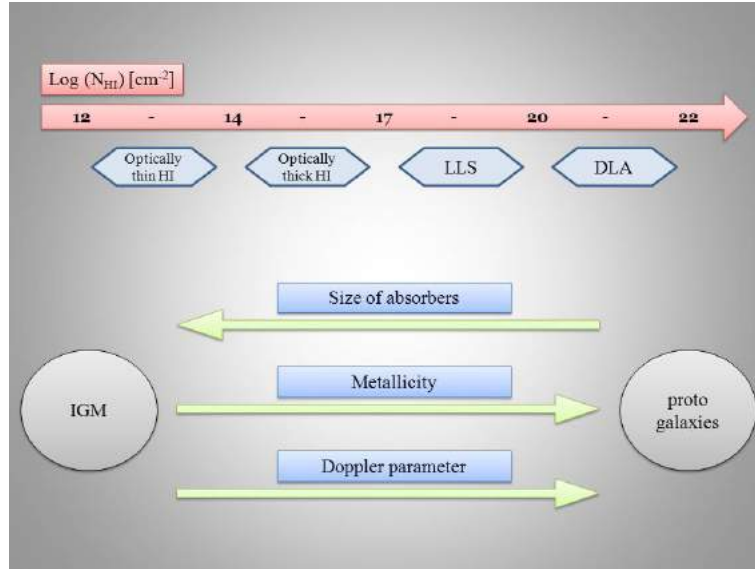
Figure 1.8: This is a schematic representation of the different classes of absorbers and the trend of their main properties varying with the column density.

involved in the study of IGM and their relations with galaxies.

In Fig. 1.8 a clarifying picture of the main properties varying with column density can be seen to better introduce and schematically represents the main scenario of quasar absorptions.

For each class of absorbers, if possible, main properties will be briefly discussed on the basis of measurements obtained from the observational techniques described before as the column density distribution function, the Doppler parameters, derived size and clustering properties of absorbers.

1.5.1   Lyman alpha forest

The *Lyman α* (*Lyα*) forest is the ensemble of lines arising from the absorption of *Lyα* resonance photons by neutral hydrogen along the line of sight towards a QSO. Since QSOs are at cosmological distances and intervening absorbing clouds of neutral hydrogen are located at different redshifts along the path of radiation, as light travels from a QSO to the observer each absorber encountered leaves its fingerprint on the spectrum at different wavelengths blueward to the *Lyα* emission line of the quasar itself.

The *Lyα* forest is characterized by narrow lines with column densities between $10^{12}cm^{-2} < N_{HI} < 10^{17}cm^{-2}$ as seen in 1.2.3. For a QSO at redshift $z \sim 3$, *Lyα* lines can be observed over a range $\Delta z \sim 1$. Since this redshift
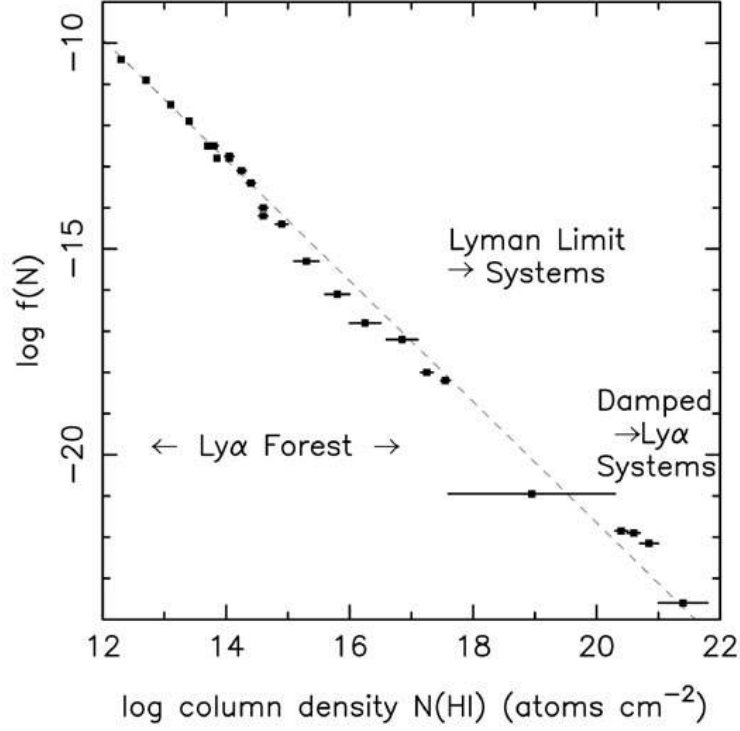
Figure 1.9: The column density distribution function for all QSO absorbers. The entire column density range from $\log N_{HI} = 12$ to 22 is represented, the dashed line shows a power law with coefficient $\beta = 1.5$ (Storrie-Lombardi and Wolfe, 2000).

range covers a significant fraction of the Hubble time at $z \sim 3$, evolution in the properties of the absorbers over this range may be expected.

Evolution of the number density of lines is usually expressed in the form

$$\frac{d\mathcal{N}}{dz} = A(1+z)^{\gamma} \tag{1.42}$$

For $2 \lesssim z \lesssim 4$ and $N_{HI} \geq 10^{14} cm^{-2}$ observations give $A \sim 3.5$ and $\gamma \sim 2.7$, those values imply significant evolution in the absorbing clouds that make up the $Ly\alpha$ forest at $z > 2$.

In Fig. 1.9 can be seen the representation of the absorption lines of the $Ly\alpha$ forest in terms of column density distribution function, observational data are best described by a simple power law of the form

$$\mathcal{F}(N_{HI}) \propto N_{HI}^{-\beta} \tag{1.43}$$

For the $Ly\alpha$ forest at $z > 2$ for example, (Hu et al., 1995) obtained

$$\mathcal{F}(N_{HI}) = 4.9 \times 10^7 cm^2 \left(\frac{N_{HI}}{cm^{-2}} \right)^{-1.46} \quad \text{for} \quad 12.3 < \log N_{HI} < 14.5 \tag{1.44}$$

Even if at present the column density distribution function in this range has been calculated with accuracy, uncertainties remain in the lower part of the curve where modern spectrographs are limited by insufficient spectral resolution and SNR. This limitation does not make clear yet if the number density of absorbers with very low column density decreases (as shown by the steepening of the CDDF) or if we simply are not yet able to "see" all the absorbers at this low column density regime ($N_{HI} \lesssim 10^{12} cm^{-2}$).

Studies on the temperature and kinematics (turbulence) of the IGM have been possible with the advent of high resolution spectroscopy since they are based on the calculation of the Doppler parameter of $Ly\alpha$ forest lines (see 1.3.3). Typical Doppler parameters have been found to be of the order of $b \sim 20km/s$, consistent with the assumption that the absorbing clouds are heated by photo-ionization ($T \sim 10^{4.5}K$), although some lines have b values as large as $100km/s$.

Another fundamental property of the absorbers is the size, which could give a hint on the nature of absorbers by direct comparison with known astrophysical objects. In principle cloud sizes can be inferred from the observed line density if one knows the number density of the absorbing clouds. If one starts with a simplified model in which the comoving number density of clouds is assumed to be the same as the number density of local galaxies, the number density can be expressed in terms of the galaxy luminosity function $\phi(L)dL$ as

$$\frac{d\mathcal{N}}{dz} = \frac{(1+z)^2}{E(z)} \int \sigma(L)\phi(L)dL \tag{1.45}$$

so using a luminosity function of the Schechter form

$$\phi(L)dL = \phi^* \left(\frac{L}{L^*} \right)^\alpha \exp\left(-\frac{L}{L^*} \right) \frac{dL}{L^*} \tag{1.46}$$

and assuming $\sigma(L) \equiv \pi R^{*2}(L/L^*)^\beta$ with $\alpha = -1$, $\beta = 0.8$ and $\phi^* = 1.5 \times 10^{-2}h^3 Mpc^{-3}$ following (Holmberg, 1975), can be found that the observed line density for $z = 3$ requires that $R^* = 1h^{-1}Mpc$.

This calculation, although should be considered just an approximation, show that the majority of the $Ly\alpha$ forest clouds at high redshift cannot be closely

associated with individual galaxies.

Another technique involving close quasar pairs has been used to infer typical sizes of $Ly\alpha$ clouds. By analyzing the frequency of line coincidences in the spectra of quasar pairs with various angular separations, typical sizes of these clouds appear to be large, of the order of $200 - 500h^{-1}kpc$, strengthening the hypothesis that $Ly\alpha$ clouds cannot be directly related to galaxies. Other important clues on the nature of $Ly\alpha$ clouds can be derived from their clustering properties: are they randomly distributed or clustered? As discussed in 1.4.1 the clustering properties of the $Ly\alpha$ lines can be studied by means of the two point correlation function. Observations from (Cristiani et al., 1995, 1997) show significant clustering only at small separations $\Delta v \lesssim 300km/s$, with a clustering amplitude $\xi(100km/s) \sim 0.5$ much lower if compared with the one from local galaxies at similar separations, showing no apparent connection between them.

Experimental evidences show that the majority of absorbing clouds responsible for the $Ly\alpha$ absorption lines at $z \gtrsim 2$ are of extragalactic origins. Nevertheless despite their distance from star forming regions, has been found experimentally that more than half of all systems with $N_{HI} \gtrsim 3 \times 10^{14}cm^{-2}$ have detectable CIV lines associated with them. The column densities of the observed CIV are in the range $10^{12} < N_{CVI} < 10^{14}cm^{-2}$ which result in a metallicity of about $Z \sim (10^{-4} - 10^{-2})Z_\odot$ for the absorbing clouds (Simcoe et al., 2004). All these observations together with results from simulations agree with the most accounted picture in present cosmology where $Ly\alpha$ clouds represents the IGM, i.e. the diffuse homogeneous medium composed by walls and filaments alternated with large voids between galaxy clusters expected in the model of structure formation.

1.5.2 LLS

Lyman-limit systems are narrow line absorption systems characterized by their capability to absorb all radiation with wavelength shorter than the Lyman limit 912Å. QSO spectrum will show a flux decrement (from $I = I_0$ to $I < 0.368I_0$) for frequencies higher than corresponding Lyman limit at the redshift of the absorber. Absorption lines associated with these systems are recognizable between thin Lyman alpha forest lines since they show a strong, heavily saturated core due to their relatively high column density.

Defining the optical depth of an absorbing cloud with HI column density

N_{HI} to Lyman-limit photons as

$$\tau(\lambda) = N_{HI}\sigma_{pi}(\lambda) = \left(\frac{N_{HI}}{1.6 \times 10^{17}cm^{-2}}\right)\left(\frac{\lambda}{912\text{Å}}\right)^3 \quad \text{for} \quad \lambda \leq \lambda_L \quad (1.47)$$

where σ_{pi} is the photo-ionization cross section. Lyman-limit systems can be defined as the ones with $\tau > 1$ for Lyman continuum radiation, which from the previous equation gives the lower column density limit for such systems $N_{HI} > 1.6 \times 10^{17}cm^{-2}$.

The column density regime covered by LLS is the most uncertain and by now represents the missing link between Lyman forest and DLAs in the column density distribution function representation (Fig. 1.9). As shown in 1.3.3 absorbers like LLS laying on the central part of the COG cannot be easily measured in column density due to saturation of the core, as a direct consequence the exact position of those systems in the column density distribution function cannot be understood and metallicity measurements had to rely on photo-ionization models.

An alternative way to measure their column densities rely on higher orders Lyman series absorption lines, like for example the $Ly\beta$ and $Ly\gamma$ which have lower oscillator strengths and so are less saturated with respect to $Ly\alpha$ lines. However those lines often appear in the UV region of the spectrum which can be measured only from space telescopes due to atmospheric absorption. As was done in the previous section the typical absorption radius can be obtained, under the assumption that the comoving number density of LLS is the same as the number density of local galaxies, from Eq. 1.45 considering that the number of LLS per unit redshift can be represented by the power law

$$\frac{d\mathcal{N}}{dz} = 0.27(1+z)^{1.55} \quad \text{for} \quad 0.01 < z < 5 \tag{1.48}$$

In this case typical radius are $R_* \sim 70h^{-1}kpc$ at $z \sim 1$ and $R_* \sim 80h^{-1}kpc$ at $z \sim 3$. These values of R_* are roughly consistent with the absorbers being clouds in galactic halos.

1.5.3   DLA and sDLA

As discussed in 1.2.3 by definition a DLA is an absorber with HI column density $N_{HI} \geq 2 \times 10^{20}cm^{-2}$, such absorbers are characterized by an absorption line profile dominated by the damping wings characteristics of the
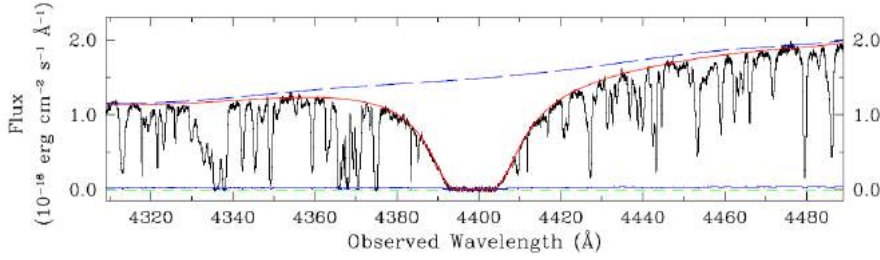
Figure 1.10: The figure shows a DLA at $z = 2.62$ toward the QSO Q0913+072, the dashed blue line represents the supposed continuum level emphasizing the large absorption with respect to surrounding $Ly\alpha$ lines (Cooke et al., 2014).

Lorentzian profile and large width which make them easily recognizable as can be seen in Fig. 1.10.

The lower limit defining the column density range of DLA was first introduced by (Wolfe et al., 1986) and it has mainly historical reason, in fact from section 1.3.3 can be seen that almost all absorbers with $N_{HI} \gtrsim 10^{19} cm^{-2}$ are characterized by line profiles with Lorentzian wings similar to the DLAs ones. So in order to take into account absorbers with column density in the range $10^{19} cm^{-2} \leq N^{HI} \leq 2 \times 10^{20} cm^{-2}$ a subclass was introduced called $subDLA^4$. DLA and subDLA are pretty much similar with the only difference, besides the column density, that hydrogen in DLAs at high redshift is largely neutral, while that in subDLAs may still be significantly ionized.

Due to their placement in the last square root part of the COG, DLAs and subDLAs will have very good measurements of their column density (see Eq. 1.34), and so their column density distribution function can be calculated. Even if in first approximation the DLAs regime can be described with the same power law used in the Lyman forest regime, a better description shows that the best approximation is represented by a function of the Schechter type

$$\mathcal{F}(N_{HI})dN_{HI} = \mathcal{F}^* \left(\frac{N_{HI}}{N_{HI}^*}\right)^{\beta} \exp\left(-\frac{N_{HI}}{N_{HI}^*}\right) \frac{dN_{HI}}{N_{HI}^*} \qquad (1.49)$$

often referred to as the Γ *function*. Best-fit values of β and N_{HI}^* typically fall in the range $-2 < \beta < -1$ and $10^{21} cm^{-2} < N_{HI}^* < 10^{21.5} cm^{-2}$ (Storrie-Lombardi and Wolfe, 2000).

It is also interesting to study the redshift evolution of the DLA population.

---

[4]Note that both DLAs and subDLAs are also Lyman-limit systems

DLAs are relatively rare, for $N_{HI} \geq 10^{20.3} cm^{-2}$ the number of lines per unit redshift interval has been found (Storrie-Lombardi et al., 1996a)

$$\frac{d\mathcal{N}}{dz} = 0.04(1+z)^{1.3 \pm 0.5} \quad \text{for} \quad 0.1 < z < 4.7 \qquad (1.50)$$

Using again Eq. 1.45 a radius $R^* \sim 30h^{-1} kpc$ at $z \sim 3$ was derived under the assumption that the comoving number density of DLAs is the same as that of local galaxies. Although this radius is somewhat larger than typical radius for HI disks of galaxies at $z = 0$, it suggests a close connection between DLAs and proto-galaxies.

From those data it is possible to derive (Storrie-Lombardi et al., 1996b) the cosmic density parameter of baryonic material associated with DLAs, Ω_{DLA}, and find that at $z \sim 3$ DLAs contain roughly half the amount of baryons that at $z = 0$ is locked up in stars and so, since star formation in the Universe occurs mainly in high-density gas clouds, we expect that DLAs are a major source of cold gas for star formation in the Universe.

Regarding the metallicities of these objects very precise and relatively easy measurements can be made thanks to

- well determined HI column density

- no ionization correction (hydrogen is almost all neutral)

- relatively high column density of metal lines that can therefore easily detected

Low-ionization species are most commonly observed, such as SiII, MnII, FeII, NiII, ZnII, CrII. Precise measurements over a large sample of DLAs with $0.4 \lesssim z \lesssim 3.4$ (Pettini et al., 1999) show that the majority of DLAs have metallicities in the range $-2 \lesssim [Zn/H] \lesssim -0.5$ i.e. about 1/100 to 1/3 of the solar metallicity. These low metallicities are consistent with the picture where DLAs represent gas clouds that have not yet experienced much star formation.

1.6 Metal absorption line systems

Metal lines more frequently encountered in QSO spectra are usually the ones with higher cosmic abundances and higher oscillator strengths, like the strong resonance doublets of CIV and SiIV, or OVI, FeII, MgII and following

CII, CaII, AlIII, NV. As a rule of thumb low ionization states like CII, MgII and FeII will trace colder regions of the IGM ($T \sim 10^4 K$) since their ionization thresholds are lower, while higher ionization states like SiIV, CIV and OVI characterized by higher ionization threshold, will indicate the presence of a hotter medium with temperature $T \gtrsim 10^5 K$.

Among different species CIV is one of the best metal tracers owing to its strong resonance doublet at $CIV\lambda\lambda 1548, 1550$Å, and relatively short frequency that allow for spectroscopy in the visible range starting from redshift $z \sim 1.2$. Moreover CIV lines fall outside the Lyman alpha forest redward the Lyman alpha emission of the QSO and so their detection is relatively easy and does not suffer from line crowding in the Lyman forest.

In 1.5.1 has been observed that most of the absorption lines of the $Ly\alpha$ forest show associated CIV lines which for this reason could traces the metallicity of the IGM. On the other side in regions where the density is near the universal mean, CIV could loses its sensitivity as a metallicity tracer, since in such regions temperature will be very high and most of the CIV will turn in CV, in this way the bulk of carbon could be missed.

Owing to its higher ionization potential OVI instead can still be considered as a good metal tracer in those regions. Nevertheless its detectability will be more difficult since the resonance frequency of the observed transition at $OVI\lambda\lambda 1031, 1037$Å will fall inside the Lyman alpha forest and will be more subject to line blending. Moreover UV spectroscopy from space based observatory (HST-Hubble Space Telescope) is required to recover a complete sample of absorbers along the line of sight.

Finally MgII and FeII are often detected in QSO spectra from doublets $MgII\lambda\lambda 2976, 2803$Å and $FeII\lambda\lambda 2600, 2586$Å but their low resonance frequencies will take their absorption lines in the infrared region of the spectrum. This region is subject to contamination from dust depletion, indeed typical sizes of dust grains are of the same order of the infrared wavelengths which are scattered by dust particles along the line of sight. The net effect of this process is the loss of sensitivity to the metal content of the IGM which makes those systems less efficient metal tracers.
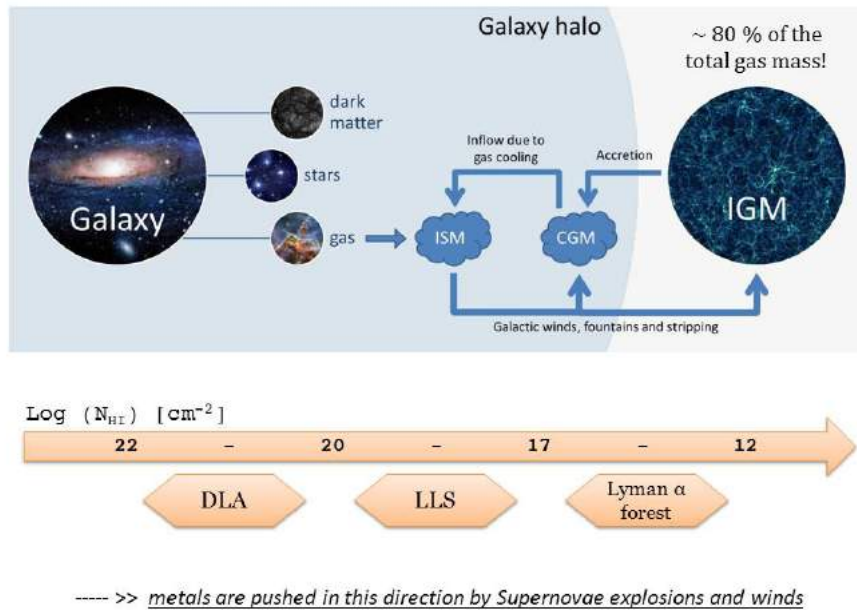
Figure 1.11: This image has been resumed from Fig. 1.2 and expresses visually the basic concept behind metals enrichment of the IGM. Since metals are produced in stars one expects that the concentration of metals decreases from the left to the right. Assuming that high column density lines as DLA traces galaxies and that Lyman forest lines traces the IGM, in this view strong metal lines should be associated with DLA while metal absorptions should become weaker going towards lower column densities.

1.6.1   Observing metals in the IGM

The Inter-Galactic Medium represents the largest reservoir of baryons in the Universe, its main constituents are Hydrogen and Helium and it contains approximately the 80% of the total gas mass of the Universe.

The study of its chemical composition across different cosmic epochs and at different density regimes represents a precious source of information on star formation history and feedback mechanisms inside galaxies.

This can be better understood by looking at Fig. 1.11. Since metals are produced in stars, it comes natural to think that greater is the distance from stars lower is the probability to find metals. In order to push metals in the IGM one has to invoke very efficient mechanisms like supernovae, galactic winds or galactic fountains (Ferrara et al., 2000; Aguirre et al., 2001). Obviously to go farther more energy and more time is required.

Now if we assume, as represented in the picture, that the column density of

QSO absorption lines decreases as the distance from galaxies increases, we expect to find strong metals lines associated with DLA and, on the other hand, metals lines which are weaker as the column density of the associated $Ly\alpha$ lines decreases.

This basic picture is confirmed by observations which show that in correspondence of high density neutral hydrogen clouds such as DLA or subDLA there are generally higher concentrations of metals, at variance with very low density regions traced by $Ly\alpha$ lines characterized by small concentrations of metals.

Taking the CIV doublet as metal tracer and focusing on the IGM, typical CIV absorptions associated to the Lyman forest lines show column densities going from $\log N_{CIV} \sim 14.5 cm^{-2}$ down to the detection threshold $\log N_{CIV} \leq 12.0 cm^{-2}$.

Those very low column densities, together with the small Doppler parameters typical of the IGM (1.5.1), translates from the observational point of view into narrow and very weak absorption lines. In order to be able to observe those lines there are two main technical requirements that an instrument should have for a successful observation:

1. **High spectral resolution:** the ability to clearly distinguish two narrow absorption lines very close each other

2. **High sensitivity:** the ability to distinguish very weak absorption lines from the surrounding noise

At this point our knowledge is essentially limited by technical requirements since increasing the sensitivity allows for lowering the detection threshold for the minimum detectable column density.

To go deep in column density and probe the $\log N_{CIV} < 12.0 cm^{-2}$ regime it is equivalent, in the same schematic view of Fig. 1.11, to go even farther from galaxies to probe the most representative regions of the IGM, i.e. regions closer to the mean density of the Universe.

This subject will be discussed in Chapter 4, in the next chapter will be described in detail the technical requirements to push observations down to $\log N_{CIV} < 12.0 cm^{-2}$ in order to probe the deep IGM and very weak associated metals.

Chapter 2

Observation of the IGM

In the previous chapter it has been pointed out that a successful observation of the IGM requires both high spectral resolution and high sensitivity. These two requirements become even more compelling if one want to push in depth observations down to the lowest density regions of the IGM.

At present high resolution can be best achieved using the so called *echelle spectrographs*. In the following the main characteristics of echelle spectroscopy will be introduced in order to better understand the impact of resolution and sensitivity on real spectra.

On the other hand to obtain high sensitivity measurements is not so straightforward. The photons collecting area strongly influences the sensitivity of the measurement but as will be discussed in this chapter sometimes even the largest available telescopes could not be enough. The final result should be a compromise between the collecting area of the telescope and the exposure time.

In this chapter will be discussed as it has been possible to obtain a sample of QSO spectra with very high sensitivity (SNR > 100) from the combination of archival data of two similar echelle spectrographs: the HIRES at the Keck telescopes and UVES at the Very Large Telescopes.

The sample presented here has unique characteristics in term of sensitivity and wavelength coverage which have been imposed by the scientific aim of this thesis: the study of the lowest density regions of the IGM.
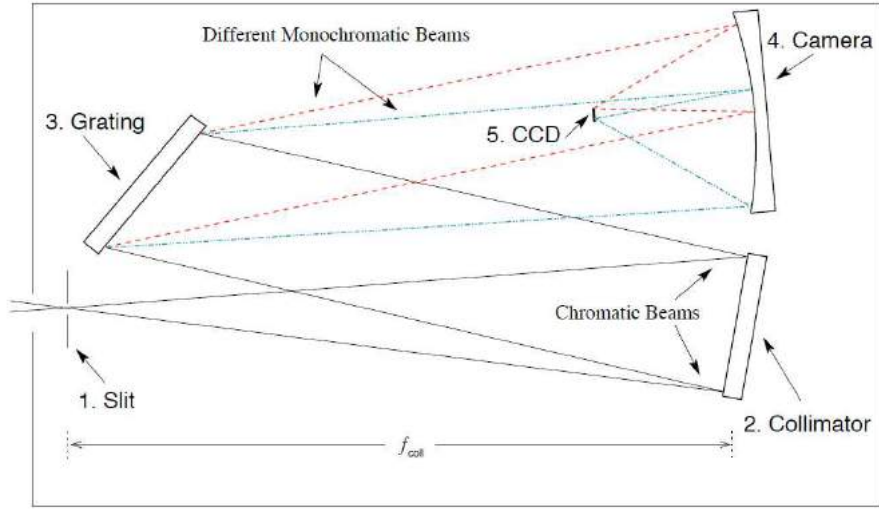
Figure 2.1: Schematic representation of the main components of a spectrograph and light path from slit to detector. Chromatic beams (solid lines) are parallelized by the collimator toward the grating which reflects the light at different angles depending from wavelength. Different monochromatic beams (dash-dash and dash-dot lines) are then focused at different spatial locations on the detector plane by the camera. The distance of the slit from the collimator is the collimator focal length, f_{coll}, and the distance of the detector from the camera mirror is the camera focal length, f_{cam} (Churchill, 2016).

2.1 The high resolution spectrograph

An Echelle spectrograph can be considered as a variant of standard grating spectrograph. The most common grating spectrograph (schematics shown in Fig. 2.1) is composed basically by an entrance slit (1) located at the focal plane of the telescope, a light beam collimator (2), a ruled grating (3), which is the dispersive element, and a camera (4) which focuses the beam onto the detector (5) (usually a Charged Coupled Device or CCD).

The light enters the spectrograph through the slit, then hits the collimator which is an optical element whose distance from the slit is exactly equal to its focal length f_{coll} in order to collimate and reflect the light beam. The collimated beam comes diffracted by the grooves of the grating. The grating is tilted with respect to the incoming light path. The reflected light beam, still collimated, is then focused onto the detector by the camera which represents the last optical element of the spectrograph and is characterized by a focal length f_{cam} which also corresponds to the detector distance from the camera. Dispersed photons can be finally counted by the CCD pixels

which discretized the spectrum measurement.

In the following the basic properties will be discussed to better understand how high resolution can be achieved in practice and its important role in the measurement of a spectrum, throughout the entire section we will refer to the upcoming book of Chris Churchill (Churchill, 2016).

2.1.1   Slits

The slit receives the focused image of the source and represents the entrance of the spectrograph. Slits are typically rectangular apertures characterized by an height h and a width w chosen by taking into account the specific requirements of the observation. In particular the width of the slit is very important since as will be shown in 2.1.7 it is related to the spectral resolution: narrower slits mean higher resolution.

On the other hand one has to take into account the effects of the atmosphere which make a point like source to appear as a blur with finite extension on the focal plane of the telescope, this effect is known as *seeing* and the resulting image on the focal plane is called *Point Spread Function* or PSF of the telescope image.

The PSF can be represented in good approximation by a Gaussian of the form

$$\mu_\lambda(r) = \frac{1}{2\pi\sigma_\lambda^2} \exp\{-\frac{r^2}{2\sigma_\lambda^2}\} \qquad (2.1)$$

where r is the physical coordinate in the telescope focal plane and σ_λ is the profile dispersion width which is wavelength dependent. The seeing can be measured from the full width at half maximum of the PSF defined as $FWHM_\lambda = 2.35\sigma_\lambda$ and has to be taken into account for the slit width selection. A width smaller than the seeing disk will results in a loss of light so usually the slit width is a compromise between the desired resolution and the efficiency determined by the slit losses due to the seeing.

The effects of seeing could be avoided in principle by eliminating the atmosphere, this is possible nowadays owing to space telescope like the *Hubble Space Telescope* or HST. In these cases the PSF reduces to the *Airy disk*[1] of the telescope whose width can be calculated in radians as $\sin\theta \approx 1.22\frac{\lambda}{D}$ where λ is the wavelength and D is the diameter of the telescope. This

---

[1]The Airy disc of a telescope is the particular image produced by the diffraction of a beam of light passing through the hole represented by the entrance of the telescope itself.
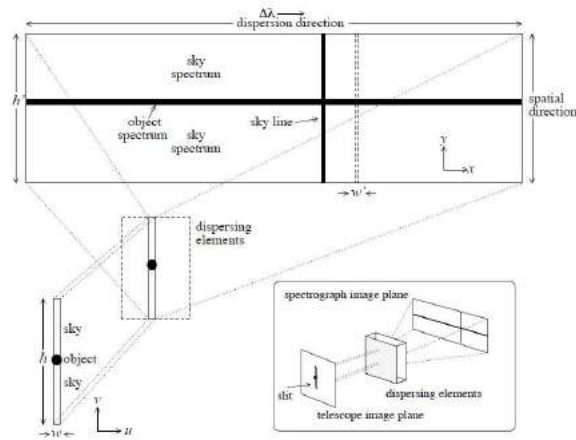
Figure 2.2: A simplified image showing the projection of the slit on the detector plane. In the focal plane of the telescope slit has dimension w and h in coordinate u and v respectively. After passing through the dispersive element light is dispersed in the x coordinate (dispersion direction) so that for each x a monochromatic image of the slit is produced. Along the y coordinate (spatial direction) slit is imaged such that sky and object are spatially mapped. Dimensions of the slit on the detector plane become h' and w' (Churchill, 2016).

situation represents the so called *diffraction limit.*

The height of the slit is typically larger than the width in order to include not only the PSF of the image, but also a significant portion of the sky. The observed sky spectrum will be useful for identification of the background and atmospheric emission lines which has to be removed in order to recover the exact spectrum of the source.

By looking at Fig. 2.2 can be understood another important characteristics of a grating spectrograph: they are usually configured such that the dispersion occurs perpendicularly to the height of the slit, therefore this direction is called the *dispersion direction* while the other is called the *spatial direction.* To better understand one has to take a look at the image of the slit after it comes dispersed on the spectrograph plane: by fixing an arbitrary coordinate on the dispersion direction, each one of the vertical cut parallel to the spatial direction represents a monochromatic image of the slit at the wavelength established by the fixed coordinate in the dispersion direction. However even if this represents the optimal configuration, depending on the specifics of the instrument design it cannot be always achieved. This is the case for example of the medium resolution spectrograph *X-SHOOTER* (SPI,
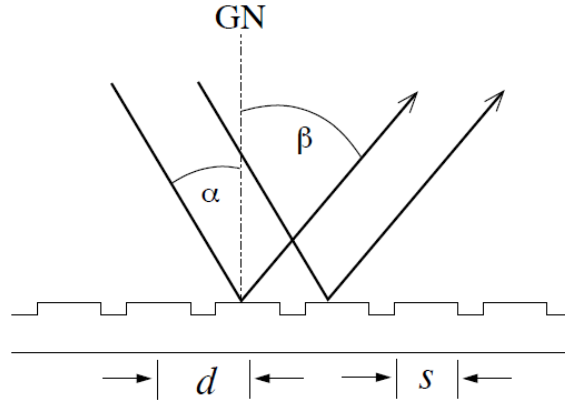
Figure 2.3: Schematic representation of a grating made of grooves, or facets, characterized by a edge-to-edge length s and a center-to-center separation d. Incidence angle α and wavelength dependent reflection angle β are measured with respect to the grating normal GN (Churchill, 2016).

2006) whose dispersion occurs in a slightly transversal direction with respect to the perpendicular of the slit.

2.1.2   Dispersing element: diffraction gratings

The diffraction grating could be considered as the main component of the spectrographs since is the one responsible for dispersion of the light. A typical grating is represented schematically in Fig. 2.3 and is composed by equally spaced grooves called *facets*. There are three main parameters which characterize a grating: its length L, the distance between the centers of the facets d and their width s, the number of facets N_f can be easily retrieved as L/d.

Basically a monochromatic beam of light coming from an angle α is diffracted and reflected to a wavelength dependent angle β, the combined effect of all frequencies produce the wavelength dependent *intensity function* which is focused on the CCD to produce the final spectrum.

The angular dependence of the intensity function is governed by two distinct interference pattern: the first one is due to phase differences φ_{cc} between the centers of adjacent facets and is called the *interference function*, the second one is due to phase differences φ_{ce} between the center and the edge of each facet and is called the *blaze function*, lets try to better understand how the intensity function comes out from this two interference patterns (see
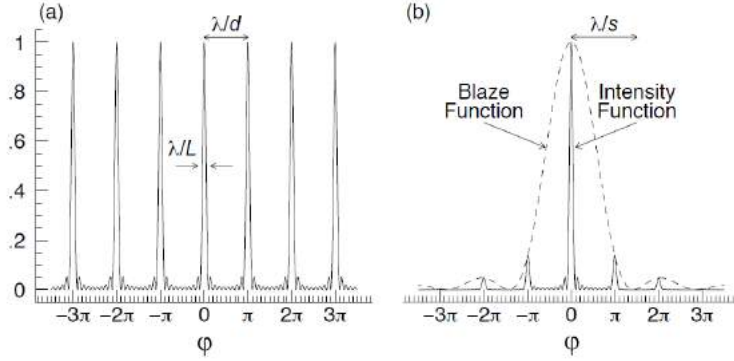
Figure 2.4: (a) The periodic interference function for a monochromatic beam of light plotted as a function of the phase difference. Note that each peak, located at $n\pi$ with n integer, satisfy the grating equation and is thus diffracted at different β (b) The dashed line represents the blaze function which modulates the interference function giving the overall intensity pattern. The two curves are plotted as functions of the center-to-center phase difference (Churchill, 2016).

Fitzpatrick (2013); Morin (2010) for a detailed derivation).

The interference function illustrated in Fig. 2.4(a) can be described by the formula

$$I_i(\varphi_{cc}) = \frac{\sin^2(L/d)\varphi_{cc}}{\sin^2\varphi_{cc}} \tag{2.2}$$

for a given λ constructive interference occurs when the condition $\varphi_{cc} = n\pi$ is satisfied, where the integer n is called the diffraction order. So by considering the relation between the phase φ_{cc} and λ described by the equation

$$\varphi_{cc} = \frac{\pi d}{\lambda}(\sin\alpha + \sin\beta) \tag{2.3}$$

imposing the condition $\varphi_{cc} = n\pi$ is it possible to find the angle β where to place the camera to observe the peak of the intensity by the formula

$$\frac{n\lambda}{d} = \sin\alpha + \sin\beta \tag{2.4}$$

this equation is called the *grating equation* and tell us the relationship between the diffraction angle and the wavelength for the individual orders n of the grating. The grating equation is very important because can easily explain the basilar behavior of a grating as illustrated in Fig. 2.5. Moreover since there are multiple $n\lambda$ that satisfy the grating equation, wavelengths in recursive orders could be diffracted in the same direction. This results in spatial overlap of wavelengths satisfying $(n-1)\lambda_{n-1} = n\lambda_n$ which is obtained by the ratio of the grating equation for adjacent orders.
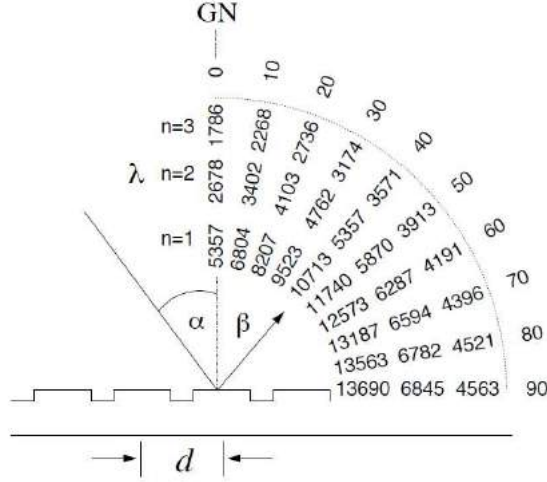
Figure 2.5: Basilar behavior of a grating with $1/d = 1200 \; facets \; mm^{-1}$ for an incident angle $\alpha = 40°$. Wavelength for orders $n = 1, 2$ and 3 are tabulated at the corresponding diffraction angle as obtained by the grating equation. Note that for $n = 0$ there is no dispersion and chromatic beam of light is diffracted at $\beta = -\alpha$ (Churchill, 2016).

The spatial overlap can cause confusion in the recorded spectrum, as will be discussed in the next sections the solution to this problem depends on the spectrograph configuration. The region of the spectrum in individual orders which does not overlap with adjacent orders is called the *free spectral range*. By differentiating the grating equation

$$\frac{d\beta}{d\lambda} = \frac{n}{d \cos \beta} = \frac{\sin \alpha + \sin \beta}{\lambda \cos \beta} \tag{2.5}$$

becomes clear that the angular dispersion (which is responsible for the spectral resolution) increases with larger n and smaller d.

The interference function is not enough to understand the final intensity function behavior, indeed the presence of phase differences φ_{ce} due to the finite extension of the facets are responsible for the presence of the blaze function described by equation

$$I_b(\varphi_{ce}) = \frac{\sin^2 \varphi_{ce}}{\varphi_{ce}^2} \tag{2.6}$$

where the relation between angle β and wavelength λ is represented by the formula

$$\varphi_{ce} = \frac{\pi s}{\lambda}(\sin \alpha + \sin \beta) \tag{2.7}$$

At a variance with the interference function Eq. 2.6 does not represent a periodic function, this can be understood by solving Eq. 2.4 for λ and substituting into Eq. 2.7 which yields the relative blaze intensity for order n

$$I_b(n) = \frac{\sin^2 n\pi(s/d)}{[n\pi(s/d)]^2} \tag{2.8}$$

The intensity peaks for the order $n = 0$ where there is no dispersion while is significatively reduced for subsequent orders, in the next section will be shown how this problem can be avoided by tilting the facets, a process called *blazing* which shift the peak of the blaze function.

In Fig. 2.4(b) is shown the combined effect of the blaze function which modulates the interference function to give the intensity function

$$I(\lambda) = I_i(\varphi_{cc}) \cdot I_b(\varphi_{ce}) = \frac{\sin^2(L/d)\varphi_{cc}}{\sin^2 \varphi_{cc}} \cdot \frac{\sin^2 \varphi_{ce}}{\varphi_{ce}^2} \tag{2.9}$$

From the picture can be seen that for a given λ appearing in multiple orders, the relative intensity is dominated by the blaze function and governed by Eq. 2.8.

2.1.3   Blazing

Equation 2.5 shows that when a certain angular dispersion is desired, the ratio d/n has to be properly fixed and the resulting relative intensity from Eq. 2.9 needs to be maximized in order to fully exploit the spectrograph capabilities. On the other hand Eq. 2.8 shows that the intensity is maximized for $n = 0$, which means no dispersion, while higher orders show significantly reduced intensity. This comes directly from the shape of the blaze function and the only way to overcome this problem is to modify it in some way.

The blaze peak can be shifted by tilting by an angle ϕ the grating facets, a process called *blazing*. In this way the maximum of the blaze function of Fig. 2.4(b) can be conveniently placed on the desired order n. The net effect of blazing a grating is to modify the blaze function such that a peak for all n occurs when the incident and diffraction angles are symmetric about the facets normal (see Fig. 2.6) which occurs at different λ for each n.

This can be clearly understood by redefining the equations for the blaze function in the case of a blazed grating, defining the facets length as $s = d\cos^2 \phi$ and the incident angle with respect to the facets normal FN as $\bar{\alpha} = \alpha - \phi$
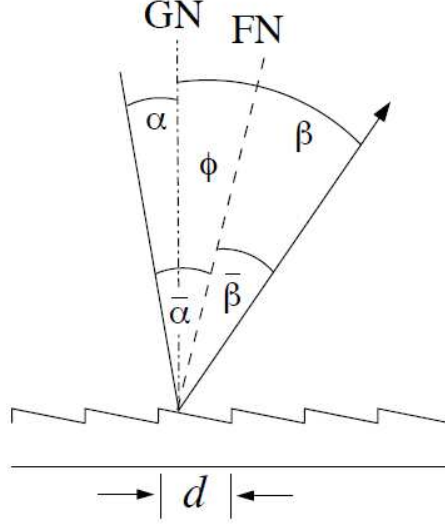
Figure 2.6: Tilting the facets at an angle ϕ a phase shift is introduced and the maximum of the intensity function can be maximized for $n > 0$. The blaze function then peaks when $\bar{\beta} = -\bar{\alpha}$ with respect to the facets normal FN (Churchill, 2016).

and the diffraction angle $\bar{\beta} = \beta - \phi$, Eq. 2.7 can be rewritten as

$$\varphi_{ce} = \frac{\pi s}{\lambda}(\sin \bar{\alpha} + \sin \bar{\beta}) = \frac{\pi d \cos^2 \phi}{\lambda}[\sin(\alpha - \phi) + \sin(\beta - \phi)] \qquad (2.10)$$

Since the condition for constructive interference is $\varphi_{ce} = 0$, maxima will occur when $\bar{\beta} = -\bar{\alpha}$ or with respect to the grating normal GN when $\alpha - \phi = -(\beta - \phi)$ or $\phi = (\alpha + \beta)/2$.

It can be demonstrated that this will occur for all orders n starting from the grating equation rewritten as

$$\frac{n\lambda}{d} = \sin \alpha + \sin \beta = \sin(\bar{\alpha} + \phi) + \sin(\bar{\beta} + \phi) \qquad (2.11)$$

and applying the condition $\bar{\beta} = -\bar{\alpha}$ which yields

$$\frac{n\lambda_b}{d} = 2 \sin \phi \cos \bar{\alpha} = 2 \sin \phi \cos(\alpha - \phi) \qquad (2.12)$$

where λ_b represents the blaze wavelength for order n which can be used in Eq. 2.10 to obtain

$$\varphi_{ce} = \frac{n\pi \cos \phi[\sin \bar{\alpha} + \sin \bar{\beta}]}{2 \tan \phi \cos \bar{\alpha}} \qquad (2.13)$$

From this equation it can be seen that the maxima condition $\varphi_{ce} = 0$ is satisfied for all n unless $\bar{\beta} = -\bar{\alpha}$.

Again recalling that the intensity function comes from the interference function modulated by the blaze function, the relative intensity for order n at each angle $\bar{\beta}$ becomes

$$I(n,\bar{\beta}) = I_i(n) \cdot I_b(n,\bar{\beta}) = \frac{\sin^2\left(\frac{n\pi\cos\phi[\sin\bar{\alpha}+\sin\bar{\beta}]}{2\tan\phi\cos\bar{\alpha}}\right)}{\left(\frac{n\pi\cos\phi[\sin\bar{\alpha}+\sin\bar{\beta}]}{2\tan\phi\cos\bar{\alpha}}\right)^2} \qquad (2.14)$$

where

$$\lambda(n,\bar{\beta}) = \frac{d}{n}(\sin\bar{\alpha} + \sin\bar{\beta}) \qquad (2.15)$$

is the wavelength at each $\bar{\beta}$ for the order n.

Eq. 2.14 and Eq. 2.15 represent the basilar equations for the spectrograph design. These equations show that a central choice for the design is represented by the spectral dispersion (which finally defines the grating resolving power) which increases with increasing n/d. Higher order means higher dispersion while lower order means lower dispersion.

2.1.4   Low order spectrographs

As a first example let's examine a grating with a blaze angle $\phi = 18°$, $1/d = 600$ facets mm^{-1} and $\bar{\alpha} = 22°$, in Fig. 2.7 the solutions to Eq. 2.14 and Eq. 2.15 for the first orders $n = 1, 2$ and 3 are shown.

This configuration is for a spectrograph covering the wavelength range of $4000 - 8500$ in order $n = 2$ for a camera centered at $\bar{\beta} = 22°$ which subtends roughly $20°$ ($10° \leq \bar{\beta} \leq 30°$).

As can be seen from the picture, blaze functions peaks at $\bar{\beta} = 22°$ for all n, but they have different width. Moreover the plot in polar coordinates shows that for each order the wavelengths are diffracted to different angles $\bar{\beta}$ such that the peaks of the blaze functions of adjacent orders will not overlap in the wavelengths space (Fig. 2.7(b)), in this way it is possible to select only the light coming from an order by means of appropriate band-pass filters.

In this case filters have to be used in order to block radiation with $\lambda \leq 4860$ from order $n = 3$ and $\lambda \geq 9140$ from order $n = 1$ and eliminate potential confusion (as can be deduced from Fig. 2.7(b) in this example a little contamination from secondary maxima of orders 1 and 3 still remains but their intensities are so low that become significant only near the edges of the filter where the intensity of the order 2 rapidly goes down).
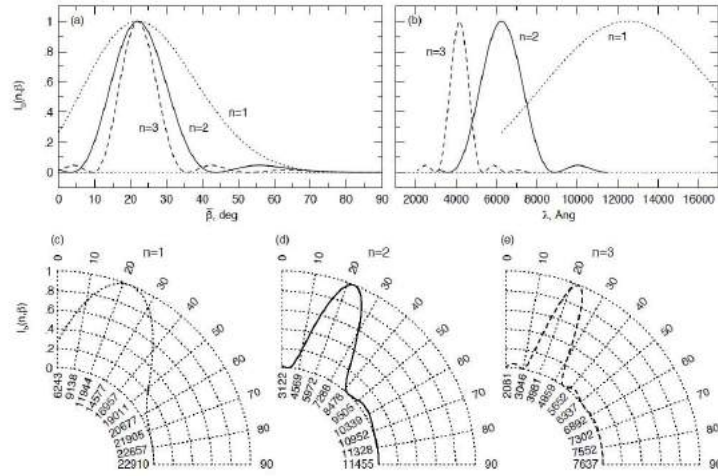
Figure 2.7: Solutions to Equations 2.14 and 2.15. (a) The dependence of the blaze functions on the diffraction angle $\bar{\beta}$ (b) The blaze functions with wavelength λ (c, d, and e) The blaze functions of the first three orders represented in polar coordinates. This representation provides the spatial relationship between λ, $I_b(n, \bar{\beta})$ and $\bar{\beta}$ (Churchill, 2016).

2.1.5   Echelle spectrographs

When higher resolution is required it has been shown (Eq. 2.5) that higher diffraction orders must be exploited. For a configuration with a blaze angle $\phi = 52°$, $1/d = 80$ facets mm^{-1} and $\bar{\alpha} = 70°$ solutions to the equations 2.14 and 2.15 are shown in Fig. 2.8 for orders $n = 32 - 70$.

The situation is similar to the previous case but this time each blaze function covers a narrow range of wavelengths due to the higher dispersion, and the most striking aspects is that each one of them are aligned at $\bar{\alpha} = 70°$.

To recover a spectrum which goes from ~ 4000 to ~ 8500 becomes nearly impossible with an approach similar to the low orders one because, as becomes clear from the polar coordinates plot, each one of the blaze function of the 38 considered orders will spatially overlaps at the blaze peak generating to much confusion.

In this case since the use of filters is not possible, diffraction orders can be spatially separated by mean of a second grating or prism, called *cross disperser*, placed along the light path following the *echelle* grating, the spectrograph in this configuration is called *echelle spectrograph*.

The cross disperser is used in low order configuration, usually $m_X = 1$, and is oriented such that facets direction is at a right angle with respect to the
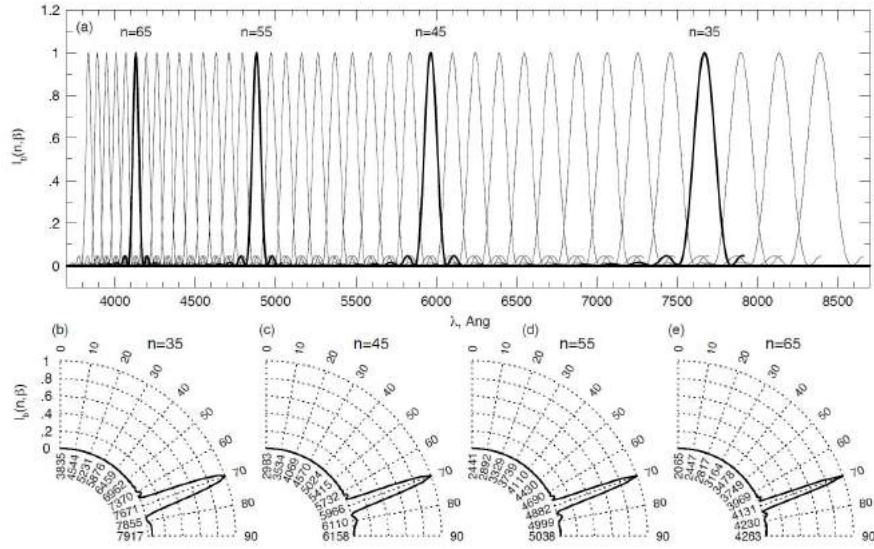
Figure 2.8: Solutions to Equations 2.14 and 2.15 for orders $n = 32 - 70$ of an echelle spectrograph with a blaze angle $\phi = 52°$, $1/d = 80$ facets mm^{-1} and $\bar{\alpha} = 70°$. (a) The blaze functions with wavelength λ. (b, c, d, and e) The blaze functions shown in polar coordinates representation (for $n = 35$, 45, 55, and 65), which provides the angular relationship between λ, $I_b(n, \bar{\beta})$ and $\bar{\beta}$ (Churchill, 2016).

direction of echelle facets, this is to ensure that dispersion in the two gratings occurs perpendicularly to each other. The net result of this configuration is to "spread" wavelength over a bi-dimensional plane where each dimension is represented by the dispersion angle of the two grating: $\bar{\beta}$ for the echelle and $\bar{\beta}_X$ for the cross disperser (Fig. 2.9(c)). At such a high dispersion orders of the echelle grating along the $\bar{\beta}$ dimension come separated along the $\bar{\beta}_X$ dimension by the low dispersion relation of the cross disperser.

In this case the slit height has to be properly chosen to avoid overlap in the spatial direction (see 2.1.1). Indeed in the blue region, where the separation between adjacent orders along the $\bar{\beta}_X$ dimension becomes smaller, an excessively high slit will result in spatial overlap of the lower part of the slit in one order with the upper part of the slit in the following order.

The final relative intensity of the spectrum comes from the echelle blaze functions modulated by the cross disperser blaze function as shown in Fig. 2.9(a), the grating equation solution for the cross disperser is also shown in a polar coordinates plot (Fig. 2.9(b)).

Main advantages of echelle spectroscopy are the higher resolution with the
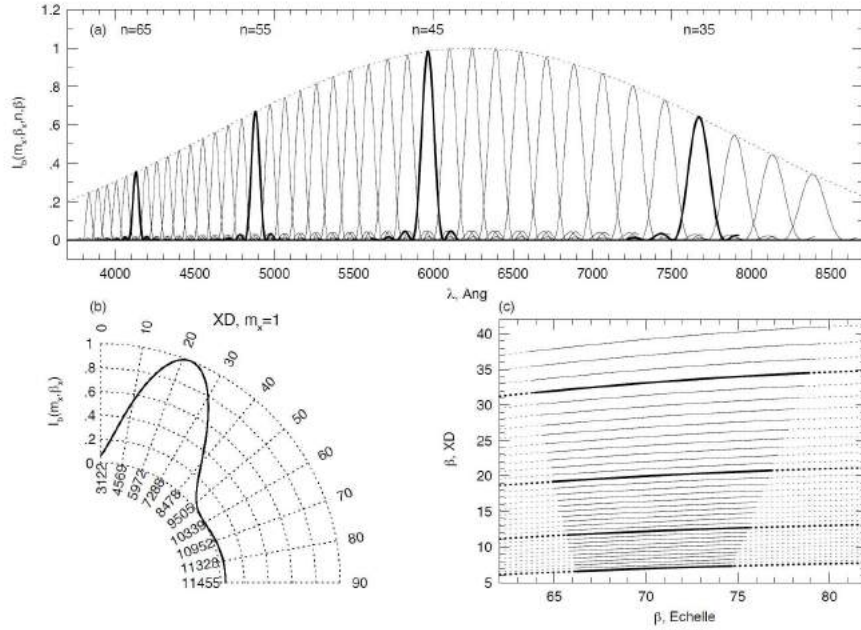
Figure 2.9: Same as Fig. 2.8 but this time cross dispersed in order $m_X = 1$ with a grating blazed at $\phi = 14°$, $1/d = 1200$ facets mm^{-1} and $\bar{\alpha} = 22°$. (a) The dependence of the echelle blaze functions on wavelength λ. The dotted curve is the cross disperser blaze function. (b) The cross disperser blaze function in polar coordinates representation. (c) The mapping of the echelle orders n, in the $\bar{\beta}_X - \bar{\beta}$ plane. Wavelength increases from the lower left to the upper right, orders $n = 35$, 45, 55, and 65 are presented as thick curves. The solid curves provide the free spectral range of each order (Churchill, 2016).

maximum wavelength coverage and less confusion between adjacent orders. However, the high number of separated orders makes difficult to recover the final complete and calibrated mono-dimensional spectrum from the diffracted bi-dimensional image, a subject which will be discussed in more detail in the next chapter.

2.1.6   Grating resolving power

In section 2.1.2 the intensity function of a monochromatic beam of light has been discussed, as can be seen from Fig. 2.4(b) the resulting intensity is not a $\delta - function$[2] but has a finite width resulting from smearing of

---

[2]In spectroscopy the theoretical spectrum of a monochromatic beam of light can be represented by the $\delta(\lambda)$ function which will be zero everywhere except for the point λ_0 corresponding to the single wavelength of that light.

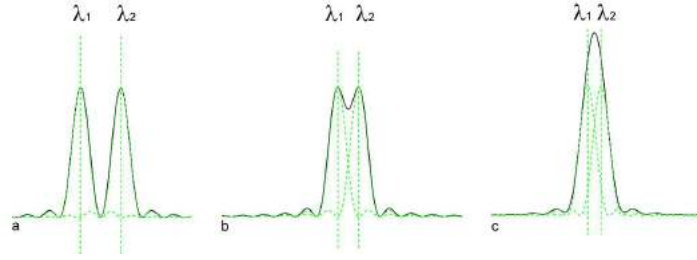*Figure 2.10:* The intensity function for two wavelengths λ_1 and λ_2. Three situations are represented: (a) the distance between the two functions is greater than $\Delta\lambda_g$ and their peaks are clearly resolvable, (b) the distance between the two functions is equal to $\Delta\lambda_g$ and their peaks are barely resolvable, (c) the distance between the two functions is less than $\Delta\lambda_g$ and their peaks are not resolvable anymore.

intensity over a small range of phase differences which can be obtained from the grating equation

$$\Delta\varphi_{cc}(n) = \frac{\lambda}{L} \tag{2.16}$$

This result can be converted into the wavelength space by differentiating the grating equation yielding

$$\Delta\varphi_{cc}(n) = \frac{n}{d}\Delta\lambda_g \tag{2.17}$$

Equating Eq. 2.16 and Eq. 2.17 gives

$$\Delta\lambda_g = \frac{\lambda}{n}\frac{d}{L} \tag{2.18}$$

This result implies that in real situations the full shape of the intensity function of chromatic light for wavelength closer than $\Delta\lambda_g$ cannot be *resolved* by the grating. As illustrated in Fig. 2.10 the intensity functions for two wavelength λ_1 and λ_2 closer than $\Delta\lambda_g$ begin to merge each other masking their original shapes.

For this reason it can be defined the so called *resolving power* of the grating, for each λ, as

$$R_g = \frac{\lambda}{\Delta\lambda_g} = n\frac{L}{d} \tag{2.19}$$

The resolving power or *resolution* shown in Eq. 2.19 is a characteristics of the grating, but the final *spectrograph resolving power* comes from the combined effect of the grating and the entrance slit.

2.1.7   Spectrograph resolving power

The spectrograph resolving power depends from the so called *spectral purity*. The spectral purity represents a measure of the width of a monochromatic image of the slit.

In practice considering the light path from the slit, hitting the collimator and reflected to the grating, it is a straightforward conclusion that light coming from one edge strikes the grating with a slightly different angle α with respect to the light coming from the opposite edge of the slit. So if w is the width of the slit, the range of all incident angles on the grating is

$$d\alpha = \frac{w}{f_{coll}} \tag{2.20}$$

this formula can be derived by geometrical considerations recalling that f_{coll} represents the focal length of the collimator (Fig. 2.1). Similarly the width w' of the image of the slit on the camera focal plane in physical coordinates, x, can be derived from the focal length of the camera f_{cam} by the relation

$$dx = f_{cam} d\beta \tag{2.21}$$

The relation between $d\alpha$ and $d\beta$ can be obtained differentiating the grating equation

$$\frac{n}{d} d\lambda = \cos\alpha \, d\alpha + \cos\beta \, d\beta \tag{2.22}$$

which for a monochromatic source $(d\lambda = 0)$ gives

$$d\beta = -\frac{\cos\alpha}{\cos\beta} d\alpha \tag{2.23}$$

Finally using equations 2.20, 2.21 and 2.23 the width w' of the image of the slit on the camera focal plane can be obtained as

$$w' = f_{cam} d\beta = -w \frac{f_{cam}}{f_{coll}} \frac{\cos\alpha}{\cos\beta} \tag{2.24}$$

where the minus sign derives from the fact that the image is inverted.

The conversion of the slit image width from physical coordinates, x, into dispersion coordinates, λ, is given by

$$\Delta\lambda = w' \frac{d\lambda}{dx} = w' \frac{d\lambda}{d\beta} \frac{d\beta}{dx} = \frac{w'}{f_{cam}} \frac{d\lambda}{d\beta} \tag{2.25}$$

where Eq. 2.21 has been substituted for $d\beta/dx$. From Eq. 2.22 one can obtain the angular dispersion

$$\frac{d\lambda}{d\beta} = \frac{d}{n} \cos\beta \tag{2.26}$$

which substituted into Eq. 2.25, together with Eq. 2.24 for w', gives the spectral purity

$$\Delta\lambda = -\frac{w}{f_{coll}}\frac{d}{n}\cos\alpha \tag{2.27}$$

which effectively represents the *full width at half maximum* or FWHM of the intensity function produced by a monochromatic beam of light on the detector plane after passing trough the spectrograph elements.

The spectrograph resolving power can be defined from the ratio of the grating equation 2.4 to the spectral purity 2.27 as

$$R = \frac{\lambda}{\Delta\lambda} = \frac{f_{coll}}{w}\frac{n^2}{d^2}\left(\frac{\sin\alpha + \sin\beta}{\cos\alpha}\right) \tag{2.28}$$

by comparison with the grating resolving power of Eq. 2.19 becomes clear that the spectrograph resolving power will be lower as a result of the degradation of purity of the incoming spectrum due to the finite width of the slit. The last equation shows that in order to maximize spectrograph resolution narrow slits are required, and the trade off is that too narrow slits will obscure a portion of the source's image. As already discussed in section 2.1.1, this will occur if the seeing disc is larger than the width of the slit producing the so called *slit losses* which will be discussed in **??**.

2.1.8 The Instrumental Spread Function

In the previous paragraph the spectral purity $\Delta\lambda$ given by Eq. 2.27 has been defined as the FWHM of the intensity function produced by a monochromatic beam of light on the detector plane. In practical cases this function, sometimes called the *Instrumental Spread Function* or ISF, is very important since it characterizes the instrument and its basic behavior.

Suppose for example to have a narrow absorption line in a spectrum, if the width of this line is smaller than the FWHM of the ISF, it will appear in the spectrum with about the same shape of the ISF. Only lines characterized by a width larger than the FWHM of the ISF will be resolved. For this reason the spectral purity $\Delta\lambda$ can also be called the *resolution element*.

In general the ISF of a spectrograph can be approximated by a normalized and symmetric Gaussian function as

$$\Phi_\lambda()\Delta\lambda = \Phi_\lambda(\lambda' - \lambda) = \frac{1}{\sqrt{2\pi}\Delta\sigma}\exp[-\frac{(\lambda' - \lambda)^2}{2(\Delta\sigma)^2}] \tag{2.29}$$
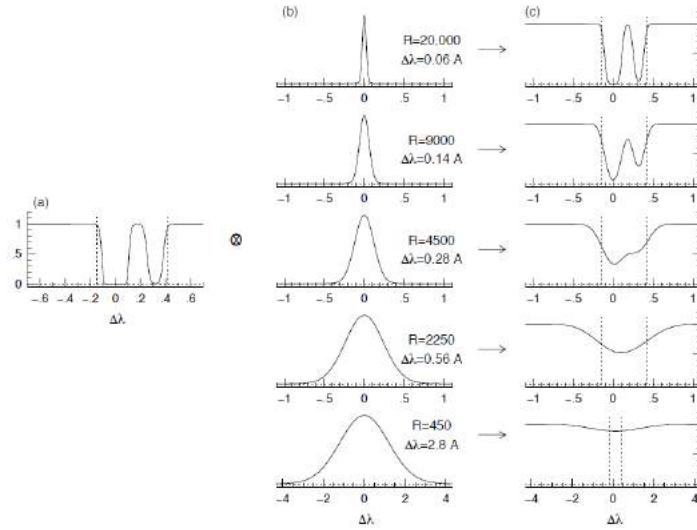
Figure 2.11: Illustrations of the reduction in spectral purity via convolution of the ISF for five typical spectroscopic resolutions, $R = 450$, 2250, 4500, 9000, and 20000. (a) The observed flux illustrating the intrinsic $Ly\alpha$ profile entering the spectrograph in relative wavelength units, centered at 1215Å. (b) The Gaussian ISF profiles for the five resolutions, with $\Delta\lambda$ (fwhm) labeled for each. (c) The apparent flux profiles for each resolution after the convolution with the relative ISF. Vertical dashed lines provide wavelength range of the intrinsic profile (Churchill, 2016).

where $\Delta\sigma = \Delta\lambda/2.35$ and the FWHM of the function correspond to the resolution element $\Delta\lambda$.

To further emphasize the important role of the ISF and resolution it will be enough to observe that the measured spectrum of a chromatic source will be the convolution of the real spectrum with the ISF of the spectrograph used for the measurement. The resulting effect of this operation for different ISF, and so for different resolutions, can be seen in Fig. 2.11. It is clear from the figure that a wrong resolution choice will be paid in terms of confusion between adjacent wavelengths which will inevitably makes unfeasible a correct measurement.

It is also worth mentioning the problem of the ISF sampling. Indeed, one of the issues in the design of modern spectrograph concern not only the shape of the final ISF but also the number of pixels in the detector required to be appropriately sampled. Too many pixels will results in oversampling which must be paid for with longer observations, on the other side fewer pixels will undersample the shape of the ISF with consequent loss of information.
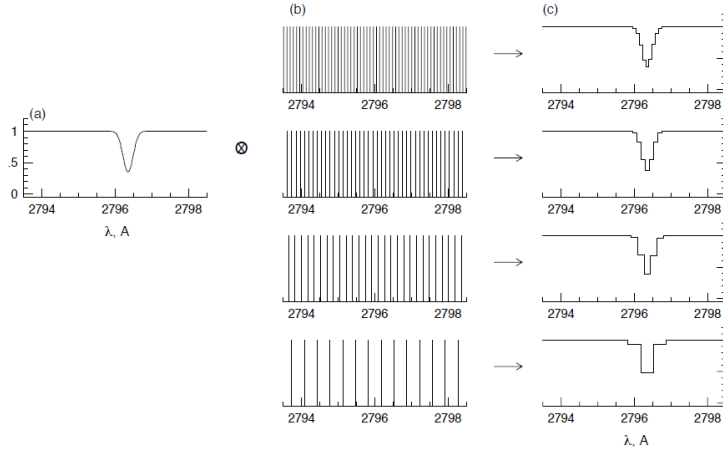
Figure 2.12: Illustration of the pixelization process for an $R = 8000$ unresolved absorption line at $\lambda = 2796.35$ (panel a) for four different sampling rates. (b) The Shah functions for three different sampling rates corresponding to $p = \Delta\lambda/\Delta\lambda_{pix} = 1, 2, 3$, and 4 pixels for the resolution element $\Delta\lambda = \lambda/R$. (c) The sampled spectra after the convolution with the three different Shah functions (Churchill, 2016).

The aforementioned process, often referred as *pixelization* or *discretization*, can be better understood by looking at Fig. 2.12: the discrete ISF in the right column of the picture, can be considered as the product of the theoretical ISF defined in Eq. 2.29, shown in the left column as an unresolved absorption line, for the *Shah function* illustrated in the central column. This particular function defined as a "train" of *Dirac* $\delta - function$, will have as much δ peaks as the number of pixels available for sampling, in Fig. 2.12(c) the resulting effect of this pixellization has been shown for 4 different Shah functions.

The right number of pixels to appropriately sample the ISF can be derived from the *sampling theorem* which states that the sampling frequency must be great or equal to the *Nyquist frequency* defined as the double of the maximum frequency one wish to include in the final spectrum. In terms of wavelength one has to start from the shortest one in the spectrum λ_{min} and, recalling that resolution R is constant, from the Eq. 2.28 obtain the minimum resolution element

$$\Delta\lambda_{min} = \frac{\lambda_{min}}{R} \tag{2.30}$$

Following the Nyquist definition the minimum distance $\Delta\lambda_{pix}$ between the Shah $\delta - functions$ (which represents the individual pixels) to sample cor-

rectly the ISF, can be defined as $\Delta\lambda_{pix} = \Delta\lambda_{min}/2 = \frac{\lambda_{min}}{2R}$, which is equivalent to sample the ISF at λ_{min} with two pixels. Below this value a portion of the original signal will be lost producing an undesired effect in the measured data called *aliasing*.

2.2 Telescopes and state of the art instruments

As already mentioned, in order to measure the metallicity of the lowest density regions of the IGM high resolution itself is not enough. Often quasar spectra are characterized by very low column density absorption lines, as low that will be difficult to distinguish the absorption features from the surrounding noise.

The characteristic which allow to clearly distinguish the signal from the noise is the *sensitivity* of the spectrum, in real measurement this characteristic can be quantified by the so called *Signal to Noise Ratio* or SNR of the spectrum, which is simply represented by the ratio of the source flux measured in a certain region of the spectrum to the correspondent statistical errors.

2.2.1 Practical estimate of the SNR

To give a practical estimate of the SNR it will be necessary to examine all the contributions to the measured signal and the corresponding errors. For this purpose it will be useful to define some basic properties of the detectors used in the measurement, for the sake of simplicity the following discussion will be limited only to the modern CCD detectors which, owing to their high efficiency in the visible and near-infrared range, are the most used.

Charged Coupled Devices are made of an array of pixels each of which should give in principle an independent measure of the incident flux. The basic concept behind is simple: for each photon hitting the surface of the pixel an electron is liberated inside the material (Fig. 2.13).

In reality things are not so simple, one has to take into account the *quantum efficiency* and the readout electronics which will affect the final value of the CCD pixels usually measured in *counts*. The former represents a characteristic of the CCD material and defines the efficiency of the electrons pairs production process as a function of the wavelength of the hitting photons. The latter can be characterized by a constant scale factor called *gain* which
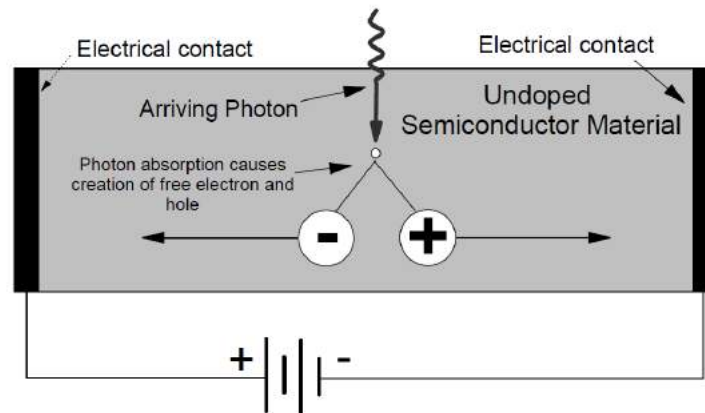
Figure 2.13: Basic of photoconductive cells which are the building elements of a CCD detector. It is a properly doped semiconductor with an electrical tension applied on the opposite sides. The circuit allows to set an uniform electrical field inside the semiconductor. The flux of photons generates free electrons inside the material which will move due to the electric field generating a photo-current proportional to the incident flux (Dutton, 1998).

relates the number of free electrons produced and the measured counts and accounts for the ADU (Analog-to-Digital Units) conversion and amplification of the measured charge.

The gain value has to be properly chosen in order to minimize the *read-out noise* or RON. The readout noise can be considered as the "zero level" of the error and is generated by the random oscillations of the measured charge due to statistical fluctuations introduced by the amplifiers and the readout process itself (DeBernardis, 2005).

So ultimately the measured charge is proportional to the incident flux. The measured charge should be considered as the sum of different contributions which should be taken into account in a detailed analysis and can be summarized as follow

$$
\begin{aligned}
S \quad &= \quad \text{The flux from the target} \\
BKG \quad &= \quad \text{The flux from the background in the slit which usually comes} \\
&\qquad \text{from the atmospheric emission} \\
DK \quad &= \quad \text{The } dark\ current \text{ which represents the number of free electrons} \\
&\qquad \text{randomly produced by the internal structure of the material} \\
&\qquad \text{even when no photons enter the device} \\
BIAS \quad &= \quad \text{The number of electrons produced by thermal excitation due to} \\
&\qquad \text{the fact that detectors operate at environment temperature, it} \\
&\qquad \text{represents the zero level of the observation}
\end{aligned}
$$

The incident flux itself does not represent the flux S from the source: during its interaction with the atmosphere, telescope and spectrograph, the original flux S gets modified such that the observed signal S_o will be described as

$$S_o = (S + BKG) \cdot \Delta\lambda_{pix} A\eta(\lambda)T_{exp} + DK \cdot T_{exp} + BIAS \ (counts) \quad (2.31)$$

where $\Delta\lambda_{pix}$ has been defined in 2.1.8 and represents the wavelength interval spanned by a pixel, A is the photon collecting area, $\eta(\lambda)$ the overall efficiency of the system (see **??** for detail) and T_{exp} the exposure time.

As discussed in the next chapter one of the aims of data reduction is the quantification and removal of the background BKG, dark current DK and zero level $BIAS$, in order to estimate only the source signal S. What remains after this operation represents the measured signal S_m, which can be expressed as

$$S_m = S \cdot \Delta\lambda_{pix} A\eta(\lambda)T_{exp} \ (counts) \quad (2.32)$$

To obtain a practical estimate of the SNR will be sufficient to give an estimation of the errors related to the observed signal of Eq. 2.31 where each term gives its contribution to the final error. Such an estimate can be derived from statistic recalling that the number of photons arriving on the detector's surface in a fixed time follows the *Poisson distribution*. The statistical error will be represented by the sigma of the distribution simply obtained by the square root of the measured counts. Finally the measured errors E_m per pixel will be

$$E_m = \sqrt{[(S + BKG) \cdot \Delta\lambda_{pix} A\eta(\lambda) + DK]T_{exp} + RON^2} \quad (2.33)$$

It is important to notice that while for the signal the contribution from background, dark current and bias can be removed, in the case of errors different contributions cannot be deleted but can only be "*propagated*" throughout the data reduction processes. This is the case for example of Eq. 2.33 where the contribution from the RON has been propagated in order to obtain the error estimate. A more detailed discussion of errors propagation will be presented in the next chapter.

From Eq. 2.32 and Eq. 2.33 the estimate of the SNR per pixel can be finally derived as

$$SNR = \frac{S_m}{E_m} = \frac{S \cdot \Delta\lambda_{pix} A\eta(\lambda)T_{exp}}{\sqrt{[(S + BKG) \cdot \Delta\lambda_{pix} A\eta(\lambda) + DK]T_{exp} + RON^2}} \quad (2.34)$$

By observing this equation some important conclusions can be derived. For example by fixing the efficiency $\eta(\lambda)$ of the system and DK and RON, which are properties of the CCD, for a source with flux S and a desired dispersion $\Delta\lambda_{pix}$ remains only two possible way to increase the SNR of the measurement: increase the photon collecting area A, which means bigger telescopes, or increase the exposure time T_{exp}.

Since unfortunately astronomers resources are not unlimited, it should be clear that in real cases the choice of a telescope and the required exposure time is the result of practical considerations coming from the science case itself. For example when the same observation can be accomplished with a smaller telescope by increasing the exposure time, to use a bigger telescope would be a waste of resources. On the other hand sometimes the required SNR is so high that also with the biggest available telescopes very long exposure time (of the order of many hours) is required.

Besides of the SNR, key factors are also the desired resolution and the apparent magnitude of the sources. Indeed higher resolution means smaller $\Delta\lambda_{pix}$ and so the need of a large collecting area becomes compelling even more if the proposed target appear very faint due to its cosmological distance. This is why bigger telescopes are usually employed in studies involving high resolution measurements of quasar spectra.

2.2.2 Largest ground based telescopes

One of the advantages of the ground based telescopes is the possibility to make primary mirrors with very large diameter as for example will be for the next generation telescope *E-ELT* or *European Extremely Large Telescope* which will have a diameter of 39 meters and which will view the first light in the 2024 (McPherson et al., 2012).

To build large telescopes today is still a very difficult and expensive task, sometimes the resources available for a single country or institution are not enough to accomplish it. To overcome this issue the scientific community needs to join the resources together and for this reason modern big telescopes are the result of the contribution given by several nations which in turn will gain a certain amount of granted observation time. It is clear that such precious resources cannot be wasted and so there are international commissions which evaluate the scientific proposals and select the most interesting ones. Telescopes with large collecting areas allow to collect more photons from

Figure 2.14: (a) VLT telescopes at Cerro Paranal - Chile (source: ESO website). (b) Keck telescopes on the top of Mauna Kea - Hawaii (source: Keck website)

the source and consequently reduce the integration time. However can be demonstrated from Eq. 2.34 that an observation of an object of apparent magnitude down to 18.0 (typical of distant quasars) reaching a $SNR > 100$ will require several hours of integration time even with a 10 meters telescope. For this reason it is very difficult to perform observations with very high SNR (orders of hundreds) and scientific proposals need very strong motivations to be not rejected by the scientific commissions.

This work will be focused on two main sites for observations from the ground: the *Very Large Telescope* or *VLT* (Fig. 2.14(a)) located at an altitude of 2635 meters on the top of the Cerro Paranal mountain (Chile) and the *Keck Observatory* (Fig. 2.14(b)) at an altitude of 4145 meters on the top of the Mauna Kea (Hawaii).

The latter presents two unit telescopes with a 10 meters primary mirror while the former presents four unit telescopes with primary mirrors of 8.20 meters of diameter. The two sites are similar: both can operate with single units or in interferometric mode joining together the light from all units, both are characterized by an alt-azimuth mount and by a *Ritchey-Chrétien* optical configuration which are illustrated in Fig 2.15.

Each one of them lodges various instruments at the different foci, in the following our attention will be concentrated on the two high resolution echelle spectrograph: *HIRES* for the Keck and *UVES* for the VLT.
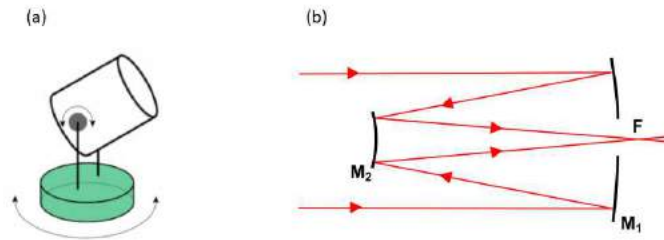
Figure 2.15: (a) This sketch represents the basic movements of the altitude-azimuth mount, the base is tight to the ground. (b) Representation of a Ritchey-Chrétien configuration. Rays of light (in red) enter the telescope in parallel and come reflected by the primary mirror M_1. The secondary mirror M_2, which also contain corrector plates to remove aberrations, reflects again the rays which focuses behind the primary mirror.

2.2.3   HIRES - The high-resolution spectrograph at KECK

At the right nasmyth focus of the Keck-1 telescope is mounted one of the most powerful high resolution echelle spectrographs in the world: the *HIRES (High Resolution Echelle Spectrograph)* instrument. In Fig. 2.16 is illustrated a schematic view of the optical configuration and main components of the instrument.

Light from the telescope enters from the left in this schematic, coming to focus at the $f/15$ nasmyth focus. An entrance hatch at the front of HIRES seals off the slit area such that the instrument can be run during the day under high light-level conditions in the dome. The beam then proceeds to focus at the HIRES slit plane. The slit plane is tilted such that light can be reflected up at an angle and re-imaged onto a CCD TV acquisition and guiding camera.

At the Nasmyth focal plane, a bi-parting precision slit is provided for adjusting spectral resolution. A series of decker plates just above and in very close proximity to the slit are provided for defining the entrance slit length. Some of these decker apertures are for defining slit length only and are used in conjunction with the slit, while all other decker apertures define both slit length and width and are used in place of the slit jaws. Available slits and decker allow for spectral resolution up to 86600 which is a very high resolution if compared with typical resolution of standard grating spectrographs of the order of few thousands.

An Iodine absorption cell can be moved into position directly in front of the slit for very precise wavelength calibration. Quartz/halogen incandescent
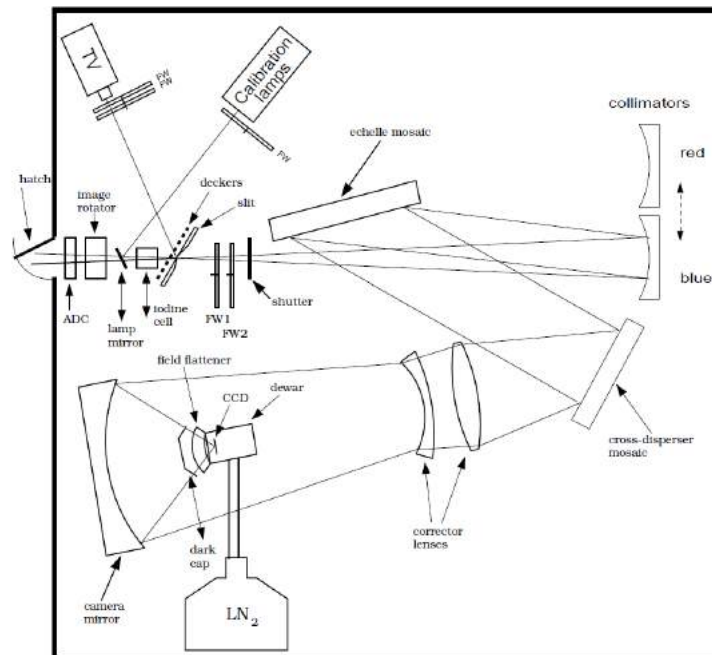
Figure 2.16: Schematic of the HIRES spectrograph and its main components (Vogt, 1994).

lamps and hollow-cathode lamps, located up near the ceiling of the slit area, provide for flat fielding and wavelength calibration. Light from these calibration lamps gets fed into the HIRES optical axis by reflection off a feed mirror which slides into place when calibration is desired. The calibration lamp system has a filter wheel for chromatic and intensity control.

Immediately behind the slit are two filter wheels, mainly for filters required for blocking unwanted cross-disperser orders. Behind the filter wheels is a shutter for controlling the start and stop of an exposure.

The beam then expands and gets collimated by either of two red/blue optimized collimator mirrors. The collimated beam is then sent to an echelle grating (1×3 mosaic), and then to a cross-disperser grating (2×1 mosaic). HIRES has two cross dispersers blazed at different wavelengths, the instrument can operates in two different modes depending on the selected cross disperser and collimator: HIRESb and HIRESr, optimized for short and long wavelengths respectively. The efficiency of the two is equal at approximately 4200Å. Switching between the two is not possible during the night.

After passing the cross-disperser the beam is redirected into a large ($30 arcsec$
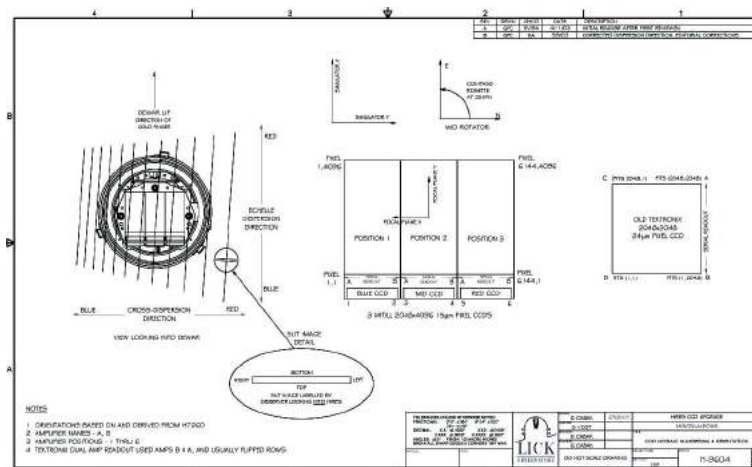
Figure 2.17: Layout of the new HIRES upgraded CCD mosaic (Vogt, 2006).

diameter entrance aperture) prime focus camera where the detector is placed. On 18 August 2004, HIRES underwent a major upgrade of its detector. The CCD upgrade resulted in significant improvements in the HIRES detector, dewar, and CCD electronics.

The new detector consists of a mosaic of three $2048x4096$ $15\mu m$ pixel *MIT - Lincoln Labs* CCDs, compared to a single $2048x2048$ $24\mu m$ pixel *Tektronix* CCD of the old detector. The new layout and a comparison with the old one are represented in Fig. 2.17.

The basic format for all data after 2004/08 is represented by a standard FITS file containing three extensions, one for each of the new detectors. Looking at Fig. 2.17 from the left to the right of the camera focal plane, following the cross-dispersion direction, there will be the "blue", "green" and "red" CCDs, each one of them measuring a different portion of the spectrum. The setting of the instrument is decided by the observer depending on the desired resolution and spectral coverage. Central choices before the observation are the choice between UV or RED cross-disperser, the angles of the echelle and cross-disperser, the decker and the CCD binning. All these parameters define univocally the setup of the particular observation and are registered in the header of the FITS file.

The main characteristics of the HIRES spectrograph are resumed in Table 2.1, readers can refer to Vogt (1994) for further details.

Scale	$0.725 mm/arcsec$
Echelle	$12'' \times 48''$ mosaic, $52.68 gr/mm$, $70.5°$ blaze
Cross Dispersers	red: $16'' \times 24''$ mosaic, $250 gr/mm$, 5600Å
	blue: $16'' \times 24''$ mosaic, $400 gr/mm$, 3800Å
CCD readout noise	2.8, 3.1, 3.1 electrons (B, G, R)
CCD dark current	3.8, 4.4, 2.2 e-/hour (B, G, R) in low gain
	5.5, 7.7, 4.2 e-/hour (B, G, R) in high gain
Spectral range	$300 - 1100 nm$
Spectral resolution	To 86600 ($0.057\text{Å} = 2.5 pixel$ at 5000Å)
Spatial resolution	$0.119 arcsec/pixel$
Slit projection	$0.575 arcsec = 3.2 pixel$
Slit dimensions	B decker: $0.574 arcsec$
	C decker: $0.861 arcsec$
	D decker: $1.148 arcsec$
	E decker: $0.400 arcsec$ (added after CCD upgrade)

Table 2.1: Main characteristic of the HIRES instrument (Vogt, 2002).

2.2.4   UVES - The ESO spectrograph at VLT

UVES, the *Ultraviolet and Visual Echelle Spectrograph* located at Nasmyth platform B of the second Unit Telescope (Kueyen) of the VLT, is a cross-dispersed echelle spectrograph designed to operate with high efficiency from the atmospheric cut-off at $300 nm$ to the long-wavelength limit of the CCD detectors ($\sim 1100 nm$).

The main difference with respect to HIRES is the presence of two different light paths, the light beam coming from the telescope is split into two arms, the UV-Blue and the Visual-Red, within the instrument. The two arms can be operated separately or in parallel with a dichroic beam splitter. The resolving power is ~ 40000 when a $1 arcsec$ slit is used. The two-pixel resolution (to be obtained with a narrower slit or with the use of an image slicer) is 80000 or 110000 in the Blue and the Red arm, respectively.

A schematic view of the optical design and main component of the instrument are represented in Fig. 2.18. The light from the telescope enters from the top and passes through the calibration-mirror unit, which allows the light from calibration lamps entering the spectrograph, the iodine cell, and the image-slicer unit, respectively. The last one allows to choose between
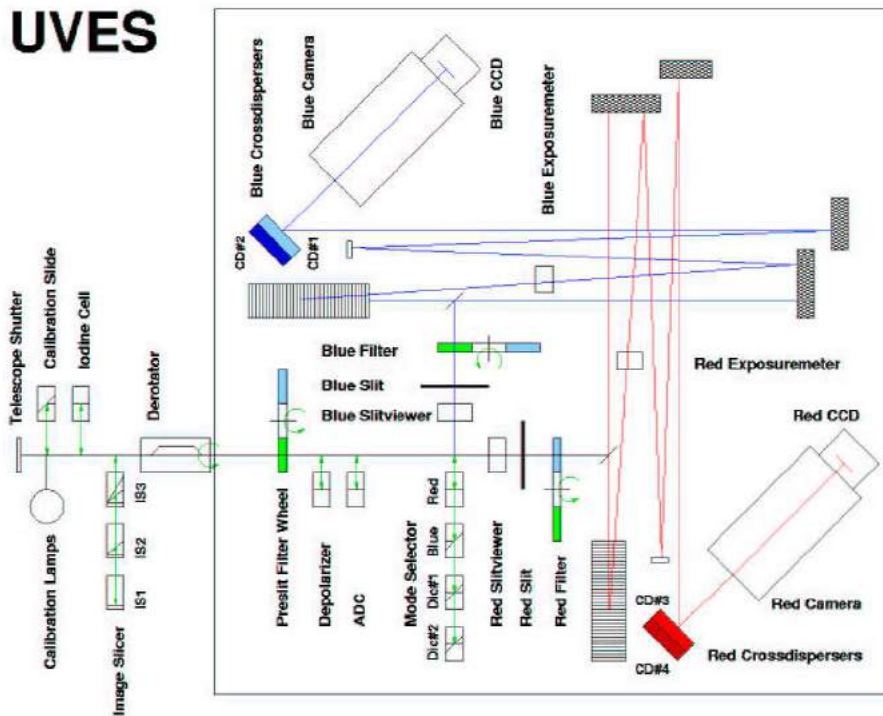
Figure 2.18: Schematic overview of the UVES spectrograph (Monaco, 2014).

three different image slicers which reformat the 2D-image of a rectangular area in the $f/15$ focal plane of the telescope into a narrow slit which is imaged on the spectrograph entrance slits, in this way it is possible to reduce slit losses without limiting the resolution (see 2.1.1).

After passing through the field derotator, the light beam from the telescope is focussed at the position of the mode selector (which includes the dichroic filters) and the beam is split into a red (going straight) and a blue path (reflected to the right). On the fixed table in the pre-slit area, additional optical components are available for insertion in the optical beam: filters, a depolarizer, an Atmospheric Dispersion Compensator (ADC) and two pupil stops of different size.

Two dichroics are available to work in parallel with the two arms. The blue arm ($\lambda\lambda 300 - 500nm$) and the red arm ($\lambda\lambda 420 - 1100nm$) have an identical layout. They are folded and cross each other to minimize the size of the table on the platform. The two-arm solution gives high efficiency because it permits to optimize the spectral response of coatings, gratings and detec-

tors. In both arms, the respective beams that enter the spectrograph are reflected by a folding mirror to the main collimators. From here, the collimated beams are dispersed by the echelle gratings and sent back to the main collimators. The echelle $R4$ gratings are the largest ever made of this type. They are operated in quasi-Littrow mode, that is with the angle of incidence and diffraction equal but in a different plane, to maximize efficiency.

Via the intermediate spectrum mirrors and the transfer collimators the dispersed beams fall on the cross-disperser units which offer two different cross disperser for each path. From here the echelle spectra enter the cameras and are recorded on the CCD detectors.

The detector in the blue camera consists of one 2048×4096 pixels EEV CCD (EEV 44-82), windowed to 2048×3000. The detector in the red camera consists of a mosaic of one 4096×2048 pixels EEV (EEV 44-82) also known as "Sting" and one 4096×2048 pixels MIT-LL (CCID-20) $4k \times 2k$ also known as 'Zeus'. Both detectors are oriented such that the direction of the spectral dispersion is along the larger dimension of the CCDs.

Zeus is the replacement of the old MIT CCD that was also known as "Nigel" and that took place in early July 2009. Zeus is a thicker chip than Nigel was, hence the Quantum Efficiency is improved in the far red. A gain of a factor ~ 2 at $900 - nm$ has been measured both in the lab and on-sky. The fringing on Zeus is also reduced.

The EEV and MIT CCDs are designed to optimize the detector response as a function of wavelength and to reduce fringing at far-red wavelengths. The gap between the two red CCDs is $\sim 0.96mm$ (loss of one order in the gap). This and the non-perfect alignment of the two chips require that the spectra on the two chips of the mosaic are extracted separately.

Differences in the detectors from the blue and red paths are reflected in different data formats. For the blue side the standard output is a single FITS file with no extensions. For the red side instead two extensions are present each for one of the two chips of the mosaic: the redder MIT CCID-20 in the first extension and the bluer EEV-44 in the second one. Prior to August 2002 the standard data format for the red side was similar to the blue side with the disadvantage that files had to be properly split before applying a standard echelle reduction package.

The CCD parameters are periodically remeasured as part of the UVES calibration plan. The updated values are entered in the instrument database

| **UVES standard settings** | | | | | | | |
Mode	Cross Disp.	Below slit filter	Min. Wav.	Central Wav. (nm)	Max. Wav.	Decker Height	Gap (nm)
Using dichroic							
DIC1	CD#1	HER_5	303	346	388	10."	=
	CD#3	SHP700	476	580	684	12."	5
DIC1	CD#2	HER_5	326	390	445	8."	=
	CD#3	SHP700	458	564	668	11."	2
DIC1	CD#1	HER_5	303	346	388	10."	=
	CD#3	SHP700	458	564	668	11."	2
DIC1	CD#2	HER_5	326	390	445	8."	=
	CD#3	SHP700	476	580	684	12."	5
DIC2	CD#1	HER_5	303	346	388	10."	=
	CD#4	BK7_5	565	760	946	8."	7
DIC2	CD#2	HER_5	326	390	445	8."	=
	CD#4	BK7_5	565	760	946	8."	7
DIC2	CD#2	HER_5	373	437	499	10."	=
	CD#4	BK7_5	565	760	946	8."	7
DIC2	CD#2	HER_5	373	437	499	10."	=
	CD#4	OG590	660	860	1060	12."	10
DIC2	CD#1	HER_5	303	346	388	10."	=
	CD#4	OG590	660	860	1060	12."	10
DIC2	CD#2	HER_5	326	390	445	8."	=
	CD#4	OG590	660	860	1060	12."	10
Using blue arm only							
BLUE	CD#1	HER_5	303	346	388	10."	=
BLUE	CD#2	HER_5	373	437	499	10."	=
Using red arm only							
RED	CD#3	SHP700	414	520	621	8.9"	1
RED	CD#3	SHP700	476	580	684	12."	5
RED	CD#3	SHP700	500	600	705	12."	5
RED	CD#4	OG590	660	860	1060	12."	10

Table 2.2: The UVES standard settings, 2 in the blue, 4 in the red, and 7 in dichroic mode (Monaco, 2014).

	Blue arm	Red arm
Wavelength range	$300 - 500 nm$	$420 - 1100 nm$
Resolving power-slit product	41400	38700
nm/pixel	$0.0019 nm$ at $450 nm$	$0.0025 nm$ at $600 nm$
Max. Resolving power (2-pixel sampling)	~ 80000	~ 110000
Echelle (R4 mosaic)	$41.59 g/mm$	$31.6 g/mm$
CCDs	2048×4096 (windowed to 2048×3000)	two 2048×4096 (mosaic of different types)
Pixel ($15\mu m$) scale		
disp. dir. (varying along order)	$0.215'' \pm 20\%$	$0.155'' \pm 20\%$
along slit (dep. on cross-disp.)	$0.25''$ (CD1 and CD2)	$0.18''$ (CD3), $0.17''$ (CD4)
CCD Read-out noise fast read-out, low gain (slow read-out, high gain)	4.1 (2.1) e^- rms	EEV 4.2 (2.8) e^- rms MIT 3.7 (2.1) e^- rms
CCD Dark current	$0.4 e^-/pix/h$	EEV $0.5 e^-/pix/h$ MIT $1.5 e^-/pix/h$

Table 2.3: Main characteristics of the two arms of the UVES instrument (Monaco, 2014).

and are recorded in the FITS headers, for use in the data reduction. The Blue and Red CCDs are operated at a temperature of $\sim 153 K$ and $135 K$ respectively by means of liquid-nitrogen cryostats.

At a variance with the HIRES spectrograph which allow the user to freely chose the echelle and cross disperser angles to cover the desired wavelength region, UVES instrument offers a series of "standard settings" (resumed in Table 2.2) identified by the central wavelength of the cross disperser blaze function, moreover the user can chose between single or double arm mode. Also in this case the standard settings in addition to all others parameters like the CCD binning, the slit width and the adopted filters, define univocally the setup of the particular observation and are recorded in the header of the FITS file.

The main capabilities of the two UVES arms are summarized in Table 2.3, readers can refer to Monaco (2014) for further details.

2.3   The online archive databases

Nowadays it is very difficult to obtain observational time on instrument like HIRES or UVES due to their huge over subscription. On the other hand a wealth of data had been gathered along the years, HIRES being operative from July 1993 and UVES from October 1999.
As already pointed out since every observation represents a huge investment in terms of money and time, it became clear that data would be better exploited by making them available to the entire scientific community after one year of proprietary period of the approved proposals PIs. This became possible with the advent of databases. Nowadays all these data are generously offered to the public thanks to online archives

- KOA (Keck Observatory Archive) - for HIRES

- ESO (European Southern Observatory) Archive - for UVES

Those projects aim to put together all data acquired on the entire instruments operational life and make them available by a search trough an user friendly frontend. Data are stored in the $24h$ following the observation together with all the calibration frames. The database capabilities allow to search the huge amount of data by mean of simple selection criteria and to generate automatically a list of appropriate calibration files for the selected observation. In the following some of the main parameters useful in archive query are reported (note that they are pretty the same between UVES and HIRES archives)

- coordinates (right ascension - RA and declination - DEC)

- observation date

- name of the Principal Investigator

- program title

- wavelength coverage

- Signal-to-Noise-Ratio

- slit width and height

- CCD binning

just to enumerate only the main criteria, nevertheless more detailed criteria are available on the archives query forms. Once the search yields the results, data can be selected and downloaded on a local computer. Even if extracted spectra can be downloaded directly, usually people prefer to download the raw data, i.e. data which need yet to be calibrated and extracted with the process called *data reduction*. In this way the user can perform the data reduction by himself step by step to avoid issues and bad extracted spectra usually encountered in automatically extracted spectra. All the processes involved in the data reduction of echelle spectra will be discussed in details in the next chapter.

Considering the operational life of the two instruments, after more than a decade the online archives represent a precious source of data which in most cases are not yet fully exploited. From this consideration starts one of the main ideas of this work: to push even more the exploitation of the archives by joining together data observed with both UVES and HIRES.

2.3.1   Maximize sensitivity and wavelength coverage

HIRES and UVES are very similar instruments and could be considered complementary to each other since they are looking at the opposite hemispheres. Nevertheless there is a small declination range in common between them which allows to search for objects that have been observed with both instruments.

Focusing on HIRES and UVES archival data only, to maximize sensitivity and wavelength coverage means to find all available observations of one target independently from the instrument and the particular instrumental setups. Online archives usually contain multiple exposures of the same target performed with different setups which often cover different wavelength regions of the available range.

Merging all the setups and all available exposures from both instruments allows not only to improve the final SNR of the spectrum but also to obtain complete spectra which in most cases extend over the entire available wavelength range of the spectrographs.

In practice one has to search both archives for multiple recurrences of the same set of coordinates (declination δ and right ascension RA) and then crossing the results to find occurrences in both UVES and HIRES.

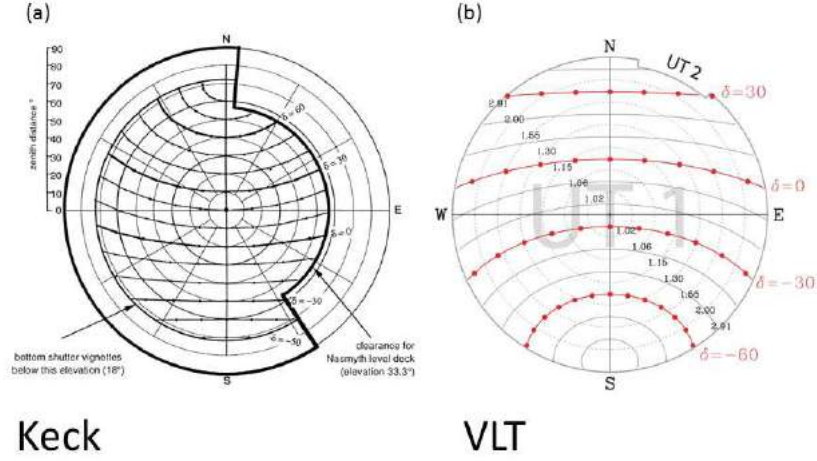The first consideration to be done concerns the pointing limits of the two

Figure 2.19: The plots above show the area of the sky visible by the Keck telescopes (a) and by the VLT telescopes (b). The outermost circle marks the telescope safety limit. Concentric circles denote zenith distance intervals of 10 degrees. An object will move along a curves of constant declination, with solid dots marking movement in interval of 1 hour. The maximum visibility period of an object can be read directly by counting intervals between dots (source: VLT web pages and Vogt (2002)).

telescopes which are located in the opposite hemispheres. Keck Telescope is located at latitude 20 degrees North and rejecting object with Zenithal distance greater than ~ 45 degrees (concentric circle in Fig. 2.19) its view is limited to the declination range

$$-30 \leq \delta_{Keck} \leq +60$$

as illustrated in Fig. 2.19(a). For VLT, which is located at latitude of 25 degrees South, rejecting again object with Zenithal distance greater than ~ 45 degrees view is limited to the declination range

$$-60 \leq \delta_{VLT} \leq +30$$

as illustrated in Fig. 2.19(b). From intersecting those two ranges, the objects that can be seen by those telescopes must have declination between -30 to $+30$.

Looking at Fig. 2.19 can be understood that at the extreme end of this range either telescopes will have difficulties to observe objects for a sufficiently long time. For example higher declination means more time for Keck but less time
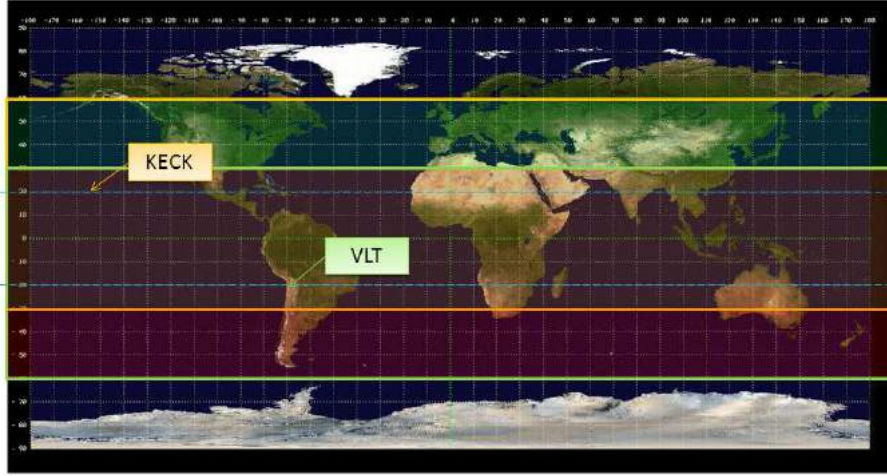
Figure 2.20: Intersection of the accessible declinations range from the Keck Observatory (orange box) and the accessible declinations from the Very Large Telescope (green box). The area delimited by the horizontal dashed blue lines represents the declinations range in which objects can be observed from both telescopes for at least six hours.

for VLT. This reduces even more the useful declination range, if for example we want to limit the search to objects which can be observed for at least 6 hours from both telescopes, one has to pick the part of sky with declination in between

$$-20 \leq \delta_\star \leq +20$$

of the visibility range as illustrated in Fig. 2.20.

The second consideration concerns the basic policy of observatories. If for example ten hours of observation are requested in one proposal, the requested time is divided in many smaller observations in order to better organize available time and avoid to vaste resources if for example something goes wrong during an observing run. Moreover systematic errors or shifts can be better controlled on short exposures. For this reasons it is difficult to find in the archives exposures longer than 2 hours and if one searches for targets with high number of available hours the main criteria should be the available number of exposures rather than the exposure time.

On the basis of those considerations we searches both archives using the following parameters

- EXPOSURE TIME: $\geq$ 2000sec

- TYPE: science only

- DATE (HIRES): 2004 (CCD upgrade) - present

- DATE (UVES): entire operational life

For practical reasons our search on the HIRES archive has been limited in time considering only data after the CCD upgrade (August 2004). The search has been restricted to scientific frames only and in order to consider only the highest SNR images, an exposure time $\geq 2000s$ has been imposed. Using these criteria a total of 2839 objects have been found for HIRES and 18758 for UVES.

At this point an algorithm has been created to search for multiple recurrences of the same coordinates on individual lists first and crossing the two lists later. In this way all the available exposures of one target can be identified and gathered together. The algorithm works as follow

1. Reads the results of the search (simple text files) and stores the object's parameters into arrays

2. Sorts the results in order of ascending declinations

3. Converts coordinates from hours to degrees

4. Finds all exposures in both instruments whose declination does not differ more than 0.01 degrees and associates them to a target

5. Finds all exposures in both instruments whose right ascension does not differ more than 0.1 degrees and associates them to a target

6. Counts the number of exposures available for each target identified

7. Saves the final list of tagets in order of decreasing number of available exposures

The tolerances have been chosen in order to include little pointing variations and to exclude false matching of near objects.

At this point the log file has been visually checked in order to select quasars only (archives also include observations of stars and galaxies). The list of quasars obtained in this way has been further reduced selecting only QSO candidates with at least 20 hours of observation from the sum of HIRES and UVES. The result of this operation is a list of 35 quasars candidates spanning a redshift range $1.7 \leq z \leq 6.3$.

2.3.2   The very high SNR QSO sample

The ideal sample should have at least SNR > 100 for each QSO and should cover homogeneously the four redshift bins centered on $2 - 2.5 - 3 - 3.5$ in order to allow to study the redshift evolution of physical quantities.

In order to obtain such a sample QSO candidates have been selected by redshift, apparent magnitude and total available hours of observation. Using these selection criteria the final QSO sample obtained consists of 12 QSO from $z = 1.71$ to $z = 3.78$ where each one of the redshift bins just defined contains three QSO.

The sensitivity achieved by the final spectra resulting from the merging of all available exposures is really very high: SNR > 100 for QSO at $z > 3$ and SNR > 150 for QSO at $z < 3$, six of them having SNR > 200 and one of them SNR > 250. All SNR are expressed as "per pixel" values[3] and calculated from the average over a 100Å interval in the CIV region of each QSO. The sample characteristics are summarized in Table 2.4 while the list of figures can be found at the end of the thesis.

The characteristics of this sample are unique in terms of sensitivity and spectral extension. Just to have an idea here we report in Fig. 2.21 the distribution of the average SNR per resolution element in a ± 5Å interval at different rest frame wavelengths for all the 247 QSO spectra of the *KODIAQ* survey (O'Meara et al., 2015). This picture shows clearly that the most common spectra used today for science have SNR between ~ 2 to ~ 50 where spectra with SNR ~ 50 are yet considered "high SNR" and only in few cases a SNR of the order of 100 can be reached.

On the other hand Table 2.4 shows that to reach SNR ≥ 100 at least 20 hours of observation are required with 10 meters class telescopes. To obtain such a large amount of time for the observation of one target only it is very difficult, in practice we overcame this issue at "zero cost" only relying on archival data.

This very high SNR QSO sample allows to push our search of metals down to the lowest column densities ever probed with the added value, with respect to other similar studies, of an improved statistical significance provided by the

---

[3]The SNR can be expressed "per pixel" if it is evaluated from a single pixel only, or "per resolution element" if it is evaluated from the weighted mean of all the pixels included in the interval.
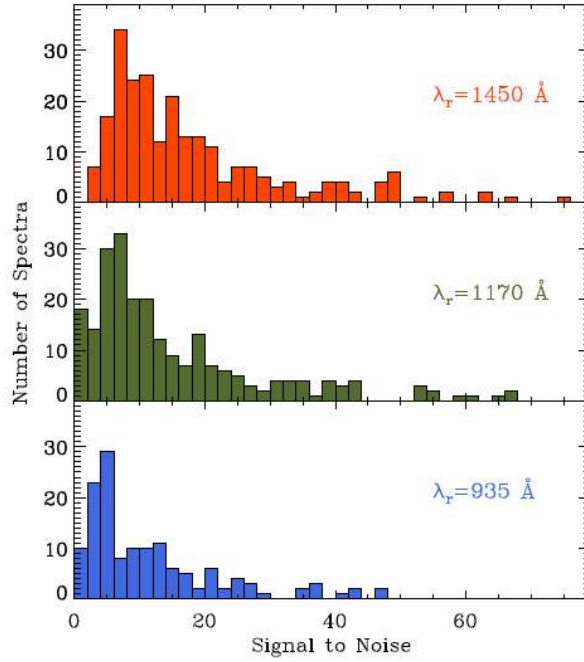
Figure 2.21: Median signal to noise per $2.6km/s$ interval in a window $\pm 5\text{Å}$ about $\lambda = 935, 1170, 1450\text{Å}$ for the 247 spectra in the KODIAQ sample (O'Meara et al., 2015).

multiple available lines of sight. This topics will be discussed in Chapter 4 dedicated to this sample analysis.

All the QSO of the sample presented here have been reduced, flux calibrated and merged with a specifically developed generalized pipeline. This pipeline is characterized by a robust and reliable approach to data reduction which will be described in the next chapter. Procedures for the flux calibration and for the final merging will be discussed in Chapter **??**.

All these procedures have been developed with particular attention to the treatment of errors trying to be the most rigorous possible and for this reason they represent an added value with respect to the final product.

Table 2.4: Table of the main characteristic of the high SNR QSO sample. Available hours of observations in the CIV regions for both HIRES and UVES instruments are listed. The SNR is calculated as the average pixel value over a 100Å interval in the CIV region.

HIGH SNR QSO SAMPLE							
NAMES	REDSHIFT	m_V	m_R	HIRES (h)	UVES (h)	TOT (h)	SNR
QSO B1937-1009	3.787	19.00	17.19	14.20	7.50	21.70	134.93
QSO B1618-0035	3.710	17.97	17.00		24.05	24.05	120.89
QSO B0055-26	3.662	17.47	17.99		20.25	20.25	101.51
QSO B2126-15	3.268	17.30	17.30	2.00	20.05	22.05	143.75
HE0940-1050	3.076	16.60	16.60		65.88	65.88	231.31
QSO J0407-4410	3.021	17.60	17.10		33.48	33.48	142.89
QSO J0103+1316	2.721	16.57		23.53	4.27	27.80	181.76
QSO B1347-2457	2.534	16.30		1.66	48.29	49.95	210.27
QSO J2220-2803	2.414	16.00	16.10	0.66	44.68	45.34	221.32
QSO B1122-168	2.397	16.50	16.10	4.33	43.83	48.16	243.00
QSO B0237-2322	2.225	16.63			44.65	44.65	242.78
QSO J1344-1035	2.134	17.10	17.00	3.63	41.33	44.96	154.28
QSO B0515-4414	1.713	14.90	14.00		42.76	42.76	253.68

Chapter 3

Echelle data reduction

The process which allows to obtain a full mono-dimensional spectrum start-ing from a raw fits file is called *data reduction*. In this process can be included all the operations required to remove instrumental features and contamina-tions from atmospheric emission, to calibrate in wavelength and extract the spectrum from the CCD pixels counts.

Due to its complexity and the large number of operations to be performed on thousands of pixels, the data reduction is often performed automatically through dedicated computer programs called *pipelines*. Usually each instru-ment has its dedicated pipeline which accounts for the specific features of that instrument, nevertheless for similar instruments the basic procedures are the same.

In this chapter the procedures involved in the data reduction process will be described in detail, moreover a new generalized pipeline will be presented able to reduce in the same way both HIRES and UVES data. Following the discussion in Bernstein et al. (2015), this new HIRES/UVES pipeline has been obtained starting from the "XIDL" package developed by Jason X. Prochaska (Prochaska, 2006) for the HIRES spectrographs which represents an extension of the MIKE pipeline (Bernstein et al., 2003). The detailed description that will be presented in this chapter could be also considered as a user manual for the new pipeline.

It is important to emphasize that the consistency between datasets of dif-ferent instruments, required for the merging of HIRES and UVES data, is only one of the motivation of this work. As discussed in this chapter the particular approach to data reduction implemented by the XIDL package is preferable to other ones since it preserves the identity of pixels along
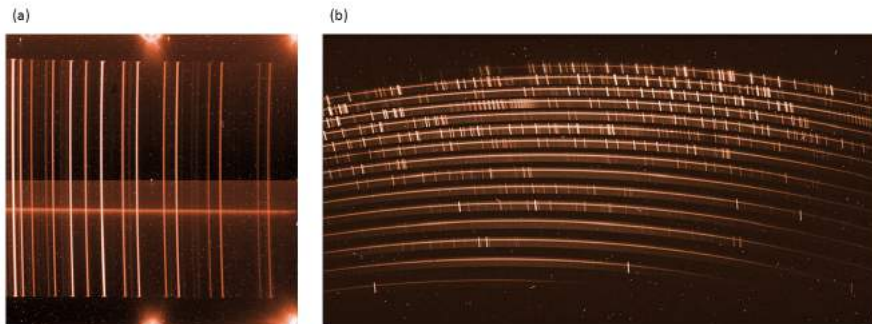
Figure 3.1: A comparison of a long-slit spectrograph raw image from the decommissioned instrument ISAAC (a) and an echelle spectrograph raw image from the medium resolution instrument X-SHOOTER (b). In each frame is visible the faint spectrum of the source (horizontal) and the emission lines of the sky (vertical). The most evident difference, besides the number of orders, is that echelle orders are curved and emission lines tilted, this represents a further complication in the reduction process (credits: ESO website).

the reduction steps and allows for a more realistic and robust estimation of the final errors. One of the main reasons for the development of the new HIRES/UVES pipeline is the extension of the same approach to UVES data.

3.1 Data reduction overview

Using an echelle spectrograph allows us to reach a resolution higher than $R \sim 50000$, but the price to pay is represented by the complexity of the data reduction process itself. Indeed how seen in 2.1.5 the output of an echelle spectrograph is not a single extended order as for a standard long-slit configuration but is represented by a series of separated orders each one covering only a small wavelength range. Fig. 3.1 shows for comparison the images produced by the detectors of a long-slit and an echelle spectrographs, an evident difference is the strong curvature of the orders which represents a further complication in the reduction process as explained in 3.4.

Despite those differences the main steps of data reduction are basically the same even if they are executed in a different way. Before entering in details it will be useful to briefly summarize the main steps of data reduction in a more general way. The discussion will follow the schematic representation of Fig. 3.2, the detailed description of each step will be discussed in section 3.4.
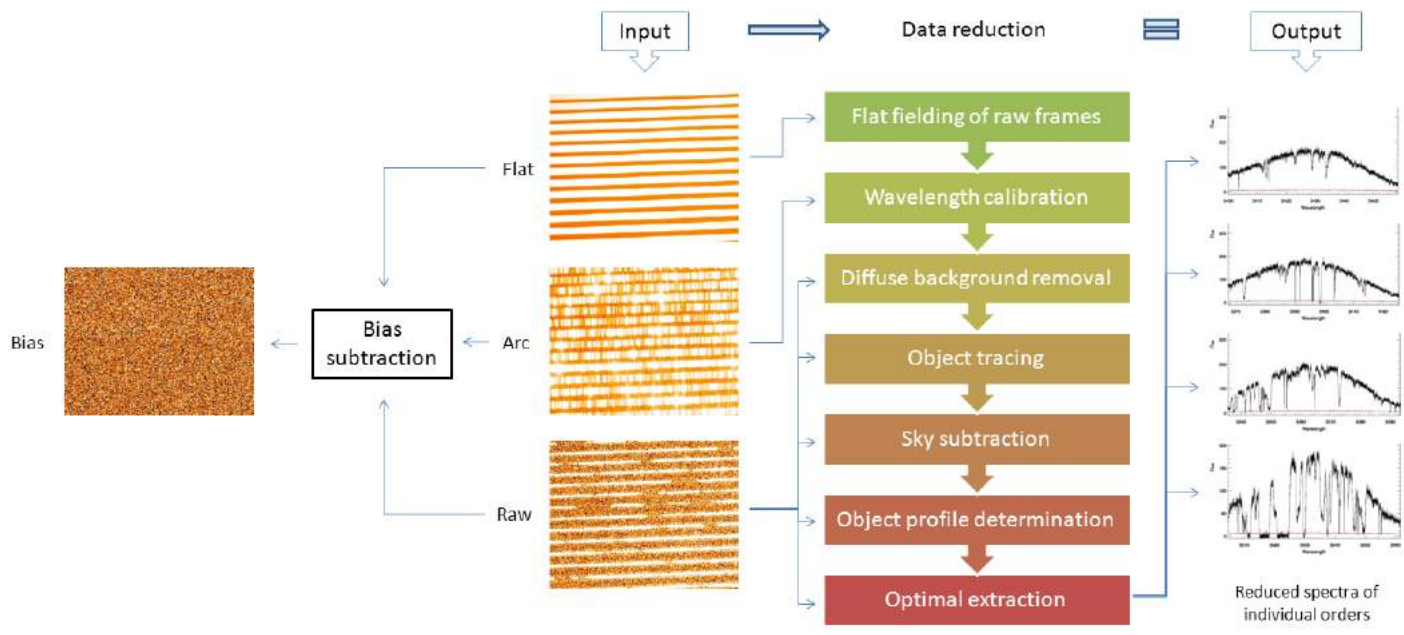
Figure 3.2: The figure shows a schematic representation of the standard approach to data reduction. The input is represented by the calibration frames (arcs and flat) and the raw frame previously bias subtracted. The output of the concatenated processes listed in the colored boxes is represented by the extracted spectra of the single orders as showed in the rightmost part of the figure. The figure of different frames are just portion of UVES images taken only for illustration purposes.

3.1.1   Required input data

During the observation every single CCD pixel counts the number of photons coming from a certain direction depending on the particular instrument setup (slit width, angles of cross disperser and echelle grating, etc...). Among all this photons only a fraction comes from the target while the others are from background sources like the emission of the sky or the internal scattered light, moreover additional counts come from the CCD dark current.

The process of data reduction basically removes the contribution of CCD and background sources and leaves only those counts the quasar's photons are responsible for. The wavelength calibration allows to assign a precise wavelength to every single CCD pixel and only after those procedures will be possible to put together all the contributions from the pixels array in order to make a full quasar spectrum.

In order to properly calibrate and reduce a raw data frame (the image of the spectrum as recorded by the CCD) at least four other frames, called "calibration frames", are required

1. **BIAS** - can be obtained with a zero seconds exposure acquired with all the shutters closed. It represents the "zero level" of the CCD pixels i.e. the pixels counts produced by thermal excitation even if no photons enter the device (see 2.2.1)

2. **DARK** - this frame is required to measure the rate of free electrons spontaneously produced inside the crystal structure of the material which compound the CCD (2.2.1). It is obtained very similarly to the bias frame, with the only difference that the exposure time is usually fixed to $1h$ to obtain the rate in counts/h.

3. **ARCS** - are usually obtained placing a Thorium-Argon lamp in the light path of the spectrograph. The spectrum of the ThAr lamp is recorded with a few seconds exposure and will serve as a reference for the wavelength calibration

4. **FLAT FIELD** - to obtain the flat field image it is important to illuminate uniformly all the CCD surface. There are different ways to make a flat field image, the most common ones are essentially two: to make a few seconds exposure of the telescope dome illuminated with an incandescent lamp or use the twilight sky which represents an even

more stable and homogeneous source. As will be discussed in 3.4 a flat field image can be used for a variety of purposes, the most important one is the characterization of the pixel-to-pixel variation of the intrinsic detector response and of all the imperfections of the CCD surface as blemishes, pock marks or, in the worst cases, dust grains

For a successful reduction it is very important that all these frames are acquired with the same instrument configuration and as close as possible in time to the raw scientific frame. This is to avoid shifting or deformation of the spectral orders due to thermal distortions of the instrument which inevitably occur during the night and day cycle.

3.1.2   Reduction cascade

The first steps required in a data reduction process are the *bias subtraction* and the *dark subtraction*. The first one is simply accomplished subtracting the bias image from the other ones (raw, arcs, flat and dark also) in order to refer the measured counts to the zero level of the CCD. Once the bias has been subtracted one can proceed to measure the dark current from the mean of the values measured in all the pixels of the dark frame. This value has to be subtracted from the other frames after being multiplied for the respective exposure times.

As illustrated in Fig. 3.2 the bias and dark subtracted flat, arc and raw frames represent the standard inputs for the data reduction process which can be defined by the following basic steps

flat fielding The input for this procedure is represented by the flat frame. Since the flat field frame has been obtained by uniform illumination of the slit, orders will be characterized by the low frequency response of the spectrograph (see blaze function in 2.1.2). The only higher frequencies features will be represented by the pixel-to-pixel response variations and by the CCD defects. To correct the raw frame for these defects the basic approach is to normalize to unity the flat field frame by fitting the low frequency blaze functions of the orders with low order polynomials (usually a second degree). This will produce a frame usually called "pixel flat" whose average value is one but which preserves the high frequency pixel-to-pixel variations. Once the pixel

flat has been obtained the flat fielding can be performed by simply dividing the raw frame by the pixel flat.

wavelength calibration The arcs frame is used as input for the wavelength calibration. The basic procedure is to extract the spectrum of the ThAr lamp from the 2D image and to cross-correlate the extracted spectrum with an archived solution. The archived solution is represented by the spectrum of the ThAr lamp already wavelength calibrated. The advantage of this technique resides in the high number of emission lines of the ThAr lamp whose wavelength is known with extreme accuracy. By locating the pixels on the arcs frame which correspond to the peaks of the emission lines it is possible to create an array of pixel-wavelength pairs which can be interpolated to find the wavelength solution for all pixels of the CCD.

Owing to the fact that calibration and raw frames have been acquired with the same setup, we can be confident that the wavelength map created can be applied on the raw scientific frame to obtain the wavelength value corresponding to the measured flux in each pixels.

scattered background removal During the acquisition of the scientific image is it possible that stray light enters the spectrograph and been scattered on the internal components of the instrument producing a smooth background which illuminates the CCD. To remove this scattered light is sufficient to make a 2D fit of the gaps between echelle orders to produce a full image of the scattered background which has to be subtracted to the raw frame.

object tracing This procedure is required to identify pixels on the individual orders which are associated with the center of the object along the spatial direction. The easiest way to accomplish this task is to slice the order such that each slice represents approximately an image of the slit at a different wavelength as explained in 2.1.1. Later each slice can be searched for a peak to identify the exact object center.

Main issue with this approach is that cosmic rays events can produce peaks in the slices not associated with the object. To overcome this problem only peaks identified unambiguously in more of the 50% of the order are preserved and the remaining points of the trace are interpolated.

sky subtraction Once the trace has been identified it is possible to use a fixed width window centered on the trace for masking the object spectrum along the orders. In other words this is equivalent to "cut" from the slit the portion which includes the object such that only sky surrounding the object will fall in the remaining portion of the slit. The "strips" of the order resulting from this operation can be used to identify the spectrum of the sky background which can be interpolated along the entire height of the slit (including the masked object pixels) and subtracted from the raw frame.

optimal extraction The net result of the previous operations is a 2D array of pixels each one of them having an associated value of the object flux and the corresponding wavelength. To convert this 2D image into a 1D spectrum is possible with a procedure called "extraction" of the spectrum. There are multiple ways to perform this task but for echelle spectra the basic approach is the *optimal extraction* as originally introduced by Keith Horne in the 1986 (Horne, 1986).

This approach is based on the definition of an object profile and an object variance used together for weighting the contribution from pixels recording different portion of the PSF. For example pixels on the wings of the PSF will contribute less to the final spectrum with respect to pixels on the central part of the PSF. The optimal extraction not only improves the SNR of the final spectrum but also allows for an efficient clipping of cosmic rays easily individuated in pixels that exceed for more than 3σ the expected flux value from the profile. The object profile can be approximated with a Gaussian when the SNR of the image is not so high, otherwise it can be fitted along the orders with an approach similar to the one described for the object tracing. The final outputs of this process are the extracted mono-dimensional spectra of the individual orders as showed in the rightmost part of Fig. 3.2.

One can notice that the output of this process is not yet a complete 1D spectrum in physical units, orders are still curved due to the blaze functions of the grating and still need to be merged each other to form the complete spectrum. In order to perform a complete reduction two additional steps are still missing and intentionally omitted: the *flux calibration* and the *orders merging*. These last two steps are very difficult to be performed with

echelle spectra and represent a big difference with respect to standard long-slit spectroscopy, for this reason these will be discussed separately in the next chapter.

The discussion so far has been very simplistic, in practice the accomplishment of each one of the reduction steps is very difficult and expensive in terms of calculations. For this reason nowadays is preferable to perform all the data reduction in an automatic fashion with complex program codes called pipelines.

Pipelines are designed to take into account the individual features of each instrument and to execute in the appropriate way all the steps of the data reduction automatically so that the users do not need to care about the details of the reduction. In practice each instrument has its own pipeline developed specifically, so usually operations are performed in different way such that in most cases final results could be different.

3.2 UVES pipeline vs HIRES pipeline

Even if HIRES and UVES are very similar instruments their dedicated pipelines show significant differences in the approach of some of the reduction steps described previously. It is clear therefore that sometimes different approaches come also from choices of the software developers not always related to the instrument design.

The UVES pipeline has been developed by the Data Flow Systems Department (DFS) of the ESO Software Development Division, basically it has been obtained from a series of procedures or "recipes" written in *ISO-C* programming language which are part of a library called *Common Pipeline Library (CPL)* which is a standard for ESO data reduction packages. As many of the ESO pipelines the different modules can be called from different host applications, namely *Esorex*, which is basically a command line tool, or *Reflex*, which is the recommended application and is represented by a graphical tool which also take care of the data organization and interdependencies between different recipes.

For the HIRES instrument two choices are available: the MAKEE (Mauna Kea Echelle Extraction) pipeline developed by Tom Barlow and the HIRES Redux pipeline developed by Jason X. Prochaska. The former is based on a

series of C and FORTRAN procedures and has been designed to run non-interactively using a set of default parameters, the latter is based on the XIDL library specifically developed for HIRES starting from the routines created for the MIKE Redux pipeline.

The MIKE Redux pipeline Bernstein et al. (2015) is composed by a set of IDL[1] based routines which take advantage of some of the procedures developed for the reduction of the SDSS (*Sloan Digital Sky Survey*) spectroscopic data.

As will be explained in the next section the MIKE package has two main characteristics which make it preferable with respect to other packages: a rigorous approach to the reduction steps which preserve the CCD pixels identity and an implementation of the source code general enough to be extendable to different echelle spectrographs. The HIRES Redux pipeline is one example: the XIDL package indeed represents a set of procedures developed for HIRES but derived directly from the MIKE Redux routines.

In the next paragraph will be presented a comparison between different approaches of UVES and HIRES pipelines to some processes of data reduction in order to emphasize main requirements and motivations for the development of an extended UVES/HIRES pipeline.

3.2.1 Different approaches comparison

One of the first and most evident differences between the UVES pipeline and the HIRES pipeline concern the number of calibration frames required which in turn reflects the different approaches to some of the procedures. ESO pipeline requires two additional frames called *FORMAT CHECK* and *ORDER DEFINITION*.

These two frames can be essentially considered as an arc and a flat field respectively acquired with a short slit and so with respect to the other frames they are characterized by a reduced spatial extension of the spectrum projected on the CCD plane (see 2.1.1).

For a better clarity the following discussion will be focused on the single procedures limiting the comparison only to the most important ones. All the discussion will remain in a general character, further details on these

---

[1]IDL stand for *Interactive Data Language* and is a vector oriented programming language most used for interactive processing of images and large amount of data

procedures will be discussed in the next section with respect to the new pipeline approach.

Wavelength calibration

One of the most difficult processes in echelle data reduction is the wavelength calibration. The main difficulty rise from the typical curvature of the emission lines of the ThAr lamp along the spatial extent of the slit (as can be clearly seen in Fig. 3.1a) which vary not only between different orders but along the same order too.

In order to properly associate a wavelength to each one of the pixels of the CCD recording the spectrum one has to take into account this curvature, the accuracy achieved in this process strongly influences the final results of the following steps like the optimal extraction and in particular the sky subtraction.

First step in wavelength calibration consists in a first identification of the known ThAr emission lines in the spectrum of the ThAr lamp. To automatically perform this step the basic procedure is to compare the extracted arc spectrum with an already known wavelength solution.

The UVES pipeline approach provides a first guess line table using the FORMAT CHECK frame. The procedure is to use a physical model to predict the position on the CCD of a series of reference arc lines. All the input parameters required by the physical model, such the UVES instrumental settings and a series of ambient parameters like the temperature and the humidity, are read from the input raw frame FITS header.

The predicted central positions of these lines are projected onto the FORMAT CHECK frame. The lines are found in the narrow ThAr frame by a two-dimensional centering procedure. The initial dispersion relation, usually based on about a hundred initial detections is refined with successive iterations on the spectrum until most lines are found. A fit of the arc line positions retrieved from this procedure is used to trace orders and obtain a first guess order table useful for the second step of the wavelength calibration.

The HIRES pipeline approach does not require the FORMAT CHECK frame since the first identification of the arc emission lines is performed using a set of archived wavelength solutions available for multiple instrument configurations.

The basic procedure is to make a boxcar extraction of the central column of each order on the arc frame, the ThAr spectra obtained are then compared to the archived solution which corresponds to the current instrumental setup and it is used as a template for the first guess. The procedure cross-correlate with an FFT the extracted ThAr spectrum with the template and corrects for pixels shifts, then a peak finder identifies and records lines matching the wavelength of archived reference arc lines. Arc lines identified with this procedure are used to perform a 2D polynomial fit on the entire CCD to extend the wavelength solution in region which are not covered by arc emission lines.

At this point the next step is to extend the mono-dimensional solution retrieved at the center of the orders to the entire extension of the slit. As anticipated, this is the most difficult task since wavelength solution varies continuously along the slit as showed by the tilt of the arc lines.

UVES approach to solve this problem is to subdivide the slit in three different windows, the wavelength extraction is then performed independently on each window and the wavelength solutions obtained at different slit offsets are interpolated (using linear interpolation) to the measured object offset position. The wavelength solution obtained with this procedure is then applied to the entire height of the slit such that, for each order, all the pixels in one column corresponds to the same wavelength. This approach is equivalent to "rectify" the orders which can be a good approximation for the extraction of the object (whose limited extension occupies only a small fraction of the slit), but does not allow for the exact reconstruction of the profile of sky emission lines since sky occupies the entire height of the slit.

To avoid this issue HIRES pipeline uses a brute force approach fitting on the 2D plane of the CCD all the identified arc lines. Arc lines are well approximated by a straight line so that the procedure calculates the slope of the lines varying along the orders and interpolates measured values to retrieve lines tilt in all the regions of the orders. This approach allow to reconstruct a 2D map of the exact wavelengths associated with each one of the original pixels of the CCD. This map can be used in the optimal extraction to retrieve the exact shapes of the sky and object profiles as a function of wavelength.

Trace definition

First important requirement for a good optimal extraction is the exact definition of the object trace along the orders.

For this purpose the UVES pipeline uses the ORDER DEFINITION frame, which traces the center of the slit, to obtain a first guess of the object position. With this approach the pipeline assumes that the object spatial profile varies slowly with wavelength. This might not be true if the object offset varies significantly with respect to the order location defined by the ORDER DEFINITION frame. Moreover object has to be well centered on the slit otherwise a failure of the automatic procedures should be expected.

The HIRES approach in this sense is much more robust since it rely solely on the detection of "good" trace points detected directly on the processed raw science frame and extrapolated on the entire CCD with Principal Component Analysis (PCA), the detailed procedure is described in 3.4.7. HIRES pipeline approach allows to recover the exact trace of the object independently from its shape and offset with respect to the central portion of the slit.

Sky subtraction

UVES pipeline approach to the sky subtraction is to deal with the sky emission as it is a "pedestal" for the object emission, so assuming a simple model where the sky background is constant it is evaluated together with the spatial profile during optimal extraction. Sky emission lines, since filling the whole slit, are removed during the extraction as they contribute to the linear (constant) term of the spatial profile. A first guess to the sky emission is obtained from a boxcar extraction of the spectrum masking the central region where the object is located.

HIRES pipeline approach is to deal with the sky background in exactly the same way as the object emission. First the slit profile is determined using the flat field image, after masking the object position it is used together with the 2D wavelengths map to obtain a first guess of the sky background. As in the UVES pipeline further refinement of the sky measurement is obtained fitting together the sky background and the object emission. At a variance with respect to UVES pipeline 2D wavelengths map is used in the fit to obtain the measured quantities as a function of wavelengths and no approximation are made on the sky level which is properly modeled by the slit profile. This

kind of rigorous approach allows to maintain pixels identity producing a
final sky model which gives an estimate of the sky background for each one
of the pixels in the slit.

Optimal extraction

The optimal extraction is strongly dependent from the determination of the
object profile. The basic procedure is to fit together the sky background
and object profile imposing the minimization of the χ^2 obtained from the
comparison with real data and errors. This procedure is required for the
determination of the optimal profile which better describes the real shape of
the observed Point Spread Function (see 2.1.1). This procedure is usually
iterated in order to remove cosmic rays pixels, identified by deviations from
the profile, and improve the determination of the final profile.

UVES approach is to apply the profile p_i measured in the $i-th$ pixel to the
not sky subtracted image to obtain a measure of the optimal flux F from
the formula

$$F = \frac{(\sum_i 1/\sigma_i^2) \sum_i p_i f_i/\sigma_i^2 - (\sum_i p_i/\sigma_i^2) \sum_i f_i/\sigma_i^2}{D} \qquad (3.1)$$

and the sky S from

$$S = \frac{(\sum_i p_i^2/\sigma_i^2) \sum_i f_i/\sigma_i^2 - (\sum_i p_i/\sigma_i^2) \sum_i p_i f_i/\sigma_i^2}{D} \qquad (3.2)$$

with variances

$$V(F) = \frac{\sum_i 1/\sigma_i^2}{D} \qquad V(S) = \frac{\sum_i p_i^2/\sigma_i^2}{D} \qquad (3.3)$$

where $D = (\sum_i 1/\sigma_i^2) \sum_i p_i^2/\sigma_i^2 - (\sum_i p_i/\sigma_i^2)^2$ while f_i and σ_i^2 are the not
sky subtracted flux and the correspondent variance of the $i-th$ pixel and
the summation is performed along one column. The wavelength calibration
solution is applied during resampling from pixels to equidistant wavelength
bins recalling that the pipeline associates the same wavelength to all the
pixels of one column. One could eventually output a 2D version of the spec-
trum, nevertheless the flux is not sky subtracted and orders are rectified
loosing the original pixels identity.

HIRES approach is very different and relies on the determination of bi-
dimensional models of the principal quantities involved in the optimal ex-
traction. First the profile fitting procedure uses 2D wavelengths map and

slit profile in order to obtain a 2D model of the sky background as a function of 2D wavelengths, optimal profile too is expressed as a function of the bi-dimensional wavelength solution and it is used to derive a 2D model variance for the sky subtracted image (more details in 3.4.9). The sky model is subtracted from the preprocessed science frame to obtain an image of the object flux only. The net result of this procedure are a series of 2D images which describe wavelength, sky, measured object flux, optimal profile and variance of each one of the pixels illuminated by the slit.

This kind of approach allows for the calculation of the optimal flux for each one of the 2D pixels following the original optimal extraction formula

$$F_i = \frac{p_i f_i / \sigma_i^2}{p_i^2 / \sigma_i^2} \tag{3.4}$$

with a pixel optimal variance

$$V_i(F_i) = \frac{1}{p_i^2 / \sigma_i^2} \tag{3.5}$$

as discussed in Horne (1986) this approach can be applied on single pixels since these can be considered as many independent measures of the flux. In this case p_i, f_i and σ_i^2 represent respectively the optimal profile, the sky subtracted flux and the variance of the $i - th$ pixel. The final 1D spectrum can be obtained by a simple rebinning of the 2D data to the desired new wavelength scale.

As discussed in the next section, one of the most important aspects of the new pipeline is represented by the possibility to output these 2D images, including the 2D optimal fluxes, in order to implement the subsequent analysis with original pixels which are by definition free from resampling noise (see **??**).

3.3   A new UVES/HIRES reduction pipeline

The decision to build a generalized UVES/HIRES pipeline born from the need to extend the HIRES pipeline approach to UVES data and to reduce in the same way data from both instruments.

First this section introduces the GUI (Graphical User Interface) developed for the interactive uses of the pipeline, then the initialization procedures of the pipeline as well as the methods used to organize the data are described.

Since this section and the following one aim to provide an exhaustive but comprehensible treatment of each one of the basic procedure of the pipeline, only the main and most important features of the pipeline algorithms will be described without entering in too technical details. The basic scheme which will be applied in the following discussion is:

- to introduce the main procedures following this naming scheme "*instr_procedure.pro*" where *instr* represents the instrument which can be UVES or HIRES and remain unspecified while procedure is the name of the data reduction process to be performed

- to describe the layout of the code, its behavior and basic functionality

- to discuss main modification introduced in the original code required for the extension of the original HIRES procedure to the UVES data

- to describe most relevant options which could be activated on the main procedure

Throughout the algorithms, the detector is oriented in memory within IDL such that rows correspond to spatial information and columns correspond to spectral information.

3.3.1 Pipeline initialization

Procedures involved in the pipeline initialization are in order of appearance:

- instr_strct.pro

- px_old_uves_conv.pro

- instr_headid

- instr_editstrct.pro

Functionality

Before to start data reduction it is important to properly organize data in folders. The code of the pipeline is structured such that initialization procedure searches raw data in a `\Raw` folder. The best approach is to divide data by targets and nights of observation such that each scientific frame comes together with the correspondent set of calibration frames specific for that

night (recalling that calibration frames required for a useful data reduction have to be taken in the same night of the scientific observation).

When the pipeline is started a dialog asks the user to choose between three different options:

- UVES

- HIRES

- PX_TOOLS

first two options allow to select the instrument while the last one activates a series of post reduction procedures which will be described in the next chapter.

Once selecting the instrument the user are prompted to select the working directory (the one with the `\Raw` folder inside) and if it is the first execution all the folders which will contain the intermediate and final products of the pipeline are created.

At this point the procedure searches for FITS files contained in the `\Raw` folder, for each one of them the header is read and a series of *keywords* are used to identify specific tags in the header containing information about the exposure and the instrumental settings like the exposure time, CCD binning, read out noise, right ascension, slit width, etc.

For each frame all the tags which have been read are saved into an IDL structure (akin to a structure in C), information read from the header are also used to obtain a guess of the exposure type (e.g. BIAS, FLAT, ARC or OBJ) which is saved in the structure.

All the information saved in the structure are used to organize data and for almost all the procedures of the data reduction, for this reason the structure is archived as a binary FITS table and is continuously accessed and updated during the reduction process.

Code generalization

Since each instrument has its specific keywords this has to be taken into account in the different implementations of the procedure for HIRES and UVES, moreover different instruments design imposes different approaches in the organization of data. For these reasons the *instr_strct* procedure has

been the one which required most efforts for the generalization to UVES data.

Basically for HIRES data each frame is subdivided in three different chips: BLUE, GREEN and RED, reflecting the fact that HIRES CCD is composed by three different chips each one recorded in a different extension of the raw FITS frame. So the procedure reads each extension separately and for all the reduction processes different chips can be reduced independently as if these were separate images.

At a variance with respect to HIRES, UVES data comes in two FITS files: one for the BLUE side and the other for the RED side. Moreover while the BLUE FITS has no extension, the RED one has two extensions due to the readout mode of the red CCD which split the pixels plane in two regions identified as *RED lower (Rl)* for the bluer portion and *RED upper (Ru)* for the redder portion.

In order to take into account the different design of UVES without changing too much the basic layout of the code I decided to read the single BLUE FITS file and the multi extension RED FITS file has if these were three different chips. In analogy to HIRES the procedure splits the original files in three separated images: BLUE, REDl and REDu, which can be treated separately for all the reduction process as for the BLUE, GREEN and RED chips of HIRES.

For what concerns the different keywords it has been sufficient to update the keyword structure used by the original procedure with the specific keywords of UVES frames. The main structure used to save information on setup and configuration is also updated in order to account for differences between HIRES and UVES, for example the *decker* tag which identifies HIRES slits is substituted with the *slitlen* tag which is used to save the length of UVES slit.

To ensure compatibility with data prior to August 2002 (which have a different data format, see 2.2.4) an algorithm for the data format conversion (*px_old_uves_conv.pro*) has been implemented in the procedure.

Finally unnecessary UVES calibration frames as the FORMAT CHECK and the ORDER DEFINITION are ignored by the pipeline.

Available options

\EDIT - it is the only available option for the *instr_strct* procedure and it
is activated by default. This option allows the user to check the data
read by the procedure through an interactive window which display the
list of files read and some properties derived from the headers: frame
number, image root, obj name, type, exposure time, CCD binning,
slit width, lamp, central wavelength of the instrumental setup (only
UVES) and CCD chip (only UVES).

By visually inspecting these parameters the user can check whether
or not the initialization procedure fail to identify some of the input
frames, and rule out one or more frames from the reduction when
something goes wrong

3.3.2   Identify different setups

Procedures responsible for the identification of the different setups are in
order of appearance:

- instr_setup.pro

- instr_setup_val.pro

- instr_getarcfil.pro

Functionality

The term *setup* indicates a unique configuration of the spectrograph defined
by four instrumental settings: the slit, the CCD binning, the echelle and
the cross disperser angles. To successfully reduce scientific data, a full set
of calibration exposures for each setup is required. In this sense, one must
treat the observations and data reduction of each unique setup separately.
The *instr_setup* procedure has just the aim to perform this subdivision of
data in different setups and check whether or not all required calibration
frames are available.
The basic approach is to retrieve required information from the general
structure obtained by the *instr_strct* procedure, then the procedure *in-
str_setup_val* gathers together all the exposures which have same slit and
binning, and values of echelle and cross disperser angles compatible within
the interval defined by the tolerance values.

Setups that contains at least one object are numbered from 1 to 50 (which
means max 50 different setups allowed) and correspondent bias, flat, and
arc frames are searched for. Calibration frames not associated with any sci-
ence exposure are assigned to setups from 51 onward. Final results of the
procedure are summarized in a text file which shows setups created for each
night and associated calibrations.

At this stage the arc frame closest in time to the object exposure is also
selected by the *instr_getarcfil* procedure and saved, together with the infor-
mation on setups, in the main structure.

If more than one setup is present in the exposures observed in one night,
different setups can be selected and reduced separately if all the required
calibration are present.

Code generalization

Also in this case the extension of the original code to UVES data required to
take into account the different configuration of UVES. One of the features
of the HIRES spectrograph is that the user can freely choose the orientation
of the echelle grating and the cross disperser to cover the desired wavelength
region, this can potentially result in a very high number of different config-
urations and so different setups but the number of configuration does never
affect the number of chips which measure the spectrum (BLUE, GREEN
and RED).

With UVES things are a little bit different since it is a two arms spectro-
graph and so some configurations can be obtained in the BLUE arm only
while others can be obtained in the RED arm only, this aspect create a link
between chips (BLUE, REDl and REDu) and setups. Moreover as illus-
trated in 2.2.4 UVES offers a series of standard fixed configuration which
are identified by the central wavelength of the cross disperser blaze function
only (see 2.1.5). In most cases standard configurations are well suited to
cover all the spectral regions of interest, is it possible however to choose
configurations different from the fixed ones.

To take into account those differences the procedure *instr_setup_val* for the
exposures-setups association has been modified such that setups are defined
by the following settings: slit, CCD binning, central wavelength and chip.

The different chips selection has been introduced to implicitly eliminate the
distinction between chips and setups such that different chips represent also

different setups, in this way red arm setups appear twice and are distinguished by the letter "l" or "u" according to the respective chip. The effect on the pipeline interface is that for HIRES data is it possible to select chips and setups, while for UVES data only the setup can be chosen since to select the setup is also equivalent to select the chip.

Available options

`XTOLER=xtoler` - defines the tolerance value used for the cross disperser angle, the default value is: 0.001. For UVES data only this option has effect and defines the tolerance value used for the central wavelength

`ETOLER=etoler` - defines the tolerance value used for the echelle angle, the default value is: 0.001

Figure 3.3: Graphical User Interface of the new pipeline (UVES mode in this example). The layout is very simple and allows to perform each one of the reduction step (in the *Procedures* column) interactively checking the results in real time selecting the *chk* options, or at the end of the execution with the procedures in the *Check* column. The user has also the possibility to perform automatically all the steps in one take using the *"ALL"* button.

3.4   Building blocks of the new pipeline

Once finished the initialization procedures the pipeline produces a GUI (Graphical User Interface) illustrated in Fig. 3.3 (UVES mode). The aim of this GUI is to facilitate the interactive use of the pipeline selecting multiple options or execution modes, monitoring the state of individual procedures and checking interactively the results to assess the quality of the intermediate and final products.

Basically the GUI is very similar both for HIRES and UVES, only differences are the selection of setups for the HIRES mode and some additional options for the UVES mode.

With reference to Fig. 3.3 going from the left to the right there are

- the selectors for the chips and setups

- a column for the available options

- the "ALL" button which starts a sequence to perform all the reduction steps in one take

- the single procedures column which allow to launch individually the single procedures

- the post reduction column which includes three procedures for the flux calibration and merging of multiple exposure of the same setup (discussed in Chapter **??**)

- a column dedicated to procedures for the check of the intermediate and final results

- a text window which display the pipeline status during the execution of procedures

- a button to clear the text window and another one to close the interface and exit the reduction

Following the same scheme used for the previous section, in the present one we wish to introduce the single reduction procedures, which are also the basic building blocks of the pipeline, and all the available options; in doing so we will continuously refer to Fig. 3.3 in order to provide an user's guide to the main functionality of the pipeline.

Since there are many subroutines within each one of the main pipeline procedures, only the name of the most important ones will be reported here following for the presentation the order of execution required by the pipeline. Techniques described in this section are quite general so that can be extended to every other echelle spectrograph.

3.4.1 CCD characterization

The CCD characterization includes the calculation of gain and RON and the bias subtraction. Procedures involved are

- instr_mkbias.pro

- xcombine.pro

- instr_subbias_sngl.pro

- instr_findron.pro

- instr_findgain.pro

- x_calcgain.pro

Functionality

The new UVES/HIRES pipeline provides two different methods for the bias subtraction: the first one uses multiple bias images to create a *master bias* frame which is subtracted from the raw images (default method), the second one does not require the bias frames since it makes use of the overscan regions of the CCD.

The first method is executed by the *instr_mkbias* procedure which reads the available bias images for a given setup, all these images are then combined by the *xcombine* procedure which by default takes the median with sigma rejection of the available exposures. This results in a low noise and cosmic rays free image which represents the bias level in each pixel of the CCD, i.e. the master bias frame. The master bias is then archived as a standard FITS file which will be used by the next procedures.

The second method allows to estimate the bias level directly from the overscan regions of each raw frame. Raw frames are read in the same format as it was saved at the telescope, so images are characterized by an "active" region of native CCD pixels and "blind" regions which only measure the bias
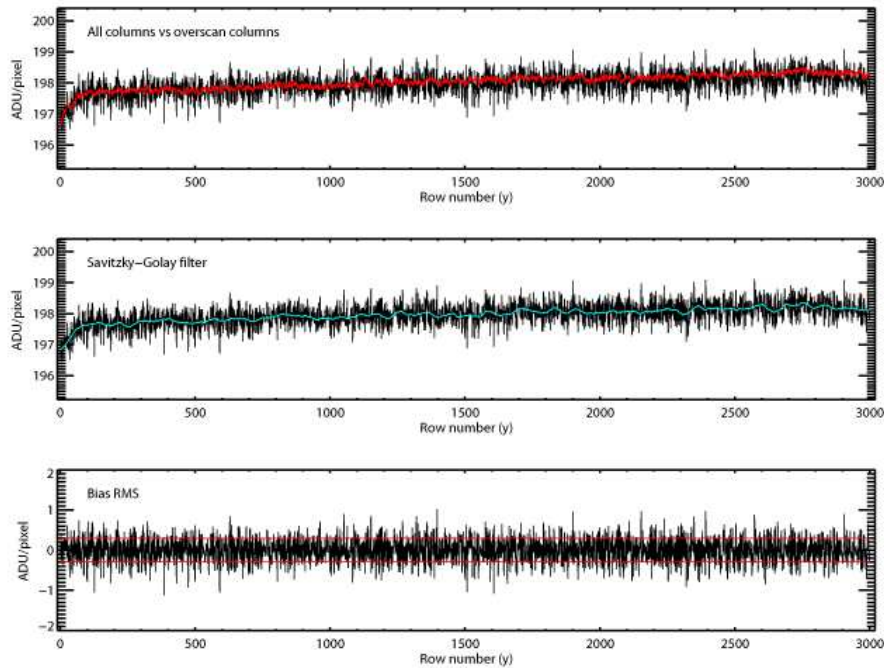
Figure 3.4: The upper panel shows a comparison between the bias spectrum obtained from the average of the overscan column only (black) and the same quantity measured from the average of all CCD columns (red). This plot shows that the two spectra have same slope, indicating that no significant trends are present along the spatial direction of the CCD. Middle panel shows the average of the overscan columns with sigma rejection and the same spectrum smoothed with a Savitzky-Golay filter (cyan). Bottom panel illustrates the bias spectrum normalized with the smoothed version, this normalization allows to measure the RMS of bias pixels (identified by the orange horizontal lines) which represents the CCD read-out noise.

and are called the overscan regions.

In order to be subtracted to the entire frame overscan regions have to be properly modeled. In this case only the overscan region parallel to the spectral direction (CCD columns) are modeled, since there are only small variations of the bias level along the spatial direction as shows the comparison between the average of the CCD columns in the overscan region only (black line in the upper panel of Fig. 3.4) and the same average extended to all the columns of the bias image (red line in the upper panel of Fig. 3.4). Basic operations performed by the *instr_subbias_sngl* procedure are: to produce a 1D overscan spectrum as a function of the CCD rows by an average of the overscan columns with 3-sigma clipping (black line in the upper

and middle panel of Fig. 3.4), to apply on this spectrum a fourth degree Savitzky-Golay filter with a symmetric window of width 121 averaged pixels (cyan line in the middle panel of Fig. 3.4), to replicate and subtract from the original image this averaged and filtered overscan model, to trim the final image to the active pixels only. Each bias subtracted image is saved in the \OV directory of the night's folder.

Once available the bias model can be used to obtain an estimate of the CCD read out noise. As explained in 3.4.6 the RON is an essential contribution to the errors model so, even if the factory value can be read directly from the header of the FITS files, a direct measurement could improve the final errors estimate.

As illustrated in the lower panel of Fig. 3.4, a measurement of the RON can be obtained from the root-mean-squared deviation of the bias pixels, this operation can be performed by the *instr_findron* procedure which measures the RON from the standard deviation of pixels for each available bias frame, the final RON will be the result of an average over all these single RON measurements.

The last step consists in a measurement of the CCD gain from pairs of flat field images assuming Gaussian statistic. The dedicated procedure is *instr_findgain*, if N flat field images are available for the selected setup this procedure makes a sequence of $N - 1$ pairs and from each one obtain an estimate of the gain from the formula

$$image = \sqrt{\frac{1}{1/i1 + 1/i2}} \ln \left[\frac{i1}{i2} \overline{\left(\frac{i2}{i1} \right)} \right]$$

(3.6)

$$1/\langle gain \rangle = variance(image)$$

the mean gain obtained from all the available pairs is recorded and used for the conversion from ADU to electrons for all of the images being processed. Finally experimental tests (Bernstein et al., 2015) had shown that master dark frames do not significantly affect the overall quality of processed images.

Code generalization

Two of the procedures described here was not available in the original code and so they have been specifically developed for the new pipeline: the *instr_mkbias* procedure for the creation of the master bias and the *instr_findron*

procedure for the determination of the CCD read out noise.

The generalization of the other procedures has only required to account for the different code layouts and for the different dimensions of UVES CCD ($2148x3000$ blue chip and $2148x4096$ red chips) with respect to HIRES CCD ($2140x4096$ all chips).

Available options

`\CHK (available from GUI)` - in this case the only effect of this option is to show an image of the processed bias frames for the calculation of the RON

`\CLOBBER (available from GUI)` - if enabled allows procedures to replace old products (if present) with the new ones

`\OVERSC (available from GUI)` - this option allows to select the bias subtraction methods. By default the pipeline searches for bias frames to construct the master bias, this option tells the program to use the overscan regions

3.4.2   Orders definition

In this step the master flat for the flat field calibration is constructed and it is used for the definition of the echelle orders boundaries. Procedures used are

- instr_mktflat.pro

- xcombine.pro

- instr_edgeflat.pro

- x_edgeflat.pro

Functionality

The flat field frames can be used for a variety of purposes. For example they can provide an estimate of the gain, as described in the previous subsection. In this pipeline the main processes which make use of the flat field frames are: the spatial definition of the echelle orders boundaries (*trace flat*), the definition of the spatial profile of an uniformly illuminated slit for each echelle
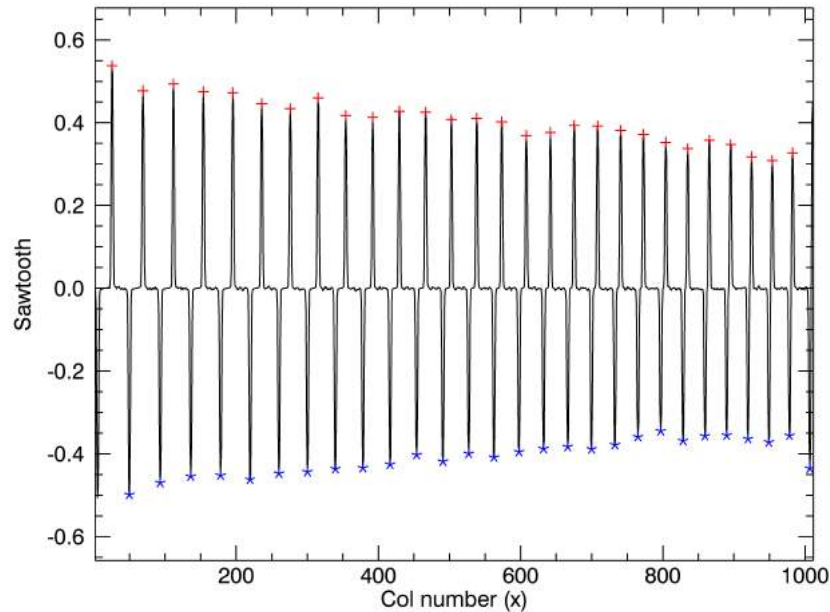
Figure 3.5: This image represents the sawtooth filtered master trace flat averaged along the spectral direction. Positive peaks (red cross) identify the left sides of orders while the right sides are identified by the negative ones (blue stars).

order (*slit flat*) and the relative pixel-to-pixel CCD response correction (*pixel flat*). Here only the former is described while the two latter are discussed separately in 3.4.4 and 3.4.5.

The pipeline first produce a *master flat* frame from the combination of all the available flat field frames. The *instr_mktflat* procedure reads the raw flat field frames, applies the bias subtraction and trim the images. Then the *xcombine* procedure uses the median value of each image to calculate a scale factor which takes all the flat field frames to the same level, converts the images from ADU to electrons using the measured gain and makes a median with 3-sigma rejection of all the images to produce a low noise and cosmic rays free flat field image, i.e. the master flat.

Pixels counts of the master flat are then used to obtain the correspondent error image following the model described in 3.4.6. The master flat is then saved to a standard FITS file and its errors image is added to this file as an extension.
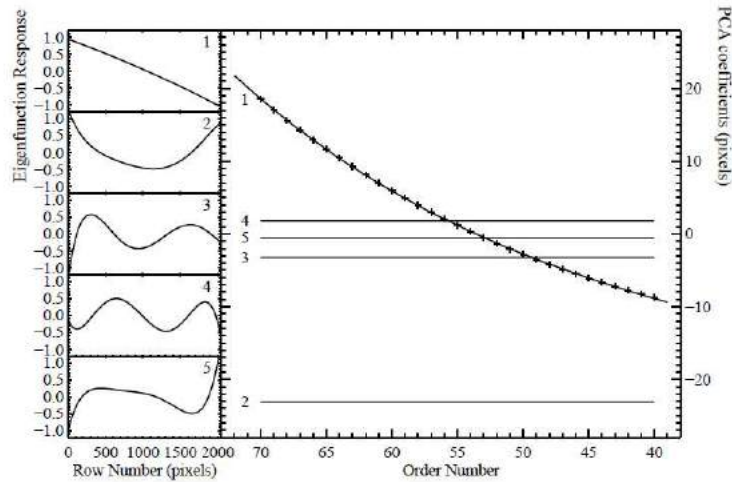
Figure 3.6: The left panel shows eigenfunctions of the PCA analysis of the Legendre polynomial coefficients which describe the center of orders. Eigenfucntions are normalized between −1 and 1. The large right panel shows correspondent PCA coefficients as a function of the order number. The code interpolates bad orders fitting a 2nd order polynomial on the first coefficient of the PCA, other coefficients are essentially constant (Bernstein et al., 2015).

At this point the saved master flat can be passed to the *instr_edgeflat* procedure which searches for the location of echelle orders boundaries. The procedure applies a sawtooth filter in the spatial direction (along rows) producing a series of positive and negative peaks as illustrated in Fig. 3.5 which shows the sawtooth averaged over the spectral direction of the CCD. Positive and negative peaks are then used to identify and trace along the length of the CCD the location of left and right edges respectively.

At this point edges that have been successfully traced in 90% or more of the respective echelle orders are used to predict the edges of the remaining orders using *Principal Components Analysis* or *PCA* (Murtagh and Heck, 1987).

The basic approach is to use the good traces to calculate the center and the width of the orders as a function of CCD row number. Can be demonstrated that the slit length along the CCD plane can be described by a 2D polynomial linear with respect to row number and quadratic with respect to order number. On the other hand the orders centers can be fitted by a Legendre polynomial as a function of the CCD row number. The PCA is then used to obtain the basis functions and coefficients which describes the
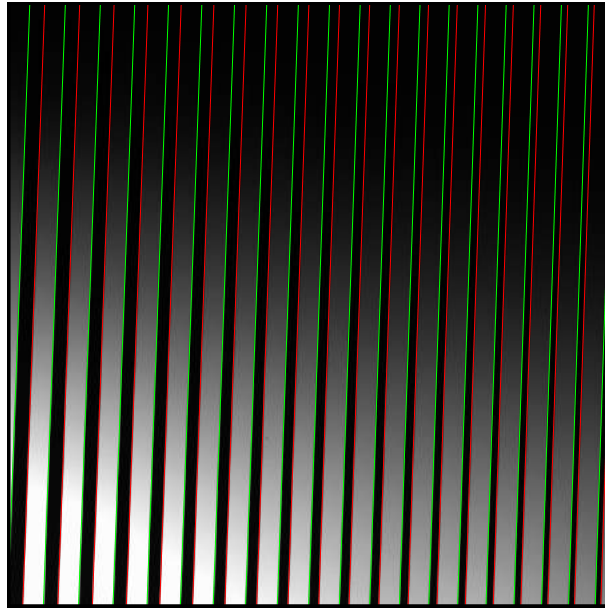
Figure 3.7: Example of the order edges (red for the left and green for the right) obtained by the *instr_edgeflat* procedure overimposed on the respective UVES master trace flat (small portion of the lower red chip for setup 520).

central positions of the good orders.

As an example of the PCA performed by this procedure in Fig. 3.6 are illustrated both the eigenfunctions (in the left panels) and the respective PCA coefficients (in the main rightmost panel).

The linear combination of these functions multiplied by the respective coefficients allows to accurately retrieve the exact location of the orders center, for all orders and for all rows.

Once the PCA basis and coefficients have been calculated, the analysis can be extended to all the orders which fall on the CCD by a simple polynomial fit of the good PCA coefficients. This allows to extrapolate in a robust and reliable way the trace of the central position, and so the edges, for orders with poorly defined edges, even in regions of the CCD where the signal is very faint.

An example of the outcome of this routine is illustrated in Fig. 3.7. All the information on orders edges and central traces are saved in a IDL structure and archived as a FITS table with the prefix *OStr_* in the \Flat directory of the main folder in order to make it accessible by the subsequent procedures. The results of this procedure can be checked interactively with the *chk_flat*,

chk_arc and *chk_obj* commands in the "Check" column of the GUI. These commands produce respectively a flat, arc and object image with the defined edges overplotted. Those procedures can be also useful to visually inspect the behavior of available frames.

Code generalization

In this case the generalization of the *instr_mktflat* and *instr_edgeflat* procedures only required to account for the different codes layout which depends from instrumental differences between HIRES and UVES as explained in 3.3.1 and 3.3.2.

Modifications of the original code include the correction of a bug identified in the *xcombine* procedure. The procedure invokes a C program for the calculation of the median of multiple images with sigma rejection. The model for the calculation of sigma was based on two wrong assumptions: that the inverse gain was supplied and that the images to combine were in unit of e^-. This bug then was propagated to the *instr_mktflat* procedure, which stored the master trace flat in units of ADU and the model variance in units of e^-, and then to all the other procedures which make use of the master trace flat.

The corrected procedures calculate both the median of the images and sigma in units of e^-, in the same way the resulting master trace flat and its variance are stored in unit of electrons.

Available options

\CHK (available from GUI) – if enabled show the sawtooth plot of Fig. 3.5 during the execution of the *instr_edgeflat* procedure

\CLOBBER (available from GUI) – if enabled allows procedures to replace old products (if present) with the new ones

\OVERSC (available from GUI) – by selecting this option the flat field frames are bias subtracted using the overscan regions method

\AVG (available from GUI) – this option allows to combine multiple flat field exposures by an average with sigma rejection rather than a median with sigma rejection

\NOSORDR - by default the pipeline try to ignore ghosts and second order light which could appear, especially in the red chip of HIRES, in the inter-order gaps. These features could produce a failure of the *instr_edgeflat* procedure and so are ignored by imposing a minimum distance between orders. This option allows to skip the second order light identification

NOOVERLAP=nooverlap - default value of this parameter is 4, this means that orders with a separation less than 4 are extrapolated by the PCA

3.4.3 Wavelength calibration

The following procedures are used to obtain a 2D wavelength solution from the available arc images. The order structure, bias and normalized master flat are required in input

- instr_allarc.pro

- instr_allarc_sngl.pro

- instr_procarc_sngl.pro

- instr_fitarc.pro

- instr_arctempl.pro

- instr_fit2darc.pro

- instr_tracearc.pro

- instr_fittrcarc.pro

- instr_mkaimg.pro

Functionality

All the procedures for the wavelength calibration are performed together automatically by the *instr_allarc* procedure. The basic steps which allow to obtain a 2D wavelength solution from the arc images are now described following the logical order of execution.

First arc frames are processed, so bias subtracted and flat fielded (see 3.4.5) by the procedure *instr_procarc_sngl*.
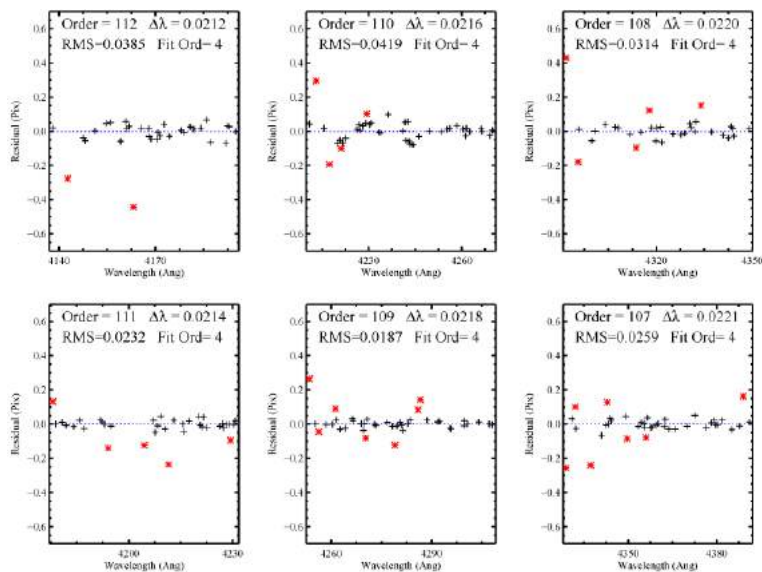
Figure 3.8: Residuals from the fit of identified ThAr lines on few orders from UVES setup 520. Red stars indicate points rejected from the fit, as can be seen only pixels whose residual lie between -1 and 1 are considered good for the fit. This kind of plots are also available in the *QA* directory of the main folder to assess the quality of the final products.

The second step is performed by the *instr_fitarc* routine and produces a list of well identified arc lines with their centroid pixels and associated wavelengths. The code first rectifies the orders and makes a pseudo-boxcar extraction of the central column, this produces a series of arc spectra correspondent to each one of the echelle orders. Then a bi-dimensional array (i.e. a matrix) is constructed where each row represents an order and is filled with the correspondent extracted spectrum. This array is then cross-correlated with an archived template using an FFT algorithm.

The template is probably the most important ingredient of this process, it is represented by a wavelength solution of the individual orders obtained with an instrumental setup similar to the one to be reduced. This solution comes in the same bi-dimensional format which has been applied to the extracted arc to be calibrated, the only additional information are the correspondent wavelengths and the physical order numbers. The FFT algorithm is required to properly measure the shift between the template and the extracted spectra and to assign to each order the correspondent physical number which is recorded in the order structure produced in 3.4.2. The pipeline uses the *instr_arctempl* procedure to find the archived template which best matches

(a) Entire CCD   (b) Individual orders
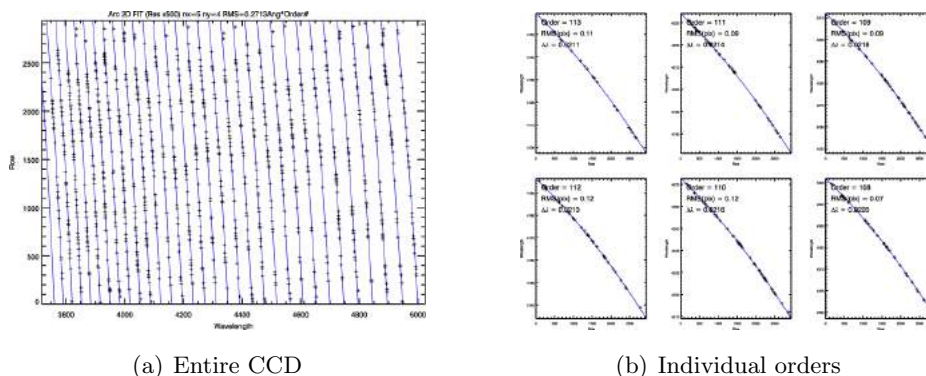
Figure 3.9: Left panel illustrates an example of the bi-dimensional fit performed on iden-
tified ThAr lines over the entire CCD. The fit (blue lines) represents the wavelengths in
the center of echelle orders vs CCD rows for UVES setup 520. Black crosses represents
the offsets of arc lines from the 2D solution expressed as (residuals × 500). The right panel
shows residual of the 2D fit on individual orders. This plot also is available in the *QA*
directory.

the present setup.

Once the shift has been applied the archived wavelengths can be used as a
first guess to identify known emission lines of the ThAr lamp. The code first
searches for high SNR peaks in the spectrum whose fluxes in the five central
pixels f_i respect the condition $f_1 < f_2 < f_3 > f_4 > f_5$. Then peaks which
lies within three pixels from a laboratory calibrated ThAr line are recorded
together with their measured centroid. At this point the centroid pixels
of identified lines are fitted with a third order Legendre polynomial with
rejection as a function of the logarithm of the correspondent wavelengths.

The solution obtained with this procedure is then used to repeat the peaks
identification process including this time even lower SNR lines. The new
set of identified lines is then fitted with a fourth or fifth order Legendre
polynomial with rejection in order to achieve residuals with $RMS < 0.1 pix$
on the most of the orders. In order to assess the overall quality of the final
products, the plot of the residuals for each order, as illustrated in Fig. 3.8,
are available in the \QA directory of the main folder.

The third step is performed by the *instr_fit2darc* routine and produces a
wavelength solution down the center of each echelle orders. This routine
performs a 2D fit to all of the good ThAr lines identified by the previous
procedure for all of the echelle orders. Basically it uses a 2D Legendre
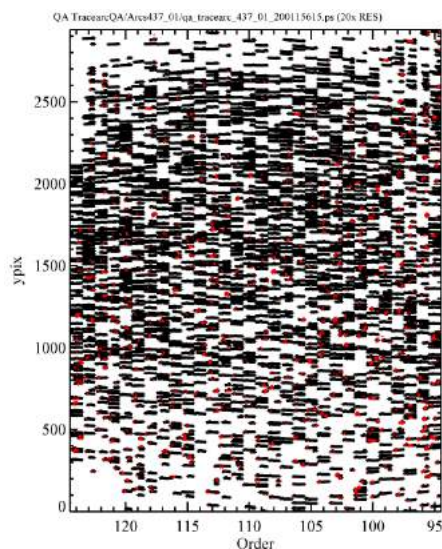polynomial to fit the $\lambda_i m_i$ products, where the λ_i are the wavelengths of the

Figure 3.10: This picture illustrates the approach to the calculation of the ThAr lines slopes across echelle orders. This image represents the arc frame in coordinates of row number vs order number, peaks identified along the entire length of the slit are fitted with a straight line used to derive the slope. Black crosses are (residuals × 20), red crosses represents rejected peaks. Image is available for the quality assessment in the *QA* directory.

arc lines identified and m_i are the correspondent echelle orders, as a function of the echelle order number and the row number y_i such that

$$m\lambda = P_{jk}(m, y) \tag{3.7}$$

A typical value of the RMS for the 2D fit performed by this procedure is $\approx 0.2 Ang * Ord\#$. The result obtained on the entire CCD is illustrated in the left panel of Fig. 3.9 while the right panels show the results of the 2D fit on individual orders. This 2D fit of the identified arc lines allows to confidently interpolate and extrapolate the central wavelength solution for each order on the CCD, even in those regions which are not properly covered by arc emission lines.

The next step is to extend the wavelength solution measured in the central region to the entire length of the slit throughout all the echelle orders. To accomplish this operation first the pipeline uses the procedure *instr_tracearc* to identify all the 5 sigma peaks in the spectrum for each column of the CCD. Then the procedure traces arc lines from the center of the order towards the edges fitting with sigma rejection a straight line over the centroid of peaks which are identified in the entire length of the slit, Fig. 3.10 shows an
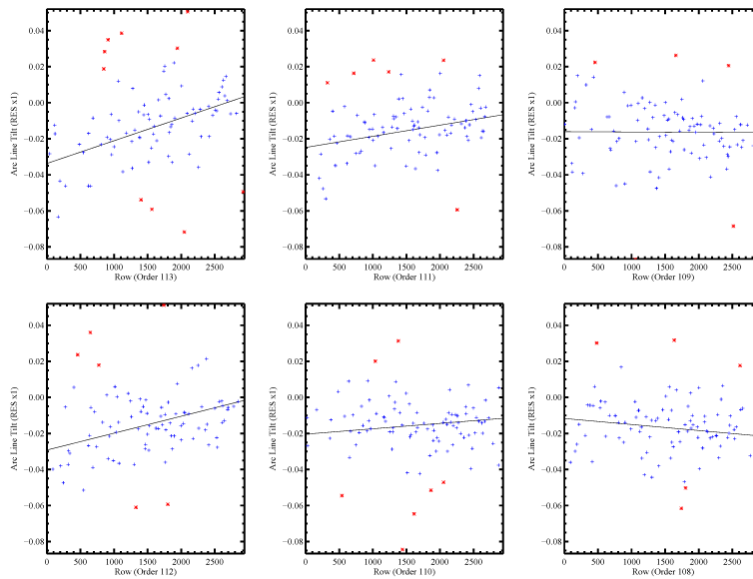
Figure 3.11: Here can be seen an example of the fit over slopes measured in each echelle order for the UVES setup 520. The linear trend of the slope in individual orders appears to change slowly from one order to the other. These plots are also available in the *QA* directory.

example of this process. The best fit calculated in this way allows to retrieve a value of the slope for each arc line.

At this point the slopes of the arc lines over the entire echelle footprint can be interpolated or extrapolated with a 2D Legendre polynomial fit with rejection of the measured slopes as a function of the CCD rows. The 2D fitting process, which is performed by the *instr_fittrcarc* procedure, proceed in a very similar way to the one described for the 2D fit of arc lines. The results of this procedure are illustrated in Fig. 3.11, which clearly shows slopes variation both between different orders and along the same order too. Given the 2D solution for the slope of the lines as a function of position and the 2D solution for the central wavelength, the *instr_mkaimg* routine can finally creates a wavelength image, i.e. assigns a unique wavelength value to the center of each pixel in each order. A simple spline interpolation is used to determine the values.

The final 2D wavelength solution is archived as a FITS file in the `\Arcs` directory of the main folder.

In case of automatic procedure failure on one or more orders, the wavelength solution can be corrected interactively using the *tweak_arc* command in the

"Check" column of the GUI. If the interactive tweak procedure goes wrong, lines can be automatically re-identified with the *auto_id* command.

Code generalization

Besides the changes required for the general layout of the code, one of the most important things required to extend the HIRES procedures for wavelength calibration to the UVES data concerns the creation of the templates used for the first guess solution.

To ensure that the *instr_fitarc* procedure works properly, the UVES templates should have the same format of the HIRES ones. The basic format used for the templates is an IDL SAVE file which contains the following variables: physical order numbers for a given setup and an extracted and wavelength calibrated arc spectrum for each order of this setup.

In order to make those templates we used the ThAr spectrum obtained by Murphy et al. (2007). This spectrum is obtained with UVES data from multiple setups covering all the available UVES wavelength range, it is provided by the authors as a single spectrum which is the result of the merging of all extracted orders from all available setups. The log-linear dispersion of the final product is set to $1.75 km/s$.

In addition to this spectrum our procedure makes also use of one of the secondary products of the ESO UVES pipeline, the file *wave_map_chip.fits*. For a given setup this file contains an image of the UVES wavelength solution for pixels in the echelle footprint mapped over the CCD.

Those wavelength maps have been used to extrapolate individual orders from the full ThAr spectrum and to convert fluxes and wavelengths to the same system produced by the boxcar extraction of the arc images. This last step is crucial to compare consistently the boxcar extracted arc spectra of the rectified orders with the templates, as described in the previous subsection. In order to cover all the possible instrument configurations a template for each one of the standard UVES settings described in 2.2.4 has been produced. For the red arm configurations templates have been splitted in two in order to account for the separate treatment of lower and upper chips.

Finally the physical order numbers for a given setup have been retrieved from the online UVES exposure time calculator.

Available options

`\CHK (available from GUI)` - in this case enabling this option allows the
user to perform a manual wavelength calibration. This option is not
recommended

`\CLOBBER (available from GUI)` - if enabled allows procedures to replace
old products (if present) with the new ones

`\OVERSC (available from GUI)` - this option tells the procedure to per-
form bias subtraction on arc frames using the overscan method

3.4.4   Slit profile correction

Once available the 2D wavelength solution it is used together with the mas-
ter trace flat to measure the slit profile as a function of wavelength, the
procedures used are

- instr_slitflat.pro

- x_slitflat.pro

- instr_fitscatt.pro

Functionality

First of all the procedure *instr_fitscatt* is used to subtract the diffuse scat-
tered light which typically affects the background of the measurements.
Since the scattered light illuminates uniformly the CCD even outside the
echelle footprint, the basic approach is to mask echelle orders to model only
the light in the inter-order gaps. Then a low order 2D b-spline is used to
fit pixels in the gaps and to interpolate over the entire CCD. The scattered
background model obtained by this procedure is then subtracted from the
image, an example of the result is shown in Fig. 3.12.
The aim of this routine is to measure the relative throughput of the spec-
trograph along the slit length (i.e. the slit profile) for a uniform and ho-
mogeneous source (as should be the dark night sky), this can be done for
example using the master trace flat calculated in 3.4.2. The *instr_slitflat*
procedure accomplishes this task in two separated steps for each order: the
determination of the mean slit profile and the measurement of any possible
trend of the profile along the spectral direction.
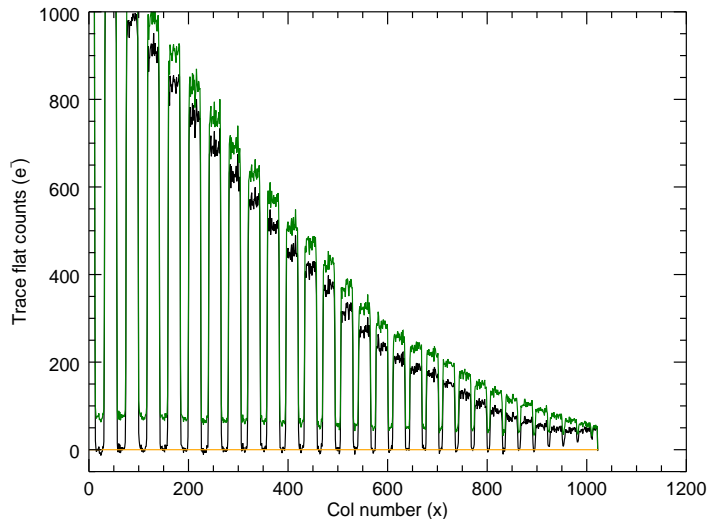
Figure 3.12: Cross section of UVES master trace flat before (green) and after (black) scattered light subtraction. The procedure makes a 2D interpolation of the inter-order gaps to obtain a full image of the scattered light background.

The first task is accomplished fitting a b-spline as a function of wavelength on a central stripe of pixels having a width corresponding to the 70% of the slit length. Then the b-spline coefficients are used to extend the fit to all the pixels which fall in the order and to normalize the trace flat along the spectral direction.

The position of pixels along the slit is expressed in terms of slit fraction and more precisely in unit of *half-slit-length*. In this system 0 indicates a pixel in the center of the slit and 1 or -1 indicates pixels on the opposite edges. This strategy allows to implicitly account for a slowly varying slit length along the orders.

Using this system the mean profile can be calculated fitting a b-spline to all the pixels of the normalized trace flat in the order as a function of the slit fraction as illustrated in Fig. 3.13. The b-spline is then evaluated from -1.25 to -1.25 half-slit-length and stored in the PROFILE0 tag of the order structure created in 3.4.2.

As can be seen in Fig. 3.13 the slit profile calculated from trace flat is characterized by a significant linear gradient across the extent of the slit due to the non-uniform illumination of the flat field lamp. This trend is removed
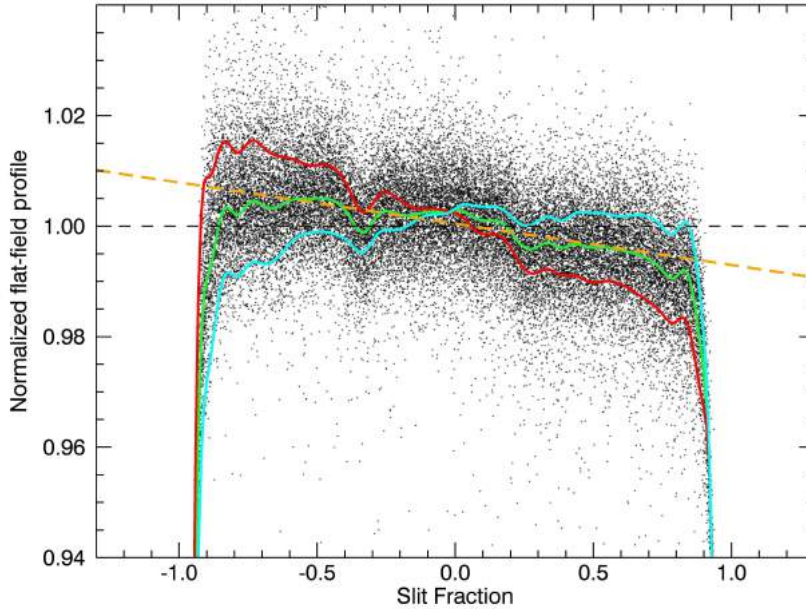
Figure 3.13: Example of the slit profile recovered for an echelle order of UVES. The scattered black points are pixels of the master trace flat plotted as a function of the slit fraction. In order to show the variation of the profile along spectral direction, green, red and cyan lines represent respectively the mean profile and the upper and bottom CCD rows profiles derived as PROFILE0+PROFILE1 and PROFILE0-PROFILE1. The orange dashed line shows the typical tilt of the profile due to non uniform illumination of the flat lamp.

by default with a linear fit of the central 70% of the mean slit profile (orange line in Fig. 3.13).

Profile variations along the spectral direction are approximated with a linear function of the row number. First the mean slope is calculated similarly to the mean profile fitting a b-spline to the residuals of the normalized trace flat about the corresponding b-spline fit calculated before. This mean slope is then re-sampled according to the mean profile and stored in the PROFILE1 tag of the order structure.

The PROFILE0 and PROFILE1 arrays can now be used to retrieve the slit profile model for each row number. The slit profile calculated in this way will be only used as a first guess to model the sky during the final fitting of the object profile which is described in 3.4.9.

Code generalization

The generalization of those procedures has not required too much changes other than the adaptations of the codes layout.

Available options

`\CHK (available from GUI)` - shows a spectrum of the flat and a plot of the mean fitted profile, 2D fitted profile and pixels of the normalized flat as a function of the slit fraction for each order

`\CLOBBER (available from GUI)` - if enabled allows procedures to replace old products (if present) with the new ones

`\NODETILT (available from GUI)` - this option skip the removal of the linear tilt along the slit length (not recommended if trace flat being used)

3.4.5 Flat field calibration

This procedure has been developed for the calculation of the pixel-to-pixel flat from the trace flat obtained in 3.4.2

- instr_nrmflat.pro

Functionality

For what concerns the pixel-to-pixel CCD response correction, the basic approach of the original pipeline was to rely on particular exposures called *milky flats* measured in optimal conditions and subsequently archived by the staff members of the HIRES spectrograph.

The milky flat exposure is obtained from multiple observations of the twilight sky, this ensure a uniform and homogeneous source. In order to fully illuminate the entire CCD and to fill the inter-order gaps, an holographic diffuser which spreads the beam of light at any wavelength is placed behind the slit. In this way it is possible to obtain a uniform illumination of all the CCD pixels.

At this point all the exposures obtained in this way are co-added to produce the final milky flat. Large scale flux variations are then recovered as a function of position on the CCD and used to normalize the milky flat. The

normalized milky flat is median filtered along the spectral direction and a 2D median filter is also applied on the entire frame to mask pixels which deviate by more than 3%. The resulting image is archived as a standard FITS file.

The package of the HIRES pipeline includes a series of milky flat exposures measured, at different times, for almost all the available chips and binning of the CCD and for the two different cross dispersers.

The pixel-to-pixel correction is simply obtained by the ratio of images to the archived milky flat closer in time and with the same binning as the image to be corrected, as explained in more details in 3.4.6.

Now the problem was that for the UVES CCD no archived milky flat solution was available, then to compensate for this lack we decided to accomplish the same steps using instead of the milky flat the master trace flat produced by the *instr_mktflat* procedure (3.4.2).

The procedure responsible for the creation of the normalized master trace flat is *instr_nrmflat*. First the scattered light background is subtracted by the *instr_fitscatt* procedure as described in 3.4.4. Then the order structure produced by the *instr_edgeflat* procedure (3.4.2) it is used to locate orders and perform the normalization on each one of them separately.

Now the routine proceed in two different ways depending on whether a slit profile is already available or not. If available the slit profile is used to correct for the large scale flux variations in the spatial direction (i.e. along the slit length).

Remaining flux variations in the spectral direction (which essentially represent the blaze function of the echelle order) are then corrected with the following procedure: pixels which belong to the same echelle order are sorted in ascending wavelengths, a spectrum is obtained from the median of the 11 columns located around the central position of the order, the value of the normalization function for each pixel in the order is obtained from a b-spline interpolation of this spectrum evaluated at every 8 pixels.

The 8 pixels interval has been chosen to remove the strong fringing which affects upper and lower chips of the UVES red CCD. The result of this operation is illustrated in Fig. 3.14 which shows the master trace flat before and after the normalization.

The procedure can be executed also before the wavelength calibration to obtain a rough pixel flat to be used for the arc frames pixel-to-pixel correction.
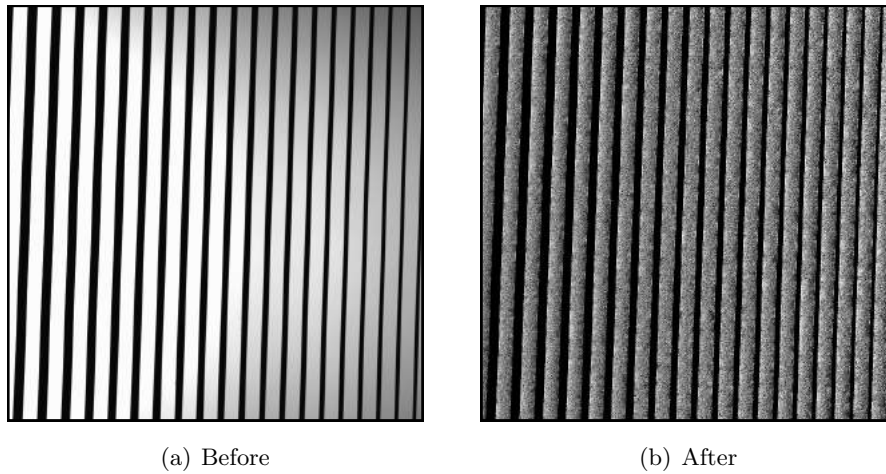
(a) Before   (b) After

Figure 3.14: A portion of the master trace flat for the lower red chip of UVES setup 520 before (left panel) and after (right panel) the normalization procedure. Low frequency variation are removed fitting a b-spline over the flat spectrum, the breakpoints spacing is large enough to preserve high frequency variations (pixel-to-pixel).

In this case the normalization in the spatial direction cannot be performed and the master trace flat is normalized only in the spectral direction.

Code generalization

The *instr_nrmflat* procedure is a routine not implemented in the original code of the HIRES pipeline, its purpose is to compensate for the lack of archived milky flats for the correction of the relative pixel-to-pixel response of the CCD.

Available options

`\CHK (available from GUI)` - if enabled shows the results of the scattered light background subtraction, the plots of the b-spline fits performed on individual orders and two images of the master trace flat: before and after the normalization

`\CLOBBER (available from GUI)` - if enabled allows procedures to replace old products (if present) with the new ones

`\AFRAME` - this option allows to select the execution mode of the procedure, if enabled the procedure searches for the arc solution and the slit profile to correct variations over the spatial direction

3.4.6 Raw frame processing

At this point raw scientific frames can be processed

- instr_proc.pro

- instr_proc_sngl.pro

- instr_subbias_sngl.pro

Functionality

The functionality of the *instr_proc* procedure is very simple since basically it reads all the images to be processed, bias subtract them with the specified method (3.4.1), converts the counts from ADU to electrons multiplying pixels by the estimated gain (3.4.1), divides out for the pixel-to-pixel flat calculated before (for HIRES data searches the `\Keck\HIRES\CALIBS\FLATS` path and select the most appropriate solution among available ones) and rules out bad CCD rows which are flagged by default with zero inverse variance (zero statistical weight).

The procedure is also responsible for the error estimate in the original pixels. As explained in 2.2.1 the sigma of each CCD pixel comes from the summation of the Poissonian noise, given by the square root of the counts of that pixel, and the read-out noise (3.4.1). The problem with this definition is the large systematic bias which one encounters when the number of counts in one pixel approaches zero. In order to account for this effect errors are modeled using the following approximation

$$\sigma^2 = |image(e^-) - 7| + 7 + RON^2 \qquad (3.8)$$

which imposes at least 7 counts in each pixel. This error measurement represents only an initial estimate since it is propagated in every reduction step which follows and for each operation performed on original pixels. The first propagation performed by default in this procedure accounts for the pixel-flat variance introduced by the flat fielding process.

The Eq. 3.8 shows clearly that when pixels counts go to zero (for example in correspondence of a strong absorption line) the reliability of the errors estimates strongly depends on the correctness of the RON value.

The processed image is saved in a FITS files and the variance image is saved in the extension #1 of the file, the file is then archived in the `\Final` directory of the main folder.

Code generalization

One of the difference of UVES with respect to HIRES is that the gain can assume values less than one, this means that each count corresponds to a less than one electron. Following the error model implementation of the ESO UVES pipeline we modified the code to account for discretization noise introduced when the gain is less than one, the modified error model is

$$\sigma^2 = |image(e^-) - 7| + 7 + [(1 - GAIN^2)/12] + RON^2 \qquad (3.9)$$

and is applied to data only if $GAIN < 1$. Another important aspect concerns the bad rows of UVES CCD. Since UVES undergoes various CCD changes during its operational life, to extend the compatibility of the pipeline we implemented a procedure which reads the observation date from the FITS header and for each CCD rules out appropriately the bad rows imposing zero inverse variance.

Available options

`\CLOBBER (available from GUI)` – if enabled allows procedures to replace old products (if present) with the new ones

`\OVERSC (available from GUI)` – this is to select the bias subtraction method, if enabled the overscan region is used rather than the master bias

`\NOFLAT` – this option allows to skip pixel-to-pixel correction in cases where the pixel flat is not available

`\IV _CORRECT` – this option is enabled by default and implements the propagation of the pixel flat variance into the estimated CCD pixels variance

3.4.7   Object tracing

The position of the object along the slit for each one of the echelle orders is traced by the procedure

- instr_fntobj.pro

- x_fntobj.pro
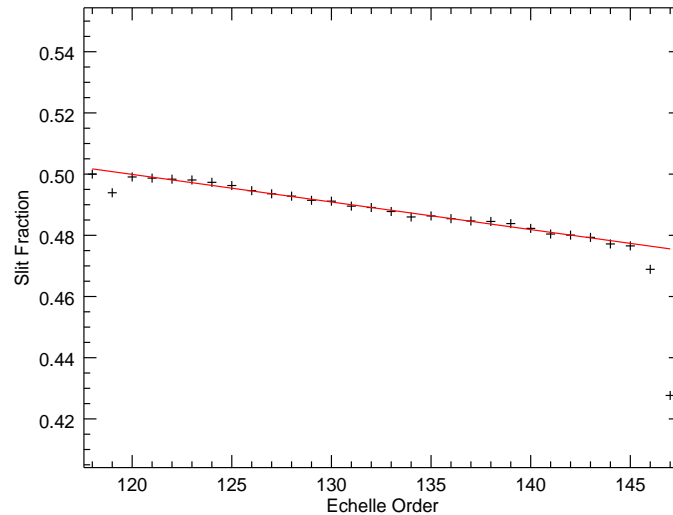
Figure 3.15: Linear fit (in red) of the object central positions found in rectified echelle orders (black crosses). Central positions are expressed in slit fractions (0.5 indicates the center of the order) and plotted versus the order number. The fit is used to retrieve an initial guess of the central position for the faintest orders. The variation of the position with respect to the center of orders has an atmospheric origin.

Functionality

A successful extraction of the object spectrum strongly depends from the estimate of the object position along the slit across the entire extension of the echelle orders, called the *object trace*.

Similarly to the procedure described in 3.4.2 for the determination of the orders edges, the *instr_fntobj* procedure accomplishes this task in a robust way using PCA analysis to extrapolate the object trace in those region where counts approach zero.

First orders are rectified and collapsed along the dispersion direction with rejection of bad pixels and cosmic rays to locate the position of the most significant peak. Peaks identified in ten or more orders are associated with the object and their central positions, expressed as a fraction of the order width, are used to obtain the object position in the remaining orders with a simple linear interpolation as illustrated in Fig. 3.15 This Figure also shows that the object trace moves relative to the center of each echelle order due to atmospheric effects.
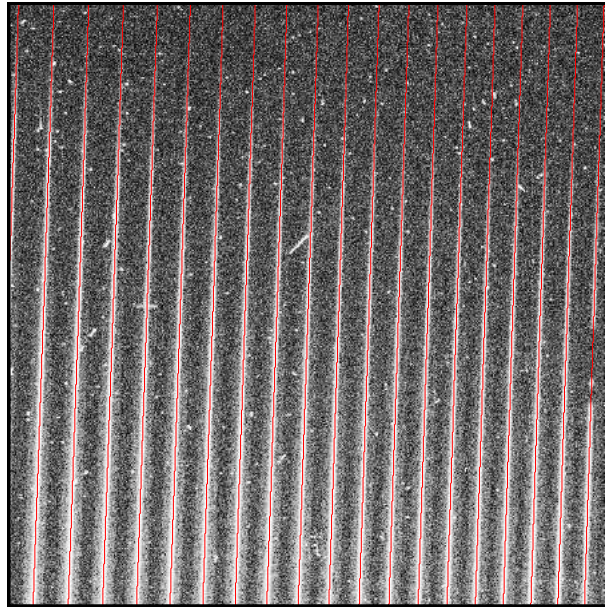
Figure 3.16: This portion of the lower red UVES chip shows an example of the traces derived by the procedure (in red). As can be seen the result is very robust since PCA extrapolation allows to retrieve the good solution even in regions of negligible flux toward the edges of the CCD.

Object positions in rectified orders are then used to retrieve a first guess of the trace in the original science image using the orders boundaries as a reference.

Subsequently a flux-weighted algorithm is used to re-center the initial guess of the trace and a 6th order Legendre polynomial is used to perform a fit with rejection of the trace centroid for each row of the order. When less of the 50% of the centroids are rejected orders are considered good and used for the extrapolation of bad orders with PCA analysis.

The basic approach is very similar to the one described in 3.4.2, PCA coefficients derived for the good traces are fitted with low order polynomials and used with the respective eigenfunctions to retrieve an analytical solution for the object trace in each order. The analytical solution is then used to perform a last iteration for the measurement and fit of the centroids and only in those cases where the SNR is sufficiently high the empirical values are retained, otherwise the extrapolated analytical solution will be used. An example of the result of this procedure is illustrated in Fig. 3.16.

At this stage is also created the *object structure* which will be used until the

end of the data reduction to save the final products of the extraction (such as the 1D wavelength and spectrum or other important properties such as the FWHM of the object profile), the structure is archived as a FITS table with the prefix *Obj_* in the `\Extract` directory of the main folder.

One of the tags of the object structure defined by this procedure is the *aper* tag. This parameter is very important since it is used to identify the boundaries of the object profile, i.e. to delimit the fraction of the slit which only contains flux from the sky. This parameter is set by default to the 60% of the slit length.

Finally the *trace* tag is used to save the object trace derived at this point of the data reduction process which does not yet represent the final trace. The final trace will be the result of small deviations correction performed during the optimal extraction procedure and is saved in the *fin_trc* tag of the object structure.

Code generalization

The generalization of those procedures has not required too much changes other than the adaptations of the codes layout. Since UVES slit lengths have usually larger values with respect to HIRES, this allowed to increase the fraction of the slit used for the profile calculation which has been changed by default from the 60% to the 65%.

Moreover to improve the trace definition on the red side of UVES we have changed the number of coefficients used for the Legendre polynomial fit and for the PCA from 7 (default value) to 5, and decreased the radius of the flux-weighted algorithm used for the re-centering of the initial guess of the trace from 0.25 (default value) to 0.20 (units of order width fraction).

Available options

`\CHK (available from GUI)` - this option shows the median row from the rectified and collapsed orders, the linear fit of the peak positions for the available orders and the final object trace overplotted on the raw scientific frame

`\STD (available from GUI)` - this option changes the default base aperture for the profile definition from the 60% (65% for UVES) to the 85%, this should be activated for standard star which are usually brighter

and whose PSF occupies a larger fraction of the slit

\NONLIN – allows to perform a polynomial fit on the secondary coefficients of the PCA (PCA1) instead of the linear fit which is the default behavior

FWIDTH=fwidth – it is the radius used for the flux-weighted algorithm which re-centers the initial trace, by default it is the 25% (20% for UVES) of the order width. To decrease this value could improve the final result but the procedure could result less reliable

NCOEFF=ncoeff – number of coefficients to use in the Legendre polynomial fit and in the PCA analysis, 7 by default

FCOEFF=fcoeff – number of coefficients to use in the final Legendre polynomial fit, 5 by default

OBJAPER=objaper – this parameter is very important since it defines the base aperture to be used for the object profile definition. This value also sets the fraction of the slit to be used for the sky subtraction (see next subsection), the default value is [0.60,0.60] ([0.65,0.65] for UVES) in units of half slit length (one value for each half of the slit)

MINPCA=minpca – minimum number of good orders required for the PCA analysis, 10 by default (7 in the red)

3.4.8 Sky subtraction

The slit profile calculated at step 3.4.4 is used to obtain a first guess of the sky background required for the optimal extraction, the procedures involved are

- instr_skysub.pro

- instr_fitscatt.pro

- x_echskysub.pro

Functionality

The sky background model obtained with this procedure it is used only as a starting point for the derivation of the object spatial profile as described
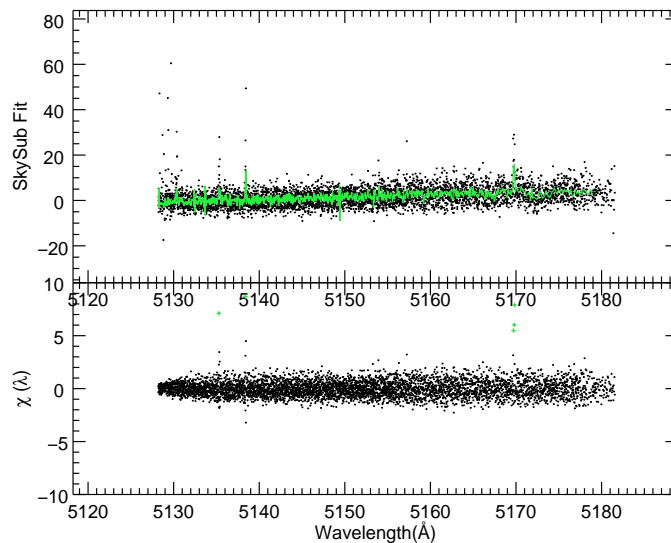
Figure 3.17: In the upper panel is illustrated a simple example of the b-spline fit (green) of sky pixels as a function of 2D wavelength. The lower panel shows the χ^2 of sky pixels while green crosses mark rejected pixels.

in the next subsection. Since it represents only an initial guess the accurate fitting of narrow emission lines is not too much important.

Before to proceed with the fit of the sky, the diffuse scattered light background is subtracted from the science frame following the procedure described in 3.4.4. Then for each order sky pixels are identified as those pixels which have a slit profile value greater than the 30% and which fall outside the boundaries defined by the *aper* tag of the object structure.

The *instr_skysub* procedure makes the implicit assumption that outside those boundaries the object flux is negligible. One should skip this step or change aperture if this assumption is not valid anymore, as for example in the case of standard stars or very bright sources (see 3.4.7).

Once sky pixels have been identified they are divided for the correspondent slit profile value (see 3.4.4) and used to fit with rejection a b-spline of grade 4 as a function of the 2D wavelength (upper panel of Fig. 3.17). The breakpoints for the spline correspond to a spacing of 1.2 time the local wavelength dispersion per pixel.

Then the set of spline parameters obtained by the fit are used to extend the result to the object pixels. At this point the fit of sky+object pixels

is multiplied by the slit profile providing a full characterization of the sky background across the entire echelle order.

The sky background obtained in this way is summed to the scattered light background and finally subtracted to the original science image in order to obtain an object model. This object model is used together with the image variance to obtain χ^2 values for the sky pixels (lower panel of Fig. 3.17). The χ^2 values are used to establish when more breakpoints in the b-spline fit are required (e.g. in correspondence of narrow sky emission lines). If there are regions where the χ^2 values exceed 4 more breakpoints are added in those regions and the b-spline fit is repeated.

The object model created by this procedure is archived in the extension #2 of the final image defined in 3.4.6 and will be used to derive a first guess of the object profile in the next process of the data reduction.

Code generalization

In addition to the modifications of the codes layout in order to account for the technical differences between HIRES and UVES and the different organization of data and setups, we report the correction of a bug in the original code which produces a failure of the process responsible for the b-spline breakpoints addition in correspondence of large χ^2 values.

Available options

\CHK (available from GUI) - this option provides the user with plots of the scattered light subtraction results, sky spectrum fit and residuals for each order and finally a plot of the sky subtracted image for a visual check

\NOMEDSCATT - this option is available for the blue chip only of the HIRES instrument and allows to select between a median image of the scattered light background and the b-spline fit of the interorder gaps (this accounts for wrong scattered light image in those region of the blue chip where usually echelle orders overlap, the effect is less serious for the UVES blue side)

SCATTTRIM=scatttrim - represents the number of pixels used to extend the orders mask for the scattered light subtraction beyond the orders

boundaries, it is 2 by default (1 for the blue side where orders most
likely overlap)

3.4.9   Optimal extraction

This is the last and most important step: sky and scattered light back-
grounds are subtracted from the science frame, then for each echelle order
the object profile is determined and the optimal extraction is performed.
Due to the complexity of this last step the description of the functionality
has been divided in two parts: one concerns the definition of the object pro-
file the other describe the extraction. The procedures involved in this step
are

- instr_extract.pro

- instr_fitscatt.pro

- x_extechopt.pro

- x_extobjopt.pro

- x_extobjbox.pro

Functionality - Object profile definition

The term *extraction* refers to the process which produces the 1D wavelength
calibrated object spectrum from the bi-dimensional science image. There
are many different ways to perform an extraction but the most efficient one
is the *optimal extraction* (Horne, 1986) which relies on the definition of the
object profile (i.e. the PSF of the object). In the following will be described
the basic steps performed by the *instr_extract* procedure in order to show the
approach of this pipeline to the optimal extraction of the object spectrum.
Before to proceed with the calculation of the optimal profile the procedure
loads all the required images and structures: the science image, the master
trace flat, the 2D wavelength solution, the sky subtracted image model from
the previous step, the order structure (see 3.4.2) and the object structure
(see 3.4.7).
Then the scattered light background is subtracted from the science image
and from the master flat, the wavelength solution is converted from air to
vacuum and the heliocentric correction is applied (this is the default format

for the output of the pipeline).

After this preliminary operations the procedure *x_extechopt* is launched. This procedure represents in some way the main core of the pipeline since it iteratively calculates the optimal profile, finds the best fit of the object and sky emissions, uses the fits to produce a model sky and a model variance which will be subsequently used for the optimal extraction.

First the sky subtracted model is extracted with a simple boxcar function whose width in unit of half slit length is defined by the *aper* tag of the object structure (see 3.4.7). The boxcar spectrum is used to calculate the median signal-to-noise ratio (MSNR) of each echelle order, then orders are processed in order of decreasing MSNR. This approach is required since the MSNR of the orders is used to discriminate between different methods for the profile determination

$2.5 < \mathbf{MSNR} < \sqrt{500}$ a full spatial profile is recovered for the object as described below

$\mathbf{MSNR}^2 > 500$ the procedure try to recover profile variation along the spectral direction approximating it as quadratic functions of wavelength, otherwise a uniform profile is fit to all pixels in the order

$\mathbf{MSNR} < 2.5$ the profile is approximated by a simple Gaussian with a fixed width. If at least 2 orders with higher MSNR are available the FWHM to be used is extrapolated from a linear fit of the measured FWHM of the empirical profiles, otherwise a default value of $FWHM = 2.358 * \sigma$ with $\sigma = aper/6$ is used

For the profile calculation first a boxcar extraction is performed on pixels inside the boundaries defined by the *aper* tag of the object structure (cyan vertical lines in Fig. 3.18), the resulting boxcar spectrum is saved in the *box_fx* tag of the object structure. A median smoothed version of the boxcar spectrum is then compared with the original one in order to find cosmic rays and pixels which are more than 5σ away from the median value. Good pixels are then used to fit with rejection a b-spline of order 4 with a spacing equal to 2 spectral pixels, then the resulting set of parameters is used to extend the b-spline fit to the entire echelle footprint using the appropriate wavelength for each pixel.

The boxcar spectrum image constructed in this way is used to normalize the sky subtracted model given in input in order to obtain an initial guess
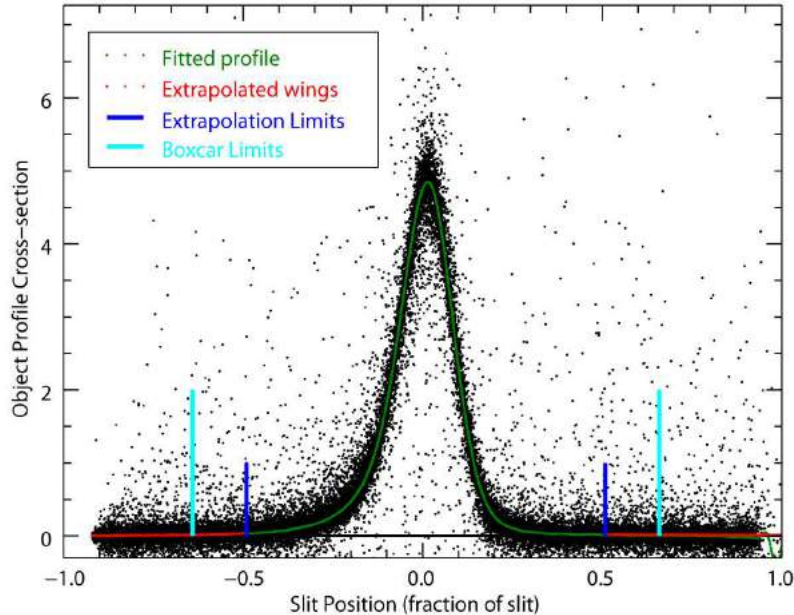
Figure 3.18: This image shows an example of the optimal profile fit in one representative order of UVES for the $2.5 < \mathrm{MSNR} < \sqrt{500}$ case. Black scattered points are the object pixels plotted as a function of their position in the slit expressed in slit fraction. The green line is the fitted profile which will be applied homogeneously to the entire order. Red lines are the exponentially decaying extrapolated wings while the blue and cyan vertical lines represent respectively the extrapolation and boxcar limits. The FWHM is measured directly from the empirical profile.

of the object profile. This kind of approach provides a profile whose spatial cross section is unity by definition as required by the Horne formula, the net effect of this operation is to take to the same level the optimal extracted flux and the boxcar extracted flux.

The resulting image from the ratio of the sky subtracted model to the boxcar image is then fitted with a b-spline as a function of the distance from the object trace as illustrated in Fig. 3.18, the rejection threshold is determined from the MSNR of the order as $(3 + \mathrm{MSNR}/20)\sigma$. This produces a mean profile (green line in Fig. 3.18) which is applied uniformly to the entire extension of the order.

One of the most important features of this code is the extrapolation of the wings of the profile which are usually too faint to be determined empirically.

The code uses the MSNR to specify the distance from the object center at which the extrapolation begins (blue vertical lines in Fig. 3.18), such that for higher MSNR there is less profile to extrapolate.

Extrapolation is performed fitting linearly the logarithm of the interested pixels, then the fit is used to produce exponentially decaying wings which are realistic and reliable (red line in Fig. 3.18). Once available the full spatial profile, the cross section is used to calculate the correspondent FWHM which is saved for lower MSNR order.

Before to proceed the profile is used to correct the object trace for small deviations. The correction of the object trace is performed shifting the mean profile one pixel left or right in order to test the correlation with respect to the central position, this operation is performed for each row and the deviation from the zero offset is fitted along the entire order. Finally the correction is forced to be not greater than one pixel and is added to the object trace, the corrected trace is then saved in the *fin_trc* tag of the object structure and subsequently used to iterate the profile fitting procedure.

Now that a full characterized object profile is available together with the slit profile (i.e. sky profile) the code performs a simultaneous b-spline fitting of the object and sky spectra imposing a distance between breakpoints which corresponds to the native width of a pixel (in km/s). The fitting routine uses profiles to separate the two components and rejects bad pixels with a threshold defined by the MSNR of the order. The fit goes on iteratively until no more points are rejected or the maximum number of iteration (25 by default) is reached.

The χ^2 of the fit is used to create a mask for the cosmic rays rejection ruling out pixels which deviates more than 10σ and their 12 closest neighbors.

The result of the fitting procedure is used to calculate a model variance (using the formula defined in 3.4.6 and adding the scattered light contribution which was previously removed from the fitted image), a model object (upper right panel of Fig. 3.19) measured as *obj_fit* $*$ *obj_profile* and a model sky (lower left panel of Fig. 3.19) measured as *sky_fit* $*$ *sky_profile*.

All the steps described so far are iterated for a default of three times, at each iteration the measurements are repeated using the corrected trace, the cosmic rays mask and the model sky calculated at the previous iteration. During the last iteration an additional b-spline fit of sky and object is performed with the final cosmic rays mask. An example of the result of this
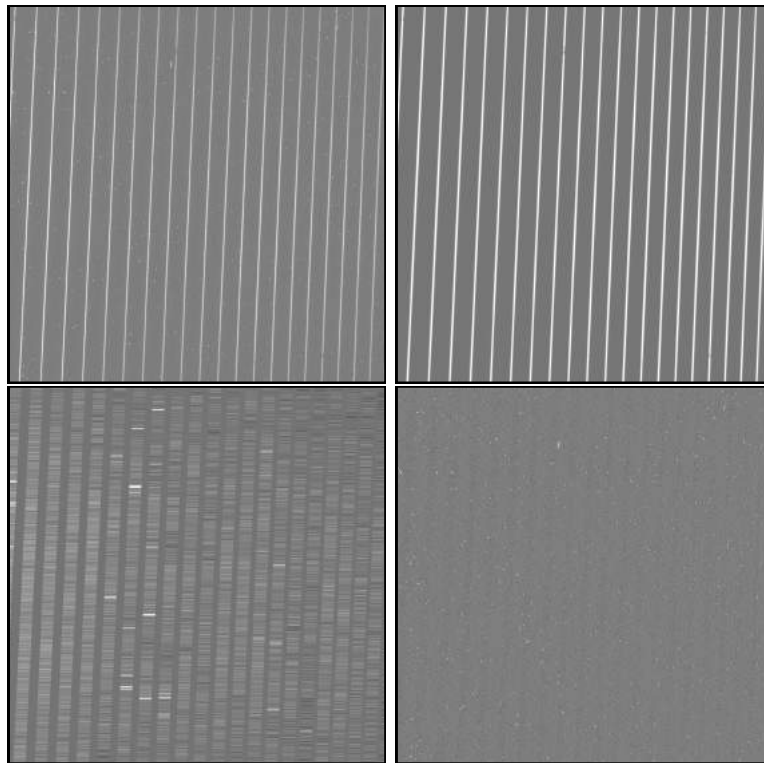
Figure 3.19: Here are represented the final models and results of the optimal profile definition procedure. Upper left panel shows the original science image with pixel-to-pixel correction and scattered light subtraction. Upper right and lower left panels show the model object and the model sky obtained from the iterated simultaneous b-spline fit and the correspondent profiles. The lower right panel shows the residuals image of the CCD pixels with respect to the best fit model. Residuals are close to the Poisson limit for the majority of the CCD pixels.

procedure is illustrated in Fig. 3.19. The residuals image illustrated in the last panel can be plotted at the end of the extraction with the *chk_res* button of the "Check" column of the GUI in order to visually inspect the goodness of the fit, in this case the residuals should be close to the Poisson limits.

Functionality - Extraction

At this point the following ingredients are available for the optimal extraction: a model sky S, a model variance V, a cosmic rays mask M, the optimal profile P and the bias and scattered light subtracted image D. All those ingredients can be used by the *x_extobjopt* procedure to derive the optimal

flux f_i and variance v_i for the $i-th$ pixel of the CCD following the original Horne derivation (Horne, 1986) which is

$$f_i = \frac{M_i P_i (D_i - S_i)/V_i}{M_i P_i^2 / V_i} \tag{3.10}$$

for the optimal flux and

$$v_i = \frac{M_i P_i}{M_i P_i^2 / V_i} \tag{3.11}$$

for the optimal variance. The result of these operations consists of two bi-dimensional images representing respectively the optimal flux and the optimal variance for each CCD pixel in the echelle footprint.

Following the original arguments presented in Horne (1986) each pixel can be viewed as an independent measure of a little portion of the flux from the object, this means that a linear combination of these measures still represents an unbiased measure of the flux. Assuming for each pixel a statistical weight $W_i = 1/V_i$ and adding the summation to equations 3.10 and 3.11 we can write

$$F_z = \frac{\sum_i W_i M_i P_i (D_i - S_i)}{\sum_i W_i M_i P_i^2} \tag{3.12}$$

for the new flux and

$$V_z = \frac{\sum_i M_i P_i}{\sum_i W_i M_i P_i^2} \tag{3.13}$$

for the new variance. The summation is performed separately on the numerator and on the denominator to avoid un-physical values which could arise from the ratio of pixels in regions where the profile is close to zero.

Usually the summation is performed along rows of the CCD, in this case the subscript z represents the row and

$$\sum_i P_i = 1 \tag{3.14}$$

since by definition the total cross section of the optimal profile is one. The wavelength associated with F_z can be constructed starting from the local dispersion of one pixel. This approach is equivalent to collapse echelle orders along the spatial dimension to obtain the mono-dimensional spectrum.

The main problem with this approach is that relies on the wrong assumption that pixels on one row have more or less the same wavelength, as we have seen in 3.4.3 this is not always true for echelle spectra.

In order to avoid this problem the pipeline uses the 2D wavelength solution to
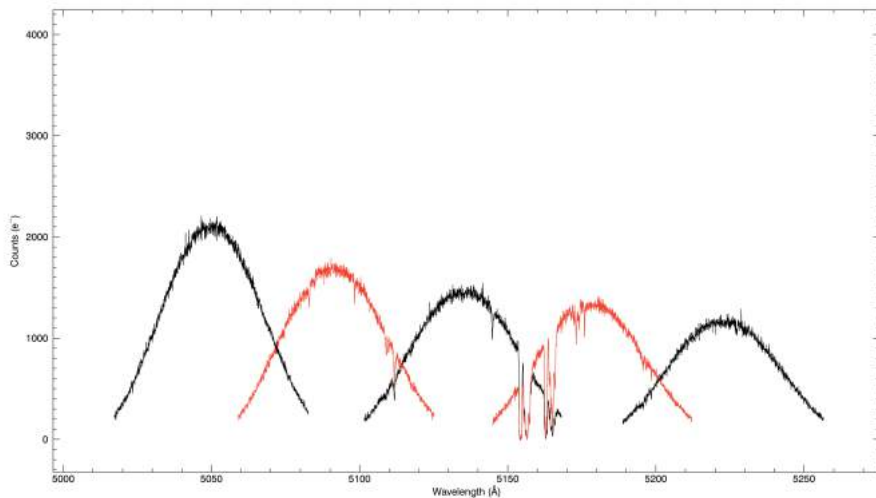
Figure 3.20: This is just an example of the output of the optimal extraction. Here are represented alternately in black and red few extracted orders of UVES setup 564 for an exposure of the QSO HE0940-1050. The typical curvature of the echelle blaze functions is still overimposed on the extracted spectra and should be removed in order to merge correctly separated orders and to obtain a full spectrum.

calculate the fraction of each pixel which contributes to the final wavelength associated with each row, which is equivalent to perform a rebinning rather then a simple summation. The output of this procedure is represented by the optimally extracted 1D spectra of the separated echelle orders as illustrated in Fig. 3.20. Flux, variance and wavelength are saved respectively in the *fx*, *var* and *wave* tags of the object structure.

This rigorous approach is possible only thanks to the 2D wavelength solution retrieved in 3.4.3 which associates a unique wavelength to each individual pixel.

Code generalization

Here will be described not only the modifications required for the generalization but also some extensions to the original code that we implemented for the flux calibration and merging procedures which will be presented in the next chapter.

The first change concerns the different velocity width of native pixels for HIRES and UVES, the former has a pixel dispersion of $1.3km/s$ versus the $1.5km/s$ for the blue side and $1.2km/s$ for the red side of UVES. This pa-

rameter is very important for the extraction since it defines the breakpoints spacing for the simultaneous b-spline fit of sky and object.

Another important thing is the different coordinates of the VLT and Keck telescopes which one have to take in consideration for the heliocentric correction.

In the *x_extechopt* procedure the breakpoints spacing used for the b-spline fit of the profile as a function of the distance from the object center was not small enough to reproduce the exact shape of the profile in correspondence of the peak. The default value has been changed from 0.1 to 0.01 which accurately reproduces the shape of the profile for UVES data.

Moreover a series of important bugs have been corrected: 1) the variance calculated for the boxcar spectrum used in the first part of the object profile definition was defined by default as an INTEGER IDL type, 2) the operation to limit the trace correction between -1 and 1 only worked for the upper limit due to wrong parenthesis in the expression and 3) there was an error in the model variance due to missing subscripts for order identification in the scattered light image.

Regarding the *x_extobjopt* procedure, the model used for the calculation of the new optimal variance has been updated taking into account also the scattered light background (in analogy to the model used by *x_extechopt*).

The original code also offers the possibility to do an optimal extraction of the master trace flat and arc. The first one can be used to remove the blaze functions of the echelle orders, while the arc spectra can be useful to measure the resolution element of the specific setup.

Nevertheless both operations were unsuccessful because the code performed the extraction using the object profile. Since the flat and the arc cross sections are in good approximation rectangular, we have found that a boxcar extraction in this case provides results good enough for our purposes. The boxcar extraction is performed using the *x_extobjbox* procedure modified in order to normalize the extracted flux with the number of pixels summed in one row. Finally two tags have been created in the object structure to save the extracted flat and arc: the *blaze* tag and the *arcs* tag.

The last new feature implemented in the code is the possibility to output for each order the 2D un-rebinned images of the wavelength, profile, optimal flux and optimal variance. Those images will be used at the end of the flux calibration procedure to execute an alternative approach for the final merg-

ing of the available exposures. Alternatively these 2D images can be directly used in the analysis in order to preserve the original identity of CCD pixels.

Available options

`\CHK (available from GUI)` - this enables a plot of the mean fitted profile, a plot of the trace deviation and a plot of the spectrum obtained from the model object. Finally the residuals are plotted

`\EXCHK (available from GUI as chk_extr)` - it shows a comparison of the boxcar extraction with the optimal extraction to asses visually any possible error in the latter. This option is mainly for debug purposes

`\SKIPSKYSUB` - skip the sky subtraction, must be enabled for bright sources as standard stars

`\EXTENBOX` - allows the procedure for the profile definition to go beyond the physical boundaries of echelle orders

`\BOXCAR` - only the boxcar spectrum will be extracted

`\STD (available from GUI)` - equivalent to the previous one

`\BBOX` - enables by default BOXCAR, SKIPSKYSUB, EXTENBOX and NOPROF options

`\SILENT (available from GUI)` - if enabled suppresses real time information during the execution of the processes

`\NOMEDSCATT` - this option is available for the blue chip only of the HIRES instrument and allows to select between a median image of the scattered light background and the b-spline fit of the interorder gaps (this accounts for wrong scattered light image in those region of the blue chip where usually echelle orders overlap, the effect is less serious for the UVES blue side)

`\NOHELIO` - skips the heliocentric correction

`\NOVAC` - skips the air to vacuum convertion

`\EXFLAT (available from GUI) -` fills the *blaze* tag of the object structure with the boxcar spectrum of the master trace flat (for blaze correction). The width of the boxcar is defined by the *aper* tag of the object structure

`\EXARC (available from GUI) -` fills the *arcs* tag of the object structure with the boxcar spectrum of the arc. The width of the boxcar is defined by the *aper* tag of the object structure

`\NO _REBIN (available from GUI) -` produces a FITS table with as many extensions as the optimally extracted orders. Each extension contains the 2D un-rebinned images of wavelength, profile, optimal flux and optimal variance of one order. The FITS is archived in the `\Extract\NoRebin` directory of the main folder

`\ADD _BOX -` performs a boxcar extraction in addition to the optimal one and saves the resulting wavelengths and fluxes in the *box_wv* and *box_fx* tags of the object structure. This option is enabled automatically when the `\EXCHK` option is activated

`\MODELOUT -` adds to the outputs the models calculated by the *x_extechopt* procedure which are saved in a multi-extension FITS and archived in the `\Extract` directory of the main folder

`\NOPROF -` skips the execution of the *x_extechopt* procedure, the calculation of the profile is leaved to the *x_extobjopt* procedure

`INPGAU=inpgau -` this parameter is read only if the `\NOPROF` option is activated and sets the σ of the Gaussian profile to be used for the optimal extraction

`HIGHSNR=highsnr -` defines the threshold used to establish the type of profile definition: if $\mathrm{MSNR}^2 > highsnr$ variations of the profile in the dispersion direction are fitted as a quadratic function of wavelength. The default value for this parameter is 500

`MIN_CUT=min_cut -` it is the maximum distance from the object center allowed for the empirical profile calculation, at greater distances wings are always extrapolated independently from the SNR. The default value for this parameter is equal to the half of the half slit length which is [0.5,0.5]

`FIN_TRC=fin_trc -` this parameter sets the number of iterations to be performed on the profile definition and object-sky fitting procedure. This value is set by default to 2 (which means that the entire process is executed for a total of 3 times)

3.5 Pipelines products comparison

Since the new pipeline is an extension of the HIRES Redux pipeline the quality assessment of the final products has been performed for the UVES mode only. In order to obtain a first impression on the quality of the spectra produced by the new pipeline, the spectrum of the QSO *J2123-0050* has been used as a benchmark for a direct comparison of the final products obtained from the two different pipelines: the ESO-UVES and the IDL pipeline in UVES mode.

Final products of different pipelines come in quite different formats, to proceed with a meaningful comparison all the differences have to be taken into account. In both cases the final result of the optimal extraction procedures are the individual optimally extracted 1D spectra of the echelle orders rectified with the blaze function retrieved from the master flat. Main differences to account for are the pixels counts in ADU for the ESO pipeline and in e^- for the IDL pipeline, the different final binning which is in constant wavelength interval for ESO and per resolution element for the new pipeline and the different technique used for the blaze normalization which requires a little scaling factor for a good matching of the fluxes. So the following operations have been performed on the output of the new pipeline before the comparison:

- to convert from e^- to ADU

- to rebin the spectrum with the same constant wavelength bin used by ESO

- to apply a constant scaling factor for the residual differences ($\approx 10\%$ order of magnitude) in flux due to the different blaze normalization method

these operations, performed on the flux arrays and propagated to the estimated variance, reported the products of the two different extractions at the same flux level within the estimated uncertainties, as expected if a successful

(a) Order 125   (b) Order 134
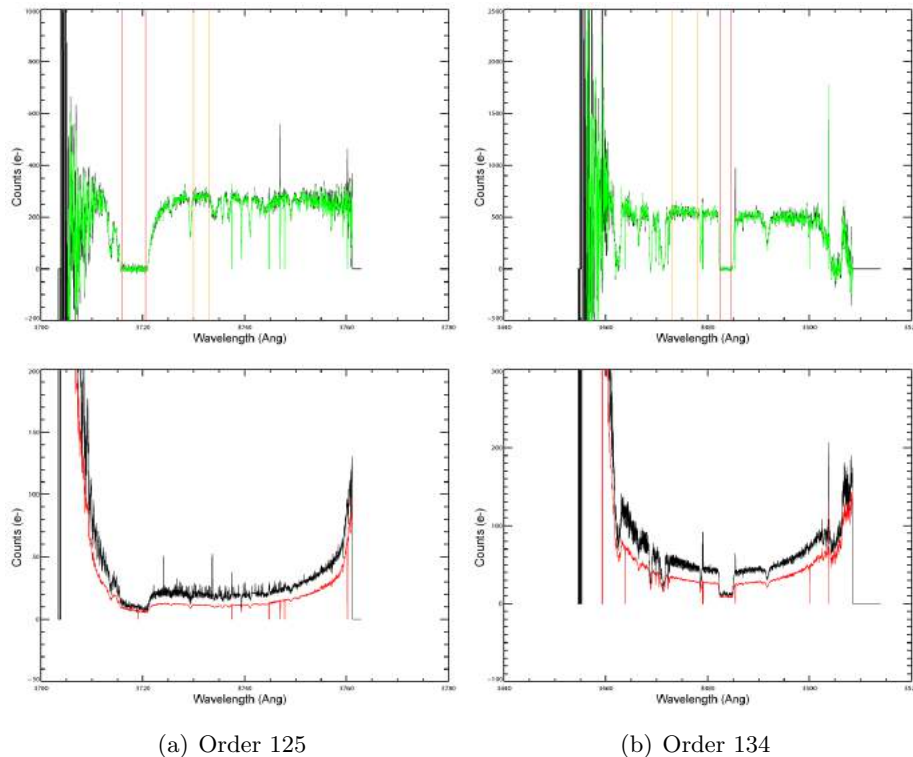
Figure 3.21: Upper panels show ESO (black) and IDL (green) extracted spectra of orders 125 (left) and 134 (right). Differences between the outputs of the two pipelines have been accounted for to take fluxes at the same level. In both orders an improvement in the cosmic rays removal is visible, vertical lines delimit continuum regions (orange) and deeply absorbed regions (red) used for the errors check test. Bottom panels show ESO (black) and IDL (red) propagated errors of orders 125 (left) and 134 (right). IDL errors appears significantly smoother and lower with respect to ESO errors which seem to have a periodic pattern overimposed.

optimal extraction has been accomplished in both cases. As an example in Fig. 3.21 are illustrated two different orders (ESO spectrum in black - IDL spectrum in green), the 125 and 134 with the respective propagated errors (ESO errors in black - IDL errors in red).

After a visual inspection of the results there are some first improvements which can be immediately distinguishable by eyes: the cosmic rays are more efficiently removed, the propagated errors appear smoother and significantly lower with respect to the ESO ones which appears to be characterized by bad superimposed periodic patterns.

Since no "a priori" assumptions can be made on the reliability of errors in

both cases, to better assess the correspondence of errors to the real data the only meaningful way is to compare them with the measured dispersion of the pixels. This operation has been done following a simple but effective method which is the same advised in Carswell and Webb (2013) by the author to do a check for the goodness of errors.

The technique consists in the selection of a portion of extracted spectrum (few Angstroms) free from lines or strange features, and to calculate the standard deviation of pixels. If the propagated errors are good the average error calculated on the same portion of spectrum should be more or less comparable with the standard deviation calculated before. Obviously closer the mean error is to the measured standard deviation more realistic are the errors estimates.

To better understand the errors behavior and its dependencies this test has been carried out separately in continuum regions (where the error is dominated by the Poissonian statistic of the object emission) and in deeply absorbed regions such for example the bottom of a subDLA absoprtion line (where the contribution to the error from other sources, like RON or residual flux from bad sky subtraction, becomes most relevant). Several orders of both spectrum (ESO and IDL) has been probed, as an example here we report the results from this test on the two orders illustrated in Fig. 3.21, regions used for the calculation are highlighted by colored vertical lines as explained in Table 3.1.

Values reported in the table show that ESO pipeline errors are overestimated in both absorbed and continuum regions (with the worst results in the continuum region), on the other hand IDL pipeline errors are not only in very good agreement with the pixels dispersion, but standard deviation of pixels seems generally lower with respect to the ESO ones suggesting an achievement of smaller noise levels (probably due to an improvement in the optimal extraction).

The net effect of an errors overestimation is a lower SNR spectrum with a consequent loss of information during the analysis process (see 4.1.1). The aforementioned measurements are a clear indication of the robustness of the errors delivered by our generalized UVES/HIRES pipeline and therefore of a realistic SNR estimate for the spectra of our sample.

Table 3.1: Comparison between propagated errors and pixels dispersion for two representative orders of UVES. The comparison has been performed with the same spectrum reduced with the ESO pipeline and the IDL pipeline in order to compare the two different products. The same results presented here have been found more or less on all orders and both arms of the spectrograph.

Order 125		
	Standard Deviation	Mean Error
Continuum region (highlighted by the orange bars in the figure)		
ESO	14.80	20.98
IDL	11.15	11.47
Absorbed region (highlighted by the red bars in the figure)		
ESO	8.15	10.58
IDL	6.99	7.09

Order 134		
	Standard Deviation	Mean Error
Continuum region (highlighted by the orange bars in the figure)		
ESO	34.01	51.47
IDL	30.05	30.70
Absorbed region (highlighted by the red bars in the figure)		
ESO	9.44	12.15
IDL	9.37	8.54

Chapter 4

CIV measurements from very high SNR QSO spectra

The high SNR QSO sample presented in Chapter 2 is nowadays the first sample to reach such an high sensitivity and this allows to explore regions of the IGM not yet probed by previous studies.

The results presented here have been obtained with Voigt profile fitting of the identified CIV absorption lines in the spectra of the sample. The outcome of this process is represented by a list of absorbers with their derived physical properties (redshift, column density and Doppler parameter).

Measured quantities have been used to obtain the Column Density Distribution Function and the redshift evolution of the CIV mass density over the $1.2 < z < 3.8$ redshift range. Then the measurements obtained from our sample have been compared with previous measurements in order to study how the improved sensitivity affects the final results.

4.1   Quality check of the final products

It is worth to emphasize that spectra of our sample of Table 2.4 are unconventional spectra in terms of SNR and wavelength coverage. The main characteristics which distinguish this sample from previous ones are the homogeneity and superior SNR. Data have all been reduced with the same pipeline and the same rigorous approach described in the previous two chapters and have all been selected to fulfill the SNR $\gtrsim 100$ condition in the CIV region.

As explained in Chapter 2 the very high SNR has been possible owing to the summation of data from multiple setups and different instruments. In order to check the overall quality of the final products, in the two images of Fig. 4.1 we want to show graphically the improvement in the SNR of the final rebinned spectrum resulting from this summation.

Taking as a representative example the spectrum of QSO B1122-168 we show a little portion of the CIV region in the middle panel of the figures, the number of original CCD pixels summed in each one of the final bin in the top panel and the comparison between the propagated errors and the RMS dispersion of the CCD pixels summed in the final bins.

The upper panel of Fig. 4.1 (b) shows that after the last step about 6000 original pixels (derived from 44 different exposures for a total of 48.16 hours of observation) are combined to provide the value of each individual final pixel. This increases enormously the statistical significance of each bin of the final spectrum.

The lower panels also show how the correspondence between the propagated errors and the RMS of CCD pixels improves as the number of summed pixels increases. The steps in the number of summed pixels (the blue curve of the upper panels) correspond to the region of the spectrum where adjacent orders overlap, their shapes and positions depend from the particular setups that we are co-adding in the final spectrum.

In order to better assess graphically the improvement that can be achieved with these very high quality spectra in terms of CIV detection, in Fig. 4.2 is illustrated the effect of the increasing SNR on a portion of the QSO B1122-168 spectrum which contains two adjacent weak CIV isolated doublets. The first doublet marked by green dashed vertical lines has a column density $\log N = 12.09 \; cm^{-2}$ and a Doppler parameter $b = 10.10 \; km/s$, the second one marked by the red dashed vertical lines has column density $\log N = 11.94 \; cm^{-2}$ and a Doppler parameter $b = 9.46 \; km/s$. Those lines are typical examples of the weak CIV absorptions that we expect to find in correspondence of the optically thin IGM.

Focusing on Fig. 4.2 (a) corresponding to a single exposure only, we see that with a SNR=32 comparable with the value of most commonly used spectra, the two CIV components are almost invisible.

Things are different if we look at Fig. 4.2 (b) which shows the same portion of spectrum as Fig. 4.2 (a), but resulting from the co-addition of all the 44
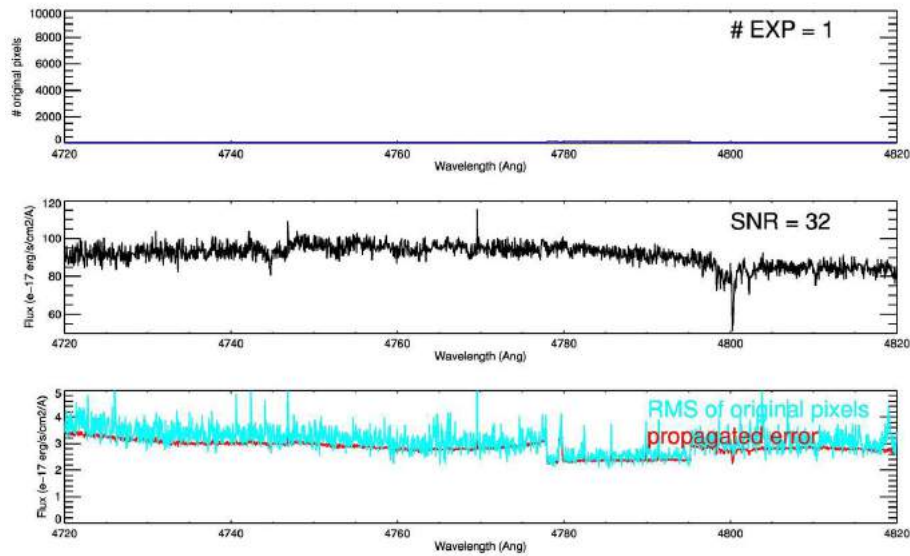
(a) 1 exposure
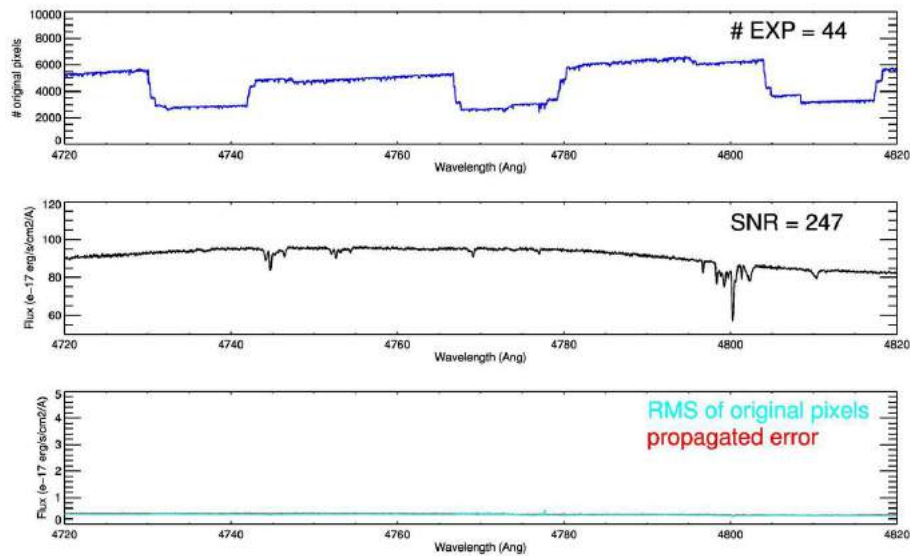
(b) 44 exposures

Figure 4.1: These two plots show the effect of the merging of several exposures on a small portion of the spectrum of QSO B1122-168. The upper panels show the number of CCD pixels summed into each one of the final bins and the number of co-added exposures. The middle panels show the co-added spectrum with the correspondent average SNR per pixel in that region. The bottom panels show the comparison between the propagated errors and the RMS dispersion of CCD pixels summed in the final bins. The upper figure (a) shows the situation for the most common SNR $\lesssim 50$ spectra. The lower figure (b) shows the final result produced by the merging of 44 different exposures.
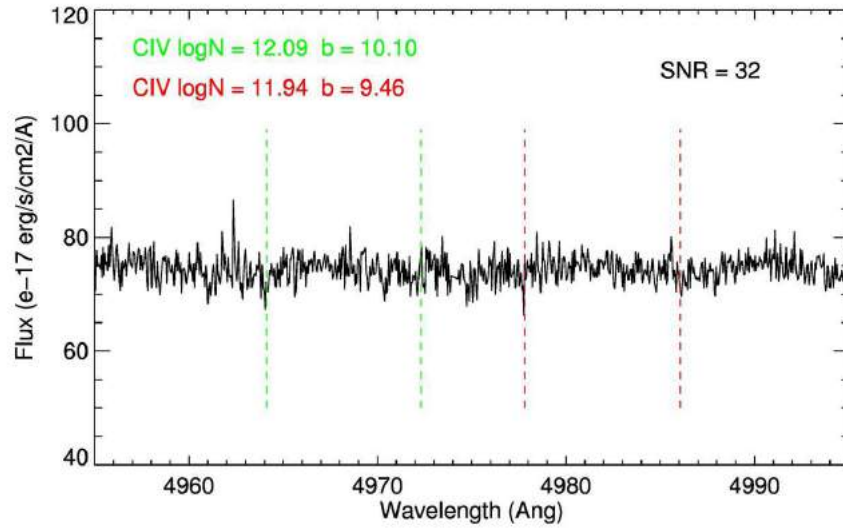
(a) conventional SNR
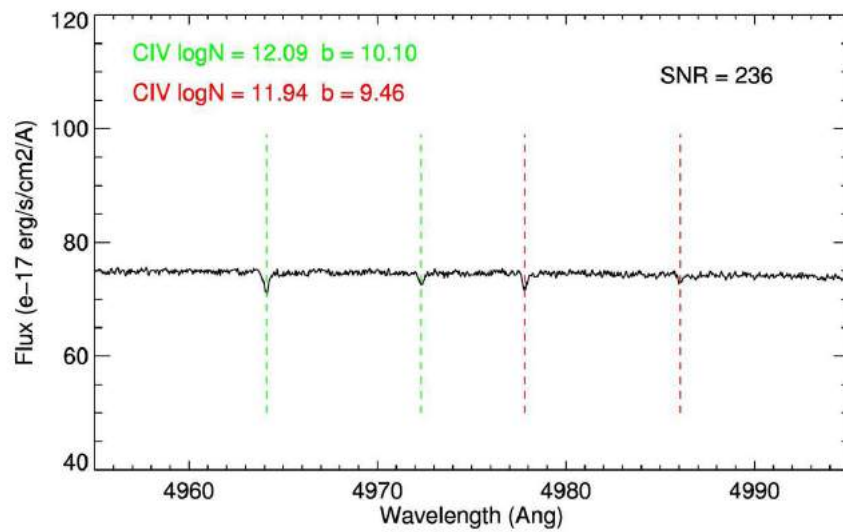
(b) very high SNR

Figure 4.2: Those two figures illustrate how the increasing SNR can improve the detection of low column density CIV absorption lines. Green and red vertical dashed lines indicate the position of two typical low column density CIV doublets whose parameters are reported in the figures. In the most commons SNR $\lesssim 50$ spectra the strongest components are barely visible (on the top figure). Increasing the SNR per pixel up to 236 both components of the two doublets become clearly distinguishable from the surrounding noise (on the bottom figure).

available exposures. This time the average SNR per pixel in this region is so high that both components of the two CIV doublets can be clearly detected and distinguished.

In order to quantify those arguments, the minimum SNR required for a $n\sigma$ detection can be calculated assuming that each wavelength in the spectral region occupied by the absorption feature satisfies the condition

$$I_{0,\lambda} - I_\lambda \geq n\sigma_{0,\lambda} \tag{4.1}$$

where $I_{0,\lambda}$ and $\sigma_{0,\lambda}$ are respectively the flux and the corresponding error of the unabsorbed spectrum at wavelength λ while I_λ represents the flux level which the spectra should have at the same wavelength consequently to the absorption. Dividing by $I_{0,\lambda}$ and integrating over all wavelengths making up the feature the last equation becomes

$$\int_{\lambda_1}^{\lambda_2} \frac{I_{0,\lambda} - I_\lambda}{I_{0,\lambda}} d\lambda \geq n \int_{\lambda_1}^{\lambda_2} \frac{\sigma_\lambda}{I_{0,\lambda}} d\lambda \tag{4.2}$$

where the integrals should be performed over the entire extension of the feature. We notice that the integral on the left side is equal to the observed equivalent width W_{obs} of the absorption line, while the integral on the right side can be approximated by the sum in quadrature of the $\frac{\sigma_i}{I_{0,i}}\Delta\lambda_i$ product over the number N_{pix} of pixels which make up the feature

$$W_{obs} \geq n\sqrt{\sum_{i=1}^{N_{pix}} \frac{\sigma_i^2}{I_{0,i}^2}\Delta\lambda_i^2} \tag{4.3}$$

where σ_i, $I_{0,i}$ and $\Delta\lambda_i$ are respectively the error, the unabsorbed flux and the wavelength interval covered by the i-th pixel.

Now if we assume that the three quantities in the summation are constant along the entire extension of the feature (which is valid for small features in normalized spectra) and define the SNR of the spectrum as I_0/σ, the previous equation becomes

$$W_{obs} \geq \frac{n}{SNR}\sqrt{N_{pix}}\Delta\lambda \tag{4.4}$$

This formula is the one most commonly used in observational astronomy, however in our case we wish to show explicitly the dependance from the width of the feature.

For this purpose we recall from section 1.3.3 that in the optically thin regime

(linear part of the curve of growth) the profile of the absorption feature can be approximated analytically by a Gaussian distribution with $\sigma = b/\sqrt{2}$ where b is the Doppler parameter of the line. Now considering that a distance of 3σ from the center of the distribution corresponds to the 99.7% of the total area subtended by the Gaussian distribution, we can write the following relation

$$N_{pix} = \frac{6b/\sqrt{2}}{\Delta V} \qquad (4.5)$$

where ΔV is the width of a pixel expressed in km/s (which in our case has a fixed value of $3\ km/s$). Using equations 4.4 and 4.5 it is possible to obtain the formula

$$SNR_{min} = \frac{n}{W_{obs}}\sqrt{\frac{\sqrt{18}b\Delta\lambda^2}{\Delta V}} \qquad (4.6)$$

which allows to calculate the minimum SNR required for the detection of an absorption feature within a $n\sigma$ confidence interval.

This formula can be linked to the the column density N of the absorption features recalling that in the linear part of the curve of growth the following relation is valid for the equivalent width (multiplied by $1+z$ since our spectra are not in rest frame coordinates)

$$W_{obs} = A\lambda^2_{trans}fN(1+z) \qquad (4.7)$$

where λ_{trans} is the wavelength of the considered transition, which in our case can be one of the two transitions (1548.20Å or 1550.78Å) of the CIV doublet, f is the oscillator strength of the transition and z is the redshift of the line to detect. The constant factor A incorporates different physical constants and can be defined as

$$A = \frac{\pi q_e^2}{4\pi\epsilon_0 m_e c^2} \qquad (4.8)$$

In this case one has to pay attention to units of measure, in our calculation we used

- $c = 299792458\ m/s$ –> speed of light

- $q_e = 1.602176565 \times 10^{-19}\ Coulomb$ –> electron charge

- $m_e = 9.10938291 \times 10^{-31}\ kg$ –> electron mass

- $\epsilon_0 = 8.854187817 \times 10^{-12}\ C^2N^{-1}m^{-2}$ –> void dielectric constant

and we have converted all wavelengths, Doppler parameters and column densities to meters.

Using the aforementioned formulas it is possible to calculate the minimum SNR required for a 3σ detection of an absorption line with a certain column density and Doppler parameter. Using for example the parameters of the two CIV doublets in figure 4.2, we can calculate that a minimum SNR ~ 35 is required to detect the CIV 1548.20 component of the green doublet while a SNR ~ 48 is required for the strongest component of the red one.

Focusing on the weaker CIV 1550.78 component of the doublets, the requirements in terms of SNR become even more stringent. For the weak component of the green doublet a minimum SNR ~ 71 is required while SNR ~ 97 for the red one.

Now we have demonstrated that with a SNR ~ 236 we are well above the detection thresholds of the CIV features illustrated in Fig. 4.2.

4.1.1 Propagated errors vs Pixels dispersion

As explained in the previous chapters all data reduction processes, flux calibration and merging processes have been specifically designed with particular attention to the propagation of errors. Each source of uncertainty has to be appropriately accounted for and the only way to understand if the final errors reflect the real behavior of data one should compare the propagated errors with the *Root Mean Square* or RMS deviation of correspondent data. This is very important since all the Voigt profile fitting procedures rely on the value of the χ^2 defined as

$$\chi^2 = \sum_{i=1}^{\nu} \frac{(x_i - \mu_i)^2}{\sigma_i^2} \tag{4.9}$$

normalized with the available degrees of freedom ν.

Here x_i and μ_i represent the pixel value and the fitted value respectively, divided by the correspondent error σ_i. This formula gives the condition for the best fit model which is the one that agrees with the experimental points within the confidence interval defined by their RMS deviation. This condition translates into $\chi^2/\nu \approx 1$.

Nevertheless the test could fail if the errors do not reflect the real dispersion of data points producing values higher than one if the errors are underestimated and viceversa lower than one if errors are overestimated.

(a) HE0940-1050   (b) B1618-0035
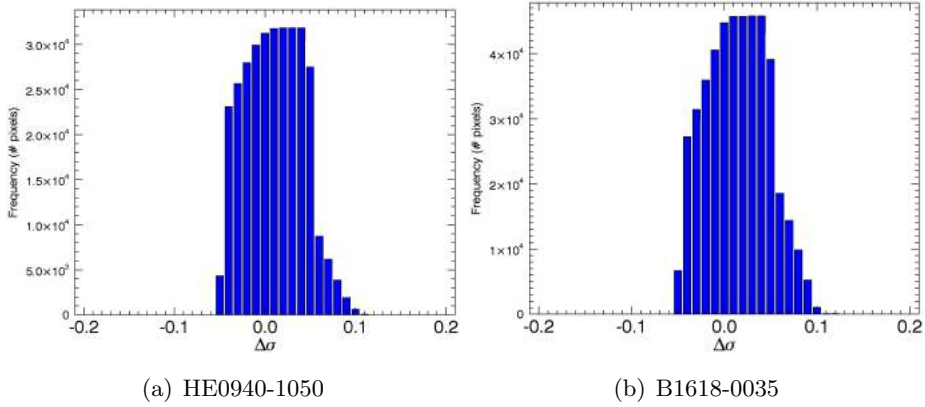
Figure 4.3: Difference between mean propagated error in a 100Å portion of spectrum free from absorption lines and the RMS dispersion of fluxes in the same correspondent region of the spectrum. Only results from spectra $HE0940-1050$ (left) and $B1618-0035$ (right) are presented as an example of the overall errors behavior.

In practice it is very difficult to obtain propagated errors which agree perfectly with the RMS dispersion, in order to account for small deviation from a realistic behavior here we present an example of the kind of analysis performed on errors of the final QSO sample spectra.

The first test concerns the direct comparison between the mean error $\overline{\sigma_i}$ and the RMS dispersion of final bin, both calculated over different portions of the normalized spectrum. Those regions have been selected to be 100Å wide and free from absorption lines.

The two images of Fig. 4.3 show an example of the distribution of the values

$$\overline{\Delta\sigma_i} = \text{RMS} - \overline{\sigma_i} \tag{4.10}$$

for the two spectra of QSO HE0940-1050 and QSO B1618-0035. As can be seen from the figures the center of the two distributions is only slightly shifted from zero indicating a small underestimation of ~ 0.015 corresponding more or less to the 5% of the error values in that region.

It is worth to mention that the final product contains all the systematics uncertainties which could arise from the reduction and merging processes. Those systematics cannot be always accounted for by the errors propagation but can only be minimized in order to not create a discrepancy between data and propagated errors. In this sense the good agreement with the RMS showed in the two distributions of Fig. 4.3 is a clear indication that there are no systematics which are not under control.
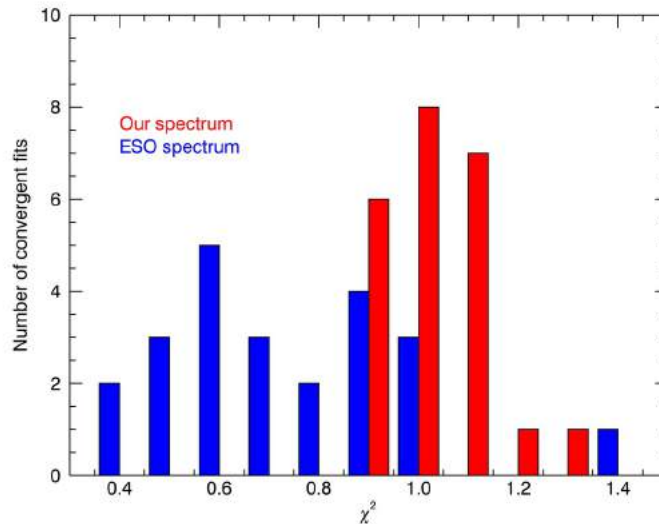
Figure 4.4: Comparison of χ^2 values of the line fitting performed on the whole list of detetcted CIV from the same spectrum (HE0940-1050) but reduced with two different pipelines: the ESO pipeline (blue data) and our dedicated experimental pipeline (red data). This plot allows to test indirectly the correspondence between errors and RMS of data: for ESO data the mean χ^2 is well below one indicating an overestimation of the final propagated errors, on the other hand our data appear to be characterized by more reliable errors as indicated by the mean χ^2 around one.

This result allows us to use the available final RMS array obtained from the merging procedures as an alternative to the propagated sigma in those cases where the resulting χ^2 from the best Voigt profile fit is not satisfactory.

The last check that we have done for the reliability of the errors estimate concerns the comparison of the errors produced by our experimental pipeline with the errors produced by the conventional approach of the ESO-UVES pipeline. As illustrated in Fig. 4.4 for this comparison we used the χ^2 values obtained from the Voigt profile fitting performed on two different version of the same spectrum of the QSO HE0940-1050: one reduced with our pipeline (in red) and the other one reduced with the ESO-UVES pipeline (in blue). Values reported in the plot correspond to the χ^2 obtained from the analysis of the whole list of CIV detected in the spectrum.

This figure clearly shows that the mean value of the blue distribution is well below one indicating a significant overestimation of the errors, at variance with the mean value of the red distribution which is pretty close to one

indicating a good correspondence of the errors with the RMS dispersion of data.

4.2 Data analysis

A preliminary step for the analysis of the spectra is the normalization to unit continuum. Continuum fitting has been performed with the *x_continuum* procedure incorporated in the auxiliary tools incorporated in the pipeline described in the previous chapter. Using an interactive interface, little chunks of continuum have been selected in the spectra and fitted with high order b-spline with a 2.5σ low and a 3.0σ high rejection thresholds.

Then normalized high SNR spectra of the sample have been used for the identification and Voigt profile fitting (1.3.3) of the CIV $\lambda\lambda 1548, 1550$Å doublet absorption lines. This operation has been accomplished using the VP-FIT software (Carswell and Webb, 2014) for the automatic Voigt profile fitting.

At present this software is distributed together with another software called RDGEN which is used for the visualization of the spectra and the first identification of absorption lines.

4.2.1 CIV lines identification with RDGEN

The basic behavior of this software is simple in principle: it offers a graphical user interface to plot regions of the spectrum that correspond to specified transitions for a given redshift, as showed in Fig. 4.5.

The user can choose which transitions of which ion to show on the main interface and scan the entire spectrum increasing the reference redshift in order to search for absorption lines. In our case we choose to plot CIV doublet transitions together with SiIV doublet, $Ly\alpha$ and $Ly\beta$ to better identify absorption systems.

Once a feature has been selected the RDGEN software allows to create the input file for the VPFIT program. This input file contains a first guess of redshift, Doppler parameter and column density which are used by VPFIT as a starting point for the automatic Voigt profile fitting process.

The basic detection criteria applied are described in the following. First we search for the presence of both components of the doublet: if the CIV 1548 and the CIV 1550 are both visible and not blended with other features we
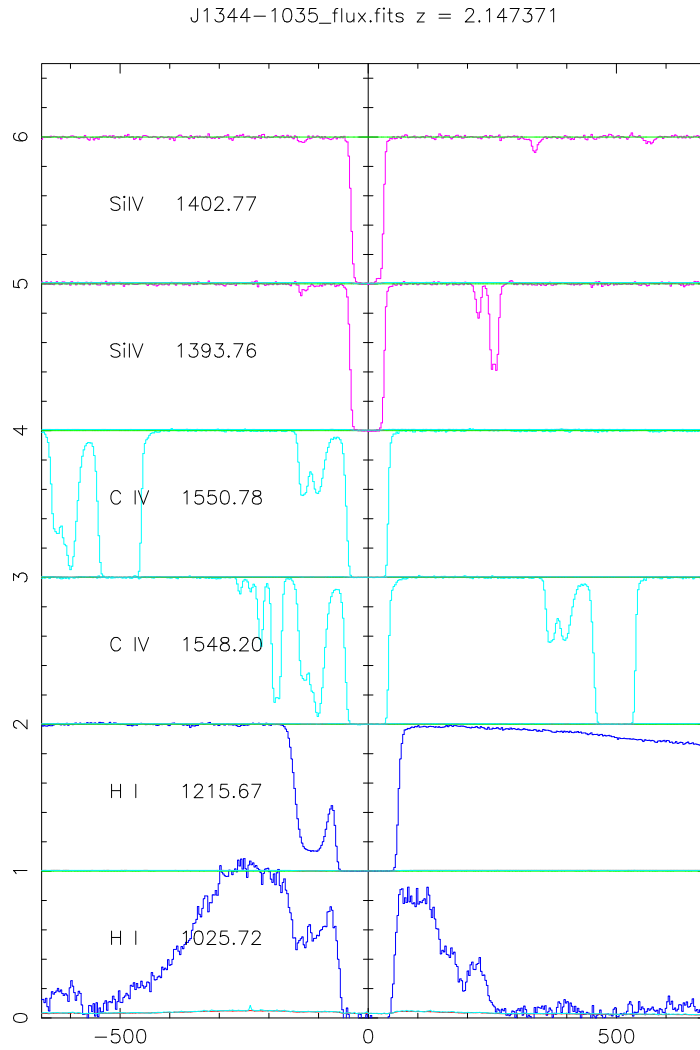
Figure 4.5: Typical interface of the RDGEN software: the user can select which portion of the spectrum to plot and which ions to search for. In this example are plotted the Lyman α, Lyman β and the two components of the CIV and SiIV doublets.

can be pretty sure that we have found a CIV absorption feature.

Since the CIV 1550 is weaker with respect to the CIV 1548 sometimes it could happen that we can only see the strongest component. Another criteria in those cases is to search, at more or less the same redshift of that feature, for an associated $Ly\alpha$ absorption or other associated ions (mainly SiIV, NV and OVI for the high ionization states and FeII, MgII, MnII, AlII or AlIII for the low ionization states).

In those cases the RDGEN software it is very useful since it allows to plot in the same velocity range the regions of the spectrum correspondent to different ions (as illustrated in Fig. 4.5).

Finally for the statistical analysis we included in our sample of absorption lines only CIV doublets whose stronger members are detected at a confidence level greater than 3σ i.e., only CIV 1548 lines with an equivalent width W larger than the detection limits defined by the formula

$$W_{obs} = \frac{n}{SNR}\sqrt{\frac{\sqrt{18}b\Delta\lambda^2}{\Delta V}} \tag{4.11}$$

where n is equal to 3, SNR is the signal to noise ratio of the spectrum in correspondence of the considered feature and the meaning of the other parameters is the same defined in 4.1.

Furthermore, as CIV absorptions is relatively common in QSO spectra, blending of systems at similar redshifts could be a problem. Therefore, when either the 1548 or the 1550 CIV component is blended with other transitions, we include the doublet in the complete sample only if the doublet ratio, defined as the ratio of equivalent widths of the stronger to the weaker component, is in the range $0.8 - 2.2$. The theoretical ratio for this doublet is 2, so the above mentioned interval is adopted to account for resolution effects and is determined from CIV narrow absorption doublets clearly not affected by blending.

4.2.2 VPFIT for Voigt profile fitting

Once the first guess input file has been produced with RDGEN it can be passed to VPFIT in order to start the Voigt profile fitting. The program tries to adjust automatically redshift, Doppler parameter and column density of the lines in order to minimize the χ^2 obtained from the fit. This
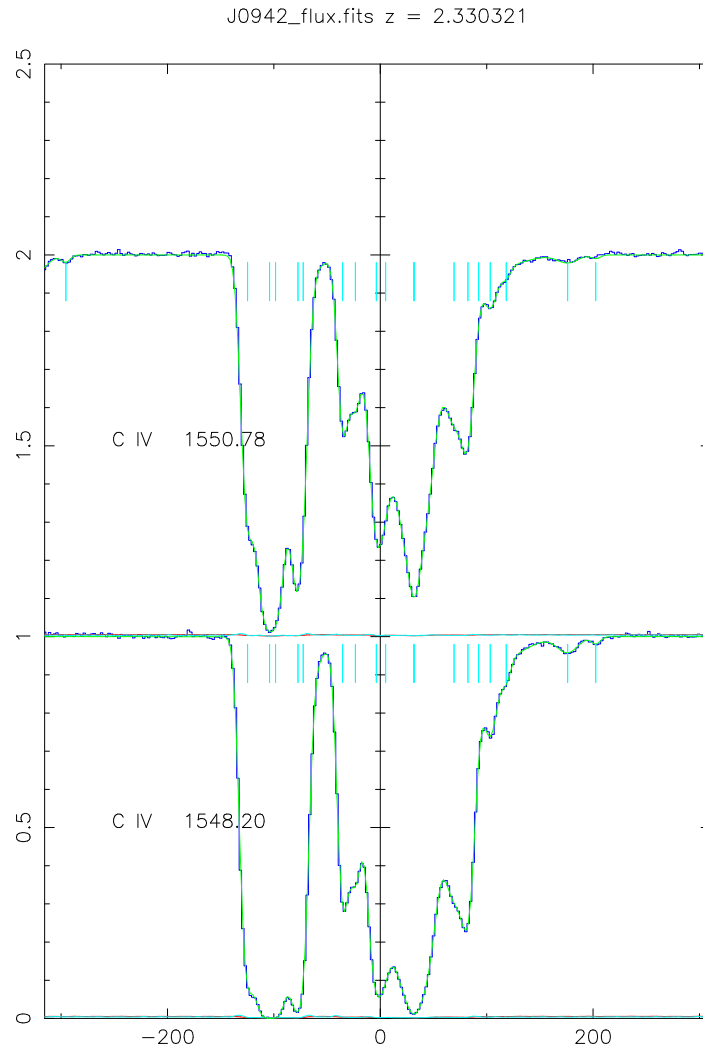
Figure 4.6: An example of the Voigt profile fit obtained with the VPFIT software on a complex CIV absorption system at redshift $z = 2.330$ in the spectrum of the QSO HE0940-1050. This image shows both transitions of the CIV, the green line represents the fitted profile while cyan ticks mark the positions of individual components of the system.

procedure is repeated iteratively until the decrease in χ^2 at each iteration is less then the default value of $\Delta\chi^2/\chi^2 < 0.001$.

VPFIT can be run in two different modes: the first one tries to adjust parameters without changing the total number of lines in the system and is useful when you want to test a certain configuration which you consider good; the second one allows to add or remove lines in order to minimize χ^2 in region where residuals are too large.

The procedure of lines removal is based on minimum threshold for the parameters which can be set appropriately by the user at the beginning of the iteration. This mode is useful when you are not sure about the system configuration and you want the program to find the best configuration in your place.

A common practice when VPFIT is executed in the "add/remove lines" mode is to use as input file the output of the previous run properly modified to reduce residuals in the worst regions. An example of the Voigt profile fit which can be achieved with this software is illustrated in Fig. 4.6 which shows the best fit solution (overplotted green continuous line) for a complex system of CIV absorption lines (marked by the cyan vertical ticks) in the spectrum of the QSO HE0940-1050.

Another important feature of VPFIT is the possibility to give in input an RMS array for the analyzed spectrum as an alternative to the propagated sigma array. In this case the RMS array produced by our merging procedures represents a precious resource to improve our line fitting process.

The last important thing to consider during the fitting process concerns the definition of the FWHM of the resulting Instrumental Spread Function (ISF) (2.1.8) of the final spectrum expressed in km/s. This parameter is important when you have a complex absorption system and the program has to decompose the system in single components which in most cases are blended.

In order to achieve a good definition of the FWHM of the ISF, it is available from the secondary products of our pipeline the extracted spectrum of the arc lamp used for the wavelength calibration. Those spectra are characterized by very narrow emission line which are very good for a determination of the FWHM of the ISF.

4.3 A case study: the CDDF and Ω_{CIV}

As a case study applied to our high SNR spectra we will present the first results from the calculation of the Column Density Distribution Function (already defined in 1.4) and of the redshift evolution of the total mass density of the CIV.

Those two statistical estimators allows to investigate basic properties of the IGM, like its metal content at different densities regimes and different cosmic epochs. With those measurements one can link observational data to important quantities such as the cosmological parameters; moreover a direct comparison with simulations allows to put constraints on feedback models in order to improve our understanding on the enrichment processes of the IGM.

Previous studies (Songaila, 2001; Boksenberg et al., 2003; D'Odorico et al., 2010; Cooksey et al., 2010; Simcoe et al., 2011; D'Odorico et al., 2013; Burchett et al., 2015) have successfully accomplished statistical measurements of the CIV Column Density Distribution Function and of the CIV mass density Ω_{CIV} over the range $0 < z < 6$ using a large sample of quasars (33 QSO for the largest sample).

Nevertheless those measurements have been limited by the SNR of the available QSO which rarely reach values larger than ~ 100 per pixel and in most cases does not allow to explore significantly the $\log N_{CIV} < 12 \; cm^{-2}$ column densities region.

Those very low column densities are crucial to probe the gas near the mean density which occupies most of the volume of the Universe and represents about the 80% of the total gas mass.

The only exception was a measurement performed more than a decade ago by Ellison et al. (2000). This measurement was based on the very high SNR spectrum of the gravitationally lensed quasar B1422+231. Today following the same steps we try to improve those results with a sample of 12 very high SNR QSO spectra.

In order to check the reliability of our data and to see the improvements with respect to previous ones, final results obtained from this work will be compared with the state of the art current data which will be presented in 4.4.

Before to proceed in the next two subsection we will briefly introduce the statistical measurements that represent the subject of this work. All the re-

sults that will be presented are based on a standard cosmological model with parameters $\Omega_m = 0.26$, $\Omega_\Lambda = 0.74$ and $h \equiv H_0/(100\ km\ s^{-1} Mpc^{-1}) = 0.72$ to be consistent with previous measurements.

4.3.1   Practical definitions - the CDDF

In practice the Column Density Distribution Function or CDDF can be discretized by simply counting, for a specified column density bin ΔN, the number of detected absorption lines n normalized with the *redshift path* covered by the available QSO sample ΔX and the column density bin as expressed in this formula

$$f(N) = \frac{n}{\Delta N \Delta X} \tag{4.12}$$

where the redshift path is defined as

$$dX = \frac{H_0}{c}(1+z)^3 dl = \frac{(1+z)^2}{E(z)} dz \tag{4.13}$$

with

$$E(z) = [\Omega_{\Lambda,0} + (1-\Omega_0)(1+z)^2 + \Omega_{m,0}(1+z)^3 + \Omega_{r,0}(1+z)^4]^{1/2} \tag{4.14}$$

defined in 1.1, and so depends from the considered cosmological model.
The error on individual points calculated with Eq. 4.12 is simply the statistical error on the number of occurrences, obtained from the square root of the number of lines in each bin, propagated throughout the formula giving

$$\sigma_f(N) = \frac{f(N)}{\sqrt{n}} \tag{4.15}$$

4.3.2   Practical definitions - the Ω_{CIV}

The mass density of the CIV can be expressed as a fraction of the critical density today with the formula

$$\Omega_{CIV} = \frac{H_0 m_{CIV}}{c \rho_{crit}} \int N f(N) dN \tag{4.16}$$

where $H_0 = 100h\ km\ s^{-1} Mpc^{-1}$ is the Hubble constant, $m_{CIV} \simeq 2 \times 10^{-23}\ g$ is the mass of a C_{IV} ion, c is the speed of light, $\rho_{crit} = 1.88 \times 10^{-29}\ h^2 g\ cm^{-3}$ is the critical density and $f(N)$ is the Column Density Distribution Function. Since the lack of statistic and incompleteness make it difficult to define

accurately the CDDF for all the column densities and for a sufficiently large redshift path, we can approximate the value of the integral in Eq. 4.16 for different redshift bins by approximating it with a sum

$$\Omega_{CIV} = \frac{H_0 m_{CIV}}{c\rho_{crit}} \frac{\sum_i N_i}{\Delta X} \qquad (4.17)$$

with a correspondent variance defined as

$$\left(\frac{\delta\Omega_{CIV}}{\Omega_{CIV}}\right)^2 = \frac{\sum_i [N_i]^2}{[\sum_i N_i]^2} \qquad (4.18)$$

the sum is over the n CIV detected lines.

4.4 State of the art/current data

The $\log N_{CIV} < 12$ regime was studied for the first time by Sara Ellison (Ellison et al., 2000) who used a very high SNR spectrum for the CIV analysis. She used the spectrum of the gravitationally lensed quasar Q1422+231 at redshift $z = 3.625$ reaching a SNR between 200 and 300 in the region redward the QSO's Lyman alpha emission. The Column Density Distribution Function obtained from this author is represented in Fig. 4.7. This image also shows the best fit linear model of the form $f(N) = BN^{-\alpha}$ where $\alpha = 1.44 \pm 0.05$ represents the slope and is derived from the best fit.

Even if this work was precursory for that time, the main limitation was the number of line of sight limited to one. Similarly to what observed for the Deep Spectrum only one line of sight is not enough to sample appropriately the sky and reduce the cosmic variance in order to obtain robust conclusions. This limitation is strongly dependent from the difficulties related to the process required to obtain a very high SNR spectrum discussed in the previous chapters.

A big improvement in the measurement of the CIV Column Density Distribution Function and of the evolution of the mass density Ω_{CIV} over the redshift range $1.5 < z < 4.0$ comes from D'Odorico et al. (2010). Prior to that work best available data in this redshift range were from Songaila (2001); Boksenberg et al. (2003).

One of the biggest improvements of D'Odorico et al. (2010) with respect to previous similar works was the number of QSOs in the sample. The sample used consists of 33 QSOs from redshift $z \sim 2.1$ to $z \sim 4.5$, the sample also
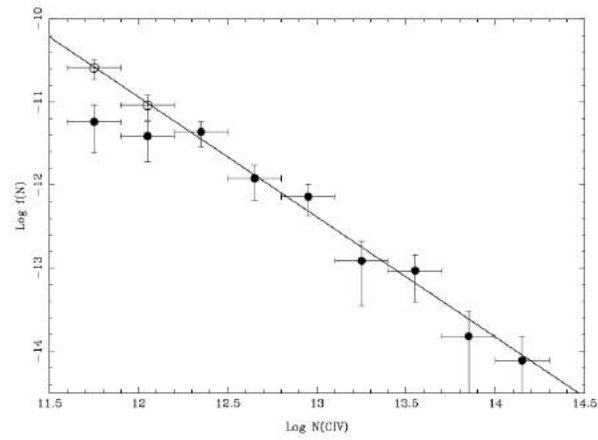
Figure 4.7: Column Density Distribution Function from the high SNR spectrum Q1422+231 (Ellison et al., 2000). Empty circles represent points corrected for incompleteness while the dashed lines is the best fit model of the form $f(N) = BN^{-\alpha}$ where $\alpha = 1.44 \pm 0.05$ (Ellison et al., 2000).
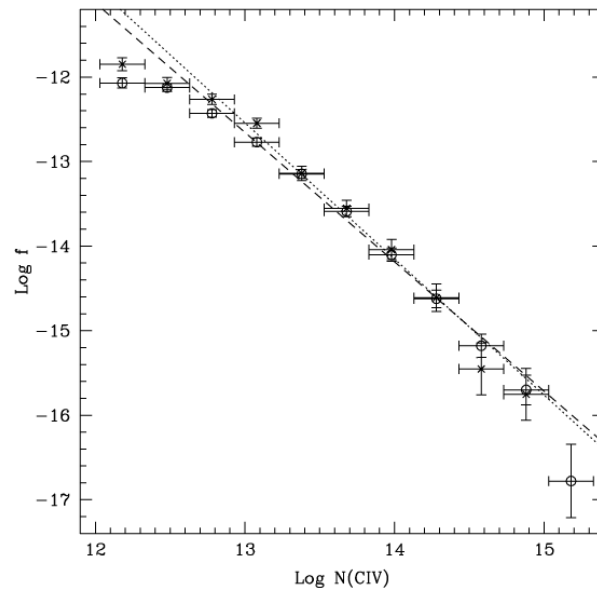
Figure 4.8: Column Density Distribution Function derived from 25 QSO (empty circles) of the sample from D'Odorico et al. (2010), and 8 QSO (crosses) of the sample from Boksenberg et al. (2003). The dotted line represents best fit model ($f(N) = BN^{-\alpha}$) for crosses ($\alpha = 1.61 \pm 0.07$) while the dashed line is the best fit model for circles ($\alpha = 1.53 \pm 0.04$) (D'Odorico et al., 2010).
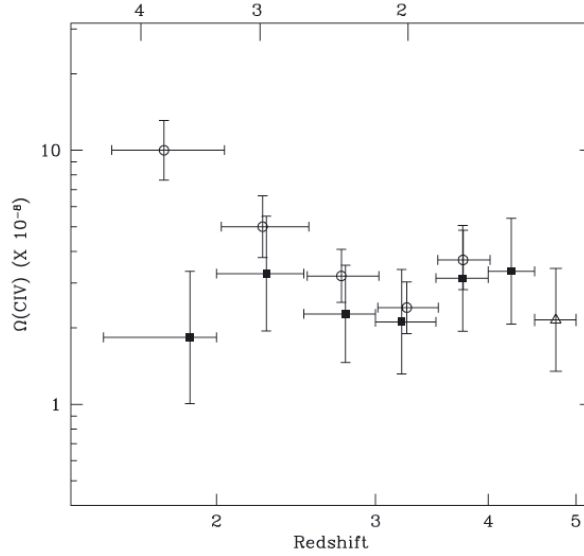
Figure 4.9: Comparison of the $\Omega_{CIV}(z)$ derived from the full sample of D'Odorico et al. (2010) (open dots) with the data from Songaila (2001) (solid squares) and Pettini et al. (2003) (empty triangle). The column densities range covered by these data is $12.0 < \log N_{CIV} < 15.0$ (D'Odorico et al., 2010).

includes 8 QSO from the previous Boksenberg et al. (2003) and is the largest one ever used for this kind of study improving much more the statistical significance of the measurements.

The CIV Column Density Distribution Function derived is illustrated in Fig. 4.8. This image shows the comparison between the CDDF derived for the 25 QSO of the D'Odorico sample and the 8 QSO from the sub-sample from Boksenberg. The slope derived from the best fit model of the form $f(N) = BN^{-\alpha}$ is $\alpha = 1.53 \pm 0.04$ obtained from the D'Odorico sample corrected for incompleteness at the low column densities.

Even with the completeness correction appears evident that the lowest two point tends to flatten deviating from the linear model. On the other hand Ellison's CDDF and the one from the Deep Spectrum tell us two things: 1) improving the SNR of spectra the CDDF can be extended down to lower column density bins 2) the flattening in the CDDF at $\log N_{CIV} \sim 12$ is only apparent and depends from the ability to detect such very faint absorption lines in the QSO spectra.

As will be discussed in section 4.6.2 our measurements represent the next step in the study of the CIV distribution ensuring both very high SNR and

(a) $13.3 < \log N_{CIV} < 15.0$ (b) $13.4 < \log N_{CIV} < 15.0$
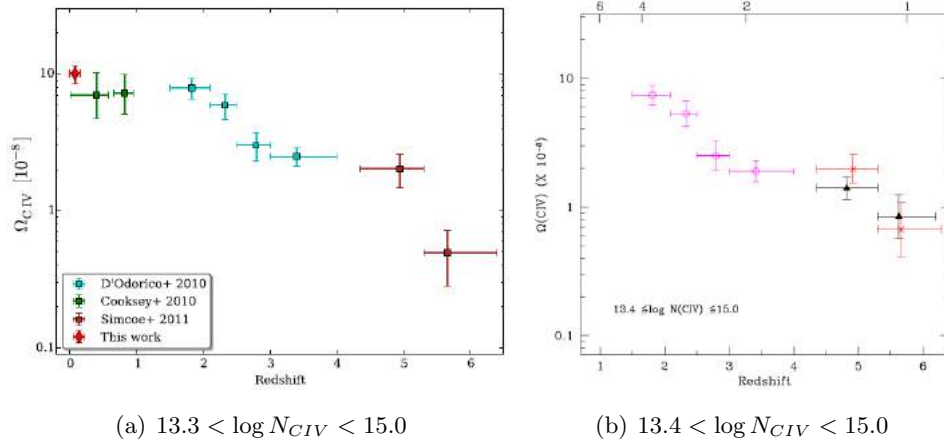
Figure 4.10: (a) Evolution of Ω_{CIV} in the full available redshift range from $z \sim 0$ to $z \sim 6$. The meaning of different colors is explained in the legend, the column density range is $13.3 < \log N_{CIV} < 15.0$ (Burchett et al., 2015). (b) Evolution of the Ω_{CIV} from $z \sim 2$ to $z \sim 6$. Data at high redshift from D'Odorico et al. (2013) (black triangles) are compared with data from Simcoe et al. (2011) (red crosses) and D'Odorico et al. (2010) (magenta open dots) at lower redshifts. The column densities range is $13.4 < \log N_{CIV} < 15.0$ (D'Odorico et al., 2013).

statistical significance provided by a number of QSOs large enough to sufficiently reduce the cosmic variance.

Thanks to the number of QSO in our sample study of the evolution of the mass density of CIV has also been possible. Indeed QSO of the sample has been selected to cover homogeneously the redshift range $1.0 < z < 4.0$. In this range state of the art data are illustrated in Fig. 4.9 from D'Odorico et al. (2010).

The evolution of the Ω_{CIV} represented in this figure shows an increasing trend down to $z \sim 1.5$ in a column density range $12.0 < \log N_{CIV} < 15.0$. Even if we expect that the major contribution to the total Ω_{CIV} should come from the higher column density lines, if one wish to appropriately trace the IGM, lower column densities threshold are required to probe the $\log N_{CIV} < 12$ regions. As will be discussed in section 4.6.3 our sample not only allows us to lowering the $\log N_{CIV}$ threshold but also to extend the present redshift coverage adding a new point in the $1.0 < z < 1.5$ range.

In order to set our results in the frame of the evolution of the Ω_{CIV} in the whole redshift range $0 < z < 6$, we will refer to data illustrated in Fig. 4.10 (a) from Burchett et al. (2015), which covers the Ω_{CIV} down to $z \sim 0$ in

the column densities range $13.3 < \log N_{CIV} < 15.0$, and in Fig. 4.10 (b) from D'Odorico et al. (2013), which extend the coverage up to $z \sim 6$ in the column densities range $13.4 < \log N_{CIV} < 15.0$.
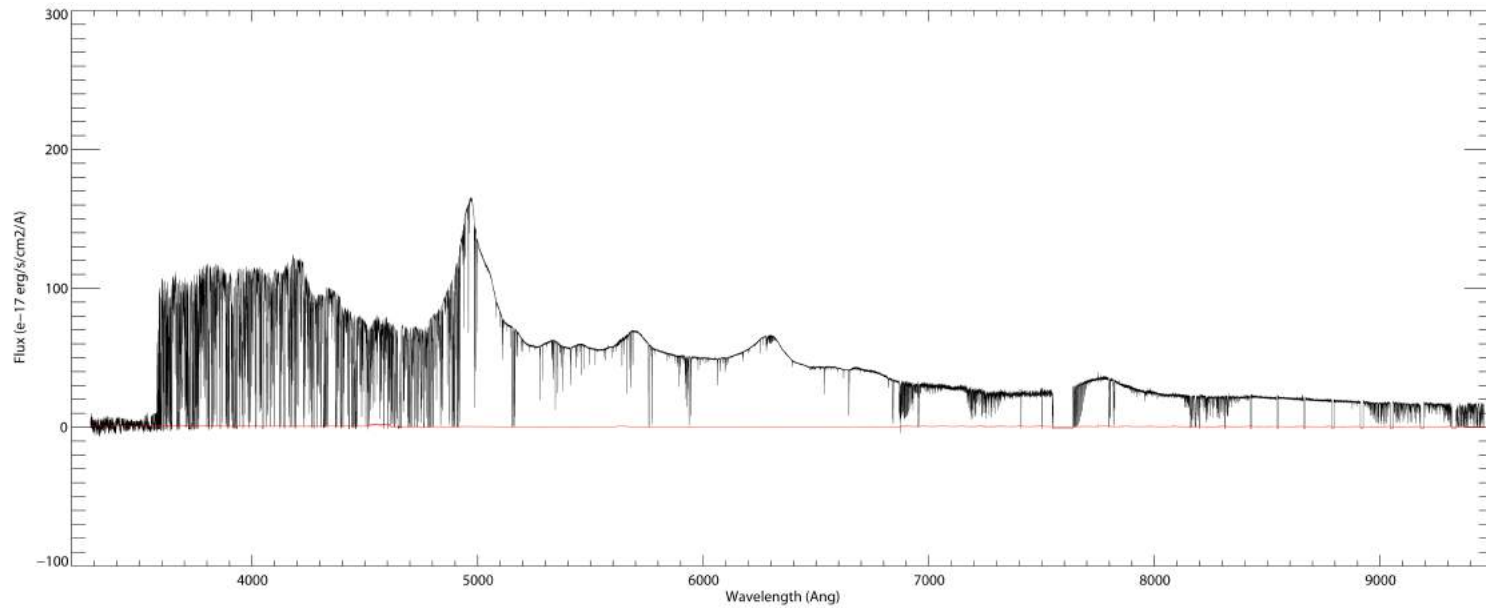
Figure 4.11: The full flux-calibrated "Deep Spectrum" of the QSO HE0940-1050. This exceptional spectrum reaches a SNR per pixel in the CIV region SNR ~ 231 and extends from ~ 3300Å to ~ 10000Å. All the data reduction, flux calibration and merging of the ~ 64 hours of exposure have been done with our new pipeline described in chapter 3.

4.5 The "Deep Spectrum" Large Program

Recently an ESO program (*Deep Spectrum - PI S. Cristiani*) has been carried out to explore deeply the $\log N_{CIV} < 12.0 cm^{-2}$ region and probe directly the metallicity of the mean density gas.

Using the UVES echelle spectrograph mounted at the ESO-VLT telescope ~ 65 hours of observations of the QSO HE0940-1050 at $z \sim 3$ have been carried out reaching the impressive SNR ~ 230 per pixel (velocity width of 3.0 km/s) in the CIV region (~ 6000Å).

This program has been in some way the precursor of the work presented in this thesis. As a co-investigator of the program I have been responsible for the data reduction, flux calibration and the co-adding of the individual exposures to produce the final spectrum which can be seen in Fig. 4.11.

Here will be presented the first results from the CIV analysis of the final Deep Spectrum concerning the calculation of the CDDF. All the results from the analysis of the Deep Spectrum, including the metallicity measurements and the study of the relation between metals and associated $Ly\alpha$ lines, will be presented in a forthcoming paper (D'Odorico et al., in preparation).

The comparison with previous measurements has been done using the FLO (From-Line-to-Overdensities) approach (Saitta et al., 2008). Basically FLO is the process of merging together lines closer than the local *Jeans Length* to form a new system with column density equal to the sum of the column densities. Further details on this technique will be provided in the next section, here it is important to point out that FLO approach helps to minimize differences between datasets which often arise from the use of different packages for the Voigt profile fitting.

In Fig. 4.12 the Deep Spectrum data are compared to the high SNR data of Ellison et al. (2000) and to the data of D'Odorico et al. (2010). None of the datasets have been corrected for completeness in order to make more evident the effect of the improved sensitivity.

The two datasets used for comparison have been recalculated with the FLO approach using the lists of CIV published by the respective authors. Both datasets are very different from each other: the former has been obtained under similar conditions with respect to Deep Spectrum data, except for the different redshift of the two QSOs, the latter differentiates for SNR and for number of lines of sight (1 vs 25).

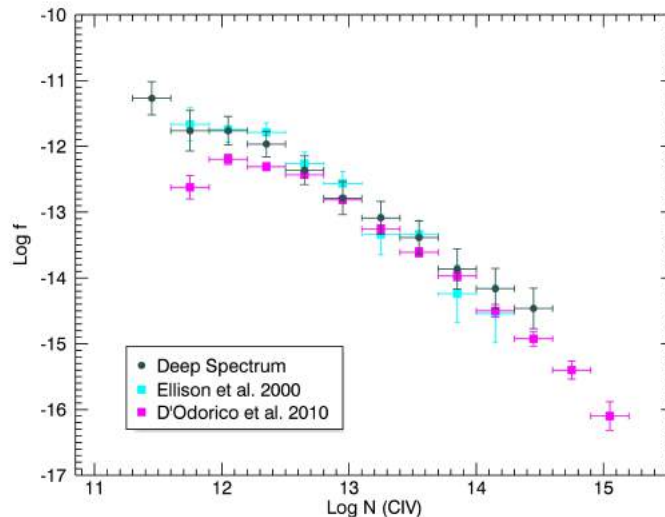As can be seen Deep Spectrum data confirm and strengthen the increasing

Figure 4.12: Column Density Distribution Function obtained from the analysis of the CIV identified in the Deep Spectrum. The lists of CIV have been recomputed using the FLO approach and a bin of $\Delta(\log N_{CIV}) = 0.3$ has been used for the calculation. Error bars are $\pm\sigma$ based on the number of points in each bin. Deep Spectrum data are compared with the high SNR data from Ellison et al. (2000) (only one QSO) and data from D'Odorico et al. (2010) (33 QSOs). The improved completeness at the low column densities and the lack of statistic (only one line of sight) at the high column densities are clearly evident.

trend found in Ellison et al. (2000). This effect is due to the improved completeness at the low column densities which is the result of the very high SNR of the Deep Spectrum. As explained in 4.1 the minimum threshold required for an efficient detection of typical $\log N_{CIV} < 12.0$ CIV features is $SNR \gtrsim 100$ which has been reached and largely exceeded by the Deep Spectrum.

On the other hand differences in SNR and in the number of available lines of sight with respect to D'Odorico et al. (2010) become evident respectively at the low column densities tail and in the high column density bins, while the central part is in agreement within errors.

The study presented from the Deep Spectrum has two main limitations: first, only one sightline is not enough to obtain robust conclusions, especially at the opposite extremes of the distribution, from a statistical point of view more sightlines are required to sample appropriately the sky and reduce the cosmic variance; second, only one QSO will cover a limited red-

shift range reducing the possibility to study the evolution of the measured physical quantities.

A further step in the exploration of the very low column densities regime is to improve present measurements, based on a single object, using a sample of very high SNR quasar including at least ten objects.

4.6 The very high SNR quasar sample

Main characteristics of our very high SNR QSO spectra have been resumed in Table 2.4. In this section we report for each QSO only the available redshift range useful for the CIV detection and so for the calculation of the redshift path covered by each QSO, the z_{min} and z_{max} values are reported in Table 4.1.

For the definition of the upper limits we followed the common use to cut from the considered paths the portion of the spectrum closer than 5000 km/s to the QSO $Ly\alpha$ emission redshift, this should eliminate the bias which could be introduced by systems nearby the QSO. In addition regions blended by very strong or complex systems, like for example telluric absorption lines, have been removed from the total redshift path.

As anticipated in the previous section, in order to ensure a meaningful comparison between different datasets all the measurements that will be presented henceforth have been computed using the FLO technique; for already published measurements publicly available CIV lists have been used to recompute the final results with the FLO technique.

This technique relies on the assumption that in general lines should not be closer than a *Jeans Length*, which represents the minimum dimension required by a system to be considered in hydrostatic equilibrium.

With this technique all lines closer than the Jeans Length are joined together in order to form a new feature whose column density is the sum of the column densities. The new Doppler parameter is the mean of the Doppler parameters weighted with respective errors and the final redshift is the mean of the redshifts of the single lines weighted with the correspondent column densities.

The main purpose of this technique is to account for all the differences which could affect the final results when different softwares, like for example FIT-LYMAN (Fontana and Ballester, 1995), have been used for the Voigt profile

Table 4.1: Redshift interval covered by each QSO of the sample available for the detection of CIV.

QSO name	z_{min}	z_{max}
B1937-1009	2.752	3.790
B1618-0035	2.699	3.710
B0055-26	2.657	3.662
B2126-15	2.351	3.268
B0940-1050	2.196	3.089
J0407-4410	2.157	3.021
J0103+1316	1.922	2.721
B1347-2457	1.813	2.575
J2220-2803	1.681	2.414
B1122-168	1.667	2.404
B0237-2322	1.534	2.239
J1344-1035	1.508	2.149
B0515-4414	1.130	1.730

fitting process. For example it has been observed that VPFIT generally requires more small components with respect to FITLYMAN in order to obtain a good value of the χ^2.

All those small components are often unphysical (they can be recognized by Doppler parameter values too small, $b < 2\ km/s$, or excessively large, as $b > 90\ km/s$) and are usually placed very close to other real lines in the system.

The final list of absorbers obtained with the FLO technique is free from software dependant spurious effects and generally more realistic. As discussed in the next subsection those characteristics will become very important even for our particular approach to the completeness correction.

4.6.1   Completeness correction

For what concern the completeness correction the usual approach is to add at random positions in the spectra absorption features with different Doppler parameters for each one of the low column density bins. The correction factor is obtained from the number of "false" lines retrieved by eye in the
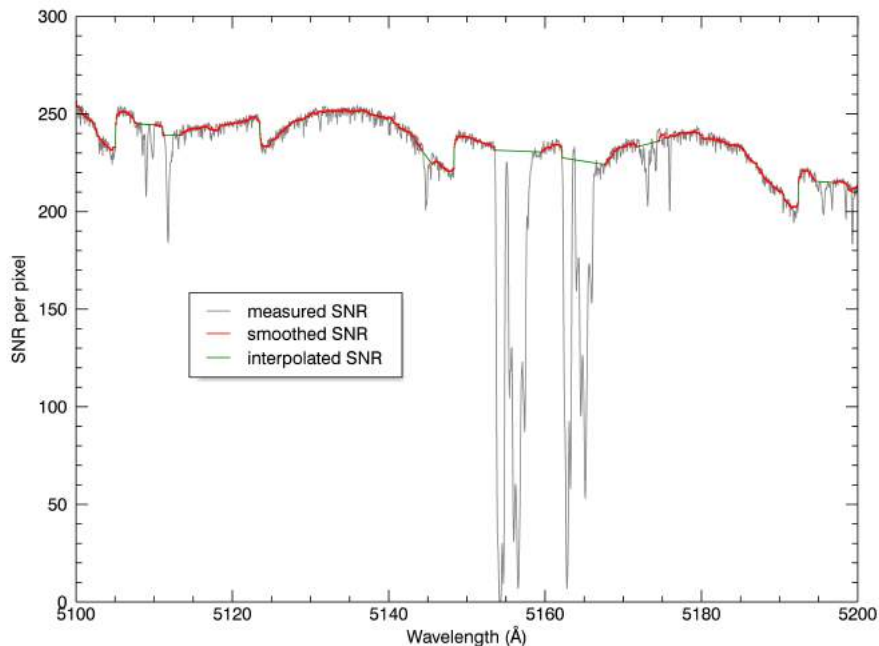
Figure 4.13: This figure shows the process followed for the calculation of the SNR used in the redshift path correction procedure. The measured SNR per pixel (gray line) has been median smoothed with a 30 pixel window (red line), then the regions corresponding to absorption lines have been identified using the normalized spectra and linearly interpolated (green line). This procedure is required since our approach to completeness correction must rely on the ideal SNR that one would have in absence of absorptions.

spectra.

This procedure has the limitation that if performed by different people it could produce different results because of the subjectivity of the identification process.

In this work we try a different approach based on the SNR of the spectra and on the calculation of the *effective redshift path*. This kind of approach, which will be described below, can be performed automatically and thus it ensures a more reproducible and robust correction.

As expressed in Eq. 4.12 each bin of the CDDF is usually calculated from the number of absorption lines with column densities in the interval ΔN_{bin} identified over the total available redshift path ΔX.

The main problem with this calculation is that the total redshift path usually includes also spectral regions with SNR lower than the detection threshold imposed by Eq. 4.6, in this way it cannot be clear if the number of detected

lines is real or if it is lower because in some regions the SNR of the spectra is not high enough for a clear detection.

In order to properly account for this effect one should use the formula of Eq. 4.6 to rule out from the calculation regions with SNR lower than the threshold. The redshift path calculated in this way provides a meaningful statistical result ensuring a realistic correspondence between the redshift path and the number of detected absorption lines, which is equivalent to a completeness correction.

Using this approach it is possible to associate an *effective redshift path* ΔX_i to each detected CIV line using its column density N and Doppler parameter b obtained from the Voigt profile fitting. In this case the corrected CDDF can be calculated as

$$f(N) = \sum_{i=1}^{n} \frac{1}{\Delta N \Delta X_i} \tag{4.19}$$

while the Ω_{CIV} can be obtained similarly as

$$\Omega_{CIV} = \frac{H_0 m_{CIV}}{c \rho_{crit}} \sum_{i=1}^{n} \frac{N_i}{\Delta X_i} \tag{4.20}$$

in both cases the summation is over the number n of CIV lines detected in the considered bin.

The practical procedure followed for the calculation of the effective redshift paths of the CIV lines is illustrated in Fig. 4.13 and Fig. 4.14. First the SNR of each spectrum is median smoothed with a 30 pixels window (red curve in Fig. 4.13). In order to remove the "holes" left in the SNR by the absorption lines the normalized spectrum has been used to identify regions of the spectrum lower than the QSO continuum by more than the 1%, then the resulting SNR has been obtained by linear interpolation of the "continuum" regions (green curve in Fig. 4.13). This preliminary procedure is very important since our approach to completeness correction has to rely on the ideal SNR that one would have in a spectrum free from absorption lines.

For each CIV doublet the smoothed SNR calculated in this way is then compared with the 3σ detection threshold of the stronger component (CIV 1548) obtained from Eq. 4.6 using the correspondent parameters of the doublet. Only spectral regions whose correspondent SNR is higher than the threshold will contribute to the calculation of the final effective redshift path.

An example of this procedure is represented in Fig. 4.14, colored dashed lines represent the limiting SNR for three representative absorption lines:
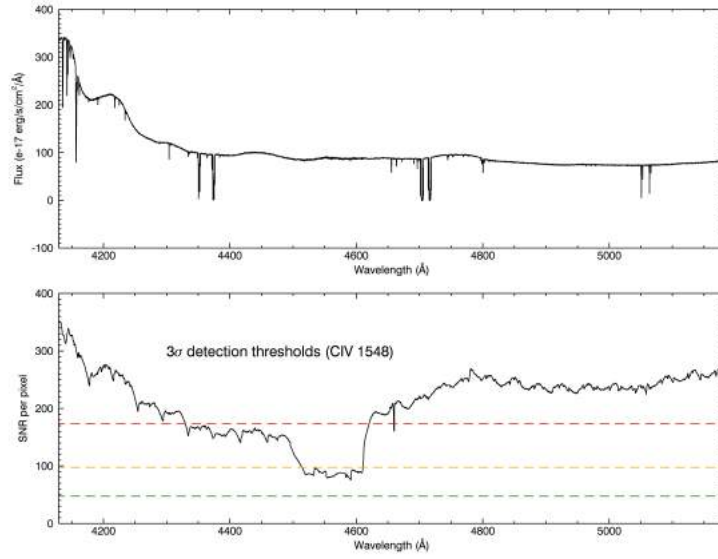
(a) CIV 1548
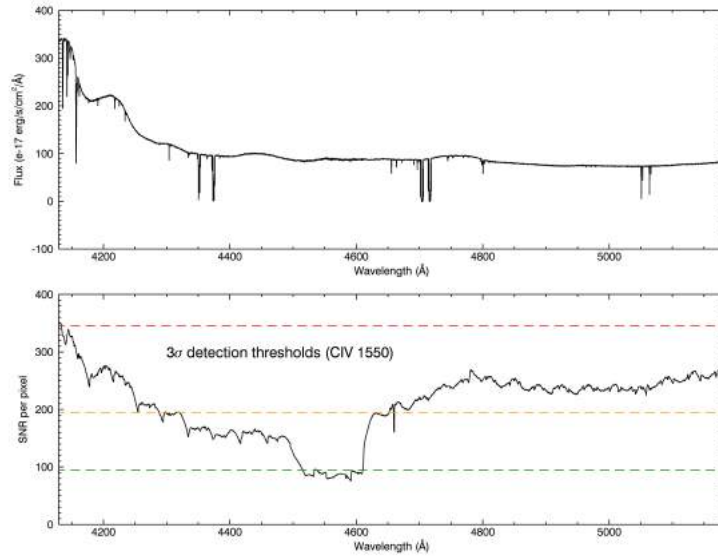
(b) CIV 1550

Figure 4.14: These two figures represent an example (spectra of QSO B1122-168) of the procedure described for the calculation of the effective redshift path. In each figure the upper panel shows the CIV region of the spectrum while the lower panel shows the correspondent SNR median smoothed with a 30 pixels window. The colored dashed lines represent an example of the limiting SNR required for a 3σ detection of three typical CIV absorption lines: $\log N_{CIV} = 11.2$ and $b = 4$ (red line), $\log N_{CIV} = 11.6$ and $b = 8$ (orange line), $\log N = 12.0$ and $b = 12$ (green line). For each one of the lines in the example the only regions of the spectrum which contribute to the effective redshift path are those one whose correspondent SNR is higher than the threshold indicated by the dashed lines. The thresholds are different depending on which one of the two CIV doublet transitions has been used in the calculation. Here for comparison both the 1548 (a) and the 1550 (b) cases are represented.

$\log N_{CIV} = 11.2$ and $b = 4$ (red line), $\log N_{CIV} = 11.6$ and $b = 8$ (orange line), $\log N = 12.0$ and $b = 12$ (green line).

The accuracy of the final result produced by the completeness correction just described is strongly dependant from the accuracy of the physical quantities (column densities and Doppler parameters) retrieved from the fitting procedure. In order to avoid unphysical values all lines which do not respect the detection criteria have been rejected prior to the execution of the completeness correction. Moreover each contribution to the CDDF and the Ω_{CIV} has been weighted with the correspondent effective redshift path to account for excessively small values at the denominator of Eqs. 4.19 and 4.20.

In this case the CDDF can be rewritten as

$$f(N) = n \sum_{i=1}^{n} \frac{1}{\Delta N \Delta X_i} \frac{\Delta X_i}{\sum_{i=1}^{n} \Delta X_i} = \frac{n^2}{\Delta N \sum_{i=1}^{n} \Delta X_i} \qquad (4.21)$$

and the Ω_{CIV} becomes

$$\Omega_{CIV} = \frac{H_0 m_{CIV}}{c \rho_{crit}} n \sum_{i=1}^{n} \frac{N_i}{\Delta X_i} \frac{\Delta X_i}{\sum_{i=1}^{n} \Delta X_i} = \frac{H_0 m_{CIV}}{c \rho_{crit}} n \frac{\sum_{i=1}^{n} N_i}{\sum_{i=1}^{n} \Delta X_i} \qquad (4.22)$$

where the $\Delta X_i / \sum_{i=1}^{n} \Delta X_i$ represent the normalized weights.

4.6.2 The CDDF down to $\log N_{CIV} \sim 11$

The first measurement performed with the list of CIV absorbers derived from the analysis of the full high SNR QSO sample concerns the Column Density Distribution Function. Our measured CDDF extends from $\log N_{CIV} = 11.0$ to $\log N_{CIV} = 15.3$ and has been calculated using a column density bin $\Delta(\log N_{CIV}) = 0.3$ with N_{CIV} in cm^{-2}.

In Fig. 4.15 it is illustrated the comparison with the two CCDF obtained from Ellison et al. (2000) and D'Odorico et al. (2010) calculated in the same way. The error bars are $\pm \sigma$ and have been obtained propagating in Eq. 4.21 the error on counts which in Poissonian statistic is represented by the square root of the number of points in each bin.

The completeness correction has been performed with the procedure described before after rejecting lines which not satisfy the detection criteria. Data without completeness correction are represented by red crosses and plotted for comparison.

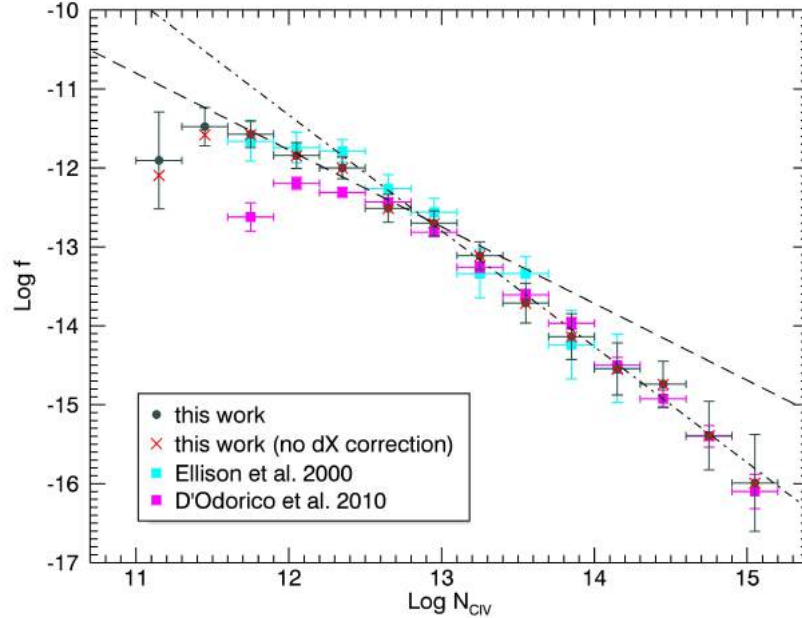After a first look clearly appears that our distribution is in pretty good

Figure 4.15: Column Density Distribution Function calculated from the whole high SNR sample and compared with data from Ellison et al. (2000) and D'Odorico et al. (2010). Our sample has been corrected for incompleteness using the effective redshift path approach, points without correction have been plotted for comparison (red crosses). The dashed and dot-dashed lines represent our "broken" power-law modeling of the form $f(N) = BN^{-\alpha}$ with $\alpha_1 = 0.97 \pm 0.17$ for the points in the interval $11.6 < \log N_{CIV} < 13.1$ (dashed line) and $\alpha_2 = 1.47 \pm 0.20$ for the points in the interval $13.1 < \log N_{CIV} < 15.2$ (dot-dashed line).

agreement with previous measurements. In particular at the high column densities the statistical significance of our calculation has been improved with respect to the Deep Spectrum alone, as indicated by the good correspondence with D'Odorico et al. (2010) data (based on 25 lower SNR spectra).

On the other side of the distribution the very high SNR of our sample allowed to push our CIV detection down to $\log N_{CIV} \sim 11.0$ which is the lowest column densities ever probed. Moreover our data confirm that going toward the lowest detectable column densities the number of CIV absorbers is still increasing as suggested by previous measurements.

Looking at Fig. 4.15 it can be noticed that our completeness correction does
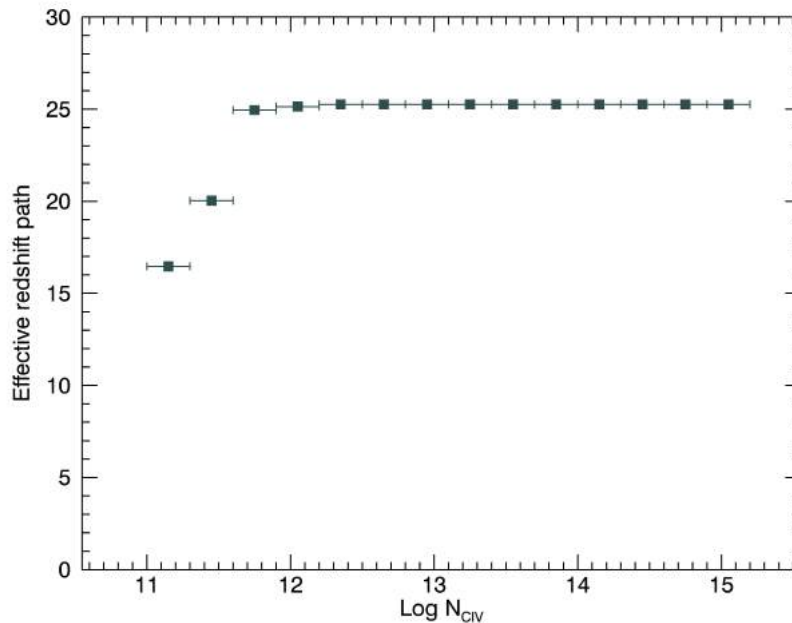
Figure 4.16: Effective redshift path which contributes to each one of the column density bins used for the calculation of the CDDF. The effective redshift paths have been calculated using the limiting SNR required for a 3σ detection of a CIV 1548 absorption line.

not change too much the position of uncorrected points (red crosses), from the correspondence of our data before and after the completeness correction it appears that our sample is complete down to $\log N_{CIV} \sim 11.6$.

This observation can be confirmed from the plot in Fig. 4.16 which shows the effective redshift path resulting from our correction for the same column density bins used for the calculation of the CDDF. This plot shows that our sample can be considered complete in the interval $11.6 < \log N_{CIV} < 15.2$ in the sense that all available spectra (that is equivalent to say "the whole available redshift path") have a SNR high enough to contribute in the same way to the detection of CIV 1548 lines in those column density bins.

To model our data in the complete interval we decided to use a broken power law of the form $f(N) = BN^{-\alpha}$ with double index to describe the two half of the distribution. In Fig. 4.15 are represented the best fits of the model, one with $\alpha_1 = 0.97 \pm 0.17$ for the points in the interval $11.6 < \log N_{CIV} < 13.1$ (dashed line) and the other with $\alpha_2 = 1.47 \pm 0.20$ for the points in the in-
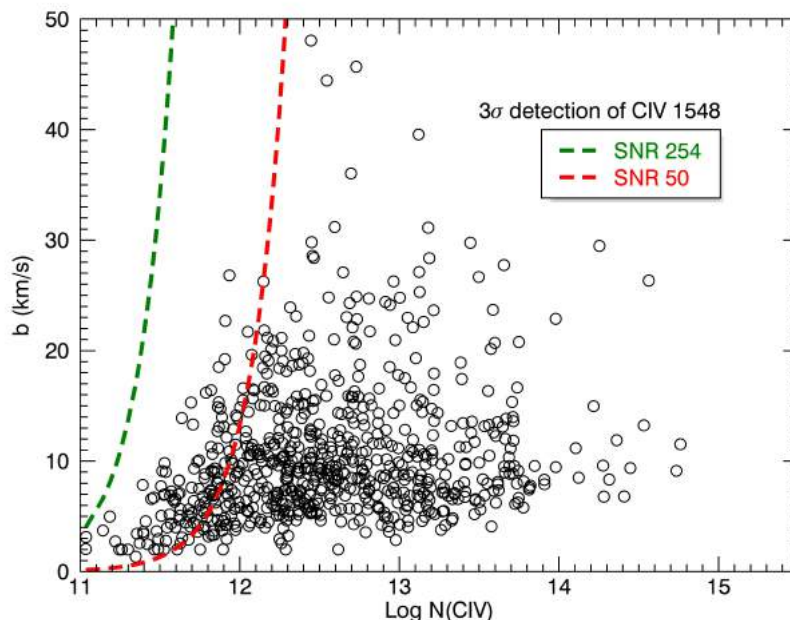
Figure 4.17: Distribution of absorption lines from the full QSO sample in the $(b - \log N)$ plane. Both values for each line have been obtained from the Voigt profile fitting, errors bars have been omitted for better clarity. This distribution shows clearly the strong degeneracy between Doppler parameters and column densities of detected lines. The dashed lines indicate the maximum Doppler parameter which can be detected at correspondent column densities with a fixed SNR. The area leftward the curve represents the "forbidden region" which cannot be observed with the specified SNR. The conventional $SNR = 50$ spectra case is also represented to show the enormous improvement of our sample.

terval $13.1 < \log N_{CIV} < 15.2$ (dot-dashed line).

Now it is important to observe that to be "complete" in a column density bin does not necessary mean that we are observing all the lines present in that bin, indeed there could be the possibility that none of the available spectra have SNR high enough to detect a certain line. This is a direct consequence of the degeneracy which exists between the column density and the Doppler parameter. For example if we chose a fixed column density allowing free Doppler parameter value, the detection threshold defined in Eq. 4.6 becomes increasingly high as the the Doppler parameter increases.

In those cases there are no way to recover a missing information and it is not possible to talk about completeness correction anymore. In order to better

understand those arguments in Fig. 4.17 it is illustrated the distribution of our detected absorption lines in the $(b - \log N)$ plane. The colored dashed lines have been obtained using Eq. 4.6 fixing a SNR and calculating for the correspondent column densities the maximum detectable Doppler parameter. In other words the region leftward the curve represents the "forbidden region" which means that lines with parameters in that region cannot be detected in spectra with SNR less then the one represented by the curve.

In our figure we have chosen to plot the threshold relative to the highest SNR available in our QSO sample (SNR ~ 254 for the spectrum of the QSO B0515-4414) in this way we can be pretty sure to separate the area of the $(b - \log N)$ plane which we are not able to see (which cannot be corrected) from the area where we have at least one spectra that could see something (where we can apply our completeness correction).

Even if a big improvement has been obtained with respect to conventional spectra (see the SNR ~ 50 red dashed curve in Fig. 4.17 for a comparison) we see that there is still a region of the $(b - \log N)$ plane which remains unaccessible to our very high SNR QSO sample. This region is just the one which corresponds to the $11.0 < \log N_{CIV} < 11.3$ column density bin. On the other hand if we look at column density $\log N_{CIV} \sim 11.6$ we see that the SNR is high enough to detect lines with Doppler parameters up to $b \sim 50 \ km/s$ which is a confirmation of our completeness interval.

In the lowest column density bin we are missing all the lines with Doppler parameter $b > 5 \ km/s$, in practice to say something about those lines there are no other ways than increase the SNR of our spectra even more. A completeness correction cannot accounts for lines which cannot be detected and therefore we still cannot say if the forbidden region is populated with line or not.

To understand how many lines we should expect in this region, in Fig. 4.18 is represented the number of lines as a function of both the Doppler parameter and the column density. This plot can be seen in two alternative ways: we can fix a column density (identified by the different colors) and see how the number of lines varies across the Doppler parameters, or we can focus on a single bin of Doppler parameters and see how the number of lines in that bin varies for different column densities.

Focusing for example on the $b = 4$ bin where the chances of detection at each column density increase for all spectra, we see that the number of lines
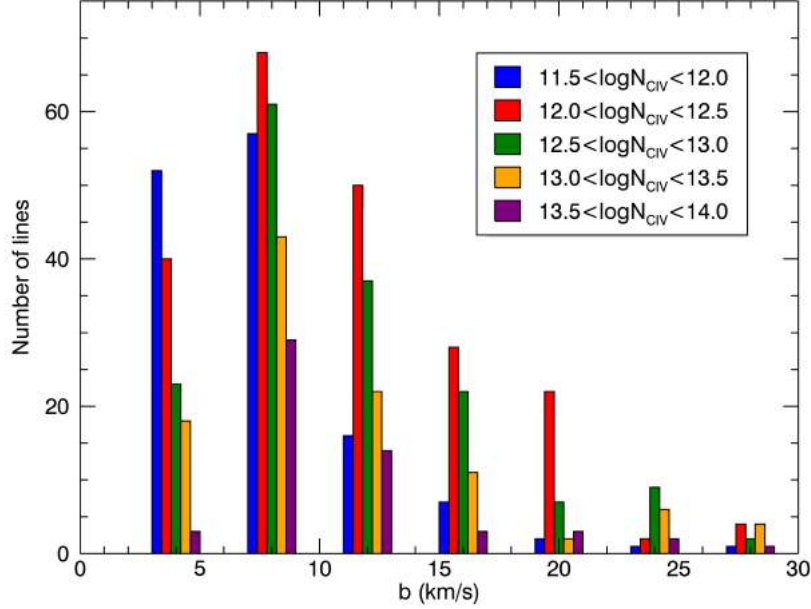
Figure 4.18: Distribution of the number of lines as a function of both the Doppler parameter and the column density. The distribution has been obtained with a $\Delta b = 4$ and with column densities intervals specified in the plot. Can be noticed that the distribution shows a maximum in correspondence of $b \sim 10 \; km/s$ and that for a fixed Doppler parameter the number of lines shows an increasing trend as the column density decreases.

increases as the column density decreases. This trend is preserved even at higher Doppler parameter except for the lowest column densities, represented by the blue bar. Moreover we see that the relative number of blue and red lines decreases as the Doppler parameter increases.

This effect can be interpreted as a manifestation of the incompleteness of our sample at the lowest column densities which becomes more important for larger Doppler parameters. Moreover there is a general trend in the distribution which shows a maximum near the value $b \sim 10 \; km/s$.

Taking into account all these aspects one would expect that our forbidden region should be the most populated one. In this case considering that the majority of CIV absorbers would have Doppler parameters $b \sim 10 \; km/s$ at $\log N_{CIV} \sim 11.1$ the threshold required for a 3σ detection of the CIV 1548 line would be SNR ~ 343 (per pixel).

Looking at Table 2.4, recalling that the SNR increases as the square root of

the observation time, it can be understood that to reach such an high SNR with present telescopes the exposure time of the highest SNR spectra has to be increased by a factor 9/4. For example if a SNR ~ 230 requires at least 50 hours of observation, to obtain a SNR ~ 345 spectrum at least 112 hours of observation would be required.

This huge amount of hours of observation can be justified only demonstrating that the contribution of all the lines in the forbidden region to the total density of CIV provides a significant amount of information with respect to the higher column densities.

In order to further investigate the contribution of the lower column densities to the total CIV density, in the next subsection will be presented the measurement of the total mass density of the CIV and the correspondent cumulative distribution including all the detected lines in the interval $11.0 \leq \log N_{CIV} \leq 15.0$ will be compared with the one from previous lower SNR sample.

4.6.3 New estimate of Ω_{CIV} in the range $1.0 < z < 4.0$

In this section it will be presented a new estimation of the total mass density of the triply ionized Carbon as a function of redshift in the range $1.0 < z < 4.0$. Owing to the extremely good SNR of our sample for the first time ever this measurement has been extended to very low column densities including all the lines in the $11.0 \leq \log N_{CIV} \leq 15.0$ interval as illustrated in Fig. 4.19 (a).

The completeness correction has been performed as discussed in 4.6.1, points without correction are represented by red crosses and plotted for comparison. As can be seen from direct comparison for this column densities interval our sample can be considered almost complete since the completeness correction is very small. This is probably due to the small contribution of $\log N_{CIV} < 12.0$ lines.

To visually inspect the statistical significance of each one of the Ω_{CIV} points of our calculation, we report in Fig. 4.19 (b) the number of lines per redshift path which have been detected in each one of the defined redshift bins.

In order to investigate the behavior of different column densities regimes we repeat the measurement of Ω_{CIV} dividing the sample into low, $11.0 < \log N_{CIV} \leq 13.0$, and high, $13.0 < \log N_{CIV} \leq 15.0$, column densities as illustrated in Fig. 4.20 (a). The statistical significance of each bin can be
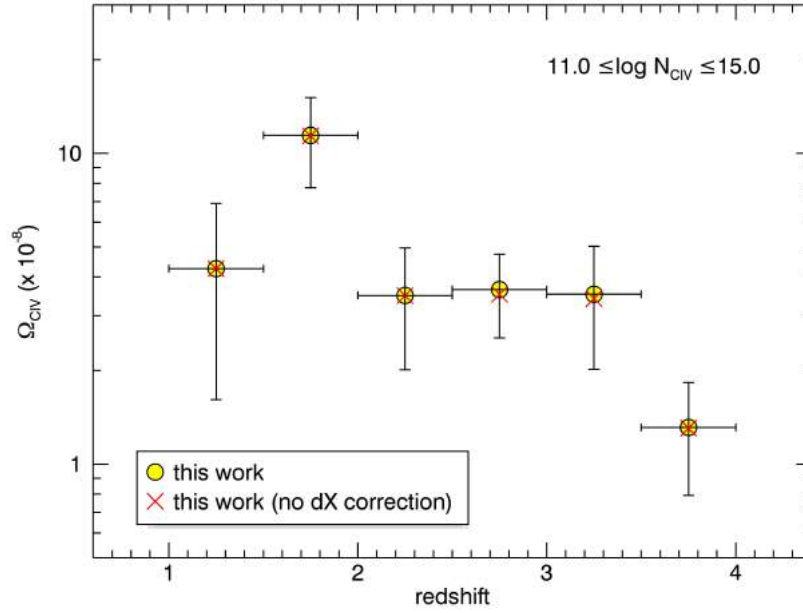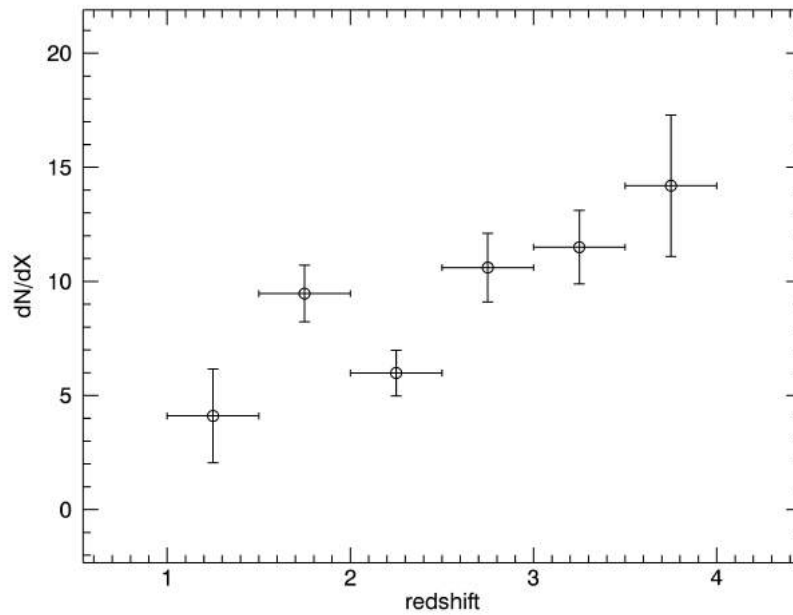
(a) Ω_{CIV} our sample

(b) dN/dX our sample

Figure 4.19: (a) The first ever measurement of the Ω_{CIV} distribution including all lines in the column density range $11.0 \leq \log N_{CIV} \leq 15.0$. (b) Number density of absorbers per redshift path covered by our sample in the different redshift bins.
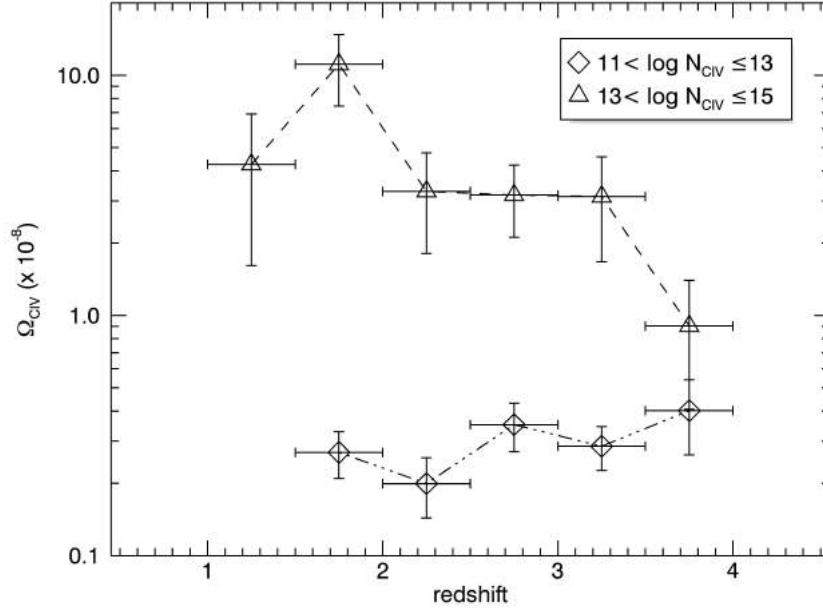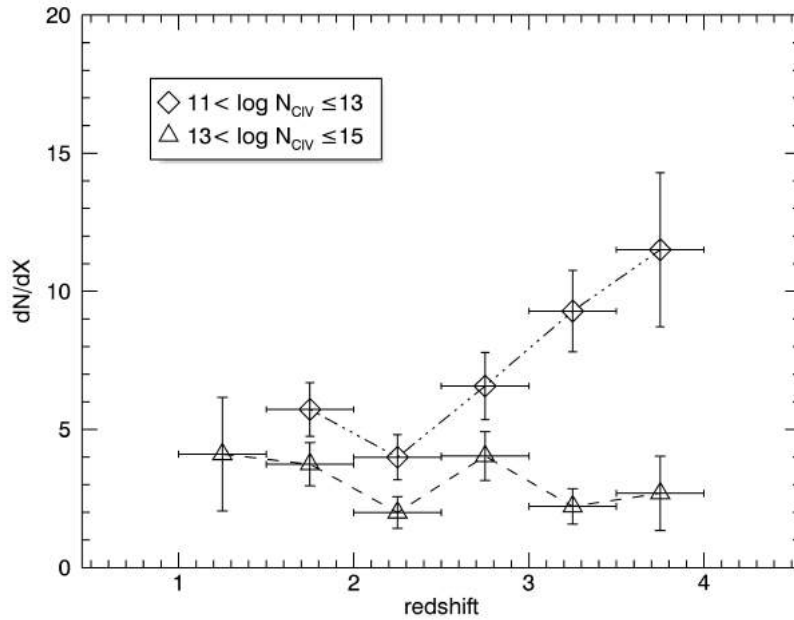
(a) Ω_{CIV} our sample (N bins)

(b) dN/dX our sample (N bins)

Figure 4.20: (a) The Ω_{CIV} distribution derived from our sample subdivided in two column densities ranges for high and low column densities. The $11.0 < \log N_{CIV} \leq 13.0$ bin traces faint absorption lines and should be more representative of the IGM. (b) Number density of absorbers per redshift path covered. Absorbers have been subdivided in the two column densities ranges indicated in the figure.
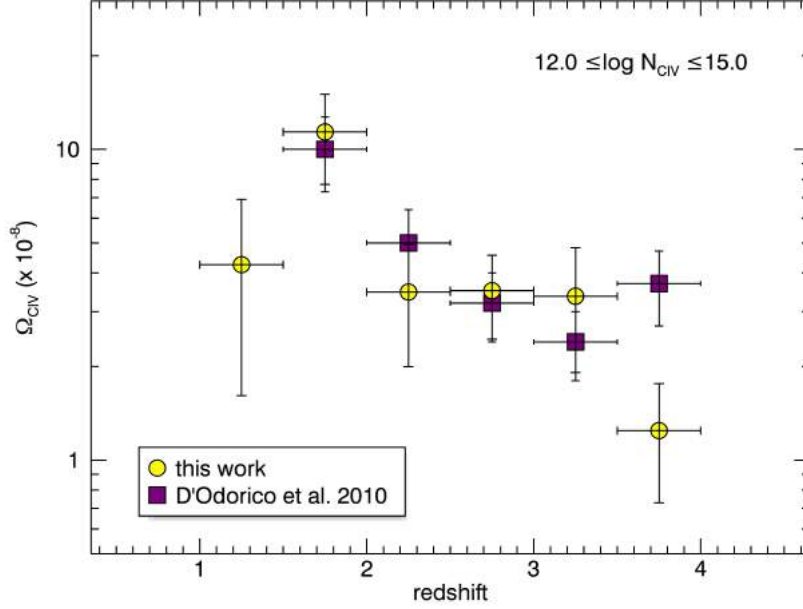
Figure 4.21: The Ω_{CIV} distribution derived from our sample compared with data from D'Odorico et al. (2010) in the same redshift range including only lines in the column density range $12.0 \leq \log N_{CIV} \leq 15.0$. A new point has been added in the $1.0 - 1.5$ redshift bin indicating an apparent decrease in the total mass density.

seen in Fig. 4.20 (b) where the number of lines per redshift path is plotted for the two different column densities regimes.

In principle, if the sample is unbiased, lower column densities should be more representative of the IGM. Following Storrie-Lombardi and Wolfe (2000) in the redshift range covered by our sample we should expect 10^{-1} DLAs per unit redshift path. Our sample contains three DLAs which is the exact number of DLAs that we expected considering our total redshift path.

As we can see the low column densities regime shows almost no evolution in the specified redshift interval, while the high column densities are in general more representative of the overall trend showed by the total mass density distribution. Furthermore this subdivision shows that the major contribution to the total mass density of CIV comes from the higher column densities.

Prior to this work the same measurement has been accomplished by D'Odorico et al. (2010) (hereafter D10) in a similar redshift range but with lower SNR spectra. In Fig. 4.21 it is illustrated a comparison of D10 data with our
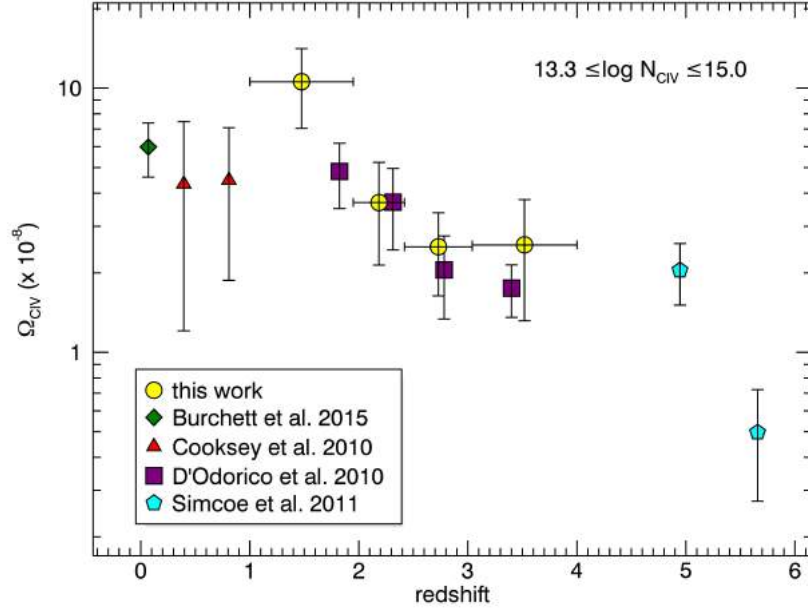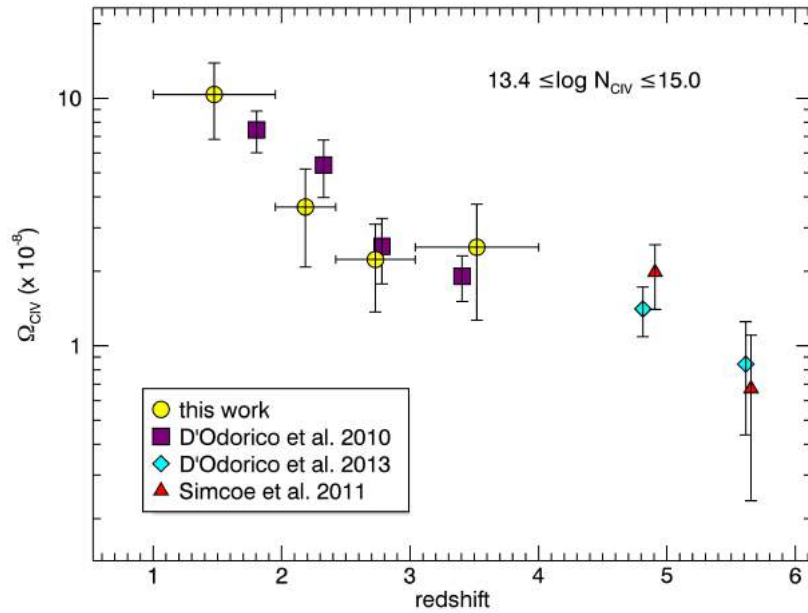
(a) $13.3 < \log N_{CIV} < 15.0$

(b) $13.4 < \log N_{CIV} < 15.0$

Figure 4.22: In those two figures we recalculate the Ω_{CIV} distribution limiting our sample to the column densities ranges $13.3 \leq \log N_{CIV} \leq 15.0$ (a) and $13.4 \leq \log N_{CIV} \leq 15.0$ (b) to compare the results with other data from lower SNR samples covering the whole redshift range $0 < z < 6$. Different datasets are expressed by the different colors as reported in the legends of the plots.

measurements. In order to make a meaningful comparison our results have been computed using the same $\Delta z = 0.5$ redshift bin and including in the calculation only lines in the $12.0 \leq \log N_{CIV} \leq 15.0$ interval.

From this comparison can be noticed that our measurements are in good agreement with D10 in the central bins, while the only significant discrepancy is in the $[3.5 - 4.0]$ redshift bin where our data seem to be lower. Moreover our sample allowed to extend the redshift coverage adding a new point in the $[1.0 - 1.5]$ redshift bin. This new point shows an unexpected decrease from the peak of Ω_{CIV} at the redshift bin $[1.5 - 2.0]$.

It is worth to notice that the statistical significance of the first and last bins is lower with respect to the other ones (as indicated by the errors bars). The low statistical significance of those bins is due to the fact that fewer QSOs contribute to the low and high ends of the redshift interval covered by our sample, as can be seen from Table 4.1.

Now if we compare the distribution of Fig. 4.21 in the interval $12.0 \leq \log N_{CIV} \leq 15.0$ with the distribution of Fig. 4.19 (a) which includes even lines at column densities $\log N_{CIV} < 12.0$, it can be noticed that the two distributions do not differ too much one from the other. This means that the major contribution to the Ω_{CIV} comes from higher column densities.

In order to obtain a better idea and to quantify the fraction of CIV that each column density contributes to the total mass of CIV we represent in Fig. 4.23 the cumulative distribution of CIV mass density. In the figure are represented both the distribution obtained from our sample (gray) and the one obtained from D10 sample (cyan).

The main purpose is to show how much CIV can be added to the total by increasing the SNR (and so the completeness) of the QSOs sample, to this end we included in the two distributions all lines available from the respective sample in the redshift interval in common between them, which is $1.5 < z < 4.0$, and no completeness correction has been applied to data. The plot shows the Ω_{CIV} obtained summing all lines from $N_{max} = 10^{1}5.0$ to N_{min} which varies along the available column densities in the respective samples.

As can be seen the two distributions agree pretty well until $\log N_{CIV} \sim 12.7$ where our distribution continues to grow up while the other one seems to converge. This comparison tells us two important things: the first is that the convergence of D10 data it is only apparent and it is due to the effect
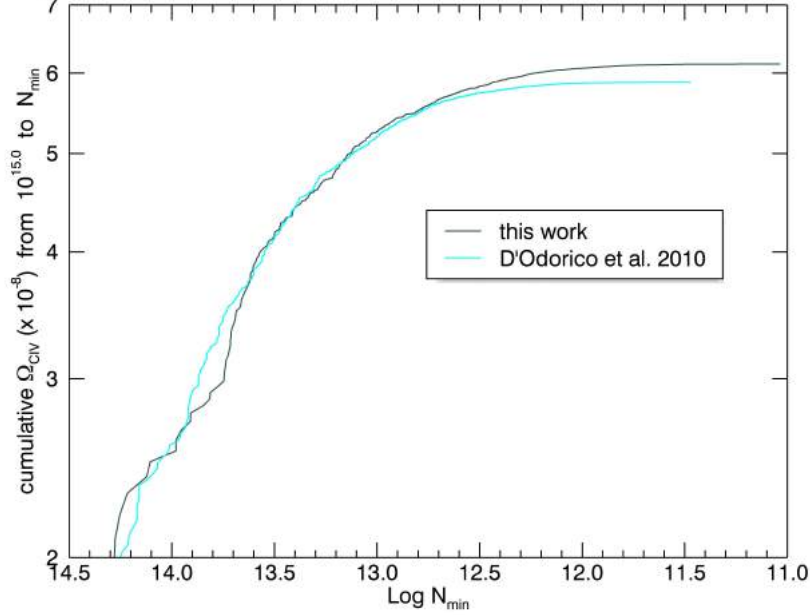
Figure 4.23: Normalized cumulative distribution of the CIV column densities for our sample and for CIV lines of D'Odorico et al. (2010). The distributions have been obtained from the summation of all column densities from N_{min} to N_{max} as a function of N_{min} and then normalized for comparison. The plot shows a zoom of the distributions near the region of convergence.

of incompleteness of D10 sample; the second thing is that, if our sample is almost complete down to $\log N_{CIV} \sim 11.7$ as discussed in the previous subsection, the convergence of our distribution seems to be real and shows that even if the number of lines with $\log N_{CIV} < 11.7$ increases no more significant CIV mass can be added to the total Ω_{CIV}.

After all even if look at the two distributions in the plot we see that our improved sample does not contribute significantly to the total mass density of CIV.

Finally, following the same argument of D10, in order to compare our results with measurements which cover the whole redshift range $0 < z < 6$ we repeated the computation of Ω_{CIV} changing the definition of the redshift bins and the column densities ranges. Details for each bin and each column density interval can be seen in Table 4.2.

The new redshift bins have been selected so that each one of them covers

Table 4.2: Measurements of the Ω_{CIV} in the specified redshift bins and column densities intervals performed on the QSOs of our sample.

z range	ΔX	# lines	Ω_{CIV} ($\times 10^{-8}$)	$\delta\Omega$ ($\times 10^{-8}$)
$11.0 \leq \log N_{CIV} \leq 15.0$				
1.00-1.50	0.97	4	4.25	2.64
1.50-2.00	6.13	58	11.42	3.68
2.00-2.50	6.01	36	3.49	1.48
2.50-3.00	4.90	52	3.64	1.09
3.00-3.50	5.04	58	3.52	1.50
3.50-4.00	1.48	21	1.31	0.52
$12.0 \leq \log N_{CIV} \leq 15.0$				
1.00-1.50	0.97	4	4.25	2.64
1.50-2.00	6.15	50	11.38	3.67
2.00-2.50	6.02	30	3.47	1.48
2.50-3.00	5.20	45	3.51	1.06
3.00-3.50	5.42	35	3.37	1.45
3.50-4.00	1.49	13	1.24	0.52
$13.3 \leq \log N_{CIV} \leq 15.0$				
1.00-1.95	6.32	19	10.57	3.53
1.95-2.42	6.26	10	3.69	1.55
2.42-3.04	6.31	16	2.51	0.87
3.04-4.00	6.36	8	2.55	1.24
$13.4 \leq \log N_{CIV} \leq 15.0$				
1.00-1.95	6.32	15	10.35	3.53
1.95-2.42	6.26	9	3.63	1.55
2.42-3.04	6.31	11	2.23	0.86
3.04-4.00	6.36	7	2.50	1.23
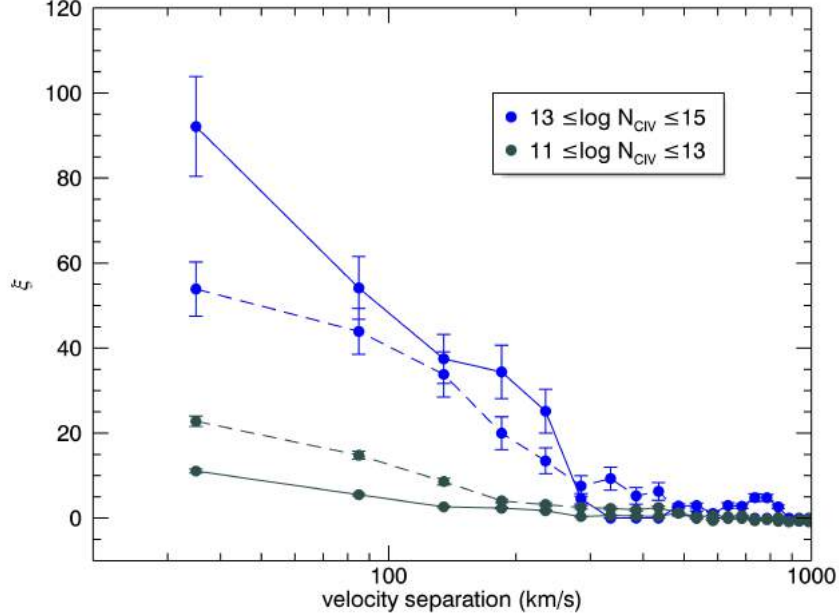
Figure 4.24: Two point correlation function of our CIV absorbers derived for low $11.0 < \log N_{CIV} < 13.0$ and high $13.0 < \log N_{CIV} < 15.0$ column densities subdivided into two redshift interval $1.5 < z < 2.5$ (dashed lines) and $2.5 < z < 4.0$ (solid lines).

more or less the same redshift path. Considering that the total redshift path is $\Delta X = 25.5$, dividing the sample in four, each bin will cover a redshift path $\Delta X_{bin} \sim 6.3$.

The column densities ranges have been chosen to match other data based on lower SNR spectra: $13.3 \leq \log N_{CIV} \leq 15.0$ to compare our results with B15 data as illustrated in Fig. 4.22 (a), while $13.4 \leq \log N_{CIV} \leq 15.0$ to make the comparison with D13 data in Fig. 4.22 (b).

Our results are in agreement with the previous data from D10 with the only exception of the first bin.

4.6.4 Two point correlation function of CIV lines

We have calculated the two point correlation function $\xi(v, v + \Delta v)$ for the detected CIV absorption lines selecting two redshift intervals and dividing the sample in low $11.0 < \log N_{CIV} < 13.0$ and high $13.0 < \log N_{CIV} < 15.0$ column densities as illustrated in Fig. 4.24. The redshift interval selected

are $1.5 < z < 2.5$ (dashed lines) and $2.5 < z < 4.0$ (solid lines) while column density intervals are distinguished by their colors.

As explained in 1.4.1 the two point correlation function can be used to study the possible origin of a set of absorbers and their connection with galaxies. Practically the measured two point correlation function can be compared to the one obtained from, for example, Lyman-break-galaxies (LBGs) to see if their clustering properties are similar.

The functions calculated and plotted in Fig. 4.24 show that CIV absorbers begin to have a significant clustering only for separation less than $\sim 400\ km/s$ for the low column densities and less than $\sim 600\ km/s$ for the high column densities. If we convert these velocity separations into comoving $Mpc\ h^{-1}$ and compare it with the clustering properties of LBGs (Cuesta et al., 2016; Anderson et al., 2014; Porciani and Giavalisco, 2002) it appears that low column densities CIV have clustering properties similar to LBGs while higher column densities show a stronger clustering which indicates a possible connection with denser environment.

It is interesting to note the different behavior of the functions at different redshift for the two column densities intervals: low column density CIV absorbers appear to be more clustered at higher redshift at variance with high column densities which appear more clustered at lower redshift.

Conclusions and future prospects

The study presented here represents a first step in the exploration of a new approach to high resolution spectroscopy. We have demonstrated that present observations are still far from the systematic limits imposed by the instrumentation. Indeed even with current telescopes and CCDs the SNR of spectra can be still increased as the square root of the observing time being essentially limited by the overall available exposure time.

In this context it becomes clear that it is of crucial importance to use softwares for the data reduction and analysis allowing to fully exploit the potential offered by the instrumentations.

When the systematics arising from the instrumentation become lower than those coming from the data reduction and analysis softwares, the direction to follow is to create specific softwares able to keep pace with the state of the art technology.

Those arguments become even more stringent if we look ahead to the next generation of 30 meters telescopes. It is clear that we are going toward an era of high precision cosmology and technical improvements are demanding appropriate softwares for handling data of such a good quality. The standard approach is not enough anymore but new solutions and approaches are required to reach the desired sensitivity.

This is the case for example of the analysis software (Cupani et al., 2015a,b) which is now under development for the ESPRESSO instrument (Pepe et al., 2013) which will be mounted at the ESO Very Large Telescopes.

This software and the data reduction pipeline presented in this thesis will represent the starting point for the development of the data reduction and analysis software for the next generation spectrograph HIRES (Zerbi et al.,

2014) which will be mounted on the E-ELT (McPherson et al., 2012) telescope and which is being designed.

Those kind of studies are not sterile technicalities but are driven mainly by the proposed science, think for example at the measurements of the fundamental physical constants using the *Sandage-Loeb test* just to cite one possibility (Cristiani et al., 2015).

As a case study here we have reported the most sensitive measurements of the weak CIV lines ever performed with high resolution spectrographs. Our sample of very high SNR QSO spectra has been obtained both from dedicated programs (like the *Deep Spectrum - LP*) and from archival data. Our attention has been focused on the two biggest telescopes available to date, the VLT and the Keck telescopes, and their state of the art instruments: UVES and HIRES respectively.

In order to maximize the SNR we have selected for the sample only those QSO observed for the longest possible time during the operational life of the two instruments. In this sense our sample represents the highest SNR QSO spectra sample available nowadays.

We have shown that our approach to data reduction allows to reduce systematics with respect to what achieved by other commonly used softwares. Moreover it is important to stress that the study presented here has been performed on echelle spectrographs constructed more than a decade ago. Next generation of instruments will allow to improve even more those results.

For example new ultra-stable spectrographs will be equipped with optical fibers. This technical solution will allow to significantly improve the optimal extraction and the flux calibration procedures which will rely on fully characterized and stable PSF all night long, independently from the particular seeing conditions.

In this thesis we have demonstrated that already available observations can provide co-added spectra of superior quality if handled in a proper way.

Looking at the first measurements that we have performed on our sample, we have seen that the high SNR allows to probe the IGM down to very low column densities providing robust data to be compared with simulations.

Moreover it has been shown that a rigorous approach to the data reduction

processes not only improves the final sensitivity of the spectrum but also provides more realistic errors producing final data which are robust and reliable.

Using our QSO sample we have calculated the Column Density Distribution Function, the total mass density of the triply ionized Carbon and the two point correlation function, pushing our calculations down to column densities $\log N_{CIV} \sim 11.0$ improving previous measurements which only account for $\log N_{CIV} \geq 12.0$.

Our particular approach to the completeness correction has shown that there is still a region of the $(b - \log N)$ plane which cannot be accessed, even with our very high SNR. This region is the one that we expect to be more populated, nevertheless it is still unclear how much those lines should contribute to the total CIV density.

After few calculation we have demonstrated that to observe the forbidden region of the $(b - \log N)$ plane a $SNR \geq 340$ would be required and that to obtain such an high sensitivity at least 110 hours of observation would be necessary.

However we have demonstrated that even with the sample presented in this work the cumulative distribution of Ω_{CIV} seems to converge indicating that even if the number of lines with $\log N_{CIV} < 12.0$ increases no more significant CIV mass can be added to the total Ω_{CIV}.

Measurements presented in this thesis are only an example of the possible measurements which could be done with those very high SNR spectra. The sample produced for this thesis represents a precious resource since it could be used for other measurements which are still limited by the SNR of present available spectra.

Some example of those measurements are the study of the distribution of the neutral hydrogen clouds with respect to the associated CIV ions as a function of the column density, the measure of the optical depth in the $Ly\alpha$ forest using the Pixel Optical Depth (POD) approach, studies of the thermal state of the IGM and of the UV background.

In the next future we plan to further exploit the spectra of our sample making some of those measurements, in particular the study of the $Ly\alpha - CIV$ correlation. Moreover the first phases of the design of the ELT-HIRES data reduction pipeline are already started and all the data reduction processes

described here are being analyzed to be implemented in this new pipeline.
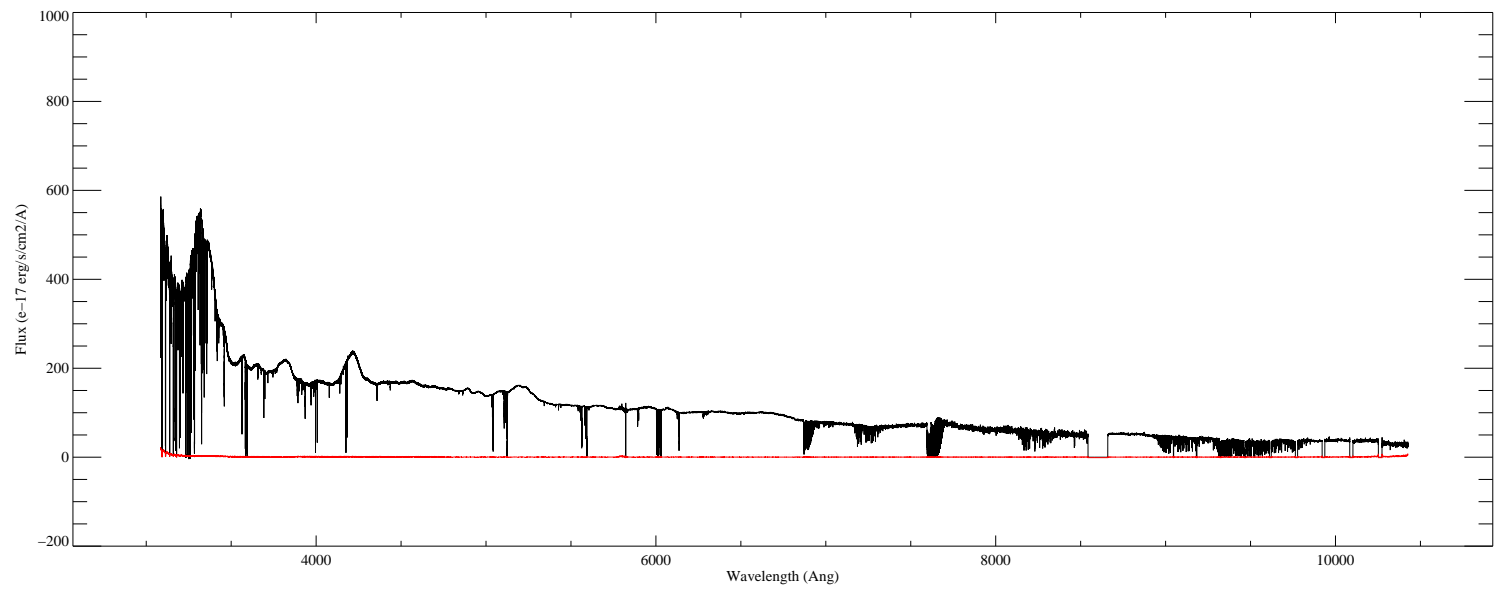
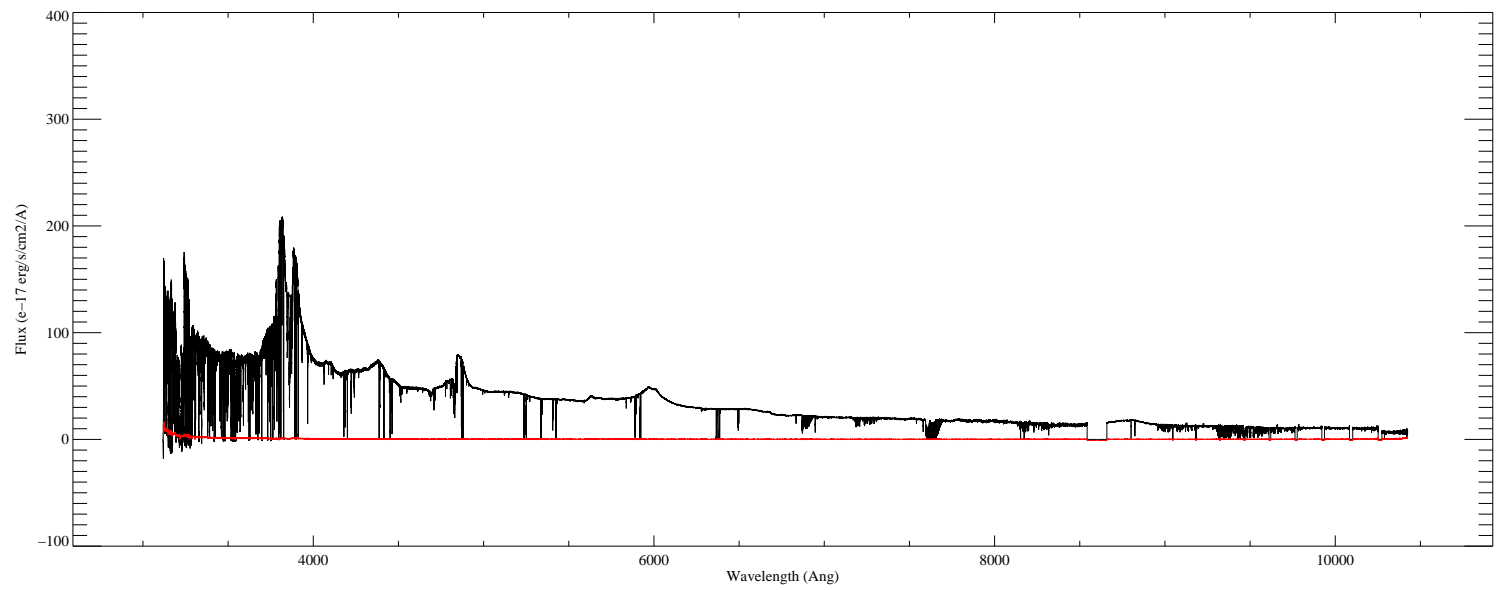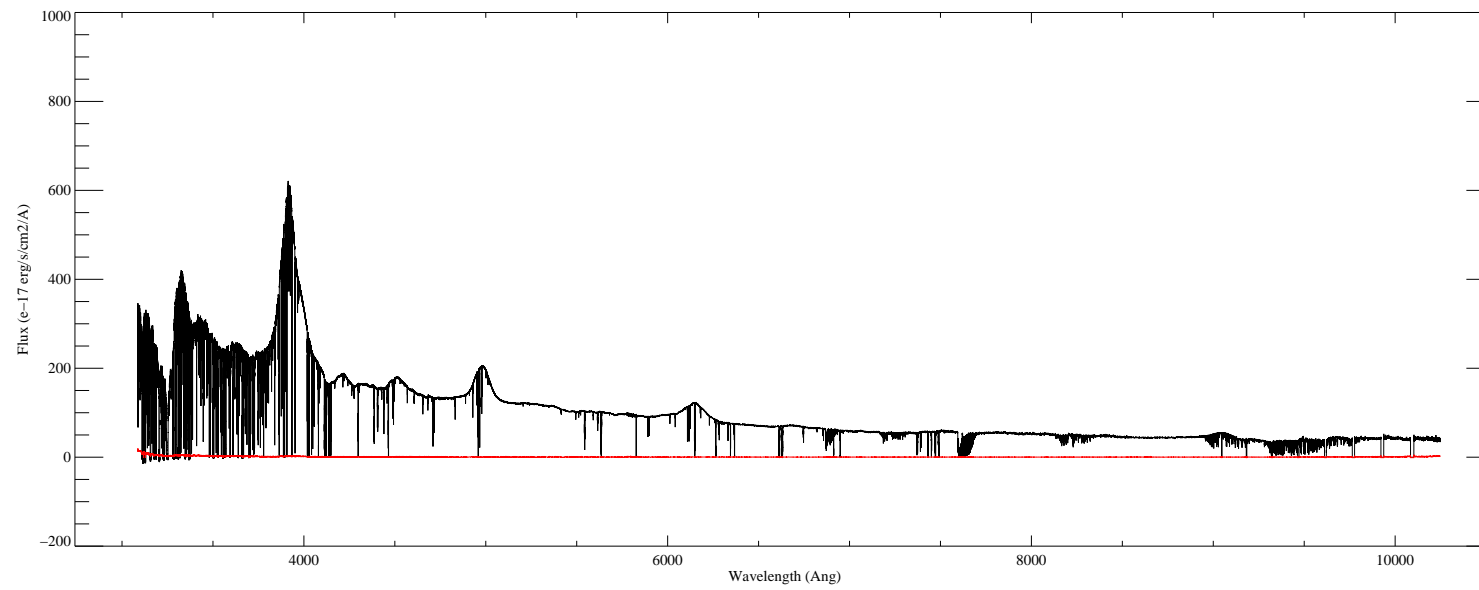# List of Spectra

Figure 4.25: B0515-4414

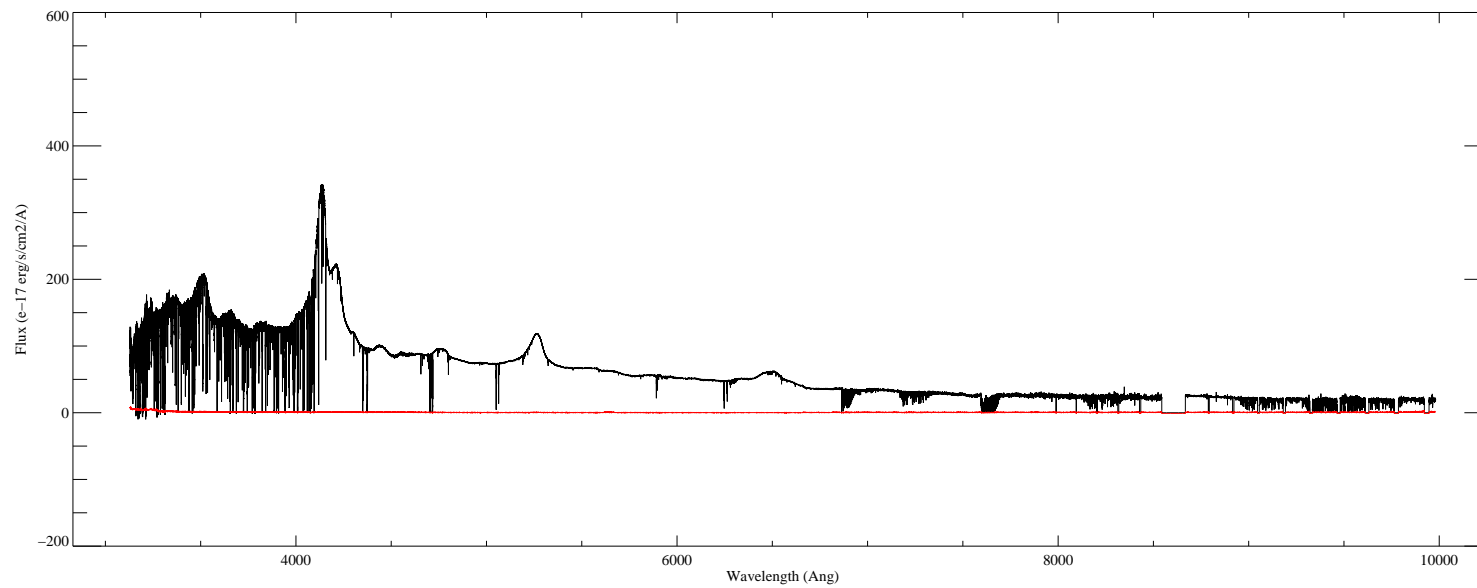*Figure 4.26:* J1344-1035

Figure 4.27: B0237-2322
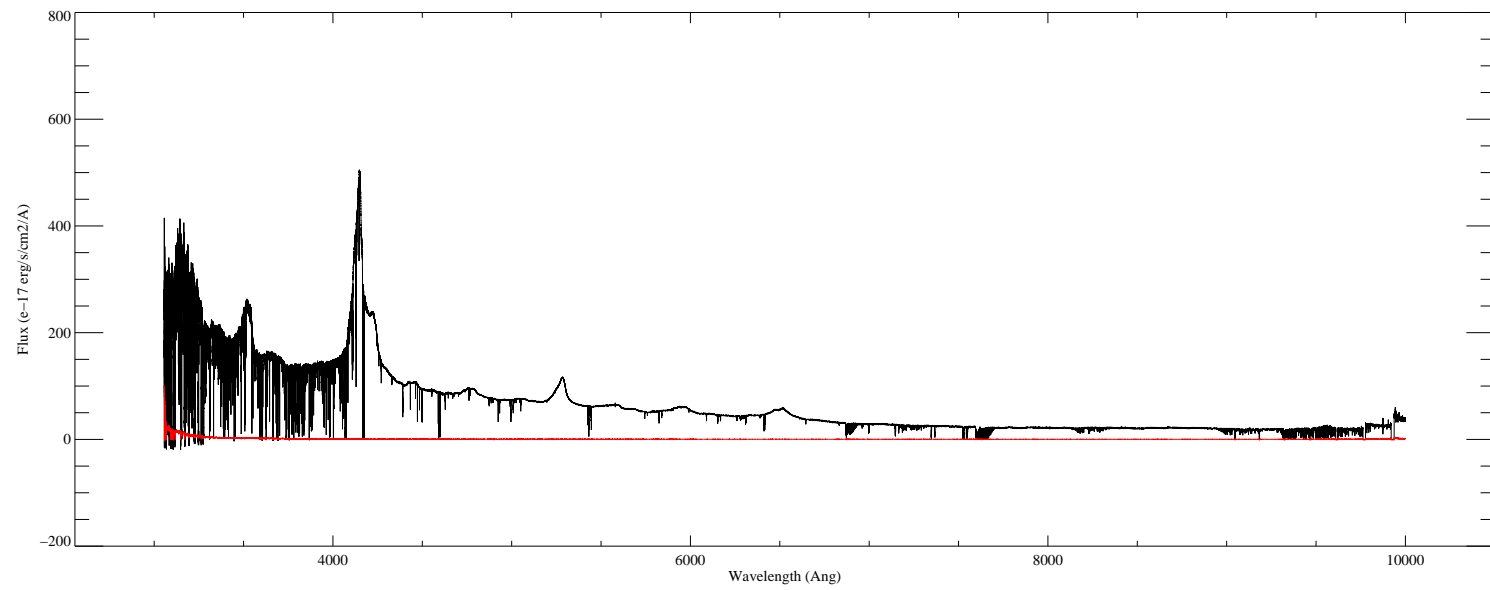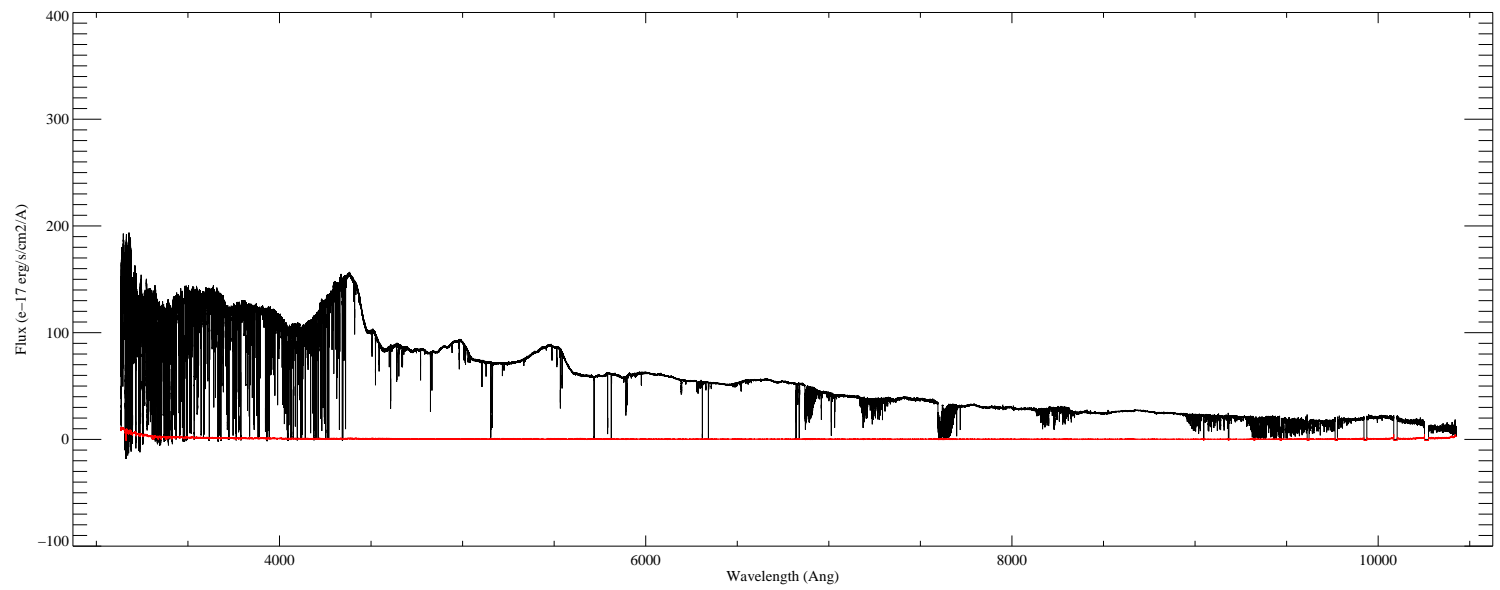
Figure 4.28: B1122-168
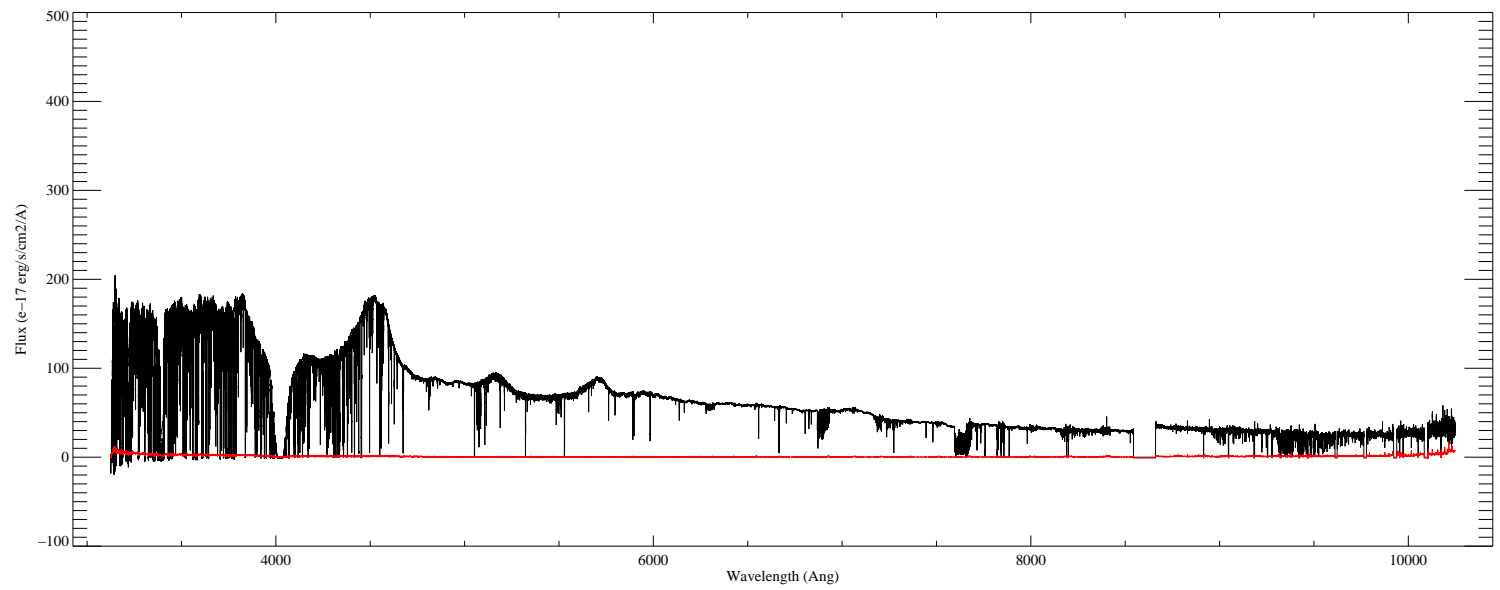
Figure 4.29: J2220-2803
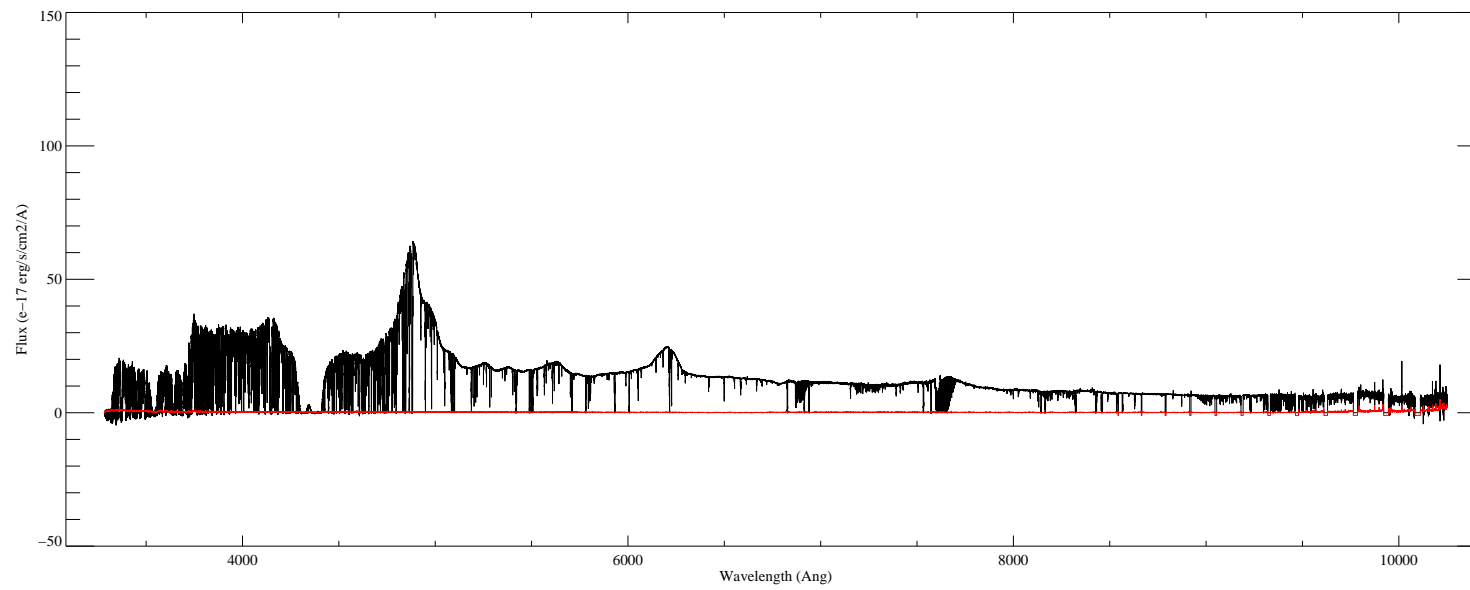
Figure 4.30: B1347-2457
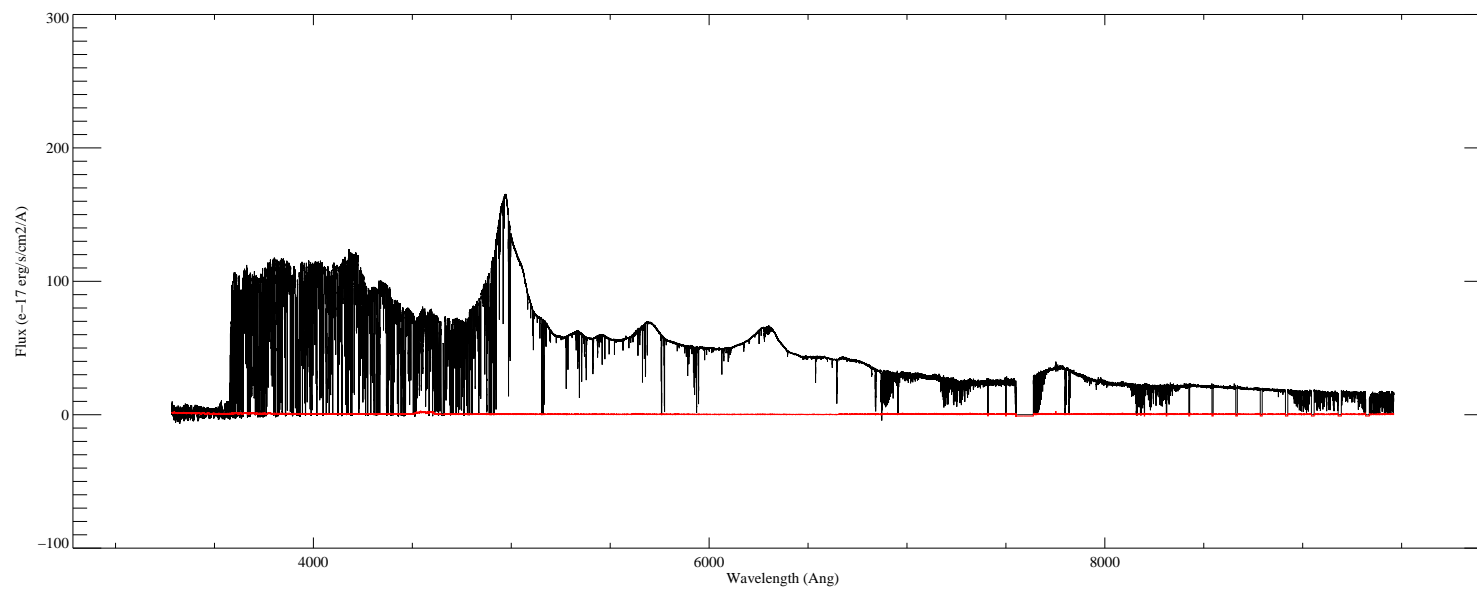
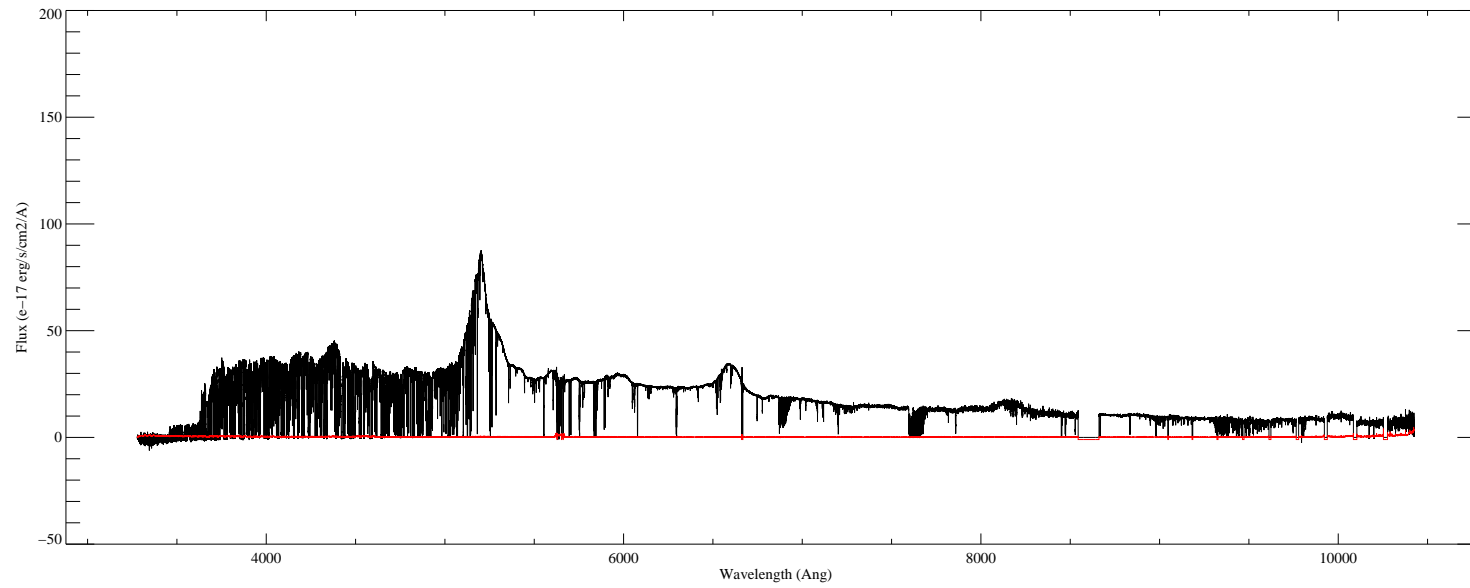*Figure 4.31:* J0103+1316

Figure 4.32: J0407-4410

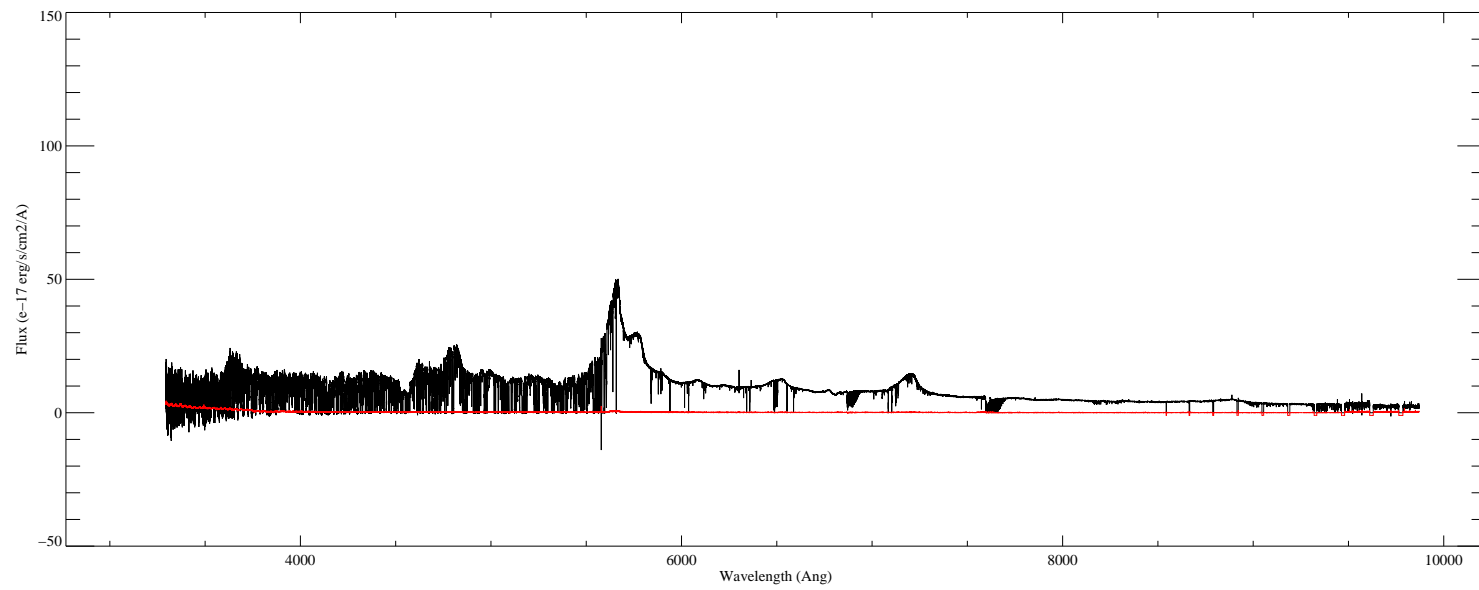*Figure 4.33:* HE0940

Figure 4.34: B2126-15
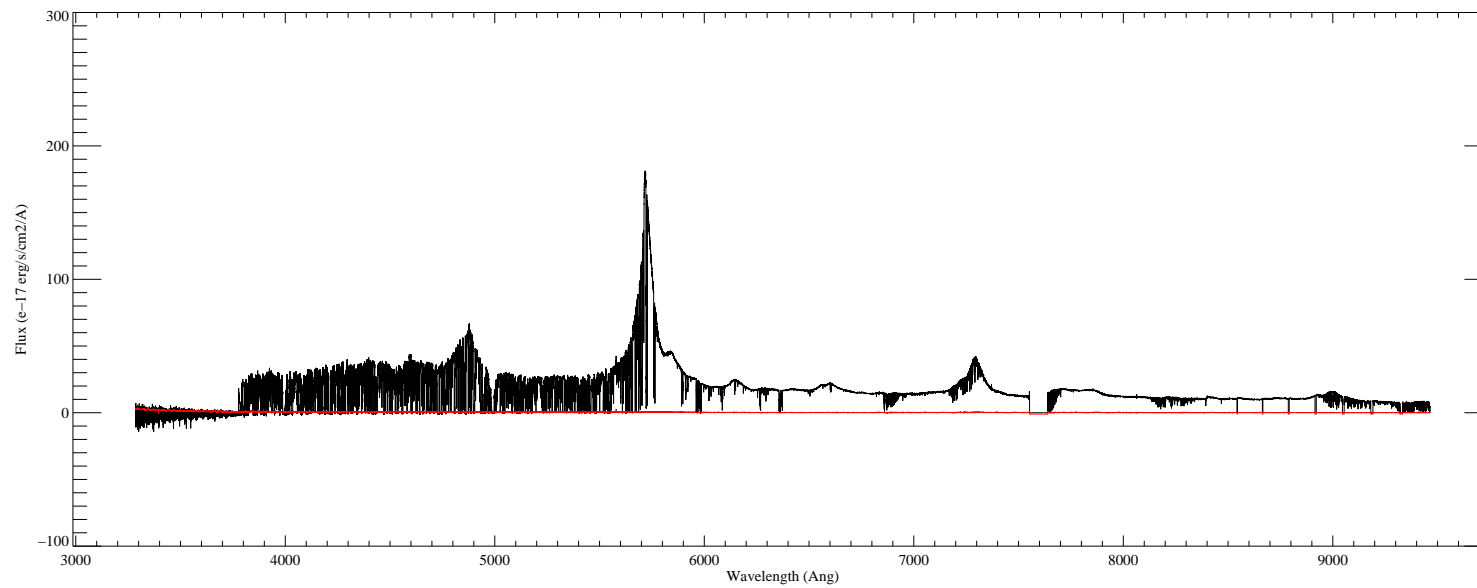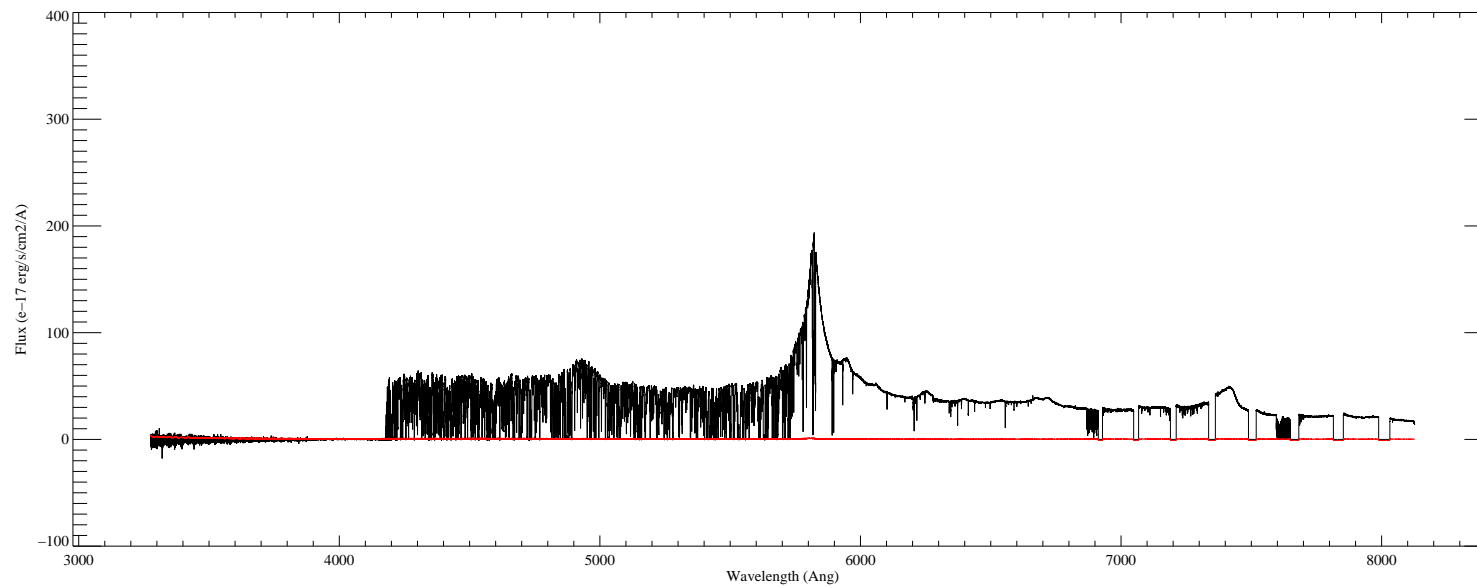
Figure 4.35: B0055-26

Figure 4.36: B1618-0035

Figure 4.37: B1937-1009

Bibliography

A. Aguirre et al. Metal enrichment of the intergalactic medium at z=3 by galactic winds. *The Astrophysical Journal*, Vol. 560:599–605, 2001.

L. Anderson et al. 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*, Vol. 441:24–62, 2014.

J. N. Bahcall and E. E. Salpeter. On the interaction of radiation from distant sources with the intervening medium. *Astrophysical Journal*, Vol. 142:1677, 1965.

R. Barkana and A. Loeb. The physics and early history of the intergalactic medium. *Reports on Progress in Physics*, Vol. 70:627, 2007.

R. A. Bernstein, S. M. Burles, and J. X. Prochaska. Data reduction with the mike spectrometer. *PASP*, Vol. 8444:9, 2015.

R. A. Bernstein et al. Mike: A double echelle spectrograph for the magellan telescopes at las campanas observatory. Proceedings of the SPIE, 2003.

A. Boksenberg, W. Sargent, and M. Rauch. Properties of qso metal line absorption systems at high redshifts: Nature and evolution of the absorbers and the ionizing radiation background. eprint arXiv:astro-ph/0307557, 2003.

J. N. Burchett et al. A deep search for faint galaxies associated with very low-redshift c iv absorbers: Ii. program design, absorption-line measurements, and absorber statistics. eprint arXiv:1510.01329, 2015.

R. Carswell and J. Webb. *VPFIT 10.0*, 2013.

R. F. Carswell and M. J. Rees. Constraints on voids at high redshifts from ly-alpha absorbers. *Monthly Notices of the Royal Astronomical Society*, Vol. 224:13–16, 1987.

R. F. Carswell and J. K. Webb. Vpfit: Voigt profile fitting program. Astrophysics Source Code Library, record ascl:1408.015, 2014.

J. Charlton and C. Churchill. Quasistellar objects: Intervening absorption lines. *Encyclopedia of Astronomy and Astrophysics*, Vol. 132:117–136, 2000.

C. W. Churchill. *Cosmological Absorption Line Spectroscopy*. Cambridge University Press: London, 2016.

B. Ciardi and A. Ferrara. The first cosmic structures and their effects. *Space Science Reviews*, Vol. 116:625, 2005.

R. J. Cooke, M. Pettini, et al. Precision measures of the primordial abundance of deuterium. *The Astrophysical Journal*, Vol. 781:16, 2014.

L. Cooksey et al. The last eight-billion years of intergalactic c iv evolution. *The Astrophysical Journal*, Vol. 708:868–908, 2010.

S. Cristiani, S. D'Odorico, et al. The space distribution of the lyman alpha clouds in the line of sight to the z=3.66 qso 0055-269. *Monthly Notices of the Royal Astronomical Society*, Vol. 273:1016–1032, 1995.

S. Cristiani, S. D'Odorico, et al. The clustering properties of the lyman alpha clouds. *Monthly Notices of the Royal Astronomical Society*, Vol. 285:209–217, 1997.

S. Cristiani et al. From CODEX to ESPRESSO to HIRES@E-ELT: a view on cosmology and fundamental physics from the IGM perspective . *MEMSAI*, 86:486, 2015.

R. A. C. Croft et al. Recovery of the power spectrum of mass fluctuations from observations of the lyalpha forest. *The Astrophysical Journal*, Vol. 495:44–62, 1998.

J. A. Cuesta et al. The clustering of galaxies in the sdss-iii baryon oscillation spectroscopic survey: baryon acoustic oscillations in the correlation

function of lowz and cmass galaxies in data release 12. *Monthly Notices of the Royal Astronomical Society*, Vol. 457:1770–1785, 2016.

G. Cupani et al. Data Analysis Software for the ESPRESSO Science Machine. In A. R. Taylor and E. Rosolowsky, editors, *Astronomical Data Analysis Software an Systems XXIV (ADASS XXIV)*, volume 495 of *Astronomical Society of the Pacific Conference Series*, page 289, September 2015a.

G. Cupani et al. Data treatment towards the ELT age. The ESPRESSO case. *MEMSAI*, 86:502, 2015b.

P. DeBernardis. Tecniche sperimentali in astrofisica, 2005.

V. D'Odorico et al. The rise of the civ mass density at z < 2.5. *Monthly Notices of the Royal Astronomical Society*, Vol. 401:2715–2721, 2010.

V. D'Odorico et al. Metals in the igm approaching the re-ionization epoch: results from x-shooter at the vlt. *MNRAS*, Vol. 435:1198–1232, 2013.

H. Dutton. *Understanding Optical Communications*. IBM Corporation, 1998.

S. L. Ellison et al. The enrichment history of the intergalactic medium-measuring the c iv/h i ratio in the lya forest. *The Astronomical Journal*, Vol. 120:1175–1191, 2000.

X. Fan et al. Constraining the evolution of the ionizing background and the epoch of reionization with z 6 quasars. ii. a sample of 19 quasars. *The Astronomical Journal*, Vol. 132:117–136, 2006.

A. Ferrara, M. Pettini, and Y. Shchekinov. Mixing metals in the early universe. *Monthly Notices of the Royal Astronomical Society*, Vol. 319: 539–548, 2000.

R. Fitzpatrick. *Oscillations and Waves*. CRC Press - Taylor & Francis Group, 2013.

A. Fontana and P. Ballester. Fitlyman: a midas tool for the analysis of absorption spectra. *The Messenger*, Vol. 80:37–41, 1995.

G. Gamow. The origin of elements and the separation of galaxies. *Physical Review*, Vol. 74:505, 1948.

J. E. Gunn and B. A. Peterson. On the density of neutral hydrogen in intergalactic space. *Astrophysical Journal*, Vol. 142:1633, 1965.

E. Holmberg. Magnitudes, colors, surface brightness, intensity distributions absolute luminosities, and diameters of galaxies. *University of Chicago Press (Stars and Stellar Systems)*, Vol. 9:123, 1975.

K. Horne. An optimal extraction algorithm for ccd spectroscopy. *Astronomical Society of the Pacific*, Vol. 98:609–617, 1986.

E. M. Hu et al. The distribution of column densities and b values in the lyman-alpha forest. *Astronomical Journal*, Vol. 110:1526, 1995.

G. Lemaitre. Contributions to a british association discussion on the evolution of the universe. *Nature*, Vol. 138:704–706, 1931.

C. R. Lynds. A sensitive test for the presence of atomic hydrogen in intergalactic space. *Nature*, Vol. 207:963, 1965.

A. McPherson et al. E-elt update of project and effect of change to 39m design. *Proceedings of the SPIE*, Vol. 8444:9, 2012.

H. Mo and F. van den Bosch. *Galaxy Formation and Evolution.* Cambridge University Press, 2010.

L. Monaco. *UV-Visual Echelle Spectrograph User manual.* ESO, 2014.

D. Morin. Interference and diffraction, 2010.

J. Mortlock et al. A luminous quasar at a redshift of z = 7.085. *Nature*, Vol. 474:616–619, 2011.

M. T. Murphy, P. Tzanavaris, J. K. Webb, and C. Lovis. Selection of thar lines for wavelength calibration of echelle spectra and implications for variations in the fine-structure constant. *MNRAS*, Vol. 378:221–230, 2007.

F. Murtagh and A. Heck. *Multivariate Data Analysis.* Springer-Verlag Berlin Heidelberg, 1987.

J. M. O'Meara et al. The first data release of the kodiaq survey. *The Astronomical Journal*, Vol. 150:11 pp., 2015.

F. Pepe et al. ESPRESSO - An Echelle SPectrograph for Rocky Exoplanets Search and Stable Spectroscopic Observations. *The Messenger*, 153:6–16, September 2013.

S. Perlmutter et al. Measurements of omega and lambda from 42 high redshift supernovae. *Astrophysical Journal*, Vol. 517:565, 1999.

M. Pettini et al. Metal abundances at z<1.5: Fresh clues to the chemical enrichment history of damped lyalpha systems. *The Astrophysical Journal*, Vol. 510:576–589, 1999.

M. Pettini et al. The c iv mass density of the universe at redshift 5. *The Astrophysical Journal*, Vol. 594:695–703, 2003.

C. Porciani and M. Giavalisco. The clustering properties of lyman break galaxies at redshift z 3. *The Astrophysical Journal*, Vol. 565:24–49, 2002.

J. Prochaska. The hires redux page. www.ucolick.org, 2006.

A. G. Riess. Observational evidence from supernovae for an accelerating universe and a cosmological constant. *Astronomical Journal*, Vol. 116: 1009, 1998.

B. Ryden. *Introduction to Cosmology*. Addison Wesley, 2002.

F. Saitta et al. Tracing the gas at redshift 1.7-3.5 with the Lyα forest: the FLO approach. *MNRAS*, 385:519–530, March 2008. doi: 10.1111/j. 1365-2966.2008.12860.x.

W. Sargent, P.J. Young, A. Boksenberg, and D. Tytler. The distribution of lyman-alpha absorption lines in the spectra of six qsos - evidence for an intergalactic origin. *Astrophysical Journal Supplement Series*, Vol. 42: 41–81, 1980.

P. A. G. Scheuer. A sensitive test for the presence of atomic hydrogen in intergalactic space. *Nature*, Vol. 207:963, 1965.

R. A. Simcoe, W. L. W. Sargent, and M. Rauch. The distribution of metallicity in the intergalactic medium at z 2.5: O vi and c iv absorption in the spectra of seven qsos. *The Astrophysical Journal*, Vol. 606:92–115, 2004.

R. A. Simcoe et al. Constraints on the universal c iv mass density at z 6 from early infrared spectra obtained with the magellan fire spectrograph. *The Astrophysical Journal*, Vol. 743:12 pp, 2011.

A. Songaila. The minimum universal metal density between redshifts of 1.5 and 5.5. *The Astrophysical Journal*, Vol. 561:L153–L156, 2001.

The optical design of the X-shooter for the VLT, volume 6269, 2006. SPIE– The International Society for Optical Engineering.

L. J. Storrie-Lombardi and A. M. Wolfe. Surveys for z>3 damped lyal-pha absorption systems: The evolution of neutral gas. *The Astrophysical Journal*, Vol. 543:552–576, 2000.

L. J. Storrie-Lombardi, M. J. Irwinn, and R. G. McMahon. Apm z> 4 survey: distribution and evolution of high column density hi absorbers. *Monthly Notices of the Royal Astronomical Society*, Vol. 282:1330–1342, 1996a.

L. J. Storrie-Lombardi, M. J. Irwinn, and R. G. McMahon. Evolution of neutral gas at high redshift: implications for the epoch of galaxy formation. *Monthly Notices of the Royal Astronomical Society*, Vol. 283: L79–L83, 1996b.

G. Thomas and K. Stamnes. *Radiative Transfer in the Atmosphere and Ocean*. Cambridge University Press, 1999.

S. Vogt. *Hires User's Manual*. UCO/Lick Observatory, 1994.

S. Vogt. *Keck Telescope and Facility Instrument Guide*. UCO/Lick Observatory, 2002.

S. Vogt. *HIRES Instrument and Data Format*. UCO/Lick Observatory, 2006.

A. M. Wolfe et al. Damped lyman-alpha absorption by disk galaxies with large redshifts. i - the lick survey. *Astrophysical Journal Supplement Series*, Vol. 61:249–304, 1986.

F. M. Zerbi et al. HIRES: the high resolution spectrograph for the E-ELT. In *Ground-based and Airborne Instrumentation for Astronomy V*, volume

9147 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, page 914723, August 2014. doi: 10.1117/12.2055329.

F. Zwicky. Die rotverschiebung von extragalaktischen nebeln. *Helvetica Physica Acta*, Vol. 6:110, 1933.