4-14-2016 12:00 AM

Insights from Unsupervised Clustering and Composite Spectral Analysis Into the Physical Properties Driving Emission and Absorption in Quasar UV/Optical Spectra

Aycha Tammour, *The University of Western Ontario*

Supervisor: Sarah Gallagher, *The University of Western Ontario*

A thesis submitted in partial fulfillment of the requirements for the Doctor of Philosophy degree in Astronomy

Abstract

Although quasar spectra appear remarkably similar, subtle systematic differences are well-known and are linked to the physical conditions in the emitting regions. This thesis characterizes the properties of quasar optical and UV spectral features and reveals some of their underlying physical causes by analyzing large spectroscopic datasets from the Sloan Digital Sky Survey.

We first examine the narrow-line region using four optical emission lines with a range of ionization potentials (IPs) using a sample of 16,027 quasars with redshift $z \leq 0.75$. We bin the sample using the continuum luminosity and the Hβ width to constrain the accretion rates and black hole masses and create composite spectra from objects in each bin. The luminosities of the high IP lines are constant with redshift consistent with no evolution in the quasar spectral energy distribution (SED) or the host galaxy ISM illuminated by the continuum. In contrast, the [O II] line becomes stronger at high er redshifts pointing to an enhanced star formation contributing to the [O II] at higher redshifts.

In the UV, we explore using unsupervised clustering (specifically the K-means algorithm) in searching for structure in the parameter space of quasar spectral measurements. We use measurements of 4,110 objects from the Pâris et al. (2014) Catalog and perform K-means on the equivalent widths (EWs), and the red- and blue-half-width at half-maxima of Mg II, C IV, and C III]and use objects in each of the clusters to generate composite spectra. We find that quasars with larger C IV blueshifts and smaller EWs have softer ionizing radiation. This result is particularly interesting when seen using C III].

Next, we use a sample of 2,683 broad absorption line quasars from the Gibson et al. (2009) catalog. We apply K-means using the EWs and minimum and maximum outflow velocities of the C IV absorption trough and generate composite spectra from objects in each of the clusters. We find that objects with deep and broad C IV troughs are more likely to be variable and to show absorption from lower ionization transitions. Conversely, objects with shallower C IV troughs are less likely to be variable and have lower fractions of absorption from lower ionization lines. We examine this behaviour in light of properties such as SED hardness and intrinsic reddening.

Keywords: Catalogues, Active Galaxies: Quasars: Emission lines, Absorption lines, Quasar winds, Methods: Unsupervised Clustering

Co-Authorship Statement

The work presented in this thesis has greatly benefited from advice by my supervisor Dr. Sarah Gallagher and collaborators: Dr. Gordon Richards, Dr. Mark Daley, Dr. Nur Filiz Ak, and Dr. Karen Leighly. While I led the projects and performed the analysis, their suggestions and feedback were very helpful for the completion of this thesis and the publication of chapters 2 and 3.

Chapter 2: *Tracing Quasar Narrow-Line Regions Across Redshift: A Library of High S/N Optical Spectra*, has been published in the Monthly Notices of the Royal Astronomical Society, Volume 448, Issue 4, 2015.

Chapter 3: *Insights into Quasar UV Spectra Using Unsupervised Clustering Analysis*, has been accepted for publication in the Monthly Notices of the Royal Astronomical Society and will appear in Volume 459, Issue 2, 2016.

Chapter 4: *Constraining the Diversity of Broad-Absorption Troughs in BAL Quasars with Unsupervised Clustering Analysis*, will be submitted in the near future.

Acknowledgements

This work would not have been possible without the support of so many amazing people. I would like first to thank my supervisor, Dr. Sarah Gallagher, for her continuous support. The guidance, and encouragement I received from her during the past few years were essential for the completion of this thesis. Working with Sarah helped me become a better scientist and for this I am always grateful.

I am grateful to Dr. Gordon Richards for his thoughtful feedback that was invaluable for the work presented in this thesis and for the publication of chapters 2 and 3.

Many thanks to my thesis committee for providing useful ideas and constructive comments that greatly improved my thesis.

I would like also to thank the members of the physics and astronomy department at Western University (faculty, staff and fellow graduate students) for creating a friendly and supportive environment. During my years as a graduate student, I always had someone to go to for advice on issues related to graduate school or professional development.

The presence of many wonderful friends during the past few years in London has been a great blessing. I am particularly grateful to Seniha Yildiz and Iman Ibrahim and their families for the friendship and encouragement.

Despite the distance, my family –parents and siblings– has been a great source of motivation throughout the years I spent pursuing my post-graduate education. Their constant faith in me and unconditional love continue to be an inspiration.

Dedication

For the children of Syria.

Contents

List of Figures

List of Tables

List of Appendices

Chapter 1

Introduction

1.1 Overview

Active Galactic Nuclei (AGN) appear as bright compact cores in the centres of a small fraction of the "normal" galaxy population. Quasars are considered among the most luminous of AGNs with typical bolometric luminosities of $\sim 10^{46}$ erg s^{-1}. The enormous amounts of energy generated within these small physical scales (a few parsecs[1]), made them a set of intriguing objects since the early days of their discovery in the mid 20th century (Schmidt, 1963). Accretion onto a supermassive black hole (SMBH) is now accepted as the source of the overwhelming radiation produced by quasars.

The presence of a SMBH in the centres of most massive galaxies has become widely accepted (e.g., Silk & Rees, 1998; Magorrian et al., 1998). This presence was initially inferred by dynamical signatures such as the sharp increase in the central velocity dispersion in massive galaxies and the "clumpiness" in the surface brightness profiles (see Kormendy, 2004, for a good review). Magorrian et al. (1998) were the first to use dynamical modelling to show that the mass of the SMBH scales with its host galaxy bulge mass M_{bulge} which –for a fixed M/L ratio— scales with the bulge luminosity (L_{bulge}). This correlation takes the form $M_{BH} \propto L_{bulge}^{1.2}$. Furthermore, Ferrarese & Merritt (2000) and Ferrarese & Ford (2005) demonstrated that the black hole mass also scales with the bulge stellar velocity dispersion: $M_{BH} \propto \sigma_*^{4.8\pm0.5}$. This rather tight correlation of M_{BH} and σ_* (which traces the bulge mass) might point to a connec-

[1] 1 parsec= 3.086×10^{16} m

tion between the growth of the SMBH and the host galaxy through feedback processes such as winds or jets (e.g., Begelman, 2004, and references therein).

Though the $M_{BH} - L_{bulge}$ and $M_{BH} - \sigma_*$ relations were found for galaxies with no active nuclei, they are likely still valid in the cases of AGNs (e.g., Ferrarese et al., 2001). However, proving the applicability of the $M_{BH} - L_{bulge}$ and $M_{BH} - \sigma_*$ relations to AGNs is non-trivial. The technique usually used to estimate BH masses for dormant galaxies relies on the dynamics inferred from the stellar absorption features. This method not only requires high spatial resolution spectra (only available for nearby objects), but also fails in the case of an AGN as the strong continuum hides the stellar features. To overcome this issue, astronomers use a technique called Reverberation Mapping which relies on a scaling relation between the radius of the central region and the continuum luminosity (Peterson, 1993). Reverberation mapping is discussed in further detail in §1.2.2.

One of the early studies of the $M_{BH} - L_{bulge}$ and $M_{BH} - \sigma_*$ in AGN done by Ferrarese et al. (2001) used a sample of 6 local AGN with BH masses determined with reverberation mapping. This sample turned out to be in very good agreement with the $M_{BH} - L_{bulge}$ and $M_{BH} - \sigma_*$ of "normal" galaxies with a slope of 5.3 ± 1.6 vis-à-vis 4.81 ± 0.55 for the galaxies with no AGN, however with a considerable scaling factor needed to match the normalization factors of the two samples (see the discussion in §1.5). The validity of the $M_{BH} - \sigma_*$ relation for AGN host galaxies supports the notion that an AGN is really a transient phase of growth that *all* supermassive black holes go through.

Indeed, it was suggested by Sołtan (1982) that quasars represent a phase that is common in most massive galaxies rather than being an anomalous stable phenomenon in the universe. The Soltan argument is based on the idea that if BHs grow through accreting material during the quasar phase, then the total energy emitted by quasars in a certain volume during their lifetime and the mass density of BHs in local galaxies should be comparable. The mass accumulated as a result of accretion during the quasar period T is given by:

$$M = \frac{1}{\eta c^2} L_{bol}\, T, \tag{1.1}$$

where η is the efficiency of converting mass into energy. Using the $M_{BH} - \sigma_*$ relation, Yu & Tremaine (2002) estimated the BH mass density in local galaxies and showed that this density

matches with the energy product of quasars estimated from their luminosity functions across redshift assuming efficiency $\eta = 0.1$. They interpreted this result as evidence that BHs in massive galaxies primarily grow during the quasar phase as material is being accreted onto them at high rates (Sołtan, 1982; Yu & Tremaine, 2002; Marconi et al., 2004).

1.2 The AGN Paradigm

1.2.1 Quasar Spectra

A typical optical quasar spectrum has strong emission lines with a large range of ionization potentials (e.g. IP=13.6 eV for Balmer series lines and IP= 126 eV for [Ne V]λ3427), and a strong non-thermal continuum that usually becomes bluer with higher luminosities (e.g., Vanden Berk et al., 2001). The emission lines are of two kinds: the broad lines with velocity widths of a few thousands of km s^{-1}, and the narrow lines with widths of a few hundreds of km s^{-1} (see Fig 1.1). Transitions that are permitted in atomic physics result in broad and narrow emission lines while forbidden transitions –which violate the dipole selection rules– produce narrow emission lines. We expand on the physical conditions of the line-emitting regions in §1.2.2.

Spectral Energy Distributions (SEDs) describe the broad-band continuum spectrum as power versus frequency or wavelength (νf_ν *vs.* ν or λf_λ *vs.* λ) and are considered major tools for understanding the energy output of astronomical objects. A quasar SED is remarkably different from that of a galaxy without an AGN in that it has considerable power over several orders of magnitudes of frequency from X-ray to radio, while a normal galaxy SED would have most of its power emitted in a much smaller range of frequency. A quasar's SED not only covers a broad range of frequencies but also appears to be roughly constant across most of the energy range indicating almost equal contributions from the various energy generation processes (see Fig. 1.2; Elvis et al., 1994; Richards et al., 2006a).

Within a restricted frequency range (e.g., optical/UV), quasar SEDs can often be described with a power law of the form:

$$f_\nu \propto \nu^\alpha, \tag{1.2}$$

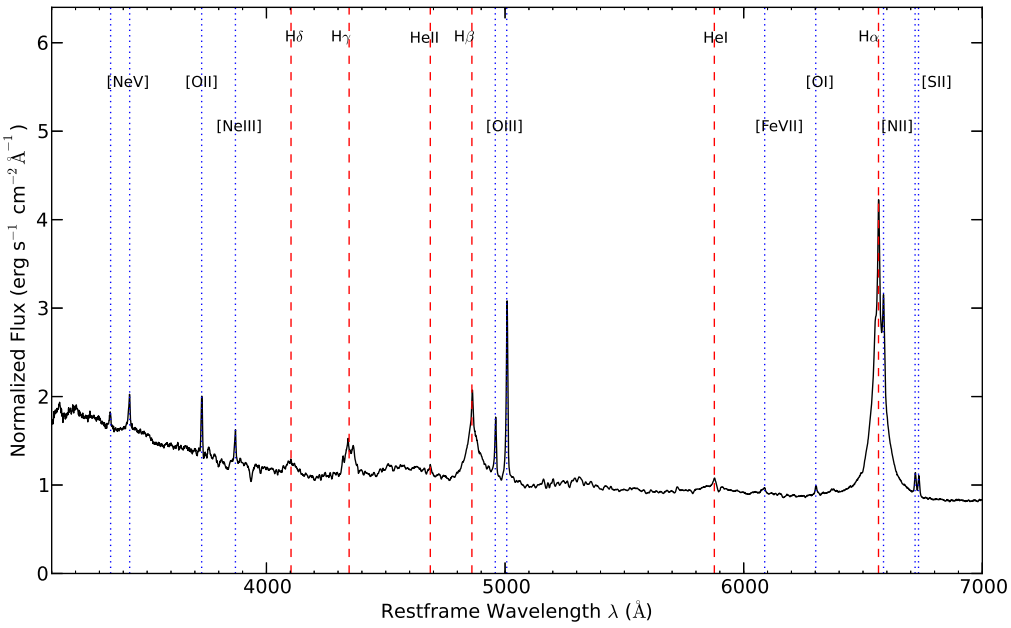

Figure 1.1: An example of an optical quasar spectrum. The blue dotted lines mark the forbidden (narrow) emission lines and the red dashed lines mark the permitted (broad) emission lines. The spectrum is a median composite of 70 quasars from the Quasar Catalogue of the SDSS-DR7 (Schneider et al., 2010).

where α is the spectral index and can take values of $\sim$ 0 and -2 (e.g., Richards et al., 2006a). However, quasar SEDs are far from simple; prominent features include a UV/optical bump –usually referred to as the Big Blue Bump or BBB, and the IR bump (see Fig. 1.2; Richards et al., 2006a). The presence of structure in quasar SEDs implies that energy generation is a result of a combination of thermal (blackbody physics) and non-thermal processes (such as inverse Compton scattering and synchrotron radiation).

Models that are most successful in reproducing the observed SED assume a Keplerian accretion disk that is optically thick and geometrically thin; i.e., $H << R$, where H is the scale height from the plane of the accretion disk and R is the radial distance from the SMBH (Shakura & Sunyaev, 1973; Longair, 1994; Blaes, 2007). The BBB is thought to be of thermal origin and is a result of the emission emerging directly from the accretion disk. Approximating the emission from the accretion disk by multi-temperature blackbody radiation results in a broad

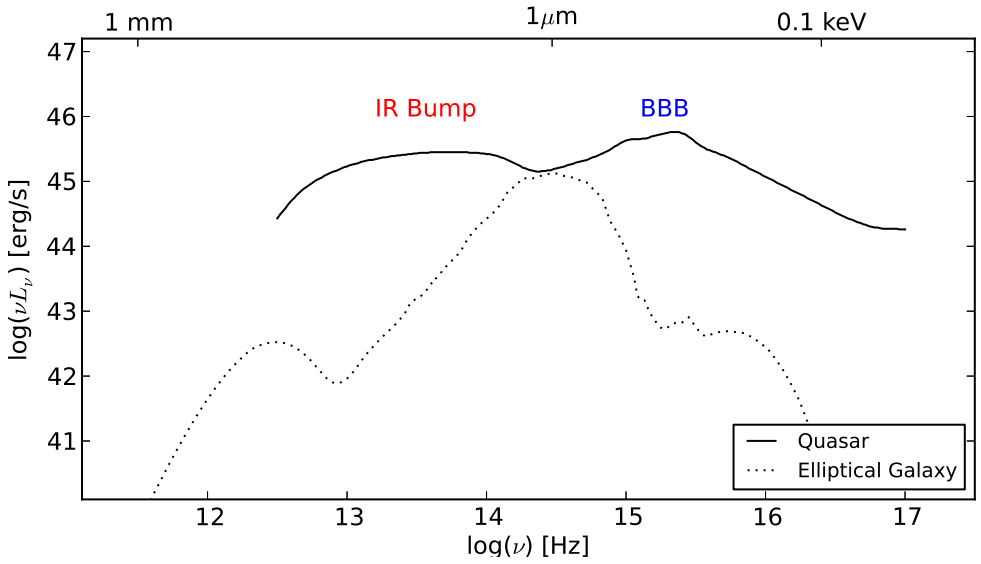

Figure 1.2: Radio quiet quasar mean SED from Richards et al. (2006a). Energy is emitted at a wide range of frequencies. The SED displays structure –the BBB at $\sim 10^{15.5}$ Hz and the IR bump shortward of 10^{14} Hz. For comparison, we plot a scaled elliptical galaxy SED from Silva et al. (1998).

feature that peaks at frequency $\sim 10^{15}$ Hz in agreement with the observed BBB (Peterson, 1997).

Another notable feature in quasar SEDs is the "smaller" near-IR bump. This enhancement around $3 - 5$ μm is also thought to be of thermal origin. The strongest candidate to produce this feature is dust which is heated by emission from the central source and is re-emitting in the IR. Emission at $\sim 3 - 5$ μm is associated with a temperature of ~ 1500 K, suitable for the presence of dust (especially graphite) which sublimates at higher temperatures (e.g. Nenkova et al., 2002).

Accretion disk models, however, cannot account for the high energy ionizing radiation in the X-ray. This radiation is assumed to originate from a hot ($\sim 10^9$ K) optically thin corona located right above the accretion disk (Longair, 1994). Near-IR and optical photons emitted by the accretion disk inverse Compton-scatter off this hot plasma which boosts them to X-ray energies. A useful quantity for characterizing accretion efficiency is the Eddington luminosity: the limit which marks the balance between the radiation pressure from electron (Thomson)

scattering and gravitational force (Longair, 1994; Peterson, 1997):

$$F_{rad} = F_{grav}. \tag{1.3}$$

For a single electron:

$$\sigma_e \frac{L_{\rm Edd}}{4\pi c r^2} = \frac{GMm_p}{r^2},$$

where σ_e is the Thomson scattering cross-section ($\sigma_e = 6.65 \times 10^{-25}\mathrm{cm}^2$). The Eddington luminosity is therefore:

$$\begin{aligned} L_{\rm Edd} &= \frac{4\pi G c m_p}{\sigma_e} M \\ L_{\rm Edd} &\simeq 1.5 \times 10^{38} (M/M_\odot)\ \mathrm{erg\ s^{-1}}. \end{aligned} \tag{1.4}$$

We can define the Eddington accretion rate by combining this last result with equation 1.1 as:

$$\dot{M}_{\rm Edd} = \frac{L_{\rm Edd}}{\eta c^2}, \tag{1.5}$$

which is just the accretion rate needed to generate Eddington luminosity. A useful quantity to describe the level of accretion is the relative accretion rate:

$$\frac{\dot{m}}{\dot{M}_{\rm Edd}} \propto \frac{L}{L_{\rm Edd}} \tag{1.6}$$

Accretion disks working in the sub-Eddington domain (i.e., $L/L_{\rm Edd} \leq 0.3$) are thought to be compatible with the thin accretion model (Longair, 1994; Peterson, 1997).

1.2.2 AGN Structure

Observational properties of quasars are generally explained in terms of a model that includes five major components: a supermassive black hole (SMBH), an accretion disk, a dusty torus, a broad-line region (BLR), and a narrow-line region (NLR) (Peterson, 1997; Krolik, 1999). This so-called "Unification Model" seems to offer feasible explanations for many of the observed AGN phenomena such as the presence or absence of broad and narrow emission lines in the spectra of different classes of AGN (Antonucci, 1993). Figure 1.3 is a schematic view of

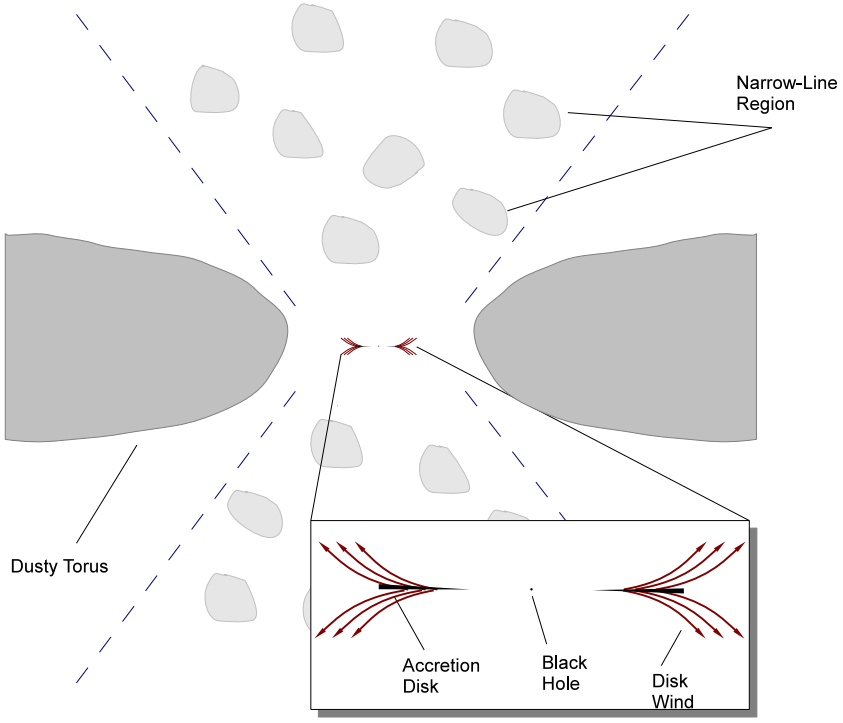

Figure 1.3: AGN structure. Matter is being accreted into a SMBH through an accretion disk. The broad lines are believed to be created in the disk wind which is powered by radiation pressure, (e.g., Murray et al., 1995), and the narrow lines are emitted in the narrow-line region –the diffuse ISM of the host galaxy photoionized by the AGN. The figure also depicts the large dusty structure known as the dusty torus.

our current understanding of the structure of AGN. According to this model, all AGN are intrinsically similar and orientation (i.e., our viewing point) *alone* is sufficient to account for the observed differences among AGN spectra. Despite the many successes the unification model has achieved in explaining the various AGN observations, it still lacks strong physical grounds that justifies all its components such as the presence of dense gas clouds (the source of the BLR emission) in the close vicinity of the central region (see below for more details on this region).

The SMBH with a mass of $10^6 - 10^9 M_\odot$ in the central region accretes material through the accretion disk. Direct imaging of this extremely small region is not possible but other observations and theoretical modelling support accretion onto the SMBH as the main source of

the energy generated by an AGN. The exact details of how the gas is being funnelled into the central region is not well known but an accretion disk seems to offer the best agreement with observations. Through this disk, viscosity causes matter to lose its angular momentum while it swirls down towards the black hole and as it gets closer and closer to the BH it releases more of its gravitational potential energy as radiation:

$$L_{acc} = \frac{dU}{dt} = \frac{GM}{r}\frac{dm}{dt} = \frac{GM\dot{m}}{r} \tag{1.7}$$

As mentioned previously, the power emitted in this process can be approximated as $L = \eta\dot{m}c^2$ where $\dot{m}$ is the mass accretion rate and η is the efficiency of converting mass into energy and is usually taken to be 0.1 (Yu & Tremaine, 2002). Though the value of η is really unknown, an order of magnitude can be inferred by estimating the gravitational potential energy released from a BH (the innermost stable orbit of a non-rotating BH) when a mass m falls from infinity to a distance of $3R_S$, where R_S is the Schwarzschild radius which defines the limit from which light cannot escape from a BH: $R_S = 2GM/c^2$ (Longair, 1994; Peterson, 1997):

$$\Delta U = \frac{GMm}{3R_S} = \frac{GMm}{6GM/c^2} \simeq 0.1mc^2 \tag{1.8}$$

Broad emission lines seen in the optical spectra of quasars are emitted by gas located in the close vicinity of the central source (within $\sim$ 1 pc). This gas is photoionized by radiation from the central source. The lack of broad forbidden lines (such as [O III]λ5007) indicates that the gas density of the BLR is high enough to suppress those forbidden transitions and provides a lower limit for electron density of $n_e \gtrsim 3 \times 10^6$ cm^{-3}. Calculations of line flux ratios (such as HeIλ5876 to HeIλ10830) give $\simeq 10^9$ cm^{-3} for upper limits (Krolik, 1999, chapter 10). The large widths of those emission lines (thousands to tens of thousands km s^{-1}) cannot be attributed to thermal motions alone but are caused by Doppler broadening by the gravitational effects of the BH on the gas. For example, line flux ratios give an average value of 10^4 K for the BLR temperature (Peterson, 1997; Krolik, 1999). But using $mv^2 = 3kT$ results in a line width of $\sim$ 10 km s^{-1} for Hydrogen! Needless to say, this width is not what we measure for emission lines in AGN spectra, and so other mechanisms must be contributing to the measured broad emission line widths.

An important feature of quasar broad lines is variability which is observed as a change in the line fluxes in response to changes in the continuum. Because of the high densities and short

recombination time scales, the time scale is set by the light travel-time: the time needed for the radiation emitted by the central source to reach the BLR (Peterson, 1997). Variability in quasars is known to be non-periodic i.e., it does not have a fixed timescale. Its amplitude is also observed to increase in the shorter wavelength part of the spectrum which makes it easier to study quasar variability in the UV and X-ray than in optical (Krolik, 1999).

AGN variability is the basis for BH mass determinations with a method called reverberation mapping. This technique assumes that the broad lines respond to the variations in the continuum radiation with a time delay $\tau = R_{BLR}/c$. The time delay τ is used to determine the size of the BLR and consequently the mass of the black hole (assuming virialized Keplerian motions) (Peterson, 1993; Kaspi et al., 2005):

$$M_{BH} = f\frac{\Delta V^2 R_{BLR}}{G} \tag{1.9}$$

Where G is the gravitational constant, and the factor f reflects our ignorance about the geometry of the BLR; it was found to be of the order of ~ 5 by matching the BH masses obtained by reverberation mapping with the $M_{BH} - \sigma_*$ for quiescent galaxies (Woo et al., 2010). ΔV is the line profile width measured as the FWHM of the broad emission line. Interestingly, reverberation mapping showed that broad lines with higher ionization potentials seem to have shorter time delays than broad lines with lower ionization potential which provides strong evidence for the stratification of the BLR, i.e., higher ionization lines are emitted at smaller scales than the lower ionization lines. In addition, reverberation mapping has revealed a strong correlation between the size of the BLR and the continuum luminosity: $R_{BLR} \propto L_{AGN}^{1/2}$ (e.g., Kaspi et al., 2005). This correlation allows us to use optical/UV spectra to determine the BH masses in objects where reverberation mapping data are not available, so-called "single epoch" BH masses. The C IV, Mg II, and Hβ lines and the continuum luminosity at 5100 Å are commonly used for BH mass estimations (e.g., Shen et al., 2011).

One of the most well-known correlations found between the broad lines and the continuum is the "Baldwin effect" (named after Baldwin 1977). The Baldwin effect is observed as an inverse correlation between the equivalent width[2] of the CIV line and the continuum luminosity. Fig. 1.4 shows the Baldwin effect from the original Baldwin (1977) paper and a more recent

[2]The equivalent width of a line is defined as: $EW_\lambda = \int \frac{f_\lambda - f_{cont}}{f_{cont}} d\lambda$

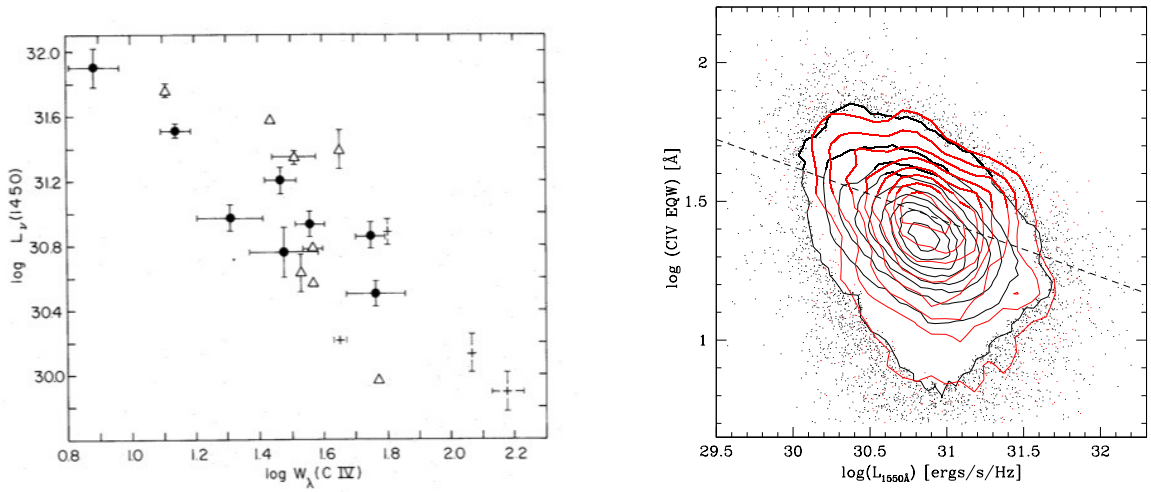

Figure 1.4: The Baldwin Effect: equivalent width of broad lines (C IV here) is anticorrelated with the continuum luminosity i.e, lower luminosity quasars have relatively stronger emission lines. Left from the original paper of Baldwin (1977), right from Richards et al. (2011). There is a considerably large scatter when a larger SDSS sample is used.

study by Richards et al. (2011). A correlation such as the Baldwin effect holds the potential for being used as a standard candle, but unfortunately the large scatter found in this correlation undermines its usefulness for such purposes. Other broad lines (such as Lyα) show a Baldwin effect but with varying slopes and strengths of the correlation (Richards et al., 2011). Interestingly, variations in the Baldwin effect of different lines are not random; higher ionization lines seem to have preferentially stronger dependence on continuum luminosity (i.e., larger slopes Shields, 2007, and references therein).

A physical account of the Baldwin effect has not been fully constructed. Variations in the strength of the Baldwin effect in different broad lines eliminates the possibility of orientation as the sole cause of the effect as we would expect different broad lines to be equally affected by the inclination of the illuminating source. Another perhaps more robust explanation uses the known correlation between the shape of the ionizing continuum and the luminosity in a sense that the ionizing continuum becomes softer (more power on lower energies) with increasing continuum luminosity which can potentially lead to weaker emission lines in more luminous objects –the Baldwin effect (Netzer et al., 1992; Shields, 2007).

Broad lines usually display structures –profile asymmetries– which are thought to be signatures of gas motions such as outflows. They are also known to be blueshifted from the systemic redshift measured from lower ionization lines such as Mg II. The measured blueshifts of the broad lines are found to be higher in high ionization lines than in low ionization lines; this correlation can be explained in terms of a stratified BLR where higher ionization lines are emitted closer to the central source than lower ionization lines (e.g. Richards et al., 2011).

The NLR –in contrast to the BLR– has a lower gas density. The lines emitted in this region are known to be "forbidden", i.e. do not obey at least one of the electric dipole selection rules which govern the changes in the quantum numbers during a radiative transition. These rules require the following (Draine, 2011):

1. Parity must change.
2. $\Delta L = 0, \pm 1$.
3. $\Delta J = 0, \pm 1$, but $J = 0 \rightarrow 0$ is forbidden.
4. Only one single-electron wave function changes, $\Delta l = \pm 1$.
5. $\Delta S = 0$.

Using line ratios, it is possible to infer some of the physical conditions in the NLR gas such as the electron densities and temperatures (e.g., Osterbrock, 1989). Determinations of the electron temperature can be done using a pair of lines that arise from the same ion (in order to avoid complications of abundance) but with different excitation potentials. This means that the population rates of these two levels is primarily determined by temperature and hence the ratio of the line intensities will reflect this temperature. The [N II] $\lambda\lambda$5755, 6548, 6583 and [O III] $\lambda\lambda$4363, 4959, 5007 lines are commonly used for electron temperature measurements (e.g., Osterbrock, 1989; Draine, 2011). For instance, using three [O III] lines (Fig. 1.2.2), Osterbrock (1989) derives the following empirical relation (his equation 5.4) which allows us to calculate T_e:

$$\frac{j_{\lambda 4959} + j_{\lambda 5007}}{j_{\lambda 4363}} = \frac{7.90 \exp(3.29 \times 10^4 / T_e)}{1 + 4.5 \times 10^{-4}\, n_e\, / T_e^{0.5}} \tag{1.10}$$

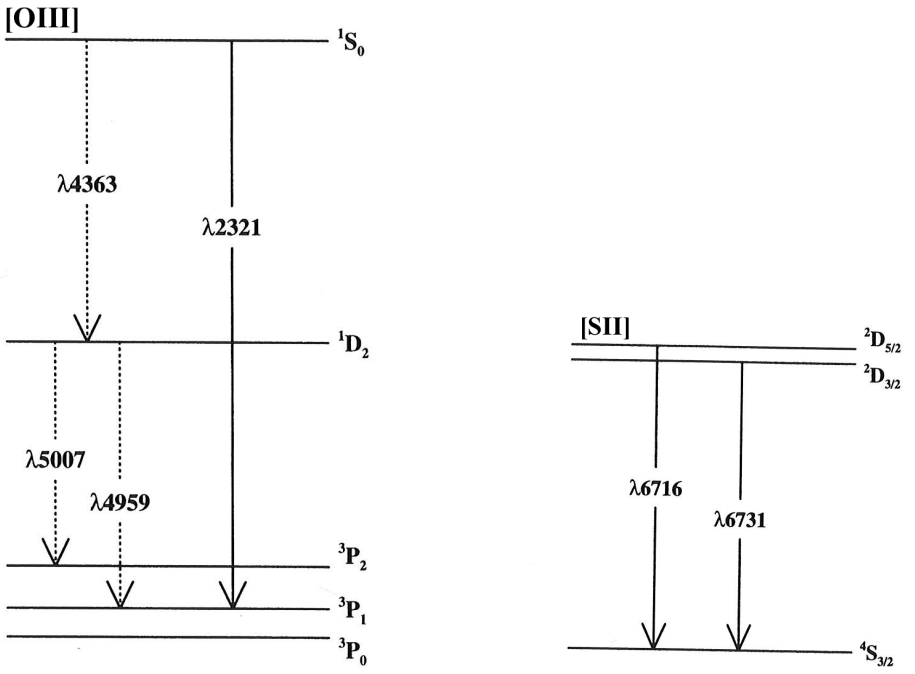

Figure 1.5: Energy levels for [O II] (left) and [S II] (right). Figure is from Peterson (1997).

For the NLR such calculations reveal average temperatures of the order of $\sim 10^4$ K (e.g. Osterbrock, 1989; Peterson, 1997; Krolik, 1999).

For calculations of electron densities we would need to use lines from the same ion that are formed due to transitions from two states with almost equal excitation energy and therefore their population rates are temperature independent. The [S II] $\lambda\lambda 6716, 6731$ and [O II] $\lambda\lambda 3726, 3729$ doublets are commonly used pairs for this purpose (see Fig 1.2.2) In the case of low density gas, line intensity is proportional to n^2 (n is the gas density) where radiative de-excitation dominates. It is proportional to n in the high density gas where de-excitation is mostly collisional. In between these two extremes, the intensity ratio is a strong function of electron density (e.g. Osterbrock, 1989, Fig. 1.6). Using such calculations gives a range of $10^2 - 10^4$ cm^{-3} for the NLR density (Peterson, 1997; Zhang et al., 2013a). One important thing to keep in mind when looking at such calculations is the fact that each line pair can result in different values for T_e and n_e. This is mainly due to the fact that different lines result from different "cloud" regions within the NLR.

Photoionization by radiation from the central engine is thought to be the main driver of line

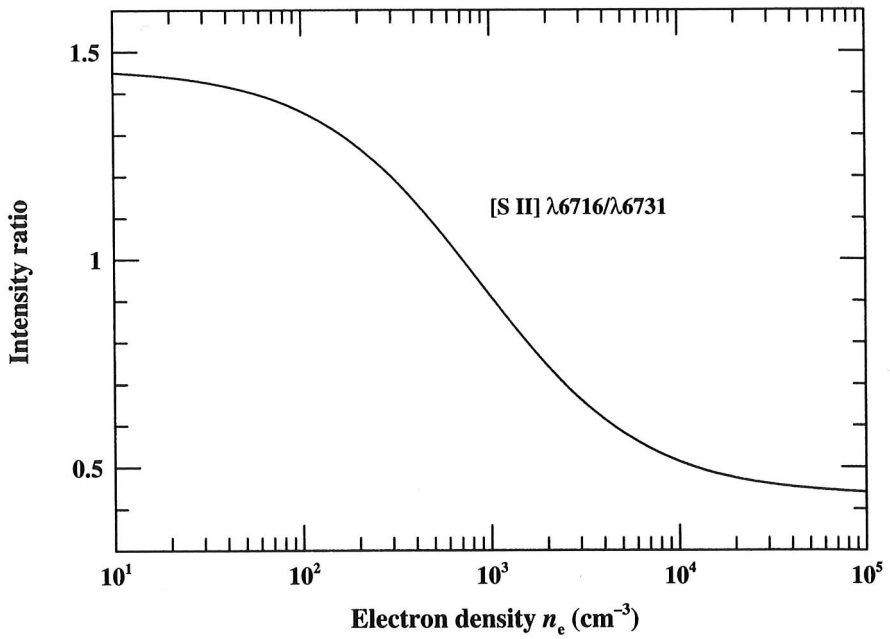

Figure 1.6: Determining the critical density n_e from the line ratio of [S II] $\lambda\lambda 6716, 6731$. The ratio is almost constant for n_e below 10^2 cm^{-3} and above 10^4 cm^{-3}. This line ratio would be a good diagnostic for density in between these two limits. Figure is from Peterson (1997).

production detected as broad and narrow lines in quasar spectra. This can be inferred from the simple fact that the energy from collisional ionization (associated with the gas temperature inferred from line ratios) is insufficient to produce some of the observed high ionization transitions (such as [Ne V] with IP= 97 eV). Other lines of evidence that support this conclusion include the observed correlated variability between the continuum and the strengths of the broad lines (with some time delay caused by light travel time). In addition, the wide range of photon energies associated with radiation from the central source can in principle account for the presence of both low and high ionization transitions in quasar spectra (Peterson, 1997; Krolik, 1999).

Line formation in quasars is sensitive to several properties such as the shape of the ionizing SED, distance from the central region, gas density, temperature, and element abundance. The ionization parameter U:

$$U = \Phi/nc, \tag{1.11}$$

is one of the key parameters used to define ionization levels in terms of the ratio between the

ionizing flux (Φ, photon cm^{-2} s^{-1}) and the gas density (n, cm^{-3}), i.e., photoionization rates to recombination rates.

Quasar photoionization models that can reproduce observed broad lines most successfully assumes that a range of density and ionization parameters exists in the BLR gas surrounding the central engine, and clouds with the appropriate values of density (which decreases with radius) and ionization parameter will emit a specific line optimally. This means that for a given line (transition), gas that is too close to the central source might be over-ionized or too dense to allow for this transition to happen, on the other hand at farther distances photons might not have enough energy to remove an extra electron and generate this transition. This is known as the locally optimally emitting cloud (LOC) model which proposes that a given line is emitted predominantly from gas in a particular portion of the BLR density-ionization parameter space that optimizes the production of this line (e.g., Baldwin et al., 1995; Krolik, 1999).

1.3 Quasar Winds and Broad Absorption-Line Quasars

Outflowing gas (in the form of equatorial winds) has been proposed to resolve some of the standing issues concerning the so-called BLR clouds confinement problem and to justify differences in velocity shifts between high and low IP emission lines seen in quasar spectra (e.g., Collin-Souffrin et al., 1988). Indeed, emission lines (specifically in high IP lines such as C IV) blueshifted from the systemic redshift (estimated from lower IP lines such as Mg II) are common in quasar spectra (e.g., Gaskell, 1982; Richards et al., 2002). Furthermore, the presence of outflows in quasars may be necessary to maintain accretion onto the SMBH by enabling the gas to rid itself of angular momentum. In addition, outflows can potentially be the means by which a SMBH and its host galaxy are linked as mentioned in the discussion of the $M_{BH} - L_{bulge}$ and $M_{BH} - \sigma_*$ relations (§1.1).

Developed theoretical models of quasar winds describe an outflow that emerges from the accretion disk (at distances of $\sim 10^{16}$ cm for a black hole of mass 10^8 $M_\odot$) and is accelerated outwards by radiation pressure (e.g., Murray et al., 1995; Proga et al., 2000). A layer of "shielding" gas between the wind and the central X-ray source prevents the wind from becoming over-ionized by filtering the high energy radiation thus allowing UV photons to accelerate

the wind to velocities up to $\sim 0.1c$ (Leighly, 2004; Leighly & Moore, 2004). The idea of a self-shielding gas is also consistent with correlations found between the EW and blueshift of the C IV and the hardness of the ionizing SED (probed by the strength of the He II emission line or by the optical-to-X-ray spectral index, α_{ox}[3]); these correlations can be seen as quasars with softer ionizing SEDs (weaker He II and $\alpha_{ox} \leq -1.8$) have weaker C IV emission lines (smaller EWs) and are more blueshifted (Wu et al., 2009; Richards et al., 2011). As such, the shape of the ionizing SED can have a significant impact on the contribution of the winds to the broad emission lines. This is also seen as C IV emission lines which appear as narrow symmetric profiles that are centered at systemic redshift in quasars with lower L_{UV} to L_{X-ray} ratios, while quasars with asymmetric blueshifted C IV profiles have higher L_{UV} to L_{X-ray} ratios.

More remarkable signatures of quasar winds are manifested in the broad (e.g., Weymann et al., 1991) or narrow (e.g., Ganguly et al., 2001) intrinsic absorption features that are often blueshifted from systemic redshift. By definition, quasars with absorption troughs (in C IV) that are broader than 2000 km s^{-1} at 10% below the continuum level are classified as broad absorption line quasars (Weymann et al., 1991). BAL quasars constitute a fraction of nearly 20% of optically selected quasar samples (e.g., Hewett & Foltz, 2003; Trump et al., 2006; Ganguly et al., 2007; Gibson et al., 2009). Intrinsic absorption in quasars is seen in many species with range of IPs; HiBAL quasars are those with absorption in high ionization potential (IP) transitions *only* such as C IV 1449 Å (IP= 47.9 eV) or Si IV 1394 Å (IP=33.5 eV), while LoBAL quasars are those with absorption from lower IP transitions such as Al III 1857Å (IP=18.8 eV) and/or Mg II 2800Å (IP= 7.6 eV) in addition to the high IP transitions.

Furthermore, measurements of α_{ox} indicates that BAL quasars are X-ray weak compared to non-BAL quasars (e.g., Gallagher et al., 2006). These observations are in agreement with the wind model and the presence of a layer of shielding gas that filters X-ray photons and allows UV photons to accelerate the wind. In addition BAL quasars are found to be more reddened that non-BALs, and LoBALs are significantly more reddened than HiBALs (e.g., Weymann et al., 1991; Reichard et al., 2003b). This is consistent with seeing BAL quasars through a line-

[3]Defined as $\alpha_{ox} = 0.384 \log(f_{2\mathrm{keV}}/f_{2500\ \text{Å}})$, this quantity measures the slope of the flux densities at rest-frame 2 keV and 2500 Å. The factor 0.384 is the logarithm of the ratio of the frequencies at which the flux densities are measured.

of-sight at a low inclination angle closer to the accretion disk and passing though the wind.

1.4 AGN Host Galaxies

Attempts to study the host galaxies of quasars are faced with the difficulty of isolating the emission of the faint host from the relatively bright nucleus. And although this remains a challenge, it has become somewhat more feasible with new advanced facilities such as the Hubble Space Telescope refurbished optical system and the large aperture ($\sim$ 8 m) ground-based telescopes equipped with adaptive optics (e.g., spatial resolution of $\sim$ 0.1 arcsec, Wright et al., 2010).

Despite many studies that show quasars are predominantly housed in massive elliptical galaxies (e.g. McLure et al., 1999; Nolan et al., 2001; Pagani et al., 2003; Kauffmann et al., 2003; Dunlop, 2004), studies of samples of low redshift quasar hosts sometimes show incompatible results regarding their properties such as morphologies or stellar contents. The main reason behind these inconsistencies appears to be related to the properties of these samples such as the nuclear luminosity (e.g. Dunlop, 2004; Kauffmann et al., 2003).

Indeed, Kauffmann et al. (2003) find that low luminosity AGN in their $\sim 20,000$ AGN from the Sloan Digital Sky Survey (SDSS) have a "normal" stellar population similar to stellar populations found in massive ellipticals with no AGN while high luminosity AGN have younger stellar populations ($\leq$ 1 Gyr).

This same distinction seems to be holding for the morphology of AGN hosts. While quasar hosts are found to be predominantly massive ellipticals in the samples of Dunlop et al. (2003), Pagani et al. (2003), and Floyd et al. (2004), both bulge-dominated and disk-dominated morphologies are present in different samples such as Bahcall et al. (1997), Jahnke et al. (2004), and Kotilainen et al. (2013). Figure 1.7, for example, shows HST images from the sample of Bahcall et al. (1997) where galaxies seem to display a variety of morphologies: ellipticals, spirals, and some even show signs of galaxy interaction. However, Dunlop (2004) argues that when AGN hosts of similar nuclear luminosity are compared, the hosts of higher nuclear luminosities are generally ellipticals while the disk component becomes more evident at lower luminosities. Another line of evidence that supports the notion that most hosts of AGNs are

massive ellipticals comes from the results of their surface brightness profile fitting (e.g., Dunlop et al., 2003). AGN host galaxies appear to have large size scales of $\sim$ 10 kpc for the effective radius, R_e, which is defined as the radius of a circle that includes half of the bulge light (Sparke & Gallagher, 2007). Furthermore, the surface brightness of both radio-quiet and radio-loud[4] quasars are found to be inversely proportional to their radii (e.g., Dunlop et al., 2003). This last result is consistent with the Kormendy relation which is mainly a projection of the Fundamental Plane (FP)[5] for bulge-dominated galaxies with no AGN (Kukula et al., 2001; Dunlop et al., 2003).

1.5 Galaxy Redshift Evolution

Compared to their higher redshift counterparts, galaxies in the local universe appear to be more massive, with old stellar populations (> 8 Gyr) (Kukula et al., 2001; Nolan et al., 2001), and have a higher fraction of bulge–dominated morphologies (Kukula et al., 2001; Dunlop, 2004, and references therein). Measurements of quasar hosts in the local universe suffer from complications that result from the bright nucleus which overwhelms the light from the host. And these measurements become more difficult for galaxies at higher redshifts as a result of the smaller angular sizes and lower surface brightnesses.

The rise in the star formation rate (SFR) with increasing redshift is known for galaxies with no AGN and is thought to peak at redshift $\sim$ 2 and decline for higher redshifts. This trend of SFR increase with redshift seems to be also present for galaxies with AGN (e.g., Madau et al., 1998; Floyd et al., 2013). Using U and V-filter images of quasar host galaxies from HST, Floyd et al. (2013) showed that indeed the mean star formation in their $z = 2$ sample is $\sim 350\ \mathrm{M_\odot yr^{-1}}$ ($1\ \mathrm{M_\odot} = 1.99\times10^{30}$ kg) for the radio-loud quasars and $\sim 100\ \mathrm{M_\odot yr^{-1}}$ for the radio-quiet quasars while at $z = 1$ the star formation falls to $\sim 150\ \mathrm{M_\odot yr^{-1}}$ for the radio-loud quasars and $\sim 50\ \mathrm{M_\odot yr^{-1}}$ for the radio-quiet quasars. This same sample also confirmed the findings of Kukula et al. (2001) of higher masses of AGN hosts in the local universe compared to those at larger

[4]Radio-loudness is often defined in the literature using the ratio between the observed radio flux density at rest-frame 6 cm and the optical flux density at rest-frame 2500 Å; R= $f_{6\ \mathrm{cm}}/f_{2500\ \text{Å}} > 10$ (e.g., Jiang et al., 2007).

[5]The fundamental plane is a correlation between three parameters in elliptical galaxies: effective radius R_e, central velocity dispersion σ, and surface brightness I_e: $R_e \propto \sigma^{1.2} I_e^{-0.8}$ (Sparke & Gallagher, 2007).

z. Their measurements estimated masses almost one order of magnitude higher for their $z = 1$ sample compared to the $z = 2$ one for the radio-quiet quasars.

The $M_{BH} - L_{bulge}$ and $M_{BH} - \sigma_*$ relations for high redshift AGN are observed to show some variations from those of local AGN and non-AGN (see discussed in §1.1) (e.g Woo et al., 2006; Treu et al., 2007; Woo et al., 2008; Bennert et al., 2011; Canalizo et al., 2012). At higher redshifts these two relations are offset from the local ones as Fig. 1.8 clearly shows. Treu et al. (2007), for instance, found that at redshift $z = 0.36$ the $M_{BH} - L_{bulge}$ and $M_{BH} - \sigma_*$ differ from the local ones by $\Delta \log M_{BH} = 0.51$, $\Delta \log L_B = 0.40$ and $\Delta \log M_{BH} = 0.54, \Delta \log \sigma_* = 0.13$ respectively, i.e., objects had either higher BH masses or lower L_{bulge} and σ_* as shown in Fig. 1.8. Similar results were found by Canalizo et al. (2012) for a sample of $z = 0.14 - 0.37$ quasars. The offsets from the local $M_{BH} - \sigma_*$ relation could point to a scenario where at higher redshifts and for a given BH mass, galaxies had less massive bulges (and hence lower σ_*). Galaxy mergers are a possible explanation for the increase in the bulge mass towards the local universe (e.g., Woo et al., 2008; Canalizo et al., 2012)

1.6 Observational Results on the NLR

The NLR is the part of the galaxy ISM where low density gas is photoionized by the central source at distances ≥ 1 kpc. We discussed the various physical processes that take place in the NLR in §1.2.2. Common forbidden lines usually seen in the optical spectra of quasars include [Ne V] $\lambda 3427$ with an ionization potential IP=97.1 eV, and [O III] $\lambda\lambda 4959, 5007$ with IP=35.1 eV.

One of the great advantages of studying the NLR is its large spatial extent which allows us to spatially resolve its structure and content. For instance, narrow-band imaging of the NLR shows that its morphology is consistent with the AGN unified model (described in §1.2.2) which suggests that an optically thick dusty structure surrounds the central source and causes the NLR to have a symmetric cone-like shape, called an ionization cone (Peterson, 1997, Fig. 1.6). Using HST narrow-band images, Bennert et al. (2002) found that the NLRs in their sample of PG radio-quiet quasars have sizes ranging from $3 - 10$ kpc. Moreover, the size of the NLR in their sample was found to scale with luminosity in a similar fashion to the scaling

relation of the BLR size with continuum luminosity: $R \propto L^{0.52}_{[\mathrm{OIII}]}$. Greene et al. (2011) found a similar scaling relation between the size of the NLR and the [OIII] luminosity with a slope of ~ 0.2. It should be noted, however, that a scaling relation of this sort means that some objects with high luminosity would wind up with unrealistically large NLR sizes (close to 100 kpc in some cases!) (Netzer et al., 2004). At such high distances the NLR would no longer be gravitationally bound to the galaxy which might explain the observed decrease in the luminosity of the narrow lines with the increase of continuum luminosity referred to as the "disappearing NLR" (Netzer et al., 2004; Croom et al., 2002; Greene et al., 2011).

The correlation between the source luminosity and the size of the NLR is also the basis for explaining why the narrow lines follow the Baldwin effect (discussed in §1.2.2) which was originally presented as an anticorrelation between the continuum luminosity and the relative strength of the broad emission lines (e.g. Croom et al., 2002; Zhang et al., 2013b). Croom et al. (2002) used a large sample of quasars from the 2dF and 6dF surveys to make composite spectra at different redshift steps and found that many narrow lines in their composites (such as [O II] λ3727 and [Ne V] λ3426) follow the Baldwin effect. They argue that the presence of this anticorrelation between the strength of the lines and the source luminosity might be a result of the NLR extending beyond the host galaxy in high luminosity sources, i.e., the so-called "disappearing NLR".

Using high resolution spectra allowed a more detailed study of the structure of NLRs in some objects. Such studies showed evidence for stratification in the NLR gas. They also revealed that the velocity of some emission lines correlates with the continuum radiation energy –high ionization lines were found to be more blueshifted than the lower ionization ones (Kraemer & Crenshaw, 2000) (see also discussion in §1.2.2 for stratification in the BLR). The observed stratification in AGN gas can be understood in light of photoionization; the farther out from the ionizing source the lower the photons energy and the less in number in an analogous way to the structure of the HII regions (Peterson, 1997; Draine, 2011).

Several empirical correlations were found between the narrow emission lines of AGN, and the broad lines and continuum. The most famous one is perhaps the anticorrelation between the broad line FeIIλ4570 optical complex and the narrow line [O III] λ5007 (e.g. Boroson & Green, 1992; Boroson, 2002). Another strong correlation with the EW of the narrow [O III] line is

with the width of the broad component of Hβ. Indeed several large quasar samples showed that objects with weaker [O III] would also have stronger Fe II and narrower Hβ (Boroson & Green, 1992; Boroson, 2002; Véron-Cetty & Véron, 2006, e.g.). However, these observed correlations still have no strong physical explanation that can fully explain them. Indeed it is unclear how, with such an enormous difference in physical scales, the NLR and the BLR can be interacting with each other and producing such correlations.

1.7 Trends in Large Datasets

The overarching theme of the work presented here is using large homogenous datasets (particularly form the Sloan Digital Sky Survey) to explore correlations among the properties of quasar spectral features. It is well-know that such trends are generally found to be multidimensional, i.e., involving correlations among several parameters (e.g., Boroson & Green, 1992; Sulentic et al., 2000). However, isolating the primary drivers of these correlations from ancillary ones can be non-trivial and requires some understanding of the physical nature of these correlations and the contributions they make to the observed spectrum. An example of this is seen in the role the luminosity plays in the observed strength of emission lines; e.g., the Baldwin effect seen as a negative correlation between the continuum luminosity and the emission line EWs. Controlling for luminosity would then allow an examination of correlations among emission-line properties free from the complications of luminosity effects.

One notable application of searching for patterns in multidimensional parameter space using spectral properties was applied to a sample of 87 quasars. This study by Boroson & Green (1992), used Principal Component Analysis to define eigenvectors that maximize the variance along orthogonal axes. Their work revealed Eigenvector 1 –a suite of correlations among several properties of quasars including the strengths of the optical Fe II multiplet and the [O III] 5007 Å line and the width of Hβ. It has been subsequently argued that Eigenvector 1 is driven by the Eddington accretion rate (e.g., Boroson, 2002; Shen & Ho, 2014).

Working with large datasets (of a few thousand objects) provides the opportunity to control for confounding effects, such as luminosity, and still maintain a reasonable number of objects that permit reliable conclusions to be drawn about underlying physical properties. This has

been done repeatedly since large spectroscopic surveys (such as the SDSS) became available and often involves binning the sample in one or two dimensions (e.g., the EW or width of a particular emission line) and generating stacked spectra to visualize the results (e.g., Croom et al., 2002; Richards et al., 2011; Hill et al., 2014; Shen & Ho, 2014; Baskin et al., 2015).

In the three projects presented here, we use large spectroscopic samples from the SDSS to trace correlations and make connections to their underlying physical roots. In chapter 2, we use a conventional approach and bin our sample in two dimensional space to control for two key parameters: Hβ width and continuum luminosity. We also use sub-bins in redshift to track the behaviour of the emission lines in each Hβ width-luminosity-redshift bin. In chapters 3 and 4, we explore using unsupervised clustering analysis (namely the K-means algorithm) to partition our datasets of UV spectral measurements and examine the resulting clusters by combining the objects in median composite spectra. In doing so, we do not limit our analysis to traditional boundaries, such as the 2000 km s^{-1} commonly used for the width of Hβ, but rather we leave it to the algorithm to trace the structure in a given multidimensional parameter space and isolate groups of interest using more natural boundaries. We elaborate on the methods and parameters used in each project in the analysis sections in chapters 2, 3, and 4.

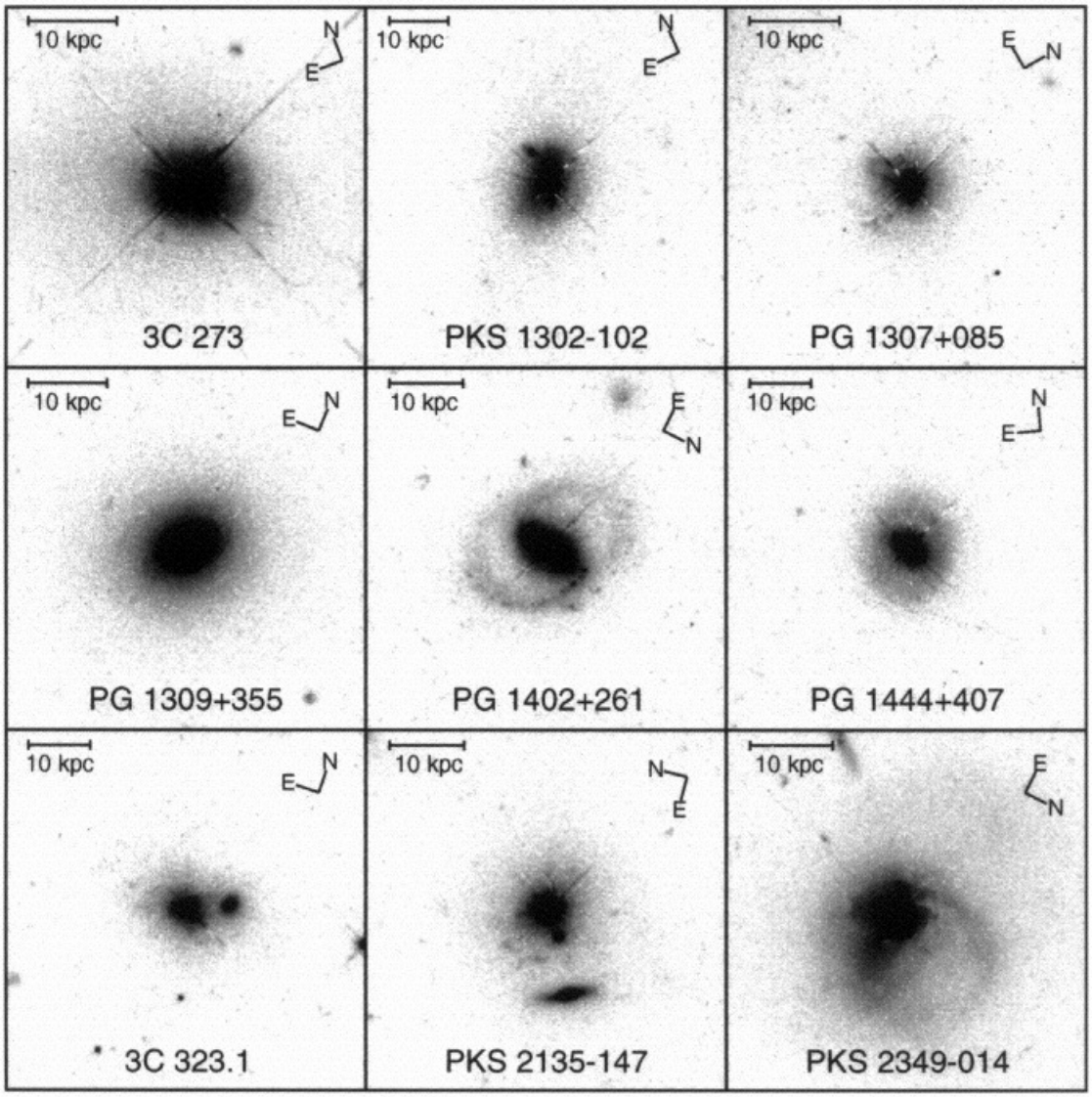

Figure 1.7: HST images of some of the host galaxy of the quasar sample from Bahcall et al. (1997). The sample contained a range of host galaxy morphologies from ellipticals (PG 1309+355) to spirals (PG 1402+261). Some seem to have nearby companions (PKS 2135-147) or show signs of interactions (PKS 2349-014).

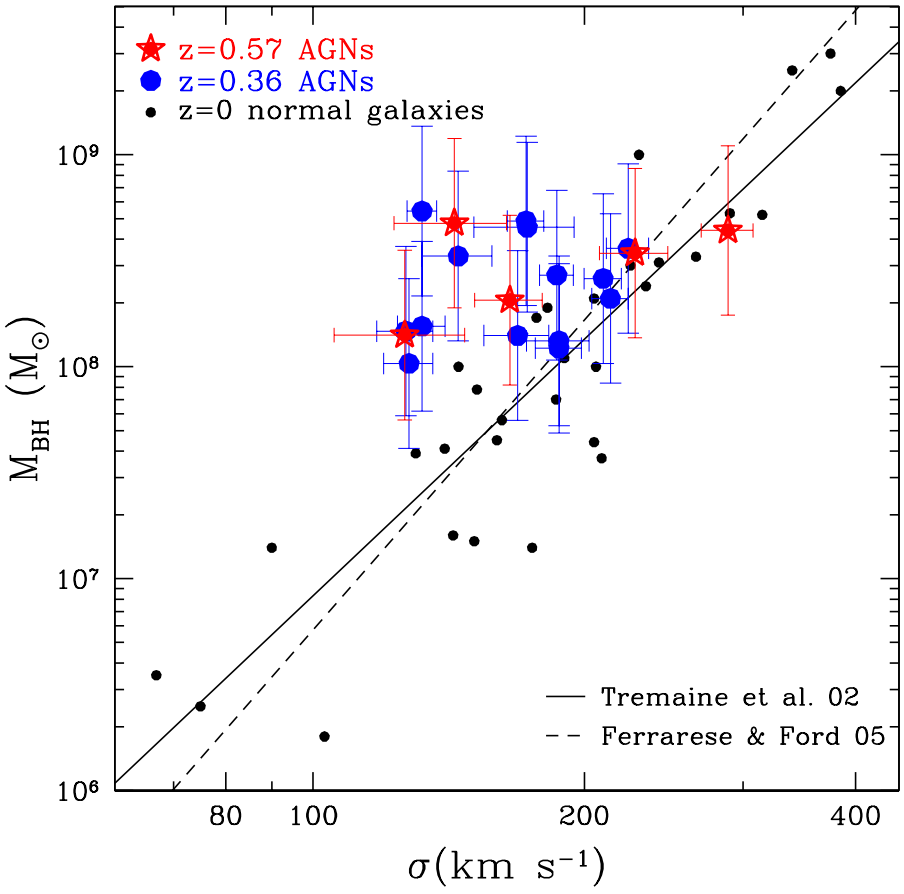

Figure 1.8: The $M_{BH}-\sigma_*$ relation at higher redshifts from Woo et al. (2008). At higher redshifts the objects seem to have higher M_{BH} and/or lower σ_*.

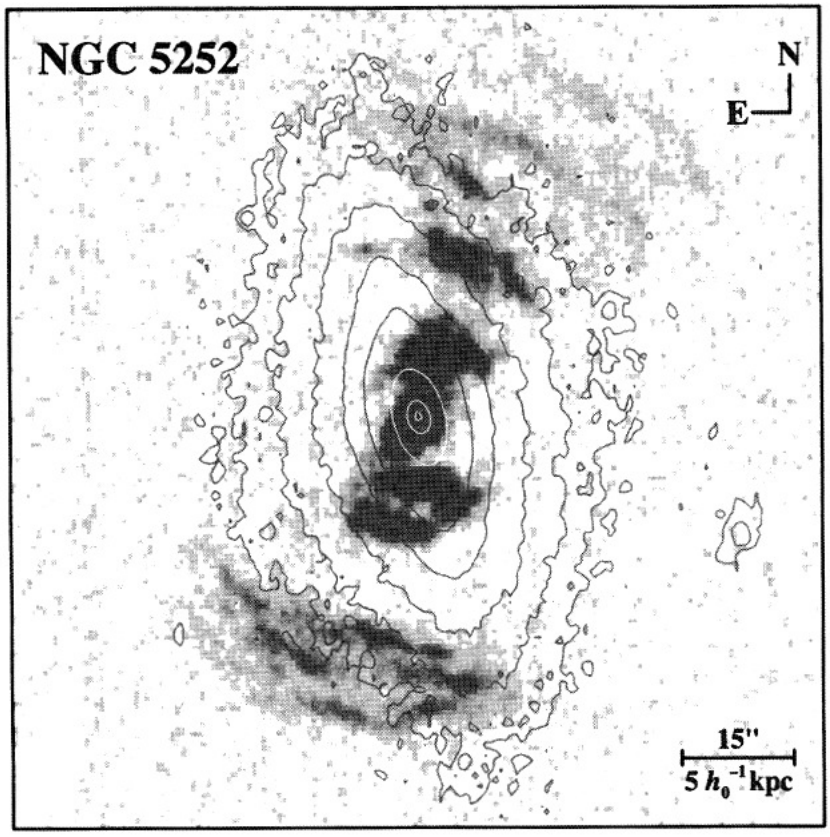

Figure 1.9: [O III]λ5007 narrow-band image that shows the symmetric ionization cones of the NLR inverted colour scheme in NGC 5252. The contours show the stellar surface brightness distribution (Wilson & Tsvetanov, 1994).

Chapter 2

Tracing Quasar Narrow-Line Regions Across Redshift: A Library of High S/N Optical Spectra

A. Tammour, S. C. Gallagher, & G. T. Richards
Monthly Notices of the Royal Astronomical Society, Volume 448, Issue 4, p.3354-3362
April 2015

2.1 Introduction

The current paradigm of active galactic nuclei (AGN) includes a central source that comprises a super-massive black hole (SMBH) surrounded by an accretion disk (Antonucci, 1993; Urry & Padovani, 1995). Through this accretion disk, material gets funnelled into the SMBH (Shakura & Sunyaev, 1973). Quasars are the most luminous class of AGN with bolometric luminosities of up to 10^{48} erg s^{-1}.

Photoionization by radiation from the accretion disk is thought to be the main mechanism through which gas in AGNs is ionized (Osterbrock & Ferland, 2006). Studying the interactions between the ionizing photons and the material they encounter allows us to study the physical conditions in the ionized regions. For example, the presence of narrow forbidden lines in quasar spectra points to a region that has low density and is located relatively far from the black hole's

sphere of influence. For the most part, the gas in the NLR region responds kinematically to the gravitational potential of the galaxy as a whole and therefore produces narrower emission lines (FWHM $\leq$ 1000 km s^{-1}) – hence its label. On the other hand, in the close vicinity of the central source ($\leq$ 1 pc), the gas mainly feels the gravitational influence of the SMBH and therefore produces emission lines that are primarily Doppler-broadened with widths of 1000s km/s. We therefore call this region the broad-line region (BLR).

Though the broad and narrow-line regions are physically distinct, correlations between properties of the two regions are found in the literature. Most well-known are the so-called eigenvector 1 relations, identified by Boroson & Green (1992, hereafter BG92) from principal component analysis of several spectral measurements of the low-redshift PG quasar sample. BG92 found that narrow Hβ width, strong Fe II emission, and weak [O III] emission tended to be found in the same objects. The strong eigenvector 1 objects have also been identified as having relatively high $L/L_{\rm Edd}$ values (e.g., Boroson & Green, 1992; Boroson, 2002; Sulentic et al., 2007). Given our understanding of the rather different origins of BLR and NLR emission, the linking of broad- and narrow-line spectral properties is puzzling, and the extended discussion in BG92 reflects this conceptual difficulty.

To first order, the NLR can be described as the diffuse interstellar medium of the host galaxy illuminated by the accretion disk continuum. Quasar NLRs are therefore the regions where we expect the interactions between AGNs and their host galaxies to take place; the narrow emission lines are among the few tracers of the host galaxy that are evident in optical quasar spectra. It is therefore important to check that any apparent links between broad and narrow-line region spectral properties do not arise from secondary and unintentional selection effects of host galaxies. For example, the strong eigenvector 1 objects are necessarily low redshift (because of the lines used in the study), and therefore low luminosity. The original BG92 sample were all PG quasars, which because of their selection are not representative of the quasar population as a whole (Jester et al., 2005).

Therefore, we are interested in studying samples of quasars with very similar central engines across a range of redshift to determine if we can isolate the effects of host galaxy evolution on NLR emission. With the large Quasar Properties Catalogue of Shen et al. (2011) with 105,783 objects, this hypothesis can be tested by selecting similar types of quasars (in terms of

$L_{\rm opt}$ and $M_{\rm BH}$), and examining their narrow-line region properties independently.

Throughout this work we use: $\Omega_\Lambda = 0.7,\ \Omega_0 = 0.3,\ h = 0.7$ (Spergel et al., 2003).

2.2 Data and Analysis

2.2.1 The Sample

The initial sample is selected from the Shen et al. (2011) catalogue of quasar properties (SDSS-DR7; Schneider et al., 2010; Shen et al., 2011). We restrict the sample to radio-quiet quasars (defined in the catalogue as objects with $R = f_{6cm}/f_{2500} < 10$; where f_{6cm} and f_{2500} are the radio and optical flux densities observed at 6 cm and 2500 Å respectively). We also impose a redshift limit of $z \leq 0.75$ to ensure $\lambda_{rest} = 5100$ Å is in our spectral range. This emission-line-free segment of the continuum is used to first normalize the spectra and later to calculate the relative line luminosities of the composites. The final sample contains 16,027 spectra. We utilize the OH-subtracted spectra of Wild & Hewett (2010) who constructed a set of the telluric-line-cleaned spectra (specifically around Hβ). We then apply IRAF's DEREDDEN tool[1] to correct the spectra for Galactic extinction using the $E(B-V)$ values provided in the catalogue (Shen et al., 2011, and references therein). Finally, the spectra are shifted to the rest frame using the Hewett & Wild (2010) redshift determination as provided in Shen et al. (2011).

2.2.2 The Composites

One of the advantages of starting from a large homogeneous dataset like the SDSS quasar sample is the ability to stack objects to create higher signal-to-noise spectra which then allows us to examine weaker features (such as the high IP [Ne V] line) that are otherwise hard to measure (e.g, Vanden Berk et al., 2001; Sulentic et al., 2002; Croom et al., 2002; Hill et al., 2014). To make sure we are identifying objects with similar AGN properties, we group objects according

[1]IRAF is distributed by the National Optical Astronomy Observatories, which are operated by the Association of Universities for Research in Astronomy, Inc., under cooperative agreement with the National Science Foundation.

to their continuum luminosities at 5100 Å (L_{5100}) and the FWHM of Hβ[2] as reported in Shen et al. (2011). These two quantities are often used in combination in the literature to estimate more physical properties of the AGN such as the mass of the SMBH and the accretion rate (Shen et al., 2011, and references therein, see also §2.2.3). We used the measured properties (luminosity and Balmer-line broad-line width) rather than derived properties ($M_{\rm BH}$ and $L/L_{\rm Edd}$) to keep the binning process closer to the data. As shown later, in practice, our choice of binning also means that objects with similar values of $M_{\rm BH}$ and $L/L_{\rm Edd}$ are grouped together. We note however that the conversion of Hβ FWHM and L_{5100} into BH masses and $L/L_{\rm Edd}$ can be influenced by several caveats such as the dependance of FWHM to the orientation and our conversion factor of L_{5100} to bolometric luminosity (e.g, Richards et al., 2006a). We start by sorting the objects according to their L_{5100} values and then dividing them into 10 groups (deciles) with an equal number of objects (1602 or 1603). We choose to look at four of these groups with the low ($L1$: $\log L_{5100} = 43.15 - 44.17$ erg s^{-1}), intermediate ($L4$: $\log L_{5100} = 44.36 - 44.45$ erg s^{-1} and $L7$: $\log L_{5100} = 44.63 - 44.72$ erg s^{-1}), and high ($L10$: $\log L_{5100} = 44.98 - 46.19$ erg s^{-1}) luminosities. Specifically, we used groups 1, 4, 7, and 10 and skipped the intermediate ones (2, 3, 5, 6, 8, and 9). This choice gives us the full dynamic range of the sample, while keeping the number of bins to a manageable number.

We do not increase the size of the bins to use all of the spectra in the sample because we are interested in creating high S/N composites of objects that are as similar as possible. The conversion of L_{5100} and Hβ FWHM into $L/L_{\rm Edd}$ and $M_{\rm BH}$ has significant uncertainties (e.g, Vestergaard & Peterson, 2006; Shen et al., 2011). For example, high energy AGN spectral energy distributions change as a function of luminosity (e.g., Steffen et al., 2006), and so derived quantities such as bolometric luminosities (required for L/L_{Edd}) are likely not properly calibrated for all luminosities (Richards et al., 2006b).

We then use the measurements of Hβ from Shen et al. (2011) to separate objects into three groups: narrow, intermediate, and broad Hβ using 2000 and 4000 km s^{-1} as our boundaries (excluding objects with Hβ FWHM $>$ 20000 km s^{-1}). We choose fewer bins in Hβ FWHM compared to L_{5100} because of the considerably smaller dynamic range of the FWHM values,

[2]Hereafter when we refer to Hβ, we mean the broad component of the line from which a narrow-line component has been subtracted.

Table 2.1: Breakdown of the number of objects in each luminosity subset with narrow (n; FWHM < 2000 km s^{-1}), intermediate (i; 2000 km s^{-1}< FWHM < 4000 km s^{-1}), and broad (b; 4000 km s^{-1}< FWHM <20,000 km s^{-1}) Hβ at each redshift step. We only use the bins with > 10 objects to create composites. The bins with < 10 objects are highlighted with boldface in this table.

log L_{5100}	$L1$			$L4$			$L7$			$L10$		
[erg s^{-1}]	43.15 – 44.17			44.36 – 44.45			44.63 – 44.72			44.98 – 46.19		
Redshift	n	i	b	n	i	b	n	i	b	n	i	b
0.1	14	29	29	**0**	**3**	**1**	**0**	**0**	**1**	**0**	**0**	**0**
0.2	55	203	249	**10**	20	23	**1**	**7**	**9**	**0**	**0**	**1**
0.3	40	168	255	22	92	126	**6**	36	28	**2**	19	**10**
0.4	34	75	91	56	238	299	27	81	85	16	37	49
0.5	12	72	68	34	141	188	19	143	216	**8**	94	108
0.6	13	38	82	**8**	85	104	37	252	306	27	178	232
0.7	**6**	21	41	14	51	84	23	137	184	30	314	475
Total	174	606	815	144	630	825	113	656	830	83	642	875

a factor of $\approx$ 20 compared to three orders of magnitude for the sample's range of L_{5100}. The boundary at 2000 km s^{-1} marks the traditional definition of narrow-line Seyfert 1s (Osterbrock & Pogge, 1985), and the intermediate bin covers the main peak of the Hβ FWHM distribution.

Once the quasars are divided by luminosity and Balmer line width, we are interested in looking at the NLR properties of these bins. Specifically, we investigate if there is any evidence for redshift evolution that might be expected if AGN host galaxies have evolved since redshift 0.75. We thus divide each of these luminosity groups into 7 subsets according to their redshifts with bins of width 0.1 centred at: 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, and 0.7. This results in 4 $L_{5100} \times 3$ Hβ FWHM $\times$ 7 z = 84 bins, 8 of which are empty (primarily the high luminosity/low redshift ones) as shown in Table 2.1. For the 76 non-empty bins, we make median composite spectra with 3σ clipping using IRAF's SCOMBINE tool. We choose to create median composites to preserve the EWs of the lines (see Vanden Berk et al. (2001) for a good discussion of the difference between median and mean composites).

To examine the distributions of objects within our binning, we look at the distributions of L_{5100}, Hβ FWHM, and redshift (Fig. 2.1, 2.2, and 2.3). The figures show that the distributions are fairly consistent among the different bins. We note, however, that there are more quasars with higher luminosity at higher redshift as expected because of the SDSS flux limit for quasar spectroscopy ($i < 19.1$) and the redshift evolution of the quasar luminosity function (Richards et al., 2006a). Table 2.1 also shows that the narrow Hβ subsets contain fewer objects than the intermediate and broad subsets. This is mainly due to the selection of objects in the SDSS quasar database requiring at least one broad line with FWHM > 1000 km s^{-1} (Schneider et al., 2010). The spectral library is given in Appendix A (available in the online-only version).

2.2.3 Measurements

Optical quasar spectra show several identifiable narrow emission lines. For this study, we focus on four narrow emission lines that are available in the wavelength range of our composites: [Ne V] λ3427, [O II] λ3728, [Ne III] λ3870, and [O III] λ5007, with ionization potentials of 97.1, 13.6, 40, 35 eV, respectively (Peterson, 1997). These are the brightest, common emission lines, and they span a large range of ionization potentials. The high S/N present in most of our composites allows us to measure some of the relatively weak features such as [Ne III] λ3728 that are typically not detected in individual spectra. We measure the flux of these narrow lines in each composite using one of two approaches. First, in the Hβ - [O III] $\lambda\lambda$4959, 5007 region, we use IRAF's SPECFIT package to fit the composite with a multicomponent model (Kriss, 1994). This method allows us to subtract the Fe II emission features which form a pseudo-continuum of blended lines that are in some cases heavily blended with other emission lines in this part of the spectrum. For the removal of the Fe II lines, we make use of the Fe II template of Véron-Cetty et al. (2004). We use a Lorentzian and up to two Gaussians to fit Hβ, and up to three Gaussians to fit each of the [O III] lines. An example of the fitting model in this spectral region is shown in Figure 2.4. In the [Ne III] - [Ne V] - [O II] ($\sim$ 3300 – 4000 Å) region we use IRAF's SPLOT tool and simply integrate the flux above the continuum (defined around the emission line by visual inspection) as we do not anticipate significant contamination from the Fe II lines.

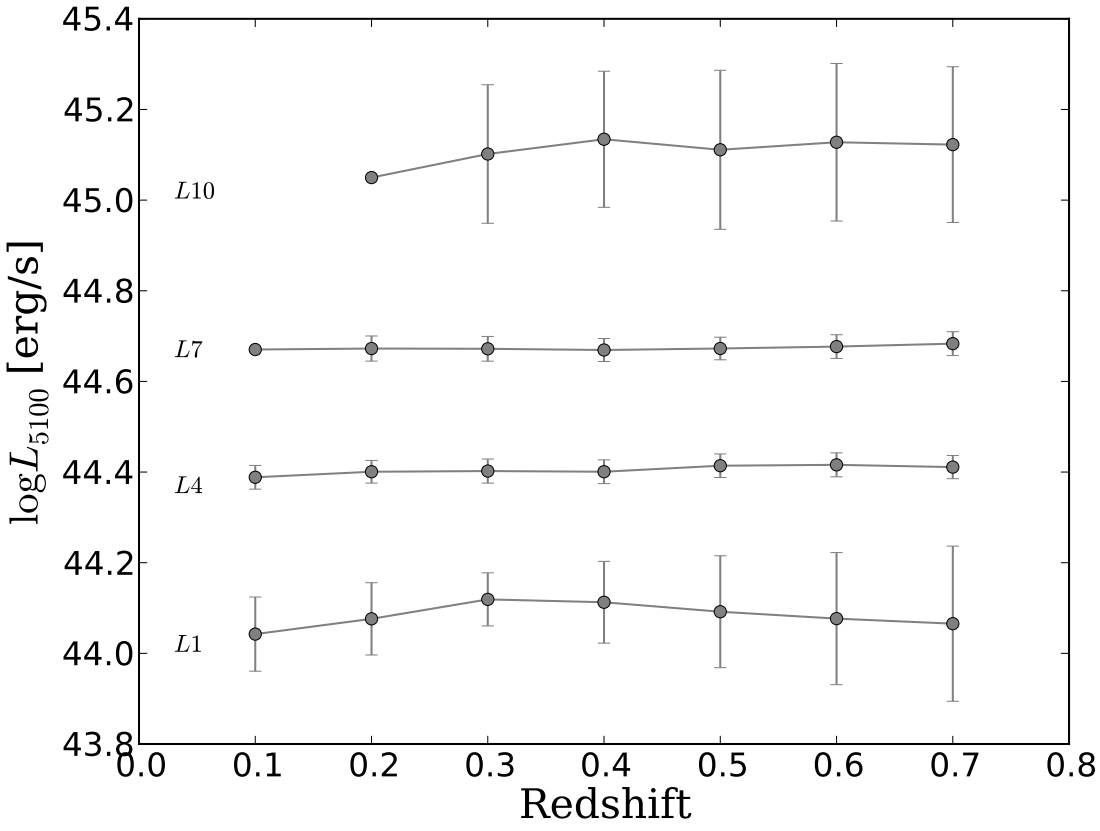

Figure 2.1: Median L_{5100} from Shen et al. (2011) vs. redshift. Within the given uncertainty, the continuum luminosity is constant within the same bin across redshift. The $L10$ group does not have objects with redshift < 0.2. The error bars show the standard deviation in the bin. The two data points without error bars represent bins with fewer than 3 objects.

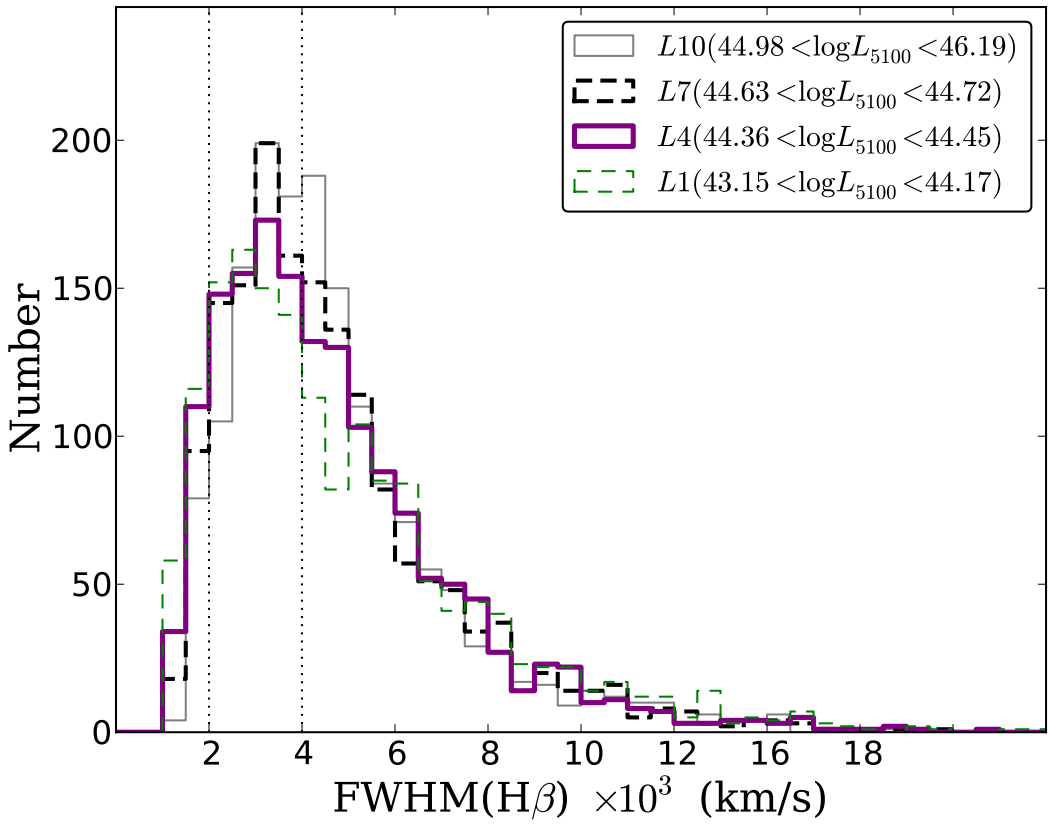

Figure 2.2: Distribution of Hβ FWHM for the four luminosity subsets. The distributions look similar except for a weak trend of rising numbers of broader Hβ objects in higher continuum luminosity bins (grey solid line). Narrower Hβ quasars are more abundant in the lowest luminosity bin (green dashed line). Objects with Hβ FWHM > 20,000 km s^{-1} are not included in making the composites as these were often found to arise from either bad measurements or low quality spectra.

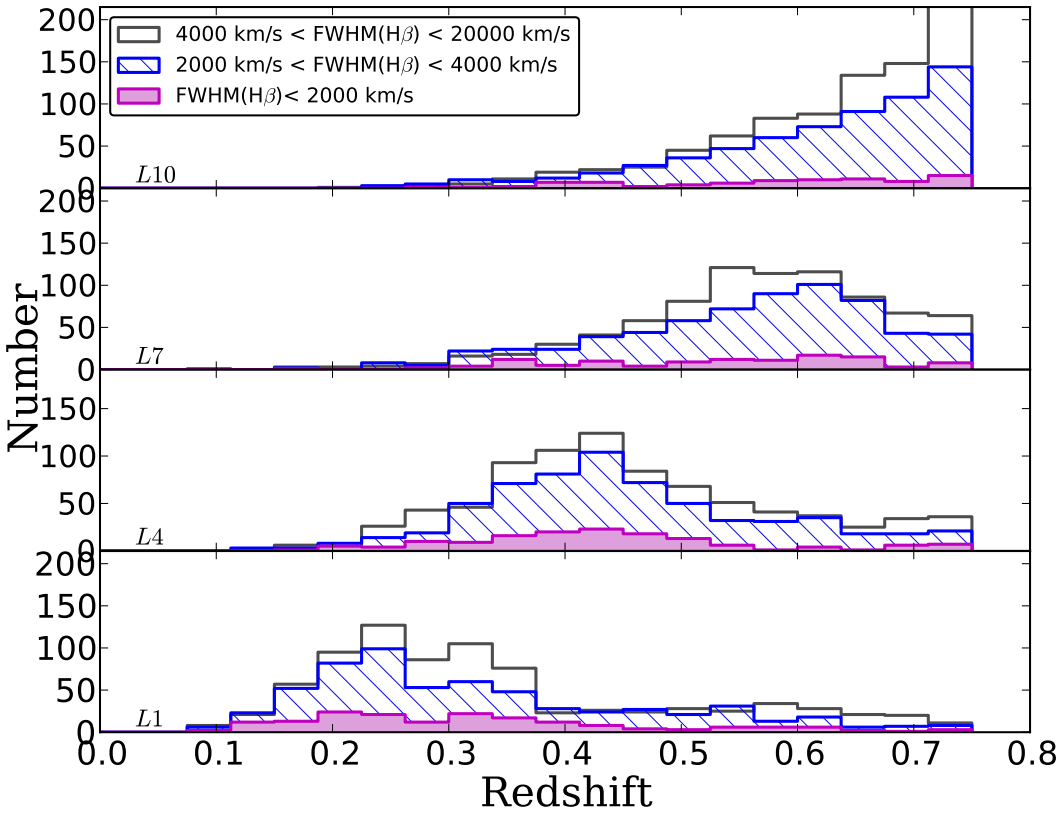

Figure 2.3: Distribution of redshift at the four continuum luminosity levels for the narrow (solid magenta), intermediate (hashed blue) and broad (open grey) Hβ subsets. There is a gradual shift in the distributions of redshift from lower luminosity to the higher luminosity subsets – higher luminosity objects have more objects at higher redshift than lower luminosity ones as expected from the SDSS flux limit for quasar spectroscopy and the redshift evolution of the quasar luminosity function.

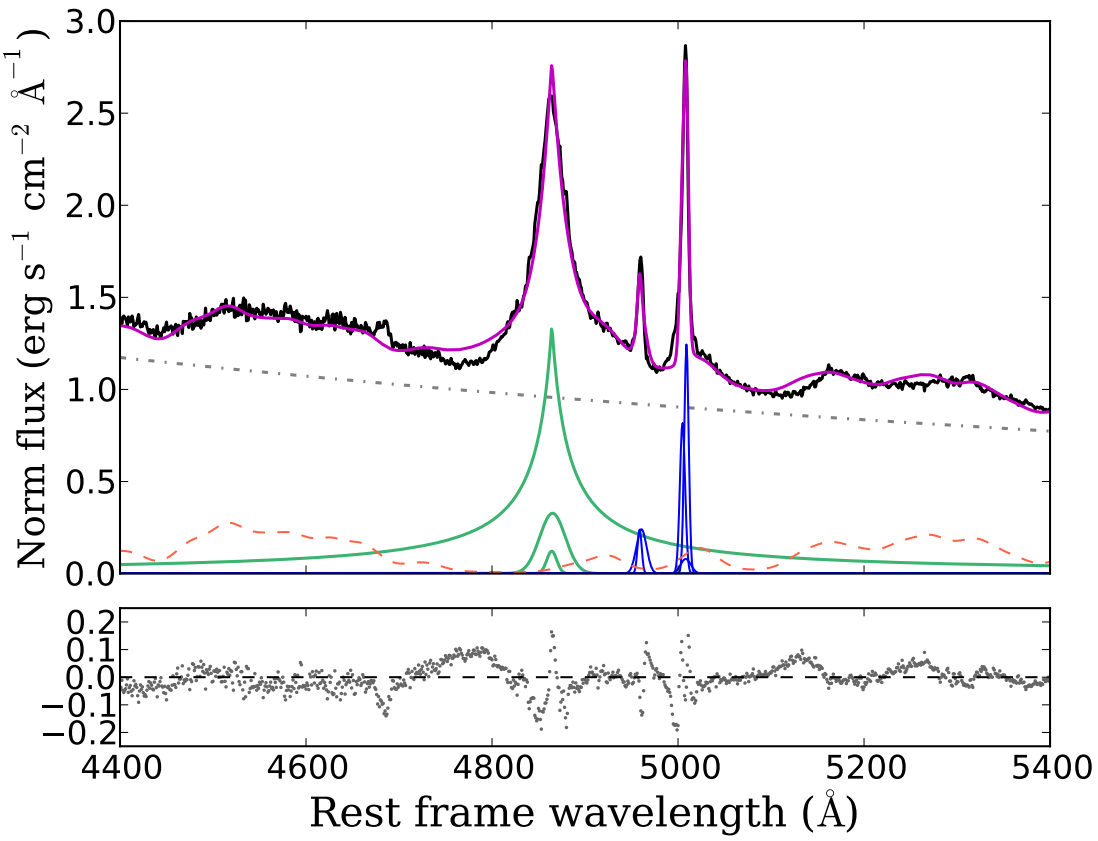

Figure 2.4: Model fit for composite iL7z2 (black) and the full fitting result (magenta). Hβ is fit with a Lorentzian and up to two Gaussians (green), the [O III] doublet is fit with up to three Gaussians per line (blue), the grey dash-dot line is a power-law continuum, and the Fe II template is shown as a dashed red line. The residuals are shown in the lower panel.

The results of the measurements are tabulated in Table 2.3.

We use the L_{5100} and Hβ width measurements to calculate the BH mass and accretion rate for each composite. For the BH masses, we use equation 5 from Vestergaard & Peterson (2006):

$$\log \frac{M_{BH}}{M_\odot} = 0.910 + 0.5 \log \left(\frac{5100 L_{5100}}{10^{44} \text{erg s}^{-1}} \right) + 2 \log \left(\frac{\text{FWHM(H}\beta)}{\text{km s}^{-1}} \right). \tag{2.1}$$

For the accretion rates, we estimate the bolometric luminosity as $L_{\text{bol}} = 9 L_{5100}$ (Peterson, 1997). The Eddington luminosity, L_{Edd}, is calculated as (Netzer, 2006):

$$L_{\text{Edd}} \simeq 1.5 \times 10^{38} \left(\frac{M_{BH}}{M_\odot} \right) \text{erg s}^{-1}. \tag{2.2}$$

Values of the calculated M_{BH}, L_{Edd}, and L_{bol} for each composite are given in Table 2.2. Richards et al. (2006b) have expressed some reservations about the potential pitfalls of applying a single bolometric correction to all quasars as it is quite likely that different quasar populations should have distinct bolometric correction values because of systematic differences in their spectral energy distributions (Krawczyk et al., 2013). This concern was among the reasons that the samples comprising each composite are binned by L_{5100} and Hβ width alone rather than the derived quantities of M_{BH} and L/L_{Edd}. We emphasize that these reported values of M_{BH}, L_{Edd}, and L_{bol} are not used for the binning of objects for the composite selection, but are simply reported for completeness.

2.3 Measured Line Luminosities

The table below (Table 2.3) includes the measured line luminosities for the four narrow lines discussed in this paper ([Ne V], [Ne III], [O II], and [O III]). The full table is available online as an electronic table[3].

[3] http://adsabs.harvard.edu/abs/2015yCat..74483354T

Table 2.2: BH masses, $L_{\rm bol}$ and $L_{\rm Edd}$ for the composites in each of the L_{5100} and Hβ width bins.

Composite	$\log(M_{BH}/M_{\odot})$	$\log L_{\rm Edd}$ [erg s^{-1}]	$\log L_{\rm bol}$ [erg s^{-1}]	L/L$_{\rm Edd}$
nL1	7.39	45.57	45.06	0.31
iL1	7.90	46.08	45.05	0.09
bL1	8.54	46.71	45.05	0.02
nL4	7.57	45.75	45.37	0.41
iL4	8.00	46.18	45.36	0.15
bL4	8.63	46.81	45.36	0.04
nL7	7.73	45.90	45.63	0.53
iL7	8.22	46.40	45.63	0.17
bL7	8.74	46.92	45.63	0.05
nL10	7.98	46.16	46.06	0.80
iL10	8.47	46.64	46.08	0.28
bL10	8.94	47.12	46.07	0.09

Table 2.3: Measured line luminosities for the four forbidden lines [O II], [O III], [Ne III], and [Ne V]. The full table is available online as an electronic table[5].

Composite	z	L([Ne V]) $\times10^{40}$erg s^{-1}	L([Ne III]) $\times10^{40}$erg s^{-1}	L([O II]) $\times10^{40}$erg s^{-1}	L([O III]) $\times10^{40}$erg s^{-1}
nL1z1	0.1	17.08 ± 0.79	5.55 ± 0.51	7.29 ± 0.50	23.68 ± 5.97
nL1z2	0.2	9.88 ± 0.41	6.68 ± 0.45	9.49 ± 0.39	21.71 ± 4.62
nL1z3	0.3	10.18 ± 0.78	11.02 ± 0.77	12.25 ± 0.72	21.09 ± 1.14
nL1z4	0.4	11.74 ± 0.96	13.11 ± 0.99	12.62 ± 0.81	25.56 ± 1.65
nL1z5	0.5	7.95 ± 1.99	5.61 ± 1.70	14.93 ± 2.07	34.09 ± 2.07
nL1z6	0.6	18.03 ± 5.41	15.21 ± 5.60	22.80 ± 4.55	72.31 ± 7.37
nL1z7	0.7	28.98 ± 9.22	8.27 ± 8.06	17.77 ± 7.73	89.13 ± 17.34

2.4 Results and Discussion

2.4.1 Line Luminosity and Redshift

We focus on the four strongest forbidden lines present in the spectral range of our composites: [Ne V] λ3427, [Ne III] λ3870, [O III] λ5007, and [O II] λ3728. In addition, we only use the spectra constructed from more than 10 objects. We trace these four lines across redshift through their luminosities and flux ratios in each of the 7 L_{5100}/FWHM(Hβ) bins. In a stratified NLR and with IPs spreading over a range of $\sim$ 13 – 100 eV, we expect the lines with low IP (e.g, [O II]; IP =13.5 eV) to be emitted farther away from the SMBH (and thus the source of the ionizing continuum) and therefore from regions with weaker AGN influence than the high IP lines (e.g, [Ne V]; IP =97 eV).

As mentioned in §2.2.2, the selection of objects that went into each of the composites is not dependent on any properties of the NLR. In addition, by dividing the spectra into bins of L_{5100} and Hβ FWHM, we aim to ensure that the objects used in making each composite are powered by similar central sources in the sense that the continuum luminosity and the Hβ FWHM are sensitive to the $L/L_{\rm Edd}$ ratios and BH masses of quasars. We choose *not* to use the derived values of Eddington ratio and $M_{\rm BH}$, but rather the measured values of 5100 Å luminosity and FWHM Hβ as these are straightforward spectral measurements that are not subject to changes in calibration or uncertain assumptions about bolometric corrections.

Figure 2.5 shows the line luminosities for the four lines plotted against redshift. Visually, we see a weak overall tendency of the luminosities of the lower IP lines to increase towards higher redshifts. To examine this apparent tend more quantitatively, we calculate the slopes of the line luminosity vs. redshift for each L_{5100} and Hβ bins along with the correlation coefficients (Table 2.4). We discuss the lines individually in the following subsections.

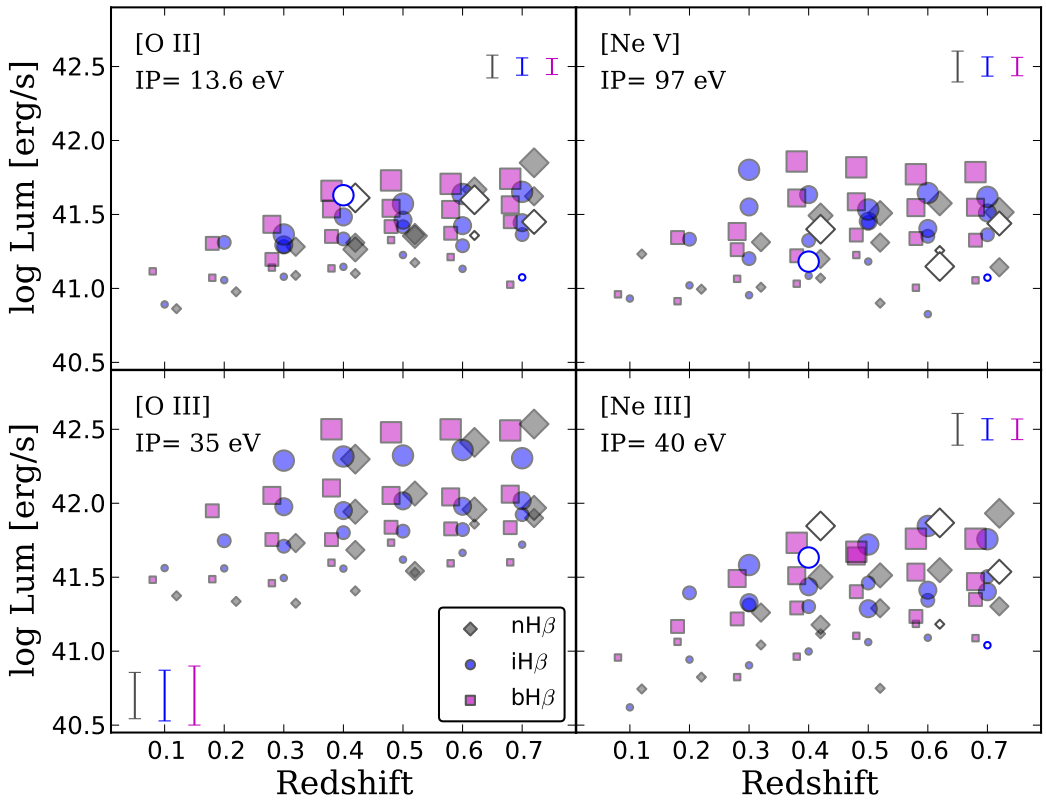

Figure 2.5: Line luminosity vs. redshift for the four narrow lines. Grey diamonds are the composites with Hβ < 2000 km s^{-1}, blue circles are the composites with 2000 < km s^{-1} Hβ < 4000 km s^{-1}, and magenta squares are the composites with 4000 km s^{-1} < Hβ< 20000 km s^{-1}. Open symbols denote non-detections. Marker size corresponds to the continuum luminosity bin: larger for higher L_{5100}. Only composites with $N > 10$ are included. The grey/magenta points are slightly offset to the left/right of their redshift value for clarity. In all luminosity bins, the [O II] line is clearly becoming stronger at higher redshifts. The overall Spearman test gave: ρ_S([O II]) = 0.532, $p < 0.001$; ρ_S([O III]) = 0.447, $p < 0.001$; ρ_S([Ne III]) = 0.479, $p < 0.001$; ρ_S([Ne V]) = 0.373, $p = 0.005$ for about 60 objects (detections only).

Table 2.4: Results of the least-squares regression and the Spearman rank coefficient for the line luminosities and redshift for the four narrow-lines in each of the 7 L_{5100} and Hβ bins.

Subset	Line	Slope	Std Error	ρ_S	p
nHβ	[O II]	1.281	0.222	0.702	0.0008
	[O III]	1.400	0.412	0.497	0.0171
	[Ne III]	1.430	0.366	0.450	0.0309
	[Ne V]	0.473	0.348	0.196	0.3811
iHβ	[O II]	0.509	0.223	0.380	0.0625
	[O III]	0.567	0.316	0.399	0.0505
	[Ne III]	0.746	0.331	0.448	0.0282
	[Ne V]	0.494	0.322	0.120	0.5571
bHβ	[O II]	0.550	0.249	0.543	0.0056
	[O III]	0.679	0.398	0.172	0.3804
	[Ne III]	0.695	0.315	0.378	0.0536
	[Ne V]	0.633	0.340	0.378	0.0542
L1	[O II]	0.297	0.114	0.510	0.026
	[O III]	0.459	0.120	0.708	0.000
	[Ne III]	0.524	0.127	0.727	0.001
	[Ne V]	0.014	0.147	0.088	0.729
L4	[O II]	0.376	0.110	0.736	0.001
	[O III]	0.176	0.149	0.450	0.080
	[Ne III]	0.238	0.133	0.443	0.086
	[Ne V]	0.033	0.121	0.209	0.438
L7	[O II]	0.343	0.199	0.385	0.174
	[O III]	-0.003	0.104	0.029	0.921
	[Ne III]	0.109	0.188	0.188	0.519
	[Ne V]	0.016	0.167	-0.045	0.884
L10	[O II]	0.588	0.195	0.608	0.036
	[O III]	0.310	0.197	0.462	0.131
	[Ne III]	0.464	0.183	0.626	0.029
	[Ne V]	-0.451	0.297	-0.607	0.083

L([O II])

Compared to the much more prominent [O III], the [O II] line is fairly weak in the spectra of radio-quiet quasars. In general, [O II] emission may have contributions from photo-ionization from the AGN and/or star-formation in the host galaxy. Photoionization models suggest that in a pure AGN, the [O II] to [O III] ratio ranges between 0.1 and 0.3 (Ho, 2005, and references therein). Similarly, Kim et al. (2006) find that in quasars, the AGN contribution to [O II] is more dominant than the host galaxy's. The [O II] luminosities in our composite spectra are shown in top-left panel of Figure 2.5. The figure shows a trend of increasing [O II] luminosity with redshift of nearly an order of magnitude. We calculated a Spearman correlation coefficient of $\rho_S = 0.54$ for [O II] luminosity vs. redshift shown in Figure 2.5, with $p < 0.0001$ (p is the probability of the null hypothesis which assumes no correlation) which indicates the high probability that the two quantities are strongly correlated. We note however that this correlation can potentially be a result of the missing high luminosity, low-redshift quasar composites in our sample. However, tracing the data from quasars of similar luminosity, the general trend is towards higher [O II] luminosities at higher redshift. To further examine this possibility, we look at the correlation between the line luminosity and redshift for each L_{5100} bin separately and we find that the correlation is still strong for most cases but the significance of the correlation becomes much lower (higher p value) as a result of fewer data points in each sub-sample (L1:$\rho_S = 0.510$, L4: $\rho_S = 0.736$, L7: $\rho_S = 0.385$, L10: $\rho_S = 0.608$). In addition to the Spearman test, we calculate a least-squares regression slope for each L_{5100} and Hβ bin (Table 2.4) and find that in most cases, [O II] has a weak but significant trend to increase towards higher redshift.

The [O II] luminosity is indeed known to be a reliable tracer of SFR in non-AGN and AGN galaxies (e.g. Gallagher et al., 1989; Kennicutt, 1998; Rosa-González et al., 2002; Hopkins et al., 2003; Kewley et al., 2004; Ho, 2005; Kim et al., 2006; Silverman et al., 2009). This line becomes particularly important for samples with $z \geq 0.5$ as the primary SFR metric of Hα has passed out of the optical bandpass (e.g, Kewley et al., 2004). To get a handle on whether star formation might be contributing to the [O II] emission, we consider the ratio of the line fluxes of the two oxygen lines. Figure 2.6 shows the flux ratio of [O II] to [O III] in our sample.

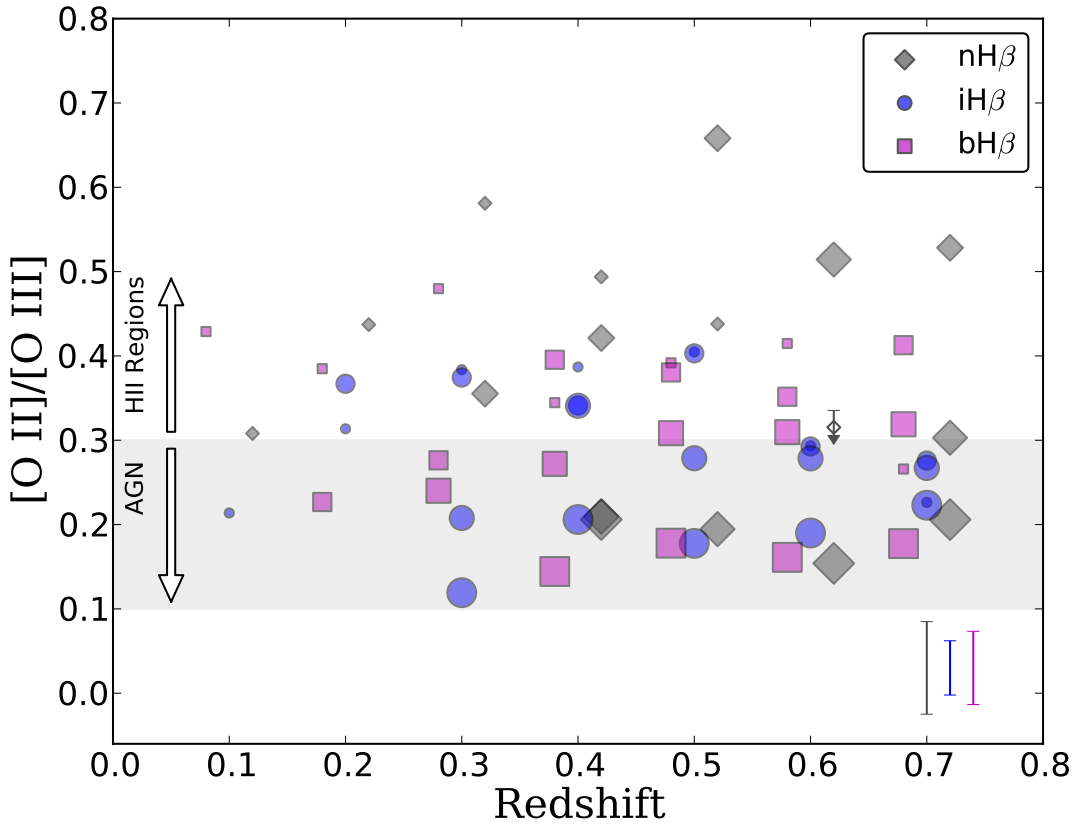

Figure 2.6: [O II]/[O III] flux ratio plotted against redshift. Markers are similar to Fig. 2.5. The shaded region marks the 0.1–0.3 boundary for an AGN-dominated line ratio; a ratio larger than 0.3 is typical of stellar H II regions). The line ratio distribution shows no evidence for evolution as a function of redshift.

The figure shows that almost 50% of the objects (primarily among the low and intermediate continuum luminosity bins; $\log L_{5100} \sim 43 - 44.5$ erg s^{-1}) have [O II]/[O III] > 0.3 consistent with [O II] emission produced in stellar H II regions (Osterbrock & Ferland, 2006). Therefore, we conclude that there is a contribution from star formation to the [O II] emission, at least in some objects.

The apparent increase in [O II] line luminosity as a function of redshift could arise from different factors other than an increase in SFR in the typical quasar host galaxy. For instance, the use of a fixed-size aperture (the SDSS uses a 3″ fiber) means that spectra potentially include

a larger fraction of the host galaxy gas at high redshifts than in lower redshift objects due to the larger projected physical size at higher redshifts. Specifically, the projected size is 5.5 kpc at $z = 0.1$ and 22 kpc at $z = 0.7$. However, we do not anticipate this to be the main reason for the observed increase in the [O II] luminosity (if due solely to photo-ionization by the quasar) as the size of the entire NLR is probably less than 10 kpc. To check, we employ the NLR scaling relation from Bennert et al. (2002, their eq. 1):

$$\log R_{NLR} = (0.52 \pm 0.006) \log L_{[\mathrm{O\ III}]} - (18.5 \pm 2.6), \tag{2.3}$$

and our measured [O III] luminosities (which are significantly higher than the [O II] luminosities), we find the size of the NLR in our composites is $\sim$ 4 kpc. If the [O II] emission is instead largely from star formation in the host galaxy, then the size of the star-forming region within the galaxy is the relevant factor. However, the size of the entire galaxy is not expected to exceed $\sim$ 10 kpc, and the star-forming regions may be considerably more compact. Ichikawa et al. (2012), for example, found in a sample of galaxies selected from the MOIRCS Deep Survey with up to redshift 3, that the half-light radius (R_{50}) in the K-band is between 5–10 kpc.

The calculation of SFRs using the [O II] line in isolation is, however, not completely straightforward. The conversion of the observed line luminosity to SFRs is complicated by effects such as reddening and metal abundance. Several attempts have been made to come up with calibrations that allow for a direct estimation of SFRs from line fluxes (e.g., Rosa-González et al., 2002; Kewley et al., 2004). These calibrations are useful for estimating SFRs without directly involving assumptions about metallicity and intrinsic extinction that are by and large the major sources of uncertainty. We use the L([O II]) calibration equation from Rosa-González et al. (2002) for non-AGN galaxies to estimate SFRs for our composite spectra:

$$SFR_{[\mathrm{O\ II}]} = 8.4 \times 10^{-41} L_{[\mathrm{O\ II}]}\ \mathrm{M_{\odot} yr^{-1}}, \tag{2.4}$$

where we have first subtracted 10% of the [O III] luminosity from the [O II] luminosity to account for the AGN contribution, i.e., we have assumed all of the [O III] emission is from the AGN, and assumed an [O II] to [O III] line ratio of 0.1. We plot the results in Figure 2.7 which shows that the SFR calculated with equation 2.4 increases with z. As interpreted at face value, the figure also shows that higher continuum luminosity composites have higher SFRs as

expected from the higher [O II] luminosities in the higher L_{5100} composites (see Fig. 2.5). The different Hβ FWHM bins overlap in terms of SFRs. We are also interested in comparing the slope of SFR vs. z for our sample and a sample of non-AGN galaxies. Rosa-González et al. (2002) found a best fit for the z = 0–1 range of SFR $\propto (1 + z)^{4.5}$; the line representing this fit is shown in Fig. 2.7 (with an offset of 0.5 to match our data). Our best fit is also shown in the figure with a slope of 2.3 ± 0.4 which is notably smaller than the Rosa-González et al. (2002) fit. Given the assumptions that have gone into converting our [O II] luminosities (which have an uncertain contribution from the quasar) into SFRs, it is not clear that this indicates a significant discrepancy between the samples.

The AGN Lines

With higher IPs, the [O III], [Ne III], and [Ne V] lines can be thought of as AGN lines (IPs: 35, 40, and 97 eV, respectively). After accounting for the lack of high luminosity, low redshift sources, the luminosities of [Ne III] and [O III] are almost constant with redshift as the two lower panels of Figure 2.5 show. With almost equal IPs of 35 and 40 eV, [O III] and [Ne III] are both expected to respond similarly to the ionizing radiation and consequently have similar behaviour across redshift.

We calculate Spearman correlation coefficients for each line vs. redshift and find: ρ_S([O III]) = 0.447 ($p < 0.0001$), ρ_S([Ne III]) = 0.479 ($p < 0.0001$). The slight increase in the luminosities of these two lines might be a result of the contribution of the host galaxy ISM which undergoes enhanced SFRs at higher redshift. However, a comparison of the apparent increase in [O III] luminosity with redshift to the error bars (lower left corner of Fig. 2.5), and tracing the trend within each luminosity bin indicates to us that this apparent correlation is not real.

As for the [Ne V] line (Fig. 2.5), the correlation is still significant but is much weaker with ρ_S([Ne V]) = 0.373 (p = 0.005). Table 2.4 lists all the slopes and the Spearman correlation coefficients calculated for the luminosity and Hβ bins separately for the line luminosity vs. redshift correlations.

This again supports the notion of a pure AGN line with luminosity that is not changing up to redshift 0.75. The behaviour of [Ne V] is as expected in a scenario where the ionizing radiation (and consequently the ionization parameter) is not changing over redshift. The [Ne III]/[Ne V]

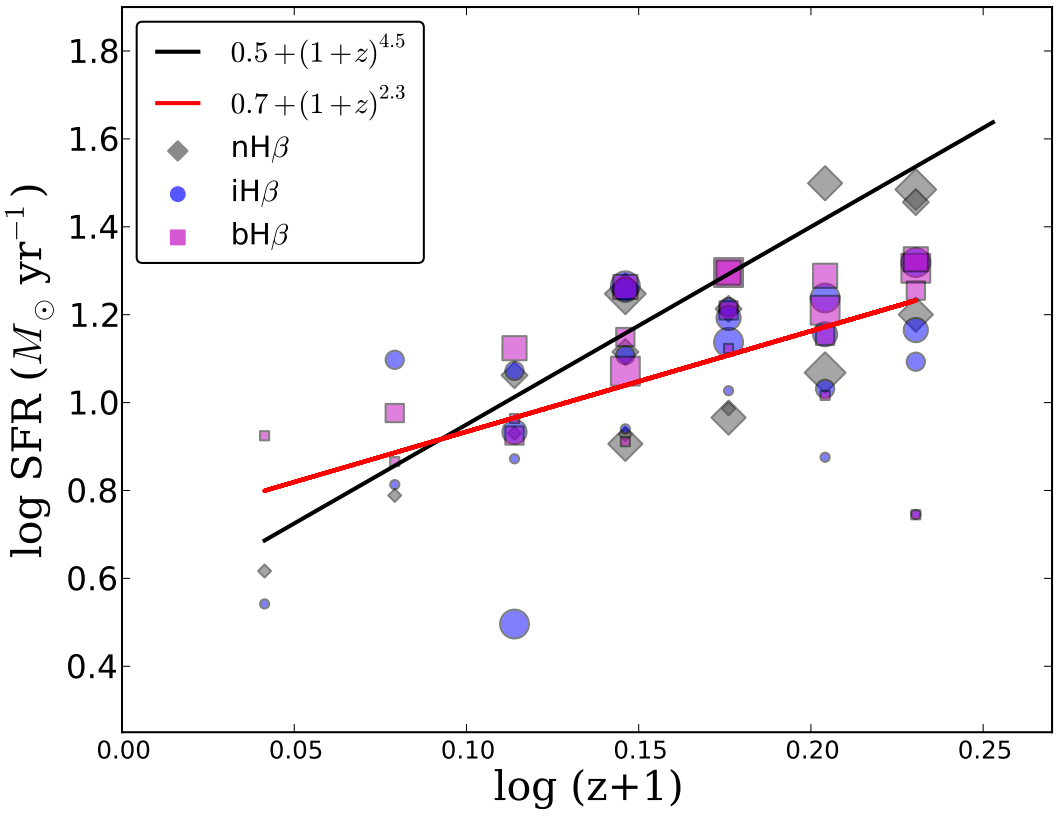

Figure 2.7: SFRs estimated with Eq. 2.4. The points are color-coded to show the narrow Hβ composites in grey, intermediate in blue, and broad in magenta. Larger points indicate higher L_{5100} bins. The solid black line is $(1 + z)^{4.5} + 0.5$ from Rosa-González et al. (2002) and the red line is our fit with a slope of 2.3±0.4 and intercept of 0.7. The figure shows a clear increase in the estimated SFRs towards higher redshift, though our sample shows less evolution than the non-AGN galaxy sample.

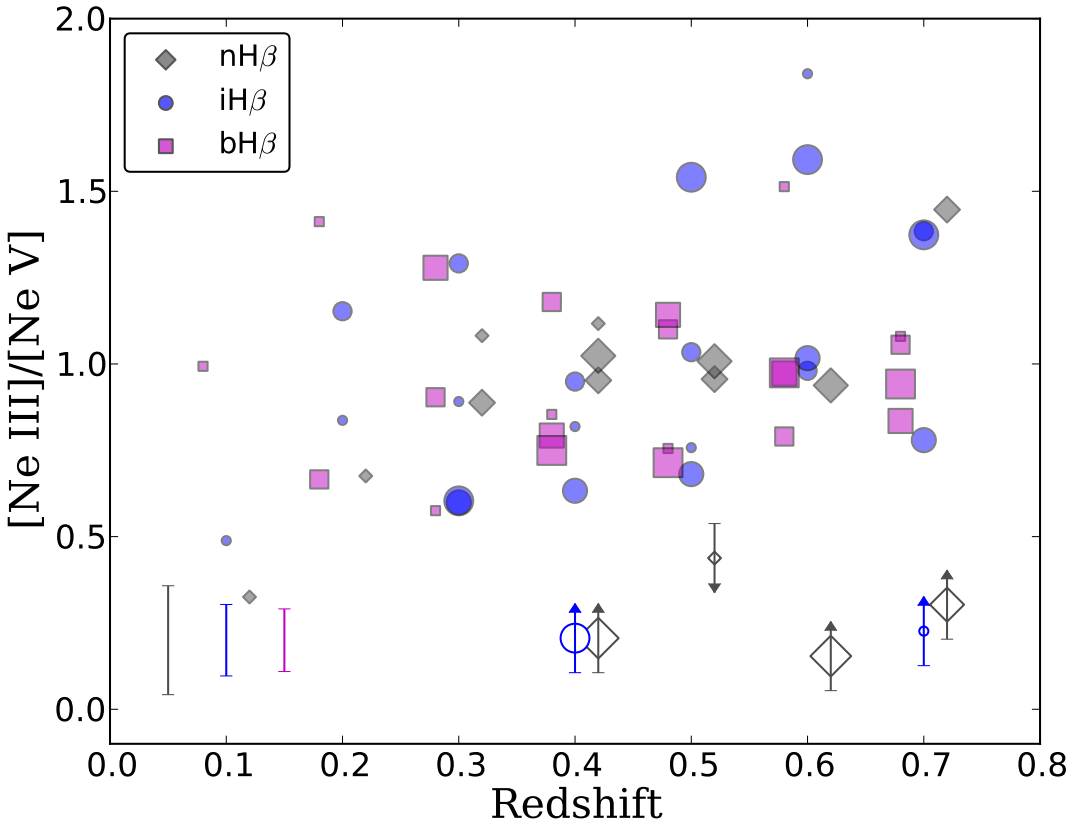

Figure 2.8: [Ne III]/[Ne V]flux ratio. The ratio does not show a clear trend with redshift indicating no evidence for a change in the ionizing spectrum with redshift.

flux ratio shown in Fig. 2.8 does indeed show evidence for no change in the hardness of the ionizing radiation in the redshift range of 0–0.75.

2.4.2 [O III], Fe II and Eigenvector 1

A suite of correlations among quasar spectral properties have been identified, most notably from Principal Component Analysis (PCA) (e.g, Boroson & Green, 1992; Boroson, 2002; Sulentic et al., 2007). One strong PCA correlation that is particularly related to the spectral range we are focusing on is referred to as Eigenvector 1 (EV1; Boroson & Green, 1992; Boroson, 2002). EV1 describes an inverse correlation between the strengths of [O III]λ5007 and the Fe IIλ4570 feature (the strength of Fe II is often quantified as the EW ratio of Fe II to Hβ,

defined as R_{FeII}). In addition to [O III] and Fe II, EV1 also includes correlations among quantities such as Hβ width, line blueshift and asymmetry, and the radio-loudness (e.g, Bachev et al., 2004). This is usually seen as objects with strong Fe II have weaker [O III], narrower, symmetric Hβ, and are more likely to be radio-quiet. In Fig. 2.9, we plot the [O III] luminosity and EW vs. R_{FeII}. The R_{FeII} values in this figure are derived from the median values of the Hβ and Fe II from the Shen et al. (2011) measurements. Figure 2.9 shows that composites with narrower Hβ do indeed have stronger Fe II. Our composite spectra (see Appendix A in the online version) also show that objects with narrower Hβ have stronger, more conspicuous Fe II. Figure 2.9 (top panel) also shows that the [O III] EW and the R_{FeII} ratio are anti-correlated –in agreement with the EV1 findings. We also examine the correlation between the [O III] and Fe II using the [O III] line luminosity (Fig. 2.9, lower panel) but we find no correlation between the two quantities. We note that the original EV1 studies (Boroson & Green, 1992; Boroson, 2002) used the equivalent width of [O III] (or its ratio to the Hβ line) to quantify its strength rather than the luminosity. The [O III] and the Fe II lines are emitted in two regions with different physical properties (e.g, density and ionization parameter) and at significantly different size scales (i.e., kpc vs. pc). Despite being repeatedly found, the presence of a correlation between these two lines has not been fully understood. The predominant factor that appears to be driving this correlation is the accretion rate (Boroson & Green, 1992; Boroson, 2002; Shen & Ho, 2014). In this picture, objects at high accretion rates have thicker accretion disks which blocks radiation from reaching the narrow-line region leading to weaker [O III] emission, however, this scenario does not account for the correlations that include Fe II.

2.5 Summary and Conclusion

In this work we examine the narrow forbidden lines in quasar optical spectra across redshift and up to $z = 0.75$. To enhance the S/N, we utilize a large sample of quasars from the SDSS-DR7 and create 62 median composite spectra (with at least 10 objects in each composite). We indirectly constrain the properties of the central source (accretion rate and BH mass) by grouping objects with similar L_{5100} and Hβ values in bins that we use to make the composites.

Our results can be summarized as follows:

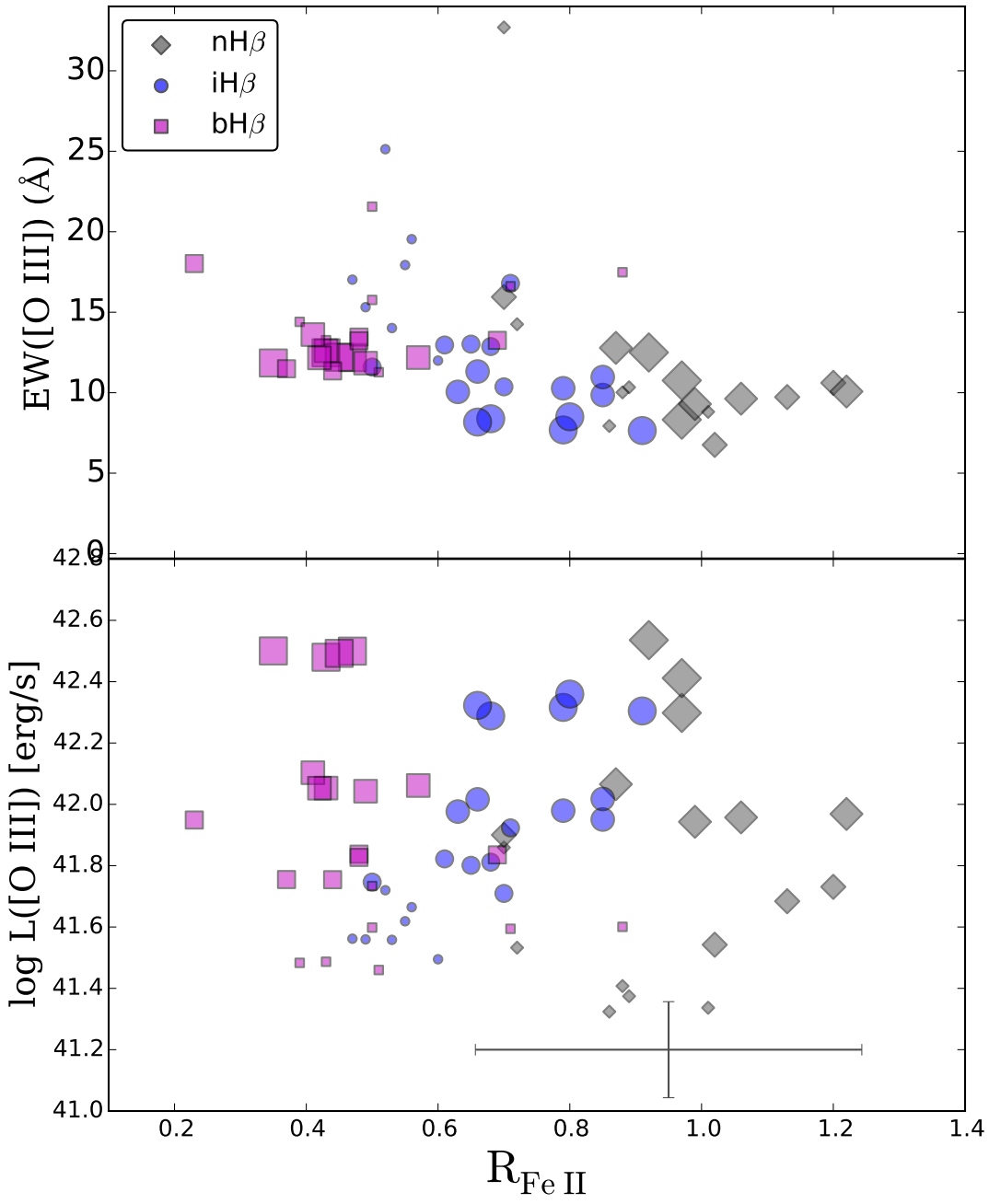

Figure 2.9: Top: The [O III] EW is plotted against the $R_{Fe\ II}$ ratio. The marker shapes correspond to Hβ width (as labeled in the legend) and the sizes are proportional to the L_{5100} bin. The two quantities are strongly anti-correlated with $\rho_S = -0.63$. Bottom: [O III] luminosity plotted against $R_{Fe\ II}$. Average error bars are shown. The two quantities appear to be uncorrelated with $\rho_S = 0.18$.

1. The luminosity of the lowest ionization narrow line, [O II], appears to increase with redshift (top-left panel of Fig. 2.5). Because the [O II] line is known to be a reliable SFR estimator in non-active galaxies, we interpret this result as evidence for higher SFRs in the host galaxies of AGNs at higher redshift, We calculate a SFR using the [O II] luminosity and the scaling relation of Rosa-González et al. (2002) and find that the estimated SFR is increasing with redshift however with a smaller power-law slope (2.3 ± 0.4) than found in non-AGN galaxies (4.5; Rosa-González et al. 2002).

2. The strength of the correlation between the line luminosity and redshift appears to depend on the IP of the line. The luminosity of the line with the lowest IP ([O II]) has the strongest correlation with redshift while the "intermediate" IP lines ([Ne III] and [O III]) have weaker dependancies and the [Ne V] with IP =97 eV appears to be independent of redshift (Fig. 2.5, see also Table 2.4). This behaviour is in line with the idea of a stratified NLR where lines with higher IPs are emitted closer to the central source (which is similar for $z < 0.75$) with negligible contribution from the host galaxy gas while lines with low IPs are emitted farther out and also have more contribution from stellar sources of ionizing radiation.

3. We find that the strength of the broad Fe II emission is anti-correlated with both the [O III] EW and the width of Hβ, i.e., objects with weaker [O III] (lower EWs) tend to have narrower Hβ and stronger Fe II (larger R_{FeII}, see Fig. 2.9). This result is in agreement with the definition of EV1 (Boroson & Green, 1992; Boroson, 2002). The presence of a correlation between these lines has been attributed to the accretion rate of the central source (L/L_{Edd}).

The study of the impact of AGNs on their host galaxies allows us to probe galaxy evolution and the possible role that SMBHs play in it. AGN narrow-line regions are the zones where the interaction between the ionizing radiation from the accretion disk and the host galaxy ISM takes place and can therefore serve as excellent probes of the AGN–host galaxy interface. The increase in the SFRs at higher redshift (up to $z \sim 1$–2) is well-known in non-AGN galaxies and appears to show similar behaviour in AGN hosts. Examining higher redshifts to explore this trend further will require near-IR spectroscopy.

Chapter 3

Insights into Quasar UV Spectra Using Unsupervised Clustering Analysis

A.Tammour, S. C. Gallagher, M. Daley, G. Richards

Monthly Notices of the Royal Astronomical Society, Volume 459, Issue 2, p.1659-1681

June 2016

3.1 Introduction

Among the different types of Active Galactic Nuclei (AGN), quasars stand out as the most luminous with typical bolometric luminosities $> 10^{46}$ erg s^{-1}. Accretion onto a supermassive black hole is now accepted as the main mechanism that powers quasars (Shakura & Sunyaev, 1973). Despite having remarkably similar spectra that typically show a distinct strong blue continuum and broad emission lines with velocity widths of 1000s km s^{-1}, quasar spectra exhibit subtle but strong trends among their emission lines and continua that persist over a wide range of wavelengths and luminosities. These repeated patterns allow us to probe the complex physical processes taking place in quasars' inner regions.

For example, the profiles of some of the broad high ionization lines (e.g., C IV λ1550) exhibit some structure which indicates that in addition to the Doppler broadening of the line by the central black hole's gravitational field, a non-virial component of motion in the gas (inflow or outflow) is at work creating this structure seen as blue asymmetries or blueshifts of the peak

from the systemic redshift measured in lower ionization lines (such as Mg II, e.g., Richards et al., 2002). Another notable example is the Baldwin Effect (Baldwin, 1977) seen as an anti-correlation between the strength of C IV λ1550 (measured by its equivalent width, EW) and the continuum luminosity at 1550 Å.

More recently it has been demonstrated that C IV EW is anti-correlated with its blueshift and both of those quantities are tied to the X-ray hardness of the quasar in a sense that objects that are soft in the X-ray (characterized by the spectral index α_{ox}[1]) are likely to have weaker C IV and larger blueshifts (e.g., Richards et al., 2002; Leighly & Moore, 2004; Leighly, 2004; Richards et al., 2011).

Moreover, a fraction of $\sim$ 20% of quasars of optically selected samples display broad ($\Delta V > 2000$ km s^{-1}) UV absorption features (broad absorption line quasars, BALQs) (e.g., Weymann et al., 1991; Hewett & Foltz, 2003). BALQs exhibit relatively weak X-ray emission (e.g., Gallagher et al., 2006) and show blueshifted broad absorption troughs with large velocity offsets indicating an absorbing medium moving outward with velocities up to 25,000 km s^{-1} (e.g., Weymann et al., 1991).

A two-component origin of quasar broad-lines can successfully account for the blueshift-EW anti-correlation of high-ionization lines (e.g., C IV) and its connection with the SED hardness. In this two-component disk-wind model, the observed broad-lines are a combination of emission from the accretion disk and emission from a fast outflowing gas that emerges close to the accretion disk and is accelerated by radiation pressure from UV photons (e.g., Collin-Souffrin et al., 1988; Murray et al., 1995; Proga et al., 2000).

3.1.1 Patterns in Multi-dimensional Space

Looking for systematic patterns appearing repeatedly among the measured variables in large multi-dimensional datasets of quasars can help elucidate the physical driver behind those trends. It is often the case in spectral studies that finding those patterns requires binning the dataset by one or two important parameters then comparing the properties of objects among the different

[1]The spectral index $\alpha_{ox} = 0.384 \log(f_{2\mathrm{keV}}/f_{2500})$, measures the slope of the flux densities at 2 keV and rest frame wavelength 2500 Å. The quantity 0.384 is the logarithm of the ratio of the frequencies at which the flux densities are measured.

bins by stacking the spectra to create (median or mean) composite spectra (e.g., Croom et al., 2002; Richards et al., 2011; Hill et al., 2014; Shen & Ho, 2014; Tammour et al., 2015b). The binning has been done in most cases in a two-dimensional space with fixed boundaries which follow "traditional" cuts such as 2000 or 4000 km s^{-1} for the width of Hβ (e.g., Sulentic et al., 2007; Tammour et al., 2015b) which is not an unreasonable choice as these parameters are found to best constrain the properties of the objects under study and to be separating meaningful classes of AGN (e.g., Boroson & Green, 1992; Boroson, 2002; Sulentic et al., 2000).

One notable method used to study multidimensional parameter space in quasar samples is Principal Component Analysis –a statistical method that takes a multidimensional dataset and finds the orthogonal axes (i.e., Eigenvectors) that minimize the variance along each projection (Hastie et al., 2009; Ivezić et al., 2014). This approach was able to uncover interesting correlations among quasar properties (e.g., Boroson & Green, 1992; Boroson, 2002; Yip et al., 2004). Boroson & Green (1992), for example, found a strong inverse correlation between the strength of the Fe II λ4570 complex and the strength of [O III] λ5007. This correlation, has been consistently found in other quasar samples and is thought to originate from the Eddington accretion rate $L_{\rm bol}/L_{\rm Edd}$ (e.g., Boroson & Green, 1992; Sulentic et al., 2000; Boroson, 2002; Shen & Ho, 2014).

In this work , we explore the use of unsupervised clustering analysis to find patterns among quasar spectral properties in the UV regime. Unlike supervised learning which uses a labelled training set to assign labels to unlabelled data, unsupervised learning does not require previous knowledge of labels — it is "learning without a teacher" (Hastie et al., 2009, Ch. 14). Clustering analysis is one of the unsupervised learning techniques that aims to find structure in multidimensional data space (Bishop, 2009; Hastie et al., 2009). In unsupervised clustering, the data is assigned to clusters according to a chosen metric of similarity such as the Euclidian distance (e.g, K-Means algorithm) or non-Euclidian metrics (e.g., Density Based Spatial Clustering, DBSC) (e.g., Hastie et al., 2009; Ivezić et al., 2014). Despite its robustness in revealing important properties about quasars, PCA assumes orthogonality among the axes and its eigenvectors are a linear combination of the input parameters which makes interpreting them physically rather non-intuitive. We therefore find that K-Means has the advantage of producing a more instinctive output that can be more readily linked to its empirical roots.

In §3.2.1 we describe our samples and in §3.2.2 we discuss in detail the K-Means algorithm and how we apply it to our quasar samples. §3.3 includes our results and in §3.4 we discuss some of the caveats of this technique. A summary and conclusion are given in §3.5.

Throughout this work we use: $H_0 = 70$ km s^{-1}, $\Omega_M = 0.3$, $\Omega_\Lambda = 0.7$ (Spergel et al., 2003).

3.2 Analysis

3.2.1 Sample Selection

Our starting point in selecting this sample is the Pâris et al. (2014) Quasar Catalog of the 10^{th} Data Release of the Sloan Digital Sky Survey (SDSS-DR10). The large size of this catalog and its improved measurements from those of the SDSS automated pipeline make it ideal for our statistical study of quasar UV spectra. The catalog contains 166,583 quasars with measurements of emission and absorption lines such as the Gaussian amplitude (AMP), equivalent width (EW), full-width at half-maximum (FWHM) and the red and blue half-width at half-maxima (RHWHM and BHWHM). These measurements were obtained with a set of principal components (eigenspectra) that were used to fit C IV, the C III] blend (modelled as a single Gaussian), and Mg II. We choose to restrict the selection to objects within the redshift range of 1.6 to 2.2 to allow us to study some of the interesting UV features such as: Lyα λ1215, Si IV λ1396, C IV λ1550, the He II λ1640 and O III] λ1664 blend, the Al III λ1857, Si III] λ1892 and C III] λ1908 blend, and Mg II λ2800 (Vanden Berk et al., 2001).

We select the initial sample (main sample, §3.3.1, 3.3.2, and 3.3.3) based on the following criteria: EW is > 0 for each of the lines C IV, C III] and Mg II, the EW error is $< 10\%$ of the measured EW for all of the three lines, the BAL flag from visual inspection (BAL_FLAG_VI) is set to 0 (no BAL quasars), and finally we use a signal-to-noise minimum of S/N > 3 at 1700 Å.

This selection gives us a sample of 4110. Fig. 3.1 shows the distribution of M_i vs. redshift in the main sample. We visually examine the individual spectra of these objects and remove 7 objects with bad spectra (missing flux) and one object that looked like a misidentified BAL quasar. In addition to this main sample, we select two other samples: the mixed sample with

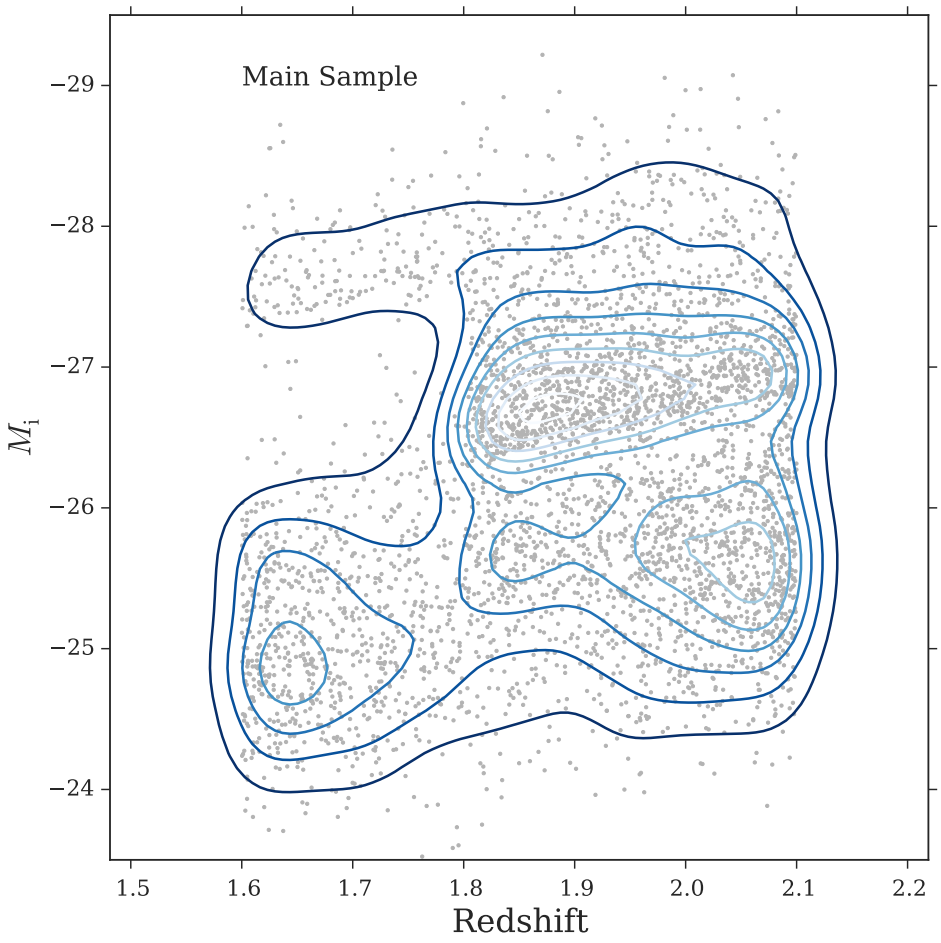

Figure 3.1: Distribution of M_i vs. z in our main sample. Given the narrow range of redshift and the SDSS selection, the luminosity distribution is relatively flat with z.

the EW > 0 and the BAL flag condition relaxed (contains 6463 objects, see §3.3.4), and the BALQ sample with the BAL quasars *only* (contains 1533 objects, see §3.3.5).

We finally note that there were many cases of heavy narrow absorption in C IV and sometimes in Mg II that were not flagged as BAL quasars in the catalog. We decide to keep those objects in our samples because we wanted to test if the algorithm is able to isolate those objects from the larger quasar population; see §3.3 for more discussion about this point.

3.2.2 Unsupervised Clustering with the K-Means Algorithm

K-Means is one of the most widely used unsupervised clustering algorithms (Hastie et al., 2009; Bishop, 2009). Part of the popularity of this algorithm is due to its simple and intuitive design. K-Means aims to minimize the "inertia criteria" (i.e., within-cluster sum of squares) given by:

$$J = \sum_{n=1}^{N} \sum_{k=1}^{K} r_{nk} \|x_n - \mu_k\|^2, \tag{3.1}$$

where μ is the mean in cluster k, N is the total number of data points (samples) and K is the number of clusters. For each point x_n, r_{nk} is an indicator that takes the values $r_{nk} = 1$ if x_n is a data point in cluster k, and $r_{nk} = 0$ otherwise. In a nutshell, the algorithm starts by assigning a fixed number of random points to serve as centroids for the clusters to be found. It then proceeds to assign data points that are closer to each centroid to a cluster. Those new cluster members are then used to calculate a new mean which becomes the new cluster centroid. The distances between the new centroids and the cluster members are calculated again and the members are reassigned to the clusters with the smallest distance to their centroids. The algorithm continues to iterate this calculating and reassigning process until it converges. A good visual illustration of this process is given in Bishop (their Fig. 9.1; 2009).

To perform the K-Means clustering, we use `scikit-learn`[2] (Pedregosa et al., 2011) applied to the EW, BHWHM, and RHWHM measurements provided by the Pâris et al. (2014) catalogue for each of the C IV, C III], and Mg II lines separately. After extensive experimentation, we chose to focus on these three features because they adequately describe the lines

[2]Using the sklearn.cluster.KMeans package: `http://scikit-learn.org/stable/modules/generated/sklearn.cluster.KMeans.html#sklearn.cluster.KMeans`

in terms of strength and structure (asymmetry). The blue-HWHM of C III] also serves as a measure of the strength of the Si III] and Al III lines blue-ward of and often heavily blended with the C III] line (we discuss this further in §3.3.3).

The only parameter that this algorithm requires beforehand is K: the number of clusters that we want (expect) our data to be grouped into. Determining this K is not necessarily a clear-cut exercise as the "ground truth" in our case is not known, i.e., the data points are not labeled. However, heuristic approaches coupled with a good understanding of the dataset and its features are sufficient to determine a range of optimal values of K.

In Fig. 3.2, we show two metrics that describe the "goodness of fit" for different values of K starting from $K = 2$ to $K = 8$. The top panel gives the sum of the square distances (errors) between data points (samples) and their cluster centroid for each attempt. K at the "elbow" of Fig. 3.2 is the one that minimizes the inertia given in equation 3.1. Ideally, one can get a good estimate of K by looking for the point where the curve bends (forms an elbow), but in most "real" datasets the curve is smooth and one can only find a range of possible K values that minimizes the within-cluster some of squares. The lower panel gives the Silhouette Score, a metric that measures the isolation of clusters by comparing the mean of the distances between a sample and its fellow cluster members, a, and the mean of distances between this sample and the ones in the nearest neighbour cluster, b (Pedregosa et al., 2011; Han et al., 2012):

$$s = \frac{b - a}{max(a, b)}. \tag{3.2}$$

This score, s, ranges from 1 for a dense, well-defined cluster to -1 for a less dense cluster that overlaps with the ones surrounding it.

Fig. 3.2 suggests that 3 or 4 clusters are reasonable for C IV and Mg II while more clusters might be needed for C III]. We proceed to use $K = 3$ to 6 for all 3 lines and we discuss in §3.3 cases where adding an extra cluster sometimes serves to isolate outliers and create a "cleaner" cluster. Table 3.1 shows a breakdown of the clustering results for each emission line or blend for $K = 3$, 4, 5 and 6.

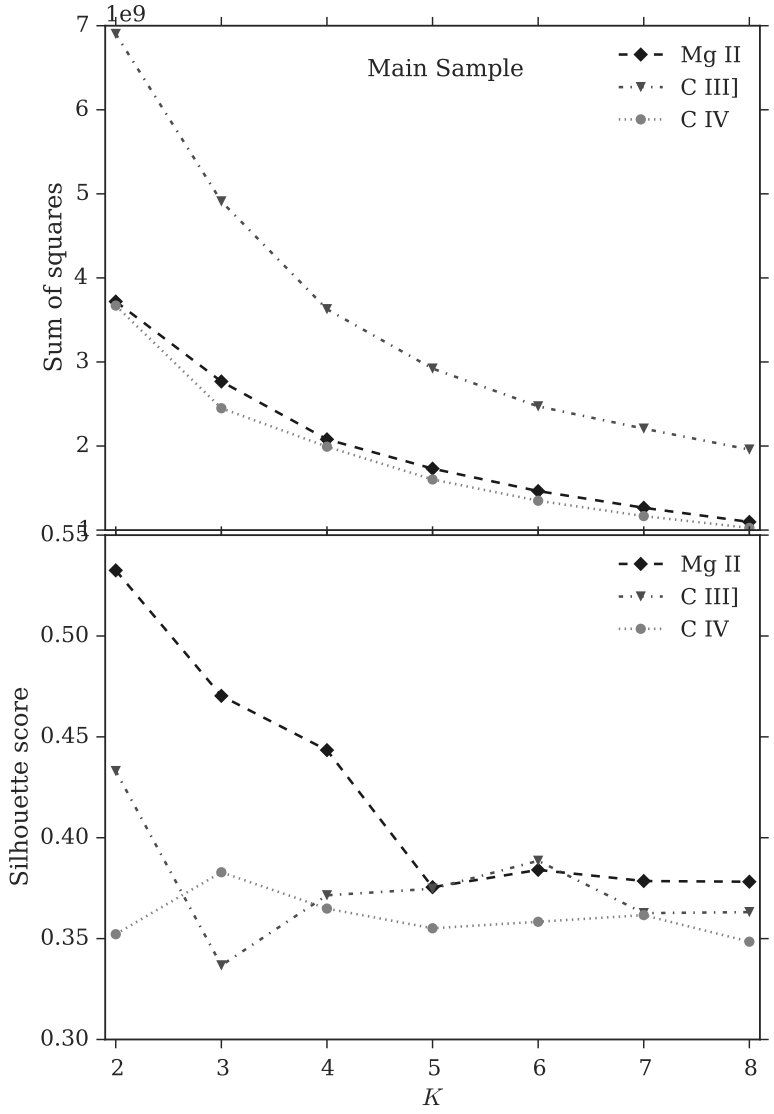

Figure 3.2: Top: The sum of the squared distances between data points and their cluster centroid (eqn. 3.1) calculated for $K = 2$ to 8 and using EW, RHWHM, and BHWHM for each of the C III], C IV and Mg II line (blend) separately. For our samples, a range of 3 to 6 is sufficient to capture the curve minimum. Bottom: Average Silhouette scores (eqn. 3.2) for all data points calculated for the clustering done on each of the three emission lines separately using $K = 2$–8. Higher scores indicate less overlap among clusters. For Mg II, K =2, 3 and 4 give the best separation, while C III] requires $K = 2$ or $K > 4$, and for C IV $K > 3$ gives the best separated clusters.

3.2.3 Median Composite Spectra

To better visualize the results of the clustering and to examine the properties of the other emission lines which were not used in the clustering analysis, we create median composite spectra from the objects in each of the clusters we find. We start by correcting all individual spectra for Galactic extinction[3] and redshift using the extinction at the g-band and the redshift values estimated from PCA as quoted in the Pâris et al. (2014) catalogue. We normalize the individual spectra using the median of the flux between 2360 Å and 2390 Å and apply 3σ clipping[4] before they are median-combined.

3.3 Results and Discussion

3.3.1 Mg II Clusters

The Mg II clusters generated using EW, BHWHM and RHWHM of Mg II are shown in Fig. 3.3 for $K = 3$, 4, 5, and 6. The figure shows the 2-D plane of the blue- and red-HWHM as the x and y-axes. The figure also shows the mean EWs for each cluster. While the EW is still contributing to the clustering (see values of the cluster centroid coordinates in Table. 3.1), the two width parameters turned out to weigh more in determining the clustering results. In Fig. 3.4, we show the $K = 5$ run only along with the BALQ quasars (not used in the clustering analysis) over-plotted and the distributions of both the full main and BALQ samples. Mg II is largely symmetric and the red and blue-HWHM of Mg II have similar distributions in the BAL and non-BAL quasar samples. The average EW of Mg II in the clusters increases with its width as Fig. 3.4 shows.

We median-stack objects in each one of the clusters shown in Fig. 3.3 and show the results in Fig. 3.5. The projected width of Mg II is the predominant quantity covering differences among clusters and is potentially probing differences in the black hole masses of objects in each cluster. However, with this change in the widths of Mg II, we see no significant change

[3]Using the Cardelli, Clayton & Mathis (1989) law.

[4]Each median spectrum only includes points contributing to the spectrum that are inside the range of 3σ standard deviations in each wavelength bin.

Table 3.1: Number of objects in each cluster formed with K-Means applied in the main sample to measurements of EW, BHWHM, and RHWHM of C III], C IV, and Mg II separately using K = 3, 4, 5 and 6. For each cluster, we also give the coordinates of the cluster centroids in the (EW,BHWHM,RHWHM) space. The labels "a", "b", "c", etc. refer to the same clusters and composites shown for the rest of this paper. Clusters are ordered according to the increasing values of BHWHM of their centroids. Clusters/composites are labeled with the name of the line used for the clustering and the number of clusters used. For example, C III]-5d is cluster d from the clustering done using K = 5 in the C III] parameter space. In Appendix B.2 we show that the algorithm is able to find the same unique centroids for the clusters after up to 50 repeats of running K-Means.

		Number of objects EW (Å), BHWHM (km s^{-1}), RHWHM (km s^{-1})					
Line	K	a	b	c	d	e	f
C III]	3	1437 (24.35,1783.97,2534.25)	2309 (26.85,2754.39,3821.80)	356 (29.68,3376.13,6692.33)	– –	– –	– –
	4	980 (23.76,1711.87,2226.60)	1919 (26.75,2177.15,3857.84)	328 (29.66,3230.03,6828.22)	875 (26.50,3668.96,3455.78)	– –	– –
	5	392 (22.90,1489.56,1471.68)	1556 (25.44,1988.50,3133.32)	1059 (27.47,2365.86,4354.74)	288 (29.67,3268.35,7016.40)	807 (26.46,3718.25,3425.02)	– –
	6	375 (22.91,1473.23,1437.42)	1468 (25.40,1937.23,3126.74)	1063 (27.52,2345.50,4337.00)	209 (29.65,2770.46,7281.64)	774 (25.91,3460.89,3191.11)	213 (28.90,4591.70,5247.12)
C IV	3	1713 (43.39,1746.90,1958.73)	1313 (31.04,2044.52,3507.01)	1076 (38.35,3304.51,2698.20)	– –	– –	– –
	4	1284 (43.97,1562.68,1880.82)	1270 (30.62,1971.40,3453.93)	1020 (41.03,2775.14,2248.06)	528 (36.29,3582.87,3349.97)	– –	– –
	5	977 (45.94,1527.30,1688.75)	1280 (33.91,1945.03,2872.95)	486 (29.37,1999.76,4103.41)	791 (42.30,2892.92,2122.82)	568 (35.79,3504.51,3279.19)	– –
	6	856 (46.92,1507.34,1617.35)	1060 (34.27,1769.15,2746.57)	417 (28.99,1923.20,4188.69)	756 (33.70,2683.73,3194.66)	730 (43.22,2859.02,2060.24)	283 (37.99,4060.17,3185.96)
Mg II	3	2503 (38.14,1795.43,1739.41)	1258 (46.39,3129.40,2591.14)	341 (51.15,3946.39,4451.96)	– –	– –	– –
	4	2176 (37.42,1757.32,1630.73)	1271 (44.83,2585.56,2628.85)	241 (51.47,3629.85,4850.80)	414 (49.41,4327.26,2591.17)	– –	– –
	5	1468 (35.36,1589.28,1498.43)	1371 (42.53,2262.72,2087.34)	707 (46.65,2818.55,3085.96)	159 (53.23,3799.77,5284.70)	397 (49.38,4393.77,2639.92)	– –
	6	1482 (35.42,1596.65,1496.80)	1366 (42.54,2235.09,2131.36)	622 (46.98,2830.91,3192.47)	77 (51.61,3100.29,6065.12)	366 (48.02,3989.87,2222.78)	189 (52.35,4747.26,3891.82)

in the strengths, widths or asymmetries of any of the other lines in the composites. C IV for example appears to have nearly identical profiles in all the composites regardless of the width of Mg II. And because of the small range of luminosity in the sample, the Mg II clusters are likely picking out BH masses. This lack of correlation between the line widths of Mg II and C IV is consistent with the known discrepancy in the estimates of black hole masses frequently done using single epoch C IV measurements (e.g., Baskin & Laor, 2005; Shen & Liu, 2012).

The absorbed peak of C IV in composite Mg II-d5 (Fig. 3.3) appears to be a result of many narrow absorption lines in C IV in this cluster that are evident in the median composite due to the low number of objects (159 objects). We visually examine the spectra of individual objects in this cluster and find that most of them have very broad Mg II profiles and frequent narrow absorption in Mg II and/or C IV. Objects in cluster Mg II-e5 (the cluster with the highest blue-HWHM) also have very broad Mg II profiles and in many cases their spectra show very weak or no emission lines. In some cases C IV and C III] are blueshifted which could point to inaccurate redshift determinations for those objects due to the extremely weak emission lines.

3.3.2 C IV Clusters

We repeat the same analysis done in (§3.3.1) but this time using the C IV parameters as the features fed to K-Means with K =3 and up to 6. In Fig. 3.6, we show the clusters for the $K = 5$ case with the BALQs over-plotted and the distributions of their blue- and red-HWHM compared to the distributions of the main sample. The C IV clusters are significantly more compact than those of Mg II with a dynamical range of $\sim$ 1000–5000 km s^{-1} for the blue- and red-HWHM. The average EW is overall smaller than that of the Mg II clusters and appears to be larger in objects with narrower C IV (Table 3.1).

The blue-HWHM of the main and BALQ samples appear to follow similar distributions (top panel in Fig. 3.6) but the distribution of the red-HWHM in the BALQ sample (right panel of Fig. 3.6) is skewed towards higher values than that of the main sample, i.e., the red-HWHM of C IV is larger in most BAL quasars than that of non-BAL quasars.

We point out, however, that the C IV emission-line measurements for BALQs can be sometimes unreliable because the absorption feature is usually fully or partially superimposed on the

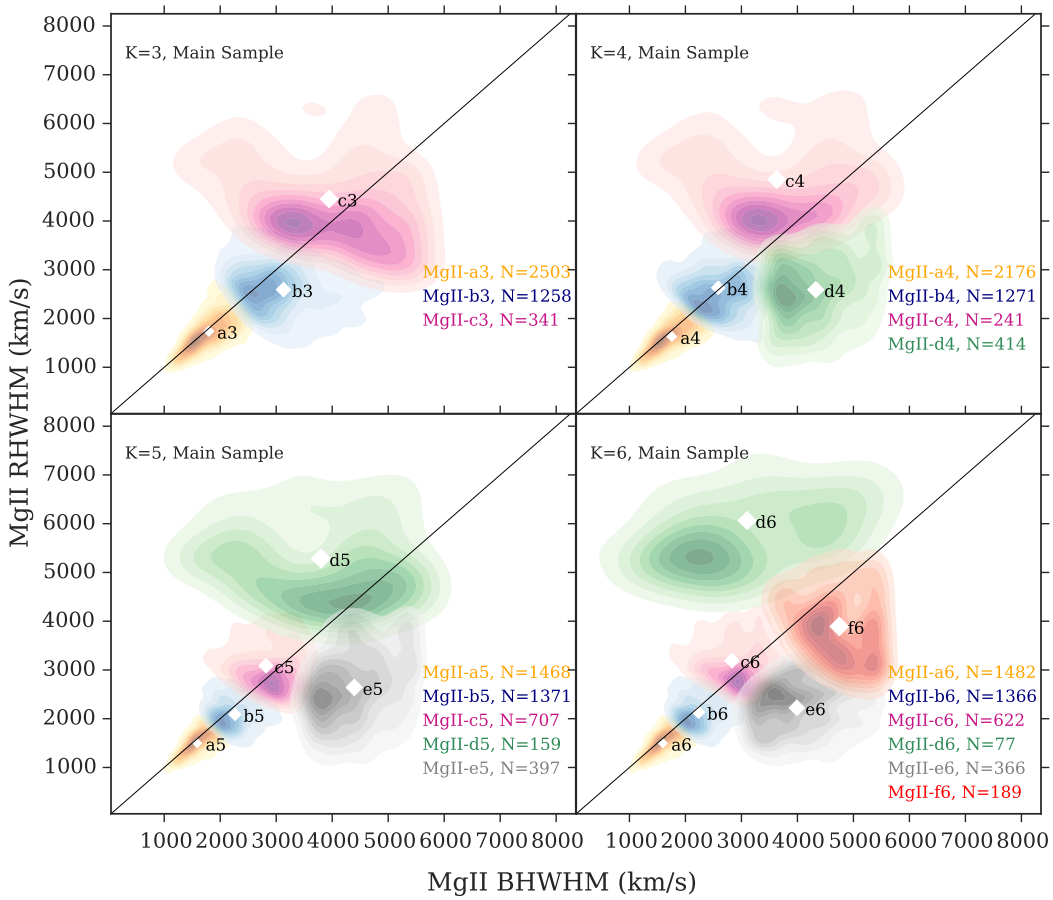

Figure 3.3: Red- vs. Blue-HWHM plane for clusters found using $K = 3$ to 6 (as labeled in the top-left corner of each panel) and the EW, BHWHM and RHWHM measurements of Mg II. The cluster labels match the ones given in Table 3.1. The white diamonds show the projected locations of the cluster centroids in this plane and their sizes are proportional to the mean EW of each cluster. The diagonal black lines mark the 1 to 1 ratio (symmetric line).

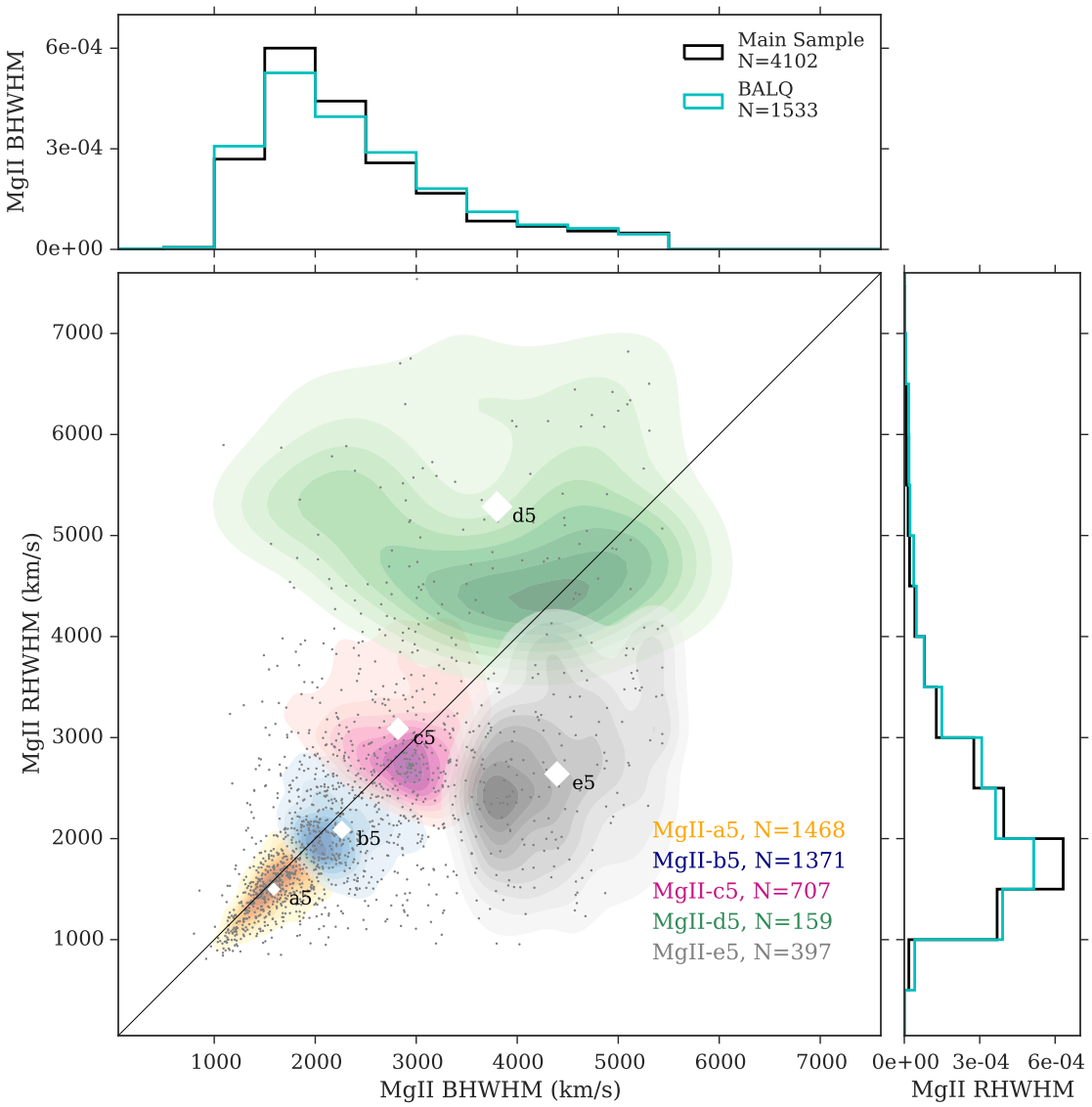

Figure 3.4: Red- vs. Blue-HWHM of Mg II for the $K = 5$ clustering run. We over-plot BALQs (not included in the main sample) as grey dots and show the fractional distributions of the Mg II blue- and red-HWHM of the main and BALQ samples in the upper and left panels respectively. It is clear that Mg II has a fairly symmetric profile except for the d5 and e5 clusters which we discuss further in the text. Both the blue- and red-HWHM of Mg II of BALQs have distributions similar to those in the main sample. The white diamonds show the projected locations of the cluster centroids and their sizes correspond to the mean EW for each cluster. The average EW of Mg II increases with its width as the sizes of the white diamonds appear to increase diagonally with the RHWHM and BHWHM.

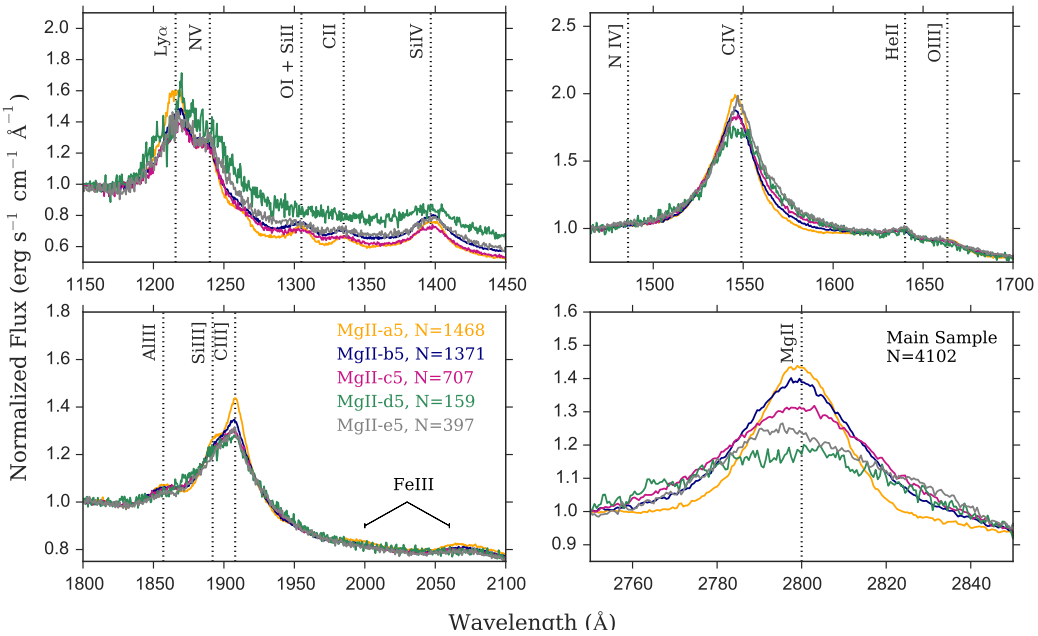

Figure 3.5: Median composite spectra made from the objects in the Mg II clusters with K = 5 shown in Fig. 3.4. The profiles have been normalized locally at the starting wavelength of each panel. The numbers in the lower-left panel refer to the number of objects in each composite/cluster (see also Table 3.1). The width of Mg II appears to be the main driver for the clustering and is potentially probing average black hole masses in each cluster. Composites Mg II-d5 and Mg II-e5 have objects with relatively high red- and blue-HWHM respectively as shown in Fig. 3.4. The weak dip in the C IV profile in composite Mg II-d5 betrays the presence of narrow absorption features in many objects in this cluster.

emission-line. This undermines the fitting of the spectral features and increases the uncertainty in the measurement of the emission-line.

Notably, Fig. 3.7 shows that despite using C IV properties *only* to generate the clusters, the other emission lines in the composite spectra generated from the clusters are corresponding to C IV in a way that is expected from observations of quasar UV spectra. For example, composite C IV-a5 in Fig. 3.7 has narrow, symmetric C IV which shows no shift from the systemic redshift and has a smaller Si III]/C III] ratio and a prominent He II (also strong Si IV, OI and SiII and narrower Mg II). While composites with blueshifted and weaker C IV (e.g., composite C IV-c5) have larger Si III]/C III] ratios and weaker He II. We point out here that the presence of a strong He II line coupled with a higher Si III]/C III] ratio is an indication of a harder ionizing SED -although other interpretations of a high Si III]/C III] ratio alone (such as a higher density) are also possible (Leighly & Moore, 2004; Casebeer et al., 2006). These observations are in line with the disk-wind model of the broad-lines in quasars which predicts that blueshifted high ionization lines are a result of a fast-moving wind emerging from the accretion disk and accelerated outward by radiation pressure, while intermediate and lower ionization lines are emitted closer to the base of the wind at the accretion disk and are therefore seen to be symmetric and at systemic redshift (Leighly, 2004; Leighly & Moore, 2004; Richards et al., 2011). The shape of the ionizing SED in the disk-wind model plays a major role in the contributions of each of the disk and wind components to the broad emission lines; quasars with higher UV luminosity relative to the X-ray have a stronger wind component (more blueshift), while quasars with stronger X-rays are over-ionized and have therefore less contribution from the wind to the broad lines (e.g., Casebeer et al., 2006; Leighly & Moore, 2004; Leighly et al., 2007; Kruczek et al., 2011; Richards et al., 2011).

We also see in Fig. 3.7 that clustering done using C IV properties results in Mg II profiles that are rather similar regardless of the C IV profiles (similar to what we see in the Mg II clusters in Fig. 3.5). This again cautions that estimates of black hole masses using C IV could result in values that do not necessarily reflect the ones estimated by the less biased Mg II. We finally point out that the weak absorption seen in the C IV profiles of composites C IV-d5 and C IV-e5 might indicate a higher number of narrow absorption among objects in these two clusters. This feature is unlikely to be due to low numbers as cluster C IV-c5 has fewer

objects but its C IV profile shows no sign of absorption. Fig. 3.6 shows that these two clusters include objects with extremely high BHWHM. We visually examine the spectra of objects from clusters C IV-d5 and C IV-e5 with the highest BHWHM and find that most of these objects have either absorption in their C IV profiles that is not flagged as BAL or have C IV profiles that are highly skewed blue-ward (could be a blueshift or absorbed flux on the red side). In Fig. 3.8 we show examples of C IV profiles for four objects with extremely high BHWHM in clusters C IV-d5 and C IV-e5. We see no clear trend in these two clusters with respect to the properties of their SED as seen by the He II and the Si III]/C III] ratio in the corresponding composites.

3.3.3 C III] Clusters

We repeat the clustering analysis here in a similar fashion to the previous two lines but with K = 3, 4, 5, and 6 and the EW, BWHM and RHWHM of C III]. Fig. 3.9 shows the results of those runs in the BHWHM–RHWHM plane for the K = 5 case. The figure also shows the distributions of the blue- and red-HWHM of C III] for the main and BALQ samples and the average EW for each cluster. The blue- and red-HWHM of C III] in BAL and non-BAL quasars is fairly consistent and the EWs of objects in the different clusters do not appear to have a major role in driving the clustering. Furthermore, the C III] measurements in the Pâris et al. (2014) catalog refer to the C III], Si III], and Al III complex measured without any deblending. This means that the EW of this blend can be similar in two objects but with significantly different contributions from the lines in the blend. The blue- and red-HWHM of the blend is also affected by this as can be seen in Fig. 3.9 –objects with extremely high RHWHM are simply objects with weaker C III] relative to the other two lines in the blend (e.g., cluster C III]-d5 vs. cluster C III]-b5). This becomes clear when we examine the composite spectra in Fig. 3.10 generated from median-combining objects in each cluster in Fig. 3.9. Again in this case, the clusters made using the C III] complex properties are reflecting similar trends to those found in the composites in Fig. 3.7 and are perhaps even more pronounced. Indeed objects with large EWs and narrow, symmetric profiles for C IV (also Lyα, He II, and Mg II; composite C III]-a5 in Fig. 3.10) have a lower Si III]/C III] ratio while objects with lower EWs and broader blueshifted

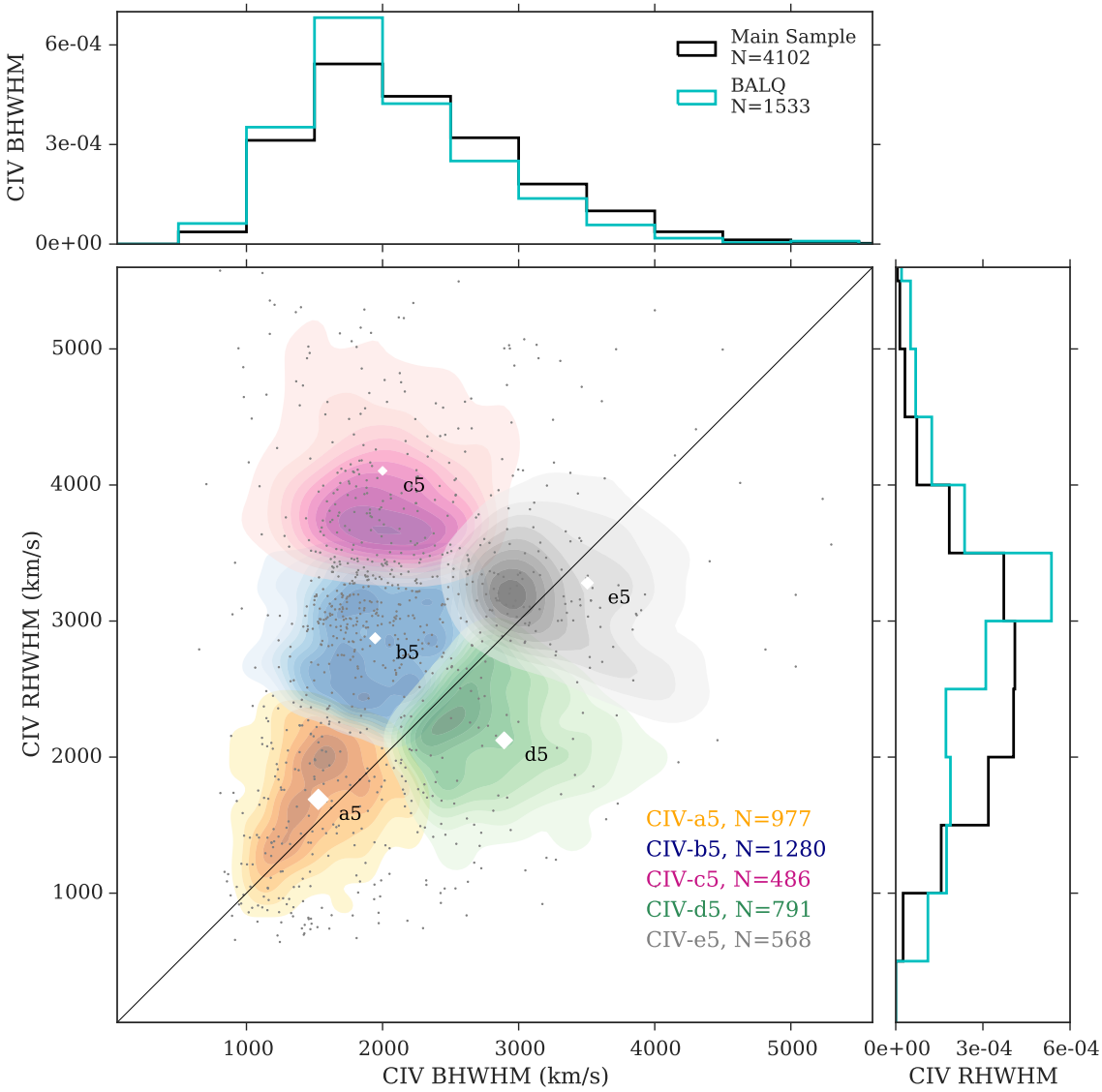

Figure 3.6: Red- vs. Blue-HWHM of C IV for the $K = 5$ clustering. We overplot BALQs as grey dots and show the fractional distributions of the C IV blue- and red-HWHM of the BALQs and the full sample in the upper and left panels respectively. The distribution of the red-HWHM in the BALQ sample (right panel) peaks at higher values than that of the main sample indicating the red-HWHM of C IV is larger in most BALQs compared to that of non-BALQs. We note that the BALQ sample contains many objects with measured HWHM of -1. We discuss these objects further in §3.3.4 and 3.3.5 and show that the algorithm was able to isolate them from the rest of the objects in the sample.

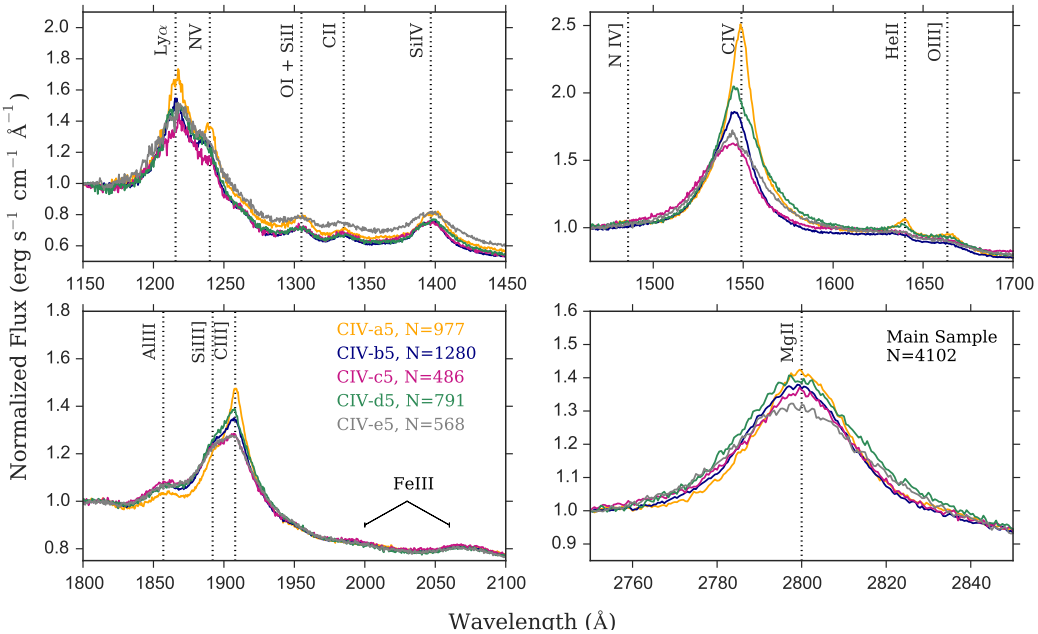

Figure 3.7: Median composite spectra made from the main sample objects in the C IV clusters with $K = 5$. The profiles have been normalized locally at the starting wavelength of each panel. The labels in the lower-left panel refer to the number of objects in each composite/cluster (see also Table 3.1). C IV profiles shift from a large EW, strongly peaked and symmetric line (C IV-a5) to a broader, weaker and blueshifted one (C IV-c5) at the extremes. C IV-a5 also shows narrow, peaked profiles in Lyα and the C III] blend with a low Si III]/C III] ratio while C IV-c5 has a higher Si III]/C III] ratio. The prominence of He II and C III] in C IV-a5 are signs of a hard ionizing continuum relative to C IV-c5. Mg II lines have nearly similar widths regardless of the shape of the C IV. The weak absorption seen in the C IV profiles on C IV-d5 and C IV-e5 is discussed in the text (§3.3.2 and Fig. 3.8).

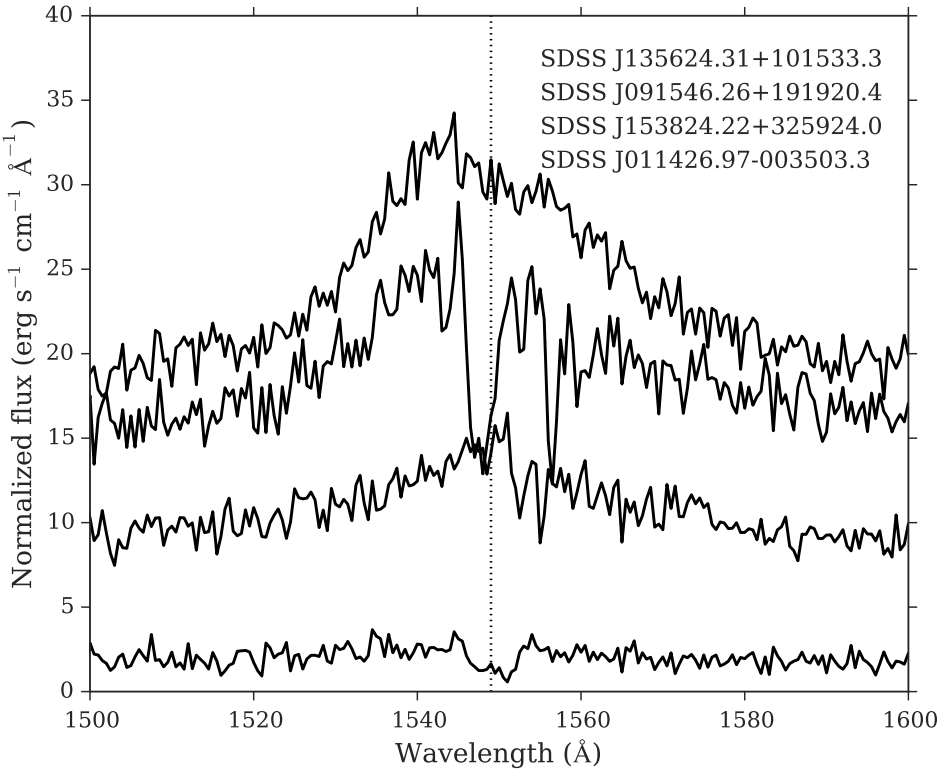

Figure 3.8: Example spectra of individual objects from the C IV-d5 and C IV-e5 composites in the main sample (Fig. 3.7). Objects in these clusters have either highly blueshifted C IV and/or have absorption that is missed from the BAL visual inspection or is not broad enough to be classified as BAL quasar.

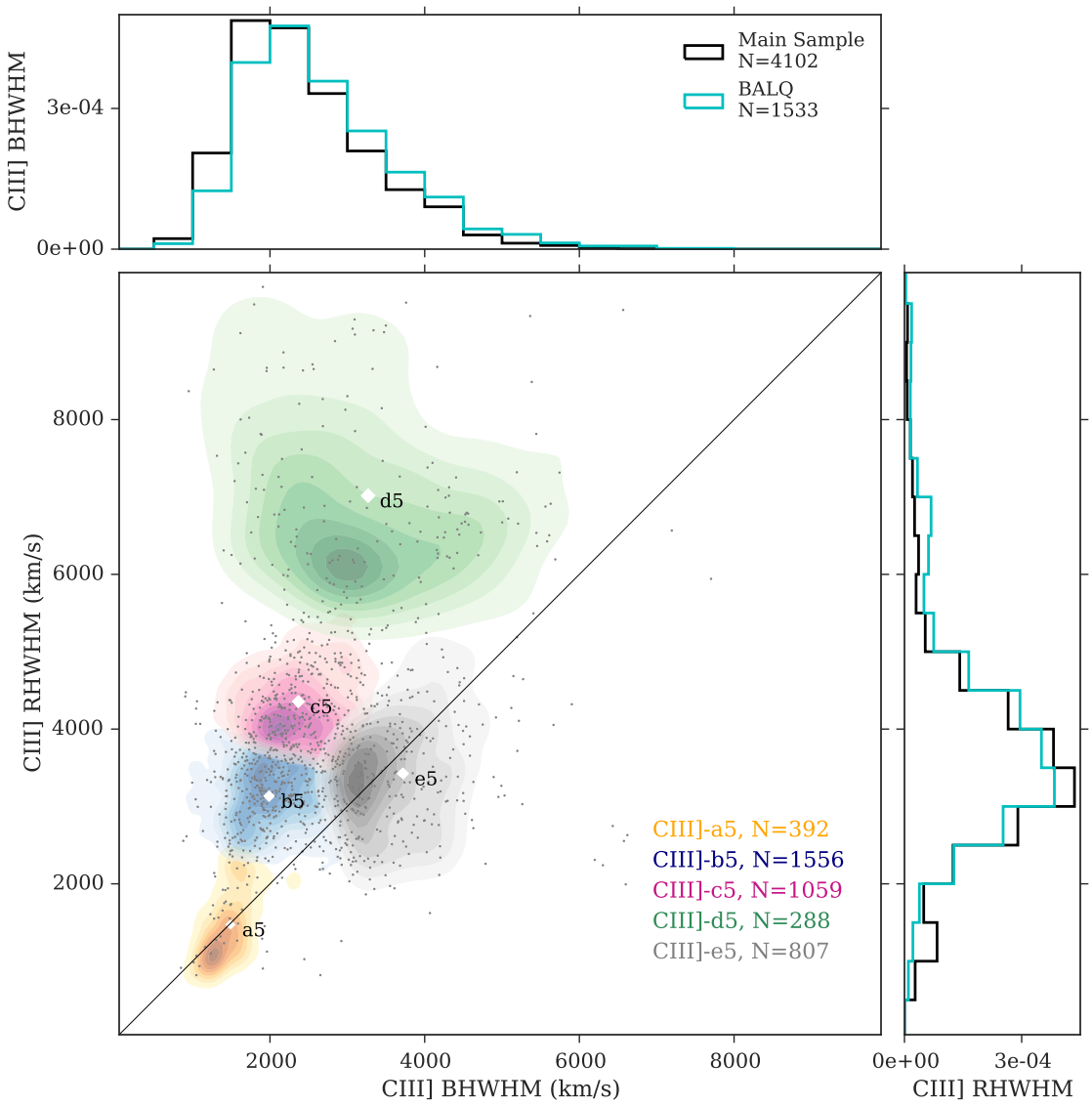

Figure 3.9: Red- vs. blue-HWHM of C III] for the $K = 5$ clustering in the main sample. The diagonal black line marks to 1:1 ratio. The white diamonds mark the projected locations of the cluster centroids and their sizes correspond to the average EW of the C III] blend of objects in each cluster which appear to be consistent among clusters. We over-plot the BAL quasars (not used in the clustering) with grey dots and show the fractional distributions of the C III] blend blue- and red-HWHM of the BALQs in the upper and right panels respectively along with those of the main sample. We see no significant differences in distributions of both samples. The larger values of the red-HWHM of the C III] blend (cluster C III]-d5) reflect a lower C III] line peak relative to the other two lines in the blend as becomes clear when we look at the composites (Fig. 3.10).

lines (composites C III]-d5 and C III]-e5 in the same figure) have a higher Si III]/C III] ratio. This suite of observations are in agreement with the disk-wind model discussed in §3.3.2 in which the shape of the ionizing SED has strong effects on the lines in a sense that a quasar with a soft ionizing SED (weak He II, high Si III]/C III] ratio, for example C III]-d in Fig. 3.10) exhibits stronger wind component (weaker C IV with larger blueshifts), while a harder SED over-ionizes the wind resulting in a stronger C IV that is centred at zero velocity (composite C III]-a; e.g., Casebeer et al., 2006; Leighly et al., 2007; Richards et al., 2011).

Another notable feature in Fig. 3.10 is the relatively strong N IV] 1486 Å line which becomes prominent in composite C III]-a5. The low Si III]/C III] ratio and the strong He II in this composite indicate a harder ionizing SED. This nitrogen is rather uncommon in quasar spectra –Bentz et al. (2004) found less than 1% of their SDSS sample has enhanced nitrogen and our visual examination of selected objects from cluster C III]-a5 confirms the scarcity of N IV] 1486 Å in this cluster. Jiang et al. (2008) studied a sample of 293 quasars with strong nitrogen emission lines (N IV] λ1486 or N III] λ1750) and found that quasars with higher nitrogen abundance share similar properties with the overall quasar population though higher nitrogen abundance is associated with higher fraction of radio-loud objects. We look at the fraction of objects detected by FIRST and find that indeed, cluster C III]-a5 includes a slightly higher fraction of quasar detected by FIRST (12.8%) compared to the other clusters which contain a range of 7.1% to 8.7% FIRST-detected objects.

3.3.4 Clustering on the Mixed Sample

Next, we set the BAL flag to 1 in the selection process mentioned in §3.2.1, allowing BALQs to be included in the sample. We also remove the constraint that EW > 0 and keep the redshift limit (1.6–2.1) and the S/N and EW uncertainty cutoffs: S/N(λ1700) > 3, EW error < 10% of measured EW. Adding BAL quasars to the sample increased its size to 6463 (not including the 15 objects we previously removed from the main sample). The purpose of defining this mixed sample is to examine the C III] properties of the entire quasar population (BAL and non-BAL quasars). We do not use C IV and Mg II measurements for the clustering analysis here because (by definition) at least one of these lines shows absorption in BALQs and can therefore bias

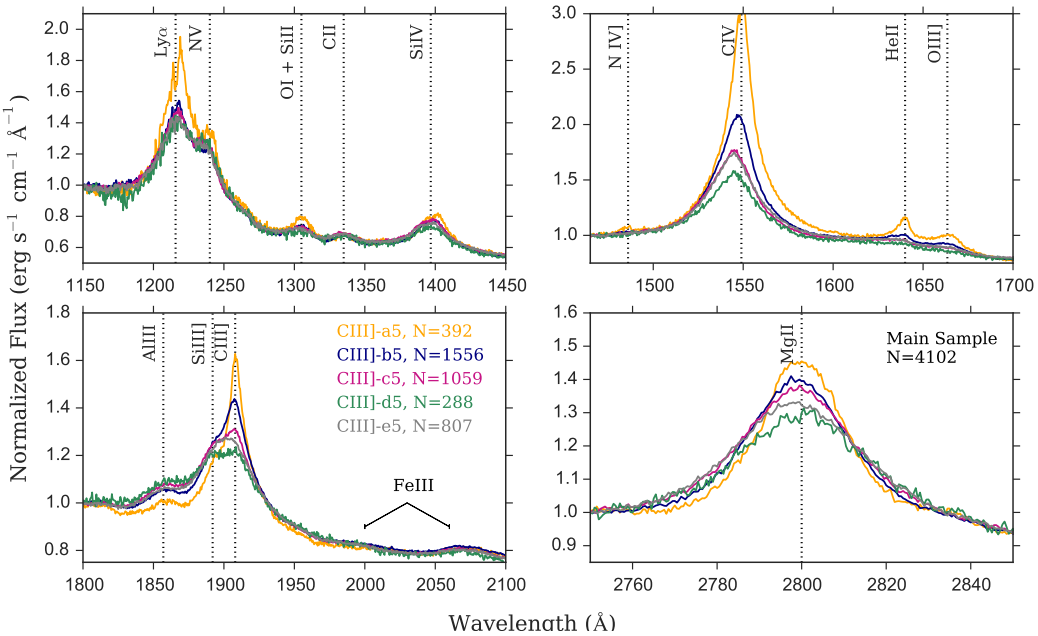

Figure 3.10: Median composite spectra made from the objects in the C III] clusters with $K = 5$ shown in Fig. 3.9. The profiles have been normalized locally at the starting wavelength of each panel. The numbers in the lower-right panel refer to the number of objects in each composite/cluster (see also Table 3.1). Composite C III]-a5 with large EWs and narrow, symmetric profiles for C IV, Lyα, He II, and Mg IIhas lower Si III]/C III] ratio while composite C III]-d5 has objects with smaller EWs and broader blueshifted lines with higher Si III]/C III] ratio indicating a softer ionizing continuum and a larger wind contribution to the C IV profile.

the results.

Fig. 3.11 shows the clusters using the C III] blend properties in the mixed sample for $K = 6$ run. We find that in the mixed sample, the algorithm is picking up objects with negative values for the three clustering features (i.e., the EW and the blue- and red-HWHM of the C III] blend). It is only when $K = 6$ that cluster C III]-a6 is almost purely made of the "negative" objects (971 objects and only one object with non negative values (22.6 Å, 741 km s^{-1}, 695.5 km s^{-1})). Of these objects $\sim$ 33% are flagged as BALQs in the Pâris et al. (2014) catalog while the rest are not marked as BALQs. Visual examination of the spectra of individual objects in cluster C III]-a6 shows that many of them have "flat" spectra (no emission lines and in some cases what looks like weakly absorbed lines) and it is possible that the negative measurements are cases where the automated fitting routine was not able to find lines to fit either because the lines are very weak or because of the strong BAL troughs. Objects with BAL troughs could be easier to misfit as the absorption troughs could be highly blueshifted and cause the fitting routine to fail. Some of the objects in cluster C III]-a6 with the weak emission lines might potentially meet the definition of the so-called weak-lined quasars (e.g., Luo et al., 2015; Shemmer & Lieber, 2015, and references therein). This is supported by the presence of a more prominent Fe III feature blue-ward of C III] known to be stronger in weak-lined quasars as can be seen in Fig. 3.12 which shows the composite spectra generated from objects in each of the clusters.

It is also intriguing to see that the objects collected in cluster C III]-a6 (with the -1 values) are generating a weak, strongly blueshifted C IV which may contain absorption features as can be seen from the dip in Lyα. This weak C IV line is also associated with the largest Si III] to C III] ratio indicating a soft ionizing SED on average in those objects compared to the other clusters. Similar to what we see in the main sample (§3.3.2 and §3.3.3), the cluster with the narrow, strong C IV line which is centered at the systemic redshift (composite C III]-b6) has the strongest He II and the smallest Si III] to C III] ratio consistent with a hard ionizing continuum which inhibits wind formation.

Finally, it is worth mentioning that the fraction of BAL quasars in each cluster appears to support the notion that BAL quasars are more likely to have softer (X-ray weaker) SEDs as we find that the numbers of BALQs in each mixed cluster decreases gradually with the Si III]/C III] ratio; cluster C III]-b6 in Fig. 3.12 (with the strongest He II and lowest Si III]/C III] ratio) has

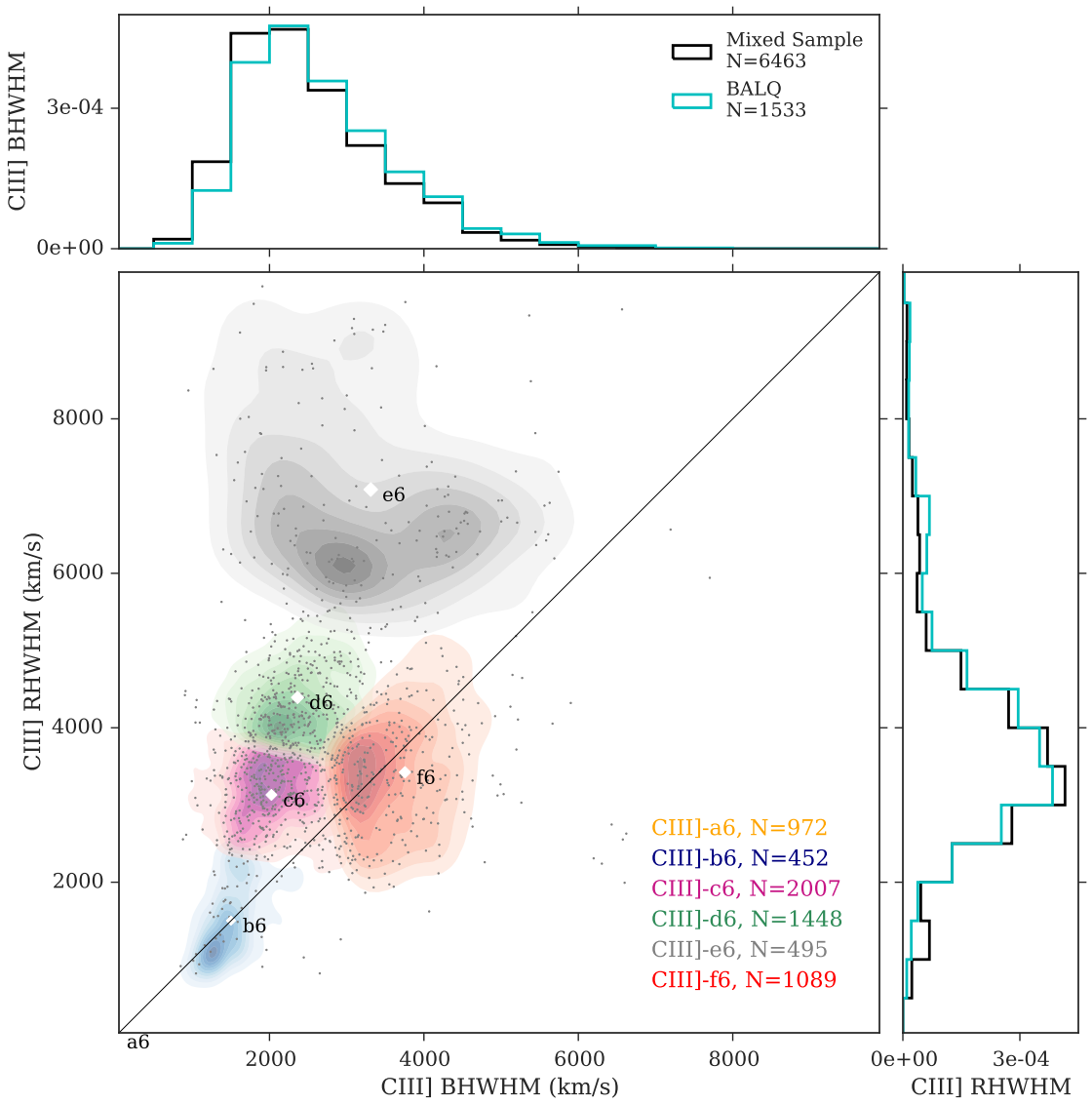

Figure 3.11: The red- vs. the blue-HWHM of the C III] blend in the mixed sample for the clustering done using $K = 6$. The white diamonds show the projected position of the cluster centroids and their sizes correspond to the the average EW of the C III] blend in each cluster. Cluster C III]-a6 (with the negative measurements) is not plotted. The BAL quasars (included in this sample) are over-plotted as grey dots. The histograms show the fractional distributions of the BHWHM and RHWHM in the mixed and the BALQ samples and show no significant difference. The fraction of BALQs varies among the clusters: C III]-a6 (33%), C III]-b6 (11%), C III]-c6 (19%), C III]-d6 (24%), C III]-e6 (33%), C III]-f6 (23%). Those differences in the BLAQ fractions potentially reflect differences in the hardness of the SED (see §3.3.4 and Fig. 3.12).

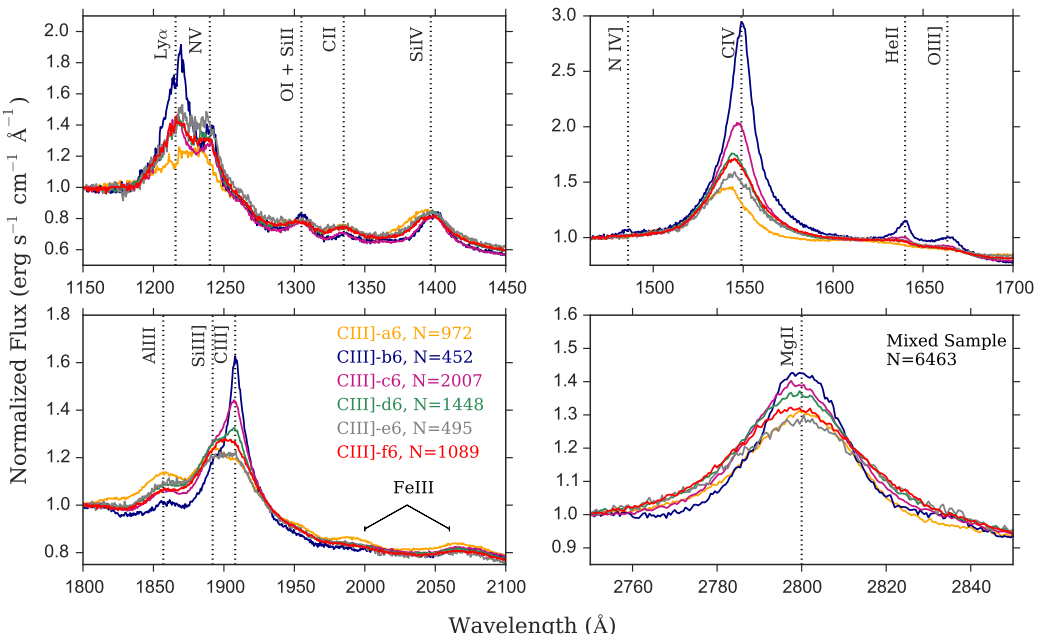

Figure 3.12: Median composite spectra made from the objects in the C III] clusters shown in Fig. 3.11 with $K = 6$ in the mixed sample. The profiles have been normalized locally at the starting wavelength of each panel and the numbers in the lower-left panel refer to the number of objects in each composite (cluster). Composite C III]-b6 has prominent narrow features such as C IV, Lyα, and C III] and weak Al III and Si III] (low Si III]/C III] ratio), while composites C III]-e6 and C III]-f6 have weaker C IV that is blueshifted and stronger Si III]/C III] ratios. Composite C III]-a6 is created from the objects with the -1 measurements for C III] reported in the catalog. The relatively strong Fe III feature in this cluster might indicate that weak-line quasars comprise a large portion of objects in this cluster.

only 11% BALQs while clusters C III]-a6 and C III]-e6 (with relatively weaker He II and higher Si III]/C III] ratio) have each 33% of its objects flagged as BALQs in the catalog. The percentage in the rest of the clusters is: C III]-c6 has 19% BALQs, C III]-d6 has 24% BALQs and C III]-f6 has 23% BALQs.

3.3.5 Sample with BAL Quasars Only

We now look at the properties of 1533 BALQ selected as described in §3.3.4 but with the BAL flag set to 1. Here, too, the algorithm is separating objects with -1 measurements for the EW and widths. Up to $K = 6$, the cluster includes 8 objects with > -1 values with the remaining 327 objects with -1 for the measurements of C III] EW and blue- and red-HWHM.

Figure 3.13 shows the median composite spectra generated from objects in the $K = 5$ run. We find that the C IV profiles (both in emission and absorption) are changing with the properties of the C III] blend used to generate the clusters. Composite C III]-b5, for example, with the lowest Si III]/C III] ratio, has a strong, narrow C IV emission-line centred at the systemic redshift and a deep absorption trough with low V_{min}. On the other hand, composites with larger Si III]/C III] ratios (C III]-d5 and e5) have blueshifted C IV emission-lines and higher V_{min} for their absorption troughs. Composite C III]-5a –mostly comprised of objects with the -1 measurements of C III]– has the highest Si III]/C III] ratio and shows weak, blueshifted C IV emission-line and a shallower trough and a very high V_{min}. The absorption troughs in Si IV appear to follow similar trends of those seen in C IV. The lack of any absorption showing in the Mg II profiles in those composites indicates that objects in these clusters contain either only high ionization BAL quasars or a mixture of both high and low ionization BAL quasars that averages out in the Mg II profiles.

In future work (Tammour et al., in prep), we explore the properties of the C III] complex and the absorption lines in detail for a sample of BAL quasars.

3.4 Caveats

The K-Means algorithm is an unsupervised clustering method. This means that, by definition, there are no fixed labels to the objects in the clusters defined by K-Means and it is the investi-

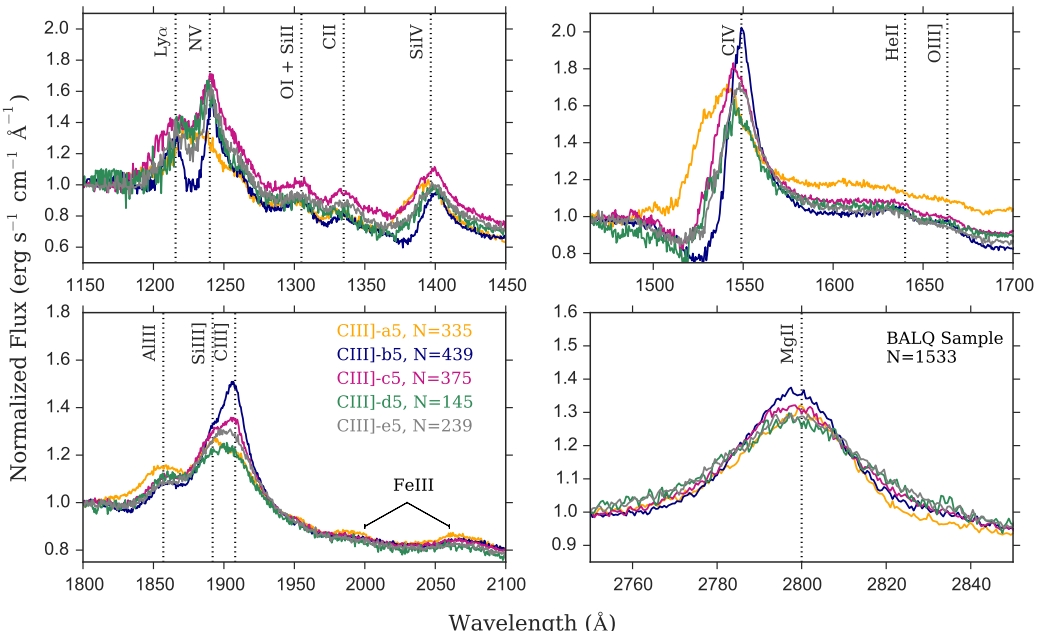

Figure 3.13: Median composite spectra made from the objects from the BALQ sample in the C III] clusters with $K = 5$. The profiles have been normalized locally at the starting wavelength of each panel and the numbers in the lower-left panel refer to the number of objects in each composite (cluster). Composites made from clusters of the C III] blend properties appear to have different emission and absorption in C IV and Si III], e.g., composite C III]-b5 has the lowest Si III]/C III] ratio and a narrow C IV emission-line that is not blueshifted and a low V_{min} BAL though, while composites C III]-d5 and C III]-e5 have a higher ratio and are both blueshifted in their C IV emission and their BAL troughs have higher V_{min}. Composite C III]-a5 mostly contains objects with -1 measurements for the C III] blend and it shows a weaker C IV emission and a shallow BAL trough with the highest V_{min}. Si IV shows similar trends to those seen in C IV.

gator's judgement whether or not to accept the clusters found by the algorithm. The two main decisions to be made beforehand are what features to use to define the parameter space, and how many clusters to group objects into. Deciding which features to use is limited to what is available and is strongly related to the problem in hand. In this work, we use the EW and the blue- and red-HWHM of individual lines or line blends. We decide to use these three parameters as they adequately represent the relative strength and structure (asymmetry) of the emission lines. Determining K (the number of clusters) is a heuristic exercise and requires familiarity with the dataset and the scientific question in hand, and a few iterations between choosing a K and examining the output. In our case, we find that even though $K = 3$ gives nicely separated clusters, increasing K to 5 or 6 allowed us to isolate some of the outliers that were otherwise "hiding" in the clusters and became only visible to us when we created composite spectra from the objects in clusters. The two metrics we use (discussed in §3.2.2) are heuristic and are useful for guidance –this is why we opt to use a range of K's rather than one single value. In choosing $K = 5$ (and $K = 6$ for the mixed sample) we try to find a balance between a low number of clusters that could hide some of the interesting objects and an unnecessarily high number which would only fracture the clusters into smaller bits that do not hold any interesting properties from the rest.

Finally, we note that results based on machine learning techniques strongly rely on the input data fed to the algorithm. For this work we use measurements from the Quasar properties catalog of Pâris et al. (2014). This catalog includes results of automated measurements of 166,583 quasars and cases where the fitting routine failed to find a proper fit or where broad absorption was not properly flagged are not surprising. We in fact realized that one additional advantage of using K-Means is its ability to pick those objects and isolate them from the rest of the dataset. Indeed one of the challenges of working with large datasets is identifying outliers that could potentially be interesting peculiar cases or simply inaccurate measurements that went unnoticed. An example of this is isolating a group of objects with -1 measurements of their EW, blue- and red-HWHM that were not flagged in the Pâris et al. (2014) as we show in §3.3.4 and 3.3.5.

3.5 Summary and Conclusion

In this work we explore the use of unsupervised clustering analysis in searching for patterns among a large number of quasar UV spectra in a multidimensional space. The K-Means algorithm allows us to find clusters in the EW, BHWH, and RHWHM parameter space of three different UV emission lines and line blends: C IV 1450Å, C III] 1903Å, and Mg II 2800Å (see §3.2.2). We combine objects in each of the clusters we find in median composite spectra to examine the properties of the emission lines in each cluster. We note that we are not claiming that the 5 clusters we use in the main analysis represent distinct classes of quasars but rather our composite spectra show in all of the discussed cases that the line properties move gradually between two extremes. The interesting part of the analysis comes from using one line's strength and asymmetries to probe physical properties of the objects (such as the case for the C III] blend probing the hardness of the SED) and seeing the effects of this on other lines not used in the clustering (such as C IV).

We summarize our findings as follows:

- K-Means is a simple yet powerful algorithm which, instead of binning data at fixed pre-determined boundaries, allows us to more freely explore the structure in a multidimensional parameter space and to find clusters of objects with similar properties in this space.

- When stacked in composite spectra, the clusters we find show some of the well-known trends in quasar UV spectra such as the correlations among the C IV blueshift and its EW and the shape of the ionizing SED probed by the strength of He II and the Si III]/C III] ratio (Figures 3.7 and 3.10).

- More interestingly, we find this same inverse correlation in the C IV line using emission lines that are seemingly not part of this correlation such as the C III] blend (Fig. 3.10). Because C III] is generally not contaminated by absorption, it is perhaps more useful for finding like objects in a quasar sample.

- We find that, unlike C IV and C III], the properties of Mg II are not strongly correlated with those of the other lines in the spectra; the width of C IV for example, does not show

any clear correlation with that of Mg II. This could lead to potential discrepancy in the determination of black hole masses using broad lines in single epoch spectra (Fig. 3.5).

- We use the technique to examine the properties of C III] in a mixed sample of BAL and non-BAL quasars and find that the properties of C III] in the mixed sample recovered similar trends to those we find in the main non-BALQ sample (Figures 3.11 and 3.12).

- In the mixed sample, we find a higher fraction of BALQs in composites with weaker He II and higher Si III]/C III] ratio (softer SEDs), while composites with stronger He II and lower Si III]/C III] ratio (harder SEDs) have a lower fraction of BALQs (§3.3.4, Fig. 3.11).

- We find that a somewhat more prominent NIV] 1486 Å feature is seen in some of the composites generated from the C III] clusters (composite C III]-a5 in Fig. 3.10, and composite C III]-a6 in Fig. 3.12). The stronger HeII and higher SiIII/CIII] ratio indicate a hard(er) SED in those objects and the appearance of stronger NIV] is not surprising given that the IP for the NIV] λ1486 line is 47.4 eV (Karen Leighly, private communication). Higher nitrogen abundances have been inferred before in radio-loud objects and we find a slightly higher fraction of FIRST-detected objects (12%) in those clusters compare to the ones with weaker nitrogen (7-8%).

- When we apply the clustering to a sample of BALQs only using the measurements of C III], we find evidence that the properties of C III] are able to separate objects into groups with different properties of their C IV emission and BAL troughs (Fig. 3.13). We investigate this result further in Tammour et al. (in prep).

Chapter 4

Constraining the Diversity of Trough Profiles in Broad Absorption Line Quasars with Unsupervised Clustering Analysis

A. Tammour, S. C. Gallagher, N. Filiz Ak, M. Daley, K. Leighly, G. T. Richards

To be submitted to the Monthly Notices of the Royal Astronomical Society

4.1 Introduction

Quasars are the most luminous class of active galactic nuclei with typical bolometric luminosities on the order of 10^{46} erg s^{-1}. Intrinsic broad absorption in high-ionization transitions is seen frequently in quasar UV spectra: nearly 10-20 per cent of the objects in optical quasar samples show broad blueshifted absorption in their UV spectra (e.g., Hewett & Foltz, 2003; Trump et al., 2006; Ganguly et al., 2007; Gibson et al., 2009). Formally, an object is identified as a broad absorption-line (BAL) quasar when the width of its C IV 1449 Å absorption trough is greater than 2000 km s^{-1} at least 10% below the continuum level (Weymann et al., 1991).

BAL quasars with absorption in high ionization potential (IP) transitions *only* such as C IV 1449 Å (IP= 47.9 eV) or Si IV 1394 Å (IP=33.5 eV) are referred to as HiBALs. A

smaller subset of BAL quasars display, in addition to the high IP transitions, absorption from lower IP transitions such as Al III 1857Å (IP=18.8 eV) and/or Mg II 2800Å (IP= 7.6 eV). This subset is referred to as LoBALs. The similar properties of continuum and emission lines in both BAL quasars and non-BAL quasars indicate that objects with broad-absorption lines are in essence drawn from the same population as those with no BALs (e.g., Weymann et al., 1991). BAL quasars however are found to be more reddened on average (e.g., Weymann et al., 1991; Sprayberry & Foltz, 1992; Ganguly et al., 2007; Gibson et al., 2009) and X-ray weaker than non-BAL quasars (e.g., Gallagher et al., 2006; Gibson et al., 2009), with LoBAL quasars being significantly more reddened (e.g., Weymann et al., 1991) and X-ray weak than HiBAL quasars (e.g., Gallagher et al., 2006).

Current quasar models suggest that these absorption features result from fast-moving outflows that emerges from the accretion disk (at a distance of $\sim 10^{16}$ cm for a black hole of mass 10^8 $M_\odot$) and are accelerated outwards by radiation pressure (e.g., Collin-Souffrin et al., 1988; Murray et al., 1995; Proga et al., 2000). The presence of the wind despite the high energy radiation field in the close vicinity of the central source is explained by a layer of shielding gas that filters energetic photons preventing the gas from becoming overionized, and allowing UV line-driving to accelerate the winds radially up to velocities of $0.1c$ (e.g., Murray et al., 1995; Proga et al., 2000; Leighly, 2004; Leighly & Moore, 2004).

Despite the many successes quasar wind models have made in explaining observational properties of BAL troughs, questions about the geometry and driving mechanism still persist. One standing issue is the location at which the winds are launched (at what radii) and how does the launching radius affect the observed wind properties (e.g., column density, ionization structure, and geometry). Robust quasar wind models suggest that the onset radius of the outflow is $\sim 10^{16}$ cm (e.g., Murray et al., 1995; Proga et al., 2000). Constraining the properties of winds within the BAL quasar population and the types of objects hosting the different types of outflows is therefore essential to a better characterization of this phenomenon and to the overall physical nature of quasar systems.

Studies of variability, observed frequently in BAL quasars on timescales of months to years (e.g., Barlow, 1994; Gibson et al., 2008; Filiz Ak et al., 2013), are essential to understanding the structure and dynamics of quasar winds. On rare occasions, variability events have been

observed as a complete appearance or disappearance of BAL troughs (e.g., Leighly et al., 2009; Hall et al., 2011; Filiz Ak et al., 2012). Changes in the ionizing radiation could result in changes in the BAL troughs. However some lines of evidence, such as the presence of P V $\lambda\lambda$1118, 1128 absorption which is significantly less abundant than C IV (Hamann, 1998), suggest that BAL troughs are often saturated (even if the BALs do not reach zero intensity), and it is unlikely that changes in ionization are fully responsible for the observed BAL quasar variability (e.g., Leighly et al., 2009). Another plausible scenario suggests that changes in the covering fraction (i.e., how much of the ionizing continuum source is covered by the outflows) cause variations in BAL troughs. Changes in the covering fractions can result from changes in the geometry of the outflows as a result of rotation which moves the wind into or out of our line-of-sight (e.g., Hall et al., 2013, and references therein). The observed variability in BAL quasars is likely a result of a combination of ionization levels and dynamics (e.g., Capellupo et al., 2012).

In this work, we analyze the spectroscopic properties of BAL quasars using a large sample from the Gibson et al. (2009) BAL Quasar catalogue. We are interested in examining how the properties of a single BAL feature (i.e., the C IV trough) can probe the underlying physical conditions in BAL quasars –in particular those related to the winds generating the UV broad absorption in quasar spectra. To do this, we use an unsupervised clustering approach, namely K-means which allows us to search for structure in a multidimensional parameter space.

4.2 Analysis

The Gibson et al. (2009) catalogue contains 5,604 objects identified from the Sloan Digital Sky Survey, Data Release 5 (SDSS-DR5) with measurements of four absorption features: Si IV λ1400, C IV $\lambda\lambda$1400,1549, Al III $\lambda\lambda$1400,1857, and Mg II λ2800. The measurements in the catalogue include quantities such as the absorption trough equivalent width (EW), the outflow minimum and maximum velocities (V_{min}, and V_{max}) and the modified[1] balnicity index (BI_0) which measures the strength of the trough (similar to the EW) taking into account the width of the absorption feature at at least 10% of the continuum level. In this catalogue, BI_0 is

[1]In the original BI_0 definition, the minimum velocity is -3000 km s^{-1}.

defined as:

$$\mathrm{BI}_0 = \int_0^{25000} \left(1 - \frac{f(-V)}{0.9}\right) C dV, \tag{4.1}$$

where $f(V)$ is the continuum normalized flux as a function of velocity V from the line's systemic velocity, C equals 1 only when the trough velocity-width at 10% below the continuum is ≥ 2000 km s^{-1} otherwise C is set to zero (Weymann et al., 1991; Gibson et al., 2009).

We use measurements of the EW, V_{min}, and V_{max} of the C IV trough –typically the strongest absorption feature in BAL quasars– to isolate objects that are "similar" in this parameter space. Used together, these three parameters are able to uniquely capture the shape of the trough in terms of its width, depth and velocity shift, etc. We elaborate on our sample selection in §4.2.1 and discuss the K-means algorithm in §4.2.2.

4.2.1 The Samples

Starting from the Gibson et al. (2009), we choose our dataset to have a redshift range $1.79 > z > 3.7$ which allows us to examine the Si IV, C IV, Al III, and C III] region. This part of the spectrum includes both high and low IP lines and so it enables us to assess the effects of different C IV trough profiles on the other lower IP lines; their presence or absence, and their shapes and strengths. In addition to the redshift limits, we also require a signal-to-noise of at least 3 at 1700 Å and a balnicity index $\mathrm{BI}_0 > 0$ for the BAL trough of C IV. This results in 2,683 objects. Fig. 4.1 shows the $M_\mathrm{i} - z$ space of the catalogue with our dataset highlighted. Our sample traces the higher density region in the catalogue around redshift $\sim 1.5 - 2.5$ which is likely a result of using absorption in C IV line as the main criterion in defining BAL quasars.

4.2.2 Unsupervised Clustering with the K-means Algorithm

Unlike supervised clustering, such as classification which uses labelled training sets to classify unlabelled data, unsupervised clustering can be utilized when the data points are not expected to belong to formal classes. Unsupervised techniques can, for example, be useful when searching for structures in multi-dimensional space where distributions may be continuous. K-means –the algorithm we use in this work– is an unsupervised clustering algorithm that groups data points that are alike according to a given measure of similarity. To assign classes, it uses an

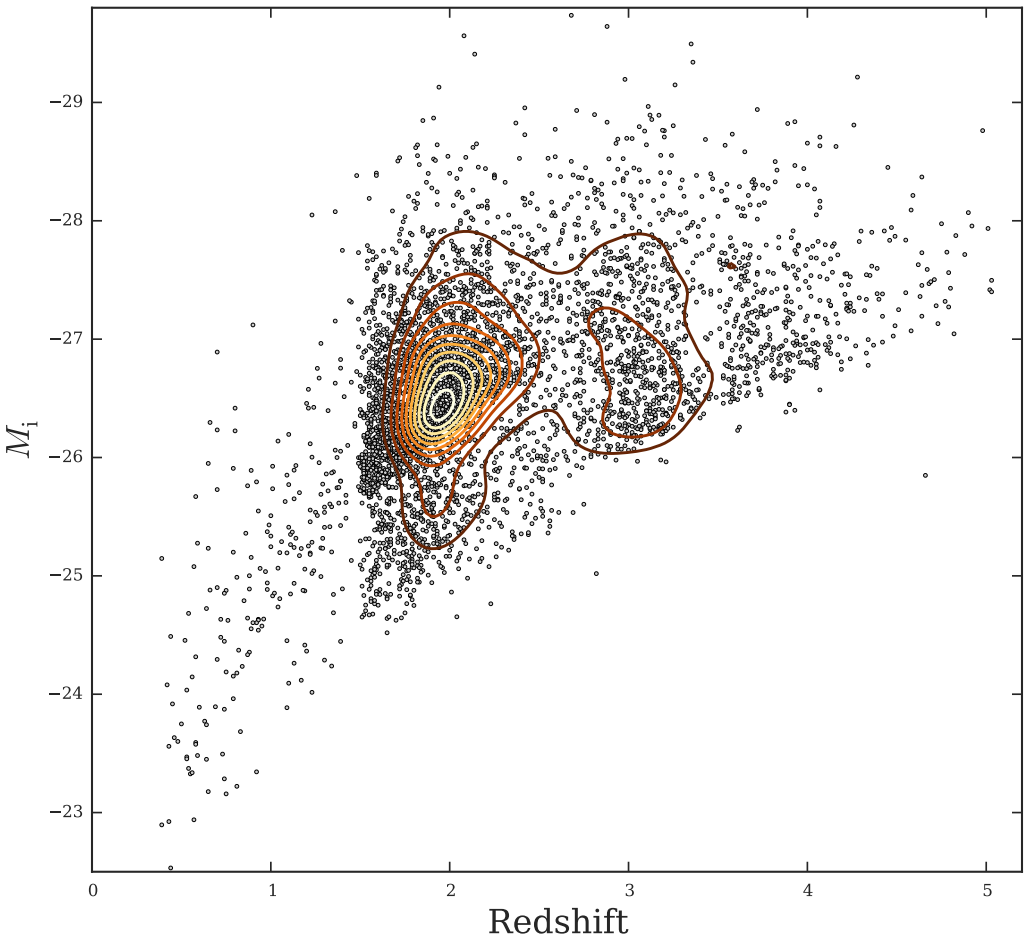

Figure 4.1: M_i vs. z for data in the Gibson et al. (2009) catalogue including 5,604 objects in grey dots. Our sample (2,683 objects) is shown with red contours and has a redshift limit of $1.79 < z < 3.7$.

iterative process to minimize the within-cluster sum of squares (Bishop, 2009; Hastie et al., 2009) given by:

$$J = \sum_{n=1}^{N} \sum_{k=1}^{K} r_{nk} ||x_n - \mu_k||^2, \tag{4.2}$$

where μ is the mean in cluster k, N is the total number of data points, and K is the number of clusters. For each point x_n, r_{nk} is an indicator that takes the values $r_{nk} = 1$ if x_n is a data point in cluster k, and $r_{nk} = 0$ otherwise.

Given a specific number of clusters K, the algorithm randomly assigns K seeds as initial values for the cluster centroids. Data points that are close to each seed are then assigned to a specific cluster. For each cluster a new centroid (mean) is now calculated from the data points belonging to this cluster and again data points closer to these new calculated centroids are assigned to their new clusters. This process of calculating new centroids (means) and (re)assigning members to clusters is repeated until it converges. Fig. 9.1 from Bishop (2009) gives a clear illustration of how K-means works for a toy dataset.

In this work, we use `scikit-learn's K-means` module (Pedregosa et al., 2011) to perform the clustering analysis. There is only one input parameter needed to run `K-means` which is K: the number of clusters for partitioning the sample. Determining an optimal value for K is by and large a heuristic exercise and depends on the dataset and the question at hand. In our case, we want to use a number of clusters that is high enough to reveal interesting differences among the objects but also low enough to prevent unnecessary fragmentation of the sample and to maintain reasonable signal-to-noise ratio when we make composite spectra. As we discuss in Tammour et al. (2015a), this is best done by experimenting on a range of values for K and examining the results. Based on previous experience, we choose to run `K-means` for $K = 3, 4, 5,$ and 6 and use only one of those cases to illustrate the results. We point out that there are other available metrics used to determine K such as the Silhouette score which describes the amount of overlap among the clusters, however, these metrics are generally found to be inconclusive (Hastie et al., 2009). We find that in our case a visual examination of the resulting clusters is sufficient particularly with only 3 dimensions used in the clustering analysis.

As a starting point, we standardize (normalize) the clustering parameters, i.e., for each data point we subtract the mean and divide by the standard deviation of that array (e.g., V_{min}), to

ensure the quantities in each dimension entering the clustering contribute at a comparable level. This is important because some of the parameters are many orders of magnitudes different than other ones used in the analysis. V_{min} for example is on the order of a few thousands km s^{-1} while the EW has values of a few 10's Å which gives the EW much less weight when calculating the cluster mean without prior normalization.

Finally, we use the spectra of objects from each of the clusters we find to create median composite spectra. The resulting composite spectrum is the median of the flux values of all the spectra in that cluster at each wavelength bin. We also apply a 3σ clipping which cuts data points (flux measurements at each wavelength) that are more than 3σ standard deviations from the median.

4.3 Results and Discussion

4.3.1 Exploring The V_{min}, V_{max} and EW Parameter Space

Starting with the 2,683 objects in the sample, we apply K-means using the V_{min}, V_{max}, and EW of the C IV BAL trough for $K = 3, 4, 5,$ and 6. Table 4.1 summarizes the results of these four runs and lists the number of objects in each cluster along with the centroid coordinates in the parameter space of V_{min}, V_{max} and EW. We also use ancillary data from Filiz Ak et al. (2014), Baskin et al. (2015), and Krawczyk et al. (2015) to examine the properties of the objects in our sample that are measured in their work such as variability, He II EW and intrinsic reddening.

In Fig. 4.2, we show the clusters for the $K = 6$ run in the C IV V_{min}, V_{max}, and EW parameter space and the composite spectra generated from objects in each of the identified clusters (we provide similar plots for $K = 3, 4,$ and 5 in Appendix C.1). We choose to focus on this case ($K = 6$) because it allows us to break down the sample into groups that reveal clear differences in their C IV toughs with little overlap among the groups in the parameter space of V_{min}, V_{max} and EW.

We see in Fig. 4.2 that the two extreme ends of the outflow velocity space (top-left panel) are occupied by clusters C IV-a and C IV-f. Both clusters, however, have very small EWs

Table 4.1: Clustering results for the 2,683 objects in the C IV sample. We show the number of objects in each cluster with the centroid coordinates in the V_{min}, V_{max} and EW space. The clusters are ordered according to the V_{min} of their centroid. C IV_{S0} refers to the fraction of objects with both C IV and Si IV absorption detected in the Gibson et al. (2009) catalogue . C IV_{SA} refers to the fraction of objects with absorption in Al III in addition to absorption from higher IP lines (C IV and Si IV) in the Gibson et al. (2009) catalogue. ΔBI refers to the fractions of the 637 objects from Filiz Ak et al. (2014) in our sample with detected variability in the C IV trough. In addition to data from Filiz Ak et al. (2014), we cross-matched our sample with data from Baskin et al. (2015, 984 objects), Krawczyk et al. (2015, 871 objects), and Shen et al. (2011, 2548 objects). In the last three columns, we list a breakdown of the number of all objects in each cluster that have measurements in these three studies. The number of the objects in each cluster with ancillary data within each cluster is not driving the calculated fractions.

K	Cluster	Num in cluster	V_{min} km s^{-1}	V_{max} km s^{-1}	EW Å	C IV_{S0} %	C IV_{SA} %	$\|\Delta BI\| > 0$ %	Num Filiz Ak[a]	Num Baskin[b]	Num Krawczyk[c]	Num Shen[d]
3	CIV-a	683	-14507.77	-19842.50	-07.85	6.1	0.3	23.3	159	208	247	668
	CIV-b	712	-4188.28	-18668.68	-33.61	48.0	15.3	39.3	280	237	252	653
	CIV-c	1288	-2996.59	-7443.97	-11.05	14.3	2.5	15.4	198	539	372	1227
4	CIV-a	522	-15985.11	-21068.89	-07.56	5.9	0.6	22.2	116	151	191	508
	CIV-b	666	-6752.10	-14259.33	-12.61	17.9	1.5	27.2	181	252	206	655
	CIV-c	553	-3907.34	-18962.38	-37.36	51.4	19.0	40.0	221	183	205	499
	CIV-d	942	-1856.26	-5787.15	-11.17	14.2	2.7	12.6	119	398	269	886
5	CIV-a	496	-16191.03	-21301.06	-07.62	6.0	0.4	22.2	110	143	181	483
	CIV-b	455	-7944.67	-12729.35	-08.31	8.4	0.7	22.2	101	169	150	449
	CIV-c	573	-4330.64	-16644.97	-24.64	40.8	7.5	37.3	214	227	200	542
	CIV-d	269	-3823.62	-20702.71	-45.88	55.0	27.9	40.9	110	67	92	239
	CIV-e	890	-1792.01	-5582.69	-10.68	13.3	2.2	11.5	102	378	248	835
6	CIV-a	470	-16451.64	-21209.60	-06.92	5.1	0.4	21.1	99	132	170	459
	CIV-b	437	-7916.63	-12406.34	-08.00	8.2	0.9	21.5	94	164	143	431
	CIV-c	363	-6377.94	-20578.79	-22.79	33.6	3.3	41.9	152	125	143	345
	CIV-d	246	-3472.74	-20435.06	-46.86	54.1	30.9	40.2	99	61	82	217
	CIV-e	388	-2547.71	-11666.33	-25.17	46.9	9.8	29.6	115	174	121	356
	CIV-f	779	-1798.00	-5158.69	-09.26	9.1	1.4	10.0	78	328	212	740

[a] Filiz Ak et al. (2014)

[b] Baskin et al. (2015)

[c] Krawczyk et al. (2015)

[c] Shen et al. (2011)

(top-right panel). Clusters C IV-c and C IV-d overlap in the low V_{min} - highV_{max} corner[2] but C IV-d includes objects with significantly larger EWs. The cluster overlap in the 2D V_{max}-V_{min} space is thus broken when the EW dimension is taken into account.

In the lower panel of Fig. 4.2, we show the median composite spectra generated from objects in each of the clusters. The composites clearly reflect the properties of the C IV troughs seen in the upper panels in addition to other spectral features that are potentially related to the C IV trough behaviour. For instance, composite C IV-f has a prominent He II emission-line which is an indication of a harder ionizing SED. A harder SED can potentially overionize the wind resulting in such a weak trough with a low launch velocity (V_{min}). Composite C IV-d on the other hand appears highly reddened, with a remarkably broad and deep C IV (and Si IV) trough which goes along with a high number of LoBALs in this cluster.

4.3.2 Presence of Lower IP transitions

The presence or absence of lower IP lines in BAL quasar spectra can be mainly seen as an indicator of the line-of-sight passing through gas with different ionization levels (e.g, Proga et al., 2000). Examining the composite spectra in the lower panel in Fig. 4.2, we see that absorption from Si IV generally follows the behaviour of the C IV trough but in some cases the composite spectra show no sign of absorption in Si IV nor Al III (e.g., composite C IV-a). We calculate the fractions of objects with lower IP transitions reported in the Gibson et al. (2009) catalogue and list them in Table 4.1. We use the notation from Filiz Ak et al. (2014) and refer to objects with absorption in Si IV (but not from Al III) as C IV_{S0} and objects with absorption from both Si IV and Al III with C IV_{SA}, however, the data are all from Gibson et al. (2009) catalogue and are therefore complete for our sample.

Table 4.1 shows that clusters C IV-d and C IV-e (and to a lesser extent C IV-c) have a significantly higher fraction of objects with absorption from lower IP transitions. Cluster C IV-d, for example, has a 54% fraction with Si IV troughs (C IV_{S0}) and a 31% fraction with both Si IV and Al III (C IV_{SA}). The presence of absorption from lower IP species in objects with stronger C IV troughs has been reported in previous work (e.g., Weymann et al., 1991;

[2]The triangle shape of the velocity space in Fig. 4.2 is due to the velocity upper and lower limits in the definition of BI_0 eqn. 4.1.

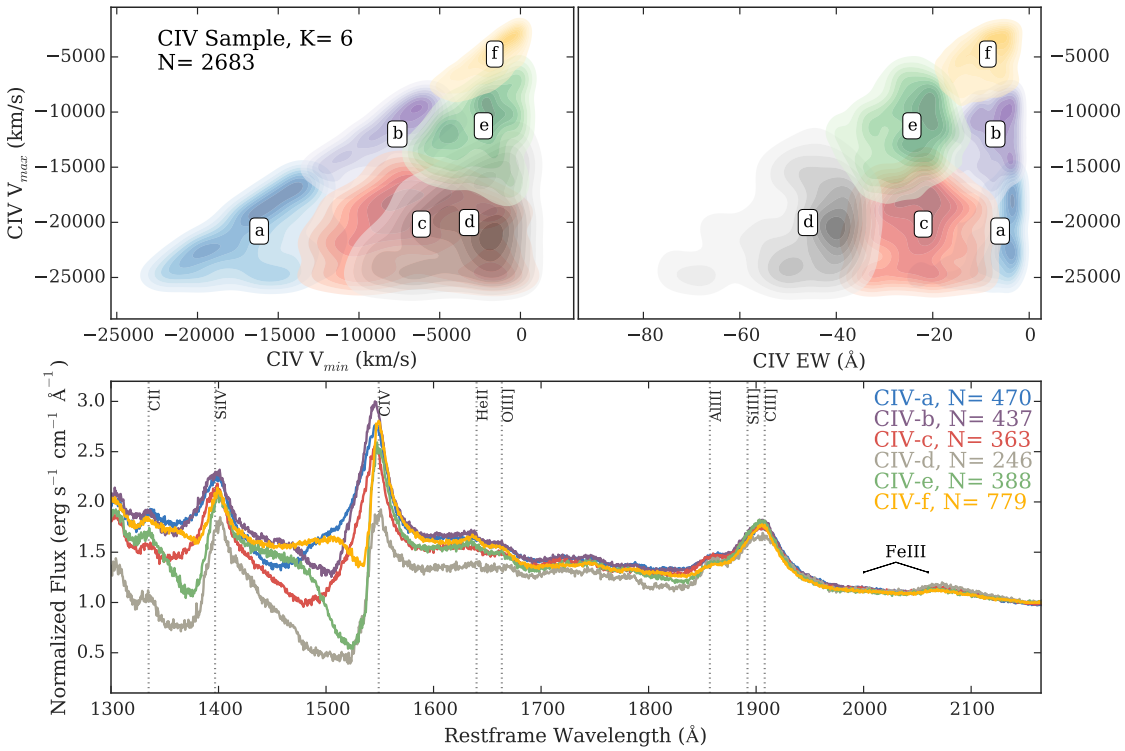

Figure 4.2: Projections in the clustering parameter space of C IV V_{max}vs. V_{min} (top-left) and V_{max} vs. EW (top-right) using $K = 6$. The white labels follow the cluster designation given in Table 4.1. Cluster C IV-c and C IV-d overlap in the V_{min}-V_{max} plane (top-left panel) but C IV-d includes objects with much larger EWs (top-right panel). The lower panel shows the median composite spectra generated from objects in each of the clusters. Composite C IV-a has the highest outflow velocities and shows no sign of absorption from lower IP transitions (Si IV or Al III). Composite C IV-d has a broad and deep C IV and Si IV troughs, displays Al III absorption, and is strongly reddened. He II appears more prominent in C IV-f.

Reichard et al., 2003a; Filiz Ak et al., 2014) and is often attributed to a line-of-sight passing through higher inclination angles with respect to the normal of the accretion disk and passing through a wind component that is slower and more dense (e.g., Proga et al., 2000). Indeed, Table 4.1 shows that on average, clusters C IV-d, C IV-e and C IV-c have the largest EWs among the clusters (−47 Å, −25 Å and −22.8 respectively; Fig. 4.2) indicating higher column densities along the line-of-sight in these objects. They also have lower launch velocities (V_{min}) consistent with a slower component of the wind.

Using extinction, $E(B-V)$, and the intrinsic continuum spectral index, α_ν[3], estimates from Krawczyk et al. (2015), we examine the distribution of these two properties in our clusters (Figures4.3, and 4.4). We see that a common feature of clusters C IV-d and C IV-e is that they are heavily reddened (C IV-e appears to be also more intrinsically red) as we show in Fig. 4.3 (panels C and D).

On the other hand, clusters with a lower fraction of C IV_{S0} and C IV_{SA} objects (clusters C IV-a, C IV-b, and C IV-f) all have significantly shallower C IV troughs with V_{min} velocities spanning the entire range in the sample (Fig. 4.2). Those clusters also show low dust extinctions (Fig. 4.3, panel C).

4.3.3 Variability

Variability in BAL quasar spectra is commonly observed on moderate and short timescales (e.g., Gibson et al., 2009; Filiz Ak et al., 2014). As we discuss in §4.1, such behaviour is likely due to changes in ionization level and/or due to transverse motion of gas through the line-of-sight. Here, we examine the fraction of objects with variability in their C IV troughs as reported in Filiz Ak et al. (2014, their Table 1) which includes variability on timescales of $\sim 1-4.5$ years in the quasars' restframe. We list the fractions of objects with |Δ BI| greater than zero calculated for each cluster in Table 4.1.

We see in Table 4.1 that clusters C IV-c and C IV-d have ∼40% of their objects with |Δ BI| > 0 (indicating positive or negative changes in their balnicity index values) while the

[3]We calculate the spectral index from their equation 4: $\alpha_\nu = -0.28 + \Delta(\alpha_\lambda)$, where $\Delta(\alpha_\lambda)$ is the difference between the intrinsic spectra index and the observed modal spectral index, and α_ν is given by the power law: $f_\nu = \nu^{\alpha_\nu}$.

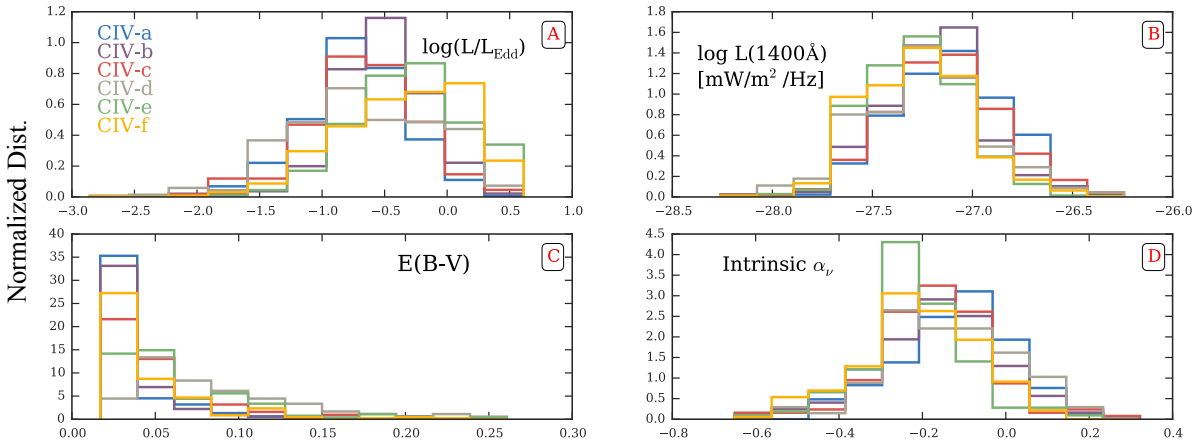

Figure 4.3: Distributions of the properties of objects in each of the clusters from the $K = 6$ run. The labels in the top-left panel match those in Table 4.1 and Fig. 4.2. **A:** the $L/L_{\rm Edd}$ for 2548 objects from the Shen et al. (2011) quasar catalogue (calculated using their fiducial BH masses). Cluster C IV-e is slightly skewed towards higher $L/L_{\rm Edd}$ while C IV-a is skewed to the opposite side. Both clusters have weak C IV troughs but their outflow velocities are significantly different with C IV-a having the highest and C IV-e the lowest. **B:** Continuum luminosity at 1400 Å for 2683 (entire sample) from the Gibson et al. (2009) catalogue . **C and D:** Dust extinction and intrinsic spectral index (respectively) for 871 objects from Krawczyk et al. (2015); clusters with deeper C IV troughs (C IV-d, C IV-e, and to some extent C IV-c) are more heavily reddened (panel C) but only C IV-e is intrinsically red on average (panel D). Panels E, F, G, and H show distributions for 984 objects from Baskin et al. (2015) in our sample.

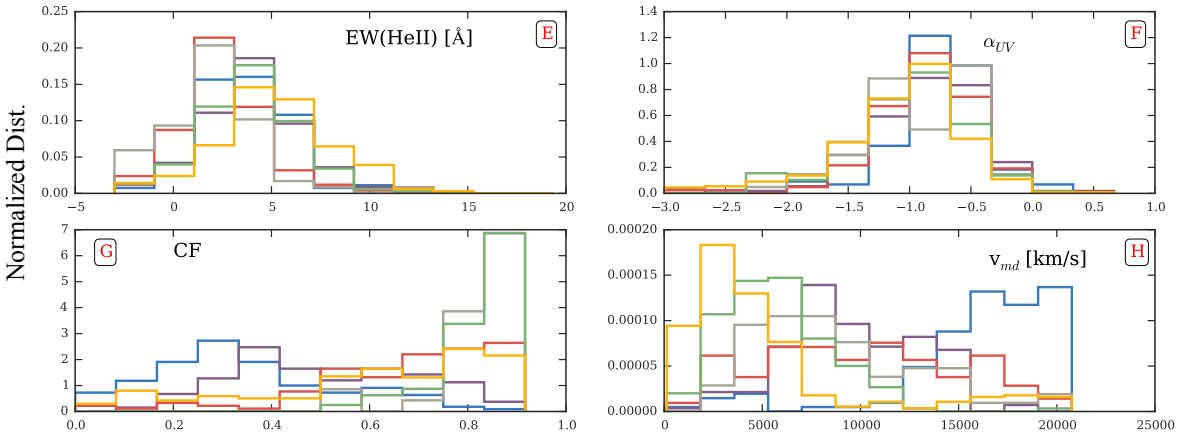

Figure 4.4: Distributions of the properties of objects in each of the clusters from the $K = 6$ run. The labels in the top-left panel match those in Table 4.1 and Fig. 4.2. **E:** He II EW; cluster C IV-f clearly has stronger He II than the other clusters which indicates a harder ionizing SED. **F:** α_{UV} (measured). Note that these values do not include correction for the intrinsic extinction, and so the meaning of these spectral indices is somewhat ambiguous. **G:** the maximum absorption depth (interpreted as absorption covering factor). **H:** v_{md} (the velocity of the maximum absorption depth).

fraction in C IV-e is 30%. Cluster C IV-f has the lowest fraction with only 10% variable objects. Examining the composite spectra of these clusters in Fig. 4.2, we see that clusters with the broad and relatively deeper C IV troughs have a larger fraction of variable objects than clusters with shallow and narrower troughs. Narrower troughs with lower EWs can form if the absorbing gas in those objects is located farther out which means the variability could occur at longer timescales and is therefore unaccounted for in our dataset.

We see in Fig. 4.5 (lower panel) that the fraction of variable objects decreases with the strength of He II in a sense that clusters with harder ionizing SEDs (higher He II EW) are on average less likely to be variable. Panel E in Fig. 4.4 also shows that the distributions of the He II EW in clusters C IV-c and C IV-d as measured by Baskin et al. (2015) are skewed towards lower values while cluster C IV-f, on the other hand, has significantly larger He II EW. Investigating the effects of the hardness of the SED on the distance to the absorbing gas and the variability timescale is beyond the scope of this work and could be addressed in future work. The low extinction for this cluster and the low outflow velocities indicate that we might be viewing the wind in this case through a low inclination line-of-sight that likely does not pass through the shielding gas. We also point out that the Eddington ratio ($L/L_{\rm Edd}$, as reported by Shen et al. (2011)) for objects in cluster C IV-f is significantly larger compared to the other clusters (panel A in Fig. 4.3). However, we note that these $L/L_{\rm Edd}$ values can potentially be problematic because, for most of the objects in our sample, the C IV *emission* lines have been used to calculate this ratio and it is unclear how the absorbed portion of the C IV affects the measurments of the emission line width used in calculating the Eddington ratio.

4.4 Summary and Conclusion

In this work, we use a large sample of BAL quasars to characterize the diversity in their broad C IV troughs. We use measurements of the C IV trough from the Gibson et al. (2009) catalogue to look for groups of similar objects in the parameter space of V_{min}, V_{max} and EW of C IV using an unsupervised clustering method (K-means). We examine the properties of objects in each of the groups we identify and make connections with the underlying physics driving the nature of the absorption troughs in these objects.

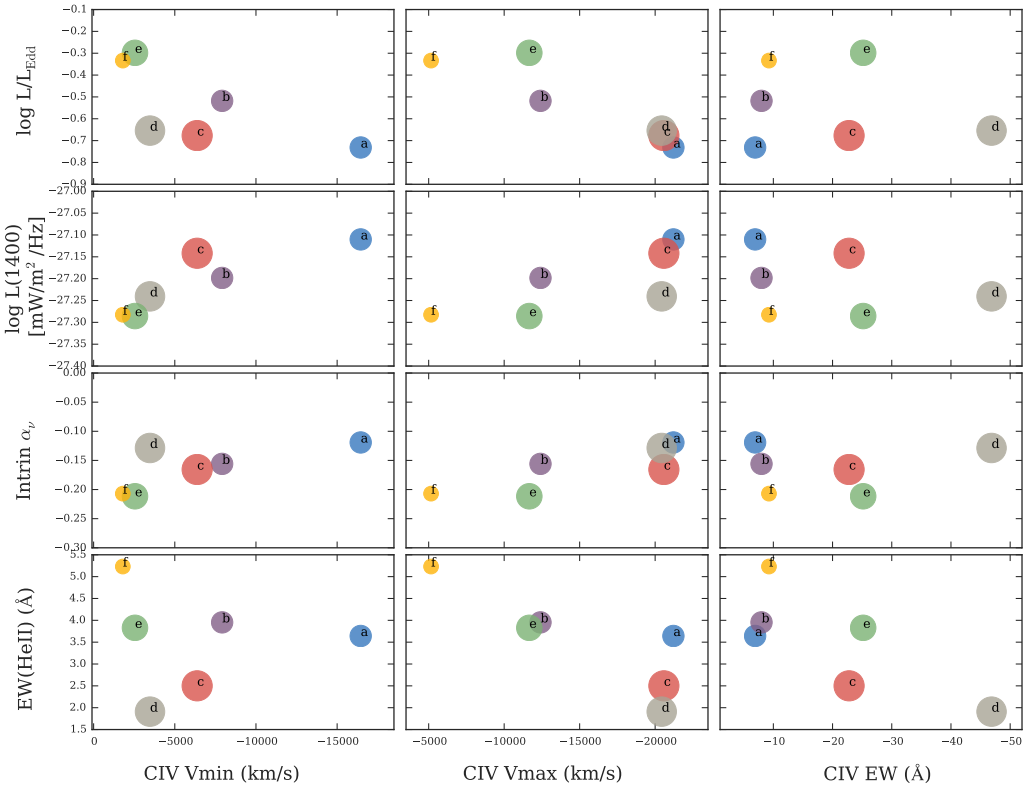

Figure 4.5: Average properties of each cluster. The colors and labels are similar to those in Fig. 4.2 and the size of the bubble indicates the fraction of variable objects in this cluster (Table 4.1) as reported in Filiz Ak et al. (2014). Top row: mean Eddington accretion rate $L/L_{\rm Edd}$ from Shen et al. (2011) for each cluster plotted against C IV trough V_{min}, V_{max} and EW (2,548 objects). Second row from top: the continuum luminosity at 1400 Å from Gibson et al. (2009). Second row from bottom: the intrinsic α_ν for 871 objects from Krawczyk et al. (2015). Bottom row: He II EW for 984 objects from Baskin et al. (2015). Variability is clearly higher in objects with lower He II EW (softer SEDs). Cluster C IV-f stands out with very high He II EW. Only 10% of the objects in this cluster are found to be variable -the smallest fraction in our sample. The C IV troughs in this group are significantly weaker (smaller EWs) and have extremely low outflow velocities. Cluster C IV-a has C IV V_{min} significantly higher than the other clusters. It also has lower C IV EW, and it is slightly bluer than the other clusters.

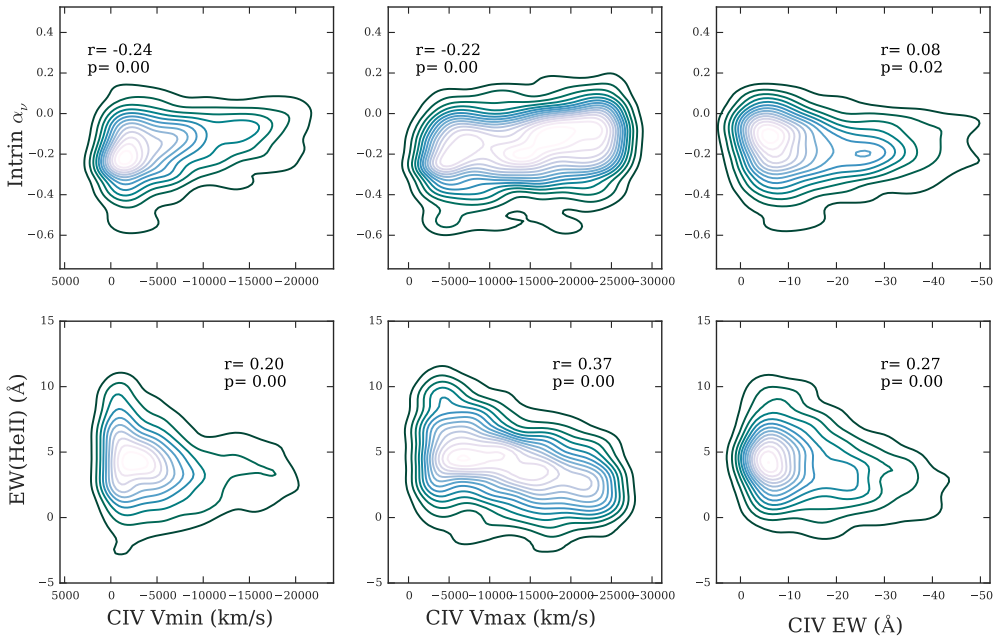

Figure 4.6: For the sample we examine correlations between the clustering parameters (V_{min}, V_{max}, and EW of C IV trough) and: (top row) the intrinsic α_ν, 871 objects from Krawczyk et al. (2015), and (bottom row) He II EW, 984 objects from Baskin et al. (2015). We find weak correlations between the intrinsic spectral index and outflow velocities; bluer objects have slightly higher velocities (see also cluster C IV-a in Fig. 4.5). The He II EW is also correlated with the trough properties; objects with larger He II EW (harder SEDs) are more likely to have smaller terminal velocities and weaker absorption troughs (see also cluster C IV-f in Fig. 4.5).

We summarize our findings as follows:

- We find that objects with strong C IV troughs (high EW) and low V_{min} tend to have higher fractions of absorption from lower IP transitions; for example 54% in C IV-d and 47% in C IV-e have also absorption in Si IV (Table 4.1 and Fig. 4.2). On the other hand, objects with weaker C IV troughs are less likely to display lower IP absorption in their spectra regardless of their outflow velocities (clusters C IV-a, C IV-b and C IV-f in Table 4.1 and Fig. 4.2).

- We find a higher fraction of variable objects ($\sim 40\%$) in clusters with strong C IV troughs (broader and deeper, clusters C IV-c and C IV-d in Table 4.1 and Fig. 4.2). These objects also have weaker ionizing SEDs as can be seen from their low He II EWs (Fig. 4.5). While clusters with the lowest fraction of detected variability (10% in cluster C IV-f, Table 4.1) tend to have shallow and narrow C IV troughs and have harder SEDs (significantly stronger He II, Fig. 4.5).

- We identify two groups with similarly very weak C IV troughs but on the extreme ends of the velocity space: cluster C IV-a has exceptionally high minimum and maximum velocities (average V_{min} is double the next closest group) while C IV-f has exceptionally low outflow velocities. In their composite spectra, clusters C IV-a and C IV-f both have very shallow and narrow C IV troughs (Fig. 4.2). We find that these two groups are significantly different in their intrinsic reddening with C IV-a containing objects that have weaker SEDs, are intrinsically blue(r), and have very high outflow velocities, while C IV-f contains objects that have harder SEDs, are intrinsically red(der), and have lower velocities. C IV-f potentially includes objects with winds that are launched from larger radii because on small scales the gas is overionized by the hard radiation from the central source while C IV-a objects may contain winds similar to those seen in C IV-e but viewed from smaller inclination angles.

- In the full sample, we find weak but statistically significant correlations between the hardness of the SED (probed by the He II EW) and the C IV trough properties (bottom panel in Fig. 4.6); objects with harder SEDs tend to have lower outflow velocities and

weaker troughs (smaller EWs). Weaker correlations are also present between the intrinsic reddening and the outflow velocities (bottom panel in Fig. 4.6): intrinsically blue objects tend to have higher minimum and maximum outflow velocities. This correlation might be driven by the two extreme cases (C IV-a and C IV-f) we discuss in the previous point.

One important caveat to keep in mind, is that the properties of objects in each cluster shown in Figures 4.3, 4.4, and 4.5 do not describe the same set of objects, i.e., only a small subset of objects in our sample has measurements for their reddening, He II EW and variability. This ultimately means that when we describe a cluster as intrinsically red, this applies for the subset of objects in this cluster that have α_ν measurements from Krawczyk et al. (2015). However, we do not anticipate that this should affect our discussion of the characteristics of the clusters as the numbers within each measured property (e.g., α_ν) are evenly distributed among the clusters as we show in Table 4.1. Therefore we do not expect that selection effects are driving the conclusions that we draw from the ancillary data sets.

Chapter 5

Conclusions And Future Directions

Quasar spectra carry the signatures of the physical processes at work in the quasar system. Such signatures are often detected as correlations among properties of the emission and absorption lines and/or the continuum. Large homogeneous spectral samples such as those from the SDSS allow us to trace some of the systematic trends in optical and UV spectral features that can be linked to the underlying physics of the accretion system and the surrounding gas.

It has been well established that the continuum luminosity and the BH mass of a quasar (often combined as the Eddington ratio ($L/L_{\rm Edd}$) are of fundamental importance to the observed appearance of quasars in the rest-frame optical (e.g., Shen & Ho, 2014). Some of the most revealing observations are the Eigenvector 1 correlations which include an anti-correlation in the optical regime between the broad emission features of Fe II emitted in close proximity of the central source and the narrow [O III] doublet emitted at much larger physical size scales (~kpc's). The presence of a link between gas emitted from two regions separated by such large distances is indeed puzzling. To address potential selection effects and mitigate the confounding influence of the host galaxy, we design a study to look at the NLR independent of the properties of the central engine. To investigate this further, in chapter 2, we use a sample of optical quasar spectra from the Shen et al. (2011) catalog of the 7^{th} Data Release of the SDSS to track the behaviour of the narrow forbidden lines constraining the continuum luminosity and the width of the broad Hβ line. We focus on four lines with a wide range of ionization potentials (13.6 to 97 eV) and find that the luminosity of the higher ionization lines is consistent up to redshift 0.75 indicating no evolution in the quasar spectral energy distribution (SED) or

the host galaxy ISM illuminated by the continuum up to this redshift. In contrast, we find that the [O II] line becomes stronger at higher redshifts, and we interpret this as a consequence of enhanced star formation contributing to the [O II] emission in host galaxies at higher redshifts. We also confirm the eigenvector 1 correlations between the width of Hβ, and the strength of [O III] and Fe II in our sample.

In the remainder of this thesis, we work in the rest-frame UV regime focusing on questions related to the role of the ionizing SED in the formation and observed properties of quasar winds. The wind model –invoked as the ultimate source of the broad emission and absorption features in quasar spectra– has been successful in addressing some of the pressing problems regarding the BLR. It can, for instance, account for the blueshifts and asymmetries commonly seen in UV features specifically in C IV. In chapter 3, we use the properties of three UV emission lines to probe the conditions in the emitting gas and deduce their correlations with the winds. We specifically use the properties of three strong UV lines (C IV, C III], and Mg II) to bin our large sample, which we derive from the Pâris et al. (2014) quasar catalog, into similar groups and examine the median composites of each group and what they convey about the shape of the SED and the characteristics of the winds. We find that our results are in agreement with the wind model, in particular, objects with softer ionizing SEDs (weaker He II broad emission lines and larger Si III]/C III] ratio) have blueshifted and weaker C IV emission lines, while objects with harder ionizing SEDs (stronger broad He II emission and smaller Si III]/C III] ratio) have stronger symmetric C IV that peaks at the systemic redshift. These trends were particularly evident when we used the measurements of C IV or C III] to partition our sample while they became less clear when the Mg II line was used, i.e., the C IV profiles did not show significant changes with the width of Mg II. This has implications for the direct estimates of BH masses routinely done using the widths of Mg II and C IV which generally shows discrepancy between the values determined from these lines. We apply the same analysis to the measurements of the C III] blend in a sample of BAL quasars and find evidence for changes in the BAL properties with the hardness of the SED. We expand on this last result in a separate project (chapter 4) using a large, well-defined sample from the Gibson et al. (2009) BAL quasar catalogue.

In chapter 4 we investigate quasar winds more directly via the broad absorption features in quasar UV spectra, particularly that of the C IV trough. These features appear in $\sim$ 20% of

optically selected samples and they are thought to result from viewing the absorbing medium through a line-of-sight passing through the wind. Our goal in this work is to constrain the diversity commonly seen in the broad absorption trough of BALQs and to link their different profiles to the underlying physical conditions producing them. The wind, for instance, is known to be sensitive to the shape of the ionizing SED in a sense that UV photons are essential to accelerate the wind but too much X-ray radiation will suppress the wind by over-ionizing the gas. We explore the properties of the C IV troughs using their EWs and initial and terminal outflow velocities and partition our sample into groups that are similar in this parameter space. We identify groups of objects with C IV troughs that are deep and broad (low launch and high terminal velocities) that also tend to have higher fractions of lower IP troughs (e.g., Si IV and Al III). On the other hand, objects with shallow and generally narrow(er) troughs (small EWs and either both high or both low initial and terminal outflow velocities) have significantly lower fractions of lower IP troughs. We interpret these findings as resulting from views through a line-of-sight passing through high column density gas close to the accretion disk in the former case and passing through a lower column density and lower inclination angle in the latter. In addition, we utilize extra measurements from the literature such as the He II EW, intrinsic reddening and absorption line variability estimates. These data in conjunction with our analysis show that some of our groups are more likely to be variable (on time scales of 1 to 4.5 years) than others. The variability characteristics appear to correlate with He II strength seen as objects with harder SEDs (stronger He II) have a significantly lower fraction of variable objects while more variable objects have on average softer SEDs. More work is needed to explore the connection between an increased fraction of variable quasars and the shape of the SED.

Moving forward, we would like to extend our analysis of the C IV troughs to lower IP lines and investigate how the parameter space of these troughs probes the properties of BAL quasars. Another important addition to this work includes X-ray data for selected objects from the emission-line groups in chapter 3 and the BAL quasars groups we identify in chapter 4. Such data will allow us to constrain the ionizing radiation properties and gauge their influence on the observed behaviour of the C III] blend and C IV line shapes and the C IV troughs. Detailed measurements of the C III] blend and the He II line can also contribute to our understanding of the effects of the shape of the SED on the properties of both emission and absorp-

tion features in quasars. Other techniques such as manifold learning algorithms can be used to explore quasar properties in multi-dimensional parameter space. Such methods are powerful ways to work in high dimensional space and have the potential to uncover interesting correlations among quasar properties and ultimately the underlying physical processes producing such correlations.

Bibliography

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

Bachev R., Marziani P., Sulentic J. W., Zamanov R., Calvani M., Dultzin-Hacyan D., 2004, ApJ, 617, 171

Bahcall J. N., Kirhakos S., Saxe D. H., Schneider D. P., 1997, ApJ, 479, 642

Baldwin J. A., 1977, ApJ, 214, 679

Baldwin J., Ferland G., Korista K., Verner D., 1995, ApJL, 455, L119

Barlow T. A., 1994, PASP, 106, 548

Baskin A., Laor A., 2005, MNRAS, 356, 1029

Baskin A., Laor A., Hamann F., 2015, MNRAS, 449, 1593

Begelman M. C., 2004, Coevolution of Black Holes and Galaxies, p. 374

Bennert N., Falcke H., Schulz H., Wilson A. S., Wills B. J., 2002, ApJL, 574, L105

Bennert V. N., Auger M. W., Treu T., Woo J.-H., Malkan M. A., 2011, ApJ, 726, 59

Bentz M. C., Hall P. B., Osmer P. S., 2004, AJ, 128, 561

Bishop C., 2009, Pattern Recognition and Machine Learning. Springer, New York, NY, USA

Blaes O., 2007, in Ho L. C., Wang J.-W., eds, Astronomical Society of the Pacific Conference Series Vol. 373, The Central Engine of Active Galactic Nuclei. p. 75 (arXiv:astro-ph/0703589)

Boroson T. A., 2002, ApJ, 565, 78

Boroson T. A., Green R. F., 1992, ApJS, 80, 109

Canalizo G., Wold M., Hiner K. D., Lazarova M., Lacy M., Aylor K., 2012, ApJ, 760, 38

Capellupo D. M., Hamann F., Shields J. C., Rodríguez Hidalgo P., Barlow T. A., 2012, MNRAS, 422, 3249

Cardelli J. A., Clayton G. C., Mathis J. S., 1989, ApJ, 345, 245

Casebeer D. A., Leighly K. M., Baron E., 2006, ApJ, 637, 157

Collin-Souffrin S., Dyson J. E., McDowell J. C., Perry J. J., 1988, MNRAS, 232, 539

Croom S. M., et al., 2002, MNRAS, 337, 275

Draine B. T., 2011, Physics of the Interstellar and Intergalactic Medium. Princeton University Press, New Jersey, USA

Dunlop J. S., 2004, Coevolution of Black Holes and Galaxies, from the Carnegie Observatories Centennial Symposia, p. 341

Dunlop J. S., McLure R. J., Kukula M. J., Baum S. A., O'Dea C. P., Hughes D. H., 2003, MNRAS, 340, 1095

Elvis M., et al., 1994, ApJS, 95, 1

Ferrarese L., Ford H., 2005, SSRv, 116, 523

Ferrarese L., Merritt D., 2000, ApJL, 539, L9

Ferrarese L., Pogge R. W., Peterson B. M., Merritt D., Wandel A., Joseph C. L., 2001, ApJL, 555, L79

Filiz Ak N., et al., 2012, ApJ, 757, 114

Filiz Ak N., et al., 2013, ApJ, 777, 168

Filiz Ak N., et al., 2014, ApJ, 791, 88

Floyd D. J. E., Kukula M. J., Dunlop J. S., McLure R. J., Miller L., Percival W. J., Baum S. A., O'Dea C. P., 2004, MNRAS, 355, 196

Floyd D. J. E., Dunlop J. S., Kukula M. J., Brown M. J. I., McLure R. J., Baum S. A., O'Dea C. P., 2013, MNRAS, 429, 2

Gallagher J. S., Hunter D. A., Bushouse H., 1989, AJ, 97, 700

Gallagher S. C., Brandt W. N., Chartas G., Priddey R., Garmire G. P., Sambruna R. M., 2006, ApJ, 644, 709

Ganguly R., Bond N. A., Charlton J. C., Eracleous M., Brandt W. N., Churchill C. W., 2001, ApJ, 549, 133

Ganguly R., Brotherton M. S., Cales S., Scoggins B., Shang Z., Vestergaard M., 2007, ApJ, 665, 990

Gaskell C. M., 1982, ApJ, 263, 79

Gibson R. R., Brandt W. N., Schneider D. P., Gallagher S. C., 2008, ApJ, 675, 985

Gibson R. R., et al., 2009, ApJ, 692, 758

Greene J. E., Zakamska N. L., Ho L. C., Barth A. J., 2011, ApJ, 732, 9

Hall P. B., Anosov K., White R. L., Brandt W. N., Gregg M. D., Gibson R. R., Becker R. H., Schneider D. P., 2011, MNRAS, 411, 2653

Hall P. B., et al., 2013, MNRAS, 434, 222

Hamann F., 1998, ApJ, 500, 798

Han J., Kamber M., Pei J., 2012, Data Mining: Concepts and Techniques. Elsevier, MA, USA

Hastie T., Tibshirani R., Friedman J., 2009, The Elements of Statistical Learning. Springer, New York, NY, USA

Hewett P. C., Foltz C. B., 2003, AJ, 125, 1784

Hewett P. C., Wild V., 2010, MNRAS, 405, 2302

Hill A. R., Gallagher S. C., Deo R. P., Peeters E., Richards G. T., 2014, MNRAS, 438, 2317

Ho L. C., 2005, ApJ, 629, 680

Hopkins A. M., et al., 2003, ApJ, 599, 971

Ichikawa T., Kajisawa M., Akhlaghi M., 2012, MNRAS, 422, 1014

Ivezić Z., Connolly A. J., VanderPlas J., Gray A., 2014, Statistics, Data Mining, and Machine Learning in Astronomy: A Practical Python Guide for the Analysis of Survey Data. Princeton University Press, New Jersey, USA

Jahnke K., Kuhlbrodt B., Wisotzki L., 2004, MNRAS, 352, 399

Jester S., et al., 2005, AJ, 130, 873

Jiang L., Fan X., Ivezić Ž., Richards G. T., Schneider D. P., Strauss M. A., Kelly B. C., 2007, ApJ, 656, 680

Jiang L., Fan X., Vestergaard M., 2008, ApJ, 679, 962

Kaspi S., Maoz D., Netzer H., Peterson B. M., Vestergaard M., Jannuzi B. T., 2005, ApJ, 629, 61

Kauffmann G., et al., 2003, MNRAS, 346, 1055

Kennicutt Jr. R. C., 1998, ARA&A, 36, 189

Kewley L. J., Geller M. J., Jansen R. A., 2004, AJ, 127, 2002

Kim M., Ho L. C., Im M., 2006, ApJ, 642, 702

Kormendy J., 2004, Coevolution of Black Holes and Galaxies, p. 1

Kotilainen J., Falomo R., Bettoni D., Karhunen K., Uslenghi M., 2013, preprint, (arXiv:1302.1366)

Kraemer S. B., Crenshaw D. M., 2000, ApJ, 532, 256

Krawczyk C. M., Richards G. T., Mehta S. S., Vogeley M. S., Gallagher S. C., Leighly K. M., Ross N. P., Schneider D. P., 2013, ApJS, 206, 4

Krawczyk C. M., Richards G. T., Gallagher S. C., Leighly K. M., Hewett P. C., Ross N. P., Hall P. B., 2015, AJ, 149, 203

Kriss G., 1994, Astronomical Data Analysis Software and Systems, 3, 437

Krolik J. H., 1999, Active Galactic Nuclei. Princeton Series in Astrophysics

Kruczek N. E., et al., 2011, AJ, 142, 130

Kukula M. J., Dunlop J. S., McLure R. J., Miller L., Percival W. J., Baum S. A., O'Dea C. P., 2001, MNRAS, 326, 1533

Leighly K. M., 2004, ApJ, 611, 125

Leighly K. M., Moore J. R., 2004, ApJ, 611, 107

Leighly K. M., Halpern J. P., Jenkins E. B., Casebeer D., 2007, ApJS, 173, 1

Leighly K. M., Hamann F., Casebeer D. A., Grupe D., 2009, ApJ, 701, 176

Longair M. S., 1994, High energy astrophysics. Vol.2: Stars, the galaxy and the interstellar medium. Cambridge: Cambridge University Press, Cambridge, UK

Luo B., et al., 2015, ApJ, 805, 122

Madau P., Pozzetti L., Dickinson M., 1998, ApJ, 498, 106

Magorrian J., et al., 1998, AJ, 115, 2285

Marconi A., Risaliti G., Gilli R., Hunt L. K., Maiolino R., Salvati M., 2004, MNRAS, 351, 169

McLure R. J., Kukula M. J., Dunlop J. S., Baum S. A., O'Dea C. P., Hughes D. H., 1999, MNRAS, 308, 377

Murray N., Chiang J., Grossman S. A., Voit G. M., 1995, ApJ, 451, 498

Nenkova M., Ivezić Ž., Elitzur M., 2002, ApJL, 570, L9

Netzer H., 2006, in Alloin D., ed., Lecture Notes in Physics, Berlin Springer Verlag Vol. 693, Physics of Active Galactic Nuclei at all Scales. p. 1, doi:10.1007/3-540-34621-X˙1

Netzer H., Laor A., Gondhalekar P. M., 1992, MNRAS, 254, 15

Netzer H., Shemmer O., Maiolino R., Oliva E., Croom S., Corbett E., di Fabrizio L., 2004, ApJ, 614, 558

Nolan L. A., Dunlop J. S., Kukula M. J., Hughes D. H., Boroson T., Jimenez R., 2001, MNRAS, 323, 308

Osterbrock D. E., 1989, Astrophysics of gaseous nebulae and active galactic nuclei. University Science Books, Mill Valley, CA, USA

Osterbrock D. E., Ferland G. J., 2006, Astrophysics of gaseous nebulae and active galactic nuclei. University Science Books, CA, USA

Osterbrock D. E., Pogge R. W., 1985, ApJ, 297, 166

Pagani C., Falomo R., Treves A., 2003, ApJ, 596, 830

Pâris I., et al., 2014, A&A, 563, A54

Pedregosa F., et al., 2011, Journal of Machine Learning Research, 12, 2825

Peterson B. M., 1993, PASP, 105, 247

Peterson B. M., 1997, An Introduction to Active Galactic Nuclei. Cambridge University Press

Proga D., Stone J. M., Kallman T. R., 2000, ApJ, 543, 686

Reichard T. A., et al., 2003a, AJ, 125, 1711

Reichard T. A., et al., 2003b, AJ, 126, 2594

Richards G. T., Vanden Berk D. E., Reichard T. A., Hall P. B., Schneider D. P., SubbaRao M., Thakar A. R., York D. G., 2002, AJ, 124, 1

Richards G. T., et al., 2006a, ApJS, 166, 470

Richards G. T., et al., 2006b, ApJS, 166, 470

Richards G. T., et al., 2011, AJ, 141, 167

Rosa-González D., Terlevich E., Terlevich R., 2002, MNRAS, 332, 283

Schmidt M., 1963, Nature, 197, 1040

Schneider D. P., et al., 2010, AJ, 139, 2360

Shakura N. I., Sunyaev R. A., 1973, A&A, 24, 337

Shemmer O., Lieber S., 2015, ApJ, 805, 124

Shen Y., Ho L. C., 2014, Nature, 513, 210

Shen Y., Liu X., 2012, ApJ, 753, 125

Shen Y., et al., 2011, ApJS, 194, 45

Shields J. C., 2007, in Ho L. C., Wang J.-W., eds, Astronomical Society of the Pacific Conference Series Vol. 373, The Central Engine of Active Galactic Nuclei. p. 355 (arXiv:astro-ph/0612613)

Silk J., Rees M. J., 1998, A&A, 331, L1

Silva L., Granato G. L., Bressan A., Danese L., 1998, ApJ, 509, 103

Silverman J. D., et al., 2009, ApJ, 696, 396

Sołtan A., 1982, MNRAS, 200, 115

Sparke L. S., Gallagher J. S., 2007, Galaxies in the Universe. Cambridge University Press

Spergel D. N., et al., 2003, ApJS, 148, 175

Sprayberry D., Foltz C. B., 1992, ApJ, 390, 39

Steffen A. T., Strateva I., Brandt W. N., Alexander D. M., Koekemoer A. M., Lehmer B. D., Schneider D. P., Vignali C., 2006, AJ, 131, 2826

Sulentic J. W., Zwitter T., Marziani P., Dultzin-Hacyan D., 2000, ApJL, 536, L5

Sulentic J. W., Marziani P., Zamanov R., Bachev R., Calvani M., Dultzin-Hacyan D., 2002, ApJL, 566, L71

Sulentic J. W., Bachev R., Marziani P., Negrete C. A., Dultzin D., 2007, ApJ, 666, 757

Tammour A., Gallagher S. C., Daley M., Richards G., 2015a, Submitted to MNRAS, 000

Tammour A., Gallagher S. C., Richards G., 2015b, MNRAS, 448, 3354

Treu T., Woo J.-H., Malkan M. A., Blandford R. D., 2007, ApJ, 667, 117

Trump J. R., et al., 2006, ApJS, 165, 1

Urry C. M., Padovani P., 1995, PASP, 107, 803

Vanden Berk D. E., et al., 2001, AJ, 122, 549

Véron-Cetty M.-P., Véron P., 2006, A&A, 455, 773

Véron-Cetty M.-P., Joly M., Véron P., 2004, A&A, 417, 515

Vestergaard M., Peterson B. M., 2006, ApJ, 641, 689

Weymann R. J., Morris S. L., Foltz C. B., Hewett P. C., 1991, ApJ, 373, 23

Wild V., Hewett P. C., 2010, preprint, (arXiv:1010.2500)

Wilson A. S., Tsvetanov Z. I., 1994, AJ, 107, 1227

Woo J.-H., Treu T., Malkan M. A., Blandford R. D., 2006, ApJ, 645, 900

Woo J.-H., Treu T., Malkan M. A., Blandford R. D., 2008, ApJ, 681, 925

Woo J.-H., et al., 2010, preprint, (arXiv:1004.0252)

Wright S. A., Larkin J. E., Graham J. R., Ma C.-P., 2010, ApJ, 711, 1291

Wu J., Vanden Berk D. E., Brandt W. N., Schneider D. P., Gibson R. R., Wu J., 2009, ApJ, 702, 767

Yip C. W., et al., 2004, AJ, 128, 2603

Yu Q., Tremaine S., 2002, MNRAS, 335, 965

Zhang Z., Liang Y., Hammer F., 2013a, preprint, (arXiv:1302.3013)

Zhang K., Wang T.-G., Gaskell C. M., Dong X.-B., 2013b, ApJ, 762, 51

Appendix A

Optical Spectral Library

This spectral library consists of 62 quasar spectra generated from a homogeneous set of quasars that share similar continuum luminosity, broad line width, and are at similar redshifts. Starting from a sample of $\sim 16,000$ quasars, we separate the objects into 4 groups of continuum luminosity (high, two intermediate, and low) and further separate each of these luminosity groups into 3 groups depending on their Hβ width (using 2000 km s^{-1} and 4000 km s^{-1} as boundaries). We further bin these groups at redshift steps of 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, and 0.7. The results are shown below. A full coloured version of the figures as well as the FITS files are available online[1].

[1] http://vizier.cfa.harvard.edu/viz-bin/VizieR?-source=J/MNRAS/448/3354

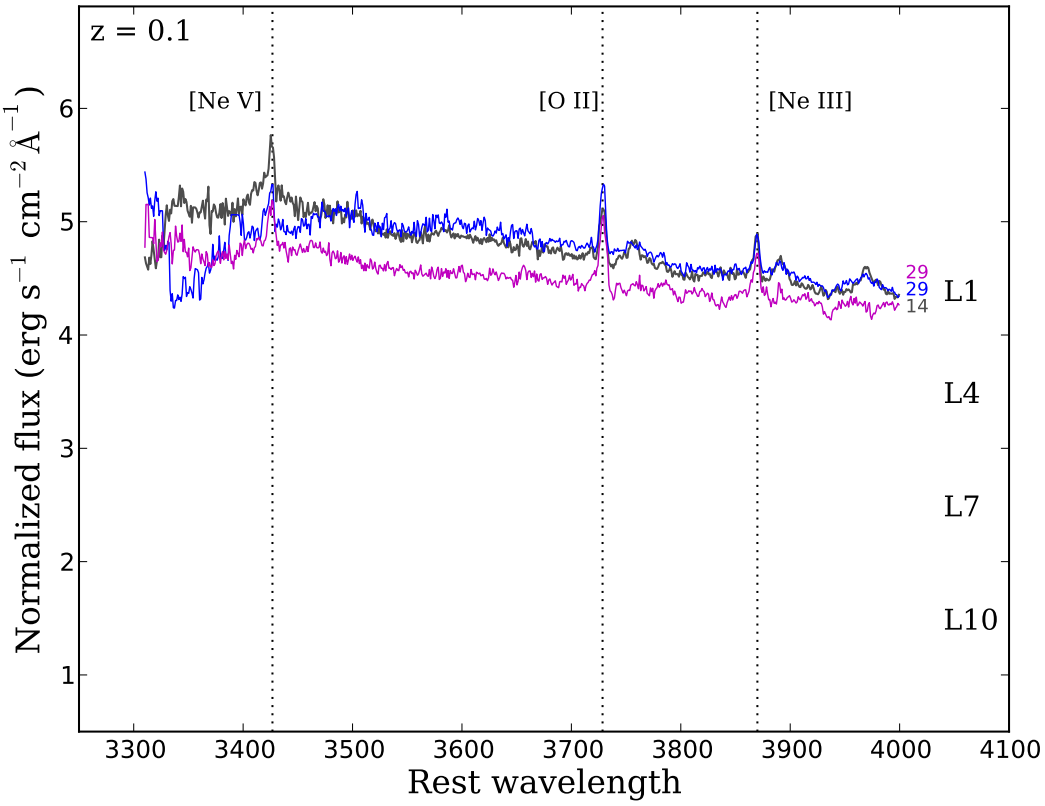

Figure A.1: Median composite spectra at redshift $z = 0.1$. Top is the high continuum luminosity (L10) and bottom is the low luminosity (L1). Spectra at the same luminosity level but with different widths of Hβ are overplotted with narrow Hβ in grey, intermediate Hβ in blue, and broad Hβ in magenta. To the right side, we show the number of individual spectra used in making each composite. Only composites with > 10 objects are shown. L1, L4, L7 and L10 indicate the continuum luminosity subsets with L1 being the lowest and L10 the highest. At $z = 0.1$, only the lowest luminosity bin has data.

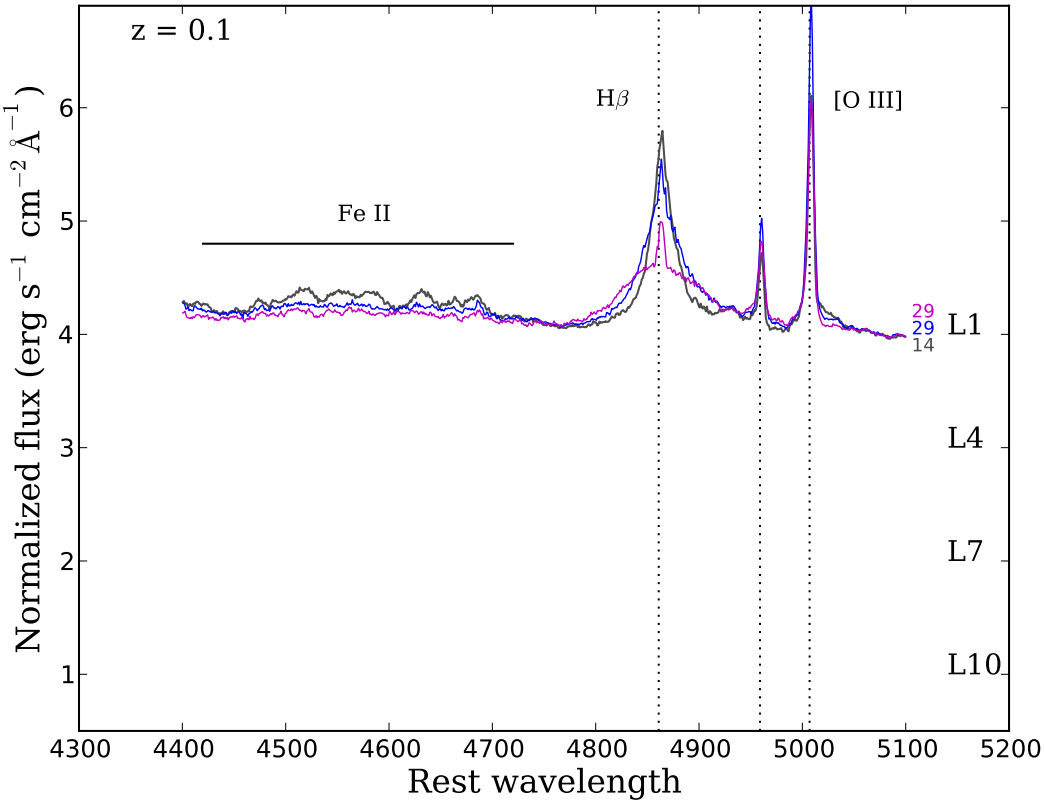

Figure A.2: Similar to Fig. A.1 but for the wavelength range covering Hβ and [O III].

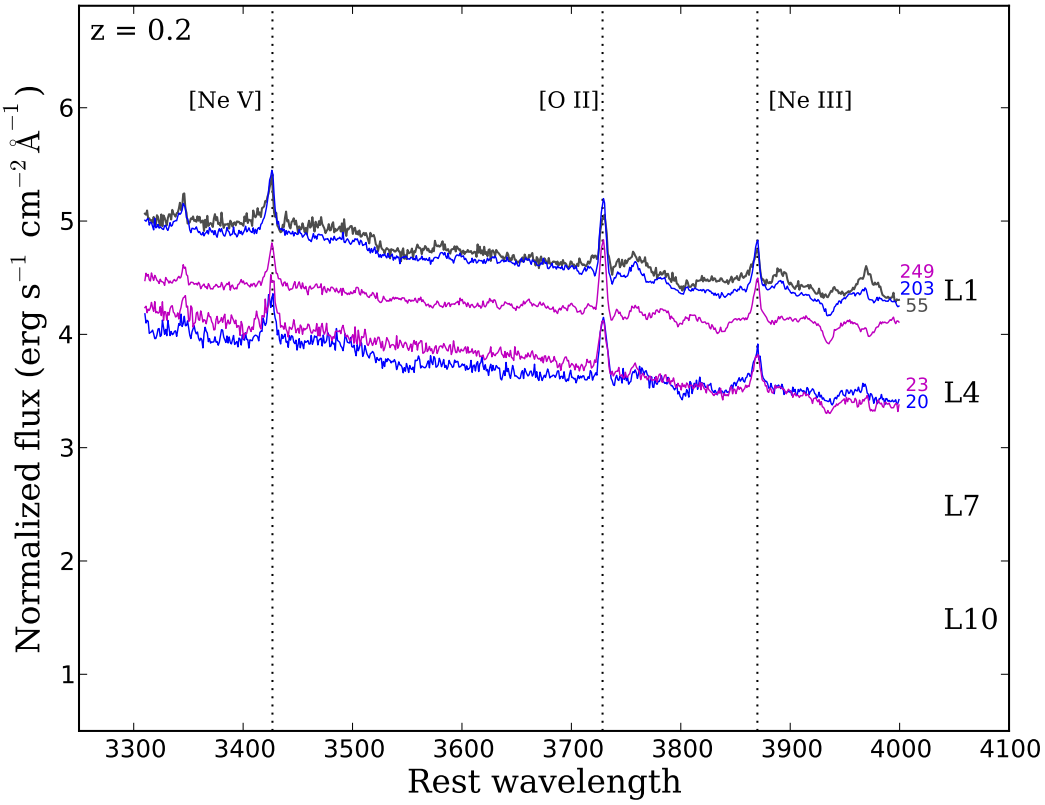

Figure A.3: Same as Fig. A.1 but for redshift bin centred at $z = 0.2$.

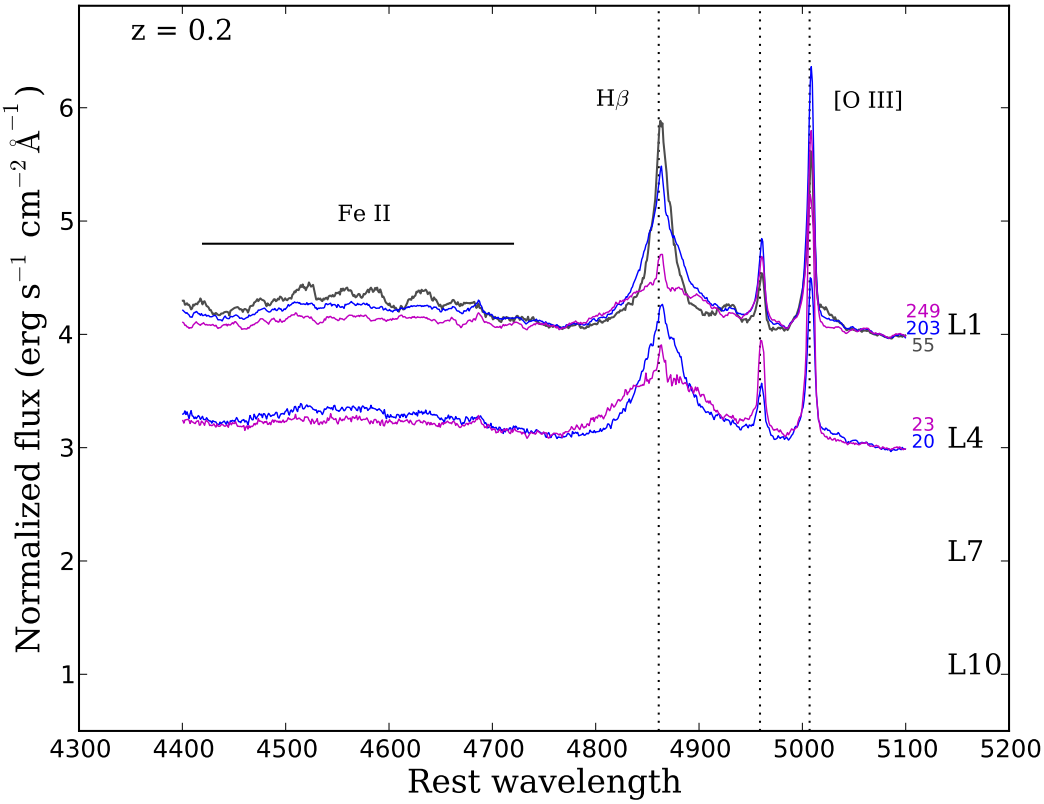

Figure A.4: Same as Fig. A.2 but for redshift bin centred at $z = 0.2$.

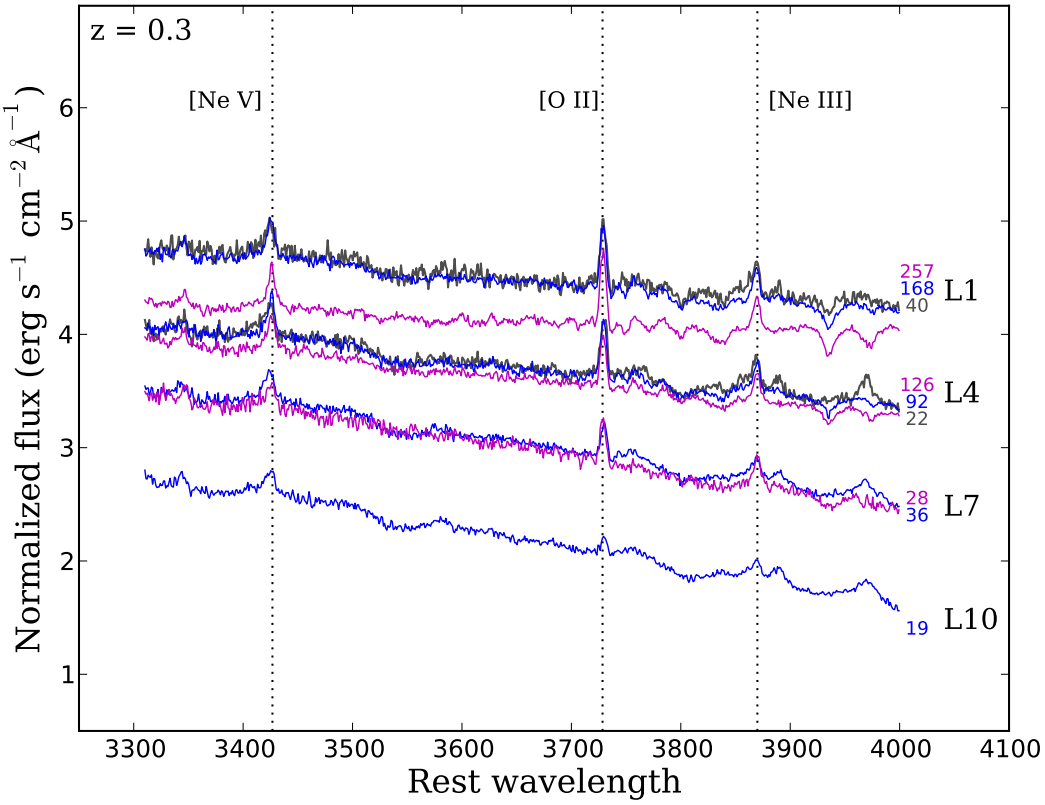

Figure A.5: Same as Fig. A.1 but for redshift bin centred at $z = 0.3$.

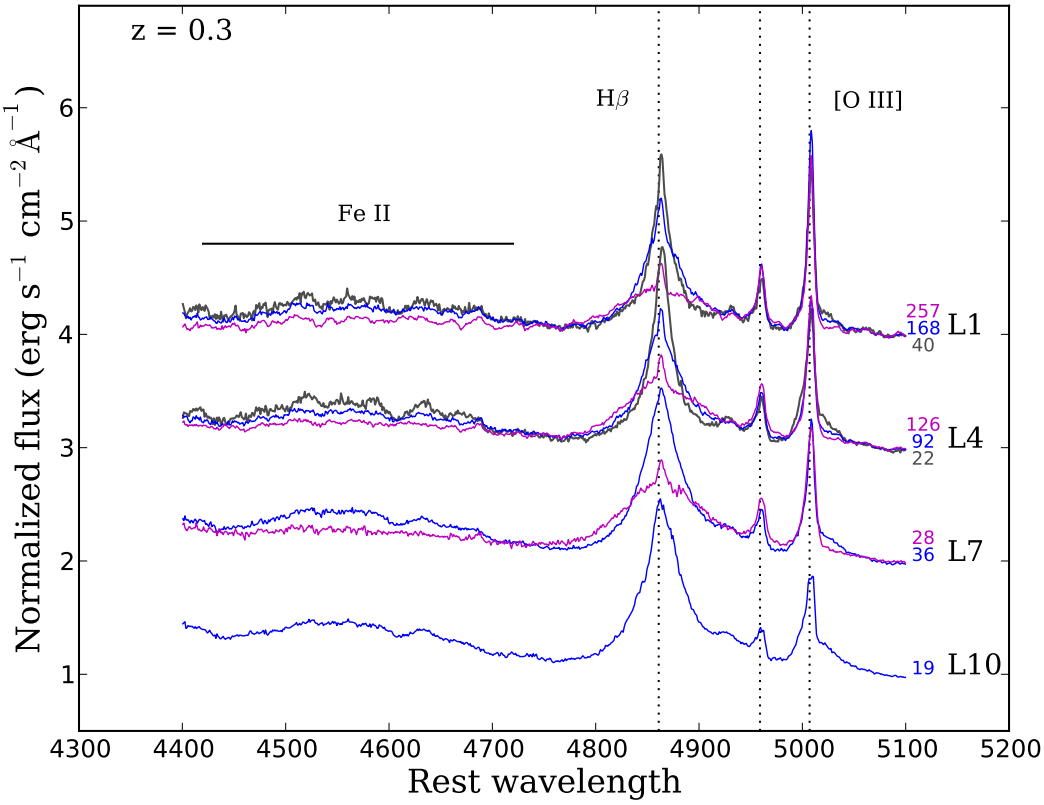

Figure A.6: Same as Fig. A.2 but for redshift bin centred at $z = 0.3$.

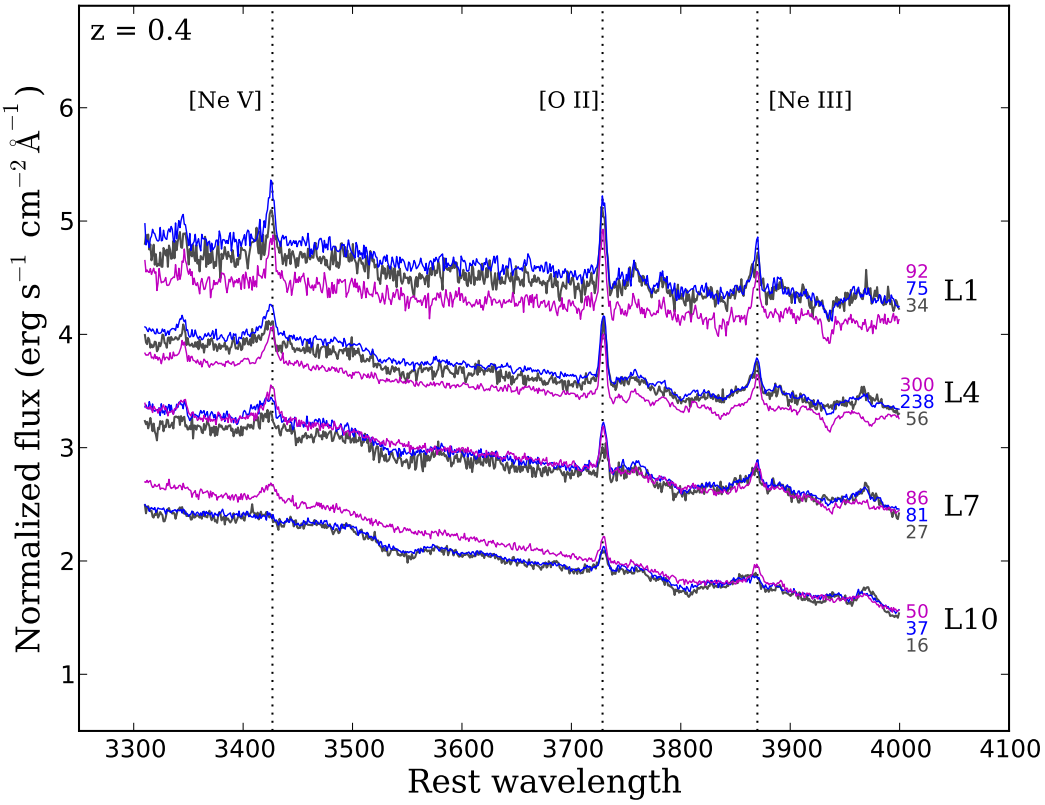

Figure A.7: Same as Fig. A.1 but for redshift bin centred at $z = 0.4$.

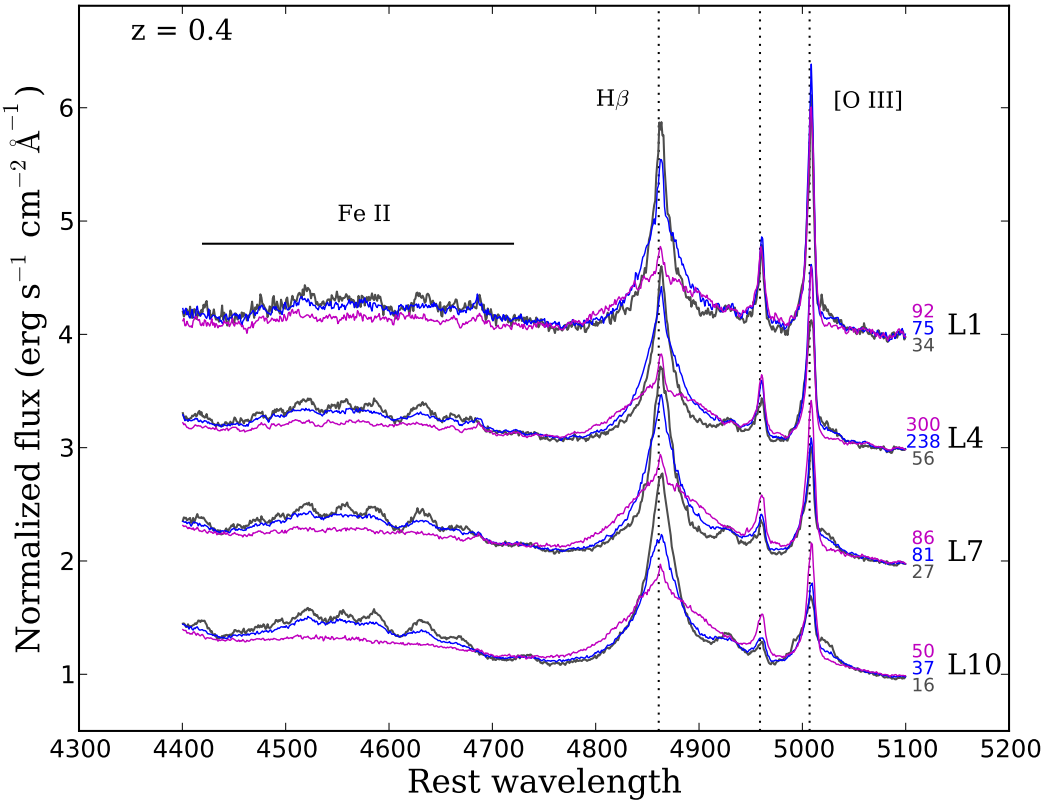

Figure A.8: Same as Fig. A.2 but for redshift bin centred at $z = 0.4$.

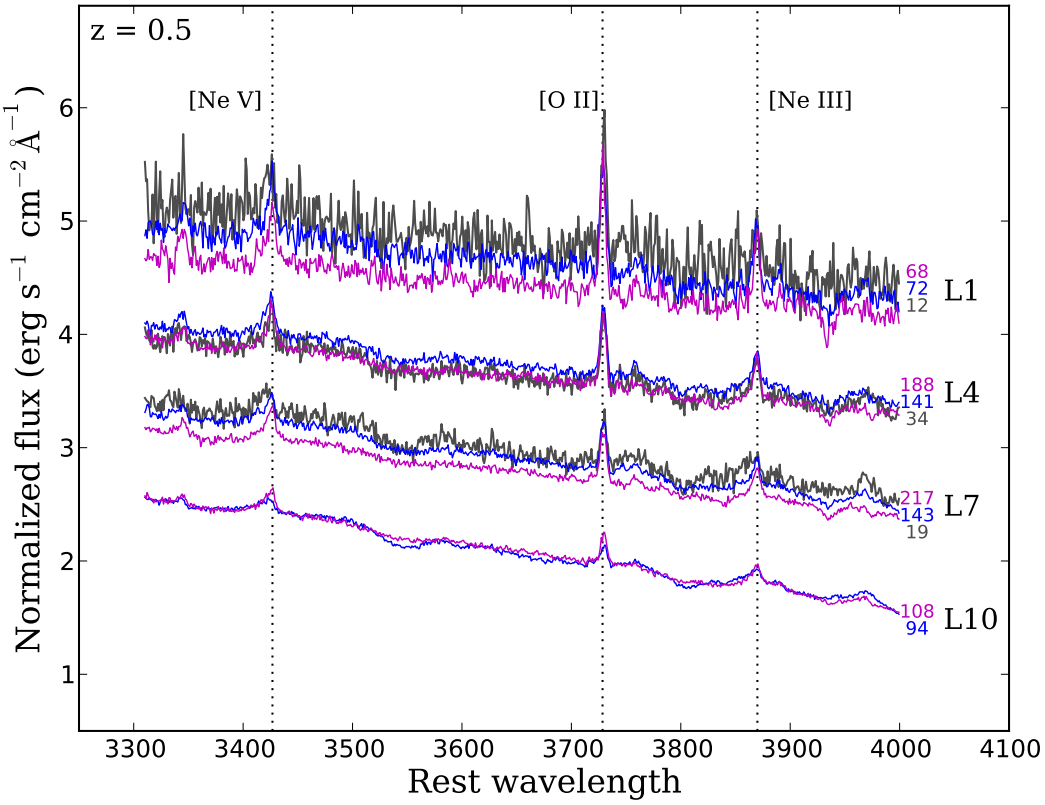

Figure A.9: Same as Fig. A.1 but for redshift bin centred at $z = 0.5$.

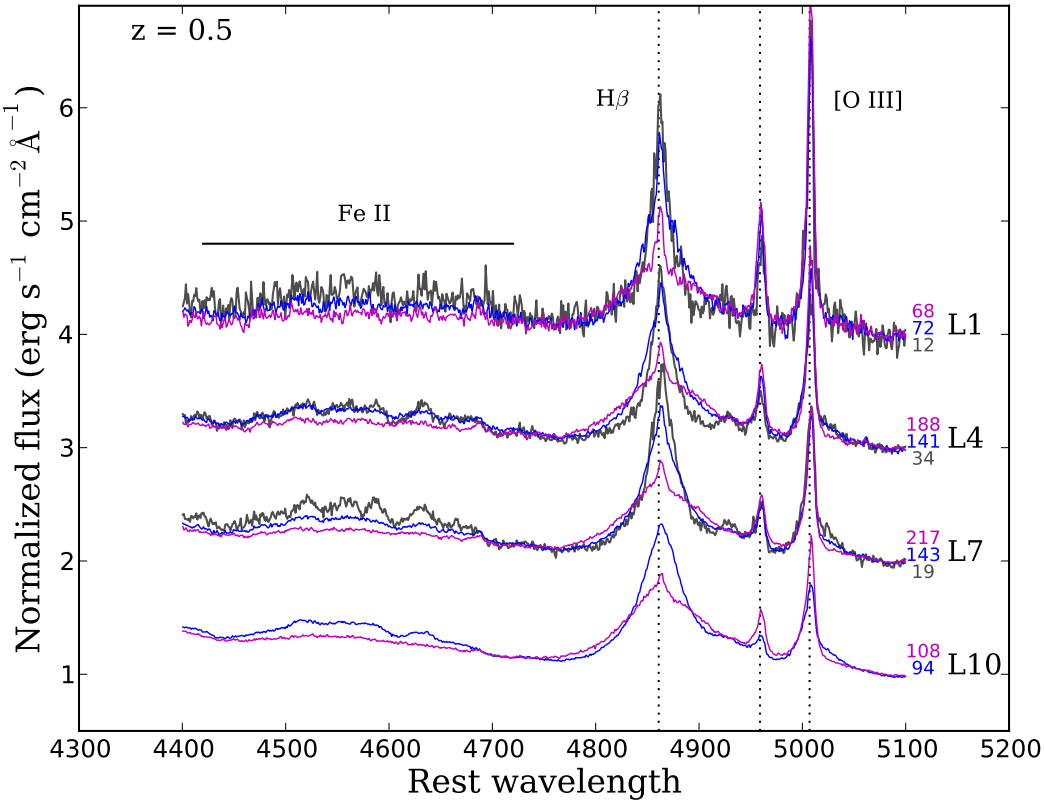

Figure A.10: Same as Fig. A.2 but for redshift bin centred at $z = 0.5$.

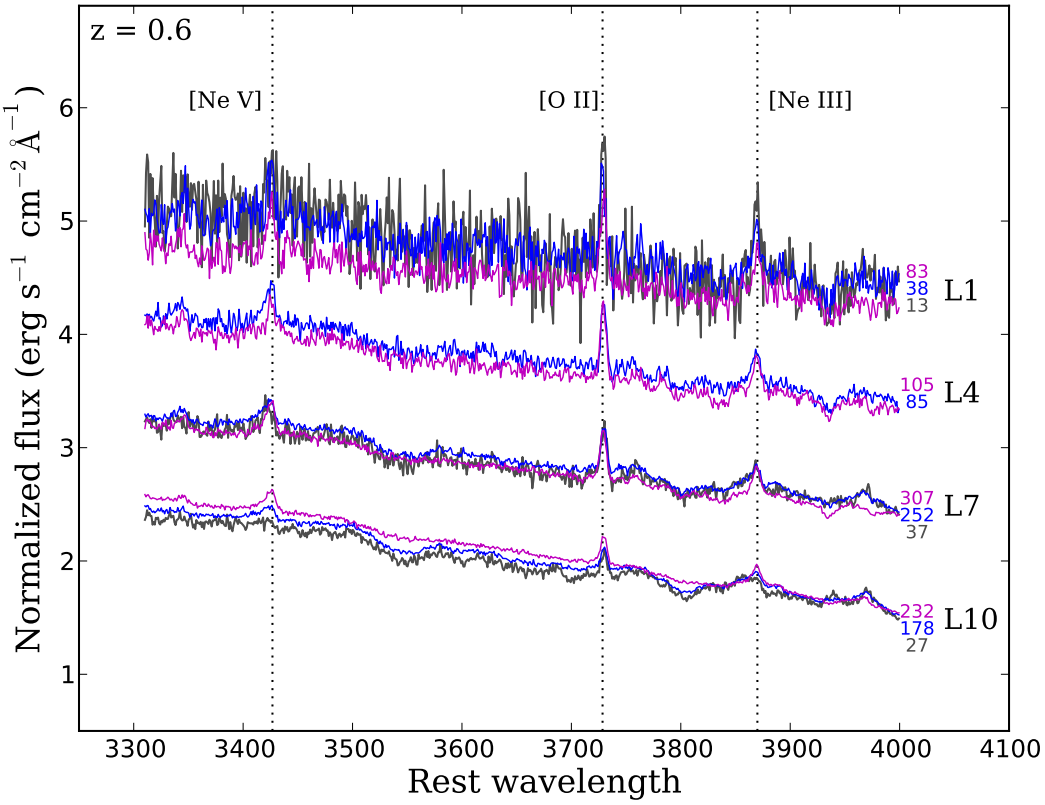

Figure A.11: Same as Fig. A.1 but for redshift bin centred at $z = 0.6$.

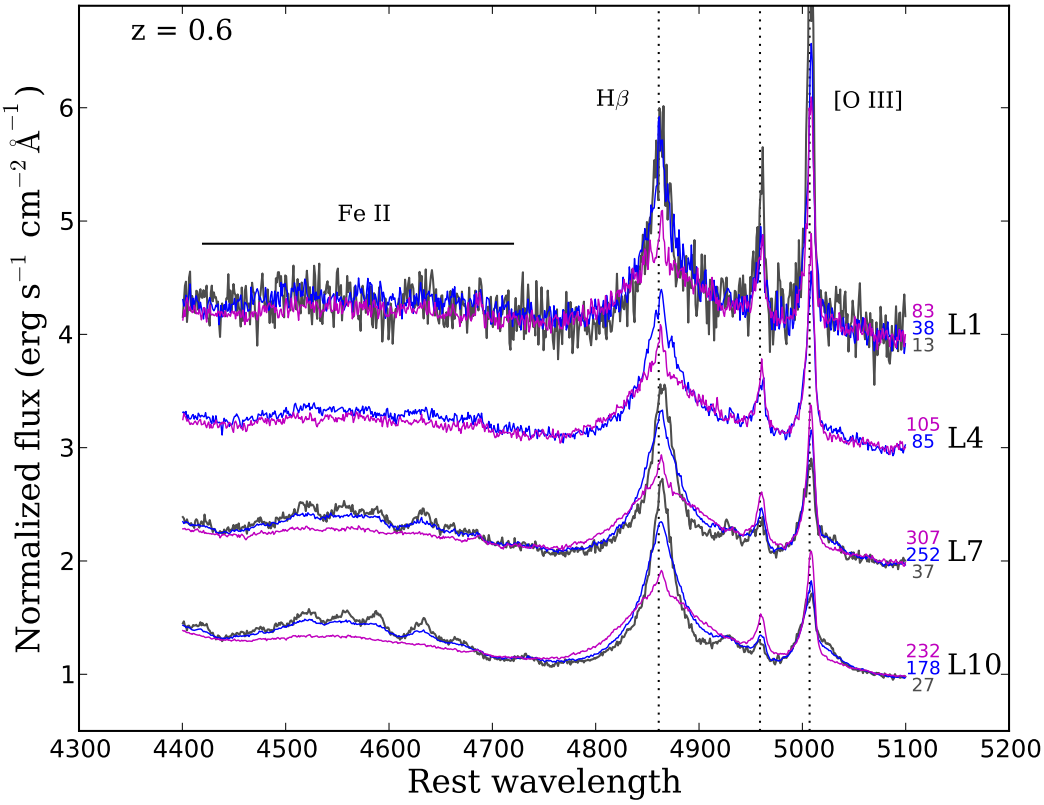

Figure A.12: Same as Fig. A.2 but for redshift bin centred at $z = 0.6$.

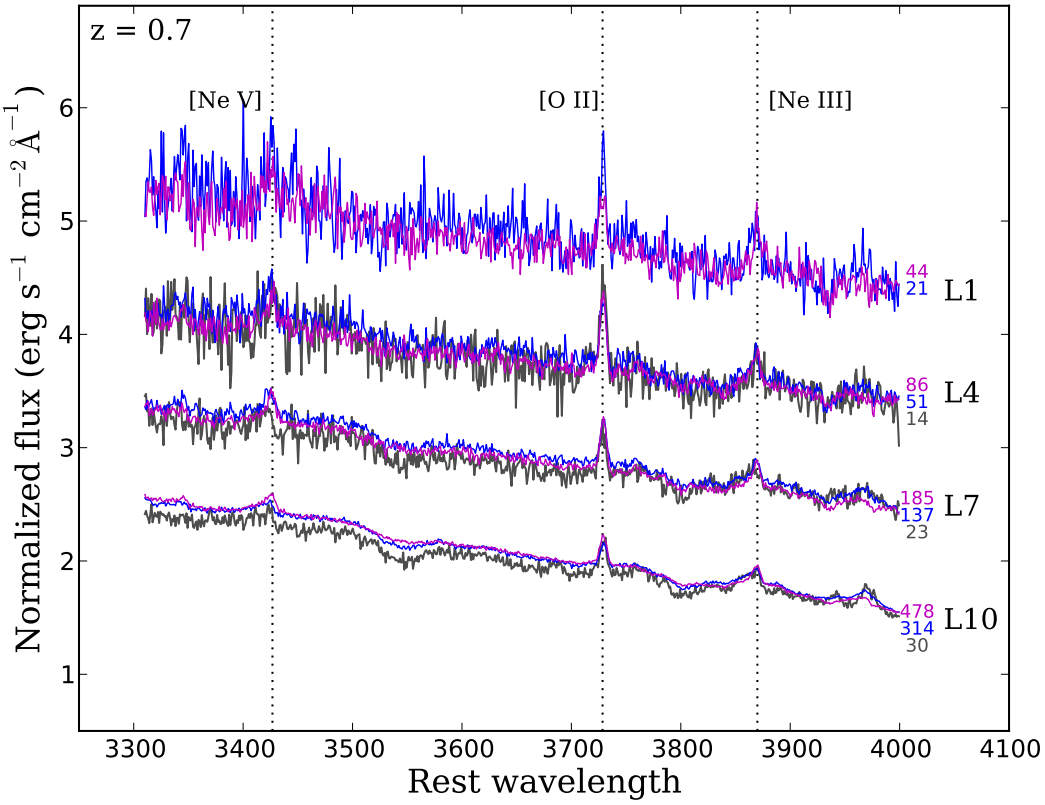

Figure A.13: Same as Fig. A.1 but for redshift bin centred at $z = 0.7$.

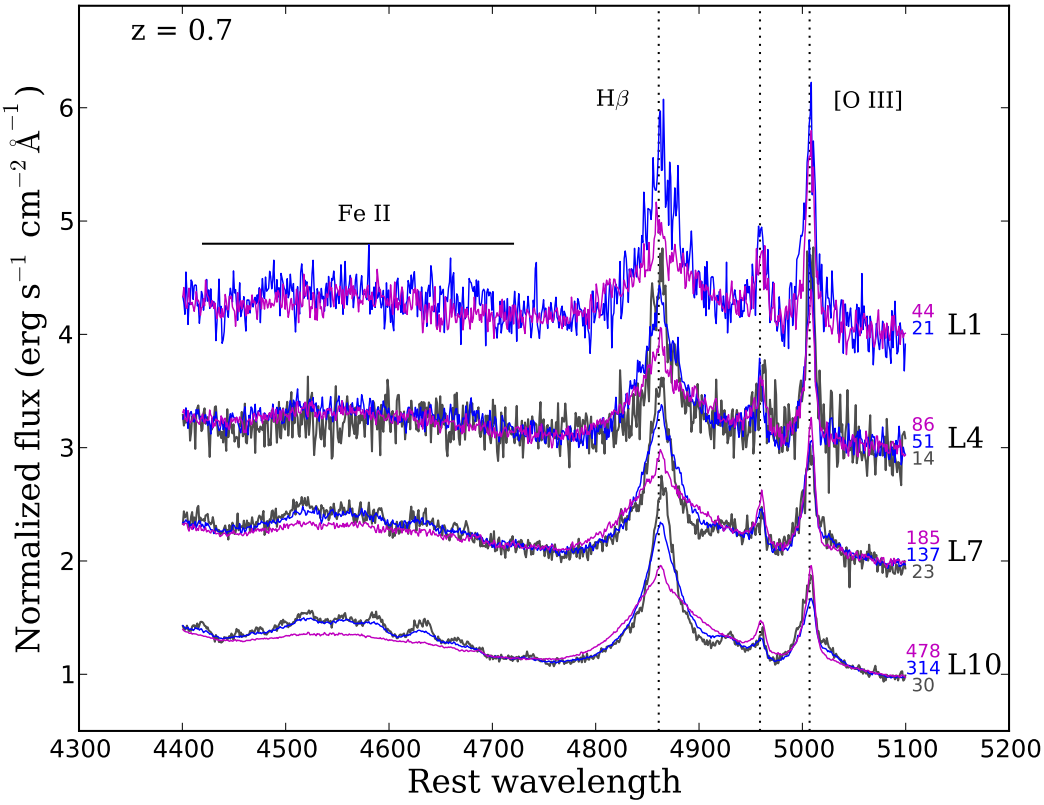

Figure A.14: Same as Fig. A.2 but for redshift bin centred at $z = 0.7$.

Appendix B

UV Composite Spectra & Reproducibility of Clusters

B.1 Spectra

For completeness, we show here the composite spectra for the rest of the *K* runs that we did not include in the analysis in §3.3 for the main sample as well as the mixed and broad absorption line quasars (BALQ) samples. The median composite spectra are generated in a similar manner to what we describe in §3.2.3.

B.1.1 Main sample -Mg II

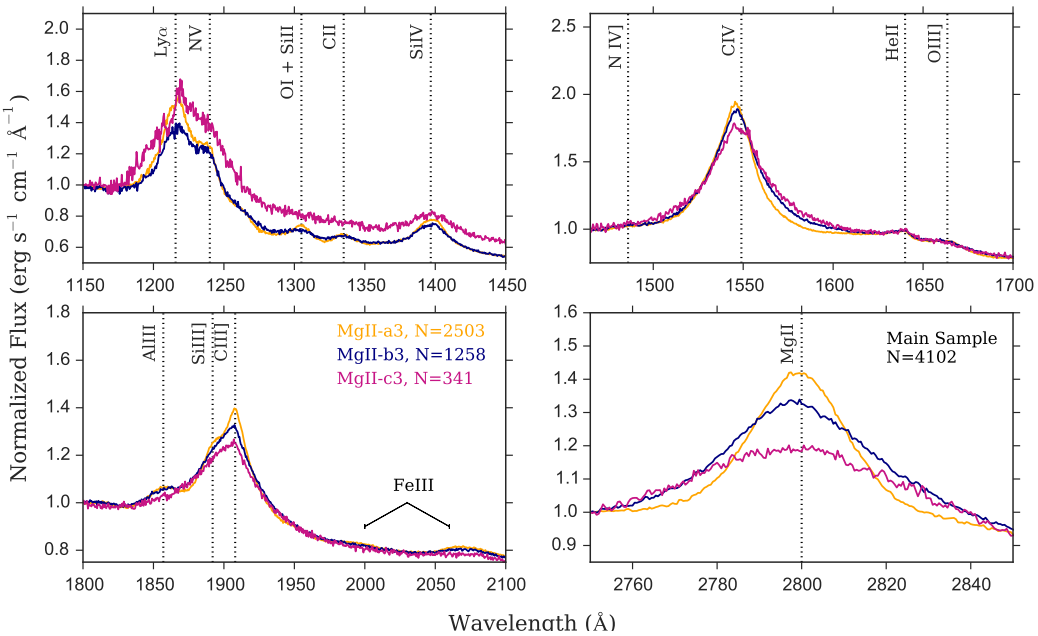

Figure B.1: Median composite spectra made from the objects in the Mg II clusters in Fig. 3.3. Similar to Fig. 3.5 but for $K = 3$.

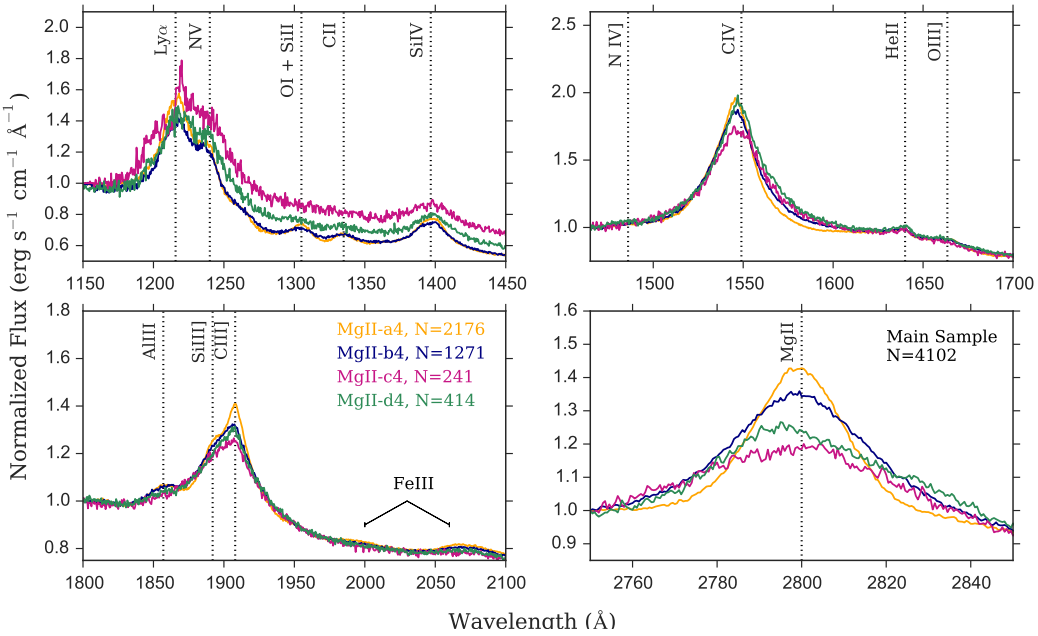

Figure B.2: Median composite spectra made from the objects in the Mg II clusters in Fig. 3.3 Similar to Fig. 3.5 but for $K = 4$.

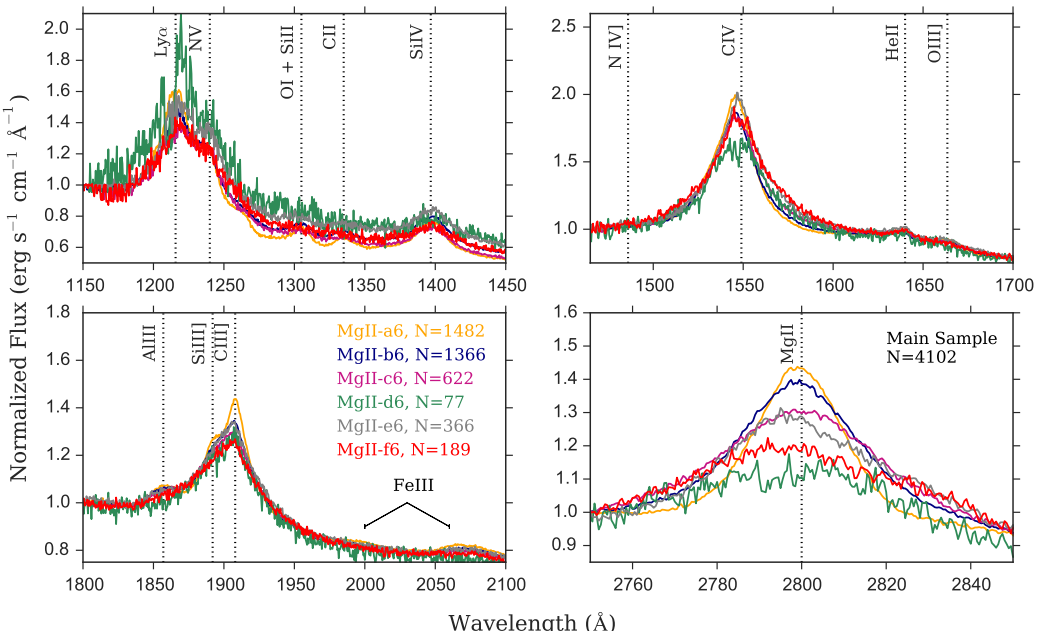

Figure B.3: Median composite spectra made from the objects in the Mg II clusters in Fig. 3.3 Similar to Fig. 3.5 but for $K = 6$.

B.1.2 Main Sample -C IV

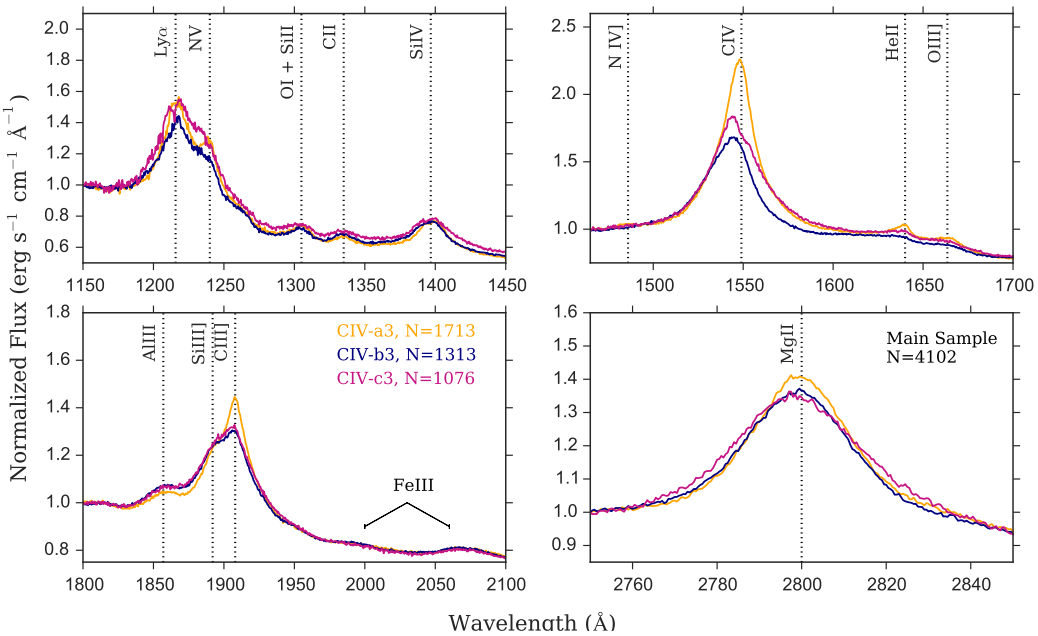

Figure B.4: Median composite spectra made from the objects in the C IV clusters. Similar to Fig. 3.7 but for $K = 3$.

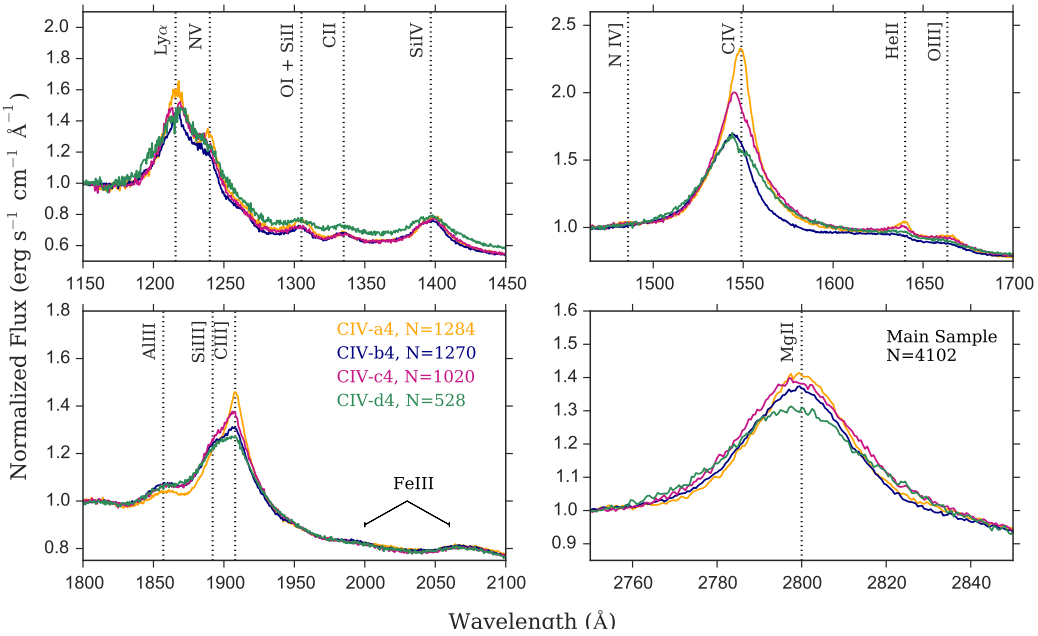

Figure B.5: Median composite spectra made from the objects in the C IV clusters. Similar to Fig. 3.7 but for $K = 4$.

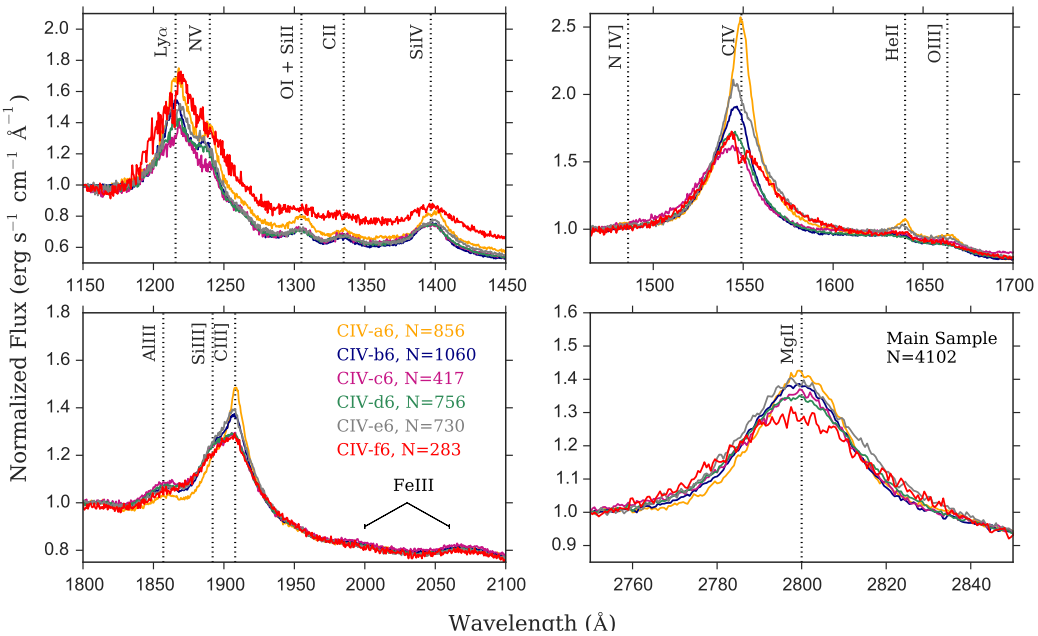

Figure B.6: Median composite spectra made from the objects in the C IV clusters. Similar to Fig. 3.7 but for $K = 6$.

B.1.3 Main sample -C III]

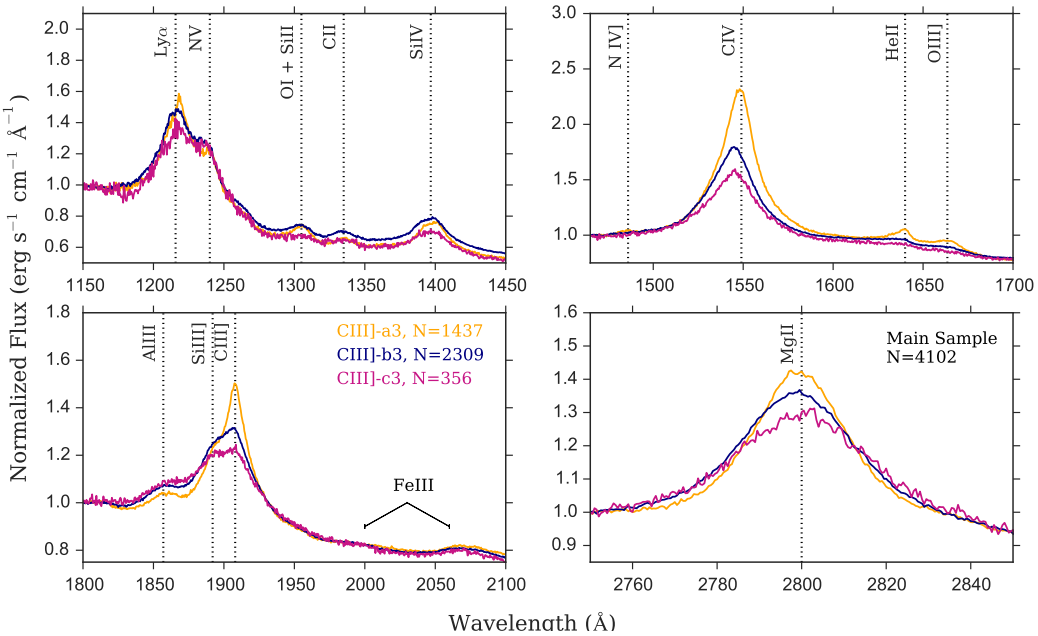

Figure B.7: Median composite spectra made from the objects in the C III] clusters. Similar to Fig. 3.10 but for $K = 3$.

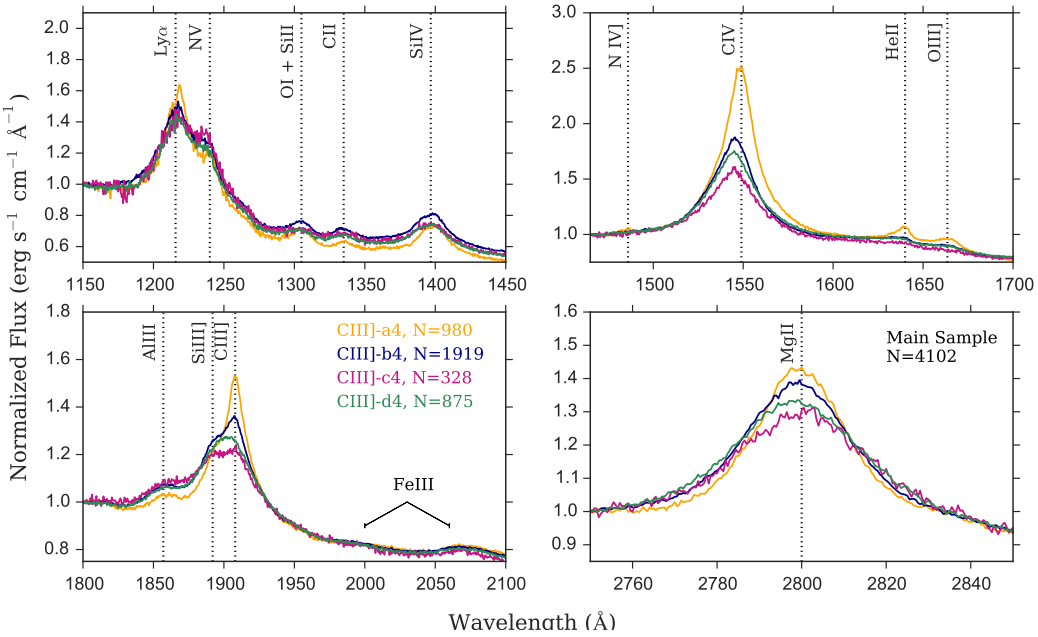

Figure B.8: Median composite spectra made from the objects in the C III] clusters. Similar to Fig. 3.10 but for $K = 4$.

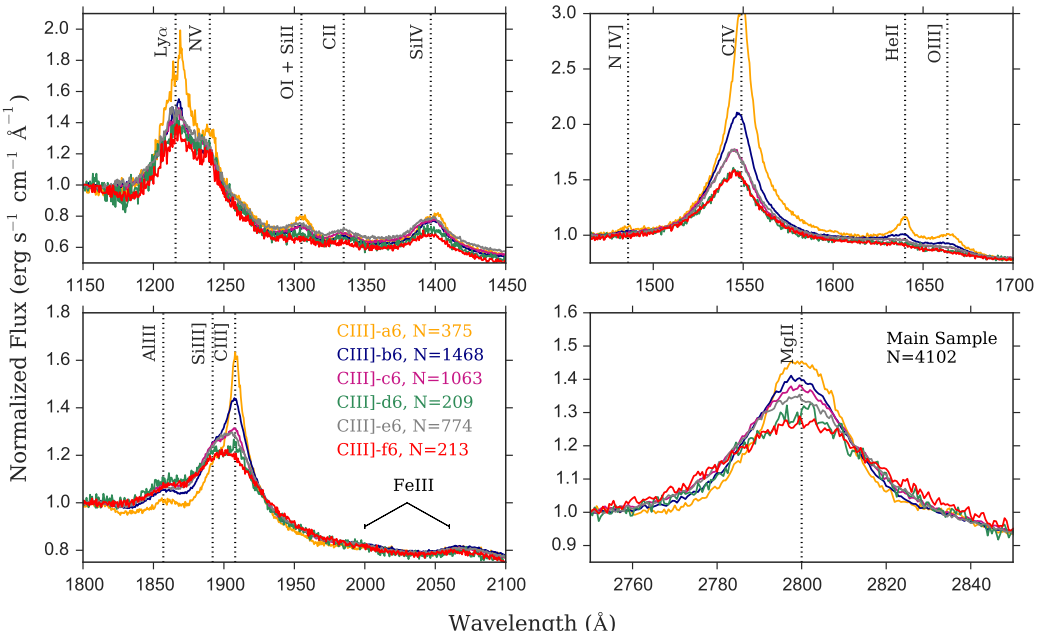

Figure B.9: Median composite spectra made from the objects in the C III] clusters. Similar to Fig. 3.10 but for $K = 6$.

B.1.4 Mixed Sample

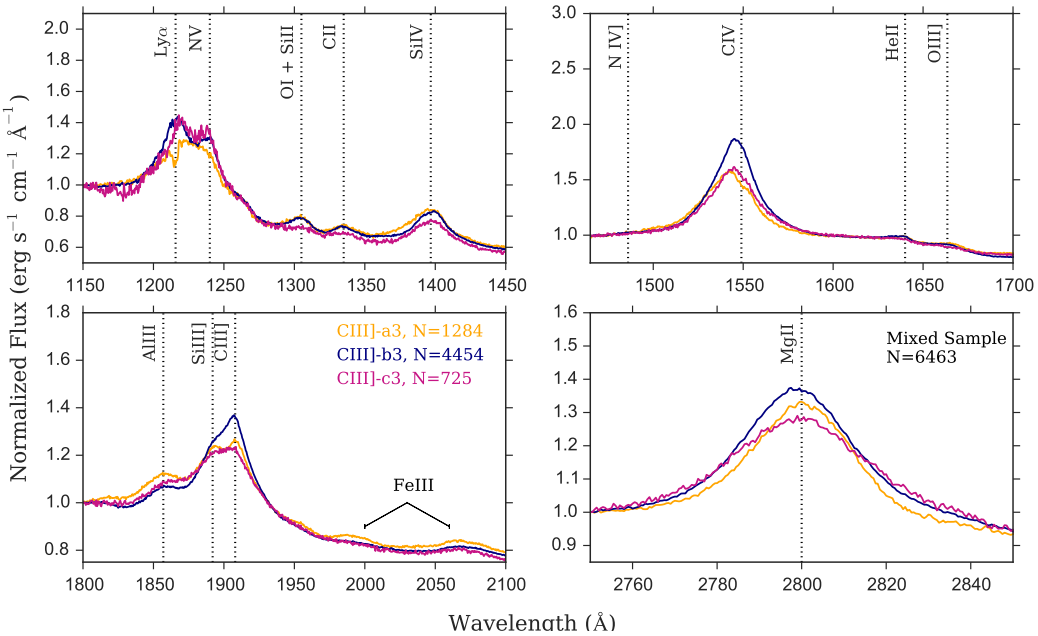

Figure B.10: Median composite spectra made from the objects in the C III] clusters in the mixed sample. Similar to Fig. 3.12 but for $K = 3$.

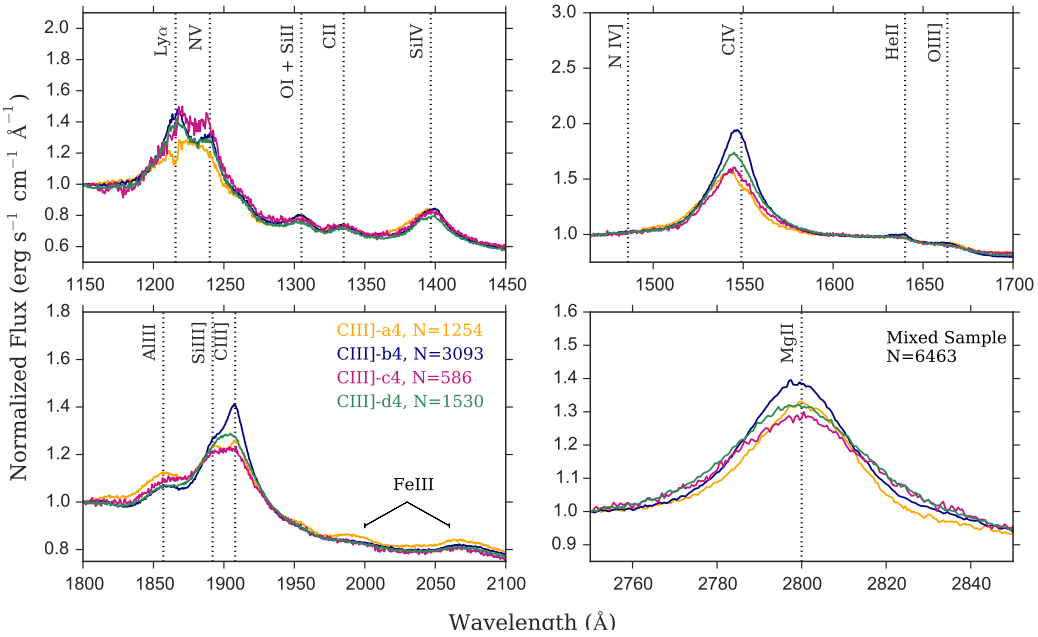

Figure B.11: Median composite spectra made from the objects in the C III] clusters in the mixed sample. Similar to Fig. 3.12 but for $K = 4$.

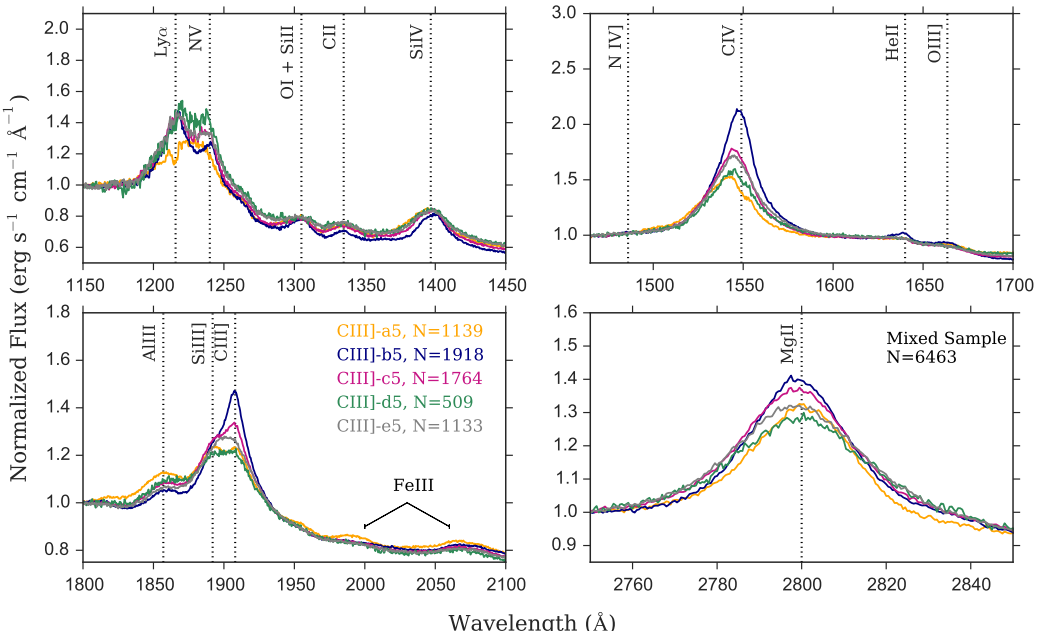

Figure B.12: Median composite spectra made from the objects in the C III] clusters in the mixed sample. Similar to Fig. 3.12 but for $K = 5$.

B.1.5 BAL Quasars Only

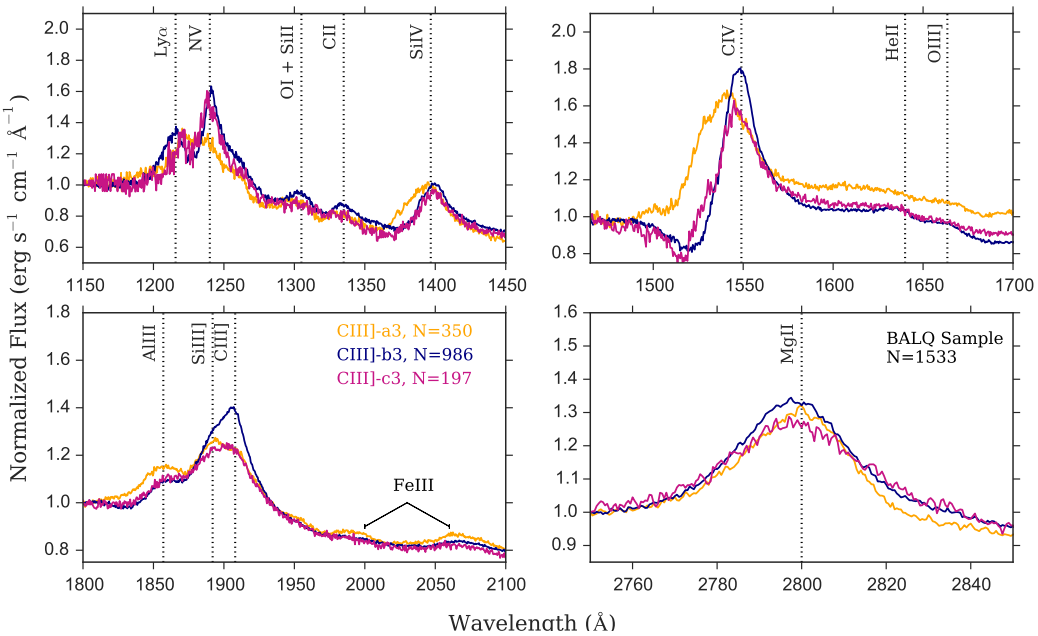

Figure B.13: Median composite spectra made from the objects in the C III] clusters in the BALQs sample. Similar to Fig. 3.13 but for $K = 3$.

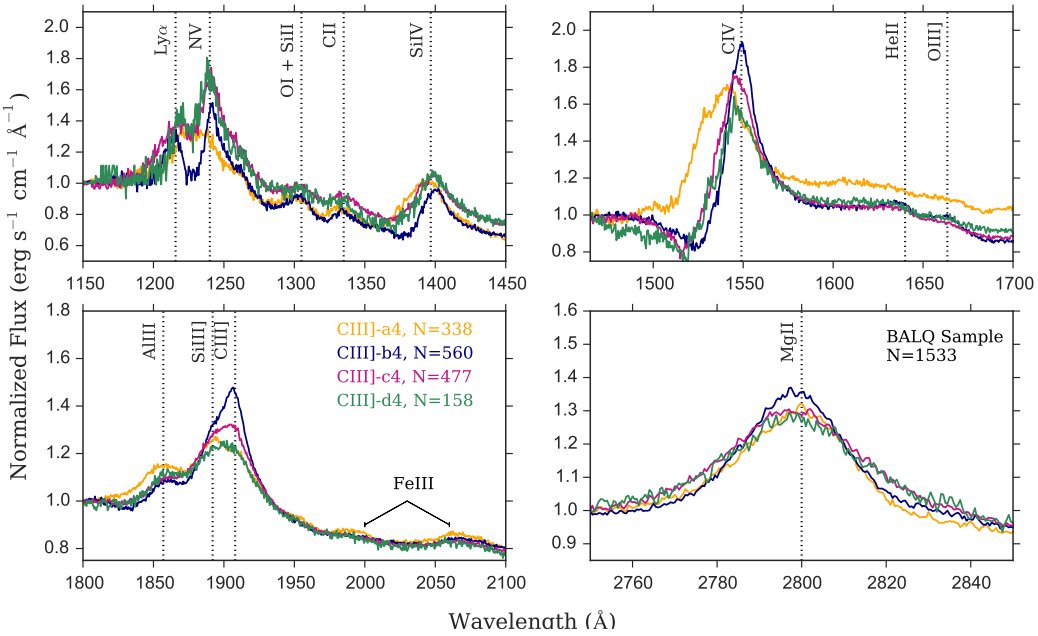

Figure B.14: Median composite spectra made from the objects in the C III] clusters in the BALQs sample. Similar to Fig. 3.13 but for $K = 4$.

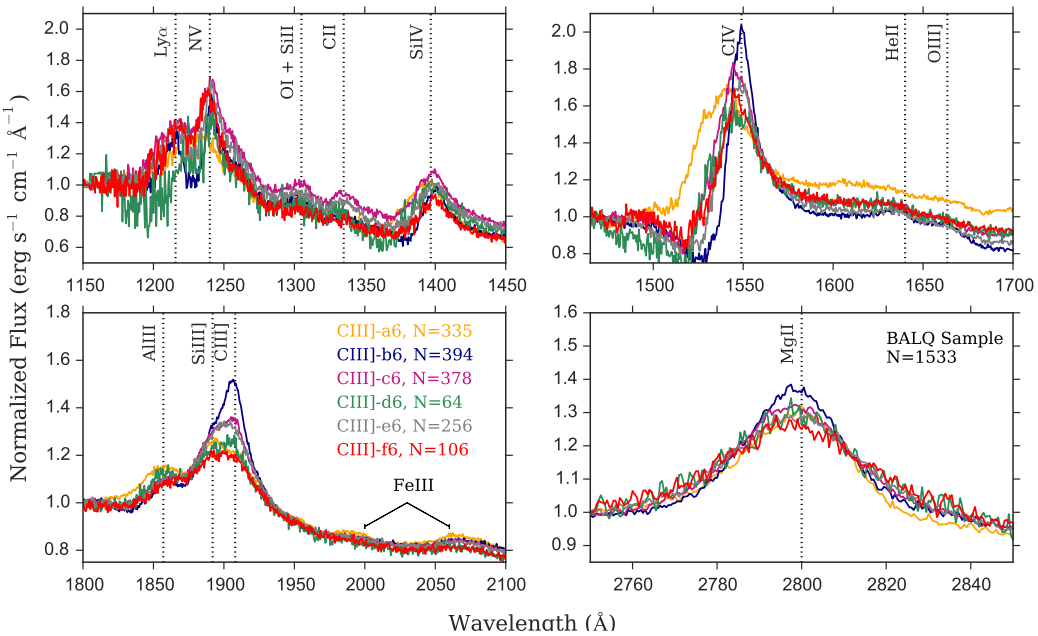

Figure B.15: Median composite spectra made from the objects in the C III] clusters in the BALQs sample. Similar to Fig. 3.13 but for $K = 6$.

B.2 Reproducibility of Clusters

We tested whether the algorithm is able to find the same unique set of centroids when the clustering is repeated. We find that for the clusters generated in the main sample, the algorithm is finding the same set of clusters even after 50 repeats. We show here one example for $K = 5$.

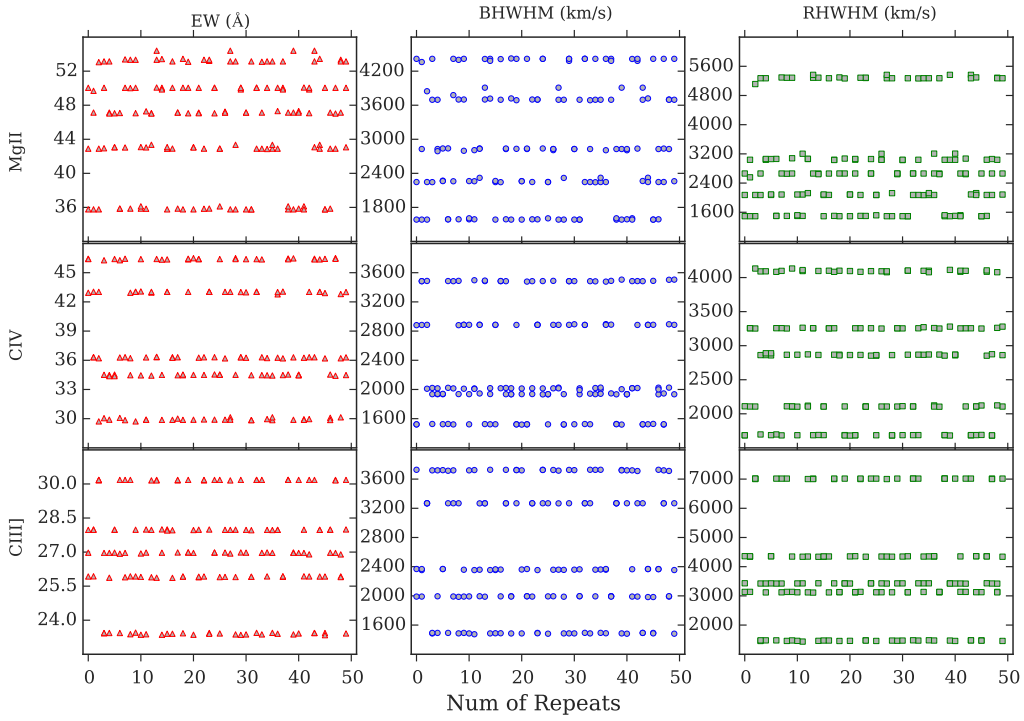

Figure B.16: The centroids for three different clustering attempts one for each of the lines (Mg II, C IV, and C III]) in the main sample repeated 50 times using $K = 5$. This shows that K-Means is finding clusters that are consistent and that the clusters we use in our analysis reflect meaningful groupings of like objects in the parameter space of EW, RHWHM and BHWHM.

Appendix C

BAL Quasars: Clusters and Composite Spectra

C.1 Appendix C: Clusters and Composite Spectra

For completeness, we show here the clusters and composite spectra generated in a similar fashion as described in §4.2.2 for the cases not discussed in the main text (i.e., $K = 3, 4,$ and 5). The general trends revealed in the composite spectra remain, and so our conclusions are not driven by the input K value.

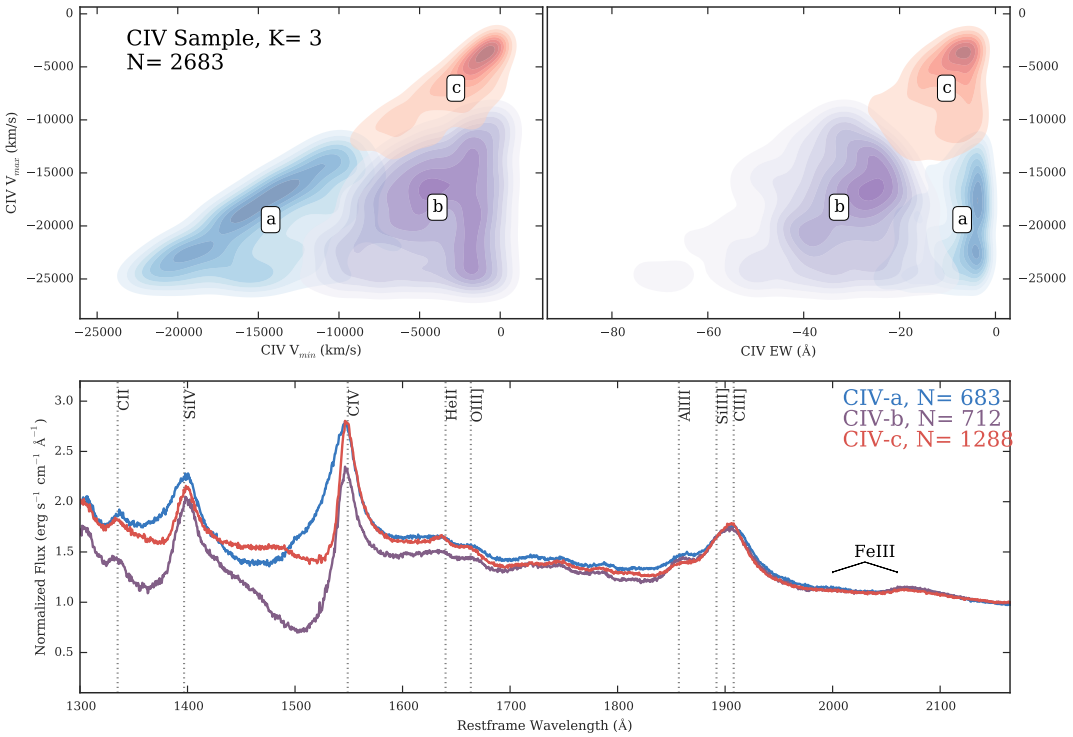

Figure C.1: Projections in the clustering parameter space of C IV V_{max}vs. V_{min} (top-left) and V_{max} vs. EW (top-right) using $K = 3$. The white labels follow the cluster designation given in table 4.1.

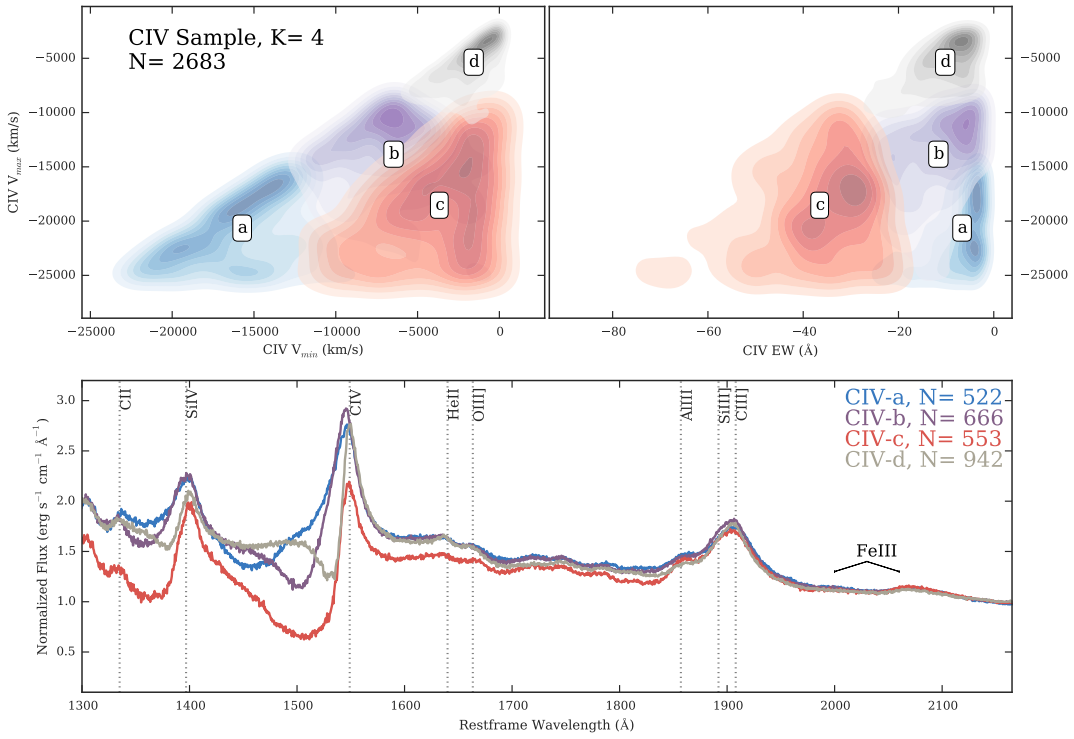

Figure C.2: Similar to Fig. C.1 but for $K = 4$.

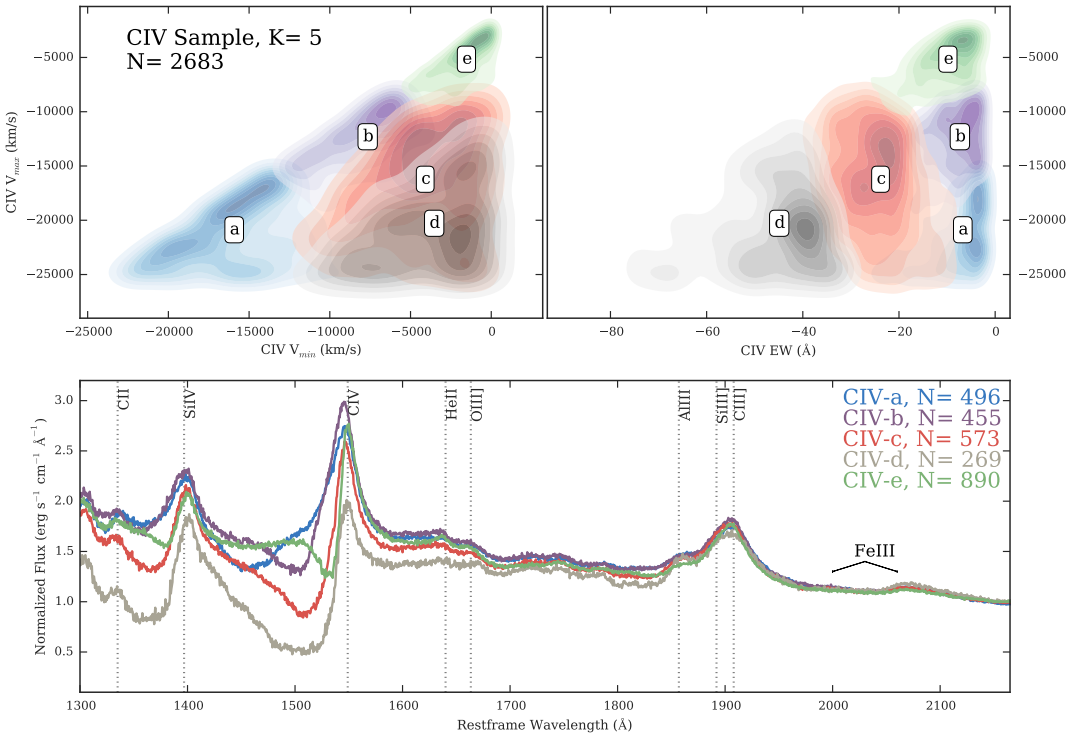

Figure C.3: Similar to Fig. C.1 but for $K = 5$.

Curriculum Vitae

Name:	Aycha Tammour
Post-Secondary	The University of Western Ontario, London, ON, Canada
Education &	2011 - 2016 Ph.D., Astronomy
Degrees:	Minnesota State University, Mankato, MN, USA
	2008 - 2011 M.Sc., Physics
	Al-Baath University, Homs, Syria
	2000 - 2004 B.Sc., Physics
Related Work	Graduate Research Assistant
Experience:	University of Western Ontario, Canada, 2011 - 2016
	Graduate Teaching Assistant
	University of Western Ontario, Canada, 2011 - 2015
	Minnesota State University, USA, 2008 - 2011
	Physic Instructor
	Tishreen University, Latakia, Syria, 2006 - 2008
Honours &	Award of Excellence in Teaching
Awards:	Dept of Phys & Astro, University of Western Ontario – 2014
	Schlumberger Faculty for the Future Fellow
	The Schlumberger Foundation – 2012 - 2013
	Fulbright Fellow
	US Department of State – 2008 - 2011
	Al-Basil Award of Excellence
	Al-Baath University, Syria – 2002 and 2001

Publications:

- Tammour, A., Gallagher, S. C., & Richards, G., Tracing Quasar Narrow- Line Regions Across Redshift: A Library of High Signal-to-Noise Optical Spectra, 2015, MNRAS, 448, 3354

- Tammour, A., Gallagher, S. C., Daley, M., & Richards, G., Insights into Quasar UV Spectra Using Unsupervised Clustering Analysis, 2016, MNRAS, 459, 1659