

# BLACK HOLE MASSES

## 1.1 INTRODUCTION

The goal of better understanding the origin of the correlation between the masses of super-massive black holes (BHs) and the masses of host-galaxy spheroids has led to much work focussing on the properties of quasars and active galactic nuclei (AGN) at relatively high redshifts,  $z \gtrsim 2$ . Extensive reverberation-mapping campaigns have been used to calibrate single-epoch virial-mass estimates which use the velocity widths of the hydrogen Balmer emission lines and the nuclear continuum luminosity to provide reliable BH masses. Single-epoch virial BH mass estimates using  $H\beta$  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-ultraviolet 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 the Sloan Digital Sky Survey (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 ultraviolet luminosity is not established definitively when extrapolating to high-redshifts and luminosities. While some authors have found good agreement between BH mass-estimates based on  $C\ IV$  and  $H\beta$  (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

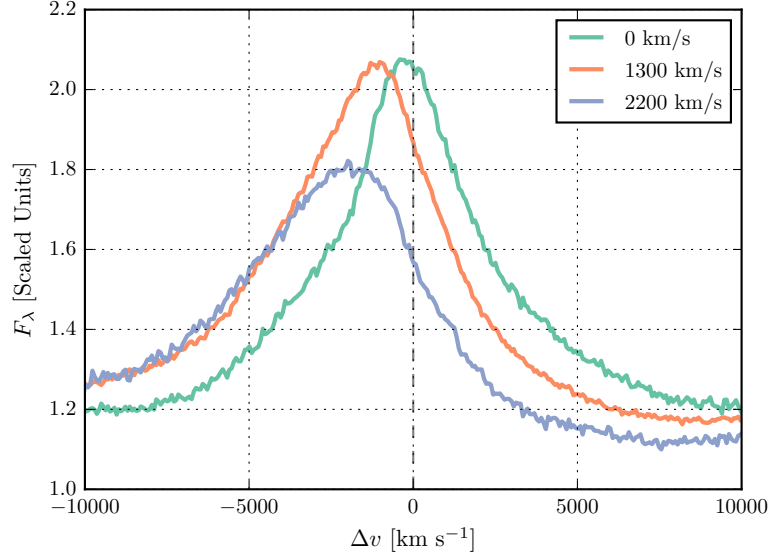


Figure 1.1: Composite spectra of the C IV-emission line as a function of C IV blueshift for SDSS DR7 quasars. Quasars classified as BALs, or possessing strong associated absorbers have been excluded, and the composite-spectra shown are derived using an arithmetic mean of a minimum of 200 spectra at each 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  $\text{km s}^{-1}$  and EW 33.1, 31.6, and 28.8 Å.

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

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

full-width at half maximum (FWHM) and the line dispersion ( $\sigma$ ). 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; Gagnaud et al., 2008; Vestergaard et al., 2008; Vestergaard and Osmer, 2009; Kelly et al., 2010; Kelly and Shen, 2013). Figure 1.1 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 ( $\gtrsim 2000 \text{ km s}^{-1}$ ) the C iv-profile is displaced to the blue by amounts comparable to the FWHM of the profile. This indicates that non-virial motions, very likely due to outflows, are having a significant effect on the observed C iv emission velocity profile (e.g. Gaskell, 1982; Baskin and Laor, 2005a; Sulentic et al., 2007; Richards et al., 2011; Wang et al., 2013). At fixed emission-line EW, virtually the entire C iv-profile appears to shift blueward and the change in line shape is not simply an enhancement of flux in the blue wing of a still identifiable symmetric component. While gravity almost certainly plays a key role, determining the escape velocity for out-flowing material for example, it is clear that the virial assumption, on which single-epoch BH-mass measurements are predicated, is not straightforwardly applicable for the C iv-emission line in quasars exhibiting large blueshifts. As a consequence, BH-masses derived from C iv emission line velocity-widths are systematically biased compared to masses from the Balmer lines (e.g. Shen et al., 2008; Shen and Liu, 2012; Coatman et al., 2016).

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

In recent literature, attempts have been made to minimise the influence of the systematic non-virial contribution to the C iv emission on estimates of the BH mass. Strategies include (i) significantly reducing the dependence of the derived masses on the emission-line velocity width (e.g. from the  $V^2$  dependence predicted assuming a virialized broad line region to just  $V^{0.56}$  in Park et al. 2013; see also Shen and Liu 2012), (ii) adopting a measure of emission-line velocity-width that is relatively insensitive to changes in the core of the emission-line profile (e.g. Denney et al., 2013) and (iii) estimating the amplitude of the non-virial contribution to the C iv emission-line via comparison with other ultraviolet emission lines (e.g. Si iv+O iv  $\lambda 1400$  in Runnoe et al. 2013a and Brotherton et al. 2015). The increased number of quasars

with high-quality spectra that cover both the observed-frame optical (where the redshifted C IV appears) and near-infrared (where 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.

## 1.2 QUASAR SAMPLE

We have compiled a sample of 307 quasars at redshifts  $1.5 < z < 4$  with both optical and near-infrared spectra. Reliable emission line properties were measured for 230 quasars (Section 1.3.5), 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 1.5.3, the quasars have C IV blueshifts of up to  $\sim 5000 \text{ km s}^{-1}$ , and span the full range observed in the population.

### 1.2.1 Near-infrared data

The near-infrared data has been described in Chapter ?? and the telescopes/spectrographs used are summarised in Table 2.1. We have sub-divided our sample into two overlapping groups: quasars with reliable H $\alpha$  line measurements (the ‘H $\alpha$  sample’) and quasars with reliable H $\beta$  measurements (the ‘H $\beta$  sample’).

### 1.2.2 Optical data

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

Table 1.1: The numbers of quasars with reliable  $H\alpha$  and  $H\beta$  line measurements, and the spectrographs and telescopes used to obtain the near-infrared spectra

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

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

### 1.3 SPECTRAL MEASUREMENTS

Conventionally, single-epoch virial estimates of the BH mass are a function of the line-of-sight velocity width of a broad emission line and the quasar luminosity. The velocity width is a proxy for the virial velocity in the broad line region (BLR) and, as revealed in reverberation-mapping studies, the luminosity is a proxy for the typical size of the BLR (the  $R - L$  relation; e.g. Kaspi et al., 2000; Kaspi et al., 2007). Most reverberation mapping campaigns have employed  $H\beta$  time-lags and velocity widths, but the line-widths of  $H\alpha$  and  $\text{Mg II } \lambda 2800$  have been shown to yield consistent BH masses (e.g. McLure and Jarvis, 2002; Greene and Ho, 2005; Onken and Kollmeier, 2008; Shen et al., 2008; Wang et al., 2009; Rafiee and Hall, 2011; Mejía-Restrepo et al., 2016). In Section 1.4.1 we verify that the  $H\alpha$  and  $H\beta$  line-widths yield consistent BH for the 99 quasars in our sample with measurements of both.

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

Balmer emission velocity-widths in the same quasars. Our process is as follows. First, an accurate measure of the quasar’s systemic redshift is required, for which we adopt the centre of the Balmer emission, where the centre,  $\lambda_{\text{half}}$ , 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 1.4.2). 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  $\text{km s}^{-1}$ ) is defined as  $c \times (1549.48 - \lambda_{\text{half}}) / 1549.48$  where  $c$  is the velocity of light and  $1549.48 \text{ \AA}$  is the rest-frame wavelength for the C IV doublet<sup>1</sup>. Positive blueshift values indicate an excess of emitting material moving towards the observer and hence out-flowing from the quasar.

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

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

### 1.3.1 C IV

We first define a power-law continuum,  $f(\lambda) \propto \lambda^{-\alpha}$ , with the slope,  $\alpha$ , determined using the median values of the flux in two continuum windows at 1445-1465 and 1700-1705  $\text{\AA}$ . 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-

<sup>1</sup> 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  $\sim 80 \text{ km s}^{-1}$ .

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

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

We find that using the commonly employed three-Gaussian component model, rather than the GH polynomials, resulted in only marginal differences in the line parameters. Our best-fit parameters are also in good agreement with Shen et al., (2011), who employ a multi-Gaussian parametrization. In Fig. 1.2 we compare our measurements of the C IV FWHM from the 71 SDSS DR7 spectra in our sample with the measurements published in Shen et al., (2011). There is a very strong agreement between our measurements, with a scatter of 0.05 dex ( $200\text{ km s}^{-1}$ ).

### 1.3.2 $H\alpha$

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  $H\alpha$  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  $H\beta$  and  $H\alpha$  frequently include a significant narrow component from the physically more extended narrow line region (NLR). Additional Gaussian components were included in our parametric model to fit the narrow component of  $H\alpha$  as well as [N II]  $\lambda\lambda 6548, 6584$  and [S II]  $\lambda\lambda 6717, 6731$ . 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



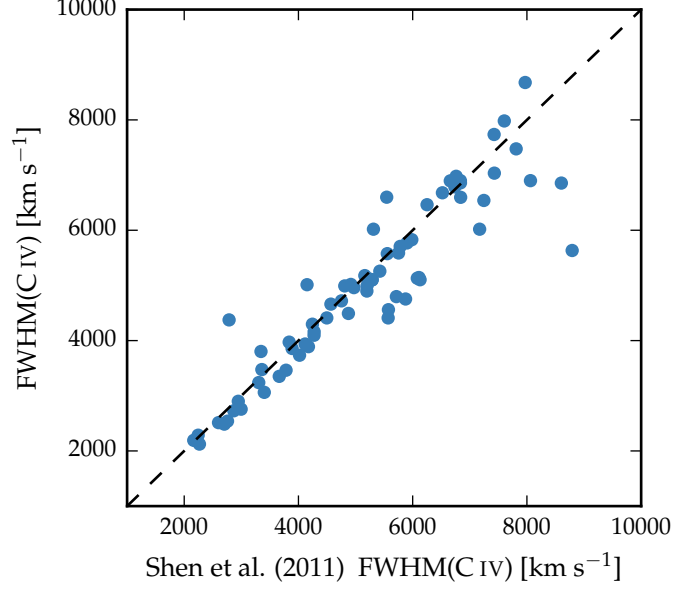


Figure 1.2: Demonstration of the effectiveness of our line parameter estimation scheme via a comparison of the C iv FWHM with Shen et al., (2011).

amplitudes of all components must be non-negative. The relative flux ratio of the two [N II] components is also fixed at the expected value of 2.96. In 70 per cent of the spectra the [O III]  $\lambda\lambda 4960, 5008$  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 Fig. 1.3 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 just 0.07 dex.



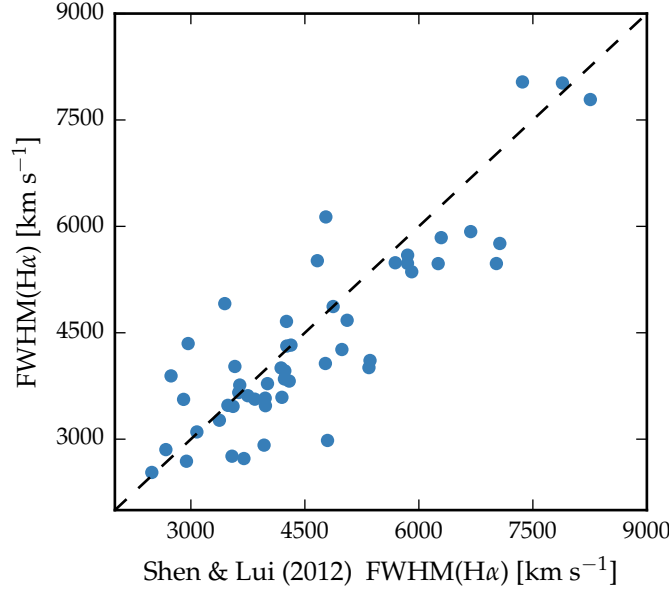


Figure 1.3: Demonstration of the effectiveness of our line parameter estimation scheme via a comparison of the  $H\alpha$  FWHM with Shen and Liu, (2012).

### 1.3.3 $H\beta$ and $[O\text{ III}]$

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  $[O\text{ III}]$  doublet is fit with two Gaussians, to model both the systemic and any outflow contributions. The peak flux ratio of the  $[O\text{ III}]$  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  $1200\text{ km s}^{-1}$  is placed on the FWHM.

The parametric model we fit to the  $H\beta/[O\text{ III}]$  emission region was very similar to the model employed by Shen, (2016). In Fig. 1.4 we plot our  $H\beta$  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 0.04 dex.

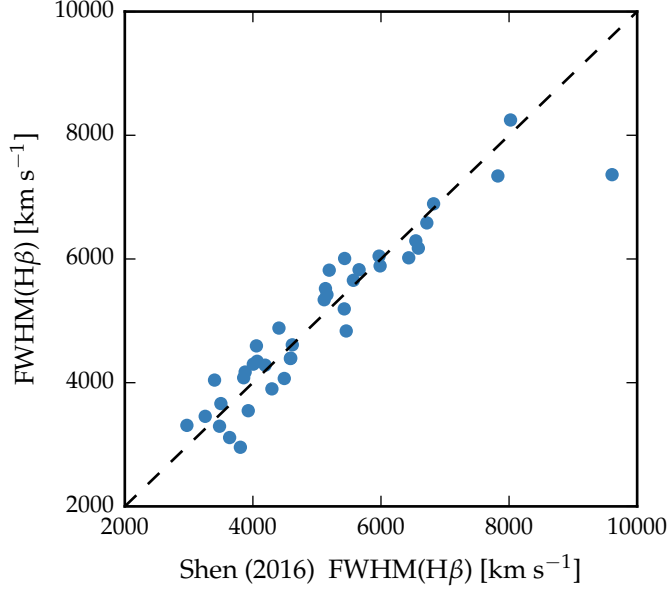


Figure 1.4: Demonstration of the effectiveness of our line parameter estimation scheme via a comparison of the  $H\beta$  FWHM with Shen, (2016).

#### 1.3.4 Fitting procedure

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

In Fig. 1.5 we present our parametric fits to the C IV,  $H\alpha$  and  $H\beta$  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  $H\alpha$  emission line centroid is at  $0 \text{ km s}^{-1}$ . The y-axes of the data-minus-model residual plots have been scaled by the spectrum flux errors. The mean reduced chi-squared values in our  $H\alpha$ ,  $H\beta$  and C IV fits are 1.69, 1.62, and 1.77 respectively and, in general, there are no strong features observable in the spectrum minus model residuals. The only significant features seen in the residual C IV spectra correspond to the location of narrow absorption lines which were excluded in the fitting procedure.

Table 1.2 includes the line parameters of our best-fitting model for each line. The reported line-width measures are corrected for instrumental broadening by subtracting the resolution of the spectrograph in quadrature. The spectrograph resolutions, which we estimate from the line widths in the observed sky spectra, range from  $25 \text{ km s}^{-1}$  for XSHOOTER to  $477 \text{ km s}^{-1}$  for the low-resolution LIRIS grism and are therefore small relative to the quasar broad line widths.

Table 1.2: The format of the table containing the emission line properties from our parametric model fits. The table is available in machine-readable form in the online version of Coatman et al., (2017).

	Units	Description
NAME		Catalogue name
FWHM_BROAD_HA	$\text{km s}^{-1}$	FWHM of broad $\text{H}\alpha$ line
FWHM_BROAD_HA_ERR	$\text{km s}^{-1}$	
SIGMA_BROAD_HA	$\text{km s}^{-1}$	Dispersion of broad $\text{H}\alpha$ line
SIGMA_BROAD_HA_ERR	$\text{km s}^{-1}$	
Z_BROAD_HA		Redshift from broad $\text{H}\alpha$ line
FWHM_BROAD_HB	$\text{km s}^{-1}$	FWHM of broad $\text{H}\beta$ line
FWHM_BROAD_HB_ERR	$\text{km s}^{-1}$	
SIGMA_BROAD_HB	$\text{km s}^{-1}$	Dispersion of broad $\text{H}\beta$ line
SIGMA_BROAD_HB_ERR	$\text{km s}^{-1}$	
Z_BROAD_HB		Redshift from broad $\text{H}\beta$ line
FWHM_CIV	$\text{km s}^{-1}$	FWHM of C iv doublet
FWHM_CIV_ERR	$\text{km s}^{-1}$	
SIGMA_CIV	$\text{km s}^{-1}$	Dispersion of C iv doublet
SIGMA_CIV_ERR	$\text{km s}^{-1}$	
BLUESHIFT_CIV_HA	$\text{km s}^{-1}$	Blueshift of C iv relative to $\text{H}\alpha$
BLUESHIFT_CIV_HA_ERR	$\text{km s}^{-1}$	
BLUESHIFT_CIV_HB	$\text{km s}^{-1}$	Blueshift of C iv relative to $\text{H}\beta$
BLUESHIFT_CIV_HB_ERR	$\text{km s}^{-1}$	
LOGL <sub>5100</sub>	$\text{erg s}^{-1}$	Luminosity at $5100\text{\AA}$
LOGL <sub>1350</sub>	$\text{erg s}^{-1}$	Luminosity at $1350\text{\AA}$

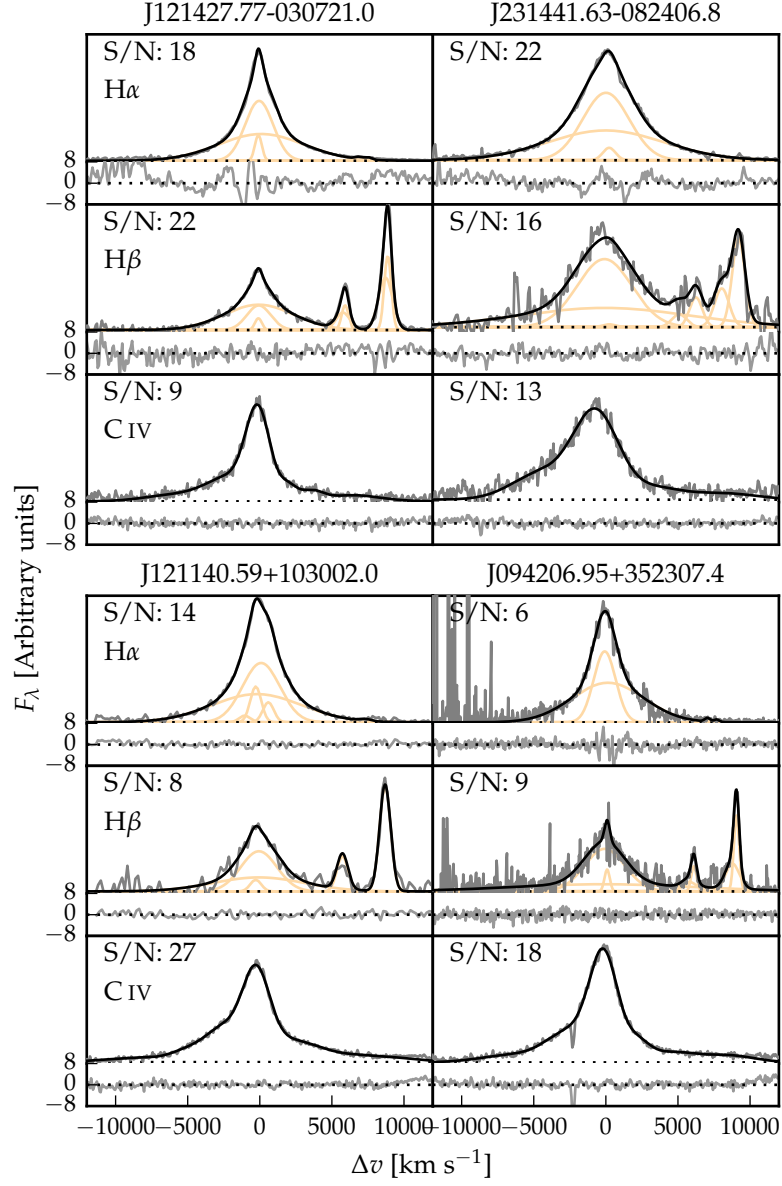


Figure 1.5: Model fits to continuum-subtracted H $\alpha$ , H $\beta$ , and C IV emission in four quasars, chosen to represent the range of S/N (indicated in the figure and given per  $150\text{km s}^{-1}$  pixel in the continuum) and line shapes present in the catalogue. The data is shown in grey, the best-fitting parametric model in black, and the individual model components in orange. The centroid of the broad H $\alpha$  emission is used to set the redshift, and  $\Delta v$  is the velocity shift from the line rest-frame transition wavelength. Below each fit we plot the data minus model residuals, scaled by the errors on the fluxes.

Table 1.3: The number of spectra removed from our sample by the cuts described in Section 1.3.5.

		H $\alpha$ sample	H $\beta$ sample
Total		194	279
H $\alpha$ /H $\beta$	Wavelength	6	27
	S/N	8	83
C IV	Wavelength	6	5
	S/N	4	12
	Absorption	6	8
Total remaining		164	144

### 1.3.5 Spectra removed from sample

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

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

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

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

The number of spectra that are removed by each cut is given in Table 1.3 and the distribution in redshift and luminosity is shown in Fig. 1.6. 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-band. 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.

### 1.3.6 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 sub-

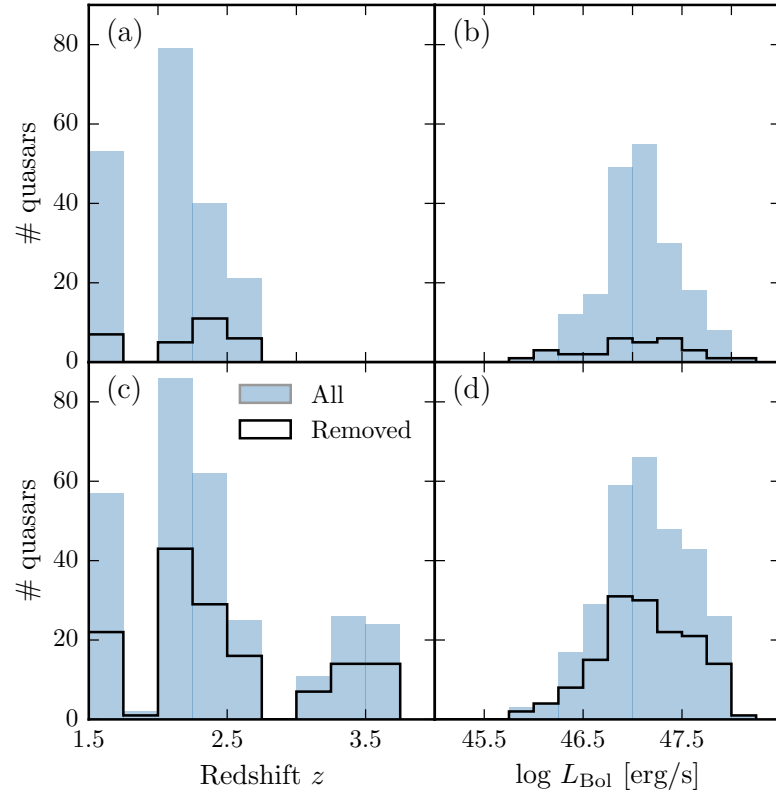


Figure 1.6: The redshift and luminosity distributions of the spectra removed from our  $\text{H}\alpha/\text{C iv}$  (a, b) and  $\text{H}\beta/\text{C iv}$  (c, d) samples.

traction procedure. To calculate more realistic uncertainties on our fitted variables we employed a Monte Carlo approach. One thousand artificial spectra were synthesised, with the flux at each wavelength drawn from a Normal distribution (mean equal to the measured flux and standard deviation equal to the known error). Our emission-line fitting recipe was then implemented on each of these mock spectra. The uncertainty in each parameter is given by the spread in the best-fitting values from the one thousand realisations of the fitting routine. In some cases the standard deviation of the parameter distribution was biased by extreme values caused by bad fits<sup>2</sup>. 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.

### 1.3.7 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  $\sim 14$  years after the SDSS spectra, and the VLT SINFONI spectra 20 years or more after the Hamburg/ESO observations<sup>3</sup>. 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  $\sim 8$  years. The root-mean-square difference in the C iv FWHM measured from the BOSS and SDSS spectra is a modest  $\simeq 500 \text{ km s}^{-1}$ . 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  $500 \text{ km s}^{-1}$ . We conclude therefore that any intrinsic changes with time do not materially affect the emission line measurements.

### 1.3.8 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

<sup>2</sup> In the analysis of the real spectra such fits are identified via visual inspection.

<sup>3</sup> Time differences in the quasar rest-frame are reduced by a factor of  $(1+z)$ .



BLR. The monochromatic continuum flux is generally measured at  $1350 \text{ \AA}$  for C IV and  $5100 \text{ \AA}$  for H $\alpha$  and H $\beta$ .

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

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

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

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

Comparison of the 5100 Å luminosity, computed using the photometry- and spectrum-based methods for 177 quasars, showed a scatter 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 ultraviolet 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. 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.

### 1.3.9 *Characterising the emission-line widths*

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

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

As such information is not yet available, there is a strong rationale for investigating whether the systematic changes in the C iv-emission line profile can be used to improve the single-epoch BH-mass estimates derived using the C iv line. In Fig. 1.7 we show how the C iv FWHM, line dispersion,  $\sigma$ , and line shape, FWHM/ $\sigma$ , vary as a function of the blueshift. The C iv FWHM is correlated with the blueshift, with the median FWHM of quasars with the largest blueshifts a factor

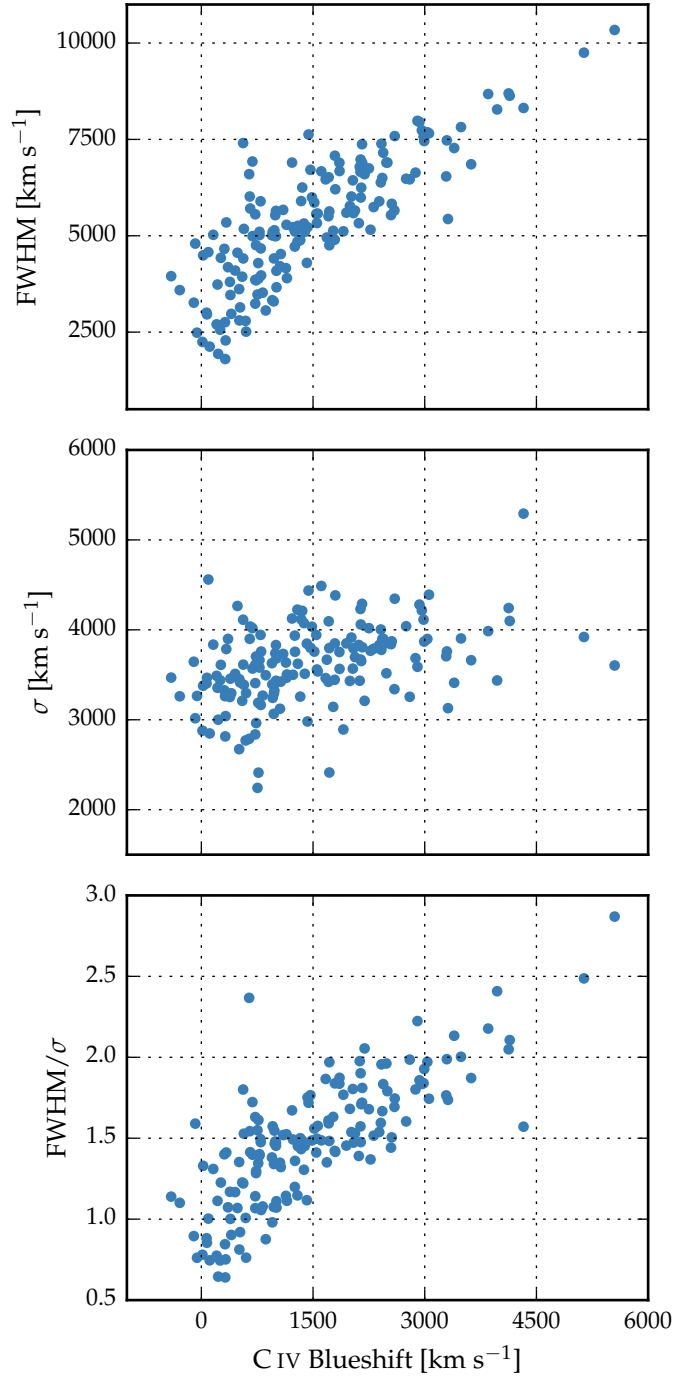


Figure 1.7: The FWHM, dispersion ( $\sigma$ ) and shape ( $\text{FWHM}/\sigma$ ) of C IV as a function of the C IV blueshift.

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

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

To emphasise this point, in Fig. 1.8 we overlay the C iv line profiles of SDSSJ1236+1129 and SDSSJ1525+2928, whose dispersions are indistinguishable ( $4168 \pm 271$  and  $4303 \pm 128 \text{ km s}^{-1}$  respectively). Notwithstanding the very similar dispersion values, the emission-line velocity fields differ dramatically and, therefore, the dispersion values cannot be measuring accurately the virial-induced velocity spread of the C iv emission in both quasars.

The analysis here, building on earlier work (including Shen and Liu, 2012; Sulentic et al., 2007), confirms a link between C iv emission-line shape and blueshift, raising the prospect of developing a blueshift-dependent correction to single-epoch BH-mass estimates based on the C iv line. Expressed in another way, we are interested in testing if the significant systematic change in line shape as a function of C iv blueshift can be used to provide improved single-epoch BH-masses from the C iv emission line. The tightness of the correlation we observe between the C iv FWHM and blueshift implies that such an approach may be more effective than using the C iv emission-line velocity dispersion without reference to blueshifts. A further practical advantage is that, given the typical S/N of current survey-quality spectra, virial BH mass estimates for high-redshift quasars are usually based on the FWHM rather than the dispersion (e.g. Shen et al., 2011), which, being strongly affected by the continuum placement, is often found to be difficult to measure robustly (e.g. Mejía-Restrepo et al., 2016).

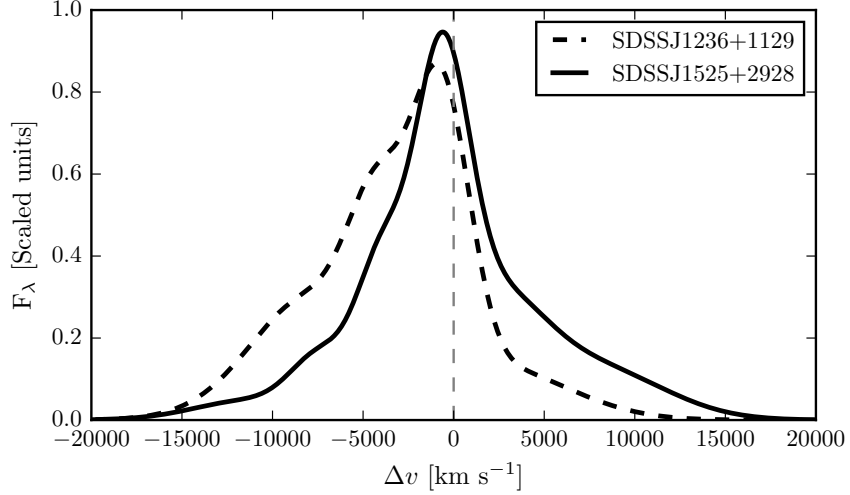


Figure 1.8: Comparison of the C IV line profiles of SDSSJ1236+1129 and SDSSJ1525+0426. Notwithstanding the essentially identical dispersion values, the emission-line velocity fields differ dramatically and, therefore, the dispersion values cannot be measuring accurately the virial-induced velocity spread of the C IV emission in both quasars.

#### 1.4 AN EMPIRICAL CORRECTION TO CIV-BASED VIRIAL BH-MASS ESTIMATES

##### 1.4.1 $H\alpha/H\beta$ FWHM comparison

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

In our sample, we have 99 quasars with reliable measurements of both  $H\alpha$  and  $H\beta$  lines. The 99 objects include 21 quasars which were excluded from the main 308-object catalogue because the C IV FWHM and/or blueshift could not be measured reliably. The line widths are compared in Fig. 1.9 and, as expected, a tight correlation is observed. Greene and Ho, (2005), using a sample of 162 quasars with high S/N SDSS spectra at  $z < 0.35$ , established the following relation between the  $H\alpha$  and  $H\beta$  FWHMs

$$\text{FWHM}(H\beta) = (1.07 \pm 0.07) \times 10^3 \left( \frac{\text{FWHM}(H\alpha)}{10^3 \text{ km s}^{-1}} \right)^{(1.03 \pm 0.03)} \quad (1.1)$$

The relation is shown as the dashed line in Fig. 1.9. The root-mean-square scatter about this relation is 0.07 dex, compared to the  $\sim 0.1$  dex

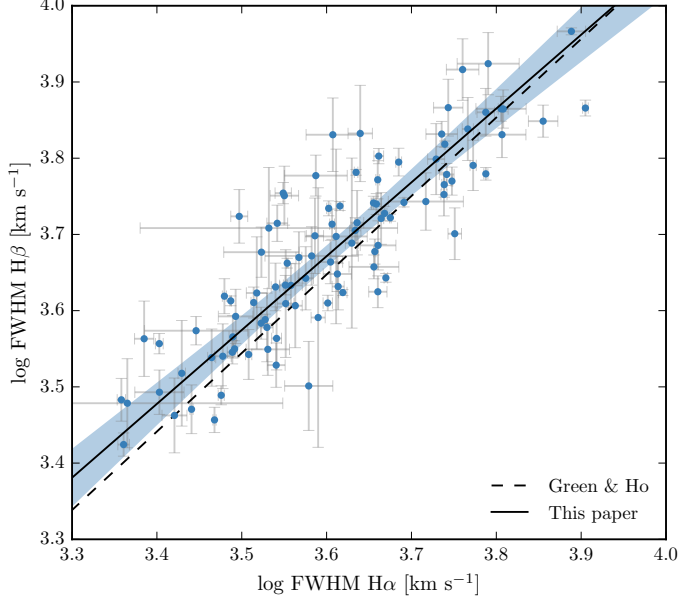


Figure 1.9: Comparison of H $\alpha$  and H $\beta$  FWHM measurements for 99 quasars. The solid line is our best-fitting power-law model, and the blue-shaded region shows the 2- $\sigma$  uncertainties on the model parameters. The dashed line is the relation found by Greene and Ho, (2005) using a sample of  $z < 0.35$  SDSS AGN.

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

We assume a relation of the same form used by Greene and Ho, (2005), 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, of variance  $\Sigma_i^2$ , with a Gaussian of variance  $V$  representing the intrinsic variance in the data. The log-likelihood is then given by

$$\ln \mathcal{L} = - \sum_{i=1}^N \frac{1}{2} \ln [2\pi (\Sigma_i^2 + V)] - \sum_{i=1}^N \frac{\Delta_i^2}{2[\Sigma_i^2 + V]} \quad (1.2)$$

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 reason-

able assumption that there is an intrinsic scatter in the relationship between the variables that is independent of the measurement errors. Following the suggestion by Hogg, Bovy, and Lang, (2010), the linear model was parametrized in terms of  $(\theta, b_{\perp})$ , where  $\theta$  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 Fig. 1.10. The solid line in Fig. 1.9 is the maximum likelihood solution

$$\text{FWHM}(\text{H}\beta) = (1.23 \pm 0.10) \times 10^3 \left( \frac{\text{FWHM}(\text{H}\alpha)}{10^3 \text{km s}^{-1}} \right)^{0.97 \pm 0.05} \quad (1.3)$$

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

As discussed above, our relation is displaced to slightly higher  $\text{H}\beta$  FWHM than the Greene and Ho, (2005) relation – the offset is  $210 \text{km s}^{-1}$  for a quasar with  $\text{H}\alpha$  FWHM  $4500 \text{km s}^{-1}$ . We infer a power-law index that, although slightly shallower, is consistent with the Greene and Ho, (2005) index within the quoted uncertainties. The intrinsic scatter in the data,  $\sigma_1$ , we infer from the fit is 0.04 dex. This is smaller than the total scatter seen in Fig. 1.9 (0.06 dex), which suggests that measurement errors make a significant contribution to the total scatter in the relation.

We constructed composite spectra of the  $\text{H}\alpha$  and  $\text{H}\beta$  regions from 217 and 171 quasars respectively. Spectra were first de-redshifted to the quasar rest-frame, and then interpolated on to a common wavelength grid with a  $1\text{\AA}$  resolution. The spectra were scaled by the mean flux in the interval  $4700\text{--}5100\text{\AA}$  ( $\text{H}\beta$ ) and  $6400\text{--}6800\text{\AA}$  ( $\text{H}\alpha$ ). The composite was then defined as the median flux from all of the normalised spectra in each wavelength bin. The  $\text{H}\alpha$  and  $\text{H}\beta$  lines in the composite spectra are shown in Fig. 1.11. The cores of the two lines are very similar, but  $\text{H}\beta$  has more flux in the wings of the line.

For 19 of the 99 quasars with  $\text{H}\beta$  and  $\text{H}\alpha$  emission profiles, one of the two Gaussians used to reproduce the  $\text{H}\beta$  profiles has a FWHM greater than  $20000 \text{km s}^{-1}$  and a fractional contribution to the total  $\text{H}\beta$  broad line flux of  $>0.3$  (Marziani et al., 2009; Marziani et al., 2013). Such a broad component is not seen in the  $\text{H}\alpha$  profiles and the very broad  $\text{H}\beta$ -component may be an artifact of the fitting scheme. A particular issue for  $\text{H}\beta$  is the presence of  $\text{Fe II}$  emission, often at a significant level. Furthermore, additional lines could be contributing to the underlying continuum (e.g. the  $\text{He I}\lambda\lambda 4922, 5017$  doublet; Véron, Gonçalves, and Véron-Cetty, 2002; Zamfir et al., 2010).



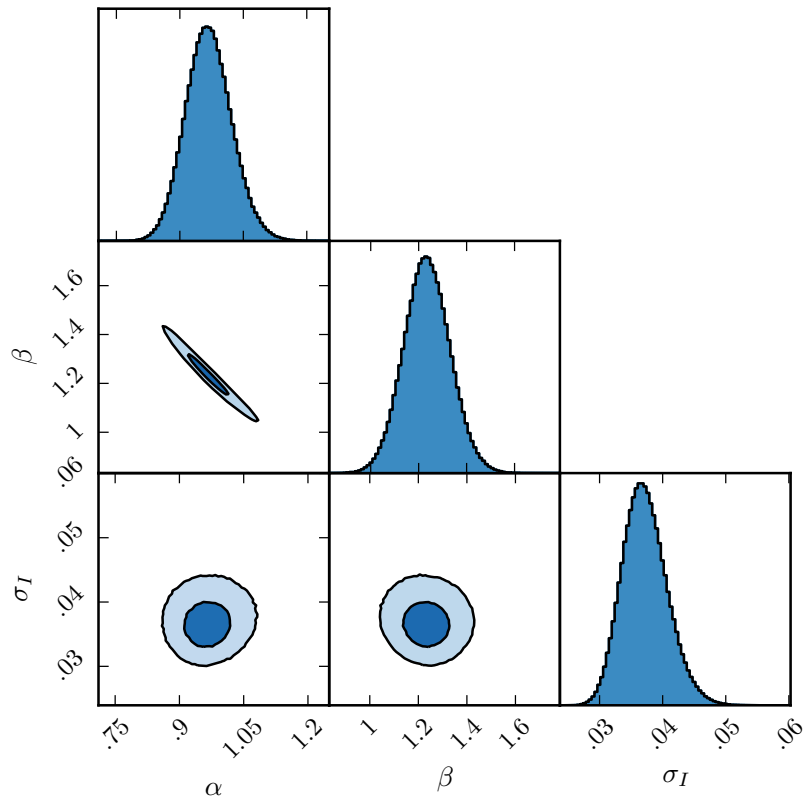


Figure 1.10: One- and two-dimensional projections of the MCMC sampling of the posterior distribution from the fit in Fig. 1.9.  $\alpha$  is the power-law index,  $10^\beta$  is the normalisation, and  $\sigma_I$  is the intrinsic scatter. In the two-dimensional projections, 1- and 2- $\sigma$  contours are shown.

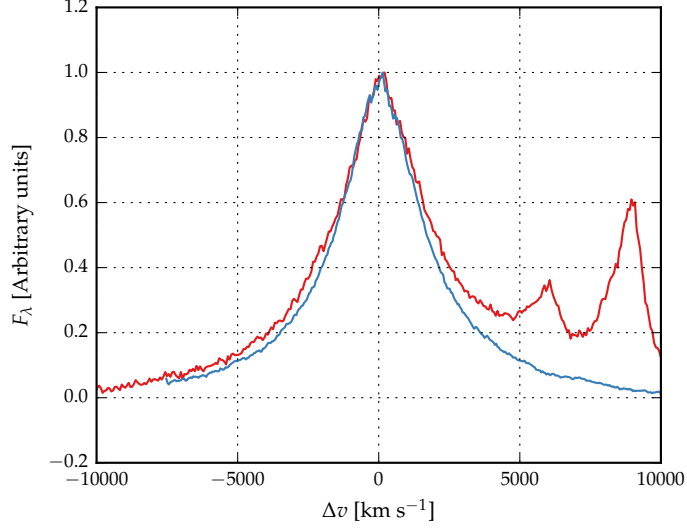


Figure 1.11: The H $\alpha$  (blue) and H $\beta$  (red) emission line regions in the median composite spectrum, shown as function of the velocity shift from the respective predicted line peak wavelengths. The background continuum and optical Fe II emission has been modelled and subtracted. The line fluxes have been scaled in order for the profile shapes to be readily compared.

In Sec. 1.4.3 we use the whole of the H $\beta$  profile to derive an un-biased BH mass. If, instead, the FWHM is calculated from the narrower of the two Gaussian components rather than the composite profile, then the H $\beta$  FWHM decreases by  $630 \text{ km s}^{-1}$  on average. This effectively removes the average offset to broad H $\beta$  profiles evident in Fig. 1.9. This will enhance the C IV FWHM relative to the H $\alpha$ /H $\beta$  FWHM by  $\sim 15$  per cent and increase the size of the correction which must be applied to the C IV-based BH masses by  $\sim 30$  per cent.

#### 1.4.2 Measuring the quasar systemic redshift

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

For 62 and 86 quasars in the H $\alpha$  and H $\beta$  samples respectively narrow [O III] emission is also detected with sufficient S/N to measure the line properties. In the model fit to the H $\beta$  region the velocity centroids of the broad H $\beta$ -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 H $\alpha$  and H $\beta$  emission and the narrow [O III] emission to have a dispersion of 300 and 400  $\text{km s}^{-1}$ , 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  $250 \text{ km s}^{-1}$  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  $L_{\text{Bol}} \sim 10^{44} \text{ erg s}^{-1}$ , the centroids of the broad component of H $\beta$  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 1.3, the broad components of H $\alpha$  and H $\beta$  were modelled with up to two Gaussians, with identical velocity centroids. If there is any significant asymmetry in these lines then the emission will be poorly fit by our model and the redshift derived from the peak of the best-fitting model could be biased. To investigate this possibility we relaxed the requirement for the centroids of the two broad Gaussians to be the same, and measured the systemic redshift from the peak of the composite profile. With this new, more flexible model, the mean absolute difference between the centroids of the two Gaussian components used to model H $\alpha$  and H $\beta$  was 480 and 780  $\text{km s}^{-1}$  respectively. With these adjustments, we found the mean difference between the [O III]- and H $\alpha$ (H $\beta$ ) based redshift estimates to be -100(-120)  $\text{km s}^{-1}$  and the scatter to be 290(320)  $\text{km s}^{-1}$ . Therefore, the shift between the Balmer and [O III] velocities is reduced, suggesting that there might be a  $\sim 100 \text{ km s}^{-1}$  systematic bias in our measurements of the quasar systemic redshift. Regardless, since both the systematic offset and the scatter are small in comparison to the dynamic range in C IV blueshifts, the blueshift-based empirical correction we will derive does not depend on whether the broad Balmer emission or the [O III] centroid is used to define the systemic redshift, or how the broad Balmer emission is parameterized.

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

#### 1.4.3 Balmer/C IV line widths as a function of C IV-blueshift

In this section we directly compare the C IV and H $\alpha$ /H $\beta$  line widths as a function of the C IV blueshift. Because virial BH mass estimates are generally based on the H $\beta$  FWHM, we first convert our H $\alpha$  FWHM measurements to equivalent H $\beta$  FWHM using Eq. 1.3. In Figs. 1.12 and 1.13 we show the C IV FWHM relative to both the (H $\beta$ -scaled) H $\alpha$  FWHM and the H $\beta$  FWHM, as a function of the C IV blueshift.

Employing the same Bayesian fitting framework described in Section 1.4.1, we fit independent linear models to the C IV FWHM relative to the H $\alpha$  and H $\beta$  FWHM as a function of the C IV blueshift. As before, our model has an additional parameter representing any intrinsic scatter in the relationship between the variables which is independent of measurement errors. We also tested a model where some fraction of the data points (which is free to vary) are drawn from an outlier distribution, represented by a broad Gaussian centered on the mean of the data. We found, however, that the inferred outlier fraction was very low (0.004, corresponding to  $\sim 0.7$  data points) and so did not include such a component in our model.

In Fig. 1.14 we show the one- and two-dimensional projections of the posterior distribution from the linear fit to the FWHM C IV/H $\alpha$  ratio. The projections from the FWHM C IV/H $\beta$  fit, which we do not show, have very similar appearances. In Fig. 1.12 we plot the maximum likelihood model and the  $2\sigma$  uncertainties on the model parameters. The maximum likelihood line is given by

$$\text{FWHM}(\text{C IV}, \text{Corr.}) = \frac{\text{FWHM}(\text{C IV}, \text{Meas.})}{(0.41 \pm 0.02) \left( \frac{\text{C IV Blueshift}}{10^3 \text{ km s}^{-1}} \right) + (0.62 \pm 0.04)} \quad (1.4)$$

for the C IV/H $\alpha$  fit and

$$\text{FWHM}(\text{C IV}, \text{Corr.}) = \frac{\text{FWHM}(\text{C IV}, \text{Meas.})}{(0.36 \pm 0.03) \left( \frac{\text{C IV Blueshift}}{10^3 \text{ km s}^{-1}} \right) + (0.61 \pm 0.04)} \quad (1.5)$$

for the C IV/H $\beta$  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 H $\alpha$  and H $\beta$  fits respectively. The intrinsic scatter for the H $\alpha$  fit is represented by the Normal probability density distribution shown in Fig. 1.15. In the same figure we show the distribution of the orthogonal displacement of each data point from the best-fitting linear relationship. The two distributions are well-matched, which demonstrates that our model is a good representation of the data and the measurement errors on the data points are small relative to the intrinsic scatter.

The overall (intrinsic and measurement) scatter about the best-fitting model is slightly higher when the C IV line-widths are compared to H $\beta$  (0.12 dex) than when compared to H $\alpha$  (0.10 dex). This is likely due, at least in part, to the generally higher S/N of the H $\alpha$  emission. In addition, contributions from the strong [O III] doublet in the vicinity of H $\beta$  make de-blending the H $\beta$  emission more uncertain. As a consequence, for quasars where H $\alpha$  and H $\beta$  are both measured, the mean uncertainty on the H $\alpha$  FWHM is  $130 \text{ km s}^{-1}$ , compared to  $340 \text{ km s}^{-1}$  for H $\beta$ .

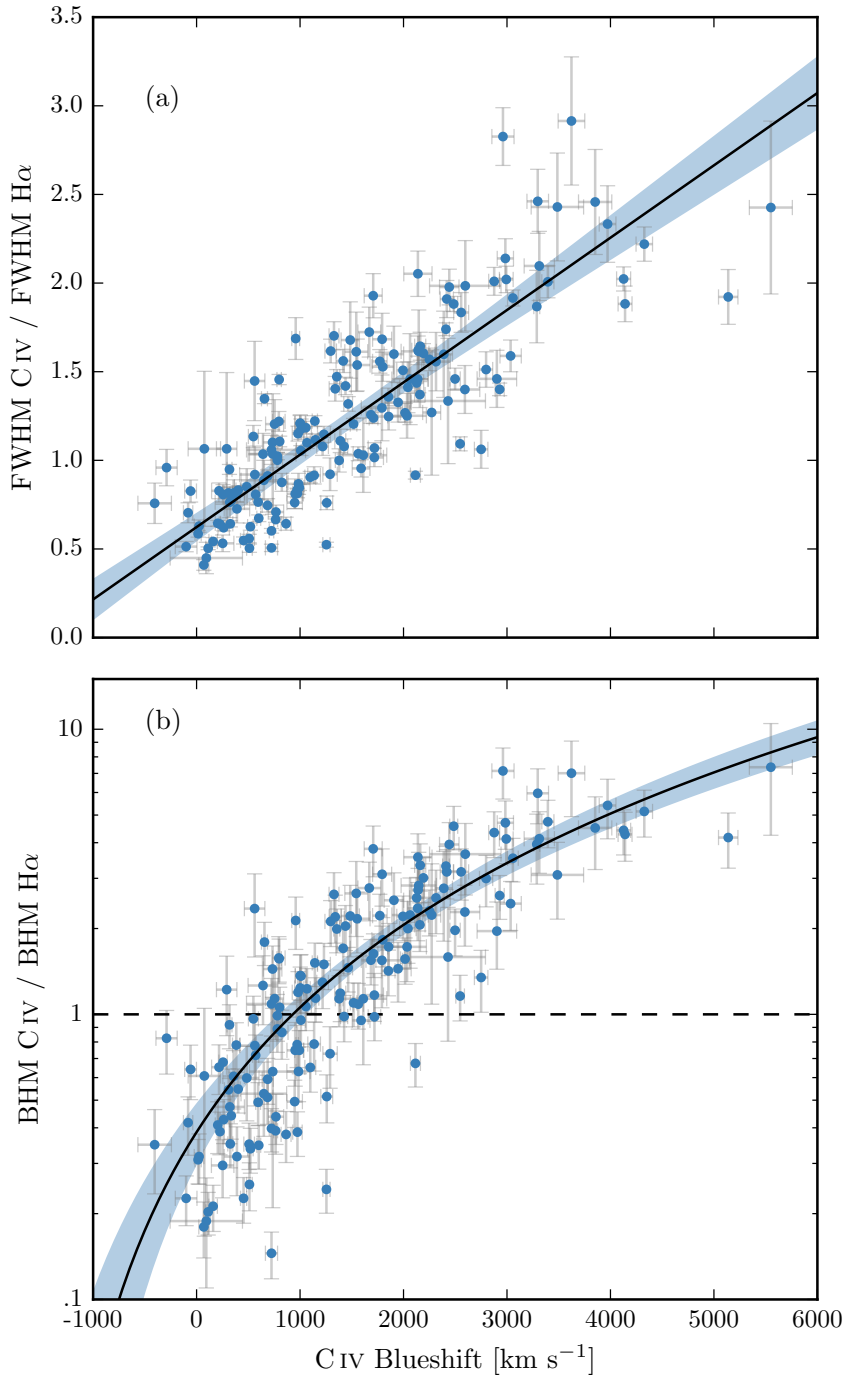


Figure 1.12: C iv FWHM relative to H $\alpha$  FWHM (a), and C iv based BH mass (BHM) compared to H $\alpha$  based mass (b), both as a function of the C iv blueshift. The black line is our best-fit linear model, and the shaded region shows the 2- $\sigma$  uncertainties on the slope and intercept. The H $\alpha$  FWHM have been scaled to match the H $\beta$  FWHM using Eq. 1.3.

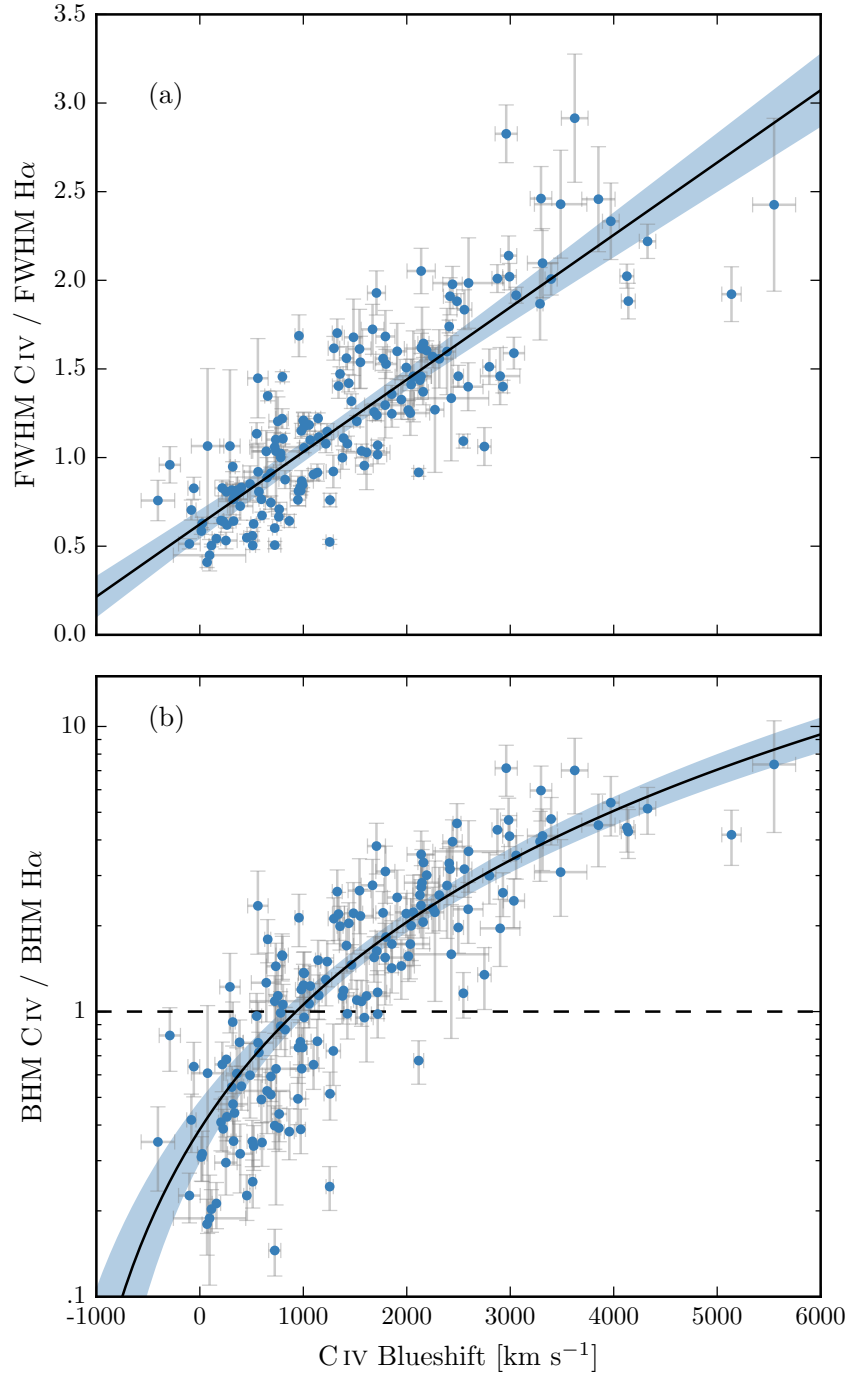


Figure 1.13: C IV FWHM relative to H $\beta$  FWHM (a), and C IV based BH mass (BHM) compared to H $\beta$  based mass (b), both as a function of the C IV blueshift.

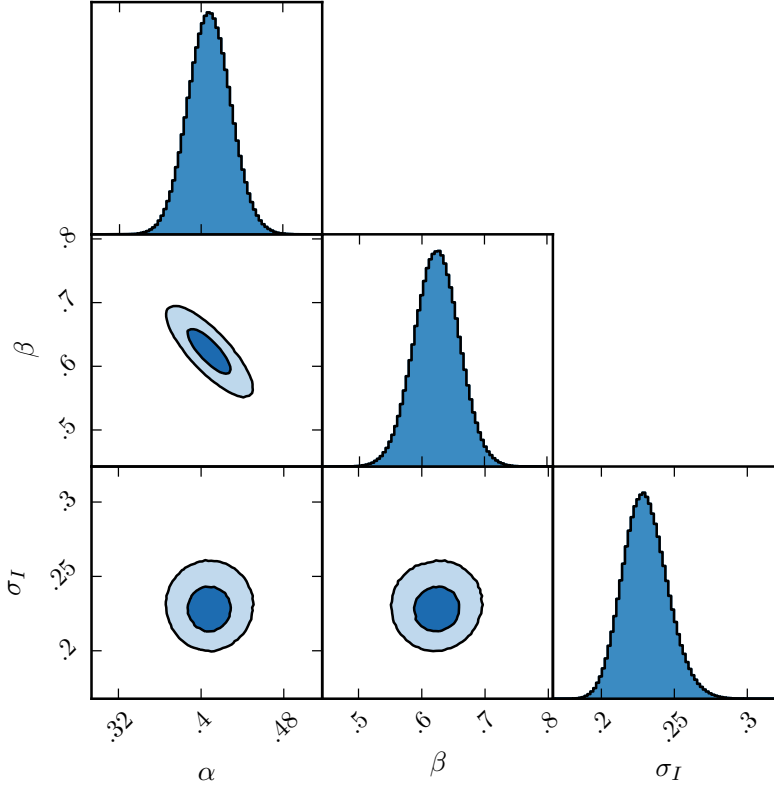


Figure 1.14: One- and two-dimensional projections of the MCMC sample of the posterior distribution for a linear fit to the FWHM C IV/H $\alpha$  ratio as a function of the C IV blueshift. In the two-dimensional projections we show 1- and 2- $\sigma$  contours. The posterior distribution for the linear fit to the FWHM C IV/H $\beta$  ratio, which we do not show, has a very similar appearance.

In the next section we use both the H $\alpha$  and H $\beta$  lines to calculate unbiased BH masses. However, we use the H $\alpha$  measurements to derive an empirical C IV blueshift based correction to the C IV masses (Eq. 1.6) because of the issues related to the accurate modelling of the H $\beta$ -profile just described. An extra advantage, which is evident in Figs. 1.12 and 1.13, is that the H $\alpha$  sample has a better C IV blueshift coverage. However, as can be seen from the similarity of Equations 1.4 and 1.5, our results would not change significantly were we instead to use the H $\beta$  sample.

#### 1.4.4 C IV based virial BH mass estimates

Virial BH masses were calculated using the widely adopted Vestergaard and Peterson, (2006) calibrations. The Vestergaard and Peterson, (2006) C IV FWHM calibration uses the monochromatic contin-



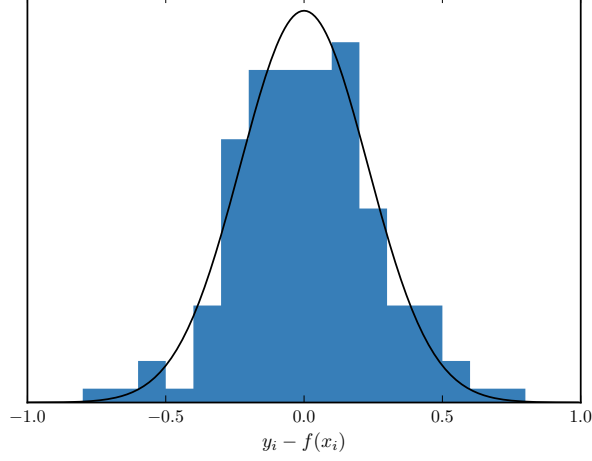


Figure 1.15: The distribution of the orthogonal displacement of each data point from the best-fitting linear relationship in the fit to  $\text{FWHM}(\text{C IV})/\text{FWHM}(\text{H}\alpha)$  as a function of the C IV blueshift (blue histogram). The black curve is a Normal distribution with a width equal to the intrinsic scatter in the population inferred from the fit. The two distributions are well-matched, which demonstrates that our model is a good representation of the data and the measurement errors on the data points are small relative to the intrinsic scatter.

uum luminosity at  $1350 \text{ \AA}$  to predict the BLR radius and corresponds to  $(a = 6.66, b = 2, c = 0.53)$  in Eq. ???. For the  $\text{H}\beta$  calibration, Vestergaard and Peterson, (2006) use the monochromatic continuum luminosity at  $5100 \text{ \AA}$  and calibration coefficients corresponding to  $(a = 6.91, b = 2, c = 0.5)$ . BH masses are computed using the line and continuum properties given in Table 1.2, and we convert our  $\text{H}\alpha$  emission-line velocity-width measures to predicted  $\text{H}\beta$  widths using Eq. ??.

In the lower panels of Figs. 1.12 and 1.13 the C IV-based estimates are compared to the  $\text{H}\alpha/\text{H}\beta$  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  $3000 \text{ km s}^{-1}$  in C IV emission blueshift and the overestimate of the BH-masses reaches a factor of 10 for quasars exhibiting the most extreme blueshifts,  $\gtrsim 5000 \text{ km s}^{-1}$ .

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 Eq. 1.4 as an indicator of the virial velocity, and adopt the same  $R - L$  relation for the  $1350 \text{ \AA}$  continuum luminosity as Vestergaard and Peterson, (2006) (i.e.  $R \propto L^{0.53}$ ). To find the constant scaling factor necessary to transform

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

$$MBH(C\text{ iv, Corr.}) = 10^{6.71} \left( \frac{FWHM(C\text{ iv, Corr.})}{10^3 \text{ km s}^{-1}} \right)^2 \left( \frac{\lambda L_\lambda(1350\text{Å})}{10^{44} \text{ erg s}^{-1}} \right)^{0.53} \quad (1.6)$$

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

#### 1.4.5 C iv-derived BH masses at low C iv blueshift

In this section, we consider why the C iv based masses of quasars with modest C iv blueshifts ( $\lesssim 1000 \text{ km s}^{-1}$ ) are systematically underestimated relative to masses derived from the Balmer lines (Figs. 1.12 and 1.13).

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

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

As a direct consequence of the empirically small C iv/ $H\alpha$  FWHM ratio, the C iv-derived BH mass estimates are systematically lower than the corresponding  $H\alpha$ -derived masses when the blueshift is small. This can be seen in Fig 1.12, where for almost every quasar with a C iv blueshift  $< 1200 \text{ km s}^{-1}$ , the C iv-derived BH mass is smaller

*Might work better  
later in discussion.*

than the corresponding H $\alpha$ -derived mass. The median fractional difference between the two estimates is 0.60.

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

### 1.5.1 Recipe for unbiased C IV based BH masses

#### *Measuring the systemic redshift*

Equations 4 and 6 together provide an un-biased estimate of the virial BH mass given the FWHM and blueshift of C IV, together with the continuum luminosity at 1350 Å. The FWHM is readily obtained, either directly from the data, or, via the fitting of a parametric model to the C IV-emission line. The blueshift – defined as the bisector of the cumulative line flux – is also straightforward to measure and our preferred procedure is described in Section 1.5.1. 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  $\mu\text{m}$  emission in the sub-millimetre band (e.g. Venemans et al., 2016). However, in general, for example in determining the BH-masses of quasars at redshifts  $z > 2$ , if only the rest-frame ultraviolet region is available determining a reliable systemic redshift is non-trivial.

The SDSS DR7 pipeline redshifts are not sufficiently reliable to measure the C IV blueshift accurately because, in part, the C IV emission line itself contributes to the determination of the quasar redshifts. This is demonstrated in Fig. 1.18a, in which we plot the C IV-blueshift versus C IV-emission equivalent width (EW) using the SDSS pipeline redshifts and the blueshifts calculated by Shen et al., (2011). A strong trend in the blueshift values as a function of line EW is not evident in Fig. 1.18a; 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 1.18b shows SDSS DR7 quasars in the same C IV parameter space as Figure 1.18a, but now using Hewett and Wild, (2010) redshifts. The improved redshift estimates

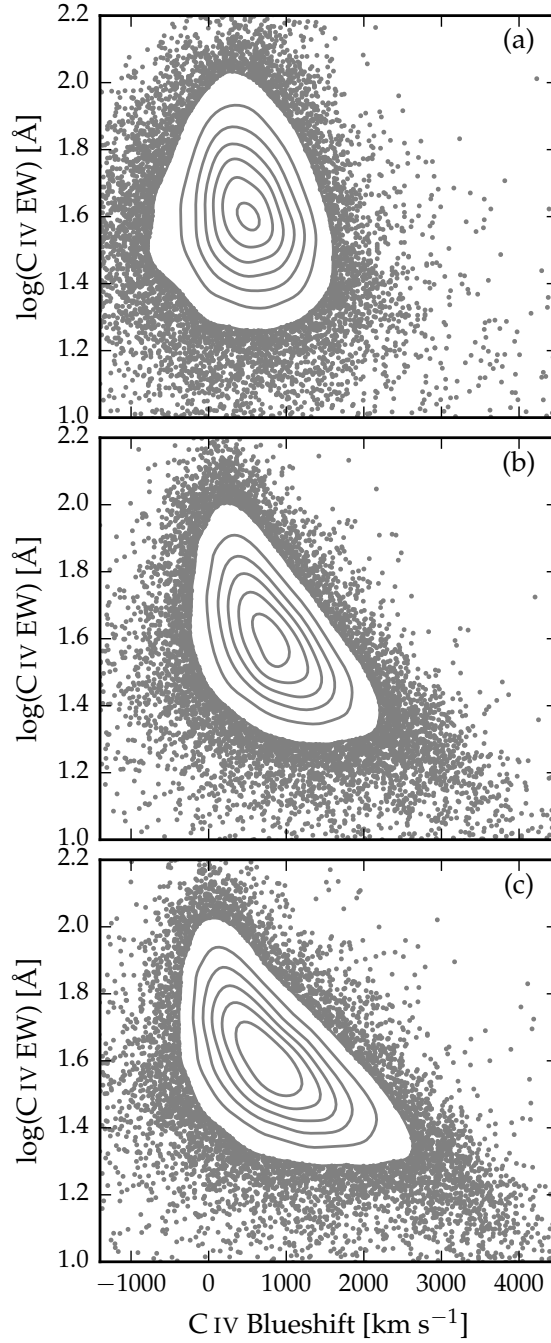


Figure 1.16: Rest-frame EW versus blueshift of the broad C IV-emission line for 32,157 SDSS DR7 quasars at  $1.6 < z < 3.0$ . Panel (a) uses C IV line parameters from Shen et al., (2011) and SDSS pipeline systemic redshifts. Panels (b) and (c) use systemic redshifts from Hewett and Wild, (2010) and Allen & Hewett (2017, in preparation) respectively, and C IV line measurements described in Sec. 1.5.1. In regions of high point-density, contours show equally-spaced lines of constant probability density generated using a Gaussian kernel-density estimator.

Table 1.4: 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 <sup>-1</sup> )	C iv blueshift (km s <sup>-1</sup> )			
	0	1000	2000	4000
250	0.33	0.20	0.14	0.09
500	0.65	0.39	0.28	0.18
1000	1.30	0.79	0.57	0.36

are predominantly responsible for the differences seen in Fig. 1.18a and b; the appearance in Fig. 1.18b of the extension to high blueshift for quasars with low C iv EW is particularly evident.

Shen et al., (2016) and our own work shows that there is an intrinsic variation of  $\sigma \simeq 220 \text{ km s}^{-1}$  in the velocity centroids of the broad-line region relative to a systemic-frame defined by the quasar narrow-line regions. The redshifts for quasars in the SDSS DR10 and DR12 catalogues (Pâris et al., 2014; Pâris et al., 2017) possess errors of  $\simeq 500\text{--}750 \text{ km s}^{-1}$  (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 1.4 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  $1000 \text{ km s}^{-1}$  C iv blueshift when there is a  $500 \text{ km s}^{-1}$  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 ultraviolet SED. The large systematic variation in the C iv emission-line profile within the population is evident from figures 11 and 12 of Richards et al., (2011). The plots and analysis in Richards et al., (2011) employ the quasar redshifts from Hewett and Wild, (2010) but, as is evident from the figures, the systematic variation in the C iv shape is correlated with changes in the quasar SEDs, including the strengths of the Si iii  $\lambda 1892$  and C iii  $\lambda 1908$  emission lines in the rest-frame ultraviolet. 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, of the C iv emission line. For the Hewett and Wild, (2010) redshifts, and ultraviolet emission-line based redshifts in general, quasars with large C iv EW and modest blueshifts have relatively small ( $\simeq 300 \text{ km s}^{-1}$ ) SED-dependent redshift errors. Redshift uncertainties as large as  $\simeq 1000 \text{ km s}^{-1}$  for such quasars are unusual and the large relative error in the corrected C iv BH-mass given in Table 1.4 is pessimistic.

Conversely, systematic redshift errors are greatest for quasars with large blueshifts, reaching  $\sim 750 \text{ km s}^{-1}$  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  $\text{MBH}(\text{C IV})/\text{MBH}(\text{Balmer})$  relation at large C IV blueshift (see Figs. 1.12 and 1.13). A definitive quantification of any systematic SED-dependent errors present in the quasar redshifts contained in the SDSS DR12 catalogue is not yet available but the principal component analysis (PCA) based redshift estimates are expected to be largely free of SED-dependent systematics.

Using published redshift estimates, notably those from Hewett and Wild, (2010) for the SDSS DR7 quasars and the BOSS PCA-based redshifts from Pâris et al., (2017) for SDSS DR12, the correction formula given in Section 1.4.3 produces significant improvements to C IV-based BH mass estimates. In a forthcoming work, Allen & Hewett (in preparation) will present a new redshift-estimation algorithm that produces redshifts independent of the C IV blueshift and other variations in the ultraviolet SEDs of luminous quasars. The low-ionization emission lines visible in the rest-frame ultraviolet (over wavelengths from  $\text{Mg II } \lambda 2796, 2803$  down to the  $\text{O I } \lambda 1304 + \text{Si II } \lambda 1307$  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  $\text{H}\beta$  and  $\text{H}\alpha$ . SED-dependent systematic errors are below the apparent inherent dispersion of  $\simeq 220 \text{ km s}^{-1}$  associated with broad emission line redshifts (Shen et al., 2016).

Figure 1.18c shows the C IV emission line parameters calculated using the Allen & Hewett redshift-estimation algorithm. The systematic trends seen in Fig. 1.18b, in particular the extension to high blueshift at low C IV EW, become more apparent in Fig. 1.18c, as expected from consideration of the known SED-related errors in the redshifts from Hewett and Wild, (2010). A population of quasars with only modest blueshifts and low EW is also apparently still present.

#### *C IV emission line blueshift measurements*

The differences in the distribution of C IV emission line properties seen in the three panels of Fig. 1.18 are due primarily to the change in the systemic redshift estimates. It is also necessary, however, to obtain a measure of the C IV emission line ‘location’ in order to calculate the blueshifts. When working with moderately-sized samples, parametric fits to the emission-line profile may be undertaken using careful mask-definition to minimise the effect of absorption features on the profiles used for the parametrization, and this is the approach we followed in Section 1.3.

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

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

We therefore chose to use a non-parametric scheme to measure the blueshift of the C iv line, which we will now describe. A continuum is first defined as a power-law of wavelength,  $f(\lambda) \propto \lambda^{-\alpha}$ , with the slope,  $\alpha$ , determined using the median<sup>4</sup> 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 C iv emission line is taken to lie within the wavelength interval 1500–1600 Å, a recipe that is commonly adopted (e.g. Shen et al., 2011; Denney et al., 2013). To reduce the impact of narrow absorption systems on the emission-line profile a ‘pseudo continuum’ is defined by applying a 41-pixel median filter to the quasar spectrum. Pixels within the C iv profile that lie more than  $2\sigma$  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. The wavelength that bisects the cumulative total line flux is recorded and the blueshift is defined in exactly the same way as in Section 1.3.

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

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



### 1.5.2 Systematic trends in residuals

The scatter about the best-fitting line in the C IV/H $\alpha$  FWHM versus C IV-blueshift relation is  $\sim 0.1$  dex, an order of magnitude smaller than the size of the C IV-blueshift dependent systematic but, nevertheless, still significant. With a view to reducing the scatter further, we searched for measurable parameters which correlate with the scatter at fixed C IV blueshift, including the luminosity, redshift, [O III] equivalent width (EW), and Fe II EW. The only significant correlation we find is with the H $\alpha$  FWHM (Fig. 1.17). Quasars with broad H $\alpha$  lines tend to lie below the relation while quasars with narrow H $\alpha$  tend to lie above it. One possibility is that this correlation is simply due to random scatter (either intrinsic or measurement error) in the H $\alpha$  FWHM which, with the other quasar properties fixed, would naturally produce a correlation between FWHM(C IV)/FWHM(H $\alpha$ ) and FWHM(H $\alpha$ ). However, the fact that we see no such correlation between the model residuals and the C IV FWHM suggests that the H $\alpha$  FWHM correlation could be revealing something more fundamental. The H $\alpha$ /H $\beta$  FWHM is part of ‘eigenvector 1’ (EV1), the first eigenvector in a principal component analysis which originated from the work of Boroson and Green, (1992). While a number of parameters have been considered within the EV1 context (e.g. Brotherton and Francis, 1999), Fig. 1.17 suggests that part of the scatter between the Balmer and C IV velocity widths might be attributed to differences in the spectral properties which are correlated with EV1 (Marziani et al., 2013).

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

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

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

<sup>5</sup> Strictly FWHM/ $\sigma \rightarrow 0$  for a Lorentzian profile, but values close to unity are typical when the dispersion is calculated over a velocity range,  $\simeq \pm 10\,000\text{km s}^{-1}$ , used to parametrize broad emission lines in quasar spectra.

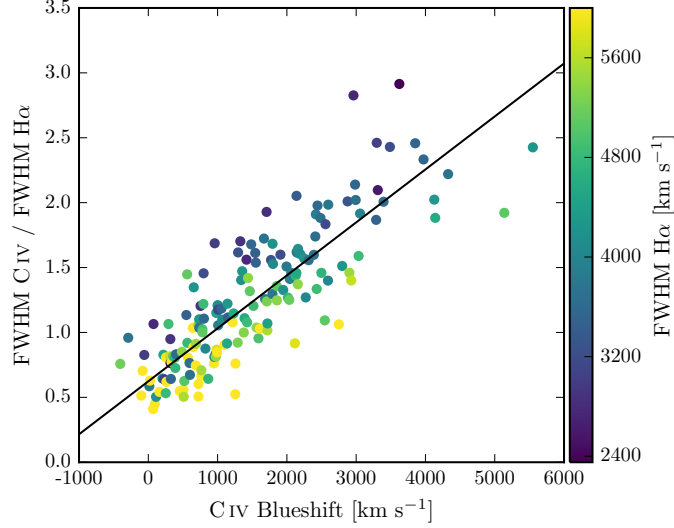


Figure 1.17: Same as Fig. 1.12a, with the marker colour representing the  $H\alpha$  FWHM. At fixed  $C\text{ IV}$  blueshift, there is a clear  $H\alpha$  FWHM dependent systematic in the model residuals.

to identify orientation-dependent signatures. Following Shen et al., (2011), we classified quasars with matches within 5 arcseconds as core-dominated, while, if multiple matches were found within 30 arcseconds, quasars were classified as lobe-dominated. Twenty core-quasars and six lobe-dominated quasars resulted but no statistically significant differences in the  $H\alpha$  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).

There are currently very few reverberation-mapping measurements of quasars with large  $C\text{ IV}$  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  $C\text{ IV}$  blueshifts and lead to a greater understanding of the relation between the Balmer line profile and the BH mass.

### 1.5.3 Effectiveness of the $C\text{ IV}$ blueshift based correction to BH masses

Figure 1.18 demonstrates that our sample has an excellent coverage of the EW-blueshift parameter space in relation to SDSS DR7 quasars at redshifts  $1.6 < z < 3.0$ . The systematic offset to higher  $C\text{ IV}$  blueshifts for our catalogue relative to the SDSS quasars as a whole is a result of the higher mean luminosity relative to the SDSS sample (Fig. 2.1). Our

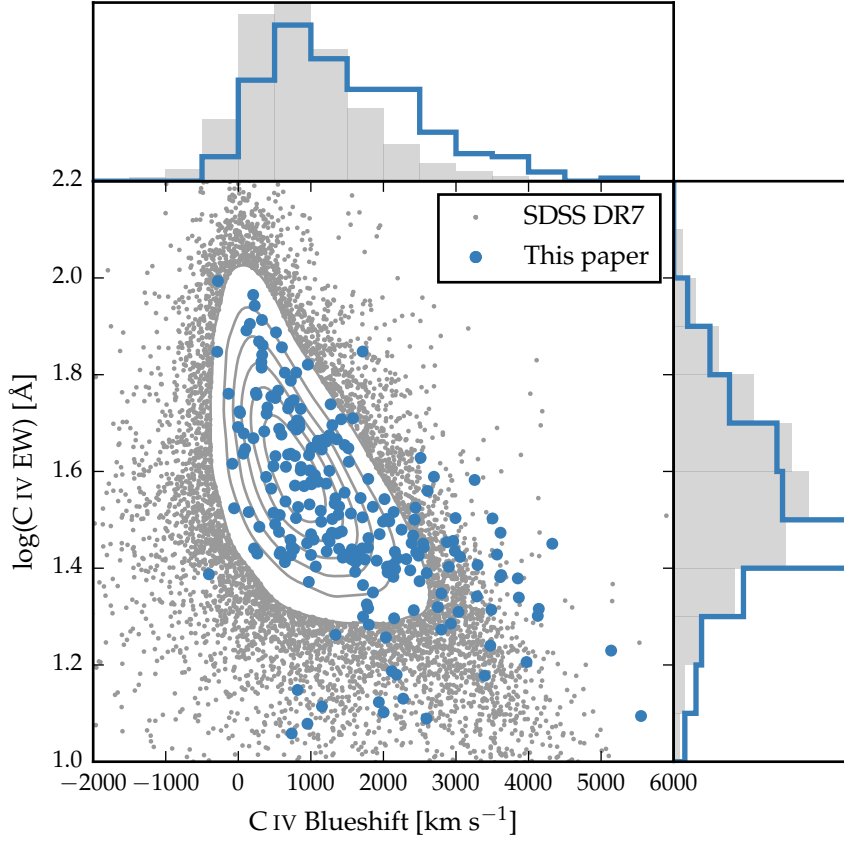


Figure 1.18: Rest-frame EW versus blueshift of the broad C IV-emission line for 32,157 SDSS DR7 quasars at  $1.6 < z < 3.0$  (grey) and our sample (blue). For the SDSS quasars, the systemic redshifts used to calculate the blueshifts are from Hewett and Wild, (2010) and C IV emission properties are described in Paper I. In regions of high point-density, contours show equally-spaced lines of constant probability density generated using a Gaussian kernel-density estimator. Our sample has very good coverage; the shift to high blueshifts is a result of the high luminosity of our sample in relation to the SDSS sample and the correlation between luminosity and blueshift.

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

Figure 1.19 compares the C iv- and H $\alpha$ -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 H $\alpha$ -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 H $\alpha$ -based masses is reduced to 0.2 dex. The scatter is 0.24 dex at low C iv blueshifts ( $\sim 0\text{km s}^{-1}$ ) and 0.10 dex at high blueshifts ( $\sim 3000\text{km s}^{-1}$ ).

There has been a considerable amount of attention regarding the relative merits of using the FWHM or dispersion to characterise the velocity width (e.g. Denney et al., 2013). The existence of a trend in the C iv-dispersion values with C iv blueshift is evident from inspection of the bottom left panel of Fig. ?? but the systematic trend relative to the spread at fixed blueshift is significantly smaller than when using C iv FWHM. Therefore, without the blueshift information, using the line dispersion would yield a more accurate BH mass than the FWHM (Fig. 1.20).

The correlation between the H $\alpha$  and C iv line dispersion is, however, weak. The Pearson coefficient for the correlation is 0.36 (and just 0.15 when the H $\beta$  measurements are used in place of H $\alpha$ ). Furthermore, there is little dynamic range in the line dispersion: the scatter is just 480 and  $460\text{km s}^{-1}$  for H $\alpha$  and C iv respectively. The observation suggests that the line dispersion does not fully trace the dynamic range in BH mass present in the quasar population. At least part of the reason is that the line dispersion is difficult to measure reliably in current survey-quality data, particularly because of the sensitivity to flux ascribed to the wings of the emission line (e.g. Mejía-Restrepo et al., 2016). Figures 1.19 and 1.20 demonstrate that estimating a reliable BH mass from the C iv FWHM and blueshift line is substantially more effective than using the C iv line dispersion with, or without, the line blueshift.

#### 1.5.4 Comparison to previous prescriptions

In Fig. 1.21 we compare the C iv blueshift-based correction presented in this chapter to various prescriptions which have been proposed in the literature to derive BH masses from the C iv line which are consistent with the masses derived from the Balmer lines. In each

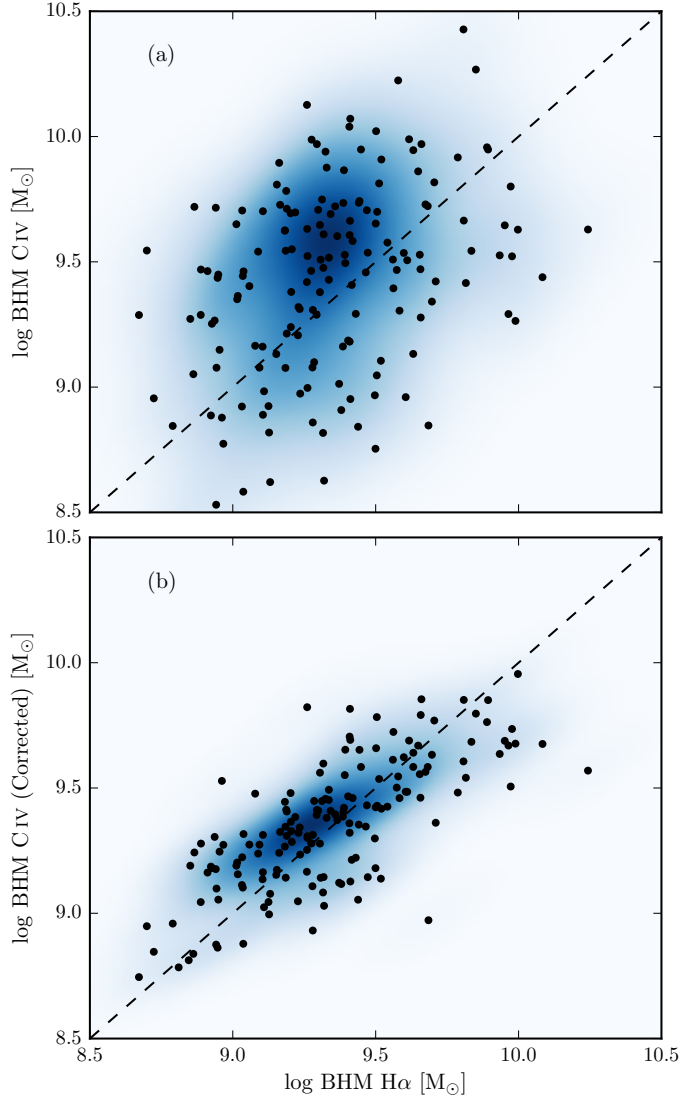


Figure 1.19: Comparison of the C iv- and H $\alpha$ -based BH masses before (a) and after (b) applying the C iv blueshift-based correction to the C iv FWHM. The density of the plotted points (estimated using a Gaussian kernel density estimator) is represented by the colour. The correction to the C iv BH masses decreases the scatter by from 0.4 to 0.2 dex. **Should definitely include some empirical validation of ICA redshifts since that is what we are telling people to use.**

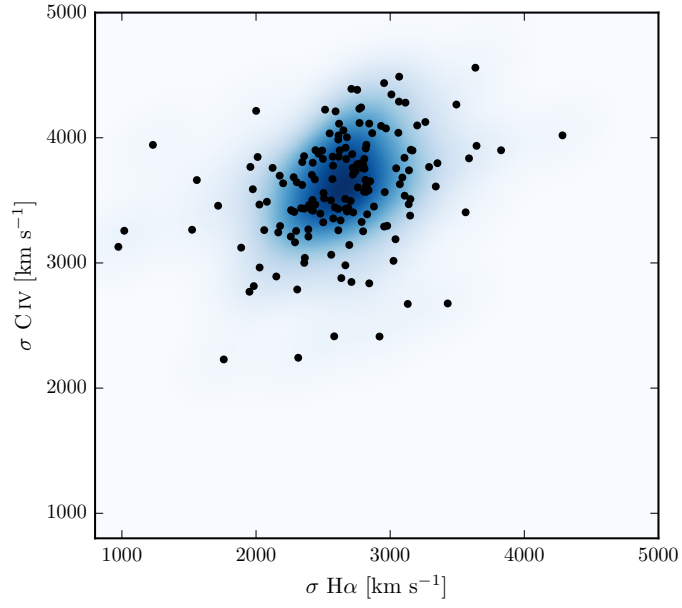


Figure 1.20: Comparison of the C IV and H $\alpha$  line dispersion,  $\sigma$ . The density of the plotted points (estimated using a Gaussian kernel density estimator) is represented by the colour. Estimating a reliable BH mass from the C IV FWHM and blueshift line is substantially more effective than using the C IV line dispersion with, or without, the line blueshift. The C IV dispersion values are larger than the corresponding H $\alpha$  measurements by a factor of 1.4 on average, which is consistent with reverberation mapping measurements (Vestergaard and Peterson, 2006).

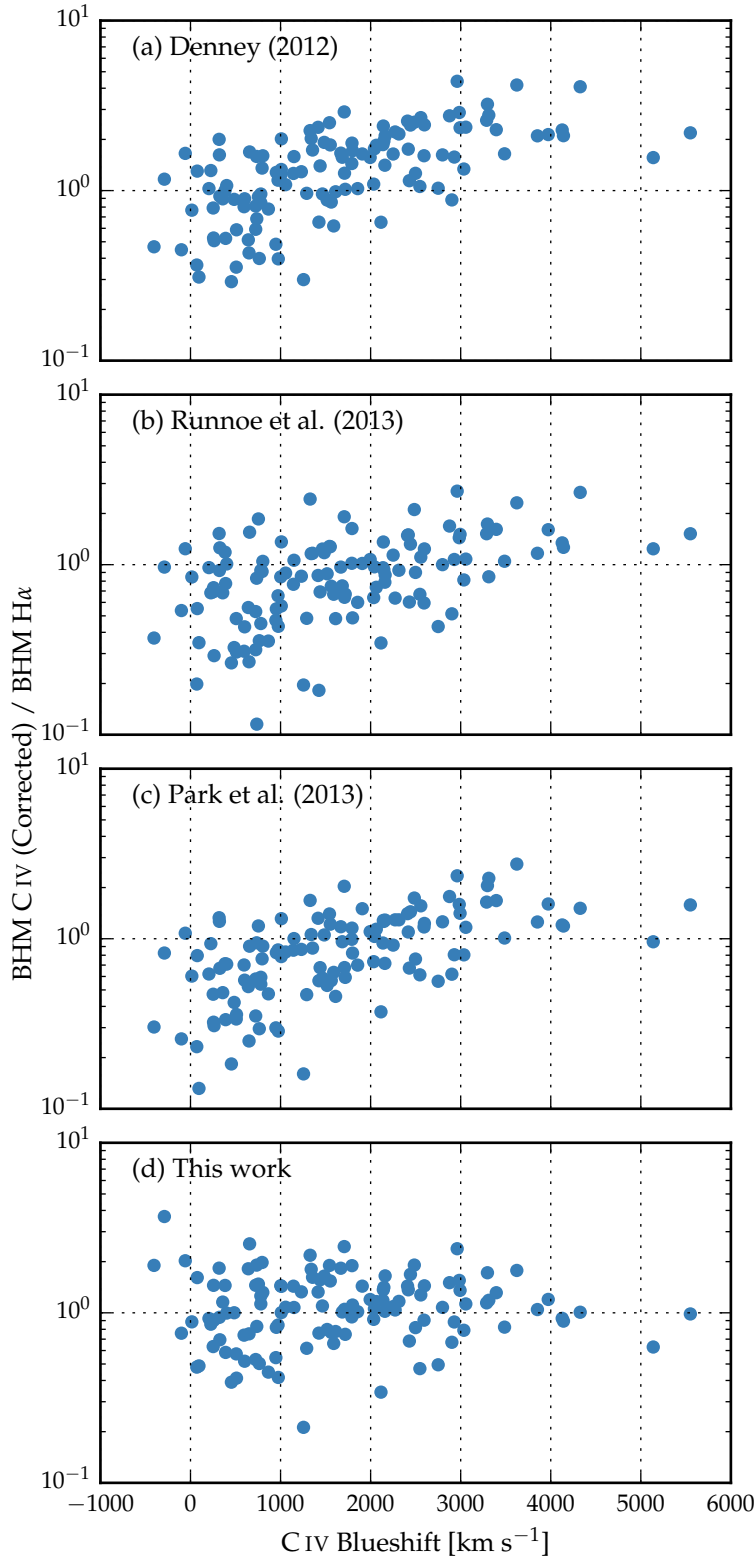


Figure 1.21: Comparison of BH mass estimates derived from C IV and H $\alpha$  as a function of the C IV blueshift. Corrections to the C IV-based masses have been applied based on the shape (FWHM/ $\sigma$ ) of the C IV emission line (a; Denney, 2012), the peak flux ratio of the Si IV+O IV blend relative to C IV (b; Runnoe et al., 2013a), by significantly reducing the dependence of the derived BH mass on the C IV velocity-width (c; Park et al., 2013), and based on the C IV blueshift (d; this chapter).

case we compare the corrected C iv-based masses to the H $\alpha$ -based masses as a function of the C iv blueshift. The correction proposed by Runnoe et al., (2013a) is based on the spectral region at rest-frame wavelengths of  $\sim 1400 \text{ \AA}$  (see below). Therefore, our analysis is based on the 123 quasars with spectra covering this region.

In Fig. 1.21a 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 (Fig. 1.7); between the two extremes in the C iv blueshift distribution the line shape changes from FWHM/ $\sigma \sim 1$  to 2.5. The investigation of Denney, (2012) was based on a sample of reverberation mapped quasars, which have a narrow range of C iv-emission line shapes, including the absence of any objects with large C iv blueshifts. The correction is not applicable at large C iv blueshifts. Therefore, while the consistency between the 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 UV region of the spectrum has proved difficult. However, the situation is improved dramatically by the new scheme developed by Allen & Hewett (2017, in preparation). Given the difficulty of measuring reliable C iv blueshifts without the Allen & Hewett scheme, Runnoe et al., (2013a) opted instead to use the continuum-subtracted peak flux ratio of the ultraviolet emission-line blend of Si iv+O iv (at  $1400 \text{ \AA}$ ) to that of C iv to correct for non-virial contributions to the C iv velocity width. This parameter was chosen because it showed the strongest correlation with the FWHM C iv/H $\beta$  residuals, as well as with the strengths of optical O iii and Fe ii.

Following Runnoe et al., (2013a), we measure the peak flux by fitting a model with four Gaussian components (two for each emission line) to the continuum-subtracted flux. As is evident from Fig. 1.18, a correlation exists between the blueshift and equivalent width of C iv: C iv emission which is strongly blueshifted is typically weak. The Si iv+O iv emission-line blend, however, shows significantly less systematic variation. Therefore, the Si iv+O iv-based correction is quite effective in practice: the systematic bias in the C iv BH masses at large C iv blueshifts is reduced to a factor of  $\sim 2$  (Fig. 1.21b). However, the C iv based masses are still systematically overestimated at large C iv blueshifts.

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



predicted assuming a virialized BLR to just  $V^{0.56}$ ). As a consequence, the C IV based masses of the quasars with large C IV blueshifts are much reduced (Fig. 1.21c). However, the systematic error in the C IV-based BH masses as a function of C IV blueshift remains.

As a comparison, the C IV-based masses shown in Fig 1.21d have been corrected using to the C IV blueshift-based procedure presented in this chapter. No systematic in the BH masses as a function of the C IV blueshift is evident.

## 1.6 POPULATION TRENDS WITH C IV BLUESHIFT

As shown in Fig. 1.22, there are systematic variations in the H $\alpha$  line profile as a function of the C IV blueshift. At C IV-blueshift  $< 1200 \text{ km s}^{-1}$ , the H $\alpha$  FWHM range is  $\simeq 2000 - 8900 \text{ km s}^{-1}$ , with mean  $\simeq 4300 \text{ km s}^{-1}$ . However, amongst the quasars with C IV-blueshift  $> 2000 \text{ km s}^{-1}$ , the mean H $\alpha$  FWHM  $= 3500 \text{ km s}^{-1}$ , with a scatter of just  $700 \text{ km s}^{-1}$ . 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 $\alpha$  EW 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, in to a bolometric luminosity we use the bolometric correction factor given by Richards et al., (2006) ( $L_{\text{bol}} = 9.26 L_{5100}$ ). 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 Fig. 1.23, demonstrate that at large blueshifts quasars are accreting at around their Eddington limits (Fig. 1.23). This finding is in accord with our interpretation that the blueshifting of C IV is evidence for strong outflows resulting from the presence of a radiation-driven accretion-disc wind. Richards et al., (2002) found that quasars with large C IV blueshifts have weak He II. This is evidence for weak soft X-ray continuum emission (Leighly, 2004), which would allow a strong line-driven wind to form. The strength of such a wind is predicted to be related to the quasar far-ultraviolet SED, which, in turn, could be related to the mass-accretion rate.

All of the objects in our sample which exhibit large C IV blueshifts would be classified as population A in the Sulentic et al., (2000) scheme based on the H $\alpha$  FWHM. Our results therefore support the

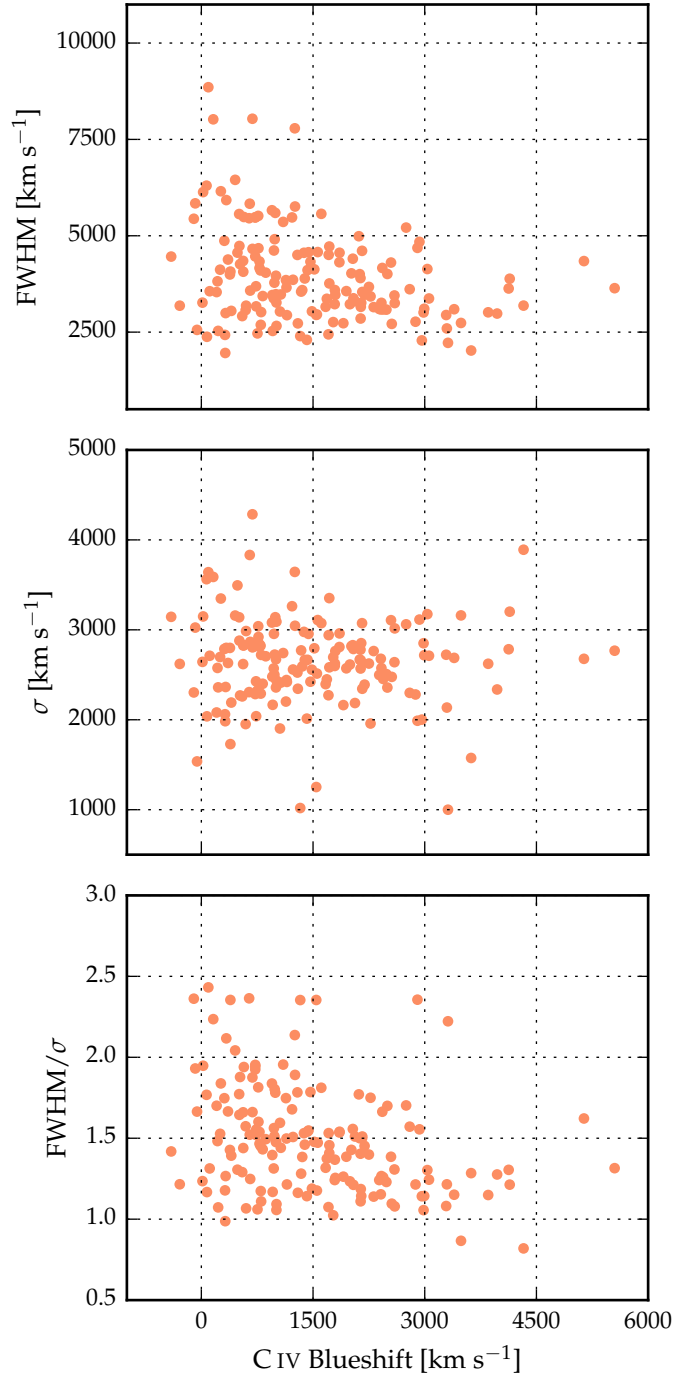


Figure 1.22: The FWHM, dispersion ( $\sigma$ ) and shape (FWHM/ $\sigma$ ) of H $\alpha$  as a function of the C IV blueshift.

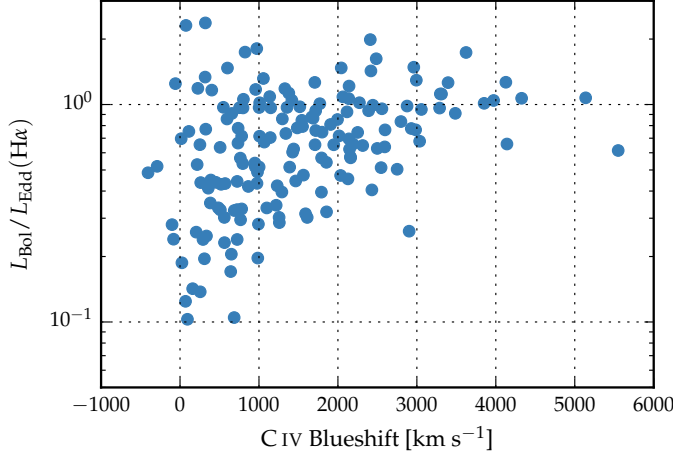


Figure 1.23:  $H\alpha$ -derived Eddington ratio versus C IV blueshift. At blueshift  $\gtrsim 2000 \text{ km s}^{-1}$  all quasars have high accretion rates ( $L/L_{\text{Edd}} \simeq 1$ ). This is in agreement with Kratzer and Richards, (2015), but in contrast to what one would derive from naive use of C IV-based BH mass scaling relations.

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

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

#### 1.6.1 Systematic biases in Balmer-based masses

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

*Can move most of  
this to earlier section  
on residuals and  
then refer back to  
that.*

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

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

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

Although the EV1-trends (Sulentic et al., 2000; Shen and Ho, 2014) are most likely driven by the accretion rate, orientation may also have a role to play in determining the observed properties of the BLR. Shen and Ho, (2014) argue that a large part of the scatter observed in the 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. For this to be true, the BLR would need to be in a flattened disc-like geometry, in which case the observed line width would increase with the inclination of the disc relative to the line of sight. Brotherton, Singh, and Runnoe, (2015) found that the core-dominance of radio-loud quasars, which is believed to be a reliable proxy for orientation, at least in a statistical sense, is significantly correlated with the H $\beta$  FWHM and hence with the BH-mass estimates. This raises the question of whether the narrow H $\alpha$  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 II $\lambda$ 1640 emission-line properties of quasars with large C iv blueshifts are more consistent with differences in the SED rather than differences in the

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

#### 1.6.2 The BAL parent population

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

*Do I actually say  
BALs are removed  
from sample?*

#### 1.6.3 The frequency of quasars with high accretion rates

Quantifying the frequency of quasars producing outflows as a function of key parameters, e.g. quasar luminosity, BH-mass, redshift,... will be important to constrain models of quasar-galaxy evolution. At fixed BH mass, the intrinsic and the observed fraction of quasars exhibiting properties that depend on the Eddington ratio can differ significantly. As an illustration, we consider the implications for the intrinsic fraction of quasars possessing large C iv blueshifts given the observed numbers in the  $m_i < 19.1$  flux-limited sub-sample of the SDSS DR7 quasar catalogue. In order to estimate the size of the selection effect, we considered the detection probability for a much-simplified quasar population. We assume that all quasars with C iv blueshifts  $> 1200 \text{ km s}^{-1}$  have enhanced accretion rates relative to the ‘normal’ population (with C iv blueshifts  $< 1200 \text{ km s}^{-1}$ ). If the accretion rate of the high-blueshift population is double the rate of the low-blueshift population (which is true in an average sense – see Fig. 1.23), then the high-blueshift population will be brighter by  $\simeq 0.75$  magnitude. Under the assumption that the BH mass distribution is

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

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<sup>6</sup> 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  $< 1200 \text{ km s}^{-1}$  and 2,484 with blueshifts  $> 1200 \text{ km s}^{-1}$  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 ultraviolet-selected non-BAL quasar population).

## 1.7 CONCLUSIONS

The main results of this chapter are as follows:

- We have analysed the spectra of 230 high-luminosity ( $10^{45.5} - 10^{48} \text{ erg s}^{-1}$ ), redshift  $1.5 < z < 4.0$  quasars for which spectra of the Balmer emission lines and the C iv emission line exist. The large number of quasars in our spectroscopic catalogue and the wide range in C iv blueshifts the quasars possess has allowed us to directly investigate biases in C iv-based BH mass estimates which stem from non-virial contributions to the C iv emission as a function of the C iv blueshift, which, in turn, depends directly on the form of the quasar ultraviolet SEDs (Richards et al., 2011).
- The C iv emission-based BH-masses are systematically in error by a factor of more than five at  $3000 \text{ km s}^{-1}$  in C iv emission blueshift and the overestimate of the BH-masses reaches a factor of 10 for quasars exhibiting the most extreme blueshifts,  $\gtrsim 5000 \text{ km s}^{-1}$ .
- We have derived an empirical correction formula for BH-mass estimates based on the C iv emission line FWHM and blueshift. The correction may be applied using equations 4 and 6 in Section 1.4.3. 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

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

and the  $H\alpha$ -based masses is 0.24 dex at low C iv blueshifts ( $\sim 0 \text{ km s}^{-1}$ ) and 0.10 dex at high blueshifts ( $\sim 3000 \text{ km s}^{-1}$ ). This is a significant improvement on the 0.40 dex scatter observed between the un-corrected C iv and  $H\alpha$  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  $z \sim 5$ .

## 1.8 FUTURE WORK

- Clustering
- Data-driven mapping





## NARROW LINE REGION PROPERTIES

---

### 2.1 INTRODUCTION

Correlations between the masses of super-massive black holes and properties of the host galaxy bulges in nearby galaxies (Gebhardt et al., 2000; Ferrarese and Merritt, 2000) and the similarity of the cosmic evolution of star formation and black hole activity (Boyle and Terlevich, 1998; Madau and Dickinson, 2014) suggest that the formation and evolution of supermassive black holes and their host galaxies are linked. Active galactic nuclei (AGN) and quasar feedback, in which star formation in the host galaxy is suppressed by the energy output from the quasar, is a favoured model. This has motivated a considerable amount of observational work searching for feedback signatures (for recent reviews, see Alexander and Hickox, 2012; Fabian, 2012; Heckman and Best, 2014).

The [O III] $\lambda\lambda 4960, 5008$  doublet is the strongest narrow line found in optical quasar spectra. It is forbidden line which traces gas on kiloparsec scales. High velocity dispersions or strong asymmetries in forbidden lines are evidence of high velocity extended ionized outflows. Recent studies have provided constraints on the prevalence of ionized outflows traced by [O III] emission in low-redshift type 2 AGN (e.g. Mullaney et al., 2013; Zakamska and Greene, 2014) using spectra from the Sloan Digital Sky Survey (SDSS; York et al., 2000). However, the signatures of feedback are expected to be stronger at higher redshifts ( $z \sim 2$ ) when the black hole accretion and star formation rates both peak.

In this paper we analyse the [O III] properties of a sample of 358 high-luminosity, redshift  $1.5 < z < 4$  quasars. In particular, we search for signs of outflowing gas, which include asymmetries and broad velocity widths. Recent near-IR spectroscopy of  $z > 1.5$  quasars often report exceptionally large [O III] widths, with FWHM  $> 1000 \text{ km s}^{-1}$  (e.g. Netzer et al., 2004; Nesvadba et al., 2008; Kim et al., 2013; Brusa et al., 2015; Carniani et al., 2015; Perna et al., 2015). **Add reference to WIISH paper.**

To date, the focus has been extreme, rare, AGN sub-populations, e.g. highly reddened and obscured AGN (Banerji et al. 2012, 2013, 2014, Eisenhardt et al. 2012, Ross et al. 2014, Zakamska et al. 2016, WIISH paper). These populations provide insight into relatively short-lived phases in the AGN activity cycle. A more complete understanding of AGN-evolution and the physical link with the host galaxies requires

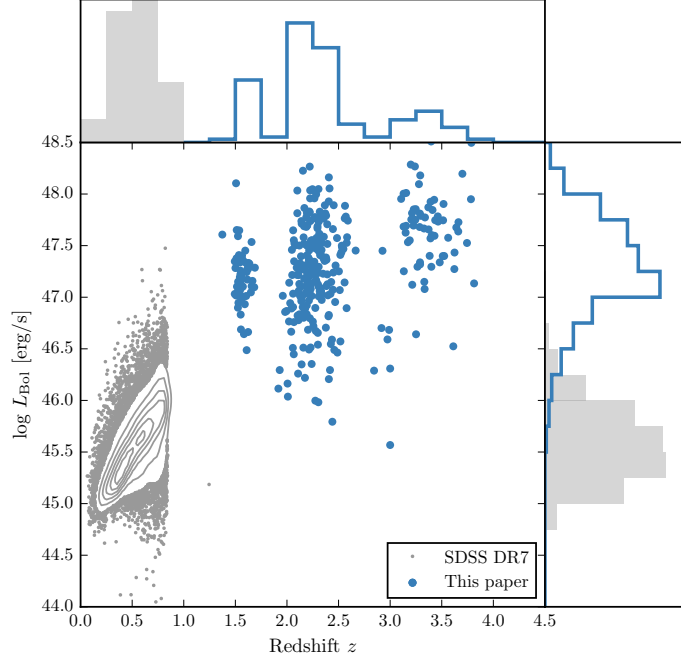


Figure 2.1: The ranges in redshift and luminosity covered by our sample, relative to the redshift-luminosity distribution of the SDSS DR7 quasar catalogue with measured [O III] line properties (Shen et al., 2011). The gaps in our sample coverage at  $z \sim 1.8$  and  $z \sim 3$  are due to the near-infrared transmission windows. In regions of high point-density, contours show equally-spaced lines of constant probability density generated using a Gaussian kernel-density estimator.

the rare, extreme, sub-populations to be placed in the context of the AGN population as a whole.

For a subsample where we have spectra covering the rest-frame ultra-violet line C IV, we can test whether the strong outflows, inferred the blueshifting of C IV relative to the systemic redshift, have any effect on the host galaxy on larger scales.

Throughout this paper we adopt a flat  $\Lambda$ CDM cosmology with  $h_0 = 0.69$  and  $\Omega_M = 0.29$ .

The paper is structured as follows.

## 2.2 QUASAR SAMPLE

We have assembled a catalogue of 358 high-luminosity, redshift  $1.5 < z < 4$  quasars, with spectra taken in the near-infrared. At these redshifts the spectra cover the rest-frame optical region, which includes the broad Balmer H $\beta$  line and the strong, narrow [O III] doublet. There is a significant overlap between the sample presented here and the one used in Coatman et al. (2016b), hereafter paper I. We briefly

describe the catalogue below, but refer the reader to Coatman et al. (2016b) for further details, including the reduction procedure.

Two hundred are selected from the SDSS Seventh Data Release (DR7; Schneider et al., 2010). These are observed with LIRIS (Machado et al., 1998) on the William Herschel Telescope (WHT), Triple-Spec (Wilson et al., 2004) on the Astrophysics Research Consortium (ARC) 3.5 m telescope, FIRE (Simcoe et al., 2010) on the Magellan-Baade telescope, the SINFONI integral field spectrograph (Eisenhauer et al., 2003; Bonnet et al., 2004) on the European Southern Observatory (ESO) Very Large Telescope (VLT), SOFI (Moorwood, Cuby, and Lidman, 1998) on the New Technology Telescope (NTT), and TRIPLE-SPEC on the Palomar 200-inch Hale telescope (P200). One hundred and seventeen were observed to measure accurate redshifts of quasar pairs at very close projected separations (Prochaska and Hennawi, 2009; Lau, Prochaska, and Hennawi, 2015; Hennawi et al., 2015). Spectra were taken with GNIRS (Elias et al., 2006) on the Gemini North telescope, ISAAC (Moorwood et al., 1998) on the European Southern Observatory (ESO) Very Large Telescope (VLT), NIRSPEC (McLean et al., 1998) on the Keck-II telescope, NIRI (Hodapp et al., 2003) also on Gemini North, and XSHOOTER (Vernet et al., 2011), again, on the VLT. Seventy-nine are bright Hamburg-ESO quasars observed with SINFONI, ninety are quasars with archival high-resolution optical spectra observed with SOFI, and four are observed with XSHOOTER and presented in the XQ-100 catalogue (Lopez et al., 2016).

This is the largest study of the narrow line region properties of high- $z$  quasars ever undertaken. The redshift and luminosity coverage of the quasars in our sample is shown in Fig. 2.1, and the quasar sample is summarised in Table 2.1. Our sample covers much higher redshifts and luminosities than the SDSS sample.

### 2.3 PARAMETERIC MODEL FITS

In this section we describe how we measure the [O III] velocity-width and parameterize asymmetries in the line. We first fit a parameteric model to the [O III] emission, and the nearby  $H\beta$  peak. This step is taken solely to enhance the signal-to-noise (SNR) of the spectra and to decompose the emission from  $H\beta$  and the [O III] doublet. We then calculate non-parameteric measures of the line profile from the best-fitting model.

Before the model can be fit, the spectra must first be transformed into the (approximate) quasar rest-frame. This transformation is only for practical purposes and the emission line parameters we will go on to derive are independent of the exact redshift used. The redshift will later be refined using the results from the fit. The redshift used in this transformation is either derived from a multi-component Gaussian fit to the broad  $H\alpha$  emission ( $\sim 40$  per cent of our sample) or, when this

Table 2.1: The numbers of quasars with [O III] line measurements and the spectrographs and telescopes used to obtain the near-infrared spectra. Further details on the instrumental configurations are given in paper I. **Check numbers.**

Spectrograph	Telescope	Number
FIRE	MAGELLAN	32
GNIRS	GEMINI-N	29
ISAAC	VLT	8
LIRIS	WHT	5
NIRI	GEMINI-N	29
NIRSPEC	Keck II	3
SINFONI	VLT	76
SOFI	NTT	78
TRIPLESPEC	ARC-3.5m	27
TRIPLESPEC	P200	46
XSHOOTER	VLT	25
		144

is not possible, from a preliminary fit to the broad H $\beta$  emission ( $\sim 40$  per cent) or narrow [O III] emission (20 per cent). **Should say whether we use peak/median etc.**

The continuum and optical Fe II emission is modelled and subtracted using a recipe that is identical to the one described in paper I. The following model is then fit to the spectra in the wavelength interval 4700–5100 Å. The fit is done as a function of the Doppler velocity shift, and we adopt the wavelength 4862.721 Å (the laboratory H $\beta$  wavelength) to transform wavelengths into equivalent Doppler velocities.

H $\beta$  is modelled by two Gaussians with non-negative amplitudes and FWHM greater than  $1200 \text{ km s}^{-1}$ . A handful we fix the centroids of the two Gaussians to be the same, normally because of very low S/N or because the blue wing is below the lower wavelength limit of the spectrograph. Any contribution to the H $\beta$  emission from the narrow-line region is weak in the vast majority of our sample, and so we do not include an additional Gaussian component to model this emission. Note: we do in fact include narrow components for eight objects in our sample. Although, for some objects, this could bias our estimate of the velocity-width of the broad component, this information is not used in the analysis presented in this paper.

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. If the addition of the second

Table 2.2: Approximate physical origin of the ICA components.

Component	Origin
$w_1$	Fe II
$w_2$	H $\beta$
$w_3$	H $\beta$
$w_4$	[O III] core
$w_5$	[O III] core
$w_6$	[O III] wing

Gaussian decreases the reduced  $\chi^2$  by more than 5 per cent then the double-Gaussian model is accepted. One hundred and twenty-five are fit with a single Gaussian, 147 with two Gaussians, and [O III] is undetected in a further 78 quasars. When a single Gaussian is used to model each line, 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. In the double Gaussian fit, the peak flux ratio of the second components is again fixed at 1:3, and the width and velocity offsets are again set to be equal.

In six quasars a significantly better fit was obtained by allowing the flux ratio between the two components to vary. In these quasars the best-fitting peak ratio varies from 0.50 to 0.84, with mean 0.70. **Check ICA component fits to see if this looks real.**

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

**However, the fits aren't always good. How to be quantative about this? Chi-squared values indicating bad fits? Also the fact that we sometimes model Fe as OIII. Discuss the limitations of using the Boroson & Greene template? Even if don't use results, should discuss Gaussian fits and their limitations, because that's what most people will be using.**

### 2.3.1 Derived Parameters

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

We do not to attach any physical meaning to the individual Gaussian components used in the model. While it is true that in some quasars the [O III] emission can be clearly separated in to a narrow component at the systemic redshift and a lower-amplitude, blueshifted broad component (e.g. Shen, 2016), often this decompo-

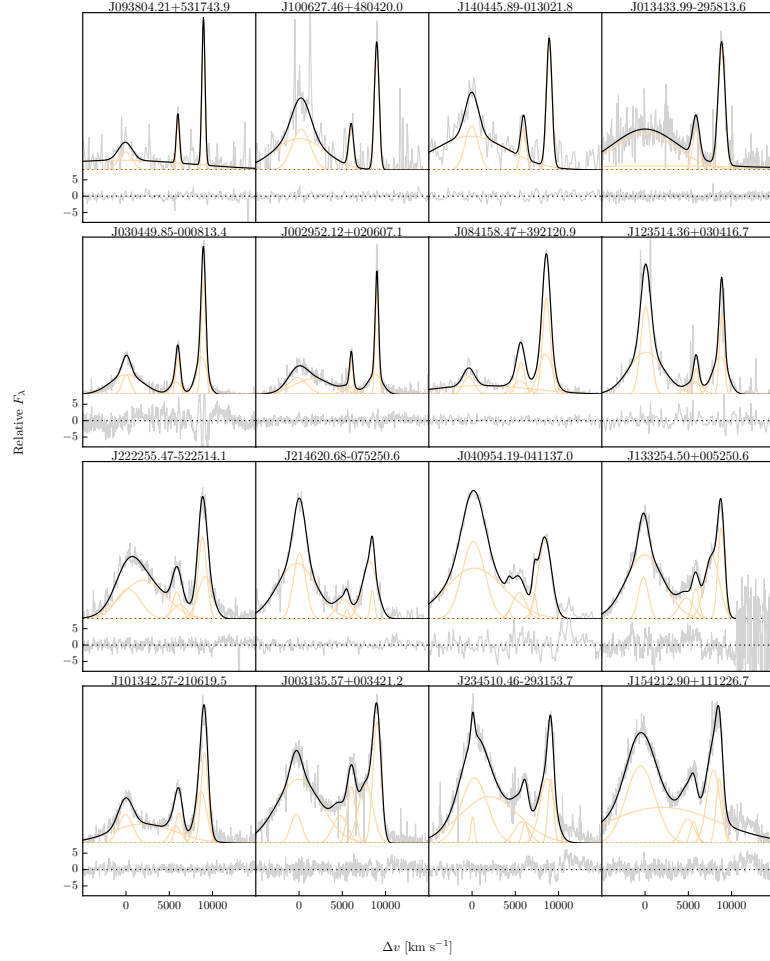


Figure 2.2: Multi-component Gaussian fits to the continuum-subtracted  $H\beta/[O\text{III}]$  emission in 16 quasars, chosen to be the representative of the wide range of  $[O\text{III}]$  line widths we measure in our sample. The data is shown in grey, the best-fitting parametric model in black, and the individual model components in orange. The broad  $H\beta$  centroid is used to measure the systemic redshift, and  $\Delta v$  is the velocity shift from the line rest-frame transition wavelength for  $H\beta$ . Below each fit we plot the data minus model residuals, scaled by the errors on the fluxes. **Resample model at higher resolution.**

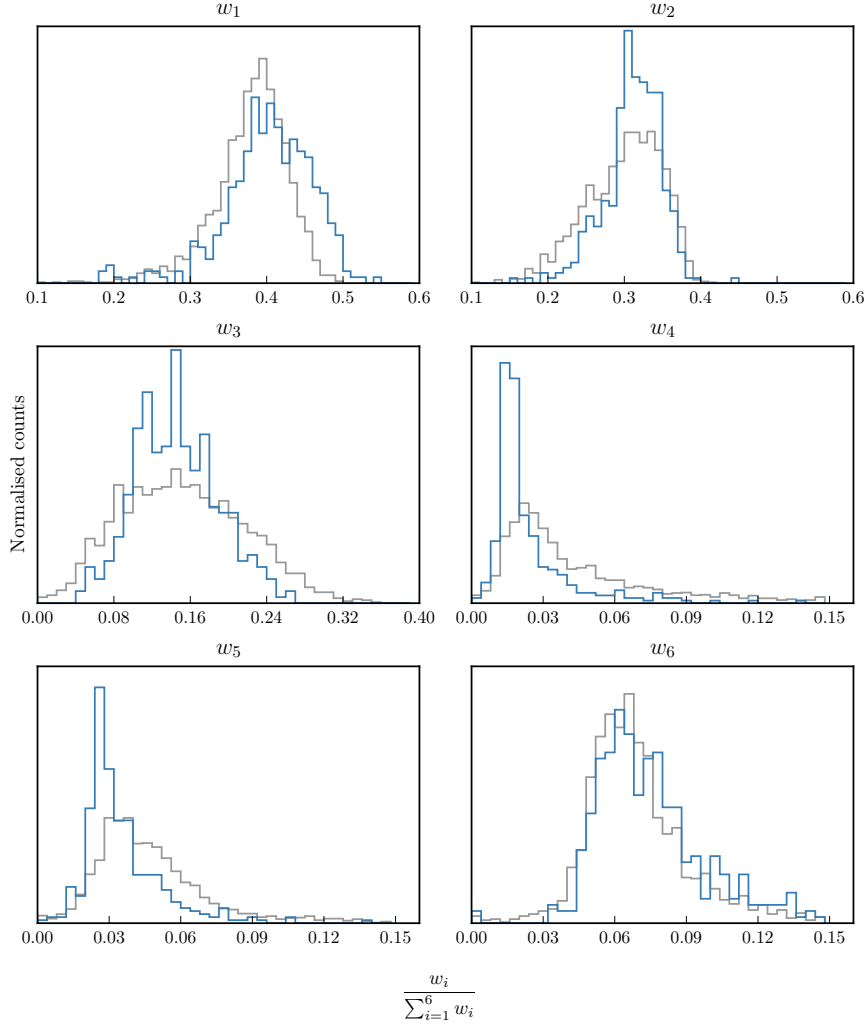


Figure 2.3: The relative weight in each of the six positive ICA components for the high-luminosity (blue) and low luminosity samples (grey). In the high-luminosity sample Fe II emission is stronger (component  $w_1$ ). The core [O III] emission is weaker (components  $w_4$ ,  $w_5$ ) but the strength of the blueshifted wing is the same ( $w_6$ ).

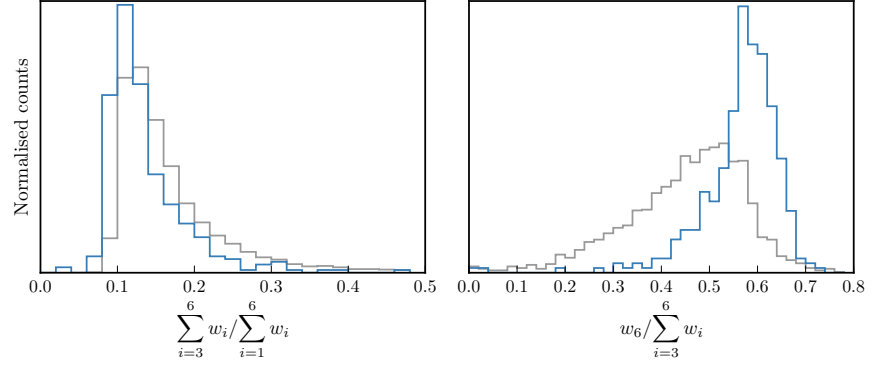


Figure 2.4: The relative weight in the three ICA components corresponding to [O III] emission (*left*) and the relative weight of the component most closely related to blueshifted [O III] emission relative to all three [O III] components (*right*). [O III] emission is weaker in the high-luminosity sample, but the relative contribution but the fractional contribution from the blueshifted component to the total [O III] emission is higher. Hence [O III] is weaker, broader, and more asymmetric in the high-luminosity sample. See Zakamska discussion.

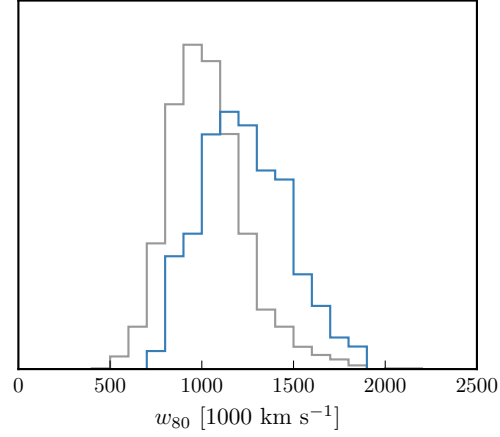


Figure 2.5: Comparison of [O III] velocity-widths in the high and low luminosity samples using the ICA component fits. If keep this need to explain in text how  $w_{80}$  is calculated from ICA component fits.



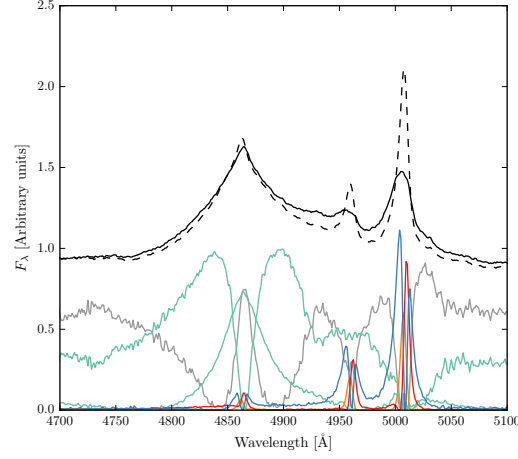


Figure 2.6: Black solid line is the median from the ICA fits to the high-luminosity sample. Black dashed line shows the median from the low-luminosity sample. The six positive ICA components are also shown.

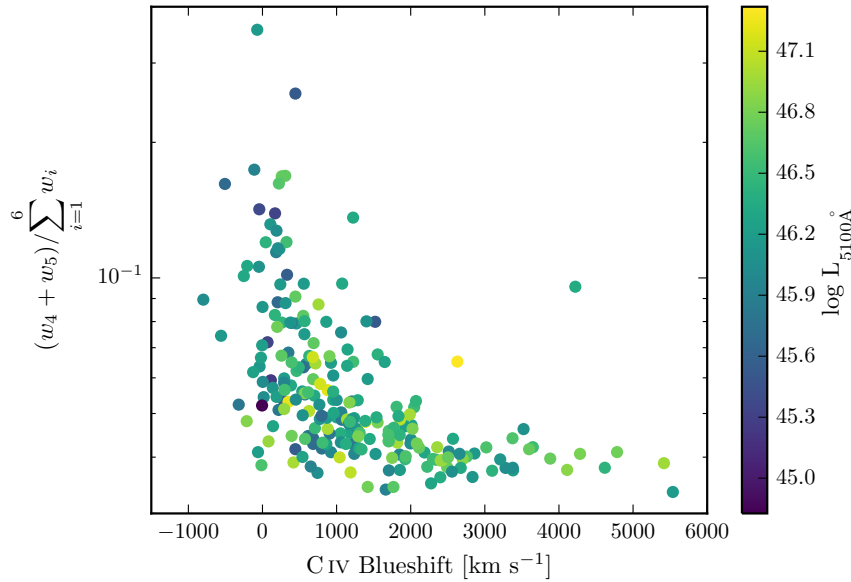


Figure 2.7: [O III] strength decreases as the C IV blueshift increases, I run in to problems comparing the C IV blueshift to the [O III] blueshift / velocity-width. See similar thing if I use [O III] EQW instead. Need to fix y ticks. Only showing the core components here. The C IV blueshift is now measured relative to the NIR ICA redshift. I think this trend is mostly being driven by the Eigenvector 1 correlations: as the blueshift increases the Fe II strength increases and the [O III] strength decreases. Doesn't appear to be driven by the luminosity. Is this tighter than EV1 trend shown with Fe/OIII strength by other authors? Is the AGN NLR absent in objects where outflows have reached kiloparsec scales, sweeping up the low-density material responsible for the [OIII]-emission?

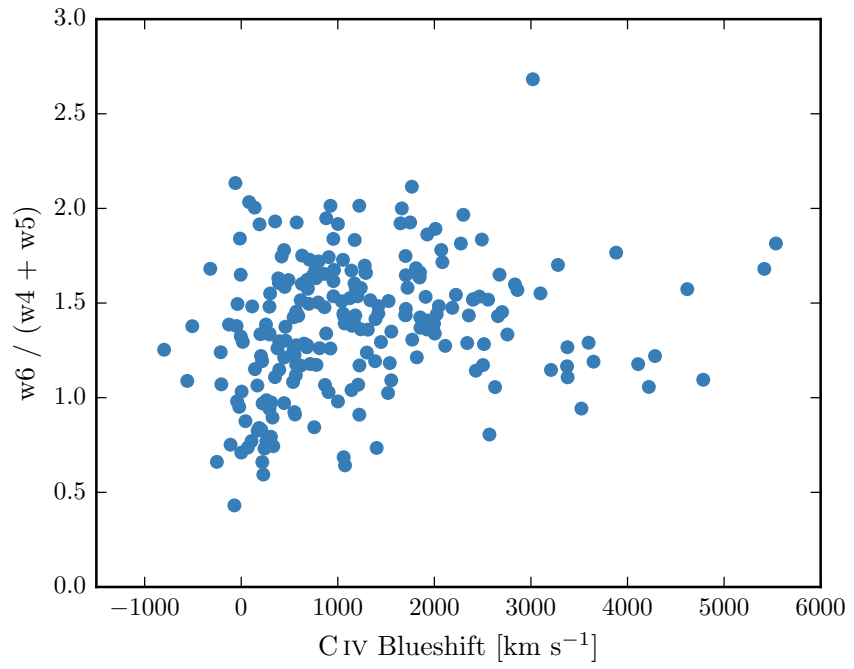


Figure 2.8: I think there is a trend here but at high blueshifts the OIII is undetected / very low S/N. Need to determin when we can believe OIII paramters. Why at low CIV blueshift is there a much bigger dynamic range than in [O III] blueshifts in Fig. 13. Is it just because we have more objects?

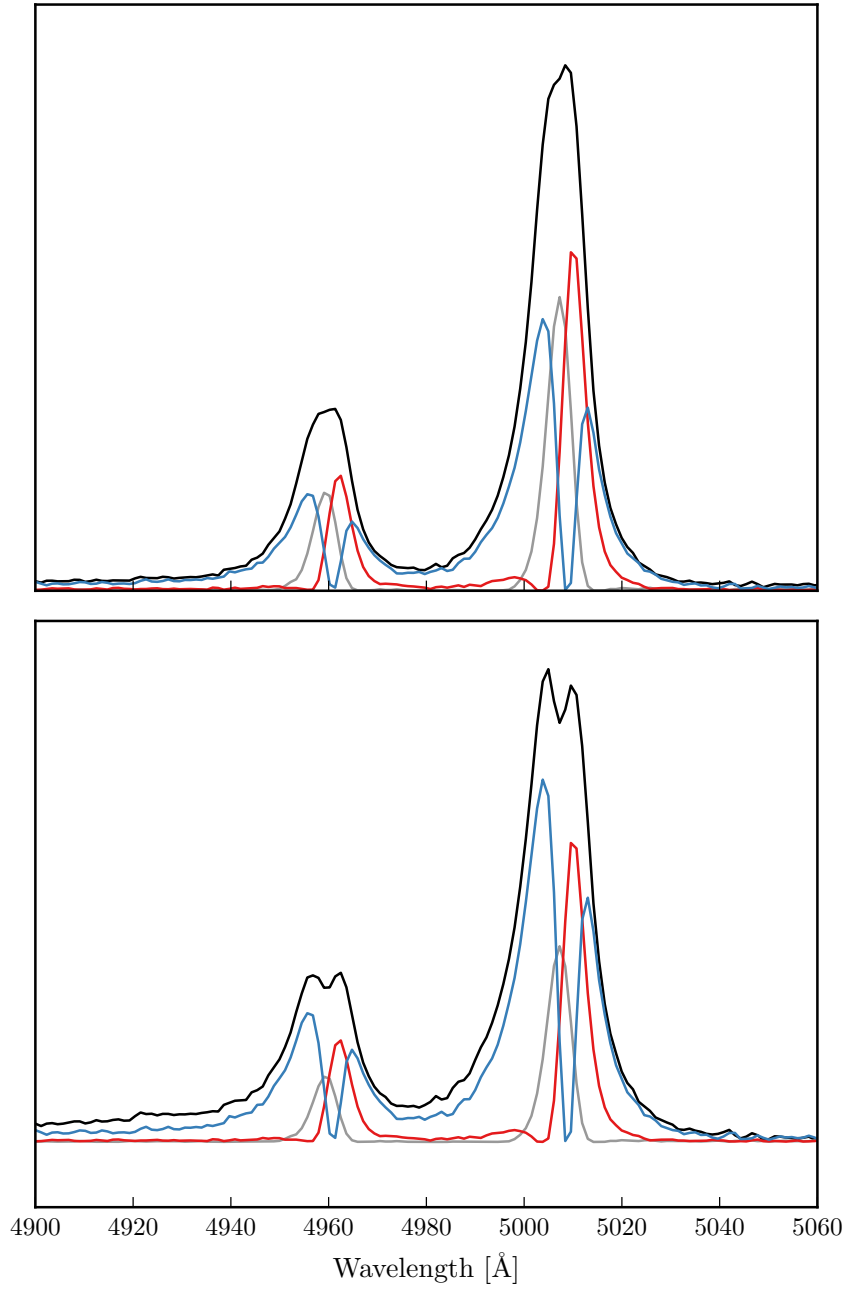


Figure 2.9: Comparison of median [O III] profiles from ICA fits to low- and high-luminosity samples.

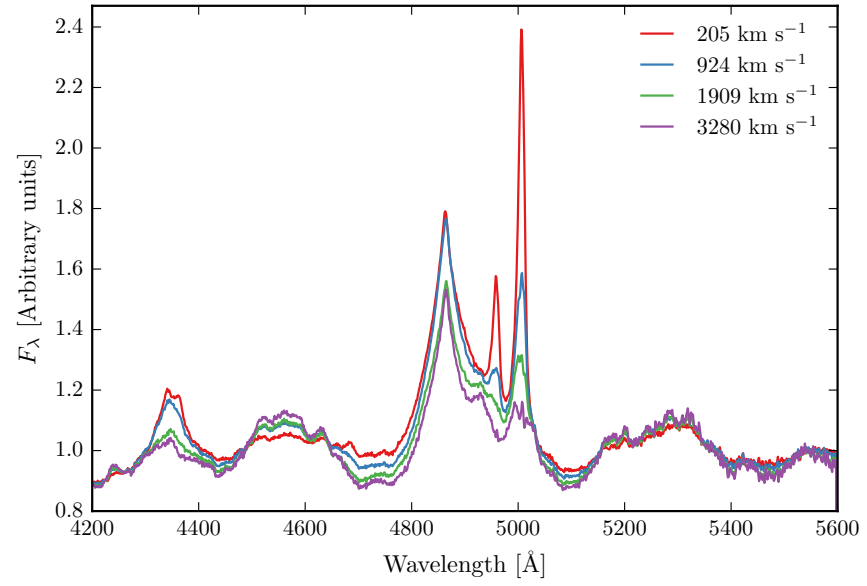


Figure 2.10: ICA median weights as a function of the CIV blueshift.

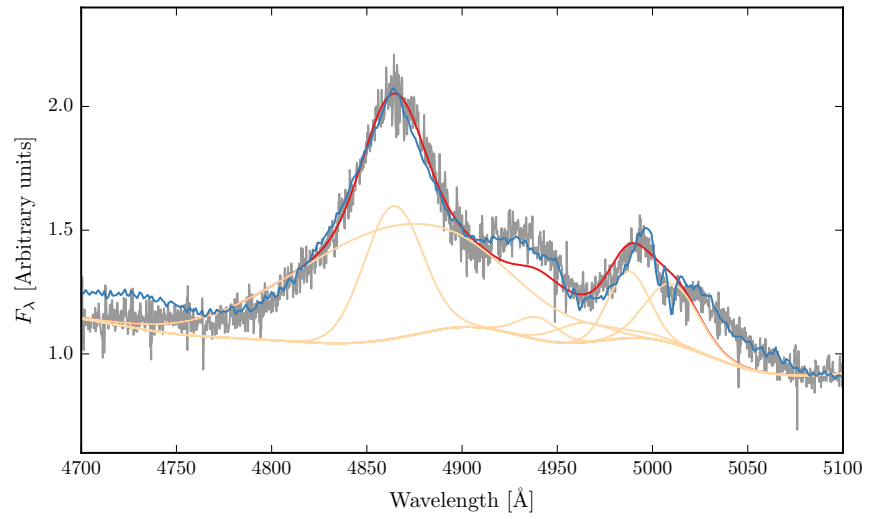


Figure 2.11: Example where poorly subtracting the iron can be confused with [O III].

sition is highly uncertain. Furthermore, often the S/N is not sufficient to statistically justify the addition of a second Gaussian component. Instead, we 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 read off. The width of the emission line can then be defined, for example, using  $w_{80} = v_{90} - v_{10}$ . The absolute asymmetry in the line profile A is defined as  $((v_{95} - v_{50}) - (v_{50} - v_5)) / (v_{95} - v_5)$  (Zakamska and Greene, 2014).

We also define the blueshift of the [O III] emission, which is a measure of the velocity shift of the profile from the expected position. This requires a measure of the observed line position, and an accurate measurement of the quasar systemic redshift. We use  $v_{10}$  to measure the location of the [O III] emission. Note that  $v_{50}$  is not suitable because when [O III] is low S/N we fit with single Gaussian.

**The line width measures are not at present corrected for instrumental broadening, but this can easily be done.**

### 2.3.2 Other flags

#### Flag 2

Low S/N. Includes some of the strong iron emitters.

#### Flag 3

74 objects where [O III] is undetected, although I don't have a rigorous definition of what this means. Merge with 2?

## 2.4 ICA COMPONENT FITS

What is the segue here? Sometimes Gaussians give poor fit? Not clear what is Hb, what is OIII, what is Fe? Show some examples of when the multi-Gaussian fits fail. Need to describe sample used in ICA component decomposition and briefly describe method (or refer to Allen & Hewett).

## 2.5 MEASURING THE QUASAR SYSTEMIC REDSHIFT

Explain flags Need to send paul redshifts

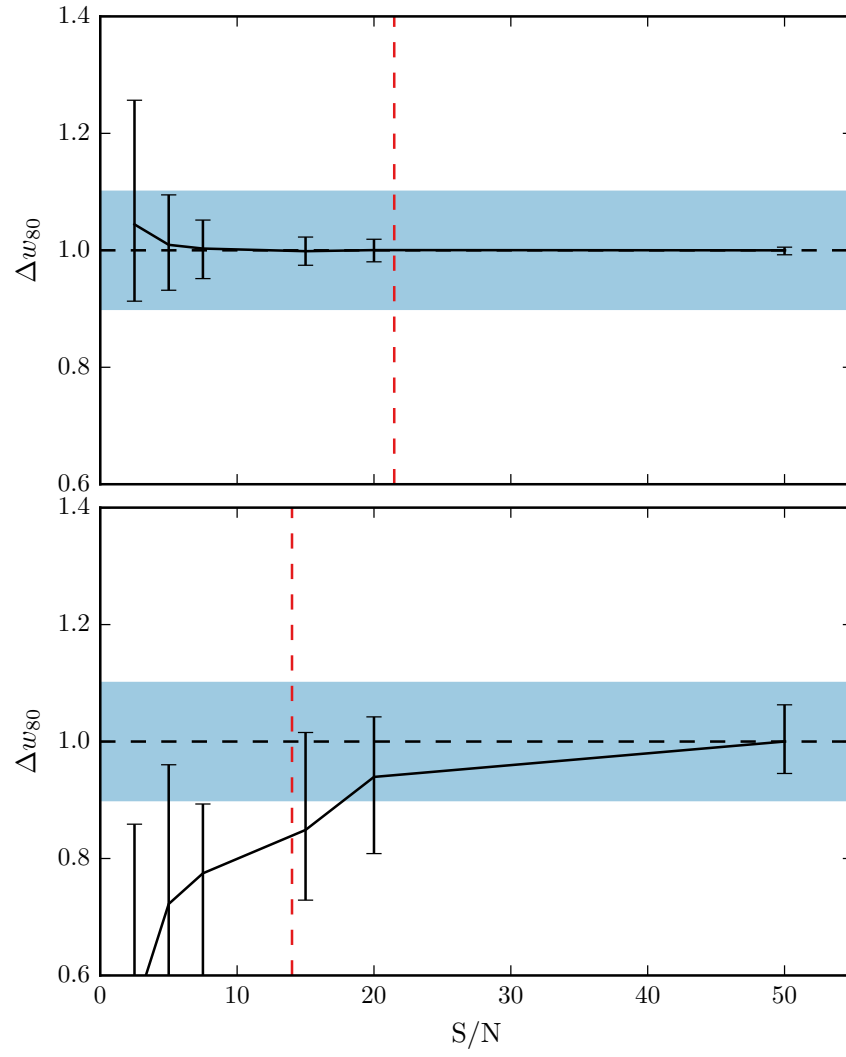


Figure 2.12

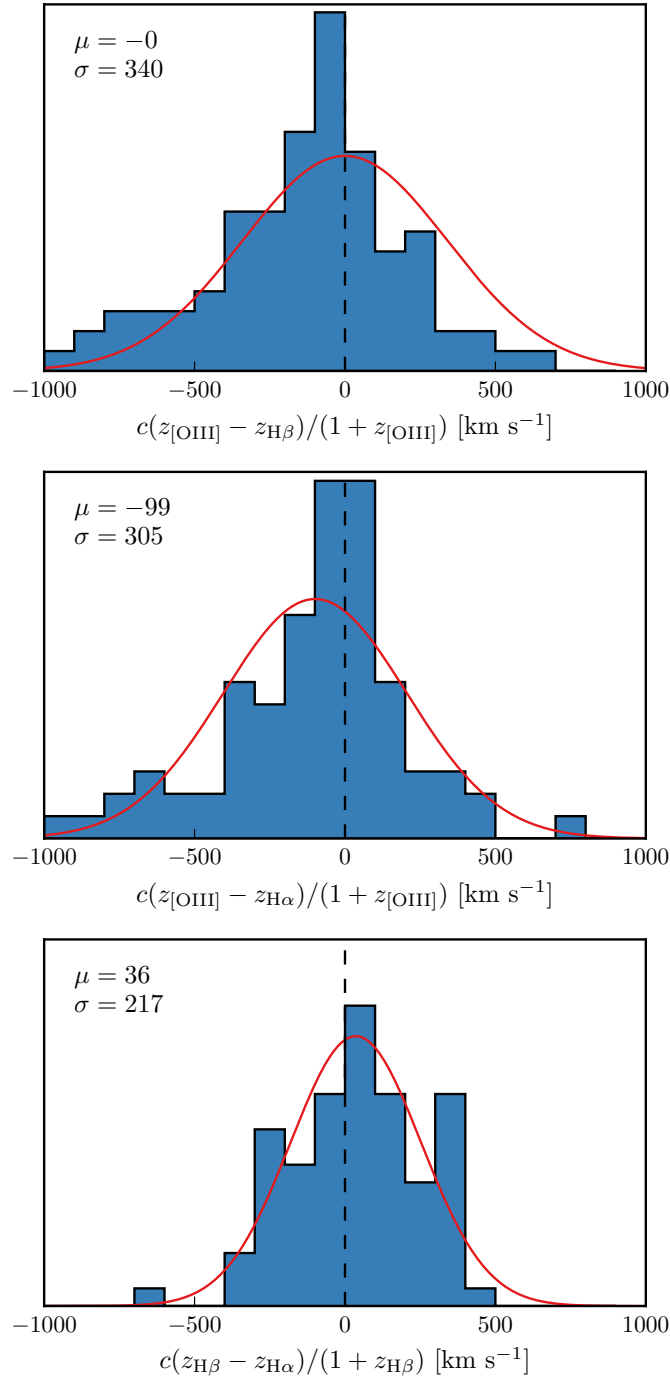


Figure 2.13: Redshift comparisons. Lots have been excluded from Ha/Hb so need to look at flags greater than one. What is the big peak? Gaussian fit to the first one has failed. Find out why these plots look different to ones in paper.

### 2.5.1 $H\alpha$

There are 224 quasars in our sample with spectra covering the  $H\alpha$  emission line. We discard seven of these from our sample because of very low S/N ( $<2.5$  measured in the  $H\alpha$  line), leaving 217. To measure the position of the line we fit a parameteric model, which is very similar to the model described in Paper I. The continuum emission is first modeled and subtracted using the procedure described in Paper I. We then test five different models with increasing degrees of freedom to model the  $H\alpha$  emission. The model we select is the simplest model for which the fractional change in the reduced chi-squared from the model with the lowest reduced chi-squared is less than ten per cent.

The models we test are: (1) a single broad Gaussian; (2) two broad Gaussians with identical velocity centroids; (3) two broad Gaussians with different velocity centroids; (4) two broad Gaussians with identical velocity centroids, and additional narrower Gaussians to model the narrow  $H\alpha$  emission, and the narrow components of  $[N\text{ II}]\lambda\lambda 6548, 6584$  and  $[S\text{ II}]\lambda\lambda 6717, 6731$ ; (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  $[N\text{ II}]$  components is fixed at the expected value of 2.96. The number of quasars fit by each model is: model 1 - 10; model 2 - 71; model 3 - 32; model 4 - 51; model 5 - 53. The redshift is then measured at the peak flux of the  $H\alpha$  model, including both the broad and narrow components of  $H\alpha$  if appropriate.

### 2.5.2 ICA

The only sensible way to measure the systemic redshift is using the NIR ICA fit.  $H\alpha$  and  $H\beta$  seem to give no systematic offset but large scatter, and they are often asymmetric, so should only use peak. And  $H\alpha$  isn't always available, and have other narrow components need to decompose.  $O\text{ III}$  peak is mostly fine to use, but then there are some objects where the whole line is blueshifted. And critically, at large CIV blueshifts the  $[O\text{ III}]$  emission is often undetected. Can show comparison of NIR ICA redshifts to... Optical ICA? Hewett & Wild?

Should emphasise that most people use  $[O\text{ III}]$  to get the most reliable systemic redshift. While this is fine at low luminosities, at high luminosities this can result in large errors (profile can be dominated by blueshifted component, Fe emission can be improperly subtracted, or  $[O\text{ III}]$  might not be detected at all. Publish ICA components with this paper?

Can also describe what I found trying to get redshifts from broad  $H\alpha$ ,  $H\beta$ ? (Narrow components generally very weak at these luminosities)



ties so can't be used.) Generally find no systematic errors but large ( $\sim 1000 \text{ km s}^{-1}$  scatter). Comparing NIR ICA to [O III] for the [O III] with high S/N I find small (few hundred  $\text{km s}^{-1}$ ) scatter.

Should publish [O III] redshifts with this paper for people to use.

### 2.5.3 *Parameter uncertainties and upper limits*

Describe how uncertainties on best-fitting parameters were calculated.

In 78 quasars, or approximately 25 per cent of or sample, the [O III] is undetected, or detected with very low S/N. In this section we describe how upper limits on the [O III] equivalent width were calculated. Firstly, the best-fitting model comprising the continuum, Fe II, and H $\beta$  emission is subtracted from the spectra, leaving behind only emission due to [O III]. From this spectra we generate 100 mock spectra, where the flux at each wavelength is randomly drawn from a Normal distribution with a mean equal to the flux convolved with a Gaussian of width  $200 \text{ km s}^{-1}$  and a width equal to the known error. We then perform an error-weighted linear least-squares regression with an [O III] template derived from a fit to a very high S/N low redshift SDSS composite spectra. The equivalent width of the best-fitting model is recorded for each of the 100 realisations of the spectra. The error in the equivalent width is defined as the root-mean-square of these values.

Calculated uncertainties using Monte Carlo. Uncertainties on  $v_{10}$  are very large, which I think makes sense since the wing Gaussian will be appearing and disappearing, giving a large dispersion in  $v_{10}$ . Or regardless  $v_{10}$  is just very sensitive to the noise. Maybe I should be using  $v_{25}$  instead?

### 2.5.4 *Absolute flux calibration of spectra and continuum luminosities*

Relative flux-calibration of the infrared spectra as a function of wavelength has been achieved, to  $\simeq 10$  per cent, through observations of appropriate flux standards. The absolute flux levels, however, can be in error by large factors due to variable atmospheric conditions combined with the narrow slit widths. For the majority of the quasars we have, therefore, established the absolute flux scale for each near-infrared spectrum using the same quasar SED-model fitting scheme employed in Paper I. Briefly, the SED-model was fit, with the normalisation and E(B-V) as free variables, to optical/infrared magnitudes, or SDSS/BOSS spectra (check order I do this.) This allows us to extrapolate from the optical when we do not have photometric data in the near-infrared. The spectra were then normalised to the SED model using a linear error-weighted least-squares regression in the the regions of the spectra covered by the H/K bands. The monochromatic contin-

uum luminosity at 5100Å was calculated directly from the normalised SED-model. **If this sounds strange can also calculate from fit to normalised spectra. Check if any missing normalisation / monochromatic luminosities.**

## 2.6 RESULTS FROM ICA FITS

Need to convince the reader that the ICA components approximately correspond to real components. Explain how non-parameteric measures derived from ICA reconstructions.

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

## 2.7 LUMINOSITY/REDSHIFT-EVOLUTION OF [O III] PROPERTIES

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

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

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

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

In Figure 2.14 we show the [O III] velocity width as a function of the [O III] luminosity and the quasar redshift. The lack of any redshift-evolution between  $z = 0$  and  $z = 1.5$  was reported by Harrison et al., (2016). Our additional data suggests that this continues to  $z \sim 2.5$ . On the other hand, at fixed redshift, we see a significant correlation between the [O III] velocity width and the luminosity.

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

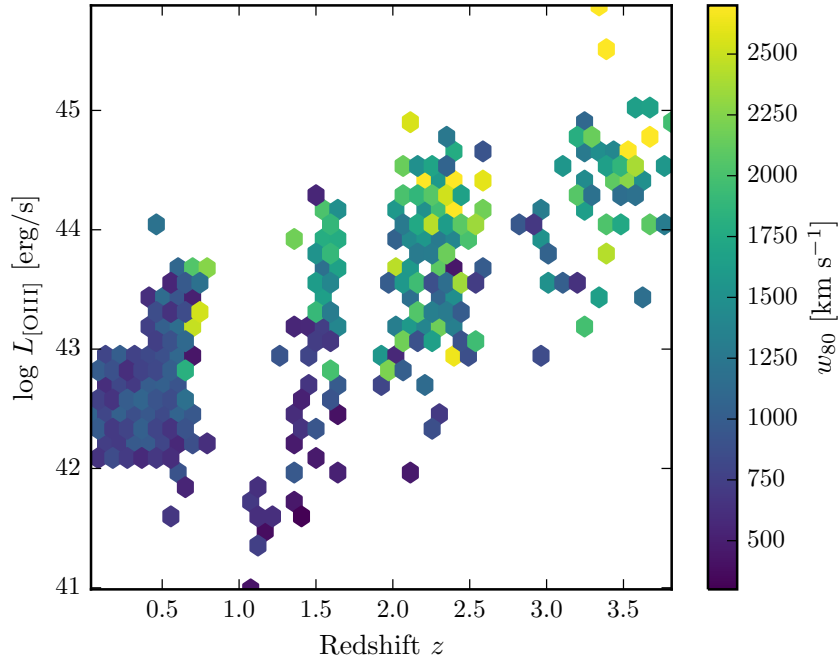


Figure 2.14: The [O III] velocity-width, characterised by  $w_{80}$ , as a function the [O III] luminosity and the quasar redshift. The color of each hexagon denotes the mean  $w_{80}$  for the objects in that luminosity-redshift bin. We have supplemented our sample with low- $z$  objects from Zakamska and Greene, (2014) and medium ( $z \sim 1.5$ ) redshift objects from Harrison et al., (2016). If I keep this plot make sure its clear which points belong to which sample.

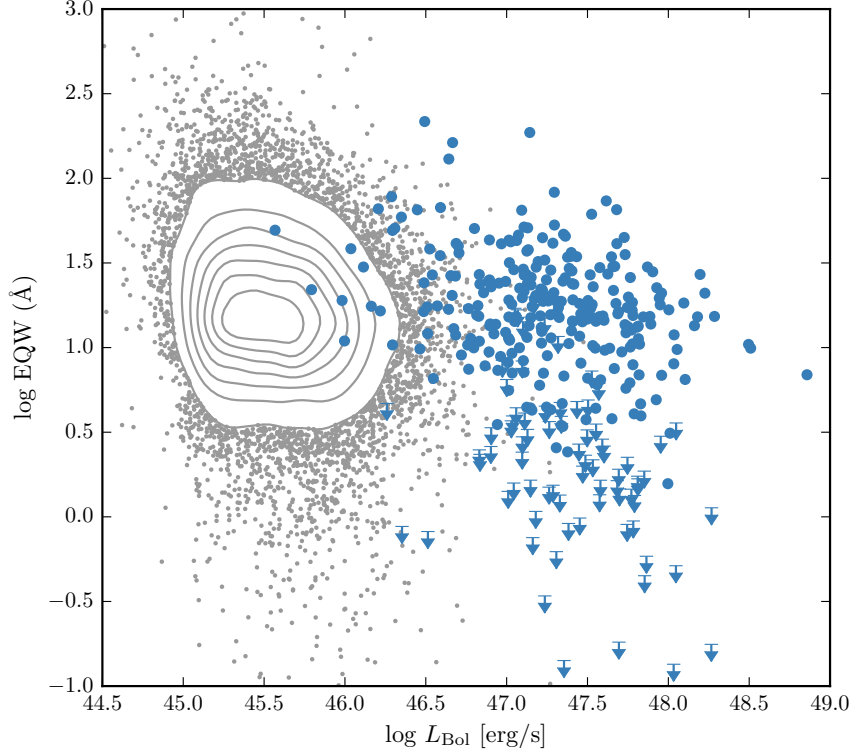


Figure 2.15: The [O III] EW as a function of the quasar bolometric luminosity for the sample presented in this paper (blue circles) and the low- $z$  SDSS sample (grey points and contours). Upper limits are denoted by the downward arrows.

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

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

## 2.8 EQUIVALENT WIDTH

In Fig. 2.15 we show the [O III]5008 EW as a function of the quasar bolometric luminosity. Bolometric luminosity is estimated from the monochromatic continuum luminosity at 5100Å using the correction

factor given by Richards et al., (2006). For comparison, we also show the low- $z$  sample from Shen et al., (2011).

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

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

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

### 2.8.1 OIII outflows

Our best-fitting profiles show a strongly blue-asymmetric profile (Fig. 2.16), with a significant fraction of the total emission in a blue wing. The luminous blueshifted broad wing and the extremely broad profile reveals high-velocity outflowing ionized gas. This can be explained if the far-side of any outflowing gas, that is moving away from the line of sight, is obscured by dust in the host galaxies (e.g. Heckman et al. 1981; Vrtilek 1985). Observations at both low and high redshifts commonly observe this blueshifted component. Our results, and those of other authors, suggest that kpc-scale outflows in ionized gas are common among the most luminous high-redshift actively accreting SMBHs.

The situation is very different in nearby AGN, where the [O III] velocity width is dominated by the galactic potential and correlates well with the stellar velocity dispersion. H I, CO and absorption line measures of the host galaxy rest frame suggest that [O III] usually gives consistent results within 200 km/s (de Robertis 1985; Whittle 1985; Wilson & Heckman 1985; Condon et al. 1985; Stripe 1990; Alloin et al. 1992; Evans et al. 2001).

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

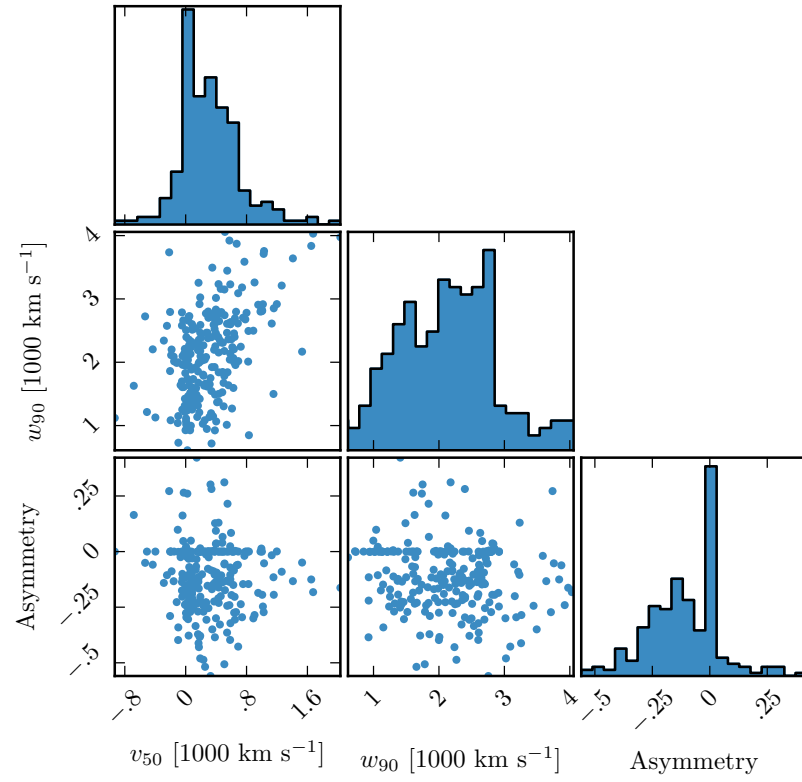


Figure 2.16: The distributions of and correlations between a subset of the non-parameteric measures we made of the best-fitting [O III] models.

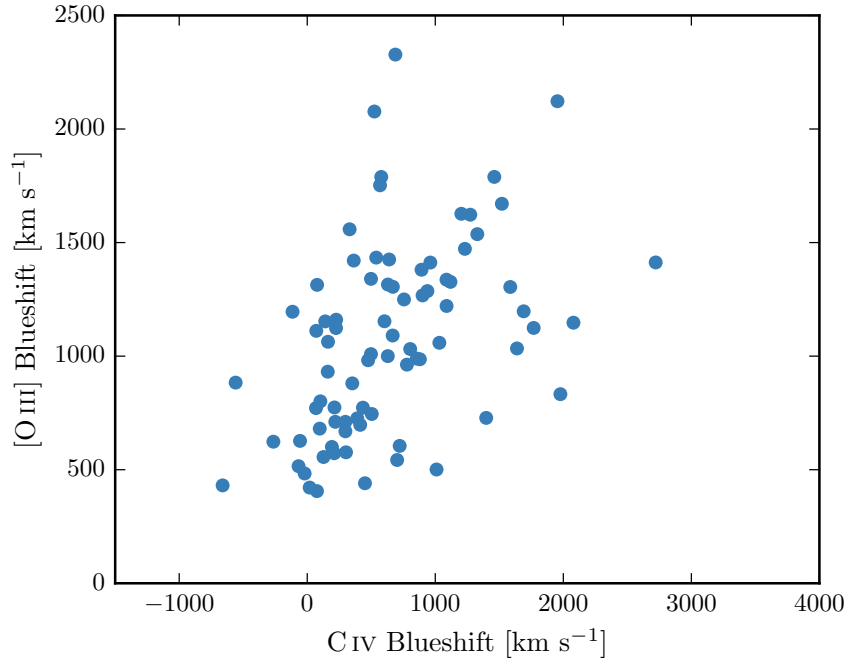


Figure 2.17: The relation between the blueshifts of C IV and [O III]. Equivalent to Fig 8. We use the H $\beta$  peak in this figure, which I think is responsible for some of the trend. However, we do see a correlation (albeit noisier) using the NIR ICA redshifts. Not a sensible to use the [O III] redshifts, since these become much more unreliable at the high C IV blueshift end (when [O III] is weaker: figure 7. Note that we are using  $v_{10}$  for the [O III] position and  $v_{50}$  for the C IV position. We can't use  $v_{50}$  for [O III] because sometimes we are using a single Gaussian, especially if the [O III] is weaker and we miss the broad component. Need to remake this plot / don't use at all because I don't believe some of the Gaussian fits to [O III], especially at high C IV blueshifts when [O III] is weak and Fe II is strong.) Only objects where fit with two components.

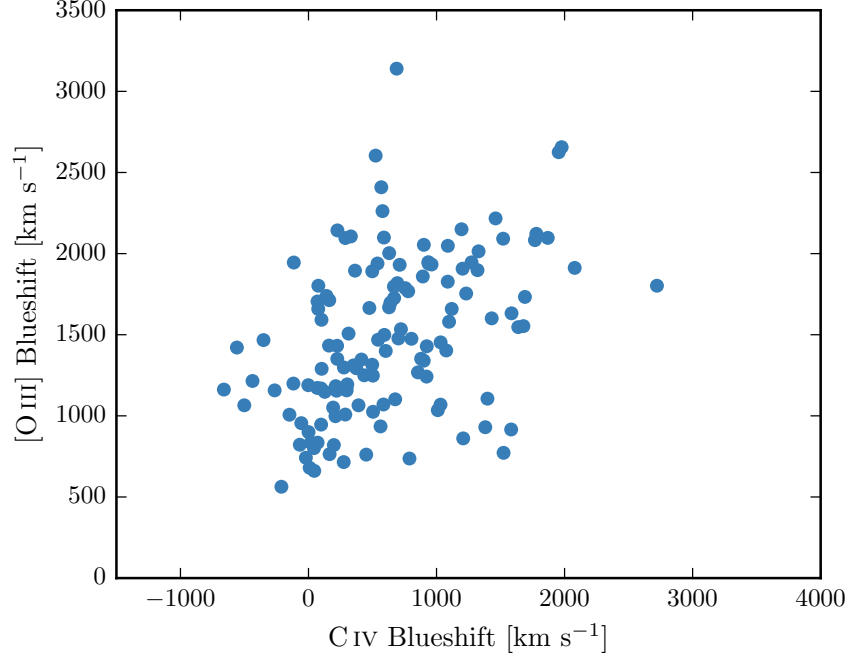


Figure 2.18

### 2.8.2 $[O\text{ III}]$ and $C\text{ IV}$ outflows are linked

As described in Paper I, we have searched for optical counterparts to our near-infrared spectra. Optical spectra are available for XXX quasars in our catalogue, and cover the broad  $C\text{ IV}$  doublet. As we described in Paper I and Coatman et al., (2016),  $C\text{ IV}$  is often blueshifted, which almost certainly signal the presence of strong outflows, most likely originating in a disc wind. In Paper I we demonstrated that the quasars in our sample cover the full range of  $C\text{ IV}$  blueshifts seen in the SDSS quasar population, which makes our sample unique in that it allows us to study properties of the quasar across the full parameter range.

The  $C\text{ IV}$  velocity centroid measurements are taken directly from paper I. We define the ‘location’ of the  $[O\text{ III}]$  emission using  $v_{10}$ , although the results are the same if  $v_{20}$ ,  $v_{50}$  etc. are used instead.

In Figure 2.17 we show the  $C\text{ IV}$  blueshifts against the  $[O\text{ III}]$  blueshifts. This comparison is done for a sub-sample of 146 objects where we have good measurements of the  $C\text{ IV}$ ,  $[O\text{ III}]$ , and  $H\beta$  (to measure the systemic redshift) profiles. Objects with  $S/N > 3$  are shown as blue filled circles and objects with  $S/N < 3$  as open grey circles. We calculated the median  $S/N$  per pixel in the best-fitting model for the  $[O\text{ III}]5008$  emission.

There is a clear and strong correlation. Similar correlations have been tentatively found in lower redshift quasars and AGN (Zamanov et al., 2002).



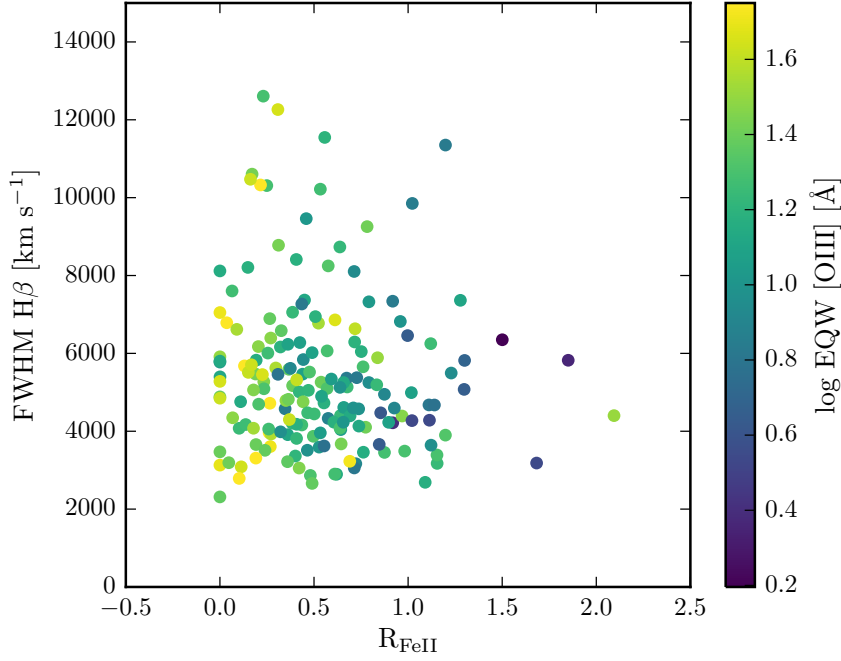


Figure 2.19: The [O III] EQW as a function of the H $\beta$  FWHM and the optical Fe II strength (EQW Fe II / EQW H $\beta$  ).

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

## 2.9 EIGENVECTOR ONE CORRELATIONS

In Figure 2.19 we show the [O III] EQW as a function of the H $\beta$  FWHM and the optical Fe II strength. The optical Fe II strength is defined as the ratio of the Fe II and H $\beta$  EQW, where the Fe II EQW is measured between 4434 and 4684Å. These parameters form part of ‘eigenvector 1’ (EV<sub>1</sub>), the first eigenvector in a principal component analysis which originated from the work of Boroson and Green, (1992). In our sample, these parameters follow very similar correlations to what is observed at low- $z$  (e.g. Shen and Ho, 2014). In particular, the anti-correlation between the [O III] and Fe II EQWs.

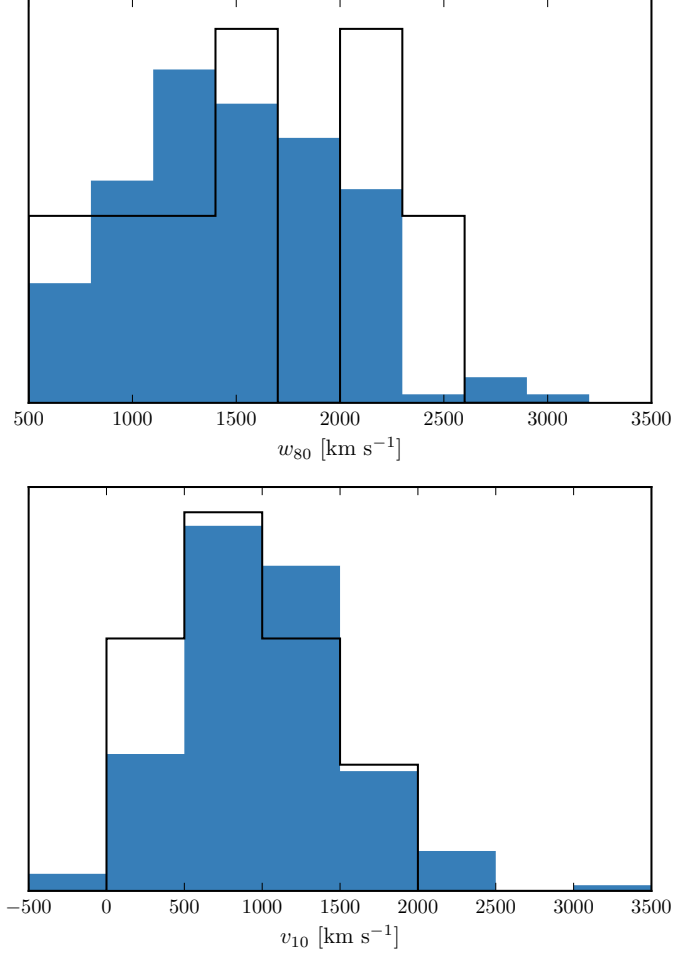


Figure 2.20: The distribution of  $w_{80}$  and  $v_{10}$  for the 19 BALs are compared to the distribution for the non-BALs. These look rubbish. Cumulative distributions instead? Try doing gaussian kernel density estimator

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

#### 2.10 MAPPING EV1 TO CIV BLUESHIFT AND EQW

#### 2.11 SIGNAL TO NOISE TESTS

#### 2.12 BROAD ABSORPTION LINE QUASARS

19 quasars in our catalogue are classified as broad absorption line (BAL) quasars, using the either the SDSS classification flags or the Allen et al., (2011) catalogue. We find that the BAL quasars have typ-

ically broader [O III] than the rest of the sample. Note that in the Zakamska et al., (2016) sample of very red quasars, the incidence of BALs is very high, and these objects have extremely broad [O III] profiles. A two-sided Kolmogorov-Smirnov statistic on the  $w_{80}$  distributions returned a p-value of 0.10. What does this mean? Try with different parameters? Histograms look rubbish so maybe just give the numbers.

## 2.13 DISCUSSION

### 2.13.1 *Type II quasars*

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

