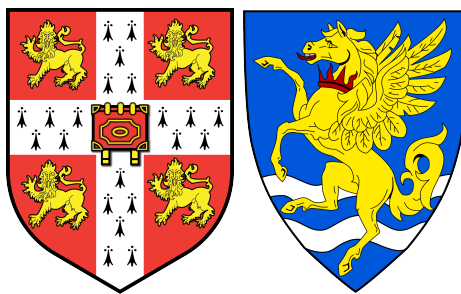LIAM COATMAN

BLACK HOLE MASSES AND OUTFLOWS IN HIGH REDSHIFT QUASARS

A DISSERTATION SUBMITTED TO THE
UNIVERSITY OF CAMBRIDGE FOR THE DEGREE
OF DOCTOR OF PHILOSOPHY

BLACK HOLE MASSES AND OUTFLOWS IN HIGH REDSHIFT QUASARS

LIAM COATMAN
UNIVERSITY OF CAMBRIDGE
INSTITUTE OF ASTRONOMY
ROBINSON COLLEGE
SUBMITTED TO THE BOARD OF GRADUATE STUDIES APRIL 2017
UNDER THE SUPERVISION OF
PROF. PAUL C. HEWETT
DR. MANDA BANERJI

CONTENTS

LIST OF TABLES

LISTINGS

ACRONYMS

INTRODUCTION

1.1 THE AGN-HOST GALAXY CONNECTION

Super-massive black holes (BHs) are found at the centres of most nearby massive galaxies and the BH mass and mass of the host galaxy spheroid are strongly correlated (Ferrarese and Merritt, 2000; Gebhardt et al., 2000; Kormendy and Ho, 2013). Although any underlying causal mechanism(s) responsible for the correlation is yet to be conclusively identified, there is considerable observational and theoretical support for models that involve BH-fuelling, outflows and a 'feedback' relationship (e.g. King and Pounds, 2015). The number density of quasars, which evolves strongly with redshift, peaks at redshifts $2 \lesssim z \lesssim 3$ (e.g. Brandt and Hasinger, 2005; Richards et al., 2006b) and the most massive ($M_{BH} \gtrsim 10^9 M_\odot$) present-day BHs experienced much of their growth during this epoch. The star formation rate, which closely follows the cosmological evolution of the quasar luminosity function, also peaks during this epoch (e.g. Boyle and Terlevich, 1998). Quantifying the growth-rate of massive BHs at $2 \lesssim z \lesssim 3$ would therefore help significantly in understanding the role quasars play in galaxy evolution.

There is now considerable observational and theoretical support for models of galaxy formation that involve black hole-fuelling, outflows and a 'feedback' relationship between active black holes and star formation in the host galaxy. Super-massive black holes accreted most of their mass and galaxies formed most of their stars at redshifts $z \gtrsim 2$ (e.g. Madau & Dickinson 2014 for star formation; find quasar reference.) During this key cosmological epoch star formation is believed to be suppressed by the energy output from the quasar, establishing the tight relationship between BH mass and host galaxy spheroid mass observed in the local Universe (e.g. Kormendy & Ho 2013).

Correlations between the masses of super-massive black holes and properties of the host galaxy bulges in nearby galaxies (Gebhardt et al., 2000; Ferrarese and Merritt, 2000) and the similarity of the cosmic evolution of star formation and black hole activity (Boyle and Terlevich, 1998; Madau and Dickinson, 2014) suggest that the formation and evolution of supermassive black holes and their host galaxies are linked. Active galactic nuclei (AGN) and quasar feedback, in which star formation in the host galaxy is supressed by the energy output from the quasar, is a favoured model. This has motivated a considerable amount of observational work searching for feedback signatures

(for recent reviews, see Alexander and Hickox, 2012; Fabian, 2012; Heckman and Best, 2014).

The energy released by accretion on to supermassive black holes is far in excess of the gravitational binding energy in their host galaxies. If just a few per cent of this energy couples to the gas in the galaxy, then quasars must significantly affect the evolution of their host galaxies.

Quasar feedback invoked to explain absence of verly luminous galaxies, black hole-bulge scaling relations, and the similarity of black hole accretion and star formation histories (Tabor & Binney 1993; Silk & Rees 1998; Springel et al. 2005; Hopkins et al. 2006).

While supermassive black holes have masses only 0.1% of their host galaxies, they are now thought to exert a significant controlling effect on galaxy evolution (Tabor & Binney 1993; Silk & Rees 1998; Springel et al. 2005).

One possibility for such a connection, known as "quasar feedback", is that the energy output of the black hole in its most luminous (quasar) phase becomes cou- pled to the gas from which stars in the host galaxy would otherwise form. The gas is then reheated or pushed out of the galactic potential (Hopkins et al. 2006), result- ing in suppression of star formation and in a "quenched" galaxy.

1.2 MEASURING BLACK HOLE MASSES

The goal of better understanding the relationship between super-massive BH accretion and star formation has led to much work fo-cussing on the properties of quasars and active galactic nuclei at these redshifts. Accurate BH mass estimates for quasars are essential in these studies. Furthermore, as one of just two fundamental quantities describing a black hole on astrophysical scales, the mass is of cru-cial importance to virtually all areas of quasar science, including the evolution and phenomenology of quasars, and accretion physics.

1.2.1 *Reverberation Mapping*

Reliable estimates of BH masses are a prerequisite for investigating the relationship between BHs and their host galaxies. If the line-emitting clouds in the broad line region (BLR) are assumed to be virialized and moving in a potential dominated by the central BH, then the BH mass is simply a product of the BLR size and the square of the virial velocity. The reverberation-mapping technique uses the time lag between variations in the continuum emission and correlated variations in the broad line emission to measure the typical size of the BLR (Peterson, 1993; Peterson, 2014). The full width at half maximum (FWHM) or dispersion (σ; derived from the second moment) velocity

of the prominent broad emission line of Hβ (4862.7Å)[1] is used as an indicator of the virial velocity, with extensions to other low-ionization emission lines such as Hα (6564.6Å) and Mg II$\lambda\lambda$2796.4,2803.5 (e.g. Vestergaard, 2002; McLure and Jarvis, 2002; Wu et al., 2004; Kollmeier et al., 2006; Onken and Kollmeier, 2008; Wang et al., 2009; Rafiee and Hall, 2011). Extensive reverberation mapping campaigns have provided accurate BH masses for ∼50 active galactic nuclei (AGN) at relatively low redshifts and of modest luminosity (e.g. Kaspi et al., 2000; Kaspi et al., 2007; Peterson et al., 2004; Bentz et al., 2009; Denney et al., 2010). [See galaxies talk for a few more details]

1.2.2 *Single-Epoch Virial Estimates*

Single-epoch virial BH mass estimates normally take the form

$$M_{BH} = 10^a \left(\frac{\Delta V}{1000 \text{ km s}^{-1}} \right)^b \left[\frac{L_\lambda}{10^{44} \text{ erg s}^{-1}} \right]^c \tag{1.1}$$

where ΔV is a measure of the line width (from either the FWHM or dispersion), L_λ is the monochromatic continuum luminosity at wavelength λ, and a, b, and c are coefficients, determined via calibration against a sample of AGN with reverberation-mapping BH mass estimates. Several calibrations have been derived using different lines (e.g. Hβ, Mg II, C IV) and different measures of the line width (FWHM or dispersion) (e.g. Vestergaard, 2002; McLure and Jarvis, 2002; Vestergaard and Peterson, 2006; McGill et al., 2008; Wang et al., 2009; Rafiee and Hall, 2011; Park et al., 2013).

Reverberation mapping campaigns have also revealed a tight relationship between the radius of the BLR and the quasar optical (or ultraviolet) luminosity (the $R - L$ relation; e.g. Kaspi et al., 2000; Kaspi et al., 2007). This relation provides a much less expensive method of measuring the BLR radius, and large-scale studies of AGN and quasar demographics have thus become possible through the calibration of single-epoch virial-mass estimators using the reverberation-derived BH masses (e.g. Greene and Ho, 2005b; Vestergaard and Peterson, 2006; Vestergaard and Osmer, 2009; Shen et al., 2011; Shen and Liu, 2012; Trakhtenbrot and Netzer, 2012). The uncertainties in reverberation mapped BH masses are estimated to be ∼ 0.4 dex (e.g. Peterson, 2010), and the uncertainties in virial masses are similar (e.g. Vestergaard and Peterson, 2006). Since the structure and geometry of the BLR is unknown, a virial coefficient f is introduced to transform the observed line-of-sight velocity inferred from the line width in to a virial velocity. This simplification accounts for a significant part of the uncertainty in virial BH masses (in addition to, for example, describing the BLR with a single radius R and scatter in the $R - L$ relation;

---

1 Vacuum wavelengths are employed throughout the thesis.
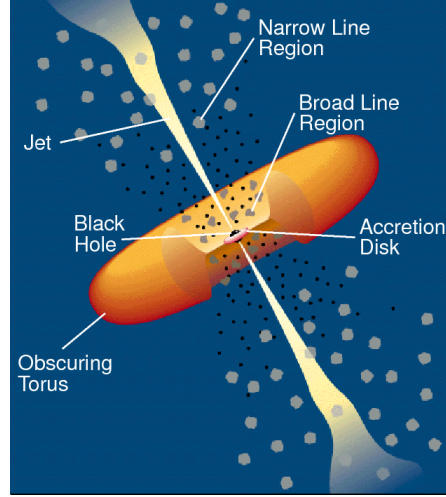
Figure 1.1: Illustration of the physical structure of an AGN in a simple orientation-based unification model. From Urry and Padovani, (1995).

Shen, 2013). Furthermore, if the BLR is anisotropic (for example, in a flattened disk; e.g. Jarvis and McLure, 2006) then the line width will be orientation-dependent (e.g. Runnoe et al., 2013b; Shen and Ho, 2014; Brotherton et al., 2015).

For example, single epoch estimates have been used to calculate black hole masses in the highest redshift quasars to study the growth of SMBHs. This figure shows a compilation of SE mass estimates for quasars over a wide redshift range from different studies. These studies show that massive, 10^9 BHs are probably already in place by $z \sim 7$, when the age of the Universe is less than 1 Gyr. This places strong constraints on BH growth models. Single epoch masses have also been used to study the distribution of quasars in the BH mass-luminosity plane, which conveys important information about the accretion process of these active black holes (e.g. Kollmeier et al. 2016). Reshift evolution of BH-bulge scaling relations (e.g. Bennert et al. 2011). Clustering (Shen & Ho 2014; Timins et al.?).

An *active galactic nucleus*, or AGN, is an energetic, non-stellar phenomena in the central region of a galaxy. AGN are powered by the accretion of gas, primarily through an accretion disk, onto a central *super-massive black hole* (SMBH) of mass 10^{6-9} M$_\odot$ (Lynden-Bell, 1969). With bolometric luminosities in the range 10^{44-48} ergs^{-1}, they are the most luminous persistent sources of radiation in the Universe.

1.3   ORIENTATION-BASED UNIFICATION MODELS

AGNs are divided into numerous classes and sub-classes based on their observational properties. AGN unification models (Antonucci, 1993; Urry and Padovani, 1995) attempt to explain the diversity in

their observational properties using as few physical parameters as possible. In many unification models the optical and radio luminosities are considered to be intrinsic parameters. Variations in the radio luminosity explains the difference between the $15 - 20\%$ of AGN which are *radio-loud* (i.e. have radio to optical flux ratios $\gtrsim$ 10) and the remainder which are *radio-quiet*. The optical luminosity explains, for example, the difference between low-luminosity *Seyfert Galaxies* and high-luminosity *quasars*[2]. In a typical quasar the optical emission may be brighter than the combined emission from all of the stars in the host galaxy by a factor of 100 or more. The large brightness contrast between the nucleus and the host galaxy makes the host galaxy difficult to detect, and the quasar is observed as an unresolved stellar-like object[3].

Unification models attempt to explain all further observational differences as being apparent differences due to *orientation* effects. The basic physical structure of an AGN in this model is illustrated in Figure 1.1. Material is pulled towards the SMBH at the centre and sheds angular momentum through viscous and turbulent processes in an accretion disk, which radiates primarily at ultraviolet (UV) to soft-X-ray wavelengths. Strong optical and UV emission lines are produced in photo-ionised gas clouds moving rapidly in close proximity to the SMBH. The doppler-broadened emission line widths imply gas cloud velocities of thousands of km s^{-1} in this *broad-emission line*. Further out are dusty, molecular clouds, the geometry of which is often modelled as a torus co-planar with the accretion disk. Along some lines of sight from the observer to the accretion disk / broad line region the dusty torus obscures the UV/optical radiation. In this case, an observer would see a weak UV/optical continua and no broad emission lines and classify the AGN as being *Type II*. On the other hand, if the line of sight is unobscured by the dusty torus then a broad emission line component would be observable in the spectrum and the AGN would be classified as being *Type I*. Further away from the central black hole and beyond the dusty torus are slower moving clouds of gas which are photo-ionised by the continuum emission from the accretion disk and produce forbidden emission lines of narrower widths (typically hundreds of km s^{-1}). Outflows of energetic particles occur along the poles of the accretion disk and form collimated radio-emitting jets and in some cases giant radio-emitting lobes. A strong, relativistically beamed component with large variations in brightness on very short timescales (e.g. $\Delta m \gtrsim 0.1$ and $\Delta t \lesssim 1$

---

2 The term 'quasar' is sometimes reserved for radio-loud objects and 'quasi-stellar object', or 'QSO', for radio-quiet objects. Here, 'quasar' is used to refer to all luminous AGN.

3 The name 'quasar' is shortened from 'quasi-stellar radio source', since quasars were originally discovered as optical point-source counterparts to a newly discovered population of radio sources.

day) is observed and a source with these properties is classified as a *blazar*.

While an orientation-based unification scheme such as this is somewhat successful at explaining many of the observational properties of AGNs, other factors such as the host galaxy morphology and gas/dust content may also be important (Peterson, 1997). It is also doubtful whether the geometry of the dusty torus is the same in all AGNs, and the fraction of obscured quasars has been shown to decrease with increasing nuclear luminosity (Lawrence, 1991). As we will now discuss, quasars might play an important role in a broader cosmological context, affecting the formation and evolution of the galaxies, groups, and clusters in which they reside. In this scenario of galaxy/quasar co-evolution the quasar is expected to transition from a highly active obscured phase to an unobscured phase as it clears out the dust surrounding it. If this picture is true then we should expect to find variations in the observational properties of the quasar and host galaxy as the system transitions through the different stages of its evolution.

1.4   THE TORUS

Elitzur & Shlosman (2006):

Recent high-resolution IR observations indicate that the torus size might be no more than a few parsecs (Elitzur 2005 and references therein); in particular, VLTI observations of NGC 1068 show that the FWHM size of the 12 mm emission is only 4 pc (Jaffe et al. 2004). The compact sizes place the torus inside the region where the SBH gravity dominates over the galactic bulge.

Two approaches have been taken for the torus dynamic origin. A hydrostatic scenario depicts the torus as a doughnut-like structure populated by molecular clouds accreted from the galaxy (Krolik & Begelman 1988). However, the origin of vertical mo tions capable of sustaining the clouds in a hydrostatic structure with H~R was recognized from the start as problematic and has eluded solution thus far (e.g., Davies et al. 2006). The other scenario, based on the seminal work by Blandford & Payne (1982), involves the outflow of clouds embedded in a hydromagnetic disk wind (Emmering et al. 1992, hereafter EBS92; Konigl & Kartje 1994; Kartje & Konigl 1996; Bottorff et al. 1997, 2000; Kartje at al. 1999; Everett 2005). In this approach the torus is merely a region in the wind that happens to provide the required toroidal obscuration; i.e., it is that region wherein the clouds are dusty and optically thick.

1.5   EVOLUTIONARY MODELS

A number of observations link the growth and evolution of quasars to the growth and evolution of galaxies. These include the following:

1. SMBHs appear to be a ubiquitous feature at the centres of all massive galaxies (e.g. Kormendy and Ho, 2013).

2. SMBH masses are proportional to the mass/velocity dispersion of their host spheroid (the $M - \sigma$ relation; Ferrarese and Merritt, 2000; Gebhardt et al., 2000).

3. The cosmological evolution of the star formation rate and the quasar luminosity function are very similar (e.g. Wall et al., 2005).

4. Cosmological simulations of galaxy formation and evolution require feedback from SMBH growth in order to reproduce the galaxy luminosity function (Kauffmann and Haehnelt, 2000).

These observations suggest that all galaxies may have gone through a 'quasar phase' during which the SMBH accretes most of its mass and the stellar-bulge forms most of its stars. This evolutionary phase could be triggered by a major merger or by instabilities in the galactic disc or bulge. In a galaxy merger large amounts of gas can shed sufficient angular momentum to settle into dense clouds and form stars or be funnelled to the centre of the galaxy to grow the existing SMBH. The large amounts of gas and dust funnelled inward to the galactic nucleus is predicted to obscure the quasar until the dust is cleared out either by quasar-driven or stellar-driven processes. An unobscured quasar then emerges, and is active until all of the available material has been accreted (Hopkins et al., 2006; Narayanan et al., 2010). The feedback processes involved are also thought to be responsible for shutting down star formation in the galactic bulge (Silk and Rees, 1998) and establishing the $M - \sigma$ relation.

Such scenarios have been invoked to explain the presence of buried AGN seen in ultra-luminous infra-red galaxies (ULIRGs; Sanders et al., 1988), a high fraction of which also show evidence of merging and interaction. However, the full picture is likely to be more complicated. Although there is evidence that mergers dominate at high luminosities (Treister et al., 2012), stochastic accretion may be more important at low luminosities (e.g. Hopkins and Hernquist, 2006).

Luminous unreddened quasars show few signs of interaction (e.g. Dunlop et al., 2003) which, if the quasar-galaxy co-evolution model is true, suggests that indications of an interaction disappear during a transitional phase. Quasars in this transitional phase would be highly reddened, as the dust enshrouding the nucleus will not have been fully cleared, but not completely obscured. A population of quasars with these properties may therefore represent a link between ULIRGs and unobscured quasars.

1.6   INTERESTING SUB-POPULATIONS

1.6.1   *Red and Reddened Quasars*

Magnitude limited optical surveys of quasars are biased against se-
lecting red and reddened quasars. Richards et al., (2003) studied a
large sample of optically selected Sloan Digital Sky Survey (SDSS;
York et al., 2000) quasars and showed the mean reddening to be
$E(B-V) = 0.03$ at the redshift of the quasar. They estimated that
~ 15% of the population was missing from the survey due to dust ex-
tinction. The missing fraction, and it's dependence on luminosity and
redshift, could help to determine whether the reddened population is
best explained in the context of orientation-based unification models
with non-spherical geometry or as an evolutionary stage in a quasars
lifetime.

Populations of heavily dust-reddened quasars have been identified
using radio surveys (e.g. Glikman et al., 2012), by using the 'K-band
excess' in the spectra of quasars relative to stars (Maddox et al., 2012),
and using near-IR colour selection (Banerji et al., 2012; Banerji et
al., 2013). Recently, Ross et al., (2014) identified a small sample of
very red SDSS quasars based on their extreme IR to optical lumi-
nosity ratios. It is yet to be determined whether these extreme ob-
jects are simply the tail of a population dominated by less reddened
quasars, or whether the distribution is bi-modal with reddening. A
population of quasars with intermediate amounts of dust reddening
$(0.1 \lesssim E(B-V) \lesssim 0.5)$ would help to address this question.

1.6.2   *Broad Absorption Line Quasars*

Broad absorption line quasars (BALQSOs) are a sub-population of
quasars exhibiting blue-shifted absorption troughs broader than
2000kms^{-1} (Weymann et al., 1991) which are unambiguously asso-
ciated with AGN-driven out-flowing gas. As well as showing high
rates of mergers, an anomalously large fraction of heavily reddened
objects exhibit broad blue-shifted absorption troughs in their spectra
(Urrutia et al., 2009; Glikman et al., 2012). This observation suggests
that the BAL phenomenon may be related to a 'blow-out' phase of
a quasars lifetime as it transitions from a dusty, obscured objected
to a luminous blue quasar, at the same time quenching star forma-
tion. Since outflows are believed to be fundamental to AGN feed-
back, a better understanding of their properties could shed light on
the outflow phenomenon. Alternatively, whether a quasar is observed
to have broad absorption lines could depend only on the orientation
of the observer in relation to an intrinsically anisotropic system.

1.6.3 *Hot-Dust-Poor Quasars*

The near-IR emission from AGN is generally explained by thermal emission from dust grains at the edge of the dusty torus closest to the accretion disk. The dust is heated to its sublimation temperature (1300-2000K Barvainis, 1992) by emission from the accretion disc. However, Hao et al., (2010) reported that 6% (at $z \lesssim 2$) to 20% (at $2 \lesssim z \lesssim 3.5$) of the quasars in the X-ray selected XMM-COSMOS Type 1 AGN sample (Brusa et al., 2010) have an unusually small amount of hot dust emission, despite having normal accretion disc spectra. They infer a torus covering factor of ~2% to 30% for these 'hot dust poor' (HDP) quasars, well below the ~75% predicted by unified models (e.g. Krolik and Begelman, 1988). Hao et al., (2011) found that HDP quasars were just as common in the Richards et al., (2006a) Spitzer/S-DSS sample (8.7% ± 2.2%) and the Elvis et al., (1994) Palomar-Green-quasar-dominated sample (9.5% ± 5.0%). Either the hot dust is destroyed (dynamically or by radiation), or the dust is not centred on the SMBH, which could happen during a major merger (e.g. Blecha et al., 2011). Alternatively, misaligned accretion disks, which will result from discrete isotropic accretion events (Volonteri, Sikora, and Lasota, 2007), will lead to a wider range of covering factors (Lawrence and Elvis, 2010).

At higher redshifts, Jiang et al., (2010) found two HDP quasars in a sample of 21 at $z \sim 6$. They find that at $z \sim 6$ the hot dust abundance is roughly proportional to the black hole mass, indicating that the two grow at about the same rate. The two HDP quasars also have the smallest SMBH masses, and may be too young to have formed a significant amount of hot dust.

1.6.4 *Type II Quasars*

As well as lacking a broad-line spectral component, Type II AGN tend to have high IR to optical light ratios, hard X-ray spectra, and be strongly polarised, consistent with dusty torus based unification schemes. The detection of unobscured continuum emission that is scattered and polarised by dust above the torus has confirmed the orientation-based unification of Type I and Type II Seyfert Galaxies. Their higher luminosity analogues, Type II quasars, have been much more difficult to detect and study. It is possible that the orientation-based Type I/II unification scheme may break down at high-luminosities, and that instead all quasars could pass through a Type II phase before the obscuring dust is cleared out by the quasar-driven outflows and a Type I quasar emerges.
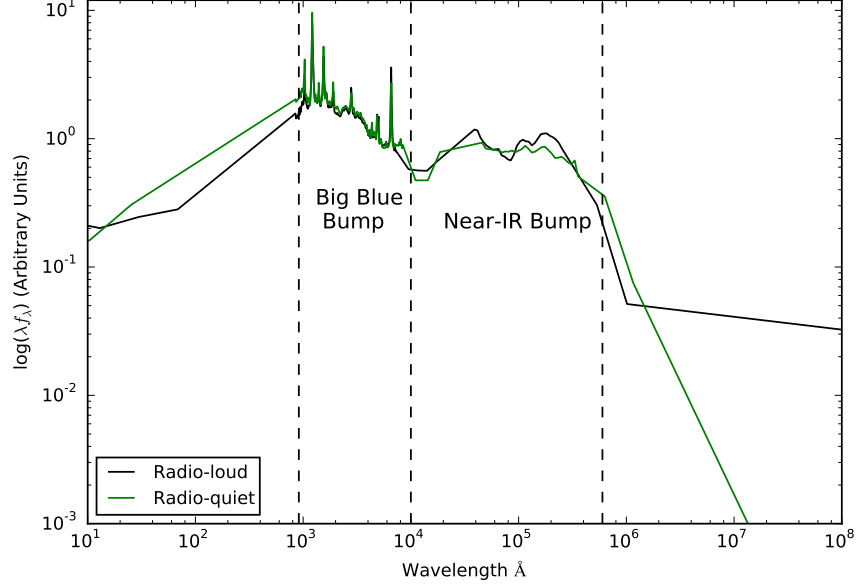
Figure 1.2: Median SEDs for radio-loud and radio-quiet quasars from Shang et al., (2011).

1.7 SPECTRAL ENERGY DISTRIBUTIONS

AGNs emit strongly over many decades in frequency of the electromagnetic spectrum and the energy emitted as a function of frequency is described by a *spectral energy distribution* (SED). As we will describe below, the broad features in the SED originate from processes which occur in different regions of the AGN. In the preceding sections, we have described how some interesting sub-populations of AGNs might relate to the broader population in the context of orientation-based unification schemes and evolutionary schemes. Comparing SEDs of different sub-populations can help to shed light on these relationships and the physical processes which drive them. Possible correlations between the SED shape and luminosity, redshift, and other properties of the AGN such as black hole mass and Eddington ratio can also constrain models of AGN structure and evolution.

Since the physical processes involved are generally understood only qualitatively, almost all AGN SED templates are empirical. The empirical template of Elvis et al., (1994), constructed using photometric observations from the radio to the hard X-rays of 29 radio-quiet and 18 radio-loud Type I quasars, is still the most commonly cited, despite many additions and updates (e.g. Polletta et al., 2000; Kuraszkiewicz et al., 2003; Risaliti and Elvis, 2004; Richards et al., 2006a; Polletta et al., 2007; Lusso et al., 2010; Shang et al., 2011; Marchese et al., 2012; Trichas et al., 2012). In Figure 1.2 we show the median spectrum from the radio-loud and radio-quiet samples of Shang et al.

(2011). Short-ward of the radio region, the radio-loud and radio-quiet spectra are almost indistinguishable.

A large amount of energy is emitted in the UV/optical region short-ward of ~ 4000Å: the *Big Blue Bump*. In the X-ray region, the *soft X-ray excess* may be the high-energy end of this feature. The Big Blue Bump is generally attributed to thermal emission from the accretion disk. In Type II AGNs, the continuum emission from the accretion disk is obscured, and so the Big Blue Bump in the SED of a Type II AGN would be less prominent than is seen in Figure 1.2.

The feature at wavelengths long-ward of $\sim 1\mu$m is the *IR Bump*, and is generally attributed to thermal emission from dust at a wide range of temperatures ($\sim 50 - 1000$ K). The amount, geometry, ionisation and optical depth of absorbing dust and gas and its inclination determines the shape of the IR Bump and the absorption of the optical/UV continua. The relative strengths of the IR Bump and the Big Blue Bump are generally comparable, although they do vary from object to object. In particular, for the HDP objects we described above the IR Bump appears to be missing entirely. The minimum between the two peaks is at $\sim 1\mu$m, which reflects the sublimation of dust at T $\gtrsim 2000$ K (Sanders et al., 1989).

Emission in the hard X-ray region of the spectrum is believed to be due to Compton up-scattering of accretion disk photons by hot electrons forming a corona in the vicinity of the disk (e.g. Sunyaev and Titarchuk, 1980). The radio emission, which originates from synchrotron emission in relativistic jets, contributes very little to the total energy output. However, the mechanical energy provided by the jets is an important component of AGN feedback models (e.g. Fabian, 2012).

Many parameters might be expected to affect the shape of the AGN SED (e.g. the black hole mass, the accretion rate, the physical properties of the accretion disk, the properties of the absorbing dust) and many of these properties might be expected to change as the quasar evolves (e.g. as dust is expelled from the nuclear regions). Given this, it is perhaps surprising that many authors have found no significant dependence of the mean SED on properties such as redshift, bolometric luminosity, SMBH mass, or accretion rate (e.g. Elvis et al., 2012; Hao et al., 2013) and that quasars up to redshift 7 have been shown to have similar UV spectra to low redshift quasars (e.g. Mortlock et al., 2011).

Throughout this thesis we adopt a ΛCDM cosmology with $h_0 = 0.71$, $\Omega_M = 0.27$, and $\Omega_\Lambda = 0.73$. All wavelengths and equivalent width measurements are given in the quasar rest-frame, and all emission line wavelengths are given as measured in vacuum.

Everett et al. (2005):

A variety of observational signatures point to the importance of outflowing gas within many types of active galactic nuclei (AGNs).

Blueshifted absorption features (in broad absorption line quasars, or BALQSOs; see, e.g., Weymann et al. 1991) are seen in approximately 15% (Reichard et al. 2003) of radio-quiet quasars, with velocities up to 0.1c. In addition, radio-loud quasars display relativistic, collimated outflows. There has also been observational evidence that suggests the mass outflow rate in AGNs is nearly equal to the mass inflow rate (see, e.g., Crenshaw et al. 2003; Chartas et al. 2003).

2

A NEAR-INFRARED SPECTROSCOPIC DATABASE OF HIGH-REDSHIFT QUASARS

2.1 INTRODUCTION

With the exception of a handful of very nearby objects, the inner regions of AGN cannot be resolved. Spectroscopic data is therefore invaluable to all AGN-related science. The optical region includes a number of strong emission features, including the broad lines of Hαλ6565 and Hβλ4863 and the narrow [O III]$\lambda\lambda$4960,5008 doublet. As we will see in Chapter 3, the low-ionisation Balmer lines are routinely used to derive black hole masses and quasar accretion rates. As the strongest narrow emission line, [O III] is used to measure the systemic redshift, and to probe quasar-driven outflows on galactic scales (see Chapter 4).

Large optical surveys have provided spectra for hundreds of thousands of AGN and quasars. With it's twelfth data release in 2016, the number of quasar spectra in the Sloan Digital Sky Survey (SDSS; York et al., 2000) catalogue alone reached almost 300,000. However, the rest-frame optical region is redshifted beyond the reach of optical spectrographs at redshifts $z \gtrsim 0.4$ and, at redshifts $z \sim 2$, near-infrared spectroscopy is required in order to access the rest-frame optical lines.

The number density of quasars in the Universe rises sharply as a function of redshift, and peaks at redshifts $2 \lesssim z \lesssim 4$. The star formation rate follows a similar evolutionary path. Therefore, understanding supermassive black hole accretion over cosmic time and quasar feedback critically depends on the availability of near-infrared spectra for high-redshift quasars. Spectroscopic observations are more challenging at infrared wavelengths than in the optical. The Earth's atmosphere is both bright and highly variable at infrared wavelengths. As a result, the number of high-redshift quasars with near-infrared spectra is limited. Previous investigations of the rest-frame optical spectra of quasars at redshifts $z \sim 2$ have typically used samples of a few dozen (e.g. Shen and Liu, 2012; Shen, 2016).

Other references? Sulentic?

In this chapter I will describe how I have constructed a database containing 462 high-redshift quasars. In later chapters, I will describe how I have used this data to derive un-biased virial black hole mass estimates for quasars at redshifts $z \gtrsim 2$ (Chapter 3) and to study quasar-drive galaxy-wide outflows (Chapter 4). The unprecedented size and quality of this dataset make a number of other exciting investigations possible, some of which are described in Section 6.

Figure 2.1: The ranges in redshift and luminosity covered by our sample, relative to the redshift-luminosity distribution of the SDSS DR7 quasar catalogue. In regions of high point-density, contours show equally-spaced lines of constant probability density generated using a Gaussian kernel-density estimator. For the SDSS sample we use Hewett and Wild, (2010) redshifts and bolometric luminosities measured by Shen et al., (2011). For the quasars in out sample the redshift is defined using the peak of the Hα/Hβ emission and the luminosity is measured in the continuum at 1350Å and converted to a bolometric quantity using the same conversion factor employed by Shen et al., (2011). Eight objects are missing because we do not have enough information to calculate the bolometric luminosity.

In Fig. 2.1 we show the luminosities and redshifts of the quasar sample relative to the redshift-luminosity distribution for the Seventh Data Release (DR7; Schneider et al., 2010) of the SDSS spectroscopic quasar catalogue. Our sample spans a redshift range $1.5 < z < 4.0$ and a bolometric luminosity range $10^{45.5} - 10^{48}$ erg s^{-1}. Spectra were obtained within one or more of the JHK pass-bands and the gaps in our sample coverage at $z \sim 1.8$ and $z \sim 3$ are due to the presence of atmospheric absorption. Obtaining near-infrared spectra of adequate resolution and signal-to-noise ratio (S/N) of even moderately bright quasars remains resource intensive. As a consequence, at fixed redshift, the luminosities of the quasars are brighter than the average luminosity of the SDSS sample, although the dynamic range in luminosity is a full 1.5 decades.

Table 2.1: Summary of near-infrared spectroscopic database.

Instrument	Number
FIRE/Magellan	36
GNIRS/Gemini	29
ISAAC/VLT	13
LIRIS/WHT	21
NIRI/Gemini	31
NIRSPEC/Keck	3
SINFONI/VLT	84
SofI/NTT	111
TRIPLESPEC/ARC	38
TRIPLESPEC/Hale	60
XSHOOTER/VLT	36
Total	462

2.2 DATA

The near-infrared spectra in our database are taken from published catalogues, by downloading and reducing archival spectra, and by reducing spectra acquired in programmes led by Prof. J. Hennawi (UCSB) and Prof. X. Prochaska (UCO/LICK). As the P.I. of two programmes, I filled in an under-sampled region of the C IV blueshift parameter space by targeting quasars with the most extreme C IV blueshifts. The telescopes and instruments used to observe these spectra are summarised in Table 2.1 and the information on the individual spectra is provided in Table 2.2. In the remaineder of this chapter I will describe each of these sub-samples in turn.

2.2.1 *Coatman et al. (2016) Quasars*

2.2.1.1 *Defining sample*

We selected quasars from the SDSS DR7 spectroscopic quasar catalogue. The sample was restricted to objects with redshifts $2.14 < z < 2.51$ (7,258 quasars), to ensure that the Hβ and Hα emission lines fall within the H- and K-bands respectively, allowing us to observe both simultaneously with the appropriate grism configuration. Given the limited number of quasars for which near-infrared spectra could be obtained, the quasar sample was further restricted to objects that are radio-quiet (5,980 quasars), show no evidence of broad absorption lines (BALs) in their spectra (5,299 quasars), and are free from significant dust extinction. We removed radio-loud objects from our sample using the same radio-loud classification as Shen et al., (2011), and BAL quasars using the classifications of both Shen et al., (2011) and

Allen et al., (2011). The removal of quasars with significant dust extinction was achieved by identifying quasars with $i - K$ colours redder than a parametric spectral energy distributions (SED) model + SMC-like extinction curve with $E(B - V)$=0.05 (the SED model is described in Chapter 5).

The K-magnitude was taken from the UKIRT Infrared Deep Sky Survey (UKIDSS; Lawrence et al., 2007) Large Area Survey (ULAS). The requirement to be in the ULAS footprint and have reliable K band photometry reduced our sample of possible targets to 1,683, and the $E(B - V)$ cut left 1,204 in our sample. Finally, a flux-limit of $K < 18.5$ (AB) was applied to ensure that spectra of sufficient signal-to-noise ratio (S/N) could be obtained (412 quasars).

We were able to obtain new infra-red spectra for 19 quasars from this sample of 412 possible targets. The quasars included in this subsample were selected to have C IV-emission shapes which span the full range observed in the population. Reliably quantifying the distribution of C IV-emission shapes has been made possible thanks to improvements in the estimation of systemic redshifts from ultraviolet spectra. The Allen & Hewett (2017, in preparation) redshift estimation algorithm generates redshifts which are independent of the C IV-emission shape. This has been a crucial factor in allowing us to quantify the distribution of C IV-blueshifts in the observed quasar population as a whole, and thus select a sample of quasars with a range of C IV blueshifts (see Section XX).

This paragraph could be more succinct

2.2.1.2 *Observations*

Near-infrared spectra were obtained with the Long-slit Intermediate Resolution Infrared Spectrograph (LIRIS; Manchado et al., 1998) mounted on the 4.2m William Herschel Telescope (WHT) at the Observatorio del Roque de los Muchachos (La Palma, Spain). Observations took place over four non-contiguous nights from 2015 March 31 to April 4. Approximately one night was lost due to poor weather and a further half-night was affected by poor transparency due to cloud. A one arcsecond slit-width was employed and the LIRIS $H + K$ low-resolution grism was selected, which covers the spectral ranges 1.53–1.79 μm and 2.07–2.44 μm with a dispersion of 9.7Å/pixel. The spatial scale of the instrument is 0.25 arcsec/pixel. Observations were divided into 60 s sub-exposures and performed in an ABBA nodding pattern, with the object placed at two positions along the slit 12 arcsec apart. Bright A0-5V stars were observed at similar air-masses to the targets in order to provide both telluric absorption corrections and a flux calibration of the quasar spectra.

2.2.1.3   *Data reduction*

The raw LIRIS data frames incorporate a known 'pixel shift' which was first removed from all frames using the LIRIS data reduction package `LIRISDR`. Subsequent data reduction was undertaken with standard `IRAF`[1] procedures. The flat-field images, which were taken at the beginning of each night via illumination of the dome, were averaged and normalised to remove any wavelength-dependent signature. Each individual two-dimensional spectrum was then flat-field corrected. Consecutive AB and BA pairs of two-dimensional spectra were subtracted to remove the sky background. All the subtracted AB/BA-pairs for a target were then averaged to give the final two-dimensional spectrum.

The size of the one-dimensional spectrum extraction windows, in the slit direction, varied from 6-10 pixels. To increase the S/N, optimal variance-weighted extraction with sigma clipping was employed. For the fainter objects in our sample we were unable to trace the spectrum across the dispersion axis reliably and the trace from a telluric standard-star observation, observed at a similar air mass and time, was used instead. The wavelength calibration, using argon and xenon lamp exposures, resulted in root mean square errors in the range 1.01–1.71 Å, with a mean of 1.47Å. The telluric standard star observations were reduced using the same steps described above. The stellar continuum was divided out of the standard star spectrum, which was then divided into the quasar spectrum to remove telluric absorption features. The spectral type and magnitude of the standard star were used to flux calibrate the quasar spectrum both in a relative and absolute sense. Variable atmospheric conditions combined with the narrow slit width resulted in a significant level of uncertainty in the absolute flux calibration for the quasar observations. The use of the UKIDSS broadband magnitudes (H and K) to normalise the spectra results in a significantly improved calibration. [2]

2.2.2   *Shen & Liu (2012) and Shen (2016) Quasars*

Shen, (2016) and Shen and Liu, (2012) obtained near-infrared spectroscopy for a sample of 74 luminous, $1.5 < z < 3.5$ quasars selected from the SDSS DR7 quasar catalogue. Targets had to possess good optical spectra covering the C IV line and have redshifts $z \sim 1.5$, 2.1, and 3.3 to ensure that the Hβ-[O III] region was covered in one of the near-infrared JHK bands. Thirty-eight of the quasars were observed with TripleSpec (Wilson et al., 2004) on the Astrophysics Research Consortium (ARC) 3.5 m telescope, and 36 with the Folded-port InfraRed

---

[1] IRAF is distributed by the National Optical Astronomy Observatory, which is operated by the Association of Universities for Research in Astronomy (AURA) under a cooperative agreement with the National Science Foundation.

[2] The data reduction pipeline is available at github.com/liamcoatman/SpectraTools

Echellette (FIRE; Simcoe et al., 2010) on the 6.5 m Magellan-Baade telescope. The reduction of the spectra is described in Shen, (2016) and Shen and Liu, (2012).

2.2.3 *Quasar Pairs*

A large part of our catalogue was observed as part of an ongoing effort to identify quasar pairs at very close projected separations (Quasars Probing Quasars[3] (QPQ); Hennawi et al., 2006a; Hennawi et al., 2010). The primary science driver of this work is to study the circum-galactic medium of the foreground quasars in absorption (Hennawi et al., 2006b). Very accurate systemic redshift measurements are a requirement and a large amount of effort has gone into obtaining near-infrared spectra which cover low-ionisation broad lines or features from the quasar narrow line region (Prochaska and Hennawi, 2009; Lau, Prochaska, and Hennawi, 2015; Hennawi et al., 2015). Twenty-nine quasars were observed with the Gemini Near-Infrared Spectrograph (GNIRS; Elias et al., 2006) on the 8.1 m Gemini North telescope, thirtenn using the Infrared Spectrometer And Array Camera (ISAAC; Moorwood et al., 1998) on the European Southern Observatory (ESO) Very Large Telescope (VLT), thirty-one with the Near InfraRed Imager and Spectrometer (NIRI; Hodapp et al., 2003) also on Gemini North and thirty-six with XSHOOTER (Vernet et al., 2011), again, on the VLT.

The XSHOOTER spectra were reduced with a custom software package developed by George Becker (for details, see Lau, Prochaska, and Hennawi, 2015). The remaining data was processed with algorithms in the LowRedux[4] package (see Prochaska and Hennawi, 2009).

2.2.4 *VLT SINFONI Quasars*

We performed a search of the ESO archive for high-redshift quasars observed with the SINFONI integral field spectrograph (Eisenhauer et al., 2003; Bonnet et al., 2004) at VLT/UT4. We found 79 quasars with redshifts $1.5 < z < 3.7$ which have H and/or K SINFONI spectroscopy, covering the Hβ and Hα lines respectively. Seventy-two of the quasars are from a large programme led by L. Wisotzki (programme 083.B-0456(A)) to study the mass function and Eddington ratios of active BHs at redshifts $z \sim 2$ drawn from the Hamburg/ESO survey (Wisotzki et al., 2000). A further seven SINFONI spectra are from a programme led by J. D. Kurk (programme 090.B-0674(B)) to obtain reliable BH mass estimates from Hα/Hβ for a sample of radio-loud/radio-quiet SDSS quasars.

---

3 www.ucolick.org/~xavier/QPQ/Quasars_Probing_Quasars
4 www.ucolick.org/~xavier/LowRedux

The SINFONI spectra were reduced using the package EASYSINF[5]. The package, which is based on the ESO-SINFONI pipeline, is described in Williams et al., (2016).

2.2.5 *ESO NTT SOFI Quasars*

One quarter of the quasar catalogue derives from a large programme (programme 187.A-0645; PI: J. Hennawi) to combine near-infrared spectra from SOFI (Moorwood, Cuby, and Lidman, 1998) on the 3.6 m New Technology Telescope (NTT) with archival high-resolution optical spectra from the UV-Visual Echelle Spectrograph (UVES; Dekker et al., 2000) at VLT/UT2 and the High Resolution Echelle Spectrometer (HIRES; Vogt et al., 1994) at Keck to construct a legacy database of bright, high-redshift ($2 < z < 4$) quasars with both rest-frame optical spectra, covering the Hβ-[O III] complex, and high-resolution restframe ultraviolet spectra. The main science goal is to obtain precise systemic redshifts which are crucial for the study of absorption line systems. Observations were undertaken over 16 nights from September 2011 to March 2013. I reduced these spectra using a custom pipeline using algorithms in the LowRedux package and, in Coatman et al., (2017), published a subset of the data for the first time.

Over five nights from 2015 August 31 to September 4 we obtained near-infrared SOFI spectra for a further 26 quasars (programme 095.B-0644(A); PI: L. Coatman). These quasars were selected from the SDSS DR7 quasar catalogue using criteria very similar to those described above for the WHT sample. In particular, we selected quasars with large C IV blueshifts to improve the statistics in this region of the C IV emission-line parameter space. The spectra were reduced using the same LowRedux pipeline.

Expand section?

2.2.6 *Hale TripleSpec Quasars*

A further sixty quasars in our catalogue are bright SDSS quasars which were observed with the TRIPLESPEC spectrograph on the Palomar 200-inch Hale telescope (P200). The objects were observed with the same science goals as the SOFI NTT large programme. The spectra were reduced using a custom pipeline, again using algorithms in the LowRedux package.

---

5 www.mrao.cam.ac.uk/~rw480/easysinf

Table 2.2: Quasars in our near-infrared spectroscopic database. Only the first 15 entries are shown. The full table (including 462 objects) is available online. Columns are as follows: (1) identifier, (2) date near-infrared spectra acquired, (3)-(4) coordinates, (5) instrument/telescope, (6) wavelength coverage, (7) velocity per pixel, (8) S/N per pixel, (9)-(11) redshifts from peak of [O III], Hα, and Hβ - see Chapter 4 for details.

ID	Date	Ra	Dec	Instr.	$\Delta\lambda$ [μm]	Δv [km s^{-1}]	S/N	z([O III])	z(Hα)	z(Hβ)
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
J000039-001804	2015-09-02	+00h00m39.00s	-00d18m03.90s	SofI/NTT	1.50-2.54	154.0	4.9		2.1412	2.1391
J000345-232353	2009-07-07	+00h03m45.00s	-23d23m53.40s	SINFONI/VLT	1.44-1.87	36.0	12.7	2.2657		2.2653
J000345-232353	2011-09-18	+00h03m45.00s	-23d23m53.40s	SofI/NTT	1.48-1.83	63.0	36.0	2.2644		2.2776
J000451-084450	2013-07-12	+00h04m50.66s	-08d44m49.63s	XSHOOTER/VLT	0.31-2.28	15.0	10.3	3.0038		3.0052
J000451-084452	2013-08-08	+00h04m50.91s	-08d44m51.98s	XSHOOTER/VLT	0.31-2.28	15.0	5.4	2.9991		
J000500-003348	2015-09-01	+00h05m00.42s	-00d33m48.20s	SofI/NTT	1.50-2.54	154.0	8.2		2.1842	2.1850
J000501+010221	2015-09-02	+00h05m00.53s	+01d02m20.80s	SofI/NTT	1.50-2.54	154.0	6.8		2.1334	2.1317
J001016+001228	2015-09-04	+00h10m16.49s	+00d12m27.60s	SofI/NTT	1.50-2.54	154.0	8.9		2.2855	2.2828
J001247+001239	2013-06-06	+00h12m47.12s	+00d12m39.49s	ISAAC/VLT	1.52-1.60	15.0	19.1			2.1618
J001708+813508	2012-08-04	+00h17m08.48s	+81d35m08.10s	TRIPLESPEC/Hale	0.94-2.80	39.0	36.5	3.3934		
J001919+010152	2015-09-04	+00h19m19.31s	+01d01m52.20s	SofI/NTT	1.50-2.54	154.0	6.5	2.3120	2.3158	2.3154
J001955-091316	2004-11-26	+00h19m54.67s	-09d13m16.45s	GNIRS/Gemini	0.60-2.61	88.0	9.9		2.1207	2.1308
J002018-233654	2009-07-07	+00h20m18.41s	-23d36m53.80s	SINFONI/VLT	1.44-1.87	36.0	16.9	2.2975		2.2931
J002023-414639	2009-07-08	+00h20m23.38s	-41d46m38.90s	SINFONI/VLT	1.09-1.41	35.0	33.4	1.5733		1.5730
J002111-242247	2009-07-16	+00h21m10.90s	-24d22m47.20s	SINFONI/VLT	1.44-1.86	36.0	11.1	2.2622		2.2595

BLACK HOLE MASSES

3.1 INTRODUCTION

The goal of better understanding the origin of the correlation between the masses of super-massive black holes (BHs) and the masses of host-galaxy spheroids has led to much work focussing on the properties of quasars and active galactic nuclei (AGN) at relatively high redshifts, $z \gtrsim 2$. Extensive reverberation-mapping campaigns have been used to calibrate single-epoch virial-mass estimates which use the velocity widths of the hydrogen Balmer emission lines and the nuclear continuum luminosity to provide reliable BH masses. Single-epoch virial BH mass estimates using Hβ are possible up to redshifts $z \sim 0.7$, and the technique has been extended to redshifts $z \sim 1.9$ via the calibration of the broad Mg $\textsc{ii}\lambda\lambda$2796,2803 emission line (McLure and Jarvis, 2002; Onken and Kollmeier, 2008; Wang et al., 2009; Rafiee and Hall, 2011). At redshifts $z \gtrsim 2$, however, ground-based statistical studies of the quasar population generally have no access to the rest-frame optical and near-ultraviolet spectral regions.

The C $\textsc{iv}\lambda\lambda$1548,1550 emission doublet is both relatively strong in the majority of quasars and visible in modern optical spectra, such as those provided by the Sloan Digital Sky Survey (SDSS), to redshifts exceeding $z \sim 5$. C $\textsc{iv}$-derived BH masses have therefore become the standard (e.g. Vestergaard and Peterson, 2006; Park et al., 2013) for both individual quasars and in studies of quasar population demographics.

Currently, the number of reverberation mapped quasars is small (~50 quasars; Park et al., 2013) and restricted to low redshifts and luminosities. The luminosities of quasars at redshifts $z \gtrsim 2$ are much greater than in the reverberation mapped sample, and the reliability of the existing calibration involving C $\textsc{iv}$ FWHM velocity measurements and ultraviolet luminosity is not established definitively when extrapolating to high-redshifts and luminosities. While some authors have found good agreement between BH mass-estimates based on C $\textsc{iv}$ and Hβ (e.g. Vestergaard and Peterson, 2006; Assef et al., 2011; Tilton and Shull, 2013), others have questioned the consistency (e.g. Baskin and Laor, 2005a; Trakhtenbrot and Netzer, 2012; Shen and Liu, 2012).

In contrast to a number of low-ionisation emission lines, such as Mg $\textsc{ii}$, the C $\textsc{iv}$ emission has long been known to exhibit significant asymmetric structure, with an excess of flux to the blue of the predicted rest-frame transition wavelength (Gaskell, 1982). More recent
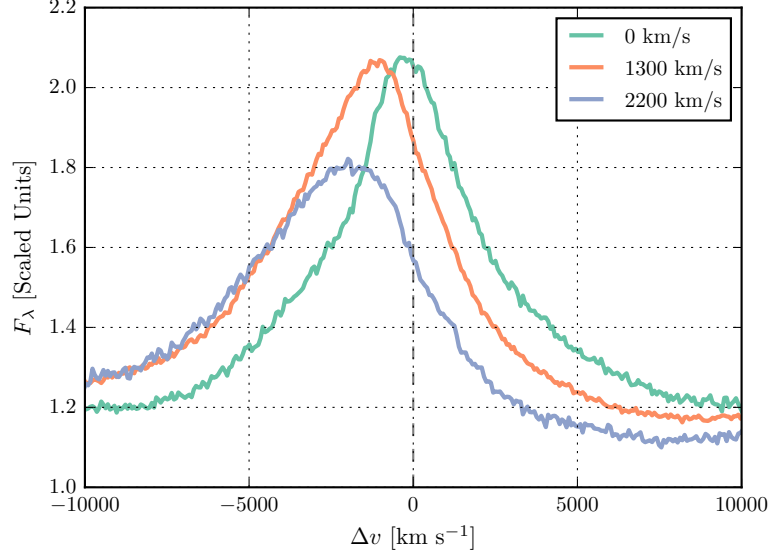
Figure 3.1: Composite spectra of the C ɪᴠ-emission line as a function of C ɪᴠ blueshift for SDSS DR7 quasars. Quasars classified as BALs, or possessing strong associated absorbers have been excluded, and the composite-spectra shown are derived using an arithmetic mean of a minimum of 200 spectra at each blueshift. Virtually the entire C ɪᴠ-profile appears to shift blueward and the change in line shape is not simply an enhancement of flux in the blue wing of a still identifiable symmetric component. In order of increasing C ɪᴠ blueshift, the composite spectra have FWHM 4870, 5610, and 6770 km s^{-1} and EW 33.1, 31.6, and 28.8 Å.

work (e.g. Sulentic, Marziani, and Dultzin-Hacyan, 2000; Richards et al., 2011) has established that the extent of 'blueshifts' in the C ɪᴠ emission correlates with a number of properties of quasar spectral energy distributions (SEDs). A fundamental assumption on which single-epoch virial BH-mass estimates are based is that the widths of the broad emission lines are directly related to the virial motions of the emitting clouds moving in the gravitational potential of the central BH. While the physical origin of the blueshifted emission has not been established there is a consensus that the associated gas is not tracing virial-induced velocities. A favoured interpretation associates the blueshifted emission with out-flowing material (see Netzer, 2015, for a recent review), reaching velocities significantly larger than virial-induced velocities associated with the BH (e.g. Sulentic et al., 2007; Richards et al., 2011). These outflows, most likely, result from the presence of a radiation line-driven accretion-disc wind (e.g. Konigl and Kartje, 1994; Murray et al., 1995; Proga, Stone, and Kallman, 2000; Everett, 2005; Gallagher et al., 2015; Higginbottom and Proga, 2015).

Excess emission-line flux in the blue wing of the C ɪᴠ emission increases commonly employed measures of the line-width, notably the

full-width at half maximum (FWHM) and the line dispersion (σ). In general, researchers studying quasar demographics at high-redshift adopt estimates of BH masses based on the width of C IV-emission, without reference to the blueshift of the C IV-emission (e.g. Vestergaard, 2004; Kollmeier et al., 2006; Gavignaud et al., 2008; Vestergaard et al., 2008; Vestergaard and Osmer, 2009; Kelly et al., 2010; Kelly and Shen, 2013). Figure 3.1 shows the shape of the C IV-emission in composite spectra constructed from SDSS DR7 quasars as a function of C IV blueshift. The profiles show how, at large values of blueshift ($\gtrsim 2000 \mathrm{km\,s}^{-1}$) the C IV-profile is displaced to the blue by amounts comparable to the FWHM of the profile. This indicates that non-virial motions, very likely due to outflows, are having a significant effect on the observed C IV emission velocity profile (e.g. Gaskell, 1982; Baskin and Laor, 2005a; Sulentic et al., 2007; Richards et al., 2011; Wang et al., 2013). At fixed emission-line EW, virtually the entire C IV-profile appears to shift blueward and the change in line shape is not simply an enhancement of flux in the blue wing of a still identifiable symmetric component. While gravity almost certainly plays a key role, determining the escape velocity for out-flowing material for example, it is clear that the virial assumption, on which single-epoch BH-mass measurements are predicated, is not straightforwardly applicable for the C IV-emission line in quasars exhibiting large blueshifts. As a consequence, BH-masses derived from C IV emission line velocity-widths are systematically biased compared to masses from the Balmer lines (e.g. Shen et al., 2008; Shen and Liu, 2012; Coatman et al., 2016).

As highlighted by Richards et al., (2011), the sample of reverberation mapped quasars includes a restricted range of the C IV emission line shapes seen in the quasar population. In particular, the reverberation mapped objects generally possess high C IV equivalent widths and low C IV-blueshifts. Nevertheless, the derived scaling relations based on the reverberation-mapped sample are regularly applied to the quasar population with low C IV EWs and/or large C IV-blueshifts, where any non-virial outflow-related contribution to the dynamics is significant.

In recent literature, attempts have been made to minimise the influence of the systematic non-virial contribution to the C IV emission on estimates of the BH mass. Strategies include (i) significantly reducing the dependence of the derived masses on the emission-line velocity width (e.g. from the V^2 dependence predicted assuming a virialized broad line region to just $V^{0.56}$ in Park et al. 2013; see also Shen and Liu 2012), (ii) adopting a measure of emission-line velocity-width that is relatively insensitive to changes in the core of the emission-line profile (e.g. Denney et al., 2013) and (iii) estimating the amplitude of the non-virial contribution to the C IV emission-line via comparison with other ultraviolet emission lines (e.g. Si IV+O IVλ1400 in Runnoe et al. 2013a and Brotherton et al. 2015). The increased number of quasars

with high-quality spectra that cover both the observed-frame optical (where the redshifted C IV appears) and near-infrared (where Hβ and Hα lie) enables us to take a rather different approach in this chapter. We will use properties of the C IV emission line itself to reduce, or even remove, the systematic bias in the BH-mass estimates. Specifically, using the low-ionisation Balmer lines Hα and Hβ as reliable proxies for the virial velocity, we will measure empirically the systematic bias in C IV-based virial BH mass estimates as a function of the C IV emission-line blueshift.

3.2   QUASAR SAMPLE

We have compiled a sample of 307 quasars at redshifts $1.5 < z < 4$ with both optical and near-infrared spectra. Reliable emission line properties were measured for 230 quasars (Section 3.3.5), with 164 possessing Hα line measurements and 144 Hβ line measurements. This will allow us to directly compare virial BH mass estimates based on the C IV line-width with estimates based on the line-widths of the low-ionisation Balmer lines Hα and Hβ. The sample is considerably larger than previous studies of the rest-frame optical spectra of high-z quasars (e.g. Shen and Liu, 2012). As we demonstrate in Section 3.5.3, the quasars have C IV blueshifts of up to ~5000km s^{-1}, and span the full range observed in the population.

3.2.1   *Near-infrared data*

The near-infrared data has been described in Chapter 2 and the telescopes/spectrographs used are summarised in Table 3.1. We have sub-divided our sample into two overlapping groups: quasars with reliable Hα line measurements (the 'Hα sample') and quasars with reliable Hβ measurements (the 'Hβ sample').

3.2.2   *Optical data*

Optical SDSS DR7 spectra are employed for 70 quasars in the full catalogue. The SDSS DR7 spectra are moderate resolution (R≃ 2000) and S/N (S/N≃ 20) and cover the observed-frame wavelength interval ~ $3800 - 9180$ Å. Many of the quasars in the SDSS DR7 catalogue have been re-observed as part of the Sloan Digital Sky Survey-III: Baryon Oscillation Spectroscopic Survey (SDSS-III/BOSS; Dawson et al., 2013). As the BOSS-spectra typically have higher S/N than the SDSS DR7 spectra, we have used the BOSS spectra when available (126 quasars). We also use high-resolution optical spectra taken with VLT/UVES (11 quasars) and VLT/XSHOOTER (8 quasars), and Hamburg/ESO spectra for a further 15 quasars. The reduced and fluxed UVES spectra were made available to us by A. Dall'Aglio (a descrip-

Table 3.1: The numbers of quasars with reliable Hα and Hβ line measurements, and the spectrographs and telescopes used to obtain the near-infrared spectra

Spectrograph	Telescope	Hα Sample	Hβ Sample
FIRE	MAGELLAN	18	19
GNIRS	GEMINI-N	22	17
ISAAC	VLT	0	4
LIRIS	WHT	15	0
NIRI	GEMINI-N	0	12
SINFONI	VLT	2	25
SOFI	NTT	47	23
TRIPLESPEC	ARC-3.5m	33	20
TRIPLESPEC	P200	23	19
XSHOOTER	VLT	4	7
Total		164	144

tion of the reduction procedure is contained in Dall'Aglio, Wisotzki, and Worseck, (2008)). The spectral resolution of the UVES observations is very high (R~40 000) and the S/N of the spectra re-binned to a resolution of $\simeq 2000$ is S/N$\simeq$ 300. The Hamburg/ESO optical spectra have a typical ~ 400km s^{-1} spectral resolution and S/N $\gtrsim$ 10 per pixel. The XSHOOTER spectra are moderate resolution (~ 6000) and conver the full optical-near-infrared spectral region (0.30 − 2.50µm).

3.3 SPECTRAL MEASUREMENTS

Conventionally, single-epoch virial estimates of the BH mass are a function of the line-of-sight velocity width of a broad emission line and the quasar luminosity. The velocity width is a proxy for the virial velocity in the broad line region (BLR) and, as revealed in reverberation-mapping studies, the luminosity is a proxy for the typical size of the BLR (the $R − L$ relation; e.g. Kaspi et al., 2000; Kaspi et al., 2007). Most reverberation mapping campaigns have employed Hβ time-lags and velocity widths, but the line-widths of Hα and Mg II λ2800 have been shown to yield consistent BH masses (e.g. McLure and Jarvis, 2002; Greene and Ho, 2005b; Onken and Kollmeier, 2008; Shen et al., 2008; Wang et al., 2009; Rafiee and Hall, 2011; Mejía-Restrepo et al., 2016). In Section 3.4.1 we verify that the Hα and Hβ line-widths yield consistent BH for the 99 quasars in our sample with measurements of both.

In our work, a robust measure of the C IV emission-line 'blueshift' provides the basis for the corrected C IV velocity-width measurements, and hence BH masses. The effectiveness of the scheme is validated via a direct comparison of the C IV velocity-widths to the

Balmer emission velocity-widths in the same quasars. Our process is as follows. First, an accurate measure of the quasar's systemic redshift is required, for which we adopt the centre of the Balmer emission, where the centre, λ_{half}, is the wavelength that bisects the cumulative total flux. Balmer emission centroids are available for all quasars in the catalogue but we verify that the measure is relatively unbiased through a comparison of the centroids to the wavelengths of the peak of the narrow [O III]$\lambda\lambda$4960,5008 doublet for the subset of spectra where both are available (Section 3.4.2). Second, the blueshift of the C IV emission line is determined. Again, we adopt the line centroid to provide a robust measure of the C IV emission blueshift. The blueshift (in $km\,s^{-1}$) is defined as $c\times(1549.48-\lambda_{half})/1549.48$ where c is the velocity of light and 1549.48 Å is the rest-frame wavelength for the C IV doublet[1]. Positive blueshift values indicate an excess of emitting material moving towards the observer and hence out-flowing from the quasar.

Emission-line velocity widths are derived from the full-width-at-half-maximum (FWHM) of the lines but we also compute the line dispersion (calculated from the flux-weighted second moment of the velocity distribution) as some authors have claimed this provides a better estimate of the virial velocity (Denney et al., 2013).

To minimise the impact of the finite S/N of the quasar spectra and the presence of absorption features superposed on the broad emission lines we first fit a parametric model to the continuum and the emission lines. The particular form of the model parametrizations is not important and the fits are used only to provide robust line parameters, such as the centroid λ_{half}, and FWHM, which are measured non-parameterically from the best-fitting model. The models used and the fitting procedure are described below. The issues involved in deriving parameters for broad emission lines from spectra of modest S/N – for example, subtraction of narrow line emission, subtraction of Fe II emission – have been covered comprehensively by other authors (e.g. Shen et al., 2011; Shen and Liu, 2012; Denney et al., 2013; Shen, 2016) and, as far as possible, we follow standard procedures described in the literature.

3.3.1   C IV

We first define a power-law continuum, $f(\lambda) \propto \lambda^{-\alpha}$, with the slope, α, determined using the median values of the flux in two continuum windows at 1445-1465 and 1700-1705Å. The continuum emission is subtracted from the spectra, which is then transformed from wavelength units into units of velocity relative to the rest-frame line-

---

1 The adopted C IV rest-frame wavelength assumes an optically thick BLR, in which case the contribution from each component is equal. Adopting a 2:1 ratio (appropriate for an optically thin BLR) changes the blueshifts by $\sim$80$km\,s^{-1}$.

transition wavelength for the C IV doublet. The parametric model is ordinarily fit within the wavelength interval 1500-1600Å (corresponding to approximately $\pm 10\,000$ km s^{-1} from the rest-frame transition wavelength), a recipe that is commonly adopted (e.g. Denney et al., 2013). The line-window was extended if more than 5 per cent of the total flux in the profile was present blueward of the short wavelength limit. Narrow absorption features, which are frequently found superimposed on C IV emission, were masked out during the fit.

The C IV emission was fit with sixth-order Gauss-Hermite (GH) polynomials, using the normalisation of van der Marel and Franx, (1993) and the functional forms of Cappellari et al., (2002). We allowed up to six components, but in many cases a lower order was sufficient (40 and 45 per cent were fit with second- and fourth-order GH polynomials respectively). GH polynomials were chosen because they are flexible enough to model the often very asymmetric C IV line profile. The flip-side of this flexibility, however, is that the model has a tendency to over-fit when spectra possess low S/N. The fits were therefore carefully checked visually and the number of components reduced if over-fitting was evident.

We find that using the commonly employed three-Gaussian component model, rather than the GH polynomials, resulted in only marginal differences in the line parameters. Our best-fit parameters are also in good agreement with Shen et al., (2011), who employ a multi-Gaussian parametrization. In Fig. 3.2 we compare our measurements of the C IV FWHM from the 71 SDSS DR7 spectra in our sample with the measurements published in Shen et al., (2011). There is a very strong agreement between our measurements, with a scatter of 0.05 dex (200km s^{-1}).

3.3.2 *Hα*

A power-law continuum is fit using two continuum windows at 6000-6250 and 6800-7000 Å. The continuum-subtracted flux is then fit in the wavelength interval 6400-6800 Å. We adopt a rest-frame transition wavelength of 6564.89 Å to transform wavelengths into equivalent Doppler velocities. The broad component of Hα is fit using one or two Gaussians, constrained to have a minimum FWHM of 1200km s^{-1}. When two Gaussians are used, the velocity centroids are constrained to be the same.

The emission-line profiles of both Hβ and Hα frequently include a significant narrow component from the physically more extended narrow line region (NLR). Additional Gaussian components were included in our parametric model to fit the narrow component of Hα as well as [N II]λλ6548,6584 and [S II]λλ6717,6731. This resulted in a better fit to the observed flux in 50 per cent of cases. We impose a 1200km s^{-1}upper limit on the FWHM of all narrow lines and the
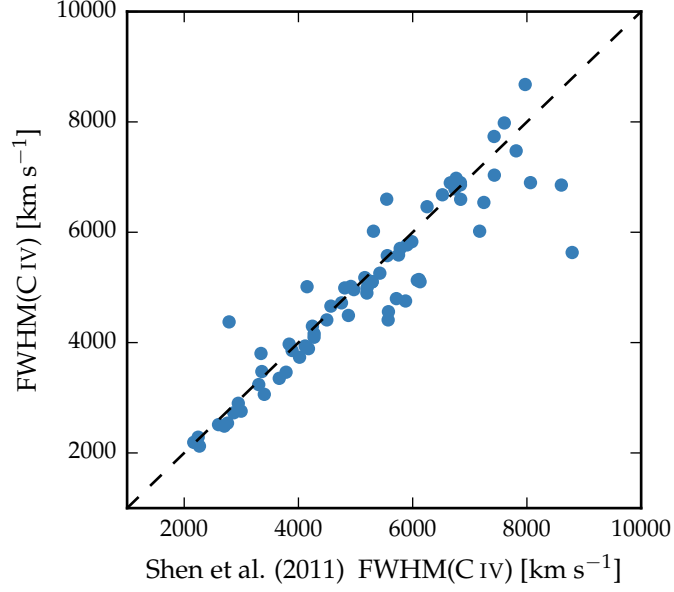
Figure 3.2: Demonstration of the effectiveness of our line parameter estimation scheme via a comparison of the C IV FWHM with Shen et al., (2011).

amplitudes of all components must be non-negative. The relative flux ratio of the two [N II] components is also fixed at the expected value of 2.96. In 70 per cent of the spectra the [O III]$\lambda\lambda4960,5008$ doublet is detected at moderate S/N in the Hβ region. In these cases the peak of the [O III] is used to fix the velocity offsets and the FWHMs of the narrow line components in the Hα region. For spectra where the [O III] doublet does not constrain the velocity and FWHM accurately, the narrow emission in the Hα and Hβ regions are fitted independently but, for each region, the individual narrow-line velocity offsets and the FWHMs are constrained to be identical. In these objects the narrow line contribution is generally weak, and so does not have a large effect on the line parameters we measure for the broad component.

The model described above is very similar to the one described in Shen and Liu, (2012) and Shen et al., (2011), the only major differences being that we do not fit the Hα and Hβ emission regions simultaneously and we fix the centroids of the Gaussian components used to fit the broad emission. In Fig. 3.3 we plot our Hα FWHM measurements against the measurements published in Shen and Liu, (2012), for 51 quasars in common to both samples. There is a strong correlation and a scatter of just 0.07 dex.
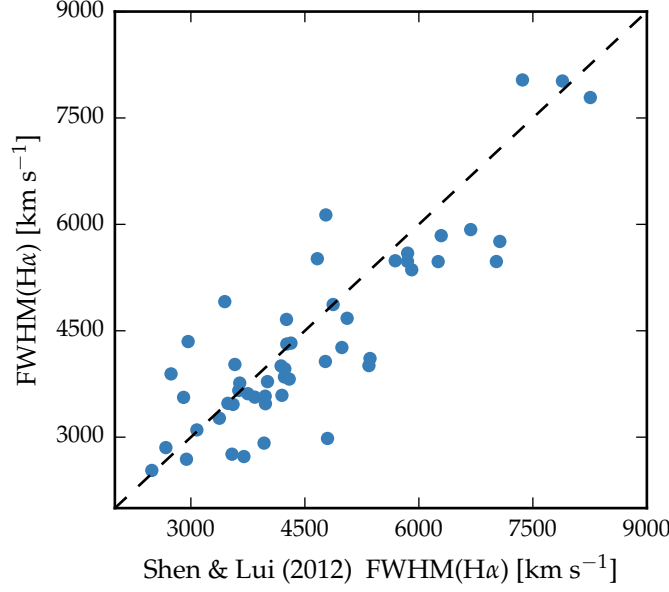
Figure 3.3: Demonstration of the effectiveness of our line parameter estima-
tion scheme via a comparison of the Hα FWHM with Shen and
Liu, (2012).

3.3.3   *Hβ and [O III]*

Emission from optical Fe II is generally strong in the vicinity of Hβ.
We therefore fit a combination of a power-law continuum and an
optical Fe II template – taken from Boroson and Green, (1992) – to
two windows at 4435-4700 and 5100-5535 Å. The Fe II template is con-
volved with a Gaussian, and the width of this Gaussian, along with
the normalisation and velocity offset of the Fe II template, are free
variables in the pseudo-continuum fit. We use the same model to fit
the broad and narrow components of Hβ as was used with Hα. Each
line in the [O III] doublet is fit with two Gaussians, to model both
the systemic and any outflow contributions. The peak flux ratio of
the [O III] 4960 Å and 5008Å lines is fixed at 1:3. As for the fit to the
narrow lines in the spectral region around Hα, the width and velocity
offsets of all the narrow components are set to be equal, and an upper
limit of 1200km s^{-1} is placed on the FWHM.

The parametric model we fit to the Hβ /[O III] emission region was
very similar to the model employed by Shen, (2016). In Fig. 3.4 we
plot our Hβ FWHM measurements against the measurements pub-
lished in Shen, (2016), for 39 quasars in common to both samples. As
expected, we observe a very tight correlation, with a scatter of 0.04
dex.
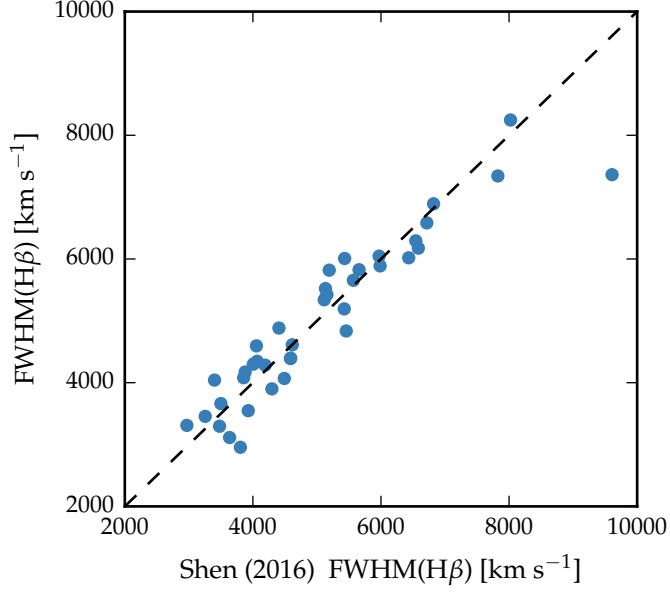
Figure 3.4: Demonstration of the effectiveness of our line parameter estimation scheme via a comparison of the Hβ FWHM with Shen, (2016).

3.3.4 *Fitting procedure*

Model parameters were derived using a standard variance-weighted least-squares minimisation procedure employing the Levenberg-Marquardt algorithm. Prior to the fit, the spectra were inspected visually and regions significantly affected by absorption or of low S/N were masked out.

In Fig. 3.5 we present our parametric fits to the C IV, Hα and Hβ emission lines in a handful of quasars, which have been chosen to illustrate the range of spectrum S/N and line shapes in the sample. The Doppler velocities have been shifted so that the Hα emission line centroid is at $0\,\mathrm{km\,s^{-1}}$. The y-axes of the data-minus-model residual plots have been scaled by the spectrum flux errors. The mean reduced chi-squared values in our Hα, Hβ and C IV fits are 1.69, 1.62, and 1.77 respectively and, in general, there are no strong features observable in the spectrum minus model residuals. The only significant features seen in the residual C IV spectra correspond to the location of narrow absorption lines which were excluded in the fitting procedure.

Table 3.2 includes the line parameters of our best-fitting model for each line. The reported line-width measures are corrected for instrumental broadening by subtracting the resolution of the spectrograph in quadrature. The spectrograph resolutions, which we estimate from the line widths in the observed sky spectra, range from $25\mathrm{km\,s^{-1}}$ for XSHOOTER to $477\mathrm{km\,s^{-1}}$ for the low-resolution LIRIS grism and are therefore small relative to the quasar broad line widths.

Table 3.2: The format of the table containing the emission line properties from our parametric model fits. The table is available in machine-readable form in the online version of Coatman et al., (2017).

	Units	Description
NAME		Catalogue name
FWHM_BROAD_HA	$\mathrm{km\,s^{-1}}$	FWHM of broad Hα line
FWHM_BROAD_HA_ERR	$\mathrm{km\,s^{-1}}$	
SIGMA_BROAD_HA	$\mathrm{km\,s^{-1}}$	Dispersion of broad Hα line
SIGMA_BROAD_HA_ERR	$\mathrm{km\,s^{-1}}$	
Z_BROAD_HA		Redshift from broad Hα line
FWHM_BROAD_HB	$\mathrm{km\,s^{-1}}$	FWHM of broad Hβ line
FWHM_BROAD_HB_ERR	$\mathrm{km\,s^{-1}}$	
SIGMA_BROAD_HB	$\mathrm{km\,s^{-1}}$	Dispersion of broad Hβ line
SIGMA_BROAD_HB_ERR	$\mathrm{km\,s^{-1}}$	
Z_BROAD_HB		Redshift from broad Hβ line
FWHM_CIV	$\mathrm{km\,s^{-1}}$	FWHM of C IV doublet
FWHM_CIV_ERR	$\mathrm{km\,s^{-1}}$	
SIGMA_CIV	$\mathrm{km\,s^{-1}}$	Dispersion of C IV doublet
SIGMA_CIV_ERR	$\mathrm{km\,s^{-1}}$	
BLUESHIFT_CIV_HA	$\mathrm{km\,s^{-1}}$	Blueshift of C IV relative to Hα
BLUESHIFT_CIV_HA_ERR	$\mathrm{km\,s^{-1}}$	
BLUESHIFT_CIV_HB	$\mathrm{km\,s^{-1}}$	Blueshift of C IV relative to Hβ
BLUESHIFT_CIV_HB_ERR	$\mathrm{km\,s^{-1}}$	
LOGL5100	$\mathrm{erg\,s^{-1}}$	Luminosity at 5100Å
LOGL1350	$\mathrm{erg\,s^{-1}}$	Luminosity at 1350Å
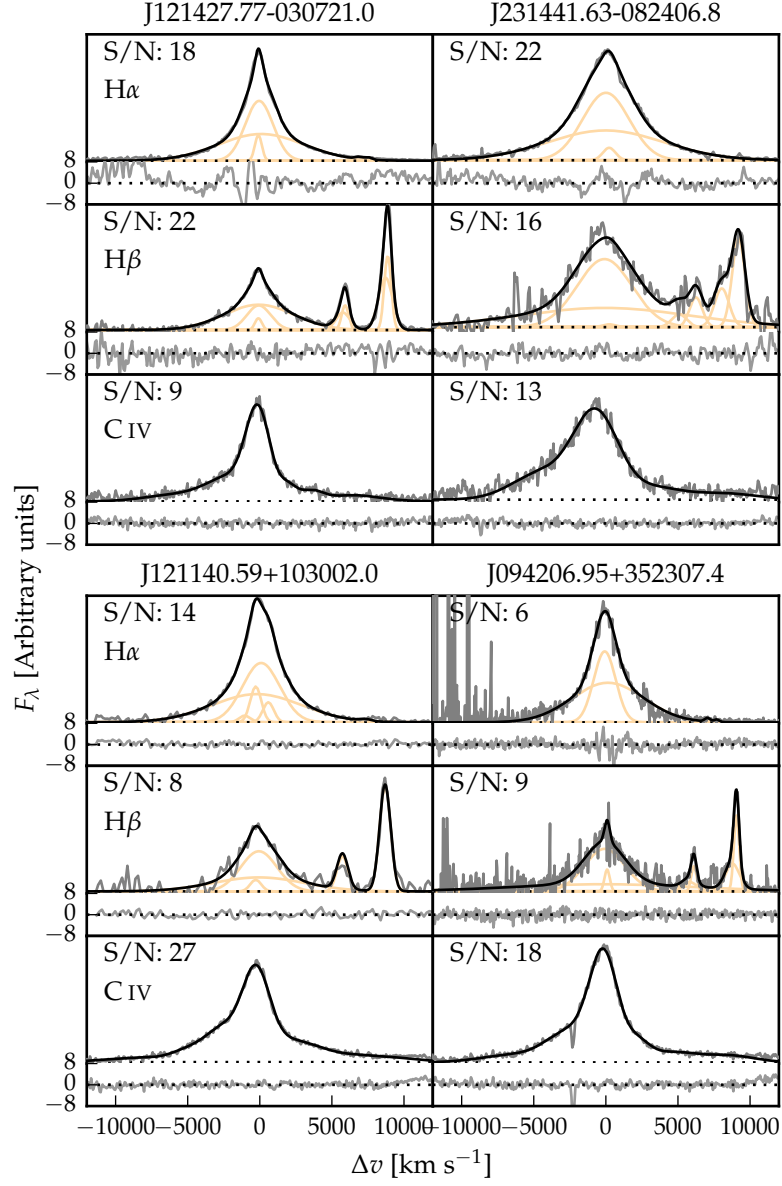
Figure 3.5: Model fits to continuum-subtracted Hα, Hβ, and C IV emission in four quasars, chosen to represent the range of S/N (indicated in the figure and given per 150km s^{-1} pixel in the continuum) and line shapes present in the catalogue. The data is shown in grey, the best-fitting parametric model in black, and the individual model components in orange. The centroid of the broad Hα emission is used to set the redshift, and Δv is the velocity shift from the line rest-frame transition wavelength. Below each fit we plot the data minus model residuals, scaled by the errors on the fluxes.

Table 3.3: The number of spectra removed from our sample by the cuts described in Section 3.3.5.

		Hα sample	Hβ sample
Total		194	279
Hα/Hβ	Wavelength	6	27
	S/N	8	83
C IV	Wavelength	6	5
	S/N	4	12
	Absorption	6	8
Total remaining		164	144

3.3.5 *Spectra removed from sample*

Through visual inspection we flagged and discarded the spectra of quasars for which reliable emission line parameters could not be obtained.

First, we flagged emission lines in spectra that possessed insufficient S/N. A single minimum S/N threshold was not entirely effective and, instead, spectra were flagged when it was judged conservatively that no meaningful constraints could be placed on the velocity centroid and/or width of the emission-line.

Second, we flagged emission lines where significant regions of the continuum and/or emission line fell outside of the wavelength coverage of the spectra. Reliable continuum definition and subtraction is not straightforward for emission lines so affected.

Third, we flagged C IV emission lines because of strong, narrow absorption close to the peak of the line where reliable interpolation across the absorption, using our parametric model, was not possible.

The number of spectra that are removed by each cut is given in Table 3.3 and the distribution in redshift and luminosity is shown in Fig. 3.6. Unsurprisingly, there is a preferential removal of intrinsically faint quasars, whose spectra can be of poorer S/N, and a loss of quasars at redshifts $z \sim 2.6$ where the Hα emission falls at the edge of the K-band. Hβ is much weaker than Hα, and the Hβ spectra are generally of lower S/N. As a result, the fraction of Hβ spectra that are flagged – 39 per cent – is particularly high.

3.3.6 *Emission-line parameter uncertainties*

The 1σ error bars calculated from the covariance matrix in least-squares minimisation will underestimate the true uncertainties on the line parameters, since they do not account for systematic errors such as the significant uncertainty introduced in the continuum sub-
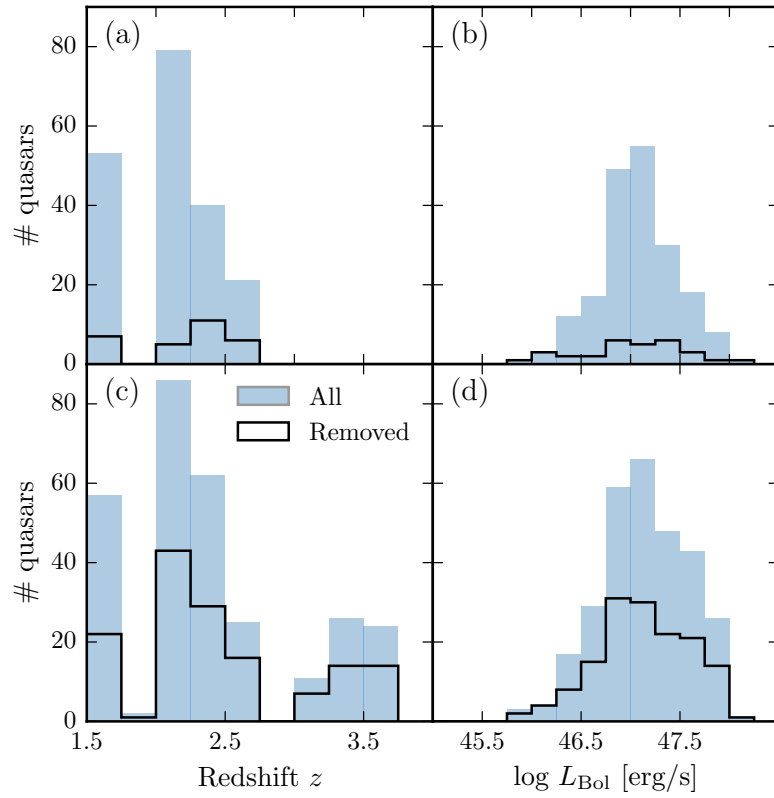
Figure 3.6: The redshift and luminosity distributions of the spectra removed from our Hα/C IV (a, b) and Hβ/C IV (c, d) samples.

traction procedure. To calculate more realistic uncertainties on our fitted variables we employed a Monte Carlo approach. One thousand artificial spectra were synthesised, with the flux at each wavelength drawn from a Normal distribution (mean equal to the measured flux and standard deviation equal to the known error). Our emission-line fitting recipe was then implemented on each of these mock spectra. The uncertainty in each parameter is given by the spread in the best-fitting values from the one thousand realisations of the fitting routine. In some cases the standard deviation of the parameter distribution was biased by extreme values caused by bad fits[2]. We therefore chose to measure the spread in the parameter distribution by fitting a composite model with two Gaussian components – one to model uncertainty in the parameter and the other any possible outlier component. The uncertainty in each line parameter was then taken to be the width of the narrower Gaussian. The uncertainties on all derived quantities, such as the BH mass, are propagated through by assuming that the uncertainties are uncorrelated and independent.

3.3.7 *Contemporaneity of spectra*

The epochs of the near-infrared and optical spectra can differ by many years. For example, the NTT SOFI spectra were taken ~14 years after the SDSS spectra, and the VLT SINFONI spectra 20 years or more after the Hamburg/ESO observations[3]. If the broad emission line profiles varied significantly on these time-scales the relation between the C IV and Balmer line-width measurements could be blurred.

Cases do exist of dramatic changes in quasar spectra over short time-scales, but this phenomenon is rare (MacLeod et al., 2016). In our spectroscopic catalogue there are 112 SDSS DR7 quasars which are re-observed in BOSS and included in the DR12 quasar catalogue. The mean time elapsed between the two sets of observations is ~8 years. The root-mean-square difference in the C IV FWHM measured from the BOSS and SDSS spectra is a modest $\simeq 500 \mathrm{km\,s}^{-1}$. Differences in the S/N of the spectra will make a substantial contribution and the scatter due to true variations in the C IV velocity-width will be significantly smaller than $500 \mathrm{km\,s}^{-1}$. We conclude therefore that any intrinsic changes with time do not materially affect the emission line measurements.

3.3.8 *Quasar monochromatic luminosity*

Computing virial BH masses also requires the quasar luminosity in an emission-line free region of the continuum adjacent to the broad line being used. The luminosity is used as a proxy for the size of the

---

2 In the analysis of the real spectra such fits are identified via visual inspection.
3 Time differences in the quasar rest-frame are reduced by a factor of $(1 + z)$.

BLR. The monochromatic continuum flux is generally measured at 1350 Å for C IV and 5100 Å for Hα and Hβ.

Relative flux-calibration of the infrared spectra as a function of wavelength has been achieved through observations of appropriate flux standards. The absolute flux levels, however, can be in error by large factors due to variable atmospheric conditions combined with the narrow slit widths. For the majority of the quasars we have, therefore, established the absolute flux scale for each near-infrared spectrum by fitting an SED-model to the optical-infrared photometry. The SED model, described in Chapter 5, gives a very good fit to the SDSS and UKIDSS magnitudes of SDSS DR7 quasars, reproducing the individual magnitudes with a σ <0.1 mag. For 207 quasars, (Y)JHK passband magnitudes from the UKIRT Infrared Deep Sky Survey (UKIDSS; Lawrence et al., 2007) Large Area Survey, the Two Micron All Sky Survey (2MASS; Skrutskie et al., 2006) and the Visible and Infrared Survey Telescope for Astronomy (VISTA) Hemisphere Survey (VHS; McMahon et al., 2013) and Kilo-Degree Infrared Galaxy (VIKING; Edge et al., 2013) survey are available. The SED model was fit to the infrared magnitudes; integrating the SED model through the pass-band transmission functions, to give model magnitudes, and performing a variance weighted least-squares fit to the observed magnitudes. The flux at 5100 Å was then taken from the normalised model.

For 19 of the remaining 23 quasars, where near-infrared photometry was not available, the quasar SED model was fit to the SDSS spectra, the flux calibration of which are known to be excellent. The fit was done using a simple variance-weighted chi-squared minimisation procedure in emission line-free intervals of the optical spectra. The model includes a reddening, $E(B - V)$, based on a Small Magellanic Cloud-like extinction curve and described in detail in Section 5.2.7, and an overall normalisation as free parameters. In practice, the quasars possess only very modest reddenings, with $E(B - V) \simeq$0.0-0.1. The flux at 5100 Å was then, again, taken from the normalised SED model.

For the four remaining quasars, which possess neither near-infrared photometry nor SDSS DR7 spectra, we fit the SED model to the BOSS DR12 spectra. To avoid the known issues in the flux calibration of the BOSS DR12 quasar spectra at observed-frame blue wavelengths (Lee et al., 2013), our fitting was confined to rest-frame wavelengths long-ward of 1275Å.

The monochromatic luminosity at 1350Å was also measured by fitting our quasar SED model to the SDSS/BOSS spectra. For 26 quasars in the catalogue the optical spectra come from surveys other than SDSS/BOSS and optical magnitudes from recent epochs are not available. In order to obtain an estimate of the luminosity at 1350 Å for the 26 quasars, we normalise the quasar SED model to the near-infrared photometric data, and read off the flux at 1350 Å.

Comparison of the 5100 Å luminosity, computed using the photometry- and spectrum-based methods for 177 quasars, showed a scatter of just ~0.1 dex. We therefore assume 0.1 dex to be the measurement uncertainty on the 5100 Å luminosities. We expect the uncertainties on the 1350 Å luminosities to be at similar level. For all the catalogue quasars, the optical and near-infrared spectra as well as the near-infrared photometry were obtained at different epochs, with rest-frame time differences of up to ~5 years. Intrinsic quasar photometric variability in the rest-frame ultraviolet and optical will therefore add additional scatter of ~0.2 mag (e.g. MacLeod et al., 2010) to the derived 1350- and 5100 Å-luminosities. Given that the luminosity enters into the calculation of BH-mass only as the square-root, the uncertainty on the luminosities does not make a large contribution to the uncertainties in the BH mass estimates.

3.3.9 *Characterising the emission-line widths*

There has been a considerable degree of attention paid to the effectiveness of different velocity-width measures of the C IV-emission; specifically, the line FWHM and the dispersion, σ, derived from the second-moment velocity (e.g. Assef et al., 2011; Denney et al., 2013). The FWHM and line dispersion trace different parts of the broad line velocity field, with the FWHM relatively more sensitive to any low-velocity core present and the line dispersion relatively more sensitive to the high velocity wings. In practice, the line dispersion is almost certainly a more robust velocity indicator when the assumptions underlying the virial-origin of the emission-line velocity width are true and the spectral S/N and resolution are adequate. This was demonstrated by Denney et al., (2013) for a sample of quasars possessing a significantly smaller range in C IV-blueshift than investigated here.

In reality, however, as highlighted by Denney, (2012), contributions to the C IV-emission line profile from gas where virial motions do not dominate can be significant. Looking to the future, the results of the new reverberation-mapping projects (Shen et al., 2015; King et al., 2015) will show what fraction of the C IV-emission line, as a function of velocity, does reverberate for quasars with an extended range of C IV-emission shapes. The derivation of quantitative corrections to transform velocity-width measures from single-epoch to reverberation-only line profiles should then be possible.

As such information is not yet available, there is a strong rationale for investigating whether the systematic changes in the C IV-emission line profile can be used to improve the single-epoch BH-mass estimates derived using the C IV line. In Fig. 3.7 we show how the C IV FWHM, line dispersion, σ, and line shape, FWHM/σ, vary as a function of the blueshift. The C IV FWHM is correlated with the blueshift, with the median FWHM of quasars with the largest blueshifts a factor
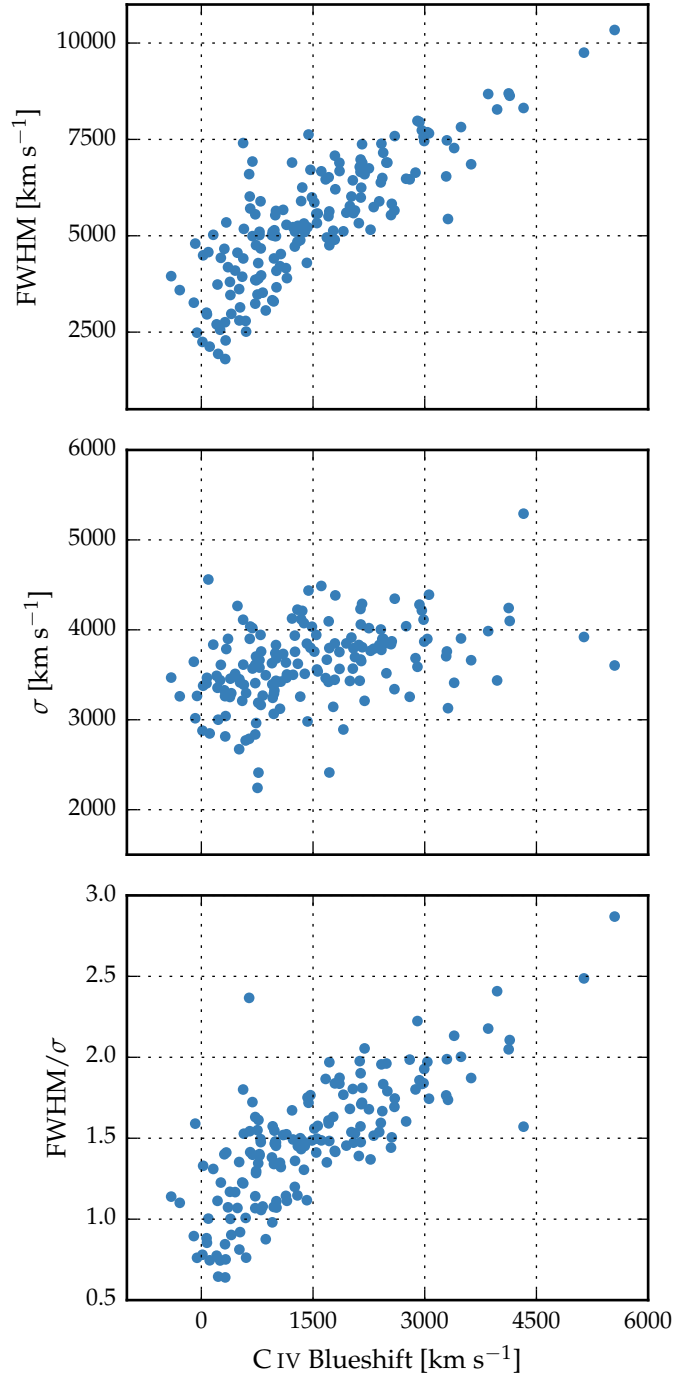
Figure 3.7: The FWHM, dispersion (σ) and shape (FWHM/σ) of C IV as a function of the C IV blueshift.

of 2-3 higher than quasars with only moderate blueshifts. The dispersion, however, does not show a similarly strong systematic variation.

Without knowledge of the C IV-blueshifts, the dynamic range present in the FWHM and line dispersion measurements accords with the expectations from the study of Denney et al., (2013); the factor of $\simeq 4$ spread in the FWHM measurements indicating greater sensitivity to the emission-line profile shape than is the case for the dispersion, which varies by a factor of only $\lesssim 2$. Adopting a value of $1200\,\mathrm{km\,s^{-1}}$ to define 'low' and 'high' blueshift, the median C IV-emission dispersion for the low and high-blueshift samples differ by only 10 per cent. It follows, therefore, that while the dispersion provides a relatively line-profile independent measure of the velocity width for quasars where the underlying assumption regarding the virial-origin of the velocity width applies, quasars where the assumption is not true can be assigned apparently normal velocity-widths and hence potentially incorrect BH-masses.

To emphasise this point, in Fig. 3.8 we overlay the C IV line profiles of SDSSJ1236+1129 and SDSSJ1525+2928, whose dispersions are indistinguishable (4168 ± 271 and $4303\pm128\,\mathrm{km\,s^{-1}}$ respectively). Notwithstanding the very similar dispersion values, the emission-line velocity fields differ dramatically and, therefore, the dispersion values cannot be measuring accurately the virial-induced velocity spread of the C IV emission in both quasars.

The analysis here, building on earlier work (including Shen and Liu, 2012; Sulentic et al., 2007), confirms a link between C IV emission-line shape and blueshift, raising the prospect of developing a blueshift-dependent correction to single-epoch BH-mass estimates based on the C IV line. Expressed in another way, we are interested in testing if the significant systematic change in line shape as a function of C IV blueshift can be used to provide improved single-epoch BH-masses from the C IV emission line. The tightness of the correlation we observe between the C IV FWHM and blueshift implies that such an approach may be more effective than using the C IV emission-line velocity dispersion without reference to blueshifts. A further practical advantage is that, given the typical S/N of current survey-quality spectra, virial BH mass estimates for high-redshift quasars are usually based on the FWHM rather than the dispersion (e.g. Shen et al., 2011), which, being strongly affected by the continuum placement, is often found to be difficult to measure robustly (e.g. Mejía-Restrepo et al., 2016).
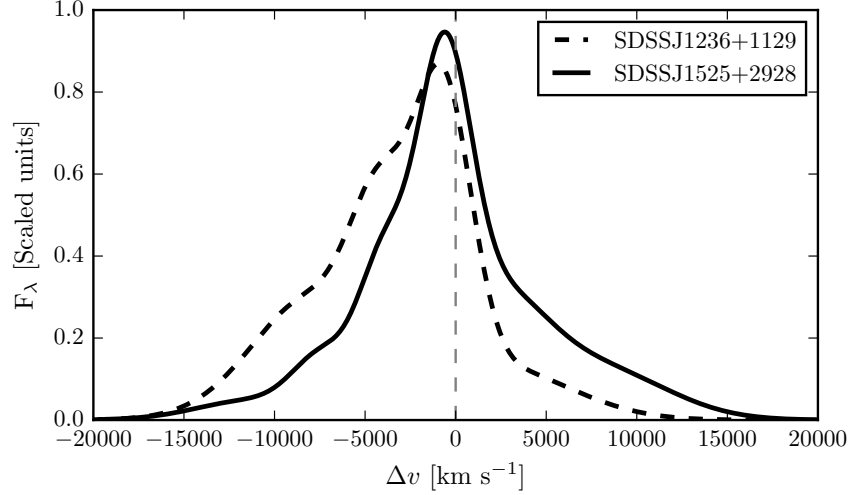
Figure 3.8: Comparison of the C IV line profiles of SDSSJ1236+1129 and SDSSJ1525+0426. Notwithstanding the essentially identical dispersion values, the emission-line velocity fields differ dramatically and, therefore, the dispersion values cannot be measuring accurately the virial-induced velocity spread of the C IV emission in both quasars.

3.4 AN EMPIRICAL CORRECTION TO C IV-BASED VIRIAL BH-MASS ESTIMATES

3.4.1 *Hα/Hβ FWHM comparison*

BH-mass calibrations which use the width of the broad Hβ emission line as a proxy for the virial velocity are widely regarded as the most reliable, since most reverberation mapping employs the Hβ line and the R − L relation has been established using Hβ. When Hβ is not available, Hα has been shown to be a reliable substitute (e.g. Greene and Ho, 2005b; Shen et al., 2011; Shen and Liu, 2012).

In our sample, we have 99 quasars with reliable measurements of both Hα and Hβ lines. The 99 objects include 21 quasars which were excluded from the main 308-object catalogue because the C IV FWHM and/or blueshift could not be measured reliably. The line widths are compared in Fig. 3.9 and, as expected, a tight correlation is observed. Greene and Ho, (2005b), using a sample of 162 quasars with high S/N SDSS spectra at $z < 0.35$, established the following relation between the Hα and Hβ FWHMs

$$\text{FWHM}(\text{H}\beta) = (1.07 \pm 0.07) \times 10^3 \left(\frac{\text{FWHM}(\text{H}\alpha)}{10^3 \text{ km s}^{-1}} \right)^{(1.03 \pm 0.03)} \tag{3.1}$$

The relation is shown as the dashed line in Fig. 3.9. The root-mean-square scatter about this relation is 0.07 dex, compared to the ∼0.1 dex
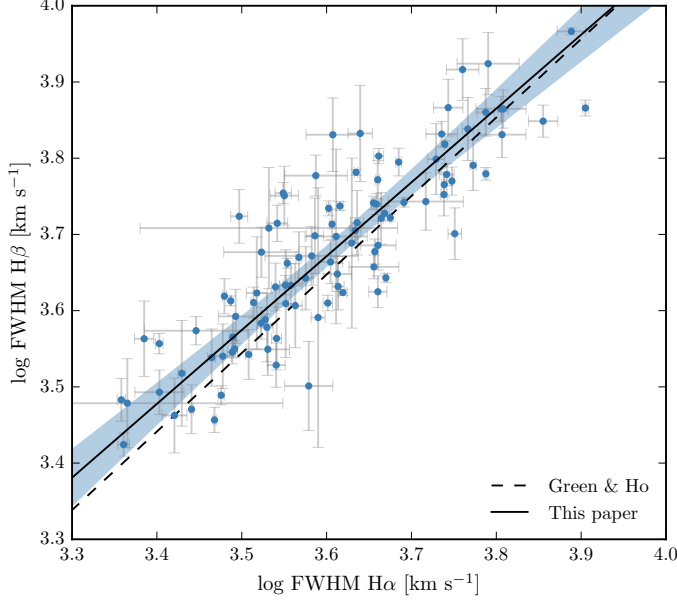
Figure 3.9: Comparison of Hα and Hβ FWHM measurements for 99 quasars. The solid line is our best-fitting power-law model, and the blue-shaded region shows the 2-σ uncertainties on the model parameters. The dashed line is the relation found by Greene and Ho, (2005b) using a sample of $z < 0.35$ SDSS AGN.

found by Greene and Ho, (2005b). However, we find a systematic offset, in the sense that the Hβ line-widths we measure are on average larger by $270 \mathrm{km\,s^{-1}}$ than predicted by the Greene and Ho, (2005b) relation. As our sample covers higher redshifts and luminosities than the sample in Greene and Ho, (2005b), we derive a new relation between the Hα and Hβ FWHMs.

We assume a relation of the same form used by Greene and Ho, (2005b), i.e. a simple power-law, and infer the model parameters by fitting a linear model (with slope α and intercept β) in log-log space. The fit is performed within a Bayesian framework described by Hogg, Bovy, and Lang, (2010). Each data point is treated as being drawn from a distribution function that is a convolution of the projection of the point's covariance tensor, of variance Σ_i^2, with a Gaussian of variance V representing the intrinsic variance in the data. The log-likelihood is then given by

$$\ln \mathcal{L} = -\sum_{i=1}^{N} \frac{1}{2} \ln \left[2\pi \left(\Sigma_i^2 + V \right) \right] - \sum_{i=1}^{N} \frac{\Delta_i^2}{2[\Sigma_i^2 + V]} \tag{3.2}$$

where Δ_i is the orthogonal displacement of each data point from the linear relationship. An advantage of this approach is that it allows a proper treatment of the measurement errors on both variables, which in this case are comparably large. The model also makes the reason-

able assumption that there is an intrinsic scatter in the relationship between the variables that is independent of the measurement errors. Following the suggestion by Hogg, Bovy, and Lang, (2010), the linear model was parametrized in terms of $(\theta, b_\perp)$, where θ is the angle the line makes with the horizontal axis and $b_\perp$ is the perpendicular distance from the line to the origin. Uniform priors were placed on these parameters, and the Jeffreys prior (the inverse variance) was placed on the intrinsic variance. The posterior distribution was sampled using a Markov Chain Monte Carlo (MCMC) method using the Python package emcee (Foreman-Mackey et al., 2013).

The one- and two-dimensional posterior distributions are shown in Fig. 3.10. The solid line in Fig. 3.9 is the maximum likelihood solution

$$\mathrm{FWHM(H\beta)} = (1.23 \pm 0.10) \times 10^3 \left(\frac{\mathrm{FWHM(H\alpha)}}{10^3 \mathrm{km\,s^{-1}}} \right)^{0.97 \pm 0.05} \quad (3.3)$$

and the shaded region shows the 2σ uncertainties on the model parameters.

As discussed above, our relation is displaced to slightly higher Hβ FWHM than the Greene and Ho, (2005b) relation – the offset is $210\,\mathrm{km\,s^{-1}}$ for a quasar with Hα FWHM $4500\,\mathrm{km\,s^{-1}}$. We infer a power-law index that, although slightly shallower, is consistent with the Greene and Ho, (2005b) index within the quoted uncertainties. The intrinsic scatter in the data, σ_I, we infer from the fit is 0.04 dex. This is smaller than the total scatter seen in Fig. 3.9 (0.06 dex), which suggests that measurement errors make a significant contribution to the total scatter in the relation.

We constructed composite spectra of the Hα and Hβ regions from 217 and 171 quasars respectively. Spectra were first de-redshifted to the quasar rest-frame, and then interpolated on to a common wavelength grid with a 1Å resolution. The spectra were scaled by the mean flux in the interval 4700-5100Å (Hβ) and 6400-6800Å (Hα). The composite was then defined as the median flux from all of the normalised spectra in each wavelength bin. The Hα and Hβ lines in the composite spectra are shown in Fig. 3.11. The cores of the two lines are very similar, but Hβ has more flux in the wings of the line.

For 19 of the 99 quasars with Hβ and Hα emission profiles, one of the two Gaussians used to reproduce the Hβ profiles has a FWHM greater than $20\,000\,\mathrm{km\,s^{-1}}$ and a fractional contribution to the total Hβ broad line flux of >0.3 (Marziani et al., 2009; Marziani et al., 2013). Such a broad component is not seen in the Hα profiles and the very broad Hβ-component may be an artifact of the fitting scheme. A particular issue for Hβ is the presence of Fe II emission, often at a significant level. Furthermore, additional lines could be contributing to the underlying continuum (e.g. the He I$\lambda\lambda$4922,5017 doublet; Véron, Gonçalves, and Véron-Cetty, 2002; Zamfir et al., 2010).
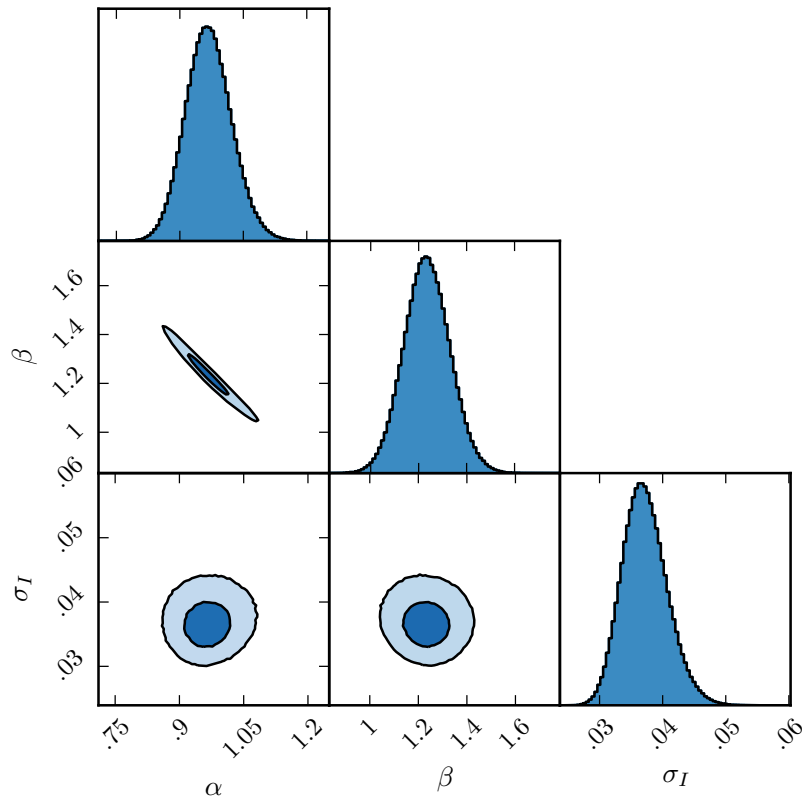
Figure 3.10: One- and two-dimensional projections of the MCMC sampling of the posterior distribution from the fit in Fig. 3.9. α is the power-law index, 10^{β} is the normalisation, and σ_I is the intrinsic scatter. In the two-dimensional projections, 1- and 2-σ contours are shown.
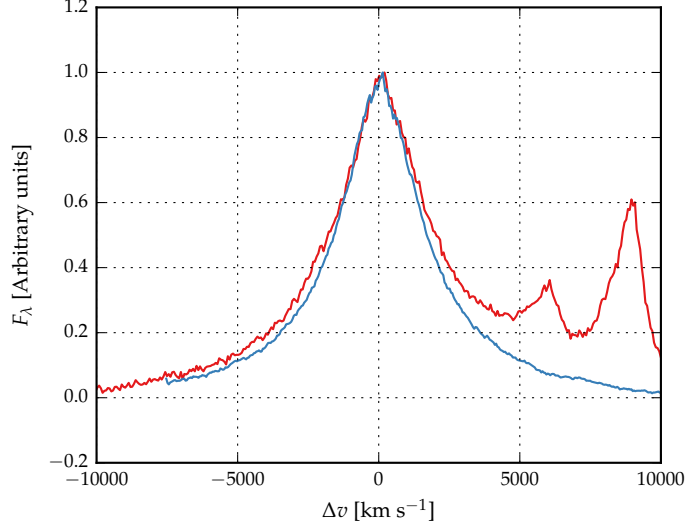
Figure 3.11: The Hα (blue) and Hβ (red) emission line regions in the median composite spectrum, shown as function of the velocity shift from the respective predicted line peak wavelengths. The backgroud continuum and optical Fe II emission has been modelled and subtracted. The line fluxes have been scaled in order for the profile shapes to be readily compared.

In Sec. 3.4.3 we use the whole of the Hβ profile to derive an un-biased BH mass. If, instead, the FWHM is calculated from the narrower of the two Gaussian components rather than the composite profile, then the Hβ FWHM decreases by 630 km s^{-1} on average. This effectively removes the average offset to broad Hβ profiles evident in Fig. 3.9. This will enhance the C IV FWHM relative to the Hα/Hβ FHWM by ~15 per cent and increase the size of the correction which must be applied to the C IV-based BH masses by ~30 per cent.

3.4.2 *Measuring the quasar systemic redshift*

An accurate measure of the quasar's systemic redshift is required in order for the blueshift of the C IV emission line to be determined. Balmer emission centroids are available for all quasars in the catalogue and so we use this to define the systemic redshift.

For 62 and 86 quasars in the Hα and Hβ samples respectively narrow [O III] emission is also detected with sufficient S/N to measure the line properties. In the model fit to the Hβ region the velocity centroids of the broad Hβ-line and the core component of the [O III] emission were deliberately determined separately. We find the intrinsic difference in the velocity centroids of the Hα and Hβ emission and the narrow [O III] emission to have a dispersion of 300 and 400 km s^{-1}, which is very similar to the value found by Shen et al., (2016).

However, the median velocity centroid of the narrow component of the [O III] emission is blueshifted by $250 \mathrm{km\,s}^{-1}$ relative to the centroid of the broad Balmer line. Applying our parametric model fitting routine to the composite spectrum from Hewett and Wild, (2010), which is constructed using relatively low redshift SDSS quasars with $L_{\mathrm{Bol}} \sim 10^{44} \mathrm{erg\,s}^{-1}$, the centroids of the broad component of Hβ and the narrow component of [O III] are found to be at essentially identical velocities, suggesting that the blueshifting of narrow [O III] could be luminosity dependent.

As described in Section 3.3, the broad components of Hα and Hβ were modelled with up to two Gaussians, with identical velocity centroids. If there is any significant asymmetry in these lines then the emission will be poorly fit by our model and the redshift derived from the peak of the best-fitting model could be biased. To investigate this possibility we relaxed the requirement for the centroids of the two broad Gaussians to be the same, and measured the systemic redshift from the peak of the composite profile. With this new, more flexible model, the mean absolute difference between the centroids of the two Gaussian components used to model Hα and Hβ was 480 and $780 \mathrm{\ km\,s}^{-1}$ respectively. With these adjustments, we found the mean difference between the [O III]- and Hα(Hβ) based redshift estimates to be -100(-120) $\mathrm{km\,s}^{-1}$ and the scatter to be 290(320) $\mathrm{km\,s}^{-1}$. Therefore, the shift between the Balmer and [O III] velocities is reduced, suggesting that there might be a $\sim$100 $\mathrm{km\,s}^{-1}$ systematic bias in our measurements of the quasar systemic redshift. Regardless, since both the systematic offset and the scatter are small in comparison to the dynamic range in C IV blueshifts, the blueshift-based empirical correction we will derive does not depend on whether the broad Balmer emission or the [O III] centroid is used to define the systemic redshift, or how the broad Balmer emission is parameterized.

Later, in section XX, we demonstrate how improvements in the estimation of systemic redshifts from ultraviolet quasar spectra means that it is now possible to quantify the distribution of C IV-blueshifts in the observed population as a whole. Clearly, this is a crucial development in making a blueshift-based correction viable.

3.4.3   *Balmer/C IV line widths as a function of C IV-blueshift*

In this section we directly compare the C IV and Hα/Hβ line widths as a function of the C IV blueshift. Because virial BH mass estimates are generally based on the Hβ FWHM, we first convert our Hα FWHM measurements to equivalent Hβ FWHM using Eq. 3.3. In Figs. 3.12 and 3.13 we show the C IV FWHM relative to both the (Hβ-scaled) Hα FWHM and the Hβ FWHM, as a function of the C IV blueshift.

Employing the same Bayesian fitting framework described in Section 3.4.1, we fit independent linear models to the C IV FWHM relative to the Hα and Hβ FWHM as a function of the C IV blueshift. As before, our model has an additional parameter representing any intrinsic scatter in the relationship between the variables which is independent of measurement errors. We also tested a model where some fraction of the data points (which is free to vary) are drawn from an outlier distribution, represented by a broad Gaussian centered on the mean of the data. We found, however, that the inferred outlier fraction was very low (0.004, corresponding to ∼0.7 data points) and so did not include such a component in our model.

In Fig. 3.14 we show the one- and two-dimensional projections of the posterior distribution from the linear fit to the FWHM C IV/Hα ratio. The projections from the FWHM C IV/Hβ fit, which we do not show, have very similar appearances. In Fig. 3.12 we plot the maximum likelihood model and the 2σ uncertainties on the model parameters. The maximum likelihood line is given by

$$\text{FWHM}(\text{C\,IV}, \text{Corr.}) = \frac{\text{FWHM}(\text{C\,IV}, \text{Meas.})}{(0.41 \pm 0.02)\left(\frac{\text{C\,IV Blueshift}}{10^3\,\text{km\,s}^{-1}}\right) + (0.62 \pm 0.04)} \quad (3.4)$$

for the C IV/Hα fit and

$$\text{FWHM}(\text{C\,IV}, \text{Corr.}) = \frac{\text{FWHM}(\text{C\,IV}, \text{Meas.})}{(0.36 \pm 0.03)\left(\frac{\text{C\,IV Blueshift}}{10^3\,\text{km\,s}^{-1}}\right) + (0.61 \pm 0.04)} \quad (3.5)$$

for the C IV/Hβ fit. The intercepts of the two relations are consistent, while the difference between the slopes is only marginally inconsistent given the quoted uncertainties.

The intrinsic scatter in the data about the linear relation we infer is 0.23 ± 0.02 and 0.25 ± 0.02 for the Hα and Hβ fits respectively. The intrinsic scatter for the Hα fit is represented by the Normal probability density distribution shown in Fig. 3.15. In the same figure we show the distribution of the orthogonal displacement of each data point from the best-fitting linear relationship. The two distributions are well-matched, which demonstrates that our model is a good representation of the data and the measurement errors on the data points are small relative to the intrinsic scatter.

The overall (intrinsic and measurement) scatter about the best-fitting model is slightly higher when the C IV line-widths are compared to Hβ (0.12 dex) than when compared to Hα (0.10 dex). This is likely due, at least in part, to the generally higher S/N of the Hα emission. In addition, contributions from the strong [O III] doublet in the vicinity of Hβ make de-blending the Hβ emission more uncertain. As a consequence, for quasars where Hα and Hβ are both measured, the mean uncertainty on the Hα FWHM is 130 km s^{-1}, compared to 340 km s^{-1} for Hβ.
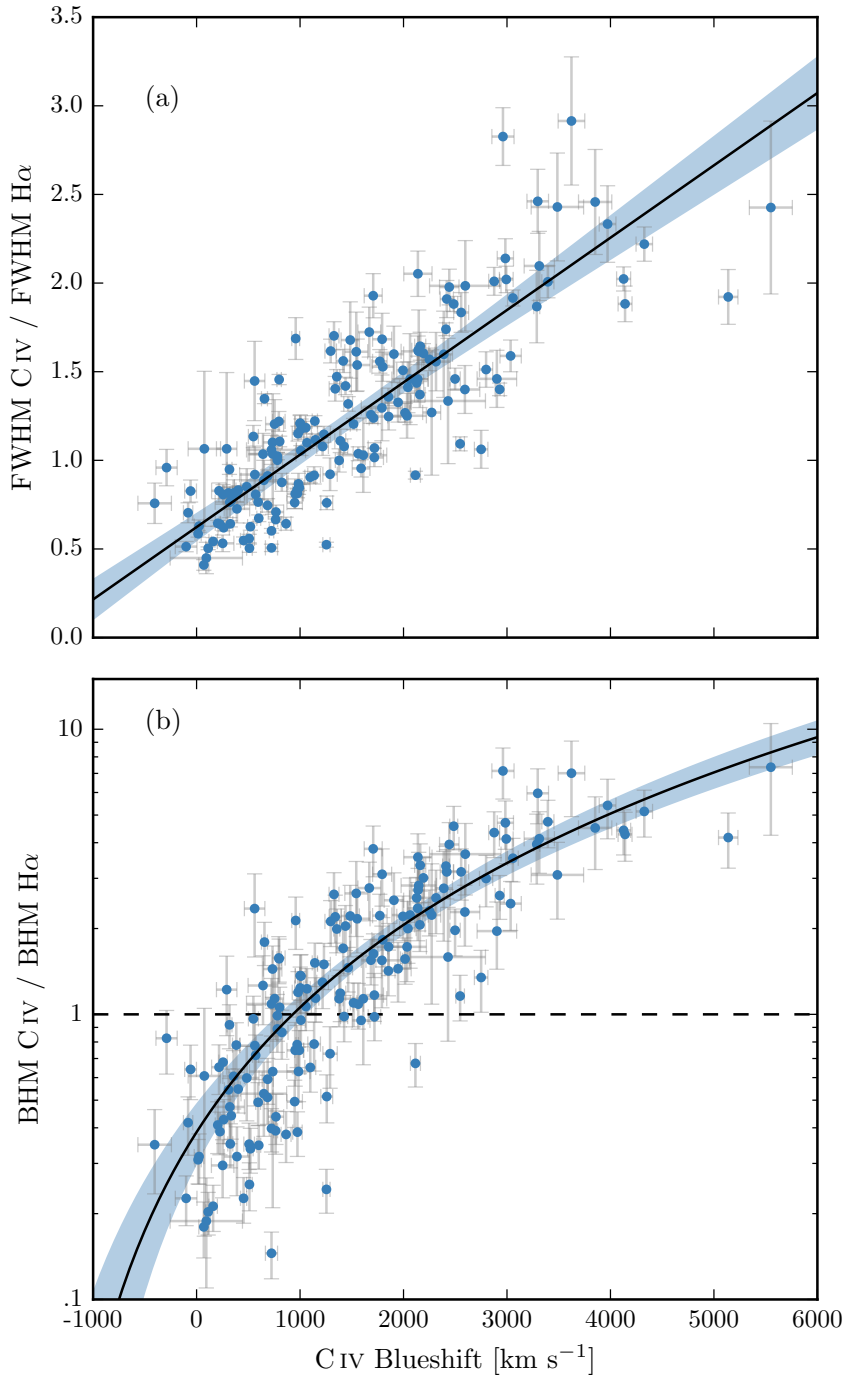
Figure 3.12: C IV FWHM relative to Hα FWHM (a), and C IV based BH mass (BHM) compared to Hα based mass (b), both as a function of the C IV blueshift. The black line is our best-fit linear model, and the shaded region shows the 2-σ uncertainties on the slope and intercept. The Hα FWHM have been scaled to match the Hβ FWHM using Eq. 3.3.
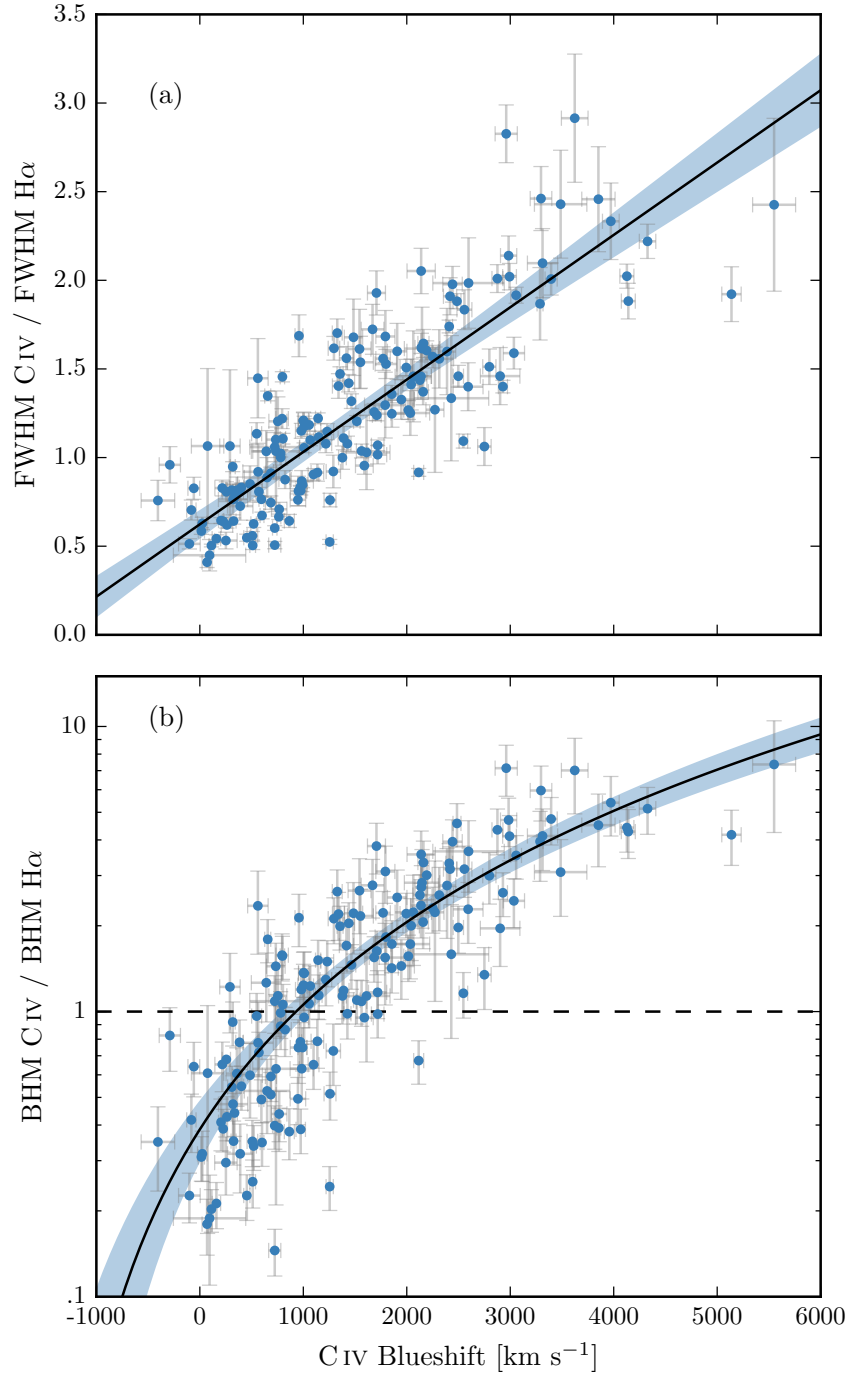
Figure 3.13: C IV FWHM relative to Hβ FWHM (a), and C IV based BH mass (BHM) compared to Hβ based mass (b), both as a function of the C IV blueshift.
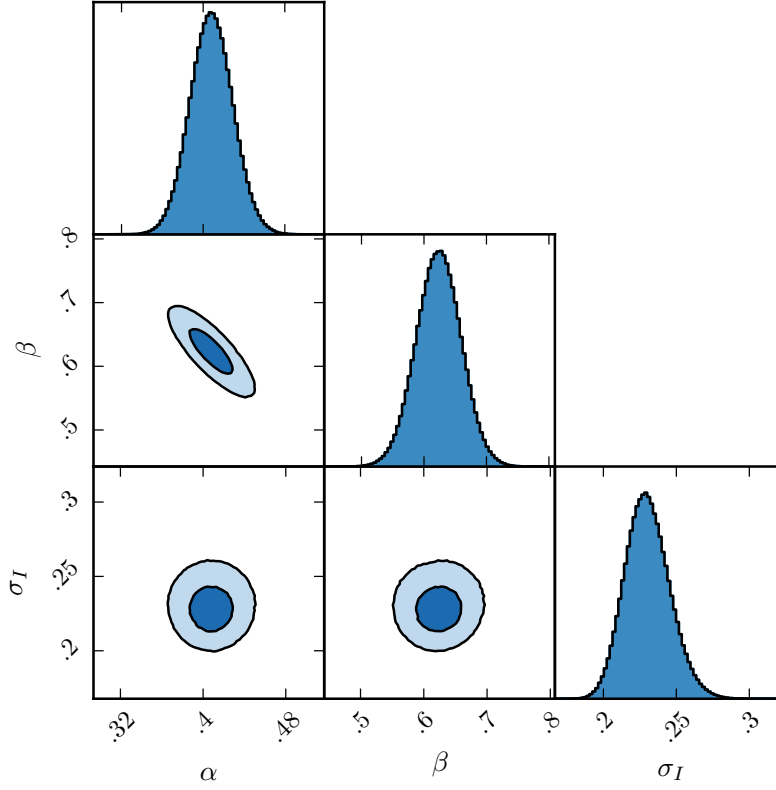
Figure 3.14: One- and two-dimensional projections of the MCMC sample of the posterior distribution for a linear fit to the FWHM C IV/Hα ratio as a function of the C IV blueshift. In the two-dimensional projections we show 1- and 2-σ contours. The posterior distribution for the linear fit to the FWHM C IV/Hβ ratio, which we do not show, has a very similar appearance.

In the next section we use both the Hα and Hβ lines to calculate unbiased BH masses. However, we use the Hα measurements to derive an empirical C IV blueshift based correction to the C IV masses (Eq. 3.6) because of the issues related to the accurate modelling of the Hβ-profile just described. An extra advantage, which is evident in Figs. 3.12 and 3.13, is that the Hα sample has a better C IV blueshift coverage. However, as can be seen from the similarity of Equations 3.4 and 3.5, our results would not change significantly were we instead to use the Hβ sample.

3.4.4 *C IV based virial BH mass estimates*

Virial BH masses were calculated using the widely adopted Vestergaard and Peterson, (2006) calibrations. The Vestergaard and Peterson, (2006) C IV FWHM calibration uses the monochromatic contin-
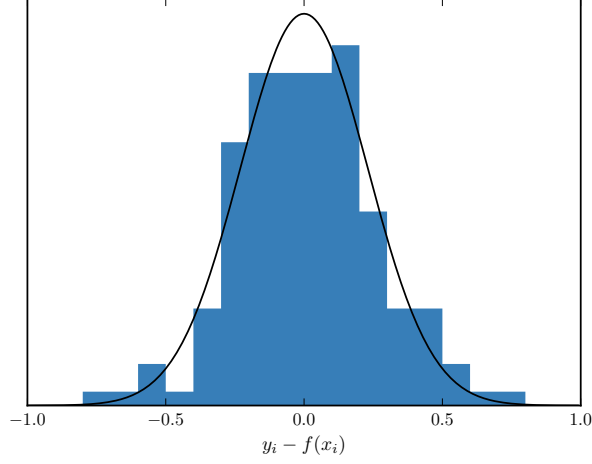
Figure 3.15: The distribution of the orthogonal displacement of each data point from the best-fitting linear relationship in the fit to FWHM(C IV)/FWHM(Hα) as a function of the C IV blueshift (blue histogram). The black curve is a Normal distribution with a width equal to the intrinsic scatter in the population inferred from the fit. The two distributions are well-matched, which demonstrates that our model is a good representation of the data and the measurement errors on the data points are small relative to the intrinsic scatter.

uum luminosity at 1350 Å to predict the BLR radius and corresponds to (a = 6.66, b = 2, c = 0.53) in Eq. 1.1. For the Hβ calibration, Vestergaard and Peterson, (2006) use the monochromatic continuum luminosity at 5100 Å and calibration coefficients corresponding to (a = 6.91, b = 2, c = 0.5). BH masses are computed using the line and continuum properties given in Table 3.2, and we convert our Hα emission-line velocity-width measures to predicted Hβ widths using Eq. 3.3.

In the lower panels of Figs. 3.12 and 3.13 the C IV-based estimates are compared to the Hα/Hβ estimates as a function of the C IV blueshift. There is a strong systematic error in the C IV-based masses as a function of blueshift, which is a direct consequence of the FWHM trend described in the previous section. The C IV emission-based BH-masses are in error by a factor of more than five at 3000km s^{-1} in C IV emission blueshift and the overestimate of the BH-masses reaches a factor of 10 for quasars exhibiting the most extreme blueshifts, $\gtrsim$5000km s^{-1}.

The virial product is the product of the virial velocity squared and the BLR radius (e.g. Shen, 2013), and is proportional to the BH mass. We use the corrected C IV FWHM given by Eq. 3.4 as an indicator of the virial velocity, and adopt the same $R - L$ relation for the 1350 Å continuum luminosity as Vestergaard and Peterson, (2006) (i.e. $R \propto L^{0.53}$). To find the constant scaling factor necessary to transform

the virial product in to a BH mass we compute the inverse-variance weighted mean difference between the virial products and the Hα-based masses. The virial BH mass can then be expressed in terms of the corrected C IV FWHM and monochromatic continuum luminosity at 1350 Å

$$MBH(\text{C\,IV, Corr.}) = 10^{6.71} \left(\frac{FWHM(\text{C\,IV, Corr.})}{10^3 \,\text{km\,s}^{-1}} \right)^2 \left(\frac{\lambda L_\lambda (1350\text{Å})}{10^{44}\,\text{erg\,s}^{-1}} \right)^{0.53}$$

$$(3.6)$$

Given measured C IV emission line FWHM and blueshift, equations 4 and 6 can then be used to provide an unbiased estimate of the quasar BH mass.

3.4.5 *C IV-derived BH masses at low C IV blueshift*

Might work better later in discussion.

In this section, we consider why the C IV based masses of quasars with modest C IV blueshifts ($\lesssim 1000 \text{km\,s}^{-1}$) are systematically underestimated relative to masses derived from the Balmer lines (Figs. 3.12 and 3.13).

Reverberation mapping measurements of nearby AGN have revealed the BLR to be stratified, with high-ionisation lines, including C IV, emitted closer to the BH than low-ionisation lines, including Hα and Hβ (e.g. Onken and Peterson, 2002). Vestergaard and Peterson, (2006) found that the C IV-emitting region is at approximately half the radius of the Hβ/Hα emitting region. Given the $\Delta V \propto R_{\text{BLR}}^{-0.5}$ virial relation, this leads to the prediction that the C IV line widths should be $\simeq 1.4$ times broader than Hα for a given BH mass. More recently, Denney, (2012) found that there is a significant contribution from gas at larger radii to the C IV emission line, enhancing the profile at lower-velocity and leading to smaller FWHM or dispersion values. The ratio of the line widths is therefore predicted to be lower than the factor of $\simeq 1.4$.

The Hα and C IV FWHM of the 77 quasars with C IV blueshifts $<1200 \,\text{km\,s}^{-1}$ are linearly correlated, as expected if the dynamics of the BLR clouds are dominated by virial motions. The median C IV/Hα FWHM ratio is 0.97 with standard deviation 0.31. Thus, as predicted by considering the contribution from low-velocity gas at large radii, the FWHM-based comparison results in a systematically lower median C IV/Hα.

As a direct consequence of the empirically small C IV/Hα FWHM ratio, the C IV-derived BH mass estimates are systematically lower than the corresponding Hα-derived masses when the blueshift is small. This can be seen in Fig 3.12, where for almost every quasar with a C IV blueshift $<1200 \,\text{km\,s}^{-1}$, the C IV-derived BH mass is smaller

than the corresponding Hα-derived mass. The median fractional difference between the two estimates is 0.60.

3.5   PRACTICAL APPLICATION OF THE C IV-BASED CORRECTION TO VIRIAL BH-MASS ESTIMATES

3.5.1   *Recipe for unbiased C IV based BH masses*

3.5.1.1   *Measuring the systemic redshift*

Equations 4 and 6 together provide an un-biased estimate of the virial BH mass given the FWHM and blueshift of C IV, together with the continuum luminosity at 1350 Å. The FWHM is readily obtained, either directly from the data, or, via the fitting of a parametric model to the C IV-emission line. The blueshift – defined as the bisector of the cumulative line flux – is also straightforward to measure and our preferred procedure is described in Section 3.5.1.2. The only potential complication arises in establishing the quasar systemic redshift and hence defining the zero-point for the C IV-blueshift measurement, since both the blueshift and the systemic redshift cannot be determined from C IV alone. In practice, when rest-frame optical lines are accessible, as is the case for the quasar sample here, an accurate systemic redshift can be obtained. The [O III] doublet and the Balmer lines all have velocity centroids very close to systemic, and the same is true for the broad Mg II doublet. For quasars at very high redshifts, $z \sim 6$, systemic redshifts can also be derived using the [C II] 158 μm emission in the sub-millimetre band (e.g. Venemans et al., 2016). However, in general, for example in determining the BH-masses of quasars at redshifts $z > 2$, if only the rest-frame ultraviolet region is available determining a reliable systemic redshift is non-trivial.

The SDSS DR7 pipeline redshifts are not sufficiently reliable to measure the C IV blueshift accurately because, in part, the C IV emission line itself contributes to the determination of the quasar redshifts. This is demonstated in Fig. 3.16a, in which we plot the C IV-blueshift versus C IV-emission equivalent width (EW) using the SDSS pipeline redshifts and the blueshifts calculated by Shen et al., (2011). A strong trend in the blueshift values as a function of line EW is not evident in Fig. 3.16a; structure in the parameter space is being masked because the C IV emission line is itself being used in the determination of the quasar redshifts.

The redshift-determination scheme of Hewett and Wild, (2010) provided much improved redshifts, not least because the redshift estimates for the majority of quasars were derived using emission-lines other than the C IV-line itself. Figure 3.16b shows SDSS DR7 quasars in the same C IV parameter space as Figure 3.16a, but now using Hewett and Wild, (2010) redshifts. The improved redshift estimates
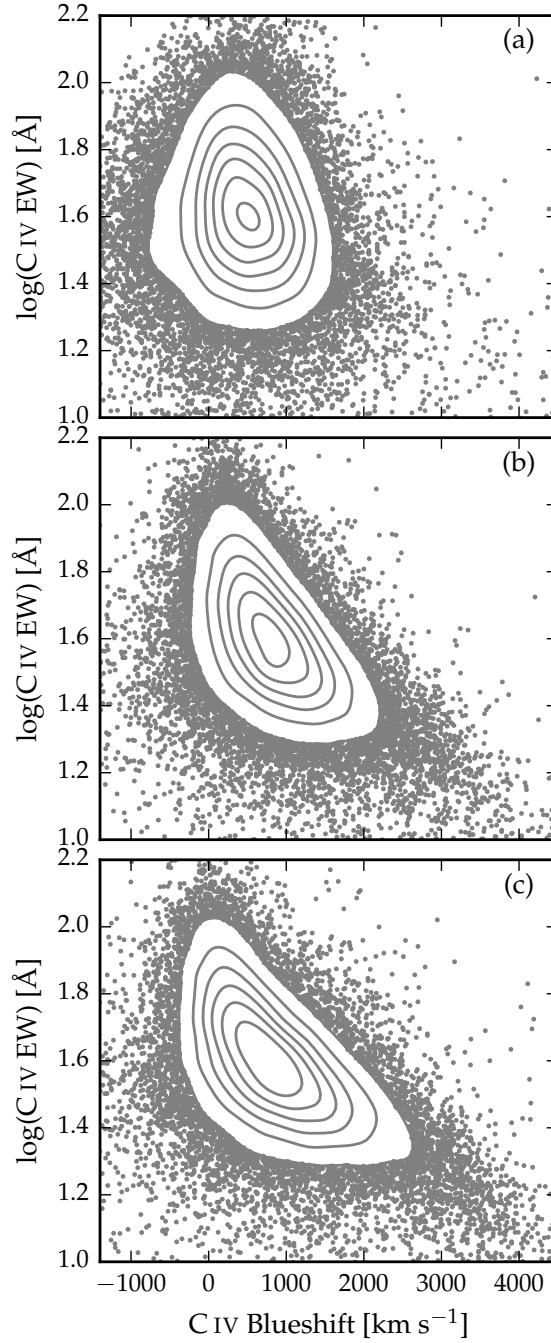
Figure 3.16: Rest-frame EW versus blueshift of the broad C IV-emission line for 32,157 SDSS DR7 quasars at $1.6 < z < 3.0$. Panel (a) uses C IV line parameters from Shen et al., (2011) and SDSS pipeline systemic redshifts. Panels (b) and (c) use systemic redshifts from Hewett and Wild, (2010) and Allen & Hewett (2017, in preparation) respectively, and C IV line measurements described in Sec. 3.5.1.2. In regions of high point-density, contours show equally-spaced lines of constant probability density generated using a Gaussian kernel-density estimator.

Table 3.4: The fractional error on the corrected BH mass as a function of
C IV blueshift for different uncertainties in the quasar systemic
redshift.

δv (km s^{-1})	C IV blueshift (km s^{-1})			
	0	1000	2000	4000
250	0.33	0.20	0.14	0.09
500	0.65	0.39	0.28	0.18
1000	1.30	0.79	0.57	0.36

are predominantly responsible for the differences seen in Fig. 3.16a
and b; the appearance in Fig. 3.16b of the extension to high blueshift
for quasars with low C IV EW is particularly evident.

Shen et al., (2016) and our own work shows that there is an intrinsic
variation of $\sigma \simeq 220$km s^{-1} in the velocity centroids of the broad-line re-
gion relative to a systemic-frame defined by the quasar narrow-line
regions. The redshifts for quasars in the SDSS DR10 and DR12 cat-
alogues (Pâris et al., 2014; Pâris et al., 2017) possess errors of $\simeq 500$-
750km s^{-1}(Pâris et al., 2012; Font-Ribera et al., 2013). The impact of
low spectrum S/N for fainter quasars in all the SDSS data releases
increases the uncertainty further. Table 3.4 includes the values for the
fractional error in the corrected BH-mass that result from a given er-
ror in the determination of the systemic rest-frame. For example, the
fractional error in the corrected BH mass is 0.39 for a quasar with a
1000km s^{-1} C IV blueshift when there is a 500km s^{-1} uncertainty in
the quasar systemic redshift.

Of potentially more significance for studies of BH-masses as a func-
tion of quasar and host-galaxy properties are redshift errors that de-
pend on the form of the quasar ultraviolet SED. The large systematic
variation in the C IV emission-line profile within the population is
evident from figures 11 and 12 of Richards et al., (2011). The plots
and analysis in Richards et al., (2011) employ the quasar redshifts
from Hewett and Wild, (2010) but, as is evident from the figures,
the systematic variation in the C IV shape is correlated with changes
in the quasar SEDs, including the strengths of the Si IIIλ1892 and
C IIIλ1908 emission lines in the rest-frame ultraviolet. As a conse-
quence, the redshifts from Hewett and Wild, (2010) still suffer from
systematic errors that are correlated with the shape, and particularly
the blueshift, of the C IV emission line. For the Hewett and Wild,
(2010) redshifts, and ultraviolet emission-line based redshifts in gen-
eral, quasars with large C IV EW and modest blueshifts have relatively
small ($\simeq 300$km s^{-1}) SED-dependent redshift errors. Redshift uncer-
tainties as large as $\simeq 1000$km s^{-1}for such quasars are unusual and the
large relative error in the corrected C IV BH-mass given in Table 3.4 is
pessimistic.

Conversely, systematic redshift errors are greatest for quasars with large blueshifts, reaching ~750km s^{-1} in the extreme for the Hewett and Wild, (2010) values. The associated error in the corrected C IV BH-masses is, however, mitigated somewhat due to the smaller gradient of the MBH(C IV)/MBH(Balmer) relation at large C IV blueshift (see Figs. 3.12 and 3.13). A definitive quantification of any systematic SED-dependent errors present in the quasar redshifts contained in the SDSS DR12 catalogue is not yet available but the principal component analysis (PCA) based redshift estimates are expected to be largely free of SED-dependent systematics.

Using published redshift estimates, notably those from Hewett and Wild, (2010) for the SDSS DR7 quasars and the BOSS PCA-based redshifts from Pâris et al., (2017) for SDSS DR12, the correction formula given in Section 3.4.3 produces significant improvements to C IV-based BH mass estimates. In a forthcoming work, Allen & Hewett (in preparation) will present a new redshift-estimation algorithm that produces redshifts independent of the C IV blueshift and other variations in the ultraviolet SEDs of luminous quasars. The low-ionization emission lines visible in the rest-frame ultraviolet (over wavelengths from Mg II$\lambda\lambda$2796,2803 down to the O Iλ1304+Si IIλ1307 blend) using the new redshift-algorithm are located at rest-frame wavelengths in excellent agreement with the systemic redshift defined using the rest-frame narrow-line optical O III doublet and broad-line Hβ and Hα. SED-dependent systematic errors are below the apparent inherent dispersion of $\simeq$ 220km s^{-1} associated with broad emission line redshifts (Shen et al., 2016).

Figure 3.16c shows the C IV emission line parameters calculated using the Allen & Hewett redshift-estimation algorithm. The systematic trends seen in Fig. 3.16b, in particular the extension to high blueshift at low C IV EW, become more apparent in Fig. 3.16c, as expected from consideration of the known SED-related errors in the redshifts from Hewett and Wild, (2010). A population of quasars with only modest blueshifts and low EW is also apparently still present.

3.5.1.2 *C IV emission line blueshift measurements*

The differences in the distribution of C IV emission line properties seen in the three panels of Fig. 3.16 are due primarily to the change in the systemic redshift estimates. It is also necessary, however, to obtain a measure of the C IV emission line 'location' in order to calculate the blueshifts. When working with moderately-sized samples, parametric fits to the emission-line profile may be undertaken using careful mask-definition to minimise the effect of absorption features on the profiles used for the parametrization, and this is the approach we followed in Section 3.3.

Effective analysis of the tens of thousands of spectra from SDSS DR7, and now DR12, however, requires a more robust scheme to de-

termine a C IV-blueshift estimate that is not very sensitive to the range of S/N among the spectra or the presence of narrow absorption systems within the C IV-emission profile. Shen et al., (2011) provide a discussion (their section 3) of the factors that effect the measurement of broad emission lines in quasar spectra of modest S/N. Their careful analysis of the C IV emission properties employed the results of parametric fits of three Gaussians to the spectra. Our own experiments in quantifying the C IV emission properties of SDSS spectra showed that a simple non-parametric measure of the C IV emission location reduced the number of outliers significantly. Visual inspection of spectra demonstrated that the improvement is due primarily to the identification of, and interpolation over, associated and outflow absorption systems, which forms part of the non-parametric measurement scheme.

We therefore chose to use a non-parametric scheme to measure the blueshift of the C IV line, which we will now describe. A continuum is first defined as a power-law of wavelength, $f(\lambda) \propto \lambda^{-\alpha}$, with the slope, α, determined using the median[4] values of the flux in two continuum windows at 1445–1465 and 1700–1705Å (the same wavelengths as adopted by Shen et al., (2011)). The C IV emission line is taken to lie within the wavelength interval 1500-1600Å, a recipe that is commonly adopted (e.g. Shen et al., 2011; Denney et al., 2013). To reduce the impact of narrow absorption systems on the emission-line profile a 'pseudo continuum' is defined by applying a 41-pixel median filter to the quasar spectrum. Pixels within the C IV profile that lie more than 2σ below the pseudo-continuum are deemed to be affected by absorption and added to an 'absorber'-mask. Two pixels on either side of each such pixel are also included in the mask. For each masked pixel, the flux values in the spectrum are replaced by values from the pseudo-continuum. The wavelength that bisects the cumulative total line flux is recorded and the blueshift is defined in exactly the same way as in Section 3.3.

Allen & Hewett will publish improved redshifts for all quasars in the SDSS DR7 and DR12 catalgoues. At the same time we will publish catalogues of unbiased BH masses for both SDSS DR7 and DR12 based on the Allen & Hewett redshifts. The components from the mean-field independent component analysis (see Allen et al., 2013, for an application to astronomical spectra) used in the Allen & Hewett redshift algorithm will also be published. With these components, if a rest-frame ultraviolet spectrum is available, it will be straightforward to determine the systemic redshift, via a simple optimisation procedure, and hence calculate the C IV blueshift.

---

4 The median is used to improve the robustness of the continuum estimate from the relatively small wavelength intervals.

3.5.2 *Systematic trends in residuals*

The scatter about the best-fitting line in the C IV/Hα FWHM versus C IV-blueshift relation is ∼0.1 dex, an order of magnitude smaller than the size of the C IV-blueshift dependent systematic but, nevertheless, still significant. With a view to reducing the scatter further, we searched for measurable parameters which correlate with the scatter at fixed C IV blueshift, including the luminosity, redshift, [O III] equivalent width (EW), and Fe II EW. The only significant correlation we find is with the Hα FWHM (Fig. 3.17). Quasars with broad Hα lines tend to lie below the relation while quasars with narrow Hα tend to lie above it. One possibility is that this correlation is simply due to random scatter (either intrinsic or measurement error) in the Hα FWHM which, with the other quasar properties fixed, would naturally produce a correlation between FWHM(C IV)/FWHM(Hα) and FWHM(Hα). However, the fact that we see no such correlation between the model residuals and the C IV FWHM suggests that the Hα FWHM correlation could be revealing somthing more fundamental. The Hα/Hβ FWHM is part of 'eigenvector 1' (EV1), the first eigenvector in a principal component analysis which originated from the work of Boroson and Green, (1992). While a number of parameters have been considered within the EV1 context (e.g. Brotherton and Francis, 1999), Fig. 3.17 suggests that part of the scatter between the Balmer and C IV velocity widths might be attributed to differences in the spectral properties which are correlated with EV1 (Marziani et al., 2013).

The shape of the line can be characterised by the ratio (FWHM/σ, where σ is the dispersion, derived from the second moment velocity; e.g. Kollatschny and Zetzl, 2011; Kollatschny and Zetzl, 2013). FWHM/σ ≃ 2.35 for a Gaussian profile, while FWHM/σ ≃ 1 for a peakier Lorentzian profile[5]. In our sample, we find the residuals and the Hα FWHM correlate with the shape of the line. The narrow lines are, on average, 'peakier' (with FWHM/σ ≃ 1) than the broader lines (with FWHM/σ ≃ 2). The origin of the Balmer-line shape correlation is not clear but one possibility is an orientation-dependence of the Hα FWHM (e.g. Shen and Ho, 2014). In this scenario quasars with broader emission lines are more likely to be in an edge-on orientation relative to our line of sight.

I think I mentioned this earlier so okay to delete there.

At radio wavelengths, the morphology of the radio structure, parametrized in terms of 'core dominance' is believed, at least in a statistical sense, to be a proxy for the orientation of the accretion disk (e.g. Jackson and Browne, 1991). We matched our sample to the FIRST radio catalogue (White et al., 1997) in an attempt

---

5 Strictly FWHM/σ → 0 for a Lorentzian profile, but values close to unity are typical when the dispersion is calculated over a velocity range, ≃ ±10 000km s^{-1}, used to parametrize broad emission lines in quasar spectra.
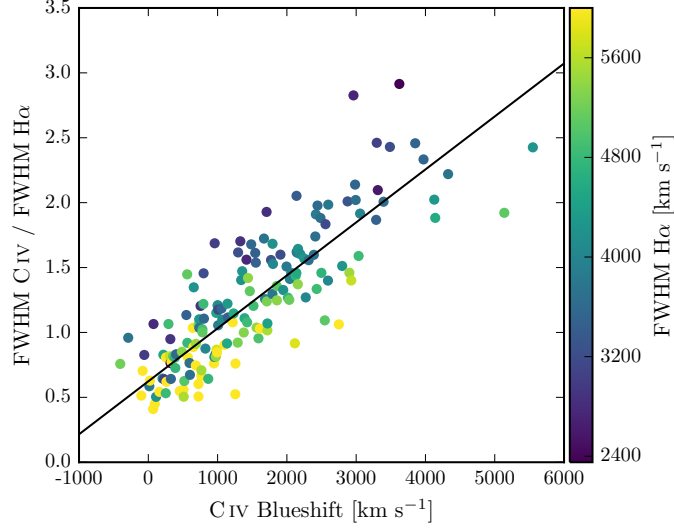
Figure 3.17: Same as Fig. 3.12a, with the marker colour representing the Hα FWHM. At fixed C IV blueshift, there is a clear Hα FWHM dependent systematic in the model residuals.

to identify orientation-dependent signatures. Following Shen et al., (2011), we classified quasars with matches within 5 arcseconds as core-dominated, while, if multiple matches were found within 30 arcseconds, quasars were classified as lobe-dominated. Twenty core-quasars and six lobe-dominated quasars resulted but no statistically significant differences in the Hα line-widths of the two samples were found. It should be noted that the sub-sample of radio-detected quasars is small and the effectiveness of the test is further compromised by the lack of radio-detected quasars at large blueshifts (see figure 14 of Richards et al., 2011, for example).

There are currently very few reverberation-mapping measurements of quasars with large C IV blueshifts. Looking to the future, the results of the large on-going statistical reverberation mapping projects (e.g. Shen et al., 2015; King et al., 2015) for luminous quasars at high-redshift will shed new light on the Balmer line emitting region of the BLR for quasars with a range of C IV blueshifts and lead to a greater understanding of the relation between the Balmer line profile and the BH mass.

3.5.3 *Effectiveness of the C IV blueshift based correction to BH masses*

Figure 3.18 demonstrates that our sample has an excellent coverage of the EW-blueshift parameter space in relation to SDSS DR7 quasars at redshifts $1.6 < z < 3.0$. The systematic offset to higher C IV blueshifts for our catalogue relative to the SDSS quasars as a whole is a result of the higher mean luminosity relative to the SDSS sample (Fig. 2.1). Our
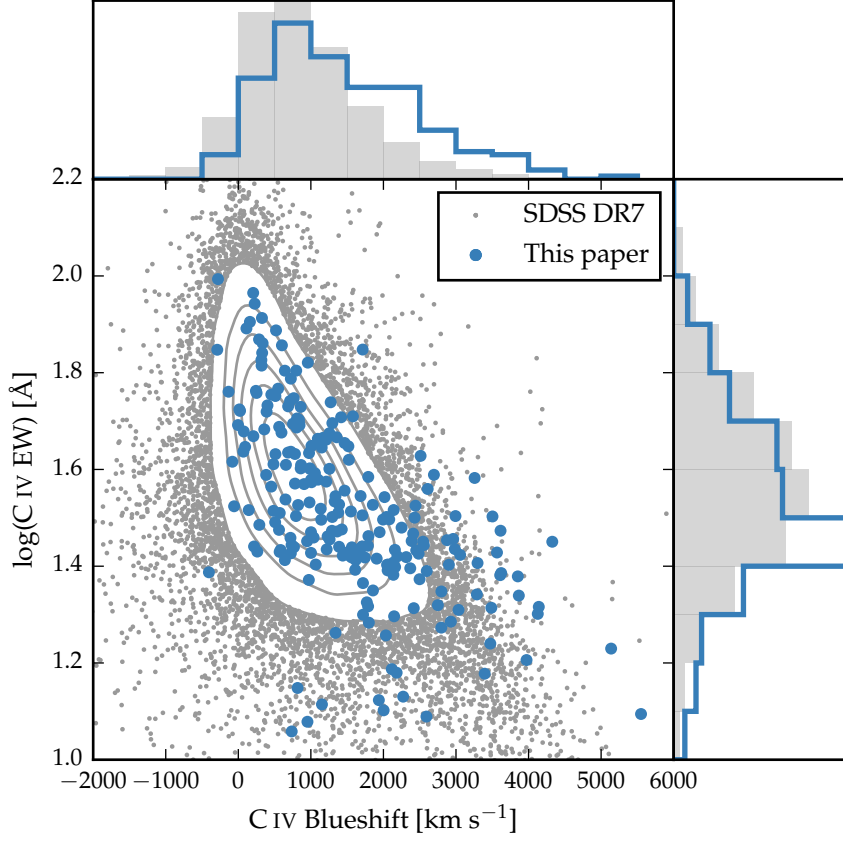
Figure 3.18: Rest-frame EW versus blueshift of the broad C IV-emission line for 32,157 SDSS DR7 quasars at $1.6 < z < 3.0$ (*grey*) and our sample (*blue*). For the SDSS quasars, the systemic redshifts used to calculate the blueshifts are from Hewett and Wild, (2010) and C IV emission properties are decribed in Paper I. In regions of high point-density, contours show equally-spaced lines of constant probability density generated using a Gaussian kernel-density estimator. Our sample has very good coverage; the shift to high blueshifts is a result of the high luminosity of our sample in relation to the SDSS sample and the correlation between luminosity and blueshift.

sample includes 21 quasars with C IV blueshifts $>3000 \mathrm{km\,s}^{-1}$, and extends to $\sim 5000 \mathrm{km\,s}^{-1}$, i.e. at the very extreme of what is observed in this redshift and luminosity range. Our investigation thus demonstrates that the C IV-blueshift based correction derived in this chapter is applicable to very high blueshifts. Conversely, there are no quasars in our catalogue with C IV blueshifts $\lesssim 0 \mathrm{km\,s}^{-1}$ and we caution against extrapolating the correction formula to negative blueshifts.

Figure 3.19 compares the C IV- and Hα-based BH masses before and after applying the blueshift-based correction to the C IV FWHM. Before the correction, the correlation between the C IV- and Hα-based BH masses is very weak, and the scatter between the masses is 0.4 dex. After correcting the C IV FWHM for the non-virial contribution, the correlation improves dramatically. The scatter between the corrected C IV-based masses and the Hα-based masses is reduced to 0.2 dex. The scatter is 0.24 dex at low C IV blueshifts ($\sim 0 \mathrm{km\,s}^{-1}$) and 0.10 dex at high blueshifts ($\sim 3000 \mathrm{km\,s}^{-1}$).

There has been a considerable amount of attention regarding the relative merits of using the FWHM or dispersion to characterise the velocity width (e.g. Denney et al., 2013). The existence of a trend in the C IV-dispersion values with C IV blueshift is evident from inspection of the bottom left panel of Fig. 3.7 but the systematic trend relative to the spread at fixed blueshift is significantly smaller than when using C IV FWHM. Therefore, without the blueshift information, using the line dispersion would yield a more accurate BH mass than the FWHM (Fig. 3.20).

The correlation between the Hα and C IV line dispersion is, however, weak. The Pearson coefficient for the correlation is 0.36 (and just 0.15 when the Hβ measurements are used in place of Hα). Furthermore, there is little dynamic range in the line dispersion: the scatter is just 480 and 460 $\mathrm{km\,s}^{-1}$ for Hα and C IV respectively. The observation suggests that the line dispersion does not fully trace the dynamic range in BH mass present in the quasar population. At least part of the reason is that the line dispersion is difficult to measure reliably in current survey-quality data, particularly because of the sensitivity to flux ascribed to the wings of the emission line (e.g. Mejía-Restrepo et al., 2016). Figures 3.19 and 3.20 demonstrate that estimating a reliable BH mass from the C IV FWHM and blueshift line is substantially more effective than using the C IV line dispersion with, or without, the line blueshift.

3.5.4   *Comparison to previous prescriptions*

In Fig. 3.21 we compare the C IV blueshift-based correction presented in this chapter to various prescriptions which have been proposed in the literature to derive BH masses from the C IV line which are consistent with the masses derived from the Balmer lines. In each
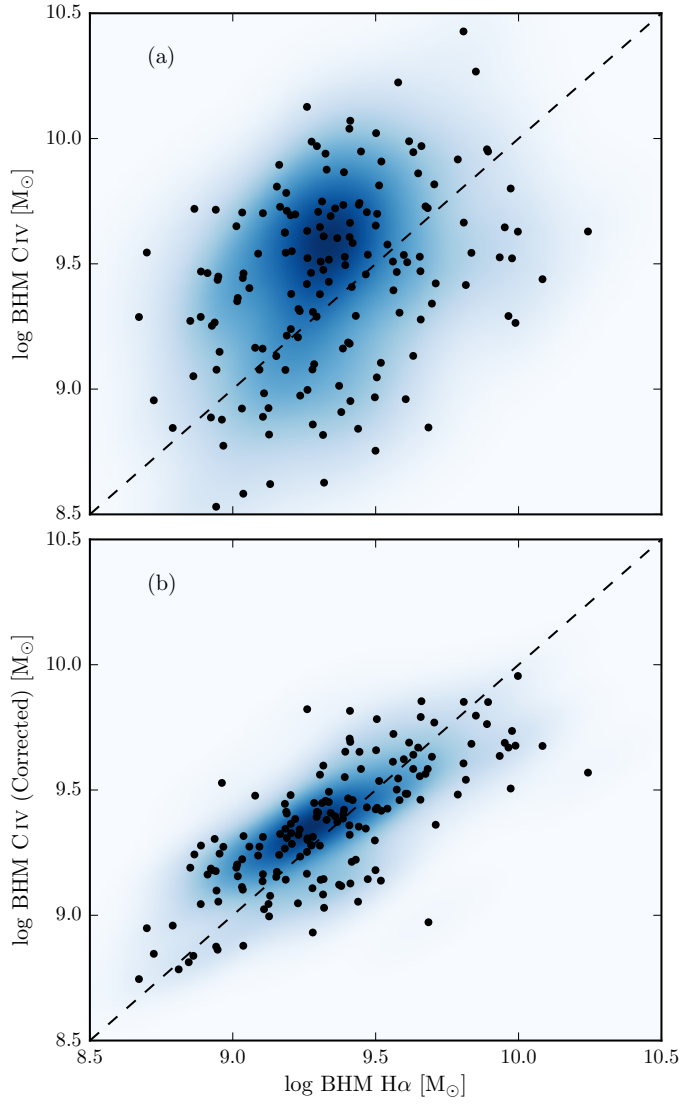
Figure 3.19: Comparison of the C IV- and Hα-based BH masses before (a) and after (b) applying the C IV blueshift-based correction to the C IV FWHM. The density of the plotted points (estimated using a Gaussian kernel density estimator) is represented by the colour. The correction to the C IV BH masses decreases the scatter by from 0.4 to 0.2 dex. Should definitely include some empirical validation of ICA redshifts since that is what we are telling people to use.
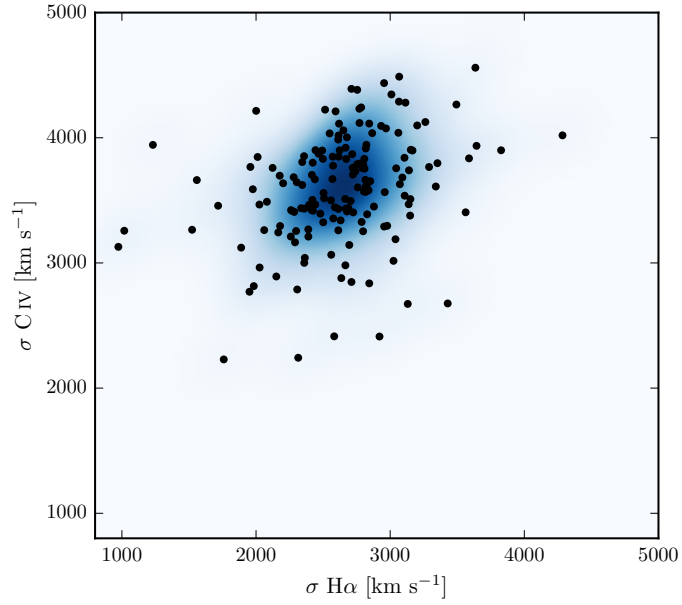
Figure 3.20: Comparison of the C IV and Hα line dispersion, σ. The density of the plotted points (estimated using a Gaussian kernel density estimator) is represented by the colour. Estimating a reliable BH mass from the C IV FWHM and blueshift line is substantially more effective than using the C IV line dispersion with, or without, the line blueshift. The C IV dispersion values are larger than the corresponding Hα measurements by a factor of 1.4 on average, which is consistent with reverberation mapping measurements (Vestergaard and Peterson, 2006).
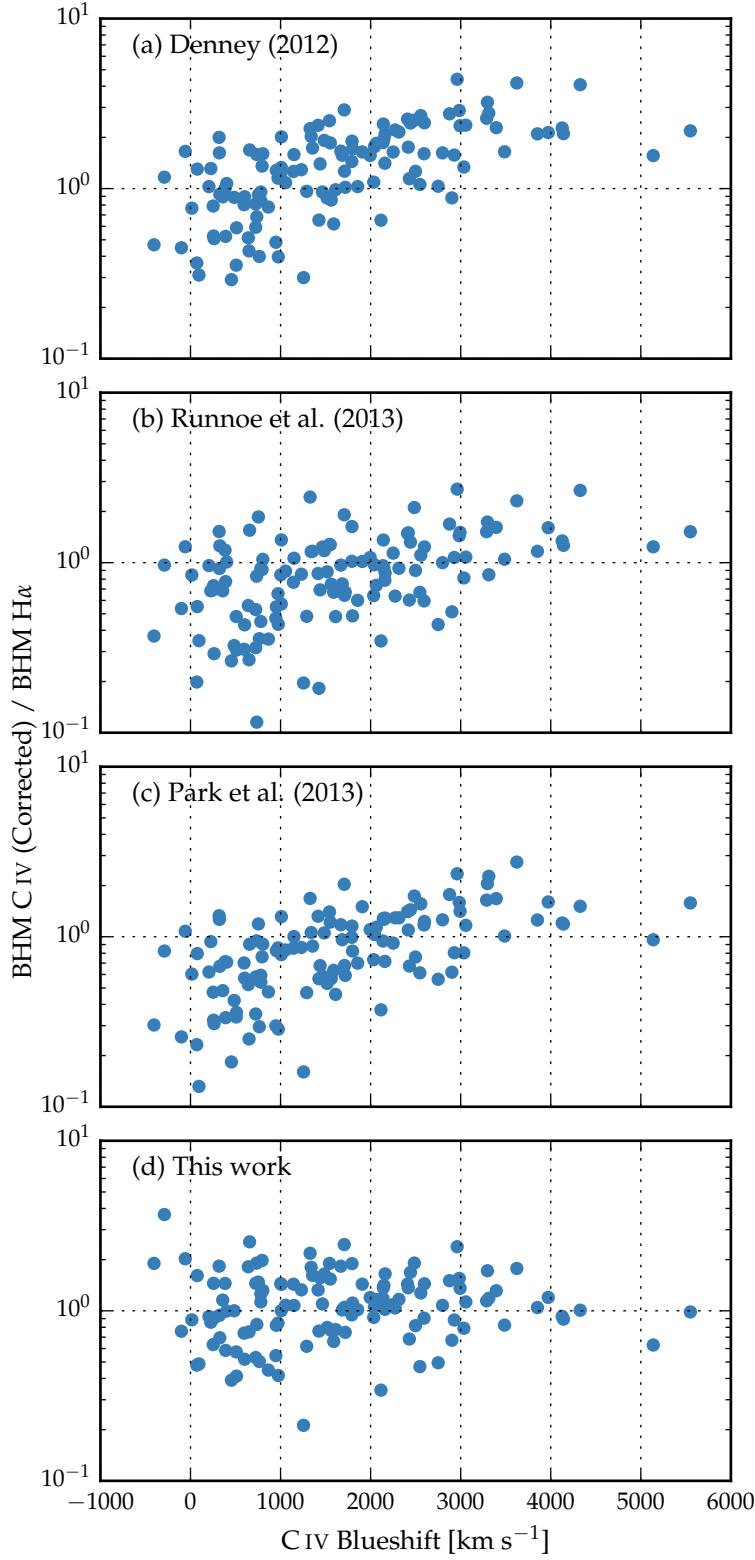
Figure 3.21: Comparison of BH mass estimates derived from C IV and Hα as a function of the C IV blueshift. Corrections to the C IV-based masses have been applied based on the shape (FWHM/σ) of the C IV emission line (a; Denney, 2012), the peak flux ratio of the Si IV+O IV blend relative to C IV (b; Runnoe et al., 2013a), by significantly reducing the dependence of the derived BH mass on the C IV velocity-width (c; Park et al., 2013), and based on the C IV blueshift (d; this chapter).

case we compare the corrected C IV-based masses to the Hα-based masses as a function of the C IV blueshift. The correction proposed by Runnoe et al., (2013a) is based on the spectral region at rest-frame wavelengths of ~1400 Å (see below). Therefore, our analysis is based on the 123 quasars with spectra covering this region.

In Fig 3.21a the C IV BH masses have been corrected using the C IV shape (FWHM/σ) based correction proposed by Denney, (2012). Denney, (2012) found the level of contamination in single-epoch spectra from non-reverberating gas to be correlated with the shape (FWHM/σ) of the C IV profile. In our sample, we observe a strong correlation between the shape of the C IV line and its blueshift (Fig. 3.7); between the two extremes in the C IV blueshift distribution the line shape changes from FWHM/σ ~1 to 2.5. The investigation of Denney, (2012) was based on a sample of reverberation mapped quasars, which have a narrow range of C IV-emission line shapes, including the absence of any objects with large C IV blueshifts. The correction is not applicable at large C IV blueshifts. Therefore, while the consistency between the Hα- and C IV-based masses at low C IV blueshifts is improved, at high C IV blueshifts the C IV-based masses remain seriously overestimated.

As explained above, reliably measuring the quasar systemic redshift from the UV region of the spectrum has proved difficult. However, the situation is improved dramatically by the new scheme developed by Allen & Hewett (2017, in preparation). Given the difficulty of measuring reliable C IV blueshifts without the Allen & Hewett scheme, Runnoe et al., (2013a) opted instead to use the continuum-subtracted peak flux ratio of the ultraviolet emission-line blend of Si IV+O IV (at 1400 Å) to that of C IV to correct for non-virial contributions to the C IV velocity width. This parameter was chosen because it showed the strongest correlation with the FWHM C IV/Hβ residuals, as well as with the strengths of optical O III and Fe II.

Following Runnoe et al., (2013a), we measure the peak flux by fitting a model with four Gaussian components (two for each emission line) to the continuum-subtracted flux. As is evident from Fig. 3.18, a correlation exists between the blueshift and equivalent width of C IV: C IV emission which is strongly blueshifted is typically weak. The Si IV+O IV emission-line blend, however, shows significantly less systematic variation. Therefore, the Si IV+O IV-based correction is quite effective in practice: the systematic bias in the C IV BH masses at large C IV blueshifts is reduced to a factor of ~ 2 (Fig. 3.21b). However, the C IV based masses are still systematically overestimated at large C IV blueshifts.

In contrast to the widely-used Vestergaard and Peterson, (2006) C IV-based virial BH mass calibration, the more recent Park et al., (2013) calibration significantly reduces the dependence of the derived masses on the emission-line velocity width (from the V^2 dependence

predicted assuming a virialized BLR to just $V^{0.56}$). As a consequence, the C IV based masses of the quasars with large C IV blueshifts are much reduced (Fig. 3.21c). However, the systematic error in the C IV-based BH masses as a function of C IV blueshift remains.

As a comparison, the C IV-based masses shown in Fig 3.21d have been corrected using to the C IV blueshift-based procedure presented in this chapter. No systematic in the BH masses as a function of the C IV blueshift is evident.

3.6 POPULATION TRENDS WITH C IV BLUESHIFT

As shown in Fig. 3.22, there are systematic variations in the Hα line profile as a function of the C IV blueshift. At C IV-blueshift $<1200 \, \mathrm{km \, s^{-1}}$, the Hα FWHM range is $\simeq 2000 - 8900 \, \mathrm{km \, s^{-1}}$, with mean $\simeq 4300 \, \mathrm{km \, s^{-1}}$. However, amongst the quasars with C IV-blueshift $>2000 \, \mathrm{km \, s^{-1}}$, the mean Hα FWHM=$3500 \, \mathrm{km \, s^{-1}}$, with a scatter of just $700 \, \mathrm{km \, s^{-1}}$. The apparent trend of peakier Hα-emission, with FWHM/σ close to unity, at large C IV-blueshift is enhanced by the modest increase in Hα EW with blueshift. Amongst the low-C IV-blueshift population there are in addition quasars with broader and more Gaussian-like Hα line profiles, with FWHM/σ $\simeq 2$.

The change in the Hα emission-line profiles as a function of C IV-blueshift means that the Hα-FWHM derived BH masses at high-blueshift are smaller than the sample mean. We transformed the observed luminosity into a mass-normalised accretion rate (Eddington ratio). To convert the monochromatic luminosity, which is observed, in to a bolometric luminosity we use the bolometric correction factor given by Richards et al., (2006a) ($L_{bol} = 9.26L_{5100}$). Although there is evidence that the bolometric correction factor is a function of the luminosity, as well as of other parameters including the C IV blueshift (Krawczyk et al., 2013), the differences are small over the parameter range covered by our sample, and for simplicity we adopt a constant factor.

The results, shown in Fig. 3.23, demonstrate that at large blueshifts quasars are accreting at around their Eddington limits (Fig. 3.23). This finding is in accord with our interpretation that the blueshifting of C IV is evidence for strong outflows resulting from the presence of a radiation-driven accretion-disc wind. Richards et al., (2002) found that quasars with large C IV blueshifts have weak He II. This is evidence for weak soft X-ray continuum emission (Leighly, 2004), which would allow a strong line-driven wind to form. The strength of such a wind is predicted to be related to the quasar far-ultraviolet SED, which, in turn, could be related to the mass-accretion rate.

All of the objects in our sample which exhibit large C IV blueshifts would be classified as population A in the Sulentic et al., (2000) scheme based on the Hα FWHM. Our results therefore support the
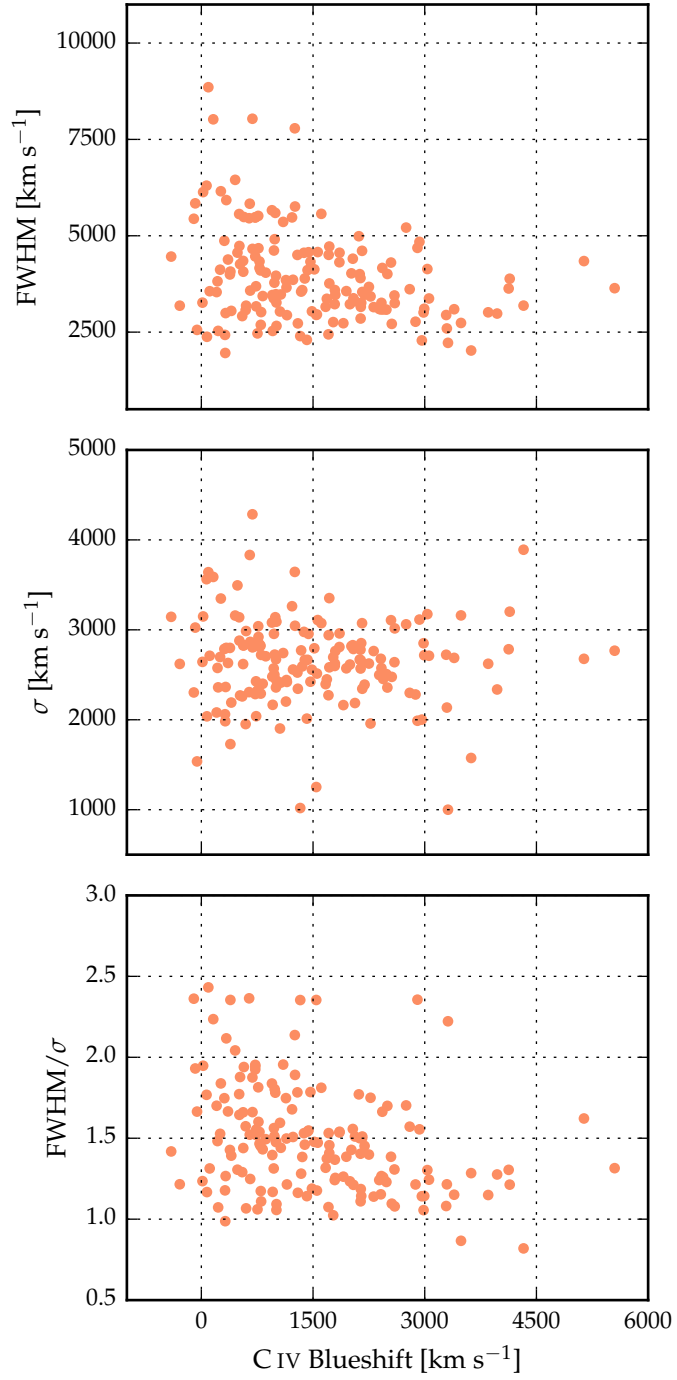
Figure 3.22: The FWHM, dispersion (σ) and shape (FWHM/σ) of Hα as a function of the C IV blueshift.
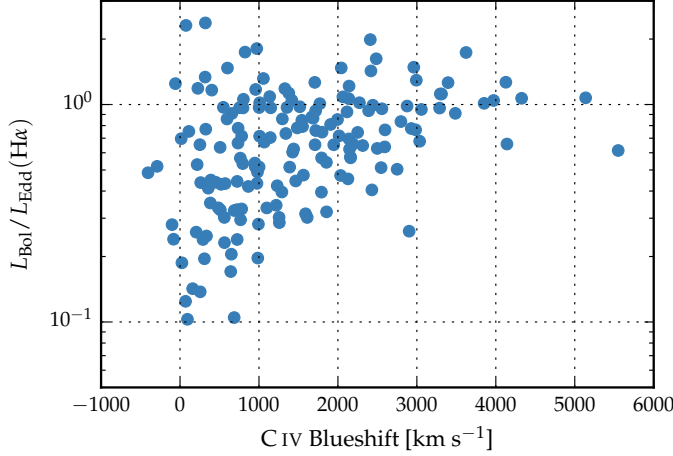
Figure 3.23: Hα-derived Eddington ratio versus C IV blueshift. At blueshift $\gtrsim 2000 \text{km s}^{-1}$ all quasars have high accretion rates (L/L$_{\text{Edd}} \simeq$ 1). This is in agreement with Kratzer and Richards, (2015), but in contrast to what one would derive from naive use of C IV-based BH mass scaling relations.

idea of the Sulentic et al., (2000) A/B division being driven by the Eddington ratio, with population A sources possessing higher accretion rates. However, we also observe a number of quasars which have high Eddington ratios but do not have line profiles suggestive of strong outflows in the C IV BLR. This suggests that a high accretion rate is a necessary but not sufficient condition for the existence of outflows (Baskin and Laor, 2005a).

The two-dimensional nature of the C IV emission line parametrization and the apparent anti-correlation between C IV EW and C IV blueshift suggests that the quasar population exhibits a continuum of properties. As such, more accurate C IV blueshift measurements for SDSS-quasars should allow an improved mapping between the C IV-emission properties and key physical parameters of the quasars. This includes improving our understanding of the origin of quasars with exceptionally weak, blueshifted C IV emission (weak emission line quasars; Luo et al., 2015) which could be exotic versions of wind-dominated quasars (Plotkin et al., 2015).

3.6.1 *Systematic biases in Balmer-based masses*

The interpretation described in the preceding section requires some caution since the emission-line shape (characterized by the value of FWHM/σ) of Hα is also changing as a function of the C IV blueshift (Fig. 3.22). At low C IV blueshifts there are a range of shapes, but all of the quasars exhibiting large C IV blueshifts have peaky Hα profiles with FWHM/σ ≃ 1. This raises the question of whether the

Hα FWHM is a reliable proxy for the virial-induced velocity dispersion for the full range of Hα line shapes we have in our sample.

When calibrating the virial-product to masses derived independently using the BH mass – stellar velocity dispersion ($M_\odot - \sigma$) relation, Collin et al., (2006) find that the scaling factor, f, is a factor ~ 2 larger for their Population '1' sources (with FWHM/$\sigma < 2.35$ and essentially equivalent to population A of Sulentic and co-workers and to the high-blueshift quasars here) than for their Population 2 (with FWHM/$\sigma > 2.35$). For single-epoch BH-mass estimates, assuming a constant value of f, as is normally done (e.g. Vestergaard and Peterson, 2006), means that Population 1 masses will be underestimated and Population 2 will be overestimated. In the context of this result from Collin et al., (2006), our high-blueshift objects all possess peaky Hα-lines and, while our quasar sample probes much higher luminosities and masses, the true BH-masses may also be underestimated. Adopting such an interpretation, the amplitude of the trend seen in Figs. 3.12 and 3.13 might not be so pronounced.

As mentioned in Section 1 and discussed in Richards et al., (2011), quasars with current reverberation mapping measurements have a restricted range of C IV-line shapes. There are currently very few reverberation-mapping measurements of quasars with large C IV blueshifts but the results of the large on-going statistical reverberation mapping projects (e.g. Shen et al., 2015) for luminous quasars at high-redshift will go some way to establishing whether the quasar broad line regions producing Balmer emission look the same for objects with very different C IV-emission blueshifts.

Although the EV1-trends (Sulentic et al., 2000; Shen and Ho, 2014) are most likely driven by the accretion rate, orientation may also have a role to play in determining the observed properties of the BLR. Shen and Ho, (2014) argue that a large part of the scatter observed in the Hβ FWHM relates not to a spread in BH masses, but rather to the orientation of the BLR relative to the line-of-sight. For this to be true, the BLR would need to be in a flattened disc-like geometry, in which case the observed line width would increase with the inclination of the disc relative to the line of sight. Brotherton, Singh, and Runnoe, (2015) found that the core-dominance of radio-loud quasars, which is believed to be a reliable proxy for orientation, at least in a statistical sense, is significantly correlated with the Hβ FWHM and hence with the BH-mass estimates. This raises the question of whether the narrow Hα emission lines observed in the quasars with the largest C IV blueshifts could be an orientation effect. However, there is no evidence that the C IV blueshift is dependent on the orientation (inferred from the radio core-dominance; Richards et al., 2011; Runnoe et al., 2014). Furthermore, Leighly, (2004) showed that the He IIλ1640 emission-line properties of quasars with large C IV blueshifts are more consistent with differences in the SED rather than differences in the

orientation. Collin et al., (2006) showed that orientation effects were also sub-dominant to the Eddington ratio in determining the shape of the Hβ line and the Hα line shape trend we observe is consistent with the finding of Marziani et al., (2003) that the Hβ emission profiles of high/low Eddington ratio low-z quasars and type 1 Seyfert nuclei are well fit by Lorentzian and double Gaussian profiles respectively. Overall, therefore, orientation does not appear to be the dominant effect in determining the C IV blueshift and correlated changes in the Hα line profile.

3.6.2 *The BAL parent population*

Classical high-ionization BAL (HiBAL) quasars are also predominantly Population A objects in the scheme of Sulentic et al., (2000). There are no HiBAL quasars in our sample by design but it is generally accepted that quasars which show high-ionisation BALs are likely to be radiating with relatively high L/L_{Edd} (e.g. Zhang et al., 2014). We therefore propose that the subset of the quasar population that exhibits large C IV-emission blueshifts, with high-EW and narrow-Hα emission lines, may be directly related to the HiBAL quasar population – perhaps even the 'parent' population (Richards, 2006). A prediction of such a linkage is that near-infrared observations of the rest-frame optical spectra of HiBAL quasars will show strong, relatively narrow, Balmer emission lines, very similar to those of the quasars with high C IV-blueshifts presented in this chapter (see Runnoe et al., 2013b, for such a study).

Do I actually say BALs are removed from sample?

I must have a few BALs. Make a composite spectra of Balmer lines of BALs?

3.6.3 *The frequency of quasars with high accretion rates*

Quantifying the frequency of quasars producing outflows as a function of key parameters, e.g. quasar luminosity, BH-mass, redshift,... will be important to constrain models of quasar-galaxy evolution. At fixed BH mass, the intrinsic and the observed fraction of quasars exhibiting properties that depend on the Eddington ratio can differ significantly. As an illustration, we consider the implications for the intrinsic fraction of quasars possessing large C IV blueshifts given the observed numbers in the $m_i < 19.1$ flux-limited sub-sample of the SDSS DR7 quasar catalogue. In order to estimate the size of the selection effect, we considered the detection probability for a much-simplified quasar population. We assume that all quasars with C IV blueshifts $>1200\,\mathrm{km\,s^{-1}}$ have enhanced accretion rates relative to the 'normal' population (with C IV blueshifts $<1200\,\mathrm{km\,s^{-1}}$). If the accretion rate of the high-blueshift population is double the rate of the low-blueshift population (which is true in an average sense – see Fig. 3.23), then the high-blueshift population will be brighter by $\simeq 0.75$ magnitude. Under the assumption that the BH mass distribution is

independent of the C IV blueshift, the high-blueshift population will then be over-represented in a flux-limited sample. To estimate the size of the bias, we need to know how many more quasars, at redshifts $2 < z < 2.5$, there are with $m_i < 19.1 + 0.75 = 19.85$ relative to $m_i < 19.1$. This is the fraction of the population which, as a consequence of having enhanced accretion rates, are boosted above the survey flux limit. The main colour-selected SDSS DR7 quasar catalogue extends only to $m_i = 19.1$ and, assuming the luminosity function is continuous[6] we thus use the number counts at $m_i < 19.1$ and $m_i < 18.35$, which differ by a factor of $\simeq 4$.

At redshifts $2 < z < 2.5$, there are 3,834 quasars with C IV blueshifts $<1200\,\mathrm{km\,s^{-1}}$ and 2,484 with blueshifts $>1200\,\mathrm{km\,s^{-1}}$ in the SDSS DR7 $m_i < 19.1$ quasar sample, a ratio of ~2:1. The above calculation, although much idealised, suggests that the intrinsic fraction of high-blueshift quasars is a factor of four smaller than in the flux-limited sample (i.e. ~15 per cent of the ultraviolet-selected non-BAL quasar population).

3.7 CONCLUSIONS

The main results of this chapter are as follows:

- We have analysed the spectra of 230 high-luminosity ($10^{45.5} - 10^{48}\,\mathrm{erg\,s^{-1}}$), redshift $1.5 < z < 4.0$ quasars for which spectra of the Balmer emission lines and the C IV emission line exist. The large number of quasars in our spectroscopic catalogue and the wide range in C IV blueshifts the quasars possess has allowed us to directly investigate biases in C IV-based BH mass estimates which stem from non-virial contributions to the C IV emission as a function of the C IV blueshift, which, in turn, depends directly on the form of the quasar ultraviolet SEDs (Richards et al., 2011).

- The C IV emission-based BH-masses are systematically in error by a factor of more than five at $3000\mathrm{km\,s^{-1}}$ in C IV emission blueshift and the overestimate of the BH-masses reaches a factor of 10 for quasars exhibiting the most extreme blueshifts, $\gtrsim 5000\mathrm{km\,s^{-1}}$.

- We have derived an empirical correction formula for BH-mass estimates based on the C IV emission line FWHM and blueshift. The correction may be applied using equations 4 and 6 in Section 3.4.3. The large SED-dependent systematic error in C IV-based BH-masses is removed using the correction formulae. The remaining scatter between the corrected C IV-based masses

---

6 The luminosity function and number-counts vary only smoothly (e.g. Ross et al., 2013) for the magnitude and redshift range used here.

and the Hα-based masses is 0.24 dex at low C IV blueshifts ($\sim$0km s^{-1}) and 0.10 dex at high blueshifts ($\sim$3000km s^{-1}). This is a significant improvement on the 0.40 dex scatter observed between the un-corrected C IV and Hα BH masses. The correction depends only on the C IV line properties - i.e. the FWHM and blueshift - and allows single-epoch virial BH mass estimates to be made from optical spectra, such as those provided by the SDSS, out to redshifts exceeding $z \sim 5$.

3.8 FUTURE WORK

- Clustering

- Data-driven mapping

4

NARROW LINE REGION PROPERTIES

4.1 INTRODUCTION

AGN are very efficient at driving outflows; X-ray and UV spectroscopy reveal high velocity outflows to be nearly ubiquituous on sub-parsec scales in high accretion rate quasars. In recent years, a huge amount of resources have been devoted to searching for observational evidence of galaxy-wide, AGN-driven ouflows, which could provide the feedback mechanism nessecary to quench star formation in massive galaxies. This has resulted in recent detections of winds in AGN-host galaxies using tracers of atomic, molecular, and ionised gas (e.g. Nesvadba et al., 2006; Arav et al., 2008; Nesvadba et al., 2008; Moe et al., 2009; Dunn et al., 2010; Alexander et al., 2010; Harrison et al., 2012; Harrison et al., 2014; Nesvadba et al., 2010; Rupke and Veilleux, 2013; Veilleux et al., 2013; Nardini et al., 2015; Feruglio et al., 2010; Alatalo et al., 2011; Cimatti et al., 2013; Cicone et al., 2014).

One particularly successful technique has been observations of forbiden emission lines, which trace warm (T$\sim 10^4$K) ionised gas in the narrow line region (NLR). Because of its high equivalent width, [O III]λ5008 is most studied of the narrow quasar emission lines. In general, the [O III] emission appears to consist of two components: a narrow, 'core' component, with a velocity close to the systemic redshift of the host galaxy, and a broader 'wing' component, which is normally blueshifted. The general consensus is that the core component traces the gravitational potential of the host galaxy, as the width correlates well with the stellar velocity dispersion. On the other hand, the wing is tracing outflowing gas. This can be explained if the far-side of any outflowing gas, that is moving away from the line of sight, is obscured by dust in the host galaxies (e.g. Heckman et al., 1981; Vrtilek, 1985).

Observations of broad velocity-widths and asymmetries in narrow emission lines stretch back several decades (e.g. Weedman, 1970; Stockton, 1976; Heckman et al., 1981; Veron, 1981; Feldman et al., 1982; Heckman, Miley, and Green, 1984; Vrtilek, 1985; Whittle, 1985; Boroson and Green, 1992). However, the small sample sizes make it difficult to know how representative these observations are. More recently, the advent of large optical spectroscopic surveys (e.g. SDSS) have facilitated studies of the NLR in tens of thousands of AGN (e.g. Boroson, 2005; Greene and Ho, 2005a; Zhang et al., 2011; Mullaney et al., 2013; Zakamska and Greene, 2014; Shen and Ho, 2014). This has provided constraints on the prevalence of ionised outflows and,

by measuring outflow properties as a function of AGN properties, on the drivers of these outflows. At the same time, there is strong evidence from spatially resolved spectroscopic observations of kinematically disturbed gas extended over galaxy scales (e.g. Greene et al., 2009; Greene et al., 2011; Hainline et al., 2013; Harrison et al., 2012; Harrison et al., 2014).

However, these studies do not cover the redshift range when star formation and black hole accretion peaked, and consequently when feedback is predicted to be strongest. At these redshifts the bright optical emission lines are redshifted to near-infrared wavelengths, where observations are much more challenging compared to optical wavelengths. As a consequence, studies at high redshifts have typically relied on relatively small numbers of objects, which might not be representative of the properties of the population (e.g. Netzer et al., 2004; Sulentic et al., 2004; Shen, 2016). Other recent studies have looked at the [O III] emission properties of rare sub-samples - e.g. heavily obscured quasars (Zakamska et al., 2016) and the most luminous quasars (Bischetti et al., 2016). These studies often report exceptionally large [O III] widths, with FWHM $> 1000 \mathrm{km\,s^{-1}}$ (e.g. Netzer et al., 2004; Nesvadba et al., 2008; Kim et al., 2013; Brusa et al., 2015; Carniani et al., 2015; Perna et al., 2015; Bischetti et al., 2016). This could suggests that AGN efficiency in driving galaxy-wide outflows increases with AGN luminosity. In addition, [O III] is often very weak, or is missing entirely (e.g. Netzer et al., 2004).

In this chapter we analyse the [O III] properties of a sample of 356 high-luminosity, redshift $1.5 < z < 4$ quasars. The large sample size will help to put these observations in context of the AGN population as a whole. We will analyse the [O III] emission properties as a function of key properties of the quasar, e.g. BH mass, luminosity, and accretion rate.

4.2 QUASAR SAMPLE

We have assembled a catalogue of 356 high-luminosity, redshift $1.5 < z < 4$ quasars. These are selected from our near-infrared spectroscopic databse (Chapter 2) to have spectra covering the strong, narrow [O III] doublet. The broad Balmer Hβ line is also observed for all but two of the sample. In 165 the spectra extend to the broad Hα emission at 6565Å, and in 260 optical spectra including C IV are also available (mostly from SDSS/BOSS). This is the largest study of the narrow line region properties of high-z quasars ever undertaken. The quasar sample is summarised in Table 4.1.

Table 4.1: The numbers of quasars with [O III] line measurements and the spectrographs and telescopes used to obtain the near-infrared spectra.

Spectrograph	Telescope	Number
FIRE	MAGELLAN	31
GNIRS	GEMINI-N	28
ISAAC	VLT	9
LIRIS	WHT	7
NIRI	GEMINI-N	29
NIRSPEC	Keck II	3
SINFONI	VLT	80
SOFI	NTT	76
TRIPLESPEC	ARC-3.5m	27
TRIPLESPEC	P200	45
XSHOOTER	VLT	21
		356

4.3 PARAMETERIC MODEL FITS

In this section we describe how parameters of the [O III] emission are derived. Our approach is to first model the spectra, and then derive parameters of interest from the best-fitting model. This enhances the useful information that can be extracted from spectra with finite signal-to-noise (S/N).

Two different models are considered. The first consists of a power-law continuum, an empirical Fe II template and multiple Gaussian components to model the emission from the broad and narrow components of Hβ and the [O III] doublet. This is a model which is commonly adopted in the literature (e.g. Shen et al., 2011). The second model consists of six spectral components derived from an independent component analysis (ICA) of a large sample of low-redshift AGN with SDSS spectra covering the same spectral region. As we will demonstrate, a linear combination of these spectral components is able to reproduce the spectra around Hβ/[O III] to a high degree of precision.

4.3.1 *Model One: Multiple Gaussians*

The first step in our procedure is to fit a combination of a power-law continuum and an optical Fe II template – taken from Boroson and Green, (1992) – to two windows at 4435-4700 and 5100-5535 Å. The Fe II template is convolved with a Gaussian, and the width of this

Gaussian, along with the normalisation and velocity offset of the Fe II template, are free variables in the pseudo-continuum fit.

This requires the spectra to be transformed to within ~1000km s^{-1} of the quasar rest-frame. The redshift used in this transformation is either derived from the peak of the broad Hα emission (~ 40 per cent of our sample), from the peak of the broad Hβ emission (~ 40 per cent) or from the peak of the narrow [O III] emission (20 per cent). In later sections, emission line locations will always be quoted as relative velocities, and so do not explicitly depend on how the quasar rest-frame is defined at this stage in the fitting procedure.

Once the continuum and Fe II emission has been modelled and removed, the following model is fit in the wavelength interval 4700-5100 Å. The fit is done as a function of the Doppler velocity shift, and we adopt the wavelength 4862.721Å (the laboratory Hβ wavelength) to transform wavelengths into equivalent Doppler velocities.

In general, Hβ is modelled by two Gaussians with non-negative amplitudes and FWHM greater than 1200km s^{-1}. In 10 objects Hβ is modelled with a single Gaussian and in 41 objects Hβ is modelled with two Gaussians, but the velocity centroids of the two Gaussians are constrained to be equal. These spectra generally have low signal-to-noise (S/N), and adding extra freedom to the model does not significantly decrease the minimised reduced chi-squared. In addition there are cases where the blue wing is below the lower wavelength limit of the spectrograph; in these cases models with more freedom are insufficiently constrained by the data with limited wavelength converage.

Contributions to the Hβ emission from the narrow-line region is weak in the vast majority of our sample, and in general we do not include an additional Gaussian component to model this emission. In 9 objects features in the model - data residuals suggest that a narrow emission component is significant, and an additional narrow Gaussian is included for these quasars. It is likely that there is some not insignificant contribution from the narrow line region in other quasars. If this is the case then measures of the Hβ velocity width will be biased to lower values on average. However, in this chapter we are only interested in location of the peak of the Hβ emission (to infer the quasar redshift). This is unlikely to be biased by not decomposing the narrow and broad emission components.

Each component of the [O III] doublet is fit with one or two Gaussians, depending on the fractional reduced χ^2 difference between the one- and two-component models. If the addition of the second Gaussian decreases the reduced χ^2 by more than 5 per cent then the double-Gaussian model is accepted. One hundred and thirty-one are fit with a single Gaussian and 154 with two Gaussians. When a single Gaussian is used to model each line, the peak flux ratio of the [O III] 4960 Å and 5008Å components are fixed at the expected 1:3 ratio and

the width and velocity offsets are set to be equal. In the double Gaussian fit, the peak flux ratio of the second components is again fixed at 1:3, and the width and velocity offsets are again set to be equal.

In 71 objects [O III] is undetected, or is very low S/N. In these cases we do not attempt to measure the width of the [O III] emission, but instead fit a fixed [O III] template, only the overal normalisation of which is allowed to vary. The template is derived by running our fitting routine on a very high S/N composite spectra of low redshift AGN.

Composite details?

Model parameters were derived using a standard variance-weighted least-squares minimisation procedure employing the Levenberg-Marquardt algorithm. Prior to the fit, the spectra were inspected visually and regions significantly affected by telluric absorption or of low S/N were masked out. Some example fits are shown in Figure 4.1

4.3.2 *Derived parameters*

All [O III] line properties are derived from the [O III]5008 emission, but, as described above, the kinematics of the peak at 4960Å are constrained by our fitting routine to be identical.

We do not to attach any physical meaning to the individual Gaussian components used in the model. While it is true that in some quasars the [O III] emission can be clearly separated in to a narrow component at the systemic redshift and a lower-amplitude, blueshifted broad component (e.g. Shen, 2016), often this decomposition is highly uncertain and dependent on the spectral S/N, resolution etc. In addition, there is no theoretical justification that wing component should have a Gaussian profile.

We therefore choose to characterize the [O III] line profile using a number of non-parameteric measures, which are commonly used in the literature (e.g. Zakamska and Greene, 2014; Zakamska et al., 2016). A normalised cumulative velocity distibution is constructed from the best-fitting model, from which the velocities below which 5, 10, 25, 50, 75, 90, and 95 per cent of the total flux accumulates can be read off. The width of the emission line can then be defined, for example, using $w_{80} = v_{90} - v_{10}$. The absolute assymetry in the line profile A is defined as $((v_{95} - v_{50}) - (v_{50} - v_5))/(v_{95} - v_5)$ (Zakamska and Greene, 2014). We use the peak of the full [O III] profile to define the systemic redshift, and verify below that this is unbiased.

The line width measures are not corrected for instrumental broadening

4.3.3 *Absolute flux calibration of spectra and continuum luminosities*

Relative flux-calibration of the infrared spectra as a function of wavelength has been achieved, to $\simeq 10$ per cent, through observations of appropriate flux standards. The absolute flux levels, however, can be

Add outline of table of derived properties for this chapter
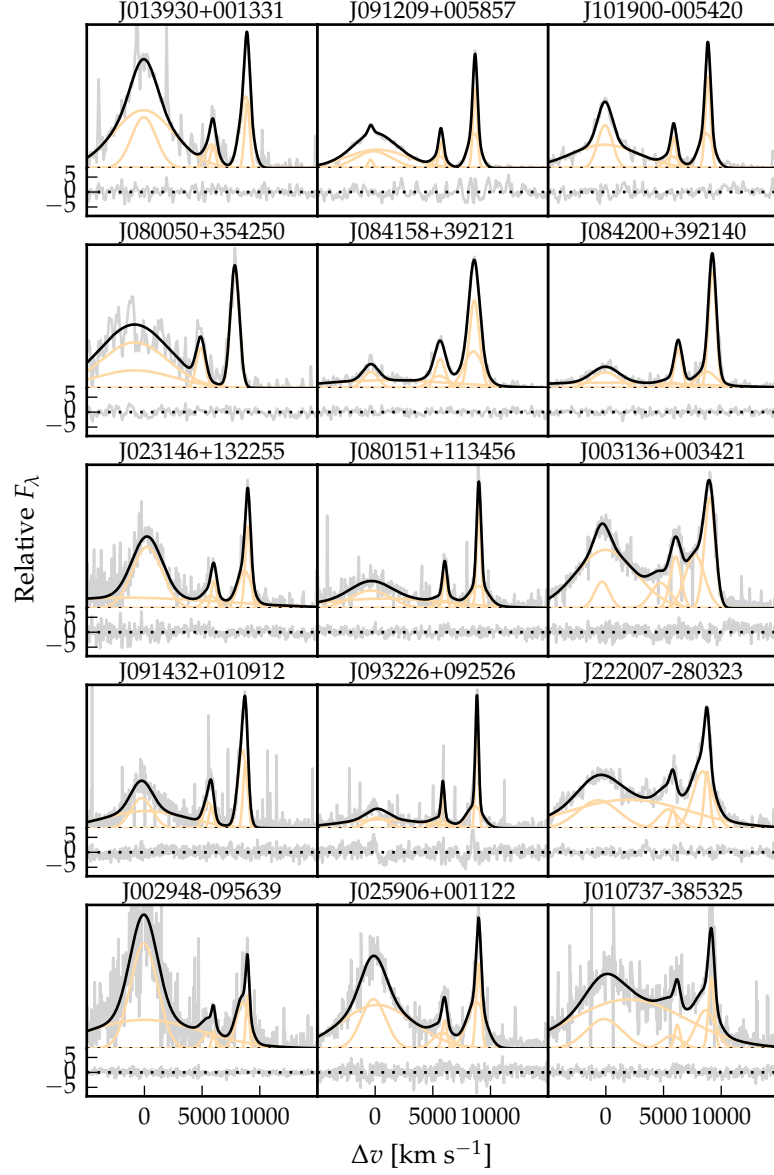
Figure 4.1: Multi-component Gaussian fits to the continuum-subtracted Hβ/[O III] emission in 12 quasars, chosen to be the representative of the wide range of [O III] line widths we measure in our sample. The data is shown in grey, the best-fitting parametric model in black, and the individual model components in orange. Below each fit we plot the data minus model residuals, scaled by the errors on the fluxes.

in error by large factors due to variable atmospheric conditions combined with the narrow slit widths. For the majority of the quasars we have, therefore, established the absolute flux scale for each near-infrared spectrum using the same quasar SED-model fitting scheme employed in Chapter 5. Briefly, the SED-model was fit, with the normalisation and E(B-V) as free varibales, to optical/infrared magnitudes, or SDSS/BOSS spectra (check order I do this.) This allows us to extrapolate from the optical when we do not have photometric data in the near-infrared. The spectra were then normalised to the SED model using a linear error-weighted least-squares regression in the the regions of the spectra covered by the H/K bands. The monochromatic continuum luminosity at 5100Åwas calculated directly from the normalised SED-model.

Check if any missing normalisation / monochromatic luminosities.

I think a lot of this is repeated from chapter 2/3

Check factor of (1+z) in luminosity calculation

4.3.4 *Reliability of derived parameters*

4.3.4.1 *Removal of Fe II emission*

While we were able to satisfactorily model the Fe II emission in the vast majority of cases, we encountered a number of cases where the relative strengths of the Fe II lines appear to differ significantly from those of I Zw 1 on which the Fe II template is based. As a result, significant Fe II flux remained in the spectrum after the removal process. This emission is at rest-frame wavlengths very close to the [O III] emission, and so could potentially lead to large errors in the inferred [O III] line parameters.

In Figure 4.2 we plot the spectral region around [O III] for 23 quasars we have identified where significant features remain following the subtraction of the continuum and Fe II. The vertical lines indicate the expected positions of the [O III] doublet, with zero velocity defined using the peak of the broad Hβ emission. [O III] is generally exteremely weak in these objects. As a result, the widely-adopted procedure of fitting multiple Gaussians will tend to fit the Fe II emission as broad, shifted [O III]. For example, J125141+080718 was studied by Shen, (2016), and assigned an extremely large [O III] blueshift. Our analysis suggests that this emission is more likely to be Fe II. Because of the difficulty measuring the [O III] properties of these objects, they are excluded from subsequent analysis.

4.3.4.2 *Low EQW [O III]*

In Figure 4.3 we show the distribution of the [O III] rest-frame EQW distribution for the 330 objects in our sample (objects where Fe II emission has been sub-optimally removed are excluded). In many objects [O III] is undetected. In others it is detected, but is too weak for its shape (i.e. the width and asymmetry) to be measured reliably. We define EQW= 8Åas the limit below which we can no longer reliably de-
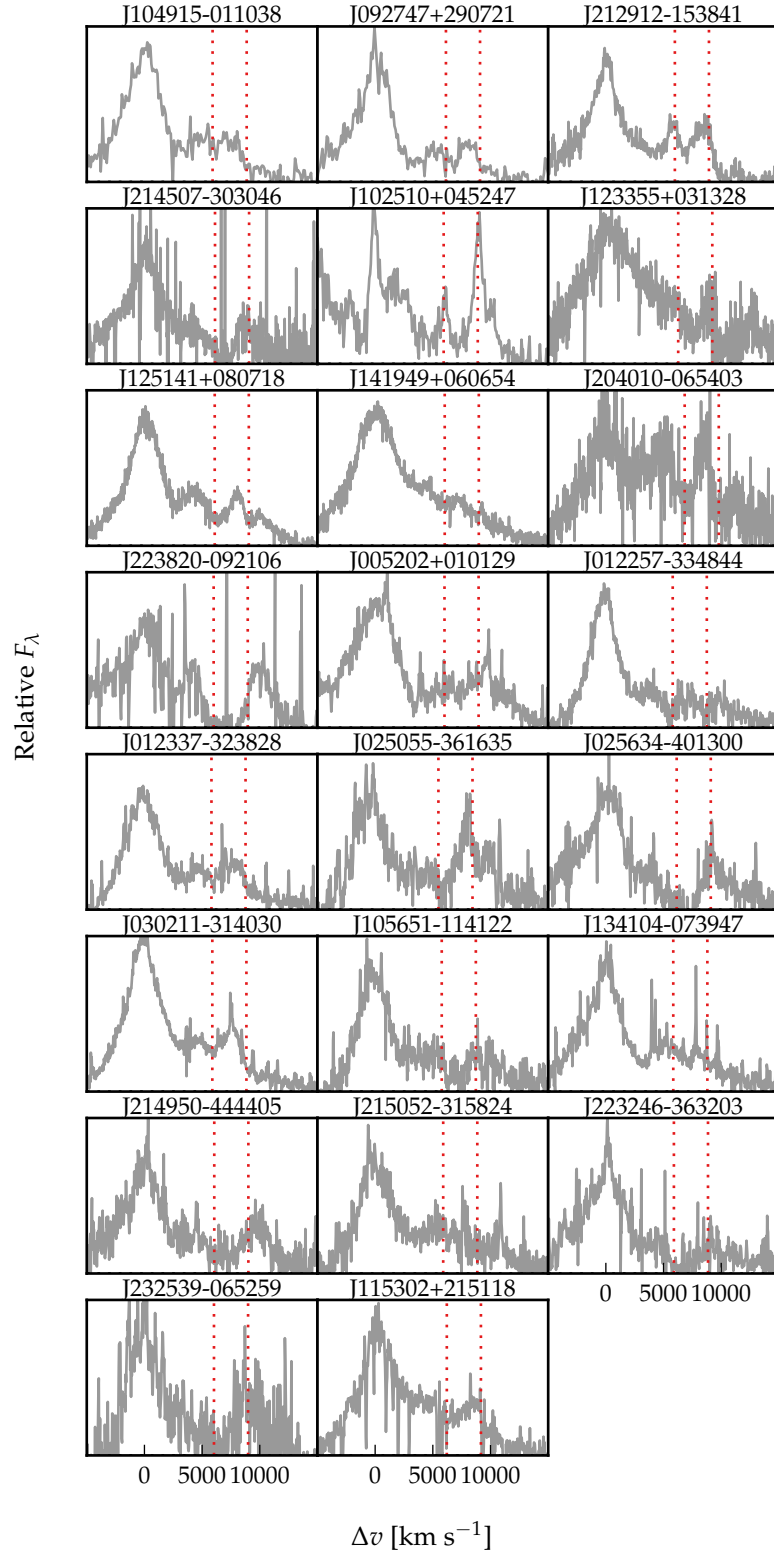
Figure 4.2: The continuum and Fe II spectra of the 23 objects we identified where the Boroson and Green, (1992) empirical template is a poor match to the Fe II emission. The vertical lines indicate the expected positions of the [O III] doublet (which is generally very weak) with a systemic redshift defined using the peak of the broad Hβ emission.
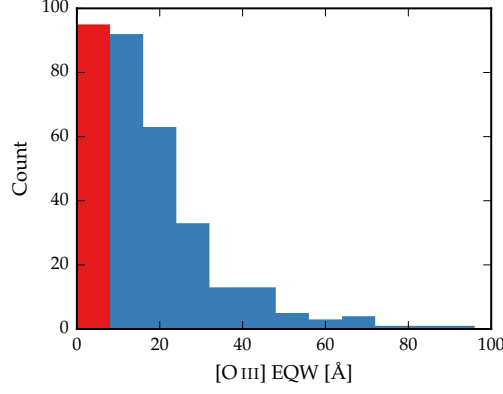
Figure 4.3: [O III] Rest-frame equivalent width of 330 quasars in our sample. The [O III] profiles of the 120 objects in the red bin (EQW < 8Å) cannot be measured reliably, and so these objects are excluded from our analysis.

termine the shape of the [O III] emission. Objects with EQW< 8Å(120) are excluded in subsequent analysis of the [O III] shape.

4.3.4.3 *Low S/N [O III]*

In this section we flag objects with poor spectral S/N. A single S/N cut is not adequate: for a given S/N, it is much easier to measure the properties of a strong line than a weaker one. Our approach is therefore as follow:

1. For each object use the best-fitting parametric model as a high S/N representation of the spectra.

2. Scale the error spectrum so that the S/N (measured in the continuum and quoted per pixel) is {2.5, 5, 7.5, 10, 15, 20, 50}.

3. At each S/N, generate 100 mock spectra by randomly drawing the flux in each pixel from a Normal distribution with mean μ equal to the model flux and width σ equal to the scaled error.

4. Run line-fitting procedure on each mock of the 100 mock spectra; record the value of w_{80} in the best-fitting model.

5. Calculate the 16th, 50th and 84th percentiles of the distribution of w_{80} values.

6. A low S/N flag is assigned to the object if the value of w_{80} changes by more than 10 per cent from the highest S/N representation to the S/N of the data.

Two examples are shown in Figure 4.4. The marker denotes the 50th percentile, and the lower and upper error bars the 16th and 84th percentiles respectively. As expected, the uncertainty on w_{80} increases
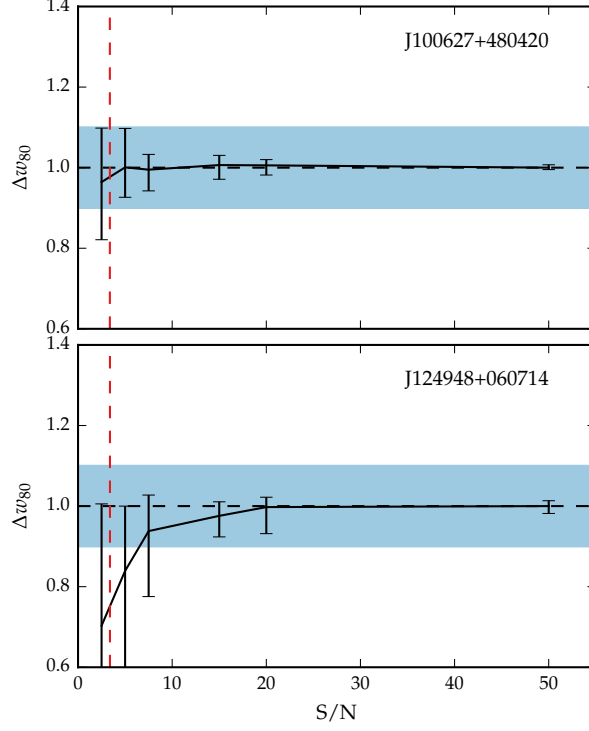
Figure 4.4

as the S/N of the spectrum decreases. The S/N in to two spectra are similar, but the [O III] line in the first object is stronger. Hence it can still be measured reliably even at low S/N. The first object would pass our S/N cut, whereas the latter would fail.

4.3.5 *Model Two: Independent Component Analysis*

Independent component analysis (ICA) is a blind source separatation tecnique for separating a signal in to linearly mixed statistically independent subcomponents. Unlike the more widely-used principle component analysis technique, ICA produces non-negative components which allows for a physical interpretation of the components and weights. ICA has been succesfully applied to model the spectra of emission-line galaxies (Allen et al., 2013) and BAL quasars (Allen et al., 2011). The quasar spectra can be thought of as a set of observations, x, which are made up of statistically independent components, c, that are combined by some mixing matric, W:

$$x = Wc \tag{4.1}$$

ICA reverses this process and describes how the observed data are generated. Both the independent components and the mixing matrix are unknown, but can be found by solving:

$$c = W^{-1}x. \tag{4.2}$$

The components were solved for using a sample of 2,154 SDSS quasars at redshifts XX. At these redshifts the SDSS spectrograph covers the rest-frame region XX-XXÅ where Hβ and [O III] lie. The individual spectra were first adjusted to give the same overall shape as a model quasar template spectrum. Six positive independent components and four additional components that could be negative were found to be sufficient to reconstruct the spectrum, without overfitting. Each quasar spectrum can then be represented as a linear combination of the independent components:

Ask Paul for details.

$$x_j = \sum_{i=1}^{10} c_{ij} W_{ij} \tag{4.3}$$

4.3.5.1 *Fitting procedure*

For each quasar in our NIR sample we perform a variance-weighted least-squares minimisation to determine the optimum value of the components weights. The fitting procedure we employ is as follows. A power-law is first fit to the quasar template spectrum in emission line free windows at 4200-4230, 4435-4700 and 5100-5535 Å. Each of the ICA components is then divided by the power-law. An identical process is performed on each of the spectrum to be fitted, so that there is essentially zero shape in both the components and the spectrum to be fitted. The first six component weights are constrained to be non-negative, and the fit is done in logarithmic wavelength space, so that each pixel corresponds to a fixed velocity width.

4.3.5.2 *Quality of fits*

In general, the ICA components do a remarkably good job at reconstructing the spectra of the objects in our sample. For example, in J125141+080718 (discussed above), it does much better job at modelling the Fe II emission than the Boroson and Green, (1992) template. It is less sensitive to the spectral S/N, and the component weights do not need to be constrained. It is therefore much simpler to apply than fitting multiple Gaussians.

However, it does have its limitations. The components were calculated using a set of lower-redshift, lower-luminosity AGN, and quasar spectra are known to vary systematically as a function of luminosity. For example, the [O III] line is typically broader in more luminous
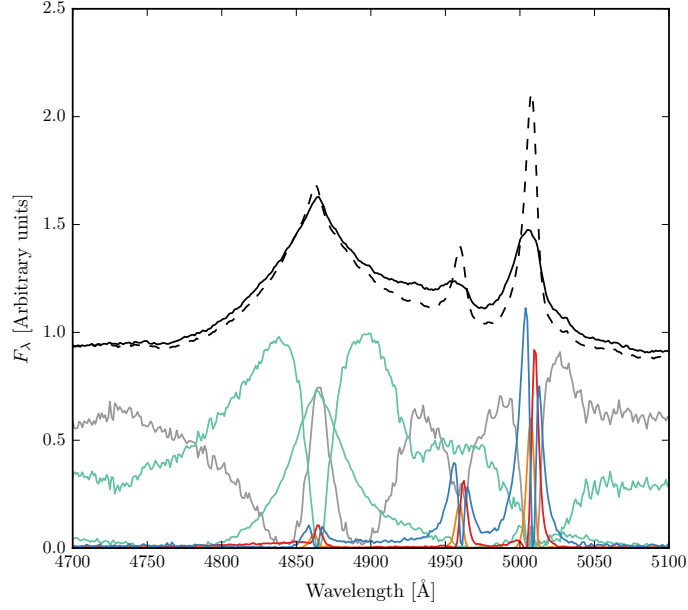
Figure 4.5: ICA reconstruction of spectrum. The total profile is shown in black, and the six individual positive components are shown with different colours. The component weights are the median values from our fits to the low- and high-luminosit samples.

quasars. Because there are so few objects with very broad [O III] in the low-redshift sample, the ICA reconstuction fails to reproduce the broadest [O III] profiles in our sample.

4.3.6   *Physical interpretation of ICA components*

Although the ICA is analysis is not based on any physics, there appears to be a direct correspondence between the individual components and the different emission features which contribute to the spectra (Fig. 4.5). This correspondence is summarised in Table 4.2. The component w_1 seems to correspond to Fe II emission, the components w_2 and w_3 to broad Hβ emission, the components w_4 and w_5 to narrow [O III] emission at the systemic redshift, and the component w_6 to broad, blueshifted [O III] emission.

4.3.6.1   *Reconstructing the [O III] profile*

In order to measure non-paramteric line parameters, e.g. v_10, we must first reconstruct the [O III] emission. It is fortunate that most of the [O III] emission is in just three of the ICA components; the remaining three contribute very little. Therefore, we can set the first three weights to zero to leave only the [O III] emission.

We define the boundaries of [O III]$\lambda5008$ as being between 4950 and 5500Å. The blue limit is close to the peak of the [O III]$\lambda4960$ line,

Table 4.2: Approximate physical origin of the ICA components.

Component	Origin
w_1	Fe II
w_2	Hβ
w_3	Hβ
w_4	[O III] core
w_5	[O III] core
w_6	[O III] wing
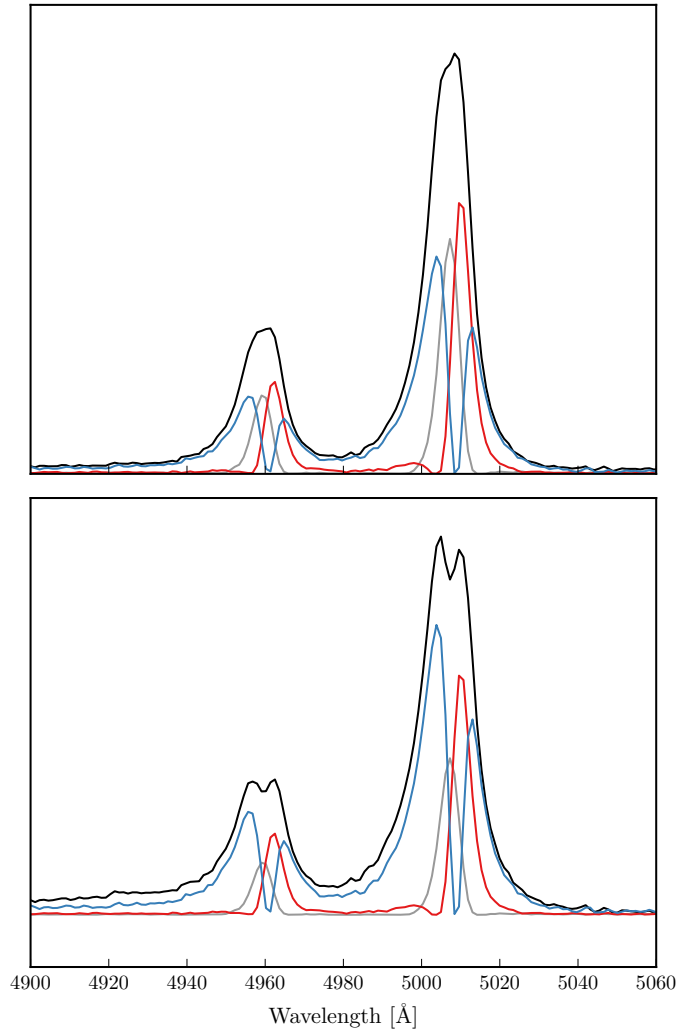
Figure 4.6: Comparison of median [O III] profiles from ICA fits it low- and high-luminosity samples.
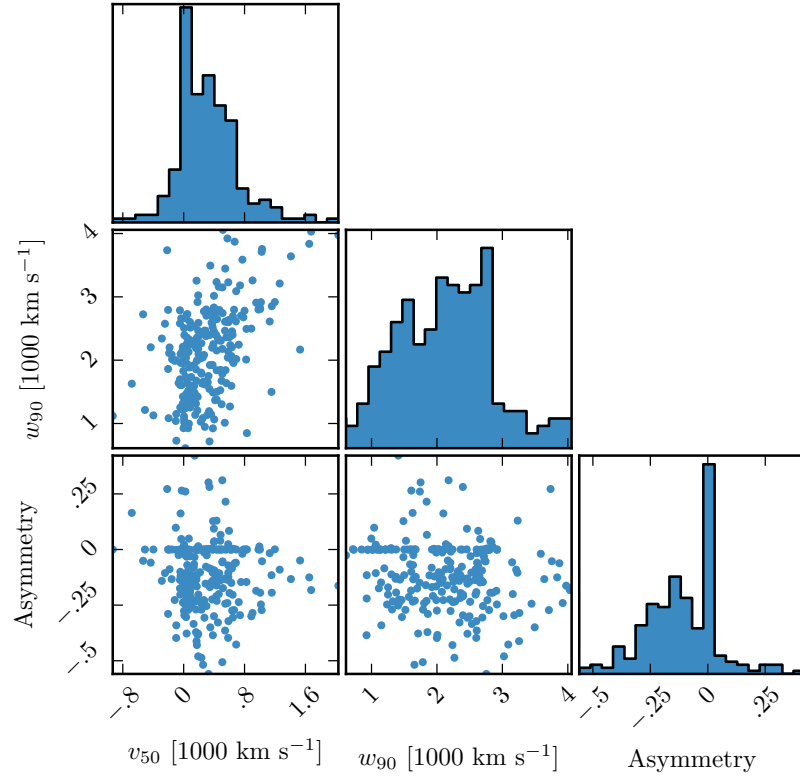
Figure 4.7: The distributions of and correlations between a subset of the non-parameteric measures we made of the best-fitting [O III] models. Remake adding EQW and blueshift? Expain reason for zero asymmetry (single Gaussian)

and so to recover the intrinsic profile we instead use the blue wing of [O III]λ4960. We use the emission from 4980-5050Å, and from 4900-(4980-(5008.2-4960.3)). The blue window is then shifted by (5008.2-4960.3) to reconstruct the blue wing of the [O III]λ5008 line.

The reconstructed spectra are shown in Figure 4.6. At present I am summing the flux all the way from 4950Å. However, this is quite a lot of flux to sum up, and we can't ascribe this flux to the wing of the [O III] emission with any certainty. This is bourne out by the fact that there are quite large differences between, for example, v_{10} measured from the Gaussian fit and v_{10} measured from the ICA fit.

4.4 RESULTS

4.4.1 *Gaussian fits*

Our best-fitting profiles show a strongly blue-asymmetric profile (Fig. 4.7), with a significant fraction of the total emission in a blue

wing. The luminous blueshifted broad wing and the exteremely broad profile reveals high-velocity outflowing ionized gas. Our results, and those of other authors, suggest that kpc-scale outflows in ionized gas are common among the most luminous high-redshift actively accreting SMBHs.

We see a correlation between the [O III] velocity width and blueshift. As the blueshift of the line increases it gets broader. This is consistent with Shen and Ho, (2014), where the strength of the narrow core is decreasing, leading to a broader and more blueshifted profile.

4.4.1.1 *Eigenvector 1 correlations*

At low redshifts, it is standard to use principle component analysis (PCA) to define quasar extrema. This has led to the so-called EV1 correlates (Boroson and Green, 1992). Specifically, much of the variance in quasar spectra is reflected in an anti-correlation between the relative strength of Fe II and the Hβ FWHM. These emission line trends in the optical (for low-z quasars) can be extrended to UV emission lines observed at higher redshifts. The C IV blueshift and EQW is a diagnostic that similarly spans the diversity of broad emission line properties in high redshift quasars (dominated by a virialized component at one extreme and a wind driven component at the other) (Richards et al., 2011; Sulentic et al., 2007). The similarity of the C IV EQW-blueshift parameter space at high redshift to EV1 parameter space at low redshift suggetss that these trends are connected.

Can we calculate a mapping between the two parameter spaces? As a first step we show how the EV1 parameters change as a function of position in the C IV EQW-blueshift parameter space in Figure 4.8. Most of the diversity in C IV properties seems to be driven by the [O III] EQW. On the other hand, properties of the C IV line cannot be used to predict the Hβ FWHM. This is similar to what we found in Chapter 3: objects with large C IV blueshifts have narrow Balmer emission lines, but objects with modest C IV blueshifts have a wide range of Balmer line widths.

Can do something similar with ICA weights, but I'm not sure what the best question to ask is. Could make a linear model of the ICA component weights and fit to the C IV blueshift?

4.4.1.2 *Extreme [O III]*

Because of this diversity, it is the dominant variable in the set of correlations making up EV1, which is believed to be linked to certain fundamental parameters of the accretion process. In Figure 4.8 we show the [O III] EQW as a function of the Hβ FWHM and the optical Fe II strength. The optical Fe II strength is defined as the ratio of the Fe II and Hβ EQW, where the Fe II EQW is measured between 4434
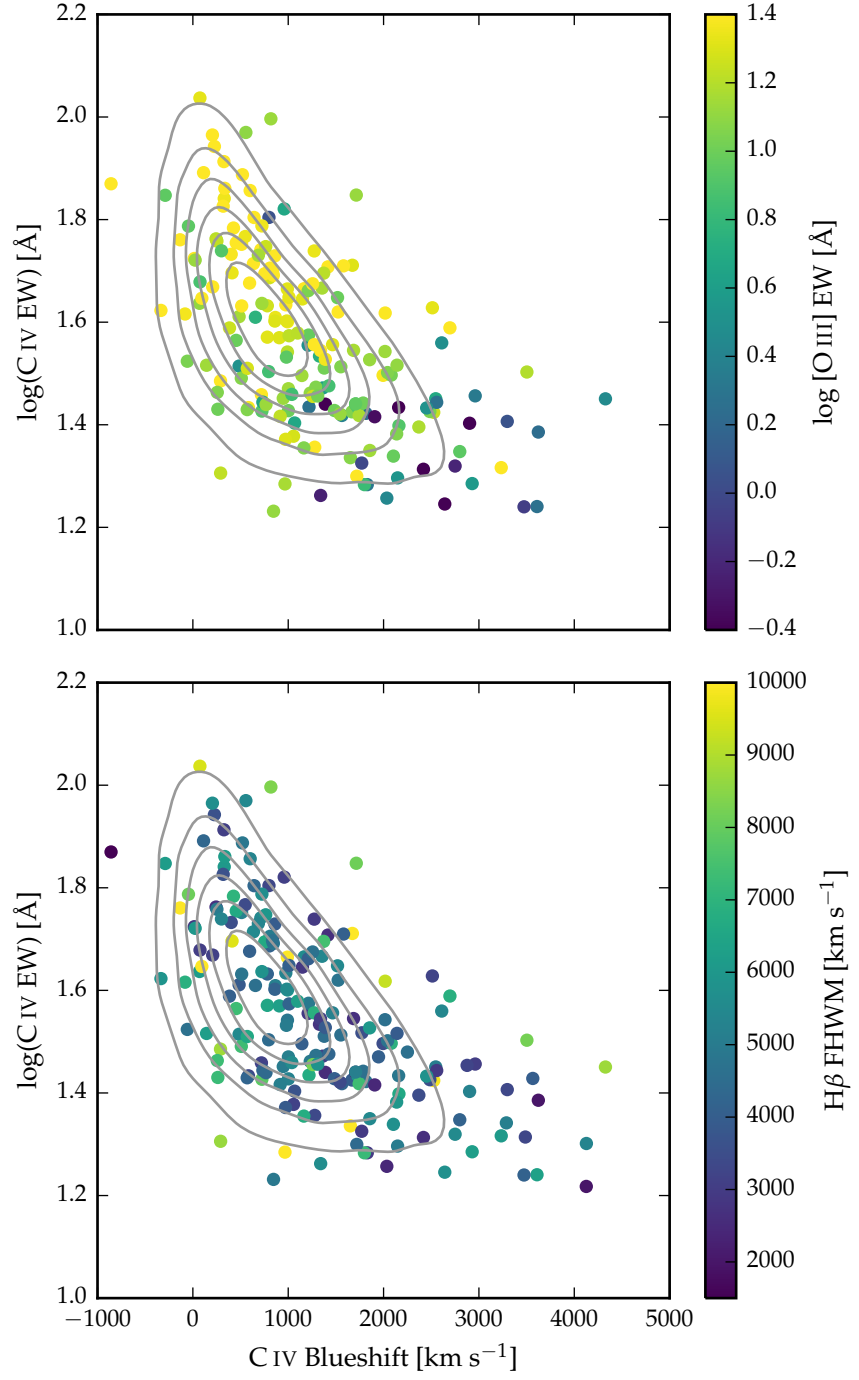
Figure 4.8: The [O III] EQW as a function of the Hβ FWHM and the optical Fe II strength (EQW Fe II/ EQW Hβ).
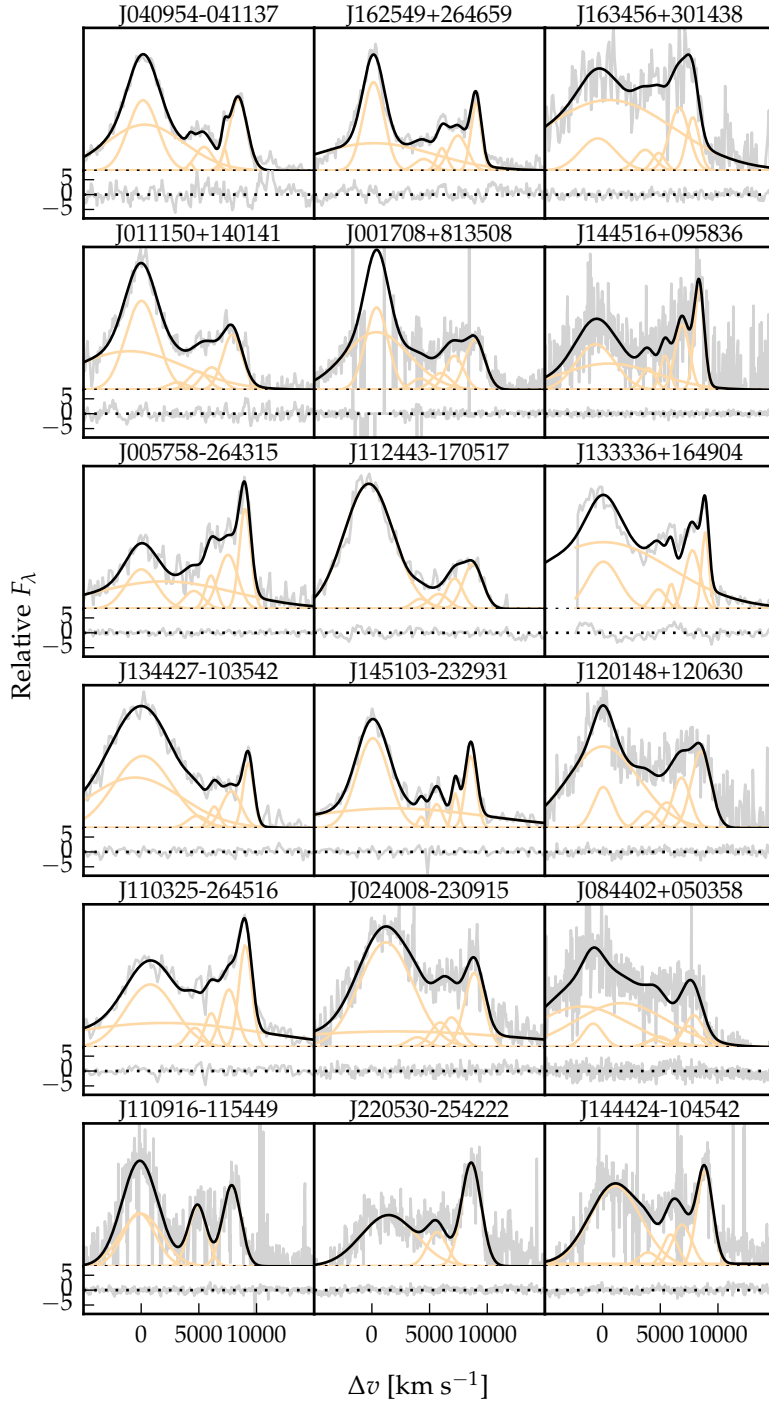
Figure 4.9: Extreme [O III] emission profiles.

and 4684Å. These parameters form part of 'eigenvector 1' (EV1), the first eigenvector in a principal component analysis which originated from the work of Boroson and Green, (1992). In our sample, these parameters follow very similar correlations to what is observed at low-z (e.g. Shen and Ho, 2014). In particular, the anti-correlation between the [O III] and Fe II EQWs.

Same as Shen, (2016), we confirm that the EV1 correlations hold at high luminosities/redshifts. See also Sulentic et al. 2004, 2006; Runnoe et al. 2013. Make sure it's clear that Shen, (2016) quasars make up a significant chunk of our sample.

4.4.2   *ICA fits*

We find there is a decreasing symmetric component at high luminosities. Relates directly to Shen and Ho, (2014). A stable narrow line region is removed by the outflowing material. Shen and Ho, (2014) showed that the strength of the core [O III] component decreases with quasar luminosity and optical Fe II strength faster than the wing component, leading to overall broader and more blueshifted profiles as luminosity and Fe II strength (or C IV blueshift) increases.

Zhang et al. 2011: Blueshift of [O III] correlates significantly with the EQW of the core. The more the peak of the line is blueshifted, the more the core component decreases dramatically, while the blue wing changes much less. We see this clearly in Figure 4.12. This is similar to behaviour of C IV? i.e. is there a mapping from this to the C IV space diagram? This would suggests that the mechamism producing the two corerlations is the same. Consistent with the core coming from the canonical extended NLR where the gas is dominate by gravity of the bulge while the wing arises in an outflow. And C IV explained by wind. Suggests intimate connection between BLR and NLR.

4.4.2.1   *EV1 correlations*

In Figure 4.13 we show how the [O III] strength varies as a function of the C IV blueshift. There is a very well defined relation: when C IV is strongly blueshifted [O III] is very weak.

This is shown in a different way in Figure 4.14. Here we divide our sample in to four bins according to the C IV blueshift. From the quasars in each C IV blueshift bin we then find then generate an ICA spectrum using the median weights from each quasar. The differences in the spectra as a function of the C IV blueshift are dramatic. [O III] becomes progressively weaker and more blueshifted. The anti-correlation with Fe III and the blueward Fe II also clear, but there is no change in the redward Fe II.

This seems to really connect high-z EV1 with low-z EV1, which I don't think has really been done before. At low redshifts we have the strengths of [O III], Fe II, and the Hβ FWHM. At high redshifts
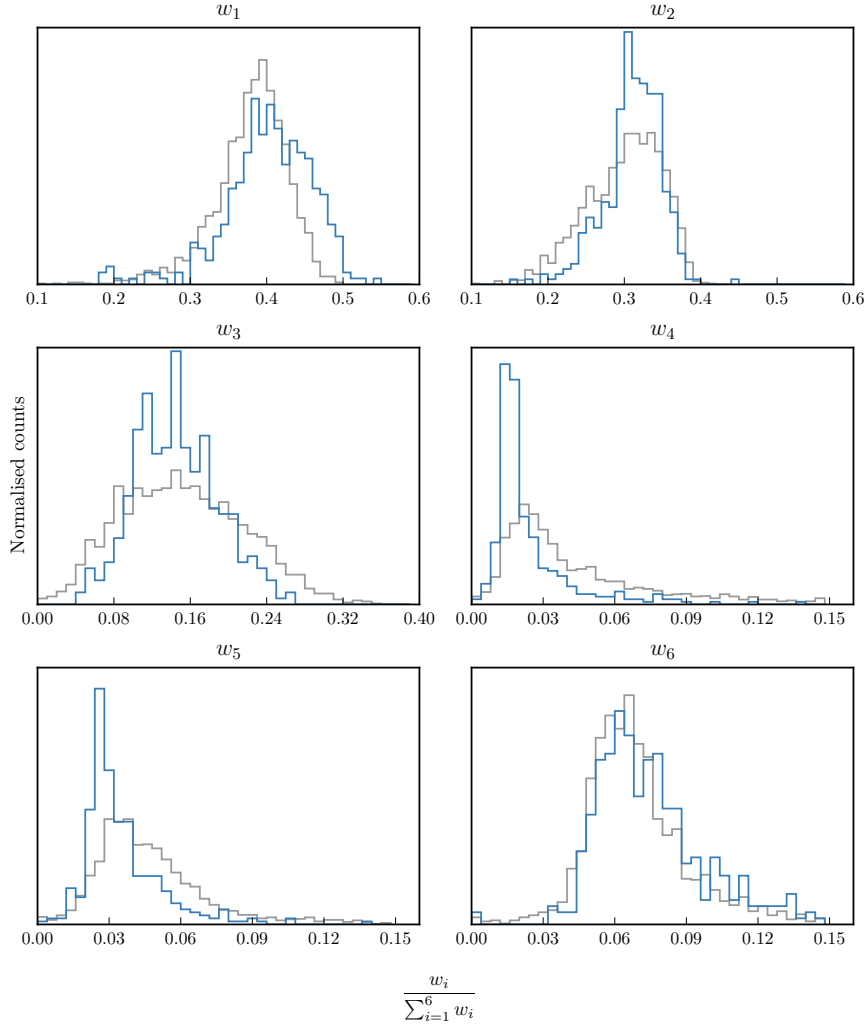
Figure 4.10: The relative weight in each of the six positive ICA components for the high-luminosity (blue) and low luminosity samples (grey). In the high-luminosity sample Fe II emission is stronger (component w_1). The core [O III] emission is weaker (components w_4, w_5) but the strength of the blueshifted wing is the same (w_6).
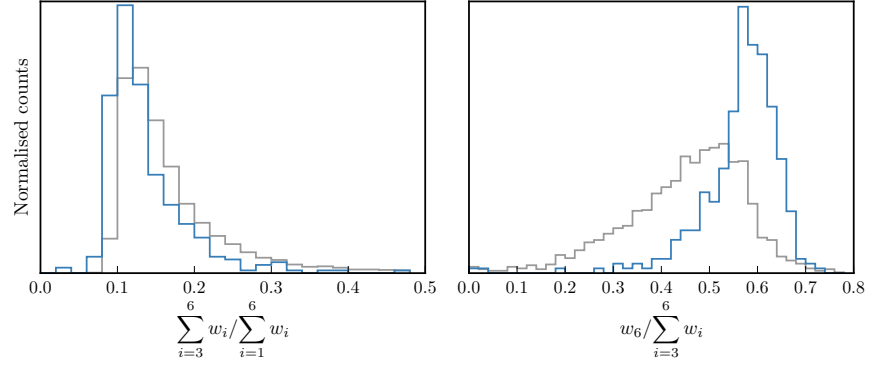
Figure 4.11: The relative weight in the three ICA components corresponding to [O III] emission (*left*) and the relative weight of the component most closely related to blueshifted [O III] emission relative to all three [O III] components (*right*). [O III] emission is weaker in the high-luminosity sample, but the relative contribution but the fractional contribution from the blueshifted component to the total [O III] emission is higher. Hence [O III] is weaker, broader, and more asymmetric in the high-luminosity sample.
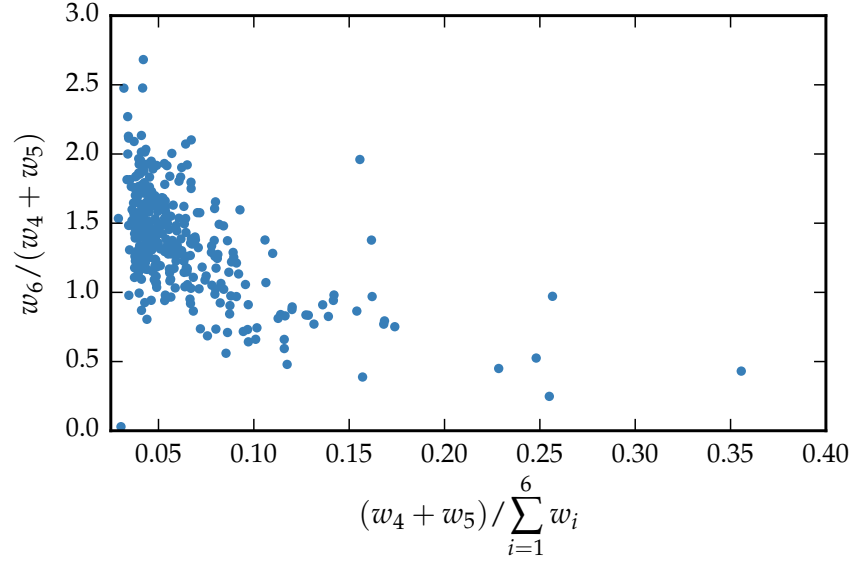
Figure 4.12
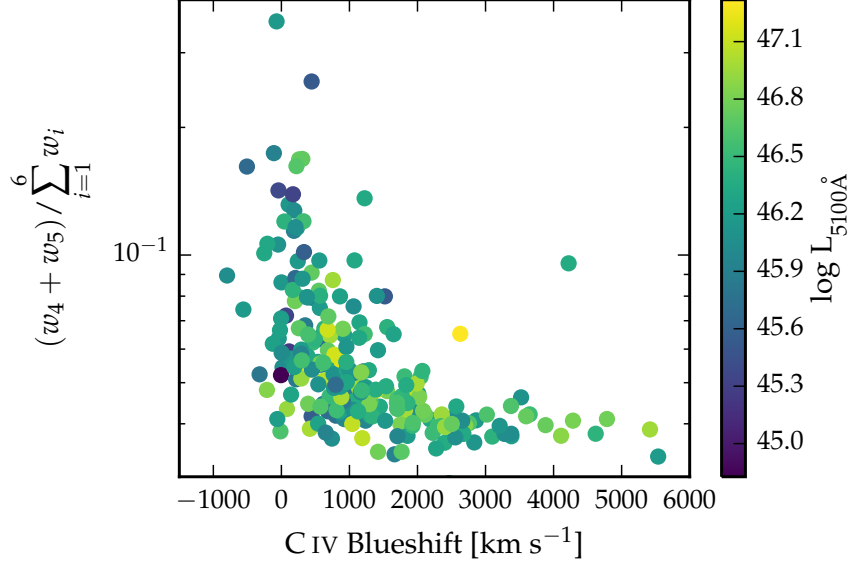
Figure 4.13: [O III] strength decreases as the C IV blueshift increases. See similar thing if I use [O III] EQW instead. Only showing the core components here. The C IV blueshift is now measured relative to the NIR ICA redshift. I think this is trend is mostly being driven by the Eigenvector 1 correlations: as the blueshift increases the Fe II strength increases and the [O III] strength decreases. Doesn't appear to be driven by the luminositiy. Is this tighter than EV1 trend shown with Fe/OIII strength by other authors? Is the AGN NLR absent in objects where outflows have reached kiloparsec scales, sweeping up the low-density material responsible for the [OIII]-emission?
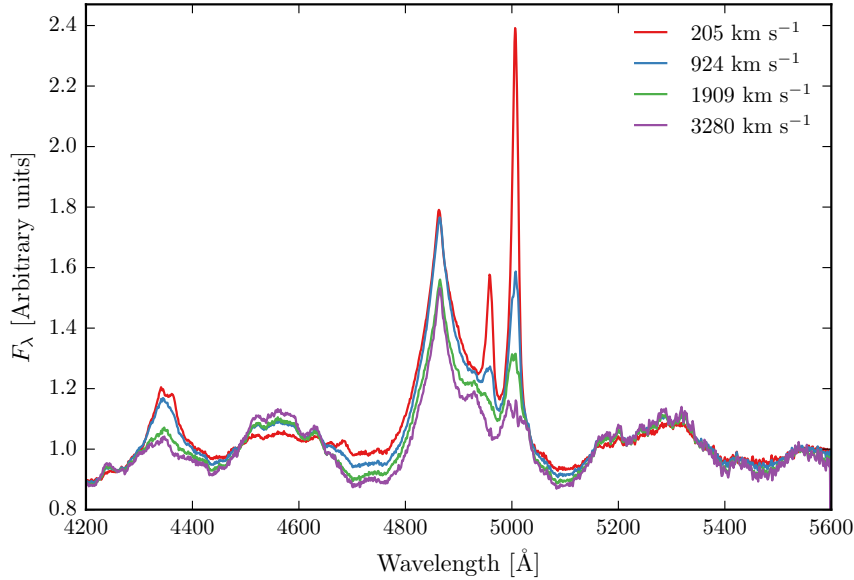
Figure 4.14: ICA median weights as a functino of the CIV blueshift.

we have the C IV blueshift and EQW. The distribution of quasars in these two planes look qualitatively similar. It's tempting to think that quasars with large C IV blueshifts should also have strong Fe II and weak [O III]. This is indeed what our results seem to suggests.

The ICA can be thought of as update on EV1. The spectral diversity is encapsulated in the EV1 components. Most of the variance in EV1 is the anti-correlation between the strengths of [O III] and Fe II. So at one end we have objects with strong Fe II and weak [O III], and at the other end objects with weak Fe II and strong [O III]. Other properties, including the C IV blueshift and the Hβ FWHM, also change systematically. Our work shows that the ICA component weights change systematically along the EV1 sequence. .

What is the best way to actually do this mapping?

4.5   MEASURING THE QUASAR SYSTEMIC REDSHIFT

In this section we do a comparison of systemic redshift estimates from [O III], broad Hβ and Hα, and from fitting the ICA component weights. This is an important issue. Accurate systemic redshift estimates are essential in a number of applications, and researchers have devoted a large amount of telescope time to obtaining near-infrared spectra to access [O III] for this purpose. HI, CO and absorption line mesaures of the host galaxy rest frame suggest that [O III] usually gives consistent results within 200 km/s (de Robertis 1985; Whittle 1985; Wilson & Heckman 1985; Condon et al. 1985; Stripe 1990; Alloin et al. 1992; Evans et al. 2001). However, our work shows that at high luminosities this can result in large errors (profile can be dominated by blueshifted component, Fe emission can be improperly subtracted, or [O III] might not be detected at all.

4.5.1   *Hα*

There are 224 quasars in our sample with spectra covering the Hα emission line. We discard seven of these from our sample because of very low S/N (<2.5 measured in the Hα line), leaving 217 To measure the position of the line we fit a parameteric model, which is very similar to the model described in Chapter 3. The continuum emission is first modeled and subtracted using the procedure described in Chapter 3. We then test five different models with increasing degrees of freedom to model the Hα emission. The models are summarised in Table 4.3. They are (1) a single broad Gaussian; (2) two broad Gaussians with identical velocity centroids; (3) two broad Gaussians with different velocity centroids; (4) two broad Gaussians with identical velocity centroids, and additional narrower Gaussians to model the narrow Hα emission, and the narrow components of [N II]λλ6548,6584 and [S II]λλ6717,6731; (5) two broad Gaussians with different velocity centroids, and additional narrower Gaussians. If used, the width and
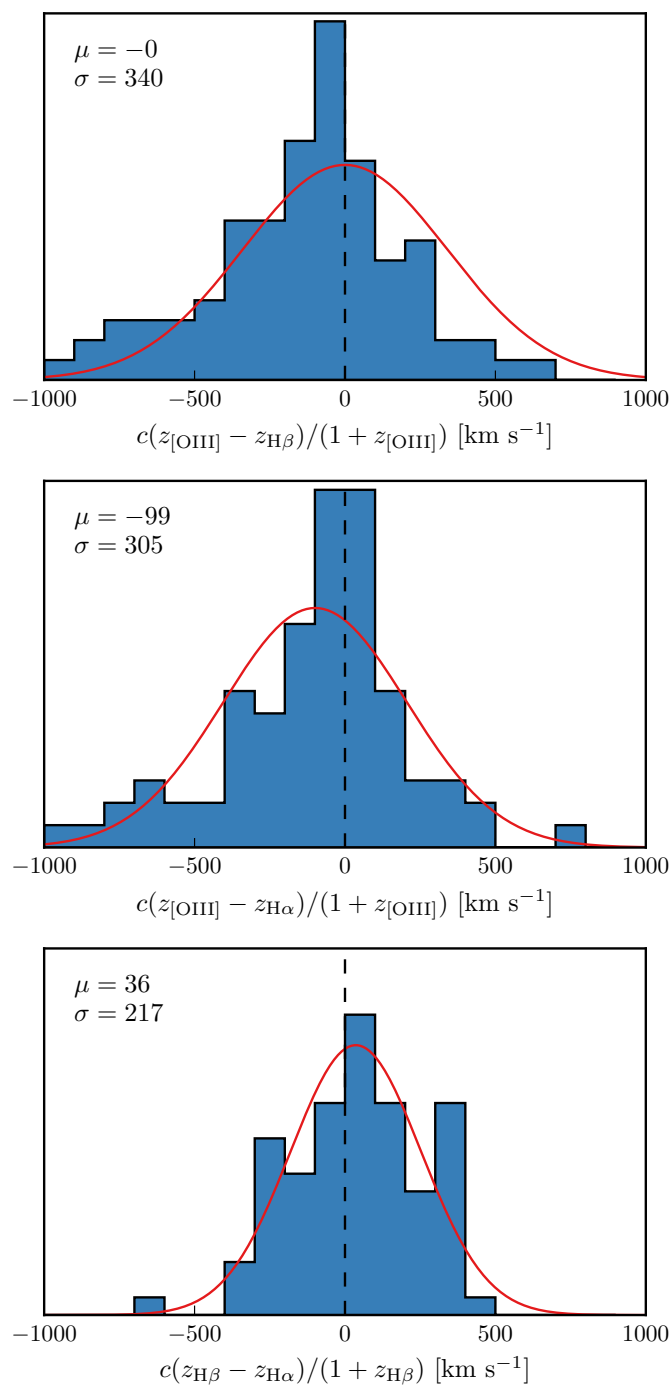
Figure 4.15: Redshift comparisons. Lots have been excluded from Ha/Hb so need to look at flags greater than one. What is the big peak? Gaussian fit to the first one has failed. Find out why these plots look different to ones in paper.

Table 4.3: Models used for Hα emission

Model	Components	Fix centroids?	Number
1	1 broad Gaussian	N/A	10
2	2 broad Gaussians	Yes	71
3	2 broad Gaussians	No	32
4	2 broad Gaussians + narrow Gaussians	Yes	51
5	2 broad Gaussians + narrow Gaussians	No	53

velocity of all narrow components are set to be equal in the fit, and the relative flux ratio of the two [N II] components is fixed at the expected value of 2.96. The model we select is the simplest model for which the fractional change in the reduced chi-squared from the model with the lowest reduced chi-squared is less than ten per cent. The redshift is then measured at the peak flux of the Hα model, including both the broad and narrow components of Hα if appropriate.

4.5.2   *ICA*

Need to recalculat ICA redshifts since shift to fit in log space

. Benefit of ICA is that it works regarless of the [O III] strength.

Can also describe what I found trying to get redshifts from broad Hα, Hβ? (Narrow components generally very weak at these luminosities so can't be used.) Generally find no systematic errors but large (~1000km s^{-1} scatter). Comparing NIR ICA to [O III] for the [O III] with high S/N I find small (few hundred km s^{-1}) scatter.

4.6   LUMINOSITY/REDSHIFT-EVOLUTION OF [O III] PROPERTIES

In this section we look for any luminosity/redshift dependent changes in the [O III] line properties. To do this we extend the dynamic range of our samples in terms of both luminosity and redshift by supplementing our sample with quasars presented by Zakamska and Greene, (2014) and Harrison et al., (2016).

The Zakamska and Greene, (2014) objects are a sample of 568 obscured luminous quasars selected from SDSS (Reyes et al., 2008; Yuan, Strauss, and Zakamska, 2016). They are selected to have [O III] luminosities above $10^{8.5}$L$_\odot$ and have a median redshift $z = 0.397$.

We also include 40 quasars at redshifts $1.1 \leqslant z \leqslant 1.7$) from the KMOS AGN Survey at High redshift (KASHz) with [O III] line measurements.

We also have the same information for ~20 000 SDSS spectra from Mullaney et al., (2013).

In Figure 4.16 we show the [O III] velocity width as a function of the [O III] luminosity and the quasar redshift. The lack of any redshift-evolution between $z = 0$ and $z = 1.5$ was reported by Harrison et
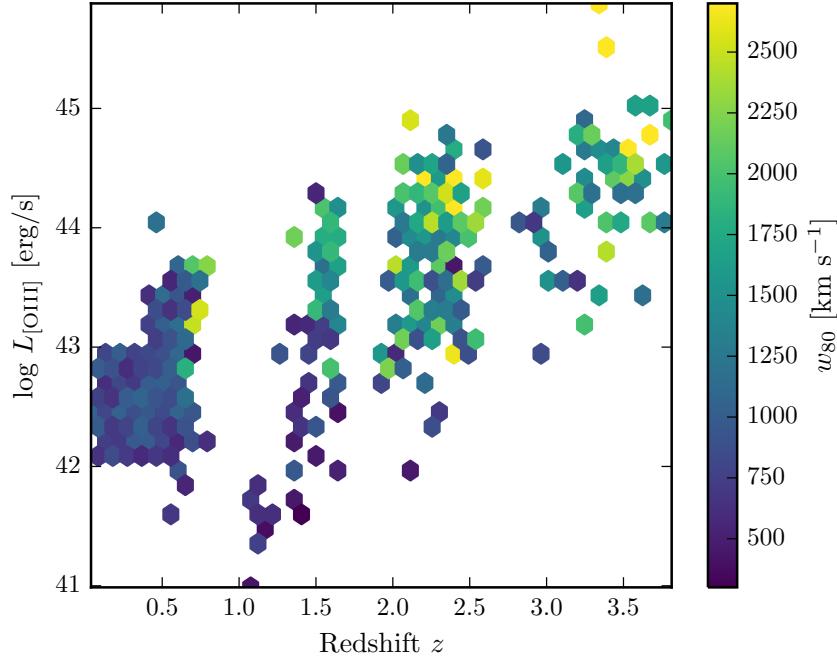
Figure 4.16: The [O III] velocity-width, characterised by w_{80}, as a function the [O III] luminosity and the quasar redshift. The color of each hexagon denotes the mean w_{80} for the obects in that luminosity-redshift bin. We have suplemented our sample with low-z objects from Zakamska and Greene, (2014) and medium ($z \sim 1.5$) redshift objects from Harrison et al., (2016). If I keep this plot make sure its clear which points belong to which sample.
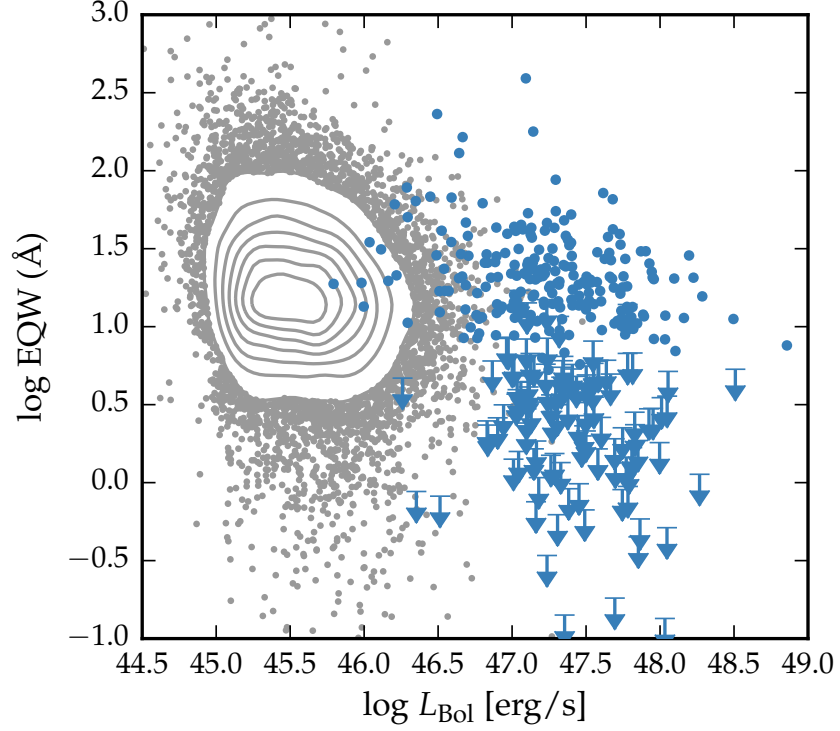
Figure 4.17: The [O III] EW as a function of the quasar bolometric luminosity for the sample presented in this chapter (blue circles) and the low-z SDSS sample (grey points and contours). Upper limits are denoted by the downward arrows.

al., (2016). Our additional data suggests that this continues to $z \sim 2.5$. On the other hand, at fixed redshift, we see a significant correlation between the [O III] velocity width and the luminosity.

The fact that we don't see many broad lines in the Zakamska and Greene, (2014) objects even at luminosities >43 erg/s could be due to the fact that these are all type II quasars, whereas the sample presented in this chapter are all type I. Mullaney et al., (2013) showed that the [O III] lines of type I quasars are typically broader than in type II quasars.

4.7   EQUIVALENT WIDTH

Firstly, the best-fitting model comprising the continuum, Fe II, and Hβ emission is subtracted from the spectra, leaving behind only emission due to [O III]. From this spectra we generate 100 mock spectra, where the flux at each wavelength is randomly drawn from a Normal distribution with a mean equal to the flux convolved with a Gaussian of width 200km s^{-1} and a width equal to the known error. We then perform an error-weighted linear least-squares regression with

an [O III] template derived from a fit to a very high S/N low red-shift SDSS composite spectra. The equivalent width of the best-fitting model is recorded for each of the 100 realisations of the spectra. The error in the equivalent width is defined as the root-mean-square of these values.

In Fig. 4.17 we show the [O III]5008 EW as a function of the quasar bolometric luminosity. Bolometric luminosity is estimated from the monochromatic continuum luminosity at 5100Åusing the correction factor given by Richards et al., (2006a). For comparison, we also show the low-z sample from Shen et al., (2011).

The equivalent width of [O III] has been found to strongly decrease as a function of redshift and/or luminosity (e.g. Brotherton, 1996; Netzer et al., 2004; Sulentic et al., 2004; Baskin and Laor, 2005b).

The size of the narrow line region is roughly expected to scale as $L^{0.5}$ (e.g. Netzer et al., 2004). However, for high luminosity quasars with strong [O III] this gives NLR sizes which are unreasonably large (~100 kpc; Netzer et al., 2004).

Netzer et al., (2004) found 1/3 of their high luminosity sample had very weak [O III], whereas quasars with weak [O III] are very rare for nearby AGN. We find that [O III] is undetected/very weak in XX per cent of our sample, which is very similar to the fraction reported by Netzer et al., (2004). Netzer et al., (2004) claim that for the population of strong [O III] emitters there is no reduction of EW with increasing source luminosity. On the other hand, there are many weak or no [O III] emitters at high luminosity that could give the impression that the line EQW decreases with increasing source luminoisity.

4.7.1 *[O III] and C IV outflows are linked*

Optical spectra are available for XXX quasars in our catalogue, and cover the broad C IV doublet. As we described in Chapter 3, C IV is often blueshifted, which almost certainly signal the presence of strong outflows, most likely originating in a disc wind. In Chapter 3 we demonstrated that the quasars in our sample cover the full range of C IV blueshifts seen in the SDSS quasar population, which makes our sample unique in that it allows us to study properties of the quasar across the full parameter range.

The C IV velocity centroid measurements are taken directly from Chapter 3. We define the 'location' of the [O III] emission using v_{10}, although the results are the same if v_{20}, v_{50} etc. are used instead.

In Figure 4.18 we show the C IV blueshifts against the [O III] blueshifts. This comparison is done for a sub-sample of 146 objects where we have good measurements of the C IV, [O III] profiles.

There is a clear and strong correlation. Note that our EQW cut removes most of the quasars with large C IV blueshifts, since [O III] is on average very weak in these qusars. Similar correlations have been
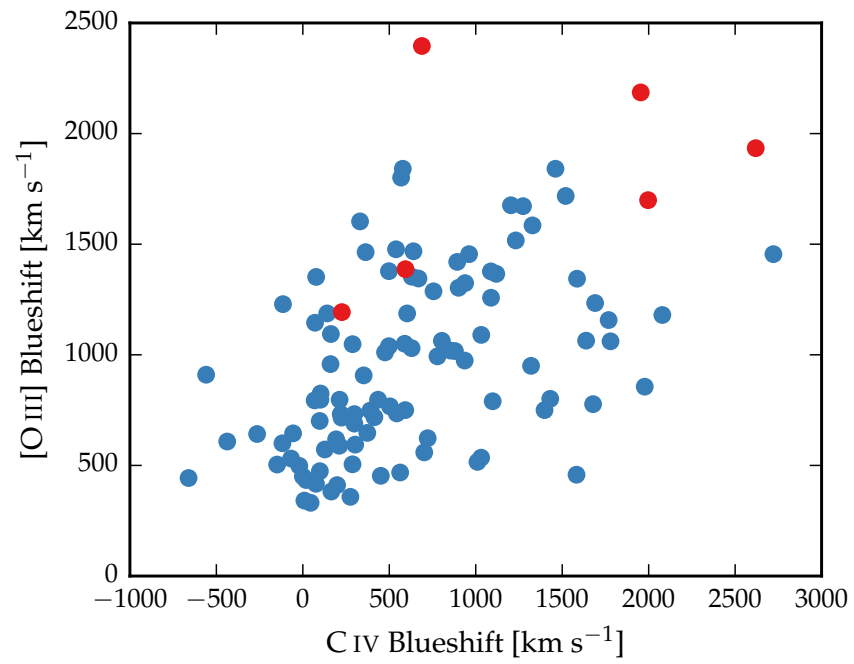
Figure 4.18: The relation between the blueshifts of C IV and [O III]. Note that we are using v_{10} for the [O III] position and v_{50} for the C IV position. We can't use v_{50} for [O III] because sometimes we are using a single Gaussian, especially if the [O III] is weaker and we miss the broad component.

tentatively found in lower redshift quasars and AGN (Zamanov et al., 2002).

The blueshifting of C IV is known to correlate with luminosity (Richards et al., 2011). In [O III], the blueshifted wing becomes relatively more prominent as the luminosity of the quasar increases (Shen and Ho, 2014). Therefore, it is plausible that the correlation between the C IV and [O III] blueshifts is a secondary effect that is driven by the correlation of each with the luminosity. However, no strong luminosity-dependent trends are apparent in Figure 4.18. We find that both the [O III] and C IV blueshifts are correlated with the luminosity, but that these correlations are much weaker than the correlation between the [O III] and C IV blueshifts.

4.8   EIGENVECTOR ONE CORRELATIONS

4.9   BROAD ABSORPTION LINE QUASARS

Check all of this

19 quasars in our catalogue are classified as broad absorption line (BAL) quasars, using the either the SDSS classification flags or the Allen et al., (2011) catalogue. We find that the BAL quasars have typically broader [O III] than the rest of the sample. Note that in the Zakamska et al., (2016) sample of very red quasars, the incidence of BALs is very high, and these objects have extremely broad [O III] profiles. A two-sided Kolmogorov-Smirnov statistic on the w_{80} distributions returned a p-value of 0.10. What does this mean? Try with different parameters? Histograms look rubbish so maybe just give the numbers.

4.10   DISCUSSION

Looking at the [O III] velocity width as a function of luminosity tells us about the physical drivers of the outflows observed in [O III]. The correlation with luminosity suggests that the highest velocity outflows are associated with the most luminous AGN. This has been reported for low-redshift AGN, for both ionized and molecular outflows (e.g. Westmoquette et al. 2012; Veilleux et al. 2013; Arribas et al. 2014; Cicone et al. 2014; Hill & Zakamska 2014).

This suggests that the outflows are driven by radiative forces. On the other hand, Mullaney et al., (2013) find that once the correlation between the [O III] luminosity and the radio luminosity has been taken in to account, the [O III] velocity width is more strongly related to the radio luminosity of the AGN.

4.10.1 *Type II quasars*

Implications of our findings on searches for high-redshift type 2 quasars. It could be that type II quasars exist. If you look at CIV/MgII the narrow line components are very weak. So the contribution from the narrow line region is very weak in luminous quasars, and you just won't see it even if the broad line region is obscured. Findings in this paper seem to suggest that the startic narrow line region is very weak in luminous quasars.

SED PROPERTIES

While many authors have focused on studies of specific sub-sets of active galactic nuclei (AGN) with extreme observational properties, what is missing is an understanding of how these extreme subsets relate to the population as a whole. I have addressed this problem using multi-wavelength spectral energy distributions (SEDs) of large samples of quasars. I have constructed an SED model which is able to reproduce the average optical to near-infrared (NIR) colours of 10,000s of AGNs spanning a broad range in redshift and luminosity.

5.1 DATA

The systematic study of the dependence of the SED shape on physical parameters has, until very recently, been limited by the difficulty in obtaining a large sample of quasars with good multi-wavelength coverage and large dynamic range in luminosity and redshift. In this work, we take advantage of a number of recent, sensitive, wide-field surveys, covering the UV to mid-IR spectral region.

5.1.1 *The Sloan Digital Sky Survey*

We use the Seventh Data Release (DR7) of the Sloan Digital Sky Survey (SDSS; York et al., 2000) spectroscopic quasar catalogue (Schneider et al., 2010), which includes 105,783 objects across 9380 deg^2. The SDSS obtained images in five broad optical passbands: u (λ_{eff} = 3543Å), g (λ_{eff} = 4770Å), r (λ_{eff} = 6231Å), i (λ_{eff} = 7625Å), and z (λ_{eff} = 9134Å). We use BEST point-spread function (PSF) magnitudes, correcting for Galactic extinction using the maps of Schlegel, Finkbeiner, and Davis, (1998), assuming a Milky Way (MW) extinction curve (Pei, 1992) and an extinction to reddening ratio $A(V)/E(B-V)$ = 3.1. Although the SDSS asinh magnitude system is intended to be on the AB system (Oke and Gunn, 1983), the photometric zero-points are known to be slightly off the AB standard. To account for this we add 0.03 mag to the u, g, r and i magnitudes, and 0.05 mag to the z magnitude.

Where did these numbers come from?

DR7Q quasar targets were primarily selected to have $i \leqslant 19.1$ if the colours were consistent with being at redshift $z < 3$, and $i \leqslant 20.2$ if consistent with $z > 3$ (Richards et al., 2002). The survey is sensitive to the most luminous quasars at a given redshift. The large number of objects at $z < 3$ with $i > 19.1$ were selected by algorithms other than the main quasar selection. For example, quasar targets were also

selected if they matched within $2''$ of an object in the Faint Images of the Radio Sky at Twenty-cm (FIRST) catalogue of radio sources (Becker, White, and Helfand, 1995).

5.1.2 *UKIDSS Large Area Survey*

We use the UKIRT Infrared Deep Sky Survey (UKIDSS; Lawrence et al., 2007) Large Area Survey (ULAS) which has observed $\sim 3,200$ deg^2 in four near-IR passbands: Y ($\lambda_{\text{eff}} = 1.0305\mu$m), J ($\lambda_{\text{eff}} = 1.2483\mu$m), H ($\lambda_{\text{eff}} = 1.6313\mu$m), and K ($\lambda_{\text{eff}} = 2.2010\mu$m)). We used the ninth data release (DR9) of the ULAS. Cross-matching (with a $2''$ radius and picking only the nearest neighbour) the SDSS DR7Q catalogue with the ULAS catalogue, which covers only $\sim 38\%$ of the SDSS footprint, resulted in 37,893 matches. The ULAS magnitudes are aperture corrected magnitudes in a $2''$ diameter aperture and are not corrected for Galactic extinction.

5.1.3 *WISE All-WISE Survey*

The Wide-field Infrared Explorer (WISE; Wright et al., 2010) mapped almost the sky in four mid-IR band-passes: $W1$ ($\lambda_{\text{eff}} = 3.4\mu$m), $W2$ ($\lambda_{\text{eff}} = 4.6\mu$m), $W3$ ($\lambda_{\text{eff}} = 12\mu$m), and $W4$ ($\lambda_{\text{eff}} = 22\mu$m). The WISE AllWISE Data Release ('AllWISE') combines data from the nine month cryogenic phase of the mission that led to the 'AllSky' data release with data from the NEOWISE program (Mainzer et al., 2011). Cross-referencing the SDSS DR7Q catalogue with the AllWISE catalogue resulted in 102,734 matches. Two objects were matched to multiple AllWISE objects, and were discarded from the sample. Vega to AB conversion factors for WISE photometry are given in the WISE Explanatory Supplement (Cutri et al., 2013)

5.1.4 *Completeness of Photometry*

Objects which are faint in the SDSS i band-passes are more likely to have magnitudes which fall below the limiting magnitudes of the UKIDSS and WISE band-passes at longer wavelengths. For a given i magnitude, a quasar with a blue spectrum is more likely to be undetected at longer wavelengths than a quasar with a red spectrum. Therefore, as we allow fainter quasars in to our sample we will be biased towards objects with redder spectra. We impose an observed i magnitude lower limit of 19.1 mag, which is the magnitude limit of the main SDSS colour-selection algorithm. We verified that above this limit the DR7Q-matched sample is 95% complete in all band-passes with $S/N > 5$ (excluding WISE $W3$ and $W4$) and that this fraction is not changing rapidly with the brightness of the sample.
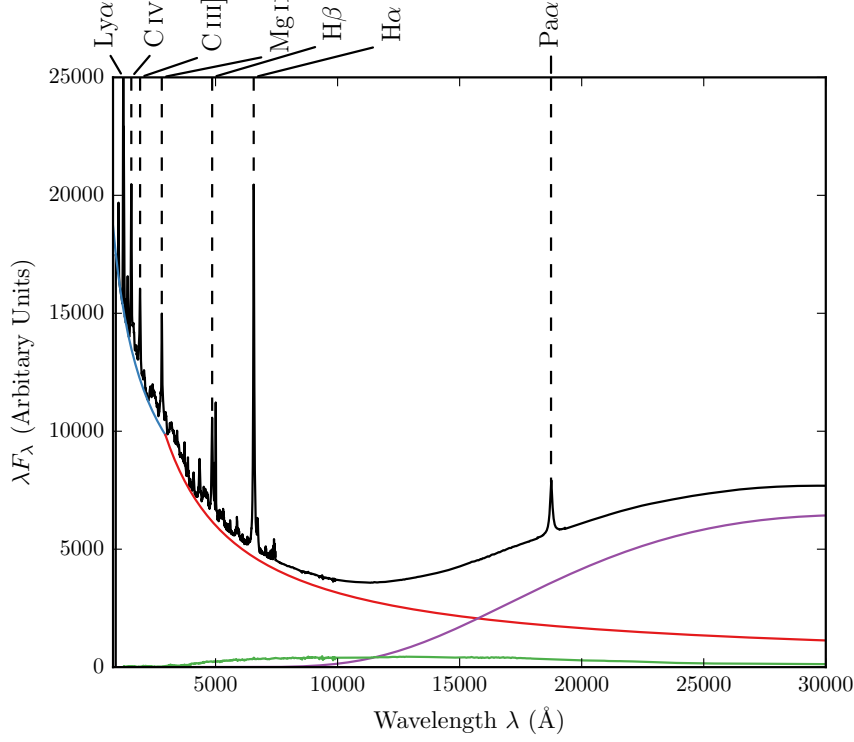
Figure 5.1: Model spectrum at $z = 1$, showing the contributions to the total flux from the blue power-law slope, red power-law slope, blackbody and host galaxy. The locations of the most prominent emission lines in the spectrum are also indicated.

5.1.5 *Final Sample*

We exclude objects flagged as BALQSOs by Shen et al., (2011), since our model is unable to reproduce the broad absorption troughs that appear in the spectra of these objects. The final sample contains 61,411 objects in the redshift range $0.2 < z < 3.8$.

5.2 SED MODEL

I have constructed a new SED model which reproduces the SEDs of AGNs from the rest-frame UV ($\sim 0.1\,\mu$m) to the rest-frame near-IR ($\sim 3\,\mu$m). In this section, I will describe how I have modelled the emission from the various components contributing to the emission in this spectral region. The model spectrum is shown in Figure 5.1, with each of the main components indicated.

5.2.1   *Accretion Disk*

More than half the bolometric luminosity of an unobscured AGN is emitted in the Big Blue Bump, which extends from the near-IR at 1 μm to past 0.1 μm in the UV, and possibly all the way to the soft X-ray region. The Big Blue Bump emission is thought to arise from an accretion disc. In the 0.1 - 1 μm region the spectrum is generally characterised by a power-law of the form $F_\nu = C\nu^{-\alpha}$ where α is the power-law index, C is a constant, and F_ν is the flux per unit frequency, usually measured in units of erg s^{-1} cm^{-2} Hz^{-1}. Equivalently this can be expressed as $F_\lambda = C'\lambda^{\alpha-2}$ where F_λ is the flux per unit wavelength, usually measured in units of s^{-1} cm^{-2} Å^{-1}.

The value of the power-law index is uncertain. From a theoretical perspective, models of geometrically thin accretion discs (Shakura and Sunyaev, 1973) assume, in particular, that the disc is stationary, axisymmetric, and extends down to the innermost stable circular orbit, and that angular momentum is transported by local 'viscous' stresses that convert gravitational energy entirely into heat. This gives the dependence of the effective temperature on radius as $t_{\text{eff}} \propto r^{-3/4}$. A spectrum is then calculated by dividing the disc into concentric annuli, calculating the spectrum emitted by each annulus and then summing them all together. Assuming that each annulus radiates like a blackbody, the $r^{-3/4}$ effective temperature distribution gives $F_\nu \propto \nu^{1/3}$ (Peterson, 1997), although it is unclear whether this is consistent with observations.

In our model we characterised the Big Blue Bump from $\sim 0.1 - 1\mu\text{m}$ as a broken power-law with three free parameters: a break-wavelength λ_{break}, a blue power-law index α_{blue} for wavelengths shorter than the break wavelength, and a red power-law index α_{red} for wavelengths longer than the break wavelength.

5.2.2   *Hot Dust*

At wavelengths longer than 1μm, emission from hot dust begins to dominate over emission from the accretion disc. The SED in this region is generally characterised either by a power-law ($\propto \lambda^{\beta_{\text{NIR}}}$), with $\beta \simeq 0.5$ (e.g. Richards et al., 2006a; Zhang et al., 2014), or by a blackbody at ~ 1300 K, thus peaking in the near-IR (e.g. Leipski et al., 2014). We modelled the hot dust emission using a simple blackbody:

$$F_\lambda = \frac{2hc^2}{\lambda^5} \frac{1}{e^{\frac{hc}{\lambda k_B T}} - 1} \tag{5.1}$$

The blackbody component has two free parameters: the temperature of the blackbody T_{BB} and the overall normalisation.

5.2.3   *Emission Lines*

Hundreds of emission lines are present in a typical AGN spectra. Some of the most prominent lines are shown in Figure 5.1. The emission line spectrum is taken from Maddox and Hewett, (2006), who extend the composite of Francis et al., (1991) to include the Hα (6560Å) and Paα (18750Å) emission lines. A single parameter, EL_{scale}, scales the equivalent widths of all emission lines equally:

$$F_\lambda = EL_{scale} \times \frac{F_{\lambda,el}}{F_{\lambda,cont}} \times F_\lambda \qquad (5.2)$$

where $F_{\lambda,el}$ is the line flux in the template, $F_{\lambda,cont}$ is the continuum flux in the template, and F_λ is the continuum flux in the model.

5.2.4   *Host Galaxy*

Emission from the host galaxy is important, particularly in the region around the 1μm inflection point in the quasar SED. While the host galaxies of bright quasars tend to be massive, bright ellipticals, the hosts of lower luminosity AGN can have disc components (e.g. Dunlop et al., 2003). Our model incorporates $z = 0$ Sa, Sb, Sc and elliptical-type templates from Mannucci et al., (2001), which for simplicity do not evolve with redshift. We characterise the relationship between the luminosity of the AGN L_{AGN} and the luminosity of the host galaxy L_{Gal} as a power-law

$$L_{Gal} = L_{AGN}^{\beta} \qquad (5.3)$$

with power-law index $\beta = 0.42$ (Maddox and Hewett, 2006). Dividing both sides of Equations 5.3 by the luminosity of the AGN gives the luminosity of the host galaxy relative to the luminosity of the AGN

$$\frac{L_{Gal}}{L_{AGN}} = L_{AGN}^{\beta-1} \qquad (5.4)$$

which for $\beta < 1$ decreases with increasing AGN luminosity. In a flux limited sample, the AGN luminosity will tend to increase with redshift and so the luminosity of the host galaxy relative to the luminosity of the quasar will decrease with increasing redshift. Hence, the contribution from the host galaxy to the total flux is important at low redshift, but becomes gradually less significant towards higher redshifts.

Since the contribution from the host galaxy to the flux changes as a function of AGN luminosity, and hence redshift, we choose a reference redshift z_{nrm} where we set the fractional contribution of the

host galaxy to the total flux, η. In an arbitrary region of the spectrum (we use $4000 - 5000$ Å) we calculate both the AGN continuum flux $F_{AGN}(z_{nrm})$ and the flux from our host galaxy template spectrum $F_{Gal}(z_{nrm})$. The fractional contribution from the host galaxy to the total flux is then:

$$\eta = \frac{CF_{Gal}(z_{nrm})}{F_{AGN}(z_{nrm}) + CF_{Gal}(z_{nrm})} \tag{5.5}$$

where the constant C is the factor by which we must multiply the unnormalised galaxy spectrum in order for Equation 5.5 to hold true. Rearranging for the constant C we find

$$C = \frac{\eta}{1 - \eta} \frac{F_{AGN}(z_{nrm})}{F_{Gal}(z_{nrm})} \tag{5.6}$$

Hence at redshift z_{nrm} the host galaxy flux we add to our rest frame quasar continuum is

$$F_\lambda = \frac{\eta}{1 - \eta} \frac{F_{AGN}(z_{nrm})}{F_{Gal}(z_{nrm})} F_{\lambda,Gal} \tag{5.7}$$

where $F_{\lambda,Gal}$ is our host galaxy template spectrum in the quasar rest frame. The contribution from the host galaxy at a different redshift z is given by

$$
\begin{aligned}
F_\lambda &= \frac{\eta}{1 - \eta} \frac{F_{AGN}(z)}{F_{Gal}(z)} \frac{F_{AGN}(z_{nrm})}{F_{Gal}(z_{nrm})} \left(\frac{F_{AGN}(z)}{F_{Gal}(z)} \right)^{-1} F_{\lambda,Gal} & (5.8) \\
&= \frac{\eta}{1 - \eta} \frac{F_{AGN}(z)}{F_{Gal}(z)} \frac{F_{AGN}(z_{nrm})}{F_{Gal}(z_{nrm})} \frac{F_{Gal}(z)}{F_{AGN}(z)} F_{\lambda,Gal} & (5.9) \\
&= \frac{\eta}{1 - \eta} \frac{F_{AGN}(z)}{F_{Gal}(z)} \frac{L_{AGN}(z_{nrm})}{L_{Gal}(z_{nrm})} \frac{L_{Gal}(z)}{L_{AGN}(z)} F_{\lambda,Gal} & (5.10) \\
&= \frac{\eta}{1 - \eta} \frac{F_{AGN}(z)}{F_{Gal}(z)} \frac{L_{AGN}(z_{nrm})}{L_{AGN}(z_{nrm})^\beta} \frac{L_{AGN}(z)^\beta}{L_{AGN}(z)} F_{\lambda,Gal} & (5.11) \\
&= \frac{\eta}{1 - \eta} \frac{F_{AGN}(z)}{F_{Gal}(z)} \left(\frac{L_{AGN}(z)}{L_{AGN}(z_{nrm})} \right)^{\beta - 1} F_{\lambda,Gal} & (5.12)
\end{aligned}
$$

We need to know how the luminosity of the AGN depends on redshift. This is given by:

$$\frac{L_{AGN}(z)}{L_{AGN}(z_{nrm})} = 10^{-0.4(M_{AGN}(z) - M_{AGN}(z_{nrm}))} \tag{5.13}$$

where $M_{AGN}(z)$, the absolute magnitude of an AGN at redshift z, is given by

$$M(z) = m - 5(\log_{10} D_L(z) - 1) \tag{5.14}$$

and $D_L(z)$ is the luminosity distance to a source at redshift z in parsecs. Hence:

$$\frac{L_{AGN}(z)}{L_{AGN}(z_{nrm})} = 10^{-0.4(M_{AGN}(z) - M_{AGN}(z_{nrm}))} \tag{5.15}$$

$$= 10^{\left(\log_{10}\left(\frac{D_L(z)}{D_L(z_{nrm})}\right)^2\right)} \tag{5.16}$$

5.2.5 *Lyman-α Forest Absorption*

The optical spectra of high redshift quasars show hundreds of sharp absorption lines, which mostly correspond to the redshifted neutral hydrogen Lyα 1216Å transition. These absorption features are collectively referred to as the *Lyman-α forest*. To simulate the effect of Lyman-α forest absorption on our model SED we use the parametrisation of Becker et al., (2013), who derived an analytic function for the effective optical depth τ_{eff} over the redshift range $2 < z < 5$ made using 6065 quasar spectra from SDSS DR7. In their model the effective optical depth τ_{eff} is given by

As Paul how he implements this. My code for this can easily be sped up.

$$\tau_{eff} = \tau_0 \times \left(\left(\frac{1+z}{1+z_0}\right)^b + C\right) \tag{5.17}$$

where,

$$
\begin{aligned}
t_0 &= 0.751 \\
b &= 2.9 \\
C &= -0.132 \\
z_0 &= 3.5
\end{aligned}
$$

The transmitted flux $F_{\lambda,trans}$ at redshift z is then given by

$$f_{\lambda,trans} = F_\lambda \times e^{-\tau_{eff}} \tag{5.18}$$

An absorption line at λ_{abs} in the rest-frame of an AGN at redshift z_{AGN} has wavelength

$$(1 + z_{AGN})\lambda_{abs} \tag{5.19}$$

in the rest frame of an observer on Earth. In the rest-frame of a cloud of neutral hydrogen at redshift z_{cloud} the absorption line has wavelength

$$\frac{(1 + z_{\mathrm{AGN}})\lambda_{\mathrm{abs}}}{(1 + z_{\mathrm{cloud}})} \tag{5.20}$$

and so to absorb Lyman-α at 1216 Å the gas cloud must be at a redshift

$$z_{\mathrm{cloud}} = \frac{(1 + z_{\mathrm{AGN}})\lambda_{\mathrm{abs}}}{1216\text{Å}} - 1 \tag{5.21}$$

For every wavelength $\lambda_{\mathrm{abs}} < 1216$ Å in the rest-frame of an AGN at redshift $z > 2$ we calculate z_{cloud} using Equation 5.21 and then calculate the transmitted flux at λ_{abs} by substituting z_{cloud} in to Equations 5.17 and 5.18.

5.2.6 *Lyman-Limit Systems*

Ask Paul if this should be implemented

Lyman-limit systems are clouds of HI which are optically thick at the Lyman limit (912Å), which generally implies a neutral hydrogen column density $N(\mathrm{HI}) > 10^{17}\mathrm{cm}^{-2}$. Photons at wavelengths shorter than the Lyman-limit will be absorbed, which creates a sharp break in the observed continuum. We model the effect of a Lyman-limit system at the redshift of the quasar by setting the flux at wavelengths less that 912Å in the quasar rest frame to zero.

5.2.7 *Dust Extinction*

The selection criteria of the SDSS DR7Q catalogue, and particularly the DR10Q catalogue, are sensitive to quasars with moderate amounts of dust reddening (possibly as high as E(B-V) ~ 0.5; Richards et al., 2003) at the redshift of the quasar, and so we included the effect of dust extinction in our model. We considered four types of extinction curve: the Large Magellanic Cloud (LMC), Small Magellanic Cloud (SMC), Milky-Way (MW) extinction curves from Pei, (1992) and an extinction curve appropriate for the quasar population which has been derived by Paul Hewett. To derive the quasar extinction curve, UKIDSS photometry was used to provide an E(B-V) estimate, via the magnitude displacement of each quasar from the locus of unreddened objects. At redshifts $2 < z < 3$ the reddening measure is made at rest-frame wavelengths 3500-7000Å, where Galaxy, LMC and SMC extinction curves are very similar. The SDSS spectra of the same objects are then employed to generate an empirical extinction curve in the ultraviolet, down to 1200Å. The resulting curve has no 2200Å feature and rises rapidly with decreasing wavelength but is not as steep as the SMC curve. The extinctions curves give the colour excess $E(B - \lambda)$ relative to the colour excess $E(B - V)$ as a function of

wavelength λ. The colour excess $E(B - V)$ is related to the extinction in the V band, $A(V)$, via a parameter R,

$$A(V) = R \times E(B - V) \tag{5.22}$$

where $R = 3.1$ in the MW and $R \simeq 3$ in the Magellanic Clouds. Hence the extinction at a wavelength lambda $A(\lambda)$ is

$$A(\lambda) = E(B - V) \times \left[\frac{E(\lambda - V)}{E(B - V)} + R \right] \tag{5.23}$$

where the colour excess $E(B - V)$ is a free parameter in our model. The attenuation of the flux at a given wavelength is then:

$$F_\lambda = F_\lambda 10^{-A(\lambda)/2.5} \tag{5.24}$$

in the rest frame of the quasar.

5.3 THE 'STANDARD' SED MODEL

We will begin by deriving a 'standard' SED model by constraining a single set of parameters with a large sample of $0.2 < z < 4$ quasars encompassing a range of luminosities, accretion rates etc. The free parameters in our model are the blue power-law slope, the red power-law slope, the power-law break wavelength, the blackbody temperature, the blackbody normalisation, the emission line equivalent width scaling, and the fractional contribution from the host galaxy to the total flux. The reddening E(B-V) is fixed to zero, since a large fraction of SDSS quasars have very small amounts of dust reddening (Richards et al., 2003). For the host galaxy we use a Sb-type template derived by Mannucci et al., (2001). With some choice of initial parameters, we generate a set of model observed spectra at redshifts from $z = 0.25$ to $z = 3.75$ in intervals of $\Delta z = 0.1$. We then transform our set of model spectra into a set of model $ugrizYJHKW1W2$ SEDs

The throughput functions of the SDSS $ugriz$, UKIDSS YJHK and WISE $W1W2W3$ band-passes are shown in Figure 5.2, along with our model AGN spectra at three different redshifts. The mean flux density in a band-pass P is given by

$$f_\lambda(P) \quad = \quad \frac{\int P(\lambda) f_\lambda(\lambda) \lambda d\lambda}{\int P(\lambda) \lambda d\lambda} \tag{5.25}$$

where $P(\lambda)$ is the dimensionless throughput function of the band-pass. The corresponding magnitude, $m_\lambda(P)$, is then

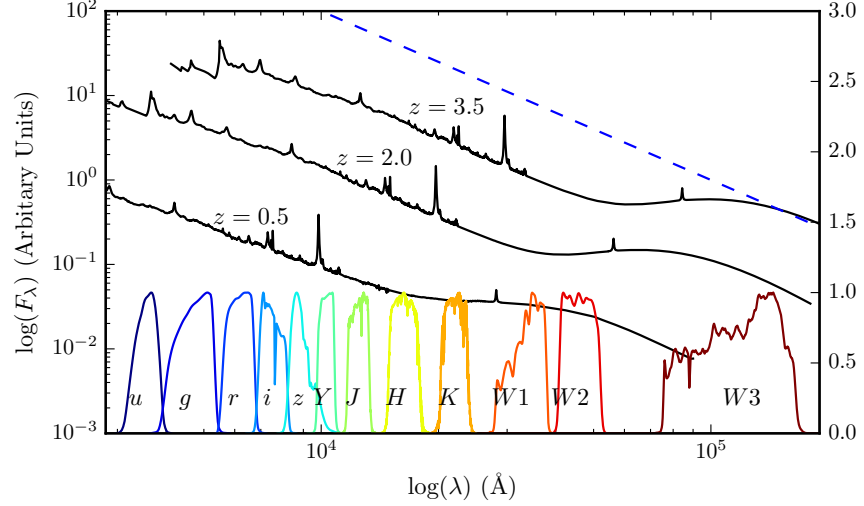$$m_\lambda(P) \quad = \quad -2.5 \log(f_\lambda(P)) - m_0(P) \tag{5.26}$$

Figure 5.2: Model spectrum at three different redshifts (each arbitrarily scaled), and throughput functions for SDSS, UKIDSS and WISE band-passes (scaled so that the peak transmission is equal to one.) The dashed line indicates the slope of the AB magnitude system zero point.

where $m_0(P)$ is the zero-point magnitude of band P. In the AB magnitude system, the zero-point flux per unit wavelength is

$$\frac{f_\lambda(\lambda)}{\text{erg cm}^{-2}\text{ s}^{-1}\text{Å}^{-1}} = 0.1087 \left(\frac{\lambda}{\text{Å}}\right)^{-2}. \tag{5.27}$$

This is substituted into Equation 5.25 to give a zero-point mean flux density which is then converted into a corresponding magnitude.

The model SEDs are normalised such that the i magnitude of each model SED is 18.0 mag. This gives us an array of model magnitudes as a function of redshift and band-pass. We generate an equivalent data array by dividing our quasar sample into redshift bins from $z = 0.2$ to $z = 3.8$ with bin width $\Delta z = 0.1$. We normalise the individual quasar SEDs such that the observed i magnitude is equal to 18.0 mag, and then calculate a median SED in each redshift bin.

To fit the model to the data we minimise the sum of the squares of the differences between the elements in the model magnitude array and the elements in the data magnitude array. The minimisation is done using the 'nelder-mead' algorithm. Our SED model is valid only up to $\lambda \sim 3\mu m$ in the quasar rest frame (the approximate wavelength of the peak in hot dust emission); beyond this additional contributions to the total flux from cooler dust will become significant. This prevents us from using the two highest wavelength WISE bands in the fit. We also exclude the SDSS u and g band-passes from the fit at $z > 2.7$ and $z > 3.7$ respectively, where absorption in the Lymanα forest becomes large.

Parameter	Symbol	Before Correction	After Correction
Blue power-law index	α_{blue}	0.58	0.58
Red power-law index	α_{red}	-0.04	-0.05
Power-law break	λ_{break}	2945	2957
Blackbody temperature	T_{BB}	1216 K	1186 K
Blackbody normalisation	C_{BB}	0.22	0.21
Emission line scaling	C_{EL}	0.63	0.73
Galaxy fraction	η	0.29	0.28
E(B-V)	E(B-V)	0.00	0.00

Table 5.1: Best-fitting parameters from fit to DR7Q-matched sample. Only give best-fit values after correction.

The best-fitting parameters from the fit are shown in Table 5.1. The colours ($u - g$, $g - r$, etc.) of the median SED, the individual quasars, and the best-fitting model are plotted as a function of redshift in Figs. 5.3 and 5.4. Most of the large variations that can be seen in the median colours of the quasars as a function of redshift are due to strong emission lines being redshifted in to and out of the bandpasses of the band-passes being used.

Re-do fit

5.4 DISCUSSION OF FIT

In Figure 5.5 we show the difference between the magnitudes from the best-fitting model and the median magnitudes from the sample. We have transformed the effective wavelengths of the band-passes to the rest frame of the quasars in each redshift bin, to give to the residuals as a function of rest-frame wavelength. We represent the residuals measured in each band-pass using a different coloured line. Differences between residuals from different band-passes at the same rest-frame wavelength could indicate redshift evolution of the typical quasar SED.

The residuals indicate that over a large redshift range the model does a fairly good at reproducing the median observed colours of the DR7Q-matched sample. Most discrepancies are at the < 0.1 mag level. It is remarkable that a single model is so effective; the properties of a typical quasar to not change significantly over a wide range of redshifts and luminosities. On the other hand, for the individual objects there is a significant scatter about the mean. In general, our goal is to use this intrinsic spread in SED properties in order to understand the diversity in physical quasar properties.
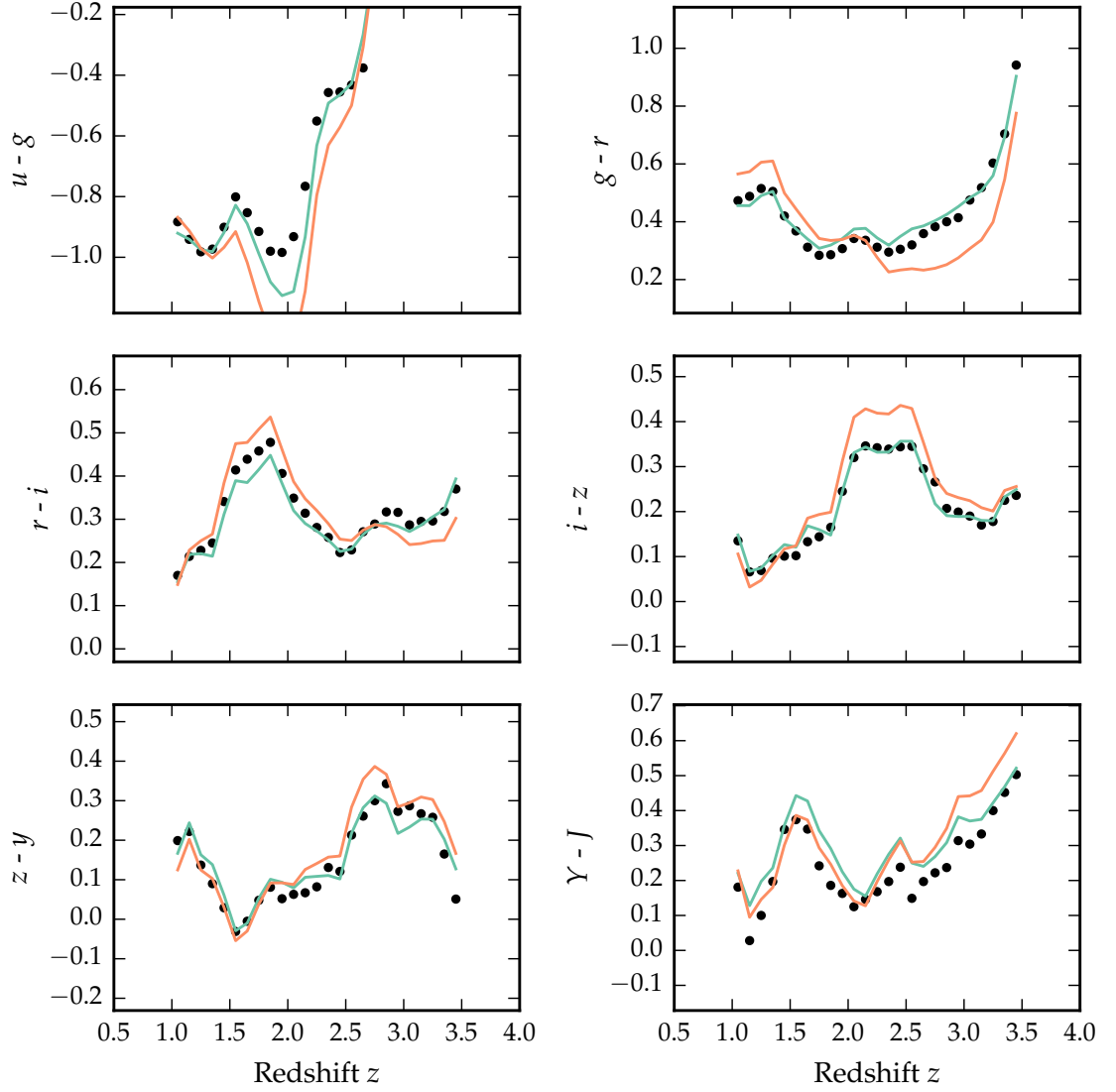
Figure 5.3: Colours of median SED (*black circles*), individual objects (*grey points*), best-fitting model (*black line*) as a function of redshift.
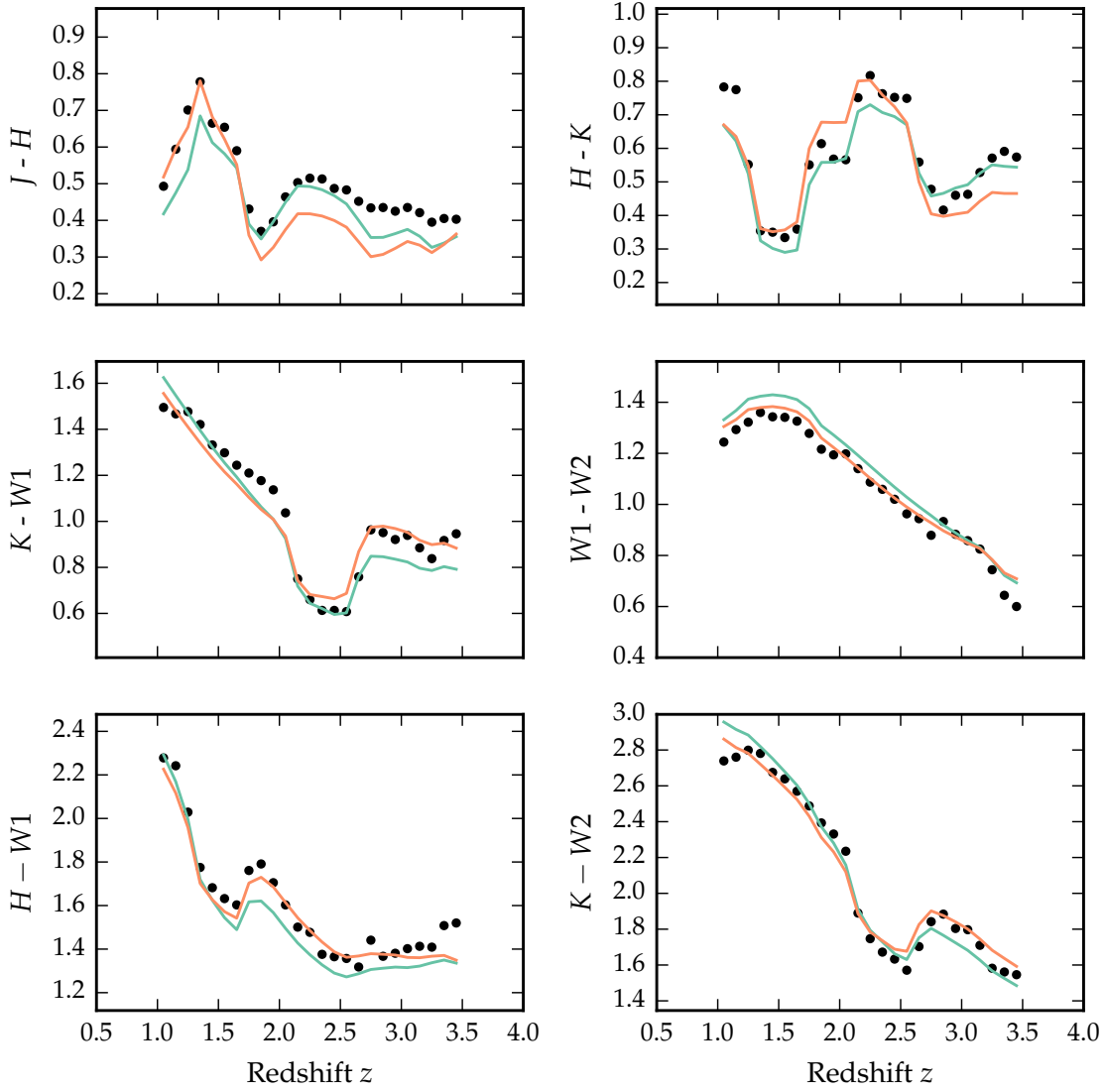
Figure 5.4: Colours of median SED (*black circles*), individual objects (*grey points*), best-fitting model (*black line*) as a function of redshift.
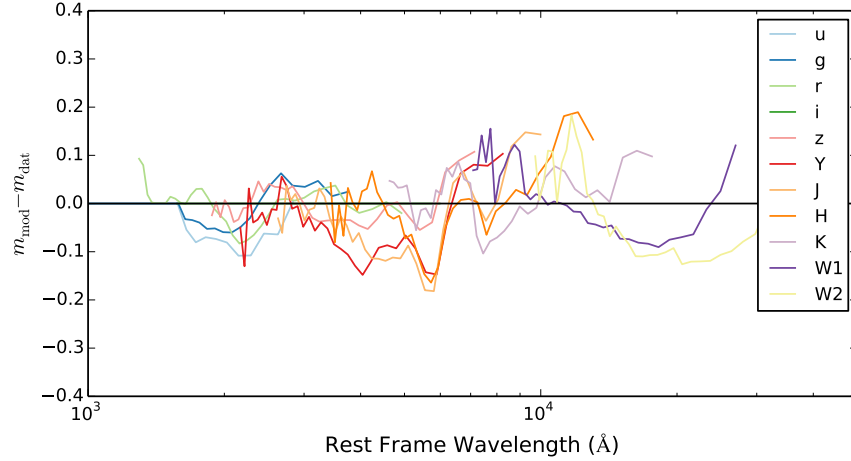
Figure 5.5: Residuals from fit to DR7Q-matched catalogue as a function of rest-frame wavelength.

5.4.1 *Flux Correction*

5.5 HOT DUST

Including a black-body with T~1250K, a simple parametric model matches the ugrizYJHKW1W2 (SDSS+UKIDSS+WISE) median colours of luminous quasars at redshifts $0.2 < z < 0.4$ extraordinarily well. The spread in the KW1W2 colours (Figure 5.8), probing the rest-frame ~1-2 micron region, is significant and strongly suggests presence of real variation in the "hot dust" temperature and luminosity among the quasars. Several other investigations have drawn attention to the rest-frame near-infrared SEDs, with populations of 'dust free' objects postulated (Hao et al., 2010; Hao et al., 2011; Jiang et al., 2010; Mor and Trakhtenbrot, 2011)

Reverberation measurements of nearby AGNs suggest that hot dust very close to the central source (few tens of light days; e.g. Minezaki et al., 2004; Suganuma et al., 2006). The hot dust signature could contain information about inner face of an obscuring torus structure and/or constrain the dust content of an accretion disk wind. Several studies have shown that the luminosity of the NIR excess emission correlates with that of the central engine with a slope close to unity e.g. Gallagher et al., 2007, suggesting that the dust is reprocessing radiation from the accretion disc.

Outflows may emerge from the outer region of the accretion disc or even the innermost region of the torus, in which the gas clouds are dusty and relatively cold. Indeed, there is observational evidence for dusty outflows close to the central engine (e.g. Bowler et al., 2014). The dust is heated by the central engine, and radiates in the near-infrared band. Wang et al., (2013), fitting the NIR emission with a

single power-law, found that objects with strong outflow signatures (blue-shifted C IV) have more hot dust emission relative to the accretion disc emission in a large sample of $z \sim 2$ non-BAL quasars. It could be that this correlation is induced by a third factor that simultaneously affects outflows and dust emission, for instance the inclination angle or metallicity. Alternatively the dust could be intrinsic to outflows and may have a non-trivial contribution to the outflow acceleration.

5.5.1   *Parameterising the hot dust emission*

We characterise the hot dust properties of our sample in terms of the temperature and luminosity of a blackbody. We choose to parameterise the luminosity in terms of the NIR to UV luminosity ratio (which is proportional to the covering factor of hot dust (L_{NIR}/L_{Bol}) used in other studies (Roseboom et al., 2013). The UV and NIR luminosity are calculated between 2000 and 9000Åand 1 and 3 μm respectively.

In Figure 5.6 we see that the two parameters are clearly correlated. For a lower temperature black-body the NIR to UV luminosity ratio is larger. Such a correlation is to be expected: as the black-body temperature is lowered, the peak shifts to longer-wavelengths (following Wien's displacement law). Because of this degeneracy we need to be very careful to seperate out real trends of $R_{NIR/UV}$ with other quasar properties from indirect trends resulting from a mutual dependence on T_{BB}.

Some previous studies (e.g. Wang et al., 2013; Zhang et al., 2014) have instead paramterised the near-IR emission using a power-law. We tested this parameterisation, and evaluated it's effectiveness relative to using a black-body. The power-law is normalised at 9000A, where its flux is set equal to the flux of the UV/optical model. The NIR power-law slope is fit between ∼1 and 2.4μm (with the exact wavelength region being fit depending on the redshift of the quasar). We found large residuals in the best-fitting model which varied systematically as a function of $\lambda_{eff}/(1+z)$. This suggests that the power-law model is a poor fit to the shape of the near-IR emission. One needs to take care in looking at trends with luminosity given the observed-frame passband information on the rest-frame SED can produce some strong systematics with redshift, particularly if the SED-model is not a good fit to the actual SED. A similar conclusion was reached by Gallagher et al.

5.5.2   *Sample*

Our goal is to determine the temperature and abundance of the hot dust component in individual quasars. These properties will be mea-
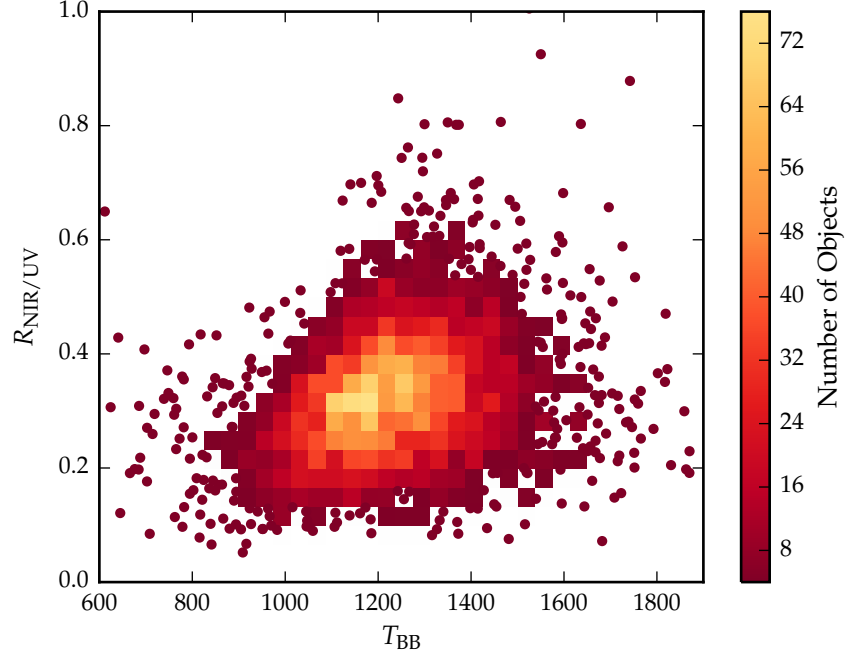
Figure 5.6: Ratio of NIR to UV luminosity ($R_{NIR/UV}$) against temperature (T_{BB}) for low-z sample. The density of points is shown in more dense regions of the space, and individual objects in less dense regions.

sured by fitting a model to the SDSS-UKIDSS-WISE photometry. Constraing a T~1200K blackbody component in the SED model requires photometric data covering ~1-3µm in the rest-frame of the quasar.

The observed-frame wavelength coverage of the available passbands limits the redshift range of the quasars which can be used. One does need to take care in looking at trends with luminosity given the observed-frame passband information on the rest-frame SED can produce some strong systematics with redshift, particularly if the SED-model is not a good fit to the actual SED. We consider only quasars at redshifts $z > 1$ where the relative host galaxy contribution to the SED is negligible. At redshifts $1 \lesssim z \lesssim 1.5$ the available ugrizYJHKW1W2 photometry provides good coverage of the rest-frame SED up to ~2µm. At $z \sim 1.5$ the W2 passband is shifted to ~1.8µm; at higher redshifts the wavelength coverage of the W2 band becomes much less than the peak wavelength of a T~1200K blackbody and experiments showed that such a component can not be adequatly constrained by the available photometry. Quasars in the redshift interval $1.5 < z < 2$ are therefore excluded from our sample.

For the quasars at $z \sim 1$ the WISE W3 band is probing rest-frame wavelengths of $\sim 5 - 6$µm. This region of the SED is dominated by emission from cooler, more distant dust, which is not accounted for in our model. However, at redshifts $z \gtrsim 2$ the WISE W3 passband

probes sufficiently short wavelengths to be useful in constraining the shape of the hot blackbody component. Therefore for quasars at redshifts $z > 2$ we again have sufficient constraints from the ugrizYJHKW1W2W3 photometry to determine the temperature and normalisation of the blackbody component. There are few objects in our sample with redshifts $z > 2.7$, and so we set this as an upper limit on the redshift of our sample. Because of these constraints, our sample is divided in to two parts: one at low redshifts ($1 < z < 1.5$) and the other at higher redshifts ($2 < z < 2.7$).

We include only quasars with observed magnitudes brighter than 19.1 in the i band-pass, i.e. the quasars selected by the main SDSS quasar selection algorithm (70,214 quasars). Cross-matching (with a $2''$ radius and picking only the nearest neighbour) the SDSS DR7Q catalogue with the ULAS catalogue, which covers only $\sim 38\%$ of the SDSS foot-print, resulted in 37,886 matches. Of these 36,628 have been detected in one or more of the WISE band-passes. We exclude quasars flagged as broad-absorption line quasars from the sample (leaving 35,272 quasars). We impose a lower-limit signal-to-noise ratio (S/N) > 5 magnitudes in the K, $W1$ and $W2$ band-passes for the low-z sample and S/N > 5 in the $W1$, $W2$, and $W3$ band-passes for the high-z sample to ensure reliable photometry. This gives us 5,910 quasars in our low-z sample and 1,989 quasars in our high-z sample.

We will hold most parameters fixed, and vary only those we are interested in, i.e. the blackbody parameters which parameterise the NIR emission. Therefore we need to define a sub-sample of objects which we know are well fit by our standard SED model in the UV/optical region. This means excluding objects with extreme emission line equivalent widths or significant dust extinction. We use the $i - K$ colours of the quasars as a measure of the overall colour of the quasars as it provides the longest baseline in wavelength without being affected by absorption in the Lyα forest at high redshifts. We discarded from our sample quasars with $i - K$ colors redder than our standard model with dust reddening E(B-V) = 0.075 and bluer than E(B-V) = -0.075 (Figure 5.7). Following this cut we are left with 4,615 quasars in our low-z sample and 1,692 quasars in our high-z sample.

5.5.3 *Diversity in hot dust properties*

In Figure 5.8 we plot the $W1 - W2$ colors of the DR7Q-matched sample as a function of redshift at z < 3. In this redshift range the $W1$ and $W2$ band-passes are probing the 1.2 - 2.8µm and 1.6 - 3.8 µ region of the rest frame SED respectively. For reference, the peak wavelength is at 2.4µm for a black-body radiating at 1200K (close to the sublimation temperature of dust grains). At any given redshift we see a ~ 0.5 mag dispersion in the $W1 - W2$ colors.
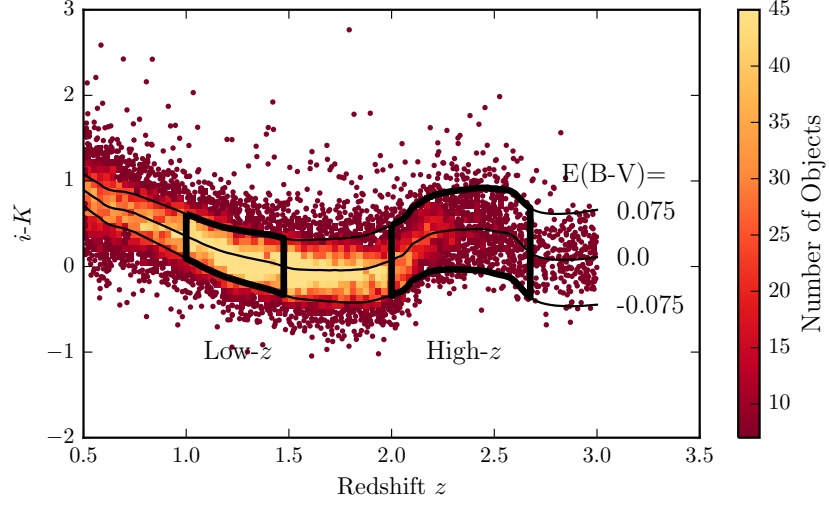
Figure 5.7: $i - K$ vs z. Demonstrates how sample was defined. The grey points show, as a function of redshift, the $i - K$ colours of all DR7Q quasars which are not classified as broad-absorption line quasars by Shen et al. and i magnitude > 19.1. The black line shows the $i - K$ colour of our standard, unreddened SED model as a function of redshift. The red and blue lines show the $i - K$ colours of our SED model with dust reddening E(B-V) = 0.075 and E(B-V) = -0.075 respectively. A significant amount of this reddening can be attributed to intrinsics variations in the UV power-law slopes of the indidual quasars, which is why we allow a negative reddening. However, there is a clear 'red tail' to the colour distribution which can be explained by dust reddending at the redshift of the quasar. We definined two samples, at low ($0.5 < z < 1.5$) and high ($2 < z < 2.7$) redshift, which are shown in the figure.
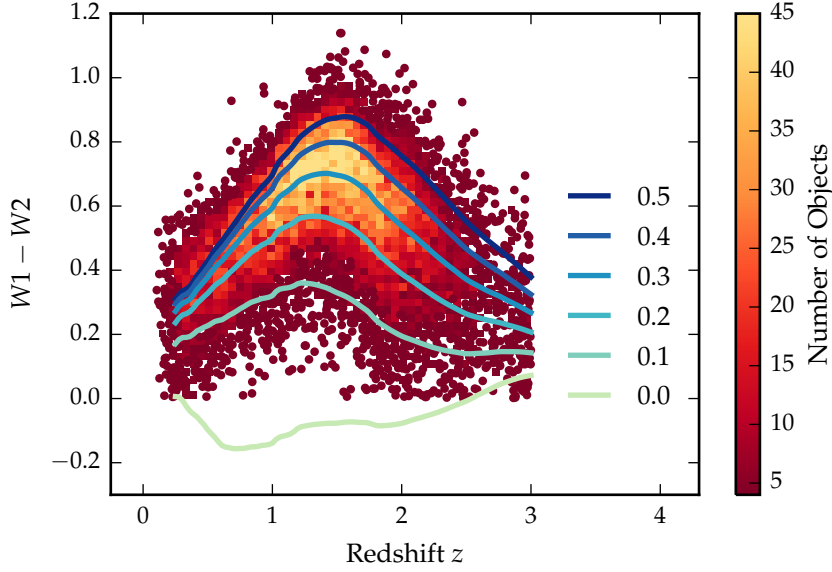
Figure 5.8: $W1 - W2$ colours of DR7 sample as a function of redshift. Above a certain density threshold points are represented by a density plot. On top we plot the colours of our standard SED model, with a fixed temperature and a varying NIR (1 - 3 μm) to UV ratio.

On the same axes in the Figure we have plotted the $W1 - W2$ colors derived from our SED model with a fixed blackbody temperature (1216K) and a ratio of NIR to UV luminosity ranging from 0.0 to 1.0, with the other model parameters held constant. We conclude that even with the sample restricted to be fairly uniform in its UV/optical properties, we still get an interesting spread in W1-W2 colors, which we can use to learn about the diversity of NIR properties in our sample. In the rest of this chapter we will characterise the hot dust properties of our sample, and test its relation to quasar properties such as luminosity, black-hole mass and normalised accretion rate, and outflow-properties.

In Figure 5.6 we show that there is quite a range of temperature and normalisation present in our sample. However, we need to check how much of this is due simply to uncertainties in the fits stemming from uncertainties in the photometry. In order to achieve this we took our standard SED model with a single temperature and normalisation black-body component, and generated 200 mock SEDs with a brightness distribution similar to that of our real sample. We estimated the mean uncertainty of the magnitudes in the K, W1, and W2 band-passes as a function of apparent brightness. We then sampled the K, W1, and W2 magnitudes from Gaussian distributions, with a mean equal to the magnitude of the model SED, and the width equal to the mean uncertainty at the appropriate brightness. Finally,
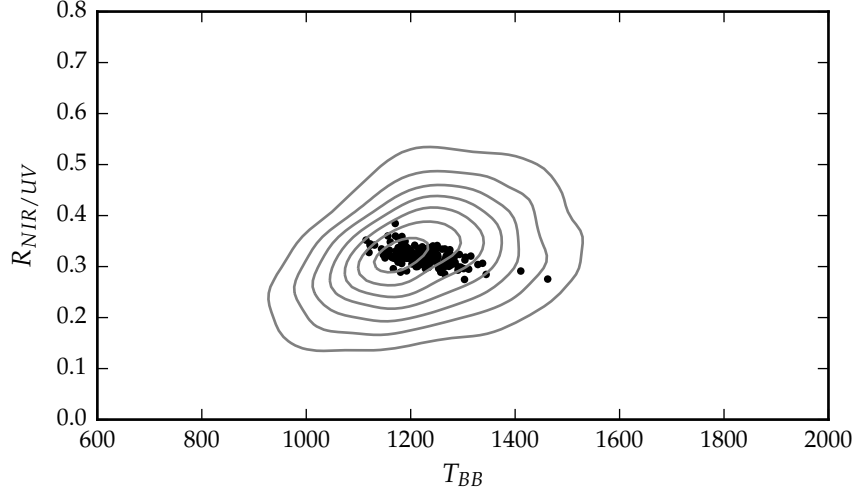
Figure 5.9: Ratio of NIR to UV luminosity ($R_{NIR/UV}$) against temperature (T_{BB}). The grey contours show equally-spaced lines of constant probability density generated using a Gaussian kernal-density estimator on our data sample. The black points are for our mock data.

we fit these mock SEDs using our standard fitting procedure. The results are shown in the Figure below, on top of the results from our real sample (shown as grey contours). We can see that uncertainty in the photometry introduces a significant scatter to the temperature, but that this scatter is less than the intrinsic scatter in the data. This demonstrates that there is a real distribution of hot dust temperatures and luminosities in our sample.

5.6   FITTING PROCEDURE

We will fit a model to the individual quasar SEDs, allowing the temperature and normalisation of the black body component to vary. The model spectrum is redshifted to the redshift of the quasar being fit and is then multiplied by the $ugrizYJHMW1W2W3$ throughput functions and normalised appropriately to give AB magnitudes. To fit the model to the data we minimise the sum of the squares of the differences between the elements in the model magnitude array and the elements in the data magnitude array. To avoid significant absorption in the Lyman-α forest at high-z, we restrict our fitting to wavelengths greater than 2000A; when the effective wavelength of a band-pass falls below this limit the band-pass is excluded from the fit. The minimisation is done using the 'nelder-mead' method, as implemented in the `minimize` function from the Python module `scipy`.

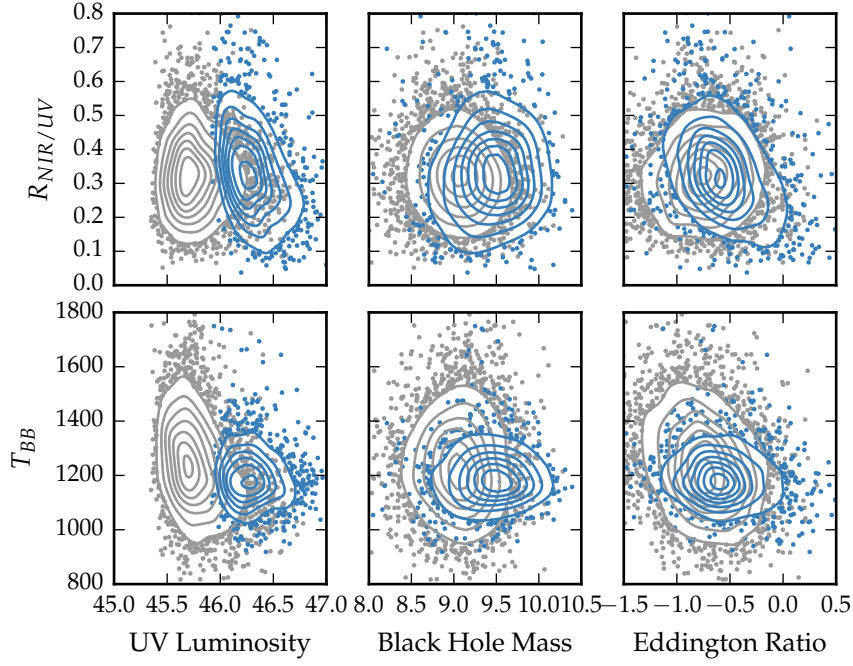*2000A is quite large given the Lyman-alpha forest impacts from 1216A.*

Figure 5.10: Best-fit black-body temperature against UV luminosity (left), black-hole mass (center) and Eddington ratio (right) for $1 < z < 1.5$ sample (black) and $2 < z < 2.7$ sample (black). In region of high-density we represent the density with contours generated using a Gaussian kernel density estimation. Needs re-making with new BH masses.

5.7 RESULTS

5.7.1 *Correlations with quasar properties*

We now look for correlations between the properties of the black-bodies we have fitted to the hot dust emission and other properties of the quasar such as redshift, black-hole mass, and normalised accretion rate (Eddington ratio).

Calculate new BH masses and redo this sectin.

5.7.2 *Spectral properties*

In the dusty wind model - first proposed by Konigl and Kartje, (1994) and later developed by, amongst others, Everett, (2005), Elitzur and Shlosman, (2006), Keating et al., (2012) - the 'torus' is the dusty part of a magneto-hydrodynamic wind beyond the dust sublimation radius. The MHD wind is roughly polar, and so the hot dust forms a vertical 'wall' around the accretion disk. UV photons from the accretion disk accelerate the wind via radiation line driving. That flattens the geometry of the wind and exposes more surface area that is viewable on a relatively face-on line of sight. The radiation pressure is increased
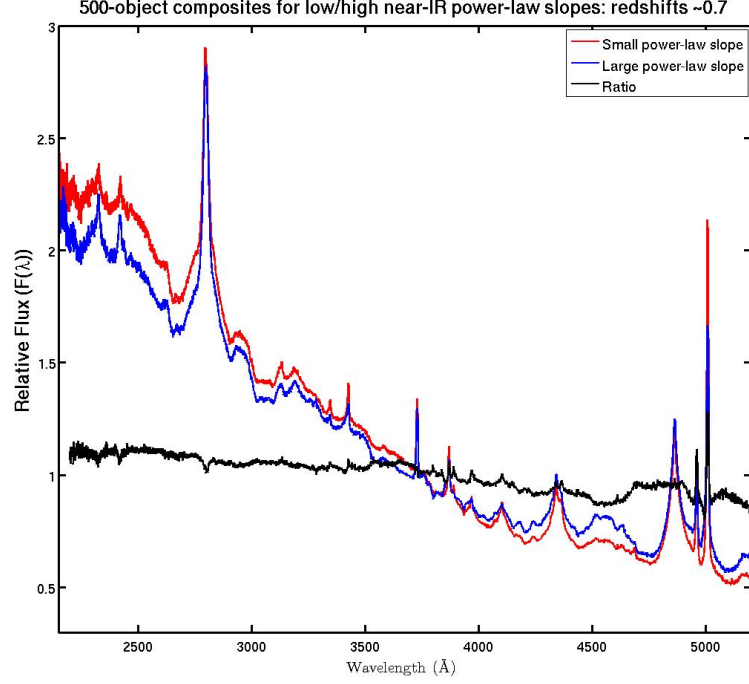
Figure 5.11: Composite SDSS spectra for objects at $z \sim 0.7$. We have divided sample into objects with objects best-fit by small (red line) and large (red line) values of β. Change this to select by $R_{NIR/UV}$ / T_{BB}. Label prominent emission lines.

at higher luminosities and/or accretion rates. This can flatten the geometry of the wind, thereby increasing the range of angles for which the inner edge of the dusty wind - where dust is at it's sublimation temperature - can be observed. A direct prediction is therefore that the in a quasars with high accretion rates and strong outflows, the emission from hot dust should be enhanced.

5.7.2.1 *Low-z*

The $z < 0.8$ SDSS spectrum composite comparison for the small and large β_{NIR} sub-samples is a very direct illustration of the Boroson and Green, (1992) Eigenvector 1 describing the spectral variation in the optical spectra of quasars; as Fe II EW increases the [O III] EW decreases. Hot dust emission increases with Fe II EW (Shen and Ho, 2014). We also note that the amount of hot dust correlates with the Si III/C III] emission ratios. The Si III/C III] ratio is generally considered to be a good indicator of density and is one of the primary EV1 correlates. The relative flux ratio of Si III to C III] increases when C IV is more blue-shifted (Richards et al., 2011). The Mg II emission line has exactly the same profile/shape for the two samples (apparent changes in Mg II seen in Fig. 5.11 are the result of changes in Fe II at
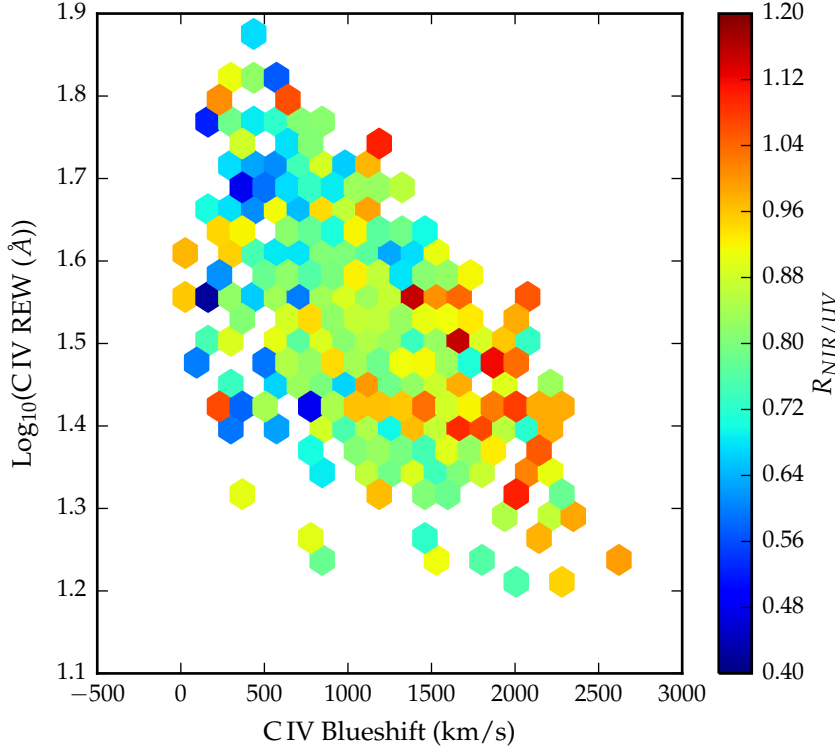
Figure 5.12: Rest-frame equivalent width and blueshift of the C IV line for 7,115 SDSS DR7 quasars. The colours of the hexagons denote the median hot dust (T≃1200 K) abundance for all quasars at a given equivalent width and blueshift. Quasars with the most extreme outflow signatures are predominantly hot-dust rich. Only bins containing a minimum of two objects are plotted.

wavelengths just shortward of the line). Finally, we note that objects with more hot dust are slightly redder.

5.7.2.2 *High-z*

In Fig. 5.12 we show how the ratio of NIR to UV luminosity depends on the blueshift and rest-frame equivalent width of the C IV line. C IV blueshifts are calclulated as in Section XX. We see that the NIR to UV luminosity ratio is strongly correlated with the blue-shift of the C IV emission line. A similar trend was noted by Wang et al., (2013). Interestingly, we note strong similarities to the object subsets selected according to their C IV-emission properties in Richards et al., (2011) (see Figures 11 & 12). We note that the correlation between the hot dust and the C IV emission properties will lead to apparent correlations between the host dust and the BH mass.

Need to re-do this and understand why beta-related trend is apparently stronger than with the blackbody parameters.
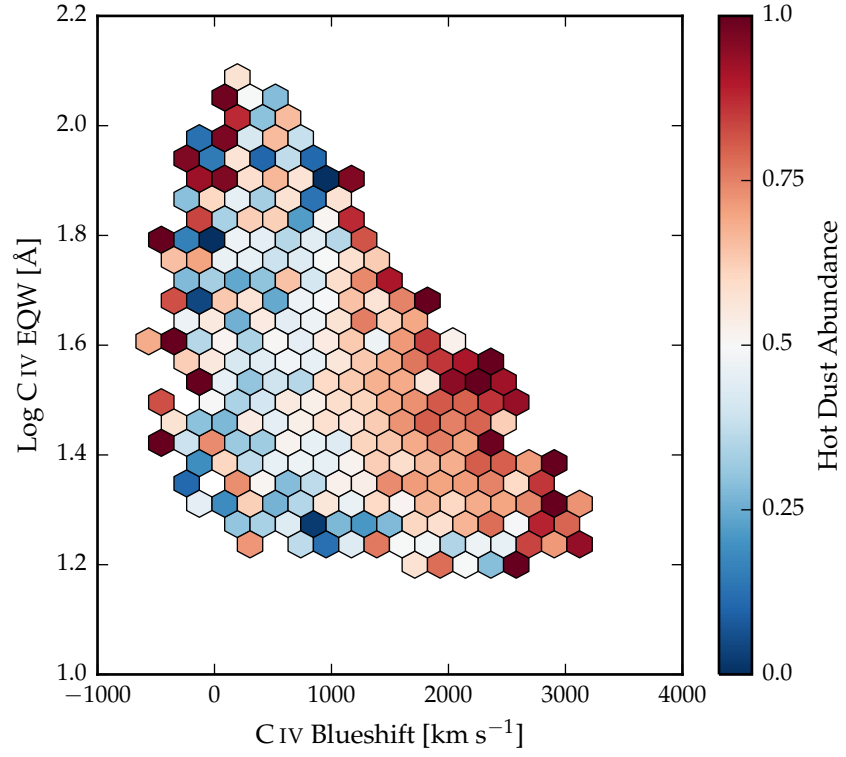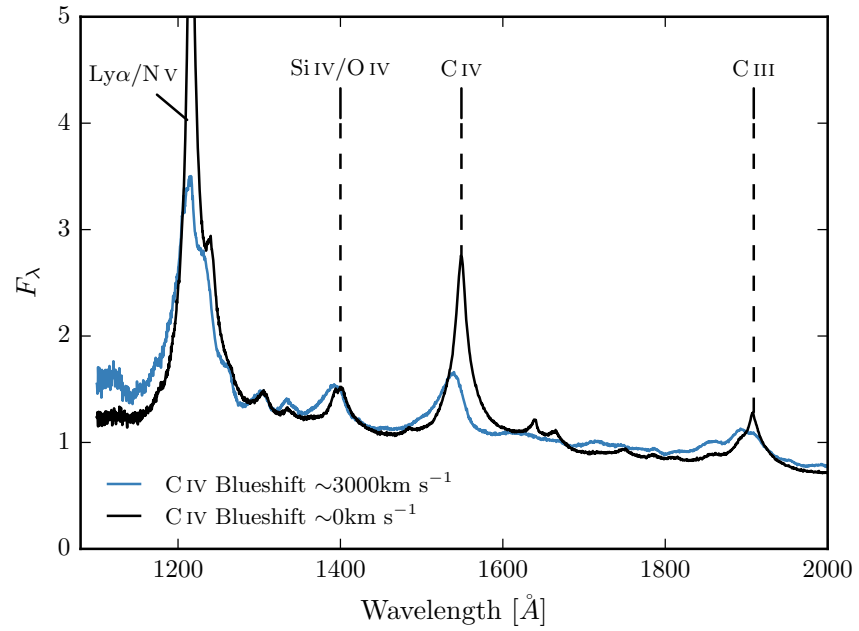
Figure 5.13

Figure 5.14

5.7.3   *BALs and radio-loud/radio-quiet*

In the spectra of about 20% of all quasars we observe broad (> 2000km/s) blue-shifted absorption troughs which are associated with quasar-driven out-flowing gas. BAL quasars in general have redder UV continua than non-BAL quasars, which is interpreted as the result of dust extinction. BAL quasars, on average, also have higher Eddington ratios and luminosities than non-BAL quasars.

We defined a sample of BAL quasars using the same method we used to define our sample of non-BAL quasars. At $1 < z < 1.5$ there are very few BAL quasars in our sample. In the $2 < z < 2.7$ redshift region we have 394 HiBAL quasars (the wavelength coverage of the SDSS spectra are not sensitive to LoBALs at these redshifts). Since BAL quasars are expected to suffer more from extinction due to dust, we have allowed E(B-V) to vary for the BAL quasar sample.

We find that the black-body temperature distributions are consistent (median T_{BB} for both samples = 1180K), but the ratio of NIR to UV emission is higher in BALs ($R_{NIR/UV}$ = 0.92 and 0.83 for BAL and non-BAL quasar sample respectively). This is qualatatively consistent with the results of Zhang et al., (2014).

It is well known that the blueshift of the C IV emission line in radio-quiet AGNs is, on average, stronger than in radio-loud AGNs (Marziani et al. 1996; Sulentic et al. 2000a; Richards et al. 2002, 2011). Statistically at least, the "radio-loud" objects are thought to have high black-hole masses and there is some form of radio-mode feedback (jet related) which is very different from the much more common (almost certainly wider opening-angle) outflow objects with large CIV-blueshifts.

So we find BALs have more hot dust, radio-loud have less. This is perfectly consistent with what we know about the positions of radio-loud objects and BALs in the C IV parameter space distribution (Richards et al., 2011).

Need to justify our use of model which has been optimised to fit colours of non-BAL quasars, when we know that BALs are typically redder.

What catalogue did we use to define quasar sample?

5.8   OTHER WORKS

Roseboom et al., (2013) studied a similar sample of luminous type 1 quasars. They, like us, modelled the NIR emission using a black-body and modelled the emission at longer wavelengths using a clumpy torus model. They find that while $L_{1-5\mu m}/L_{IR}$ appears relatively insensitive to L_{bol} and L_{IR}, a strong correlation appears between $L_{1-5\mu m}/L_{IR}$ and L_{IR}/L_{bol} (i.e. the dust covering factor). As the covering factor decreases, the maximum inclination at which a type 1 quasar would be seen increases. An increase in the inclination will mean direct sight lines to more of the inner wall of obscuring material closest to the accretion disc.

Mor and Trakhtenbrot, (2011) also looked at the hot dust properties of a sample of $0.75 < z < 2$ quasars, with photometry from SDSS and WISE. They modelled the NIR emission with hot clouds of pure graphite dust. They reported an anti-correlation between the covering factor of hot dust clouds and the quasar bolometric luminosity. Like us, they neglect cooler dust components which will dominate the SED at longer wavelengths. As we have discovered (see Figure residual plot), the missing flux decreases with redshift because we observe shorter rest-frame wavelengths when the observed spectrum is redshifted to a greater degree. This will induce an anti-correlation between the luminosity of the hot dust component and the luminosity of the quasar (which is correlated with redshift). At z=0.75, the W3 band-pass (the longest in their fits) is sensitive to flux from 6.9µm; at this wavelength we expect the contribution from cooler dust to dominate over the hot dust. It is possible that this effect could explain the tension with our own result that $R_{NIR/UV}$ does not depend on the quasar luminosity in our low-z sample.

Shen and Ho, (2014) quantify the relative torus emission using the $r - W1$ colour for a sample of $0.4 < z < 0.8$ SDSS quasars. At these redshifts W1 is observing between 1.9 and 2.4 microns in the rest-frame of the quasar, which suggests that they are sensitive to the same component of hot dust which we are investigating. They observe a mild trend of decreasing relative torus emission as the quasar luminosity increases. We note that their use of the r-W1 at much higher redshifts may be problematic, as the W1 flux will be increasingly dominated by direct emission from the accretion disc.

Gallagher et al., (2007) undertook a similar investigation for a much smaller sample of 234 radio-quiet quasars.

5.8.1 *Eddington ratio*

Wang et al., Zhang et al., and Mor & Trakhtenbrot find no significant dependence of the amount of hot dust on the Eddington ratio.

Shen & Ho find that torus emission is enhanced in quasars with larger R_{FeII}. They show how EW(OIII) and other high-ionisation lines (and to a lesser extent low-ionisation lines like MgII) anti correlate with R_{FeII}. The enhancement of torus emission relative to accretion disc emission at the high-RFeII end of EV1 may be caused by more efficient disc winds that facilitate the formation of a dusty torus. From our $z \sim 0.8$ composite SDSS spectra, we observed that objects with large NIR to UV luminosity ratios on average have stronger FeII emission.

CONCLUSIONS / FUTURE WORK

Put some stuff from research proposals here

6.1 FUTURE: RED QUASARS

Punctuated fuelling episodes, e.g. driven by galaxy mergers, satellite accretion and even secular processes, almost certainly lead to AGN experiencing activity-, outflow- and obscuration-dominated cycles with some overlap between phases. However, quantitatively, it remains unclear how these phases relate to the fundamental properties of the accreting black-hole (e.g. mass (M_{BH}), bolometric luminosity (L_{bol}) and Eddington ratio (L/L_{Edd}) and the elements of the non-spherical geometry).