# CONTENTS

In Figures 2.1, 3.16, 3.19 and 4.6 regions of high point-density, contours show equally-spaced lines of constant probability density generated using a Gaussian kernel-density estimator.

## LIST OF TABLES

## ACRONYMS

AGN  Active Galactic Nuclei

NLR  Narrow Line Region

BLR  Broad Line Region

EV1  Eigenvector 1

ICA  Independent Component Analysis

PCA  Principal Component Analysis

SDSS  Sloan Digital Sky Survey

BOSS  Baryon Oscillation Spectroscopic Survey

UV  Ultra-Violet

EQW  Equivalent width

S/N  Signal-to-noise ratio

BH  Black Hole

SED  Spectral Energy Distribution

IR  Infrared

NIR  Near-infrared

FWHM  Full-Width-at-Half-Maximum

# INTRODUCTION

## 1.1 THE DISCOVERY OF QUASARS

In 1963, the powerful radio source 3C 273 was identified with a star-like, thirteenth magnitude object with a strongly red-shifted[1] optical spectrum (Schmidt, 1963). This finding implied 3C 273 was at an enormous distance (at least by the standards of the time). For an object this distant to appear so bright it must be extremely luminous: 10 times more luminous, in fact, that the largest galaxies known. At the same time, its rapid variability meant that it could be no bigger than a light-week across. 3C 273 was therefore a new kind of exotic object, at the edge of the known Universe. Understandably, its discovery caused a lot of excitement, both for astronomers and the wider public. It was quickly realised that 'quasars'[2], and other lower-luminosity classes of active galactic nuclei (AGN)[3], are powered by the release of gravitational potential energy as mass is accreted onto a super-massive[4] black hole (BH) at the centre of a galaxy (e.g. Hoyle and Fowler, 1963; Salpeter, 1964; Lynden-Bell, 1969; Lynden-Bell and Rees, 1971).

## 1.2 THE AGN-HOST GALAXY CONNECTION

Beginning in the early 1990s, inactive super-massive BHs were found in the centres of many nearby massive galaxies (e.g. Kormendy and Richstone, 1995; Ferrarese and Ford, 2005; Kormendy and Ho, 2013). This proved that, rather than being rare and exceptional objects, quasar activity was in fact a stage in the life of all massive galaxies (e.g. Lynden-Bell, 1969). Shortly after, it was discovered that the BH mass was tightly correlated with properties of the host-galaxy bulge (e.g. the stellar velocity

---

1 Redshift $z = c(\Delta\lambda/\lambda) = 0.158$, where c is the speed of light.
2 The term 'quasar' originated as a contraction of 'quasi-stellar radio source', although 90 per cent of quasars are now known to be radio-quiet.
3 Throughout this thesis we use the terms 'quasar' and 'Active Galactic Nucleus (AGN)' interchangeably to describe active super-massive black holes, although the term quasar is generally reserved for the luminous ($L_{Bol} > 10^{12}L_{\odot}$) subset of AGN.
4 Super-massive: $10^{6-9}$ $M_{\odot}$.

dispersion, σ, which is proportional to the bulge mass; e.g. Ferrarese and Merritt, 2000; Gebhardt et al., 2000; Graham et al., 2001; Tremaine et al., 2002; Marconi and Hunt, 2003; Aller and Richstone, 2007; Gültekin et al., 2009). This was an unexpected finding, given that the sphere-of-influence[5] of a BH ($\lesssim 100$ parsecs; e.g. Kormendy and Ho, 2013), is many orders of magnitude smaller than the dimensions of a typical galactic bulge. The tight correlation between their masses suggested that the BH and the host-galaxy bulge grow synchronously. Both the density of quasars and the cosmic star formation history evolve strongly with redshift and peak at $2 \lesssim z \lesssim 3$ (e.g. Boyle and Terlevich, 1998; Brandt and Hasinger, 2005; Richards et al., 2006b). The similarity of their cosmic evolution was taken as further evidence for the existence of an intimate connection between quasars and their host galaxies.

In a currently favoured model, rapid BH fuelling and star-formation are triggered by a gas-rich galaxy merger (e.g. Hopkins et al., 2006), satellite accretion or secular processes (e.g. Fanidakis et al., 2012). The energetic output of the rapidly-accreting BH couples with the gas in the host-galaxy and regulates star formation and the growth of the BH itself (e.g. Silk and Rees, 1998; King, 2003; Di Matteo, Springel, and Hernquist, 2005; King and Pounds, 2015). This process, which is referred to as 'quasar feedback', is also commonly invoked to reproduce the high-mass end of the galaxy luminosity function in cosmological simulations (e.g. Kauffmann and Haehnelt, 2000). The insight that quasars may play a crucial role in the evolution of galaxies has led to an explosion of interest in their properties in recent years.

## 1.3 AGN: THE CURRENT PARADIGM

The basic features of the current AGN paradigm are widely accepted, although many of the details are unknown. The basic features are: a hot accretion disc surrounding a super-massive BH, rapidly orbiting clouds of ionised gas, and a dusty, obscuring structure (generally referred to as the 'torus'). Collimated jets of relativistic plasma and/or associated lobes are also seen in the 10 per cent of quasars that are radio-loud (e.g. Peterson, 1997). A cartoon picture illustrating the basic structure of an AGN is shown in Figure 1.1.

---

5 Sphere-of-influence: where the gravity of the BH dominates over the other mass components.
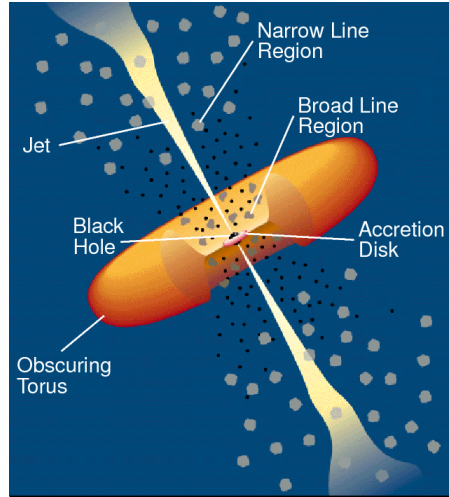
Figure 1.1: Cartoon picture of the inner regions on an AGN. Credit: Urry and Padovani, (1995).

### 1.3.1 Accretion disc

Material is pulled towards a super-massive BH and sheds angular momentum through viscous and turbulent processes in a hot accretion disc (e.g. Begelman, 1985). The accretion disc reaches temperatures of $\sim 10^6$ K, and radiates primarily at ultraviolet to soft-X-ray wavelengths.

### 1.3.2 Broad line region

One of the pre-eminent features of many AGN spectra are broad optical and ultra-violet emission-lines produced in the broad line region (BLR). The BLR consists of gas clouds at distances from several light-days to several light-months that are photo-ionised by the ultra-violet continuum emission emanating from the accretion disc. Because of the close proximity to the central super-massive BH, bulk motions are dominated by gravity and radiation pressure from the accretion disc. The very broad emission-line widths are assumed to Doppler-broadened, and imply line-of-sight velocities of many thousands of $km\,s^{-1}$.

### 1.3.3 Dusty torus

Farther out from the central engine on parsec-scales are dusty, molecular clouds which are co-planar with the accretion disc. These dusty clouds are generally referred to as the 'torus'. The torus is a central feature of orientation-based unification

schemes (e.g. Antonucci, 1993) in which differing observational properties are explained as the effect of observing anisotropic objects in different orientations. In a Type II AGN, the system is observed in an edge-on configuration and, as a result, emission from the accretion disc and BLR is obscured by the dusty torus. The orientation of a Type I AGN is such that the observe has direct sight-lines to the accretion disc and BLR. Although this simple picture (shown in Figure 1.1 as well as in countless other publications) is a useful starting point, the idea of a torus as a static, doughnut-like structure is almost certainly a gross over-simplification. For example, the problem of maintaining the large scale height required to explain the observed fraction of Type I/II AGN has long been recognized (e.g. Krolik and Begelman, 1988). In one set of more realistic models, the torus is a dusty wind blown from the accretion disc (e.g. Konigl and Kartje, 1994; Everett, Gallagher, and Keating, 2009; Gallagher et al., 2012; Everett, 2005; Keating et al., 2012; Elitzur and Shlosman, 2006).

### 1.3.4 *Narrow line region*

Farther away from the central BH and beyond the dusty torus is the spatially-extended narrow emission-line region (NLR). Like the BLR, the NLR is ionised by radiation from the accretion disc. Unlike the BLR, densities in the NLR are low enough that forbidden transitions are not collisionally suppressed. Emission-line widths are typically hundreds of $km\,s^{-1}$ in the NLR. The size of the NLR grows the AGN luminosity, and can reach kiloparsec scales in luminous quasars (e.g. Hainline et al., 2013).

## 1.4 THE ADVENT OF LARGE SURVEYS AND THE SYSTEMATIC STUDY OF AGN PROPERTIES

The Palomar-Green (PG) Bright Quasar Survey (BQS; Schmidt and Green, 1983), the first large-area quasar survey, identified 114 quasars via their ultra-violet excess relative to stars. Boroson and Green, (1992) were among the first to use the PG quasar sample to analyse quasar spectroscopic properties in a systematic way. In their landmark study, they used a principle component analysis (PCA) to identify the features responsible for the largest variance in quasar spectra. The first eigenvector of their PCA decomposition - generally referred to as 'eigenvector 1' (EV1) - is correlated with the FWHM of the broad Hβ emission-

line and the relative strengths of optical Fe II and Hβ. The underlying driver behind EV1 is thought to be the Eddington ratio[6].

With the advent of CCD[7] technology came a new generation of surveys, most notably the Sloan Digital Sky Survey (SDSS; York et al., 2000) and the 2QZ survey (Croom et al., 2004). SDSS, and the next-generation SDSS-III: Baryon Oscillation Spectroscopic Survey (BOSS; Dawson et al., 2013), now contain spectra of ~ 400 000 AGN and quasars. These large, uniform data sets have revolutionised the study of AGN and quasars by facilitating statistical studies of their properties covering wide ranges in redshift and luminosity.

Emission-lines provide a wealth of information about the properties of AGN and their environments. The optical region in particular includes a number of strong emission features, including the Balmer lines (Hα and Hβ) and the [O III] doublet that are used to measure BH masses, accretion rates, systemic redshifts and outflow properties. However, at $z \sim 2$, rest-frame optical lines are redshifted to near-infrared wavelengths. Near-infrared spectroscopy is therefore essential for a complete understanding of quasars during the peak epoch of galaxy formation. Spectroscopic observations are challenging at infrared wavelengths, and there are far fewer infrared observations of quasars than optical ones. However, in Chapter 2, we describe the construction of a near-infrared spectroscopic catalogue containing 462 redshift $1.5 < z < 4$ quasars. This is the largest sample of its kind, and has facilitated investigations of quasar BH masses and outflow properties which are described in Chapters 3 and 4.

## 1.5 MEASURING BLACK HOLE MASSES

The BH mass is one of the most important physical parameters of a quasar and considerable resources have been devoted to measuring the masses of BHs in active galaxies. Large-scale studies of AGN and quasar demographics have become possible through the calibration of single-epoch virial-mass estimators using results from reverberation-mapping campaigns (e.g.

---

6 Eddington ratio: $L/L_{Edd}$, where the Eddington luminosity $L_{Edd}$ (= 3.2 × $10^4(M/M_\odot)L_\odot$) is the maximum luminosity set by the balance of outward radiation pressure and inward gravitational force.
7 CCD: Charge coupled device.

Peterson, 2010; Vestergaard et al., 2011; Marziani and Sulentic, 2012; Shen, 2013).

### 1.5.1 *Reverberation mapping*

Under the assumptions that the BLR dynamics are virialised[8] and the gravitational potential is dominated by the BH, the BH mass is given by:

$$M_{BH} \simeq \frac{V_{Virial}^2 R_{BLR}}{G} \tag{1.1}$$

where $V_{virial}$ is the virial velocity in the BLR and $R_{BLR}$ the characteristic BLR radius. The problem of measuring the mass therefore reduces to the problem of measuring the velocity and orbital radius of the line-emitting clouds in the BLR.

Continuum variability is a common characteristic of quasars, owing to the stochastic nature of the accretion process. Because the BLR is photo-ionized by the continuum, the broad emission-lines also vary with some characteristic lag, which is related to the light travel time across the BLR. The reverberation mapping method, first proposed by Blandford and McKee, (1982), 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 (e.g. Peterson, 1993; Netzer and Peterson, 1997; Peterson, 2014).

The typical velocity in the BLR is measured from the Doppler-broadened width of an emission-line produced in the BLR. 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 into a virial velocity. Unfortunately, $f$ is unknown and likely varies from object to object. In practice, the value of $f$ is empirically determined by requiring that the reverberation-mapping masses are consistent with those predicted from the $M_{BH}$-$\sigma$ relation for local inactive galaxies.

Because reverberation mapping depends on temporal resolution rather than spatial resolution, this technique can be applied out to much greater distances than direct dynamical modelling (e.g. Kormendy and Ho, 2013). However, because reverberation

---

8 The virial theorem states that the average kinetic energy of a system is equal to half of the average negative potential energy.

mapping relies on dense spectrophotometric monitoring campaigns which span many years, the number of AGN with measured lags is limited to $\sim 50$ AGN (e.g. Kaspi et al., 2000; Peterson et al., 2004; Kaspi et al., 2007; Bentz et al., 2009; Denney et al., 2010; Barth et al., 2011; Grier et al., 2012). This sample is strongly biased to low luminosity Seyfert 1 galaxies[9], and the maximum redshift is just $z \sim 0.3$. Comprehensive statistical studies of active BHs, particularly during the epoch of peak galaxy formation ($z \gtrsim 2$), therefore require a different approach to measuring BH masses.

### 1.5.2 *Single-epoch virial estimates*

Reverberation mapping campaigns have also revealed a tight relationship between the radius of the BLR and the quasar optical (or ultra-violet) luminosity (the $R_{BLR} - L$ relation; e.g. Kaspi et al., 2000; Kaspi et al., 2007). A slope of $\simeq 0.5$ is found, which is consistent with naive predictions (e.g. Peterson, 1997). This relation provides a much less expensive method of measuring the BLR radius, and large-scale studies of AGN and quasar demographics (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) have thus become possible through the calibration of single-epoch virial-mass estimators using the reverberation-mapping measurements (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).

With single-epoch virial masses, the growth rate of massive BHs can be measured across cosmic time. This conveys important information about the accretion processes occurring in active BHs (e.g. Kollmeier et al., 2006) and is crucial in order to understand the processes responsible for establishing the $M_{BH}$-$\sigma$ relation (e.g. Bennert et al., 2011). Single-epoch virial estimates have been used to calculate BH masses in the highest redshift quasars (e.g. a $10^9 \, M_\odot$ BH in a redshift $z = 7.1$ quasar; Mortlock et al., 2011). Recent claims of a BH with mass $10^{10} \, M_\odot$ in a redshift $z = 6.3$ quasar (when the Universe is less than 1 Gyr old; Wu et al., 2015) challenges our understanding of the accretion histories of supermassive BHs (e.g. Willott, McLure, and Jarvis, 2003).

---

9 Seyfert 1: A low-luminosity ($L_{Bol} < 10^{12} L_\odot$) class of AGN with broad emission-lines and clearly detectable host-galaxies.

The uncertainties in reverberation mapped BH masses are estimated to be $\sim 0.4\,\mathrm{dex}$ (e.g. Peterson, 2010), and the uncertainties in virial masses are similar (e.g. Vestergaard and Peterson, 2006). However, the main concern and biggest unknown is the extension of the method to high redshifts. This requires that the relations calibrated for sub-Eddington BHs with $M_{BH} \sim 10^7\,M_\odot$ are valid for BHs with masses up to $10^{10}\,M_\odot$ that are radiating near the Eddington luminosity. Furthermore, the vast majority of reverberation mapping measurements are for H$\beta$ and so the $R_{BLR} - L$ relation that underpins the virial method has only been established using this line. H$\beta$ is redshift beyond the reach of optical spectrographs at redshifts $z \gtrsim 0.7$, and extending the method to higher redshifts requires the secondary-calibration of other low-ionization emission-lines such as H$\alpha$ and Mg II (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).

At redshifts of $z \gtrsim 2$ the low-ionization hydrogen and Mg II emission-lines are no longer present in the optical spectra of quasars and it is necessary to employ an emission-line in the rest-frame ultra-violet. The strong C IV emission doublet is visible in the optical spectra of quasars to redshifts of $z \sim 5$ and C IV-derived BH masses have become the standard (e.g. Vestergaard and Peterson, 2006; Park et al., 2013). However, C IV has long been known to exhibit significant displacements to the blue and these 'blueshifts' almost certainly signal the presence of strong outflows. As a consequence, single-epoch virial BH mass estimates derived from C IV velocity-widths are known to be systematically biased compared to masses from the hydrogen Balmer lines.

In Chapter 3, we use a sample of 230 high-luminosity ($L_{Bol} = 10^{45.5} - 10^{48}$ erg s$^{-1}$), redshift $1.5 < z < 4.0$ quasars with both C IV and Balmer line spectra to quantify the bias in C IV BH masses as a function of the C IV blueshift. C IV BH masses are shown to be a factor of five larger than the corresponding Balmer-line masses at C IV blueshifts of $3000\,\mathrm{km\,s^{-1}}$ and are over-estimated by almost an order of magnitude at the most extreme blueshifts, $\gtrsim 5000\,\mathrm{km\,s^{-1}}$. Using the monotonically increasing relationship between the C IV blueshift and the mass ratio BH(C IV)/BH(H$\alpha$), we derive an empirical correction to all C IV BH masses. The scatter between the corrected C IV masses and the Balmer masses is $0.24\,\mathrm{dex}$ at low C IV blueshifts ($\sim 0\,\mathrm{km\,s^{-1}}$) and just $0.10\,\mathrm{dex}$ at high blueshifts

($\sim 3000\,\mathrm{km\,s^{-1}}$), compared to 0.40 dex before the correction. The correction depends only on the C IV line properties - i.e. the full-width at half-maximum (FWHM) and blueshift - and can therefore be applied to all quasars where C IV emission-line properties have been measured, enabling the derivation of un-biased virial BH mass estimates for the majority of high-luminosity, high-redshift, spectroscopically confirmed quasars in the literature.

## 1.6 WINDS AND OUTFLOWS IN AGN

Quasars are very powerful sources of radiation, and are embedded in matter-rich environments at the centres of galaxies. Strong winds, driven by some combination of gas pressure, radiation pressure, and magnetic forces, are to be expected under these conditions (e.g. Blandford and Payne, 1982; Proga, Stone, and Kallman, 2000; Everett, 2005). In line with these expectations, evidence for outflowing gas is common in the spectra of quasars.

Perhaps the most dramatic evidence for outflows is seen in broad absorption-line quasars (BAL quasars; Weymann et al., 1991). BAL quasars are characterised by broad absorption features in the ultra-violet resonance lines of highly ionised N v, C IV and Si IV. The absorption troughs are thousands of $\mathrm{km\,s^{-1}}$ wide and significantly blueshifted relative to the quasar rest-frame[10]. The absorption is thought to occur in outflows reaching $60\,000\,\mathrm{km\,s^{-1}}$ (e.g. Turnshek, 1988). The near-universal blueshifting of the observed absorption features can be understood if the far-side of the outflow is obscured by the accretion disc, and so only the near-side, which is moving towards the observer, is detected. The observed C IV BAL fraction in radio-quiet quasars is $\sim$ 15 per cent (e.g. Hewett and Foltz, 2003; Reichard et al., 2003) and the intrinsic fraction in the quasar population has been estimated at $\sim$ 40 per cent (Allen et al., 2011). Outflows can also explain narrow ultra-violet and X-ray absorption-lines (NALs) which are seen in $\sim$ 60 per cent of Seyfert 1 galaxies (Crenshaw et al., 1999) and some quasars (e.g. Hamann et al., 1997). The blueshifting of high-ionisation lines in the BLR (including C IV) can also be understood if the lines are produced in outflowing clouds (although see, e.g., Gaskell and Goosmann 2016, for an alternative explanation).

---

10 Much rarer cases of redshifted BAL troughs do also exist (e.g. Hall et al., 2013).

The blueshifting of C IV appears to be nearly ubiquitous in the quasar population (e.g. Richards et al., 2002; Richards et al., 2011).

Together, these results suggest that outflows are very common and the energy released by quasars can have a dramatic effect on their immediate surroundings. Accretion-disc wind models have been developed to explain the wide range of emission and absorption-line phenomena which are observed (e.g. Murray et al., 1995; Elvis, 2000; Proga, Stone, and Kallman, 2000; Everett, 2005).

In models for the co-evolution of quasars and galaxies, the energy released by quasars impacts galaxies on much larger scales than is probed by the emission and absorption diagnostics described above. In recent years, a huge amount of resources have been devoted to searching for observational evidence of galaxy-wide, quasar-driven outflows (for recent reviews, see Alexander and Hickox, 2012; Fabian, 2012; Heckman and Best, 2014). This has resulted in recent detections of outflows in AGN host-galaxies using tracers of atomic, molecular, and ionised gas with enough power to sweep their host-galaxies clear of 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).

The advent of large optical spectroscopic surveys (e.g. SDSS) has facilitated studies of the NLR in tens of thousands of AGN at redshifts $z \lesssim 0.4$ which has provided constraints on the prevalence and drivers of ionised outflows (e.g. Mullaney et al., 2013; Zakamska and Greene, 2014). Because of its high equivalent width (EQW), [O III] is the most studied of the narrow AGN emission-lines. By following [O III] to near-infrared wavelengths, we are able to extend these investigations to the redshift range when star formation and BH accretion peaked. In Chapter 4, we analyse the [O III] properties of a sample of 354 high-luminosity, redshift $1.5 < z < 4$ quasars. To date, this is the largest study of the NLR properties of high redshift quasars.

Comparing our sample to SDSS AGN at lower redshifts and luminosities, we find [O III] to be significantly broader, more blue-asymmetric and weaker. The [O III] EQW is tightly correlated with the C IV blueshift, and we propose that the NLR gas is being swept away on relatively short time-scales by quasar-
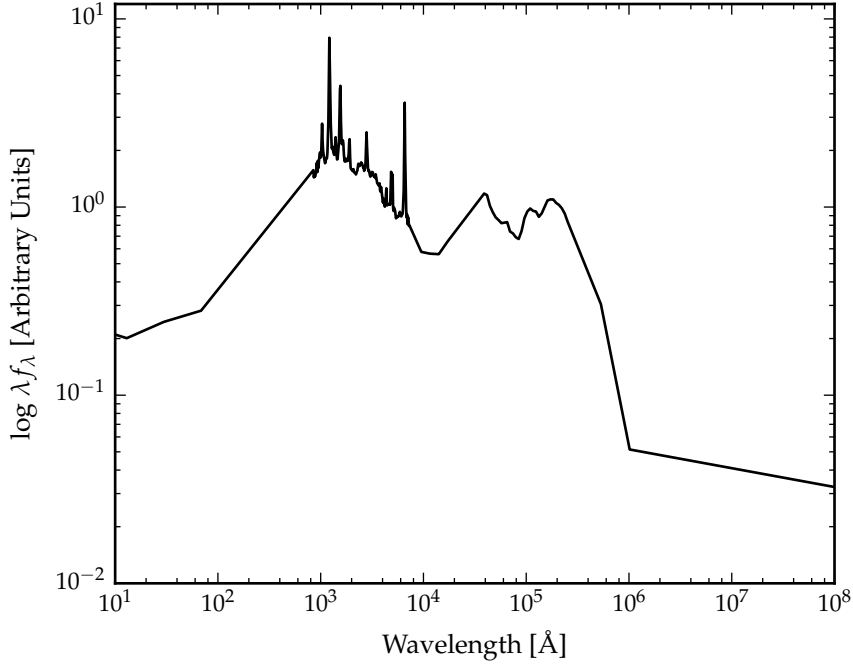
Figure 1.2: Median radio-loud AGN SED from Shang et al., (2011).

driven outflows. In quasars for which [O III] is detected, we find that the blueshifting of [O III] and C IV are correlated. This establishes a connection between quasar-driven outflows in the broad and narrow line regions. We confirm earlier results that the EV1 correlations found in low-luminosity AGN also exist in high-redshift quasars and demonstrate that Independent Component Analysis (ICA) can be used to extend these results to spectra with lower signal-to-noise.

## 1.7 AGN SPECTRAL ENERGY DISTRIBUTIONS

AGN emit strongly over many decades of the electromagnetic spectrum (Figure 1.2). Different physical processes dominate AGN spectral energy distributions (SEDs) at different frequencies. Hard X-ray emission is dominated by Compton upscattering of accretion disk photons by electrons in a hot corona (e.g. Sunyaev and Titarchuk, 1980), ultra-violet/optical by thermal accretion disc emission, IR by dust at a wide range of temperatures, and radio by synchrotron emission in relativistic jets. A complete understanding of AGN properties therefore critically depends on the availability of multi-wavelength data that spans the full SED. Progress has been made in re-

cent years with sensitive, wide-field photometric surveys, including the ultra-violet/optical SDSS, the near-infrared UKIRT Infrared Deep Sky Survey (UKIDSS; Lawrence et al., 2007) and the mid-infrared Wide-field Infrared Explorer (WISE; Wright et al., 2010).

To first order, AGN have remarkably similar SEDs that show very little dependence on luminosity, redshift, BH mass or accretion rate (e.g. Elvis et al., 2012; Hao et al., 2013). In Chapter 5, we build a simple parametric SED model that is able to reproduce the median optical-infrared colours of tens of thousands of SDSS AGN at redshifts $1 < z < 3$. On the other hand, the SED properties of individual quasars do show significant variation. In particular, we find the spread at $1 - 2$ μm is large and suggests the presence of significant diversity in hot dust properties. We find that the amount of hot dust is strongly correlated with the strength of outflows in the quasar BLR (parametrised using the C IV blueshift) and consider the implications of this result in the context of accretion disc wind models (e.g. Elitzur and Shlosman, 2006). We demonstrate that apparent correlations between the hot dust abundance and BH mass and Eddington ratio result from systematic biases in the C IV-based BH masses and disappear when new, unbiased BH mass estimates (derived in Chapter 3) are adopted.

## 1.8 NOTE ON ADOPTED CONVENTIONS

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 EQW measurements are given in the quasar rest-frame, and all emission-line wavelengths are given as measured in vacuum. Unless otherwise stated, optical (i.e. SDSS) magnitudes are given in the AB system and infrared magnitudes in the Vega system, following the conventions of the original surveys.

# 2

## A NEAR-INFRARED SPECTROSCOPIC DATABASE OF HIGH-REDSHIFT QUASARS

### 2.1 SPECTROSCOPIC SURVEYS

Emission-lines provide a wealth of information about the properties of AGN and their environments. They provide diagnostics of dynamics, temperatures, densities, dust content, elemental abundances and the shape of the ionising spectrum which are often unattainable through any other technique. The optical region includes a number of strong emission features, including the broad Balmer lines Hα (6565 Å) and Hβ (4863 Å) and the narrow [O III]λλ4960,5008 doublet. As we will see in Chapter 3, the Balmer lines are routinely used to derive BH masses and AGN accretion rates. As the strongest narrow emission-line in the optical spectrum, [O III] is used to measure systemic redshifts, and to probe AGN-driven outflows in the NLR (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 AGN and quasars in the SDSS spectroscopic catalogue alone reached almost 400 000. However, the rest-frame optical region is redshifted beyond the reach of optical spectrographs at redshifts $z \gtrsim 0.4$. Accessing the rest-frame optical lines at redshifts $2 \lesssim z \lesssim 4$, during the peak epoch of galaxy evolution, requires near-infrared spectroscopy.

Spectroscopic observations are more challenging at near-infrared wavelengths than in the optical because 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 and previous investigations of the rest-frame optical spectra of quasars at redshifts $z \sim 2$ have typically used samples containing a few dozen objects (e.g. Marziani et al., 2009; Shen and Liu, 2012; Shen, 2016).

In this chapter, we will describe the construction of a database containing 434 high-redshift quasars. In later chapters, we will describe how this data has been used to significantly reduce large systematic biases afflicting BH mass estimates for quasars at redshifts $z \gtrsim 2$ (Chapter 3) and to study the preva-

Table 2.1: Number of database objects observed with each near-infrared spectrograph/telescope.

| 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 |

lence and drivers of quasar-driven galaxy-wide outflows (Chapter 4). The unprecedented size and quality of this dataset make a number of other investigations possible, some of which are described in Chapter 6.

## 2.2 NEAR-INFRARED SPECTROSCOPIC DATA

The near-infrared spectra in our database are taken from published catalogues, by downloading and reducing archival spectra, and by reducing previously un-published spectra acquired in programmes led by Prof. J. Hennawi (UCSB) and Prof. X. Prochaska (UCO/LICK). We undertook two further observing programmes (Principle Investigator (PI): L. Coatman) to increase the number of objects in under-sampled regions of the C IV EQW and blueshift parameter space (see Section 3.1 for a detailed discussion of C IV emission properties in high-redshift quasars). The telescopes and instruments used to observe the spectra are summarised in Table 2.1 and information on individual spectra is provided in Table 2.2. There are 434 unique quasars in our catalogue. Multiple spectra exists for a number of quasars, and the total number of spectra in our catalogue is 462. The columns in Table 2.2 are as follows:

1. ID: Jhhmmss+ddmmss. ID is repeated when multiple spectra exist for the same object.

2. Unique catalogue name.

3  Date spectrum acquired.

4-5  RA and DEC (J2000; truncated coordinates).

6  Instrument and telescope used to acquire spectrum.

7  Wavelength range covered by spectrum.

8  Velocity per pixel in spectrum.

9  Signal-to-noise ratio (S/N) per pixel in spectrum.

10  Redshift.

Table 2.2: Quasars in the near-infrared spectroscopic database. Only the first 15 entries are shown. The full table (including 462 objects) is available on-line. Maybe DVD? Ask Debbie

| ID | Cat. Name | Date | Ra | Dec | Instr. | $\Delta\lambda$ [µm] | $\Delta\nu$ [km s$^{-1}$] | S/N | $z$ |
|---|---|---|---|---|---|---|---|---|---|
| (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) |
| J000039-001804 | QSO460 | 2015-09-02 | +00h00m39.00s | -00d18m03.90s | SofI/NTT | 1.50-2.54 | 154.0 | 4.9 | 2.14 |
| J000345-232353 | QSO552 | 2009-07-07 | +00h03m45.00s | -23d23m53.40s | SINFONI/VLT | 1.44-1.87 | 36.0 | 12.7 | 2.27 |
| J000345-232353 | QSO330 | 2011-09-18 | +00h03m45.00s | -23d23m53.40s | SofI/NTT | 1.48-1.83 | 63.0 | 36.0 | 2.26 |
| J000451-084450 | QSO290 | 2013-07-12 | +00h04m50.66s | -08d44m49.63s | XSHOOTER/VLT | 0.31-2.28 | 15.0 | 10.3 | 3.00 |
| J000451-084452 | QSO289 | 2013-08-08 | +00h04m50.91s | -08d44m51.98s | XSHOOTER/VLT | 0.31-2.28 | 15.0 | 5.4 | 3.00 |
| J000500-003348 | QSO454 | 2015-09-01 | +00h05m00.42s | -00d33m48.20s | SofI/NTT | 1.50-2.54 | 154.0 | 8.2 | 2.18 |
| J000501+010221 | QSO459 | 2015-09-02 | +00h05m00.53s | +01d02m20.80s | SofI/NTT | 1.50-2.54 | 154.0 | 6.8 | 2.13 |
| J001016+001228 | QSO475 | 2015-09-04 | +00h10m16.49s | +00d12m27.60s | SofI/NTT | 1.50-2.54 | 154.0 | 8.9 | 2.28 |
| J001247+001239 | QSO082 | 2013-06-06 | +00h12m47.12s | +00d12m39.49s | ISAAC/VLT | 1.52-1.60 | 15.0 | 19.1 | 2.16 |
| J001708+813508 | QSO107 | 2012-08-04 | +00h17m08.48s | +81d35m08.10s | TRIPLESPEC/Hale | 0.94-2.80 | 39.0 | 36.5 | 3.40 |
| J001919+010152 | QSO476 | 2015-09-04 | +00h19m19.31s | +01d01m52.20s | SofI/NTT | 1.50-2.54 | 154.0 | 6.5 | 2.32 |
| J001955-091316 | QSO001 | 2004-11-26 | +00h19m54.67s | -09d13m16.45s | GNIRS/Gemini | 0.60-2.61 | 88.0 | 9.9 | 2.12 |
| J002018-233654 | QSO553 | 2009-07-07 | +00h20m18.41s | -23d36m53.80s | SINFONI/VLT | 1.44-1.87 | 36.0 | 16.9 | 2.30 |
| J002023-414639 | QSO554 | 2009-07-08 | +00h20m23.38s | -41d46m38.90s | SINFONI/VLT | 1.09-1.41 | 35.0 | 33.4 | 1.57 |
| J002111-242247 | QSO555 | 2009-07-16 | +00h21m10.90s | -24d22m47.20s | SINFONI/VLT | 1.44-1.86 | 36.0 | 11.1 | 2.26 |

### 2.2.1  *Coatman et al., (2016) sample*

#### 2.2.1.1  *Target selection*

We selected quasars from the Seventh Data Release (DR7; Schneider et al., 2010) of the SDSS 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-passbands 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 BALs in their spectra (5,299 quasars), and are free from significant dust extinction. We removed radio-loud objects and BAL quasars using the classification flags described in Section 2.4.3. The removal of quasars with significant dust extinction was achieved by identifying quasars with $i - K$ colours redder than a parametric SED model combined with an extinction curve with $E(B - V)=0.05$ (a very similar procedure is described in greater detail in Section 5.5.1).

The K-magnitude (used to compute the $i - K$ colour) was taken from the UKIDSS Large Area Survey (ULAS). The requirement to be in the ULAS footprint and have reliable K passband 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 S/N could be obtained (leaving 412 quasars).

We were able to obtain new infrared spectra for 19 quasars from this sample of 412 possible targets. The quasars included in this sub-sample 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 recent improvements in the estimation of systemic redshifts from ultra-violet spectra (see Section 3.5.1 for details).

#### 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 4.2 m 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 $1.53 - 1.79\,\mu m$ and $2.07 - 2.44\,\mu m$ with a dispersion of $9.7\,\text{Å/pixel}$. The spatial scale of the instrument is $0.25\,\text{arcsecond/pixel}$. Observations were divided into $60\,\text{second}$ sub-exposures and performed in an ABBA nodding pattern, with the object placed at two positions along the slit $12\,\text{arcsecond}$ 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 given 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 $1.01 - 1.71\,\text{Å}$, with a mean of $1.47\,\text{Å}$. The telluric standard star observations were reduced using the same steps described above. The stel-

---

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.

lar 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[2].

### 2.2.2 *Shen, (2016) sample*

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 were required 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 passbands. Thirty-eight of the quasars were observed with Triple-Spec (Wilson et al., 2004) on the Astrophysics Research Consortium (ARC) 3.5 m telescope, and 36 with the Folded-port InfraRed 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 *Quasars Probing Quasars sample*

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]; 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 resources have been devoted to obtaining near-infrared spectra which cover low-ionisation broad lines or features from the quasar NLR (Prochaska and Hennawi, 2009; Lau, Prochaska, and Hennawi, 2016; 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, thirteen 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 Spectrom-

---

2 The data reduction pipeline is available at https://github.com/liamcoatman/SpectraTools.

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

eter (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 G. Becker (for details, see Lau, Prochaska, and Hennawi, 2016). The remaining data was processed with algorithms in the LowRedux[4] package (see Prochaska and Hennawi, 2009).

### 2.2.4 *VLT SINFONI*

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 (083.B-0456; PI: L. Wisotzki) to study the mass function and Eddington ratios of active BHs drawn from the Hamburg-ESO survey (Wisotzki et al., 2000). A further seven SINFONI spectra are from a programme (090.B-0674; PI: J. Kurk) to obtain reliable BH mass estimates from Hα/Hβ for a sample of radio-loud/radio-quiet SDSS quasars.

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 *NTT SofI*

One quarter of the quasar catalogue derives from a large 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 rest-frame ultra-violet spectra. The main science goal is to obtain precise

---

4 www.ucolick.org/~xavier/LowRedux
5 www.mrao.cam.ac.uk/~rw480/easysinf

systemic redshifts which are crucial for the study of absorption-line systems. Observations were undertaken over 16 nights from 2011 September to 2013 March. Both the 'red' (R $\simeq$ 1000) and the H and K (R $\simeq$ 1500) grisms were employed. The spectra were reduced using a custom pipeline built from algorithms in the LowRedux package.

Over five nights from 2015 August 31 to September 4 we obtained near-infrared SOFI spectra for a further 26 quasars (095.B-0644; PI: L. Coatman). These quasars were selected from the SDSS DR7 quasar catalogue using criteria very similar to those described above for the WHT/LIRIS sample. In particular, we selected quasars for which C IV was significantly blueshifted relative to the quasar rest-frame to improve the statistics in this region of the C IV emission-line parameter space. The 'red' grism was employed with a one arcsecond slit-width. The spectra were reduced using the same LowRedux pipeline described above.

### 2.2.6 *Hale TripleSpec*

A further 60 quasars in our catalogue are bright SDSS quasars which were observed with the TRIPLESPEC spectrograph (Herter et al., 2008) 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.

## 2.3 REDSHIFT AND LUMINOSITY DISTRIBUTION OF CATALOGUE

In Figure 2.1 we show the luminosities and redshifts of the quasar sample relative to the redshift-luminosity distribution of the SDSS DR7 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 passbands 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 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
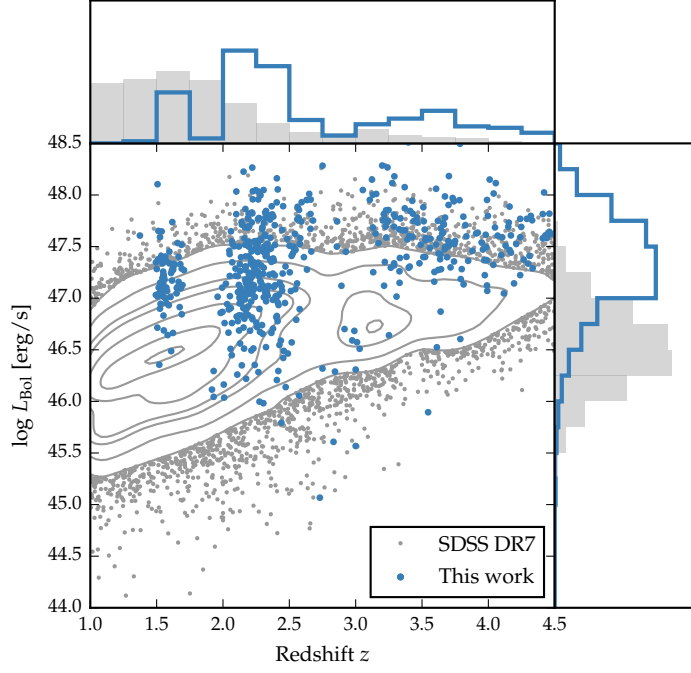
Figure 2.1: The ranges in redshift and luminosity covered by our sample, relative to the redshift-luminosity distribution of the SDSS DR7 quasar catalogue. For the SDSS sample we use Hewett and Wild, (2010) redshifts and bolometric luminosities computed by Shen et al., (2011) from $L(3000\text{Å})$ $(z < 1.9)$ and $L(1350\text{Å})$ $(z \gtrsim 1.9)$ using bolometric corrections $BC(3000\text{Å}) = 5.15$ and $BC(1350\text{Å}) = 3.81$ (Richards et al., 2006a). For the quasars in this work the redshift is defined using the peak of the $H\alpha/H\beta$ emission and the bolometric luminosity is computed from $L(5100\text{Å})$ using $BC(5100\text{Å}) = 9.26$ (Richards et al., 2006a).

luminosity of the SDSS sample, although the dynamic range in luminosity is a full 1.5 decades.

## 2.4 SUPPLEMENTARY DATA

### 2.4.1 *Optical spectroscopic data*

Optical spectra are available for 79 per cent of the catalogue. The sources of the optical spectra employed (summarised in Table 2.3) are as follows:

1. SDSS DR7 spectroscopic quasar catalogue. Spectra are moderate resolution ($R \simeq 2000$) and S/N (S/N $\simeq 20$) and

Table 2.3: Percentage of catalogue for which optical spectroscopic data is available from the given sources.

|     | Source        | %  |
|-----|---------------|----|
| (1) | SDSS          | 60 |
| (2) | BOSS          | 45 |
| (3) | HAMBURG-ESO   | 7  |
| (4) | VLT/UVES      | 4  |
| (5) | VLT/XSHOOTER  | 8  |

cover the observed-frame wavelength interval $\sim 3800 - 9180\,\text{Å}$.

2. BOSS DR12 (Pâris et al., 2017) spectroscopic quasar catalogue. Compared to SDSS spectra, BOSS spectra cover a slightly broader wavelength range and are typically higher S/N.

3. The Hamburg-ESO survey (Wisotzki et al., 2000). The spectra have a typical $R \simeq 700$ spectral resolution and $S/N \gtrsim 10$ per pixel.

4. Spectra taken with VLT/UVES. The reduced and fluxed UVES spectra were made available to us by A. Dall'Aglio (a description of the reduction procedure is contained in Dall'Aglio, Wisotzki, and Worseck 2008). The spectral resolution of the UVES observations is very high ($R \simeq 40\,000$) and the S/N of the spectra, re-binned to a resolution of $R \simeq 2000$, is $S/N \simeq 300$.

5. Spectra taken with VLT/XSHOOTER. The XSHOOTER spectra are moderate resolution ($R \simeq 6000$) and cover the full optical-near-infrared spectral region ($0.30 - 2.50\,\mu m$).

### 2.4.2  *Photometric data*

We cross-matched our catalogue with photometric data from a number of wide-field surveys. The matching was done using a five arcsecond matching radius, with only the closest neighbour retained in the case of multiple matches. The cross-matched surveys and the percentage of successful matches is summarised in Table 2.4. The columns in Table 2.4 are as follows:

1. SDSS DR9 (Ahn et al., 2012) photometric source catalogue. Point spread function magnitudes.

2. Two Micron All Sky Survey (2MASS; Skrutskie et al., 2006) Point Source Catalogue. Default magnitudes.

3. UKIDSS Large Area Survey (DR10). One arcsecond aperture corrected magnitudes ('apermag3').

4. Visible and Infrared Survey Telescope for Astronomy (VISTA) Hemisphere Survey (VHS; McMahon et al., 2013). One arcsecond aperture corrected magnitudes ('apermag3').

5. Kilo-Degree Infrared Galaxy (VIKING; Edge et al., 2013) Survey (DR4). One arcsecond aperture corrected magnitudes ('apermag3').

6. WISE AllWISE Data Release (Mainzer et al., 2011). Profile-fitting magnitudes ('mpro').

2.4.3 *Radio/BAL quasar classification*

Using the catalogues provided by Shen et al., (2011), Allen et al., (2011) and Pâris et al., (2017) and visual inspection, 19 quasars in the catalogue are identified as being C IV BAL quasars.

We cross-match our catalogue to the FIRST radio catalogue (White et al., 1997). We classify quasars with matches within five arcsecond as core-dominated, while, if multiple matches are found within 30 arcsecond, quasars are classified as lobe-dominated (e.g. Shen et al., 2011). 128 objects are outside of the FIRST footprint, 269 are not detected in FIRST, 29 are detected and are core-dominated, and 8 are detected and are lobe-dominated.

2.5 ABSOLUTE FLUX CALIBRATION OF NEAR-INFRARED SPECTRA

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 using either SDSS/BOSS spectroscopy or the available photometric data as a fiducial baseline. The methods are attempted in the order given below, and

Table 2.4: Cross-matched surveys and the percentage of successful matches.

|    | SDSS | 2MASS | UKIDSS | VHS | VIKING | WISE |
|----|------|-------|--------|-----|--------|------|
|    | (1)  | (2)   | (3)    | (4) | (5)    | (6)  |
| u  | 73   | -     | -      | -   | -      | -    |
| g  | 73   | -     | -      | -   | -      | -    |
| r  | 73   | -     | -      | -   | -      | -    |
| i  | 73   | -     | -      | -   | -      | -    |
| z  | 73   | -     | -      | -   | 9      | -    |
| Y  | -    | -     | 41     | 10  | 9      | -    |
| J  | -    | 57    | 41     | 34  | 9      | -    |
| H  | -    | 57    | 41     | 20  | 9      | -    |
| K  | -    | 57    | 41     | 33  | 9      | -    |
| W1 | -    | -     | -      | -   | -      | 97   |
| W2 | -    | -     | -      | -   | -      | 97   |
| W3 | -    | -     | -      | -   | -      | 97   |
| W4 | -    | -     | -      | -   | -      | 97   |

the method we adopt is dependent on the availability of the required data. We are unable to verify the absolute flux calibration of the near-infrared spectra for four objects because neither SDSS/BOSS spectra nor optical/near-infrared data is available. The methods used to flux calibrate the near-infrared spectra are summarised in Table 2.5.

### 2.5.1  *SDSS spectrum as a fiducial baseline*

The flux calibration of the SDSS spectra is excellent, and so these spectra can be used as a fiducial baseline to calibrate the near-infrared spectra. The quasar SED model described in Chapter 5 is used to bridge the gap between the wavelength coverage of the near-infrared and optical SDSS spectrum. This model provides a very good fit to the SDSS and UKIDSS magnitudes of SDSS DR7 quasars, reproducing the individual magnitudes to $< 0.1$ mag. The first step is to normalise the SED model to the SDSS spectrum. This is done using a variance-weighted $\chi^2$ minimisation procedure in several emission-line-free intervals of the spectrum. The second step is to normalise the near-infrared spectrum to the SED model. Again, this is done using a variance-weighted $\chi^2$ minimisation procedure, with regions of the spectrum falling between the near-infrared passbands

Table 2.5: Methods used in absolute flux calibration of near-infrared spectra.

| Method | % |
|---|---|
| SDSS | 60 |
| BOSS | 9 |
| NIR photometry | 25 |
| NIR+OPT photometry | 6 |
| None | 1 |

masked-out in the minimisation. The flux calibration procedure is demonstrated in Figure 2.2a.

### 2.5.2 *BOSS spectrum as a fiducial baseline*

This procedure is identical to one described in the previous section, the only difference being that the SDSS spectrum is substituted for a BOSS spectrum. 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 1275 Å.

### 2.5.3 *Photometric data as a fiducial baseline*

In the first step, the quasar SED model is normalised to the available optical (SDSS) and near-infrared (VHS, Viking, UKIDSS or 2MASS) photometric data. The SED model is integrated through the appropriate passband transmission functions to give model magnitudes (Equations 5.1 and 5.2), and a variance weighted $\chi^2$ minimisation procedure is performed with the observed magnitudes. The second step in the procedure is then identical to the previous two sections. This procedure is illustrated in Figure 2.2b.

### 2.5.4 *Reliability of luminosity measurements*

*Ask Manda: (1+z) in luminosity calculation.*

The flux at 1350 and 5100 Å was read off directly from the normalised SED model. These values are used to compute monochromatic continuum luminosities, given by:

$$L_\lambda = (1 + z)\, 4\pi L_D(z)^2\, \lambda F_\lambda \qquad (2.1)$$

26
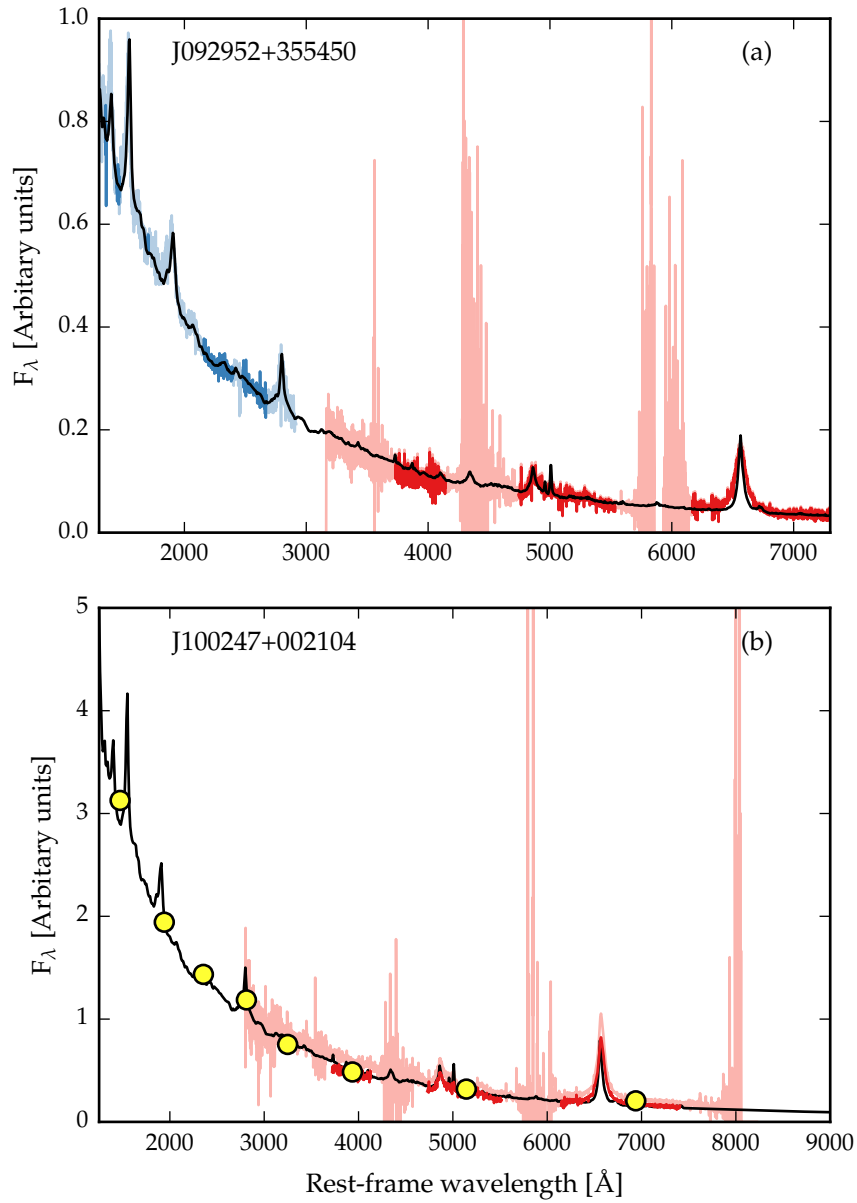
Figure 2.2: Demonstration of how the absolute flux calibration of the near-infrared spectrum (shown in red) is established using the SDSS spectrum (shown in blue, a) and photometric data (b) as fiducial baselines. An empirical quasar SED model (shown in black) is used to bridge the discontinuity in the wavelength coverage of the optical/near-infrared spectroscopic and photometric data. The darker regions of the spectra are used in the fitting procedures.

where $L_D(z)$ is the luminosity distance at redshift $z$. Monochromatic luminosities are used to compute BH masses and bolometric luminosities in Chapters 3, 4 and 5. Comparison of the 5100 Å luminosity, computed using the photometry- and spectrum-based methods for 296 quasars, showed a scatter (mean absolute deviation) of just $\sim 0.1$ 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 $\sim 5$ years. Intrinsic quasar photometric variability in the rest-frame ultra-violet and optical will therefore add additional scatter of $\sim 0.2$ mag (e.g. MacLeod et al., 2010) to the derived 1350 and 5100 Å luminosities.

The monochromatic continuum luminosity at 5 μm was also computed by linearly interpolating through the WISE photometric data points. 5 μm luminosities were derived in this way for 434 quasars up to redshift $z = 3.4$. At higher redshifts, the longest wavelength WISE passband ($W4$) is at $< 5$ μm in the quasar rest-frame and so 5 μm luminosities could not be computed using this method.

## 2.6 CORRECTING FOR INSTRUMENTAL BROADENING

Throughout this thesis, reported line-width measures are corrected for instrumental broadening by subtracting the resolution of the spectrograph in quadrature. Because the quasar emission-line profiles are typically non-Gaussian, this deconvolution procedure is only approximate. The spectrograph resolutions, which we estimate from the line widths in the observed sky spectra, are given in Table 2.6. The resolutions are generally small relative to the widths of quasar broad emission-lines (FWHM $\sim 4000$ km s$^{-1}$).

Table 2.6: Measured spectral resolutions of the spectrographs used in this thesis.

| Spectrograph | FWHM [$\mathrm{km\,s^{-1}}$] |
| --- | --- |
| FIRE | 59 |
| GNIRS | 136 |
| ISAAC | 46 |
| LIRIS | 477 |
| NIRI | 465 |
| NIRSPEC | 122 |
| SINFONI | 124 |
| SOFI (MR) | 323 |
| SOFI (LR) | 535 |
| P200 TRIPLESPEC | 88 |
| ARC TRIPLESPEC | 97 |
| XSHOOTER | 25 |
| SDSS/BOSS | 152 |
| UVES | 3 |
| HAMBURG-ESO | 400 |

# 3

# CORRECTING C IV-BASED VIRIAL BLACK HOLE MASSES

## 3.1 SINGLE-EPOCH VIRIAL BH MASSES

The goal of better understanding the origin of the correlation between the masses of super-massive BHs and the masses of host-galaxy spheroids has led to much work focussing on the properties of quasars and 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 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-ultra-violet spectral regions.

The C 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 SDSS, to redshifts exceeding $z \sim 5$. C 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 ($\sim$ 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 IV FWHM velocity measurements and ultra-violet 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 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 II, the C 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 work (e.g. Sulentic, Marziani, and Dultzin-Hacyan, 2000; Richards et al., 2011) has established that the extent of 'blueshifts' in the C IV emission correlates with a number of properties of quasar 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).

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 km s$^{-1}$) the C IV-profile is displaced to the blue by amounts comparable to the FWHM of the profile. At fixed emission-line EQW, 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. 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; Gav-
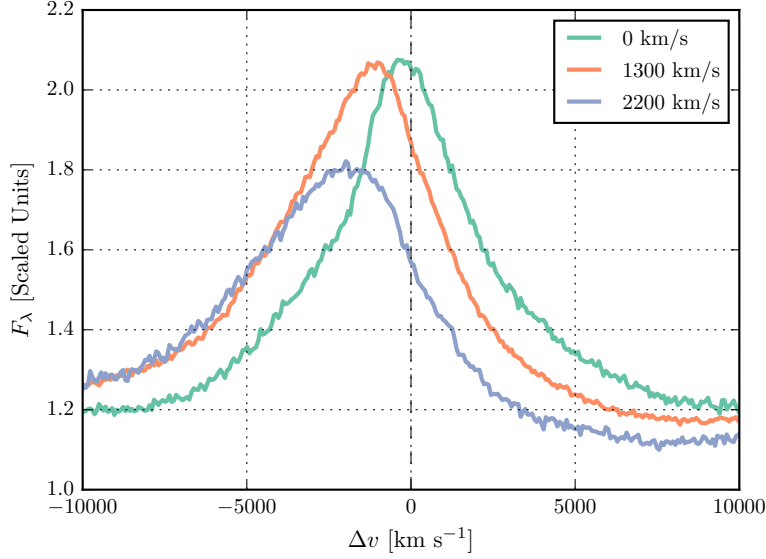
Figure 3.1: Composite spectra of the C IV emission-line as a function of C IV blueshift for SDSS DR7 quasars. Quasars classified as BALs, or possessing strong associated absorbers have been excluded, and the composite spectra shown are derived using an arithmetic mean of a minimum of 200 spectra at each blueshift. Virtually the entire C 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. In order of increasing C IV blueshift, the composite spectra have FWHM 4870, 5610, and $6770\,\mathrm{km\,s^{-1}}$ and EQW 33.1, 31.6, and 28.8 Å.

ignaud et al., 2008; Vestergaard et al., 2008; Vestergaard and Osmer, 2009; Kelly et al., 2010; Kelly and Shen, 2013). 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).

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 EQWs 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 EQWs 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

33

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 $\Delta V^2$ dependence predicted assuming a virialized BLR to just $\Delta V^{0.56}$ 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 ultra-violet emission-lines (e.g. Si IV+O IV$\lambda$1400 in Runnoe et al. 2013 and Brotherton et al. 2015). The increased number of quasars with high-quality spectra that cover both the observed-frame optical (where the redshifted C IV appears) and near-infrared (where H$\beta$ and H$\alpha$ 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 H$\alpha$ and H$\beta$ 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 non-BAL 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 H$\alpha$ line measurements and 144 H$\beta$ 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 H$\alpha$ and H$\beta$. 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 $\sim 5000\,\mathrm{km\,s}^{-1}$, and span the full range of blueshifts observed in the population.

The near-infrared data has been described in Chapter 2 and the telescopes/spectrographs used are summarised in Table 3.1. Corresponding optical spectroscopy was obtained from the SDSS (70 quasars), BOSS (126 quasars) and Hamburg-ESO surveys (15 quasars), and with VLT/UVES (11 quasars) and VLT/XSHOOTER (8 quasars). Many of the quasars in the SDSS DR7 catalogue have been re-observed as part of BOSS. As the

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 |

BOSS spectra typically have higher S/N than the SDSS DR7 spectra, we have used the BOSS spectra when available. Once more, further details are provided in Chapter 2. We have subdivided our sample into two overlapping groups: quasars with reliable Hα line measurements (the 'Hα sample') and quasars with reliable Hβ measurements (the 'Hβ sample').

## 3.3 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 BLR and, as revealed in reverberation-mapping studies, the luminosity is a proxy for the typical size of the BLR (the $R_{BLR} - L$ 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 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 masses 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, $\lambda_{half}$, 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

$$c \times \frac{1549.48\text{Å} - \lambda_{half}}{1549.48\text{Å}} \qquad (3.1)$$

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 parametrisations is not important and the fits are used only to provide robust line parameters, such as the centroid $\lambda_{half}$, and FWHM, which are measured non-parametrically 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

---

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 $\sim 80\,km\,s^{-1}$.

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, $\alpha$, determined using the median[2] values of the flux in two continuum windows at 1445-1465 and 1700-1705 Å (the same wavelengths as adopted by Shen et al. 2011). 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-transition wavelength for the C IV doublet. The parametric model is ordinarily fit within the wavelength interval 1500-1600 Å (corresponding to approximately $\pm 10\,000\,\mathrm{km\,s^{-1}}$ from the rest-frame transition wavelength), a recipe that is commonly adopted (e.g. Shen et al., 2011; 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 parametrisation. In Figure 3.2

---

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

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).

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 $190 \, \text{km s}^{-1}$.

### 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 $1200 \, \text{km s}^{-1}$. 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 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 $1200 \, \text{km s}^{-1}$ upper limit on the FWHM of

all narrow lines and the 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\lambda$4960,5008 doublet is detected at moderate S/N in the H$\beta$ 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 H$\alpha$ region. For spectra where the [O iii] doublet does not constrain the velocity and FWHM accurately, the narrow emission in the H$\alpha$ and H$\beta$ 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 H$\alpha$ and H$\beta$ emission regions simultaneously and we fix the centroids of the Gaussian components used to fit the broad emission. In Figure 3.3 we plot our H$\alpha$ 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 $300\,\mathrm{km\,s^{-1}}$.

### 3.3.3   *H$\beta$ and [O iii]*

Emission from optical Fe ii is generally strong in the vicinity of H$\beta$. 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 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. We use the same model to fit the broad and narrow components of H$\beta$ as was used with H$\alpha$. 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 5008 Å lines is fixed at 1 : 3. As for the fit to the narrow lines in the spectral region around H$\alpha$, the width and velocity offsets of all the narrow components are set to be equal, and an upper limit of $1200\,\mathrm{km\,s^{-1}}$ is placed on the FWHM.

The parametric model we fit to the H$\beta$/[O iii] emission region was very similar to the model employed by Shen, (2016).
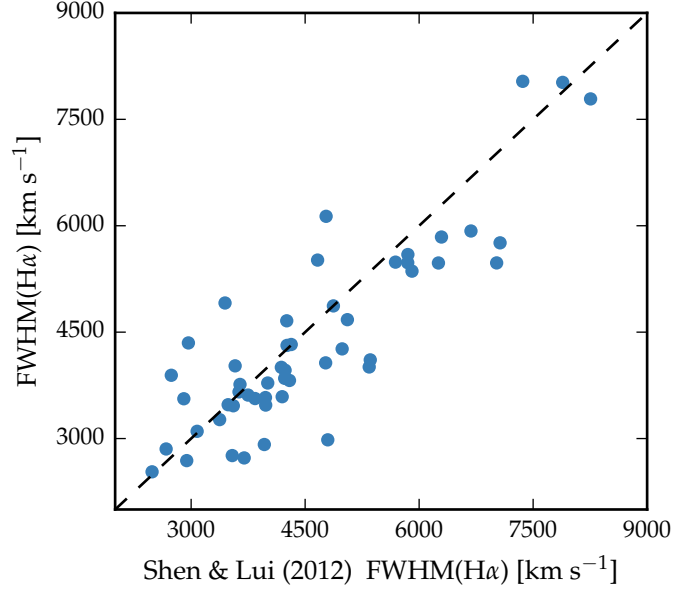
Figure 3.3: Demonstration of the effectiveness of our line parameter estimation scheme via a comparison of the Hα FWHM with Shen and Liu, (2012).

In Figure 3.4 we plot our Hβ FWHM measurements against the measurements published in Shen, (2016), for 39 quasars in common to both samples. As expected, we observe a very tight correlation, with a scatter of $270\,\mathrm{km\,s^{-1}}$.

### 3.3.4 *Fitting procedure*

Model parameters were derived using a standard variance-weighted least-squares minimisation procedure employing the Nelder-Mead 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 Figure 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 median reduced-$\chi^2$ values in our Hα, Hβ and C IV fits are 0.96, 1.58, and 0.91 respectively and, in general, there are no strong features observable in the model residuals. The only significant features seen in the residual C IV
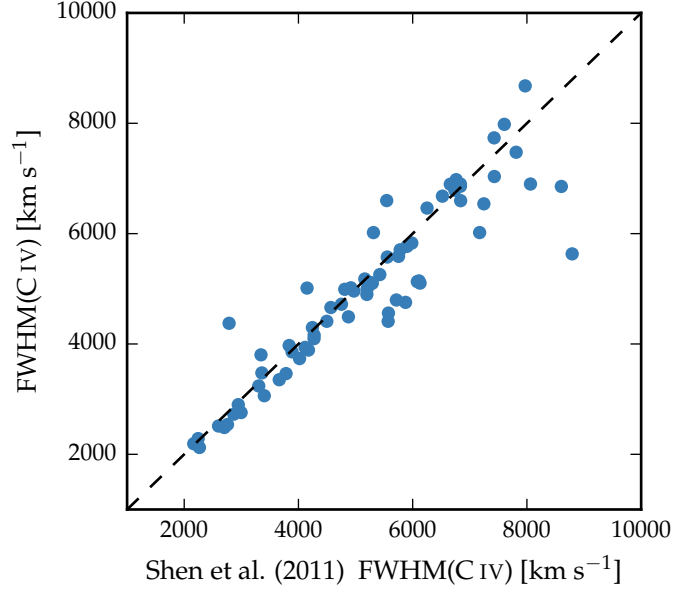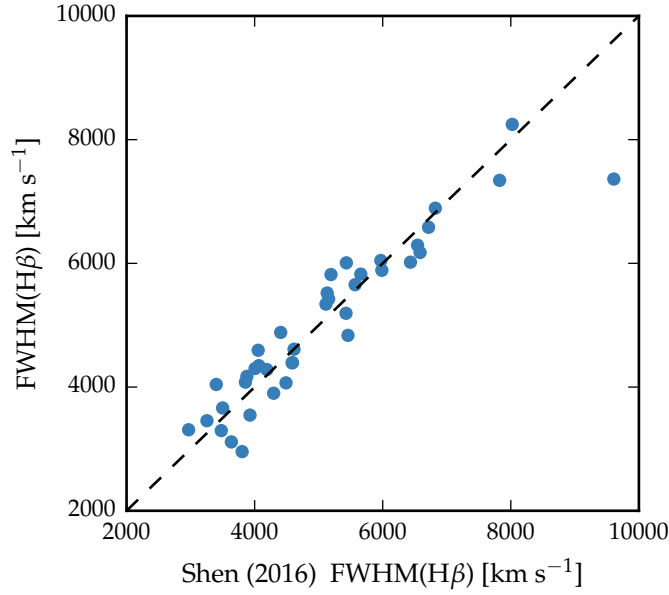
Figure 3.4: Demonstration of the effectiveness of our line parameter estimation scheme via a comparison of the Hβ FWHM with Shen, (2016).

spectra correspond to the location of narrow absorption-lines which were excluded in the fitting procedure.

Table 3.4 includes the line parameters of our best-fitting model for each line.

### 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.

Table 3.2: 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 |
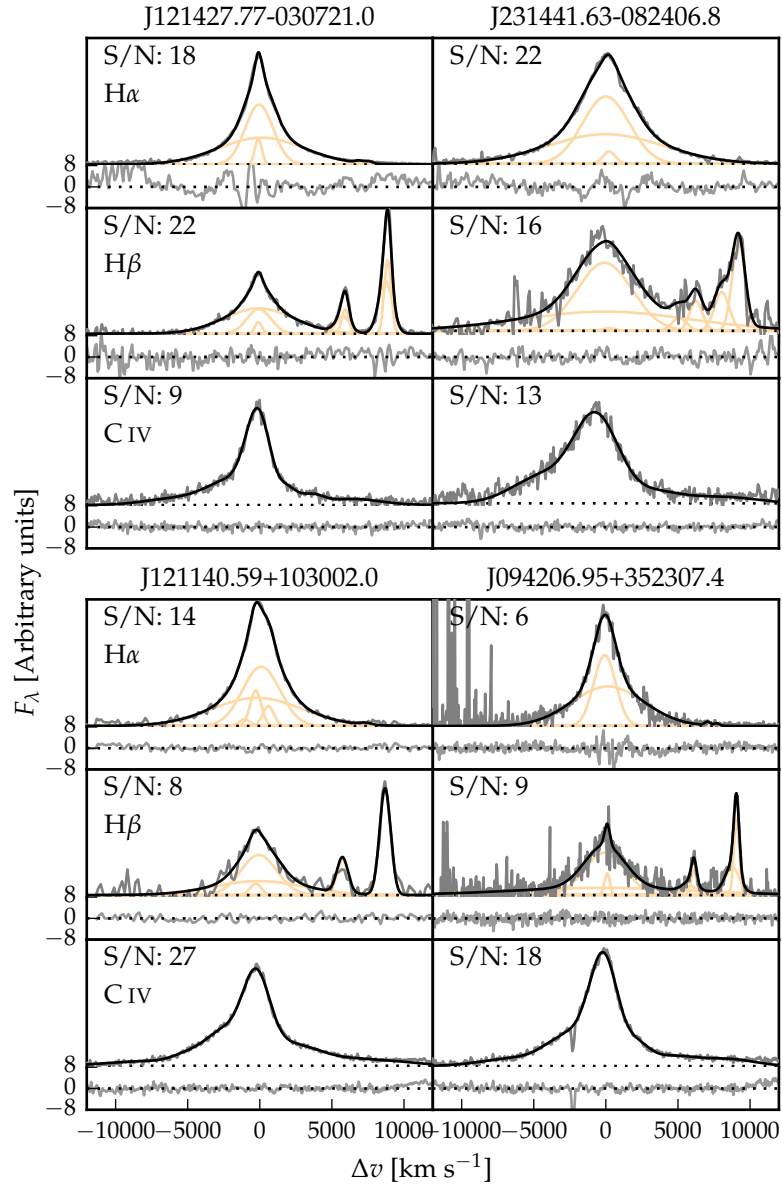
Figure 3.5: Model fits to continuum-subtracted Hα, Hβ, and C IV emission in four randomly-selected quasars. The data is shown in grey, the best-fitting parametric model in black, and the individual model components in orange. The centroid of the broad Hα emission is used to set the redshift, and $\Delta v$ is the velocity shift from the line rest-frame transition wavelength. The S/N is indicated and is measured in a region of the continuum and quoted per $150\,\mathrm{km\,s^{-1}}$ pixel. Below each line we plot the data minus model residuals, scaled by the errors on the fluxes.

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 3.2 and the distribution in redshift and luminosity is shown in Figure 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 H$\alpha$ emission falls at the edge of the K-passband. H$\beta$ is much weaker than H$\alpha$, and the H$\beta$ spectra are generally of lower S/N. As a result, the fraction of H$\beta$ spectra that are flagged – 39 per cent – is particularly high.

### 3.3.6 *Emission-line parameter uncertainties*

The 1-$\sigma$ 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 subtraction 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
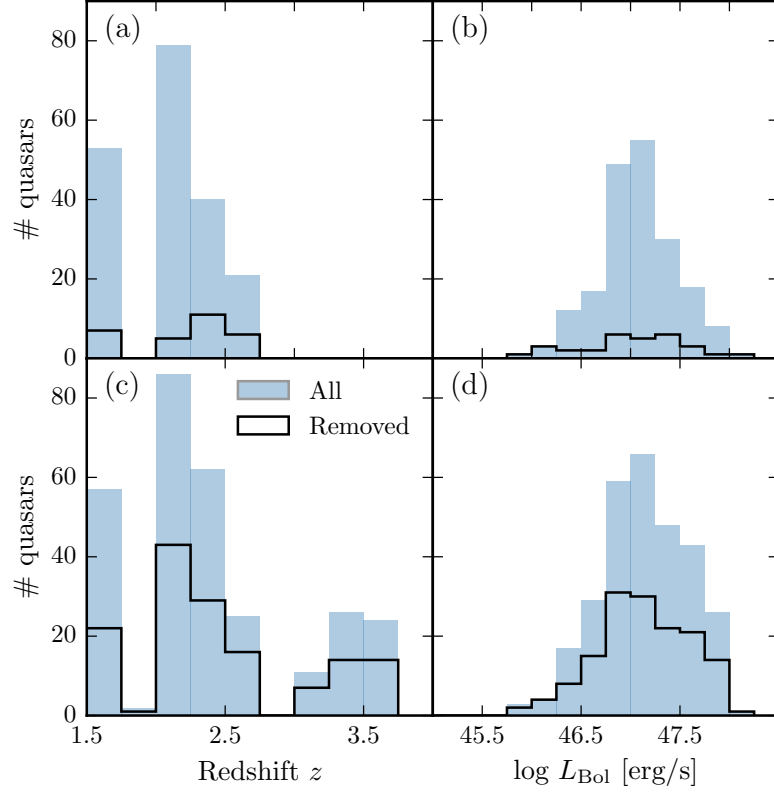
Figure 3.6: The redshift and luminosity distributions of the spectra removed from our Hα/C iv (a, b) and Hβ/C iv (c, d) samples.

fits[3]. 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 In the analysis of the real spectra such fits are identified via visual inspection.

### 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[4]. 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 BLR. The monochromatic continuum flux is generally measured at 1350 Å for C IV and 5100 Å for Hα and Hβ. The calculation of these luminosities is described in Chapter 2.

As described in Chapter 2, we estimate the uncertainties on the monochromatic luminosities to be ∼ 0.3 dex. 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.

---

4 Time differences in the quasar rest-frame are reduced by a factor of $(1 + z)$.

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 Figure 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 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
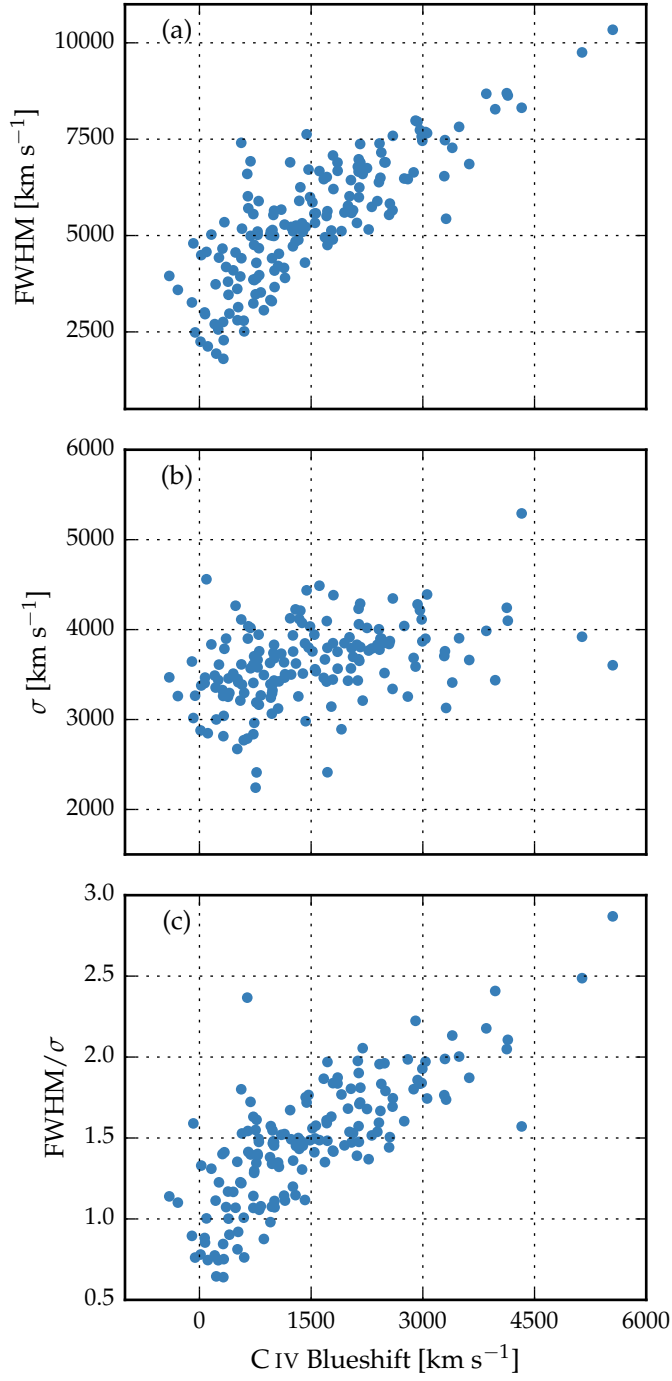
Figure 3.7: The FWHM, dispersion (σ) and shape (FWHM/σ) of C IV
as a function of the C IV blueshift.

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 Figure 3.8 we overlay the C IV line profiles of J123611+112922 and J152529+292813, whose dispersions are indistinguishable ($4168 \pm 271$ and $4303 \pm 128\,\mathrm{km\,s^{-1}}$ 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 Sulentic et al., 2007; Shen and Liu, 2012), 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).
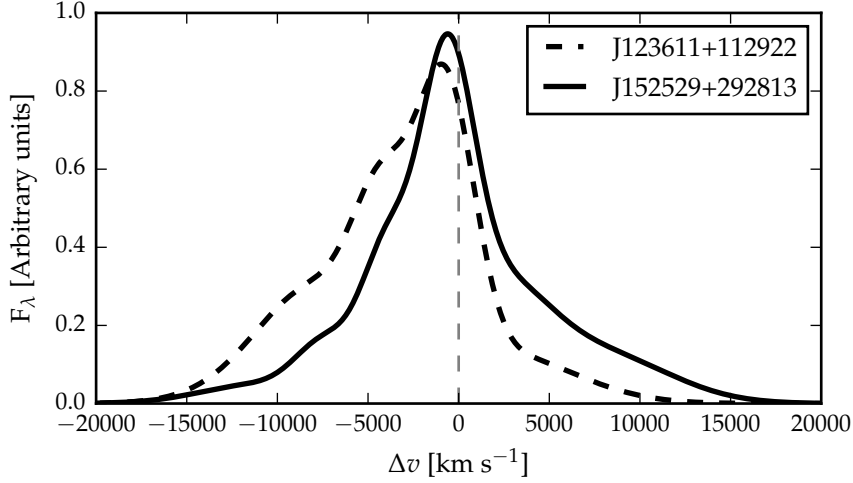
Figure 3.8: Comparison of the C IV line profiles of J123611+112922 and J152529+292813. Notwithstanding the essentially identical dispersion values, the emission-line velocity fields differ dramatically and, therefore, the dispersion values cannot be measuring accurately the virial-induced velocity spread 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 $R_{BLR} - L$ 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).

Hβ is emitted preferentially over Hα when the density and/or ionisation parameter is higher (e.g. Osterbrock, 1989). If the density or ionisation parameter in the BLR decreases with increasing radii, then Hβ would be expected to exhibit a broader profile than Hα (e.g. Greene and Ho, 2005b). We constructed composite spectra of the Hα and Hβ regions from 217 and 171 quasars respectively. This includes quasars which were excluded from the main 308-object catalogue because the C IV FWHM and/or blueshift could not be measured reliably. Spectra were first de-redshifted to the quasar rest-frame, and then

interpolated on to a common wavelength grid with a 1 Å resolution. The spectra were scaled by the mean flux in the interval 4700-5100 Å (Hβ) and 6400-6800 Å (Hα). The Hα and Hβ lines in the median composite spectra are shown in Figure 3.11. In line with our expectations, Hβ has a significantly broader profile.

In our sample, we have 99 quasars with reliable measurements of both Hα and Hβ lines. The line widths in individual objects are compared in Figure 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

$$\text{FWHM(H}\beta) = (1.07 \pm 0.07) \times 10^3 \left( \frac{\text{FWHM(H}\alpha)}{10^3 \text{ km s}^{-1}} \right)^{(1.03 \pm 0.03)} \tag{3.2}$$

The relation is shown as the dashed line in Figure 3.9. The root-mean-square scatter about this relation is 0.07 dex, compared to the $\sim 0.1$ dex 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 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 $\alpha$ and intercept $\beta$) 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, $\Sigma_i^2$, with a Gaussian of variance V representing the intrinsic variance in the data. The log-likelihood is then given by

$$\ln \mathcal{L} = - \sum_{i=1}^{N} \frac{1}{2} \ln \left[ 2\pi \left( \Sigma_i^2 + V \right) \right] - \sum_{i=1}^{N} \frac{\Delta_i^2}{2[\Sigma_i^2 + V]} \tag{3.3}$$

where $\Delta_i$ 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 reasonable assumption that there is an
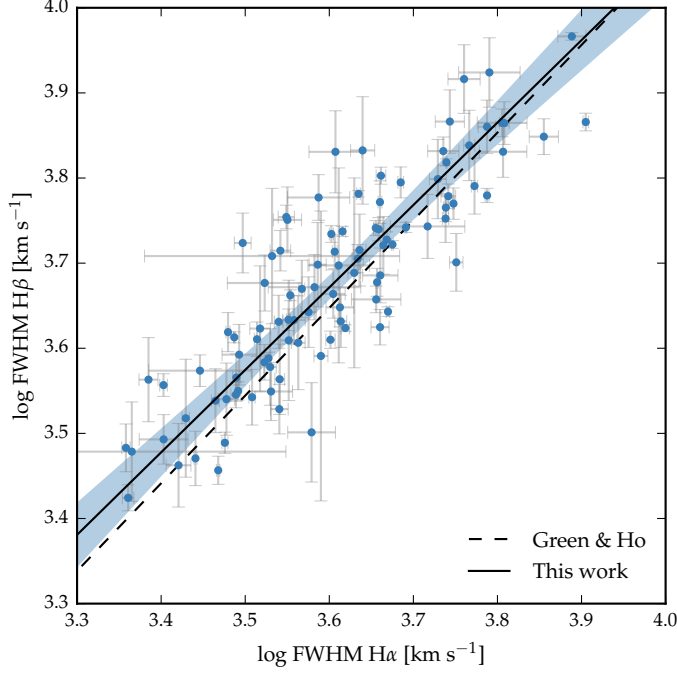
Figure 3.9: Comparison of Hα and Hβ FWHM measurements for 99 quasars. The solid line is our best-fitting power-law model, and the blue-shaded region shows the 2-σ uncertainties on the model parameters. The dashed line is the relation found by Greene and Ho, (2005b) using a sample of $z <$ 0.35 SDSS AGN.

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 $\theta$ 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 Figure 3.10. The solid line in Figure 3.9 is the maximum likelihood solution

$$\text{FWHM}(\text{H}\beta) = (1.23 \pm 0.10) \times 10^3 \left( \frac{\text{FWHM}(\text{H}\alpha)}{10^3 \text{km s}^{-1}} \right)^{0.97 \pm 0.05} \tag{3.4}$$

51
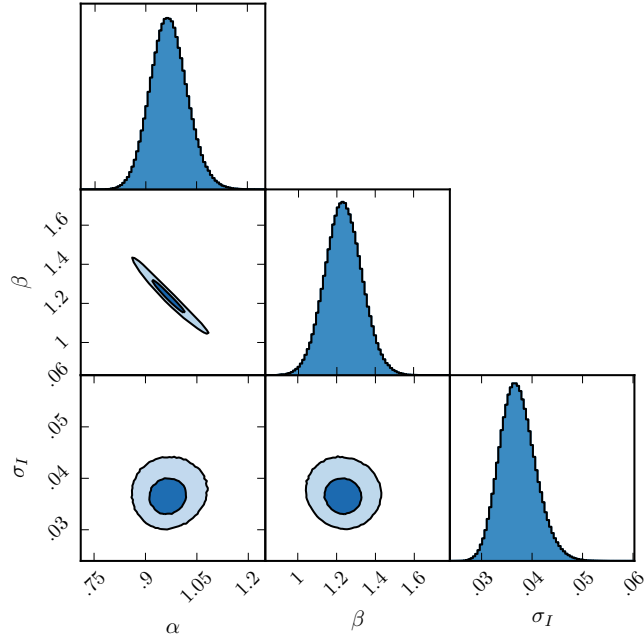
Figure 3.10: One- and two-dimensional projections of the MCMC sampling of the posterior distribution from the fit in Figure 3.9. $\alpha$ is the power-law index, $10^{\beta}$ is the normalisation, and $\sigma_I$ is the intrinsic scatter. In the two-dimensional projections, 1- and 2-$\sigma$ contours are shown.

and the shaded region shows the 2-$\sigma$ uncertainties on the model parameters.

As discussed above, our relation is displaced to slightly higher H$\beta$ FWHM than the Greene and Ho, (2005b) relation – the offset is 210 km s$^{-1}$ for a quasar with H$\alpha$ FWHM 4500 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, $\sigma_I$, we infer from the fit is 0.04 dex. This is smaller than the total scatter seen in Figure 3.9 (0.06 dex), which suggests that measurement errors make a significant contribution to the total scatter in the relation.

For 19 of the 99 quasars with H$\beta$ and H$\alpha$ emission profiles, one of the two Gaussians used to reproduce the H$\beta$ profiles has a FWHM greater than 20 000 km s$^{-1}$ and a fractional contribution to the total H$\beta$ broad line flux greater than 30 per cent (Marziani et al., 2009; Marziani et al., 2013). The very broad H$\beta$-component, which is not seen in the H$\alpha$ profiles, may be an artifact of the fitting scheme. A particular issue for H$\beta$ is the
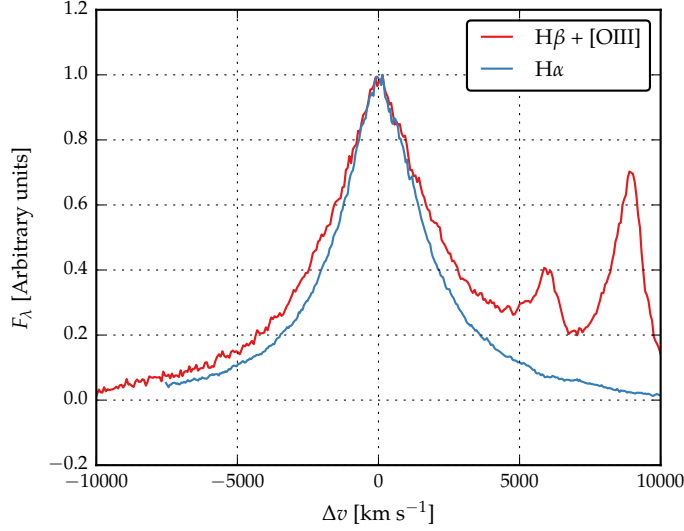
Figure 3.11: The Hα (blue) and Hβ (red) emission-line regions in the
median composite spectrum, shown as function of the ve-
locity shift from the respective predicted line peak wave-
lengths. The background continuum and optical Fe II
emission (Hβ only) has been modelled and subtracted.
The line fluxes have been scaled in order for the profile
shapes to be readily compared.

presence of Fe II emission, often at a significant level. Further-
more, additional lines could be contributing to the underlying
continuum (e.g. the He Iλλ4922,5017 doublet; Véron, Gonçalves,
and Véron-Cetty, 2002; Zamfir et al., 2010). If the Hβ FWHM
is calculated only from the narrower of the two Gaussian com-
ponents, then the Hβ FWHM decreases by $630\,\mathrm{km\,s^{-1}}$ on aver-
age, and the new relationship between the Hα and Hβ FWHM
is much closer to one-to-one. The C IV FWHM relative to the
Hα/Hβ FHWM will be enhanced by ∼ 15 per cent, and so the
C IV-based BH masses relative to the Balmer-based masses will
increase by ∼ 30 per cent.

### 3.4.2  *Measuring the quasar systemic redshift*

An accurate measure of the quasar's systemic redshift is re-
quired 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 sys-
temic 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 centroid. 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 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 $L_{Bol} \sim 10^{44}$ 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. We also tested models with no constraints on the centroids of the two broad Gaussians, and measured the systemic redshift from the peak of the composite profile. With this set of models, the median difference between the [O III]- and Hα(Hβ) based redshift estimates is reduced to $-100(-120)$ km s$^{-1}$, with a $290(320)$ km s$^{-1}$ scatter. This suggests that there is a $\sim 100$ km s$^{-1}$ systematic error in our Balmer-based redshift estimates. 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 parameterised.

### 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 Equation 3.4. In Figures 3.12a and 3.13a 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) is 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 Figure 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 Figure 3.12a we plot the maximum likelihood model and the 2-σ uncertainties on the model parameters. The maximum likelihood line is given by

$$\text{FWHM(C\,\textsc{iv}, Corr.)} = \frac{\text{FWHM(C\,\textsc{iv}, Meas.)}}{(0.41 \pm 0.02)\left(\frac{\text{C\,\textsc{iv} Blueshift}}{10^3\,\text{km\,s}^{-1}}\right) + (0.62 \pm 0.04)} \quad (3.5)$$

for the C IV/Hα fit and

$$\text{FWHM(C\,\textsc{iv}, Corr.)} = \frac{\text{FWHM(C\,\textsc{iv}, Meas.)}}{(0.36 \pm 0.03)\left(\frac{\text{C\,\textsc{iv} Blueshift}}{10^3\,\text{km\,s}^{-1}}\right) + (0.61 \pm 0.04)} \quad (3.6)$$

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 $0.23 \pm 0.02$ and $0.25 \pm 0.02$ for the Hα and Hβ fits respectively. The intrinsic scatter for the Hα fit is represented by the Normal probability density distribution shown in Figure 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.
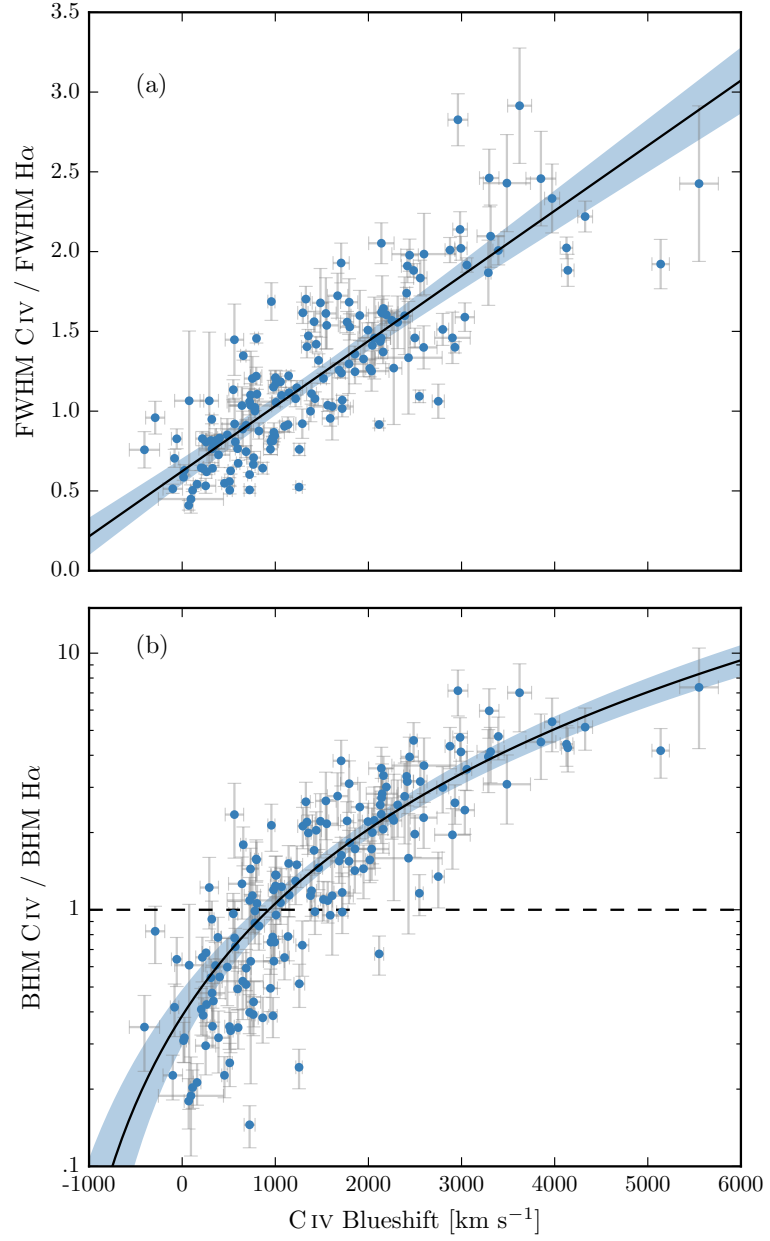
Figure 3.12: C IV FWHM relative to Hα FWHM (a), and C IV based BH
mass (BHM) compared to Hα based mass (b), both as a
function of the C IV blueshift. The black line is our best-fit
linear model, and the shaded region shows the 2-σ uncer-
tainties on the slope and intercept. The Hα FWHM have
been scaled to match the Hβ FWHM using Equation 3.4.
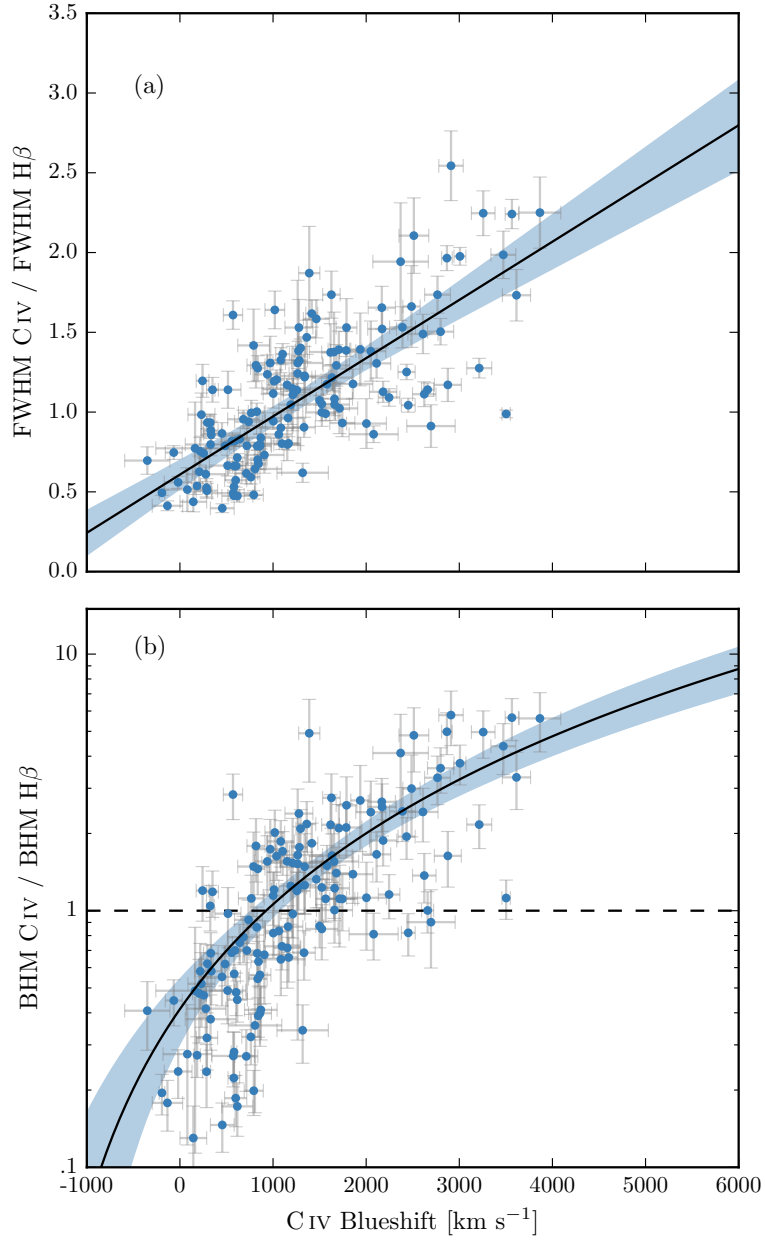
Figure 3.13: C IV FWHM relative to Hβ FWHM (a), and C IV based BH mass (BHM) compared to Hβ based mass (b), both as a function of the C IV blueshift.
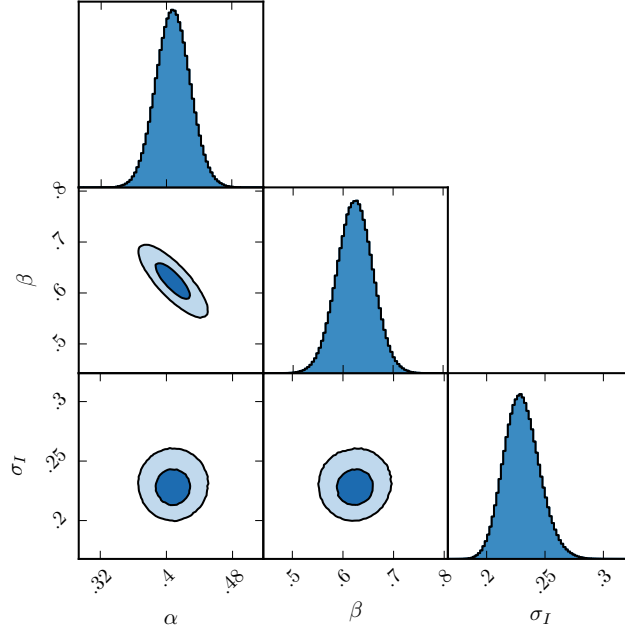
Figure 3.14: One- and two-dimensional projections of the MCMC sample of the posterior distribution for a linear fit to the FWHM C IV/Hα ratio as a function of the C IV blueshift. In the two-dimensional projections we show 1- and 2-σ contours. The posterior distribution for the linear fit to the FWHM C IV/Hβ ratio, which we do not show, has a very similar appearance.

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 $130\,\mathrm{km\,s^{-1}}$, compared to $340\,\mathrm{km\,s^{-1}}$ for Hβ.

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 (Equation 3.7) because of the issues related to the accurate modelling of the Hβ-profile just described. An extra advantage, which is evident in Figures 3.12a and 3.13a, is that the Hα sample has a better C IV blueshift coverage. How-
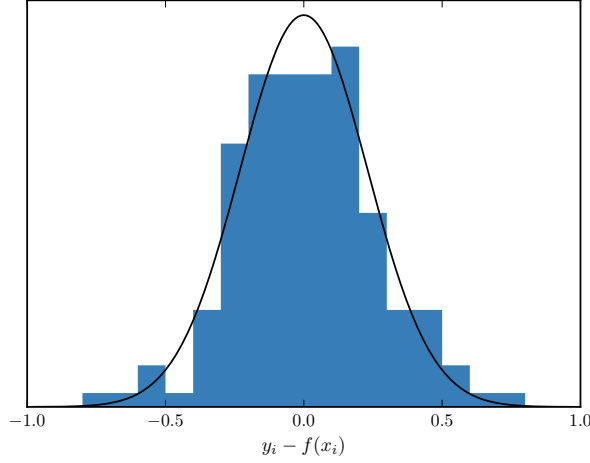
Figure 3.15: The distribution of the orthogonal displacement of each data point from the best-fitting linear relationship in the fit to FWHM(C IV)/FWHM(Hα) as a function of the C IV blueshift (blue histogram). The black curve is a Normal distribution with a width equal to the intrinsic scatter in the 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.

ever, as can be seen from the similarity of Equations 3.5 and 3.6, 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. BH masses are computed using the line and continuum properties given in Table 3.4, and we convert our Hα emission-line velocity-width measures to predicted Hβ widths using Equation 3.4.

In Figures 3.12b and 3.13b 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 $3000 \, \mathrm{km \, s^{-1}}$ in C IV emission blueshift and the overesti-

mate of the BH masses reaches a factor of 10 for quasars exhibiting the most extreme blueshifts, $\gtrsim 5000 \, \mathrm{km \, s^{-1}}$.

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 Equation 3.5 as an indicator of the virial velocity, and adopt the same $R_{BLR} - L$ relation for the 1350 Å continuum luminosity as Vestergaard and Peterson, (2006) (i.e. $R_{BLR} \propto L^{0.53}$). To find the constant scaling factor necessary to transform the virial product into 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:

$$\mathrm{MBH(C\,IV, Corr.)} = 10^{6.71} \left( \frac{\mathrm{FWHM(C\,IV, Corr.)}}{10^3 \, \mathrm{km \, s^{-1}}} \right)^2 \left( \frac{\lambda L_\lambda (1350\text{Å})}{10^{44} \, \mathrm{erg \, s^{-1}}} \right)^{0.53} . \quad (3.7)$$

Given measured C IV emission-line FWHM and blueshift, Equations 3.5 and 3.7 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

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 $\Delta V \propto R_{BLR}^{-0.5}$ virial relation, this leads to the prediction that the C IV line widths should be $\simeq 1.4$ times broader than Hα/Hβ for a given BH mass, and this expectation is reflected in the Vestergaard and Peterson, (2006) BH mass calibrations.

In our sample, the median C IV/Hα FWHM ratio is $0.97 \pm 0.31$ for the 77 quasars with C IV blueshifts $< 1200 \, \mathrm{km \, s^{-1}}$. This is significantly smaller than the predicted value of 1.4. 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 Figure 3.12b, where for almost ev-

ery quasar with a C IV blueshift $<1200\,\mathrm{km\,s^{-1}}$, the C IV-derived BH mass is smaller than the corresponding Hα-derived mass.

Denney, (2012) used multiple-epoch spectra from reverberation mapping campaigns to isolate the part of the C IV profile which is varying. In some profiles they identified a non-varying core component, possibly originating from gas at larger radii than the BLR. In single-epoch spectra, both parts of the line - varying and non-varying - are measured. The gas at larger radii will enhance the profile at lower-velocity and lead to smaller FWHM or dispersion values. This could explain the low C IV/Hα FWHM ratio observed in our sample of quasars with small C IV blueshifts.

## 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 3.5 and 3.7 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 ultra-violet 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 demonstrated in Figure 3.16a, in which we plot the C IV-blueshift versus C IV-emission EQW 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 EQW is not evident in Figure 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 are predominantly responsible for the differences seen in Figures 3.16a and 3.16b; the appearance in Figure 3.16b of the extension to high blueshift for quasars with low C IV EQW is particularly evident.

Shen et al., (2016) and our own work shows that there is an intrinsic variation of $\sigma \simeq 220 \, \mathrm{km \, s^{-1}}$ in the velocity centroids of the BLR emission-lines relative to a systemic-frame defined by emission-lines in the quasar NLR. The redshifts for quasars in the SDSS DR10 and DR12 catalogues (Pâris et al., 2014; Pâris et al., 2017) possess errors of $\simeq 500 - 750 \, \mathrm{km \, s^{-1}}$ (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 3.3 includes the values for the fractional error in the corrected BH mass that result from a given error 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 $1000 \, \mathrm{km \, s^{-1}}$ C IV blueshift when there is a $500 \, \mathrm{km \, s^{-1}}$ uncertainty in the quasar systemic redshift.

Of potentially more significance for studies of BH masses as a function of quasar and host-galaxy properties are redshift errors that depend on the form of the quasar ultra-violet SED. The systematic variation in the C IV shape is correlated with changes in the quasar SEDs, including the strengths of the Si III]$\lambda$1892 and C III]$\lambda$1908 emission-lines in the rest-frame ultra-violet (Figure 3.17). As a consequence, the redshifts from Hewett and Wild, (2010) still suffer from systematic errors that are correlated with the shape, and particularly the blueshift,
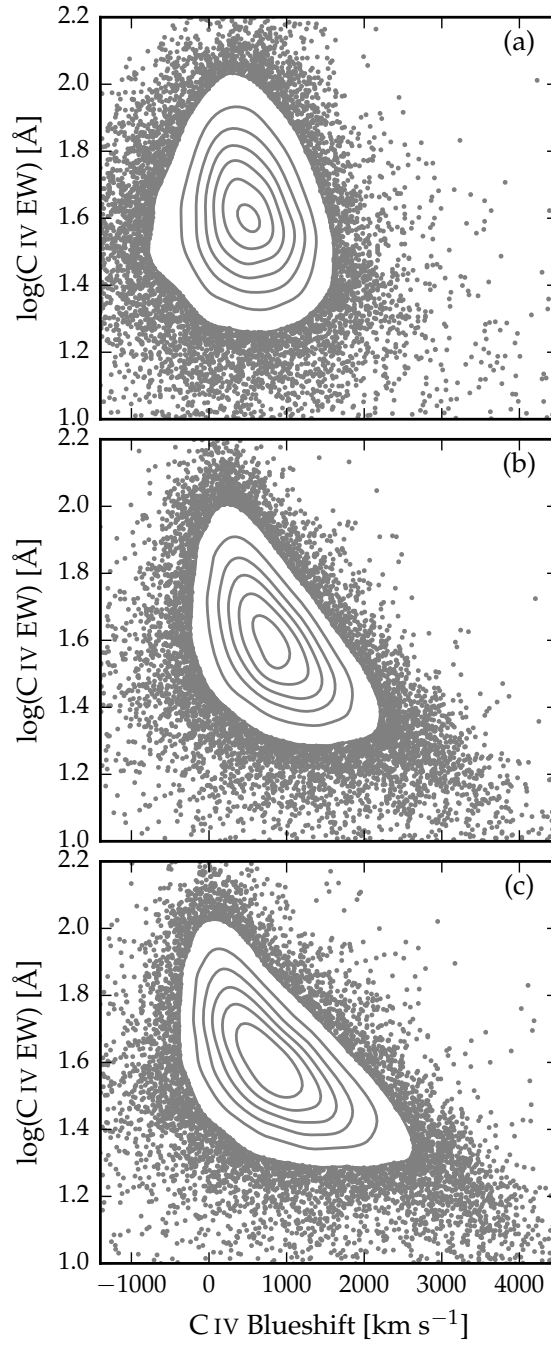
Figure 3.16: Rest-frame EQW versus blueshift of the broad C IV emission-line for 32,157 SDSS DR7 quasars at 1.6 < z < 3.0. Panel (a) uses C IV line parameters from Shen et al., (2011) and SDSS pipeline systemic redshifts. Panels (b) and (c) use systemic redshifts from Hewett and Wild, (2010) and Allen & Hewett (2017, in preparation) respectively, and C IV line measurements described in Section 3.5.1.2.

Table 3.3: The fractional error on the corrected BH mass as a function of C IV blueshift for different uncertainties in the quasar systemic redshift.

| $\delta v$ [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 |
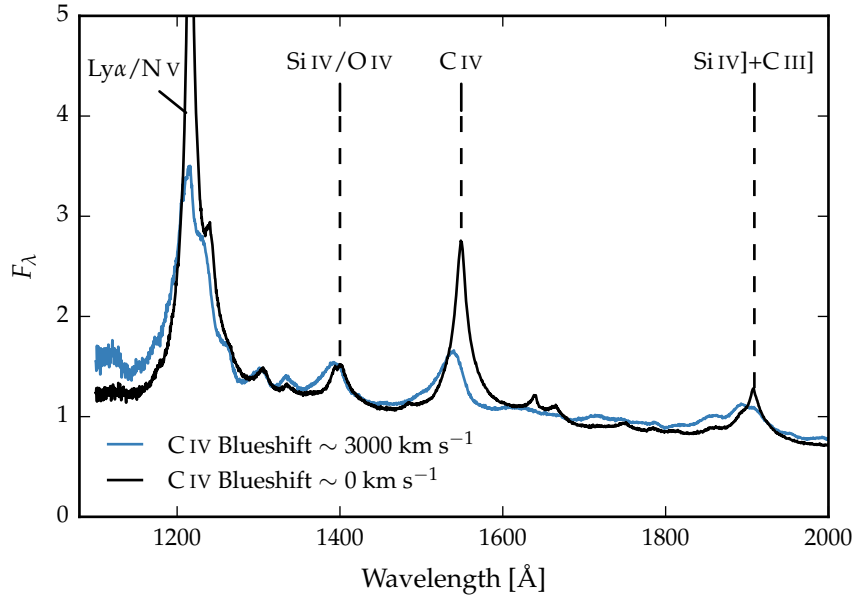


Figure 3.17: Composite spectra from Ly$\alpha$ to C III] for quasars with different C IV blueshifts. The systematic variation in the C IV shape is correlated with changes in the quasar SEDs, including the strengths of the Si III]$\lambda$1892 and C III]$\lambda$1908 emission-lines.

of the C IV emission-line. For the Hewett and Wild, (2010) redshifts, and ultra-violet emission-line based redshifts in general, quasars with large C IV EQW and modest blueshifts have relatively small ($\simeq 300$ km s$^{-1}$) SED-dependent redshift errors. Redshift uncertainties 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 3.3 is pessimistic.

Conversely, systematic redshift errors are greatest for quasars with large blueshifts, reaching $\sim 750$ km 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 Figures 3.12b and 3.13b). 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 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 (2017, 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 ultra-violet (over wavelengths from Mg II$\lambda\lambda$2796,2803 down to the O I$\lambda$1304+Si II$\lambda$1307 blend) using the new redshift-algorithm are located at rest-frame wavelengths in excellent agreement with the systemic redshift defined using the rest-frame narrow-line optical [O III] doublet and broad-line H$\beta$ and H$\alpha$. SED-dependent systematic errors are below the apparent inherent dispersion of $\simeq 220\,\mathrm{km\,s^{-1}}$ 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 Figure 3.16b, in particular the extension to high blueshift at low C IV EQW, become more apparent in Figure 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 EQW 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 Figure 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 parametrisation, 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 determine 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. The continuum is first modelled and subtracted using the procedure described in Section 3.3.1. The C IV emission-line is taken to lie within the wavelength interval 1500-1600 Å. 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 2-$\sigma$ 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[5]. The wavelength that bisects the cumulative total line flux is recorded and the blueshift computed using Equation 3.1.

---

5 While the absorption identification and interpolation scheme is not effective for spectra that possess associated absorption spread over more than $\sim 1200 \, \mathrm{km \, s^{-1}}$, the sample you are using does not contain CIV-lines with such extensive absorption. These are in SDSS but have been removed?

## 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] EQW, and Fe II EQW. The only significant correlation we find is with the Hα FWHM (Figure 3.18). 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 something 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). Figure 3.18 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 an emission-line can be characterised by the ratio FWHM/σ. FWHM/σ ≃ 2.35 for a Gaussian profile, while FWHM/σ ≃ 1 for a peakier Lorentzian profile[6]. In our sample, we find the model 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). This suggests that the BLR structure (e.g. the balance between rotation, turbulence and radial motions) is changing with the emission-line FWHM (e.g. Collin et al., 2006; Kollatschny and Zetzl, 2011; Kollatschny and Zetzl, 2013). This raises the question of whether the Balmer-line FWHM is a reliable proxy for the virial-induced velocity dispersion for the full range of Balmer line shapes we have in our sample.

---

6 Strictly FWHM/σ → 0 for a Lorentzian profile, but values close to unity are typical when the dispersion is calculated over a velocity range, ≃ ±10 000 km s$^{-1}$, used to parametrize broad emission-lines in quasar spectra.

When calibrating the virial-product to masses derived independently using the $M_{BH} - \sigma$ relation, Collin et al., (2006) find that the scaling factor, f, is a factor $\sim 2$ larger for their Population '1' sources (with FWHM/$\sigma < 2.35$ and essentially equivalent to population A of Sulentic et al. 2000) than for their Population 2 (with FWHM/$\sigma > 2.35$ and equivalent to population B of Sulentic et al. 2000). 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. This result could account for some of the remaining scatter between the C IV- and Balmer-based BH masses (Figure 3.18).

Shen and Ho, (2014) argue that a large part of the scatter observed in the H$\beta$ FWHM relates not to a spread in BH masses, but rather to the orientation of the BLR relative to the line-of-sight of the observer. This would be the case if the BLR had a flattened disc-like geometry. In this case, the observed line width would increase with the inclination of the disc relative to the line of sight. 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). Twenty core-dominated quasars and six lobe-dominated quasars were identified in our sample, but no statistically significant differences in the H$\alpha$ 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).

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

Figure 3.19 demonstrates that our sample has an excellent coverage of the C IV EQW-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 (Figure 2.1). Our 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 investiga-
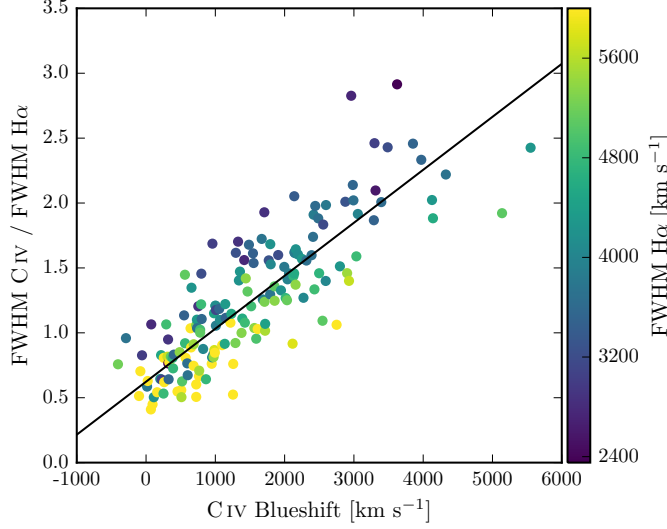
Figure 3.18: Same as Figure 3.12a, with the marker colour representing the Hα FWHM. At fixed C IV blueshift, there is a clear Hα FWHM dependent systematic in the model residuals.

tion 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. In particular, quasars with negative blueshifts as large as $\sim 1000\,\mathrm{km\,s^{-1}}$ appear in the SDSS DR7 catalogue and applying our correction in this regime boosts the derived masses by unphysical factors.

Figure 3.20 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 Figure 3.7b but the systematic trend relative to the spread at fixed blueshift is significantly smaller
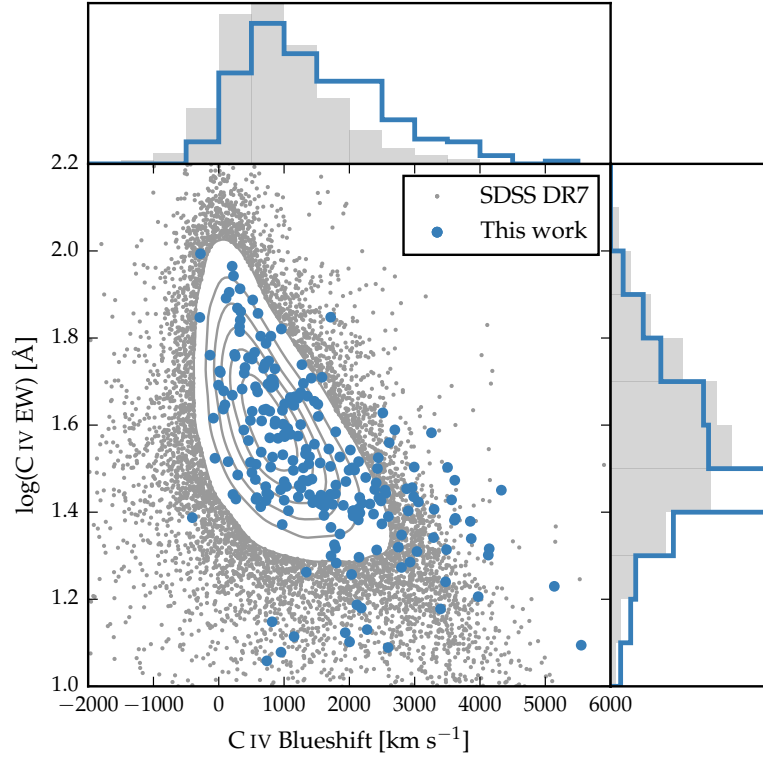
Figure 3.19: Rest-frame EQW versus blueshift of the broad C IV emission-line for SDSS DR7 quasars and our sample (see Figure 3.16b for details). Our sample has very good coverage of the parameter space; the shift to high blueshifts is a result of the high luminosity of our sample in relation to the SDSS sample and the correlation between luminosity and blueshift.
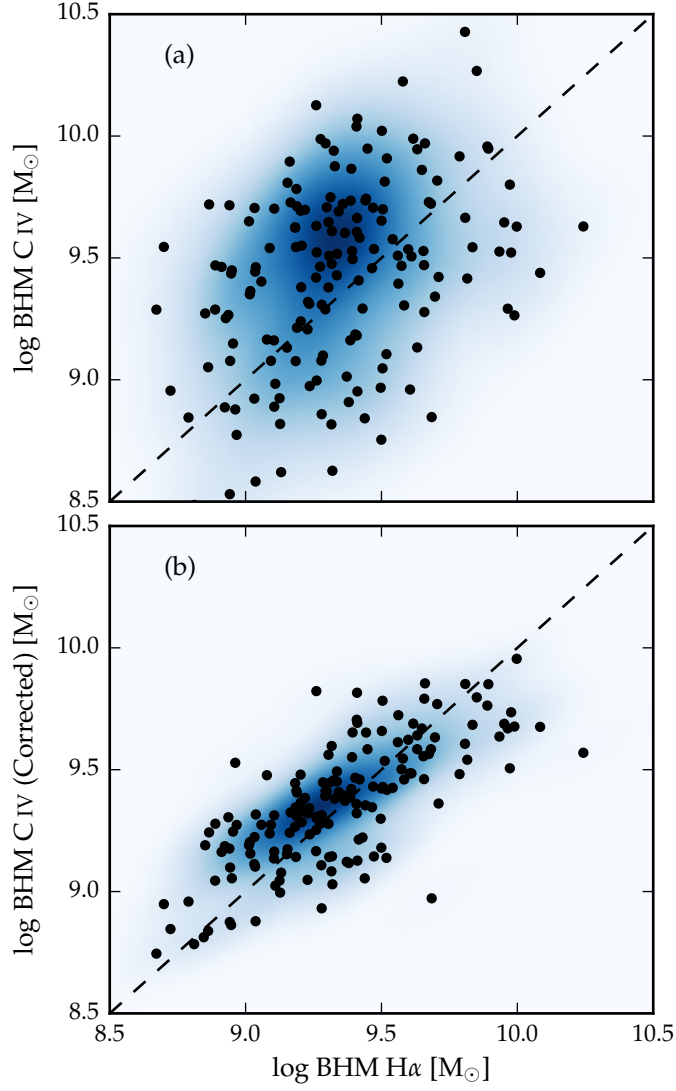
Figure 3.20: Comparison of the C IV- and Hα-based BH masses before (a) and after (b) applying the C IV blueshift-based correction to the C IV FWHM. The density of the plotted points (estimated using a Gaussian kernel density estimator) is represented by the colour. The correction to the C IV BH masses decreases the scatter from 0.4 to 0.2 dex.
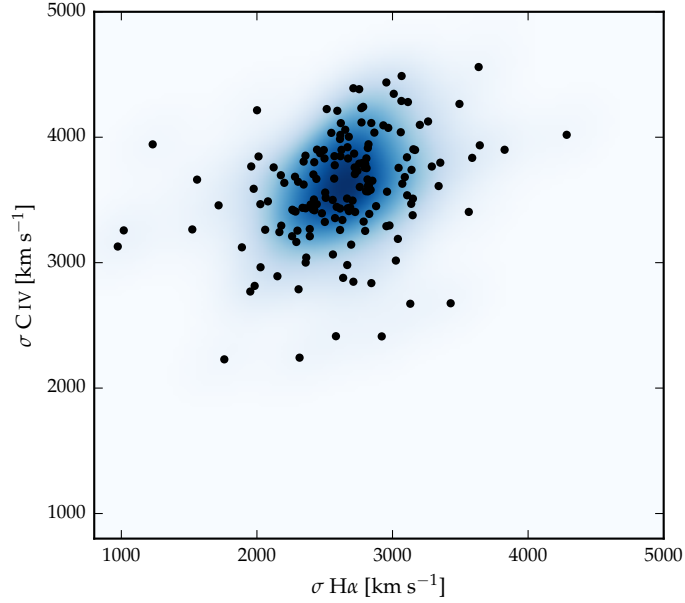
Figure 3.21: Comparison of the C IV and Hα line dispersion, σ. The density of the plotted points (estimated using a Gaussian kernel density estimator) is represented by the colour. Estimating a reliable BH mass from the C IV FWHM and blueshift is substantially more effective than using the C IV line 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).

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 (Figure 3.21).

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 km s$^{-1}$ for Hα and C IV respectively. This 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 3.20 and 3.21 demonstrate that estimating a reliable BH mass from the C IV FWHM and

blueshift 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 Figure 3.22 we compare 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 case, we compare the corrected C IV-based masses to the H$\alpha$-based masses as a function of the C IV blueshift. The C IV blueshift-based correction presented in this chapter is also tested in Figure 3.23. The correction proposed by Runnoe et al., (2013) 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 Figure 3.22a the C IV BH masses have been corrected using the C IV shape (FWHM/$\sigma$) 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/$\sigma$) of the C IV profile. In our sample, we observe a strong correlation between the shape of the C IV line and its blueshift (Figure 3.7c); between the two extremes in the C IV blueshift distribution the line shape changes from FWHM/$\sigma$ ~ $1 - 2.5$. 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. As a result, the correction is not applicable at large C IV blueshifts. Therefore, while the consistency between the H$\alpha$- 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 ultra-violet 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., (2013) opted instead to use the continuum-subtracted peak flux ratio of the ultra-violet 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
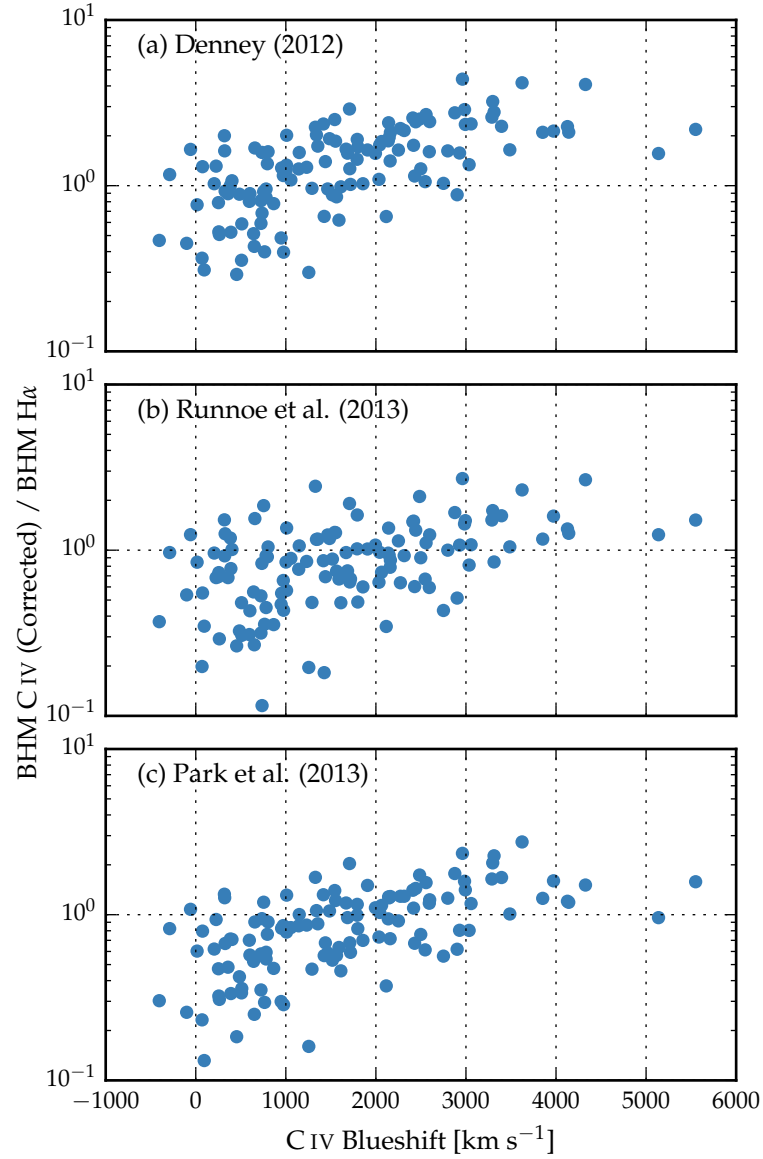
Figure 3.22: Comparison of BH mass estimates derived from C IV and Hα as a function of the C IV blueshift. Corrections to the C IV-based masses have been applied based on the shape (FWHM/σ) of the C IV emission-line (a; Denney, 2012), the peak flux ratio of the Si IV+O IV blend relative to C IV (b; Runnoe et al., 2013), and by significantly reducing the dependence of the derived BH mass on the C IV velocity-width (c; Park et al., 2013).
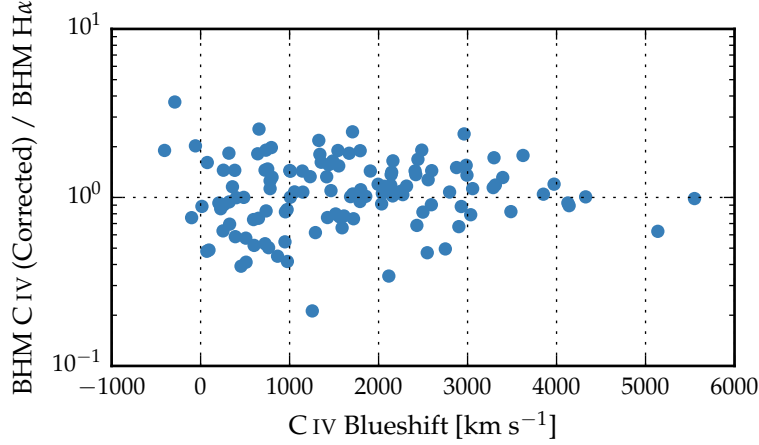
Figure 3.23: Comparison of BH mass estimates derived from C IV and Hα as a function of the C IV blueshift. C IV-based masses have been corrected using the C IV blueshift-based prescription presented in this Chapter.

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., (2013), 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 Figure 3.19, a correlation exists between the blueshift and EQW 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 (see Figure 3.17). 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 (Figure 3.22b). 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 $\Delta V^2$ dependence predicted assuming a virialized BLR to just $\Delta V^{0.56}$). As a consequence, the C IV based masses of the quasars with large C IV blueshifts are much reduced (Figure 3.22c). However, the systematic error in the C IV-based BH masses as a function of C IV blueshift remains.

For comparison, the C IV-based masses shown in Figure 3.23 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 Figure 3.24, there are systematic variations in the H$\alpha$ line profile as a function of the C IV blueshift. At C IV-blueshift $< 1200\,\mathrm{km\,s^{-1}}$, the H$\alpha$ 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 H$\alpha$ FWHM $= 3500\,\mathrm{km\,s^{-1}}$, with a scatter of just $700\,\mathrm{km\,s^{-1}}$. The apparent trend of peakier H$\alpha$-emission, with FWHM/$\sigma$ close to unity, at large C IV-blueshift is enhanced by the modest increase in H$\alpha$ EQW with blueshift. Amongst the low-C IV-blueshift population there are in addition quasars with broader and more Gaussian-like H$\alpha$ line profiles, with FWHM/$\sigma \simeq 2$ .

The change in the H$\alpha$ emission-line profiles as a function of C IV-blueshift means that the H$\alpha$-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, into a bolometric luminosity we use the bolometric correction factor given by Richards et al., (2006a) ($L_{bol} = 9.26 L_{5100}$). 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 Figure 3.25, demonstrate that at large blueshifts quasars are accreting at around their Eddington limits (Figure 3.25). 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-ultra-violet 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 H$\alpha$ FWHM. Our results there-
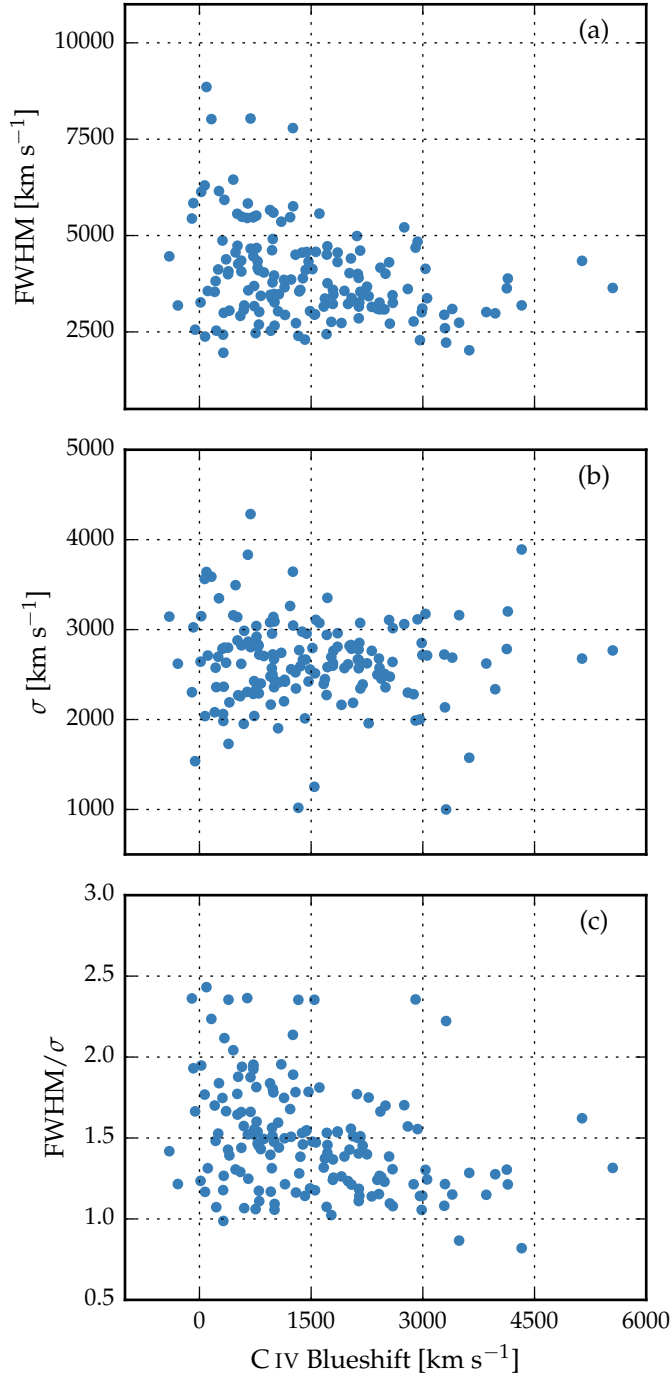
Figure 3.24: The FWHM, dispersion (σ) and shape (FWHM/σ) of Hα as a function of the C IV blueshift.
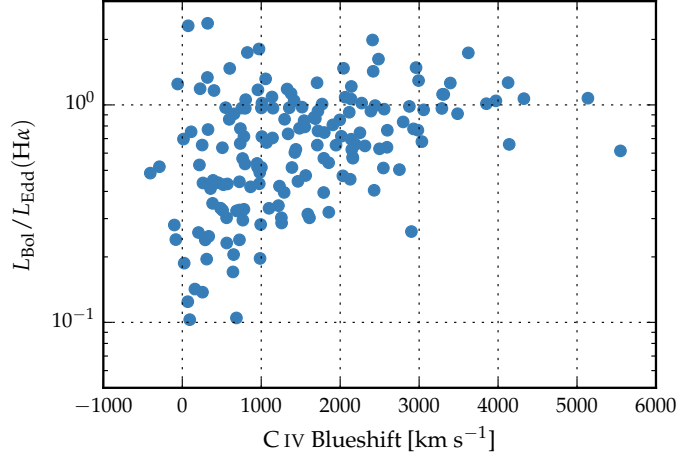
Figure 3.25: Hα-derived Eddington ratio versus C IV blueshift. At blueshift $\gtrsim 2000\,\mathrm{km\,s^{-1}}$ all quasars have high accretion rates ($L/L_{Edd} \simeq 1$). This is in agreement with Kratzer and Richards, (2015), but in contrast to what one would derive from naive use of C IV-based BH mass scaling relations.

fore support the 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 parametrisation and the apparent anti-correlation between C IV EQW 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).

As we described in Section 3.5.2, the shape of the Balmer lines (FWHM/σ) depends strongly on the FWHM. This suggests that the BLR structure is changing with the emission-line FWHM. One possibility we discussed is that the Balmer line-

width is orientation-dependent (e.g. Shen and Ho, 2014). 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. 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.

As mentioned in Section 3.1 and discussed in Richards et al., (2011), quasars with current reverberation mapping measurements have a restricted range of C IV-line shapes. In particular, 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.

### 3.6.1 *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-EQW 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.

To test this hypothesis, we selected 18 C IV BAL quasars from our catalogue with near-infrared spectra including Hβ. Using
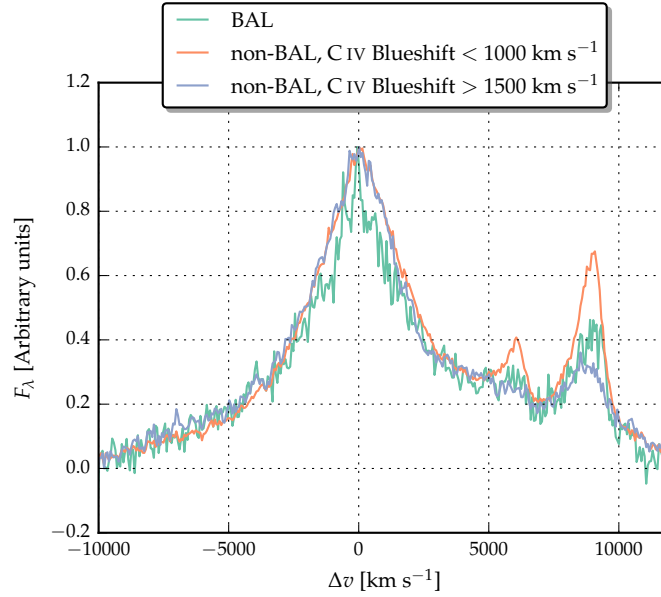
Figure 3.26: Composite spectra of the Hβ/[O iii] region made of BAL quasars, non-BAL quasars with C iv blueshifts below 1000 km s$^{-1}$, and non-BAL quasars with C iv blueshifts above 1500 km s$^{-1}$. Add sentence explaining what we see.

the same method described in Section 3.4.1, we constructed a median composite spectrum from this sample. We also constructed composite spectra for quasars with modest and large C iv blueshifts. The results are shown in Figure 3.26. We find that [O iii] is very weak and Hβ is relatively narrrow in the BAL quasars (e.g. Yuan and Wills, 2003). Therefore, as predicted, the optical emission-line properties of the BAL quasars are very similar to the properties of the non-BAL quasars with large C iv blueshifts.

*Hβ narrower? Ask Paul about this plot.*

### 3.6.2 *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, etc. 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 Figure 3.25), 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[7] 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 $\sim 2:1$. 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. $\sim 15$ per cent of the ultra-violet-selected non-BAL quasar population).

## 3.7 SUMMARY

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

---

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

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 ultra-violet 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 3.5 and 3.7 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 and the Hα-based masses is 0.24 dex at low C IV blueshifts ($\sim 0\,\mathrm{km\,s^{-1}}$) and 0.10 dex at high blueshifts ($\sim 3000\,\mathrm{km\,s^{-1}}$). 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 CATALOGUE OF DERIVED PROPERTIES

Table 3.4 includes the line parameters from our emission-line fits, and other derived properties used in this chapter. The columns in Table 3.4 are as follows:

1 Catalogue name.

2-3 Broad Hα FWHM, and its error, in $\mathrm{km\,s^{-1}}$.

4-5 Broad Hα line dispersion, and its error, in $\mathrm{km\,s^{-1}}$.

6-7 Broad Hα redshift, and its error.

8-9 Hα-FWHM-based BH mass using Vestergaard and Peterson, (2006) calibration, and its error, in $M_\odot$. Hα FWHM is first converted into equivalent Hβ FWHM using Equation 3.4.

10 Hα flag. When flag is 1 emission-line measurements are unreliable and emission-line is excluded from the main analysis in this chapter (see Section 3.3.5).

11-12 Broad Hβ FWHM, and its error, in $\mathrm{km\,s^{-1}}$.

13-14 Broad Hβ line dispersion, and its error, in $\mathrm{km\,s^{-1}}$.

15-16 Broad Hβ redshift, and its error.

17-18 Hβ-FWHM-based BH mass using Vestergaard and Peterson, (2006) calibration, and its error, in $M_\odot$.

19 Hβ flag. When flag is 1 emission-line measurements are unreliable and emission-line is excluded from the main analysis in this chapter (see Section 3.3.5).

20-21 Broad C ɪv FWHM, and its error, in $\mathrm{km\,s^{-1}}$.

22-23 Broad C ɪv line dispersion, and its error, in $\mathrm{km\,s^{-1}}$.

24-25 Broad C ɪv EQW, and its error, in Å.

26-27 Si ɪv+O ɪv/C ɪv peak flux ratio, and its error.

28 C ɪv flag. When flag is 1 emission-line measurements are unreliable and emission-line is excluded from the main analysis in this chapter (see Section 3.3.5).

29-30 C ɪv blueshift, relative to Hα, and its error, in $\mathrm{km\,s^{-1}}$.

31-32 C ɪv blueshift, relative to Hβ, and its error, in $\mathrm{km\,s^{-1}}$.

33-34 Uncorrected C ɪv-FWHM-based BH mass using Vestergaard and Peterson, (2006) calibration, and its error.

35-36 Corrected C ɪv-FWHM-based BH mass, and its error.

Table 3.4: The format of the table containing the emission-line properties from our parametric model fits. The table will be available in machine-readable form online <span style="color:red">where online?</span>.

| Column | Name | Units | Description |
|:---:|:---:|:---:|:---:|
| 1 | UID | | Catalogue name |
| 2 | FWHM_BROAD_HA | $\mathrm{km\,s^{-1}}$ | Broad H$\alpha$ FWHM |
| 3 | FWHM_BROAD_HA_ERR | $\mathrm{km\,s^{-1}}$ | |
| 4 | SIGMA_BROAD_HA | $\mathrm{km\,s^{-1}}$ | Broad H$\alpha$ $\sigma$ |
| 5 | SIGMA_BROAD_HA_ERR | $\mathrm{km\,s^{-1}}$ | |
| 6 | Z_BROAD_HA | | H$\alpha$ redshift |
| 7 | Z_BROAD_HA_ERR | | |
| 8 | LOGMBH_HA | $M_\odot$ | H$\alpha$ BH mass |
| 9 | LOGMBH_HA_ERR | $M_\odot$ | |
| 10 | FLAG_HA | | H$\alpha$ flag |
| 11 | FWHM_BROAD_HB | $\mathrm{km\,s^{-1}}$ | Broad H$\beta$ FWHM |
| 12 | FWHM_BROAD_HB_ERR | $\mathrm{km\,s^{-1}}$ | |
| 13 | SIGMA_BROAD_HB | $\mathrm{km\,s^{-1}}$ | Broad H$\beta$ $\sigma$ |
| 14 | SIGMA_BROAD_HB_ERR | $\mathrm{km\,s^{-1}}$ | |
| 15 | Z_BROAD_HB | | H$\beta$ redshift |
| 16 | Z_BROAD_HB_ERR | | |
| 17 | LOGMBH_HB | $M_\odot$ | H$\beta$ BH mass |
| 18 | LOGMBH_HB_ERR | $M_\odot$ | |
| 19 | FLAG_HB | | H$\beta$ flag |
| 20 | FWHM_CIV | $\mathrm{km\,s^{-1}}$ | Broad C IV FWHM |
| 21 | FWHM_CIV_ERR | $\mathrm{km\,s^{-1}}$ | |
| 22 | SIGMA_CIV | $\mathrm{km\,s^{-1}}$ | Broad C IV $\sigma$ |
| 23 | SIGMA_CIV_ERR | $\mathrm{km\,s^{-1}}$ | |
| 24 | EQW_CIV | Å | Broad C IV EQW |
| 25 | EQW_CIV_ERR | Å | |
| 26 | 1400_CIV | | Si IV+O IV/C IV peak flux ratio |
| 27 | 1400_CIV_ERR | | |
| 28 | FLAG_CIV | | C IV flag |
| 29 | BLUESHIFT_CIV_HA | $\mathrm{km\,s^{-1}}$ | C IV blueshift, relative to H$\alpha$ |
| 30 | BLUESHIFT_CIV_HA_ERR | $\mathrm{km\,s^{-1}}$ | |
| 31 | BLUESHIFT_CIV_HB | $\mathrm{km\,s^{-1}}$ | C IV blueshift, relative to H$\beta$ |
| 32 | BLUESHIFT_CIV_HB_ERR | $\mathrm{km\,s^{-1}}$ | |
| 33 | LOGMBH_CIV_VP06 | $M_\odot$ | Uncorrected C IV BH mass |
| 34 | LOGMBH_CIV_VP06_ERR | $M_\odot$ | |
| 35 | LOGMBH_CIV_C17 | $M_\odot$ | Corrected C IV BH mass |
| 36 | LOGMBH_CIV_C17_ERR | $M_\odot$ | |

# QUASAR-DRIVEN OUTFLOWS IN THE NARROW LINE REGION

## 4.1 INTRODUCTION

X-ray and ultra-violet spectroscopy has revealed high-velocity outflows to be nearly ubiquitous in high accretion rate quasars. Strong evidence for high-velocity outflows include BALs, NALs and blueshifts in high-ionisation broad emission-lines. This suggests that the energy released by quasars can have a dramatic effect on their immediate environment.

The BH mass and the mass of the host-galaxy spheroid are strongly correlated. This suggests that the BH and bulge grown synchronously, with the energetic output of the rapidly-accreting BH coupling with the gas in the host-galaxy and regulating star formation and the growth of the BH itself (e.g. Silk and Rees, 1998; King, 2003; Di Matteo, Springel, and Hernquist, 2005; King and Pounds, 2015). If this picture is true, then quasars should be capable of driving powerful outflows over galactic scales. In recent years, a huge amount of resources have been devoted to searching for observational evidence of these galaxy-wide, quasar-driven outflows. This has resulted in recent detections of outflows in quasar-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 using forbidden emission-lines to probe conditions in the AGN NLR. Because of its high equivalent width, [O III]$\lambda\lambda 4960, 5008$ is the most studied of the narrow AGN emission-lines. The [O III] emission is found to consist of two distinct 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 is dominated by the gravitational potential of the host-galaxy whereas the broad, blueshifted wing traces

outflowing gas. The relative balance between the core and wing components varies significantly from object to object, and may depend on properties of the AGN (e.g. luminosity; Shen and Ho, 2014).

Observations of broad velocity-widths and blueshifts 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, these studies rely on small samples, which are often unrepresentative of the properties of the AGN population. 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 and drivers of ionised outflows. At the same time, there is strong evidence from spatially resolved spectroscopic observations that these outflows are 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 BH accretion peaked ($2 \lesssim z \lesssim 4$), which is when quasar feedback is predicted to be at its most effective. At these redshifts bright optical emission-lines including the [O III] doublet are redshifted to near-infrared wavelengths, where observations are much more challenging. As a consequence, studies at high redshifts have typically relied on relatively small numbers of objects (e.g. Netzer et al., 2004; Sulentic et al., 2004; Shen, 2016). These studies find [O III] to be broader in more luminous AGN, suggesting that AGN efficiency in driving galaxy-wide outflows increases with luminosity (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). The fraction of objects with very weak [O III] emission also appears to increase with redshift and/or luminosity (e.g. Netzer et al., 2004).

Other recent studies have looked at the [O III] emission properties of extreme objects - e.g. heavily obscured quasars (Zakamska et al., 2016) and the most luminous quasars (Bischetti et al., 2016) - at redshifts $z \sim 2$. When detected, the [O III] emission in these objects is extremely broad and strongly blueshifted. These observations are consistent with galaxy formation mod-

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 | 7 |
| 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 |
| Total | | 354 |

els that predict AGN feedback to be strongest in luminous, dust-obscured quasars.

In this chapter we analyse the [O III] properties of a sample of 354 high-luminosity, redshift $1.5 < z < 4$ quasars. To date, this is the largest study of the NLR properties of high redshift quasars.

## 4.2 QUASAR SAMPLE

From our near-infrared spectroscopic catalogue (Chapter 2), we have selected 354 quasars which have spectra covering the strong, narrow [O III] doublet. The broad Balmer Hβ line has also been observed for all but two of the sample. For 165 quasars, the spectra extend to the broad Hα emission-line at 6565 Å, and in 260 objects optical spectra, including C IV, are also available (mostly from SDSS/BOSS). The sample covers a wide range in redshifts ($1.5 \lesssim z \lesssim 4$) and luminosities ($45.5 \lesssim \log L_{Bol} \lesssim 49$ erg s$^{-1}$). The spectrographs and telescopes used to obtain the near-infrared spectra are summarised in Table 4.1.

In this section, we describe how emission-line parameters are derived. Our approach is to model the spectra using a power-law continuum, an empirical Fe II template (taken from Boroson and Green 1992) and multiple Gaussian components to model the emission from the broad and narrow emission-line regions. Non-parametric properties are then derived from the best-fitting model. This approach, which is commonly adopted in the literature (e.g. Shen et al., 2011; Shen and Liu, 2012; Shen, 2016), is more robust when analysing spectra with limited S/N (in comparison to measuring line properties directly from the data) and allows different emission-lines to be de-blended.

The same approach was used to model the Hβ/[O III] complex in Chapter 3. However, a number of small adjustments have been made to the model (Section 4.3.3). Hα emission-line properties (used to estimate the quasar systemic redshift) are also re-derived in this chapter using a slightly modified model (Section 4.3.4) to the one adopted in Chapter 3. C IV emission-line properties (used to infer the strength of BLR outflows) are taken directly from Chapter 3.

### 4.3.1 *Transforming spectra to the quasar rest-frame*

Before a spectrum can be modelled, it must first be transformed to 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). The rest-frame transformation is only required to be accurate to within $\sim 1000\,\mathrm{km\,s^{-1}}$ of the systemic redshift for our fitting procedure to work. In later sections, more precise estimates of the systemic redshift will be calculated using our parametric model fits.

### 4.3.2 *Removing Fe II emission*

Fe II emission is generally strong in the vicinity of Hβ/[O III]. Therefore, before Hβ/[O III] is modelled, we first model and subtract the continuum and Fe II emission using the procedure described in Chapter 3.

We encountered 24 objects for which Fe II emission appears to be present in the spectrum even after the subtraction proce-
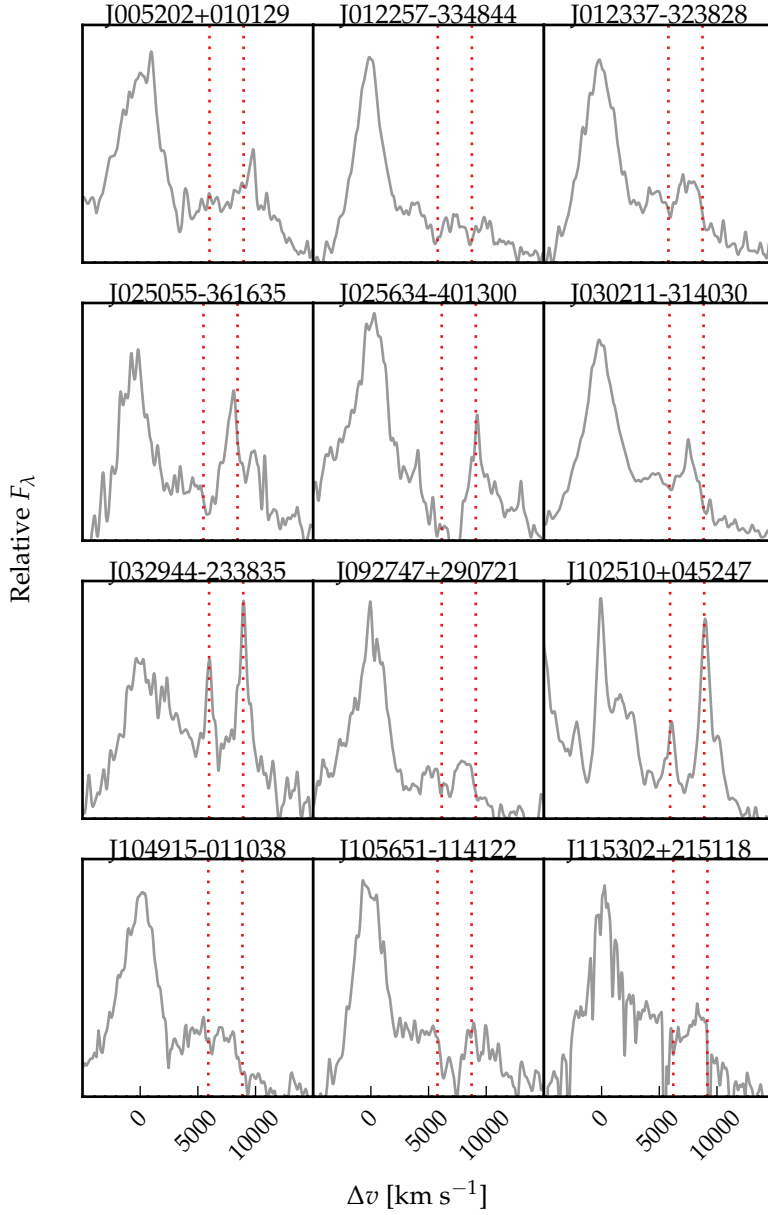
Figure 4.1: Spectra of the 24 objects for which significant Fe II emis-
sion is still visible following our Fe II-subtraction proce-
dure. Spectra have been smoothed via convolution with
a $100\,\mathrm{km\,s^{-1}}$ Gaussian kernel. The vertical lines indicate
the expected positions of the [O III] doublet (which is gen-
erally very weak) with the systemic redshift defined using
the peak of the broad Hβ emission.

J123355+031328     J125141+080718     J134104-073947

J141949+060654     J204010-065403     J212912-153841

J214507-303046     J214950-444405     J215052-315824

J223246-363203     J223820-092106     J232539-065259

Relative $F_\lambda$

$\Delta v$ [km s$^{-1}$]

Figure 4.1: Continued.

Table 4.2: Summary of models used to fit the Hβ emission, and the number of quasars each model is applied to.

| Model | Fix centroids? | Number |
|---|---|---|
| 2 broad Gaussians + 1 narrow Gaussian | No | 9 |
| 2 broad Gaussians | No | 274 |
| 2 broad Gaussians | Yes | 39 |
| 1 broad Gaussian | N/A | 8 |

dure (Figure 4.1). In these objects the relative strengths of the Fe II lines differ significantly from those of I Zw 1, on which the Boroson and Green, (1992) Fe II template we use is based. The residual Fe II emission is at rest-frame wavelengths very close to zero-velocity wavelengths of the [O III] doublet, which is generally very weak in these objects. The Gaussians we fit to the spectra to model [O III] are therefore strongly biased, and the [O III] emission properties we infer from the model are in error by large factors.

For example, J223819-092106 was analysed by Shen, (2016) using a very similar model. Shen, (2016) reported the [O III] emission in this object to be shifted by $\sim 7500 \, \mathrm{km\,s^{-1}}$ relative to the Hewett and Wild, (2010) systemic redshift. Our analysis suggests that emission which was modelled by Shen, (2016) as [O III] is more likely to be poorly-subtracted Fe II emission. Because the derived [O III] emission properties can be strongly biased in objects so-affected, these objects are flagged and are excluded from our analysis in the remainder of this chapter (leaving 330 objects in our sample).

### 4.3.3 *Modelling Hβ/[O III]*

The Hβ and [O III] emission is fit using a similar procedure to the one described in Chapter 3. However, we make a number of modifications to the parametric model employed, which we will now describe.

### 4.3.3.1 *Hβ*

In general, Hβ is modelled by two Gaussians with non-negative amplitudes and FWHM greater than $1200 \, \mathrm{km\,s^{-1}}$. 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 gen-

erally have low S/N, and adding extra freedom to the model does not significantly decrease the reduced $\chi^2$. In addition there are cases where the blue wing of the Hβ emission is below the lower wavelength limit of the spectra; in these cases models with more freedom are insufficiently constrained by the data.

Contributions to the Hβ emission from the NLR 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 nine 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 NLR in other quasars in our sample. If this is the case then measures of the Hβ velocity width will be biased to lower values on average. However, our systemic redshift estimates that use the peak of the Hβ emission (Section 4.3.8) will not be affected. The Hβ models, and the numbers of quasars each model is applied to, are summarised in Table 4.2.

### 4.3.3.2 *[O III]*

Each component of the [O III] doublet is fit with one or two Gaussians, depending on the fractional reduced $\chi^2$ difference between the one- and two-component models. Concretely, if the addition of the second Gaussian decreases the reduced $\chi^2$ by more than 5 per cent then the double-Gaussian model is accepted. One hundred and twenty-eight spectra are fit with a single Gaussian and 140 with two Gaussians. The peak flux ratio of the [O III] 4960 Å and 5008 Å components are fixed at the expected 1:3 ratio and the width and velocity offsets are set to be equal[1].

In 62 objects with very weak [O III] (mean EQ$W \sim 2$ Å) we found that the Gaussian model has a tendency to fit features to the noise. In some cases this can lead to large errors on the [O III] line properties. To avoid this problem, we instead fit a fixed [O III] template to the spectra, with the overall scaling of this template the only free-parameter in the fit. This template is generated by running our line-fitting routine on a median composite spectrum of the 268 quasars with reliable [O III] line measurements. The spectra used to construct the composite

---

1 For J003136+003421, a significantly better fit ($\Delta\chi^2_\nu \sim 25\%$) is obtained when the peak flux ratio constraint relaxed; the peak ratio of the best-fitting model is 1:2.13.
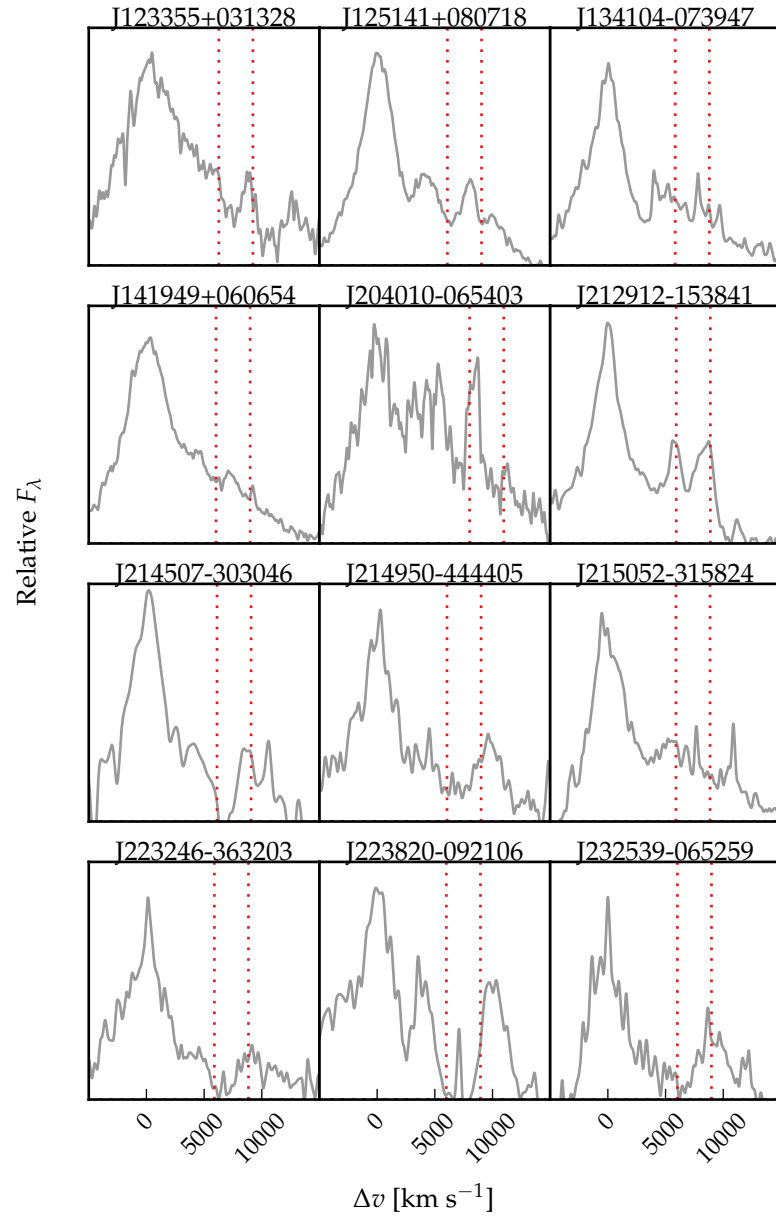
Table 4.3: Summary of models used to fit the [O III] emission, and the number of quasars each model is applied to.

| Model | Number |
|---|---|
| 2 Gaussians | 140 |
| 1 Gaussian | 128 |
| Template | 62 |

were first de-redshifted and continuum- and Fe II-subtracted. The models we use to fit [O III], and the numbers of quasars each model is applied to, are summarised in Table 4.3.

In Figure 4.2 we show example fits to 15 objects, chosen at random. The median reduced-$\chi^2$ value is 1.31 and, in general, there are no strong features observable in the spectrum minus model residuals.

### 4.3.4 H$\alpha$

There are 165 quasars in our sample with spectra covering the H$\alpha$ emission-line. In Section 4.3.8, we use the peak of the H$\alpha$ emission as one estimate of the quasar systemic redshift. In this section we describe how the H$\alpha$ emission was modelled.

The continuum emission is first modeled and subtracted using the procedure described in Section 3.3.2. We then test five different models with increasing degrees of freedom to model the H$\alpha$ emission. The models are summarised in Table 4.4. 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 narrow H$\alpha$ emission, and the narrow components of [N II]$\lambda\lambda 6548, 6584$ and [S II]$\lambda\lambda 6717, 6731$; (5) two broad Gaussians with different velocity centroids, and additional narrower Gaussians. If used, the width and 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.

In order to determine which model is selected for each spectrum, we use the following procedure. Each of the five models are fit to every spectrum and the reduced-$\chi^2$ recorded. Initially, the model with the smallest reduced-$\chi^2$ is selected. We then measure how the reduced-$\chi^2$ changes as the complexity of the model is decreased (i.e. considering the models in Table 4.4 in descending order). If it results in an increase in the reduced-$\chi^2$

Table 4.4: Summary of models used to fit the Hα emission, and the number of quasars each model is applied to.

| 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 |

which is less than 10 per cent relative to the best fitting model, then the simpler model is selected.

### 4.3.5 *Deriving emission-line properties from the best-fitting models*

All [O III] line properties are derived from the [O III]$\lambda$5008 emission, but, as described above, the kinematics of [O III]$\lambda$4960 are constrained to be identical in our fitting routine.

We do not attach any physical meaning to the individual Gaussian components used in the model. Decomposing the [O III] emission into a narrow component component at the systemic redshift and a lower-amplitude, blueshifted broad component is subject to large uncertainties and is highly dependent on the spectral S/N and resolution. Furthermore, there is no theoretical justification that the broad component should have a Gaussian profile.

We therefore choose to characterize the [O III] line profile using a number of non-parametric measures, which are commonly used in the literature (e.g. Zakamska and Greene, 2014; Zakamska et al., 2016). A normalised cumulative velocity distribution 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 calculated. These velocities are then adjusted so that the peak of the [O III] emission is at $0\,\mathrm{km\,s^{-1}}$.

The width of the emission-line can then be defined using either $w_{50}$ ($\equiv v_{75} - v_{25}$), $w_{80}$ ($\equiv v_{90} - v_{10}$) or $w_{90}$ ($\equiv v_{95} - v_5$). In terms of the FWHM, $w_{50} \simeq$ FWHM/1.746, $w_{80} \simeq$ FWHM/0.919, $w_{90} \simeq$ FWHM/0.716, assuming a Gaussian line profile. $w_{90}$ is relatively more sensitive to the wings of the line profile, whereas $w_{50}$ is relatively more sensitive to the core. We also define the relative asymmetry of the line as:
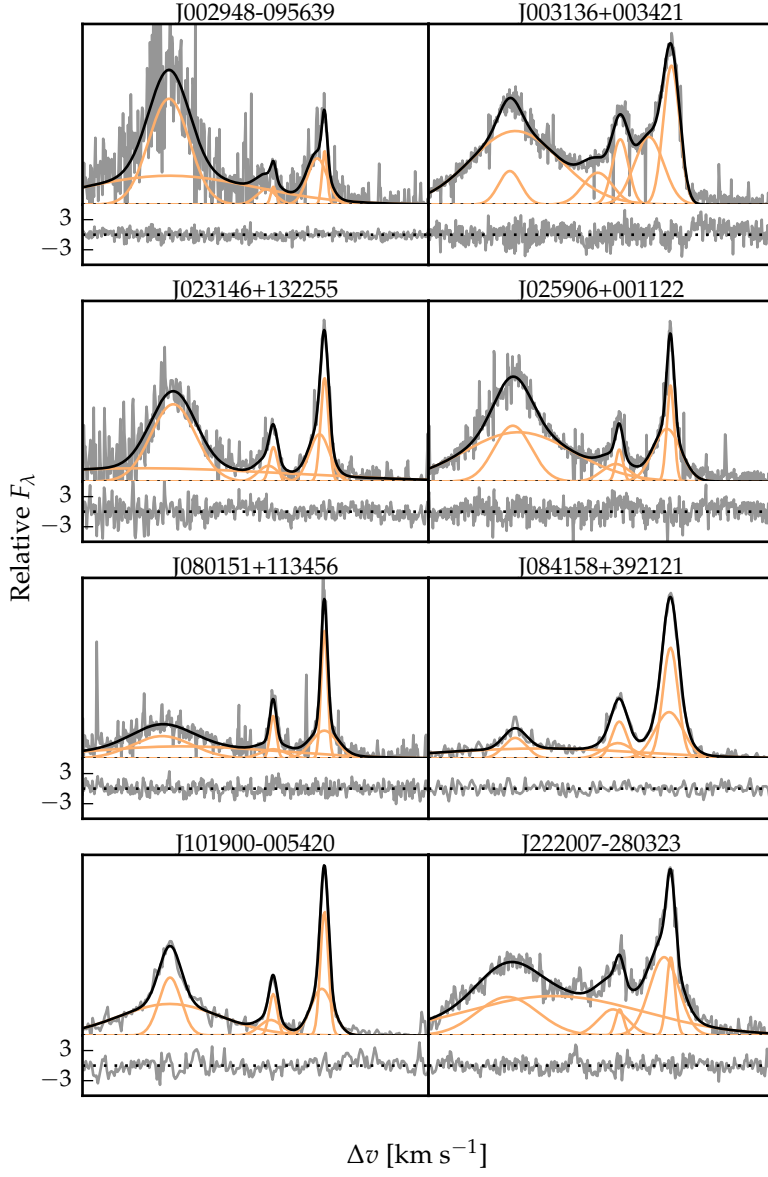
Figure 4.2: Model fits to the continuum- and Fe II-subtracted Hβ/[O III] emission in 15 quasars, chosen at random. The data is shown in grey, the best-fitting model in black, and the individual model components in orange. The peak of the [O III] emission is used to set the redshift, and Δv is the velocity shift from the rest-frame transition wavelength of Hβ. Below each spectrum we plot the data minus model residuals, scaled by the errors on the fluxes.

Table 4.5: The format of the table containing the emission-line properties from our parametric model fits.

| Column | Name | Units | Description |
|---|---|---|---|
| 1 | UID | | Catalogue name |
| 2 | OIII_V5 | $km\,s^{-1}$ | [O III] $v_5$ |
| 3 | OIII_V5_ERR | $km\,s^{-1}$ | Uncertainty in $v_5$ |
| 4 | OIII_V10 | $km\,s^{-1}$ | [O III] $v_{10}$ |
| 5 | OIII_V10_ERR | $km\,s^{-1}$ | Uncertainty in $v_{10}$ |
| 6 | OIII_V25 | $km\,s^{-1}$ | [O III] $v_{25}$ |
| 7 | OIII_V25_ERR | $km\,s^{-1}$ | Uncertainty in $v_{25}$ |
| 8 | OIII_V50 | $km\,s^{-1}$ | [O III] $v_{50}$ |
| 9 | OIII_V50_ERR | $km\,s^{-1}$ | Uncertainty in $v_{50}$ |
| 10 | OIII_V75 | $km\,s^{-1}$ | [O III] $v_{75}$ |
| 11 | OIII_V75_ERR | $km\,s^{-1}$ | Uncertainty in $v_{75}$ |
| 12 | OIII_V90 | $km\,s^{-1}$ | [O III] $v_{90}$ |
| 13 | OIII_V90_ERR | $km\,s^{-1}$ | Uncertainty in $v_{90}$ |
| 14 | OIII_V95 | $km\,s^{-1}$ | [O III] $v_{95}$ |
| 15 | OIII_V95_ERR | $km\,s^{-1}$ | Uncertainty in $v_{95}$ |
| 16 | z_OIII | | [O III] redshift |
| 17 | z_OIII_ERR | | Uncertainty in [O III] redshift |
| 18 | OIII_W50 | $km\,s^{-1}$ | [O III] $w_{50}$ |
| 19 | OIII_W50_ERR | $km\,s^{-1}$ | Uncertainty in [O III] $w_{50}$ |
| 20 | OIII_W80 | $km\,s^{-1}$ | [O III] $w_{80}$ |
| 21 | OIII_W80_ERR | $km\,s^{-1}$ | Uncertainty in [O III] $w_{50}$ |
| 22 | OIII_W90 | $km\,s^{-1}$ | [O III] $w_{90}$ |
| 23 | OIII_W90_ERR | $km\,s^{-1}$ | Uncertainty in [O III] $w_{50}$ |
| 24 | OIII_A | | [O III] asymmetry |
| 25 | OIII_A_ERR | | Uncertainty in [O III] asymmetry |
| 26 | OIII_EQW | Å | [O III] EQW |
| 27 | OIII_EQW_ERR | Å | Uncertainty in [O III] EQW |
| 28 | OIII_LUM | $erg\,s^{-1}$ | [O III] luminosity |
| 29 | OIII_LUM_ERR | $erg\,s^{-1}$ | Uncertainty in [O III] luminosity |
| 30 | EQW_FE_4434_4684 | Å | Fe II EQW |
| 31 | EQW_FE_4434_4684_ERR | Å | Uncertainty in Fe II EQW |
| 32 | HB_VPEAK | $km\,s^{-1}$ | Hβ peak velocity |
| 33 | HB_VPEAK_ERR | $km\,s^{-1}$ | Uncertainty in Hβ peak velocity |
| 34 | HA_VPEAK | $km\,s^{-1}$ | Hα peak velocity |
| 35 | HA_VPEAK_ERR | $km\,s^{-1}$ | Uncertainty in Hα peak velocity |
| 36 | HB_Z | | Hβ redshift |
| 37 | HB_Z_ERR | | Uncertainty in Hβ redshift |
| 38 | HA_Z | | Hα redshift |
| 39 | HA_Z_ERR | | Uncertainty in Hα redshift |
| 41 | OIII_FE_FLAG | | Bad Fe II subtraction |
| 42 | OIII_EXTREM_FLAG | | Extreme [O III] emission |

$$A = \frac{(v_{90} - v_{peak}) - (v_{peak} - v_{10})}{(v_{90} - v_{10})}. \qquad (4.1)$$

All of the derived parameters we have calculated are summarised in Table 4.5. The columns are as follows:

1 Unique ID: QSOXXX.

2-3 Systemic redshift measured at [O III] peak wavelength, and its error.

4-17 $v_5$, $v_{10}$, $v_{25}$, $v_{50}$, $v_{75}$, $v_{90}$ and $v_{95}$ velocity of [O III], relative to [O III] peak, and their errors, in $km\,s^{-1}$.

18-23 $w_{50}$ ($\equiv v_{75} - v_{25}$), $w_{80}$ ($\equiv v_{90} - v_{10}$) and $w_{90}$ ($\equiv v_{95} - v_5$) velocity width of [O III], and their errors, in $km\,s^{-1}$.

24-25 Dimensionless [O III] asymmetry $A$, and its error.

26-27 Rest-frame [O III] EQW, and its error, in Å.

28-29 1-$\sigma$ upper-limit on rest-frame [O III] EQW, in Å.

30-31 [O III] luminosity, and its error, in $erg\,s^{-1}$.

32-33 4434-4684 Å rest-frame Fe II EQW, and its error, in Å.

34-35 Velocity of H$\beta$ peak, relative to [O III] peak, in $km\,s^{-1}$, and its error.

36-37 Velocity of H$\alpha$ peak, relative to [O III] peak, in $km\,s^{-1}$, and its error.

38-38 Redshift of H$\beta$ peak, and its error.

40-41 Redshift of H$\alpha$ peak, and its error.

44 Fe II flag.

45 Extreme [O III] flag.

46-47 C IV $v_{50}$, relative to [O III] peak, in $km\,s^{-1}$, and its error.
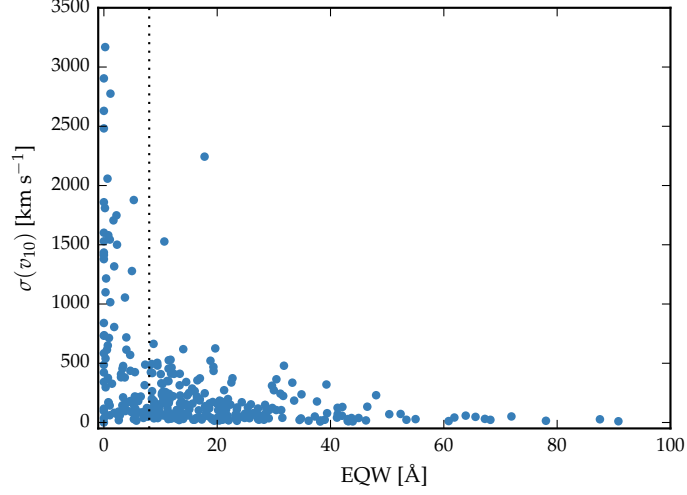
Figure 4.3: Uncertainty in $v_{10}$ as a function of the EQW, for [O III]. Uncertainties in $v_{10}$ are large to the left of the vertical line, at 8 Å. These objects are ignored in our subsequent analysis of the [O III] line shape.

### 4.3.6 *Deriving uncertainties on parameters*

Our method to estimate realistic uncertainties on emission-line properties derived from the best-fitting model is very similar to the one described in Section 3.3.6. Very briefly, random simulations of each spectrum are generated. Our fitting-procedure is run on each simulated spectrum, and the errors on the line parameters are estimated by looking at the distribution of values from the ensemble of simulations. In a slight modification of the procedure in Section 3.3.6, the error is defined as half the the 68 (84 - 16) percentile spread in the parameter values.

### 4.3.7 *Low EQW [O III]*

In Figure 4.3 we show how the uncertainty in $v_{10}$ depends on the EQW. As the strength of [O III] decreases, the average uncertainty in $v_{10}$ increases. When the [O III] $EQW > 80$ Å, the mean uncertainty in $v_{10}$ is $50\,\mathrm{km\,s^{-1}}$; this increases to $450\,\mathrm{km\,s^{-1}}$ when $10 < EQW < 20$ Å. As the EQW drops below 8 Å, typical uncertainties in $v_{10}$ become very large (exceeding $1000\,\mathrm{km\,s^{-1}}$ in many objects). Clearly, the emission-line is too weak for properties - in this case $v_{10}$ - to be reliably measured in many of these objects. Therefore, when the [O III] line properties (e.g. velocity-width, centroid) are analysed in later sections, these ob-

jects with EQW $< 8$ Å will be excluded. This leaves 226 quasars in the sample.

### 4.3.8 *Reliability of systemic redshift estimates*

In this section, we compare systemic redshift estimates based on [O III], Hβ and Hα. The wavelength of each of these lines is measured at the peak of the emission and this measurement is made using the best-fitting parametric model. In the case of the Balmer lines, this model includes both broad and (if present) narrow emission features.

We compare systemic redshift estimates based on [O III] and Hβ (Figure 4.4a), [O III] and Hα (Figure 4.4b) and Hβ and Hα (Figure 4.4c). [O III], Hβ and Hα measurements are available for 226, 418 and 226 objects respectively. We exclude [O III], Hβ and Hα measurements when the uncertainties on the peak velocities exceed 200, 300 and 200 km s$^{-1}$ respectively. This excludes 4, 6 and 12 per cent of the [O III], Hβ and Hα measurements respectively. We also exclude [O III] measurements from 16 objects with very broad, blueshifted [O III] emission that is strongly blended with the red wing of Hβ (these objects are discussed in Section 4.6) because these redshifts are almost certainly strongly biased. After these cuts, there are 182, 85 and 162 objects being compared in samples (a), (b) and (c) respectively.

We generate probability density functions using a Gaussian kernel density estimator. The bandwidth, which is optimised using leave-one-out cross-validation, is 170, 120 and 140 km s$^{-1}$ for samples (a), (b) and (c) respectively. The systematic offset between the Hα and Hβ estimates is consistent with being zero, and the scatter is 230 km s$^{-1}$. The scatter in these distributions is consistent with previous studies of redshift uncertainties from broad emission-lines (e.g. Shen et al., 2016). The [O III] redshifts appear to be systematically offset in comparison to both Hα and Hβ, in the sense that [O III] is blueshifted. This effect is strongest when [O III] is compared to Hβ, in which case [O III] is shifted by $\sim 100$ km s$^{-1}$ to the blue.

Hewett and Wild, (2010) found that [O III] was blueshifted by $\sim 45$ km s$^{-1}$ relative to a rest-frame defined using photospheric Ca IIλλ3935,3970 absorption in the host galaxies of $z < 0.4$ SDSS AGN. They also noted, as we find here, that [O III] is increasingly blue-asymmetric at higher luminosities. Therefore, our finding is consistent with Hewett and Wild, (2010), when the
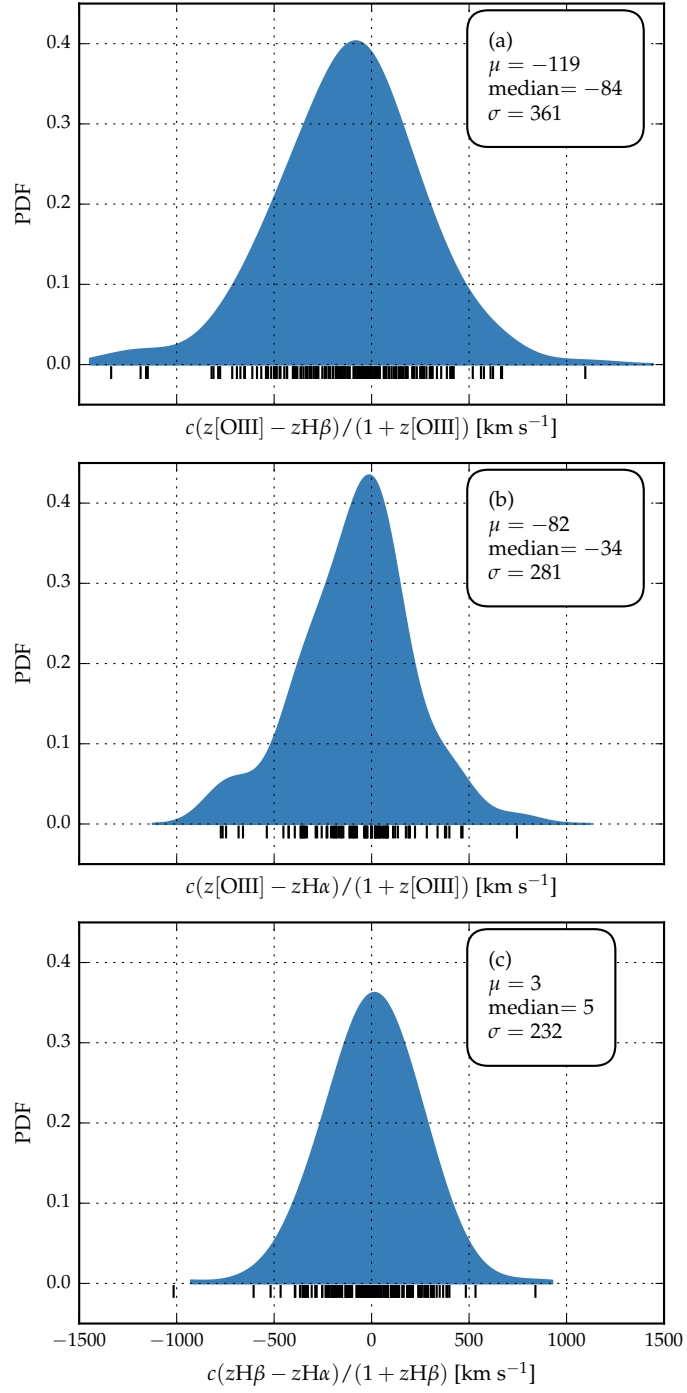
Figure 4.4: Comparison of systemic redshift estimates using [O III], broad Hβ and broad Hα. The probability density functions are generated using a Gaussian kernel density estimator with a $\simeq 150\,\mathrm{km\,s^{-1}}$ kernel width. The short black lines show the locations of the individual points.

very different luminosities of the two samples are accounted for.

## 4.4 RESULTS

In our sample of 354 quasars, there is a huge diversity in [O III] emission properties (Figure 4.2). In Figure 4.5, we present a subset of the measurements we have made of the [O III] line.

The strength of [O III] depends on the covering factor of NLR gas, its density and ionisation parameter. The [O III] EQW follows an approximately log-normal distribution, peaking at 17 Å. In 10 per cent of our sample [O III] is very weak, with EQW < 1 Å. The average [O III] strength is consistent with earlier studies on smaller samples (e.g. Sulentic et al., 2004; Netzer et al., 2004; Shen, 2016).

The mean and standard deviation of the line width (characterized by $w_{80}$) is $1535 \pm 562\,\mathrm{km\,s^{-1}}$, with a median of 1529, minimum of 206 and maximum of 3214. This is consistent with recent near-infrared spectroscopy of $z > 1.5$ quasars which often report velocity widths $\gtrsim 1000\,\mathrm{km\,s^{-1}}$ (e.g. Netzer et al., 2004; Kim et al., 2013; Brusa et al., 2015; Shen, 2016). For gas discs rotating in the potential of the massive galaxies line widths do not exceed $w_{80} \simeq 600\,\mathrm{km\,s^{-1}}$ (Liu et al., 2013). Therefore the [O III] gas cannot be in dynamical equilibrium with the host-galaxy. [O III] emission is suppressed by collisional de-excitation in higher-density environments, and so the large velocity widths cannot be due to Doppler broadening in the BLR.

The [O III] asymmetry is shown in Figure 4.5c. In 40 per cent of the sample [O III] is fit with a single Gaussian. The asymmetry is zero in this model and so these objects are excluded. For the objects fit with two Gaussians, [O III] is blue-asymmetric in 90 per cent. This indicates that there is an outflow component in the [O III]-emitting gas. We also find a weak anti-correlation between $w_{80}$ and the asymmetry (Spearman correlation coefficient: $-0.30$, p-value:L $4e - 4$), in the sense that the broaded lines tend to be more blue-asymmetric.

### 4.4.1 *Luminosity/redshift-evolution of [O III] properties*

We extend the dynamic range of our samples in terms of both luminosity and redshift by supplementing our sample with quasars presented by Mullaney et al., (2013) and Harrison et al., (2016). The Mullaney et al., (2013) catalogue con-
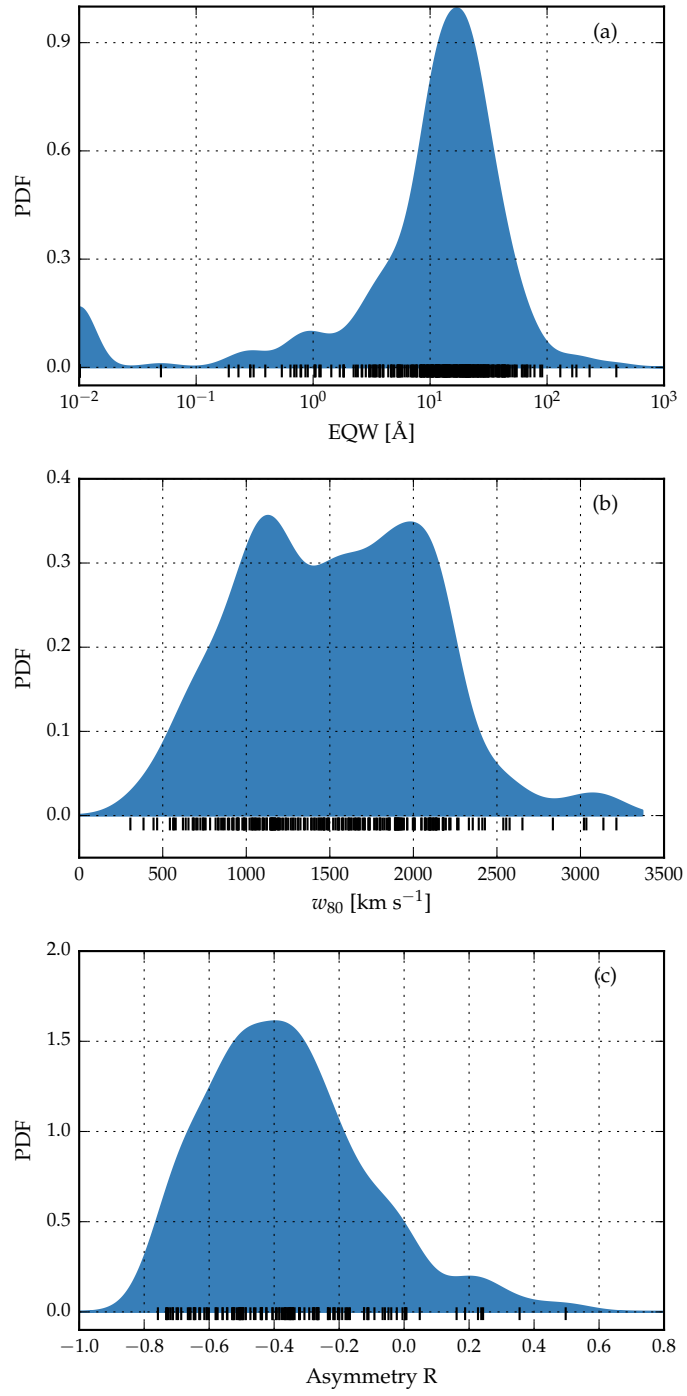
Figure 4.5: Probability density distributions of the [O III] parameters EQW (a), $w_{80}$ (b) and asymmetry R (c). The $1200\,\mathrm{km\,s^{-1}}$ upper limit on the velocity width of the Gaussian functions used to model [O III] is responsible for the peak at $1200\,\mathrm{km\,s^{-1}}$ in (b).

tains [O III] line measurements for $\sim 25\,000$ optically-selected AGN with SDSS spectra at $z < 0.4$. Mullaney et al., (2013) fit [O III] with one or two Gaussians, and then used similar non-parametric measures to the ones we adopt. We select only the Type I AGN from the Mullaney et al., (2013) catalogue. The Harrison et al., (2016) sample contains 40 X-ray selected quasars ($L_{2-10\text{keV}} \simeq 10^{43-44}$ erg s$^{-1}$) at intermediate redshifts ($1.1 < z < 1.7$) observed with the KMOS integral field unit spectograph on the VLT. We also use the SDSS DR7 quasar catalogue, with properties derived by Shen et al., (2011). [O III] is visible in SDSS spectra up to redshifts $z = 0.84$. There are $20\,663$ quasars in the Shen et al., (2011) catalogue with [O III] $EQW > 0$ Å.

### 4.4.1.1 *Equivalent width*

In Figure 4.6 we show the [O III] EQW as a function of the quasar bolometric luminosity. Bolometric luminosity is estimated from the monochromatic continuum luminosity at 5100 Å, using the correction factor given by Richards et al., (2006a). For comparison, we also show the sample from Shen et al., (2011).

Many authors have reported the [O III] EQW to decrease with quasar luminosity (e.g. Brotherton, 1996; Sulentic et al., 2004; Baskin and Laor, 2005b). The origin of this correlation - which is known as the Baldwin effect (e.g. Baldwin, 1977; Brotherton, 1996; Zhang et al., 2011; Stern and Laor, 2012) - has not been demonstrated conclusively. One interpretation is that the size of the NLR increases when the quasar luminosity increases, because more ionising photons are available. However, beyond a certain radius there will be no more gas left to ionise, and the size of the NLR will plataeu.

We do not observe a Baldwin effect in our sample, despite the luminosity range spanning three dex. This is seen more clearly if we show the distribution of EQWs in the SDSS sample ($\log L_{\text{Bol}} \sim 45.5$ erg s$^{-1}$) and the sample presented in this paper ($\log L_{\text{Bol}} \sim 47.3$ erg s$^{-1}$). Only objects where [O III] is detected with $EQW > 1$ Å are included. The two distributions are essentially identical, and peak at 13 Å.

The fraction of objects with very weak [O III] ($EQW < 1$ Å) is 10 per cent in our sample, compared to one per cent of the SDSS sample. Therefore, the fraction of objects with very weak [O III] is an order of magnitude larger in the more luminous
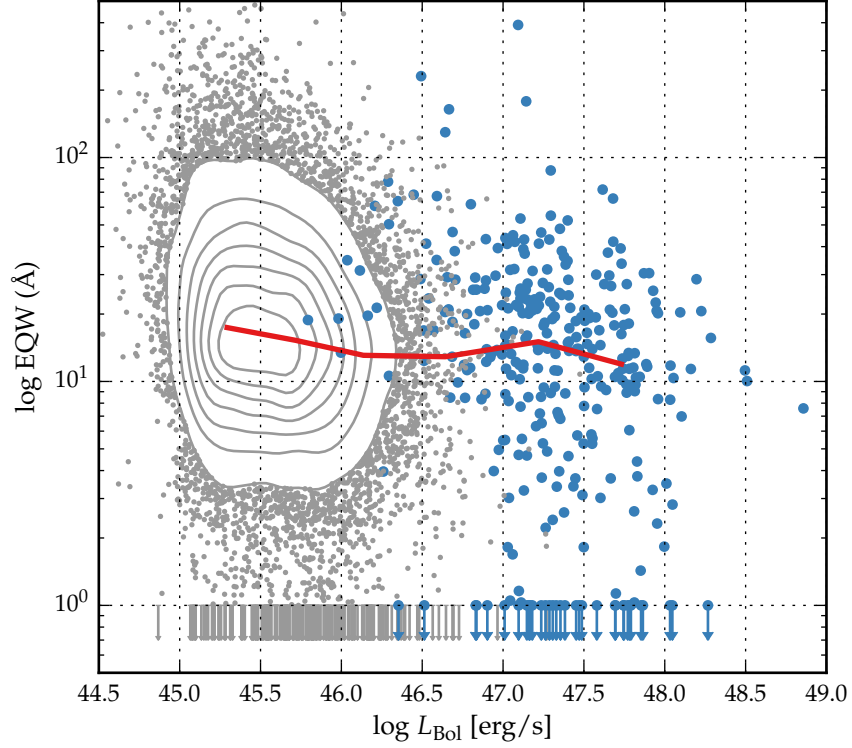
Figure 4.6: The [O III] EQW as a function of the quasar bolometric luminosity for the sample presented in this chapter (blue circles) and the low-z SDSS sample (grey points and contours). An upper limit at $EQW = 1\,\text{Å}$ indicates points with $EQW < 1\,\text{Å}$. The red line shows the median [O III] EQW in luminosity bins centred on 45.3, 45.7, 46.1, 46.6, 47.2 and 47.7 $\text{erg s}^{-1}$, considering only objects with $EQW > 1\,\text{Å}$.

sample. If we instead define 'weak' [O III] emission as having an $EQW < 5\,\text{Å}$, then 22 per cent of our high-luminosity sample is classified as such, compared to 8 per cent of the SDSS sample.

Netzer et al., (2004), comparing the [O III] properties in a much smaller sample over a comparable luminosity range, reached a similar conclusion. The conclusion reached by Netzer et al., (2004) was that the size of the NLR scaled with the square root of the luminosity of the source of ionising photons in low luminosity AGN, in line with theoretical predictions (e.g. Netzer, 1990). However, extrapolating this relationship to high luminosity quasars leads to the prediction of enormous NLRs with galactic dimensions. If these NLRs had properties similar to the ones in local Seyferts, they would quickly escape the system and disappear. Therefore, the prediction is that no
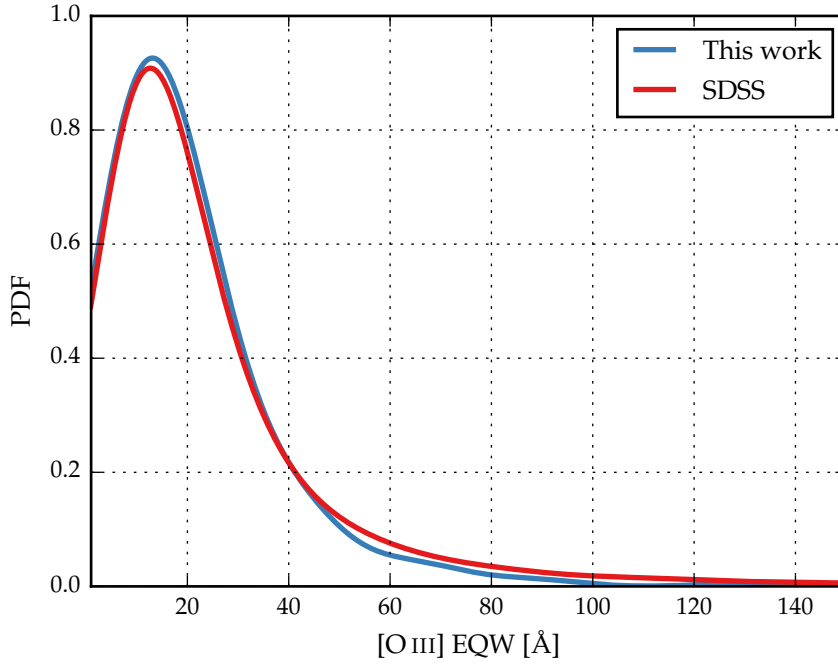
Figure 4.7: [O III] EQW distribution at EQW > 1 Å PDF generated
using a Gaussian KDE with a 8 Å bandwidth (optimized
via leave-one-out cross-validation). The two distributions
are essentially identical, and peak at 13 Å. However, 10
per cent of the sample in this work have EQW < 1 Å, com-
pared to just 1 per cent of the SDSS sample.

NLR emission should be observed in high-luminosity quasars,
which is the case for $\sim$ 10 per cent of our sample. This means
that when strong [O III] is detected, it must be from denser gas.
Netzer et al., (2004) suggest that this high density gas could
be produced in the kilo-parsec scale nuclear regions by violent
star formation. We find that that virtually all objects which have
large C IV blueshifts have very weak [O III] emission.

*Does this
support/contradict
Netzer picture? See
email from Paul.*

### 4.4.2  Velocity width

In Figure 4.8 we show the [O III] velocity width as a function
of the quasar optical luminosity. A strong correlation exists. On
the other hand, we do not find any significant correlation be-
tween the [O III] velocity width and the quasar redshift. The
lack of any evolution in typical [O III] properties between $z = 0$
and $z = 1.5$ has previously been reported (e.g. Harrison et al.,
2016); our sample demonstates that the [O III] properties do not
evolve from $z = 1.5$ all the way to $z = 4$.

Figure 4.8: [O III] velocity width $w_{80}$ as a function of quasar bolometric luminosity. Objects with extreme [O III] profiles (Section 4.6) are shown in red.

## 4.5 EIGENVECTOR 1 CORRELATIONS

### 4.5.1 *EV1 trends exist in high-redshift quasars*

The FWHM of the broad Hβ emission-line and the strength of [O III] and the relative strengths of optical Fe II and Hβ have been identified as the features responsible for the largest variance in the spectra of AGN. These parameters form part of EV1, the first eigenvector in a PCA which originated from the work of Boroson and Green, (1992). The underlying driver behind EV1 is thought to be the Eddington ratio (e.g. Sulentic et al., 2000; Shen and Ho, 2014).

In Figure 4.9 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 and 4684 Å. There are 395 objects in our sample with spectra covering Hβ . 303 with partial (> 150 Å) coverage of the 4434-4684 region to constrain the Fe II emission. 283 after removing bad fit Fe II. 230 excluding those with more than 25 per cent fractional error in Hβ FHWM or Fe II strength.
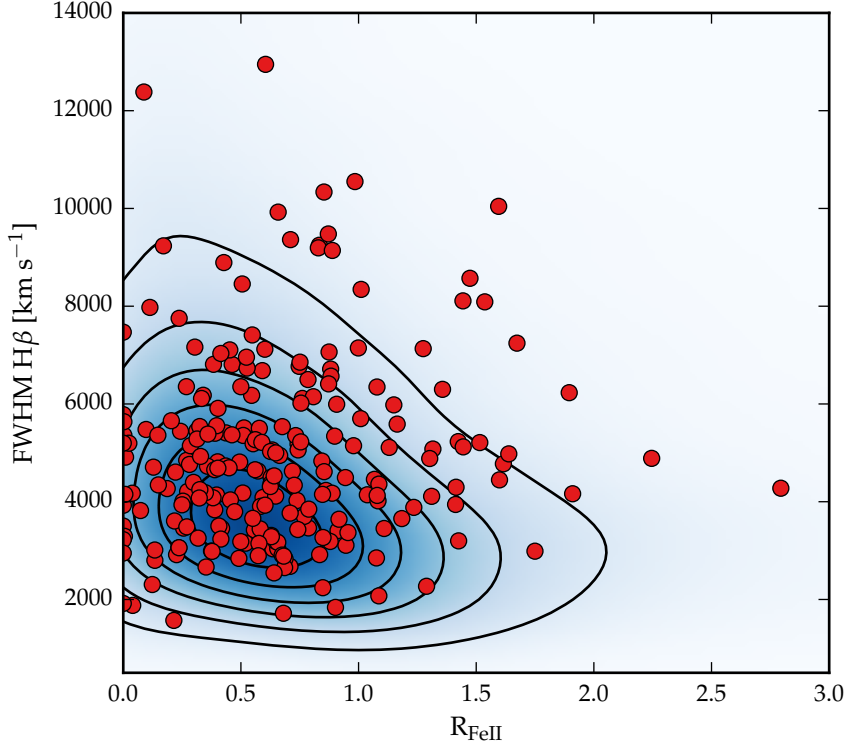
Figure 4.9: EV1 parameter space. The contours and shading show low-redshift, low-luminosity SDSS AGN (with measurements taken from Shen et al. 2011) and the red circles show the high-redshift, high-luminosity objects presented in this chapter.

The low-redshift SDSS sample, with parameters taken from Shen et al., (2011), is also shown in Figure 4.9. In our sample, these parameters follow very similar correlations to what is observed at low-redshift. In particular, we observe a strong anti-correlation between the [O III] and Fe II EQW. The Hβ FWHM are displaced to higher values, which is consistent with the high-redshift, high-luminosity sample having larger BH masses. Thus, we confirm earlier results using much smaller samples that suggest that the same EV1 correlations exist in high-redshift quasars (e.g. Netzer et al., 2004; Sulentic et al., 2004; Sulentic et al., 2006; Runnoe et al., 2013; Shen, 2016).

### 4.5.2 Connecting EV1 at low and high redshifts

The C IV blueshift and EQW is a diagnostic that similarly spans the diversity of broad emission-line properties in high red-
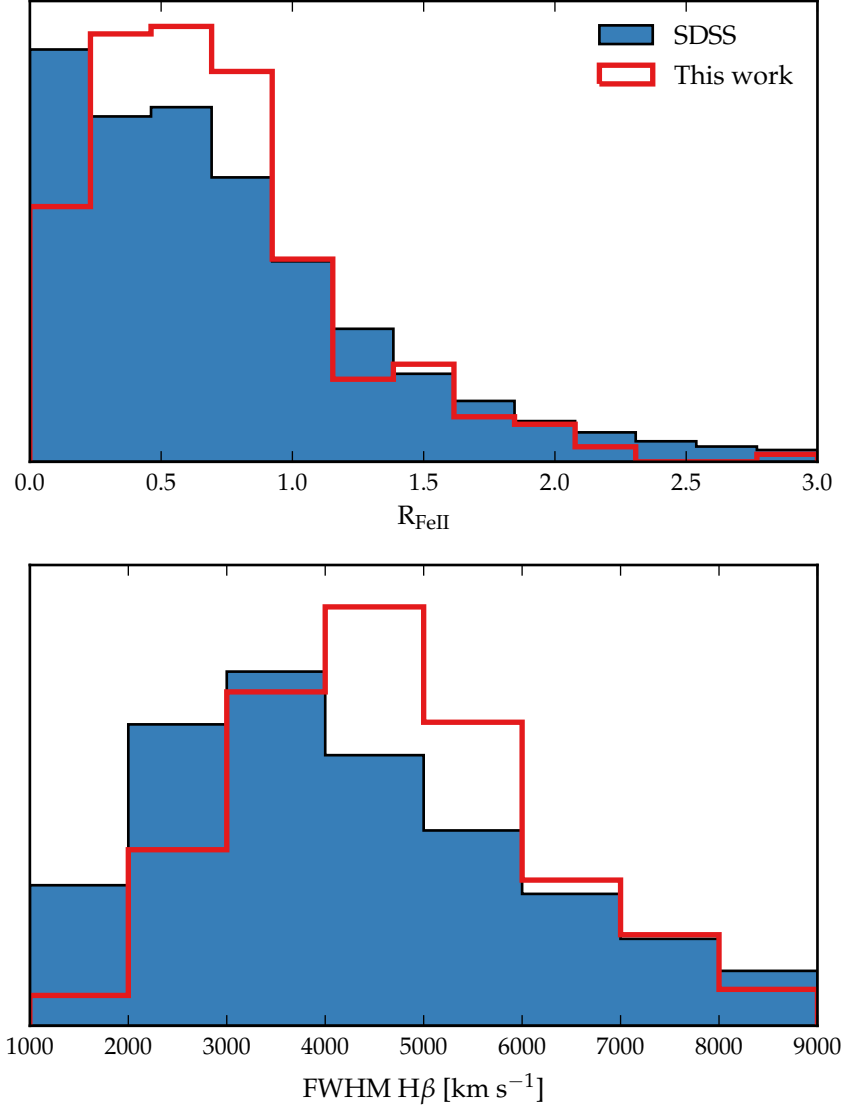
Figure 4.10: EV1 parameters for the high-luminosity quasar sample presented in this work and the low-luminosity SDSS sample. Hβ FWHM are shifted to higher values in the high-luminosity sample, which is consistent with these quasars having large BH masses. The Fe II distributions are similar.

shift quasars (Sulentic et al., 2007; Richards et al., 2011). The similarity of the C IV EQW-blueshift parameter space at high redshift to EV1 parameter space at low redshift suggests that these trends are connected. Because we have optical and near-infrared spectra for XX quasars in our sample, we are able to test how the low-redshift EV1 parameter space maps to the high-redshift C IV parameter space. In Figure 4.11 we show how

the EV1 parameters change as a function of position in the C IV EQW-blueshift parameter space.

Two hundred and thirteen objects are shown in Figure 4.11. Objects for which the C IV line properties could not be measured reliably (see Section 3.3.5) have been removed. We consider only objects for which the C IV EQW exceeds 15 Å. The C IV blueshift is measured relative to the redshift determined from the peak of [O III], Hβ or Hα. We also show the C IV line parameters of 32 157 SDSS DR7 quasars at redshifts $1.6 < z < 3.0$. The derivation of the C IV emission properties of these objects is described in Section XX.

Most of the diversity in C IV properties is correlated with the [O III] EQW. This is seen more clearly in Figure 4.12, in which we plot the [O III] EQW as a function of the C IV blueshift. The correlation see in Figure 4.12 is very strong. The mean [O III] EQW is 40 Å amongst the population of quasars with C IV blueshifts $< 500 \, \mathrm{km \, s^{-1}}$, compared to 5 Å for quasars with C IV blueshifts $> 2000 \, \mathrm{km \, s^{-1}}$. In other words, the NLR emission appears to be missing in quasars with large C IV blueshifts. One possibility is that the NLR gas is swept away on relatively short timescales by quasar outflows; this possibility is explored in more detail in Section XX.

On the other hand, the C IV blueshift and EQW cannot be used to predict the Hβ FWHM. This is consistent with 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.

## 4.6 EXTREME [O III] EMITTERS

Figure 4.13 shows the spectra of 18 objects which we visually identified as having broad, blueshifted [O III] emission which is heavily blended with the red wing of Hβ. Because the emission is so heavily-blended, it is difficult to determine unambiguously what combination of Hβ , [O III] and Fe II is responsible for the unusual plateau-like emission observed in these objects. Therefore, uncertainties on the [O III] emission properties are generally high in these objects.

In Figure 4.8 we show that the luminosities of all of these objects are larger than the sample median. The mean luminosity of the quasars with extreme [O III] emission is $10^{47.9} \, \mathrm{erg \, s^{-1}}$, compared to $10^{47.2} \, \mathrm{erg \, s^{-1}}$ for the rest of the sample. Figure 4.8 demonstrates that these quasars occupy a unique region of
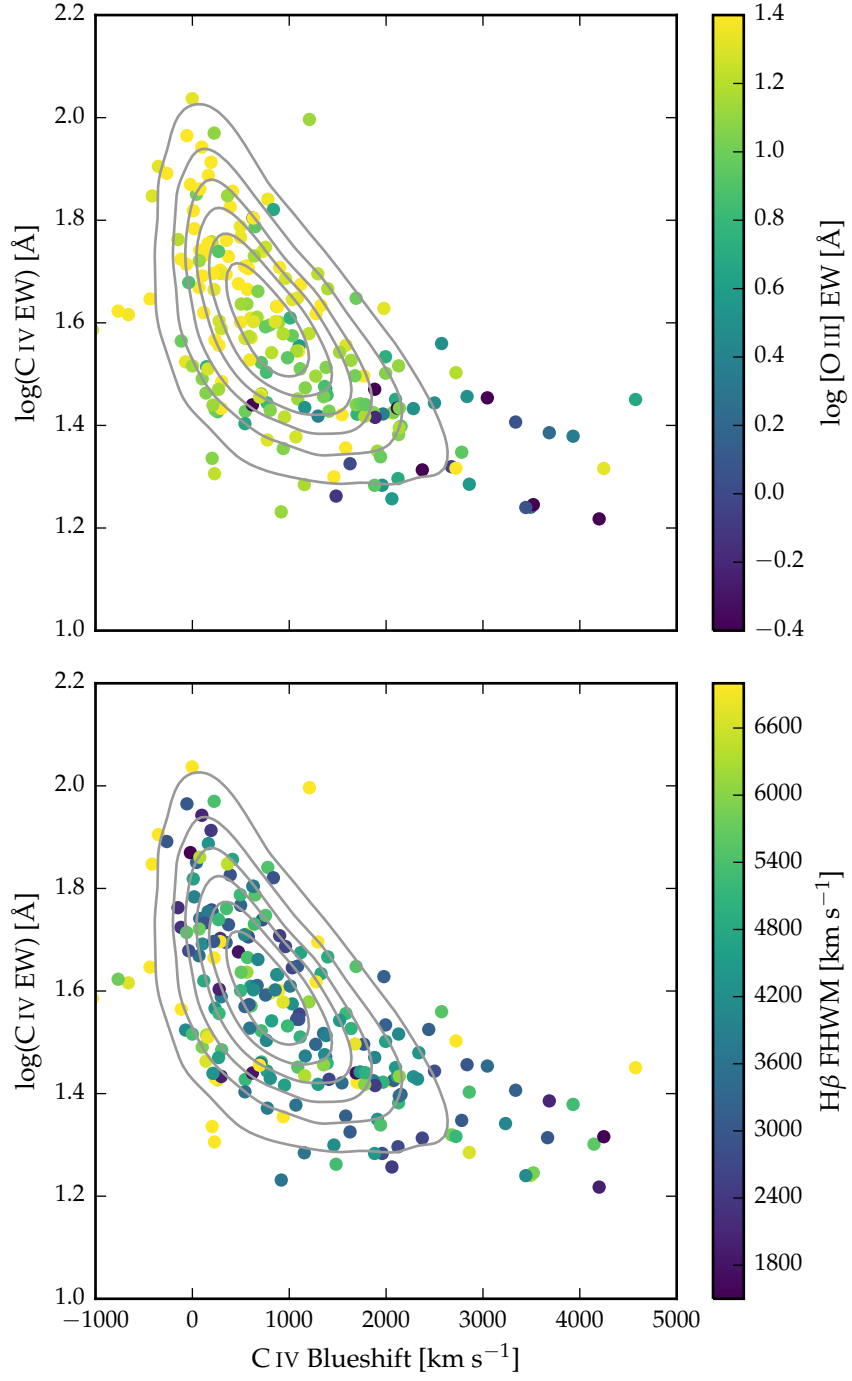
Figure 4.11: The high-redshift EV1 parameter space of C IV blueshift and EQW. Our sample is shown with points, and quasars from the full SDSS catalogue are shown with grey contours. The [O III] EQW varies systematically with position in the C IV blueshift-EQW parameter space (a) but the Hβ FWHM shows significantly less systematic variation (b).
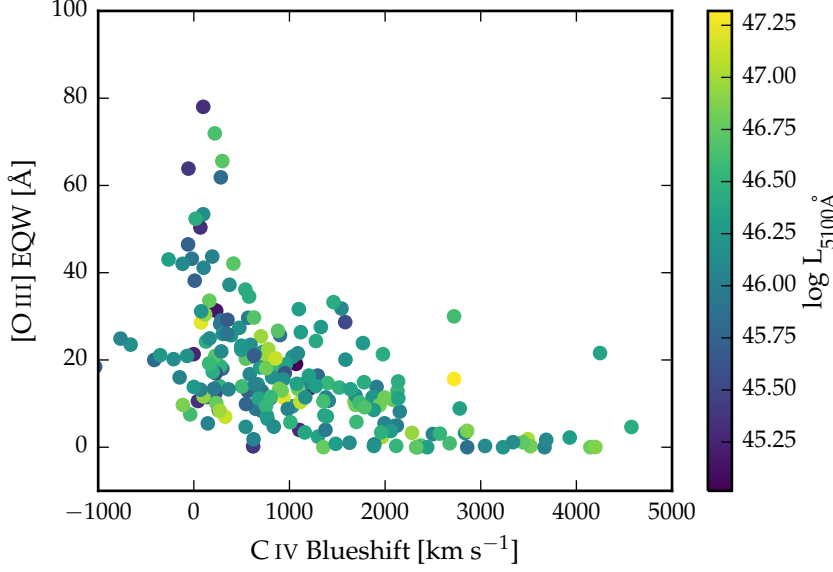
Figure 4.12: [O III] EQW as a function of the C IV blueshift. The [O III] EQW is strongly anti-correlated with the C IV blueshift. On the other hand, no strong luminosity-dependent trends (indicated by the colours of the points) is evident.

the $w_{80}$-$L_{Bol}$ parameter space: $w_{80} \gtrsim 1500\,\mathrm{km\,s^{-1}}$ and $L_{Bol} \gtrsim 10^{47.5}\,\mathrm{erg\,s^{-1}}$.

A similar [O III] emission was also observed in J1201+1206 in a sample of five quasars at redshifts $2.3 \lesssim z \lesssim 3.5$ with luminosities $10^{47.5} < L_{Bol} < 10^{48}\,\mathrm{erg\,s^{-1}}$ observed by Bischetti et al., (2016). These [O III] emission-lines are also somewhat similar to the lines observed in a sample of four extremely dust-reddened quasars at $z \sim 2$ recently identified by Zakamska et al., (2016). The four Zakamska et al., (2016) quasars have 5 μm luminosities of $\sim 10^{47}\,\mathrm{erg\,s^{-1}}$, which is comparable to the 5 μm luminosities of the brightest objects in our sample. However, the [O III] velocity widths of the Zakamska et al., (2016) objects are extreme in relation to our sample ($w_{80} \simeq 3500 - 5500\,\mathrm{km\,s^{-1}}$). The extreme nature of the [O III] emission in these objects led Zakamska et al., (2016) to propose that these objects are being observed transitioning from a dust-obscured, star-burst phase to a luminous, blue quasar (e.g. Sanders et al., 1988). 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. This suggests that these extreme outflows signatures are only associated with rare/short-lived
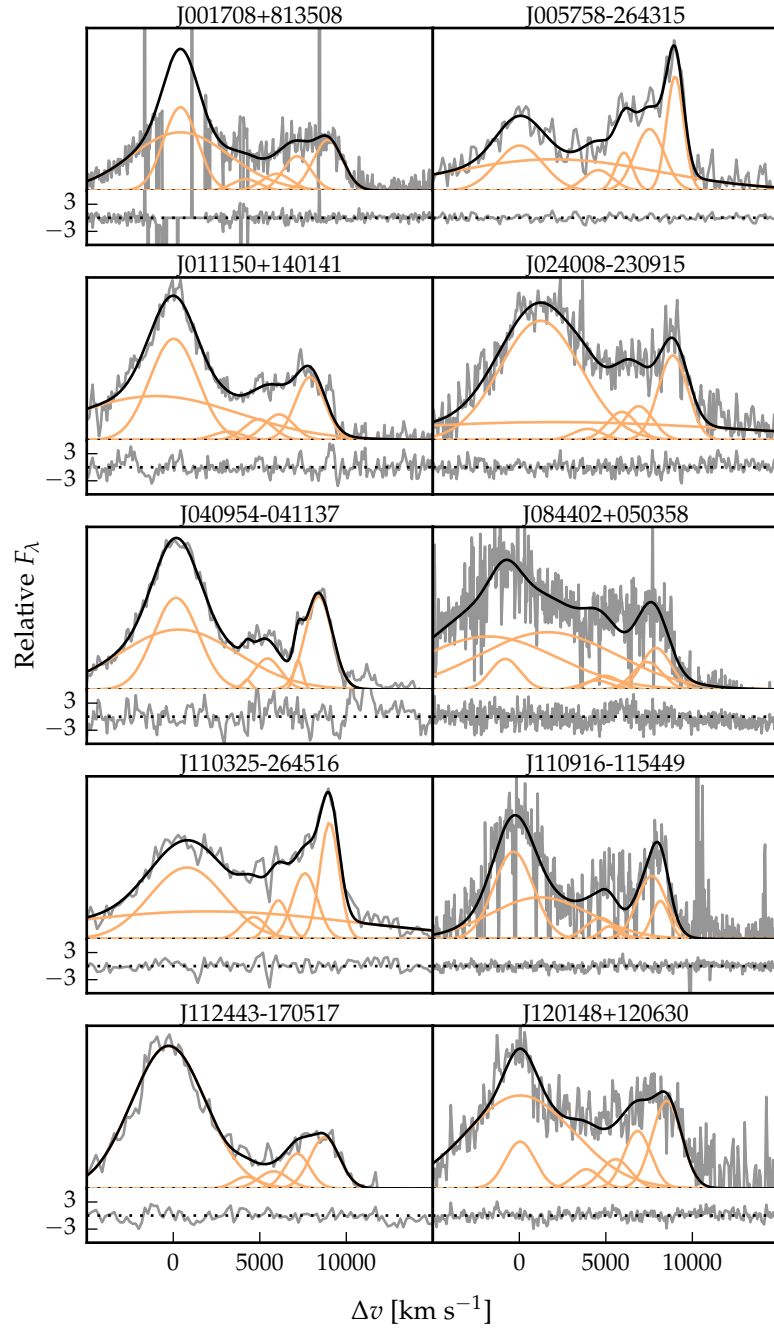
Figure 4.13: Model fits to the continuum- and Fe II-subtracted Hβ/[O III] emission in 18 quasars with extreme [O III] emission profiles. The data is shown in grey, the best-fitting model in black, and the individual model components in orange. The peak of the [O III] emission is used to set the redshift, and Δv is the velocity shift from the rest-frame transition wavelength of Hβ. Below each spectrum we plot the data minus model residuals, scaled by the errors on the fluxes.
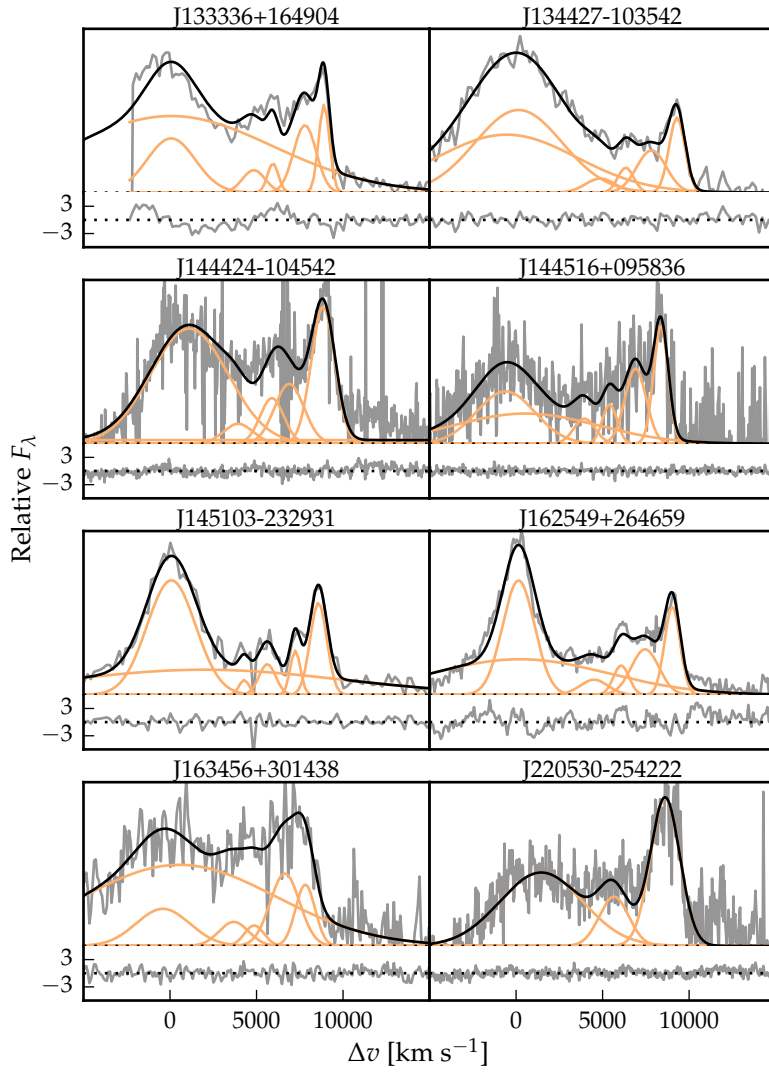
Figure 4.13: Continued.

phases in the AGN cycle represented by the reddened quasars and are not ubiquitous in the luminous AGN population. The early phase of feedback is likely to highly obscured by dusty inflowing material (e.g. Haas et al., 2003). Quasar-driven outflows will eventually blow away the obscuring dust, revealing a luminous, relatively un-obscured quasars. The fact that we do not see such extreme outflow signatures in our sample, where the mean dust reddening is just 0.03, appears to support this idea.

### 4.6.1 *Connections between [O III] and C IV outflows*

In Chapter 3 we found that C IV is blueshifted realtive to the systemic redshift in a large majority of luminous, high-redshift quasars. In a sub-set, the blueshifting of C IV can reach many thousands of $km\,s^{-1}$. This is strong evidence that quasars are capable of driving fast outflows, at least over the sub-parsec scale of the BLR. In the present chapter, we have found that outflows in the NLR, as indicated by broad velocity-widths and asymmetries in [O III], are similarly ubiquitous in the luminous quasar population.

In Section 4.5.2, we found that the [O III] EQW has a very strong dependence on the C IV blueshift. The mean [O III] EQW is 40 Å amongst the population of quasars with C IV blueshifts $< 500\,km\,s^{-1}$, compared to 5 Å for quasars with C IV blueshifts $> 2000\,km\,s^{-1}$. This establishes a connection between gas in the broad and narrow line regions.

In Figure 4.14 we show the the [O III] blueshift as a function of the C IV blueshift, for the objects where [O III] is detected with EQW$> 8$ Å. The [O III] blueshift is defined as $v_{10}([O\,III])$ - $v_{peak}([O\,III])$ whereas the C IV blueshift is defined as $v_{50}(C\,IV)$ - $v_{peak}([O\,III])$ with the C IV line measurements taken from Chapter 3. We do not show objects for which the errors on the [O III] and C IV blueshifts exceed 250 or 125 $km\,s^{-1}$ respectively. These objects, shown in the top two panels of Figure 4.14, have a similar dyamic range to the main sample, meaning our results should not be biased by their exclusion. We also remove the objects with extreme [O III] emission, because the systemic redshift determined from the peak of the [O III] emission is strongly biased in these objects.

Because of the strong anti-correlation between the C IV blueshift and the [O III] EQW, removing objects with weak [O III] eliminates most of the objects with large ($> 2000\,km\,s^{-1}$)
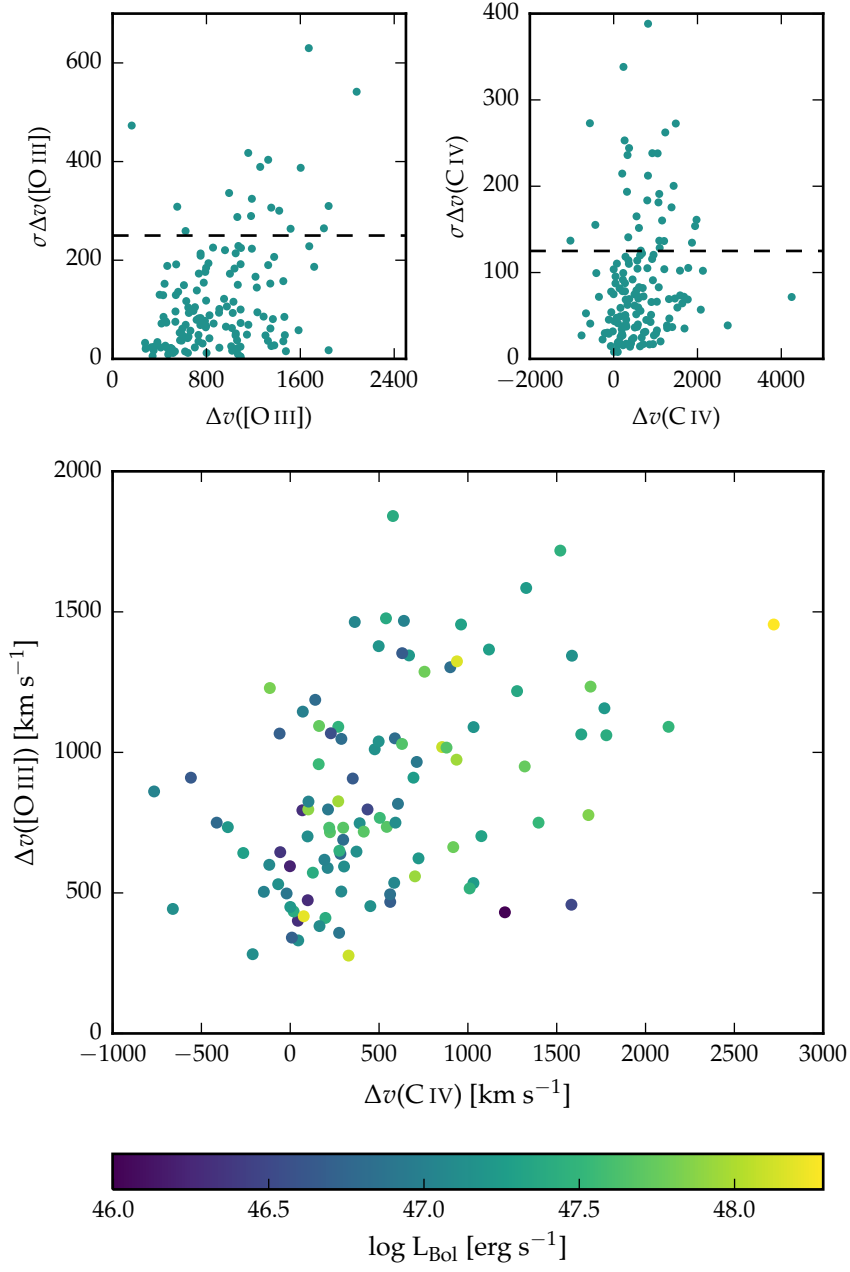
Figure 4.14: The relation between the blueshifts of C IV and [O III].

CIV blueshifts. Nevertheless, [OIII] appears to be more blueshifted in quasars with large CIV blueshifts. Although the scatter is large, the correlation appears to be significant (Spearman correlation coefficient: 0.46, p-value: 6e-7). This suggests a direct connection between the gas kinematics in the broad and narrow line regions.

We considered a number of of alternative approaches to parametrising both the [OIII] line shape and the systemic redshift. As expected, very similar trends are observed when the [OIII] line shape is parametrised using $v_{25} - v_{peak}$, $v_{50} - v_{peak}$, $w_{80} = v_{90} - v_{10}$, or the relative asymmetry R. The same trend is also observed when the systemic redshift is defined using the peak of the Hβ emission.

The blueshifting of CIV is known to correlate with luminosity (Richards et al., 2011). In [OIII], 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 CIV and [OIII] 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 4.14.

## 4.7 DISCUSSION

### 4.7.1 *Static NLR is removed by outflows*

Is the AGN NLR absent in objects where outflows have reached kilo-parsec scales, sweeping up the low-density material responsible for the [OIII]-emission? If the BLR outflows can escape, they are very fast and wouldn't need long to clear out the NLR gas. Estimate a time-scale for how long the NLR would take to be cleared given typical size of galaxy and velocity of outflow.

READ ZAKAMSKA DISCUSSION

It has been known for some time that the [OIII] EQW is anti-correlated with the strength of optical FeII, and this trend is thought to be driven by the Eddington ratio. Shen and Ho, (2014) showed that the amplitude of the core [OIII] emission decreases faster than the wing component as the Eddington ratio increases. Therefore, the [OIII] emission is weaker and more blueshifted in high accretion rate quasars. In Chapter 3 we found that all quasars with strong BLR outflows have high Eddington ratios. In this Section, we show that the CIV and

[O III] blueshifts are directly linked. This suggests a direct connection between the gas kinematics in the broad and narrow line regions. But we need IFU to find out where this gas is (See text in research proposal).

## 4.8 INDEPENDENT COMPONENT ANALYSIS

Take out any results / discussion from this section and move to Gaussians. I don't think I learn anything with the ICA components I don't already know.

ICA better with S/N - can give some examples from PCA work (ask Paul for references). If the [O III] emission is confined to a relatively small number of components can then recover emission. Can show problem with $w_{80}$ as an example. Encouraging, but we do need a better training set (training set doesn't span the diversity of properties we see in the high-luminosity sample). The content of the ICA "test" is fine but I would suggest setting up the experiment somewhat differently and orienting the reader at the start. For example I would mention a) the likely lack of sensitivity to S/N [cf standard approach], b) equating components with physical properties of interest and thus potential effectiveness of component weights (e.g. what is to come in Fig. 1.16). Would also though put in the information about using the SDSS objects to generate the component set and how one probably expects the high-luminosity sample to include spectral diversity not present in the SDSS-sample [and hence in the ICA components]. You can, as you have drafted, then say how the analysis could be improved - better training set, removing the component cross-talk (I can tell you how this can be achieved),...

In this section, we consider an alternative approach to the analysis presented in the bulk of this chapter. We use an independent component analysis (ICA) to separate the spectrum into a linear combination of statistically independent sub-components. Each individual spectrum can then be reconstructed with a linear combination of these components. The goal of this section is to determine whether or not the relative weights of the different components can be used in place of more commonly used emission-line parameters to understand the physical processes occurring in these quasars.

Issues with the parametric model fitting approach adopted above include sensitivity to S/N. We also found the empirical

template to be a poor match to the Fe II emission observed in a number of quasars.

### 4.8.1 *The technique*

ICA is a blind source separation technique for separating a signal into 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 successfully applied to model the spectra of emission-line galaxies (Allen et al., 2013). 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 matrix, $W$:

$$x = Wc \tag{4.2}$$

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:

$$c = W^{-1}x. \tag{4.3}$$

*Ask Paul for details.*

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-XX Å where H$\beta$ 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 lower-amplitude 'correction' components that could be negative were found to be sufficient to reconstruct the spectrum, without over-fitting. Each quasar spectrum $x_j$ can then be represented as a linear combination of the independent components:

$$x_j = \sum_{i=1}^{10} c_{ij} W_{ij} \tag{4.4}$$

#### 4.8.1.1 *Fitting procedure*

Each of the individual ICA components has been adjusted to give the same overall shape as a quasar template spectrum.

We approximate the overall shape of this template by fitting a single power-law to emission-line free windows at 4200-4230, 4435-4700 and 5100-5535 Å. We then flatten each of the ICA components by dividing by this power-law. An identical process is performed on each spectrum we fit, so that both the components and the spectrum to be fitted have essentially zero large-scale slope. For each quasar in our sample we perform a variance-weighted least-squares minimisation to determine the optimum value of the components weights. The first six component weights are constrained to be non-negative, and the fit is performed in logarithmic wavelength space, so that each pixel corresponds to a fixed velocity width. The relative shift of the ICA components is also allowed to vary in the optimisation procedure, to account for errors in the systemic redshifts used to transform the spectra into rest-frame wavelengths.

### 4.8.2  *Quality of fits*

In general, the ICA components are able to reconstruct the spectra of the objects in our sample. We also find that that in some cases, the ICA reconstructions are superior at modelling the Fe II emission than the Boroson and Green, (1992) template.

*Need something more quantative Look at chi-squared distribution? Doesn't seem that reliable.*

### 4.8.3  *Physical interpretation of ICA components*

Although the ICA analysis is not based on any physics, there appears to be a direct correspondence between the individual non-negative components and the different emission features which contribute to the spectra (Figure 4.15). This correspondence is summarised in Table 4.6. The component $w_1$ seems to correspond to Fe II emission, the components $w_2$ and $w_3$ to broad H$\beta$ 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. s

### 4.8.3.1  *Reconstructing the [O III] profile*

In order to measure non-parametric 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. The four correction components are also included.
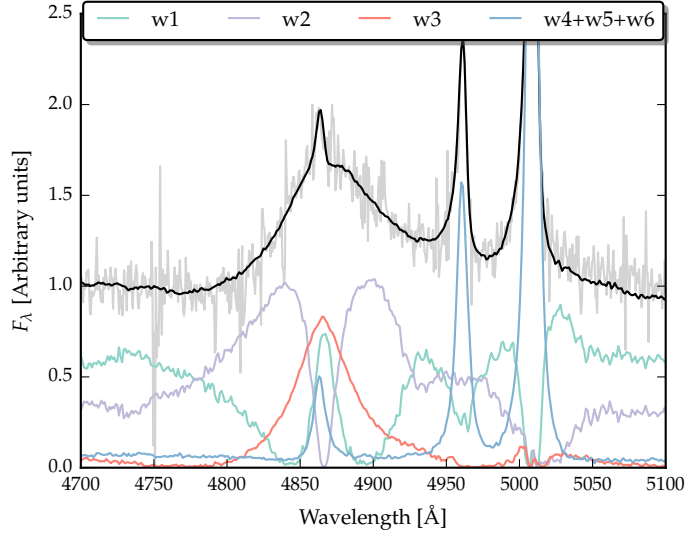
Figure 4.15: Hβ/[OIII] emission J002952+020607. The ICA reconstruction is shown in black, and the spectrum in grey. The first three components, and the sum of components four, five and six are shown individually.

Table 4.6: Physical interpretation 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 |

We define the boundaries of [OIII]$\lambda$5008 as being between 4950 and 5500 Å. The blue limit is close to the peak of the [OIII]$\lambda$4960 line, and so to recover the intrinsic profile we instead use the blue wing of [OIII]$\lambda$4960. We use the emission from 4980-5050 Å, 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 [OIII]$\lambda$5008 line. We then subtract a constant, because the flux does not always go to zero (suggests that there is probably flux which is not due to [OIII] emission in components four to six).

An examples of a reconstructed [OIII] emission-line is shown in Figure 4.16. At present I am summing the flux all the way from 4950 Å. However, this is quite a lot of flux to sum up, and we can't ascribe this flux to the wing of the [OIII] emission with
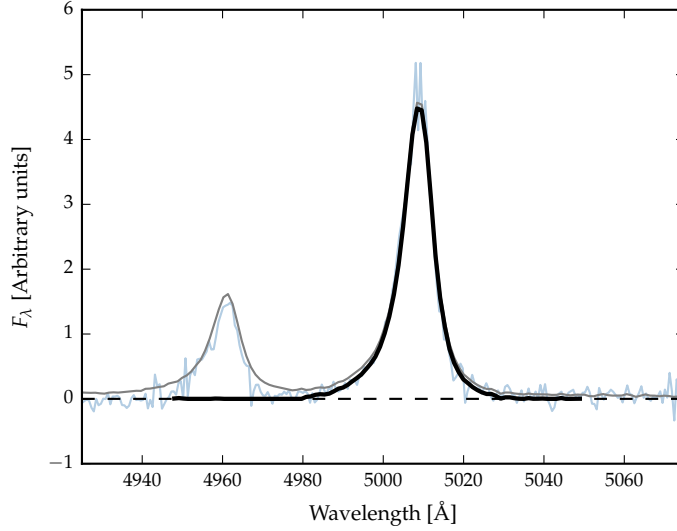
Figure 4.16: [O III] emission in J002952+020607. The data is shown in blue, and the ICA spectrum in grey. The first three ICA components have been subtracted from both the ICA composite and the data. The black curve shows the reconstructed [O III]λ5008 profile.

any certainty. This is borne 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.

Unfortunately, there are systematic differences between the line-width estimates from the Gaussian reconstructions and the ICA reconstructions, particularly for broad-line objects. The current way of doing the ICA reconstruction of the [O III] line ignores any cross-talk between the components and there is potentially flux being ascribed to the line that could be coming from some other component. We can solve this by finding some more representative broad [O III] lines in SDSS from which to derive the components as well as producing a set of components for [O III] only. Therefore we don't use these reconstructions and leave this for future work.

### 4.8.4 ICA fits

In Figure 4.17 we show the relative weights of each of the six positive ICA components. Also shown are the same measurements for a sample of low-redshift, low-luminosity AGN. We want to examine whether or not there are systematic differences between these two samples.

*Paul: Think you need to say something about the SDSS quasar sample which was used to generate the ICA components. I woul at that point herald that the limited spectral diversity in the SDSS sample is likely to prove an issue, even if the ICA aproach looks promising.*

*Paul: Paragraph belongs with material above*

$$\frac{w_i}{\sum_{i=1}^{6} w_i}$$
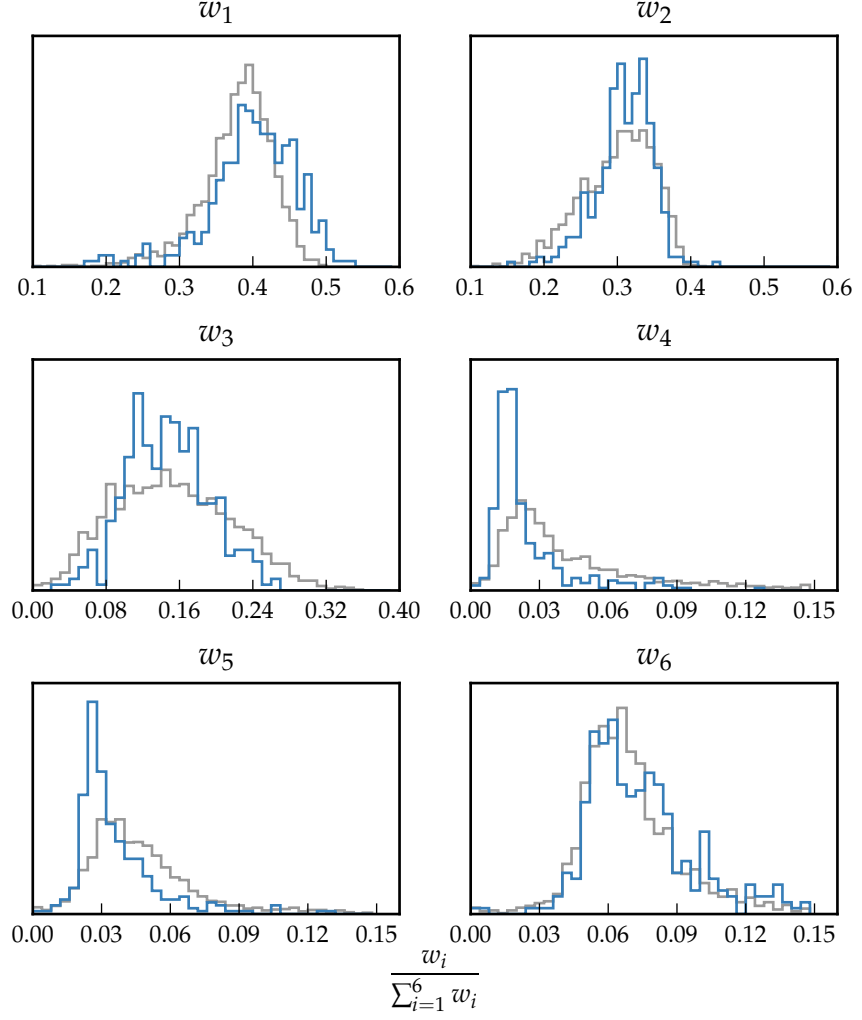
Figure 4.17: The relative weight in each of the six positive ICA components for the high-luminosity (blue) and low luminosity samples (grey). In the high-luminosity sample Fe II emission is stronger (component $w_1$). The core [O III] emission (components $w_4$, $w_5$) is weaker but the strength of the blueshifted wing ($w_6$) is the same.
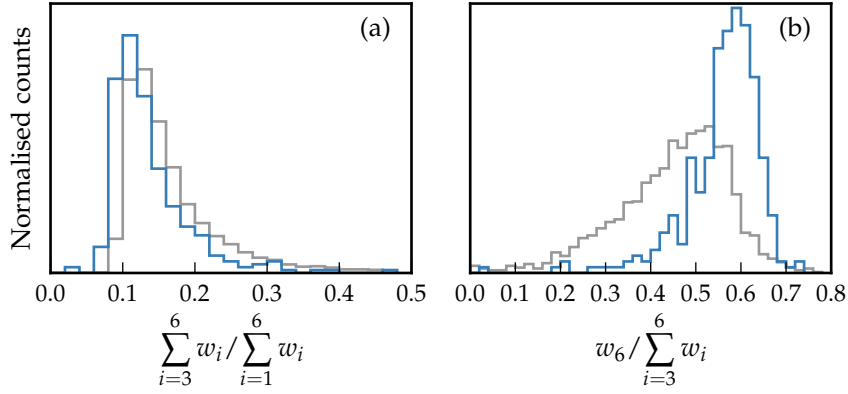
Figure 4.18: The relative weight in the three ICA components corresponding to [O III] emission (*left*) and the relative weight of the component most closely related to blueshifted [O III] emission relative to all three [O III] components (*right*). [O III] emission is weaker in the high-luminosity sample, but the relative contribution from the blueshifted component to the total [O III] emission is higher.
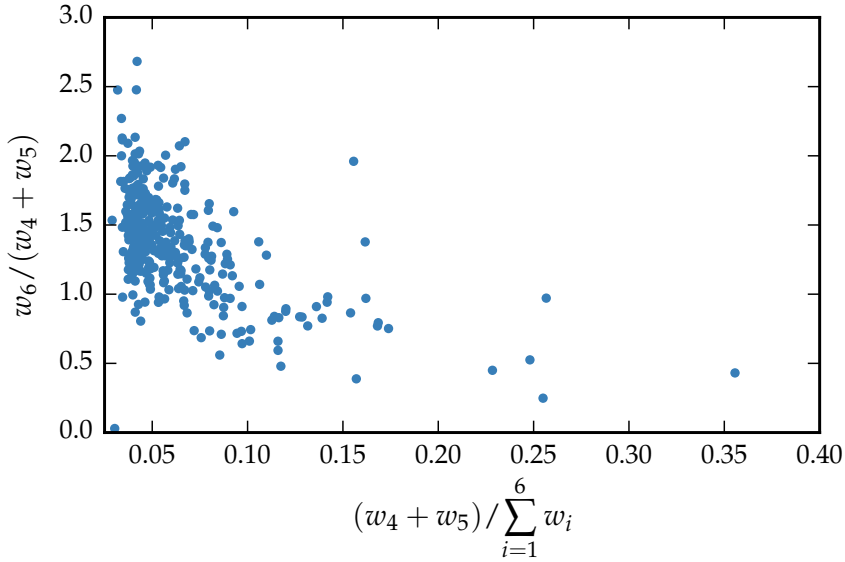


Figure 4.19: Weight in the [O III] wing relative to the weight in the [O III] core emission versus the strength of the core [O III] emission. The blue-asymmetry of the [O III] emission increases as the strength of the core component decreases.
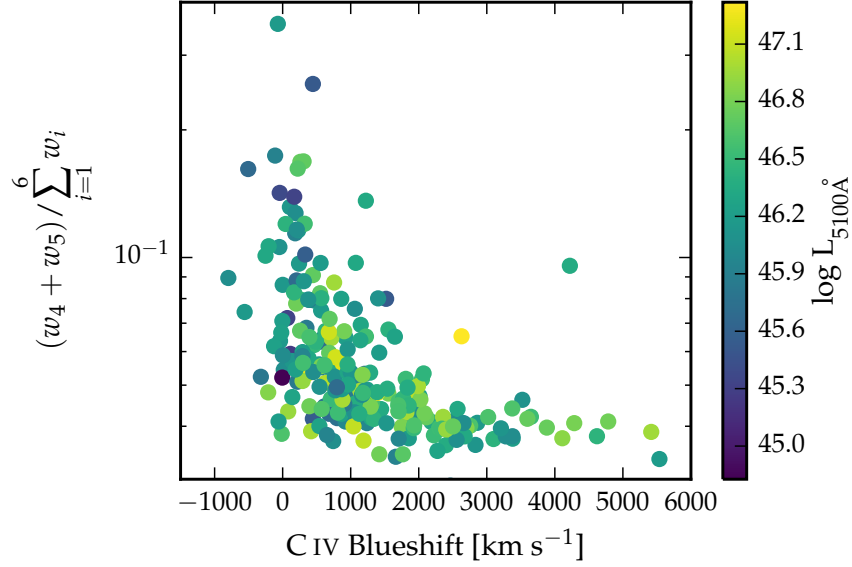
Figure 4.20: The ICA component weight $w_4$, which is a proxy for the strength of core [O III], as a function of the C IV blueshift. The C IV blueshift is measured relative to the near-infrared ICA redshift.

We see that [O III] core emission is weaker in the more luminous sample, but the strength of the wing component is similar. 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. Shen and Ho, (2014) suggested that a stable NLR is being removed by the outflowing material. Similarly, Zhang et al., (2011) found that the more the peak of the [O III] line is blueshifted, the more the core component decreases dramatically, while the blue wing changes much less. Therefore, there is an anti-correlation between the strength of the core component and the relative strength of the wing component (Figure 4.19).

To show this phenomenon more clearly, we plot the relative [O III] strength and the [O III] wing/core ratio in the high/low luminosity samples (Figure 4.19). We see that [O III] is weaker in the high luminosity sample, but that the wing component is *Similar to* much stronger relative to the core component. .

*behaviour of C IV?*
*Would suggests*
*that the mechanism*
*producing the two*
*correlations is the*
*same*
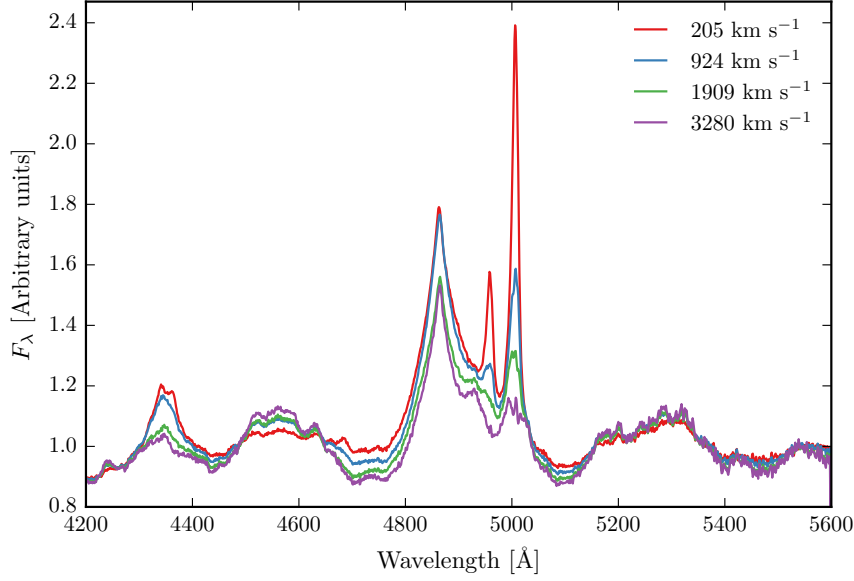
Figure 4.21: Median ICA-reconstructed spectra as a function of the
C IV blueshift.

#### 4.8.4.1 *EV1 correlations*

In Figure 4.20 we show how the [O III] strength varies as a
function of the C IV blueshift. There is a very well defined re-
lation: when C IV is strongly blueshifted [O III] is very weak.
This is very similar to what we found when we used Gaussian
functions to model the emission. The correlation between C IV
blueshift and [O III] EQW is shown in a different way in Fig-
ure 4.21. Here we divide our sample into 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 pro-
gressively weaker and more blueshifted. The anti-correlation
with Fe III and the blue-ward Fe II also clear, but there is no
change in the redward Fe II.

#### 4.8.4.2 *Updating EV1*

The ICA can be thought of as update on EV1. EV1 is a quest for
HR diagram for quasars - clearly very important. 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.

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 measures of the host-galaxy rest frame suggest that [O III] usually gives consistent results within $200\,\mathrm{km\,s^{-1}}$ (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 II emission can be improperly subtracted, or [O III] might not be detected at all. [O III] is weaker and broader so it is more difficult to detect and measure [O III] accurately for these luminous quasars (to for instance obtain reliable redshift estimates based on [O III])

### 4.8.5 *Future work*

Pros:

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.

Cons:

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

Cross-talk between components.

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 kilo-parsec; Netzer et al., 2004).

See extra text from Brotherton paper. I could be confused here, but I think the Netzer argument goes that the nlr size increase with luminosity because there are more ionising pho-

*Just present this as an idea for future work right at the end rather than having this sandwiched in the middle.*

We see a correlation between the [O III] velocity width and asymmetry. As the line gets broader it gets more blue-asymmetric. One interpretation of this is that the strength of the narrow core is decreasing, leading to a broader and more blueshifted profile (e.g. Shen and Ho, 2014).

[O III] is broader, which is consistent with these quasars having more massive BHs. [O III] also shows stronger blue asymmetries, suggesting that outflows are stronger/more prevalent at these higher luminosities/redshifts. The luminous blueshifted broad wing and the extremely broad profile reveals high-velocity outflowing ionized gas. Our results therefore suggest that kilo-parsec-scale outflows in ionized gas are common in this sample of high-luminosity, high-redshift quasars.

## 4.9 RADIO

Radio properties would be interesting. Have another look, and at least say we tried and there aren't enough objects.

## 4.10 SUMMARY

At fixed luminosity, I find that as the blueshift of the C IV emission increases, the [O III] emission becomes weaker and more blueshifted, and disappears entirely in quasars with the most extreme C IV blueshifts.

# 5

# OUTFLOWS AND HOT DUST EMISSION

## 5.1 INTRODUCTION

Many quasars and AGN show an excess in their near-infrared continuum emission. This feature is generally attributed to thermal emission from dust heated by optical/ultra-violet radiation from the accretion disc. The wavelength of the feature ($\sim 2\,\mu m$) corresponds to the spectral peak for graphite dust at its sublimation temperature (T $\sim$ 1500 K; Barvainis, 1987). Reverberation measurements of nearby AGN suggest that the hot dust is very close to the central source (few tens of light days; e.g. Minezaki et al., 2004; Suganuma et al., 2006). This places the dust at the innermost edge of the putative torus-like structure, at a radius set by the sublimation temperature of the dust grains.

Studies have fitted the near-infrared SEDs of AGN using a blackbody spectrum to represent emission from hot dust (e.g. Edelson and Malkan, 1986; Barvainis, 1987; Kishimoto et al., 2007; Mor, Netzer, and Elitzur, 2009; Riffel, Storchi-Bergmann, and McGregor, 2009; Deo et al., 2011; Landt et al., 2011; Mor and Trakhtenbrot, 2011; Roseboom et al., 2013). A hot dust component is present in the vast majority of AGN, although populations of 'dust-free' objects have also been discovered (Hao et al., 2010; Hao et al., 2011; Jiang et al., 2010; Mor and Trakhtenbrot, 2011). It is not yet clear how the hot dust properties relate to other AGN properties such as BH mass, luminosity and accretion rate.

In recent years the picture of the torus has evolved away from a static 'doughnut' towards a more general circum-nuclear, geometrically and optically thick dust distribution. As we have previously discussed (e.g. Section 1.6), winds and outflows launched from the accretion disc are very common in AGN. 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 an accretion disc wind beyond the dust sublimation radius. The dusty clouds are uplifted above the disc where they are directly exposed to the central engine. The

dust is heated, and radiates in the near-infrared band. At the same time, radiation pressure due to dust can efficiently accelerate the wind. The wind is roughly polar, and so naturally provides circum-nuclear obscuration around the accretion disc and dust-free BLR. This model is supported by recent interferometric observations of nearby Seyfert galaxies which find that the mid-infrared emission is dominated by dust in the polar regions (e.g. Raban et al., 2009; Hönig et al., 2012; Hönig et al., 2013; Tristram et al., 2014; López-Gonzaga et al., 2016).

Studying the relationship between emission from hot dust and outflow diagnostics in the BLR can help place constraints on this dusty wind model (e.g. Wang et al., 2013). This is now possible by combing data from the SDSS spectroscopic and photometric surveys with photometric surveys such as UKIDSS and WISE. At redshifts $2 \lesssim z \lesssim 3$, SDSS spectra reveal BH masses, accretion rates and diagnostics of the BLR dynamics. At the same time, the available photometric data provides full ultra-violet to infrared rest-frame coverage of the SED. In particular, the WISE photometry is sensitive to the $3\,\mu$m region of the SED which is dominated by hot dust.

In this chapter we build a simple parametric SED model that is able to reproduce the median optical-infrared colours of tens of thousands of SDSS AGN at redshifts $1 \lesssim z \lesssim 3$ (Section 5.4). We use this model to measure the hot dust properties of a large sample of $2 < z < 2.7$ quasars for which we have already measured C IV line properties, BH masses and Eddington ratios (Section 5.5).

## 5.2 DATA

### 5.2.1 *SDSS*

We use spectroscopic data from the SDSS DR7 spectroscopic quasar catalogue. The SDSS photometric survey obtained images in five broad optical passbands: $u$, $g$, $r$, $i$ and $z$ (Table 5.1). We use BEST point-spread function magnitudes from the SDSS DR7 quasar catalogue.

#### 5.2.1.1 *UKIDSS Large Area Survey*

We use the tenth data release (DR10) of the UKIDSS Large Area Survey (ULAS) which has observed $\sim 3,200$ deg$^2$ in four near-infrared passbands: Y, J, H and K. We use 'apermag3' magni-

| Survey | Passband | $\lambda_{\text{eff}}$ [μm] | AB offset | $A_{\text{filter}}/E(B-V)$ |
|--------|----------|----------|-----------|----------|
| SDSS | u | 0.3543 | 0.913 | 4.875 |
| | g | 0.4770 | −0.081 | 3.793 |
| | r | 0.6231 | 0.169 | 2.721 |
| | i | 0.7625 | 0.383 | 2.099 |
| | z | 0.9134 | 0.542 | 1.537 |
| UKIDSS | Y | 1.0305 | 0.641 | 1.194 |
| | J | 1.2483 | 0.941 | 0.880 |
| | H | 1.6313 | 1.378 | 0.569 |
| | K | 2.2010 | 1.897 | 0.352 |
| WISE | W1 | 3.4 | 2.691 | 0.182 |
| | W2 | 4.6 | 3.331 | 0.130 |
| | W3 | 12.0 | | |

Table 5.1: Available photometry, effective wavelength, Vega to AB magnitude offsets, conversion from $E(B-V)$ to extinction. Need W3 offset/ extinction. Ask Paul or just use WISE values/zero.

tudes, which are aperture corrected magnitudes in a $2''$ diameter aperture.

### 5.2.1.2 *WISE All-WISE Survey*

WISE mapped the entire sky in four mid-IR passbands: $W1$, $W2$, $W3$ and $W4$. 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 NEO-WISE program (Mainzer et al., 2011). We use the profile-fitting 'mpro' magnitudes. Only information from the first three WISE passbands are used in this work.

### 5.2.2 *Computing Vega-AB magnitude offsets*

Vega magnitudes are used throughout this chapter. This is the native magnitude system for UKIDSS and WISE. We add 0.08 mag to the UKIDSS Y passband magnitudes to bring the photometry into better agreement with the SDSS z and UKIDSS J photometry. SDSS uses an 'asinh' magnitude system (Lupton, Gunn, and Szalay, 1999) which is intended to be on the AB system (Oke and Gunn, 1983). However, the photometric zero-points are known to be slightly off the AB standard. The z pass-

band is in error by 0.02 ($z_{AB} = z_{SDSS} + 0.02$)[1]. The u passband was in error by 0.04 dex at the time of DR7. However, using an updated u throughput function (Doi et al., 2010), we find that the u zero-point is now consistent with the AB system.

Using a reference template, we calculate the AB magnitude of Vega in each passband (Table 5.1). The mean flux density $f_\lambda(P)$ in a passband defined by a throughput function $P(\lambda)$ is given by:

$$f_\lambda(P) = \frac{\int P(\lambda) f_\lambda(\lambda) \lambda d\lambda}{\int P(\lambda) \lambda d\lambda} \tag{5.1}$$

where $f_\lambda(\lambda)$ is the flux density of the object. The predicted passband is then given by:

$$m_\lambda(P) = -2.5\log(f_\lambda(P)) - m_0(P), \tag{5.2}$$

where $m_0(P)$ is the zero-point magnitude of passband $P$, given by evaluating Equation 5.1 for a reference object. In the AB system this is a constant spectral flux density of 3631 Jy. In flux per unit wavelength this is:

$$\frac{f_\lambda(\lambda)}{\mathrm{erg\ cm^{-2}\ s^{-1} Å^{-1}}} = 0.1087 \left(\frac{\lambda}{Å}\right)^{-2}. \tag{5.3}$$

In the Vega system, a spectrum of the A0V star Vega is used. Although the magnitude of Vega is by design zero in every passband, more recent measurements reveal a small magnitude offset. The Vega to AB magnitudes given in Table 5.1 assume Vega to have a magnitude 0.026.

### 5.2.3 *Galactic extinction correction*

$A(u)$, the Galactic extinction in the u passband at the position of the object, is given in the SDSS catalogue. It is computed using the maps of Schlegel, Finkbeiner, and Davis, (1998). Schlegel, Finkbeiner, and Davis, (1998) calcuted the extinction assuming a $z = 0$ elliptical template and found $A(u)/E(B - V) = 5.155$, where $E(B - V)$ is the relative extinction between the B and V passbands. Quasar and galaxy optical SEDs have very different shapes, and so we re-derived the extinction in the SDSS

---

[1] http://classic.sdss.org/dr7/algorithms/fluxcal.html.

passbands using a $z = 1.5$ quasar SED template[2]. The extinction in the UKIDSS and WISE passbands were also computed using this template. Conversions from the selective extinction $E(B - V)$ to the total extinction $A(\lambda)$ in each passband is given in Table 5.1.

### 5.2.4 *Cross-matching SDSS to UKIDSS and WISE*

There are 105 783 objects in the SDSS DR7 quasar catalogue. While WISE mapped virtually the entire sky, the UKIDSS footprint covers approximately one third of the SDSS footprint. 36 607 objects are cross-matched to the UKIDSS (with a $2''$ matching radius) and WISE (with a $3''$ matching radius) catalogues.

### 5.2.5 *Sample definition*

We include only the 20 637 quasars with i passband magnitudes brighter than 19.1, i.e. the quasars selected by the main $z < 3$ SDSS quasar selection algorithm $z < 3$ (Richards et al., 2002). We verified that above the $i = 19.1$ limit the sample is 95 per cent complete in all passbands. This suggests that our sample will not be biased towards quasars with redder spectra[3]. BAL quasars are excluded using the Allen et al., (2011) catalogue because the C IV line parameters of these quasars can not be reliably measured. This leaves 19 837 objects.

   We further limit our sample to the redshift range $1 < z < 3$. Galaxy SEDs peak at $\sim 1\,\mu$m, and fall away towards shorter wavelengths. On the other hand, AGN SEDs continues to increase short-ward of $1\,\mu$m. As a result, the contrast between the AGN and galaxy luminosity increases as the redshift increases. Imposing a $z = 1$ lower redshift limit on the redshift of our sample ensures that contributions to the SED from quasar host galaxies are negligible. The completeness of the SDSS DR7 quasar selection algorithm decreases steeply beyond $z \sim 3$, and this sets the upper redshift limit. The redshift and luminosity distribution of the final sample, containing 12 934 quasars, is shown in Figure 5.1.

*I don't fully understand this point*

---

2 Extinction corrections were derived by Prof. Paul Hewett.
3 For a given i magnitude, a redder spectrum is more likely to be detected at longer wavelengths than a bluer spectrum.
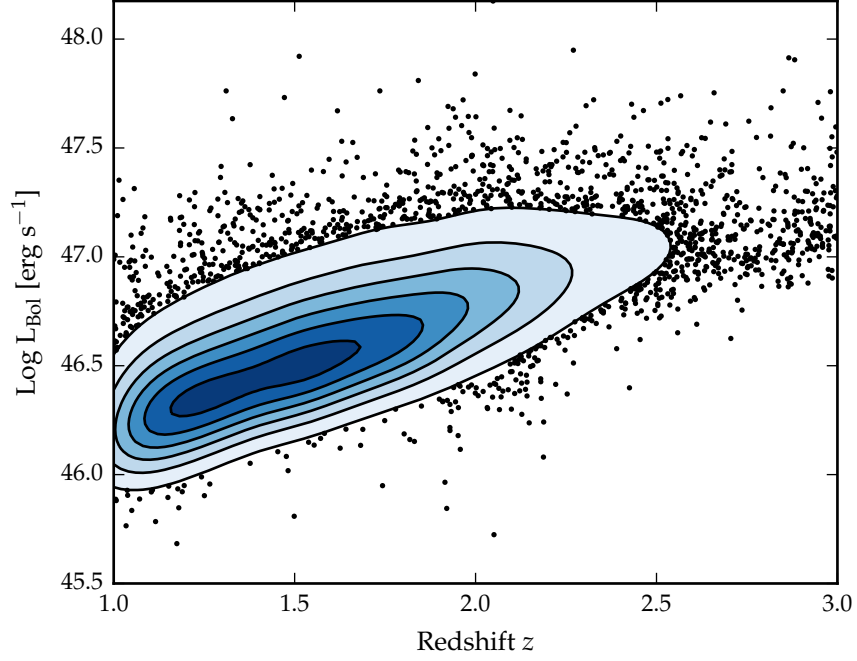
Figure 5.1: Distribution of our sample in the redshift-luminosity plane.

## 5.3 CONSTRUCTING AN AGN SED MODEL

Since the physical processes that power AGN are generally understood only qualitatively, almost all AGN SED templates are empirical. The empirical template of Elvis et al., (1994) 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). However, these composite spectra are often constructed from quasars with a huge range in luminosity as a function of wavelength. In addition, the presence of significant host-galaxy at optical wavelengths in low-redshift objects is an additional complication which has not always been taken care of adequately. There is therefore a strong rationale for taking a parametric approach to modelling quasar SEDs. This the approach adopted in this work.

*Ask Paul for details on this.*

We construct an SED model that is valid between $1216\,\text{Å}$ and $3\,\mu\text{m}$. In this region the SED is dominated by the accretion disc, broad ultra-violet/optical emission-lines and thermal emission from the hottest ($T \sim 1200\text{K}$) dust. In this section, we describe how each of these components are represented in our paramet-

ric SED model. The effect of dust extinction at the AGN redshift is also incorporated into the model. At high redshifts, Ly$\alpha$ forest absorption becomes significant. Because we do not attempt to model this effect, our model is valid only at wavelengths long-ward of 1216 Å. We model dust emission using a single temperature (T ~ 1200 K) blackbody, which peaks at ~ 2 μm. At longer wavelengths, emission from cooler dust farther from the central engine becomes increasingly important. We do not include this emission in our model, which restricts its validity to $\lesssim$ 3 μm. The model spectrum is shown in Figure 5.2, with each of the main components indicated.
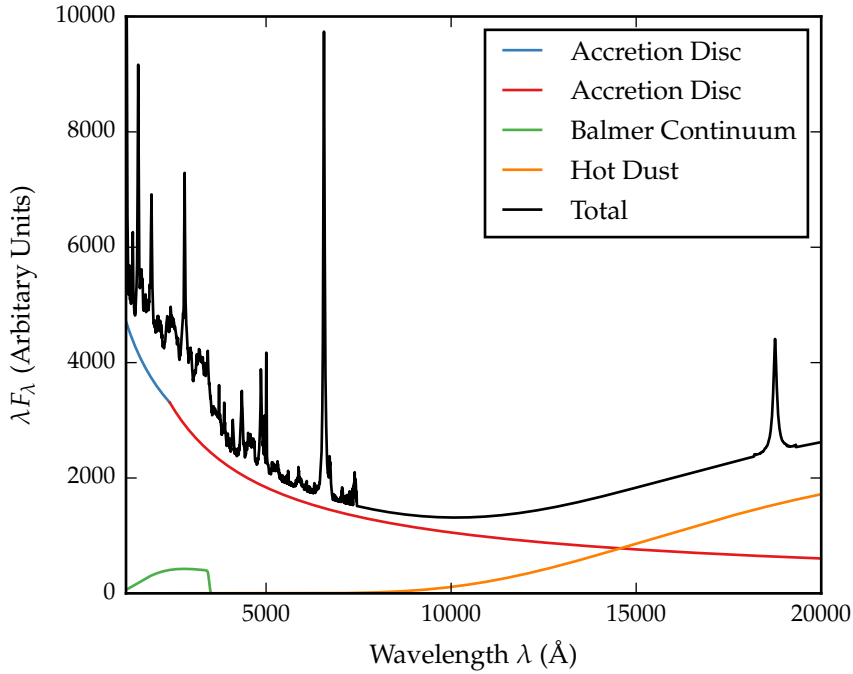


Figure 5.2: Model quasar spectrum at $z = 1$, showing the contributions to the total flux from the accretion disc, Balmer continuum, hot dust and emission-lines.

### 5.3.1 Accretion disc

Thermal accretion disc emission in the $0.1 - 1$ μm region is characterised by a broken power-law with three free parameters: a break-wavelength, $\lambda_{break}$, a blue power-law index, $\alpha_{blue}$, for wavelengths shorter than the break wavelength, and a red

power-law index, $\alpha_{red}$, for wavelengths longer than the break wavelength.

### 5.3.2 *Balmer continuum*

High order Balmer lines, optically thin Balmer continuum emission, two-photon emission and Fe II emission blend together to form the 'Balmer' continuum at $\sim 3000\,\text{Å}$. We simulate the Balmer continuum using the empirical model given by Grandi, (1982):

$$F(\lambda) = C_{BC} \times B_\lambda(T_e)(1 - e^{-\tau_\lambda}); \quad \lambda \leqslant \lambda_{BE} \tag{5.4}$$

where $C_{BC}$ is a normalisation factor, $B_\lambda(T_e)$ is the Planck function, $T_e = 13150\,\text{K}$ is the effective temperature, $\lambda_{BE} = 3460\,\text{Å}$ is the wavelength at the Balmer edge, and $\tau_\lambda = \tau_{BE}\,(\lambda_{BE}/\lambda)^{-3}$ is the optical depth with $\tau_{BE} = 45$ the optical depth at $\lambda_{BE}$. This function is convolved with a Gaussian with width $\sigma = 5000\,\text{km}\,\text{s}^{-1}$ to simulate the effect of bulk velocity shifts comparable to those present in broad AGN emission-lines.

### 5.3.3 *Hot dust*

Thermal emission from hot dust, which dominates the SED at wavelengths longer than 1 μm, is modeled using a blackbody

$$F_\lambda = C_{BB} \times \frac{2hc^2}{\lambda^5} \frac{1}{e^{\frac{hc}{\lambda k_B T_{BB}}} - 1}, \tag{5.5}$$

with two free parameters: the temperature $T_{BB}$ and normalisation $C_{BB}$ relative to the power-law continuum.

### 5.3.4 *Emission-lines*

We use an emission-line template taken from Francis et al., (1991), which has been extended by Maddox and Hewett, (2006) to include the Hα and Paα emission-lines[4]. All emission-lines, with the exception of Hα and Hβ, are scaled using a single free parameter $C_{EL}$, which preserves relative EQWs:

---

4 The spectrum is not significantly different from the Vanden Berk et al., (2001) SDSS composite.

$$F_\lambda = C_{EL} \times \frac{F_{\lambda,el}}{F_{\lambda,cont}} \times F_\lambda \qquad (5.6)$$

where $F_{\lambda,el}$ is the emission-line template, $F_{\lambda,cont}$ is the continuum flux in the template, and $F_\lambda$ is the continuum flux in the SED model. The redshifts and luminosities of the quasars contributing to the emission-line template change as a function of wavelength. To account for possible variations in the strengths of the different lines, H$\alpha$ and H$\beta$ are scaled separately:

$$F_\lambda = C_{EL} \times C_{H\alpha} \times \left(\frac{L(z)}{L(z_{nrm})}\right)^\beta \times \frac{F_{\lambda,el}}{F_{\lambda,cont}} \times F_\lambda. \qquad (5.7)$$

The luminosity dependence of the H$\alpha$ and H$\beta$ EQW (i.e. the Baldwin effect; Baldwin, 1977) is parametrised with a power-law with slope $\beta = -0.04$. The dependence of the mean AGN luminosity on redshift, $L(z)$, is determined empirically for the SDSS quasar catalogue.

### 5.3.5  *Dust extinction*

We simulate the effect of dust extinction at the quasar redshift using a custom extinction curve that is appropriate for the quasar population[5]. 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 un-reddened objects. At redshifts $2 < z < 3$ the reddening measure is made at rest-frame wavelengths $3500 - 7000\,\text{Å}$, where Galaxy, LMC and SMC[6] extinction curves are very similar. The SDSS spectra of the same objects are then employed to generate an empirical extinction curve in the ultra-violet, down to $1200\,\text{Å}$. The resulting curve has no $2200\,\text{Å}$ feature and rises rapidly with decreasing wavelength but is not as steep as the SMC curve. The extinction curve gives the colour excess $E(B-\lambda)$ relative to the colour excess $E(B-V)$ as a function of wavelength $\lambda$. The colour excess $E(B-V)$ is related to the extinction in the V passband, $A(V)$, via the ratio R:

$$R_V = \frac{A(V)}{E(B-V)} \qquad (5.8)$$

---

5 The extinction curve has been derived by Prof. Paul Hewett.
6 LMC and SMC: Large and Small Magellanic Clouds.

where we assume $R_V = 3$. Hence the extinction at a wavelength lambda $A(\lambda)$ is

$$A(\lambda) = E(B - V) \times \left[ \frac{E(\lambda - V)}{E(B - V)} + R \right] \quad (5.9)$$

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:

$$F_\lambda = F_\lambda 10^{-A(\lambda)/2.5} \quad (5.10)$$

in the rest frame of the quasar.

## 5.4 DERIVING A SED TEMPLATE FOR THE QUASAR POPULATION

In this section we determine a single set of SED model parameters for all $19\,853$ quasars, encompassing a range of redshifts, luminosities, accretion rates etc.

### 5.4.1 *Fitting procedure*

The free parameters in our SED model are summarised in Table 5.2. 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).

We divide the quasar sample into redshift bins from $z = 1$ to $z = 3$ in intervals of $\Delta z = 0.1$. In each redshift bin median passband magnitudes are calculated, and normalised such that $i = 18$. As described above, the SED model is valid between $\sim 1200$ and $30000\,\text{Å}$. We use the $rizYJHKW1W2$ passbands to constrain the model, which covers $1550 - 23000\,\text{Å}$ in the rest-frame. Model SEDs are generated at redshifts corresponding to the centres of the redshift bins. The SED model is shown at three different redshifts in Figure 5.3. Model magnitudes are calculated using Equations 5.1 and 5.2 and are normalised such that $i = 18$. We find the best-fitting model parameters by minimising the $\chi^2$ statistic for the $9 \times 21 = 181$ model and data magnitudes using the Nelder-Mead algorithm.
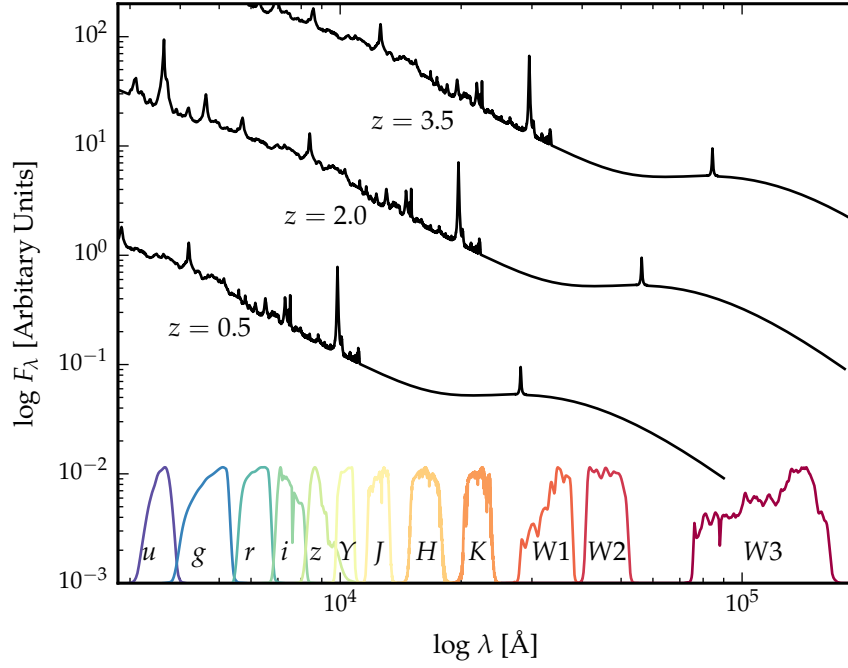
Figure 5.3: Model quasar spectrum at three different redshifts (each arbitrarily scaled), and throughput functions for SDSS, UKIDSS and WISE passbands.

| Parameter | Symbol | Value |
|---|---|---|
| Blue power-law index | $\alpha_{blue}$ | $-0.478$ |
| Red power-law index | $\alpha_{red}$ | $-0.199$ |
| Power-law break | $\lambda_{break}$ | 2402 |
| Blackbody temperature | $T_{BB}$ | 1306 K |
| Blackbody normalisation | $C_{BB}$ | 2.673 |
| Emission-line scaling | $C_{EL}$ | 1.240 |
| H$\alpha$ emission-line scaling | $C_{H\alpha}$ | 0.713 |
| Balmer continuum scaling | | 0.135 |

Table 5.2: Best-fitting SED model parameters from fit to the median colours of quasars at redshifts $1 < z < 3$.
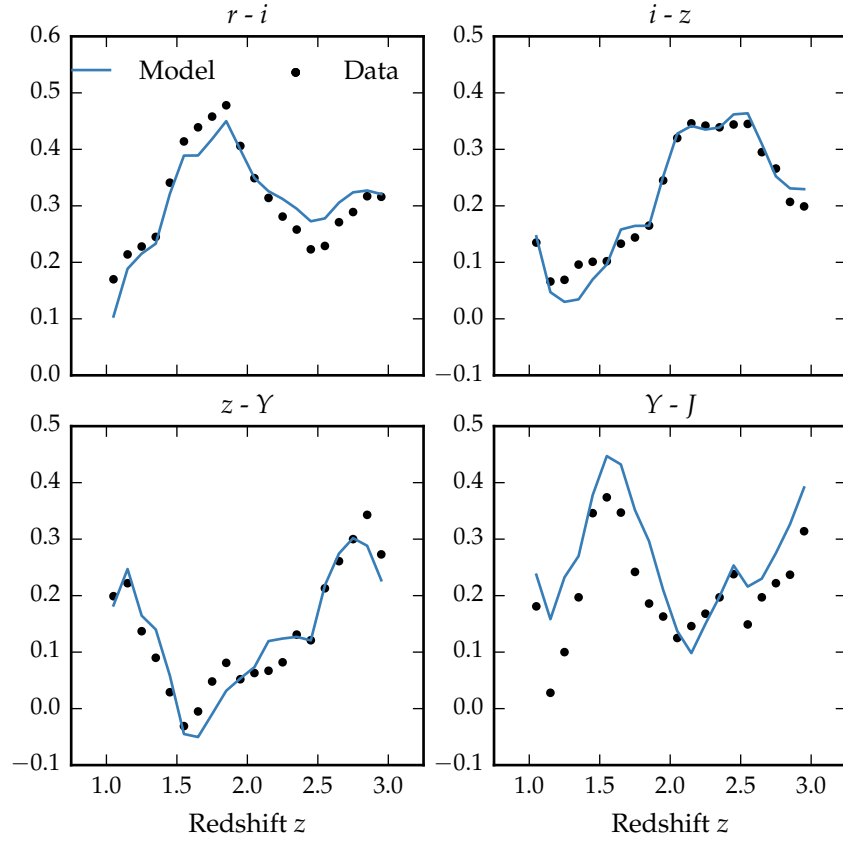
Figure 5.4: Median colours of quasars as a function of redshift and best-fitting SED model.

### 5.4.2 *Results from fit*

The best-fitting parameters from the fit are given in Table 5.2. The colours ($r - i$, $i - z$, etc.) of the median SED and the best-fitting model are plotted as a function of redshift in Figure 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 into and out of the passbands.

In Figure 5.5 we show the data minus model residuals as a function of the rest-frame wavelength. The residuals indicate that over a large redshift range the model is very effective at reproducing the median observed colours of the sample. Discrepancies are at the $< 0.1$ mag level. At a given rest-frame wavelength, there are no significant discrepancies between residuals in different passbands. This indicates that there is no significant evolution in the median SED as a function of redshift. We conclude that a single, fairly simple parametric is effective at
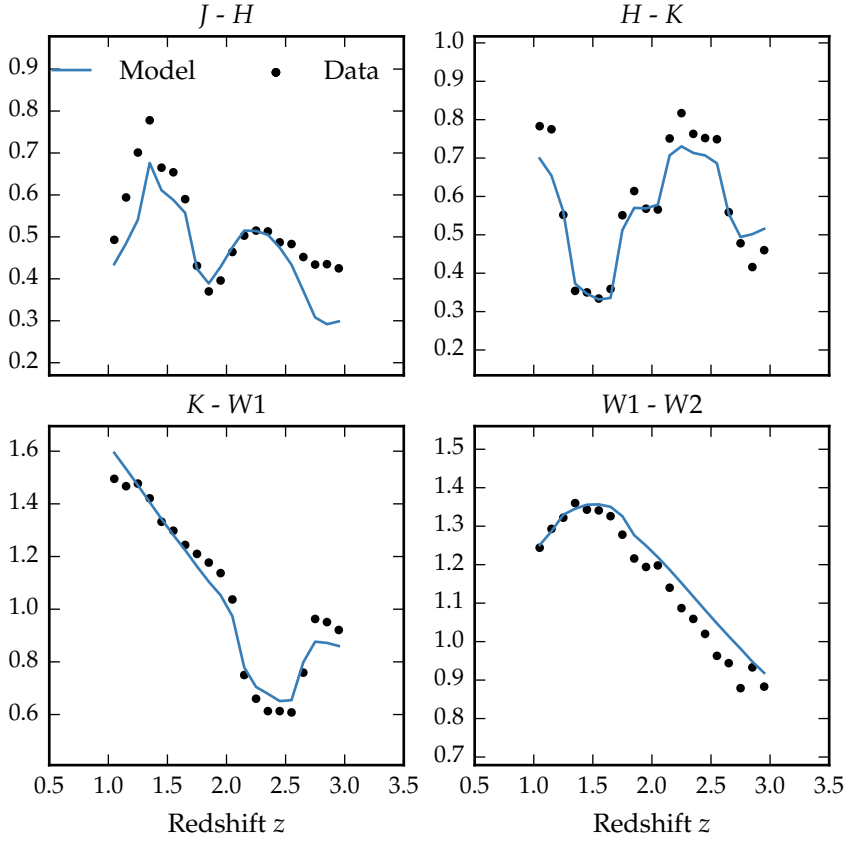
Figure 5.4: Continued.

reproducing the median colours of tens of thousands of AGN with a large dynamic range in redshift and luminosity.

## 5.5 DIVERSITY OF HOT DUST PROPERTIES

In Figure 5.6 we plot the $W1 - W2$ colours of the sample as a function of redshift. At any given redshift we see a $\sim 0.5$ mag dispersion in the $W1 - W2$ colours. In this redshift range the $W1$ and $W2$ passbands are probing the $1.2 - 2.8\,\mu m$ and $1.6 - 3.8\,\mu m$ regions of the rest frame SED respectively. The peak wavelength is at $2.4\,\mu m$ for a blackbody radiating at $1200\,K$. Therefore, the large spread in $W1 - W2$ colours is highly suggestive of a range of hot dust properties in this sample.

We characterise the hot dust properties of our sample in terms of the temperature of the blackbody component and the near-infrared to ultra-violet luminosity ratio, $R_{NIR/UV}$. The
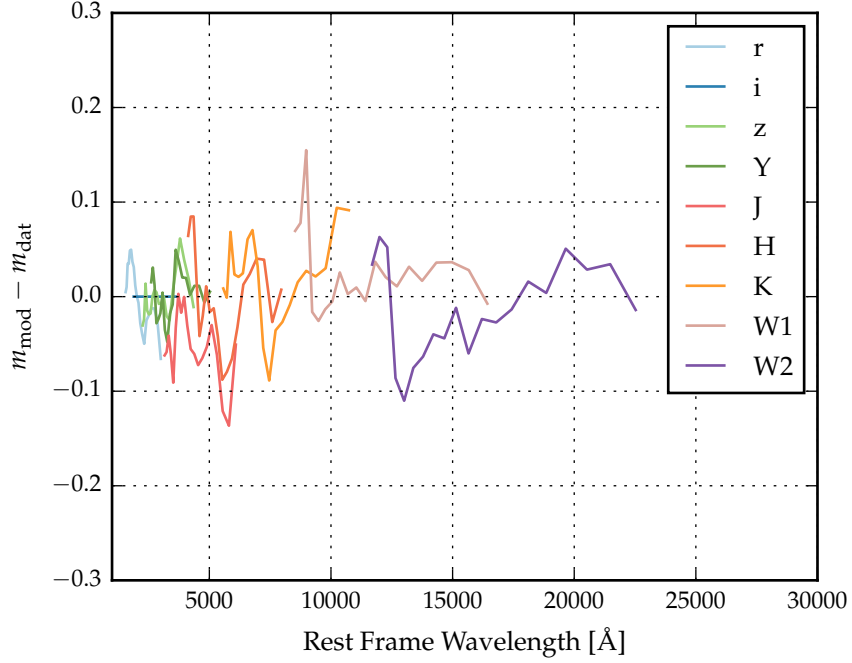
Figure 5.5: Residuals from fitting a single SED model to the colours of $1 < z < 3$ quasars as a function of rest-frame wavelength.

ultra-violet and near-infrared luminosities are calculated between 2000 and 9000 Å and 1 and 3 µm respectively in the SED.

The temperature is likely related to the distance of the dust from the central engine. Dust that is closer in will be hotter, with the sublimation temperature of the dust grains setting the minimum radius. The value of $R_{NIR/UV}$ is related to the covering factor of the hot dust. However, this simple interpretation is somewhat complicated by the fact that at large inclinations sight-lines to the hot dust may be obscured by cooler dust in the putative torus.

In Figure 5.6 we have plotted the $W1 - W2$ colours derived from our SED model with a fixed blackbody temperature (1306 K) and varying $R_{NIR/UV}$. The $W1 - W2$ colours indicate that the hot dust luminosity in this sample varies by a factor of $\sim 5$.

In the remainder of this chapter, we will measure hot dust parameters for individual quasars via SED fitting. We will characterise the range of hot dust properties present in the sample, and test its relation to quasar properties such as luminosity, black-hole mass and normalised accretion rate, and outflow-properties.
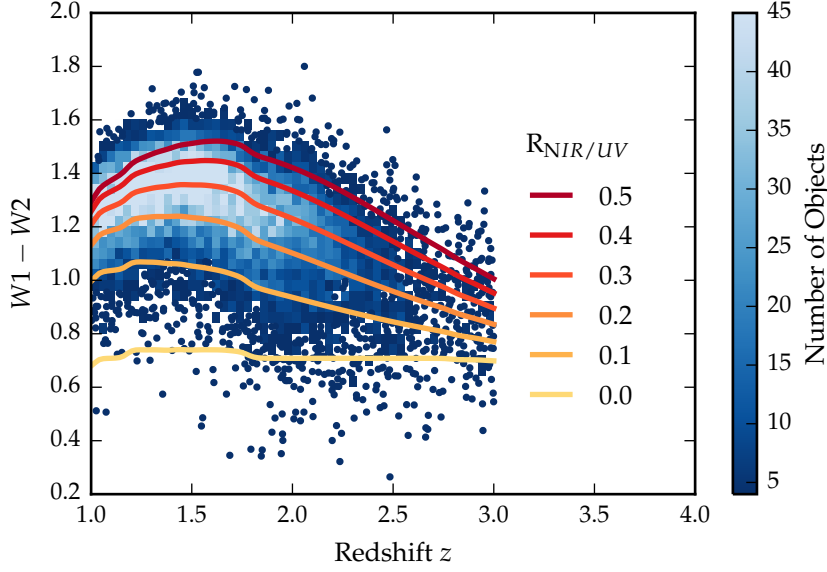
Figure 5.6: $W1 - W2$ colours of sample as a function of redshift. Above a density threshold of four points per pixel points are represented by a two-dimensional histogram. On top we plot the colours of our standard SED model, with a fixed temperature and a varying near-infrared $(1 - 3\,\mu m)$ to ultra-violet ratio.

### 5.5.1 *Defining a sample with uniform UV/optical properties*

We limit our sample to 2329 quasars with redshifts $2 < z < 2.7$. At these redshifts, C IV emission is located within the wavelength coverage of the SDSS spectra. This will allow us to test the relationship between the hot dust and BLR outflow properties.

The shifting of passbands due to redshift limits the redshift range of the quasars for which hot dust properties can be reliably constrained. Constraining a $T \sim 1200\,\mathrm{K}$ blackbody component in the SED model requires photometric data covering $\sim 1 - 3\,\mu m$ in the rest-frame of the quasar. At redshifts $2 < z < 2.7$ $W1W2W3$ is probing the $0.9 - 4\,\mu m$ region of the SED. Therefore the available data is sensitive to the hot dust component across the entirety of the redshift interval.

In general, care must be taken looking for trends with luminosity (and related properties including the BH mass and Eddington ratio) given the observed-frame passband information on the rest-frame SED can produce some strong systematics

with redshift. However, the redshift interval is narrow enough to prevent this from being a significant problem.

Holding the rest of the model parameters fixed, we will vary only the parameters of the blackbody. This requires the SED model to be a reasonable fit to the quasar SEDs in the ultra-violet/optical region. In practice, this means excluding objects with extreme emission-line EQWs and/or significant dust extinction. We use $i - K$ 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 Ly$\alpha$ forest at high redshifts. $i - K$ colours are shown as a function of redshift in Figure 5.7. In the same plot we show the quasar SED model with $E(B - V) = -0.075, 0, 0.075$. A significant amount of the scatter in $i - K$ can be attributed to intrinsic variations in the ultra-violet power-law slopes of the individual quasars, which is why we allow a negative 'reddening'.

The SDSS and UKIDSS photometry are separated by $3 - 4$ years in the source rest-frame. Therefore, some of the $i - K$ scatter could be due to temporal variations in the brightness of the targets. However, the red-asymmetry of the $i - K$ colours about the un-reddened SED model suggests that this effect is sub-dominant to intrinsic colour differences. We discarded from our sample quasars with $i - K$ colours 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 2030 quasars in our high-$z$ sample.

### 5.5.2 *Fitting procedure*

We will fit a model to the individual quasar SEDs, allowing the temperature and normalisation of the blackbody component to vary. The model spectrum is redshifted to the redshift of the quasar being fit and passband magnitudes are calculated using Equations 5.1 and 5.2. We minimise the inverse variance weighted $\chi^2$ statistic using the Levenberg-Marquardt algorithm. We impose a minimum error of 0.1 mag, corresponding to the model error for the medians colours (Figure 5.5). Data from $ugrizYJHKW1W2W3$ is used in the model. However, to avoid Ly$\alpha$ forest absorption, passbands are excluded if $\lambda_{\text{eff}} < 1400\,\text{Å}$.
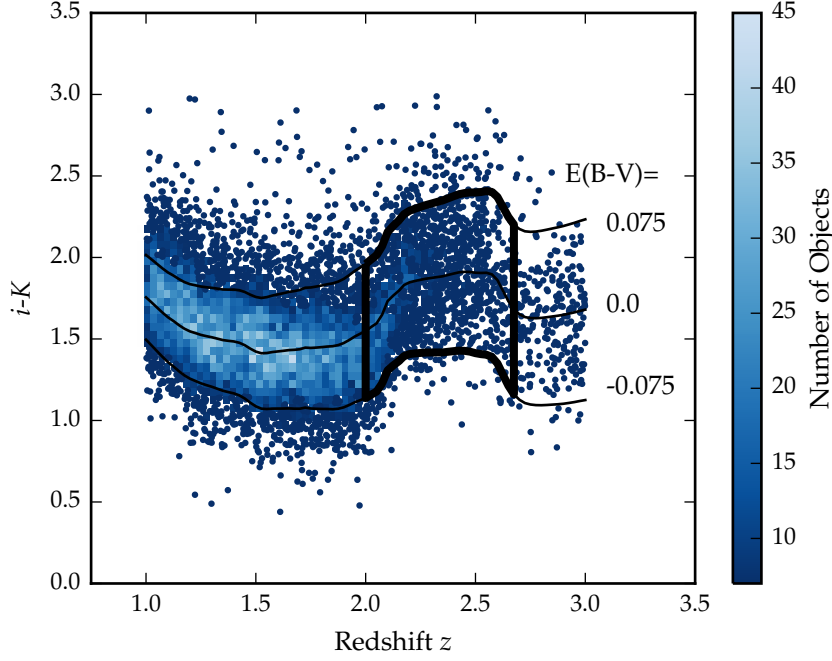
Figure 5.7: $i - K$ colours of non-BAL DR7Q quasars with $i > 19.1$ as a function of redshift. The lines show the colours of our model with varying amounts of dust extinction. Quasars with extinction $|E(B - V)| > 0.075$ are excluded.

### 5.5.3 *Distribution of hot dust parameters*

The best-fitting hot dust temperature ($T_{BB}$) and abundance ($R_{NIR/UV}$) for the individual quasars are shown in Figure 5.8. Even after restricting the sample to have a relative narrow range of ultra-violet/optical SED shapes, we see significant diversity in the hot dust abundance, with the near-infrared ultra-violet luminosity ratio having a broad range from 0.1 to 0.6. The temperature takes on a relatively narrow range of values: $1177 \pm 136$ K. This is consistent with the dust radius being set by the sublimation temperature of the dust grains.

We note a strong correlation between the temperature and $R_{NIR/UV}$. This is a result of the dependence of the blackbody peak on temperature and the fixed wavelength interval used to calculate the blackbody luminosity.
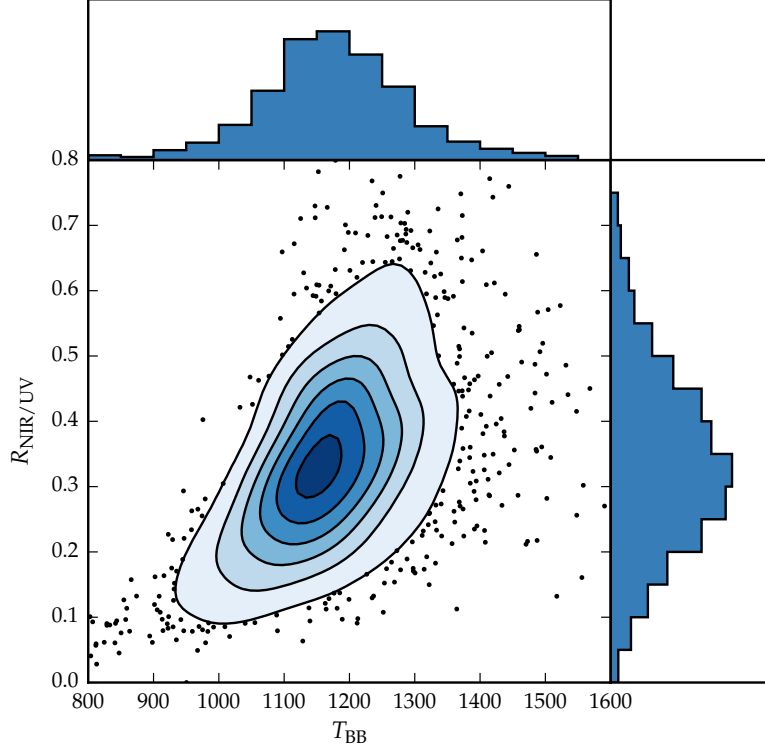
Figure 5.8: Histograms of the ratio of near-infrared to ultra-violet luminosity ($R_{NIR/UV}$) and blackbody ($T_{BB}$) and the correlation between these two parameters.

### 5.5.4 *Relationship between hot dust and BLR outflows*

In this section, we compare the hot dust parameters to the blueshift and EQW of the C IV emission. C IV blueshift measurements are described in Section 3.5. The C IV blueshift is defined with respect to a systemic redshift measured by Allen & Hewett (2017, in preparation). This information is available for 98 per cent of the objects in our sample.

In Figure 5.9 we show the that near-infrared to ultra-violet luminosity ratio is correlated with the C IV blueshift. A similar result was recently reported by Wang et al., (2013). On the other hand, we find no correlations with the hot dust temperature.

The profiles of the emission-lines with large C IV blueshifts suggest that the BLR dynamics in these objects are dominated by high-velocity outflows. As we discussed at the beginning of this chapter, outflows from further out in the accretion disc could contain significant amounts of dust. As the dusty wind is lifted above the accretion disc, it would be directly exposed
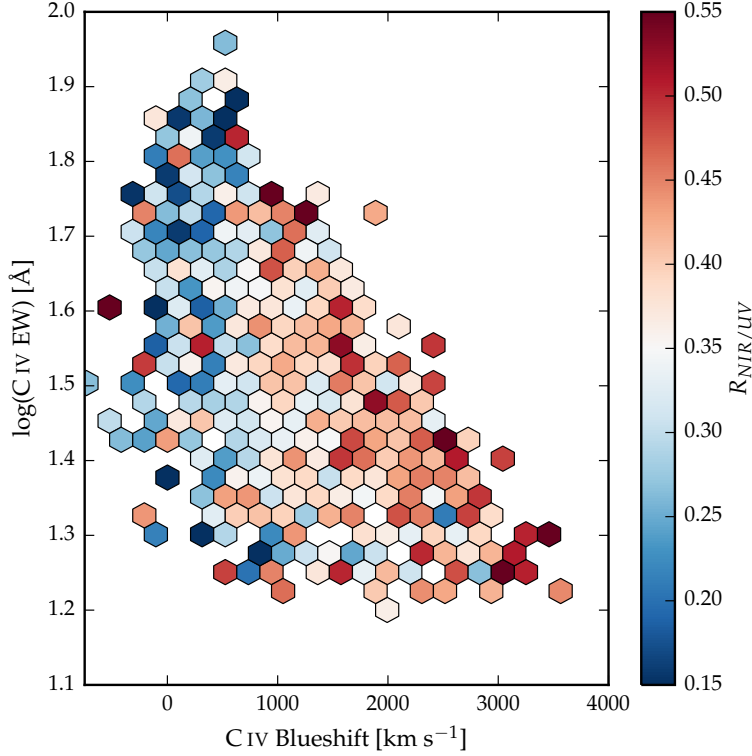
Figure 5.9: Rest-frame EQW and blueshift of the C IV line. The colours of the hexagons denote the median hot dust (T≃1200 K) abundance for all quasars at a given EQW and blueshift. Quasars with the most extreme outflow signatures are predominantly hot-dust rich.

to ultra-violet radiation from the inner accretion disc. Radiation pressure could efficiently accelerate the wind owing to the high cross-section of the dust grains (e.g. Fabian, 2012).

Radiation pressure could flatten the geometry of the wind. The greater the radiation pressure, the flatter the geometry of the wind. This leads to a wider opening angle, which exposes more surface area that is viewable on a relatively face-on line of sight. This leads to the enhanced hot dust emission we observe in the quasars with high accretion rates and strong outflows.

A prediction of this model is an anti-correlation between the torus covering factor and the hot dust abundance. The torus covering factor would be reduced by the accretion disc wind, and this would increase the maximum inclination at which a type I quasar could be seen. This would mean direct sight lines to more of the dust closest to the accretion disc. Such a correlation has been identified by Roseboom et al., (2013).

At lower redshifts, Shen and Ho, (2014) have found that the hot dust properties are correlated with EV1. 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 µm in the rest-frame of the quasar, which suggests that they are sensitive to the same component of hot dust which we are investigating. Shen and Ho, (2014) also find that torus emission is enhanced in quasars with larger $R_{FeII}$. Our work connecting the EV1 correlations observed in low-redshift AGN with the diversity of C IV emission properties in the high-redshift quasar population (Chapter 3) reveals a correlation between the C IV blueshift and the Fe II EQW. Therefore, our findings are consistent with Shen and Ho, (2014), in a different luminosity/redshift regime.

### 5.5.5 *Correlations with quasar properties*

The distribution of blackbody temperatures is relatively narrow ~ 200 K and so, as expected, we do not observe any correlations between the temperature and other quasar properties. Correlations between the hot dust abundance $R_{NIR/UV}$ and the ultra-violet luminosity, BH mass and Eddington ratio are shown in Figure 5.10.

The ultra-violet luminosity is measured at 1350 Å in the spectral modelling done by Shen et al., (2011). We do not observe any correlation between $R_{NIR/UV}$ and the ultra-violet luminosity (Figure 5.10a. However, the dynamic range in luminosity is small in this sample (~ 1 dex) because of the restricted redshift range.

In Figure 5.10b we show $R_{NIR/UV}$ as a function of the BH mass. Masses are C IV FWHM-based single-epoch virial estimates, as computed by Shen et al., (2011). $R_{NIR/UV}$ is positively correlated with the BH mass; the Spearman correlation coefficient, $\rho_S$, is 0.26. However, in the previous section, we found $R_{NIR/UV}$ to be correlated with the C IV blueshift and, in Chapter 3, we demonstrated that BH masses are severely overestimated in quasars with large C IV blueshifts. We therefore predict that the apparent correlation between $R_{NIR/UV}$ and the BH mass is due to systematic biases in the C IV-based masses. We can test this by comparing $R_{NIR/UV}$ with BH mass estimates which have been corrected using the prescription described in Chapter 3 (Figure 5.10d). As expected, the correlation vanishes ($\rho_S = 0.07$) A similar result is found when $R_{NIR/UV}$ is
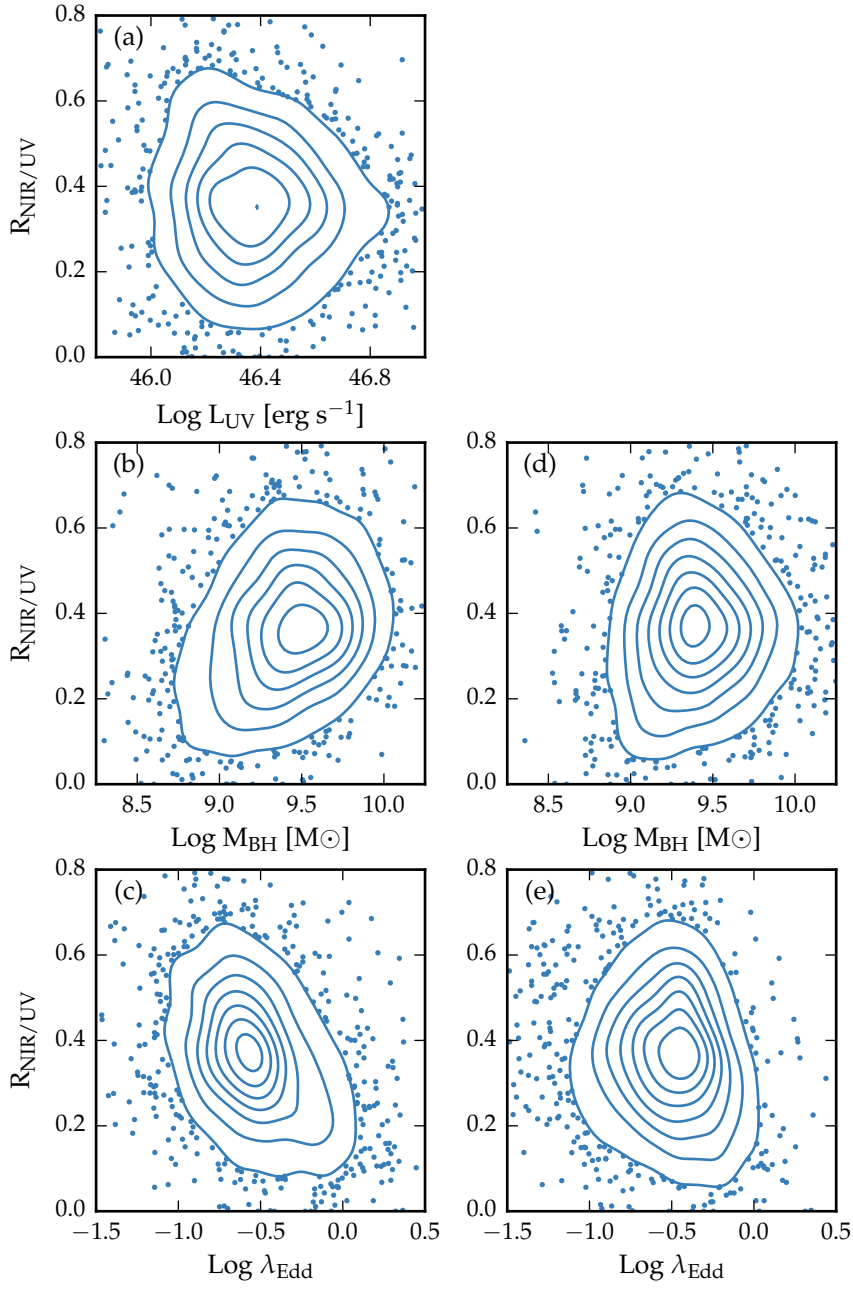
Figure 5.10: Best-fit near-infrared to ultra-violet luminosity ($R_{NIR/UV}$) as a function of ultra-violet luminosity, BH mass and Eddington ratio. In (b) and (c) BH mass estimates, which are based on the C IV FWHM, are taken from Shen et al., (2011). In (d) and (e) BH mass estimates have been corrected using the procedure described in Chapter 3. Using the corrected masses the correlations between $R_{NIR/UV}$ and the BH mass and Eddington ratio are significantly reduced.

compared to the Eddington ratio, which is inversely proportional to the BH mass. Using masses from Shen et al., (2011) an anti-correlation is observed between $R_{NIR/UV}$ and the Eddington ratio ($\rho_S = -0.36$; Figure 5.10c) but this disappears when corrected masses are employed ($\rho_S = -0.12$; Figure 5.10e). This demonstrates how using conventional BH mass estimates based on C IV can lead to spurious correlations with other quasar properties, but that this can be avoided using the improved mass estimates presented in Chapter 3.

# CONCLUSIONS / FUTURE WORK

We have explored relationships between BLR outflows, NLR outflows and hot dust emission.

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).

My research has focused on measuring fundamental properties of quasars and quasar-driven outflows in large surveys, with the ultimate goal of understanding the crucial role played by quasars in the evolution of galaxies. Building on this experience, I plan to (1) use data-driven methods to maximise the information content of spectroscopic survey data, (2) study galaxy-wide outflows traced by narrow [O III] emission, (3) use integral field unit (IFU) spectroscopy to measure the morphology and energetics of these outflows and their effect on star formation, and (4) use multi-wavelength spectral energy distribution (SED) modelling to study how these outflows are powered. My proposed projects, which are summarised below, would mesh effectively with the Galaxy Evolution and SAMI research programs.

In Coatman et al., (2017), I derived a direct empirical correction that maps the C IV line-width and blueshift (relative to the quasar rest-frame) to the Balmer emission-based black hole (BH) mass. I plan to extend this by developing a purely data-driven approach to inferring the BH mass from the rest-frame UV spectrum. Allen & Hewett (in preparation) have run an independent component analysis (ICA) on the SDSS quasar sample that compresses all of the information contained in the UV spectra into a small number of component weights. Tak-

ing the set of 230 objects with NIR spectra (and hence reliable Balmer-based BH masses), I will build a model that learns how these ICA component weights depend on the BH mass. After the training step, I can use the model to predict a BH mass (and the associated uncertainty) based only on the ICA component weights for the UV spectrum. There are numerous algorithms available to tackle this class of supervised learning problem (e.g. random forests).

Taking this one step further, I will build a data-driven model that predicts the unseen rest-frame optical spectrum from the UV region probed by SDSS spectra at redshifts $z \gtrsim 2$. This approach is inspired by the data-driven model for deriving stellar labels from spectroscopic data developed by **ness15** I will use an ICA to generate a set of 'labels' describing the optical spectrum. I will build a flexible generative model that predicts the flux at each wavelength in the SDSS rest-frame UV spectrum as a function of these weights. The coefficients in this model, which could be linear or a low-order polynomial, will be trained on the sub-sample of quasars with spectra covering the full optical to UV region. Once the coefficents have been determined, I will be able to statistically infer the optical ICA component weights based only on the UV spectrum. The optical spectrum can then straightforwardly be reconstructed from the ICA component weights and parameters of interest, such as the BH mass, can be calculated, without the need for follow-up NIR spectroscopy. I will also investigate whether an ICA decomposition (or any dimensionality reduction algorithm) can be used to infer the intrinsic quasar luminosity. The uncertainty on a single measurement would certainly be large but, with spectra for hundreds of thousands of quasars out to high redshifts, quasars could be used as standard candles to constrain cosmological parameters. I am keen to apply similar techniques to other spectroscopic surveys in which CAASTRO researchers are involved.

## 6.2 STUDYING FEEDBACK WITH IFU SPECTROSCOPY

Unfortunately, integrated spectra contain little information on the spatial extent or geometry of the outflowing gas. To study the morphology and energetics of the outflow, we must turn to spatially-resolved IFU spectroscopy. IFU spectroscopy has revealed outflowing ionised gas on scales of ~ 10kpc to be common in the hosts of quasars at low redshifts (e.g. Harrison et

al., 2014). Extended outflows have also been detected at high redshifts ($z \gtrsim 2$) in a handful of objects (e.g. Carniani et al., 2015).

Currently, I am analysing the spatially-resolved kinematics of the ionised gas in the hosts of ~120 luminous quasars at redshifts $z \sim 2$ using (mostly archival) SINFONI IFU spectra. I have already analysed the spatially-integrated spectra and in many a broad, blueshifted wing in the [O III] emission is clearly identifiable. An ICA on a sample of low-redshift SDSS quasar spectra has been used to generate individual components corresponding to the [O III] emission originating in static and outflowing gas. I can use the relative weights of these two components in each spaxel to generate velocity maps and compare the spatial distributions of the static and outflowing components. I can then test whether the outflowing component in the [O III] emission extends to ~kpc scales and estimate properties of the ionised outflows: the mass outflow rate, momentum rate, and kinetic power. I will also search for narrow Hα emission associated with star formation, using the relative strength of [N II] to discriminate between a star formation and AGN origin of the narrow Hα emission (**susie16**). I will look for anti-correlations in the spatial distributions of outflowing gas and star forming regions, which would be indicative of negative quasar feedback in action.

These observations will be challenging and, given the limited depth of our data, the [O III] emission is unlikely to be spatially resolved in every object. Therefore, I would like to observe a representative sub-set of the large sample presented in Coatman et al. (2017b, in prep.) using SINFONI or KMOS on the VLT. Because I have already carefully analysed the [O III] emission properties of these quasars, it is an ideal sample from which to select targets for IFU spectroscopy. One particular goal is to understand the disappearance of the narrow line region in quasars with strong broad line region outflows (Fig. 4.21).

The majority of the un-obscured quasars in this sample lack optical spectra. Therefore, to exploit the rich information contained in the rest-frame UV spectrum (e.g. the BLR outflow properties) I propose to observe these targets. The sample is relative bright ($b_J \sim 17.5$) and so this could easily be achieved on a moderately-sized telescope. I will then be able to extend my study of the relationship between outflows in the broad and narrow line regions.