CONTENTS

LIST OF TABLES

LISTINGS

ACRONYMS

AGN   active galactic nuclei

NLR   narrow line region

BLR   broad line region

EV1   eigenvector 1

ICA   independent component analysis

PCA   pndependent component analysis

SDSS   Sloan Digital Sky Survey

BOSS   Baryon Oscillation Spectroscopic Survey

UV   ultra-violet

EQW   equivalent width

S/N   signal-to-noise

BH   black hole

SED   spectral energy distribution

NIR   near-infrared

FWHM full-width-at-half-maximum

NARROW LINE REGION PROPERTIES

1.1 INTRODUCTION

X-ray and ultra-violet (UV) spectroscopy reveal high velocity outflows to be nearly ubiquitous on sub-parsec scales in high accretion rate AGN. Models of galaxy evolution that invoke AGN feedback require these outflows to reach galactic scales and quench star formation in the AGN host galaxy. In recent years, a huge amount of resources have been devoted to searching for observational evidence of galaxy-wide, AGN-driven outflows. This has resulted in recent detections of winds in AGN-host galaxies using tracers of atomic, molecular, and ionised gas (e.g. Nesvadba et al., 2006; Arav et al., 2008; Nesvadba et al., 2008; Moe et al., 2009; Dunn et al., 2010; Alexander et al., 2010; Harrison et al., 2012; Harrison et al., 2014; Nesvadba et al., 2010; Rupke and Veilleux, 2013; Veilleux et al., 2013; Nardini et al., 2015; Feruglio et al., 2010; Alatalo et al., 2011; Cimatti et al., 2013; Cicone et al., 2014).

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

Observations of broad velocity-widths and blueshifts in narrow emission lines stretch back several decades (e.g. Weedman, 1970; Stockton, 1976; Heckman et al., 1981; Veron, 1981; Feldman et al., 1982; Heckman, Miley, and Green, 1984; Vrtilek, 1985; Whittle, 1985; Boroson and Green, 1992). However, these studies rely on small samples, which are often unrepresentative of the properties of the population. More recently, the advent of large optical spectroscopic surveys (e.g. SDSS) have facilitated studies of the NLR in tens of thousands of AGN (e.g. Boroson, 2005; Greene and Ho, 2005; Zhang et al., 2011; Mullaney et al., 2013; Zakamska and Greene, 2014; Shen and Ho, 2014). This has provided constraints on the prevalence and drivers of ionised

outflows. At the same time, there is strong evidence from spatially resolved spectroscopic observations that these outflows are extended over galaxy scales (e.g. Greene et al., 2009; Greene et al., 2011; Hainline et al., 2013; Harrison et al., 2012; Harrison et al., 2014).

However, these studies do not cover the redshift range when star formation and black hole (BH) accretion peaked, and consequently when feedback is predicted to be strongest. At these redshifts the bright optical emission lines are redshifted to near-infrared wavelengths, where observations are much more challenging. As a consequence, studies at high redshifts have typically relied on relatively small numbers of objects (e.g. Netzer et al., 2004; Sulentic et al., 2004; Shen, 2016). These studies find [O III] to be broader in more luminous AGN, suggesting that AGN efficiency in driving galaxy-wide outflows increases with luminosity (e.g. Netzer et al., 2004; Nesvadba et al., 2008; Kim et al., 2013; Brusa et al., 2015; Carniani et al., 2015; Perna et al., 2015; Bischetti et al., 2016). In addition, [O III] is often very weak, or is missing entirely in luminous AGN (e.g. Netzer et al., 2004).

Other recent studies have looked at the [O III] emission properties of extreme objects - e.g. heavily obscured quasars (Zakamska et al., 2016) and the most luminous quasars (Bischetti et al., 2016) - at redshifts $z \sim 2$. The [O III] emission in these objects is very broad and strongly blueshifted. These observations are consistent with galaxy formation models that predict AGN feedback to be strongest in luminous, dust-obscured quasars.

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

1.2   QUASAR SAMPLE

From our near-infrared (NIR) spectroscopic catalogue (Chapter **??**), we have selected 356 quasars which have spectra covering the strong, narrow [O III] doublet. The broad Balmer Hβ line is also observed for all but two of the sample. In 165 the spectra extend to the broad Hα emission at 6565Å, and in 260 optical spectra including C IV are also available (mostly from SDSS/Baryon Oscillation Spectroscopic Survey (BOSS)). The sample, which has a redshift range $1.5 < z < 4$, is summarised in Table 1.1.

Table 1.1: The numbers of quasars with [O III] line measurements and the spectrographs and telescopes used to obtain the near-infrared spectra.

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

1.3 PARAMETRIC MODEL FITS

In this section we describe how parameters of the [O III] emission are derived. Our approach is to model the spectra using a power-law continuum, an empirical Fe II template and multiple Gaussian components to model the emission from the broad and narrow components of Hβ and the [O III] doublet. This is a model which is commonly adopted in the literature (e.g. Shen et al., 2011). We then derive emission line parameters from the best-fitting models. The model-fitting is procedure is more robust when analysing spectra with limited S/N (in comparison to measuring line properties directly from the data) and allows the emission from different transitions to be isolated.

1.3.1 *Description of model*

Before a spectrum can be modelled, it must first be transformed to the quasar rest-frame. The redshift used in this transformation is either derived from the peak of the broad Hα emission (~ 40 per cent of our sample), from the peak of the broad Hβ emission (~ 40 per cent) or from the peak of the narrow [O III] emission (20 per cent). The rest-frame transformation is only required to be accurate to within $\sim 1000 \mathrm{km\,s}^{-1}$ for our fitting procedure to work. In later sections, more precise estimates of the systemic redshift will be calculated using our parametric model fits to the [O III] emission.

The continuum and Fe II emission is first modelled and subtracted using the procedure described in Chapter **??**. The Hβ and [O III]

emission is then fit, again using the procedure described in Chapter **??**. However, we make a number of modifications to the parametric model employed in the fit, which we now describe.

In general, Hβ is modelled by two Gaussians with non-negative amplitudes and FWHM greater than $1200 \mathrm{km\,s}^{-1}$. In 10 objects Hβ is modelled with a single Gaussian and in 41 objects Hβ is modelled with two Gaussians, but the velocity centroids of the two Gaussians are constrained to be equal. These spectra generally have low S/N, and adding extra freedom to the model does not significantly decrease the minimised reduced χ^2. In addition there are cases where the blue wing of the Hβ emission is below the lower wavelength limit of the spectrograph; in these cases models with more freedom are insufficiently constrained by the data.

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

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

In 71 objects [O III] is undetected, or is detected with very low S/N. In these cases we do not attempt to measure the width of the [O III] emission, but determine only the normalisation of a fixed [O III] template. To generate this template we generate a median composite spectrum from the continuum- and Fe II-subtracted spectra of the XX quasars with reliable [O III] line measurements (see below). We then run our line-fitting routine on the composite spectrum, and use the best-fitting [O III] model as a template.

Some example fits are shown in Figure 1.1
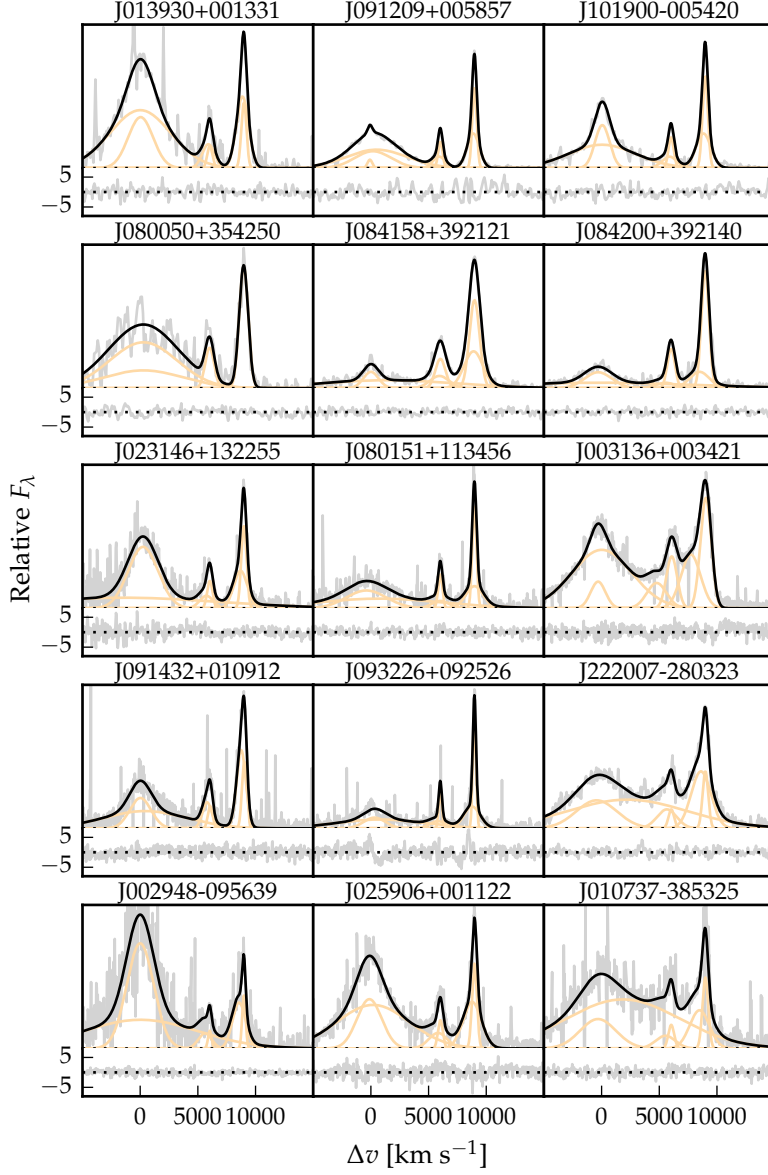
Figure 1.1: Model fits to the continuum- and Fe II-subtracted Hβ/[O III] emission in 15 quasars, chosen at random. The data is shown in grey, the best-fitting model in black, and the individual model components in orange. The peak of the [O III] emission is used to set the redshift, and Δv is the velocity shift from the rest-frame transition wavelength of Hβ . Below each spectrum we plot the data minus model residuals, scaled by the errors on the fluxes.

1.3.2 *Derived parameters*

All [O III] line properties are derived from the [O III]λ5008 emission, but, as described above, the kinematics of [O III]λ4960 are constrained by our fitting routine to be identical.

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

We therefore choose to characterize the [O III] line profile using a number of non-parametric measures, which are commonly used in the literature (e.g. Zakamska and Greene, 2014; Zakamska et al., 2016). A normalised cumulative velocity distribution is constructed from the best-fitting model, from which the velocities below which 5, 10, 25, 50, 75, 90, and 95 per cent of the total flux accumulates can be calculated. The width of the emission line can then be defined using either $w50$ ($\equiv v_{75} - v_{25}$), $w80$ ($\equiv v_{90} - v_{10}$) or $w90$ ($\equiv v_{95} - v_{5}$). In terms of the FWHM, $w_{50} \simeq \text{FWHM}/1.746$, $w_{80} \simeq \text{FWHM}/0.919$, $w_{90} \simeq \text{FWHM}/0.716$. Line-width measures are corrected for instrumental broadening by subtracting the resolution of the spectrograph (Table **??**) in quadrature. We also define the absolute asymmetry in the line profile A as $((v_{95} - v_{50}) - (v_{50} - v_{5}))/(v_{95} - v_{5})$ (e.g. Zakamska and Greene, 2014).

1.3.3 *Flux calibration of spectra*

We have established the absolute flux scale for each NIR spectrum using a similar methodology to the one described in Chapter **??**. If near-infrared photometric data was available for our whole sample then this could be used to establish the absolute flux scale of the near-infrared spectra. Because this information is missing for a sizeable fraction of our sample, we instead consider two approaches.

In the first approach, we leverage the excellent flux-calibration of the SDSS/BOSS spectra, which are available for XX objects in our sample. We use our standard quasar spectral energy distribution (SED) model (Chapter **??**) to bridge the gap between the wavelength coverage of the near-infrared and optical SDSS/BOSS spectra. The quasar SED model is fit to the SDSS/BOSS spectra, with the normalisation and extinction E(B-V) as free parameters. The near-infrared spectra are then normalised to the SED model using a linear error weighted least-squares regression. The second approach is identical, except rather than using the SDSS/BOSS spectra we fit the same SED model to the

available photometric data from the optical (SDSS) to the near-infrared (VHS, Viking, UKIDSS, 2MASS).

The monochromatic continuum luminosity - used as a proxy for the AGN bolometric luminosity - is also used below. This is calculated in the same way as in Chapter **??**.

1.3.4 *Reliability of derived parameters*

1.3.4.1 *Removal of Fe II emission*

We encountered 23 cases where the relative strengths of the Fe II lines appear to differ significantly from those of I Zw 1 on which the Fe II template we use is based. As a result, significant Fe II flux remained in the spectra after the removal process. This emission is at rest-frame wavelengths very close to the [O III] emission, and so could potentially lead to large errors in the inferred [O III] line parameters. In Figure 1.2 we plot the spectral region around [O III] these 23 objects. The vertical lines indicate the expected positions of the [O III] doublet, with zero velocity defined using the peak of the broad Hβ emission. [O III] is generally extremely weak in these objects. As a result, fitting multiple Gaussians will tend to fit the Fe II emission as broad, shifted [O III]. For example, J125141+080718 was studied by Shen, (2016), and assigned an extremely large [O III] blueshift. Our analysis suggests that this emission is more likely to be Fe II. Because of the difficulty measuring the [O III] properties of these objects, they are excluded from subsequent analysis.

1.3.4.2 *Low EQW [O III]*

In Figure 1.3 we show the distribution of the [O III] rest-frame EQW distribution for the 330 objects in our sample (objects where Fe II emission has been sub-optimally removed are excluded). In many objects [O III] is undetected. In others it is detected, but is too weak for its shape (i.e. the width and asymmetry) to be measured reliably. We define EQW= 8Å as the limit below which we can no longer reliably determine the shape of the [O III] emission. Objects with EQW< 8Å (120) are excluded in subsequent analysis of the [O III] shape.

1.3.4.3 *Low S/N [O III]*

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

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

Check if any missing normalisation / monochromatic luminosities.
Check factor of (1+z) in luminosity calculation

Can I justify quantitatively why this limit is chosen?
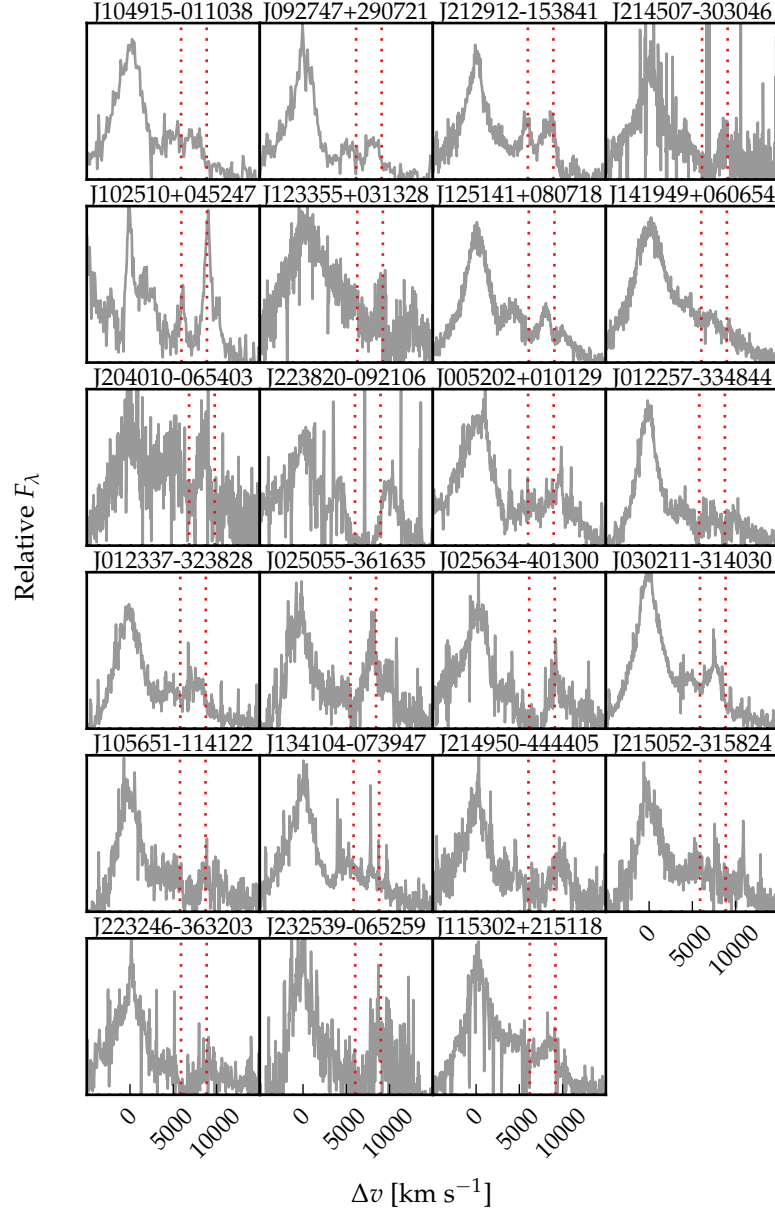
Figure 1.2: Spectra of the 23 objects for which significant Fe II emission is still visible following our Fe II-subtraction procedure. The vertical lines indicate the expected positions of the [O III] doublet (which is generally very weak) with the systemic redshift defined using the peak of the broad Hβ emission.
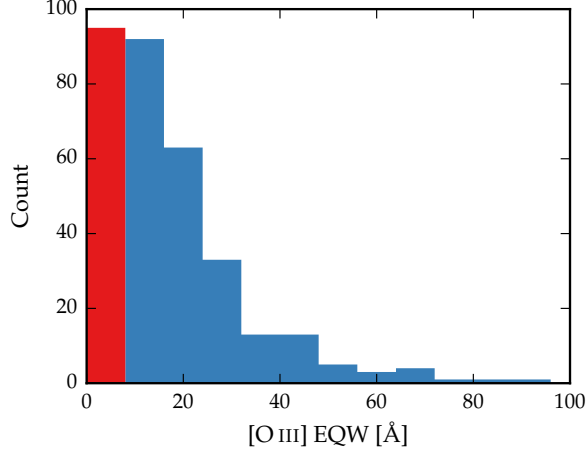
Figure 1.3: [O III] EQW. The [O III] profiles of the 120 objects in the red bin (EQW < 8Å) cannot be measured reliably using our model-fitting procedure.
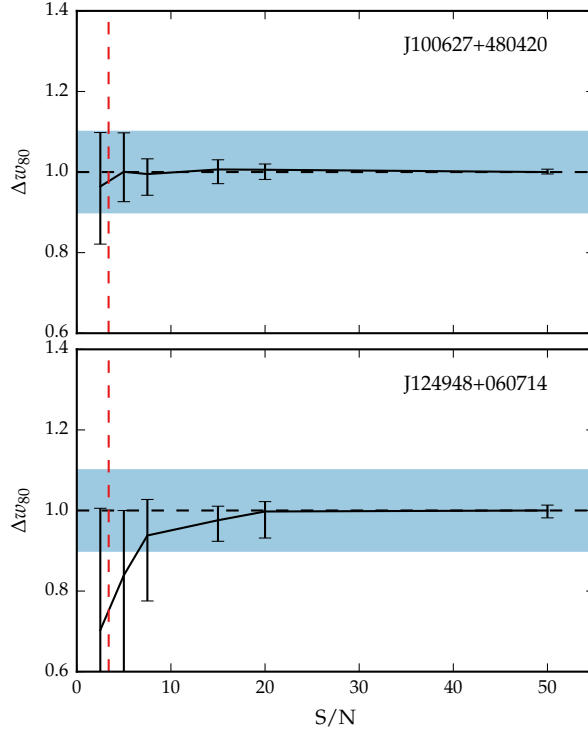
Figure 1.4: Fractional change in the [O III] line parameter w_{80} as a function of the S/N of six simulated spectra. At each S/N, our line-fitting procedure is run on 100 mock spectra, and the points and error-bars indicate the 50th percentile and 16-84 percentile range respectively. The w_{80} parameter is considered 'unreliable' if $|\Delta w_{80}| > 0.1$ at the S/N of the real spectrum (represented by the vertical line). Our w_{80} measurement is 'reliable' for J100627+480420 and 'unreliable' for J124948+060714.

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

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

4. We run our line-fitting procedure on each of the 100 mock spectra and record the value of w_{80} in the best-fitting model.

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

6. We calculate the the median w_{80} value at S/N= 50 and at the S/N of the real spectrum (by linearly interpolating between the results of our simulations). The low S/N flag is assigned to the object if the median w_{80} changes by more than 10 per cent between these two S/N realisations.

Examples of this test for two different objects are shown in Figure 1.4. The marker denotes the 50th percentile, and the lower and upper error bars the 16th and 84th percentiles respectively. As expected, the uncertainty on w_{80} increases as the S/N of the spectrum decreases. The S/N in to two spectra are similar, but the [O III] line in the first object is stronger. Hence it can still be measured reliably even at low S/N. The first object would passes our S/N cut, whereas the latter fails it.

Need some discussion on how much v95 and v05 in particular depend on SNR. You have some discussion about making sure w80 is robust to SNR changes. Did you also do similar tests for v95, v05, v10 etc?

1.4 RELIABILITY OF REDSHIFT ESTIMATES

In this section we do a comparison of systemic redshift estimates from [O III], broad Hβ and Hα. This is an important issue. Accurate systemic redshift estimates are essential in a number of applications, and researchers have devoted a large amount of telescope time to obtaining near-infrared spectra to access [O III] for this purpose. HI, CO and absorption line measures of the host galaxy rest frame suggest that [O III] usually gives consistent results within 200km s^{-1}(de Robertis 1985; Whittle 1985; Wilson & Heckman 1985; Condon et al. 1985; Stripe 1990; Alloin et al. 1992; Evans et al. 2001). However, our work shows that at high luminosities this can result in large errors (profile can be dominated by blueshifted component, Fe II emission can be improperly subtracted, or [O III] might not be detected at all.

1.4.1 *Modelling Hα*

There are 224 quasars in our sample with spectra covering the Hα emission line. We discard seven of these from our sample because
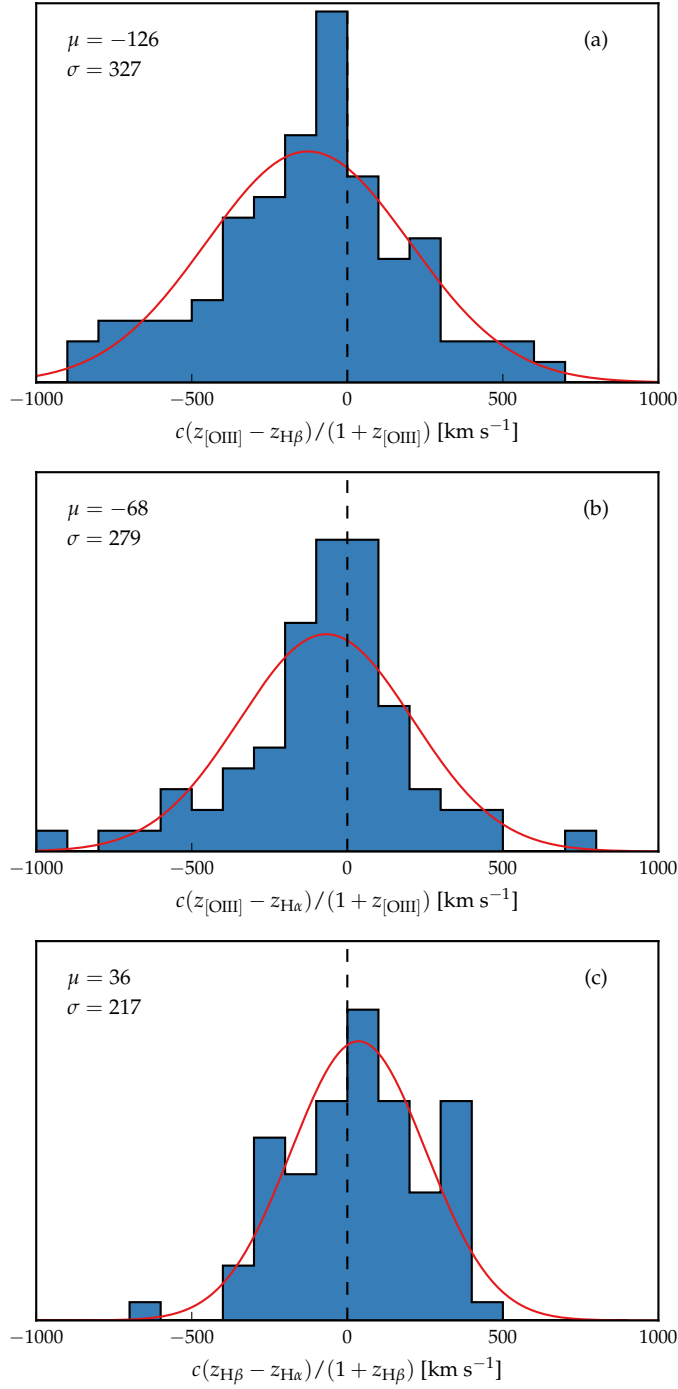
Figure 1.5: Comparison of systemic redshift estimates using [O III], broad Hβ and broad Hα. In all cases the line location is defined as the peak wavelength of the best-fitting model. There are 93, 73 and 83 objects in (a), (b) and (c) respectively. The mean and standard deviation have been calculated by fitting a Gaussian function (shown in red) to the distribution. Need to look at flags>1 for Hα and Hβ . Is peak around zero in (a) real? Is it clear in text which quasars are going in to each subfigure?

Table 1.2: Models used for Hα emission.

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

of very low S/N (<2.5 measured in the Hα line), leaving 217 To measure the position of the line we fit a parametric model, which is very similar to the model described in Chapter **??**. The continuum emission is first modeled and subtracted using the procedure described in Chapter **??**. We then test five different models with increasing degrees of freedom to model the Hα emission. The models are summarised in Table 1.2. They are (1) a single broad Gaussian; (2) two broad Gaussians with identical velocity centroids; (3) two broad Gaussians with different velocity centroids; (4) two broad Gaussians with identical velocity centroids, and additional narrower Gaussians to model the narrow Hα emission, and the narrow components of [N II]λλ6548,6584 and [S II]λλ6717,6731; (5) two broad Gaussians with different velocity centroids, and additional narrower Gaussians. If used, the width and velocity of all narrow components are set to be equal in the fit, and the relative flux ratio of the two [N II] components is fixed at the expected value of 2.96. The model we select is the simplest model for which the fractional change in the reduced χ^2 from the model with the lowest reduced χ^2 is less than ten per cent. The redshift is then measured at the peak flux of the Hα model, including both the broad and narrow components of Hα if appropriate.

1.5   RESULTS

In our sample of XX quasars, there is a huge diversity in [O III] emission properties (Fig. 1.1). In a significant fraction [O III] is undetected, whereas in others the EQW is in excess of 100Å. The median is XX, which is somewhat lower than is found in lower-redshift, lower-luminosity AGN. w_{80} varies between 400 and 3000 $\mathrm{km\,s^{-1}}$, with a median 1500 $\mathrm{km\,s^{-1}}$. When the broad wing is detected, it is almost ubiquitously blueshifted. The luminous blueshifted broad wing and the extremely broad profile reveals high-velocity outflowing ionized gas. Our results therefore suggest that kilo-parsec-scale outflows in ionized gas are common in this sample of high-luminosity, high-redshift quasars.

In Figure 1.6 we show correlations between the [O III] w_{80}, EQW, asymmetry for 119 quasars: objects with low EQW, poor S/N, poor
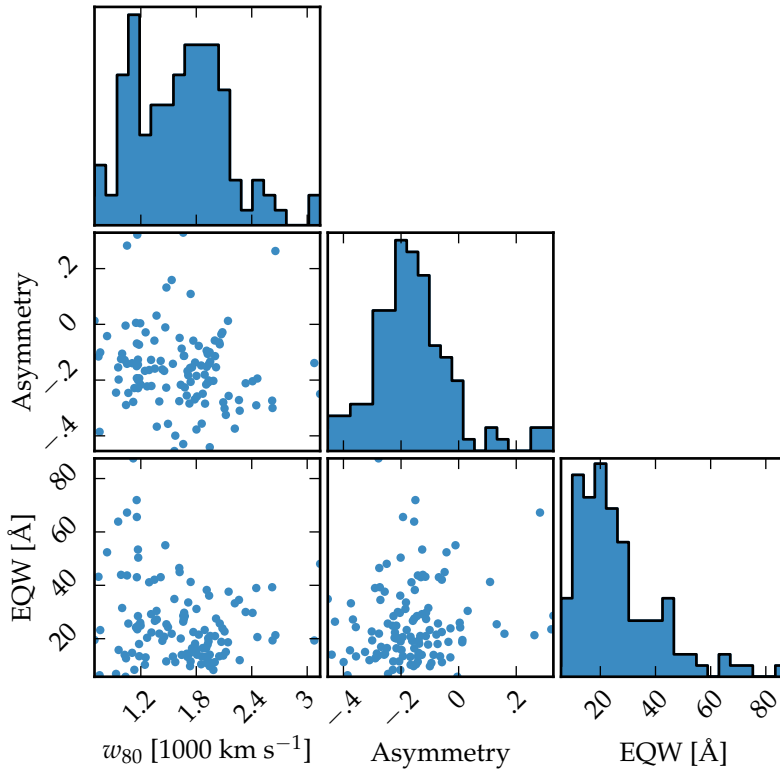
Figure 1.6: Correlations between the line width w_{80}, asymmetry R and EQW of [O III].
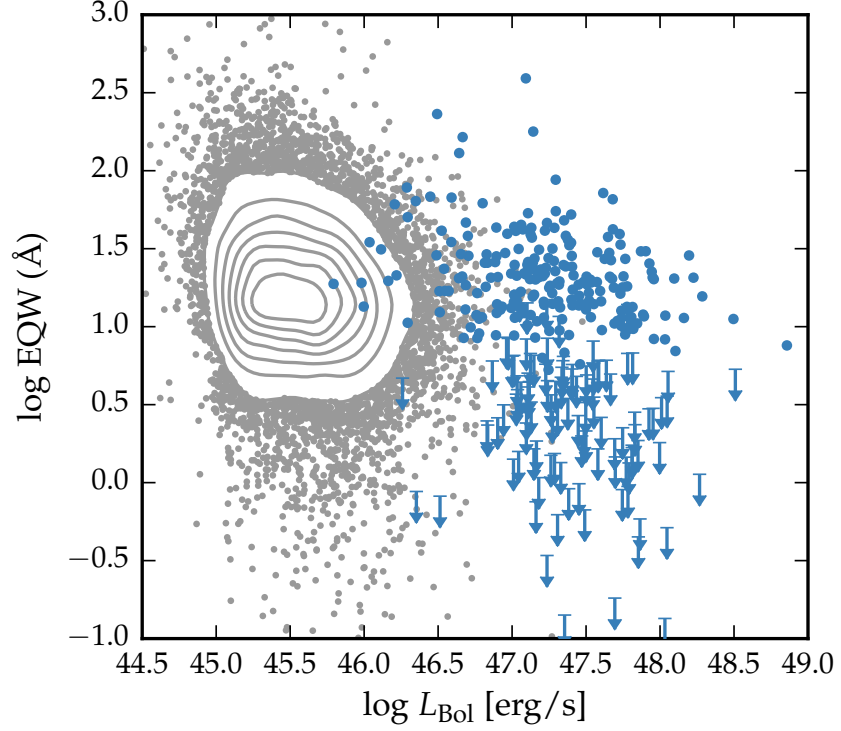
Figure 1.7: The [O III] EQW as a function of the quasar bolometric luminosity for the sample presented in this chapter (blue circles) and the low-z SDSS sample (grey points and contours). Upper limits are denoted by the downward arrows.

Fe II subtraction are not included. Objects for which [O III] is modelled using a single Gaussian are also excluded, because the asymmetry is by definition zero for these objects. We see a correlation between the [O III] velocity width and asymmetry. As the line gets broader it gets more blue-asymmetric. One interpretation of this is that the strength of the narrow core is decreasing, leading to a broader and more blueshifted profile (e.g. Shen and Ho, 2014).

1.5.1 *Equivalent width*

First we discuss how upper limits on the [O III] EQW are calculated. Firstly, the best-fitting model comprising the continuum, Fe II, and Hβ emission is subtracted from the spectra, leaving behind only emission due to [O III]. From this spectra we generate 100 mock spectra, where the flux at each wavelength is randomly drawn from a Normal distribution with a mean equal to the flux convolved with a Gaussian of width $200 \mathrm{km\,s^{-1}}$ 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 red-

shift SDSS composite spectra. The equivalent width of the best-fitting model is recorded for each of the 100 realisations of the spectra. The error in the equivalent width is defined as the root-mean-square of these values.

In Fig. 1.7 we show the [O III]5008 EW as a function of the quasar bolometric luminosity. Bolometric luminosity is estimated from the monochromatic continuum luminosity at 5100Åusing the correction factor given by Richards et al., (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, 2005).

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

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

See extra text from Brotherton paper. I could be confused here, but I think the Netzer argument goes that the nlr size increase with luminosity because there are more ionising photons. but then you run out of nlr to ionise. the luminosity of the quasar keeps increasing but the luminosity of the nlr flattens out. so the eqw starts to decrease. but we see a huge scatter in eqw at high luminosities. we can relate this to the C IV blueshift, which I don't think Netzer will have been able to.

1.5.1.1 *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 $z = 0.397$.
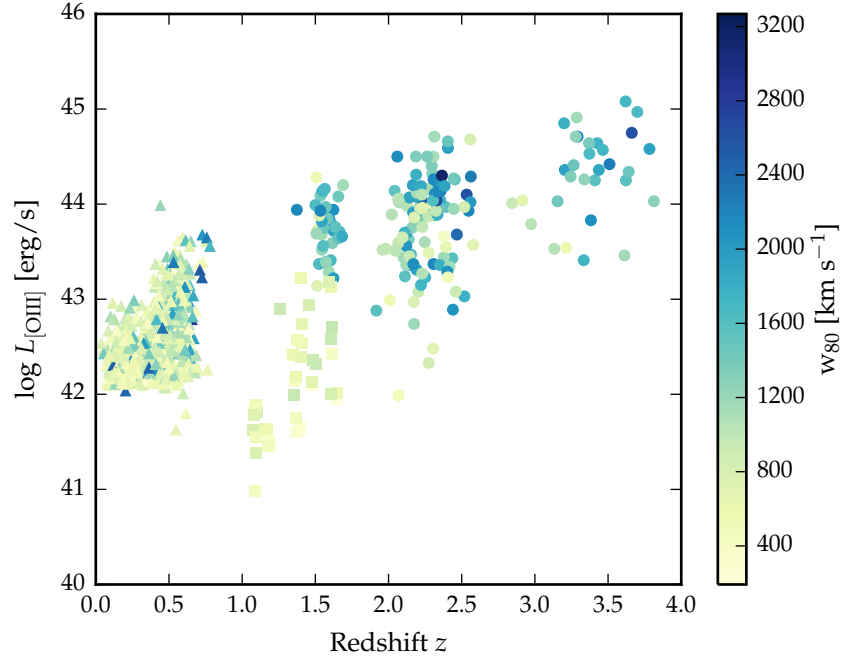
Figure 1.8: The [O III] velocity-width, characterised by w_{80}, as a function the [O III] luminosity and the quasar redshift. The colour of each hexagon denotes the mean w_{80} for the objects in that luminosity-redshift bin. We have supplemented our sample with low-z objects from Zakamska and Greene, (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.

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

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

In Figure 1.8 we show the [O III] velocity width as a function of the [O III] luminosity and the quasar redshift. The [O III] luminosity is calculated by

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 erg/s could be due to the fact that these are all type II quasars, whereas the sample presented in this chapter are all type I. Mullaney et al., (2013) showed that the [O III] lines of type I quasars are typically broader than in type II quasars.

Need some discussion about potential selection biases in these samples e.g. Zakamska & Green objects are obscured. Does this make a difference? Also, many studies only present detections and do not publish non-detections... I thought there were some Type 1 quasars at low-z with [O III] measurements from Shen? but I could be wrong.

1.5.1.2 *Eigenvector 1 correlations*

The FWHM of the broad Hβ emission line and the relative strengths of optical Fe II and Hβ have been identified as the features responsible for the largest variance in the spectra of AGN. These parameters form part of 'Eigenvector 1' (EV1), the first eigenvector in a pndependent component analysis (PCA) which originated from the work of Boroson and Green, (1992). The underlying driver behind EV1 is thought to be the Eddington ratio (e.g. Sulentic et al., 2000; Shen and Ho, 2014).

In Figure 1.9 we show the [O III] EQW as a function of the Hβ FWHM and the optical Fe II strength. The optical Fe II strength is defined as the ratio of the Fe II and Hβ EQW, where the Fe II EQW is measured between 4434 and 4684Å. Measurements of the Hβ line properties are taken from Chapter **??**. In our sample, these parameters follow very similar correlations to what is observed at low-z (see also Sulentic et al., 2004; Shen, 2016). In particular, the anti-correlation between the [O III] and Fe II EQW. However, the Hβ FWHM are displaced to higher values, which is consistent with the high-redshift, high-luminosity sample having larger BH masses.

These emission line trends in the optical (for low-z quasars) can be extended to UV emission lines observed at higher redshifts. The C IV blueshift and EQW is a diagnostic that similarly spans the diversity of
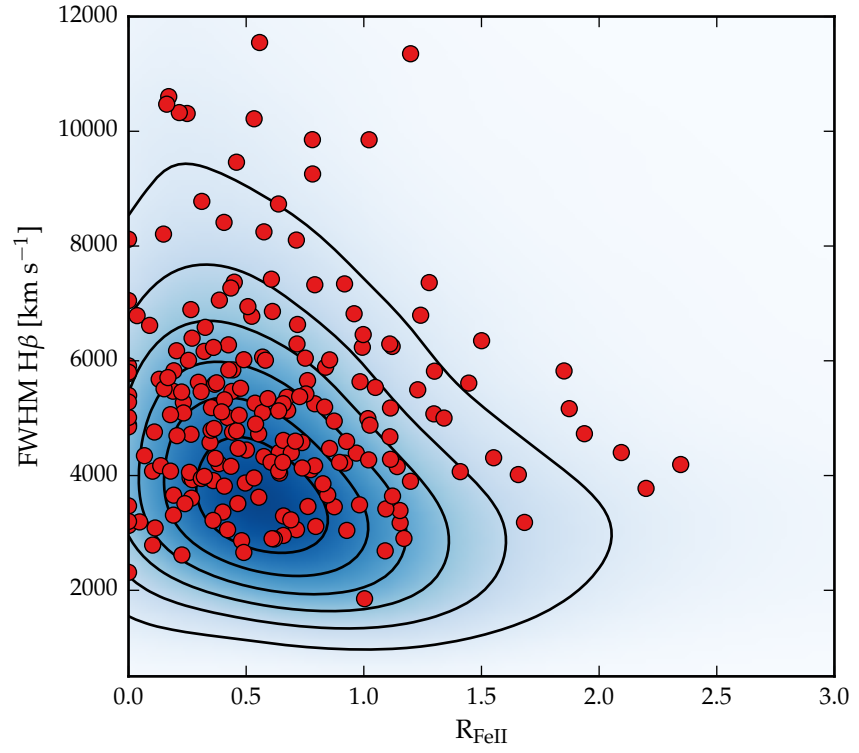
Figure 1.9: EV1 parameter space. The contours and shading show low-redshift, low-luminosity SDSS AGN (with measurements taken from Shen et al., (2011)) and the red circles show the high-redshift, high-luminosity objects presented in this chapter.

broad emission line properties in high redshift quasars (dominated by a virialized component at one extreme and a wind driven component at the other Richards et al., 2011; Sulentic et al., 2007). The similarity of the C IV EQW-blueshift parameter space at high redshift to EV1 parameter space at low redshift suggests that these trends are connected.

Can we calculate a mapping between the two parameter spaces? As a first step we show how the EV1 parameters change as a function of position in the C IV EQW-blueshift parameter space in Figure 1.10. The C IV blueshift is measured relative to the redshift determined from fitting the ICA components. Two hundred and sixty objects are shown in Figure 1.10. Objects flagged as having significant Fe II residual emission have been removed. Objects for which the Hβ or C IV line properties could not be measured reliably (see Section **??**) have also been removed. Finally, we consider only objects for which the C IV EQW exceeds 15Å.

Most of the diversity in C IV properties seems to be driven by the [O III] EQW. On the other hand, the C IV blueshift and EQW cannot be used to predict the Hβ FWHM. This is consistent with what we found in Chapter **??**: objects with large C IV blueshifts have narrow Balmer emission lines, but objects with modest C IV blueshifts have a wide range of Balmer line widths.

1.5.1.3 *Extreme [O III]*

Figure 1.11 shows the spectra of 18 objects which we visually identified as having exceptionally broad [O III] emission profiles. For all of these objects the [O III] emission is heavily blended. One consequence of this is that there is a significant degeneracy when the emission is decomposed in to individual contributions from Fe II, Hβ and [O III]. We note that there is a similarity between these objects and the four quasars presented in Zakamska et al., (2016).

Say more about this and compare to WIISH sample

1.5.1.4 *Comparison*

In Figure 1.5 we compare systemic redshift estimates made using [O III], Hβ and Hα . The scatter is around $300 \mathrm{km\,s^{-1}}$. There is a small offset, such that [O III] is normally slightly bluer than expected.

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

Optical spectra are available for XXX quasars in our catalogue, and cover the broad C IV doublet. As we described in Chapter **??**, C IV is often blueshifted, which almost certainly signal the presence of strong outflows, most likely originating in a disc wind. In Chapter **??** we demonstrated that the quasars in our sample cover the full range of C IV blueshifts seen in the SDSS quasar population, which makes our
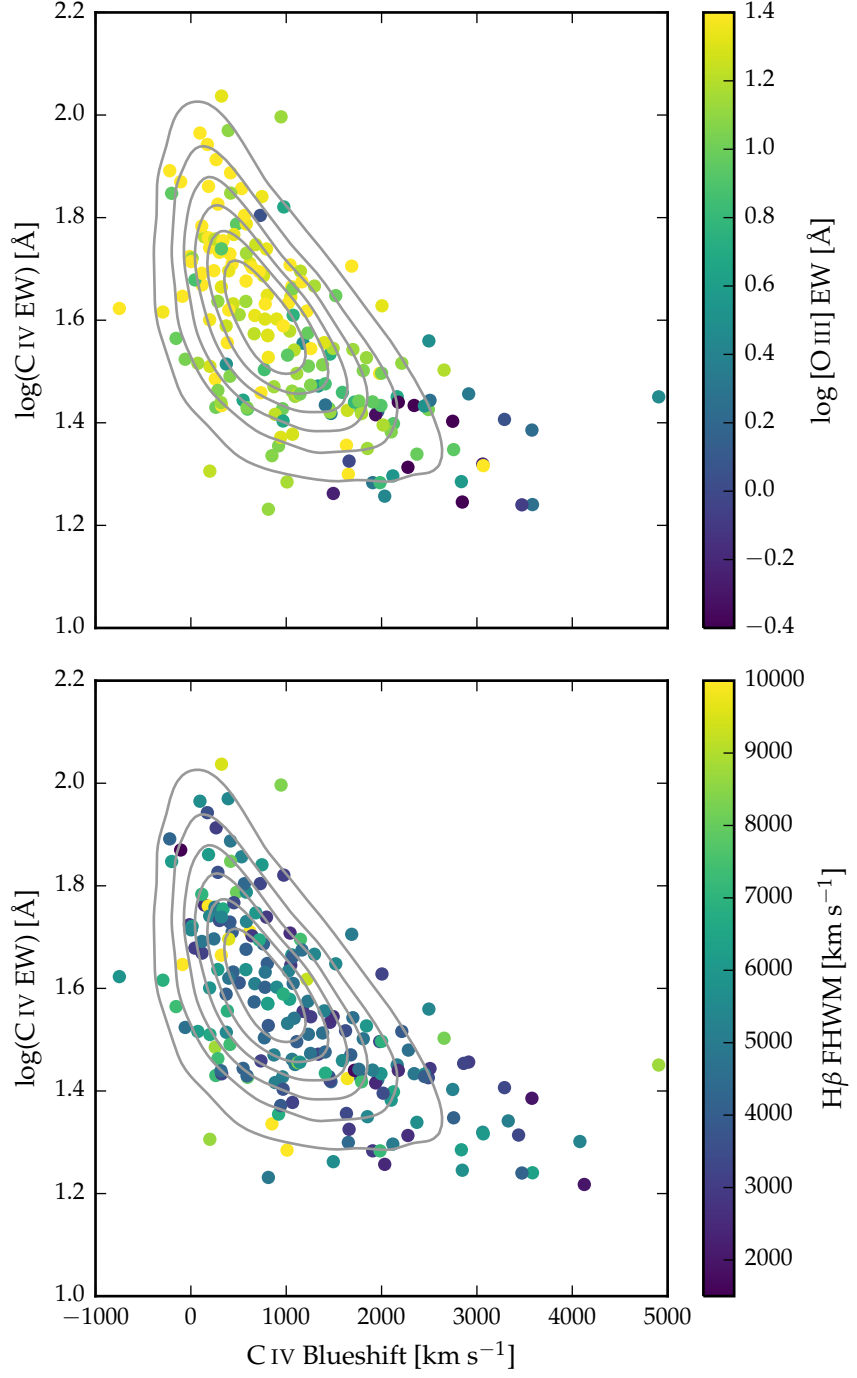
Figure 1.10: The high-redshift EV1 parameter space of C IV blueshift and EQW. Our sample is shown with points, and quasars from the full SDSS catalogue are shown with grey contours. The [O III] EQW varies systematically with position in the C IV blueshift-EQW parameter space (a) but the Hβ FWHM shows significantly less systematic variation (b).
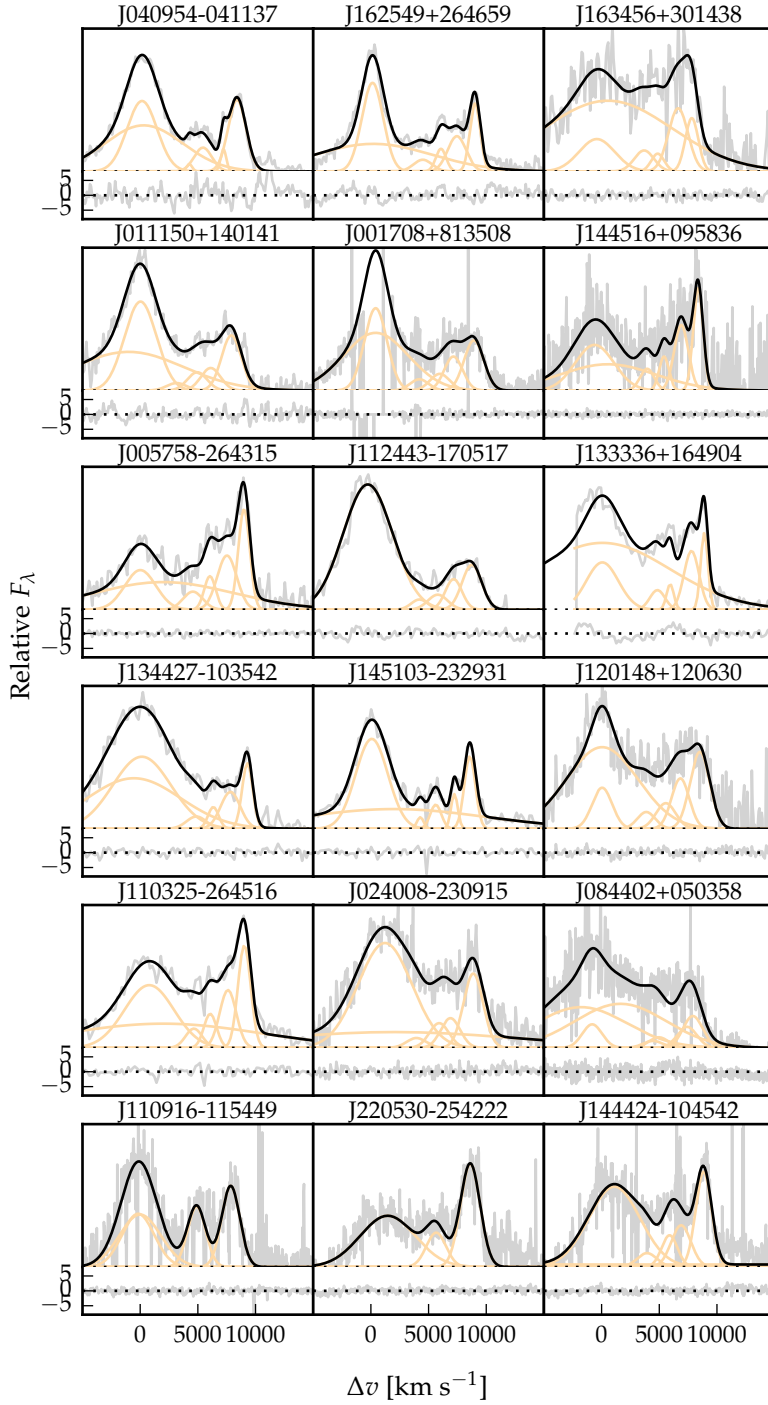
Figure 1.11: Model fits to the continuum- and Fe II-subtracted Hβ/[O III] emission in 18 quasars with extreme [O III] emission profiles. The data is shown in grey, the best-fitting model in black, and the individual model components in orange. The peak of the [O III] emission is used to set the redshift, and Δv is the velocity shift from the rest-frame transition wavelength of Hβ. Below each spectrum we plot the data minus model residuals, scaled by the errors on the fluxes.
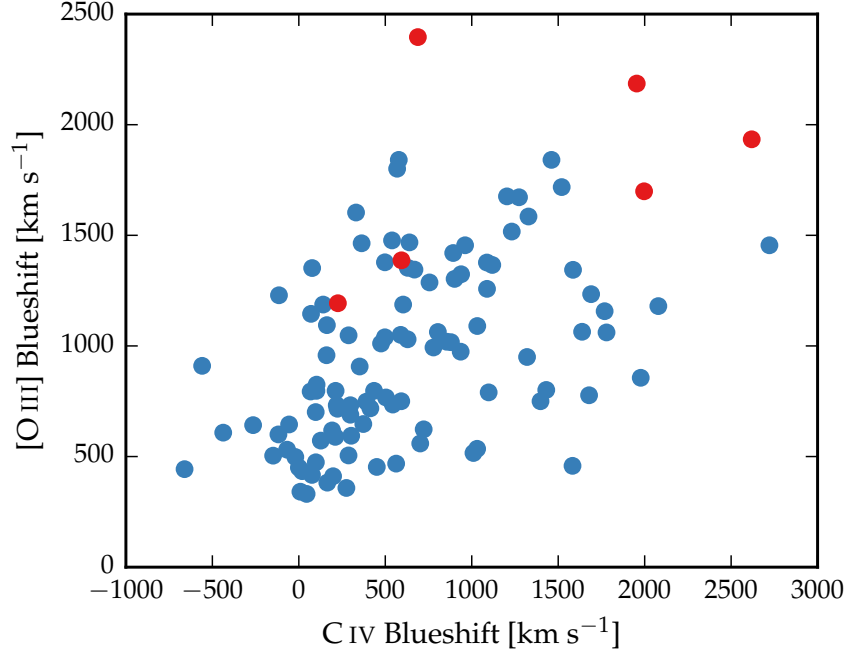
Figure 1.12: The relation between the blueshifts of C IV and [O III].

sample unique in that it allows us to study properties of the quasar across the full parameter range.

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

In Figure 1.12 we show the C IV blueshifts against the [O III] blueshifts. This comparison is done for a sub-sample of 146 objects *What is good?* where we have good measurements of the C IV, [O III] profiles.

There is a clear and strong correlation. Note that our EQW cut removes most of the quasars with large C IV blueshifts, since [O III] is on average very weak in these quasars. Similar correlations have been tentatively found in lower redshift quasars and AGN (Zamanov et al., 2002).

The blueshifting of C IV is known to correlate with luminosity (Richards et al., 2011). In [O III], the blueshifted wing becomes relatively more prominent as the luminosity of the quasar increases (Shen and Ho, 2014). Therefore, it is plausible that the correlation between the C IV and [O III] blueshifts is a secondary effect that is driven by the correlation of each with the luminosity. However, no strong luminosity-dependent trends are apparent in Figure 1.12. 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.

Also, you could put some more text and maybe a figure to explicitly demonstrate that the trend remains even after you have accounted for the trends with luminosity. This is the highlight result for me and personally I think needs a little more fleshing out.

1.6 BROAD ABSORPTION LINE QUASARS

Check all of this

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

1.7 DISCUSSION

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

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

Is the AGN NLR absent in objects where outflows have reached kiloparsec scales, sweeping up the low-density material responsible for the [O III]-emission? If the broad line region (BLR) outflows can escape, they are very fast and wouldn't need long to clear out the NLR gas. .

Might be useful to estimate a time-scale for how long the NLR would take to be cleared given typical size of galaxy and velocity of outflow

1.7.1 *Type II quasars*

Implications of our findings on searches for high-redshift type II 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 BLR 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 static NLR is very weak in luminous quasars.

1.8   ICA

Then having presented the main results, I would go on to discuss the limitations of the Gaussian approach - e.g. FeII can't be properly subtracted in many cases and sensitive to S/N - and use this as an intro to the much more flexible ICA method. You could then have a much briefer description of the ICA reconstructions and present this more as work in progress. You could show that your main results (as above) still hold with the ICA (e.g. Figs 1.15, 1.16, 1.17) and that this allows you to solve the FeII problem and push to lower SNR. Finally, you could discuss some of the potential improvements to the ICA components that would allow the derived line properties from the ICA to become even more robust. ICA works better at low S/N because we are effectively putting priors on the model parameters.

The second model consists of six spectral components derived from an ICA of a large sample of low-redshift AGN with SDSS spectra covering the same spectral region. As we will demonstrate, a linear combination of these spectral components is able to reproduce the spectra around $H\beta$/[O III] to a high degree of precision.

1.8.1   *Model Two: Independent Component Analysis*

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

$$x = Wc \tag{1.1}$$

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

$$c = W^{-1}x. \tag{1.2}$$

The components were solved for using a sample of 2,154 SDSS

quasars at redshifts XX. At these redshifts the SDSS spectrograph cov-

ers the rest-frame region XX-XXÅ where Hβ and [O III] lie. The individual spectra were first adjusted to give the same overall shape as a model quasar template spectrum. Six positive independent components and four additional components that could be negative were found to be sufficient to reconstruct the spectrum, without overfitting. Each quasar spectrum x_j can then be represented as a linear combination of the independent components:

$$x_j = \sum_{i=1}^{10} c_{ij} W_{ij} \tag{1.3}$$

1.8.1.1 *Fitting procedure*

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

1.8.1.2 *Quality of fits*

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

Is there some way to demonstrate/quantify this?

However, it does have its limitations. The components were calculated using a set of lower-redshift, lower-luminosity AGN, and quasar spectra are known to vary systematically as a function of luminosity. For example, the [O III] line is typically broader in more luminous quasars. Because there are so few objects with very broad [O III] in the low-redshift sample, the ICA reconstruction fails to reproduce the broadest [O III] profiles in our sample.
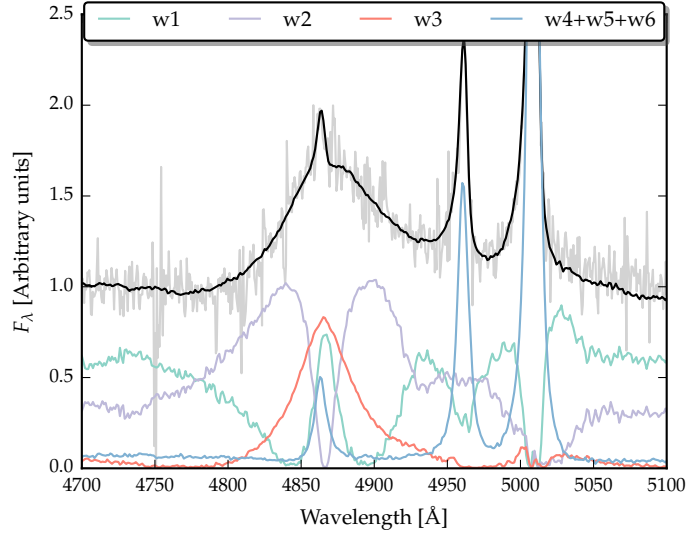
Figure 1.13: Hβ/[O III] emission J002952+020607. The ICA reconstruction is shown in black, and the spectrum in grey. The first three components, and the sum of components four, five and six are shown individually.

Table 1.3: Physical interpretation of the ICA components.

Component	Origin
w_1	Fe II
w_2	Hβ
w_3	Hβ
w_4	[O III] core
w_5	[O III] core
w_6	[O III] wing

1.8.2 *Physical interpretation of ICA components*

Although the ICA is analysis is not based on any physics, there appears to be a direct correspondence between the individual components and the different emission features which contribute to the spectra (Fig. 1.13). This correspondence is summarised in Table 1.3. The component w_1 seems to correspond to Fe II emission, the components w_2 and w_3 to broad Hβ emission, the components w_4 and w_5 to narrow [O III] emission at the systemic redshift, and the component w_6 to broad, blueshifted [O III] emission.

1.8.2.1 *Reconstructing the [O III] profile*

In order to measure non-parametric line parameters, e.g. v_{10}, we must first reconstruct the [O III] emission. It is fortunate that most of the [O III] emission is in just three of the ICA components; the remain-
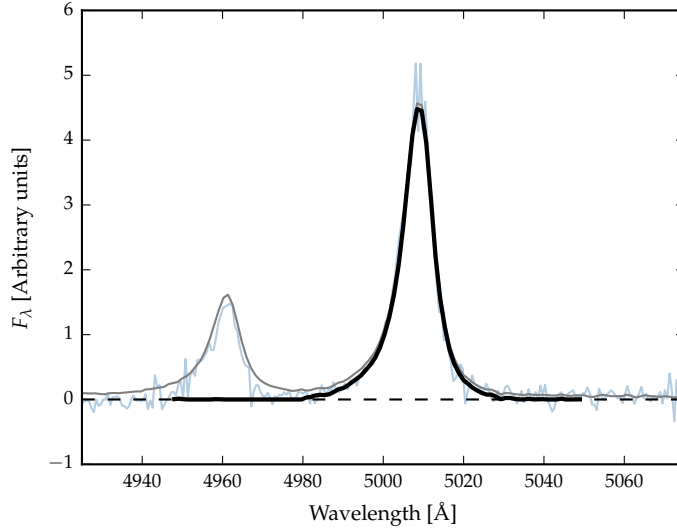
Figure 1.14: [O III] emission in J002952+020607. The data is shown in blue,
and the ICA spectrum in grey. The first three ICA components
have been subtracted from both the ICA composite and the data.
The black curve shows the reconstructed [O III] profile.

ing three contribute very little. Therefore, we can set the first three
weights to zero to leave only the [O III] emission. The four correction
components are also included.

We define the boundaries of [O III]λ5008 as being between 4950
and 5500Å. The blue limit is close to the peak of the [O III]λ4960 line,
and so to recover the intrinsic profile we instead use the blue wing
of [O III]λ4960. We use the emission from 4980-5050Å, and from 4900-
(4980-(5008.2-4960.3)). The blue window is then shifted by (5008.2-
4960.3) to reconstruct the blue wing of the [O III]λ5008 line. We then
subtract a constant, because the flux does not always go to zero (sug-
gests that there is probably flux which is not due to [O III] emission
in components four to six).

An examples of a reconstructed [O III] emission line is shown in
Figure 1.14. At present I am summing the flux all the way from 4950Å.
However, this is quite a lot of flux to sum up, and we can't ascribe
this flux to the wing of the [O III] emission with any certainty. This
is borne out by the fact that there are quite large differences between,
for example, v_{10} measured from the Gaussian fit and v_{10} measured
from the ICA fit.

Unfortunately, there are systematic differences between the line-
width estimates from the Gaussian reconstructions and the ICA re-
constructions, particularly for broad-line objects. The current way of
doing the ICA reconstruction of the [O III] line ignores any cross-talk
between the components and there is potentially flux being ascribed
to the line that could be coming from some other component. We can
solve this by finding some more representative broad [O III] lines in

SDSS from which to derive the components as well as producing a set of components for [O III] only. Therefore we don't use these reconstructions and leave this for future work.

1.8.3   *ICA fits*

In Figure 1.15 we show the relative weights of each of the six positive ICA components. Also shown are the same measurements for a sample of low-redshift, low-luminosity AGN. We want to examine whether or not there are systematic differences between these two samples.

We see that [O III] core emission is weaker in the more luminous sample, but the strength of the wing component is similar. Shen and Ho, (2014) showed that the strength of the core [O III] component decreases with quasar luminosity and optical Fe II strength faster than the wing component, leading to overall broader and more blueshifted profiles as luminosity and Fe II strength (or C IV blueshift) increases. Shen and Ho, (2014) suggested that a stable NLR is being removed by the outflowing material. Similarly, Zhang et al., (2011) found that the more the peak of the [O III] line is blueshifted, the more the core component decreases dramatically, while the blue wing changes much less. Therefore, there is an anti-correlation between the strength of the core component and the relative strength of the wing component (Figure 1.17).

To show this phenomenon more clearly, we plot the relative [O III] strength and the [O III] wing/core ratio in the high/low luminosity samples (Figure 1.17). We see that [O III] is weaker in the high luminosity sample, but that the wing component is much stronger relative to the core component. .

Similar to behaviour of C IV? Would suggests that the mechanism producing the two correlations is the same

1.8.3.1   *EV1 correlations*

In Figure 1.18 we show how the [O III] strength varies as a function of the C IV blueshift. There is a very well defined relation: when C IV is strongly blueshifted [O III] is very weak. This is very similar to what we found when we used Gaussian functions to model the emission. The correlation between C IV blueshift and [O III] EQW is shown in a different way in Figure 1.19. Here we divide our sample in to four bins according to the C IV blueshift. From the quasars in each C IV blueshift bin we then find then generate an ICA spectrum using the median weights from each quasar. The differences in the spectra as a function of the C IV blueshift are dramatic. [O III] becomes progressively weaker and more blueshifted. The anti-correlation with Fe III and the blue-ward Fe II also clear, but there is no change in the redward Fe II.
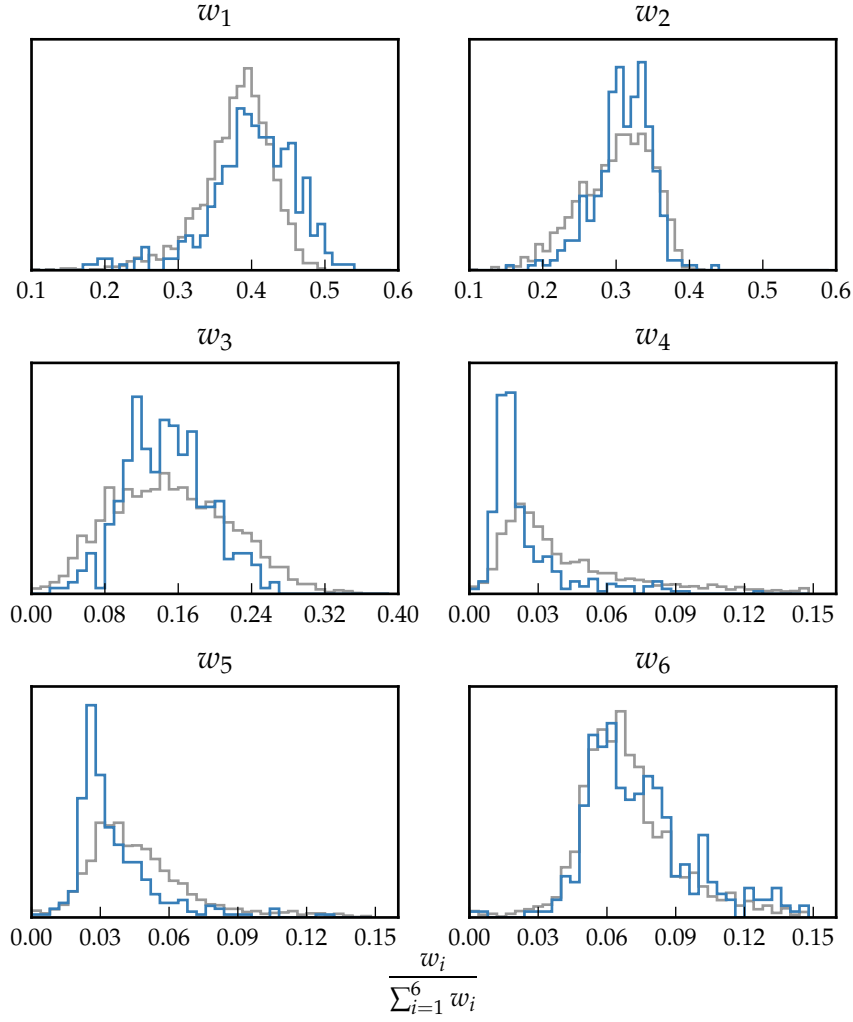
Figure 1.15: The relative weight in each of the six positive ICA components for the high-luminosity (blue) and low luminosity samples (grey). In the high-luminosity sample Fe II emission is stronger (component w_1). The core [O III] emission (components w_4, w_5) is weaker but the strength of the blueshifted wing (w_6) is the same.
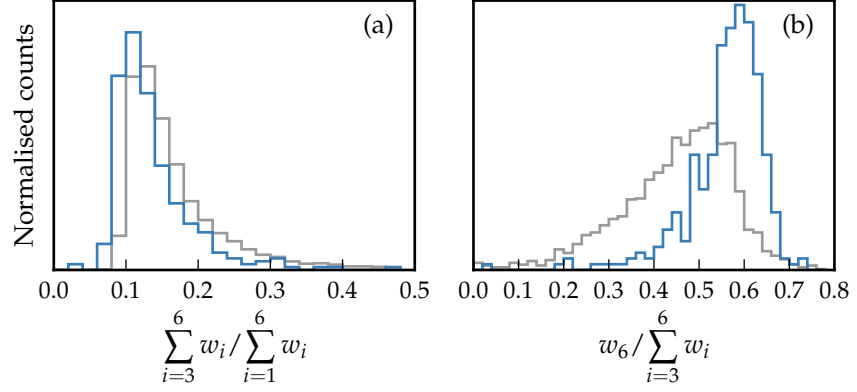
Figure 1.16: The relative weight in the three ICA components corresponding to [O III] emission (*left*) and the relative weight of the component most closely related to blueshifted [O III] emission relative to all three [O III] components (*right*). [O III] emission is weaker in the high-luminosity sample, but the relative contribution from the blueshifted component to the total [O III] emission is higher.
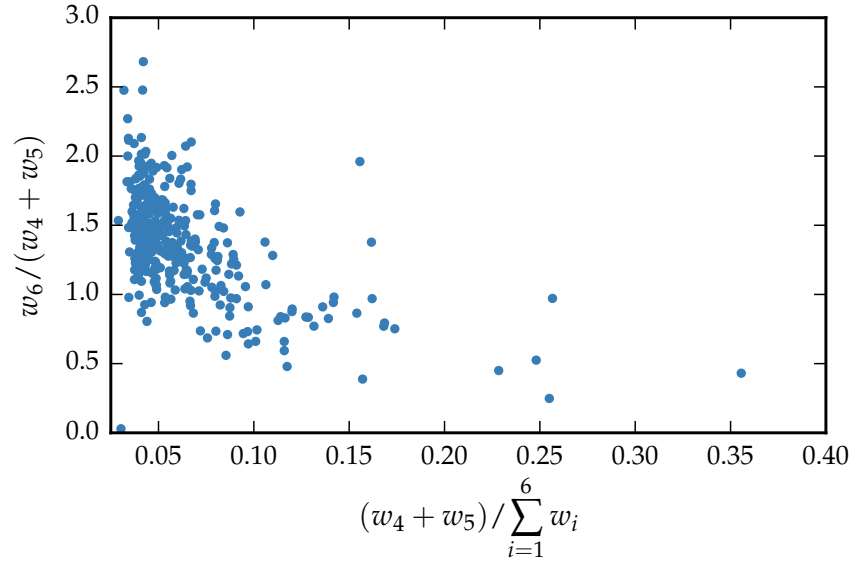

Figure 1.17: Weight in the [O III] wing relative to the weight in the [O III] core emission versus the strength of the core [O III] emission. The blue-asymmetry of the [O III] emission increases as the strength of the core component decreases.
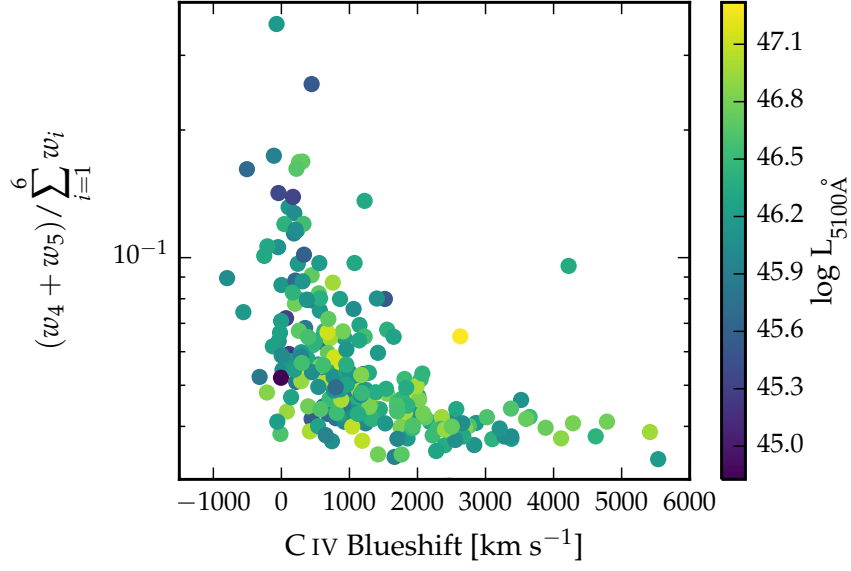
Figure 1.18: The ICA component weight w_4, which is a proxy for the strength of core [O III], as a function of the C IV blueshift. The C IV blueshift is measured relative to the NIR ICA redshift.
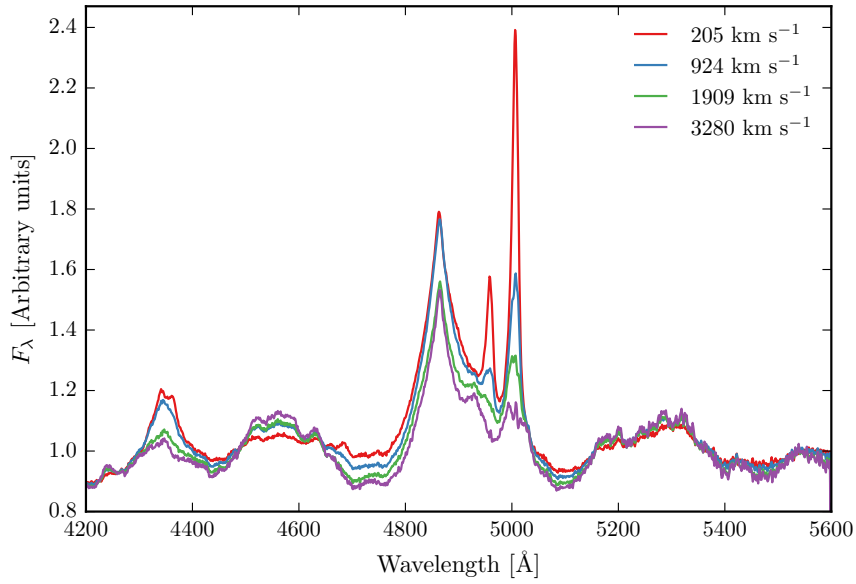

Figure 1.19: Median ICA-reconstructed spectra as a function of the C IV blueshift.

1.8.3.2   *Updating EV1*

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

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

1.8.3.2   *Updating EV1*