UNIVERSITÀ
DEGLI STUDI
FIRENZE

DOTTORATO DI RICERCA IN FISICA E ASTRONOMIA

CICLO XXIX

COORDINATORE Prof. Massimo Gurioli

The inner structure of Active Galactic Nuclei from spectroscopy of large samples of quasars

Settore Scientifico Disciplinare: FIS/05 Astronomia e Astrofisica

Dottoranda
Susanna Bisogni

Tutor
Prof. Guido Risaliti

Coordinatore
Prof. Massimo Gurioli

Anni 2013/2016

“There’s always a siren
Singing you to shipwreck
Steer away from these rocks
We’d be a walking disaster
Just ‘cause you feel it does’t mean it’s there”

— Radiohead, ***‘There There’***

Contents

Outline

Active Galaxies are peculiar galaxies, characterized by non-stellar emission from the central regions of their nuclei (*Active Galactic Nuclei*, AGN), with luminosities from 10^{40} to 10^{48} erg s^{-1}. The energy source for this radiation is the accretion onto a *SuperMassive Black Hole* (10^6-10^{10} $M_\odot$).

The discovery that Supermassive Black Holes are present in quiescent galaxies as well as in active ones, along with the finding that AGN activity peaks at z~2-3, suggests that this phase is common to all galaxies and likely influences their evolution. The study of AGN is therefore crucial to understand how galaxies formed and evolved in the Universe.

The small size and the distance of these sources imply that the emitting regions are spatially unresolved, making the spectroscopic analysis the only tool available for their study. AGN are characterised by strong emission features, coming from specific regions: the *Broad Line Region* and the *Narrow Line Region*. These lines show profiles far broader than the normal galactic emissions and are hence necessarily emitted by gas under the influence of a very strong gravitational field. Their presence in AGN spectra is therefore a confirmation that the central engine of AGN consists of a SMBH; furthermore, it enables us to give an estimation of the mass of the central body and, through the analysis of their shape, makes us able to study the geometry and the kinematics of the emitting regions.

This thesis fits in this context. We are interested in the study of emissions coming from the structures depicted in the AGN Unified Model and, particularly, in the emission lines coming from the BLR, the closest to the SMBH and therefore the one providing information on its mass.

The thesis is organised as follows: in Chapter 1 we present an introduction on AGN, describing the Unified Model and the relations with host galaxies. We then focus on the aspects we are more interested in, i.e. the measurement of the SMBH masses and the analysis of the line profile shape.

In Chapter 2 we make use of an orientation indicator, identified in the equivalent width of [O III] (due to the isotropic emission of the NLR and the disk-like shape of the continuum source) on both the broad and the narrow lines in SDSS quasar spectra. We also present the implications of the existence of such an indicator and the future perspectives for its use.

In Chapter 3 a sample of six quasars at z~2.2 with a simultaneous detection in both the UV and optical lines is examined. The goal is to address existing issues related to the use of C IV as a virial estimator, as opposed to the well tested Hβ line, avoiding the problem related

to quasar variability, thanks to observations performed with the X-shooter spectrograph at the Very Large Telescope, which provides simultaneous spectra in the (observed) wavelength range 0.3-2.5 μm.

In Chapter 4 we investigate the BLR-NLR kinematic connection in a sample of ~170 quasars in order to correct the profile of the C IV λ1549Å line for the contamination by non-virial components (such as outflows) that cause the unreliability of this line as a virial estimator in single epoch observations. In this part of the thesis we present the statistical analysis we set up to examine the connection between C IV and [O III] line profiles and to perform the correction on FWHM(C IV) in order to make it a reliable estimator. The analysis on this sample is still a work in progress, since we only attempted a first correction for a subsample with specific characteristics. Nontheless, this set-up is ready to be used not only on our complete sample but also for any other sample for which both C IV and [O III] observations are available and the first results seem to be very promising.

Finally, we summarise our findings and we present the work ahead.

1 Active Galaxies and SuperMassive Black Holes

Active Galaxies have been known since the first decades of the last century. This term refers generically to every object in which we can recognise a very energetic, non-stellar activity, located in the nuclear regions of the galaxy (*Active Galactic Nucleus* (AGN)).

It is not until the '40s, however, that AGN were considered for the first time as a class by the astronomer Carl Seyfert (Seyfert, 1943).

Seyfert proposed a general distinction between Seyfert 1 and Seyfert 2 galaxies, differing for the presence or the absence in the spectrum of "broad" emission lines ($v > 1000\,\mathrm{km\ s^{-1}}$), along with "narrow" ones (velocities between ~300 and ~800 $\mathrm{km\ s^{-1}}$, i.e. narrower than the "broad lines" but still broader than standard galactic emission and absorption lines). Starting from this first classification, in the subsequent decades several new classes of AGN were proposed, most of them based on one specific observational band (e.g. radio loud / radio quiet) or one specific property (Broad Absorption Line quasars, Narrow Line Seyfert 1s, etc.). Today we know that all these objects have a common nuclear structure and that the differences among them are mostly due to the orientation with respect the line of sight of the observer and to the rate at which the central SMBH is accreting.

In this chapter we will examine the *Unified Model* for AGN and will present the connection between the central engine and the host galaxy. We will then examine direct and indirect methods for SMBHs mass estimations and discuss the importance of spectroscopy for the study of the geometry and dynamics of the inner nucleus.

1.1 The Unified Model for AGN

The presence of an accreting *Supermassive Black Hole* (SMBH) at the center of active galaxies has been recognized as the basis of all the bizarre fenomena distinguishing these objects from the population of "normal" galaxies. The total AGN luminosity spans from 10^{40} to $10^{48}\,\mathrm{erg\ s^{-1}}$. Low-luminosity AGN ($L < 10^{44}\,\mathrm{erg\ s^{-1}}$) are the ones originally studied by Seyfert, and are usually referred to as "Seyfert Galaxies". Higher luminosity AGN ($L > 10^{45}\,\mathrm{erg\ s^{-1}}$) completely dominate the emission of the host galaxy, and are referred to as "quasars". This luminosity-based classification is still used today, even if the term "quasar"

is often used to refer to AGN in general.

The intensity, the compactness and the very broad spectral range (from γ-rays to radio wavelengths) characterising the emission of AGN can not be explained with any process related to stellar activity. In spite of the absence of a direct proof for the presence of a SMBH at the center of active galaxies, plenty of observational pieces of evidence point in this direction (Rees, 1984), while all the alternative scenarios fail at accomplishing an alternative explanation. Furthermore the scheme considering a SMBH as the central engine for AGN allows to gather together all the "zoology" of active galaxies that until the early '90s were considered as various exotic objects (Blazars, BL Lac Objects, Seyferts I and II, LINERs, Quasars etc.).

The paradigm describing all AGN as powered by the same accretion process onto SMBHs is known as *Unified Model* (Antonucci, 1993; Urry & Padovani, 1995) (see Fig. 1.1), and is by now accepted by most of the scientific community. In the Unified Model the SMBHs accrete gas from the central regions of the host galaxies through an accretion disc. This disc is geometrically thin and optically thick (Shakura & Sunyaev, 1973) and its radiation in the optical/UV band is the primary source irradiating the surrounding regions, causing their emissions. The non-spherical symmetry of the circumnuclear region then implies different observational properties depending on the orientation with respect to the observer's line of sight.

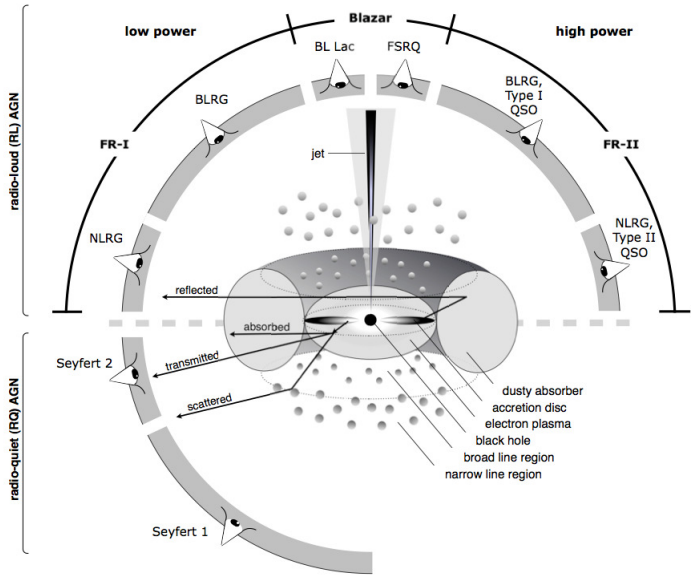


Figure 1.1: Unified Model for AGN. Credit: Beckmann&Shreder 2012

The different regions can be recognised as separate structures and their emissions can be identified in the SED (*Spectral Energy Distribution*) of AGN (Fig. 1.2):

* The BLR (*Broad Line Region*), a region of hot ($T \sim 10^4$ K) high density ($N_e \sim 10^9 - 10^{14}$ cm^{-3}) gas in the shape of clouds orbiting fast ($v > 1000$ km s^{-1}) around the central massive body. This region is responsible for the emission of broad emission lines in the UV/optical/near IR bands of AGN spectra.

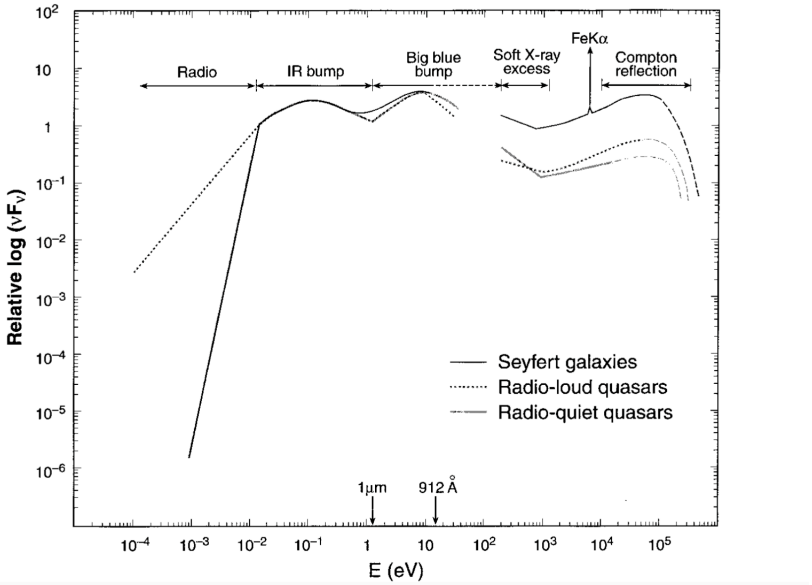


Figure 1.2: Spectral Energy Distribution for Seyfert galaxies, radio loud quasars and radio quiet quasars (Koratkar & Blaes, 1999). The SED of an object is essentially a very low resolution spectrum and is obtained with the use of photometric information in several wavelength bands. When expressed as $\log(\nu F_\nu)$ vs $\log(\nu)$, the area bounded by the SED represents the bolometric luminosity of the object, i.e. the total energy emitted by the source in all the bands. In an AGN SED we can recognise several components: the Big Blue Bump, which is due to the primary source of radiation, the optically thick and geometrically thin accretion disk, the IR bump, due to the reprocessing of the UV radiation by the dust (mainly in the TOR) and the X-ray region, emitted mainly by Inverse-Compton scattering of the UV photons by the hot corona above the accretion disk. A last component, due to synchrotron radiation of electrons in presence of magnetic fields in the jets, can be recognised in the case of radio-loud AGN.

* The TOR (Toroidal Obscuring Region), that is located beyond the dust sublimation radius, where the dust can not survive to the radiation emitted by the accretion disc, is composed by molecular and dusty gas and is characterised mostly by its thermic emission in the infrared wavelengths.

* The NLR (Narrow Line Region) is composed by less dense ($N_e \sim 10^2 - 10^4\,\mathrm{cm}^{-3}$) and slower ($300\ \mathrm{km\ s}^{-1} < v < 1000\,\mathrm{km\ s}^{-1}$) clouds and can reach up to ~kpc in size. The lower density with respect to the BLR allows the emission of both permitted and forbidden lines (the latter being instead suppressed by collisions in the BLR) in the UV/optical/near IR bands of AGN spectra.

* The jets, originating from the center of the disc, are highly collimated and emit in the radio band. These structures can reach distances of the order of several Mpc from the nucleus and can end in lobes that are typical of radio galaxies. This is the region responsible for the non-thermal emission in the radio band typical of radio-loud AGN.

Starting from this simple scheme, every AGN can be considered as the same object seen with different orientation with respect to the line of sight of the observer, provided that every source has its own peculiarities, such as the mass of the central object and the rate at which it

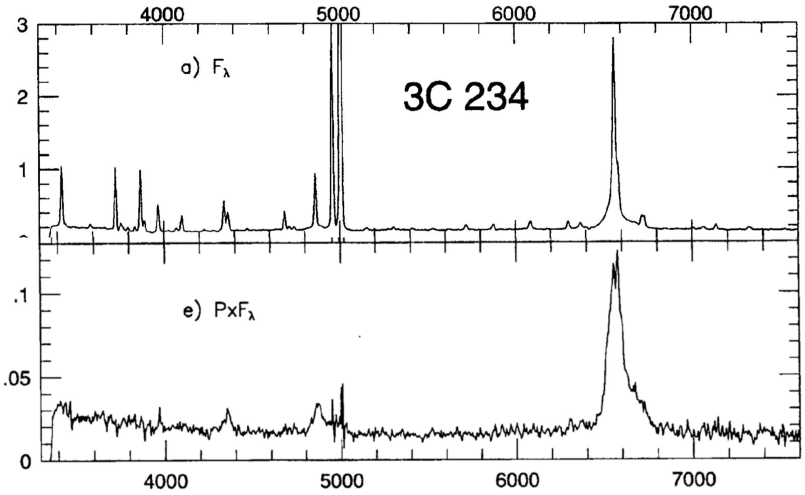


Figure 1.3: Spectra for the total (upper panel) and polarized (lower panel) spectrum of the AGN 3C 234 (Tran et al., 1995). Observations in polarized light allow to evidence the presence of the broad lines emitted by the BLR, hidden by the obscuring region in a type 2 AGN (in this case a Narrow Line Radio Galaxy), because include only the photons scattered by the scattering medium located above the torus aperture and then able to reach the observer. The difference in the total and polarized spectra is essentially the difference between a type 2 and type 1 AGN spectra. Observations confirming that type 2 AGN have a BLR as well were first performed for NGC 1068 by Antonucci & Miller (1985).

is accreting, both influencing its luminosity and, very likely, also some of the characteristics of the inner structures. For example a Blazar is an AGN seen along the radio jet axis, while a LINER is an object observed along a line of sight approximately perpendicular to it.

In particular, within this unification we can distinguish two main phenomenological families: Type 1 and Type 2 AGN. The main difference between these sources is the presence (Type 1) or lack (Type 2) of Broad Emission Lines in the spectrum, that is ascribed to the fact that the line of sight is intercepting the BLR or is being prevented from intercepting the BLR by the obscuring region (TOR) (Fig. 1.3).

1.2 AGN and host galaxies

The development of wide surveys has shown that the AGN density function peaks around $z \sim 2-3$ (Richards et al., 2006), i.e, the AGN activity in galaxies is much more common at those redshifts than in the local universe (Fig. 1.4 left upper panel).

Along with the discovery that quiescent galaxies host SMBHs at their center (Sargent et al., 1978; Kormendy, 1988), this fact indicates that the AGN activity could be a common phase for galaxies and not a peculiar behaviour of ∼10% of the galaxy population. In this scenario SMBHs in quiescent galaxies are relics of the past activity (Marconi et al., 2004) and can give indication on the formation and evolution of structures in the universe.

BH mass measurements in local quiescent galaxies, performed through the study of the resolved kinematics of stars and gas in the BH sphere of influence, have also enlighted the

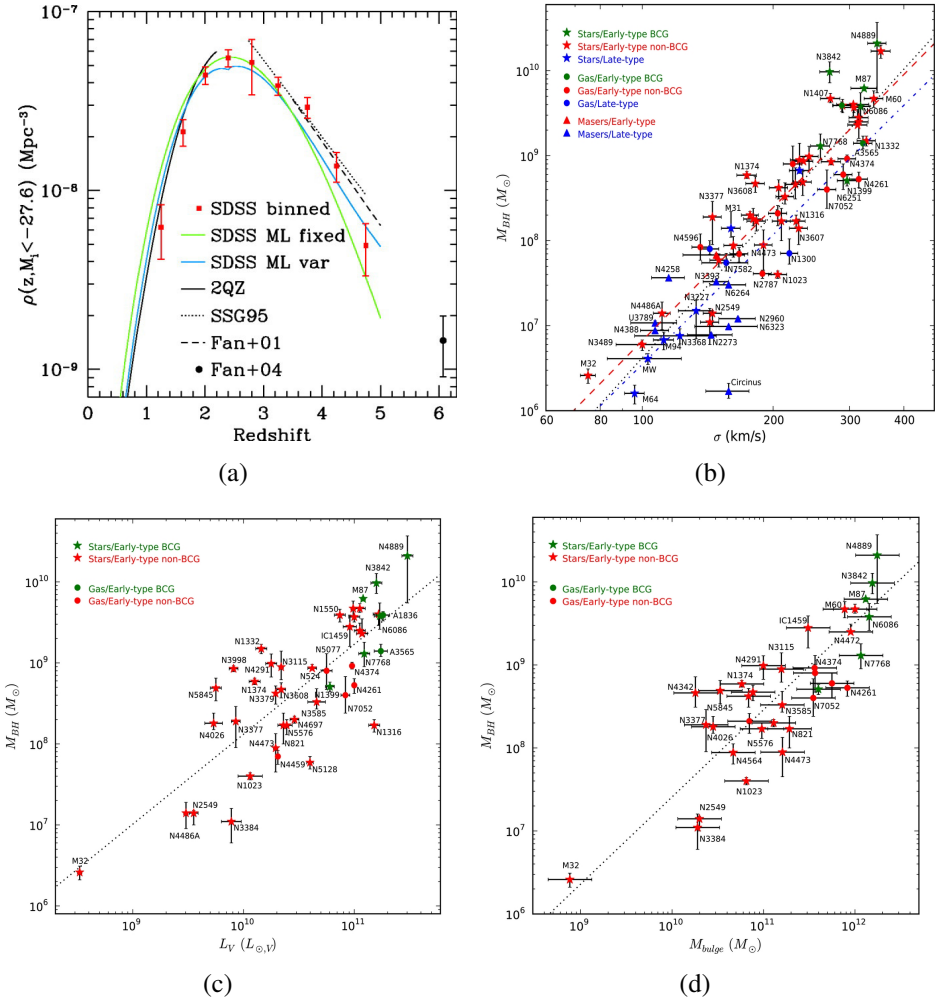


Figure 1.4: (a) Space density of luminous quasars (i.e. integrated i-band luminosity function) from Richards et al. (2006). The peak of the density is around $z \sim 2-3$; luminous quasars are quite rare at much lower and higher redshifts. (b) $M_\bullet - \sigma_*$, (c) $M_\bullet - L_V$ (luminosity in the V-band) and (d) $M_\bullet - M_{S\,ph}$ relationships from McConnell & Ma (2013).

presence of relations between BH and the host galaxy properties (Fig. 1.4 right upper panel and lower panels). SMBH mass relates with the mass of the bulge of the host galaxy (Magorrian et al., 1998), its luminosity (Kormendy & Richstone, 1995; Marconi & Hunt, 2003) and with the velocity dispersion of the stars in its bulge (Ferrarese & Merritt, 2000; Gebhardt et al., 2000). This fact confirms the hypothesis that these structures are former AGN and reveals a connection between the evolutions of BH and host galaxy.

The mechanism responsible for this connection and therefore for the realisation of the M_{BH}-host galaxies properties relations, the so-called *feedback*, is still poorly known. The general principle behind the AGN feedback is that the strength of the rate of accretion by the central SMBH could somehow be able to prevent the star formation within the spheroid of the host galaxy (Fabian, 2012) (in this case the mechanism is referred to as *negative feedback*), thus affecting the M_{sph}, through the ejection of outflowing gas. In some cases instead, the same outflow should be able to trigger or enhance the star formation in the nuclear regions of the spheroids (*positive feedback*) (Silk, 2013). Even if a deeper comprehension of how this coevolution is achieved is necessary, what is clear is that in both cases the AGN activity affects galaxies evolution. AGN then hold an important position also in the study of structures formation and evolution in the universe.

1.3 SMBHs virial mass estimation

1.3.1 Direct Techniques

Several techniques have been developed for the determination of SMBHs mass, called *virial estimations* because the underlying assumption is that the only force acting on the examined systems is the gravitational force exerted by the SMBH. The main distinction among them is due to the distance of the examined object. In local galaxies, for example, it is possible to follow the motion of individual objects in the *sphere of influence*[1] of the SMBH. A special case is represented by our own galaxy, for which the distance from the central mass is sufficiently small to allow precise measurements of the orbits of stars in the central region (Genzel et al., 2010), therefore under the only influence of the gravitational potential of the SMBH. When this level of resolution is not feasible anymore, it is however possible to study the integrated stellar dynamics or the dynamics of the gas under the influence of the gravitatonal field of the SMBH (Ferrarese & Ford (2005) and reference therein).

These techniques give a *direct* estimation of the black hole mass, which is not achievable in more distant galaxies. The small size of the central nucleus and the distance of these objects prevent a direct spatial resolution of the inner structures. In this case, therefore, a study from a spectroscopic point of view, for which the distance does not affect the resolution, is needed.

[1]The *sphere of influence* is the region of space in the proximity of the SMBH where the gravitational force exerted from the central body is dominant with respect to the gravitational force exerted from the host galaxy.

1.3.2 Indirect Techniques

Reverberation mapping

In the last decades particular interest has been devoted to the spectroscopic study of the BLR, the region from which broad emission lines are emitted, both for the proximity to the central body and for the wealth of physical information we can derive from their emission.

A fundamental technique has been developed for the study of the BLR, able to give information on the size and, in some cases, on the dynamics of this region as well. The technique, known as *Reverberation Mapping* (RM) (Blandford & McKee, 1982; Peterson, 1993), takes advantage of the fast variability of the continuum emitted by the accretion disk to get information on the BLR. RM is based on the observation of the delay between some characteristic features in the continuum light curve and the same features appearing after some time in the broad lines light curve, as a result of the primary source excitation. This delay is the time needed by the light to cross the distance between disc and BLR and provides an estimate of the BLR size (Fig. 1.5).

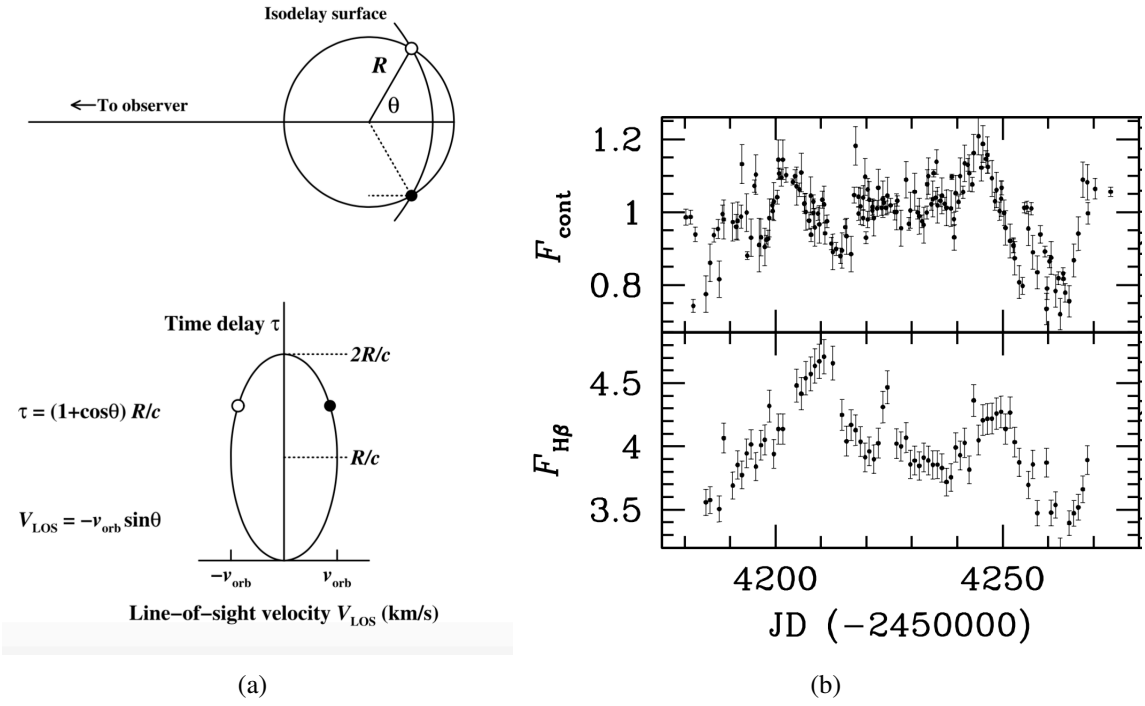
RM was developed primarily with the aim of an estimation of the BLR dimensions and of the mean response times for various emission lines. The BLR is indeed well known to be stratified, with higher ionisation lines emitted closer to the primary source (Gaskell & Sparke, 1986) with respect to lower ionisation lines and having on average a broader profile (Mathews & Wampler, 1985), as expected if the BLR gas clouds are orbiting around the central mass with Keplerian motion.

RM, however, has a further important implication: having knowledge of the size (R_{BLR}) and, from the line width, of the velocity of the gas in the emitting region (we are, again, assuming the only force present, or at least the dominant one, is the gravitational force exerted from the central body), we can obtain the SMBH mass from the equation

$$v^2 = f\frac{GM_\bullet}{R_{BLR}}\,, \tag{1.1}$$

where f is called *virial factor* and accounts for the unknows in BLR geometry. RM is then able to provide a SMBH mass estimation in objects in which we can not resolve the motion of central stars and gas (as long as they are active of course, i.e, there is an active primary source illuminating the BLR), but a "calibration" to infer the virial factor f is needed. This is usually achieved in sources in which a different BH mass measurement technique is available. In this sense RM is considered as an "indirect" SMBHs mass measurement, even if it directly measures the velocity and distance of the orbiting clouds. The spectrum used to determine the linewidth in RM is the rms residuals spectrum, obtained from the multiple monitoring observations,

$$S(\lambda) = \left\{\frac{1}{N-1}\sum_{i=1}^{N}\left[F_i(\lambda) - \overline{F(\lambda)}\right]^2\right\}^{1/2} \tag{1.2}$$



(a) (b)

Figure 1.5: (a) A very schematic explanation of *Reverberation Mapping* (Peterson et al., 2004): in the upper figure, emitting clouds of the BLR are supposed in circular orbit around the central SMBH. If there is an outburst from the central source (an event of intense emission from the accretion disc) then the clouds will respond after the time delay needed to the continuum light to travel from the disc to the clouds. The dotted line represents the path of an ionizing photon going from the central disc to a cloud at coordinates $(R, -\vartheta)$ of the BLR plus that of a BLR emitted photon until it reaches a distance from the observer which is the same as that of the primary source of the continuum. $\tau = (1 + \cos\vartheta)R/c$, where R is the distance between the cloud and the disc, is then the time delay between the cloud emitting and the continuum burst. In the lower figure instead is shown how a circular orbit (for an object in keplerian motion around a SMBH) becomes an ellipse in the velocity-time delay plane. (b) Light curves for continuum and emission line (in this case Hβ) for NGC 3227 (Denney et al., 2010). It is evident how the features in the continuum light curve, corresponding to some event increasing the luminosity of the primary source, are reproduced in the emission line light curve after a time lag. The time lag is a measure of the distance covered by the light to reach the region where, on average, the specific line is emitted.

(Peterson et al., 2004). The use of this spectrum is advisable with respect to the mean spectrum, obtainable with the same observations, because it takes into account only the part of the line which is actually varying under the effect of the continuum illumination, and is therefore more representative of the actually reverberating BLR, throwing away the constant components, emitted by other regions (such as the NLR).

Of course this technique is feasible as long as broad lines emission is detectable. RM can then target only Type 1 AGN. For Type 2 AGN the only way to get a SMBH mass measurement is again the study of the gas proper motion; in this case luminous water megamasers that can be produced by the interaction of AGN activity radiation with gas in the nucleus structures (Neufeld et al., 1994; Tarchi, 2012).

Single Epoch Mass Estimations

Once several tens of sources had been examined through RM, another important relation has emerged: BLR size, as evaluated from RM delays, is related to the luminosity of the nucleus as

$$R_{BLR} \propto L^{\alpha} \tag{1.3}$$

(Kaspi et al., 2000; Bentz et al., 2013). The presence of this relationship is not surprising: the ionisation parameter U, giving an estimation of the amount of gas in the emitting regions ionised by the primary source, can be expressed as $U \propto \frac{L}{r^2 N_e}$, i.e. it is proportional to the primary source luminosity and inversely proportional to the electronic density and to the dimension of the emitting region. If we assume then that U is similar for all AGN, then considering $r = R_{BLR}$ in the previous expression gives exactly $R_{BLR} \propto L^{0.5}$. This relationship has indeed proved to be very closed to the expected one, especially if the luminosity of the sources is corrected for the contamination by the host galaxy (Bentz et al., 2009).

The *Radius-Luminosity Relation* has a huge implication: monitoring with RM the source to have a BLR radius estimation is not needed anymore; in order to get an estimate of the central mass we only need a single luminosity measurement (Fig. 1.6). Surely this estimation is less accurate with respect to the RM measurement, but allows measurements for thousands of quasars very easily. This kind of measurement is called *Simple Epoch* (SE), in the sense that with a single spectrum (observed at a single epoch indeed) we are able to have information on the source without having to use RM, a technique that is very powerful but also extremely time-demanding.

The possibility to deal with thousands of measurements is allowing a deeper, statistical examination of the relationships between the mass of the central body and several of the properties of the host galaxy we mentioned above. The comprehension of the co-evolution mechanism (*feedback*) is one of the many uknowns that will benefit from this increase in the available data.

The mass of the SMBH can, in the case of Single Epoch estimation, be expressed as

$$M_{\bullet} = \hat{f} v^2 L^{\alpha} \ , \tag{1.4}$$

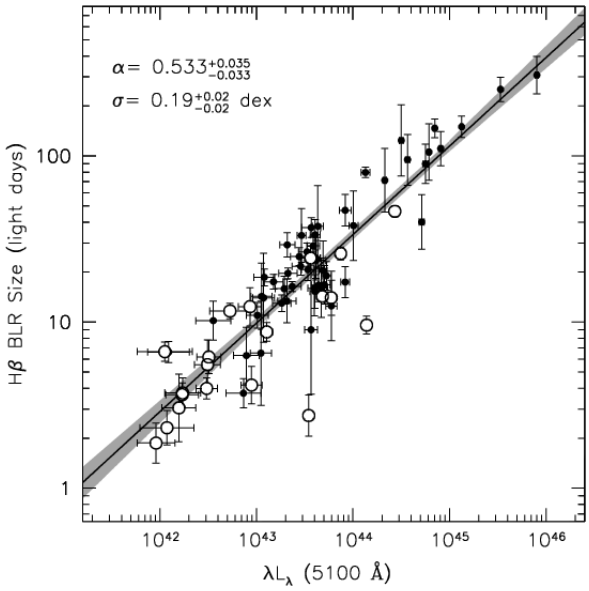


Figure 1.6: *Radius-Luminosity Relation* from (Bentz et al., 2013): the relation between the Hβ emitting BLR radius (as from the delays from RM measurements) and the specific continuum flux at 5100Å.

where, again, $\hat{f}$ accounts for the unknows of the problem and needs to be calibrated through the estimations of the masses with both this method and RM, in those sources in which this is feasable. This is a proper "indirect" method for masses estimations, since in this case we are recovering the BLR radius through the luminosity and not from direct observations of emitting gas orbiting around the central body.

Despite the obvious potential of this method, there are several problems on its usage in determining virial masses. The most serious is that a single observation gives a spectrum which is more similar to the mean one, including the non-reverberating, constant components, than to the rms spectrum. This aspect is one of the subject of this work.

The set of techniques described above forms what is called the *BH Mass Ladder* (Peterson, 2004): as in the *distance ladder* for astronomical objects, where we use primary methods to calibrate the secondary ones, the only ones available for very distant objects, here direct measurements of SMBHs mass are used to calibrate the indirect ones. A summary of the techniques and of their connection is presented in Fig. 1.7.

1.4 Line shape analysis

SMBHs mass estimation is certainly a crucial goal of broad line emissions analysis, but the broad lines shape hides important information on the geometry and the dynamics of the emitting regions as well (the most sophisticated version of Reverberation Mapping, tracing velocity resolved information, for example, has exactly this purpose).

The underlying assumption in line shape analysis is that different parts of the line are emitted by gas in different velocity regimes, i.e. coming from different regions within the

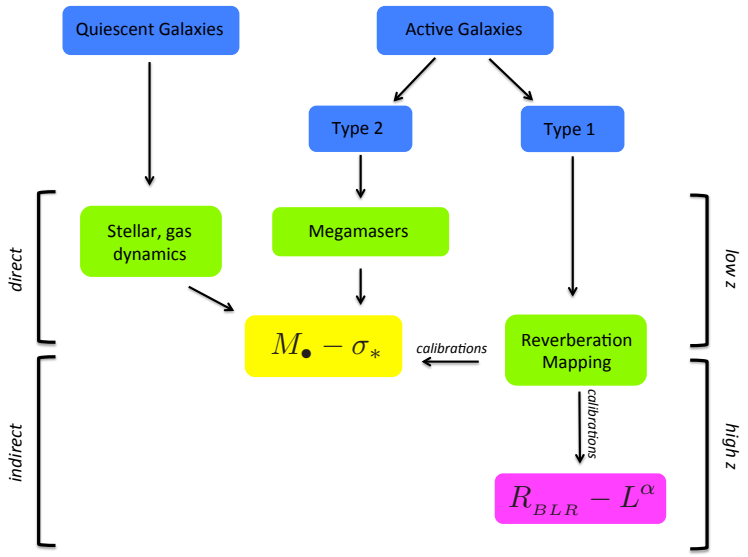


Figure 1.7: BH Mass Ladder adapted from Peterson (2004). SMBHs mass measurement techniques are distinguished as direct (or primary) and indirect (or secondary) methods. Primary methods are the study of spacially resolved kinematics of gas/stars in quiescent galaxies and water megamasers motions in Type 2 AGN. Reverberation mapping in Type 1 AGN could be considered as a direct method, since it directly measures the motion of clouds orbiting around the central BH (BLR), but it is a secondary methods in the sense that it needs calibrations with the primary methods to account for inner regions geometrical unknows. RM calibrations are performed using the $M_\bullet - \sigma_*$ empirical relation in those sources in which both the measurements are feasible. A proper secondary method is the Single Epoch mass measurement, which infers BLR dimension for the $R_{BLR} - L$ relation. This in turn is calibrated with RM measurements in those sources in which both measurements are available.

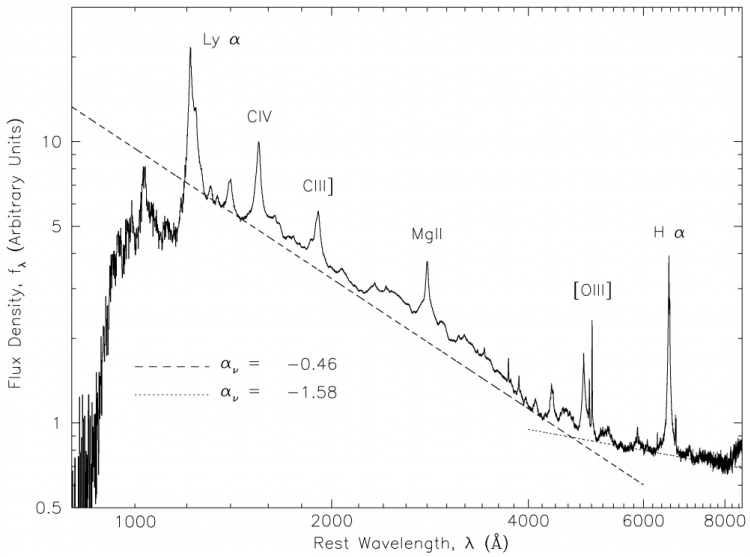


Figure 1.8: Composite quasar spectrum from Vanden Berk et al. (2001). The most prominent lines in an AGN spectrum are, going from the UV to the optical, NIR wavelengths, Lyα, C IV, C III, Mg II, Hβ and Hα, coming from both the BLR and NLR, then having both a broad and a narrow component. The [O III] $\lambda\lambda$4959, 5007Ådoublet is instead the most prominent feature emitted by the NLR.

BLR. We mentioned before that the BLR is stratified, in the sense that higher ionisation lines are emitted, on average, closer to the accretion disk, while lower ionisation lines are emitted further away. In quasars spectra we can recognise a similar behaviour in the line shape of emissions corresponding to the same ionisation potential range.

In Fig. 1.8 we report a composite quasar spectrum (Type 1 AGN) obtained rebinning to the rest-frame, scaling and stacking a large number of spectra with redshift z from 0.26 to 1.92 (Vanden Berk et al., 2001). This stacked spectrum is a representative model of a typical type 1 AGN and summarises the emissions in the wavelength range we are interested in.

The most prominent broad emission lines in quasars spectra are C IV, Mg II, Hβ (close to the [O III] $\lambda\lambda$4959, 5007Ådoublet) and Hα. All these lines have different ionisation potentials (13.6 eV for Balmer lines (Hα and Hβ), 7.6 eV for Mg II and 47.9 eV for C IV) and are therefore emitted, on average, from regions far apart. C IV, the higher ionisation line coming closer to the accretion disk, is, not surprisingly, the one diverting the most from the rest of the lines. Its asymmetry and blueshift are evidence for the presence of winds and outflows in the emitting region. This is not generally the case for the Balmer lines and Mg II. Differences in the line properties can then be used as a discering tool in the comprehension of the dynamics of the involved regions.

Moreover, since these are the lines whose width is used as an estimation of the velocity of the emitting region in virial masses estimation, it is essential to understand which part of the line is emitted from gas in a virialised condition, i.e. under the only influence of the SMBH gravitational force. Any component coming from gas in a different condition should be removed before performing the linewidth evaluation.

These aspects are examined in detail in the following chapters.

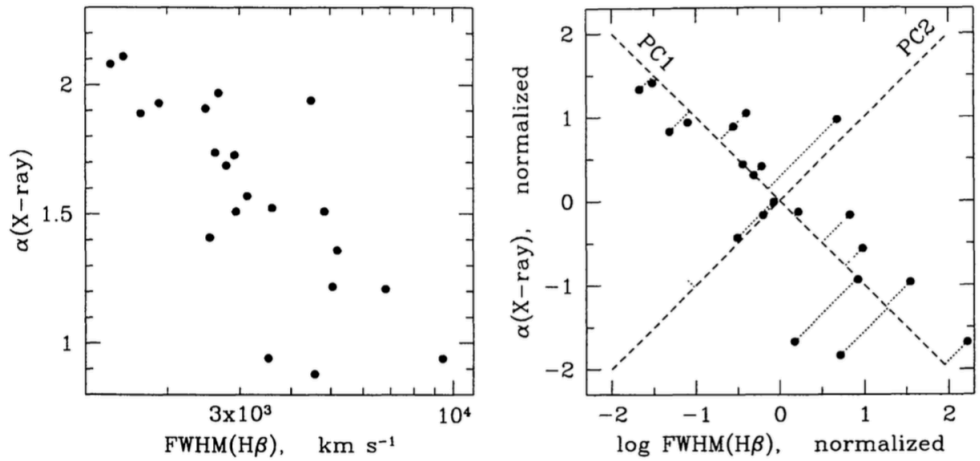


Figure 1.9: A visual representation of PCA (Francis & Wills, 1999). In the two panels the correlation between the X-ray spectral index and the FWHM of Hβ broad component is presented. In the left panel the quantities are expressed in natural units, while in the right panel data are normalised, mean subtracted and divided by the standard deviation. The dashed lines show the direction of the first and second principal components (PC1 and PC2), or Eigenvector 1 and Eigenvector 2, i.e. the first two directions along which is located the variance of the data set.

1.5 The Principal Component Analysis and the *Eigenvector 1*

Quasar spectra show similar characteristics even when redshift, luminosity and other source properties change. This is obvious when one considers the Unified Model and the presence of the same structures within all the AGN. Nonetheless quasar spectra may differ widely in other respects, despite the similarities. Assuming the Unified Model as correct, what causes the differences in quasar emissions? This is a fundamental question in AGN physics and a great deal has been written about that since the seminal work of Boroson & Green (1992). In this work they analysed a sample of 87 objects from the *Bright Quasar Survey* (part of the Palomar-Green sample (Schmidt & Green, 1983)) in order to understand what was the driver of the variance in the sample spectra, with the final aim of determining what physical processes underlie the differences in quasar emissions.

Boroson & Green (1992) analysed several spectral properties, both regarding emission lines and continuum, with the *Principal Component Analysis* (PCA), a technique aimed at dimensionality reduction. When dealing with high-dimensional data sets one can lose sight of meaningful trends, because of the large amount of data and of the presence of multiple variables that are not significant. PCA is the simplest of the dimensionality reduction techniques and is applied when linear correlations are believed to be present among some of the variables in the data set. The aim of PCA is then to identify which variables are needed to give a good representation of the system with a much lower dimensionality or, equivalently, to identify which variables explain the most part of the variance of the system. The working principle of PCA is the following: new coordinate axes, a linear combination of the original

variables, are indentified for the data set, in which the first axis (the *Eigenvector 1* (EV1)) is the one along which the most part of the variance of the system is located; the second axis explains then the most part of the variance left in the data set once EV1 has been removed and so forth (Fig. 1.9). The new coordinate axes are found through the diagonalisation of the correlation matrix of the original variables and for this reason we refer to them with the term Eigenvectors.

This technique has been massively used for several applications in astronomy (Mittaz et al. (1990); Boroson & Green (1992); Yip et al. (2004a); Vanden Berk et al. (2006) among the others) (and, in general, in all the fields in which very large data sets are available) and has been used for part of this thesis work as well (see Chapter 4, where further details on PCA are given).

In the specific case of Boroson & Green (1992), the Eigenvector 1 was identified in the anticorrelation between the strength of Fe II and [O III] emissions, the first coming from the BLR and the second one from the NLR. This finding is a crucial hint for the comprehension of the AGN inner regions and will be investigated further in the following chapter.

2 Orientation effects on spectral emission features of quasars

The Unified Model predicts the presence of the same structures in all the AGN and expects the orientation to be one of the main drivers of the achievement of all the different phenomenologies we see. Due to the morphology of Active Nuclei, the main impediment to a clear sight of the very inner structures, and therefore to the detection of lines emission, is the dusty torus; the angle at which we are observing an AGN determines if we will intercept this structure or not.

This is one of the reasons why an orientation indicator for quasars has been for a long time chased (Decarli et al., 2011; Boroson, 2011; Van Gorkom et al., 2015), even if never univocally identified: having knowledge of the orientation of the object with respect the line of sight of the observer provides a general picture of what emitting regions we are seeing properly and, in case of non isotropic emissions, allows also to study the characteristics of these emission features to get information about the geometry and kinematics of the regions themselves.

In a previous work by our group (Risaliti et al., 2011), an orientation indicator for quasars was indentified in the $EW_{[O\,III]}$. In this chapter we will give a brief introduction to this work and we will describe the resulting follow-up, which is the subject of Bisogni et al. (2017).

2.1 The [O III] equivalent width

As mentioned in the introduction, the optical-UV part of an AGN continuum emission is ascribed to an accretion disk around a supermassive Black Hole. Models developed for such a structure predict that, in order to be radiatively efficient, the disk must be optically thick and geometrically thin (Shakura & Sunyaev, 1973). In this case the geometry of the emitting region imposes a disc continuum intensity that decreases with $\cos\theta$, θ being the angle between the disk axis and the observer line of sight, i.e. the source inclination angle. This fact can be hardly directly proven due to the difficulties in intrinsic continuum measurements; the simplest way to test the behaviour of the continuum as a function of the inclination angle is therefore a comparison between this angle-dependent emission and an inclination independent one.

The [O III] line at 5007Å, emitted by the Narrow Line Region (NLR) at typically hundreds of parsecs from the central black hole, has isotropic characteristics, at least if compared with the emissions coming from accretion disk and BLR, and is considered a good indicator of bolometric luminosity of AGN (Mulchaey et al., 1994; Heckman et al., 2004) (basically all the luminosity of this line is due to the AGN activity and the relation between [O III] and AGN luminosities has a small scatter). Since line emitting regions are optically thin to line radiation, isotropy depends on their dimensions, i.e. they have to be large enough not to be significantly obscured by opaque structures, such as the accretion disk, the dusty torus and possible nuclear dust lanes. Mulchaey et al. (1994) find that the [O III] emission is isotropic in Seyfert galaxies (but see Diamond-Stanic et al. (2009) and di Serego Alighieri et al. (1997) for different results). As a consequence, the observed $EW_{[O\,III]}$, i.e., the ratio between line and local (same wavelength) continuum intensities is expected to be a function of the inclination angle θ.

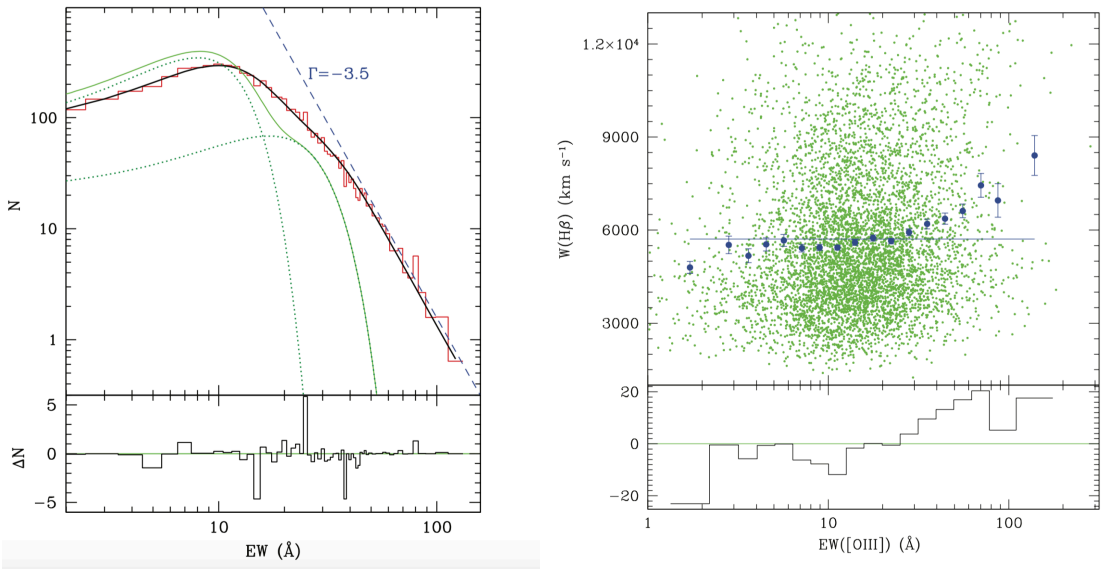
In Risaliti et al. (2011) (hereafter R11) we have studied the distribution of the observed $EW_{[O\,III]}$ for a flux limited sample of ∼7,300 quasars at redshift z< 0.8, obtained from the SDSS DR5 quasar catalog (Schneider et al., 2007; York et al., 2000). The observed $EW_{[O\,III]}$ distribution peaks at ∼10Å and appears to be dominated by the orientation effect at $EW >$ 30Å, while at lower wavelengths it mainly resembles the intrinsic distribution of EW (in general, $EW_{[O\,III]}$ is expected to depend on several geometrical and physical properties of the source, such as the dimension and shape of the NLR and the spectral characteristic of the continuum emission). The orientation signature consists in the presence, at high values, of a power-law tail ($\gamma = -3.5$) that can be reproduced only by inclination effects (Fig. 2.1 (a)).

On the other hand, an examination of the observed distribution of EWs of the main broad lines has revealed that in these cases the inclination effect is much weaker, if not totally absent. This result has been interpreted as a hint at a possible disk-like shape for the BLR: if the BLR and the accretion disk share the same geometrical anisotropy then the EWs of broad lines should not show any angle dependence. Moreover, in such a case, the width of the broad lines should exhibit a trend with $EW_{[O\,III]}$, specifically a broadening towards high $EW_{[O\,III]}$, corresponding to *edge-on* sources; indeed the dependence of broad line linewidths in terms of orientation is well known observationally (Wills & Browne, 1986).

As a preliminary test, in R11 we plotted the line width of Hβ versus $EW_{[O\,III]}$, and found that despite the expected large dispersion, the average widths of Hβ are larger in quasars with higher $EW_{[O\,III]}$, i.e. more edge-on (Fig. 2.1 (b)), a fact which is difficult to explain with other scenarios.

Based on these early results, the aim of this work is to search for $EW_{[O\,III]}$- dependent (i.e. orientation-dependent) effects in quasars spectra.

If detected they would be helpful on the one hand to test the reliability of $EW_{[O\,III]}$ as an orientation indicator and on the other one to investigate broad lines behaviour (thus BLR geometry and kinematics) taking advantage of the now available information on sources inclination with respect to the observer. Such a confirmation would therefore permit the use



(a) (b)

Figure 2.1: (a) Observed [O III] equivalent width distribution for the 2011 sample. The red histogram is the actual data distribution. The green lines represent the intrinsic distribution, the variance in the EWs for the sample due to the natural difference in the characteristics of individual sources. The solid black line is the best fit to the data and presents, in the high values tail, the power law behaviour (dashed blue line) that is possible explain only with orientation effects. (b) Hβ line width as a function of $EW_{[O\,III]}$: the green points are individual sources, while the blue ones represent the average values over logaritmic bins of $EW_{[O\,III]}$. The intrinsic dispersion of this relation is large, but for averaged values over $EW_{[O\,III]}$ bins a clear trend emerges.

of $EW_{[O\,III]}$ as a discerning tool in the study of quasars and, more generally, AGN from a morphological view-point.

For this purpose we have analyzed a large sample of quasars ($\sim$ 12000, approximately twice as many as in R11). We divided the sample in narrow bins of $EW_{[O\,III]}$, and performed a detailed spectral analysis of the staked spectra for each interval.

2.2 Sample selection and global analysis

Our sample of quasars has been selected from the 5^{th} Quasar Catalogue (Schneider et al., 2010) of the SDSS 7^{th} *DR*, making use of the measured quantities, regarding sources properties, line and continuum measurements, listed in Shen et al. (2011). We first required that the redshifts of the sources were in the range $0.001 < z < 0.8$, in order for the [O III] line to be into the instrumentation optimal response window. We further required a luminosity in the range of quasars ($M_i < -22.1$), and a signal-to-noise S/N$\geq$ 5. Finally we selected a range in equivalent width of the [O III] line, between 1 and 300 Å (R11), in order to avoid strong outliers probably due to incorrect measurements. The final sample includes 12,300 sources, and is a significant improvement with respect to the one presented in R11 (consisting of ~7,300

sources). The increase of the source number is due both to the use of the larger SDSS DR7 catalogue (105,783 quasars, to be compared with the ~80,000 of the DR5 catalogue, used in R11), and to the inclusion of all quasars, regardless of the color selection, while in R11 we selected only objects within the "Uniform" subsample (Richards et al., 2006), i.e. the one selected with the standard blue excess criterion. This less stringent selection has the advantage of significantly increasing the size of the sample, allowing a better analysis of subsamples (see next Sections), but requires a check of the global properties of this new sample. We therefore performed a global analysis, as in R11, with some relevant modifications, as summarized below.

1. We analyzed the distribution of $EW_{[O\,III]}$, fitting the histogram in Fig. 2.2 (a) with an integral function:

$$\frac{dN}{d(EW)} = \int_0^{EW} f(EW', L)g(EW')dLd(EW), \tag{2.1}$$

where L and EW' are the "intrinsic" (i.e. face-on) luminosity and EW of the source, $f(EW', L)$ is a kernel which takes into account the orientation effects, and $g(EW')$ is the intrinsic distribution of the equivalent width, which would be observed if all the objects were face-on. The derivation of $f(EW', L)$ is presented in detail in R11. Basically, under the assumptions of a flux-limited sample and assuming a typical luminosity function of optically selected quasars, the integral of $f(EW', L)$ over the luminosity produces a power law term EW'^{Γ}. If the distribution of inclination angles is isotropic, we obtain $\Gamma = 3.5$. We fitted the EW distribution with a free Γ, and assuming different functions for the intrinsic distribution $g(EW')$: a Gaussian curve, two Gaussians (as in R11), and a log-normal distribution. The latter provides the best fit to the data, with the lowest number of free parameters. The best fit values for the curve shown in Fig. 2.2, perfectly reproducing the data, are an average $\langle \log(EW_{[O\,III]}) \rangle$ of 1.04 and a standard deviation of 0.26 for the intrinsic distribution and a slope of the high-EW tail $\Gamma = 3.45 \pm 0.12$. The deviation from the intrinsic shape $g(EW)$ starts to be relevant at EW> 30Å. This implies that most of the observed sources with EW> 30Å are seen nearly edge-on.

2. We repeated the same analysis for the equivalent width of the Hβ line. The results, shown in Fig. 2.2 (b), are quite different from the previous case: the data are reproduced by a Gaussian distribution with average $< EW(H\beta) >$=58 Å, standard deviation $\sigma(H\beta)$=23 Å, convolved with a very steep high-EW tail ($\Gamma = 6.9 \pm 0.3$). A tail with $\Gamma = 3.5$ is strongly ruled out. A pure Gaussian is also statistically disfavoured, even though at a much lower significance (note the logarithmic scale in Fig. 2.2 (b)). Our conclusion is that the distribution of EW(Hβ) suggests an almost disk-like spatial dis-

Table 2.1: $EW_{[O\,III]}$ bins. Each bin is twice the previous one in a linear scale. The widths of the bins at the edges of the distribution are larger in order to have at least 150 objects in each bin. ⋆ The first two bins have been merged.

ΔEW_{oss} [Å]	EW_{oss} [Å]	n^0 objects
1 – 3 ⋆	2.0	163
3 – 6 ⋆	4.5	932
6 – 12	9.0	3443
12 – 25	18.5	4389
25 – 50	38.5	2375
50 – 100	75.0	810
100 – 250	175.0	190

tribution of the gas emitting this line, with a small deviation with respect to the perfectly planar geometry.

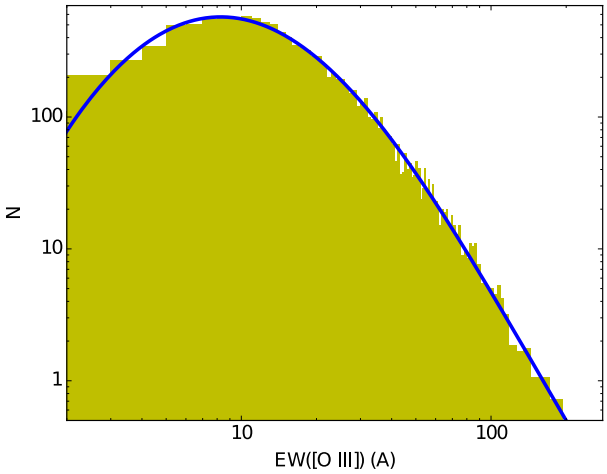
3. As a further check of our interpretation, we analyzed the distribution of the flux ratio R=[O III]/Hβ. In our scheme, since the observed Hβ flux has the same inclination effects as the continuum, the ratio R should show a similar distribution as $EW_{[O\,III]}$. This is indeed the case, as shown in Fig. 2.2 (c), where the best fitting curve is obtained from the convolution of a log-normal distribution and a high-R tail with slope $\Gamma = 3.20 \pm 0.15$. The 2σ deviation from the expected -3.5 value is probably an indication of the deviation of Hβ emission from the pure disk-line distribution of the continuum.

Based on the global analysis described above and in order to investigate the spectroscopic properties of the sample as a function of disk inclination, we divided the sample in subsets of about constant $EW_{[O\,III]}$. We chose to create 7 logarithmically spaced bins of $EW_{[O\,III]}$, with a width of 0.3 dex and starting from $EW_{[O\,III]} = 1$Å (Table 2.1). Since in the first two bins the distribution is dominated by the intrinsic dispersion rather than inclination effect (i.e we can assume that all the objects in the first 2 bins are nearly face-on), we merged them into a single sample. We end up with six subsamples, with enough objects to ensure a proper statistical analysis.

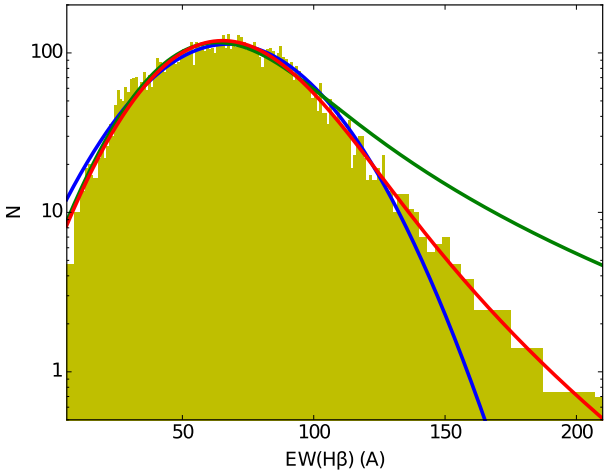
2.3 Data Analysis

2.3.1 Spectral stacking

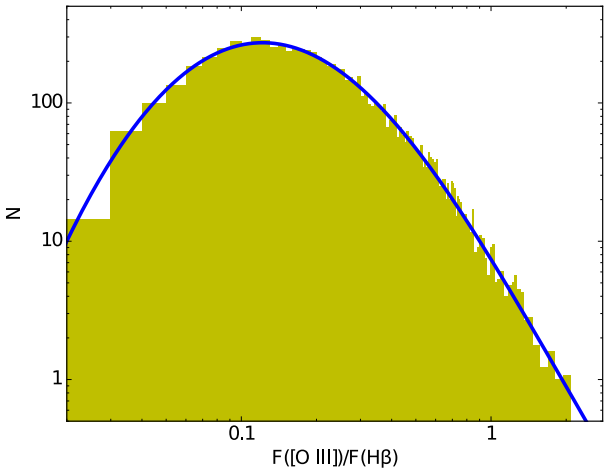
We produced a stacked spectrum for each of the subsamples described above.



(a)



(b)



(c)

Figure 2.2: Observed distribution, and best fit curves, for (a): $EW_{[O\,III]}$; (b): EW(Hβ); (c): F([O III])/F(Hβ). The curves in panels (b) show the best fit with a high-EW tail slope $\Gamma = 3.5$ (green curve), a free slope (red curve; the best fit value is $\Gamma = 6.9 \pm 0.2$), and no tail (blue curve).

The spectra of the individual sources within each $EW_{[O\,III]}$ bin have been normalized by dividing for the flux of the line under examination (e.g. [O III], Hβ, Mg II), de-redshifted according to their redshift as tabulated in Shen et al. (2011) and finally rebinned to a common wavelength grid (λ_{min} = 2000Å and λ_{max} = 7000Å with a step of $\Delta\lambda$ = 0.5Å to preserve the spectral resolution). The adopted rebinning procedure is adapted to spectroscopy from the *Drizzle* algorithm used for photometry of undersampled images (Fruchter & Hook, 2002). The drizzle algorithm is conceived to take into account the weighted flux from each input spectrum.

The code produces the matrix of redshifted and rebinned spectra from which we can obtain the composite spectrum (*stack*).

Our purpose is to obtain a precise fit of both the broad lines (Hβ, Hα, Mg II) and the narrow line [O III] 5007Å. An effect of stacking sources with different fluxes may be an alteration of the profiles of the stacked line. In order to avoid this, for each line we first normalized each spectrum to the flux of that line, and then we produced the stacked spectrum to be used on the analysis.

The average properties of each subsample are better represented by medians rather than by averages (due to the possible presence of strong outliers). However, a single median spectrum would have a too low S/N compared to the average. Based on these considerations, we obtained the stacked spectra by applying the following procedure to each spectral channel: we averaged only the fluxes between the 47^{th} and the 53^{rd} percentiles of the channel flux distribution obtained by considering all available spectra. We then checked that the stacks do not depend on the precise percentile levels, e.g., they are equivalent to those obtained by considering the 25^{th} and the 75^{th} percentiles. The number of spectra used to obtain the flux distribution in each spectral channel is approximately given by the number of objects presented in Tab. 2.1. Deviations from those numbers are due e.g. to flagged spectral channels in some of the spectra. The normalization of the spectra to the fluxes of a given line was adopted to preserve the average profile of that line. An example of the final stacks is shown in Fig.2.3; the stacks are in logarithmic scale to show how the slope is always the same for all the $EW_{[O\,III]}$ bins.

We have finally checked that including the not-uniform objects in the *SDSS* sample (Section 2.2) in our selection does not affect the final results.

2.3.2 Spectral fitting

The IDL fitting procedure has been written making use of *mpfit* (Markwardt, 2009) which fits simultaneously continuum and spectral lines.

The template we used to fit the spectra consists of the following components:

- a power law for the continuum;
- broad emissions of Fe II, obtained by convolving the individual rest frame lines inferred from the template of Véron-Cetty et al. (2004) and from simulated templates

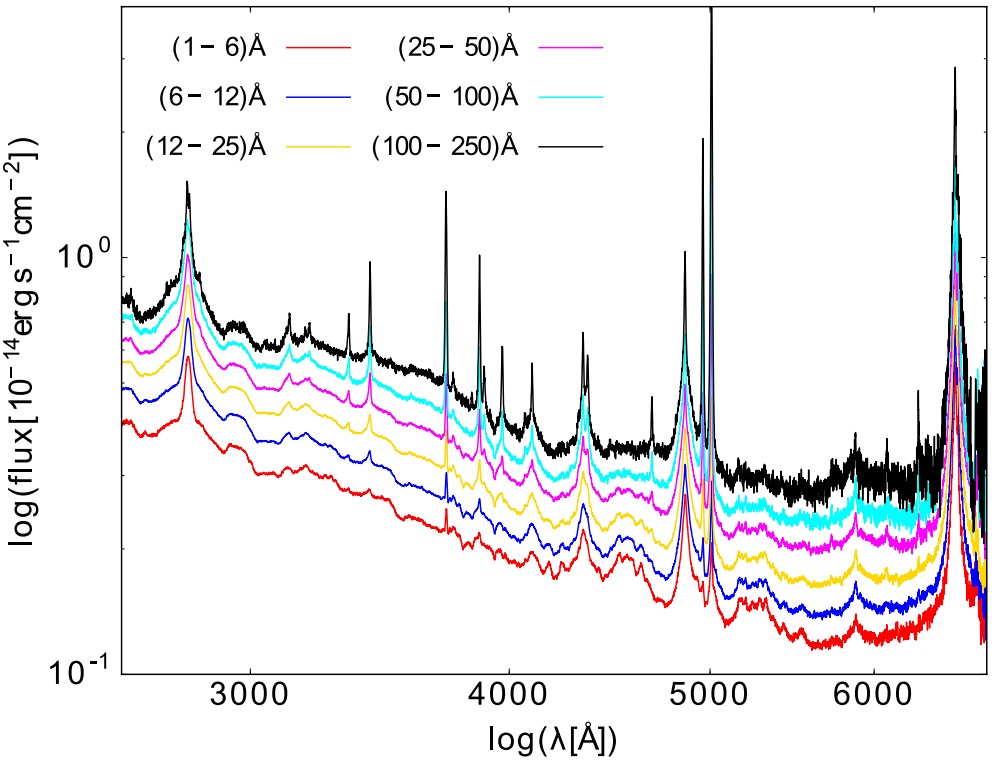


Figure 2.3: Stacks performed with spectra normalized to the 3000Å continuum; stacks are in logarithmic scale. Every color represents a different $EW_{[O\,III]}$ bin: red (1-6)Å, blue (6-12)Å, yellow (12-25)Å, magenta (25-50)Å, cyan (50-100)Å, and black (100-250)Å.

Table 2.2: Spectral ranges and constraints on the free parameters of our template. Fitting functions: BPL is the Broken Power Law and DG indicates the use of two Gaussians to fit the line profile.

line	fitting function	fitting range (Å)	starting λ_0 (Å)	α range	β range	line broadening range (km/s)
Hβ	BPL	(4400 – 5400)	4862.68	10 – 500	10 – 500	250 – 10000
Hα	BPL	(6000 – 7000)	6564.61	10 – 500	10 – 500	100 – 10000
Mg II	BPL	(2500 – 3100)	2798.75	10 – 500	10 – 500	70 – 10000
[O III]	DG	(4400 – 5400)	5008.24			60–
[O II]	DG	(3650 – 3850)	3728.48			

for different physical conditions of the BLR obtained making use of the open source plasma simulation code Cloudy (Ferland et al., 2013), with a Gaussian that accounts for the velocity of the emitting gas.

- Broad Lines, fitted with a broken power law convolved with a Gaussian function. The broken power law has the following expression

$$f(\lambda) \propto \begin{cases} \left(\frac{\lambda}{\lambda_0}\right)^{\beta} & \text{if } \lambda < \lambda_0 \\ \left(\frac{\lambda}{\lambda_0}\right)^{\alpha} & \text{if } \lambda > \lambda_0 \end{cases}, \tag{2.2}$$

where λ_0 is the central wavelength and α and β are the slopes for red and blue tail, respectively. This broken power law is then convolved with a Gaussian so to avoid the presence of a cusp (Nagao et al., 2006).

- Narrow Lines, fitted with a single Gaussian, with the exception of the high ionization lines ([O III] among them), for which an extra Gaussian is present, that takes into account the possible presence of a blue component due to outflowing gas.

The fitting ranges and constraints on the parameters for all the spectral windows of interest are listed in Table 2.2. For the broad components we report the slopes of the two power laws defining the red and blue tail (respectively α and β), besides the line broadening, that is the σ of the Gaussian function used to convolve the two power laws so to avoid the cuspy peaks. In the case of Mg II the reported central wavelength is that of the doublet, but in the fitting procedure the two lines have been actually dealt with separately. For narrow lines instead, fitted with a simple Gaussian function, the line broadening is the usual σ. The central wavelength reported in Tab. 2.2 is the nominal laboratory vacuum wavelength for each line, used as a starting guess. The central wavelength is however left free in the fitting procedure. The line width is expressed in (1) σ, the square root of the second moment of the line profile, (2) FWHM, the full width at half maximum of the profile, and (3) *Inter-Percentile Velocity* (IPV) width, defined as the difference in velocity between two reference wavelengths, λ_1 and λ_2, including a fraction f of the total flux F:

$$\int_{\lambda_1}^{\lambda_2} f_{line}(\lambda) d\lambda = fF \tag{2.3}$$

(Whittle, 1985). The values of λ_1 and λ_2 are chosen in order to have a flux of $\frac{1-f}{2}F$ in each of the excluded tails (i.e. $\lambda < \lambda_1$ and $\lambda > \lambda_2$, respectively). This additional quantity has been evaluated on broad lines best fit profiles to account for broadening and on [O III] best fit profiles with the aim of defining an asymmetry coefficient (see Section 2.4). For the broad lines the width parameters are not a direct output of the fits and, for this reason, the σ, FWHM and IPVs have been computed directly from the output fitted profiles. We evaluated the errors on these quantities through a Monte Carlo simulation. In doing this we have to take into account that the (five) parameters defining a broad component are not independent. In order to produce a reasonable Monte Carlo extraction we have instead to deal with *independent* parameters; so we made a change of parameters, choosing combinations that make the covariance matrix diagonal. The new independent parameters are distributed around their central values, estimated from the old ones making use of the transformation matrix. We performed 1000 extractions for each new independent parameter to collect 1000 parameters sets from which we built a synthetic broad line. We evaluated σ and FWHM for each realization of the line and computed the mean and the standard deviation of the 1000 synthetic profiles, finally getting an estimate of the error on each parameter of the original profile.

2.4 Results

2.4.1 Broad lines

As described in Section 2.2, the absence of the power law tail in the distribution of the EW of broad lines suggests that the BLR clouds share the same anisotropy as the accretion disk. In this case both the flux of the broad lines and that of the disk have the same dependence on the inclination angle, making the ratio between lines and continuum fluxes, i.e. the EW, no longer a function of inclination. The BLR clouds should then orbit on a disk around the SMBH and they must be optically thick to the broad emission lines. If this were the case the velocity inferred by the observer from the line widths should be only a fraction of the intrinsic one, depending on the inclination of the source with respect to the line of sight (specifically $v_{obs} = v_{int} \sin\theta$).

The subsequent step of our analysis has been the study of profiles as a function of $EW_{[O\,III]}$ for Hβ, Hα and Mg II. Fig. 2.4 shows the broad profiles normalized to their peak values for each subsample with a constant $EW_{[O\,III]}$.

The dependence of line width with $EW_{[O\,III]}$ can be interpreted, within our scenario, as due to the different source inclinations with respect to the observer line of sight in each subsample.

In Fig. 2.5, 2.6 and 2.7 the broadening can be observed more quantitatively; these figures show the increase in FWHM, σ and the Inter-Percentile Velocity (IPV) width (in this case we have evaluated in particular the difference between velocities corresponding to 5% and

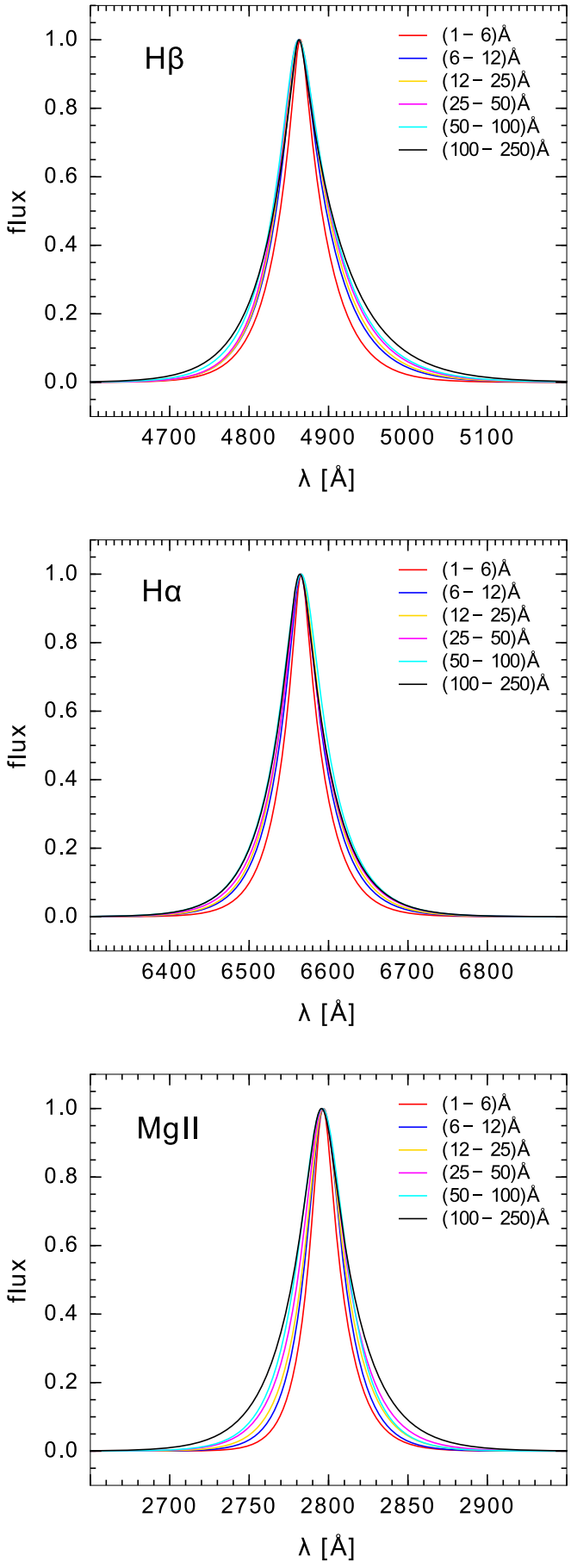


Figure 2.4: Broad components profiles; from top to bottom Hβ, Hα and Mg II. Profiles are normalized to their peak values.

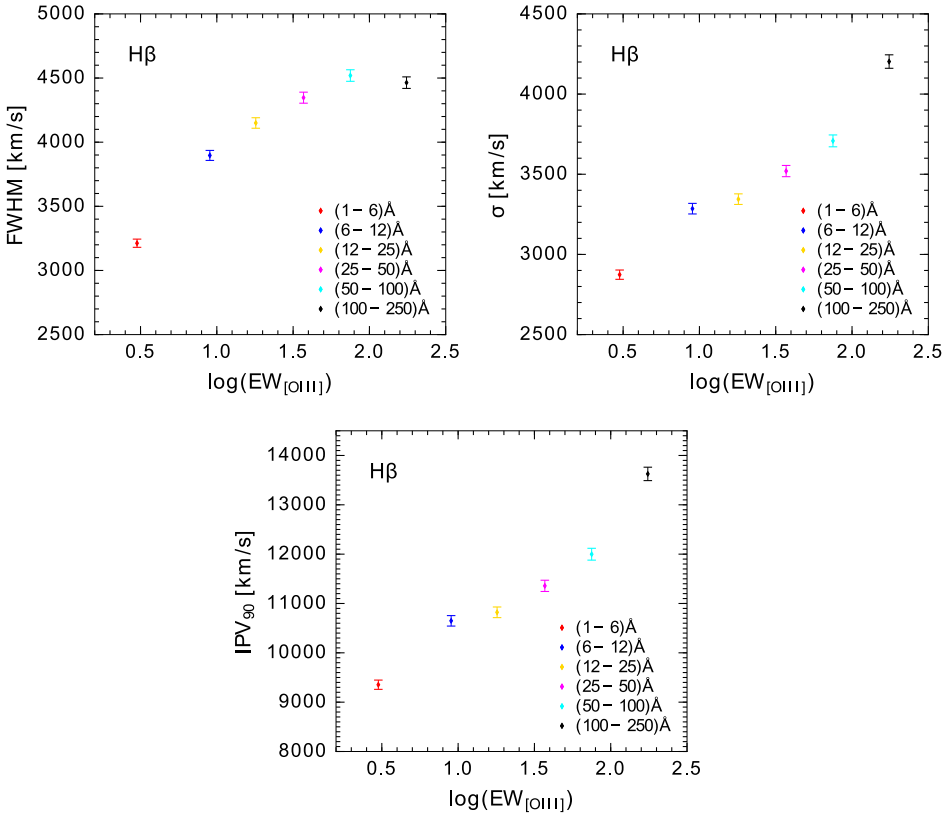


Figure 2.5: FWHM, σ and IPV (90%) of the Hβ broad component as a function of $EW_{[O\,III]}$.

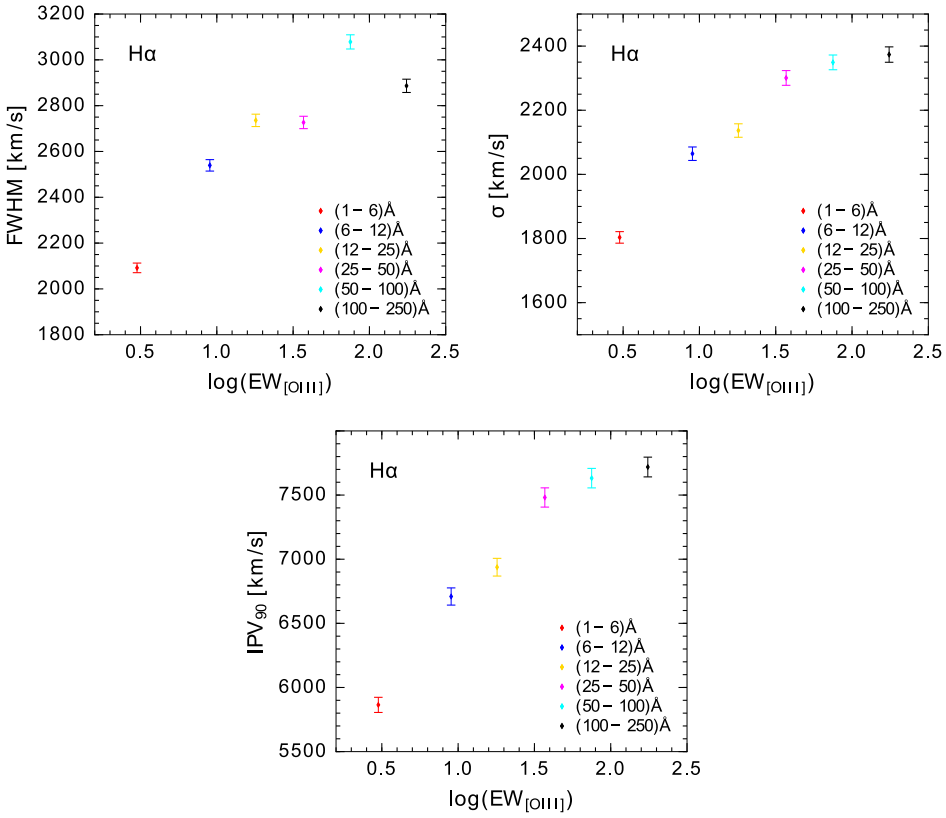


Figure 2.6: FWHM, σ and IPV (90%) of the Hα broad component depending on $EW_{[O\,III]}$.

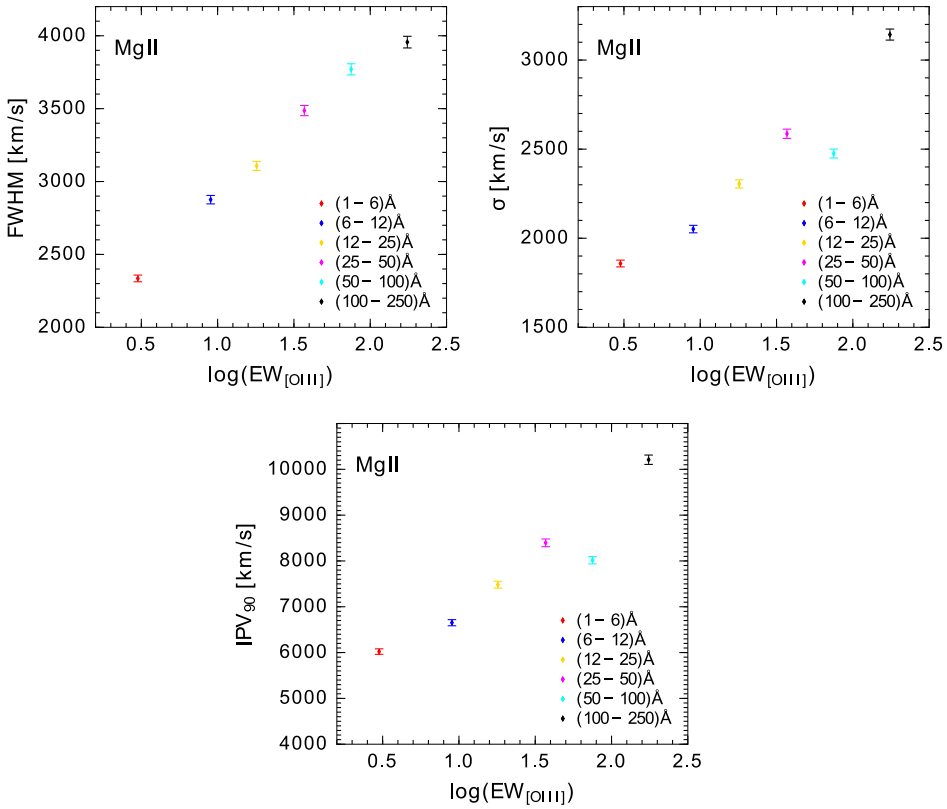


Figure 2.7: FWHM, σ and IPV (90%) of the Mg II broad component depending on $EW_{[O\,III]}$.

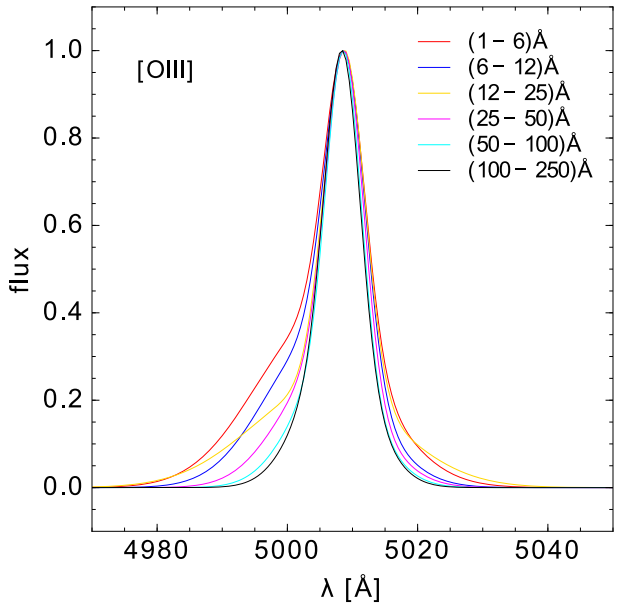


Figure 2.8: [OIII] λ5007Å profile for each $EW_{[O\,III]}$ representative spectrum. Profiles are normalized to their peak values.

95% of the integrated line flux, IPV 90%) for the broad components of the three emission lines as a function of $EW_{[O\,III]}$.

All the parameters grow rather steadily going from low (near the face-on position) to high EWs (near the edge-on position). FWHM, σ and IPV have been computed on the best fit line profiles. Statistical errors for these quantities have been evaluated as explained in Section 2.3.2, but due to their small values (ranging from a few 10^{-5} up to a few 10^{-3}) we have conservatively assigned a 1% systematic error to each of these quantities. Values and errors for each stack are listed in Tab. 2.3.

2.4.2 Narrow Lines

We then performed a detailed study of the [O III] λ5007Å line, the most prominent NLR emission feature. Fig. 2.8 shows the [O III] profile obtained for each $EW_{[O\,III]}$ stack. The blue component decreases going from low EWs ("face-on" positions) to high EWs (edge-on positions). This behaviour is what is expected if we associate the blue component to an outflow: the preferential direction of outflows is perpendicular to the accretion disk and so the observed outflow velocity will correspond to the intrinsic one only in the face-on position, while in any other position this will be decreased by the factor $\cos\theta$. An analysis of the shifts in the velocity of the two (main and blue) [O III] components has also been performed: we have examined the shift of the central velocity of the two [O III] components with respect to the [O II] doublet $\lambda\lambda$3727.092Å,3729.875Å velocity in each stack ($v_{[OIII]} - v_{[OII]}$), assumed to be the systemic velocity of the "host galaxy" of the relative stack, as first suggested by

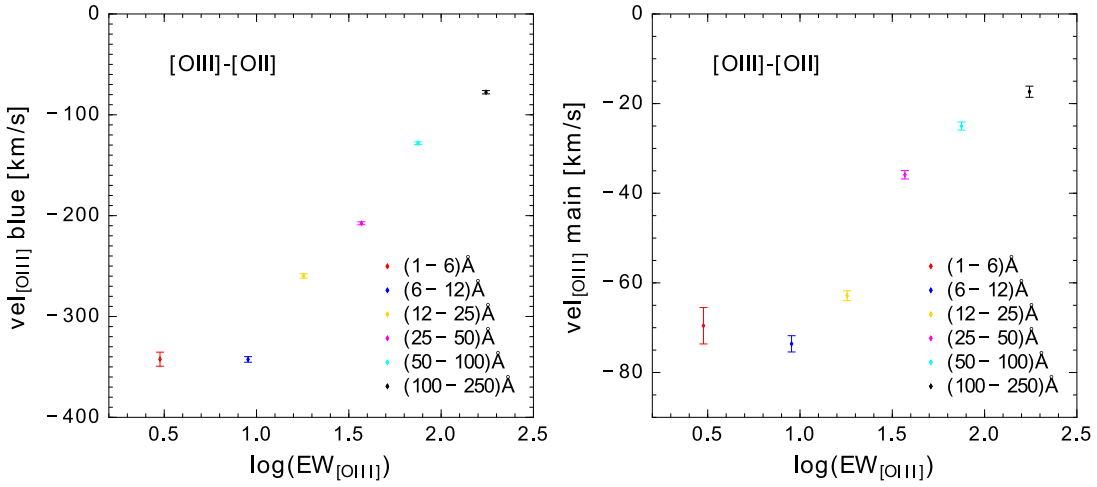


Figure 2.9: Velocity shifts for the main (left panel) and blue (right panel) [O III] components with respect to [O II] velocity, representing the systemic velocity for the host galaxy.

Table 2.3: (4400 − 5400)Å slope continuum and broad component best fit FWHM, σ and IPV(90%) of Hβ, Hα and Mg II for each $EW_{[O\,III]}$ bin.

$EW_{[O\,III]}$ (Å)	[1 − 6]	[6 − 12]	[12 − 25]	[25 − 50]	[50 − 100]	[100 − 250]
(4400 − 5400)Å slope	−1.175 ± 0.001	−1.393 ± 0.001	−1.370 ± 0.001	−1.208 ± 0.002	−1.152 ± 0.003	−1.151 ± 0.007
Hβ FWHM (km/s)	3212 ± 32	3897 ± 38	4150 ± 41	4347 ± 43	4520 ± 45	4464 ± 45
Hβ σ (km/s)	2874 ± 29	3285 ± 33	3345 ± 33	3520 ± 35	3708 ± 37	4203 ± 42
Hβ IPV(90%) (km/s)	9353 ± 94	10648 ± 106	10821 ± 108	11357 ± 114	11999 ± 120	13626 ± 136
Hα FWHM (km/s)	2092 ± 21	2539 ± 25	2736 ± 27	2727 ± 27	3078 ± 31	2886 ± 29
Hα σ (km/s)	1803 ± 18	2064 ± 21	2137 ± 21	2301 ± 23	2349 ± 23	2374 ± 24
Hα IPV (90%) (km/s)	5864 ± 59	6709 ± 67	6938 ± 69	7481 ± 75	7632 ± 76	7719 ± 77
Mg II FWHM (km/s)	2335 ± 23	2876 ± 29	3107 ± 31	3487 ± 35	3770 ± 38	3956 ± 40
Mg II σ (km/s)	1850 ± 19	2051 ± 21	2305 ± 23	2586 ± 26	2475 ± 25	3143 ± 31
Mg II IPV (90%) (km/s)	6021 ± 60	6652 ± 67	7482 ± 75	8395 ± 84	8016 ± 80	10209 ± 102

Table 2.4: Velocity shifts for main and blue [O III] components with respect to the central velocity of [O II] (representing the systemic velocity for the host galaxy), IPV(90%) and asymmetry index for the total [O III] profile for each $EW_{[O\,III]}$ bin.

$EW_{[O\,III]}$ (Å)	[1 − 6]	[6 − 12]	[12 − 25]	[25 − 50]	[50 − 100]	[100 − 250]
[O III] main shift (km/s)	−70 ± 4	−74 ± 2	−63 ± 1	−36 ± 1	−25 ± 1	−17 ± 1
[O III] blue shift (km/s)	−342 ± 7	−343 ± 3	−260 ± 2	−208 ± 1	−128 ± 1	−78 ± 2
IPV(90%) (km/s)	1606 ± 16	1320 ± 13	1722 ± 17	1091 ± 11	914 ± 9	833 ± 8
$A_{IPV_{05-95}}$	1.547 ± 0.015	1.686 ± 0.017	1.431 ± 0.014	1.513 ± 0.015	1.328 ± 0.013	1.173 ± 0.011

Boroson (2011) (Fig. 2.9). The shift in velocity clearly decreases going from low EWs to high EWs with a higher absolute value for the blue component with respect to the main one. The shifts and their errors (converted to velocity shifts starting from wavelength shifts from the Gaussian curve fits) are listed in Tab. 2.4.

2.4.3 Fe II

Fe II emission originates from the BLR. The intensity of these lines is quite weak if compared to other broad lines; in addition, because of the presence of multiplets, they are often blended and, therefore, difficult to disentangle from each other in our spectra. We paid special attention to the analysis of those features due to the importance they hold in the *Eigenvector1*, i.e. the anticorrelation between Fe II and [O III] intensities in quasars spectra, responsible for most of the spectral variance in optical spectra of quasars (Boroson & Green, 1992; Sulentic & Marziani, 2015).

This anticorrelation is commonly seen in quasar spectra so its presence in our spectra stacks is not surprising. What could be more interesting is the fact this anticorrelation has a trend with the observed $EW_{[O\,III]}$; moving from stacks corresponding to low $EW_{[O\,III]}$ values to stacks corresponding to higher values, Fe II emission becomes less and less intense, while [O III] line behaves just the opposite. This trend discloses a dependence of Eigenvector1, at least partially, on orientation. In low $EW_{[O\,III]}$ stacks ("face-on" objects) the BLR is face-on and so the Fe II emission is seen at the peak of its intensity. Recently Shen & Ho (2014) examined the relationship among the strengths of Fe II and [O III] with the Hβ linewidth in the context of Eigenvector1. They ascribe the negative correlation between the Hβ linewidth and the Fe II strength to the BH mass or, equivalently, the Eddington Ratio. In their scenario this property is the main driver of EV1, with the orientation only contributing to the dispersion of the linewidths at fixed Fe II strength. We find similar trends between these spectral characteristics, but the main difference in our interpretation is that we believe the orientation to be one of the drivers of the EV1 and explain the anticorrelation of the Hβ linewidth and Fe II strength as an orientation effect. We note that other physical drivers of the EV1, such as the BH mass, luminosity or Eddington ratio, may be relevant. In our analysis this effects, if present, are diluited by the average of the whole quasar population in each bin of $EW_{[O\,III]}$.

2.4.4 Double-peaked broad lines

If the Hβ line is emitted by a flat distribution of clouds, we expect that double-peaked lines are more frequent in edge-on sources. In order to investigate this point, we considered the $EW_{[O\,III]}$ distribution of quasars with double peaked lines in our sample. Shen et al. (2011) provide a flag indicating "unambiguous", and "possible" double-peaked lines, based on visual inspection. Given the non-homogeneous nature of this selection a large spread in the distribution is expected; however, the relatively large number of objects (about 100 "unambiguous" and 500 "possible") allows for a comparison with the global distribution of

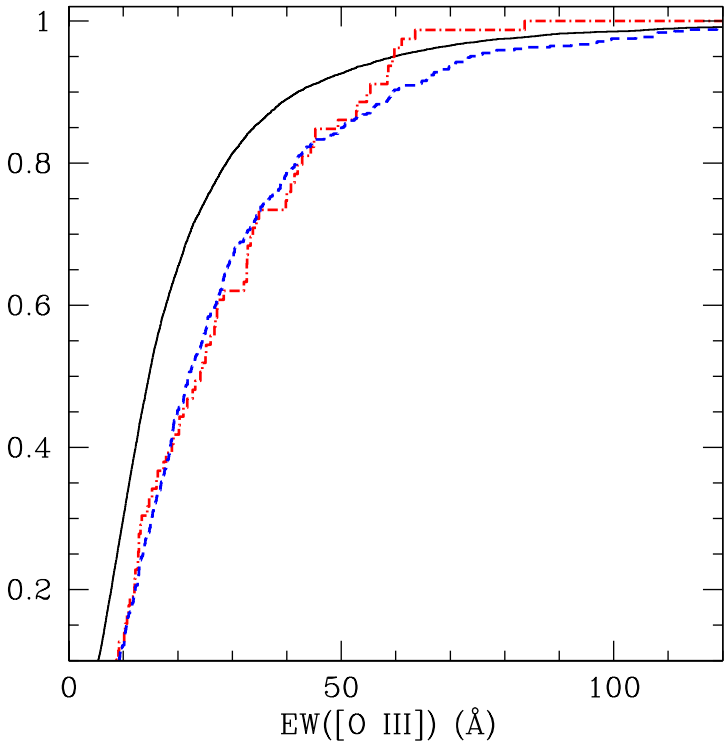


Figure 2.10: Cumulative distributions of $EW_{[O\,III]}$ for the whole sample (black, continuous line), the "unambiguous" double-peaked quasars (red, dot-dashed), and the "possible" double peaked quasars (blue, dotted line).

Table 2.5: Mean values of [O III] and continuum at 5100Å luminosities distributions for each $EW_{[O\,III]}$ bin.

$EW_{[O\,III]}$ (Å)	(1 – 6)	(6 – 12)	(12 – 25)	(25 – 50)	(50 – 100)	(100 – 250)
$\overline{log(L_{[OIII]})}$ $(ergs^{-1})$	41.5 ± 0.4	41.9 ± 0.3	42.1 ± 0.4	42.4 ± 0.3	42.6 ± 0.3	42.9 ± 0.3
$\overline{log(L_{5100})}$ $(ergs^{-1})$	44.6 ± 0.4	44.6 ± 0.3	44.6 ± 0.3	44.6 ± 0.3	44.5 ± 0.3	44.5 ± 0.3

$EW_{[O\,III]}$. In Fig. 2.10 we show the cumulative $EW_{[O\,III]}$ distributions for the global sample, and the two "unambiguous" and "possible" double-peaked subsamples. It is clear that the double peaked objects are on average shifted towards higher values of $EW_{[O\,III]}$. We performed a Kolmogorov-Smirnov test on the distributions and we found that the probability that the shift towards high $EW_{[O\,III]}$ in double peaked objects is NOT significant is below 10^{-4} for both the "unambiguous" and the "possible" double-peaked quasars.

2.4.5 $L_{[OIII]}$ and L_{5100} distributions

As a final point in examining the $EW_{[O\,III]}$ bins we consider the luminosity distributions of [O III] and continuum at 5100Å (Fig. 2.11). The flux-limited selection is visible in the continuum luminosity distribution that is steadily peaked around $log(L_{5100}) \sim 44.6$ although it can be recognized a small shift in the central values, ascribable to the expected decrease in

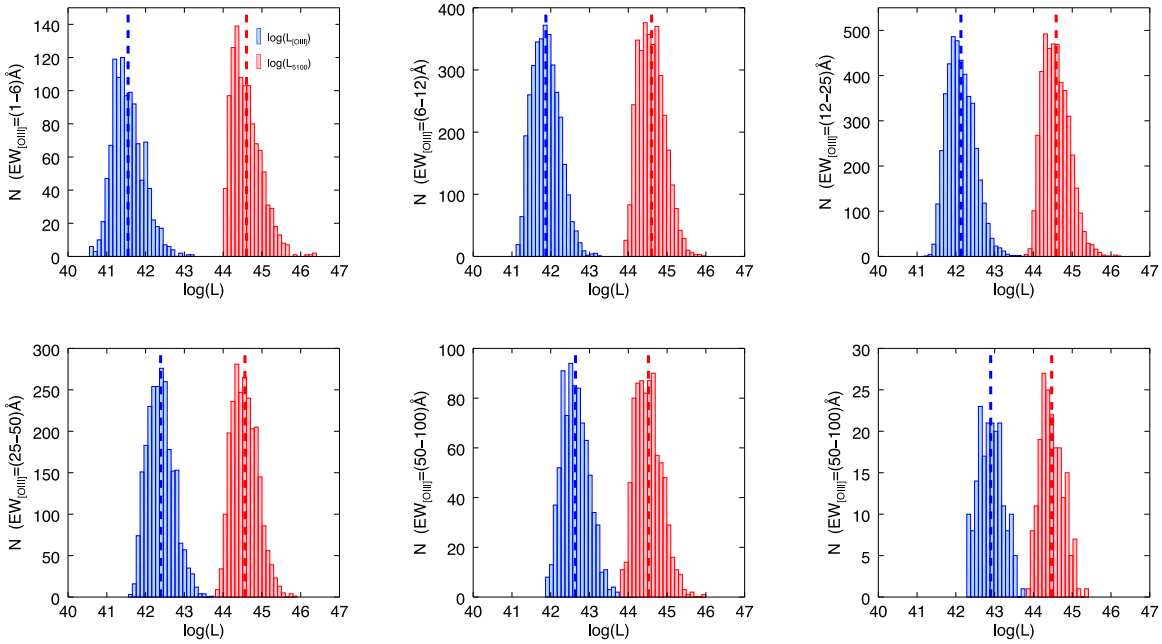


Figure 2.11: $L_{[OIII]}$ and L_{5100} distribution for each stack (the dashed lines represent the mean value for each distribution). The distribution of sources in terms of their continuum at 5100Å luminosity is stable around ~ 44.6, a consequence of the flux-limited selection. On the other hand the central value of sources distribution in terms of [O III] luminosity increases going towards high $\mathrm{EW}_{[\mathrm{O\,III}]}$ bins. This is due to the flux limit of the sample: when moving towards edge-on positions we are selecting intrinsically more luminous objects.

continuum luminosity moving from face-on to edge-on positions. A more relevant effect of the flux-limited selection is present in the [O III] luminosity distribution: sources in edge-on position are selected only if they are intrinsically more luminous. Since [O III] is an isotropic indicator of the intrinsic disk luminosity we observe an excess of high $L_{[OIII]}$ in object with high $\mathrm{EW}_{[\mathrm{O\,III}]}$. The average values of $L_{[OIII]}$ and L_{5100} for each bin are listed in Tab. 2.5.

2.5 Discussion

Our analysis of the EW distribution of [O III] and of the stacked spectra of quasars with different $\mathrm{EW}_{[\mathrm{O\,III}]}$ have revealed several relevant properties of the narrow and broad line regions:

1. The distribution of $\mathrm{EW}_{[\mathrm{O\,III}]}$ is well reproduced by an intrinsic log-normal distribution convolved with a high-EW power law tail. The maximum and dispersion of the intrinsic distribution are $\mathrm{EW}_{MAX} = 11$Å and $\sigma = 9$Å. The exponent of the power law tail is $\Gamma = 3.5$.

2. The distribution of EW of Hβ is well represented by a Gaussian distribution with $\mathrm{EW}_{MAX} = 58$ Å and $\sigma = 23$ Å and a power law tail with $\Gamma \simeq 7$.

3. The [O III]/Hβ distribution shows the same high end tail as $\mathrm{EW}_{[\mathrm{O\,III}]}$.

4. The [O III] line shows a blue tail whose intensity decreases moving from low to high $EW_{[O\,III]}$. The blueshift similarly decreases with EW.

5. The width of broad lines increases moving from low to high $EW_{[O\,III]}$.

6. Fe II emission is prominent for low $EW_{[O\,III]}$ and its intensity decreases moving to high $EW_{[O\,III]}$.

7. Double peaked broad lines objects are more frequent for high $EW_{[O\,III]}$ with respect to low $EW_{[O\,III]}$.

Here we discuss the physical consequences of our results.

2.5.1 Distribution of $EW_{[O\,III]}$ and EW(Hβ)

The distribution of $EW_{[O\,III]}$ is relevant in two respects: (1) the high-EW tail $\Gamma = 3.5$ fully confirms the scenario of an isotropic emission of the [O III] λ5007 Å line, with intensity proportional to the illumination from the ionizing source, and a disc-like continuum emission; (2) the ratio between the width of the intrinsic distribution and its peak value is an estimate of the precision of the [O III] luminosity as an indicator of the bolometric luminosity (which is supposed to be dominated by the disk emission). From our results, we conclude that an estimate of the bolometric/disk luminosity based on the [O III] line has an uncertainty of a factor of ∼2. The possible reasons for the observed intrinsic dispersion of $EW_{[O\,III]}$ are variations in the covering factor of the Narrow-Line Region clouds as seen from the disc and effects of the dispersion in the optical/UV spectral energy distribution (the emission of the [O III] line is expected to be proportional to the disk emission at the line *ionizing* frequency of [O III], i.e. $\sim$ 50 eV, while the continuum is measured at the *emission* frequency). These effects are discussed in Risaliti et al. (2011). We note that in principle an increase in $EW_{[O\,III]}$ could be due to dust reddening of the disc component. In this case, however, the effect of dust reddening should be seen also on continuum spectra, while we have shown in Section 2.3.1 that our selection ensures that only blue objects are present (see also Tab. 2.3).

The distribution of EW(Hβ) strongly suggests a disc-like emission of the line. The orientation effects found in the broad emission lines require the BLR to be not only flat, but also optically thick to these lines. This is likely to be the case for the Hβ line: for densities and column densities typical of BLR clouds ($n > 10^9 - 10^{10}$ cm^{-3} and $N_H > 10^{23}$ cm^{-2}) the optical depth of the Lyα line is expected to be higher than 10^4, and the optical depth of the Balmer lines start to be significant when τ(Lyα) is higher than a few hundred (Osterbrock & Ferland, 2006).

Moreover, the distribution of R=[O III]/Hβ confirms the suggestion of a disk-like shape for the Hβ emitting region. If the BLR geometry resembles that of the accretion disk then this ratio is a close version of the $EW_{[O\,III]}$, the difference between the two being determined by the larger height scale of the disk of the BLR with respect to that of the accretion disk, probably caused by the presence of turbulence in the gas componing this structure.

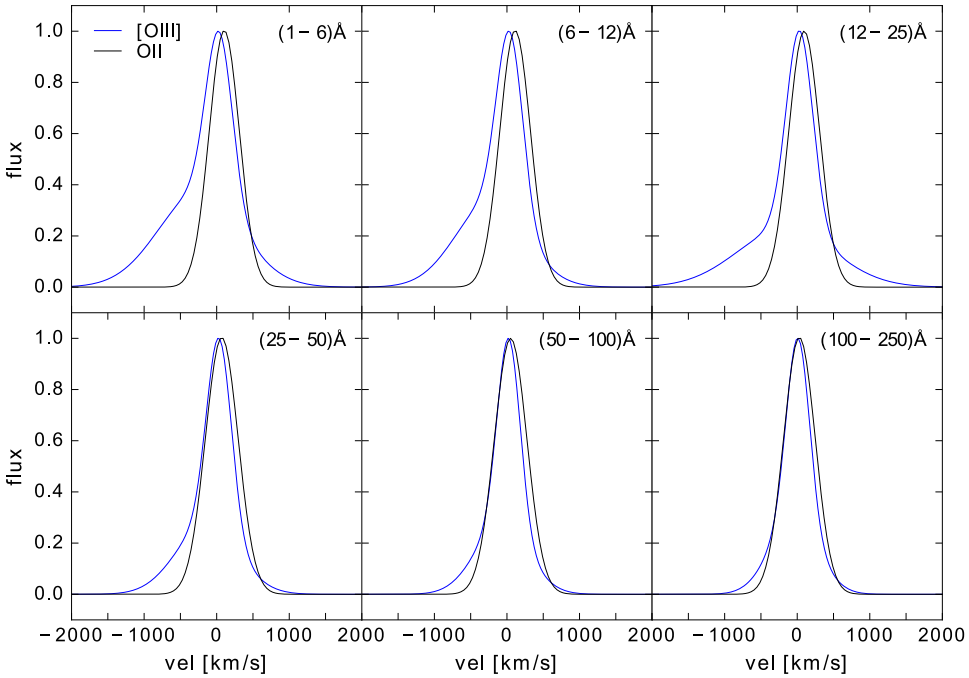


Figure 2.12: [O III] and [O II] profiles comparison for each stack; the [O III] blue component decreases moving from low to high $EW_{[O\,III]}$ stacks, i.e. from face-on to edge-on positions. From the relative position of the [O III] peak with respect to the [O II] peak we can also estimate the systematic error in sources redshift: the [O III] peak is systematically shifted with respect to the [O II] one and the shift is decreasing moving towards higher $EW_{[O\,III]}$ stacks (see Tab. 2.4).

2.5.2 The [O III] line profiles

The spectral results on the [O III] emission line can be interpreted as a simple consequence of the increasing inclination of the sources going from low EWs to high EWs.

In Fig. 2.12 [O III] and [O II] 3727.092, 3729.875Å profiles are compared for each of the six stacks. The [O III] profile shows a prominent blue tail decreasing toward high $EW_{[O\,III]}$, while [O II] holds quite steady in all the bins.

The [O III] blue component is due to gas in outflow from the NLR towards a direction mostly perpendicular to the plane of the accretion disk. In this scenario, the angle between the outflow direction and the observer line of sight is the same as the disk inclination angle. The velocity component along the line of sight is therefore

$$v_{obs} = v_{outflow} \cos\theta \simeq v_{outflow} \frac{EW[OIII]_{INT}}{EW[OIII]_{OBS}} \quad . \tag{2.4}$$

The increase in the central velocity shift of the blue [O III] component with respect to the systemic velocity of the host galaxies ($v_{[OIII]} - v_{[OII]}$) has the same explanation: the shift is more important when the object is face-on because we are observing the outflow exactly along the line of sight (Fig. 2.9 right panel) (the same result was found in Boroson (2011)).

A somewhat more surprising result is the measured blueshift in the [O III] main component. This finding can be explained as an indirect consequence of orientation effects on the

global [O III] profiles used by Shen et al. (2011) for the estimates of the redshifts. Since we use redshifts from this reference, we are obtaining an inclination dependent systematic shift: more face-on sources have a more prominent blue tail, and so a bluer central λ in the global profile. This bias is instead negligible in edge-on objects (Fig. 2.9 left panel).

To give a more quantitative measurement of the [O III] profile degree of asymmetry we evaluate an asymmetry index similar to that defined in Heckman et al. (1981) and based on differences between Inter-Percentile Velocities (IPV); the asymmetry index is defined as

$$A_{IPV_{05-95}} = \frac{v_{50} - v_{05}}{v_{95} - v_{50}} \,, \tag{2.5}$$

where v_{05}, v_{50} and v_{95} are the velocities corresponding to the wavelengths including 5%, 50%, and 95% of the line total flux. Through this definition we are able to quantify the asymmetry of a line; with a $A_{IPV_{05-95}} > 1$ the line is characterized by a blueward asymmetry, while for $A_{IPV_{05-95}} < 1$ the line is more prominent in the red part of its profile. The asymmetry index for each stack is reported in Tab. 2.4. $A_{IPV_{05-95}}$ decreases moving towards high $EW_{[O\,III]}$ in agreement with the result in Fig. 2.8 (i.e. a blue tail becoming less prominent at higher $EW_{[O\,III]}$.

2.5.3 Broad line profiles

The number of studies supporting a non-spherical shape of the BLR has constantly grown in recent years. Zhu et al. (2009) analyzed the BLR profiles in SDSS quasars and suggests that two components with different geometries and physical conditions are needed to reproduce the observed spectra; others studies also suggest the presence of two components, one of them with a spherical geometry while the outer one disk-shaped (Bon et al., 2006). In the same BLR disk-shaped scenario it has also been suggested that the kinematics of this inner region, consisting of a combination of rotational and turbulent motions, could affect its geometry, with broader lines emitted from more flattened regions (Kollatschny & Zetzl, 2011). Moreover, the proximity of the BLR to the obscuring "torus" of the Unified Model (Antonucci, 1993) suggests a smooth connection between the two structures, rather than two completely separated regions, as usually described in the standard unification model. Indeed, the BLR could represent a transition region from the outer accretion disk to the dusty region of the torus (Goad et al., 2012). Recently Pancoast et al. (2014) used direct modelling techniques on a sample of AGN for which high-quality Reverberation Mapping data were available in order to investigate the geometry and the dynamics of the BLR. They found that the geometry of the BLR, as traced by Hβ emission, is consistent with a thick disk.

All these works, despite different aims and explanations, share a common interpretation of the geometry of the BLR.

We claim that the dependence of the broad components on $EW_{[O\,III]}$ is an evidence of the disk-like shape of the BLR: moving from low to high EWs (that is from "face-on" to edge-on objects) the component of velocity of the BLR in the direction of the observer grows steadily with the cosine of the inclination angle.

Taking into account this result leads to a number of improvements in our understanding of the AGN inner regions, starting from the determination of the SMBHs virial masses. As we mentioned in Section 1.3.2, as long as we consider the BLR as composed by virialized gas, the SMBH mass can be inferred from the BLR lines width according to the relation

$$M_\bullet = f\frac{v_{oss}^2 R_{BLR}}{G} \quad , \tag{2.6}$$

where v is the BLR observed line width and R_{BLR} can be obtained from Reverberation Mapping and from the luminosity-R_{BLR} relation for single epoch observations (Kaspi et al., 2000; Bentz et al., 2013). An orientation-dependent analysis of emission lines could on the one hand improve our knowledge of the morphology of the BLR and so help us in determining more accurately the *virial factor* f (Shen, 2013), and on the other hand remove the systematic underestimate of the line widths in non edge-one sources.

2.5.4 Eigenvector 1

The orientation effects revealed using $EW_{[O\,III]}$ as an inclination indicator provide a possible interpretation of *Eigenvector 1* (EV1) (Boroson & Green, 1992), namely the anticorrelation in the intensity of the emissions of Fe II and [O III]. FeII features are BLR lines and so they show the same trend as the other broad lines, i.e. a decreasing intensity and increasing broadening going from face-on (low EWs) to edge-on (high EWs) positions. On the other hand, the [O III] emission is isotropic. EV1 can then be simply explained in terms of the orientation effects as follows: assuming a disk-like shape for the BLR the intensity of Fe II emission lines decreases from face-on to edge-on positions. This effect is clearly present in our stacks: FeII emissions are more prominent in stacks with low $EW_{[O\,III]}$ (Fig. 2.13). Fe II emissions seem to disappear moving towards high $EW_{[O\,III]}$, rather than decrease in intensity as the other broad lines do. This fact may suggest that the Fe II disk-like structure, besides being flatter than the other broad lines emitting regions, is flatter than the continuum emitting region itself, i.e. the accretion disk. Unfortunately the Fe II spectrum is characterized by the presence of several, close multiplets and its emissions are by far less intense than those of the other broad lines. This complicates a deeper investigation on this subject. It is possible, in fact, that for the two reasons mentioned above we are simply not able to detect a behaviour of FeII emissions similar to that of the other broad lines.

2.6 Summary

In this work we analyzed the spectral properties of SDSS quasars as a function of the inclination of the accretion disk with respect to the line of sight. We used the equivalent width of the [O III] λ5007Å line as an orientation indicator. Our main results are the following.

1. The $EW_{[O\,III]}$ observed distribution is the convolution of an intrinsic log-normal distribution and a power law tail $\propto EW^{-3.5}$, as expected for a randomly distributed popula-

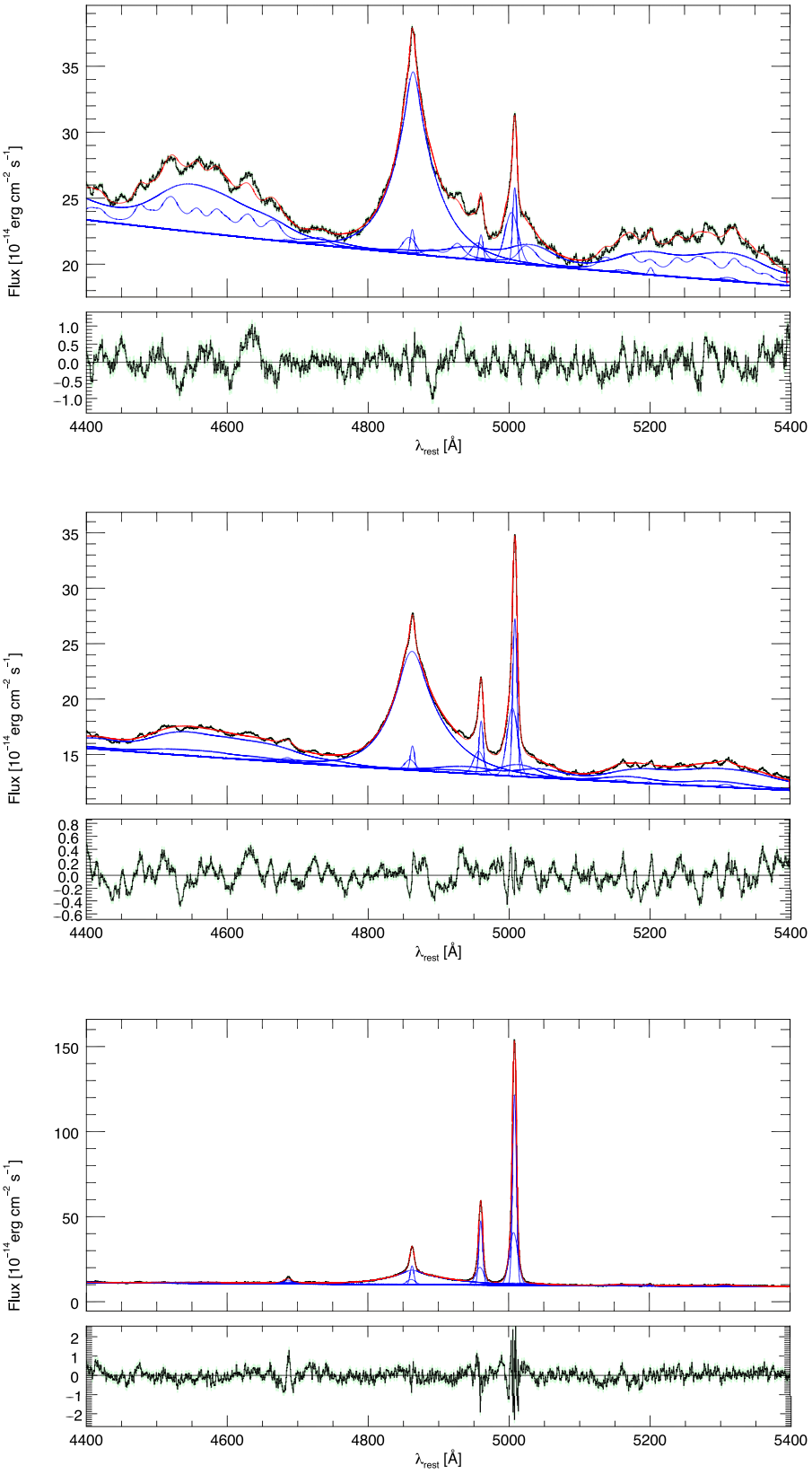


Figure 2.13: Fits of the Hβ-[O III] spectral window for the (1 – 6)Å, (12 – 25)Å, (100 – 250)Å stacks, respectively. The spectrum is fitted with several functions: for the broad component of permitted lines a double power law convolved with a Gaussian is used. Narrow lines (for both permitted and forbidden lines) are fitted with two Gaussian, the first one accounting for the main component of the line, the other one accounting for the blue tail ascribed to outflowing gas fron the NLR. For the FeII emission several templates are taken into account (see Section 2.3.2 for details).

tion of disks in a flux-limited sample (R11).

2. The EW(Hβ) instead does not show the same tail, suggesting that the BLR could be characterized by a disk-shape geometry.

3. The R = [O III]/Hβ distribution resembles the $EW_{[O\,III]}$ observed distribution, with a slightly different value for the power law tail, probably an indication of the deviation of the BLR from a pure disk-like structure.

4. The [O III] line has a blue tail, whose intensity and blueshift with respect to the rest frame wavelength both decrease moving from low to high $EW_{[O\,III]}$.

5. All the broad lines behave in the same way: the width of the line increases moving from low to high $EW_{[O\,III]}$.

6. The Eigenvector 1 too has a trend with the $EW_{[O\,III]}$; the Fe II emission is strong when $EW_{[O\,III]}$ is low and diminishes gradually while $EW_{[O\,III]}$ increases.

7. Double peaked objects are more probable for high $EW_{[O\,III]}$ than for low $EW_{[O\,III]}$.

All these findings can be uniquely and satisfactorily explained with the only use of inclination effects; we claim that is very difficult to find an alternative scenario able to account for all these evidences in a similarly simple and straightforward way. On the other hand, if the optical/UV of quasars is due to geometrically thin disks, such observational effects are unavoidable. The scheme we present makes indeed use of only two physical hypothesis: 1) the source of the UV-optical continuum is an optically thick and geometrically thin accretion disk, whose luminosity decreases with the cosine of the inclination angle, 2) the [OIII] emission is isotropic and is ascribed to the illumination of the NLR by the same accretion disk. Such hypothesis naturally arises from our current knowledge on AGN and commonly accerted hypotheses on their nature.

2.7 Future work

The identification of an orientation indicator for quasars allows to study many aspects of these sources from a different perspective. There are several implications of this finding that will be the subject of an upcoming paper. In this section I will give a short introduction about two of them.

2.7.1 Inverse distribution: the inclination of an AGN as a function of EW[O III]

Based on $EW_{[O\,III]}$ as an orientation indicator for quasars, we want to obtain the orientation angle of a source from a measurement of this quantity. Given the width of the intrinsic

$EW_{[O\,III]}$ distribution, such an estimate will also have a dispersion. Therefore, for each value of $EW_{[O\,III]}$, we can only give a probability distribution for the source inclination angle.

We start again from the expression for the differential number of sources with an intrinsic luminosity L_{int}, intrinsic $EW[O\,III]_{int}$, inclination angle θ and distance R that we used in Risaliti et al. (2011):

$$dN = \Phi(L_{int})dL_{int}g(EW_{int})dEW_{int}d(\cos\theta)R^2dR\ , \tag{2.7}$$

where $\Phi(L_{int})$ and $g(EW_{int})$ are the distribution functions of the intrinsic luminosity and intrinsic [O III] equivalent width, respectively.

To find the probability distribution of the inclination angle given an observed value of EW[O III], we need to express the intrinsic properties in terms of the observed ones, i.e. we need to perform the variable change $(\cos\theta, R) \longrightarrow (EW_{obs}, F)$. Given the anisotropic and isotropic characteristics of continuum and [O III] line respectively, we can write $\cos\theta = \frac{EW_{int}}{EW_{oss}}$ and therefore $d(\cos\theta) = \left|\frac{EW_{int}}{EW_{obs}^2}\right| dEW_{obs}$. Solving R for luminosity and flux gives instead $dR = \frac{1}{2}\sqrt{\frac{L_{obs}}{4\pi}}\frac{1}{F^{3/2}}dF$. Applying this transformation, the previous equation becomes:

$$dN = \frac{L_{int}}{F^{5/2}EW_{obs}}\left(\frac{EW_{int}}{EW_{obs}}\right)^{5/2}\Phi(L_{int})g(EW_{int})dEW_{obs}dEW_{int}dL_{int}dF\ . \tag{2.8}$$

The observed distribution for EW[O III] is then obtained from Eq. 2.8 if we assume a luminosity function of the form $\Phi(L_{int}) \sim L_{int}^{-\delta}$:

$$\frac{dN}{d(EW_{oss})} = \int_{L_{min}}^{L_{max}} L_{int}^{3/2-\delta}dL_{int} \quad \int_{F_{min}}^{F_{max}}\frac{dF}{F^{5/2}}\int_0^{EW_{oss}}\frac{EW_{int}^{5/2}}{EW_{oss}^{7/2}}g(EW_{int})d(EW_{int})\ . \tag{2.9}$$

The behaviour of the observed EW[O III] distribution for high EW_{obs} values given in Risaliti et al. (2011) is then retrieved integrating over luminosity, flux and intrinsic equivalent width. In this case instead, we perform only the first two integration in order to get the distribution of sources in terms of observed and intrinsic EW[O III]:

$$\frac{dN}{d(EW_{oss})d(EW_{int})} \propto \frac{EW_{int}^{\gamma-1}}{EW_{oss}^{\gamma}}g(EW_{int})\ , \tag{2.10}$$

where $\gamma = 7/2$ is the exponent of the power law defining the $EW_{[O\,III]}$ observed distribution behaviour at high EW. Considering again the relation between EW_{int} and EW_{oss}, we can write Eq. 2.10 as

$$\frac{dN}{d(\cos\theta)dEW_{oss}} \propto (\cos\theta)^{\gamma-1}f(EW_{oss}\cos\theta)\ , \tag{2.11}$$

i.e., as the bidimensional distribution of objects in the space of inclination angle and observed $EW_{[O\,III]}$.

It is now straightforward to infer the objects distribution with respect to the inclination angle for a fixed observed $EW_{[O\,III]}$:

$$P(\cos\theta \mid EW_{oss}) = \frac{\frac{dN}{d(\cos\theta)dEW_{oss}}}{\frac{dN}{dEW_{oss}}} = \frac{EW_{oss}^{\gamma}(\cos\theta)^{\gamma-1}f(EW_{oss}\cos\theta)}{\int_0^{EW_{oss}}(EW_{oss}\cos\theta)^{\gamma-1}f(EW_{oss}\cos\theta)d(EW_{oss}\cos\theta)}\ . \tag{2.12}$$

Table 2.6: Central values (EW_{oss}) of the ranges (ΔEW_{oss}) in which the SDSS sample was divided in Bisogni et al. (2017); for these values the $\cos\theta$ probability distribution has been computed and the corresponding mean values ($\overline{\cos\theta}$ and $\bar{\theta}$) has been evaluated.

ΔEW_{oss} (Å)	(1-6)	(6-12)	(12-25)	(25-50)	(50-100)	(100-250)
EW_{oss} (Å)	3.0	9.0	18.5	38.5	75.0	175.0
$\overline{\cos\theta}$	0.90	0.91	0.88	0.61	0.30	0.13
$\bar{\theta}$ (°)	25.5	25.1	28.1	52.1	72.5	82.6

Table 2.7: Best-fit parameters values for the double gaussian $EW_{[O\,III]}$ intrinsic distribution as reported in Risaliti et al. (2011). EW_1^*, EW_2^*, σ_1 and σ_2 represent respectively mean value and standard deviation for the two gaussians; α is the relative weight of the first gaussian in the intrinsic distribution, that is $\alpha = N_1/(N_1 + N_2)$, with N_1 and N_2 normalizations of the two gaussians.

EW_1^* (Å)	8.0 ± 0.3
σ_1 (Å)	4 ± 0.3
EW_2^* (Å)	17 ± 1
σ_2 (Å)	11 ± 0.8
α	0.67 ± 0.01

We want to know how this distribution behaves at different fixed values of observed $EW_{[O\,III]}$. The obvious choice for the observed EW values are the central values of the ranges in which we divided the SDSS sample in Bisogni et al. (2017).

We therefore considered the distribution for the values listed in Tab. 2.6 using as $EW_{[O\,III]}$ intrinsic distribution the best-fit double gaussian of Risaliti et al. (2011), whose parameters are listed in Tab. 2.7; this function had indeed the best χ^2 compared with other choices.

Fig. 2.14 shows the probability functions for each of the chosen EW values. In Risaliti et al. (2011) we noticed how the intrinsic distribution influenced in a much stronger way the first part of the observed distribution, relating to low EWs, rather than the second one, dominated by orientation effects.

The same behaviour can be recognised in the probability distribution of inclination angles: at low observed EWs there is a larger probability for the source to have a low inclination angle, all the same the probability is not zero for high inclination angle. This means that, although a low inclination angle is more probable, it is not excluded for the source to be highly inclinated with respect the line of sight; intrinsic distribution and orientation effect are mixed. For high observed EW, on the other hand, the distribution is very peaked, meaning that the orientation effect dominates; such a high value for the observed EW can not be due to the intrinsic distribution, it must necessarily be a consequence of an high inclination. For each of the observed values considered, the mean value for $\cos\theta$, evaluated as the first moment of the distribution

$$(\overline{\cos\theta})_{EW_{oss}} = \int_0^1 \cos\theta\, P(\cos\theta \mid EW_{oss})\ , \tag{2.13}$$

and the resulting $\bar{\theta}$ are listed in Tab. 2.6. According to what has been stressed so far, distri-

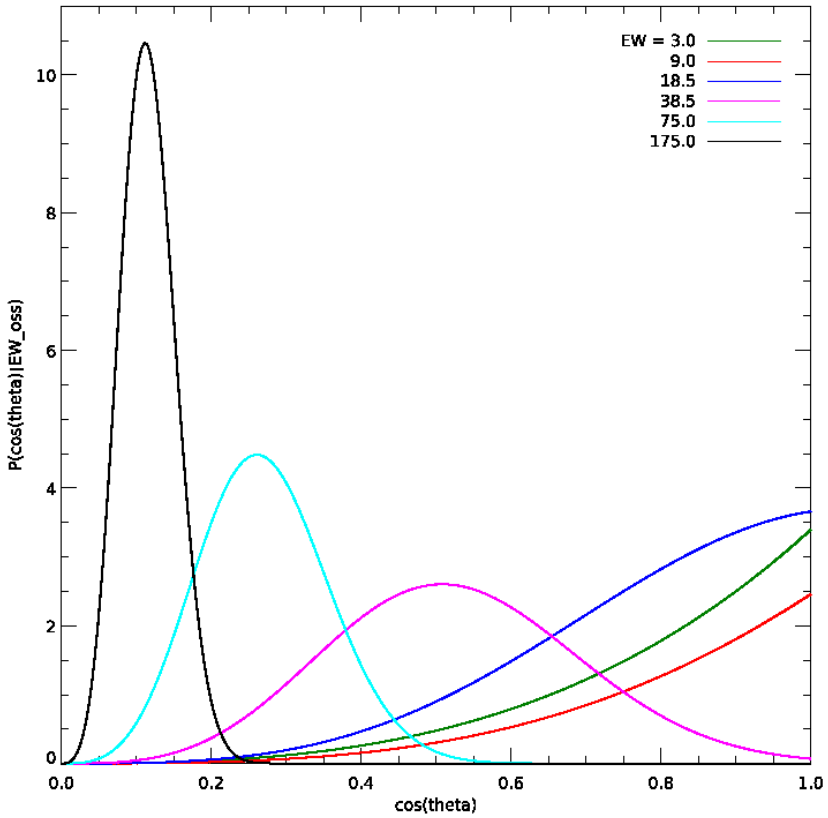


Figure 2.14: $\cos\theta$ probability distribution at fixed EW_{oss}. At low observed EW the probability of a low inclination angle is higher (or the probability to have a $\cos\theta$ close to 1 is higher) than that for an high inclination angle; even so the last one is not zero even for $\theta \sim \pi/2$. On the other hand for high EW the probability distribution gets more and more peaked and the maximum is found increasingly close to $\pi/2$.

butions related to low EWs have a mean inclination angle very similar and close to $\sim 25°$, corresponding to near face-on positions; going towards higher EWs, so getting rid of the influence of intrinsic distribution, the mean inclination angle increases and we can get more information about the source orientation with respect to the line of sight.

In an ideal situation in which it would be possible to remove the effect of the $\mathrm{EW}_{[\mathrm{O\,III}]}$ intrinsic distribution from the observed one, a measurement of the observed $\mathrm{EW}_{[\mathrm{O\,III}]}$ would allow a correct prediction of the source inclination. In the realistic case this is not feasible; in the case of highly inclinated sources it is however possible to get information about the source orientation.

Furthermore having a knowledge of the source orientation enables us to apply corrections to the virial BH mass estimation: as mentioned before, source orientation, due to the disk-like shape of the BLR, has as a main effect to reduce the width of emission lines broad components by a factor $\sin\theta$ going from edge-on to face-on positions, producing in this way an underestimation of the BH virial mass in all but the edge-on sources.

This kind of correction will moreover lead to a better calibration of the $R_{BLR} - L$ rela-

tion, reducing its dispersion and allowing a considerable improvement in single epoch mass determinations.

2.7.2 Infrared emission analysis: the TOR structure

TOR vs orientation

The study of IR data can give us information about the TOR (*Toroidal Obscuring Region*) structure, responsible for most of the differentiation between Type 1 and Type 2 AGNs (Elitzur, 2012).

Although many models have been developed for this structure (Fritz et al., 2006; Elitzur, 2008; Nenkova et al., 2008a,b; Goad et al., 2012; Stalevski et al., 2012), its real geometry and covering factor are still unknown.

In our analysis of orientation effects on quasars spectral features (Bisogni et al., 2017) we were able to observe BLR emission components even in edge-on sources. This evidence casts a doubt on the real obscuring capability of the torus or on its alignment with the accretion disk plane, as supposed in the unified model, and corroborates instead the clumpiness of the more recent models (Elitzur, 2012; Goad et al., 2012). An extreme hypothesis is the mere absence of this structure.

If such a structure presents itself in a toroidal shape however, we have to detect a variation in IR emission going from face-on to edge-on position sources.

What we expect is a more intense overall flux for sources corresponding to face-on positions, where the torus is facing the line of sight, while, for edge-on positions, a decrease in the flux, since we only have a side view of the structure in this case. As in the case of the [O III] emission line emitted by the NLR, however, also the infrared emission emitted by the torus will have an intrinsic distribution, even in this case depending on several characteristics, such as the properties of the continuum, whose emission is riprocessed by the torus, and the dimension, distance and covering factor of the tours itself. We cannot therefore connect an overall decrease in the flux to source orientation without taking into account an intrinsic dispersion of the infrared emission. The effect produced by orientation is in fact diluited by the presence of such an intrinsic distribution.

What we can instead relate with orientation is, other things being equal, a drop in the shorter wavelength part of the IR SED. In the model proposed by Nenkova et al. (2008b) the torus is represented as a clumpy structure, whose clouds angular distribution with respect to the equatorial plane is a gaussian with a width parameter σ. Such a distribution ensures both our capability to detect BLR emission even in edge-on position (being the distinction between Type1 and Type2 AGN a mere question of photon escaping probability) and an explanation of the behaviour at different wavelengths.

Nenkova et al. (2008b) associate the decrease in the blueward part of the IR SED, moving from face-on to edge-on positions, with two main reasons: an increasing number of clouds intercepted by the line of sight and a higher absorption of shorter with respect to longer

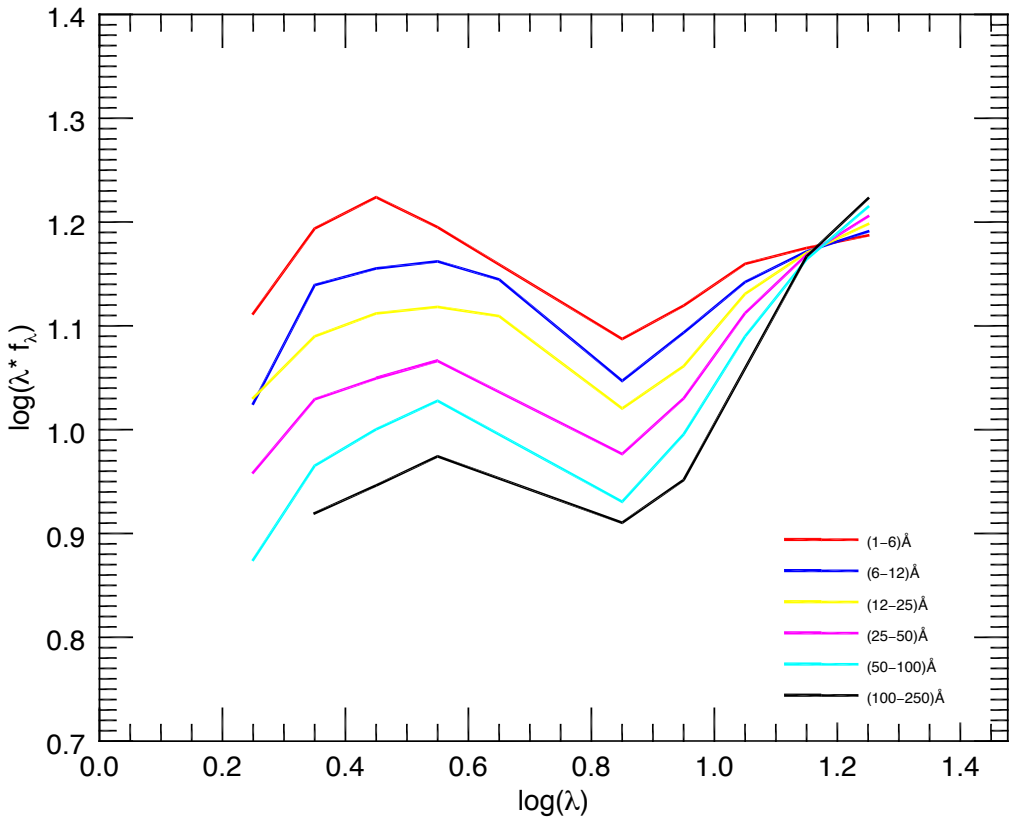


Figure 2.15: Representative infrared SEDs for each of the $EW_{[O\,III]}$ bins considered. Similarly to the case of the optical spectra, the SEDs were derived stacking the four photometric WISE points for sources pertaining to the same $EW_{[O\,III]}$ bin. After the flux normalization to the reference wavelength ($15\mu m$), it is possible to see the blueward part of the SEDs decreasing when moving from low to the high EW values, according to a change in the source orientation from *face-on* to *edge-on*.

wavelengths. The clouds distribution assures a nearly isotropic IR emission combined with a strongly anisotropic obscuration, essential to the explanation to different wavelengths behaviours.

If our interpretation of EW[O III] as an orientation indicator is correct, such a decrease in the blueward part of the SED should be seen going from low EW[O III] to high EW[O III]. We therefore examined IR data for the sources componing our SDSS quasars sample, dividing it in the same EW[O III] bins, with the aim of studying IR emissions in terms of orientation; in order to do that, we realized a representative stacked SED for each EW[O III] bin in a similar fashion to what we did for the optical spectra in Bisogni et al. (2017).

WISE data

The NASA *Wide-field Infrared Survey Explorer* (WISE) performed, during its ten months period of activity, a mid-infrared survey of the whole sky, covering four bands W_1, W_2, W_3 and W_4 centered at 3.4, 4.6, 12 and $22\mu m$ respectively (Wright et al., 2010).

The SDSS sources considered in Bisogni et al. (2017) have WISE photometric data, from which we can infer the infrared SED. The four photometric data have been de-redshifted according to the redshift as extracted cross-checking SDSS data. Once in the rest frame, each set of points has been interpolated to obtain the flux value at a reference wavelength (15μm). We choose a redward wavelength since, for an exam of orientation effects, we are interested in the variations in the blueward part of the IR SED. Since, as already pointed out, we have to deal with sources whose emission has intrinsic characteristics, the photometric points of individual sources have been normilized to the flux of the reference wavelength (15μm) to ensure that our analisys would not be influenced by intrinsic differences in the emitted flux.

As mentioned above, what we consider explanatory of the influence of orientation is, if present, a sistematic decrease in the blueward part of the SED moving from face-on to edge-on positions. Fig. 2.15 shows the representative SEDs for each considered EW bin. Moving from low to high $EW_{[O\,III]}$ we can indeed see how the flux corresponding to shorter wavelengths decreases (the redward part of the SED being normalized). This decrease is a function of the $EW_{[O\,III]}$ of the representative SED. Moving from the first to the last $EW_{[O\,III]}$ bin, we are moving from a position in which we are observing the torus *face-on*, detecting the flux emitted from the clouds componing the structure directly, to that in which the torus is seen *edge-on* and the flux emitted from the inner clouds are shielded and therefore absorbed at the shorter wavelengths from the outer ones.

Furthermore, the mere presence of an emission at these wavelenghts attests the existence of the torus. We can therefore exclude the extreme hypothesis of the absence of the torus to explain the detection of optical lines even in an edge-on position source.

3 Simultaneous detection of broad emission lines in Quasars

Broad lines enable us to give an estimation of the central SMBH. We mentioned in Chapter 1 how Single Epoch virial estimation empowers us to give a measurement of the SMBH, avoiding both the resolution and time limits imposed by kinematics and Reverberation Mapping respectively, and to have the opportunity to study quasars from a statistical point of view, given the large amount of data it makes available. Nonetheless this method has its number of issues we need to deal with, the most important being the intrinsic difference between the various emission lines. Lines with different ionisation potential are indeed emitted on average from different parts of the BLR and a comparison of their shape can reveal important information on the kinematics of the different emitting regions (see Section 1.4).

In this chapter we compare the spectra of six quasars at $z \sim 2.2$ obtained with the X-shooter spectrograph at VLT, whose spectral coverage (1200-7000Å rest-frame for these sources) allows for the simultaneous detection of all the main broad emission lines in the UV/optical spectral range. The simultaneous detection has the main advantage to avoid the fast variability of the source (on the short time-scale from days to minutes, depending on the band considered), a hallmark of AGN, and allows a proper comparison of different broad components in the same source. Such a comparison is especially important in the case of C IV λ1549, which is well known to divert the most from the behaviour of all the other lines. C IV broad component is believed to be contaminated by emission of gas in a non-virialised condition. Kinematical signatures of winds and outflows are recognised in the blueshift and asymmetry of the line. However, this broad line is the only one available for high redshift sources. An investigation on whether or not we should use it in virial estimation is therefore needed.

This projects is a follow-up of a previous work by Ho et al. (2012), who examined 7 quasar spectra at a lower redshift ($z \sim 1.5$) observed with the same instrument and for the same purpose. In this chapter we show the analysis we performed and the results we found for the new sample of 6 quasars, that are the subject of S. Bisogni, S. di Serego Alighieri, P. Goldoni, L. C. Ho, A. Marconi, G. Ponti and G. Risaliti (submitted to A&A).

3.1 Virial mass estimators comparison

We can recognise at least 4 different broad emission lines in a quasar UV/optical spectrum which are eligible as virial estimators: C IV, Mg II, Hβ, Hα. The BLR emission line most used for virial masses determination is by far Hβ λ4861, mostly because this line is known to be emitted by gas in virialised conditions and also because of its prominence in the optical spectral window. For very distant sources the optical range is no more available and we have to find a replacement for Hβ. The most promising candidates for this role are C IV λ1549 and Mg II λ2800 (McLure & Jarvis, 2002), both in the UV rest spectral frame of the sources; in using these lines we assume that they are emitted by approximatively the same region as Hβ, by gas in virialised conditions and therefore that they have widths comparable with that of this line. However, it is well known that the BLR is stratified in terms of ionization potential (Dietrich et al., 1993; Peterson & Wandel, 1999). C IV has a much higher ionisation potential than Hβ and even more so than Mg II. Moreover, C IV emission has very different behaviors depending on the source and exhibits very often a blueshift and general asymmetry. That does not fit well in a ordered keplerian motion scenario and is believed to be associated with gas in outflow or in a wind (Murray et al., 1995; Richards, 2012). It is then clear why the use of Balmer lines is more advisable in general and why Mg II can be considered a more reasonable replacement. When the redshift of the source only allows the use of C IV we are, however, compelled to find a solution. The most important point, therefore, consists in identifying which part of the line can be associated with the gas of the BLR in keplerian motion and which part we should instead consider in a non-virial state (Denney, 2012). Furthermore, the emission variability of AGN is more important for shorter wavelengths, and a comparison of C IV with optical lines is not truly reliable if these lines are not simultaneously detected. The simultaneous detection eludes the problems connected with the very fast variability of these lines, a typical signature of AGN spectroscopic emissions, and helps us in the search of possible connections between lines properties to obtain rules to use when the optical virial estimator (Hβ) is not available.

3.2 Sample selection, observation and data reduction

X-shooter (Vernet et al., 2011) is a three-arm, single-object echelle spectrograph which started operations in October 2009. The instrument covers simultaneously the wavelength range from 300 to 2400 nm in the three arms: UVB ($\Delta\lambda$ = 300-550 nm), VIS ($\Delta\lambda$ = 550-1020 nm) and NIR ($\Delta\lambda$ = 1020-2400 nm), respectively. For our observations we used slit widths of 1.3, 1.2 and 1.2 arcsec respectively for the three arms resulting in resolving powers $R = \lambda/\Delta\lambda$ = 4000, 6700 and 4300.

The sample was selected with the intent to repeat the work presented in Ho et al. (2012) (Paper I hereafter) at higher redshift. In that paper a sample of relatively bright (r$\sim$ 18-19) quasars from the SDSS DR7 release (Abazajian et al., 2009) with redshift around $\sim$

Table 3.1: Sample and observation properties. z_{Shen} are the redshifts initially used to bring the spectra to rest-frame wavelengths (Shen et al., 2011), z_{fit} is the redshift we estimate from the central wavelength of the [OIII]λ5007 line for every source, D_L is the luminosity distance, r is the SDSS r-band PSF magnitude and $E(B-V)$ the galactic extinction from Schlafly & Finkbeiner (2011). Exp Time, Airmass and Seeing are the exposure time for every source, the average airmass and the average seeing during the observations, respectively.

Name	z_{Shen}	z_{fit}	D_L [Gpc]	r [mag]	$E(B-V)$	Exp Time [s]	Airmass	Seeing ['']
J093147.37+021204.3	2.2867	2.29708 ± 0.00016	19.385	18.69	0.059	1800	1.41	1.2
J103325.92+012836.3	2.1771	2.18428 ± 0.00007	17.307	18.59	0.038	1800	1.46	0.8
J105239.38-003707.3	2.2569	2.26327 ± 0.00009	18.073	18.31	0.049	2400	1.26	1.0
J121911.23-004345.5	2.2933	2.2976 ± 0.0004	18.407	17.95	0.028	2400	1.37	1.1
J123120.55+072552.6	2.3899	2.38369 ± 0.00008	19.254	18.08	0.018	3000	1.26	1.0
J124220.07+023257.6	2.2239	2.22036 ± 0.00006	17.656	18.16	0.024	1800	1.15	1.0

1.5 was analyzed. The redshift choice ensures a simultaneous coverage from C IV to Hα with X-shooter. For this effort we selected, again from the SDSS DR7 release, QSOs with redshift around ~ 2.3 ensuring again that X-shooter would detect C IV to Hα shifted at higher wavelengths with respect to the previous sample. In order to obtain higher S/N spectra, especially in NIR where the spectra are noisier, we selected slightly brighter (r$\sim$ 17.5-18.5) QSOs observable in a single night at the VLT. The resulting sample contained eight QSOs. After selection we also checked from the SDSS spectra that the selected objects have broad emission lines suitable to BH mass estimation and that they have no obvious broad absorption features. The average broad lines FWHM of the sources in the sample is consistent with the average at these redshifts. However the average bolometric luminosity is <log L_{bol}>= 47.25, higher than the average bolometric luminosity of QSOs at this redshift, log L_{bol}=46.8 ± 0.3, but compatible within 1.5 σ. This ensures that our sample is not strongly biased.

Observations were performed in the framework of the French Guaranteed Time and took place on March the 10th 2011. For all our sources we report in Table 3.1 the properties and the characteristics of the observations.

The night was not photometric and the observing conditions were changing. Therefore during the night we monitored the spectra reduced on line and we increased or decreased the observing time of the targets depending on their quality. The night was also hampered by strong winds whose speed was near (and sometimes over) the 12 m/s limit[1] which prevents pointing towards Northern targets such as ours. These strong winds caused a loss of about two hours and a half of observing time on our program forcing us to drop two targets. The six observed targets are listed in Table 3.1 with exposure times, average airmass and seeing.

Each observation consisted of 4 different exposures of 450 sec to 750 sec each for a total of 1800 to 3000 sec. The exposures were taken using the nodding along the slit technique with an offset of 5 arcsec between exposures in a standard ABBA sequence. The slit was put at parallactic angle. Every observation was preceded by an observation of a telluric A0V

[1] http://archive.eso.org/asm/ambient-server?site=paranal

standard at similar airmass.

We processed the spectra using version 1.3.0 of the X-shooter data reduction pipeline (Goldoni et al., 2006; Modigliani et al., 2010). The pipeline performed the following actions. The raw frames were first subtracted and cosmic ray hits were detected and corrected using the method developed by van Dokkum (2001). The frames were then divided by a master flat field obtained by using day-time flat field exposures with halogen lamps.The orders were extracted and rectified in wavelength space using a wavelength solution previously obtained from calibration frames. The resulting rectified orders were then shifted and added to superpose them thus obtaining the final 2D spectrum. The orders were then merged and in the overlapping regions the merging was weighted by the errors which were being propagated during the process. From the resulting 2D merged spectrum a one dimensional spectrum was extracted at the source's position. The one dimensional spectrum with the corresponding error file and bad pixel map is the final product of the reduction.

To perform flux calibration we used different procedures for the UVB data and for the VIS-NIR data. In the UVB band we extracted a spectrum from a staring observation of the flux standard LTT3218 (Hamuy et al., 1992, 1994) taken at the beginning of the night. We then reduced the data using the same steps as above but in this case we subtracted the sky emission lines using the Kelson (2003) method. This spectrum was divided by the flux table of the same star delivered with the pipeline to produce the response function. The response was then applied to the spectrum of the sources. For the VIS and NIR arm, we used the A0V stars as flux and telluric standards. We extracted the A0V spectra with the same procedure used for the flux standard. We then used these spectra to apply telluric corrections and flux calibrations simultaneously using the Spextool software (Vacca et al., 2003). We then verified if the final spectra of the three arms were compatible in the common wavelength regions and performed a correction using the UVB spectra as reference where needed. The spectral shapes are compatible with the ones of the SDSS spectra while the fluxes are on average ~ 50 % weaker. Relying on the accurate SDSS flux calibration, we finally scaled our spectra in order to match them with SDSS spectra in common wavelenght regions reasonably free from emission lines..

3.3 Spectral Fitting

As a preliminary step we de-redshifted the sources according to their SDSS redshift spectra as reported in Shen et al. (2011) and corrected them for Galactic extinction using the E(B-V) values from Schlafly & Finkbeiner (2011) as listed in the NASA/IPAC Extragalactic Database[2] and the reddening law of Fitzpatrick (1999) with $R_V = 3.1$.

We then fitted the spectra following the same procedure described in Section 3.3 and broad lines are again fitted with the broken power law function (Eq. 2.2).

[2]The NASA/IPAC Extragalactic Database (NED) is operated by the Jet Propulsion Laboratory, California Institute of Technology, under contract with the National Aeronautics and Space Administration.

Table 3.2: Wavelengths and parameters ranges used in the fits for all the examined windows. The columns indicate: (1) the examined region or line, (2) the corresponding wavelength range on which the fit was performed, (3) nominal laboratory vacuum wavelength used as starting guess for the central wavelength of the examined line, (4-5-6) ranges of values allowed for the α, β and σ parameters (see Eq. 2.2).

region	fitting range (Å)	starting central λ (Å)	α range	β range	line broadening range (km/s)
UV	(1280 – 3500)	-	-	-	-
Optical	(4000 – 7100)	-	-	-	-
C IV	(1300 – 2000)	1549.06	10 – 500	10 – 500	40 – 10000
Mg II	(2500 – 3100)	2796.35	10 – 500	10 – 500	30 – 10000
Hβ	(4400 – 5400)	4862.683	10 – 500	10 – 500	23 – 10000
Hα	(6000 – 7000)	6564.61	10 – 500	10 – 500	22 – 10000

We want to stress how the choice of such a function allows us to reproduce with only five parameters (flux, λ_0, α, β and the σ of the Gaussian function with which the double power law is convolved) the profiles wich are commonly fitted with at least two Gaussian functions, involving 6 parameters. This is particular useful when dealing with emission line complexes, in which the use of a single component for every line helps in limiting the degeneracy in the fits. Moreover, when we do not fit lines separately, but several lines together, using a single fitting function allows us to set the same profile for all the lines with similar excitation condition (high or low ionisation). The results obtained by fitting with the function in Eq. 2.2 are consistent with those obtained by fitting with multiple Gaussians, as long as the total spectrum is well reproduced by the fit. This is particularly important, since we are making a comparison of our measurements with those of previous works that use different spectral fitting and decomposition techniques. Narrow lines are, as usual, fitted with a simple Gaussian, because their emissions are generally very well reproduced by this function. Where a blue asymmetry is present, as in the case of [O III] λ5007Å, a second Gaussian takes into account this feature.

We first obtained the slope of the continuum for fits for the entire UV (wavelength range ~1400-3500Å, containing C IV, C III] and Mg II) and for the optical window (wavelength range ~4000-7300Å, containing Hβ and Hα). We then used these slopes also in the fits for the four narrower windows, pertaining to C IV-C III], Mg II, Hβ and Hα emissions. Wavelength and parameters ranges used in the fits for all the examined windows are listed in Tab. 3.2. When necessary we applied a mask to the spectral regions contaminated by sky emissions (that was particularly required in the case of Hβ spectral window).

For the UV spectral window, we took into account several emission lines, following the prescriptions proposed in Nagao et al. (2006) : emissions are separated in two groups, high and low-ionization lines (HIL: O IV λ1402.06, N IV] λ1486.496, , C IVλ1549.06, He II λ1640.42; LIL: Si IV λ1396.76, O III] λ1663.48, Al II λ1857.40, Si IIIλ1892.03, C III] λ1908.73, Al II λ2669.95, O III λ2672.04), whose velocity profiles are known to have sistematically dif-

Table 3.3: Different combinations of the ionizing photons flux emitted by the primary source ($\Phi(H)$) and electronic density in the BLR clouds (N_e), used as input for Cloudy Fe II simulated templates for different physical condition of the BLR. The two parameters define the space of possible values for the ionization parameter U, describing the physical condition in the ionized region. For every combination we considered the case a microturbulence velocity u_{turb} = 100 km/s is present. The result is a set of 20 templates for the Fe II emissions.

$\log\left(\frac{N_e}{cm^{-3}}\right)$	$\log\left(\frac{\Phi(H)}{cm^{-2}\ s^{-1}}\right)$
8	17
10	17
10	19
12	17
12	19
12	21
14	17
14	19
14	21
14	23

ferent behaviours, so that some parameters pertaining to one group, such as the central velocity and the two power law indexes for blue and red tail, can be tied for each line. We determined these fitting parameters using C IV for the HIL group and C III] for the LIL group. The Mg II doublet was instead fitted independently of the other lines. This line should in theory be included in the LILs and therefore be tied to C III] parameters, but since Mg II is one of our investigation targets it did not make sense to tie it to another line (considering also that the C III] complex, with three emission lines present, can be degenerate). The optical window includes the Balmer lines, Hβ and Hα, and a few other lines not always present in the spectra (Hδ4103, Hγ4342, HeI λ4472, [O III] λ4960, 5007, [N II] λ6585, 6550 and He I λ7067). Hβ and Hα were fitted independently.

The fitting procedure includes Fe II emissions; they are reproduced convolving emissions templates with a Gaussian that accounts for the velocity of the emitting gas. We used two kinds of templates: the first one is the I Zw 1 Fe II template by Véron-Cetty et al. (2004), valid only for the visible band, and the second one is a series of model templates obtained with the photoionization code Cloudy (Ferland et al., 2013).

The Cloudy templates were computed with the following setup:

- we used the 371 levels Fe II model (Verner et al., 1999) instead of the simplified model (Wills et al., 1985);
- we considered a continuum emission similar to that examined in Mathews & Ferland (1987), resembling the spectrum of a typical radio-quiet AGN;
- we assumed a plane-parallel geometry with maximum cloud column density of $10^{23} cm^{-2}$;

- we used 10 combinations of ionizing photon flux and column density to consider the possible physical conditions of the BLR;
- for all models we also considered the possibility of a $100 km/s$ microturbulence velocity.

The above assumptions result in 20 different templates, which are shown in Tab. 3.3.

During the fitting procedure the templates are combined with a positive weight (free parameter of the fit) and convolved with a Gaussian function accounting for the velocity of the emitting gas (whose central value and σ are free parameters of the fit as well).

None of the examined lines exhibits an evident narrow line. We decided then to fit all of the permitted lines with a single double power law function. Also the [N II] doublet in the Hα window is not recognisable at all and we did not consider it among the emission lines in the fitting procedure. Only for Hα in J093147 the shape of the line profile reveals the presence of [N II] and, as a consequence of this, we considered the doublet in the fitting process. A special case is represented instead by J123120, for which we originally considered only broad lines, but the fit improved considerably taking into account an emission also from the NLR (see Section 3.4 and Fig. A.4). In Fig. 3.1 an example for the UV and optical windows for one of the sources is presented, along with the single windows for C IV, C III] and Mg II, Hβ and Hα lines (fits for the rest of the sources are reported in Appendix A). In the large UV window some wide regions were masked during the fitting process for the following reasons:

- the presence of strong emission blending,
- a non representativeness of the Fe II templates,
- a lack of knowlegde about what kind of emission is able to reproduce such features (this is for example the case of the red-shelf of C IV),
- the presence of noise.

The regions usually excluded from the fit are the red shelf of C IV, the bump of emission between O III] and the C III] complex, when present, and the wide spectral region between 2000 and $\sim$ 2450Å (see Nagao et al. (2006)). This choice does not affect neither the continuum slope determination nor the line analysis, since the proper examination of the lines is performed on the individual line spectral windows.

For all the sources the FWHM and the σ were estimated on the best fit profile for each line. They are connected to the shape of the line; while the FWHM is more representative of the *core*, the σ depends more on the tails of the line (Shen, 2013). Choosing one or the other leads to different results in Bh virial mass estimations, especially if we are dealing with poor quality data (Denney et al., 2016).

To give en estimate for the errors of these quantities we used a Monte Carlo approach, extracting 1000 independent values for every parameter pertaining to the line. On these synthetic profiles we evaluated 1000 values for FWHM and σ; from their distribution we were

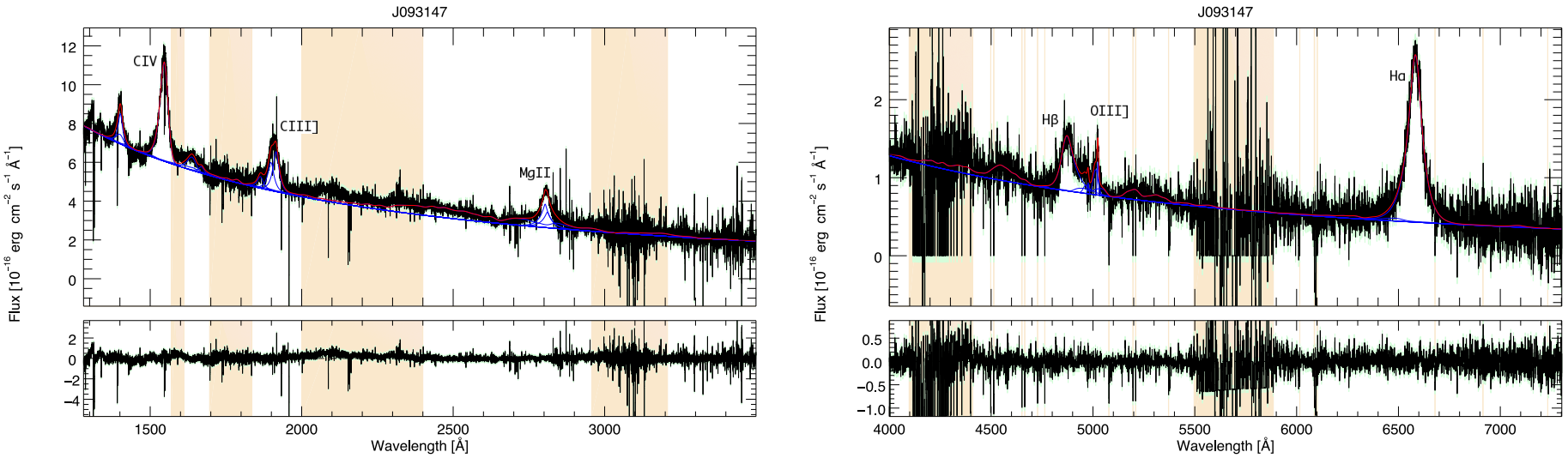


Figure 3.1: Fits for all the examined spectral windows (C IV-C III]-Mg II and Hβ-Hα large windows, C IV-C III], Mg II, Hβ and Hα small windows) for J093147. The black line is the original spectrum .The blue solid lines are the best fit models for the emissions (continuum, Fe II and emission lines) and the red solid line is the total best fit. The lower panels show residuals between best fit model and original spectrum. The colored regions are those we chose to mask. This choice can be due to the presence of strong emission blending, to a non representativeness of the Fe II templates or, in general, to a lack of knowlegde about what kind of emission is able to reproduce such features and the presence of noise.

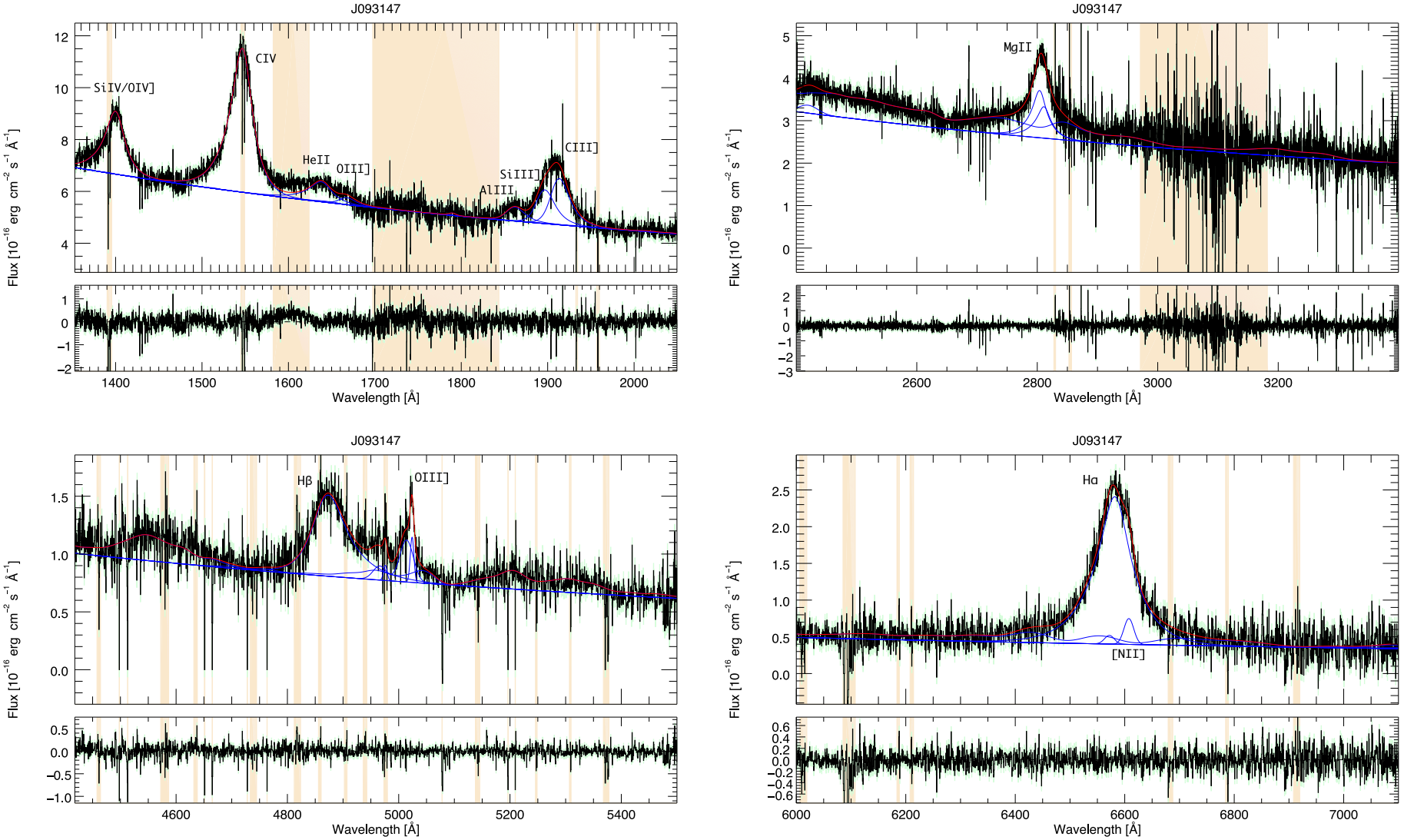


Figure 3.1 (Continued)

then able to infer the relative error to be associated with a given measurement. The errors on the slopes for UV and optical windows and on the luminosity at 1350, 1450, 3000 and 5100Å are instead computed performing, for every source, fits of various regions of the spectra and then evaluating the differences between these results and those obtained in the original fit (Pita et al., 2014). All the measured quantities are listed in Tab. 3.4.
We noticed that the central wavelength of the [OIII]λ5007Å line, from which the redshift was estimated (Shen et al., 2011), almost in every case was not close to the nominal wavelength. We then corrected the redshift using only the principal component of [OIII] (excluding the blue shifted component from the whole profile which is instead considered in the estimate of Shen et al. (2011)). The corrected redshifts are listed in table Tab. 3.1 and have been used in the following analysis.

3.4 Results

3.4.1 Line comparison

For a visual comparison we show in Fig. 3.2 the best-fit profiles for every line in all spectra. The profiles are normalized to their peak values and presented on a velocity scale. As a general trend, Hα, Hβ and Mg II behave similarly, according to what expected if the three lines are all emitted from regions in a virialized condition (McLure & Jarvis, 2002; Greene & Ho, 2005; Marziani et al., 2013). All of them show symmetric profiles and small shifts in the central wavelength, generally below 300km/s (Tab. 3.4). Surprisingly, the most asymmetric line among these (usually in the red wing) is Hβ. We suspect this is the result of a possible degeneration within the Hβ-[O III] complex, especially when He I and Fe II, whose emission are difficult to disentangle by the fitting procedure, are present. The fitting process consists in a minimization of the residuals between model and data and could therefore lead to such degeneracy. The most asymmetric line is C IV. This line frequently shows a significant shift in the central wavelength, about $-700\,\mathrm{km\ s^{-1}}$ on average, and all cases show the presence of a prominent blueshift. We notice that, in contrast with what found in Paper I, C III] does not seem to behave so differently with respect to the other lines. All sources do not present a large shift ($\sim$150 km s^{-1} on average), with the exception of J123120, for which C III] has a very large positive shift of the centroid. Furthermore, in all sources C III] does not show a prominent blueshift as does C IV instead.

3.4.2 Line widths comparison

Fig. 3.3 shows a comparison of the FWHM for every pair of lines commonly used for virial estimates, Hβ, Mg II, Hα and C IV.. The red points represent the measurements used in the final analysis.

For J123120 we decided to do two fits, one using only the broad components (black points) and one including also the narrow ones. The latter fit was done in two ways (magenta

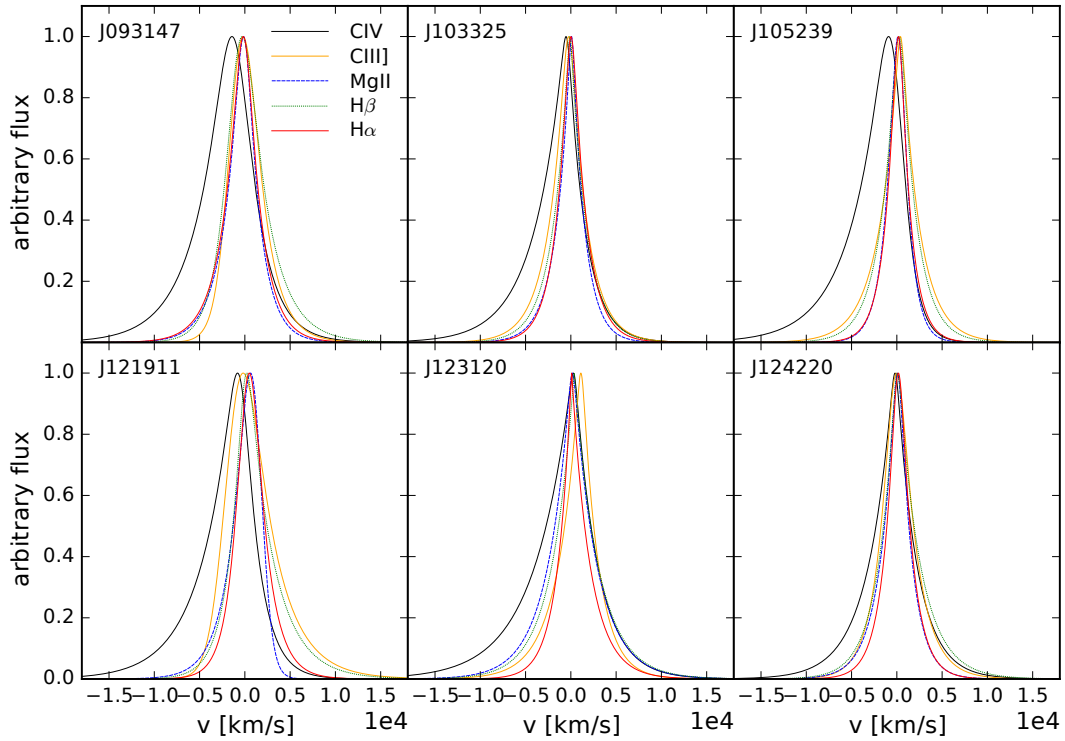


Figure 3.2: Comparison of all the examined line profiles. The profiles are normalized to their peak values and referred to their nominal wavelength on a velocity scale.

Table 3.4: Measured quantities for the examined spectral windows.

	J093147	J103325	J105239	J121911	J123120	J124220
UV slope	-1.40 ± 0.12	-1.40 ± 0.17	-1.71 ± 0.06	-1.87 ± 0.05	-1.15 ± 0.10	-1.47 ± 0.18
Optical slope	-2.19 ± 0.16	-2.15 ± 0.25	-2.38 ± 0.25	-2.44 ± 0.15	-1.53 ± 0.11	-1.76 ± 0.17
$\log(\lambda L_\lambda)(1350\text{Å})$	46.62 ± 0.04	46.61 ± 0.06	46.70 ± 0.02	46.90 ± 0.01	46.71 ± 0.01	46.73 ± 0.03
$\log(\lambda L_\lambda)(1450\text{Å})$	46.62 ± 0.04	46.59 ± 0.06	46.69 ± 0.01	46.87 ± 0.01	46.71 ± 0.01	46.71 ± 0.03
$\log(\lambda L_\lambda)(3000\text{Å})$	46.50 ± 0.01	46.46 ± 0.01	46.46 ± 0.02	46.59 ± 0.02	46.66 ± 0.02	46.56 ± 0.04
$\log(\lambda L_\lambda)(5100\text{Å})$	46.22 ± 0.02	46.21 ± 0.04	46.14 ± 0.04	46.25 ± 0.02	46.70 ± 0.01	46.26 ± 0.01
$v_{\text{C IV}}$	-1487 ± 26	-652 ± 14	-1052 ± 17	-925 ± 15	125 ± 22	-247 ± 9
$\text{FWHM}_{\text{C IV}}$	5790 ± 81	4034 ± 50	5050 ± 54	4891 ± 49	4934 ± 58	3860 ± 27
$\sigma_{\text{C IV}}$	4127 ± 68	3792 ± 30	3910 ± 57	3997 ± 46	4786 ± 61	3566 ± 13
$v_{\text{C III]}}$	-157 ± 91	-255 ± 30	377 ± 23	-83 ± 118	1057 ± 120	-11 ± 19
$\text{FWHM}_{\text{C III]}}$	4040 ± 161	3227 ± 102	3251 ± 71	5418 ± 253	3192 ± 271	3580 ± 52
$\sigma_{\text{C III]}}$	2330 ± 146	2707 ± 71	2968 ± 72	3363 ± 299	2937 ± 156	2558 ± 46
$v_{\text{Mg II}}$	-167 ± 31	54 ± 24	123 ± 18	567 ± 34	165 ± 19	53 ± 15
$\text{FWHM}_{\text{Mg II}}$	3184 ± 142	2294 ± 126	2326 ± 79	3345 ± 82	3891 ± 89	2680 ± 71
$\sigma_{\text{Mg II}}$	2466 ± 79	2120 ± 50	1592 ± 45	2111 ± 62	3641 ± 96	2272 ± 37
v_{H_β}	-239 ± 60	-109 ± 33	242 ± 28	290 ± 35	388 ± 31	101 ± 23
FWHM_{H_β}	4483 ± 174	2672 ± 158	3093 ± 107	3771 ± 127	3574 ± 84	3378 ± 111
σ_{H_β}	3040 ± 198	2542 ± 121	2535 ± 94	3261 ± 122	3604 ± 83	3087 ± 73
v_{H_α}	-185 ± 61	21 ± 11	168 ± 9	509 ± 18	139 ± 13	172 ± 9
FWHM_{H_α}	3411 ± 111	2638 ± 42	2246 ± 30	3346 ± 64	2497 ± 175	2537 ± 32
σ_{H_α}	2608 ± 161	2168 ± 34	1769 ± 25	2167 ± 91	2555 ± 183	1982 ± 26

and grey points): the first one leaving the width of the narrow component as a free parameter and the other one fixing it to the one of C III] in the UV range and the one of [O III] and Hβ in the optical range. The magenta points were adopted for the following analysis and corresponds to the case in which we tie the narrow component in the UV spectrum and to the case in which we left it free in the optical range. While in the UV we could use the obvious technique of linking together the widths of the narrow components, we could not do the same in the optical becuase the narrow Hα is much broader than the [O III] one and they cannot be reasonably linked together. The presence of a narrow component in J123120 is particularly evident for the Mg II line, although it improved the fits also for the other lines.

Even for our small sample a correlation between FWHM of Hβ, Hα and Mg II is present. Instead, C IV has a less strong correlation with the other lines: this is not surprising given the blueshift of the line in almost every source of the sample. We report the results of the linear fit assuming a linear relation between the logarithms of the linewidths in Tab. 3.5.

Concerning the parametrization of the line width, although the σ is in principle a more reliable estimator, especially when data quality is poor (Collin et al., 2006; Denney et al., 2016), most of the recent works use instead the FWHM (Ho et al. (2012); Trakhtenbrot & Netzer (2012); Ho & Kim (2015) among the others), justifying it with a smaller scatter between different lines (Mejia-Restrepo et al. (2016) for example).

In our measurements we do not observe such a larger scatter in the relationships involving the line dispersion σ with respect to the FWHM (see Fig. 3.4 and Tab. 3.5). The only exceptions are the relationships involving MgII, for which essentially the J123120 point is an outlier. Of course the smallness of our sample plays an important role in this respect, stressing the presence of outliers that would not probably be such in a larger sample. The analysis presented by Mejia-Restrepo et al. (2016) highlights the difference between measurements performed under a *global* (considering a continuum fitted on the whole SED of accretion disk, BLR and NLR emissions) and *local* approach (the more common case, in which the fit on the line is performed only on a smaller spectral window including the line). The line measurements they report are those obtained under the local approach and for which they recognize the presence of a large scatter for the line dispersion, ascribable to the subtraction of a non proper fitted continuum. We notice, however, that their local approach take into account rather narrow spectral windows, while our measurements are performed on wider wavelength ranges and, moreover, take into account a preliminar continuum evaluation, performed on even wider windows (see Tab. 3.2). Given the data quality and the spectral range our fits cover, we are confident that our line dispersions could be considered for a virial estimate. Nonetheless, since we are especially interested in the comparison with some of the works mentioned before and our data quality allows the use of the FWHM, we will focus our analysis on this quantity.

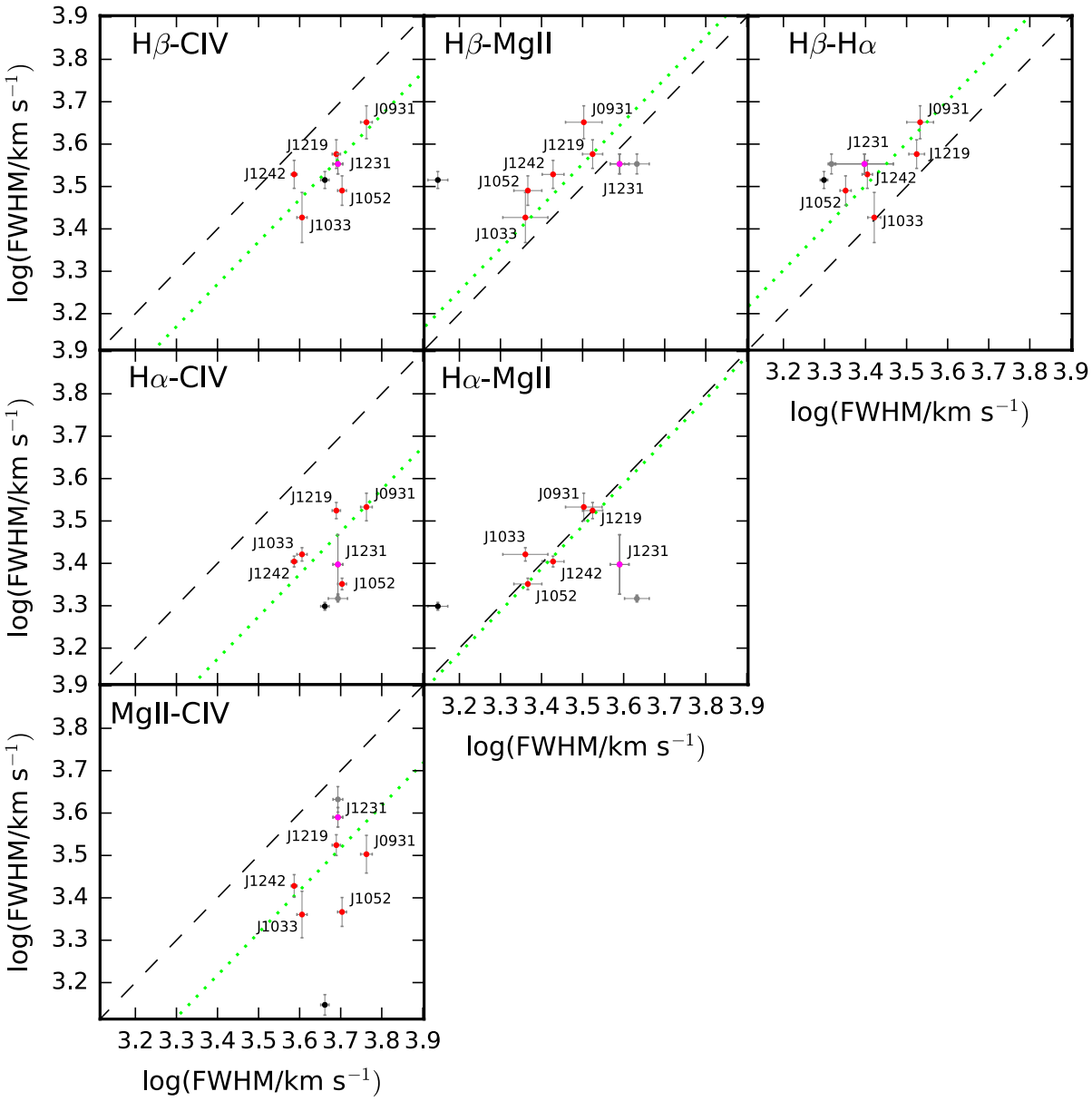


Figure 3.3: FWHM correlations between line pairs. The red and magenta points are the measurements used in the final analysis. The black point represents the measurement taking into account only broad components for J123120, while the magenta point is the measurement including also narrow components in the fit. The grey point is an alternative measurement that still considers narrow emissions: in the case of Mg II it gives the FWHM of the line for which the width of narrow component was left as a free parameter of the fit (for the magenta point instead it was tied to that of C III]), while in the case of Hα it gives the FWHM of the line for which the width of the narrow component was tied to that of [O III] and Hβ narrow component (while for the magenta point was instead left free). The black dashed line represents the 1:1 relation, while the light green one represents the best-fit to our data (red and magenta points), considering a linear relation between the logarithms of the FWHM for every pair of lines. The function we fit is $\log(FWHM_1) = m \cdot \log(FWHM_2) + c$, where the slope $m = 1$. Intercept and scatter for all the relations are reported in Tab. 3.5.

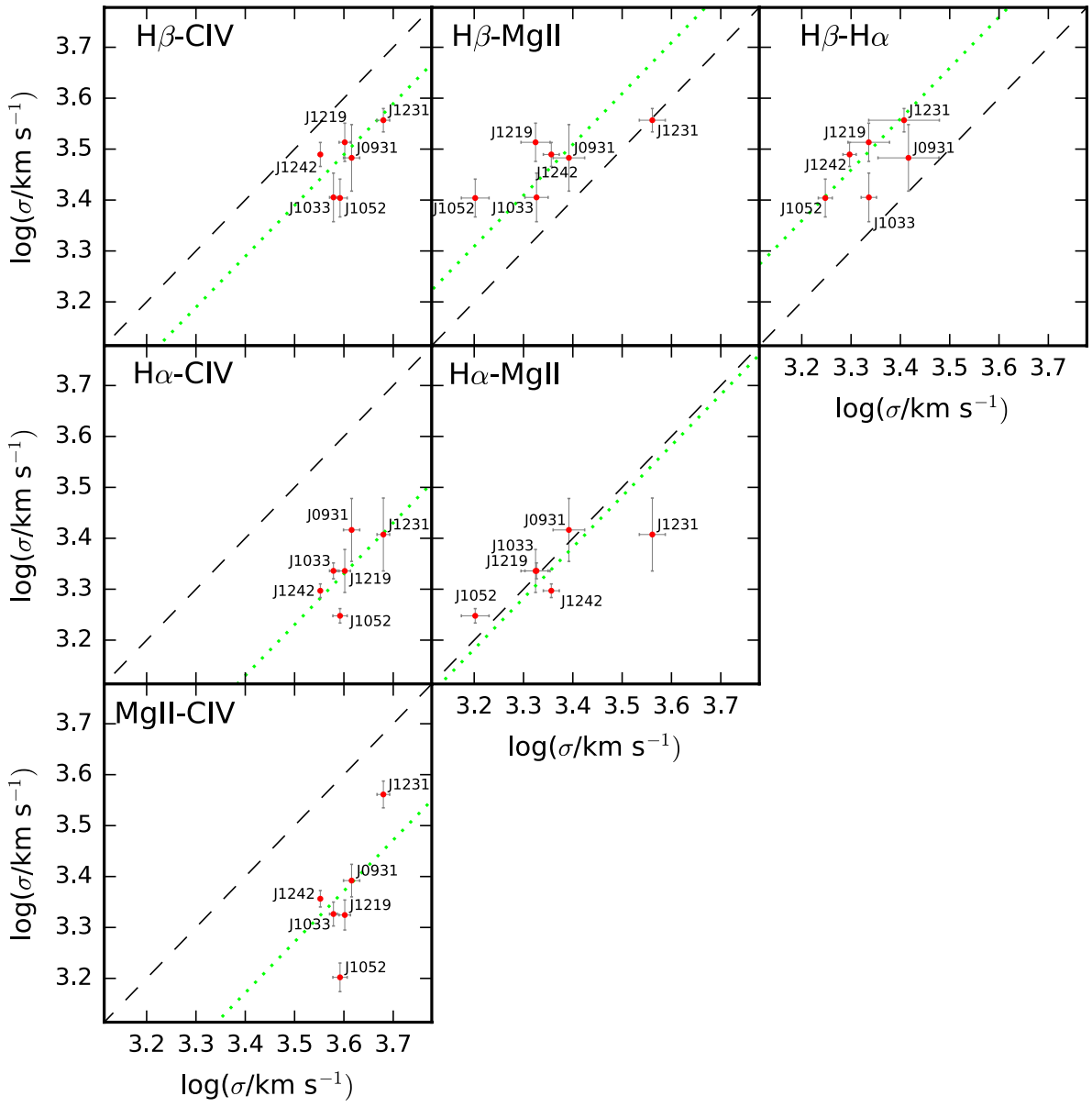


Figure 3.4: σ correlations between line pairs. The black dashed line represents the 1:1 relation, while the light green one represents the best-fit for our data (red points), considering a linear relation between the logarithms of the line dispersion for every pair of lines. The function we fit is $\log(\sigma_1) = m \cdot \log(\sigma_2) + c$, where the slope $m = 1$. Intercept and scatter for all the relations are reported in Tab. 3.5.

Table 3.5: Linewidth correlations between line pairs. For both FWHM and line dispersion σ we list the results for the linear fit between the logarithms: assuming a linear relation between the quantities, $\log(linewidth_1) = m \cdot \log(linewidth_2) + c$, we find the intercept c and the scatter Δ considering a fixed slope $m = 1$.

	log(FWHM)		log(σ)	
	c	Δ	c	Δ
Hβ-C IV	−0.131	0.048	−0.111	0.045
Hβ-Mg II	0.053	0.068	0.109	0.073
Hβ-Hα	0.103	0.045	0.159	0.044
Hα-C IV	−0.226	0.074	−0.270	0.039
Hα-Mg II	−0.012	0.046	−0.018	0.049
Mg II-C IV	−0.182	0.076	0.228	0.072
Hβ-C III]	−0.033	0.059	0.035	0.067
Hα-C III]	0.142	0.034	−0.134	0.063
Mg II-C III]	−0.131	0.057	−0.094	0.102
C IV-C III]	0.083	0.074	0.146	0.029

3.4.3 M_{BH} and Eddington ratios

Although we took into account several previous works (Vestergaard & Peterson, 2006; Bentz et al., 2013; Wang et al., 2009; Vestergaard & Osmer, 2009; Ho & Kim, 2015; Jun et al., 2015), we focus our analysis on the comparison of our data with the only two other samples with the same characteristics, i.e. whose spectra were taken with the X-shooter spectrograph and therefore cover a spectral range including all the broad lines of interest, Paper I and Mejia-Restrepo et al. (2016). We notice that, while these two samples have a redshift range $z \sim$ 1.4-1.6, ours goes to higher redshift $z \sim 2.2$ and therefore can be interesting to make a comparison in terms of mass and Eddington ratio. Since the quasars in our sample are selected to be slightly brighter than those selected in Paper I, we expect them to be characterised by higher values for at least one of these two quantities.

In Fig. 3.5 we report the measurements of M_{BH} obtained with the new prescriptions of Mejia-Restrepo et al. (2016) for their sample, our sample and for the Paper I sample, for which only the measurements pertaining to three lines out of four are present (in this work results for Hβ are not included, given the poor signal to noise in this spectral range, and all the comparisons are made with Hα). In Tab. 3.6 we report mass estimations with Mejia-Restrepo et al. (2016) prescriptions for Ho et al. (2012) and for our sources. In the same table we also report mass estimations for the Mejia-Restrepo et al. (2016) sample obtained following their Hα prescription with the luminosity at 5100Å, the same we used for our sample and for Paper I sample. For all the lines we used the third column of Mejia-Restrepo et al. (2016) Tab.7, i.e. *local approach M_{BH} calibrations, but corrected for the small systematics with respect to the global approach M_{BH} calibrations*. The C IV-based M_{BH} for the Paper I sources are computed using L_{1350} instead of L_{1450}, because this is the closest continuum luminosity available for this sample. Our sample (red data points) fits very well in all the cases and, on average, is located

in the upper part of the global distribution.

We then compute the Eddington ratios for our sources, Mejia-Restrepo et al. (2016) and Paper I samples with the same prescription used in Paper I (bolometric luminosity from McLure & Dunlop (2004)[3] and Eddinton luminosity $L_{Edd} = 1.26\ 10^{38}(M_{BH}/M_{\odot})$ erg s^{-1}) to verify if our sample is composed by higher accreting black holes. We evaluate the Eddington ratio for our sample both with Hβ and with Hα as virial estimators, while for Mejia-Restrepo et al. (2016) we re-compute the Hα, L_{5100} based values (Tab.3.6). In this way we can compare these values with those found in Paper I, for which only Hα measurements are available. We find indeed that our Eddington ratios (Tab. 3.6, Fig. 3.6) are much higher on average ($<L/L_{Edd}>= 0.719$) than those of both Paper I ($<L/L_{Edd}>= 0.252$) and Mejia-Restrepo et al. (2016) ($<L/L_{Edd}>= 0.130$)).

The higher brightness of our sample is therefore due both to the presence of more massive BHs and to the fact that they are accreting, on average, at higher rates.

3.4.4 Can C III] be used as a virial estimator?

Unlike what found in Paper I, when looking at the C III] profile we do not recognize a different behaviour of the line with respect to the others (see Fig. 3.2). We then decide to examine the relationships between the FWHM of C III] with those of the other lines. Although C III] is not commonly used, some works analyse this line (Greene et al., 2010; Shen & Liu, 2012). Greene et al. (2010) find only a slight correlation of C III] with Mg II FWHM, while Shen & Liu (2012) state that C III] linewidths correlates with C IV and therefore these lines could be emitted by the same region, then being characterized by the same issues (i.e. non virialization of the emitting region). We find this correlation (log(FWHM$_{CIV}$)-log(FWHM$_{CIII]}$)) has the largest scatter (see Tab. 3.5, Fig.3.7 and Fig.3.8) among all the relationships. However, the sample of Shen & Liu (2012) has much higher luminosity than ours and is composed by lower redshift sources. Moreover, they fit the line profiles on much narrower wavelength ranges than ours. Additionally, for the line profile model they use two Gaussians tied to give a symmetric broad component for C III]. All these differences could then contribute to the discrepancy with their results.

In Fig. 3.7 we notice the presence of only one outlier, J121911, for which an evident lack of emission is present in the C III] complex (see the figures in Appendix A for the results on the complete sample). We have checked that including or excluding this point does not affect the fit and we decide to leave it in the sample. The reason for this is that this point has larger errors, since the IDL routine we use to fit the linear relation (MPFITEXY, Williams et al. (2010), based on the MPFIT package (Markwardt, 2009)) considers the errors on both x and y variables. J121911 does not stand out evidently in the case of the line dispersion (Fig. 3.8), but for the same reason we do not consider it as a reliable measurement. Results of fits assuming a linear correlation between the quantities are listed in Tab. 3.5.

[3] $L_{bol} = 9.8\lambda L_{\lambda}(5100\text{Å})$

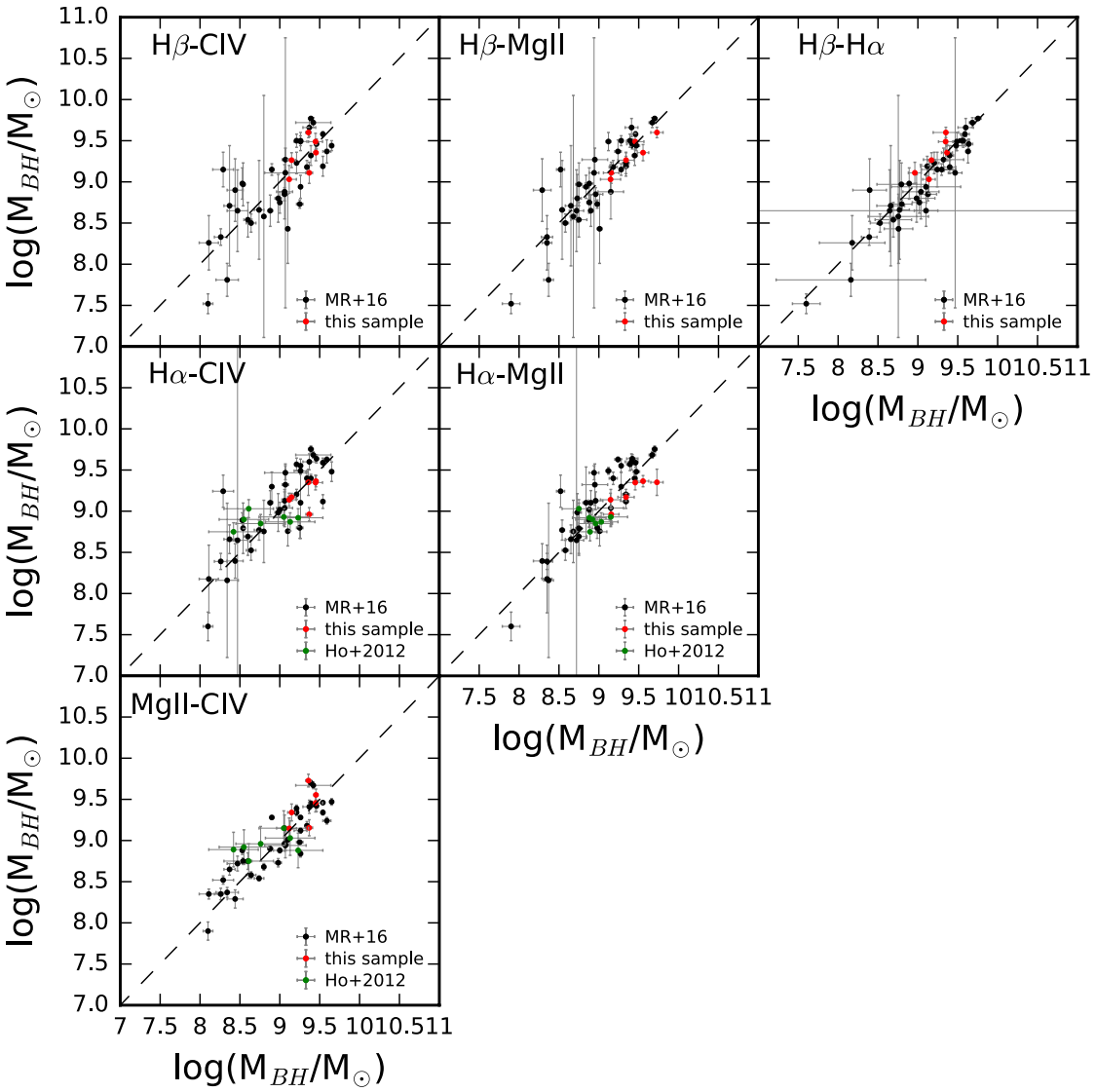


Figure 3.5: M_{BH} measurements obtained with Mejia-Restrepo et al. (2016) prescriptions for Mejia-Restrepo et al. (2016) (black points), Paper I (green points) and this work (red points) samples. The prescriptions we use are the *local approach MBH calibrations, but corrected for the small systematic with respect to the global approach MBH calibrations* that Mejia-Restrepo et al. (2016) present in the third column of their Tab.7. For Paper I only C IV, MgII and Hα are available. In the case of C IV, for Paper I sources, we had to use L_{1350} in place of L_{1450}, because this is the closest specific luminosity available for this sample. All the values, along with $M_{BH}(H_\alpha)$, $L_{5100\text{Å}}$ for Mejia-Restrepo et al. (2016) sample, are given in Tab. 3.6.

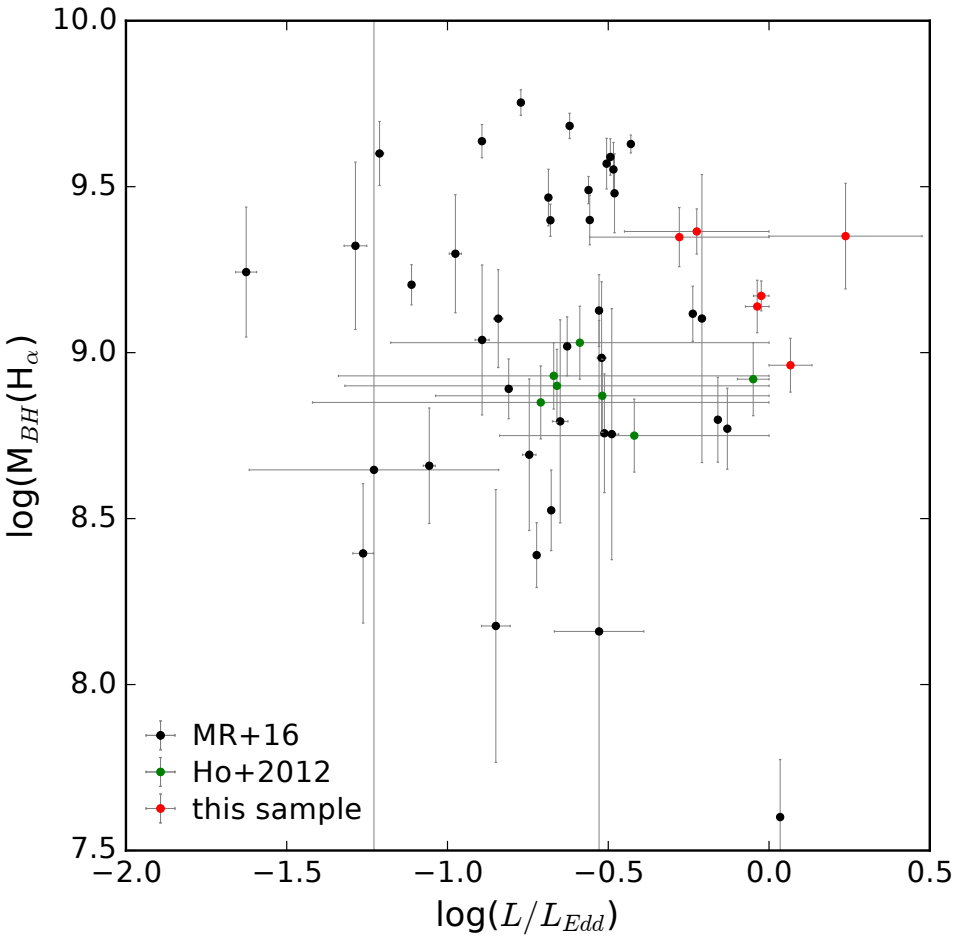


Figure 3.6: M_{BH} vs Eddington Ratios for Mejia-Restrepo et al. (2016) (black points), Paper I (green points) and this work (red points) sources. For the Eddington Ratios, we follow McLure & Dunlop (2004) for the bolometric luminosity and use $L_{Edd} = 1.26\ 10^{38}(M_{BH}/M_\odot) erg\ s^{-1}$ for the Eddington luminosity. We use M_{BH} from Hα, the only Balmer line available for all the samples. All the values are listed in Tab. 3.6.

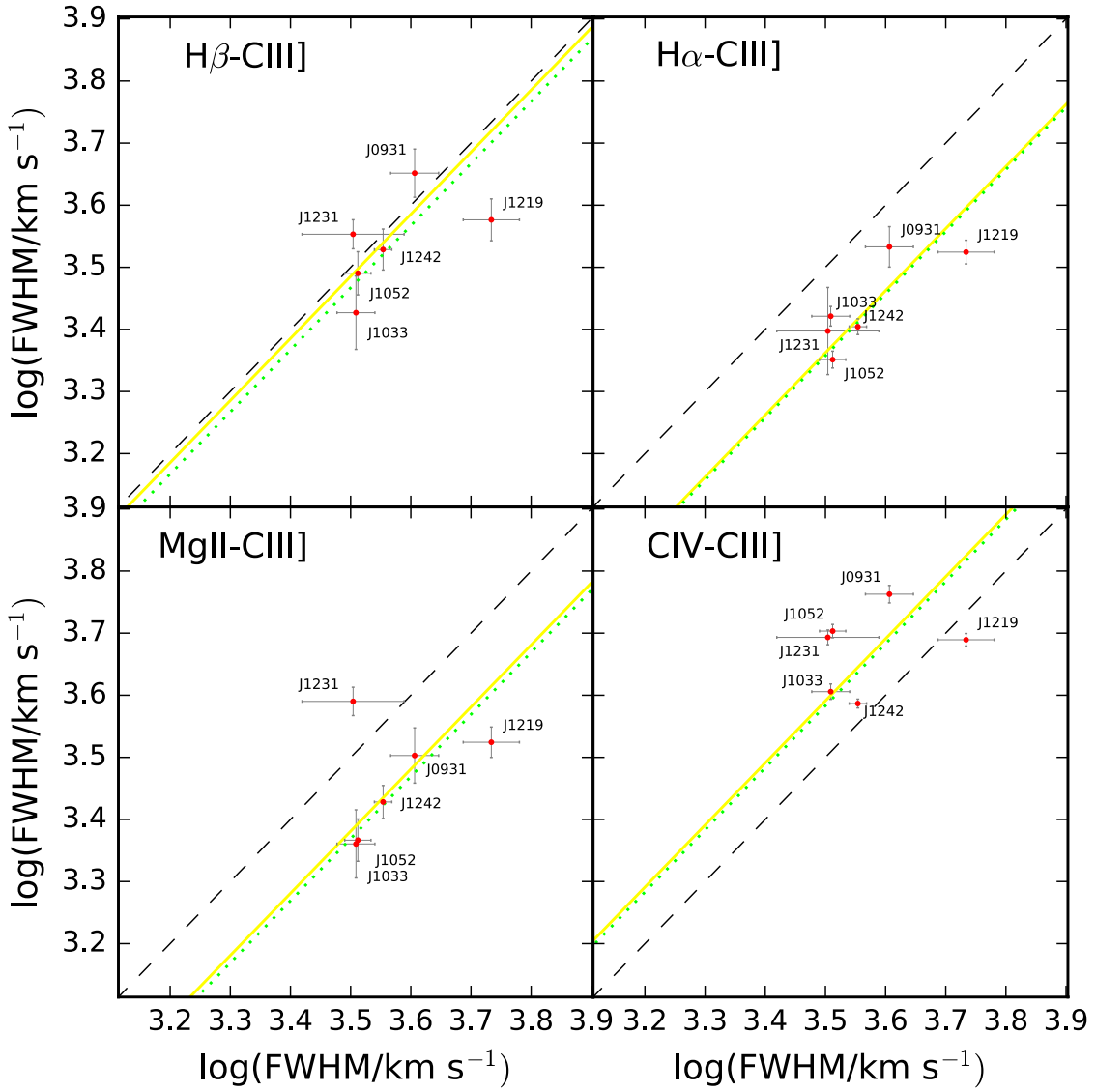


Figure 3.7: FWHM comparison for every pair of lines for all the sources in our sample (red data points). The black dashed line represents the 1:1 relation, while the light green one represents the best-fit for our data, considering a linear relation between the FWHMs for every pair of lines. The function we fit is then $FWHM_1 = m \cdot FWHM_2 + c$, where the slope $m = 1$. The solid yellow line is the best fit when not considering J121911 in the fit. Intercept and scatter for all the relations are reported in Tab. 3.5.

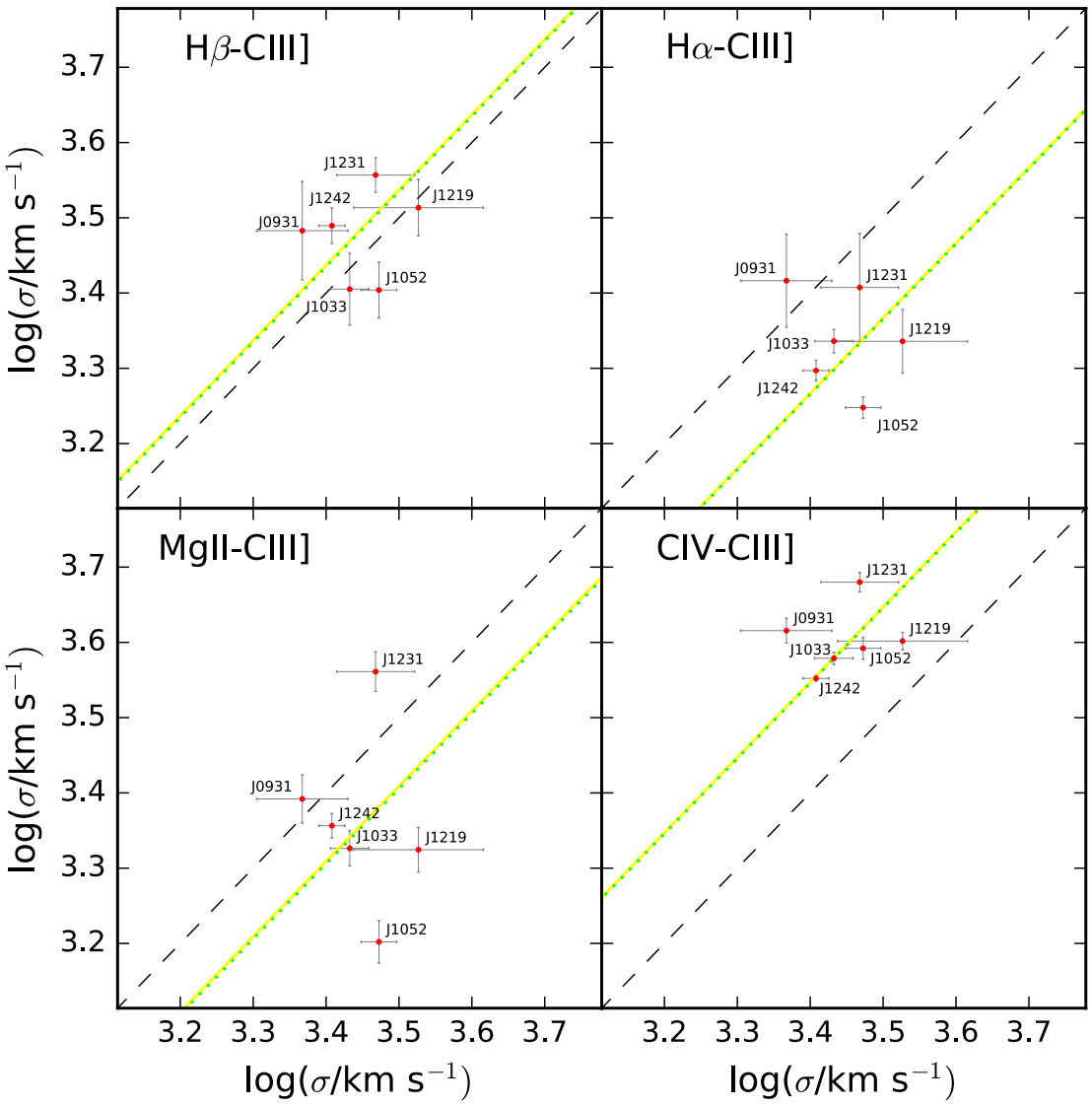


Figure 3.8: Line dispersion comparison for every pair of lines for all the sources in our sample (red data points). The black dashed line represents the 1:1 relation, while the light green one represents the best-fit for our data, considering a linear relation between the line dispersions for every pair of lines. The function we fit is then $\sigma_1 = m \cdot \sigma_2 + c$, where the slope $m = 1$. The solid yellow line is the best fit when not considering J121911 in the sample. Intercept and scatter for all the relations are reported in Tab. 3.5.

Table 3.6: Mass estimates with Mejia-Restrepo et al. (2016) prescriptions for Paper I and this work samples. For Paper I, for which Hβ is not available, Eddington Ratios are computed only with M_{BH}(Hα). To have Eddington Ratios that we can compare with Paper I, we compute $M_{BH}(H_{alpha})$,$L_{5100\text{Å}}$ also for Mejia-Restrepo et al. (2016) sample. Errors on luminosity and FWHM, described in Sec. 3.3, were propagated to give an estimate of the errors on the quantities listed in this table.

Paper I	J0856-0013	J1029+0813	J1050-0105	J1229-0307	J1238-0056	J1242+0249	J1323+0154
$\log(M_{CIV}/M_\odot)$	8.76 ± 0.31	8.42 ± 0.31	8.61 ± 0.31	9.23 ± 0.31	9.13 ± 0.31	9.05 ± 0.31	8.55 ± 0.31
$\log(M_{MgII}/M_\odot)$	8.96 ± 0.21	8.89 ± 0.21	8.75 ± 0.21	8.88 ± 0.21	9.03 ± 0.21	9.15 ± 0.21	8.92 ± 0.21
$\log(M_{H\alpha}/M_\odot)$	8.85 ± 0.11	8.75 ± 0.11	9.03 ± 0.11	8.92 ± 0.11	8.87 ± 0.11	8.93 ± 0.10	8.90 ± 0.11
R_{Edd} (Hα based)	0.195 ± 0.003	0.381 ± 0.005	0.258 ± 0.003	0.893 ± 0.011	0.303 ± 0.004	0.214 ± 0.003	0.219 ± 0.003
this work	J093147	J103325	J105239	J121911	J123120	J124220	
$\log(M_{CIV}/M_\odot)$	9.45 ± 0.09	9.12 ± 0.09	9.37 ± 0.04	9.45 ± 0.03	9.36 ± 0.03	9.15 ± 0.05	
$\log(M_{MgII}/M_\odot)$	9.46 ± 0.11	9.15 ± 0.13	9.16 ± 0.10	9.55 ± 0.07	9.73 ± 0.08	9.34 ± 0.10	
$\log(M_{H\beta}/M_\odot)$	9.49 ± 0.10	9.03 ± 0.17	9.11 ± 0.13	9.36 ± 0.10	9.60 ± 0.06	9.26 ± 0.09	
$\log(M_{H\alpha}/M_\odot)$	9.35 ± 0.09	9.14 ± 0.08	8.96 ± 0.08	9.37 ± 0.07	9.35 ± 0.16	9.17 ± 0.04	
R_{Edd} (Hβ based)	0.380 ± 0.019	1.178 ± 0.112	0.830 ± 0.086	0.609 ± 0.035	0.975 ± 0.031	0.763 ± 0.035	
R_{Edd} (Hα based)	0.526 ± 0.026	0.919 ± 0.078	1.166 ± 0.115	0.596 ± 0.032	1.730 ± 0.073	0.946 ± 0.039	
M-R (2016)	$\log(M_{Ha}/M_\odot)$	R_{Edd} (Hα based)		$\log(M_{Ha}/M_\odot)$	R_{Edd} (Hα based)		
J1152+0702	9.48 ± 0.12	0.331 ± 0.004	J0213-1003	9.13 ± 0.11	0.296 ± 0.003		
J0155-1023	9.59 ± 0.05	0.321 ± 0.002	J0341-0037	8.75 ± 0.38	0.324 ± 0.014		
J0303+0027	9.75 ± 0.04	0.1691 ± 0.0007	J0143-0056	8.79 ± 0.31	0.224 ± 0.008		
J1158-0322	9.57 ± 0.08	0.313 ± 0.003	J0927+0004	9.20 ± 0.06	0.0773 ± 0.0005		
J0043+0114	9.12 ± 0.08	0.579 ± 0.005	J0213-0036	8.89 ± 0.09	0.155 ± 0.002		
J0209-0947	9.55 ± 0.08	0.328 ± 0.003	J1050+0207	9.04 ± 0.23	0.128 ± 0.003		
J0842+0151	9.40 ± 0.05	0.2090 ± 0.0010	J0948+0137	8.69 ± 0.23	0.180 ± 0.005		
J1002+0331	9.49 ± 0.04	0.275 ± 0.001	J0042+0008	8.52 ± 0.12	0.210 ± 0.003		
J0323-0029	9.10 ± 0.43	0.618 ± 0.030	J1013+0245	9.32 ± 0.25	0.0517 ± 0.0014		
J0152-0839	9.40 ± 0.07	0.277 ± 0.002	J1021-0027	9.24 ± 0.20	0.0237 ± 0.0005		
J0941+0443	9.64 ± 0.05	0.1280 ± 0.0006	J0038-0019	8.39 ± 0.10	0.189 ± 0.002		
J0148+0003	9.68 ± 0.04	0.2398 ± 0.0009	J0912-0040	8.66 ± 0.17	0.0878 ± 0.0015		
J0934+0005	8.80 ± 0.13	0.693 ± 0.010	J1048-0019	8.16 ± 0.94	0.30 ± 0.08		
J0019-1053	9.30 ± 0.18	0.106 ± 0.002	J1045-0047	8.40 ± 0.21	0.055 ± 0.001		
J0850+0022	8.98 ± 0.23	0.302 ± 0.008	J0042-0011	7.60 ± 0.17	1.08 ± 0.02		
J0404-0446	8.77 ± 0.12	0.742 ± 0.010	J1046+0025	8.18 ± 0.41	0.141 ± 0.007		
J1052+0236	9.60 ± 0.10	0.0615 ± 0.0006	J0930-0018	8.65 ± 3.05	0.06 ± 0.02		
J0223-0007	9.10 ± 0.15	0.144 ± 0.002	J1108+0141	9.63 ± 0.03	0.372 ± 0.001		
J0240-0758	9.02 ± 0.09	0.236 ± 0.002	J1005+0245	9.47 ± 0.09	0.206 ± 0.002		
J0136-0015	8.76 ± 0.18	0.308 ± 0.006					

C III] is not usually mentioned among the feasible virial estimators. This is mostly due to (1) its small intensity if compared with other lines in the same the spectral range (C IV is by far more intense) and (2) to the blend with other emission lines in the same complex. The use of only one component to fit the broad lines, instead of two or more Gaussians, is more robust against the degeneracy in the profile fitting, also thanks to the good quality of our data. The C IV line should not be used in virial estimates because, although it is more intense, it is contaminated by non-virial components. On the contrary, the preliminar line comparison (Fig. 3.2) shows that C III] behaves very similarly to the lines that are mostly virialized (Hβ, Hα and Mg II). Therefore, we attempt to find a virial relationship for C III] comparing measurements for this line with virial masses based on the other lines.

Given the smallness of our sample, in addition to a fixed dependence of the M_{BH} on the velocity of the emitting gas according to the virial assumption, we also fix the dependence on the luminosity as $M_{BH} \propto L^{0.5}$. This choice is perfectly consistent with the hypothesis of photoionization in the BLR and with what found in previous works (Bentz et al., 2013). The BH mass is then given by the equation

$$M_{BH} = C \ \mathrm{FWHM}^2 L^{0.5} \ , \tag{3.1}$$

where the only free parameter is the scaling factor C. We chose to use the luminosity at 1450Å, as the closest to the C III] line that we can measure in a continuum window reasonably free by other emissions. In Tab. 3.7 we report our results (scaling factor C and scatter Δ) for the comparison of the C III]-based M_{BH} with all the other lines except C IV (Fig. 3.9). Since C IV, if not corrected, does not share the property of virialization of the emitting region, we do not consider this line for this comparison. Tab. 3.7 also shows M_{BH} estimates for all the sources derived from the C III] line. The scatter in these relations is comparable with those of the mass relations involving C IV (0.17, 0.19 and 0.16 dex with the only six objects in our sample and 0.26, 0.22 and 0.20 dex for the whole sample, for Hβ, Hα and Mg II, respectively). However, the strong similarity of the C III] profile with those of the lines emitted by virialized gas (Hβ and Mg II) suggests that, at least for this sample, we can use this line as a virial estimator. The use of the double power law function as a model to fit broad components helps in removing the degeneracy in the Al III, Si III] and C III] complex and in retrieving the C III] profile more accurately. Despite the scatter in the C III] virial relationships is as large as that of C IV, this line does not seem to be affected by contamination by not-reverberating components. Since we are comparing our C III]-based masses with M_{BH} obtained using the Mejia-Restrepo et al. (2016) prescription and since they do not provide a virial relationships for σ, we limit our analysis to the FWHM.

3.5 Summary

We examined a sample of six quasars at redshift $z \sim 2.2$, whose spectra were taken with the X-shooter spectrograph. This instrument covers a very large spectral range, allowing the

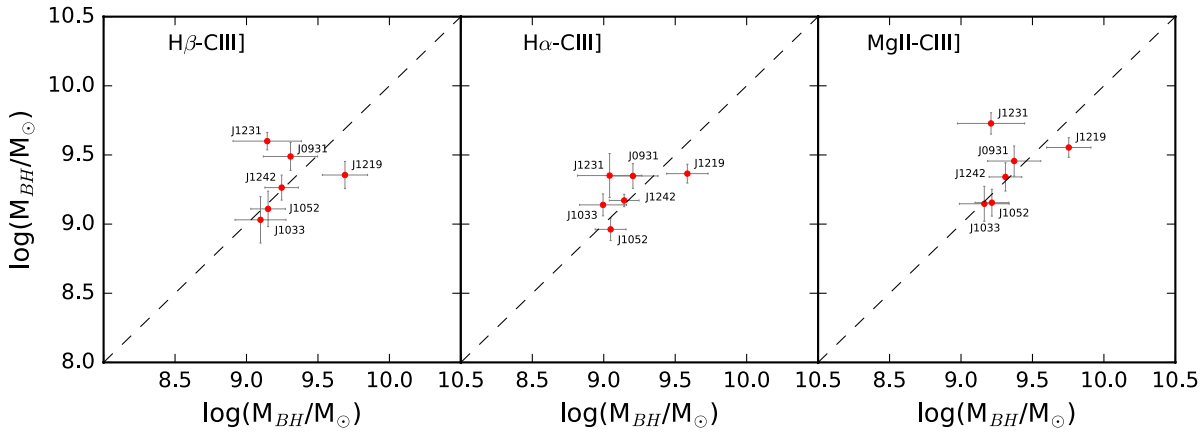


Figure 3.9: Correlations of C III]-based virial masses with those derived from the other lines. The black dashed line represents the 1:1 relation. Intercept and scatter for these relations are given in Tab. 3.7.

Table 3.7: Correlations of C III]-based virial masses with the other lines. We fit our data points to the relation $\log(M_{BH_{line}}) = C + 2\log(FWHM_{CIII]}) + 0.5\log(L_{1450\text{Å}})$ for every line pair and find the only free parameter C and the scatter for the relations, given in the first two columns of the table. We also report virial estimations for all the sources with these prescriptions.

	whole relation		$\log(M_{BH}/M_\odot)$ individual sources					
	C (km/s)	Δ (dex)	J093147	J103325	J105239	J121911	J123120	J124220
C III]-Hβ	6.78 ± 0.06	0.21	9.31 ± 0.19	9.10 ± 0.18	9.15 ± 0.12	9.69 ± 0.16	9.14 ± 0.24	9.25 ± 0.12
C III]-Hα	6.68 ± 0.05	0.13	9.20 ± 0.18	8.99 ± 0.16	9.05 ± 0.11	9.59 ± 0.14	9.04 ± 0.22	9.14 ± 0.10
C III]-Mg II	6.85 ± 0.06	0.18	9.37 ± 0.19	9.16 ± 0.17	9.22 ± 0.12	9.75 ± 0.16	9.21 ± 0.24	9.31 ± 0.11

simultaneous comparison of all the most used virial estimators. Our analysis has the merit to use a single profile model to fit every broad emission line, therefore reducing the number of free parameters and the possible degeneracy when in presence of multiple emissions. Moreover we use wider spectral windows to properly constrain the continuum before examining the lines individually. We compare our results with those of the only two other samples at high redshift whose spectra were taken with X-shooter and therefore have the same simultaneity in the line detection (Paper I and Mejia-Restrepo et al. (2016)). The analysis gives the following results:

1. The comparison of the line profiles shows that Hβ, Hα and Mg II behave in a similar way, as expected for virialized gas.

2. C IV is by far the line that most deviates from this condition because of its strong blueshifts and asymmetry.

3. We find C III] to behave consistently with the other lines, in contrast to C IV.

4. Comparison of the linewidths obtained for every line give a similar scatter for the FWHM and the line dispersion σ. However, we chose to focus our analysis on FWHM, to be consistent with the works we are comparing our sample to.

5. We compute virial masses for our sample and for the sources in Paper I, using the prescription by Mejia-Restrepo et al. (2016). All the sources follow the relations. A comparison with Paper I and Mejia-Restrepo et al. (2016) shows that our higher redshift sample has larger M_{BH} and higher accreting rates.

6. Notwithstanding the smallness of our sample, we suggest a new virial mass prescription based on the FWHM of C III], which can be considered as a valid substitute of C IV for sources in which only this spectral window is available. Unlike C IV infact, this line seems to share the behaviour of the lines emitted by virialized gas and despite of its weakness if compared to C IV, it should be used if C IV is not properly corrected in order to remove the non-virial contamination.

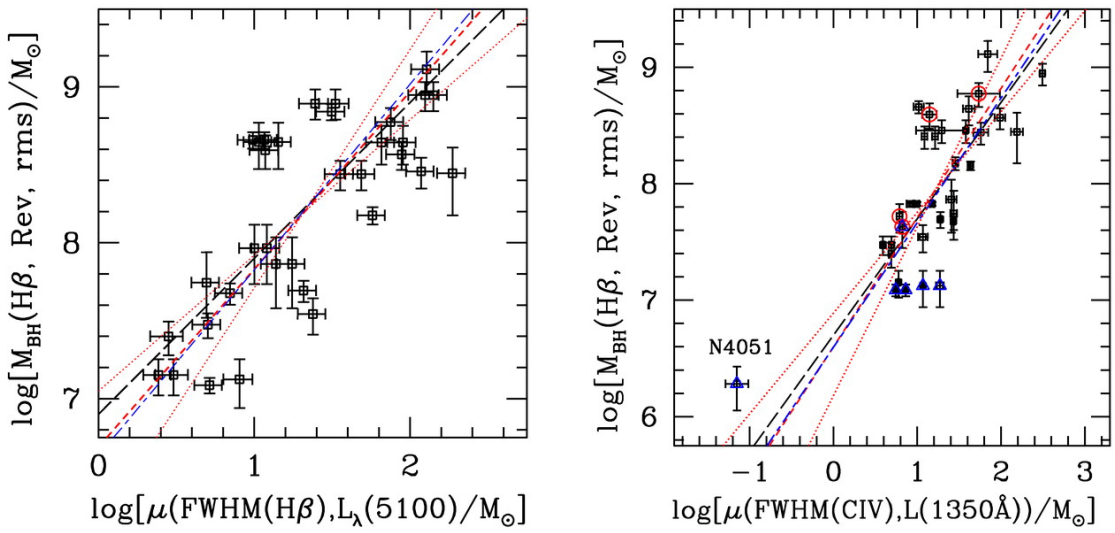
4 The BRL-NLR connection and C IV-based virial estimations

Having simultaneous detections for the different broad lines is crucial for testing their reliability as virial estimators. Nonetheless the majority of available data is not taken simultaneously and therefore we need to find a way to use C IV as a replacement for the line of choice, Hβ, even in these cases. Moreover, a simultaneous detection does not remove the contaminating component, whose presence is mainly responsible for the unreliablity of C IV as a virial estimator, from the line profile. As we have repeatedly mentioned, the major issue we have to deal with is the fact that an analysis of C IV profile reveals the presence of non-virialised gas kinematical signatures.

In order to use C IV in place of Hβ we need to identify which part of the line is emitted by gas in non-virialised conditions and to correct the line profile so that this part is not considered in the line width estimation.

We found a correlation between the strength of [O III] and the shape of C IV in quasar spectra. This pattern in the behaviour of the two lines reveals the presence of some interplay between the NLR and the BLR, in which some kinematic features of one or both the regions affect one another. This finding, important in itself to comprehend which physical processes are at work in the region connecting BLR and NLR and also how these nuclear components are shaped, can be used to identify (and remove) the part of the C IV line that is not emitted by virialised gas. What we are looking for is a correction for the C IV profile based on the C IV-[O III] correlation, so that this broad line can be eventually used safely as a virial estimator. Specifically, we want to correct FWHM(C IV), the profile property commonly used in virial estimation prescriptions.

In this chapter we will introduce the mutual behaviour of C IV and [O III] and the plausible scenarios explaining this correlation. We will then describe the detailed spectral and statistical analysis carried out on a sample of ∼170 quasar spectra. This is still a work in progress and, so far, a correction for FWHM(C IV) was found only for a small sub-sample of objects; still, the result is very promising. Moreover, we have now defined the methodology to be used for the statistical analysis of properties measured on the line profiles; this can be applied not only to our sources, but to all the samples for which both an optical ([O III]) and a UV (C IV) observation are available.



(a) (b)

Figure 4.1: (a) Scaling relation for Hβ. The unscaled BH mass μ for Single-Epoch is inferred from the FWHM of the broad component of the line and the continuum luinosity at 5100Å. The uscaled mass μ is then compared to the direct mass measured from Reverberation Mapping for the same objects. The long dashed line represents a slope of 1.0, meaning a perfect proportionality between the masses determined with the two methods. Since the proportionality is verified within the errors, the only difference between the two masses is the zero-point of the relation. (b) In the case of C IV the calibration of single-epoch measurements is still based on Hβ RM data, because too few C IV RM measurements are available to construct an indipendent C IV $R_{BLR} - L$ relation. (Vestergaard & Peterson, 2006).

This work is the subject of Bisogni, Denney & Peterson (in preparation).

4.1 Once again on virial mass calibrations

We explained in Section 1.3.2 the working principle of BH mass ladder and made already use of virial mass estimation prescriptions in literature in the previous chapter. Despite that, we never got into the details of the calibrations, i.e. on how they are actually performed, and this is instead needed in order to understand how to correct C IV profile. Specifically, we are interested in the calibration between the second and the third rungs of BH mass ladder, the calibration between the first and the second being simply the determination of a *zero-point* that takes into account the unknows of BLR geometry and kinematics. In the case of RM and Single-Epoch estimations, performing calibrations is more complicated than that, since we are relying on empirical scaling relations derived from RM rather than direct measurements of the dynamics of the system. This is also why single-epoch masses are less reliable than direct mass measurements, nonetheless essential to estimate BH masses for large samples of quasars. In this case the *unscaled* BH mass is deduced from the virial assumption and the $R_{BLR} - L$ relation by

$$M_{SE} = \left(\frac{W}{1000\ km\ s^{-1}}\right)^2 \left(\frac{\lambda L_\lambda}{10^{44}\ erg\ s^{-1}}\right)^\gamma , \tag{4.1}$$

where W is the parameter describing the line width (it can be either FWHM or σ_l) and L_λ is the monochromatic luminosity at a wavelength close to the line we are examining. The power of the linewidth is imposed by the assumption of virialisation, while the power γ is inferred from the slope of the *radius-luminosity* relation.

Under the assumption that M_{SE} is proportional to the mass we measure through RM, the two masses should be related by the equation

$$log(M_{RM}) = log(M_{SE}) + a \ , \tag{4.2}$$

the only difference between the two being the *zero-point* of the relation, a.

Fig. 4.1 (a) (Vestergaard & Peterson, 2006) shows the result of this calibration for Hβ-based single-epoch mass. A linear proportionality between the *unscaled* single-epoch measurements and RM masses is found. The only correction to be applied is therefore the offset between the two masses, i.e. zero-point, determined with a linear fit to the data. So far the calibration of the third rung on the second rung in the BH mass ladder is not dissimilar to that of the second rung on the first rung. However, to properly calibrate single-epoch mass estimations based on a line we should use Reverberation measurements available for the very specific line and Hβ is the only one for which we have a considerable number of RM data. A few RM measurements for other lines are available (Metzroth et al., 2006; Kaspi et al., 2007; Trevese et al., 2014; Cackett et al., 2015), but not enough to calibrate a separate $R_{BLR} - L$ relation. The only possibility is to consider the radius-luminosity relation for Hβ as suitable for all the other broad lines as well. So we usually calibrate single-epoch masses for all lines to the Hβ RM data, whether or not this is the line of interest.

As a result, the mass scaling relationship for C IV, the line most readily available in the optical spectrum for redshifts higher than $z \sim 1.5$, is based on Hβ measurements (Vestergaard & Peterson, 2006). Fig. 4.1 (b) shows how Vestergaard & Peterson (2006) find an overall linear proportionality for their sample and a scatter in the relationship involving C IV single-epoch masses and Hβ reverberation mapping masses that is relatively small (0.3 dex). However, other studies (Baskin & Laor (2005); Shen & Liu (2012) among the others) find little correlation between these quantities and a higher scatter (~0.4-0.5 dex). The larger scatter of the C IV single-epoch masses around the Hβ RM masses is believed to be caused by the presence of outflows and winds in the regions emitting the line.

4.2 What makes C IV different?

In a recent work Denney (2012) examined the possibility that outflows alter the estimate of BH masses based on C IV, by pointing out how reverberation mapping based C IV masses are in agreement with Hβ masses, differently from single-epoch based C IV masses. This means that the region emitting the variable part of C IV, the one used in RM measurements, does not seem to be subject to the presence of gas in non-virial motion. Denney (2012) suggests that a possible explanation of the discrepancy between C IV and Hβ masses is due

to the difference between what spectrum, *mean* or *rms*, is used to compute the line width. As mentioned in Section 1.3.2, the rms spectrum takes into account only the part of the spectrum that is variable and is reverberating the continuum emitted by the accretion disk, while the mean spectrum also shows non-variable contributions to the line and is very similar to single-epoch spectra. In RM the line width is evaluated from the rms spectrum. Single-epoch mass estimates instead are, by definition, based on the instantaneous line-of-sight velocity, because a single spectrum contains no variability information. For Hβ there is not a substantial difference between the single-epoch and the rms profile, with the exception of a contribution from the narrow line region, that is well known and subtracted before measuring the Hβ line width. However, for C IV, a comparison of the rms and the single-epoch (or mean) spectrum shows that in the latter a non-variable part that is otherwise absent from the former is present (see Fig. 4.2).

To provide a more quantitative estimate of the discrepancy between the two spectra, Denney (2012) considered the difference in the shape of the line for the RM sources where both the mean and the rms spectra were available. The single-epoch, as well as the mean, spectrum indeed shows the non-variable part as a quite central, narrower component on top of the broad one, with the result that its profile turns out to be "peakier" than the rms spectrum profile. This difference can be evaluated with a shape parameter S =FWHM/σ ($S = 2.35$ for a Gaussian profile and $S = 1$ for a Lorentian profile). Peakier profiles will then have small values for S, while less peaky ones will have larger values. The ratio of the shape parameters in the two cases (rms and mean spectra) is then an indication of the presence, or the lack thereof, of a non-virial, non-reverberating contribution to the emission line. While for Hβ the ratio between the shape of the single-epoch and rms profile is close to 1, i.e. the rms and single-epoch spectra describe similar emission components, this is not true for C IV (Fig. 4.3).

This substantial difference could be the reason of the unreliability of single-epoch mass estimates when using C IV. Having knowledge of the presence of this component is therefore important if we want to apply some kind of correction to be able to infer the proper virial mass from a single-epoch measurement of C IV line properties.

However this is not the only benefit we can achieve in characterising the presence of this C IV component. The region emitting this part of the line must have a different physical behaviour compared to the part of the line that is "reverberating" and is potentially connected in some way to the outflows and winds that some argue are affecting this emission line. Then an analysis of the C IV profile behaviour has also the potential to lead to a better understanding of the kinematics of this emitting region.

In this work we are also investigating the difference in the physical conditions, kinematics, and/or geometry in the nucleus, that leads to changes in the C IV line shape. There are several scenarios able to reproduce the presence of this "narrower" component, including both the classic NLR and regions composed by gas in outflow at the edge connecting the BLR and the accretion disk. There are also several differences between the properties of

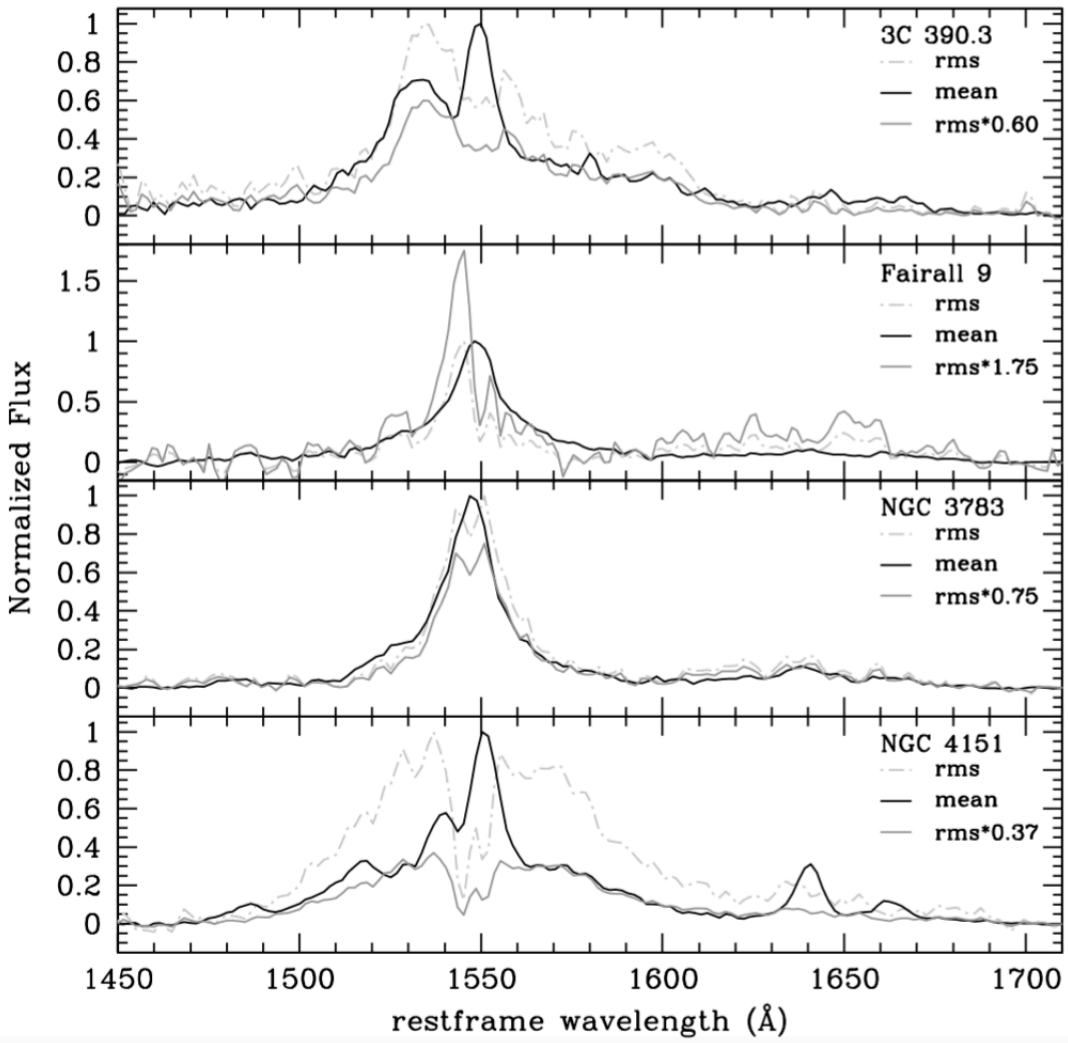


Figure 4.2: A comparison between peak-flux-normalized *rms* (grey dot-dashed line) and mean spectra (black) for some objects with C IV reverberation mapping measurements (Denney, 2012). The rms spectra are very different than the mean spectra in all the sources shown, with the *rms* spectrum broader and less peaky and the mean spectrum showing a narrower, low-velocity component on top of the component that is clearly missing in the rms spectra. The grey solid line is the rms spectrum multiplied by a factor in order to match the flux in the red wings of the mean spectrum C IV line.

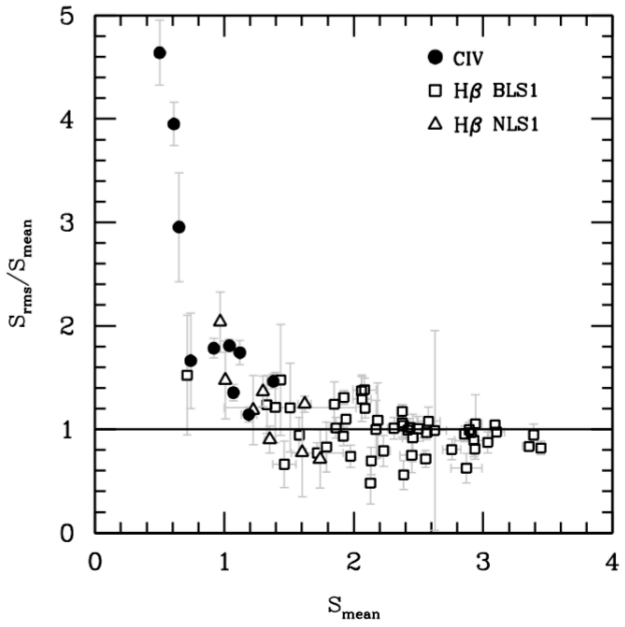


Figure 4.3: Ratios between the shape parameters S for *rms* and mean spectra for Hβ and C IV line (Denney, 2012). For Hβ the ratio is quite steadily set around 1, meaning that the line shape is similar in the mean and the rms spectra. C IV behaviour is very different, with the shape in the *rms* spectrum less peaky than in the mean spectrum, revealing the presence of the narrower, non variable component in the latter one.

the high ionisation C IV transition and lower ionisation transition, such as Hβ and Mg II, that can help us learn about physical differences of the emission behaviour in the high- and low-ionisation regions of the BLR and NLR.

4.3 C IV vs [O III]

Of particular interest for this work is to quantify for the first time the observation that the shape of C IV (characterised through several parameters, the shape parameter S among them) is connected with the behaviour of the narrow line [O III] λ5007Å (or more generally with the doublet [O III] $\lambda\lambda$4959, 5007Å), in the sense that in spectra in which C IV shows a very peaky profile we also find a very prominent [O III] emission. By contrast, when the C IV profile is more "boxy", the [O III] emission appears weak or is even not observed (see Fig. 4.4). This behaviour can have two explanations, each one not excluding the other.

The first "geometrical" explanation is related to the fact that [oiii] is a forbidden line coming only from the lower densities found in the NLR and, as discussed above (Chapter 2), is connected with the source orientation. C IV λ1549 is a permitted emission line and, in particular, is a collisionally-excited transition, that requires relatively higher densities. It is therefore more prominently seen to have only a broad component, but it has been argued (Denney (2012) and references therein) that C IV λ1549 emission can also arise in the highest-density, inner parts of the NLR. In this scheme, then, when we have a source with a prominent [O III] doublet, we could be looking at a source close to an edge-on position and C IV will

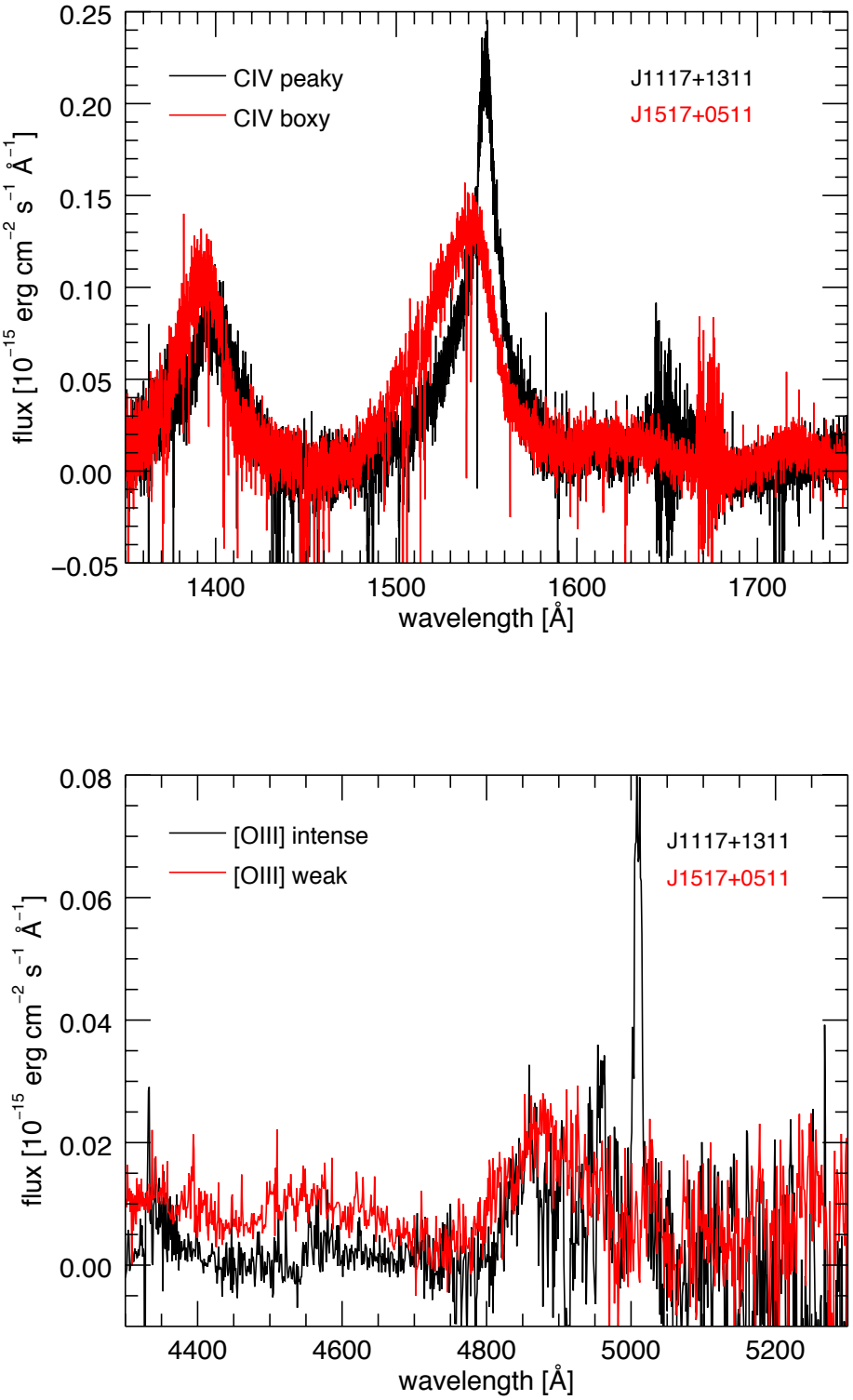


Figure 4.4: Comparison of C IV (upper panel) and Hβ+[O III] (lower panel) profiles for sources showing the two opposite behaviour: J1117+1311 (in black) exhibits a peaky and symmetric C IV profile and an intense [O III] emission; J1517+0511 instead (in red) presents a boxy and blueshifted profile and a barely noticeable [O III] emission. See Section 4.4 for a reference to the sources and for a description of the sample to which they belong.

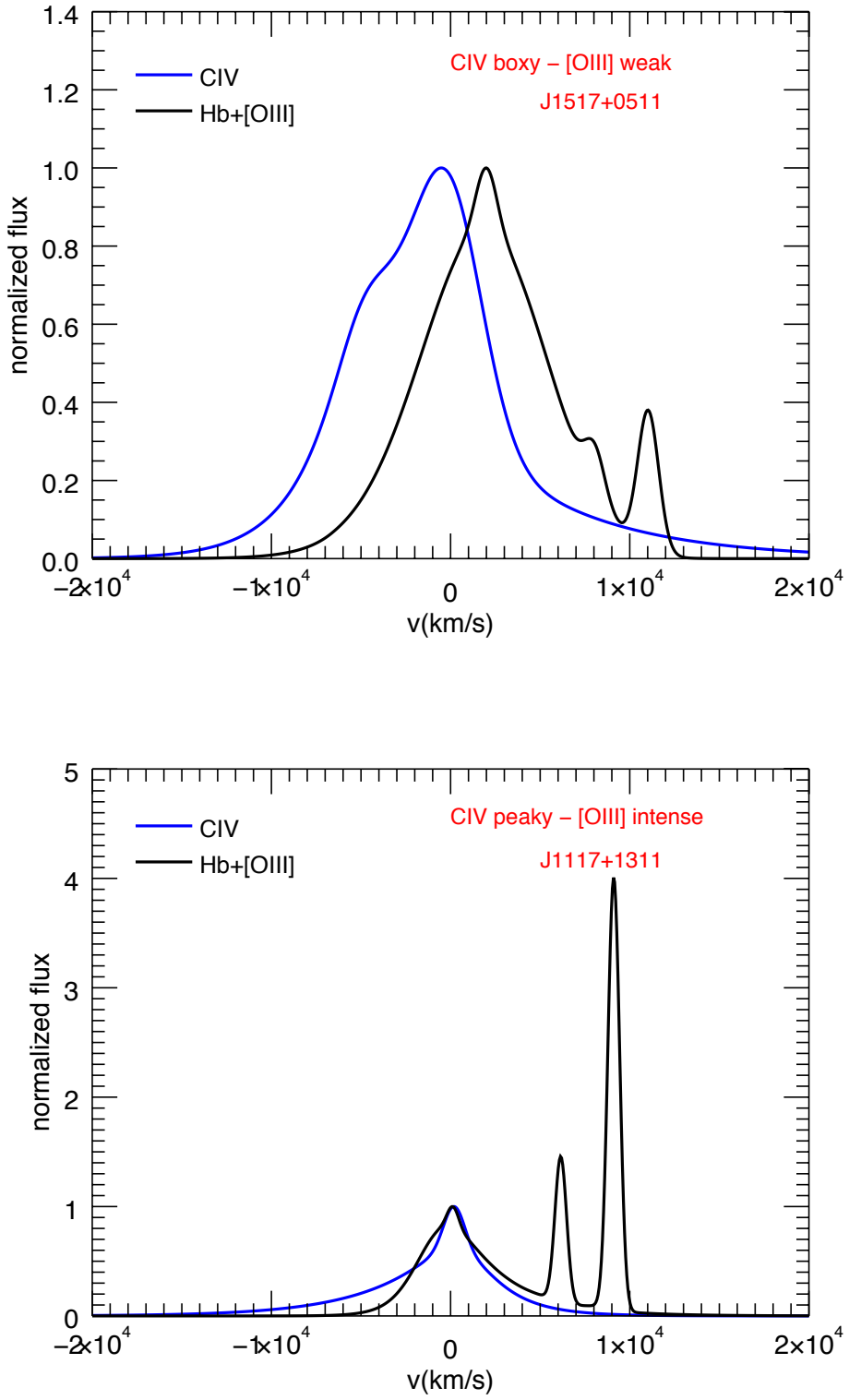


Figure 4.5: A comparison of the best-fit C IV and Hβ+[O III] profiles for the same sources as in Fig. 4.4. The C IV and Hβ profiles are normalized to their peak values in order to make the comparison easier. For J1517+0511 (upper panel) the [O III] emission is weak and the C IV profile appears to be blueshifted and very asymmetric. J1117+1311 (lower panel) exhibits an intense [O III] and a symmetric and peaky C IV. See Section 4.4 for a reference to the sources and for a description of the sample to which they belong.

show a narrow component as well. In face-on sources, in contrast, the narrow components are hidden to an extent by the bright continnum emitted by the accretion disk, that decreases as the cosine of the inclination angle. Then in this case the non-variable component should be emitted by the classic NLR.

However, there could be more than a physical structure-dependent explanation for C IV behaviour. In sources where we can see [O III] prominently, C IV also has a very peaky profile that can surely feature a narrow component, but also, the broad component appears very broad with extended wings, and relatively symmetric. In objects that lack a prominent emission of [O III], the C IV profile exhibits a generally more "boxy" shape and it is often asymmetric, showing a blueshift in its central velocity with respect to the other broad lines (see for example a comparison with Hβ in Fig. 4.5). This behaviour may be connected with the presence of gas in strong outflow in the region of the BLR. Because of the proximity of the C IV emitting region to the accretion disk, this outflow could arise from a region connecting the BLR and the accretion disk; it is not ruled out that the BLR could be a continuation of the accretion disk under different physical conditions (Emmering et al., 1992; Elitzur & Shlosman, 2006; Elitzur & Ho, 2009).

Such an outflow should be present in a source depending on its physical characteristics and activity phase, i.e. sources in different mass accretion rates and therefore with different luminosities can behave differently; the same can be true for sources hosting BHs with different masses. The presence of strong outflows could lead to the observation of gas in non-virial motion and, more importantly, outflows could inhibit emissions from the NLR or change the kinematical properties of the gas, if a NLR can be present at all in such sources. This scenario fits perfectly well with the C IV behaviour. On the one hand, when we see the [O III] emission the NLR is certainly present in the source. Consequently, the broad C IV component does not show strong evidence for a possible outflow in the BLR, i.e. blueshift peak or strongly asymmetric profile. On the other hand, when the broad C IV component exhibits a more boxy profile, larger shape parameter S, asymmetries and velocity shifts, the [O III] emission seems to be strongly reduced if not completely absent. The question we would like to ask in this work is then whether or not the presence of a strong outflow in the BLR is able to prevent the formation of the NLR in the same source.

As stated above, the geometrical and physical explanations do not exclude each other. In fact they work "in the same direction", i.e., the two effects are degenerate in the expected changes to observable spectroscopic features of AGN. It would be therefore very interesting to find a way to break such a degeneracy and figure out the way to disentangle the part pertaining to the geometrical effect and the one concerning the actual physical process.

4.4 Sample and data

To look for the correlation between C IV and [O III] behaviour, we collected a relatively large sample of quasars for which high quality UV and optical spectra were available. The sample

is initially composed as follows:

* 47 sources from the XQ-100 survey (Lopez et al., 2016) observed with XShooter (Vernet et al., 2011). The whole sample is composed of 100 quasars with redshift $z \sim 3.5 - 4.5$. We selected the 47 sources with $z < 4$, for which the coverage in the K-band was available, i.e. for which we have both the rest-frame UV (C IV) and optical ([O III]) emissions. All the details regarding data reduction can be found in Lopez et al. (2016). Unfortunately, depending on the redshift of the sources, the optical emission lines we are interested in fall in the last K-band order, known to have a vignetting problem[1]. In several cases, therefore, during the data fitting, we were unable to properly retrieve the [O III] line profile; in some cases we decided to not consider the source at all (10 objects), while in other cases we decided to flag the source as not fully reliable (19 objects). This was not the case for C IV, located in the VIS arm and not affected by signal to noise problems. More details are given in Tab. B.1.

* 70 sources out of the 85 objects from the Shang et al. (2011) sample, which is composed by three sub-samples:

  - the "PGX" sample (22 out of 23 sources selected by Laor et al. (1994, 1997) from the Bright Quasar Survey (Schmidt & Green, 1983)). For these sources Shang et al. (2007) obtained *HST* spectra for the UV wavelength and quasi-simultaneous (within a month) groud-based observation for optical spectra.
  - the "FUSE-HST" sample (17 objects out of the *Far Ultraviolet Spectroscopic Explorer* AGN program (Kriss, 2001) observed also with *HST* (UV spectra) and with the *Kitt Peak National Observatory* (KPNO) (optical spectra) (Shang et al., 2005))
  - the "RLQ" sample (originally containing 31 radio-loud quasars observed with *HST* (UV spectra) and with ground-based observatory (primarily McDonald Observatory and KPNO) (optical spectra) (Wills et al., 1995; Netzer et al., 1995)).

* 3 objects out of the 15 in the Netzer et al. (2007) sample. The optical spectra were obtained with the Gemini Near-Infrared Spectrographs (GNIRS) on Gemini South, while the UV spectra are from the SDSS.

* 6 objects out of the 10 in the Dietrich et al. (2009) sample. Optical rest-frame spectra were obtained with the spectrograph *SofI* at the 3.5 m NTT at la Silla/ESO, while UV rest-frame spectra are taken from the sample of Dietrich et al. (2002) or from the SDSS.

* 19 sources from an archival sample observed with GNIRS (GS-2004B-Q16, PI Todd Boroson; see details for data reduction in Collinson et al. (2015)). For these sources we

[1] http://www.eso.org/observing/dfo/quality/XSHOOTER/qc/problems/problems_xshooter.html

Table 4.1: MODS observation of 17 of the 19 sources of the *Boroson sample*, i.e. the sample for which we already had rest-frame optical data obtained with the GNIRS (Collinson et al., 2015). All the observations were performed with a dual-grating and clear-filter configuration. For Q1019-0054 and Q2314-0824, for which we could not obtain MODS spectra, SDSS spectra are used instead. For every source the table gives coordinates (RA and dec), the redshift (as obtained from the fit of the optical spectrum), *i*-SDSS magnitude, exposure time, airmass at the beginning of the integration, average seeing and observation date.

Name	RA	dec	z	*i* [mag]	Exp. Time [s]	Airmass	Seeing ["]	date
Q0019-0913	00 : 19 : 54.67	−09 : 13 : 16.4	2.1220	18.51	400x4	1.38	1.3	*12 Oct 2015*
Q0118-0052	01 : 18 : 27.98	−00 : 52 : 39.8	2.1932	18.01	300x3	1.20	0.7	*7 Dec 2015*
Q0139+0013	01 : 39 : 29.52	+00 : 13 : 30.8	2.1006	18.63	400x3	1.26	1.0	*13 Oct 2015*
Q0148-0010	01 : 48 : 09.65	−00 : 10 : 17.7	2.1706	17.74	300x3	1.22	1.0	*1 Jan 2016*
Q0209-0825	02 : 09 : 02.86	−08 : 25 : 31.8	2.1352	18.50	400x3	1.33	1.1	*13 Oct 2015*
Q0321-0011	03 : 21 : 58.39	−00 : 11 : 02.6	2.1573	19.31	800x4	1.19	0.8	*15 Oct 2015*
Q0409-0411	04 : 09 : 54.19	−04 : 11 : 36.9	2.1834	17.69	360x3	1.33	0.9	*11 Feb 2016*
Q0912+0058	09 : 12 : 08.64	+00 : 58 : 57.3	2.1837	18.26	400x3	1.60	1.5	*5 Dec 2015*
Q1002+0021	10 : 02 : 46.80	+00 : 21 : 04.0	2.1726	17.85	300x3	1.20	1.5	*5 Dec 2015*
Q1019-0054	10 : 19 : 00.00	−00 : 54 : 20.1	2.1818	18.36	-	-	-	-
Q1029+0205	10 : 29 : 06.72	+02 : 05 : 00.0	2.1391	18.57	600x3	1.20	0.8	*12 Feb 2016*
Q1033+0128	10 : 33 : 25.92	+01 : 28 : 36.3	2.1811	18.45	480x3	1.30	0.9	*12 Feb 2016*
Q1049-0110	10 : 49 : 15.36	−01 : 10 : 38.1	2.1227	17.61	300x3	1.44	1.0	*9 Dec 2015*
Q1214-0307	12 : 14 : 27.84	−03 : 07 : 21.0	2.1252	18.46	480x3	1.27	0.8	*12 Feb 2016*
Q1235+0304	12 : 35 : 14.40	+03 : 04 : 16.7	2.1994	18.61	600x3	1.25	0.9	*8 Feb 2016*
Q2059-0019	20 : 59 : 54.48	−00 : 19 : 17.3	2.1113	18.48	400x3	1.22	1.1	*11 Oct 2015*
Q2118+0042	21 : 18 : 32.88	+00 : 42 : 19.0	2.1684	18.01	300x5	1.20	1.0	*13 Oct 2015*
Q2259+1402	22 : 59 : 21.60	+14 : 02 : 56.0	2.1497	17.98	300x3	1.21	1.0	*11 Oct 2015*
Q2314-0824	23 : 14 : 41.52	−08 : 24 : 06.7	2.2012	18.64	-	-	-	-

also obtained observing time at the *Large Binocular Telescope* (LBT, Rothberg et al. (2016)). High-quality UV spectra with the optical spectrograph *MODS* (Pogge et al., 2010) were obtained between October 2015 and January 2016 for all the sources, with the exception of Q1019-0054 and Q2314-0824, for which SDSS spectra were used instead. The log for the observations is given in Tab. 4.1.

* 27 objects from the *Reverberation Mapping Sample* (Reichert et al., 1994; Santos-Lleó et al., 1997; Peterson et al., 1998; Dietrich et al., 1998; Kaspi et al., 2000; Peterson et al., 2000; Santos-Lleó et al., 2001; Peterson et al., 2002; Onken & Peterson, 2002; Peterson et al., 2004; Denney et al., 2006; Bentz et al., 2006; Denney et al., 2010; Grier et al., 2012; Peterson et al., 2014).

Details on the complete sample are given in Appendix B (Tab. B.1). In particular, we flag objects with 1 when we consider them reliable, 0 when they are considered not completely reliable and X when we do not consider them at all in the following analysis. The worst case is for the XQ-100 subsample, when the *K*-band data are often so much noisy we could not rely on the fitted profile.

The final sample is composed by 172 objects, of which 10 are rejected due to poor data quality or because they lack a rest-frame optical observation, and 25 are instead flagged as not completely reliable.

4.5 Data analysis

4.5.1 Spectral fitting

We performed fits for both the Mg II and Hβ+[O III] windows as described in Section 2.3.2. The only difference that is worth mentioning is the way we dealt with the C IV spectral window. The aim of this work is to isolate the "contaminating", non-reverberating component for this line that, as said before, is a combination of the narrow component, if present, with emissions coming from the Broad Line Region under a different condition with respect to the virialized one (gas in outflows/winds). To have a good description of the C IV line we then decided to fit the line to obtain a smooth representation of the profile; we are therefore not trying to decompose the line in different components with a physical meaning and to relate them to a specific emitting region, as we did with the other narrow and broad lines examined in the other spectral windows, but rather we want the best realization of the whole line profile.

In order to do that, we usually used two components to reproduce C IV, with only a few exceptions, for which we used only 1 or up to 3 components. Given the asymmetric and often peculiar shape of C IV, we used both a double power-law (Eq. 2.2) or a Gaussian function (sometimes both) to represent the upper part of the line and Gauss-Hermite polynomials to reproduce the lower part. The use of Gauss-Hermite polynomials is indeed appropriate to describe the peculiar shapes of this line; the use of these functions was first suggested by van der Marel & Franx (1993) to describe profiles showing significant deviations, both symmetric and asymmetric, from a Gaussian.

The functional form of the Gauss-Hermite series is:

$$GH_n(\lambda) = A\frac{e^{-w^2/2}}{\sigma\sqrt{2\pi}}\sum_{j=0}^{n} h_j H_j(w) \tag{4.3}$$

with

$$w = \frac{\lambda - \lambda_c}{\sigma}\ , \tag{4.4}$$

where A is the amplitude of the series, λ_c is the peak wavelength, h_j are the Gauss-Hermite moments and $H_j(w)$ are the Hermite polynomials. The first seven Hermite orthogo-

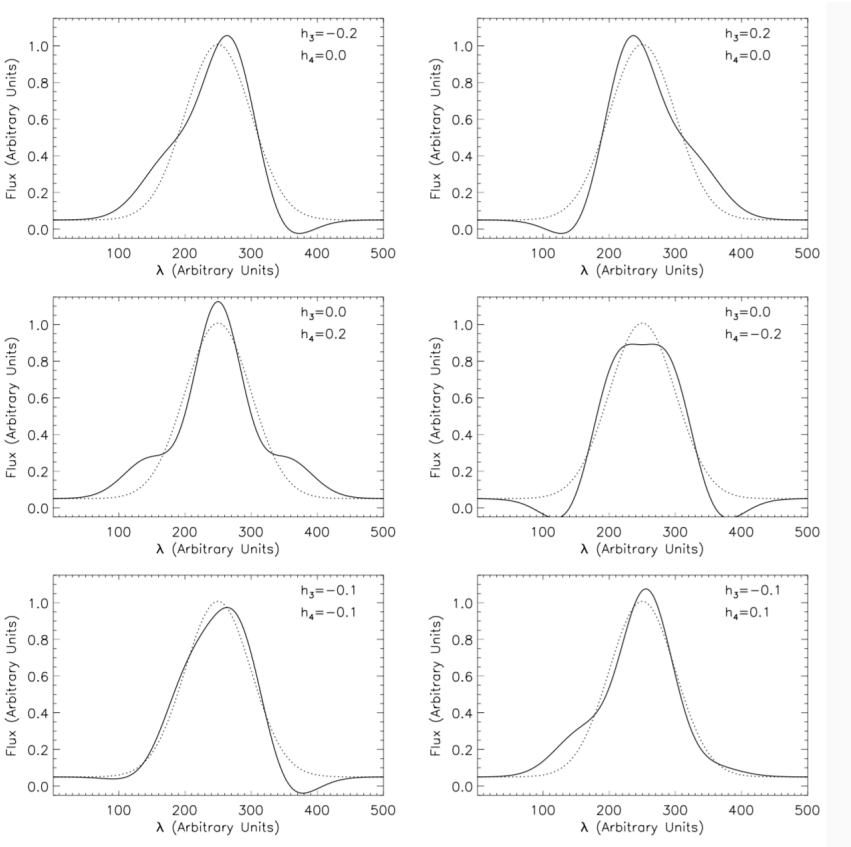


Figure 4.6: Riffel (2010): comparison of a Gaussian function (solid line) and a Gauss-Hermite series (dotted line) with the same amplitude A, σ and λ_c, for different values of h_3 and h_4. The odd h_3 moment gives the asymmetric deviation from the Gaussian function (red and blue wings), while the even h_4 moment measures the "peakiness" of the line ($h_4 < 0$ "boxier", $h_4 > 0$ "peakier" profile). When all the moments $h_j = 0$ the Gauss-Hermite series becomes a Gaussian.

nal polynomials are (Cappellari et al., 2002)

$$
\begin{aligned}
H_0(w) &= 1, \\
H_1(w) &= \sqrt{2}w, \\
H_2(w) &= \frac{2w^2 - 1}{\sqrt{2}}, \\
H_3(w) &= \frac{w(2w^2 - 3)}{\sqrt{3}}, \\
H_4(w) &= \frac{w^2(4w^2 - 12) + 3}{2\sqrt{6}}, \\
H_5(w) &= \frac{w(w^2(4w^2 - 20) + 15)}{2\sqrt{15}}, \\
H_6(w) &= \frac{w^2(w^2(8w^2 - 60) + 90) - 15}{12\sqrt{5}},
\end{aligned}
$$

where the even and odd terms describe, respectively, symmetric and asymmetric deviations from the Gaussian function (Fig 4.6).

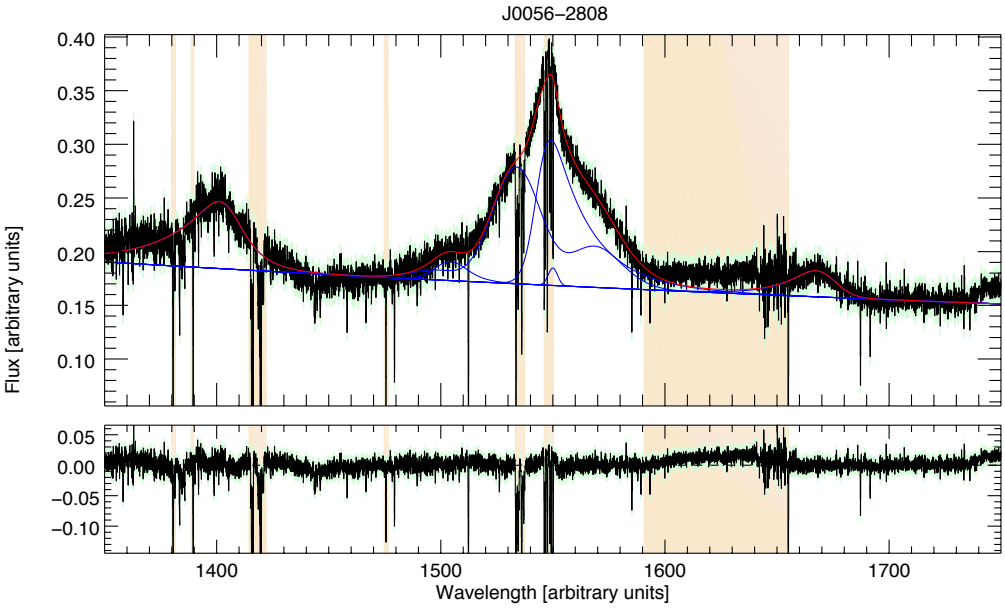


Figure 4.7: Fit of the C IV spectral window for J0056-2808. The original rest-framed spectrum is shown in black, while the blue lines indicate the model components used to fit the emission lines. In this specific case the fit to the C IV line has been obtained with (1) a GH_4, (2) a double power law, (3) a Gaussian function and (4) the model component for N IV] which shifted to cover some feature of the C IV line profile. None of them has a physical meaning, they are only used to reproduce a smooth representation of the C IV emission line. The total model, as the sum of all these individual model components, is shown in red. The colored regions indicate the spectral ranges that were masked in the fit. In particular, the region between $\sim$ 1590Å and $\sim$ 1620-30Å, known as the *red shelf* of C IV, has been extended to $\sim$ 1650Å in this case due to the presence of noise. As a result, only O III] is present in the fit of this spectrum. All the narrow absorption features of C IV were masked as well. This is an example of how asymmetric and peculiar the C IV profile can be.

Depending on the examined source, we considered both polynomials up to the forth (GH_4) or to the sixth term (GH_6), although the second case is only for a few sources, since the determination of h_5 and h_6 requires a very high S/N ratio, a condition not always verified by our spectra (S/N in the UV range from ∼8 up to ∼200, tipically more than 30).

C IV is the only line treated in a specific way. All the other permitted lines in the (rest frame) UV spectral window are usually fitted with a single broad component. Generally, in this wavelength range, a narrow component is not evident in the emissions; moreover, we are not interested in a complete description of lines other than C IV, as long as the observed spectrum is well reproduced by the fit.

The fitting range changes depending on the examined source redshift and on the spectral coverage of the instruments used for the observations. Usually, besides C IV, the lines we take into account are:

- Si IV λ1396.76 and O IV] λ1402.06: since we are not interested in resolving the two lines, we consider it as a unique complex at $\lambda \sim 1400$Å.

- N IV] λ1486.496: this emission line is not always present in the spectrum. If present, it is properly fitted as a separated component. However, since its central wavelength is not strictly constrained in the input fit, it can sometimes shift to cover some feature of C IV profile, when no actual N IV] emission is present in the spectrum. In these cases it is considered in the realization of the smoothed C IV profile.

- He II λ1640.42 and O III] λ1663.48: both lines are typically present in the spectrum. Sometimes one of the two lines reproduces the overall emission in this spectral range (1640 – 1680Å). This can also depend on the extension of the mask we chose to apply to the red-shelf of C IV in different sources.

As mentioned above, there are some regions that were masked during the fitting process. We chose to mask a region in three cases: (1) when we do not know what is emitting a given part of the spectrum (this is the case of the C IV *red-shelf*), (2) when the components we are taking into account in the fitting procedure are not able to reproduce emissions we see in the spectrum (this is instead the case of the bump of emission between O III] and the C III] complex at ∼ 1900Å) or (3) when the region is extremely noisy. Moreover, C IV emission line often shows the presence of narrow absorption features. We mask them as well, in order to achieve the best representation of the emitting line. An example of a C IV window fit is given in Fig. 4.7.

Finally, we recover the model for C IV and [O III] line profiles for all the sources for which both spectral windows are available.

4.5.2 Characterisation of the line profile

We noticed that when [O III] is strong, C IV shows a peaky, symmetric profile, while when [O III] is weak or absent in the optical spectrum then C IV is boxy, asymmetric and blueshifted

(Sec. 4.3). We want to connect the behaviours of the two lines more quantitatively, in order to indentify which part of the C IV emission is connected to [O III], that should be the constant, non-reverberating component we see in single-epoch spectra and that is instead missing in *rms* spectra. Specifically, the goal of the analysis is to find the properties describing the main features that we recognise in C IV behaviours and that are connected to [O III]: the "peakiness" (as well as the boxiness) and the "blueness" (or, otherwise, the symmetry) of the line profile.

In order to quantify the C IV-[O III] connection, the next step consists therefore in giving the most accurate description of the line profile for both the emissions. We then selected the following set of properties we want to measure.

- For the measurement of the line width we chose to use the following properties:
  - FWHM for the whole line profile and for the blue and red part only of the profile (for example, for the blue side of the line we have $\mathrm{FWHM}_{blue} = 2\ \mathrm{HWHM}_{blue}$);
  - dispersion σ (as the usual square root of the second moment of the distribution);
  - *Mean Absolute Deviation* (MAD), the mean of the absolute deviations of the distribution around the median value. This is a velocity width measurement which has the advantage to be more resilient to outliers in the data with respect to the standard deviation σ. The MAD is defined as
    $$\frac{\int |\lambda - \lambda_{\mathrm{MED}}| F(\lambda) d\lambda}{\int F(\lambda) d\lambda}\ , \tag{4.5}$$
    where $F(\lambda)$ is the line profile (i.e. the distribution) and λ_{MED} is the wavelength corresponding to the position of the median value in the distribution. This definition of MAD is slightly different from the usual one (i.e. *Median Absolute Deviation*), where the median value of the absolute deviations with respect to the median is taken.
  - *Inter-Percentile Velocity* (IPV) widths (see Section 3.3 and Eq. 2.3) for several percentile values. In particular we examined IPVs for 10, 20, 30, 40, 50, 60, 70, 80, 90% of the total flux of the line profile. IPV_{50} corresponds to the *InterQuartile Range* (IQR), also called *midspread*, which gives the distribution dispersion within the 25 and 75 percentiles.
- To describe the asymmetry of the line we chose the following parameters (in the following definitions the centroids $c(i/4)$ are defined as in Marziani et al. (1996))
  $$c(i/4) = \frac{(v_{red}(i/4) + v_{blue}(i/4))}{2}, \quad \text{for } i = 1, ..., 4\ .) \tag{4.6}$$
  - the *skewness* of the line profile, defined through the statistical moments of the distribution as
    $$Sk = \frac{\frac{\int (\lambda - \lambda_0)^3 F(\lambda) d\lambda}{\int F(\lambda) d\lambda}}{\left(\frac{\int (\lambda - \lambda_0)^2 F(\lambda) d\lambda}{\int F(\lambda) d\lambda}\right)^{3/2}}\ ; \tag{4.7}$$

profiles with a positive skewness are skewed to the right, while profiles with a negative skeweness are skewed to the left.

- several asymmetry indexes similar to $A_{IPV_{05-95}}$ defined in Chapter 2. These indexes are defined as differences between InterPercentile Velocities (see Eq. 2.5). Along with $A_{IPV_{05-95}}$, that measures the asymmetry of the line comparing velocities corresponding to wavelengths within which we find the 90% of the total flux, we considered other indexes, comparing velocities corresponding to different percentage of the total flux. The aim is that of "mapping" the asymmetry of the line at different levels, starting on the whole profile ($A_{IPV_{05-95}}$) and moving closer to the peak ($A_{IPV_{40-60}}$). Indexes > 1 describe a blueward asymmetry, while indexes < 1 a redward asymmetry.
- the asymmetry index defined in Shen & Liu (2012):

$$AS = \frac{\ln(\lambda_{red}/\lambda_{peak})}{\ln(\lambda_{peak}/\lambda_{blue})} , \tag{4.8}$$

  where λ_{peak} is the wavelength corresponding to the peak of the line and λ_{blue} and λ_{red} are those corresponding to half peak flux on the blue and red side of the line profile, respectively. This index is defined so that redward asymmetries correspond to values > 1, while blueward asymmetries to values < 1.
- the asymmetry index defined in De Robertis (1985):

$$\alpha_\beta = \frac{c(3/4) - c(1/4)}{\mathrm{FWHM}} . \tag{4.9}$$

  This parameter is negative for stronger red wings (Brotherton, 1996).
- the asymmetry index defined in Marziani et al. (1996)

$$AI_M = \frac{(v_{red}(1/4) + v_{blue}(1/4) - 2v_{peak})}{v_{red}(1/4) - v_{blue}(1/4)} ; \tag{4.10}$$

  this parameter is generally positive for red asymmetries and negative for blue asymmetries. Since it depends also on the shift of the peak v_{peak}, however, some profiles can have positive or negative values if the line peak is shifted shortward or longward of the nominal central wavelength, even if the line is symmetric.

- To give a measure of the shift of the line with respect the nominal central wavelength we used:
  - the shift of the centroid at 3/4 of the maximum flux, as defined in Boroson & Green (1992):

$$\mathrm{Shift}_{BG92} = \frac{c(3/4)}{\mathrm{FWHM}} . \tag{4.11}$$

  - the centroid corresponding to 9/10 of the maximum flux (Sulentic et al., 2007), definined accordingly to Marziani et al. (1996) definition:

$$c(9/10) = \frac{(v_{red}(9/10) + v_{blue}(9/10))}{2} . \tag{4.12}$$

- the shift of the centroid of the top of the line only (that corresponding to fluxes larger than the 80% of the maximum flux) with respect to the nominal central wavelength. In this case we computed the centroid as the first moment of the distribution:

$$\mathrm{Shift}_{c80} = \frac{\int_{80\%}^{100\%} \lambda F(\lambda) d\lambda}{\int_{80\%}^{100\%} F(\lambda) d\lambda} - \lambda_0 \ . \tag{4.13}$$

- Finally, to give a description of the shape of the line, we measure:
  - the "classical" shape parameter $S = \frac{\mathrm{FWHM}}{\sigma}$, already introduced in Section 4.2.
  - the *excess kurtosis*, defined through the statistical moments of the distribution as

$$exc_{Kr} = \frac{\frac{\int (\lambda-\lambda_0)^4 F(\lambda) d\lambda}{\int F(\lambda) d\lambda}}{\left(\frac{\int (\lambda-\lambda_0)^2 F(\lambda) d\lambda}{\int F(\lambda) d\lambda}\right)^2} - 3 \ . \tag{4.14}$$

    The excess kurtosis provides a comparison with the Gaussian distribution. From a statistical point of view this comparison is based on how much the variance of the distribution is due to large deviations from the central values (i.e. heavy or thin tails), but this also leads to differences in the shape of the distribution. This quantity can therefore be used as a shape parameter as well. In particular distributions with $exc_{kr} > 0$ have heavier tails and usually boxier peaks, while those with $exc_{kr} < 0$ have thinner tails and are usually peakier.
  - the shape parameter defined by Boroson & Green (1992):

$$\mathrm{Shape}_{BG92} = \frac{(\mathrm{FW}(1/4)\mathrm{M} + \mathrm{FW}(1/3)\mathrm{M})}{2\mathrm{FWHM}} \ ; \tag{4.15}$$

    this parameter, always positive, is higher for peaky and lower for boxy profiles.
  - and the one defined by Wills et al. (1993):

$$\mathrm{Shape}_{W93} = \frac{(\mathrm{FW}(3/4)\mathrm{M} - \mathrm{FW}(1/4)\mathrm{M})}{\mathrm{FWHM}} \ ; \tag{4.16}$$

    this paramater, always negative, has a higher absolute value for peakier profiles.
  - the kurtosis index defined by Marziani et al. (1996):

$$KrI_M = \frac{v_{red}(3/4) - v_{blue}(3/4)}{v_{red}(1/4) - v_{blue}(1/4)} \ , \tag{4.17}$$

    for which higher values correspond to peakier profiles.

- We also take into account other general characteristics and some properties chosen *ad hoc* to describe the "peakiness" of the line:
  - line centroid, defined as the first moment of the entire profile;
  - equivalent width (EW);
  - total flux of the line;

- $IPV_{1090} = IPV_{10}/IPV_{90}$, that gives a comparison of the line width of the top-part of the line and that of the base of the line;
- $EW(IPV_{10})$, the equivalent width of the central part of the line, the one containing the 10% of the total flux.

These last two quantities are meant to be representative of the peakiness of the line.

Some of these characteristics are surely redundant, but they provide slightly different descriptions of similar properties. For example, while FWHM and σ are both line width measurements, they are known to be representative of different characteristics of the line. Specifically, while FWHM is strongly dependent on the peak, σ is more sensitive to the tails of the profile. Similarly, other quantities describe comparable characteristics, but nonetheless they could convey different pieces of information on the lines behaviour. This is why we decided to measure all of them.

Before performing the line measurements, we interpolated the bestfit of the profile on a thicker wavelength grid. Specifically, since wavelength grids are different not only among different subsamples, but sometimes even within them, to be consistent we chose to use for each spectrum a new grid in which the size of the spectral channel is one hundredth of the original one. This choice does not substantially affect the measurements of quantities obtained through the integration of the line profile, such as the dispersion σ or the *Inter-Percentile Velocity widths* (see Eq. 2.3), but it significantly improves measurements related to the the location of single points (spectral channels) on the line profile, such as the FWHM, the centroid of the line or the centroid shift with respect to nominal laboratory wavelength and all the quantities derived from them. Re-sampling the bestfit profiles we are achieving a far better resolution in the determination of such properties.

In Appendix B we summarise all the quantities we measured on the complete sample (Tab. B.2). We still need to find an error estimation for the measurements. Given the complexity and the high number of the model components used to fit the C IV line profile, a reconstruction of a set of mock line profiles on which to perform measurement for all the quantities in order to infer an error estimation, as done for the analysis in chapter 2 and chapter 3, is not feasible. We are therefore estimating errors in the following way: we are reproducing 500 mock spectra for each source, adding to the best fitting spectrum a Gaussian noise proportional to the signal-to-noise that we estimated from the original one. We then obtain 500 best-fit line profiles on which we perform again all the measurements. The error on any given quantity will then be assigned as the standard deviation of the mock profiles measurements distribution. Moreover, we are in the process of correcting our profiles for spectral resolution. As pointed out in Peterson et al. (2004), this correction is very important in the case of narrow lines and should not be limited to them. In Tab. B.1 we report the resolution of the sub-samples for which we have been able to recover this information so far. Despite both the errors estimation and the resolution correction are of course needed, as explained in the following section, they do not however undermine the next step of the

analysis.

4.5.3 Correlation Analysis

Principal Component Analysis

Many literature works aim at establishing a connection between different lines (and spectroscopic features in general) in quasar spectra. Among them, we already mentioned the seminal work by Boroson & Green (1992), that had the merit to open to a completely different way to look at the available data. We gave an introduction to the statistical method known as *Principal Component Analysis* (PCA) in Section 1.5 and examined the main correlation found by Boroson & Green (1992) in the light of the orientation of the source in Section 2.4.3.

In this work we want to use the same technique to find which are the characteristics driving the variance in the data-set composed by all the properties described in the previous section, measured on both the C IV and [O III] profile.

To understand how this technique works, a few additional details are given. PCA is performed on the correlation matrix of the data-set, i.e. the data-centered and normalised covariance matrix (or the covariance matrix for the standardized variable). Since data are standardized, the total variance of the sample is equal to the number of variables involved in the analysis. This value also corresponds to the sum of the eigenvalues pertaining to the eigenvectors and representing the partial variance explained by each one of them.

Of course, since the aim of this kind of technique is the reduction of the dimensionality of the sample, when examining eigenvectors, we are considering only some of them. Usually, to chose how many eigenvectors consider and how many of them reject, the *eigenvalue-one criterion* (or Kaiser rule) is applied, i.e. we are chosing to use only the eigenvectors whose eigenvalue is more than 1. In doing this, we are discarding eigenvectors that account for less than the variance introduced in the sample by a single original variable. We are not taking this criterion literally, since, unlike in spectral PCA (Francis et al., 1992; Yip et al., 2004b; Pâris et al., 2011), we do not want to reconstruct the original sample with less dimensionality, we only want to find out which are the characteristics driving the variance in the sample, i.e., in this case, what is most representative for the connection between C IV and [O III] behaviour. We then decided to examine only 10 eigenvectors (that account for a cumulative variance in the sample of $\sim 90\%$), even if some of the following ones still have eigenvalues (slightly) larger than 1. Furthermore, we believe only the first ones among them deserve real attention (specifically the first five, explaining $\sim 70\%$ of the variance in the sample).

When performing PCA on a set of properties, it is then customary to show how each variable contributes in realizing the eigenvectors describing the variance of the sample. Eigenvectors are infact a linear combination of the original variables: the linear coefficient pertaining to a given variable, called "loading", is then a measure of the importance that variable has in the eigenvector.

We performed PCA analysis both on the complete sample, excluding only the spectra rejected due to bad data quality, and on the sample composed only by reliable spectra (the ones flagged as good in Tab. B.1). The two analysis give substantially the same results, with the existing correlations usually becoming stronger in the case in which we use only the good data-set. Therefore in the following discussion we are taking into account only the analysis for the good data-set. In Tab. 4.2 we report the loadings for the examined properties for the first 5 eigenvectors, as well as the percentage of variance explained by each one of them. In particular the eigenvector 1 accounts for $\sim 27\%$ of the variance and the first five eigenvectors alone account for $\sim 72\%$ of the total variance in the sample.

Having a look to the loadings there are a few things that stand out:

1. the eigenvector 1 is strongly dominated by properties related to [O III], even if some C IV properties are important as well.

2. the [O III] linewidth is the strongest driver of this eigenvector. In particular, IPVs seem to be the quantities explaining the most part of the variance in the sample. Specifically, IPV_{60} and IPV_{50} weight the most in the realisation of EV1, with all the other IPVs, MAD and σ following close behind. Even if to a lesser extent, also FWHM is important in EV1.

3. the other characteristics driving EV1 are the asymmetry indexes, mostly $A_{IPV_{20\text{-}80}}$, $A_{IPV_{30\text{-}70}}$ and those similarly defined, but also α_β and AI_M.

4. the third most significant driver is the shape of [O III], described by Shape_{W93}, Shape_{BG92} and KrI_M.

5. to a lesser extent, some of C IV properties are relevant as well, the most important being the shift of the line centroid Shift_{c80} (and $c(9/10)$ of course, representing a similar quantity).

6. FWHM(C IV) has a relevant loading as well, followed by the IPVs of the central part of the line (IPV_{10}, IPV_{20}), while σ seems to be less important.

7. EW(IPV_{10})(C IV) follows. Since we defined this quantity to trace the peakiness of the line, it is somehow surprising that the same loading is found for EW of the whole line.

8. important properties are also those defining the asymmetry of C IV, like $\mathrm{AI}_{05\text{-}95}$, AI_M and Sk.

9. conversely, eigenvector 2 is strongly dominated by properties describing C IV linewidth.

Therefore, we note the presence of some drivers that we already suspected to underlie the variance, but we see how, for example EW[O III], that we suspected to be important, not only is not a driver of the eigenvector 1, it does not drive anyone of the first, more important eigenvectors. An analysis of this kind is aimed to the identification of what makes the spectra

Table 4.2: Loadings of all the measured properties for the first five principal components (or eigenvectors). Values in bold are the loadings considered to be significant based on the examined eigenvectors.

		EV1	EV2	EV3	EV4	EV5
Var. explained		26.9%	17.9%	11.7%	10.1%	5.0%
Cum. var. explained		26.9%	44.8%	56.5%	66.6%	71.5%
Property						
FWHM(C IV)	[km/s]	**+0.404**	**−0.646**	−0.387	−0.320	+0.289
σ(C IV)	[km/s]	+0.216	**−0.880**	+0.221	+0.190	−0.015
Shift$_{c80}$(C IV)	[km/s]	**−0.435**	+0.214	+0.577	+0.287	+0.001
IPV$_{10}$(C IV)	[km/s]	**+0.364**	**−0.868**	−0.148	−0.026	+0.227
IPV$_{20}$(C IV)	[km/s]	**+0.342**	**−0.896**	−0.110	+0.002	+0.199
IPV$_{30}$(C IV)	[km/s]	**+0.313**	**−0.923**	−0.057	+0.044	+0.156
IPV$_{40}$(C IV)	[km/s]	+0.280	**−0.941**	+0.007	+0.085	+0.110
IPV$_{50}$(C IV)	[km/s]	+0.248	**−0.948**	+0.076	+0.124	+0.064
IPV$_{60}$(C IV)	[km/s]	+0.224	**−0.945**	+0.138	+0.151	0.025
IPV$_{70}$(C IV)	[km/s]	+0.210	**−0.934**	+0.186	+0.169	+0.001
IPV$_{80}$(C IV)	[km/s]	+0.206	**−0.914**	+0.223	+0.185	−0.010
IPV$_{90}$(C IV)	[km/s]	+0.201	**−0.883**	+0.253	+0.193	−0.020
A$_{IPV_{05\text{-}95}}$(C IV)		**+0.334**	+0.125	−0.427	+0.563	+0.205
A$_{IPV_{10\text{-}90}}$(C IV)		+0.196	+0.229	−0.277	+0.822	+0.046
A$_{IPV_{20\text{-}80}}$(C IV)		+0.202	+0.238	−0.288	+0.844	+0.039
A$_{IPV_{30\text{-}70}}$(C IV)		+0.222	+0.249	−0.303	+0.850	+0.060
A$_{IPV_{40\text{-}60}}$(C IV)		+0.263	+0.273	−0.342	+0.817	+0.108
MAD(C IV)	[km/s]	+0.229	**−0.921**	+0.186	+0.168	+0.014
EW(C IV)	[Å]	**−0.363**	−0.168	+0.382	−0.055	+0.004
AS(C IV)		−0.184	−0.231	+0.384	−0.422	−0.178
Shape$_{BG92}$(C IV)		−0.177	−0.183	+0.372	+0.578	−0.458
Shape$_{W93}$(C IV)		+0.221	+0.267	−0.436	−0.585	+0.403
α_β(C IV)		+0.258	+0.276	−0.302	+0.729	+0.113
Shift$_{BG92}$(C IV)		**−0.351**	+0.022	+0.427	+0.324	−0.076
c(9/10)(C IV)	[km/s]	**−0.426**	+0.216	+0.547	+0.310	+0.030
AI$_M$(C IV)		**−0.313**	−0.270	+0.424	−0.592	−0.248
KrI$_M$(C IV)		+0.261	+0.298	−0.500	−0.444	+0.262
total flux (C IV)	[erg s^{-1} cm^{-2}]	−0.030	+0.004	−0.059	+0.227	−0.327
Sk(C IV)		**−0.308**	−0.092	+0.360	−0.432	−0.182
exc_{Kr}(C IV)		−0.044	+0.138	+0.081	−0.006	−0.128
S(C IV)		+0.245	−0.006	−0.550	−0.499	+0.385
IPV_{1090}(C IV)		+0.269	−0.107	−0.541	−0.271	+0.420
EW$_{IPV_{10}}$(C IV)	[Å]	**−0.364**	−0.168	0.382	−0.055	+0.004
FWHM([O III])	[km/s]	**+0.643**	−0.119	−0.531	−0.077	−0.340
σ([O III])	[km/s]	**+0.918**	+0.081	−0.070	−0.044	−0.116
IPV$_{10}$([O III])	[km/s]	**+0.923**	+0.028	−0.152	−0.075	−0.269
IPV$_{20}$([O III])	[km/s]	**+0.935**	+0.036	−0.136	−0.077	−0.266
IPV$_{30}$([O III])	[km/s]	**+0.944**	+0.043	−0.122	−0.076	−0.259
IPV$_{40}$([O III])	[km/s]	**+0.949**	+0.050	−0.112	−0.074	−0.246
IPV$_{50}$([O III])	[km/s]	**+0.954**	+0.061	−0.099	−0.070	−0.225
IPV$_{60}$([O III])	[km/s]	**+0.954**	+0.070	−0.081	−0.066	−0.201
IPV$_{70}$([O III])	[km/s]	**+0.950**	+0.074	−0.068	−0.060	−0.175
IPV$_{80}$([O III])	[km/s]	**+0.940**	+0.076	−0.063	−0.054	−0.147
IPV$_{90}$([O III])	[km/s]	**+0.915**	+0.076	−0.067	−0.044	−0.108
A$_{IPV_{05\text{-}95}}$([O III])		+0.233	+0.252	+0.397	0.065	+0.518
A$_{IPV_{10\text{-}90}}$([O III])		**+0.717**	+0.293	+0.511	−0.031	+0.178
A$_{IPV_{20\text{-}80}}$([O III])		**+0.737**	+0.301	+0.520	−0.032	+0.151
A$_{IPV_{30\text{-}70}}$([O III])		**+0.728**	+0.311	+0.526	−0.016	+0.182
A$_{IPV_{40\text{-}60}}$([O III])		**+0.699**	+0.323	+0.518	+0.003	+0.245
MAD([O III])	[km/s]	**+0.942**	+0.071	−0.081	−0.055	−0.159
EW([O III])	[Å]	+0.113	−0.025	+0.383	+0.051	−0.276
AS([O III])		−0.378	−0.301	−0.240	+0.113	−0.024
Shape$_{BG92}$([O III])		**+0.640**	+0.134	+0.534	−0.008	+0.009
Shape$_{W93}$([O III])		**−0.693**	−0.169	−0.555	+0.022	−0.011
α_β([O III])		**+0.664**	+0.274	+0.551	−0.027	+0.097
Shift$_{BG92}$([O III])		+0.136	−0.198	+0.122	+0.272	+0.266
AI$_M$([O III])		**−0.531**	−0.328	−0.466	+0.054	−0.143
KrI$_M$([O III])		**−0.641**	−0.204	−0.583	+0.009	−0.128
total flux ([O III])	[erg s^{-1} cm^{-2}]	−0.112	−0.048	+0.130	+0.163	−0.133
Sk([O III])		−0.069	−0.208	−0.293	−0.042	−0.456
exc_{Kr}([O III])		−0.222	+0.005	+0.025	+0.027	+0.155
S([O III])		−0.333	−0.116	−0.446	−0.069	−0.412
IPV_{1090}([O III])		−0.006	+0.011	−0.164	−0.062	−0.410

in our sample different the ones from the others and, therefore, is meant to guide us in the choice of the quantities to use to correct the FWHM(C IV), which is itself a, less important, driver of EV1. The main point is that [O III] behaviour is driving the variance; we want to connect changes in this line to changes in C IV, with the ultimate goal of correcting C IV when the optical spectrum is not available.

PCA, at least in its basic form, does not take into account measurement errors, it only examines the variance in the system composed by the central values. This is why the lack of an estimate of measurement errors does not affect the analysis so far. Since we use this technique as a clue for the subsequent analysis and considering the errors is not going to change radically the main trends we found, we believe the errors are not crucial here. As for the lack of a correction for the resolution in our profiles, it corresponds to a certain amount of the variance introduced in the sample that can not be explained in terms of correlation among the properties. We believe, therefore, that the total variance of the sample will decrease after this correction, allowing a better explanation of the variance in terms of the first principal components.

In order to verify the presence of a connection between C IV and [O III] behaviours, we plot the main drivers of the variance in the sample, IPV_{60}([O III]) (representing the linewidth in general) and $A_{IPV_{20-80}}$([O III]) (representing the asymmetry), as a function of C IV related quantities (Fig. 4.8 and Fig. 4.9). In particular, we report C IV properties mostly describing the asymmetry, the shape and the shift of the line with respect to the nominal rest frame wavelength. The simple fact that C IV depends on [O III] behaviour is a confirmation of the existing connection between the region emitting [O III] and that emitting C IV. Some of the C IV quantities we are taking into account could mainly map the narrow component of the line (if present), such as the IPV_{10}(C IV) or the $\mathrm{EW}_{IPV_{10}}$(C IV), but many others, instead, map the whole line and therefore the broad component as well, as the $A_{IPV_{20-80}}$(C IV), Shape_{W93}(C IV) and EW(C IV). The presence of a narrow component could relate the peak of C IV to [O III] behaviour. There is instead no reason explaining the correlations of properties regarding the whole profile but a physical connection between BLR and NLR.

In examining Fig. 4.8 we notice that, for the most part, C IV properties-IPV_{60}([O III]) plots look similar in some respects: the points cover three quarters of the plot, leaving an empty quarter. Depending on the examined C IV property, peaky and boxy/blueshifted profiles occupy different quarters, but they are always separated the ones from the others. For example, in the plot for Shape_{W93}(C IV), peaky profiles occupy the lower right-hand quarter, corresponding to profiles with a small IPV_{60}([O III]), while boxy and blueshifted C IV profiles can be found in the higher left-hand quarter, corresponding to less intense but broader [O III] profiles. Looking at the plot for EW(C IV) we see that peakier C IV profiles correspond to higher EW(C IV) sources, while boxy profiles correspond to lower EW(C IV). Moreover, boxier profile are characterised by bluer asymmetries (IPV_{60}([O III]) vs $A_{IPV_{20-80}}$(C IV) and $A_{IPV_{40-60}}$(C IV) plots) and bluer shifts (Shift_{BG92}(C IV) and F_{c80}(C IV)), while peakier ones are usually less asymmetric and their shifts are moderate. The same analysis can be carried out

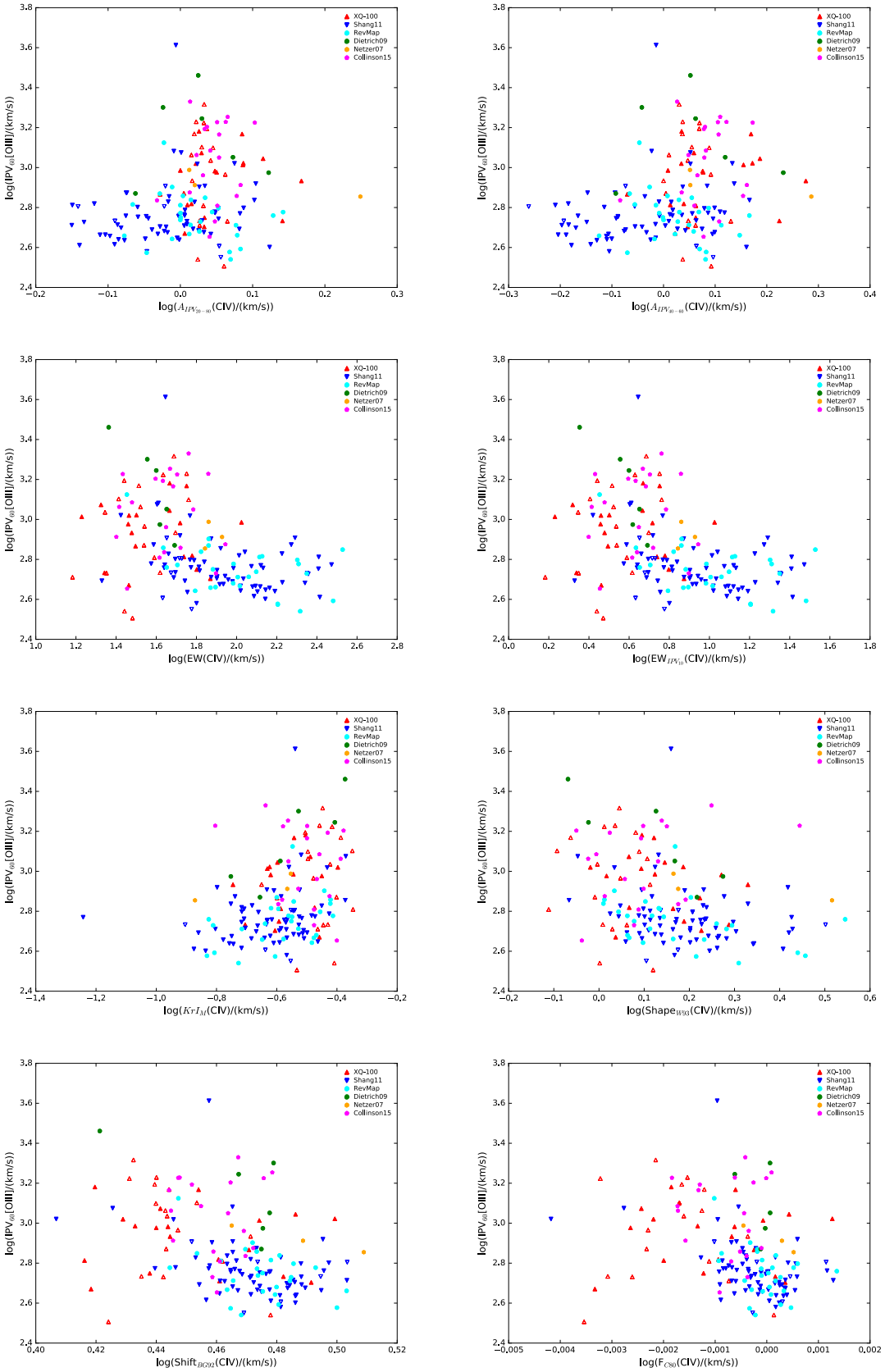


Figure 4.8: (1) $A_{IPV_{20-80}}$(C IV), (2) $A_{IPV_{40-60}}$(C IV), (3) EW(C IV), (4) EW$_{IPV_{10}}$(C IV), (5) KrI_M(C IV), (6) Shape$_{W93}$(C IV), (7) Shift$_{BG92}$(C IV) and (8) F_{c80}(C IV) as a function of IPV$_{60}$([O III]), the main driver of the eigenvector 1 (panels from top left to bottom right). Open symbols represent sources flagged as not completely reliable in the sample.

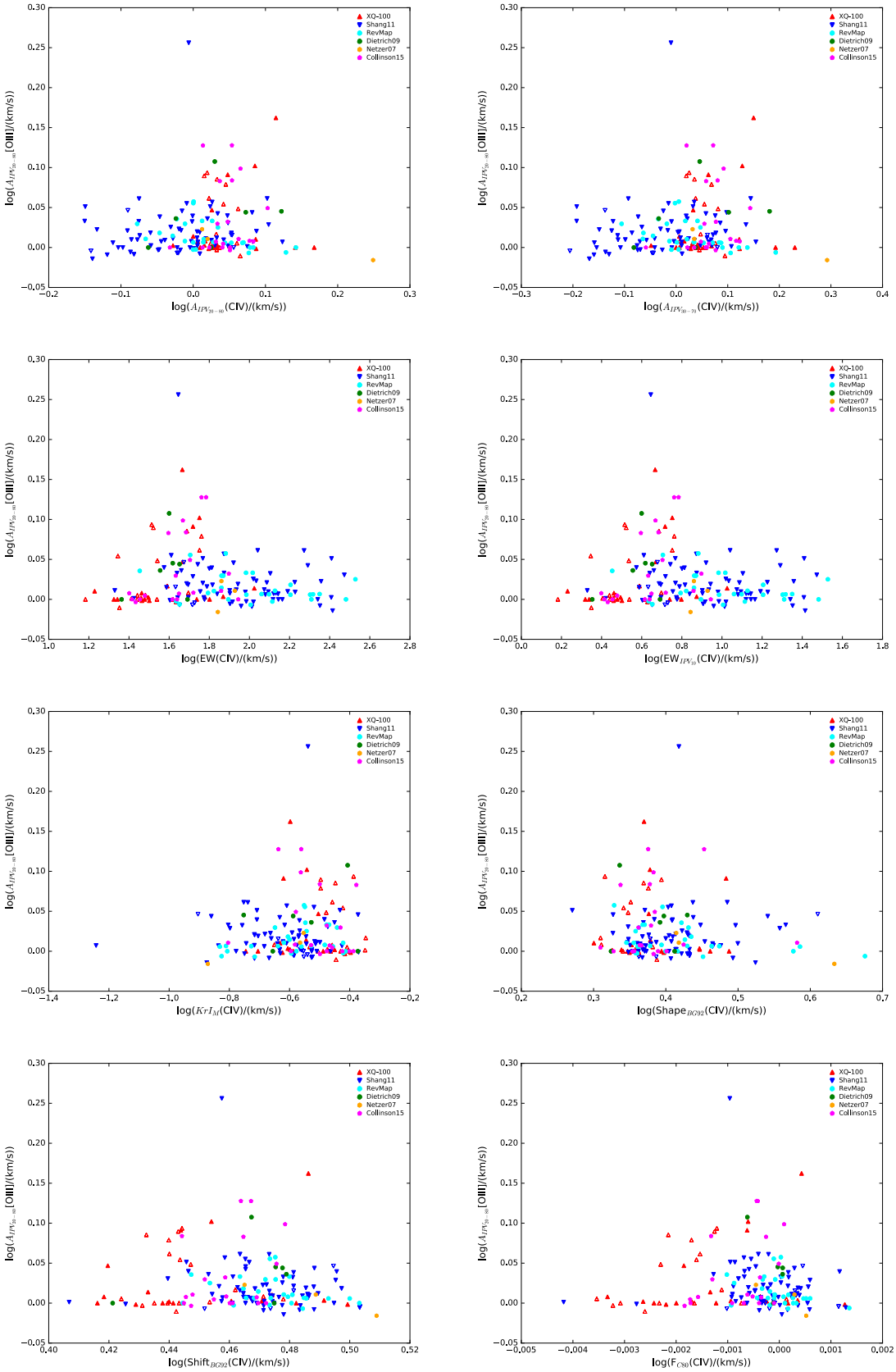


Figure 4.9: (1) $A_{IPV_{20-80}}$(C IV), (2) $A_{IPV_{30-70}}$(C IV), (3) EW(C IV), (4) EW$_{IPV_{10}}$(C IV), (5) KrI_M(C IV), (6) Shape$_{BG92}$(C IV), (7) exc_{Kr}(C IV) and (8) IPV$_{10}$(C IV) as a function of $A_{IPV_{20-80}}$([O III]), representing the asymmetry of [O III] line (panels from top left to bottom right). Open symbols represent sources flagged as not completely reliable in the sample.

for each one of the plots in which there is a dependence of the C IV profile property from the [O III] behaviour. Of course there are some quantities that do not show a particular sharp dependence from IPV_{60}([O III]); an example are all the IPVs(C IV). In the case of the correlations pertaining to $A_{IPV_{20-80}}$([O III]), the trend of C IV properties is less sharp than in the case of IPV_{60}(C IV), probably because this driver is less important in explaining the sample variance. Nonetheless we can notice a separation between the peaky and boxy C IV sources.

These findings are taken into account for the next step of the analysis.

FWHM(C IV) vs [O III] properties: Spearman correlation matrix

Besides the physical interpretation of the connection between C IV and [O III] emissions and therefore their emitting regions, what we are interested in is a validation of the virial mass estimation through C IV. As already mentioned, the quantity usually devoted to the line width measurement in viral mass estimation is FWHM. Our final goal is to correct FWHM(C IV) in order to get rid of its dependence on [O III]. We believe that the component of C IV connected to [O III] is the reason of the differentiation between C IV and Hβ, the iconic virial mass estimator.

In the PCA analysis carried out in our sample, FWHM(C IV) (along with the shift of the centroid with respect to the reference wavelength of the line and, to a lesser extent, with the EW(C IV), the IPVs regarding the top of the line and some asymmetric indexes) is the more important C IV quantity in the realisation of the eigenvector 1 (see Tab. 4.2), although significantly less important than the driving [O III] properties. We already showed in Fig. 4.8 and Fig. 4.9 the quantities more strongly dependent on the [O III] properties driving the eigenvector 1. The dependence of FWHM(C IV) on [O III] behaviour does not stand out as in those cases. Nonetheless there is evidence for correlations between this property and, mainly, characteristics representing the line width of [O III] and EW[O III] (Fig. 4.10). In particular, smaller FWHM(C IV) are mainly associated with smaller IPVs([O III]) and with higher EW([O III]). We want to examine more specifically the correlations between FWHM(C IV) and [O III] properties and, subsequentely, the correlations between FWHM(C IV) and other C IV properties, the ones we are eventually using for FWHM(C IV) correction.

PCA assumes correlations among properties to be linear. The correlation matrix we diagonalise in PCA to infer the eigenvectors corresponds to the Pearson correlation matrix, the Pearson correlation being a measure of the linearity between two variables. The Spearman correlation coefficient, unlike the Pearson correlation coefficient, gives a measure of the correlation between two variables in the hypothesis that their correlation can be described as a monotonic function. Just to check if we are overlooking something assuming linearity whitin our data-set, we also examined the Spearman correlation coefficients between the properties. This kind of analysis, as PCA, is very common in works looking for correlations among a set of properties. An important series of works examining correlations between spectroscopic properties, and involving C IV as well as [O III], is, for example, Baskin & Laor (2004, 2005).

We examined the complete Spearman correlation matrix and we found that the only [O III]

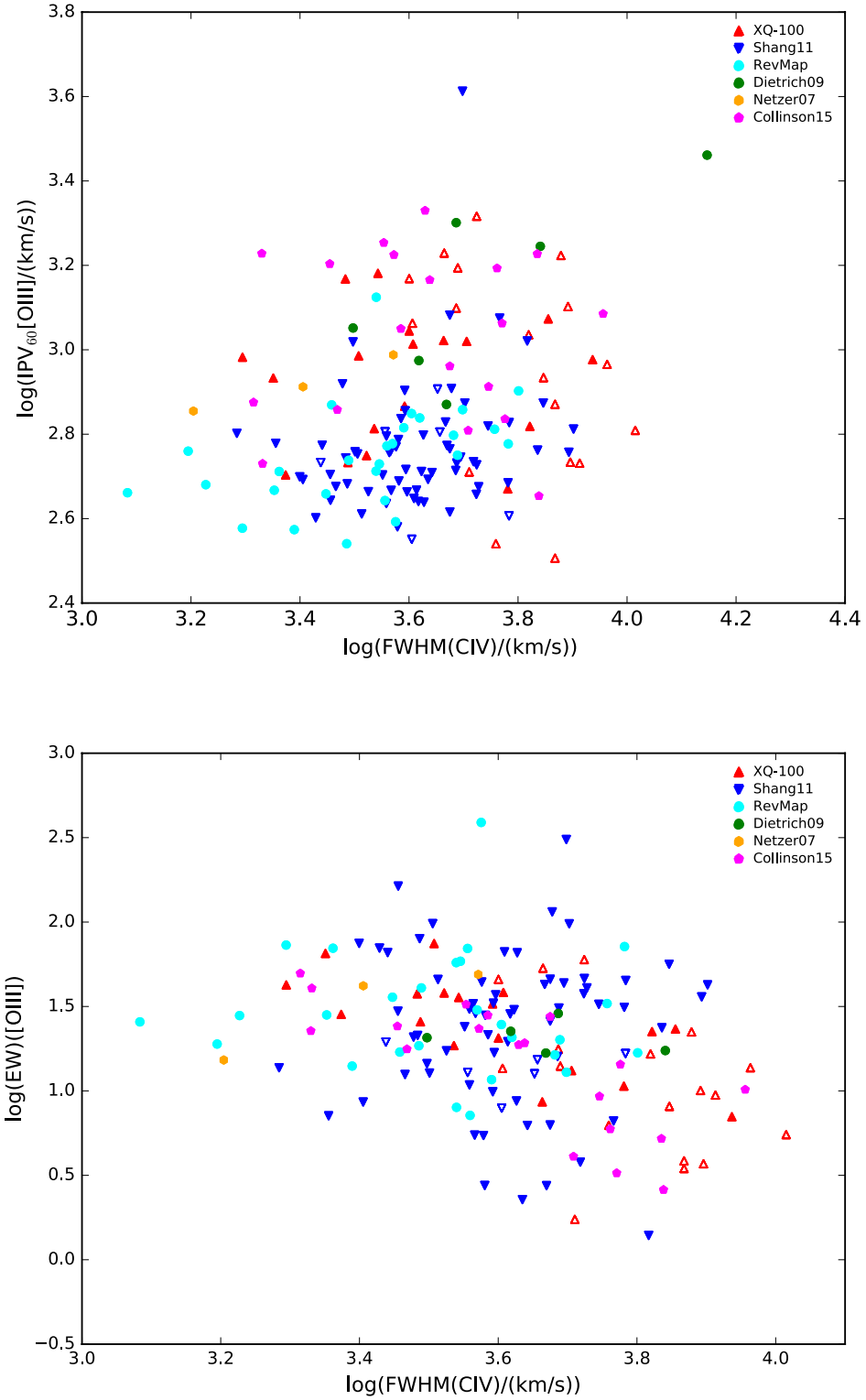


Figure 4.10: IPV60([O III]) and EW[O III] as a function of FWHM(C IV). The correlation between these properties is less strong than those presented in Fig. 4.8 and Fig. 4.9, nonetheless it is possible to recognize the presence of a trend. $r_S = 0.257$, $p{\sim}2\times10^{-3}$ and $r_S = 0.228$, $p{\sim}7\times10^{-3}$ are the Spearman correlation coefficients and the corresponding p-values for the IPV60([O III])-FWHM(C IV) and the EW[O III]-FWHM(C IV) relations, respectively.

Table 4.3: Spearman correlation coefficients matrix for the properties measured on the sample. We only report [O III] properties we found to be dominant in the EV1 from the previous PCA analysis, sorted by order of relevance. Correlations with other [O III] properties are mostly not significant. Values in bold are those for which the probability of the null hypothesis is $p < 1 \times 10^{-3}$.

	[O III]												C IV
	IPV_{60}	IPV_{50}	IPV_{70}	IPV_{40}	IPV_{30}	MAD	IPV_{80}	IPV_{20}	IPV_{10}	σ	IPV_{90}	FWHM	FWHM
C IV													
FWHM	0.257	0.252	0.265	0.251	0.250	**0.292**	0.276	0.250	0.250	0.277	**0.279**	0.206	1.000
σ	0.145	0.133	0.156	0.128	0.126	0.179	0.174	0.121	0.121	0.184	0.187	0.044	**0.478**
centroid shift	**−0.410**	**−0.412**	**−0.404**	**−0.414**	**−0.414**	**−0.394**	**−0.400**	**−0.416**	**−0.415**	**−0.372**	**−0.376**	**−0.410**	**−0.375**
IPV_{10}	0.216	0.207	0.228	0.204	0.203	0.250	0.243	0.199	0.199	0.240	0.248	0.133	**0.846**
IPV_{20}	0.198	0.189	0.210	0.186	0.184	0.232	0.225	0.181	0.181	0.222	0.231	0.116	**0.818**
IPV_{30}	0.176	0.166	0.187	0.163	0.162	0.209	0.202	0.158	0.158	0.200	0.210	0.097	**0.776**
IPV_{40}	0.143	0.133	0.154	0.130	0.128	0.175	0.168	0.123	0.124	0.168	0.178	0.063	**0.720**
IPV_{50}	0.116	0.106	0.127	0.103	0.100	0.147	0.141	0.095	0.096	0.142	0.152	0.036	**0.668**
IPV_{60}	0.096	0.085	0.106	0.082	0.079	0.127	0.121	0.074	0.075	0.125	0.134	0.013	**0.603**
IPV_{70}	0.098	0.087	0.109	0.083	0.081	0.131	0.125	0.075	0.076	0.132	0.138	0.011	**0.559**
IPV_{80}	0.115	0.103	0.126	0.098	0.095	0.148	0.142	0.090	0.091	0.152	0.156	0.021	**0.522**
IPV_{90}	0.138	0.126	0.149	0.121	0.119	0.173	0.167	0.113	0.114	0.178	0.181	0.037	**0.474**
$AI_{05\text{-}95}$	**0.368**	**0.369**	**0.366**	**0.369**	**0.367**	**0.363**	**0.362**	**0.366**	**0.363**	**0.361**	**0.371**	**0.320**	−0.046
$AI_{10\text{-}90}$	**0.284**	**0.289**	0.276	**0.288**	**0.291**	0.266	0.265	**0.288**	**0.285**	0.258	0.257	0.249	−0.187
$AI_{20\text{-}80}$	**0.306**	**0.311**	**0.297**	**0.310**	**0.312**	**0.291**	**0.287**	**0.309**	**0.306**	**0.283**	**0.284**	0.268	−0.185
$AI_{30\text{-}70}$	**0.314**	**0.316**	**0.306**	**0.315**	**0.316**	**0.302**	**0.298**	**0.313**	**0.309**	**0.299**	**0.299**	0.268	−0.190
$AI_{40\text{-}60}$	**0.356**	**0.358**	**0.349**	**0.358**	**0.359**	**0.338**	**0.339**	**0.357**	**0.352**	**0.333**	**0.335**	**0.311**	−0.188
MAD	0.138	0.126	0.149	0.122	0.119	0.172	0.166	0.114	0.114	0.174	0.179	0.040	**0.544**
EW (C IV)	**−0.558**	**−0.569**	**−0.541**	**−0.573**	**−0.579**	**−0.516**	**−0.521**	**−0.580**	**−0.582**	**−0.472**	**−0.484**	**−0.595**	−0.211
AS	−0.252	−0.257	−0.246	−0.256	−0.260	−0.239	−0.235	−0.259	−0.256	−0.223	−0.225	−0.222	−0.113
$shape_{BG92}$	**−0.291**	**−0.291**	**−0.293**	**−0.289**	**−0.288**	**−0.312**	**−0.300**	**−0.292**	**−0.289**	**−0.305**	**−0.304**	−0.262	**−0.457**
$shape_{W93}$	**0.331**	**0.331**	**0.329**	**0.328**	**0.328**	**0.358**	**0.338**	**0.332**	**0.329**	**0.360**	**0.351**	**0.312**	**0.379**
α_β	**0.319**	**0.322**	**0.311**	**0.323**	**0.323**	**0.289**	**0.295**	**0.323**	**0.319**	**0.284**	**0.279**	**0.287**	−0.182
$shift_{BG92}$	**−0.396**	**−0.394**	**−0.391**	**−0.395**	**−0.395**	**−0.393**	**−0.394**	**−0.396**	**−0.394**	**−0.380**	**−0.382**	**−0.377**	**−0.283**
$c(9/10)$	**−0.404**	**−0.405**	**−0.398**	**−0.408**	**−0.408**	**−0.389**	**−0.394**	**−0.411**	**−0.410**	**−0.368**	**−0.372**	**−0.407**	**−0.362**
AI_M	**−0.388**	**−0.390**	**−0.381**	**−0.390**	**−0.391**	**−0.369**	**−0.370**	**−0.390**	**−0.386**	**−0.361**	**−0.358**	**−0.344**	0.040
KrI_M	**0.326**	**0.327**	**0.321**	**0.327**	**0.327**	**0.353**	**0.329**	**0.332**	**0.330**	**0.361**	**0.348**	**0.327**	0.276
total flux	−0.135	−0.141	−0.131	−0.144	−0.143	−0.117	−0.121	−0.145	−0.144	−0.086	−0.100	−0.136	0.084
Sk	**−0.301**	**−0.301**	**−0.301**	**−0.302**	**−0.299**	**−0.305**	**−0.302**	**−0.298**	**−0.296**	**−0.304**	**−0.316**	−0.255	−0.011
exc_{Kr}	0.108	0.110	0.102	0.110	0.113	0.111	0.105	0.114	0.114	0.118	0.109	0.100	−0.237
S	0.213	0.218	0.209	0.221	0.221	0.214	0.205	0.225	0.227	0.187	0.193	0.250	**0.630**
IPV_{1090}	0.196	0.201	0.193	0.203	0.204	0.192	0.190	0.207	0.207	0.165	0.175	0.212	**0.621**
$EW(IPV_{10})$	**−0.560**	**−0.570**	**−0.542**	**−0.575**	**−0.580**	**−0.517**	**−0.522**	**−0.582**	**−0.583**	**−0.473**	**−0.485**	**−0.597**	−0.213

properties whose correlations with C IV are significant are those involving [O III] line width. We therefore report the correlation coefficients between all the C IV properties and these [O III] quantities only (see Tab. 4.3). We report correlations coefficients between FWHM(C IV) and all the other C IV properties as well, since, as said, these are the quantities we are eventually using to correct FWHM(C IV). What we are trying to do is to find a connection between [O III] correlations and C IV correlations in order to use the second ones to correct for a component we believe is physically connected to the first ones. In Tab. 4.3, values in bold are those for which the null hypothesis probability (p-value) is less than $p < 1 \times 10^{-3}$. Of course there are obvious correlations, such as the ones between quantities representing similar characteristics of the line (for example IPV_{40}(C IV) versus IPV_{50}(C IV) or F_{c80}(C IV) versus $c(9/10)$(C IV)), but other interesting correlations are present.

Given the purpose of this work, the most interesting correlations are those involving both C IV and [O III] properties. Specifically, there are two C IV properties relating in the same way to FWHM(C IV) and to [O III] line width:

- the centroid shift with respect to the nominal rest frame wavelength
- the shape of the line.

In Fig. 4.11 we show the two correlations with the C IV centroid shift, represented by $F_{c80} = \mathrm{Shift}_{c80}/\lambda_0$, and the shape of the line represented by Shape_{BG92}[2].

It is not surprising that these very two properties stand out from the others. These quantities are in fact related to the peakiness or boxiness (shape) and to the blueness of the line (centroid shift). They represent the specific behaviour of C IV we recognised in the first place as connected to [O III].

4.5.4 FWHM(C IV) correction through penalized regression methods

We want to use the identified connection between C IV and [O III] to correct FWHM(C IV). Correcting FWHM(C IV) means, in this specific case, to find a way to make FWHM(Hβ)-

[2]When evaluating correlations among characteristics, we move to a log-log representation of the relationships. Since many of the involved C IV characteristics are not positive, we need to translate them before log tranforming, i.e. we need to add a common constant value to all the observations for a given quantity in order to obtain definite positive arguments for all the observations. The choice of the constant is specific for every quantity, depending on the distribution of the observations for the examined characteristic and on its unit of measurement. Specifically we are performing the following translations:

- $\log \mathrm{Shift}_{BG92} \longrightarrow \log(\mathrm{Shift}_{BG92} + 3)$,
- $\log \alpha_\beta \longrightarrow \log(\alpha_\beta + 2)$,
- $\log c(9/10) \longrightarrow \log(c(9/10) + 5000)$,
- $\log AI_M \longrightarrow \log(AI_M + 1)$,
- $\log Sk \longrightarrow \log(Sk + 2)$,
- $\log exc_{Kr} \longrightarrow \log(exc_{Kr} + 1)$.

Unless otherwise specified, log transforming on characteristics which are not listed above is performed on the original quantity.

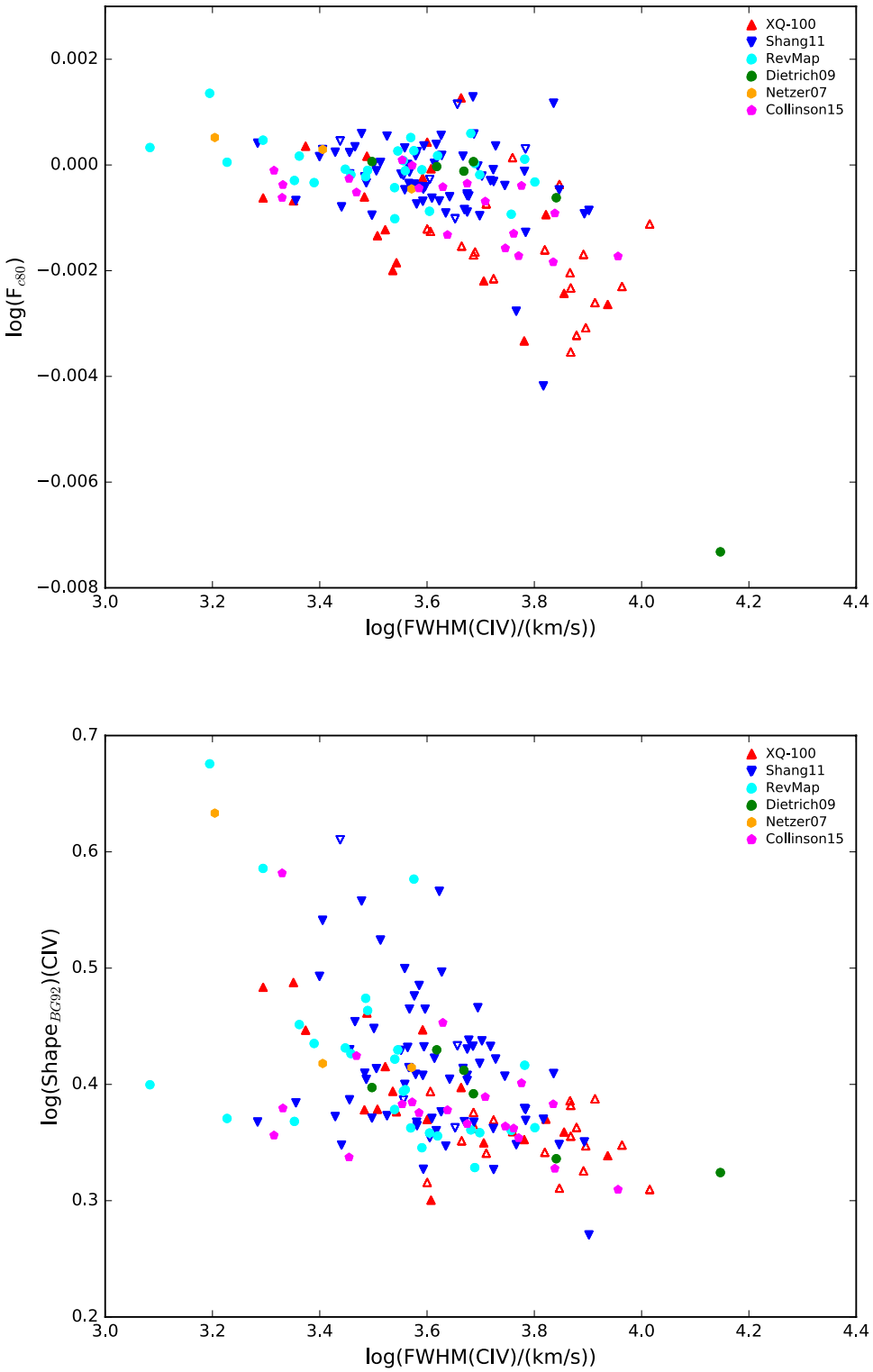


Figure 4.11: Correlations between FWHM(C IV) and C IV properties connected to [O III] characteristics, i.e. the centroid shift with respect the nominal wavelength and the shape of the line. Spearman correlation coefficients for these relationships are listed in Tab. 4.3.

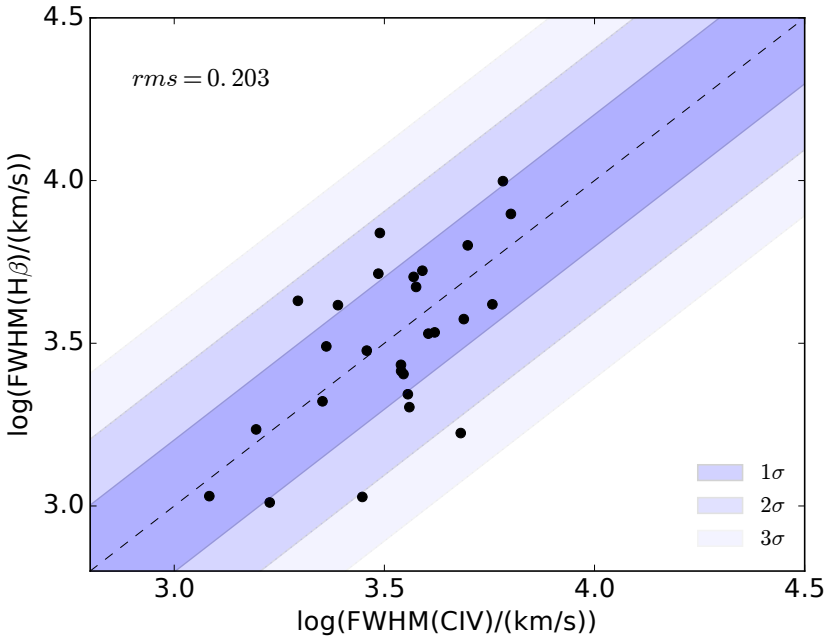


Figure 4.12: log(FWHM(Hβ))-log(FWHM(C IV)) relation for the objects composing the reverberation mapping sample.

FWHM(C IV) correlation stronger, both in the Pearson correlation coefficient and in the rms of the data points with respect to the 1:1 relationships between the two quantities. This is achieved multiplying FWHM(C IV) by a factor that is a function of those among the C IV properties we found to be correlated with [O III] behaviour. This kind of approach assumes of course Hβ to be the benchmark of virial estimation. We already mentioned how for this line, unlike for C IV, single-epoch and rms spectra are very similar, making this line more reliable for virial mass estimation even when a single measurement is available. Nonetheless reverberation mapping measurements are in general more robust for mass measurements and therefore we would like to attempt a first correction relying on this kind of measurements only. Our first attempt, then, will involve only sources for which a reverberation mapping measurement for Hβ is available. This will narrow down the sample considerably, since the RM sample is composed by 27 objects only, but this should also be the most reliable subsample for our purpose. Of course the correlations we found in the whole sample are not the same we would have found in this specific subsample. However, we are looking for a correction for FWHM(C IV) that is suitable for all the sources, representing all the aspects of the connection between the C IV and [O III] profiles. Therefore, we are considering the whole sample correlations on this subsample as well. The starting point is the initial relation between FWHM(Hβ) and FWHM(C IV) for this subsample, for which FWHM(Hβ) measurements are retrieved from *rms* spectra (Fig. 4.12).

The analysis to be performed is a of multiple linear regression in which the response (or predicted) variable is log(FWHM(Hβ)) and the predictor variables are the quantities related to C IV (their logarithm, specifically, since we want the problem to be linear), first among

them log(FWHM(C IV)). The problem consists therefore in finding the α, β, γ, δ etc. for which

$$\log \mathrm{FWHM(H}\beta) = \alpha \log \mathrm{FWHM(C\,IV)} + \beta\, f_1 + \gamma\, f_2 + \delta\, \ldots = \log \mathrm{FWHM(C\,IV)}_{corr} \quad (4.18)$$

is the tightest possible correlation and where f_i are the C IV properties we recognized as connected to [O III] behaviour. In such a case the prediction of log(FWHM(Hβ)) would be a linear combination of the independent predictor variables. In particular, since we are dealing with several C IV properties and we would like to select those reducing the rms of the relation - at the same time increasing the Pearson correlation coefficient - a stepwise multiple linear regression could seem the more suitable choice. Such an analysis selects the predictor variable to be chosen as the next entry in the linear combination by examining which one of them reduces most of the scatter in the relationship (FWHM(Hβ)-FWHM(C IV)$_{corr}$).

We are, however, facing a more complicated problem. In order to be correct, a stepwise variable selection requires the predictor variables to be independent. If this condition is not fulfilled, stepwise selection is susceptible to give highly uncorrect results and should therefore not be trusted. This is indeed true for all kind of Ordinary Least Squares (OLS): in presence of correlation among predictor variables, these analyses can produce both the wrong magnitude and very high statistical errors for the estimated regression coefficients, especially in the cases where the number of observations n is smaller than, or comparable with, the number of predictors p. Moreover, the result of the regression may change drastically with a small change in the sample, making the solution not stable (Greene, 2003). This problem is referred to as the *multicollinearity* among predictor variables and, to be addressed, a different kind of regression should be used.

Penalized regression methods were developed with the aim of reducing the problem of collinearity among predictor variables and to enhance both the prediction accuracy and the interpretation of the model with respect to classic multiple regressions. These methods, also called *Shrinkage Methods*, shrink the regression coefficients applying a penalization to their size. Doing that, they improve the interpretability of the model, because in presence of a large number of predictors they shrink their coefficients to zero, leaving only the most important ones in the model, operating therefore a variable selection. Among the penalized regression techniques, the most famous is probably the Lasso (Least Absolute Shrinkage and Selection Operator, Tibshirani (1996)), which has the advantage of assigning a null coefficient to predictor variables that are found not to be significant in the prediction of the response variable (unlike the ridge regression (Hoerl & Kennard, 1970) for example, which shrinks them but does not make them exactly equal to zero). While OLS regression methods simply minimize the residuals between data and model, the Lasso minimizes the expression

$$\hat{\beta}^{lasso} = \underset{\beta \in \mathbb{R}^p}{\operatorname{argmin}} \frac{1}{N} \sum_{i=1}^{N} \left(y_i - \sum_{j=1}^{p} x_{ij} b_j \right)^2 \quad \text{subject to} \quad \sum_{j=1}^{p} \mid b_j \mid \leq t \,, \quad (4.19)$$

which is equivalent to the optimization problem

$$\hat{\beta}^{lasso} = \underset{\beta \in \mathbb{R}^p}{\operatorname{argmin}} \left[\frac{1}{N} \sum_{i=1}^{N} \left(y_i - \sum_{j=1}^{p} x_{ij} b_j \right)^2 + \lambda \sum_{j=1}^{p} \mid b_j \mid \right] \tag{4.20}$$

where the first term represents the residuals between data and model and λ is the regularization parameter, controlling the penalization term given by the ℓ_1 norm of the coefficients. For this reason, the Lasso is referred to as a ℓ_1 penalized regression method. The introduction of the penalization term is what prevents overfitting the data and what makes the solution stable, then making these methods more reliable than OLS linear regression in presence of multicollinearity.

The Lasso is definitely an improvement with respect to stepwise multiple linear regression, because it is able to perform a variable selection while addressing the problem of multicollinearity. There is, however, even a more efficient regularized method that can be used when in presence of strong multicollinearity of the predictors: the Elastic Net (Zou & Hastie, 2005). The Elastic Net is a penalized method that minimizes the following expression:

$$\hat{\beta}^{enet} = \underset{\beta \in \mathbb{R}^p}{\operatorname{argmin}} \left[\frac{1}{N} \sum_{i=1}^{N} \left(y_i - \sum_{j=1}^{p} x_{ij} b_j \right)^2 + \lambda_1 \sum_{j=1}^{p} \mid b_j \mid + \lambda_2 \sum_{j=1}^{p} b_j^2 \right] . \tag{4.21}$$

Elastic Net is at the same time a ℓ_1 and a ℓ_2 penalization regression, i.e. it performs the regression both penalizing the ℓ_1 norm and the ℓ_2 norm of the coefficients. Since we are trying to select a significant subset from a set of variables containing quantities that describe similar characteristics of the C IV line profile, such as for example the full set of asymmetry indexes, realistically very collinear with each other, this technique seems to be more suitable for our specific case. Furthermore, this method typically applies a less strong shrinkage to the coefficients with respect to the Lasso, leaving, on average, a higher number of predictors in the linear combination predicting the response variable and therefore maintaining a more accurate description of the problem.

In all penalized regression techniques the choice of the regularization parameter λ (or parameters, λ_1 and λ_2), controlling the trade-off between the penalty and the fit, is crucial to the analysis response. The wrong choice could lead to a massive shrinkage of all the variable, giving a null coefficient for all of them, or it could not apply a shrinkage at all, leaving the same number of predictor variables in the problem, not operating a selection. For this reason the choice of the "correct" value for λ is left to a model validation technique, the *k-fold Cross Validation* (CV), that splits the sample in a train and a test subsample, the first one on which the analysis is performed and the second one on which such an analysis is validated. This process is repeated k times, in order to use every part of the sample at least once as the test subsample and of course several time as part of the train subsample.

We performed the Elastic Net with Cross Validation using the *ElasticNetCV* function provided by the Python *sci-kit learn* package (Pedregosa et al., 2011) on our data-set, composed by the response variable, log(FWHM(Hβ)), and the predictor variables (all C IV properties we decided to consider, given their relation with [O III]).

Table 4.4: Regression coefficients estimated through Elastic Net. The only variables contributing to the prediction of the response variable are log(FWHM(C IV)), log($c(9/10)$) and log(EW(IPV_{10})). The statistical uncertainties are estimated with a 10000 jackknife leave-one-out resampling. As mentioned in the text, statistical uncertainties are a mere indication in this case and should not be considered as a conventional statistical error. Penalized regression methods in fact introduce strong biases while determining which variables has to be retained in the model prediction. These biases affect by definition the possibility to give an accurate estimate of statistical errors. This fact can be recognized in the utter underestimation of the errors related to coefficients that are shrunk to zero by the penalization process. For this reason we do not report uncertainties for coefficients shrunk to zero by the analysis.

predictor	Elastic Net regression coefficient
log(FWHM(C IV))	0.52 ± 0.26
$\log(F_{c_{80}})$	...
$\log(A_{IPV_{05-95}})$	0.04 ± 0.12
$\log(A_{IPV_{10-90}})$	...
$\log(A_{IPV_{20-80}})$	...
$\log(A_{IPV_{30-70}})$	...
$\log(A_{IPV_{40-60}})$	...
log(EW(C IV))	...
$\log(\text{Shape}_{BG92})$	...
$\log(\text{Shape}_{W93})$	...
$\log(\alpha_{beta})$	...
$\log(\text{Shift}_{BG92})$	...
$\log(c(9/10))$	-0.65 ± 0.29
$\log(AI_M)$	...
$\log(KrI_M)$	...
$\log(Sk)$	...
log(EW(IPV_{10}))	0.29 ± 0.13

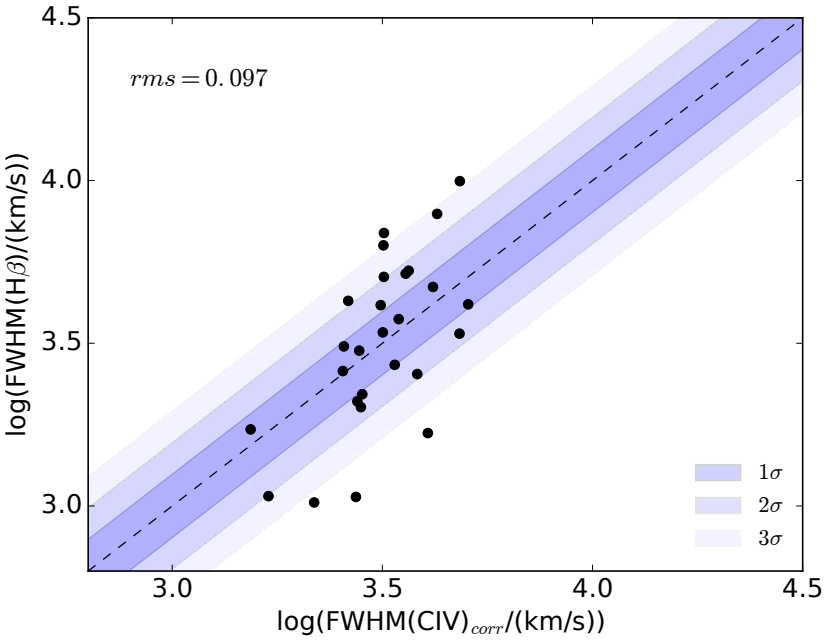


Figure 4.13: Elastic Net prediction for the response variable log(FWHM(Hβ)): the linear combination for log(FWHM(C IV)$_{corr}$) is obtained assigning the coefficients in Tab. 4.4 to the predictor variables. For this new relationship between log(FWHM(Hβ)) and log(FWHM(C IV)$_{corr}$) the scatter is $\sim$ 0.1 dex, to be compared to the $\sim$ 0.2 dex of the initial relation (Fig. 4.12).

The regression coefficients yielded by this technique are reported in Tab. 4.4. To give an evaluation of the uncertainties to be associated with our estimate of the regression coefficients, we performed the same analysis on 10000 jackknife leave-one-out data sets resampled from the original data and determined the standard deviation of the distribution obtained for each coefficient. We notice that, even if the uncertainties given by the standard deviation are very high, the mean of the distribution for each coefficient is always close to the value estimated on the true sample. It should be mentioned, however, that an estimate of statistical uncertainties on the results of a penalized regression are far from accurate, given the bias introduced *by definition* in the shrinkage process. This is why all available statistic packages do not provide an error estimation on yielded regression coefficient (see for example Goeman et al. (2016)). This well known issue can be recognised in the utter underestimation of statistical errors of the regression coefficients relating to variables that are shrunk to zero by the penalization process (Tab. 4.4). Statistical uncertainties given in Tab. 4.4 are therefore to be considered as a mere indication.

The characteristics that are not shrunk to zero by ElasticNet, and therefore the only ones to be considered in the prediction of log(FWHM(Hβ)), are log(FWHM(C IV)), log(c(9/10)) and log(EW$_{IPV_{10}}$). This result is, once again, not unexpected: these are in fact the characteristics that can be associated with the behaviour of C IV that we found to be connected with [O III] in the first place. The relationship between log(FWHM(Hβ)) and its model prediction

$$\log \mathrm{FWHM(C\,IV)}_{corr} = 0.52 \log \mathrm{FWHM(C\,IV)} - 0.65 \log c(9/10) + 0.29 \log \mathrm{EW}_{IPV_{10}} \quad (4.22)$$

is shown in Fig. 4.13. The correction decreases the rms from ∼0.2 dex to ∼0.1 dex. The increase in the Pearson correlation coefficient is less significant: the new value is $r = 0.627$ (compared to the previous $r = 0.619$) with a p-value of $p = 4.6 \times 10^{-4}$.

It should be noticed that this step of the analysis is the one where the lack of errors is substantial. Considering the errors on all the involved variables will allow us, for example, to weight differently measurements more or less reliable and will probably improve the correction. Furthermore, we have so far examined only the RM sample, which is considered the most reliable for Hβ mass estimation, but, on the other hand, is a small subsample and could be biased towards specific C IV behaviour, being not representative of all the connections between this line and [O III]. In Fig. 4.8, 4.9, 4.10 and 4.11 indeed, RM sources (in cyan) occupy the part of the plot relating to objects with an intense, narrow [O III], corresponding to peaky and symmetric C IV profiles. We expect that the examination of the whole sample would give us more results, being less biased towards a specific shape of the C IV profile.

4.6 Summary

We examined the reciprocal behaviour of the C IV and [O III] lines. These lines, although coming from different emitting regions, show a connection that can be interpreted in a mutual framework. In particular we notice that when [O III] is intense, C IV shows a peaky and symmetric profile, while when [O III] is weak or even absent, C IV is characterised by an asymmetric, boxy and blueshifted profile. We suggest that this connection can be due to the presence of a wind or outflow in the inner part of the BLR, the closest to the accretion disk, able, in the extreme cases, even to prevent the formation of the NLR. C IV is the only broad line showing such a peculiar behaviour and the one emitted closest to the accretion disk.

In order to understand this connection we examined both the line profiles, taking care of giving an accurate representation of the characteristics of both the lines, such as the shape, shifts with respect the reference wavelenght and general asymmetry. We then performed an analysis on the data-set composed by all the measured properties for both the lines, with the final aim to find a correction for FWHM(C IV) to allow us the use of this line as a valid replacement when the virial estimator of choice (Hβ) is not available, due to the redshift of the source. We believe that the unreliability of C IV as a virial estimator is due to the presence of a contaminating, non-reverberating component whose presence can be traced through [O III] behaviour.

The analysis gives us the following results:

1. The Principal Component Analysis reveals that the eigenvector 1 of the sample (therefore a large part of its variance) is driven by [O III], in particular by properties connected with the width and asymmetry of the line.

2. Some C IV properties (among them FWHM(C IV)) are, to a far lesser extent, drivers as well of eigenvector 1, but this line is instead the main driver of eigenvector 2.

3. We also examine non-linear correlations among C IV and [O III] properties. This allow us to verify that the assumption of linearity (PCA) did not affect our analysis and to have a closer look to the connections between the properties for the two lines. We find that several C IV properties show a trend both with [O III] driving properties and FWHM(C IV). These are the quantities we use to correct FWHM(C IV).

4. We examine so far only the RM sample, the most reliable in terms of Hβ virial masses. We are able through this correction to reduce the scatter in the relation and to slightly increase the Pearson correlation coefficient between the FWHM(Hβ) and FWHM(C IV)$_{corr}$.

This project is still a work in progress. We believe that the significance of the correction will be improved once the analysis on the whole sample is complete. Despite its reliability for Hβ virial masses, the RM sample appears to be biased towards objects with intense [O III] and peaky and symmetric C IV, being not representative of the wider range of connections between these two lines that are instead present in the whole sample. Moreover, along with the correction of FWHM(C IV), the comprehension of the physical connection of the regions emitting the two lines is the main goal of this work. This aspect needs to be investigated.

We have now identified the statistical methodology to be used in the investigation of the connection between C IV and [O III] line profiles. This set-up can be applied not only to our sample, but to any other sample for which both C IV and [O III] observations are available.

Conclusions

The identification of a commom paradigm, the *Unified Model*, allows us to recognise all the "zoology" of active galaxies as belonging to the same family of objects, the *Active Galactic Nuclei*. According to this model, the differentiation among the specific classes of objects is mainly ascribed to the orientation with respect to the line of sight of the observer. The mechanism responsible for the energy production is the accretion of gas onto a central *SuperMassive Black Hole*, which is achieved through an accretion disk, the primary source of the AGN. The radiation produced by the accretion disk irradiates the sourrounding regions, causing their emission, whose properties and characteristics must be greatly taken into account for the comprehension of AGN. These sources are indeed spatially unresolved and the spectroscopic analysis is the only available tool to have an insight of the inner regions.

Among the components described in the Unified Model, the BLR playes a key role: this region is the closest to the accretion disk, and an analysis of its emission lines provides information on the mass of the SMBH. This estimation, important in itself, becomes crucial when we consider two facts: (1) AGN activity peaks at z∼2-3 and it is likely to be a common phase of the evolution of all galaxies, (2) the SMBH masses relate with host galaxies properties. This implies that there is a co-evolution of SMBH growth and host galaxies and therefore that a study of AGN is essential to a deeper comprehension of how the structures formed and evolved in the universe.

Based on these considerations, this thesis work mainly focuses on the study of the profiles of broad line components in quasar spectra. Since emissions coming from the nuclear regions are the only available information on what physical processes are operating in these sources, it makes sense to study them mutually, i.e. not looking at the emission from a specific region indipendently, but trying to involve emissions relating to other components of the Unified Model as well.

In particular, in Chapter 2 we examined broad and narrow emission lines in quasars by making use of an orientation indicator, identified in the equivalent width of the most prominent narrow line in the optical range of quasar spectra, [O III]. According to the definition of the indicator, sources with a low $EW_{[O\,III]}$ correspond mostly to objects in a face-on position, while sources with a high $EW_{[O\,III]}$ to objects in an edge-on position. This finding allowed us to split a large sample of quasars (∼12,000 objects) in bins of $EW_{[O\,III]}$ (corresponding to bins of inclination) and to produce representative *stacked* spectra for each bin in order to study the spectral features. Broad lines stacked profiles show an increase in the linewidth

moving from low to high $EW_{[O\,III]}$, confirming that the BLR shares the same anysotropy as the continuum source, i.e. it is disk-shaped. As for the NLR, [O III] line shows a decrease in the blue component, ascribed to the presence of outflows in the NLR, moving from low to high $EW_{[O\,III]}$. These outflows are in fact known to be preferentially axial, i.e. directed perpendicularly to the plane of the accretion disk. The increase in the linewidth of the broad components and the decrease of the [O III] blue component can not be explained with any other simple scenario but the change in inclination of the source with respect ot the line of sight, and are therefore a confirmation of the goodness of $EW_{[O\,III]}$ as an orientation indicator. Moreover, the increase in the number of double-peaked objects in our sample when moving towards high $EW_{[O\,III]}$ further confirms our hypothesis. Finally, the well known *Eigenvector 1* explaining most of the variance in quasar spectra, i.e. the anticorrelation between the intensities of [O III] and Fe II, can be, at least partially, explained as an orientation effect. Sources with a weak Fe II and strong [O III] emissions can be quite simply interpreted as edge-on sources, while strong Fe II - weak [O III] sources as face-on sources, provided that other intrinsic properties can contribuite to the setup of such an anticorrelation, e.g. the rate at which the SMBH is accreting.

The identification of an orientation indicator could be a cornerstone in AGN studies. It allows the possibility to study quasar spectra in a brand new way, having knowledge of which inner components we are looking at and which of them are instead prevented from being intercepted by the presence of the torus. For this reason, we have derived the relation describing the distribution of sources in terms of inclination at a given $EW_{[O\,III]}$, able to give information on the orientation of the source once a measurement of $EW_{[O\,III]}$ is available. Furthermore, the TOR itself can be studied by making use of $EW_{[O\,III]}$ indicator. The geometry and covering factor of this component, usually depicted as a toroidal and continuous structure, are hardly known so far, and several evidences and recent modelizations point towards a different, clumpy representation of the structure. The analysis of the infrared emission of AGN carried out in this thesis confirms the anisotropic geometry of the TOR component. Moving from low to high $EW_{[O\,III]}$ we see a decrement of the IR emission at the shorter wavelengths, which is expected in clumpy models.

In the second part of this thesis we examined in more detail the problems related with virial mass estimations. We discussed in the first chapter how orientation effects in quasar spectra can affect SMBH measurements, if we do not take them into account when measuring linewidths. This is true for all broad components, including Hβ, the line of choice in virial estimations. There are however intrinsic characteristics that make some broad components less reliable than others in the context of virial estimation. In using broad lines for virial mass estimation, we are assuming them all to be emitted from gas in a virialised condition. While this is true for Hβ and, tipically, for Hα and Mg II, it is instead controversials for C IV, whose profile is characterised by strong shifts and asymmetries ascribed to the presence of outflows and winds in the emitting region. This fact casts some doubts on the reliability of C IV as a virial estimator. This line, however, is the most readily available for sources at

redshift higher than z∼1.5 and can be the only choice in case of very distant sources ($z \gtrsim 2$). A comparison of C IV with more reliable lines (i.e. Hβ), aimed to the rehabilitation of this line for virial mass estimations, is therefore needed.

In chapter 3 we addressed the problem through the analysis of six sources at z∼2.2, for which a simultaneous detection of all the UV and optical broad lines commonly used to give a virial estimation of SMBH mass is available thanks to X-shooter, the spectrograph at the VLT. The fast variability of AGN is one of the issues concerning the comparison between different broad lines, whose observations are tipically performed at different epochs, often very distant in time. Also, quasar variability is more important at shorter wavelengths and for higher ionization lines and can therefore affect differentially optical and UV lines. The simultaneity of the observations ensures that these additional problems are not present, allowing a comparison of the intrinsic differences between the lines.

The analysis conducted on the broad components of the six sources confirmed that Hβ, Hα and Mg II have a similar behaviour, as expected from lines emitted by gas, at least mostly, virialised. We then analysed C IV and had a confirmation of the peculiar behaviour of the line. For almost all the examined sources, this line presents a strong blueshift and asymmetry. As a result, the comparison of C IV linewidths with those of Hβ and of the other lines usually shows a larger scatter. This is somehow expected: previous studies pointed out the presence of a non-reverberating component in the C IV line, that contaminate the line profile and, therefore, linewidth measurements. Such a component is of course not removed from the line even if we are taking simultanous detection. We examined C III] λ1908.73 as well. Despite the fact that it is less intense than C IV and a part of a complex of emission lines, the behaviour of this line seems to be in agreement with other reliable lines. For this reason, we proposed a new prescription for virial masses based on C III] and suggested to use this line instead of C IV when only this spectral window is available. Of course this is an indication, since our sample is composed by six sources only. Nontheless this is a possibility that needs to be considered if we are not able to correct C IV for the contamination by gas in non-virial motion.

A correction of the C IV line profile remains however the final aim. The identification of the non-reverberating component in this line would not just make C IV a reliable estimator for SMBH masses in high-redshift sources, but would help us in the understanding of the physical processes at work in the very central structures. The stratification of the BLR in terms of ionization potential of the lines tells us that C IV is emitted in the region closest to the accretion disk. Moreover, the presence of shifts and asymmetries in this line indicates that at least part of the emitting region could be in outflow or in presence of a wind. The study of this line is therefore necessary to understand the connection between BLR and accretion disk. A possibility is that the BLR is a continuation of the accretion disk under less extreme conditions. If winds or outflows are present in the region connecting BLR and disk, they could affect the next closest component of the Unified Model, the NLR, as well. In chapter 4 we examined this aspect. We noticed that there is a correspondence between the behaviour of C IV and [O III], the most intense narrow emission in quasar spectra. Specifically, when

[O III] is intense, C IV shows a peaky and symmetric profile, while when [O III] is weak or absent, C IV is boxy, blueshifted and asymmetric. The obvious explanation is that there is a physical connection between the regions emitting the two lines. Of course part of this behaviour can be produced by the presence of an orientation effect. On the one hand, when [O III] is intense, we are looking at a source in an edge-on position, and C IV could exhibit a narrow component as well. On the other hand, when [O III] is weak, it is more evident how C IV profile is affected in its broad component: the shift and asymmetry we see in these cases does not involve surely just a narrow component. A possible explanation in this case is that an outflow is present in the region connecting BLR and accretion disk. If strong enough, such an outflow could not only affect the NLR emission, but even prevent its formation.

Starting from these considerations, we used them to correct FWHM(C IV), i.e. to find a FWHM(C IV) with a tigher correlation with FWHM(Hβ), the line used as a touchstone for virial mass estimation. The first step was the analysis for both the profiles for a sample of 172 sources, for which several quantities were measured to give an accurate description of the lines. We then performed a statistical analysis on the data-set composed by the measured characteristics for both lines. We first performed a PCA analysis in order to understand which properties drive the variance in the data-set, finding that the property leading the eigenvector 1 is by far the linewidth of [O III], followed by other [O III] properties, describing the shape and asymmetry of this line. Even if at a far lesser extent, some of the C IV properties are important as well, among them FWHM(C IV). To be sure not to have overlooked non-linear dependences in the data-set, we examined also the Spearmann correlation matrix. The aim is to identify correlations between [O III] and C IV properties, able to describe the dependence of FWHM(C IV) on [O III] behaviour: in sources in which only C IV is available, we need to correct FWHM(C IV) through properties relating to this line only. The dependences highlighted by this analysis are not unexpected: the C IV properties correlating the most with [O III], along with FWHM(C IV), are exactly those features we recognised in C IV as connected to [O III], i.e. the parameters describing the shift with respect the reference wavelength and the shape of the line. To choose which ones of them are to be used in the correction of FWHM(C IV), we need a comparison with FWHM(Hβ), the touchstone of virial mass estimator. So far, we only performed this comparison for a small subsample, the one with reverberartion mapping measurements of Hβ. The result of the analysis is promising: the scatter in the relation between FWHM(Hβ) and the corrected FWHM(C IV) is decreased with respect to the original one. Moreover, this subsample is biased towards symmetric and peaky C IV line profiles and therefore does not span the entire "space" of C IV shapes and asymmetries. We expect that the analysis for the whole sample will give us more solid results. In the process of studying our data, we have devised a statistical analysis that can be readily used for every sample of quasars for which both C IV and [O III] observations are available.

Moreover, the final goal is not only the correction for FWHM(C IV). As already mentioned, the [O III]-C IV connection has important implications on the physics and the kinematics of the inner components in AGN and needs therefore to be further investigated.

As a final note, we would like to stress the importance of the spectroscopic analysis of quasars and the perspectives it gives in the astrophysical research. The compilation of a catalogue listing the main properties of the spectra of ∼105, 000 quasars for the SDSS DR7 (Shen et al., 2011) had opened the possibility of studying spectroscopic features of huge samples of quasars (an example is given in the Chapter 2 of this thesis), something that had not been possible before.

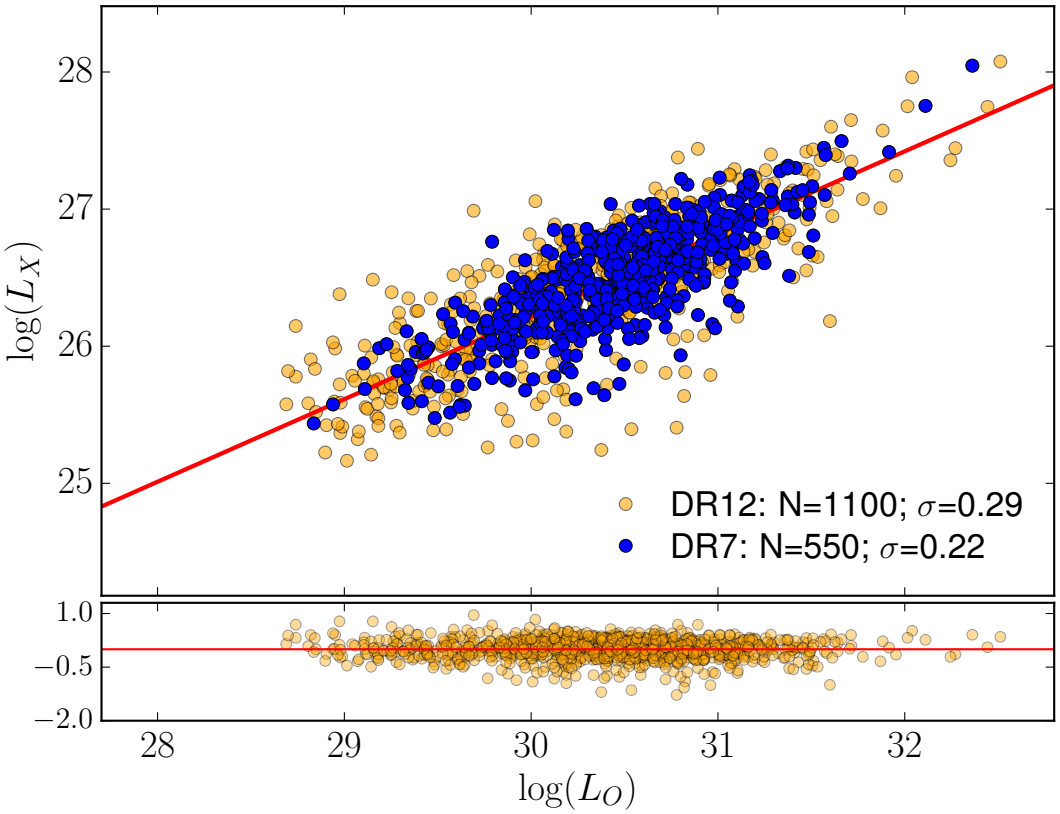


Figure C.1: Comparison between the L_{UV}-L_X relation for a sample of ∼ 1000 objects from DR12, for which only photometric measurements are available in the UV band (yellow points) and the same relation for 550 objects from DR7, for which instead the flux at 2500Å is available from the spectroscopic catalogue of Shen et al. (2011) (blue points). In the first case the dispersion in the L_{UV}-L_X relation is 0.28 dex, while the use of spectroscopic measurements narrow it down to 0.22 dex.

We are starting the process of automation of the tools we developed for the spectroscopic analysis in order to compile a similar catalogue, not available at the moment, for the ∼300, 000 quasars in the SDSS DR12 (Pâris et al., 2016). This work will take time, nontheless we can present here an immediate application of the benefit achieved. Taking a sample of ∼1000 objects from the SDSS, selected to have a serendipitous spectrum in the 3XMM-DR6 catalogue (Rosen et al., 2016) as well, it is possible to examine the L_{UV}-L_X relation with the final aim of building the Hubble Diagram for quasars as have been done in Risaliti & Lusso (2015). The proper determination of the L_{UV}-L_X relation is a significant part in the construction of the Hubble Diagram for quasars, because it enters in the accurate

determination of the distances of these objects.

Making a comparison between the L_{UV}-L_X relation available now for the Hubble Diagram, the one that uses the 2500Å flux provided by the accurate spectroscopic measurements of Shen et al. (2011), and the one for DR12, for which L_{UV} is inferred from the photometry, we can see how the spectroscopic analysis plays a keyrole (see Fig. C.1). In the case of DR12 indeed the dispersion in the L_{UV}-L_X relation is 0.28 dex, while for DR7, where spectroscopic measurements of the flux at 2500Å are available, the scatter goes down to 0.22 dex.

The spectroscopic analysis for quasars in DR12 will make it possible to more than double the sample suitable for an accurate Hubble Diagram construction.

List of Publications

Concerning this thesis:

S. Bisogni, A. Marconi and G. Risaliti,
"Orientation Effects on spectral emission features of quasars", MNRAS 464, 385-397 (2017)
http://adsabs.harvard.edu/abs/2017MNRAS.464..385B

S. Bisogni, S. di Serego Alighieri, P. Goldoni, L. C. Ho, A. Marconi, G. Ponti and G. Risaliti,
"Simultaneous detection of optical and ultraviolet broad emission lines in Quasars at z~2.2",
submitted to A&A

S. Bisogni, K. D. Denney and B. M. Peterson,
"The kinematic connection between BLR and NLR and how to use it to correct C IV*-based SMBH virial estimation"*, in preparation

Others:

L. Pei et al. (155 coauthors, including S. Bisogni),
"Space Telescope and Optical Reverberation Mapping Project. V. Optical Spectroscopy Campaign and Emission-Line Analysis for NGC 5548",
submitted to ApJ

We present the results of an optical spectroscopic monitoring program targeting NGC 5548 as part of a larger multi-wavelength reverberation mapping campaign. The campaign spanned six months and achieved almost daily cadence with observations from five ground-based telescopes. The Hβ and He IIλ4686 light curves lag that of the 5100Å optical continuum by $4.24^{+0.37}_{-0.37}$ days and $0.80^{+0.36}_{-0.35}$ days, respectively. The Hβ lag relative to the 1158Å ultraviolet continuum light curve measured by the *Hubble Space Telescope* is roughly ~50% longer than that measured against the optical continuum, and the lag difference is consistent with the observed lag between the optical and ultraviolet continua. This suggests that the characteristic radius of the broad-line region is ~50% larger than the value

inferred from optical data alone. We also measured velocity-resolved lags for Hβ and found a complex velocity-lag structure with shorter lags in the line wings, indicative of a broad-line region dominated by Keplerian motion. The responsivity of both the Hβ and He II lines decreased significantly halfway through the campaign, a phenomenon also observed for C IV, Lyα, He II(+O III]), and Si IV(+O IV]) during the same monitoring period. Finally, given the optical luminosity of NGC 5548 during our campaign, the measured Hβ lag is a factor of five shorter than the expected value implied by the R_{BLR} - L_{AGN} relation based on the past behavior of NGC 5548.

M. R. Goad et al. (100 coauthors, including S. Bisogni),

"Space Telescope and Optical Reverberation Mapping Project. IV. Anomalous Behavior of the Broad Ultraviolet Emission Lines in NGC 5548", ApJ, 824, 11 (2016)

http://adsabs.harvard.edu/abs/2016ApJ...824...11G

During an intensive Hubble Space Telescope (HST) Cosmic Origins Spectrograph (COS) UV monitoring campaign of the Seyfert 1 galaxy NGC 5548 performed from 2014 February to July, the normally highly correlated far UV continuum and broad emission line variations decorrelated for 60-70 days, starting 75 days after the first HST/COS observation. Following this anomalous state, the flux and variability of the broad emission lines returned to a more normal state. This transient behavior, characterized by significant deficits in flux and equivalent width of the strong broad UV emission lines, is the first of its kind to be unambiguously identified in an active galactic nucleus reverberation mapping campaign. The largest corresponding emission line flux deficits occurred for the high ionization, collisionally excited lines C IV and Si IV(+O IV]), and also He II(+O III]), while the anomaly in Lyα was substantially smaller. This pattern of behavior indicates a depletion in the flux of photons with $E_{ph} > 54$ eV relative to those near 13.6 eV. We suggest two plausible mechanisms for the observed behavior: (i) temporary obscuration of the ionizing continuum incident upon broad line region (BLR) clouds by a moving veil of material lying between the inner accretion disk and inner (BLR), perhaps resulting from an episodic ejection of material from the disk, or (ii) a temporary change in the intrinsic ionizing continuum spectral energy distribution resulting in a deficit of ionizing photons with energies >54 eV, possibly due to a transient restructuring of the Comptonizing atmosphere above the disk. Current evidence appears to favor the latter explanation.

T. Giannini, et al. (7 coauthors, including S. Bisogni),

"On the 2015 Outburst of the EXor Variable V1118 Ori", ApJ, 819, L5 (2016)

http://adsabs.harvard.edu/abs/2016ApJ...819L...5G

After a long-lasting period of quiescence of about a decade, the source V1118 Ori, one of the most representative members of the EXor variables, is now outbursting. Since the initial increase of the near-infrared flux of about 1 mag (JHK bands) registered on 2015 September 22, the source brightness has remained fairly stable. We estimate ŒîV 3 mag, with respect to the quiescence phase. An optical/near-IR low-resolution spectrum has been obtained with the Large Binocular Telescope instruments MODS and LUCI2 and compared with a spectrum of a similar spectral resolution and sensitivity level taken during quiescence. Together with the enhancement of the continuum, the outburst spectrum presents a definitely higher number of emission lines, in particular H Irecombination lines of the Balmer, Paschen, and Brackett series, along with bright permitted lines of several species, forbidden atomic lines, and CO ro-vibrational lines. Both mass accretion and mass-loss rates have significantly increased (by about an order of magnitude: $\dot{M}_{\rm acc}$ = 1.2-4.8 $\times$ 10^{-8}M$_\odot$yr^{-1}, $\dot{M}_{\rm loss}$ = 0.8-2 $\times$ 10^{-9}M$_\odot$yr^{-1}), with respect to the quiescence phase. If compared with previous outbursts, the present one appears less energetic. Alternatively, it could already be in the fading phase (with the maximum brightness level reached when the source was not visible), or, viceversa, still in the rising phase.

R. Cesaroni, et al. (13 coauthors, including S. Bisogni),

"Star and jet multiplicity in the high-mass star forming region IRAS 05137+3919", A&A 581, A124 (2015)

http://adsabs.harvard.edu/abs/2015A%26A...581A.124C

Context. We present a study of the complex high-mass star forming region IRAS 05137+3919 (also known as Mol8), where multiple jets and a rich stellar cluster have been described in previous works.
Aims. Our goal is to determine the number of jets and shed light on their origin, and thus determine the nature of the young stars powering these jets. We also wish to analyse the stellar clusters by resolving the brightest group of stars.
Methods. The star forming region was observed in various tracers and the results were complemented with ancillary archival data. The new data represent a substantial improvement over previous studies both in resolution and frequency coverage. In particular, adaptive optics provides us with an angular resolution of 80 mas in the near IR, while new mid- and far-IR data allow us to sample the peak of the spectral energy distribution and thus reliably estimate the bolometric luminosity.
Results. Thanks to the near-IR continuum and millimetre line data we can determine the structure and velocity field of the bipolar jets and outflows in this star forming region. We also find that the stars are grouped into three clusters and the jets originate in the richest of these, whose luminosity is $\sim 2.4 \times 10^4 L_\odot$. Interestingly, our high-resolution near-IR images allow us to resolve one of the two brightest stars (A and B) of the cluster into a double source (A1+A2).
Conclusions. We confirm that there are two jets and establish that they are powered by B-type stars belonging to cluster C1. On this basis and on morphological and kinematical arguments, we conclude that the less extended jet is almost perpendicular to the line of sight and that it originates in the brightest star of the cluster, while the more extended one appears to be associated with the more extincted, double source A1+A2. We propose that this is not a binary system, but a small bipolar reflection nebula at the root of the large-scale jet, outlining a still undetected circumstellar disk. The gas kinematics on a scale of ~0.2pc seems to support our hypothesis, because it appears to trace rotation about the axis of the associated jet.

V. Nascimbeni, et al. (10 coauthors, including S. Bisogni),

"Large Binocular Telescope view of the atmosphere of GJ1214b", A&A 579, A113 (2015)

http://adsabs.harvard.edu/abs/2015A%26A...579A.113N

The atmospheric composition and vertical structure of the super-Earth GJ1214b has been a subject of debate since its discovery in 2009. Recent studies have indicated that high-altitude clouds might mask the lower layers. However, some data points that were gathered at different times and facilities do not fit this picture, probably because of a combination of stellar activity and systematic errors. We observed two transits of GJ1214b with the Large Binocular Camera, the dual-channel camera at the Large Binocular Telescope. For the first time, we simultaneously measured the relative planetary radius $k = Rp/R_*$ at blue and red optical wavelengths ($B + R$), thus constraining the Rayleigh scattering on GJ1214b after correcting for stellar activity effects. To the same purpose, a long- term photometric follow-up of the host star was carried out with WiFSIP at STELLA, revealing a rotational period that is significantly longer than previously reported. Our new unbiased estimates of k yield a flat transmission spectrum extending to shorter wavelengths, thus confirming the cloudy atmosphere scenario for GJ1214b.

Appendices

In Appendix A, the fits for five out of the six quasars examined in Chapter 3 are shown. Fits for the first source, J093147, are shown instead in Fig. 3.1.
In Appendix B, we give details for the sample of 172 sources that we examined in Chapter 4.

Appendix A

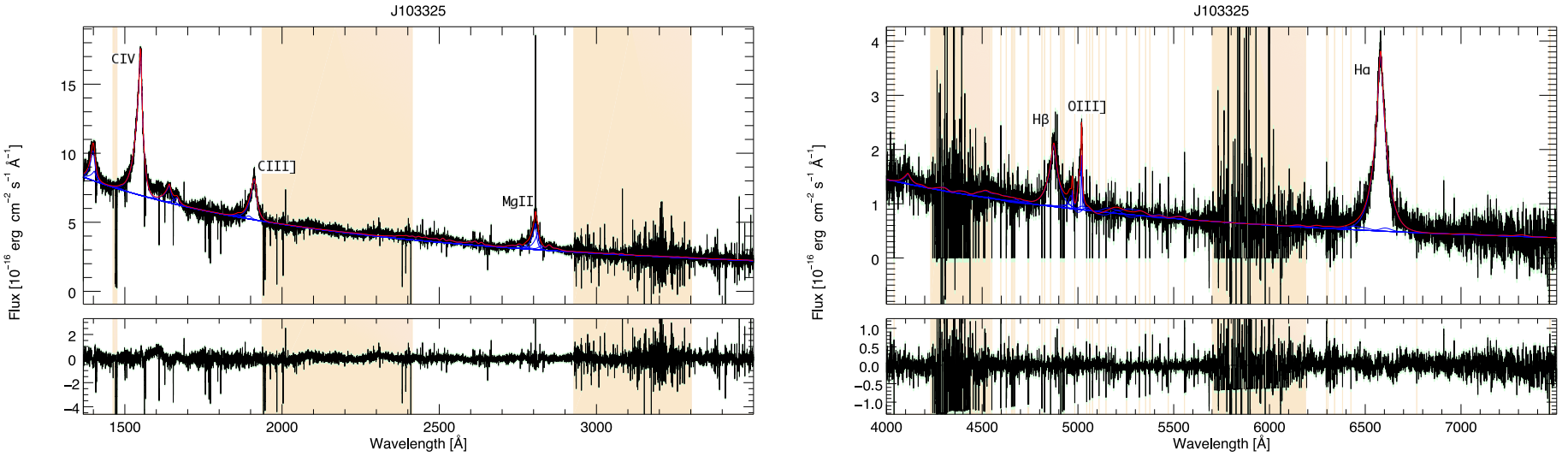


Figure A.1: Fits for all the examined spectral windows (C IV-C III]-Mg II and Hβ-Hα large windows, C IV-C III], Mg II, Hβ and Hα small windows) for J103325. The black line is the original spectrum .The blue solid lines are the best fit models for the emissions (continuum, Fe II and emission lines) and the red solid line is the total best fit. The lower panels show residuals between best fit model and original spectrum. The colored regions are those we chose to mask. This choice can be due to the presence of strong emission blending, to a non representativeness of the Fe II templates or, in general, to a lack of knowlegde about what kind of emission is able to reproduce such features and the presence of noise.

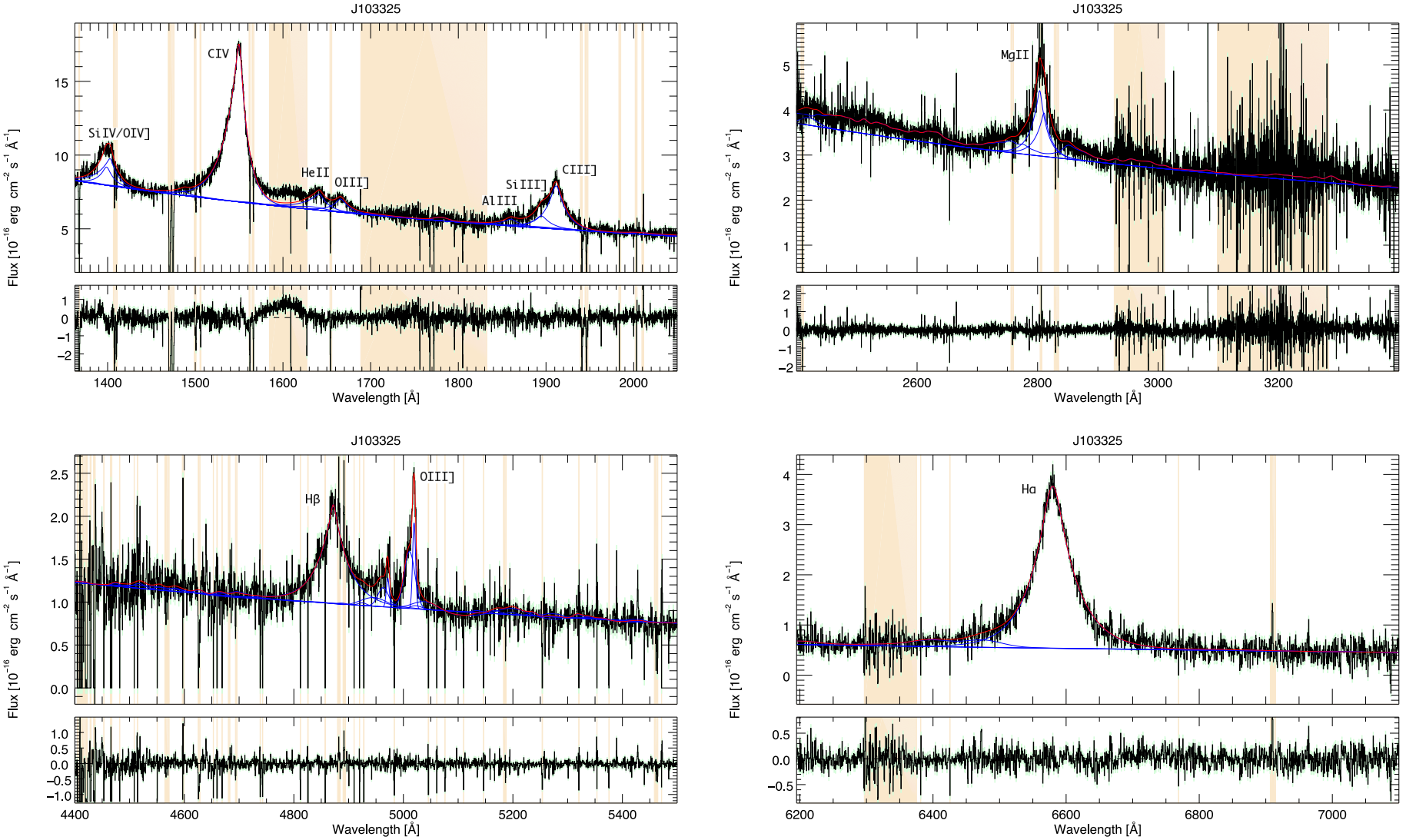


Figure A.1 (Continued)

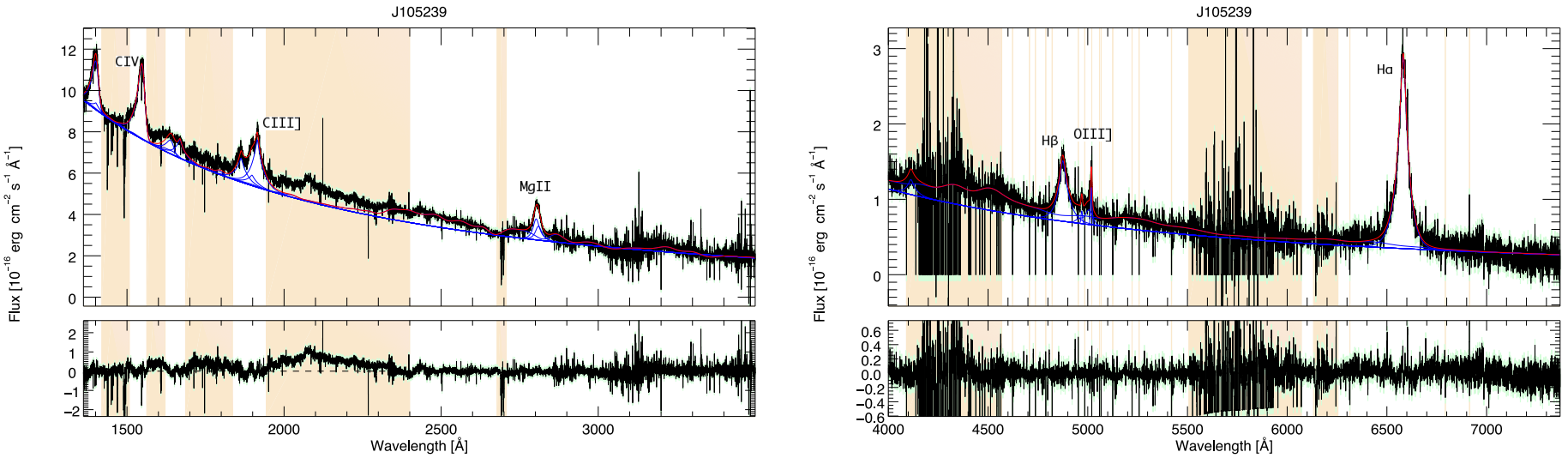


Figure A.2: Fits for all the examined spectral windows (C IV-C III]-Mg II and Hβ-Hα large windows, C IV-C III], Mg II, Hβ and Hα small windows) for J105239. The black line is the original spectrum .The blue solid lines are the best fit models for the emissions (continuum, Fe II and emission lines) and the red solid line is the total best fit. The lower panels show residuals between best fit model and original spectrum. The colored regions are those we chose to mask. This choice can be due to the presence of strong emission blending, to a non representativeness of the Fe II templates or, in general, to a lack of knowlegde about what kind of emission is able to reproduce such features and the presence of noise.

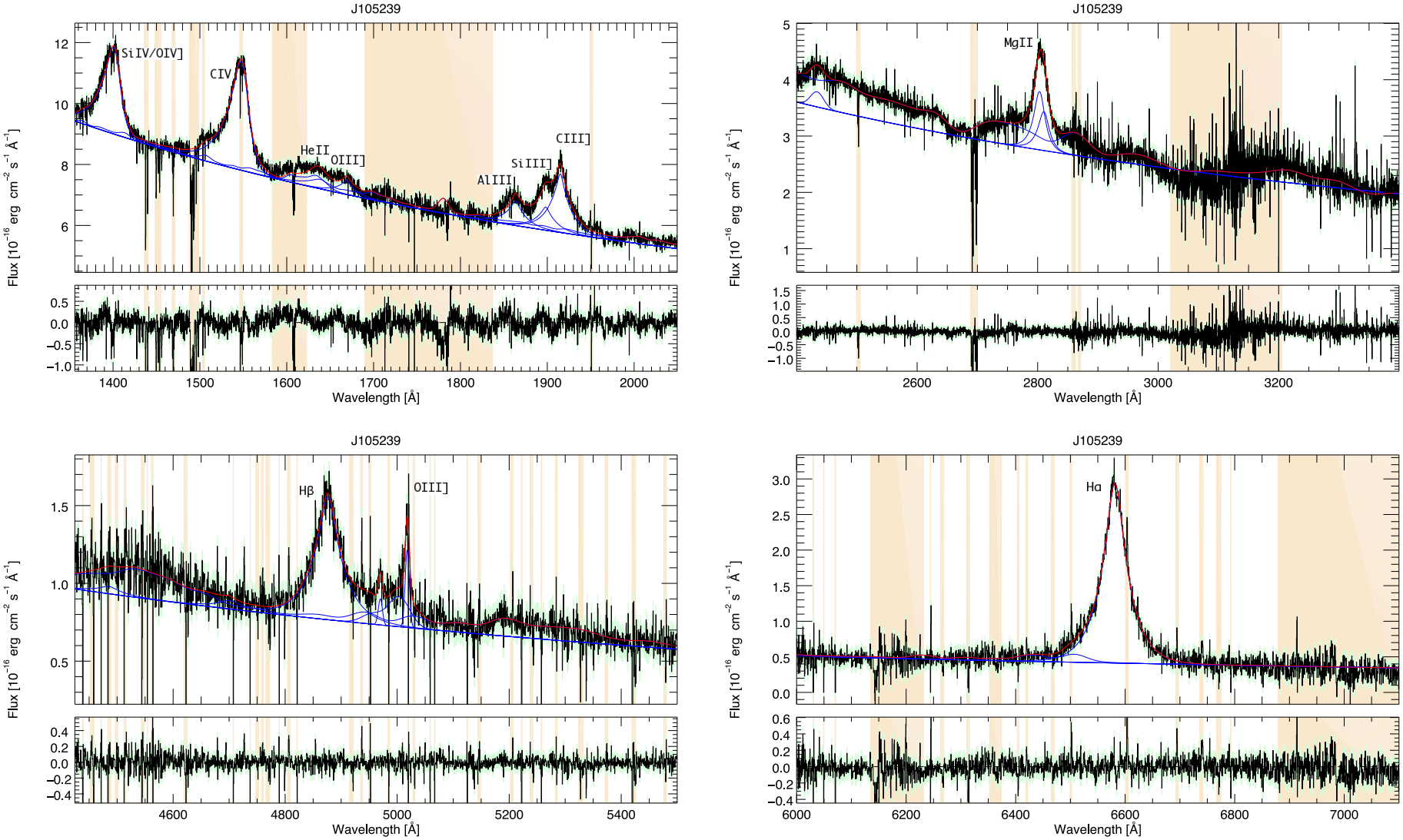


Figure A.2 (Continued)

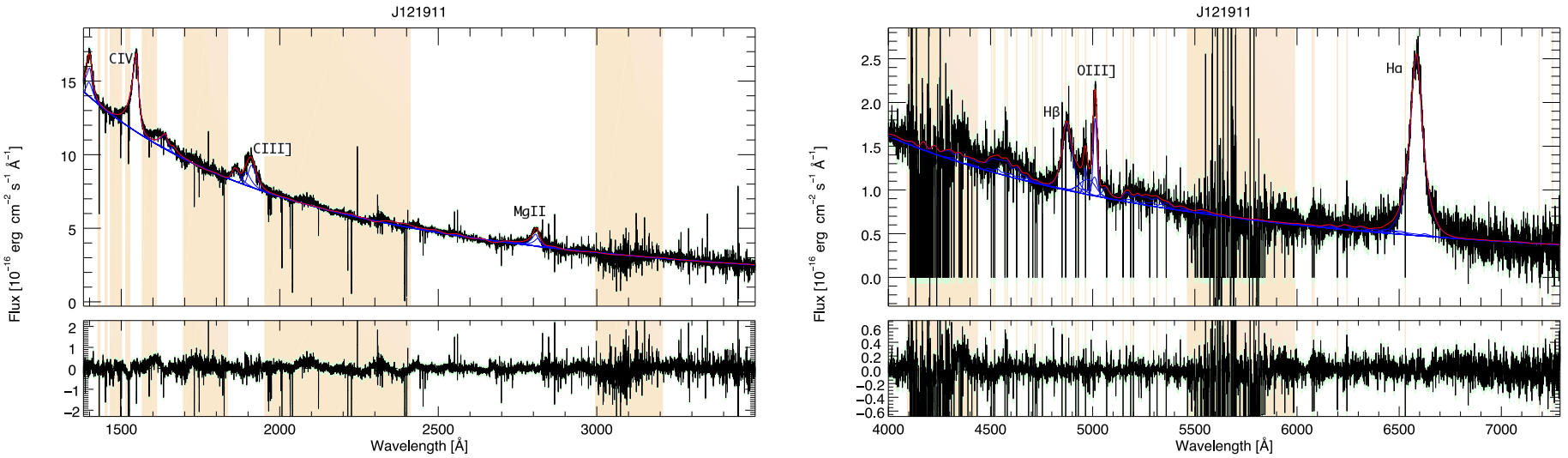


Figure A.3: Fits for all the examined spectral windows (C IV-C III]-Mg II and Hβ-Hα large windows, C IV-C III], Mg II, Hβ and Hα small windows) for J121911. The black line is the original spectrum .The blue solid lines are the best fit models for the emissions (continuum, Fe II and emission lines) and the red solid line is the total best fit. The lower panels show residuals between best fit model and original spectrum. The colored regions are those we chose to mask. This choice can be due to the presence of strong emission blending, to a non representativeness of the Fe II templates or, in general, to a lack of knowlegde about what kind of emission is able to reproduce such features and the presence of noise.

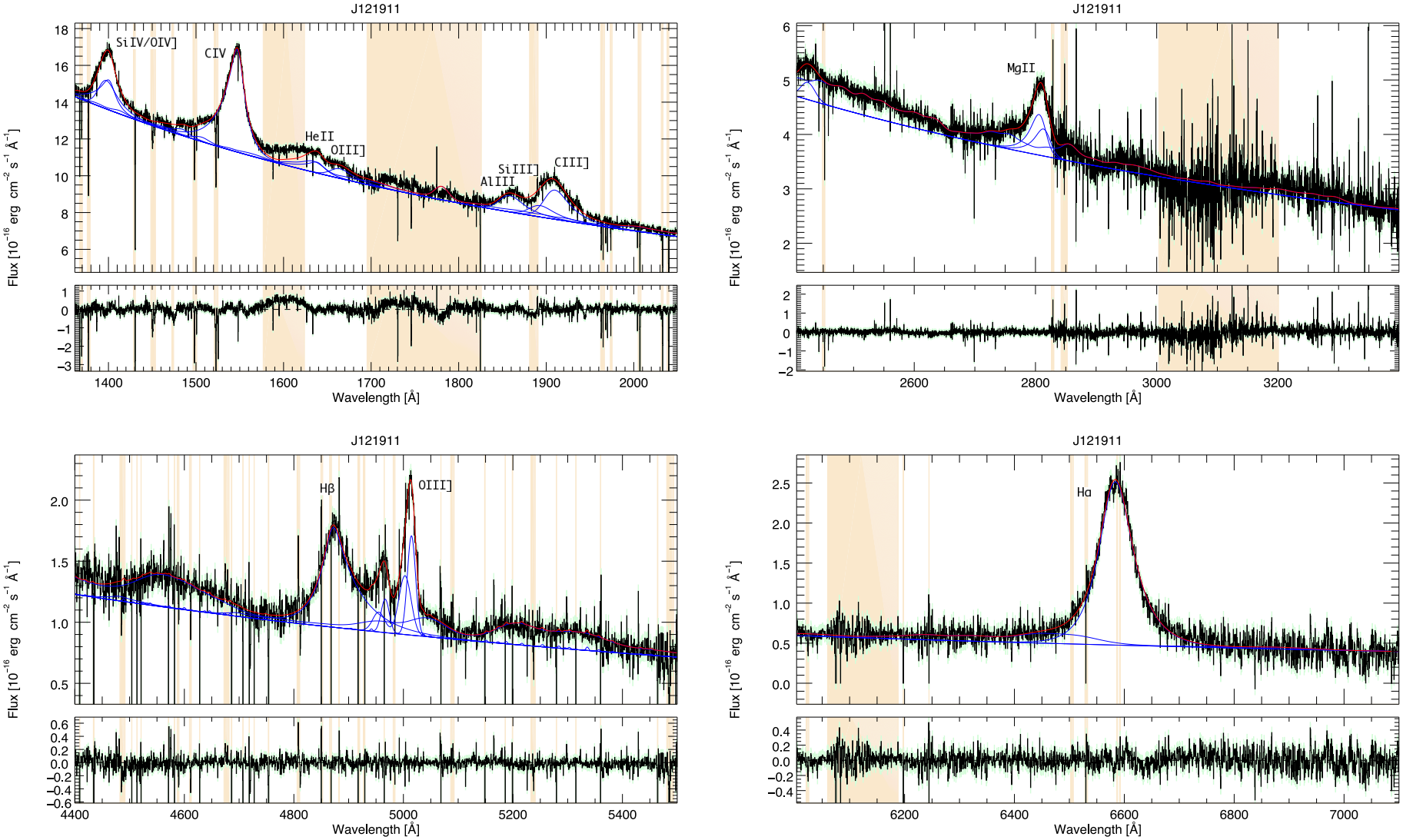


Figure A.3 (Continued)

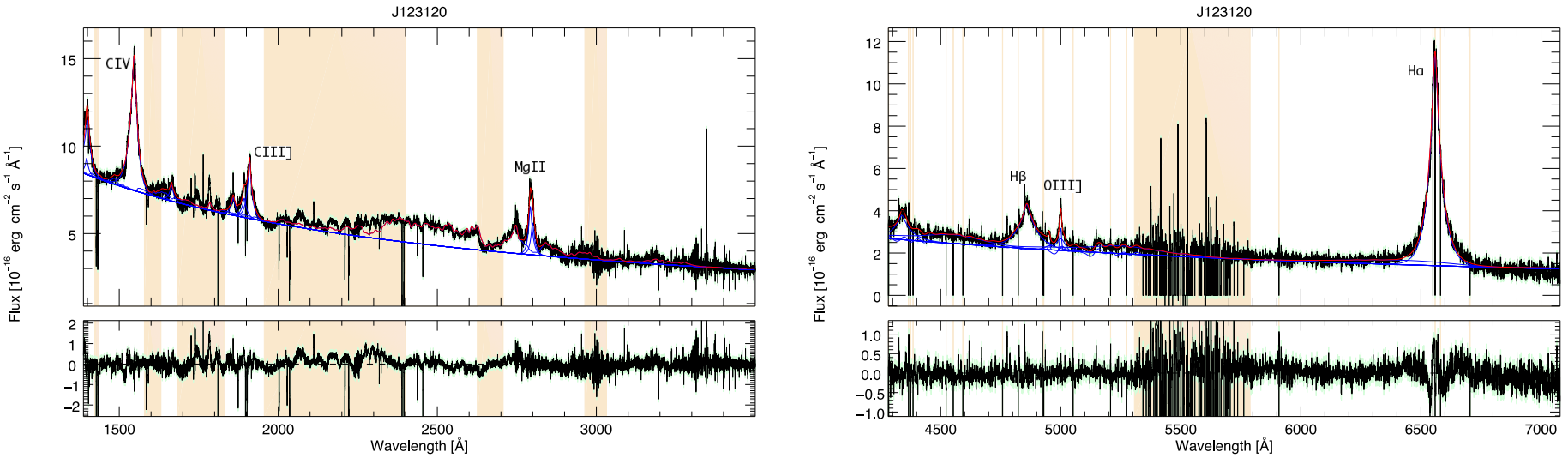


Figure A.4: Fits for all the examined spectral windows (C IV-C III]-Mg II and Hβ-Hα large windows, C IV-C III], Mg II, Hβ and Hα small windows) for J123120. The black line is the original spectrum .The blue solid lines are the best fit models for the emissions (continuum, Fe II and emission lines) and the red solid line is the total best fit. The lower panels show residuals between best fit model and original spectrum. The colored regions are those we chose to mask. This choice can be due to the presence of strong emission blending, to a non representativeness of the Fe II templates or, in general, to a lack of knowlegde about what kind of emission is able to reproduce such features and the presence of noise.

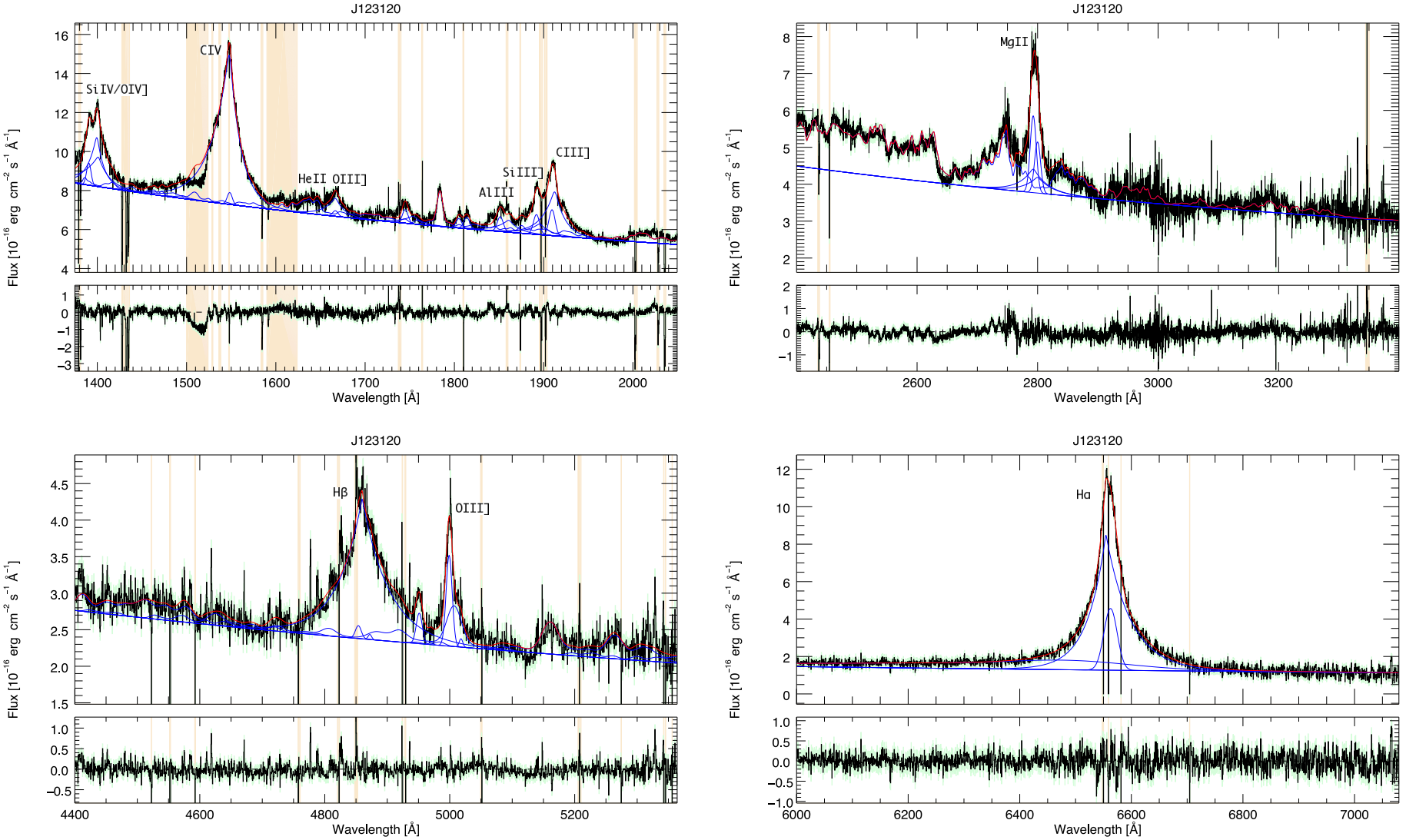


Figure A.4 (Continued)

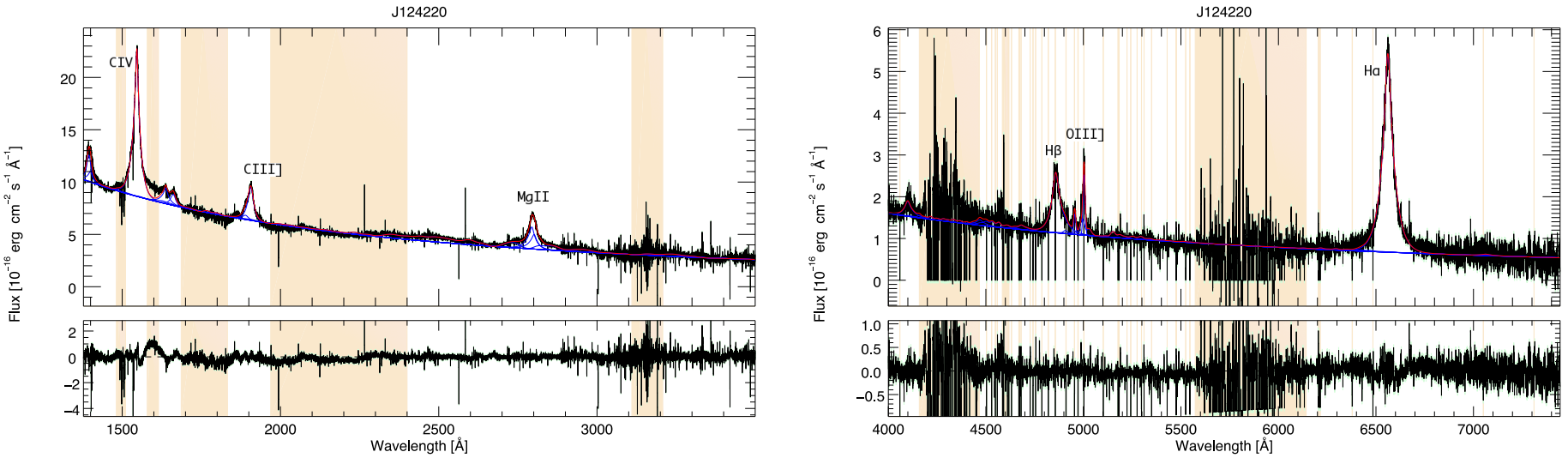


Figure A.5: Fits for all the examined spectral windows (C IV-C III]-Mg II and Hβ-Hα large windows, C IV-C III], Mg II, Hβ and Hα small windows) for J124220. The black line is the original spectrum .The blue solid lines are the best fit models for the emissions (continuum, Fe II and emission lines) and the red solid line is the total best fit. The lower panels show residuals between best fit model and original spectrum. The colored regions are those we chose to mask. This choice can be due to the presence of strong emission blending, to a non representativeness of the Fe II templates or, in general, to a lack of knowlegde about what kind of emission is able to reproduce such features and the presence of noise.

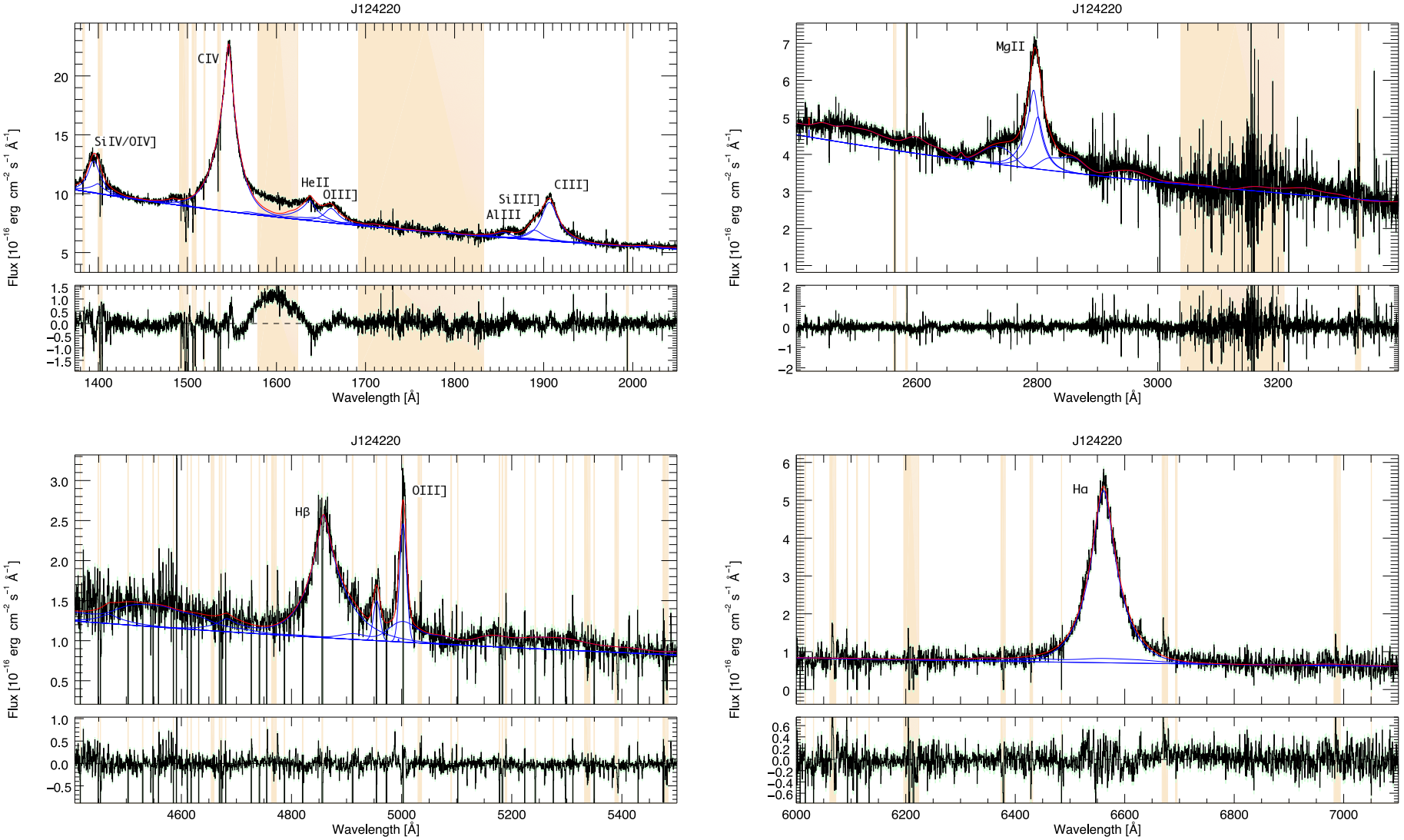


Figure A.5 (Continued)

Appendix B

Table B.1: Complete sample. Colums indicate: (1) source name, (2) right ascension, (3) declination, (4) nominal redshift initially used, if necessary, to restframe the spectrum, (5) redshift obtained by the fit, (6) UV wavelength range resolving power, (7) optical wavelength range resolving power, (8) data reference, (7) subsample name, (8) flag describing the quality of the observations.

Name	RA	dec	nominal z [a]	z fit [b]	UV R	opt R	reference	subsample	flag [c]
J0042-1020	004219.73	−102012.2	3.88161	3.864	∼ 7450	∼ 5300	Lopez et al. (2016)	XQ-100	1
J0056-2808	005624.87	−280833.3	3.624	3.627	”	”	”	”	0
J0057-2643	005758.14	−264312.9	3.655	3.660	”		”	”	X
J0100-2708	010012.47	−270852.1	3.520	3.563			”	”	0
J0124+0044	012403.97	+004431.4	3.83995	3.840			”	”	0
J0755+1345	075552.43	+134549.6	3.67176	3.666			”	”	0
J0800+1920	080050.26	+192056.3	3.9586	...			”	”	X
J0818+0958	081855.75	+095844.9	3.674	...			”	”	X
J0833+0959	083322.50	+095938.6	3.750	...			”	”	X
J0920+0725	092041.72	+072541.2	3.74951	3.750			”	”	0
J0935+0022	093556.87	+002252.8	3.74951	...			”	”	0
J0937+0828	093714.51	+082856.2	3.711916	3.713			”	”	0
J1013+0650	101347.48	+065016.6	3.79216	3.789			”	”	0
J1018+0548	101818.57	+054820.7	3.527278	3.527			”	”	1
J1020+0922	102040.74	+092253.1	3.658419	3.686			”	”	0
J1024+1819	102456.78	+181907.1	3.525	3.510			”	”	1
J1037+2135	103730.43	+213529.8	3.644856	3.630			”	”	1
J1042+1957	104234.02	+195716.3	3.639	3.633			”	”	1
J1053+0103	105340.82	+010333.5	3.6550	3.658			”	”	0
J1103+1004	110352.72	+10040.48	3.600377	3.619			”	”	X
J1108+1209	110855.56	+120951.7	3.673524	3.673			”	”	0
J1111-0804	111113.89	−080403.9	3.922	...			”	”	X
J1117+1311	111701.97	+131113.0	3.62391	3.625			”	”	1
J1126-0124	112634.42	−012438.0	3.7413	...			”	”	X
J1126-0126	112617.54	−012634.2	3.60728	3.618			”	”	1
J1135+0842	113536.55	+084217.3	3.850668	3.859			”	”	0
J1201+1206	120148.05	+120628.2	3.512	...			”	”	X
J1202-0054	120210.06	−005427.9	3.604315	3.587			”	”	1

J1248+1304	124837.39	+130439.2	3.722319	3.707	”	”	0
J1249-0159	124957.40	−015929.8	3.6286	3.692	”	”	0
J1304+0239	130452.60	+023921.8	3.65143	...	”	”	X
J1312+0841	131242.94	+084102.8	3.74341	3.735	”	”	1
J1320-0523	132030.12	−052336.3	3.70	3.72	”	”	1
J1331+1015	133150.77	+101527.5	3.84911	3.838	”	”	0
J1332+0052	133254.60	+005248.3	3.51	3.51	”	”	1
J1336+0243	133653.43	+024335.5	3.79888	3.802	”	”	1
J1352+1303	135248.09	+130309.8	3.702884	3.719	”	”	0
J1416+1811	141608.32	+181146.1	3.594	3.606	”	”	1
J1421-0643	142107.93	−064357.6	3.689	3.687	”	”	1
J1442+0920	144250.12	+091958.9	3.51448	3.558	”	”	0
J1445+0958	144516.62	+095834.9	3.552339	3.555	”	”	1
J1503+0419	150329.01	+041947.3	3.666762	3.668	”	”	1
J1517+0511	151756.20	+051100.7	3.559	3.574	”	”	1
J1524+2123	152436.17	+212307.0	3.605844	3.614	”	”	0
J1552+1005	155255.22	+100537.0	3.734	3.729	”	”	0
J1621-0042	162117.04	−004252.9	3.728456	3.705	”	”	1
J1658-0739	165844.20	−073916.4	3.742	3.736	”	”	X
3c110	041716.70	−055345.0	0.7749	–	Shang et al. (2011); Wills et al. (1995); Netzer et al. (1995)	RLQ	1
3c175	071302.40	+114614.7	0.7693	–	”	”	1
3c207	084047.59	+131223.6	0.6797	–	”	”	1
3c215	090631.90	+164611.4	0.4108	–	”	”	1
3c232	095820.95	+322402.2	0.5297	–	”	”	1
3c254	111438.48	+403720.3	0.7363	–	”	”	1
3c263	113957.04	+654749.4	0.6464	–	”	”	1
3c277.1	125226.35	+563419.7	0.3199	–	”	”	1
3c281	130754.00	+064214.3	0.6017	–	”	”	1
3c288.1	134213.18	+602142.9	0.9631	–	”	”	1
3c334	162021.92	+173624.0	0.5553	–	”	”	1
3c37	011818.49	+025806.0	0.6661	–	”	”	1
3c47	013624.40	+205727.0	0.4250	–	”	”	1
4c01.04	013957.25	+013146.2	0.2634	–	”	”	1
4c06.69	214805.46	+065738.6	1.0002	–	”	”	1
4c10.06	021707.66	+110410.1	0.4075	–	”	”	1
4c12.40	112129.79	+123617.4	0.6836	–	”	”	1
4c19.44	135704.43	+191907.4	0.7192	–	”	”	1
4c31.63	220314.97	+314538.3	0.2952	–	”	”	1

4c34.47	172320.80	+341757.9	0.2055	–	Shang et al. (2011, 2005); Kriss (2001)	*FUSE*	1
4c39.25	092703.01	+390220.9	0.6946	–	Shang et al. (2011); Wills et al. (1995); Netzer et al. (1995)	RLQ	1
4c41.21	101027.52	+413238.9	0.6124	–	”	”	1
4c49.22	115324.46	+493108.8	0.3333	–	”	”	1
4c55.17	095738.18	+552257.8	0.8990	–	”	”	0
4c73.18	192748.49	+735801.6	0.3027	–	”	”	1
B2 0742+31	074541.67	+314256.6	0.4616	–	”	”	1
B2 1611+34	161341.06	+341247.9	1.3945	–	”	”	0
MC2 0042+101	004458.72	+102653.7	0.5870	–	”	”	1
MC2 1146+111	114847.89	+105459.4	0.8614	–	”	”	0
MRK 506	172239.90	+305253.0	0.0428	–	Shang et al. (2011, 2005); Kriss (2001)	*FUSE*	1
MRK 509	204409.74	−104324.5	0.0345	–	”	”	1
OS 562	163813.45	+572024.0	0.7506	–	Shang et al. (2011); Wills et al. (1995); Netzer et al. (1995)	RLQ	1
PG 0052+251	005452.10	+252538.0	0.1544	–	Shang et al. (2011, 2005); Kriss (2001)	*FUSE*	1
PG 0844+349	084742.40	+344504.0	0.0643	–	”	”	1
PG 0947+396	095048.39	+392650.5	0.2057	–		*FUSE*,PGX	1
PG 0953+414	095652.40	+411522.0	0.2338	–		”	0
PG 1001+054	100420.09	+051300.5	0.1603	–	Shang et al. (2011, 2007)	PGX	1
PG 1100+772	110413.69	+765858.0	0.3114	–		*FUSE*,RLQ	1
PG 1103-006	110631.77	−005252.5	0.4234	–	Shang et al. (2011); Wills et al. (1995); Netzer et al. (1995)	RLQ	1
PG 1114+445	111706.40	+441333.0	0.1440	–	Shang et al. (2011, 2007)	PGX	1
PG 1115+407	111830.20	+402553.0	0.1541	–	”	”	1
PG 1116+215	111908.60	+211918.0	0.1759	–	”	”	1
PG 1202+281	120442.10	+275411.0	0.1651	–	”	”	1
PG 1216+069	121920.88	+063838.4	0.3319	–	”	”	1
PG 1226+023	122906.70	+020308.6	0.1576	–		*FUSE*,PGX,RLQ	1
PG 1259+593	130112.90	+590206.4	0.4769	–	Shang et al. (2011, 2005); Kriss (2001)	*FUSE*	1
PG 1309+355	131217.77	+351521.2	0.1823	–	Shang et al. (2011, 2007)	PGX	1
PG 1322+659	132349.54	+654148.0	0.1684	–		*FUSE*,PGX	0
PG 1351+640	135315.81	+634545.4	0.0882	–	Shang et al. (2011, 2005); Kriss (2001)	*FUSE*	1
PG 1352+183	135435.60	+180517.2	0.1510	–	Shang et al. (2011, 2007)	PGX	1
PG 1402+261	140516.19	+255534.9	0.1650	–	”	”	1
PG 1411+442	141348.30	+440014.0	0.0895	–	”	”	1
PG 1415+451	141700.80	+445606.0	0.1143	–	”	”	1
PG 1425+267	142735.54	+263213.6	0.3637	–	”	”	1
PG 1427+480	142943.00	+474726.0	0.2203	–	”	”	1
PG 1440+356	144207.46	+352622.9	0.0773	–	”	”	1
PG 1444+407	144645.90	+403505.0	0.2673	–	”	”	1

PG 1512+370	151443.04	+365050.4	0.3700	–	”	”	1
PG 1534+580	153552.36	+575409.2	0.0303	–	Shang et al. (2011, 2005); Kriss (2001)	*FUSE*	1
PG 1543+580	154530.20	+484609.0	0.4000	–	Shang et al. (2011, 2007)	PGX	1
PG 1545+210	154743.54	+205216.7	0.2642	–	Shang et al. (2011); Wills et al. (1995); Netzer et al. (1995)	RLQ	1
PG 1626+554	162756.00	+552231.0	0.1317	–	Shang et al. (2011, 2007)	PGX	1
PG 1704+608	170441.37	+604430.5	0.3730	–		*FUSE*,RLQ	1
PG 2214+139	221712.26	+141421.1	0.0657	–	Shang et al. (2011, 2005); Kriss (2001)	*FUSE*	1
PG 2251+113	225410.40	+113638.3	0.3253	–	Shang et al. (2011); Wills et al. (1995); Netzer et al. (1995)	RLQ	1
PG 2349-014	235156.13	−010913.3	0.1740	–	Shang et al. (2011, 2005); Kriss (2001)	*FUSE*	1
PKS 0403-13	040534.00	−130813.7	0.5700	–	Shang et al. (2011); Wills et al. (1995); Netzer et al. (1995)	RLQ	1
PKS 0859-14	090216.83	−141530.9	1.3320	–	”	”	0
PKS 2216-03	221852.04	−033536.9	0.8993	–	”	”	1
TEX 1156+213	115926.20	+210655.0	0.3480	–	”	”	1
3c120	043311.10	+052115.6	0.033010	0.032701	Peterson et al. (1998, 2004)	RM	1
3c390	184537.62	+095344.7	0.056100	0.055180	Dietrich et al. (1998); Peterson et al. (2004)	”	1
Akn120	051611.42	−000859.4	0.032713	0.033201	Peterson et al. (1998, 2004)	”	1
Fairall9	012345.78	−584820.8	0.047016	0.045722	Santos-Lleó et al. (1997); Peterson et al. (2004)	”	1
Mrk279	135303.45	+691829.6	0.030451	0.030373	Santos-Lleó et al. (2001); Peterson et al. (2004)	”	1
Mrk290	153552.36	+575409.2	0.029577	0.030385	Denney et al. (2010)	”	1
Mrk335	000619.52	+201210.5	0.025785	0.026020	Grier et al. (2012)	”	1
Mrk509	204409.74	−104324.5	0.034397	0.034796	Peterson et al. (1998, 2004)	”	1
Mrk590	021433.56	−004600.1	0.026385	0.026785	”	”	1
Mrk817	143622.07	+584739.4	0.031455	0.031449	Denney et al. (2010)	”	1
NGC3516	110647.49	+723406.9	0.008836	0.008654	”	”	1
NGC3783	113901.76	−374419.2	0.009730	0.009249	Reichert et al. (1994); Onken & Peterson (2002)	”	1
NGC4051	120309.61	+443152.8	0.002336	0.001980	Peterson et al. (2000)	”	1
NGC4151	121032.58	+392420.6	0.003319	0.002840	Bentz et al. (2006)	”	1
NGC4593	123939.42	−052039.3	0.009000	0.008235	Denney et al. (2006)	”	1
NGC5548	141759.53	+250812.4	0.017175	0.016424	Peterson et al. (2002, 2004)	”	1
NGC7469	230315.62	+085226.4	0.016317	0.016078	Peterson et al. (2014)	”	1
PG0026	002913.70	+131604.0	0.142000	0.144499	Kaspi et al. (2000); Peterson et al. (2004)	”	1
PG0052	005452.12	+252539.0	0.154450	0.153480	”	”	1
PG0804	081058.60	+760242.5	0.100000	0.100361	”	”	1
PG0953	095652.39	+411522.2	0.234100	0.232877	”	”	1
PG1226	122906.70	+020308.6	0.158339	0.157229	”	”	1
PG1229	123203.60	+200929.2	0.063010	0.063298	”	”	1
PG1307	130947.00	+081948.2	0.155000	0.153609	”	”	1
PG1426	142906.59	+011706.5	0.086570	0.085745	”	”	1

PG1613	161357.18	+654309.6	0.129000	0.128628			”	”	1
PG2130	213227.81	+100819.5	0.062977	0.062885			Grier et al. (2012); Peterson et al. (2004)	”	1
Q 0150-202	015227.3	−200107.0	2.147	~		~ 540	Dietrich et al. (2009)	Dietrich	1
Q 2116-4439	212011.5	−442653.8	1.504	~		”	”	”	1
Q 2154-2005	215705.9	−195113.8	2.042	~		”	”	”	1
Q 2209-1842	221210.4	−182737.8	2.098	~		”	”	”	1
Q 2230+0232	223235.3	+024755.1	2.215	~		”	”	”	1
Q 2302+0255	230445.0	+031146.1	1.062	~		”	”	”	1
J 0254+0021	025438.37	+002132.8	2.456	~	~ 1800 – 2000	~ 640	Netzer et al. (2007)	Netzer	1
J 2102+0020	210258.22	+002023.4	3.328	~	~ 1800 – 2000	~ 850	”	”	1
J 2103-0600	210311.68	−060059.4	3.336	~	~ 1800 – 2000	~ 850	”	”	1
Q0019-0913	001954.67	−091316.4	2.1223	2.1220	~ 1850	~ 1800	Collinson et al. (2015)	Boroson-MODS	1
Q0118-0052	011827.98	−005239.8	2.1883	2.1932	”	”	”	”	1
Q0139+0013	013929.52	+001330.8	2.1041	2.1006	”	”	”	”	1
Q0148-0010	014809.65	−001017.7	2.1631	2.1706	”	”	”	”	1
Q0209-0825	020902.86	−082531.8	2.1330	2.1352	”	”	”	”	1
Q0321-0011	032158.39	−001102.6	2.1481	2.1573	”	”	”	”	1
Q0409-0411	040954.19	−041136.9	2.1878	2.1834	”	”	”	”	1
Q0912+0058	091208.64	+005857.3	2.1896	2.1837	”	”	”	”	1
Q1002+0021	100246.80	+002104.0	2.1679	2.1726	”	”	”	”	1
Q1019-0054	101900.00	−005420.1	2.1838	2.1818	~ 1800 – 2000	”	”	”	1
Q1029+0205	102906.72	+020500.0	2.1370	2.1391	~ 1850	”	”	”	1
Q1033+0128	103325.92	+012836.3	2.1771	2.1811	”	”	”	”	1
Q1049-0110	104915.36	−011038.1	2.1154	2.1227	”	”	”	”	1
Q1214-0307	121427.84	−030721.0	2.1265	2.1252	”	”	”	”	1
Q1235+0304	123514.40	+030416.7	2.1981	2.1994	”	”	”	”	1
Q2059-0019	205954.48	−001917.3	2.1065	2.1113	”	”	”	”	1
Q2118+0042	211832.88	+004219.0	2.1649	2.1684	”	”	”	”	1
Q2259+1402	225921.60	+140256.0	2.1491	2.1497	”	”	”	”	1
Q2314-0824	231441.52	−082406.7	2.1965	2.2012	~ 1800 – 2000	”	”	”	1

Notes

(a) The nominal redshift used for rest-frame the spectrum, either from the reference for the source or the NED (NASA/IPAC Extragalactic Database).

(b) The fit redshift is obtained by the position of the [O III] main component in the observed-frame. We measured it for sources whose nominal redshift was obtained from UV emission lines position (usually C IV). We did not perform the measurement when: (...) the [O III] line was not reliable for such an estimation, (−) the available spectrum was already rest-framed, (~) the redshift provided by the reference was estimated from the [O III] line

(c) Quality flag for the spectra: 1 = good, 0 = not completely reliable, X = bad (excluded from the sample during the analysis)

Table B.2: Measured properties for both the line profiles (C IV and [O III]). We report as an example the measurements for the first 10 objects in the complete sample. The table for the complete sample will be available as a machine readable file in the online version of the paper.

C IV		J0042-1020	J1018+0548	J1024+1819	J1037+2135	J1042+1957	J1117+1311	J1126-0126	J1202-0054	J1312+0841	J1320-0523
FWHM	[km s^{-1}]	6042	3327	5078	3492	1970	3076	3984	3218	2244	4608
FWHM blue	[km s^{-1}]	6261	3875	6567	3984	2058	3327	5341	3699	2889	7158
FWHM red	[km s^{-1}]	5823	2780	3590	2999	1883	2824	2627	2736	1598	2058
σ	[km s^{-1}]	4304	4149	3542	4625	5484	4328	5271	3666	4944	5443
centroid shift	[km s^{-1}]	−2290	−844	−1512	−1274	−430	115	298	−923	−468	877
IPV_{10}	[km s^{-1}]	777	547	657	547	471	646	788	482	679	974
IPV_{20}	[km s^{-1}]	1565	1138	1324	1105	996	1357	1587	985	1379	1970
IPV_{30}	[km s^{-1}]	2430	1795	2014	1729	1675	2156	2430	1543	2156	3032
IPV_{40}	[km s^{-1}]	3371	2594	2758	2441	2638	3076	3360	2211	3065	4137
IPV_{50}	[km s^{-1}]	4455	3579	3590	3327	4104	4170	4466	3032	4170	5385
IPV_{60}	[km s^{-1}]	5768	4892	4619	4498	6031	5516	5877	4104	5691	6862
IPV_{70}	[km s^{-1}]	7443	6677	5976	6184	8515	7257	7869	5571	7804	8909
IPV_{80}	[km s^{-1}]	9807	9183	7957	8877	11985	9675	11131	7804	10836	12018
IPV_{90}	[km s^{-1}]	13846	13463	11361	14524	17950	13780	16976	11679	16024	17348
$AI_{05\text{-}95}$		1.1083	0.8693	0.8372	0.9176	0.8304	1.8229	1.3046	0.6646	1.3728	1.7281
$AI_{10\text{-}90}$		0.9722	1.0000	1.0000	1.0000	1.0476	1.1852	1.1176	1.0000	1.2143	1.1190
$AI_{20\text{-}80}$		1.0141	1.0800	1.0508	1.0612	1.1163	1.3846	1.3016	1.0000	1.4706	1.2222
$AI_{30\text{-}70}$		1.0367	1.1299	1.0674	1.0789	1.1549	1.5584	1.4130	1.0143	1.6986	1.3277
$AI_{40\text{-}60}$		1.0533	1.1545	1.0826	1.0841	1.1909	1.6762	1.5372	1.0200	1.8866	1.4868
MAD	[km s^{-1}]	3088	2818	2527	2917	3563	2999	3539	2424	3333	3817
EW	[Å]	29.06	63.32	29.12	46.44	52.33	21.99	46.32	105.79	30.05	31.72
AS		0.9207	0.7135	0.5420	0.7484	0.9119	0.8443	0.4885	0.7357	0.5510	0.2853
$shape_{BG92}$		2.2518	2.6020	2.2371	2.3793	3.0444	2.8932	2.3434	2.3912	3.0732	2.4964
$shape_{W93}$		−1.0888	−1.5428	−0.9569	−1.2445	−1.8667	−1.9395	−1.3984	−1.3299	−2.1366	−1.5368
α_β		0.0353	0.1497	0.0366	0.0705	0.2944	0.2331	0.3613	0.0969	0.4732	0.3955
$shift_{BG92}$		−0.3797	−0.2595	−0.3150	−0.3720	−0.2210	0.0354	0.0638	−0.2899	−0.2166	0.1572
$c(9/10)$		−2267	−820	−1310	−1206	−430	114	347	−895	−453	1162
AI_M		−0.0499	−0.1619	−0.1741	−0.1280	−0.2443	−0.1959	−0.4185	−0.1225	−0.3783	−0.4982
KrI_M		0.3482	0.2556	0.4008	0.3131	0.2398	0.1973	0.2526	0.2852	0.1798	0.2379
total flux	[10^{-15} erg cm^{-2}s^{-1}]	11.31	12.02	8.89	25.78	14.00	6.14	5.79	10.52	6.45	16.66
Sk		−0.1909	0.5627	0.6124	0.4491	0.6430	−1.1567	−0.2063	1.3356	−0.4325	−0.9422
exc_{Kr}		2.5897	4.0648	3.3246	6.5021	4.3875	3.7837	4.2082	5.9757	3.7658	3.3130
S		1.4038	0.8019	1.4337	0.7550	0.3592	0.7106	0.7558	0.8778	0.4538	0.8465
$IPV_{10\text{-}90}$		0.0561	0.0407	0.0578	0.0377	0.0262	0.0469	0.0464	0.0412	0.0423	0.0562
$EW(IPV_{10})$	[Å]	2.8932	6.2724	2.8828	4.6763	5.2057	2.1761	4.6443	10.6092	3.0058	3.1719

[O III]		J0042-1020	J1018+0548	J1024+1819	J1037+2135	J1042+1957	J1117+1311	J1126-0126	J1202-0054	J1312+0841	J1320-0523
FWHM	[km s^{-1}]	526	786	1465	2108	918	749	658	966	892	1471
FWHM blue	[km s^{-1}]	529	778	1465	2898	1116	754	752	982	885	1470
FWHM red	[km s^{-1}]	522	794	1464	1317	720	743	564	951	899	1471
σ	[km s^{-1}]	524	335	624	865	535	338	679	741	596	622
IPV$_{10}$	[km s^{-1}]	65.8	83.8	156	239	137	80.2	158	129	116	157
IPV$_{20}$	[km s^{-1}]	133	169	315	478	281	163	327	263	235	317
IPV$_{30}$	[km s^{-1}]	204	257	479	723	435	247	510	404	361	482
IPV$_{40}$	[km s^{-1}]	280	350	652	974	600	336	699	560	500	655
IPV$_{50}$	[km s^{-1}]	366	450	839	1236	776	433	896	742	661	843
IPV$_{60}$	[km s^{-1}]	468	561	1047	1517	960	541	1108	967	857	1052
IPV$_{70}$	[km s^{-1}]	605	692	1289	1835	1158	667	1352.	1276	1113	1295
IPV$_{80}$	[km s^{-1}]	843	855	1594	2220	1391	827	1665	1738	1470	1601
IPV$_{90}$	[km s^{-1}]	1593	1098	2045	2773	1710	1069	2178	2518	2024	2055
AI$_{05\text{-}95}$		1.610	1.000	0.999	1.393	1.707	1.036	1.849	1.211	0.986	0.999
AI$_{10\text{-}90}$		1.037	1.000	0.992	1.057	1.101	1.000	1.200	1.019	1.010	0.992
AI$_{20\text{-}80}$		1.018	1.000	0.996	1.114	1.233	1.000	1.453	1.032	1.000	0.996
AI$_{30\text{-}70}$		1.024	1.000	0.998	1.167	1.373	1.005	1.671	1.058	0.997	0.998
AI$_{40\text{-}60}$		1.035	1.000	0.998	1.214	1.501	1.007	1.835	1.080	0.990	0.998
MAD	[km s^{-1}]	319	266	496	699	431	262	529	534	446	498
EW	[Å]	10.67	38.01	13.19	35.77	42.42	25.66	20.55	74.61	65.17	8.61
AS		0.986	1.019	0.997	0.453	0.644	0.984	0.749	0.967	1.015	0.998
shape$_{BG92}$		2.082	2.058	2.058	1.942	2.438	2.061	3.196	2.239	2.327	2.059
shape$_{W93}$		−0.802	−0.769	−0.770	−0.790	−1.346	−0.772	−2.045	−1.004	−1.128	−0.770
α_β		2.3×10^{-3}	0.00	-2.0×10^{-4}	1.7×10^{-1}	3.8×10^{-1}	1.6×10^{-3}	6.4×10^{-1}	4.9×10^{-2}	-9.1×10^{-3}	2.0×10^{-4}
shift$_{BG92}$		−2.044	−1.166	−0.687	−0.636	−0.499	0.078	0.979	−1.148	−0.571	0.745
$c(9/10)$		−1074	−916	−1005	−1210	−445	58.1	650	−1107	−510	1095
AI$_M$		-7.9×10^{-3}	7.5×10^{-3}	-2.9×10^{-4}	-3.7×10^{-1}	-4.2×10^{-1}	-6.8×10^{-3}	$-5.0 \times 10^{e-1}$	-6.1×10^{-2}	1.3×10^{-2}	0.00
KrI$_M$		0.444	0.456	0.456	0.422	0.289	0.455	0.219	0.381	0.347	0.455
total flux	[10^{-15} erg cm^{-2}s^{-1}]	0.591	0.970	0.631	2.693	1.480	0.994	0.469	1.520	2.055	0.786
Sk		-7.6×10^{-1}	3.7×10^{-6}	0.00	-3.8×10^{-1}	-5.3×10^{-1}	-4.8×10^{-1}	-5.4×10^{-1}	-2.5×10^{-1}	1.7×10^{-2}	0.00
exc_{Kr}		6.7	-4.4×10^{-4}	-3.7×10^{-5}	-2.0×10^{-1}	-3.1×10^{-1}	2.1	3.8×10^{-1}	1.6	9.2×10^{-1}	-3.6×10^{-5}
S		1.002	2.348	2.347	2.436	1.717	2.216	0.969	1.304	1.497	2.363
IPV$_{10\text{-}90}$		0.0413	0.0763	0.0764	0.0861	0.0802	0.0750	0.0726	0.0514	0.0571	0.0766

List of Acronyms

AGN: Active Galactic Nuclei
BH: Black Hole
BLR: Broad Line Region
CV: Cross Validation
DR: Data Release
EN: Elastic Net
EV1: Eigenvector 1
EW: Equivalent Width
FUSE: Far Ultraviolet Spectroscopic Explorer
FWHM: Full Width Half Maximum
GNIRS: Gemini Near Infrared Spectrograph
HIL: High Ionization Line
HST: Hubble Space Telescope
HWHM: Half Width Half Maximum
IDL: Interactive Data Language
IPV: Inter Percentile Velocity width
IQR: Inter Quartile Range
KPNO: Kitt Peak National Observatory
Lasso: Least Absolute Shrinkage and Selection Operator
LBT: Large Binocular Telescope
LIL: Low Ionization Line
MAD: Mean Absolute deviation
MODS: Multi-Object Double Spectrographs
NLR: Narrow Line Region
NTT: New Technology Telescope
OLS: Ordinary Least Squares
PCA: Principal Component Analysis
QSO: Quasi-Stellar Object (quasar)
RM: reverberation mapping
RMS: Root Mean Square
SDSS: Sloan Digital Sky Survey
SE: Single Epoch
SED: Spectral Energy Distribution
SMBH: Supermassive Black Hole
SofI: Son of ISAAC infrared spectrograph and imaging camera
TOR: Toroidal Obscuring Region
WISE: Wide-field Infrared Survey Explorer

Bibliography

Abazajian, K. N., Adelman-McCarthy, J. K., Agüeros, M. A., et al. 2009, ApJS, 182, 543
Antonucci, R. 1993, ARA&A, 31, 473
Antonucci, R. R. J., & Miller, J. S. 1985, ApJ, 297, 621
Baskin, A., & Laor, A. 2004, MNRAS, 350, L31
—. 2005, MNRAS, 356, 1029
Bentz, M. C., Peterson, B. M., Netzer, H., Pogge, R. W., & Vestergaard, M. 2009, ApJ, 697, 160
Bentz, M. C., Denney, K. D., Cackett, E. M., et al. 2006, ApJ, 651, 775
Bentz, M. C., Denney, K. D., Grier, C. J., et al. 2013, ApJ, 767, 149
Bisogni, S., Marconi, A., & Risaliti, G. 2017, MNRAS, 464, 385
Blandford, R. D., & McKee, C. F. 1982, ApJ, 255, 419
Bon, E., Popović, L. Č., Ilić, D., & Mediavilla, E. 2006, New Astron. Rev., 50, 716
Boroson, T. A. 2011, ApJLett, 735, L14
Boroson, T. A., & Green, R. F. 1992, ApJS, 80, 109
Brotherton, M. S. 1996, ApJS, 102, 1
Cackett, E. M., Gültekin, K., Bentz, M. C., et al. 2015, ApJ, 810, 86
Cappellari, M., Verolme, E. K., van der Marel, R. P., et al. 2002, ApJ, 578, 787
Collin, S., Kawaguchi, T., Peterson, B. M., & Vestergaard, M. 2006, A&A, 456, 75
Collinson, J. S., Ward, M. J., Done, C., et al. 2015, MNRAS, 449, 2174
De Robertis, M. 1985, ApJ, 289, 67
Decarli, R., Dotti, M., & Treves, A. 2011, MNRAS, 413, 39
Denney, K. D. 2012, ApJ, 759, 44
Denney, K. D., Bentz, M. C., Peterson, B. M., et al. 2006, ApJ, 653, 152
Denney, K. D., Peterson, B. M., Pogge, R. W., et al. 2010, ApJ, 721, 715
Denney, K. D., Horne, K., Shen, Y., et al. 2016, ApJS, 224, 14
di Serego Alighieri, S., Cimatti, A., Fosbury, R. A. E., & Hes, R. 1997, A&A, 328, 510
Diamond-Stanic, A. M., Rieke, G. H., & Rigby, J. R. 2009, ApJ, 698, 623
Dietrich, M., Hamann, F., Shields, J. C., et al. 2002, ApJ, 581, 912
Dietrich, M., Mathur, S., Grupe, D., & Komossa, S. 2009, ApJ, 696, 1998
Dietrich, M., Kollatschny, W., Peterson, B. M., et al. 1993, ApJ, 408, 416
Dietrich, M., Peterson, B. M., Albrecht, P., et al. 1998, ApJS, 115, 185
Elitzur, M. 2008, New Astron. Rev., 52, 274
—. 2012, ApJLett, 747, L33
Elitzur, M., & Ho, L. C. 2009, ApJLett, 701, L91
Elitzur, M., & Shlosman, I. 2006, ApJLett, 648, L101
Emmering, R. T., Blandford, R. D., & Shlosman, I. 1992, ApJ, 385, 460
Fabian, A. C. 2012, ARA&A, 50, 455
Ferland, G. J., Porter, R. L., van Hoof, P. A. M., et al. 2013, Rev. Mex. Astron. Astrofis., 49, 137
Ferrarese, L., & Ford, H. 2005, Space Sci. Rev., 116, 523
Ferrarese, L., & Merritt, D. 2000, ApJLett, 539, L9

Fitzpatrick, E. L. 1999, PASP, 111, 63
Francis, P. J., Hewett, P. C., Foltz, C. B., & Chaffee, F. H. 1992, ApJ, 398, 476
Francis, P. J., & Wills, B. J. 1999, in Astronomical Society of the Pacific Conference Series, Vol. 162, Quasars and Cosmology, ed. G. Ferland & J. Baldwin, 363
Fritz, J., Franceschini, A., & Hatziminaoglou, E. 2006, MNRAS, 366, 767
Fruchter, A. S., & Hook, R. N. 2002, PASP, 114, 144
Gaskell, C. M., & Sparke, L. S. 1986, ApJ, 305, 175
Gebhardt, K., Bender, R., Bower, G., et al. 2000, ApJLett, 539, L13
Genzel, R., Eisenhauer, F., & Gillessen, S. 2010, Reviews of Modern Physics, 82, 3121
Goad, M. R., Korista, K. T., & Ruff, A. J. 2012, MNRAS, 426, 3086
Goeman, J., Meijer, R., & Chaturvedi, N. 2016, L1 and L2 penalized regression models
Goldoni, P., Royer, F., François, P., et al. 2006, in Presented at the Society of Photo-Optical Instrumentation Engineers (SPIE) Conference, Vol. 6269, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series
Greene, J. E., & Ho, L. C. 2005, ApJ, 630, 122
Greene, J. E., Peng, C. Y., & Ludwig, R. R. 2010, ApJ, 709, 937
Greene, W. 2003, Econometric Analysis (Pearson Education)
Grier, C. J., Peterson, B. M., Pogge, R. W., et al. 2012, ApJ, 755, 60
Hamuy, M., Suntzeff, N. B., Heathcote, S. R., et al. 1994, PASP, 106, 566
Hamuy, M., Walker, A. R., Suntzeff, N. B., et al. 1992, PASP, 104, 533
Heckman, T. M., Kauffmann, G., Brinchmann, J., et al. 2004, ApJ, 613, 109
Heckman, T. M., Miley, G. K., van Breugel, W. J. M., & Butcher, H. R. 1981, ApJ, 247, 403
Ho, L. C., Goldoni, P., Dong, X.-B., Greene, J. E., & Ponti, G. 2012, ApJ, 754, 11
Ho, L. C., & Kim, M. 2015, ApJ, 809, 123
Hoerl, A. E., & Kennard, R. W. 1970, Technometrics, 12, 55
Jun, H. D., Im, M., Lee, H. M., et al. 2015, ApJ, 806, 109
Kaspi, S., Brandt, W. N., Maoz, D., et al. 2007, ApJ, 659, 997
Kaspi, S., Smith, P. S., Netzer, H., et al. 2000, ApJ, 533, 631
Kelson, D. D. 2003, PASP, 115, 688
Kollatschny, W., & Zetzl, M. 2011, Nature, 470, 366
Koratkar, A., & Blaes, O. 1999, PASP, 111, 1
Kormendy, J. 1988, ApJ, 325, 128
Kormendy, J., & Richstone, D. 1995, ARA&A, 33, 581
Kriss, G. A. 2001, in Astronomical Society of the Pacific Conference Series, Vol. 224, Probing the Physics of Active Galactic Nuclei, ed. B. M. Peterson, R. W. Pogge, & R. S. Polidan, 45
Laor, A., Fiore, F., Elvis, M., Wilkes, B. J., & McDowell, J. C. 1994, ApJ, 435, 611
—. 1997, ApJ, 477, 93
Lopez, S., D'Odorico, V., Ellison, S. L., et al. 2016, ArXiv e-prints, arXiv:1607.08776
Magorrian, J., Tremaine, S., Richstone, D., et al. 1998, AJ, 115, 2285
Marconi, A., & Hunt, L. K. 2003, ApJLett, 589, L21
Marconi, A., Risaliti, G., Gilli, R., et al. 2004, MNRAS, 351, 169
Markwardt, C. B. 2009, in Astronomical Society of the Pacific Conference Series, Vol. 411, Astronomical Data Analysis Software and Systems XVIII, ed. D. A. Bohlender, D. Durand, & P. Dowler, 251
Marziani, P., Sulentic, J. W., Dultzin-Hacyan, D., Calvani, M., & Moles, M. 1996, ApJS, 104, 37
Marziani, P., Sulentic, J. W., Plauchu-Frayn, I., & del Olmo, A. 2013, A&A, 555, A89
Mathews, W. G., & Ferland, G. J. 1987, ApJ, 323, 456
Mathews, W. G., & Wampler, E. J. 1985, PASP, 97, 966
McConnell, N. J., & Ma, C.-P. 2013, ApJ, 764, 184

McLure, R. J., & Dunlop, J. S. 2004, MNRAS, 352, 1390

McLure, R. J., & Jarvis, M. J. 2002, MNRAS, 337, 109

Mejia-Restrepo, J. E., Trakhtenbrot, B., Lira, P., Netzer, H., & Capellupo, D. M. 2016, MNRAS, arXiv:1603.03437

Metzroth, K. G., Onken, C. A., & Peterson, B. M. 2006, ApJ, 647, 901

Mittaz, J. P. D., Penston, M. V., & Snijders, M. A. J. 1990, MNRAS, 242, 370

Modigliani, A., Goldoni, P., Royer, F., et al. 2010, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 7737, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series

Mulchaey, J. S., Koratkar, A., Ward, M. J., et al. 1994, ApJ, 436, 586

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

Nagao, T., Marconi, A., & Maiolino, R. 2006, A&A, 447, 157

Nenkova, M., Sirocky, M. M., Ivezić, Ž., & Elitzur, M. 2008a, ApJ, 685, 147

Nenkova, M., Sirocky, M. M., Nikutta, R., Ivezić, Ž., & Elitzur, M. 2008b, ApJ, 685, 160

Netzer, H., Brotherton, M. S., Wills, B. J., et al. 1995, ApJ, 448, 27

Netzer, H., Lira, P., Trakhtenbrot, B., Shemmer, O., & Cury, I. 2007, ApJ, 671, 1256

Neufeld, D. A., Maloney, P. R., & Conger, S. 1994, ApJLett, 436, L.127

Onken, C. A., & Peterson, B. M. 2002, ApJ, 572, 746

Osterbrock, D. E., & Ferland, G. J. 2006, Astrophysics of gaseous nebulae and active galactic nuclei

Pancoast, A., Brewer, B. J., Treu, T., et al. 2014, MNRAS, 445, 3073

Pâris, I., Petitjean, P., Rollinde, E., et al. 2011, A&A, 530, A50

Pâris, I., Petitjean, P., Ross, N. P., et al. 2016, ArXiv e-prints, arXiv:1608.06483

Pedregosa, F., Varoquaux, G., Gramfort, A., et al. 2011, Journal of Machine Learning Research, 12, 2825

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

Peterson, B. M. 2004, in IAU Symposium, Vol. 222, The Interplay Among Black Holes, Stars and ISM in Galactic Nuclei, ed. T. Storchi-Bergmann, L. C. Ho, & H. R. Schmitt, 15–20

Peterson, B. M., & Wandel, A. 1999, ApJLett, 521, L95

Peterson, B. M., Wanders, I., Bertram, R., et al. 1998, ApJ, 501, 82

Peterson, B. M., McHardy, I. M., Wilkes, B. J., et al. 2000, ApJ, 542, 161

Peterson, B. M., Berlind, P., Bertram, R., et al. 2002, ApJ, 581, 197

Peterson, B. M., Ferrarese, L., Gilbert, K. M., et al. 2004, ApJ, 613, 682

Peterson, B. M., Grier, C. J., Horne, K., et al. 2014, ApJ, 795, 149

Pita, S., Goldoni, P., Boisson, C., et al. 2014, A&A, 565, A12

Pogge, R. W., Atwood, B., Brewer, D. F., et al. 2010, in Proc. SPIE, Vol. 7735, Ground-based and Airborne Instrumentation for Astronomy III, 77350A

Rees, M. J. 1984, ARA&A, 22, 471

Reichert, G. A., Rodriguez-Pascual, P. M., Alloin, D., et al. 1994, ApJ, 425, 582

Richards, G. T. 2012, in Astronomical Society of the Pacific Conference Series, Vol. 460, AGN Winds in Charleston, ed. G. Chartas, F. Hamann, & K. M. Leighly, 67

Richards, G. T., Strauss, M. A., Fan, X., et al. 2006, AJ, 131, 2766

Riffel, R. A. 2010, Ap&SS, 327, 239

Risaliti, G., & Lusso, E. 2015, ApJ, 815, 33

Risaliti, G., Salvati, M., & Marconi, A. 2011, MNRAS, 411, 2223

Rosen, S. R., Webb, N. A., Watson, M. G., et al. 2016, A&A, 590, A1

Rothberg, B., Kuhn, O., Edwards, M. L., et al. 2016, ArXiv e-prints, arXiv:1608.00037

Santos-Lleó, M., Chatzichristou, E., de Oliveira, C. M., et al. 1997, ApJS, 112, 271

Santos-Lleó, M., Clavel, J., Schulz, B., et al. 2001, A&A, 369, 57

Sargent, W. L. W., Young, P. J., Lynds, C. R., et al. 1978, ApJ, 221, 731

Schlafly, E. F., & Finkbeiner, D. P. 2011, ApJ, 737, 103
Schmidt, M., & Green, R. F. 1983, ApJ, 269, 352
Schneider, D. P., Hall, P. B., Richards, G. T., et al. 2007, AJ, 134, 102
Schneider, D. P., Richards, G. T., Hall, P. B., et al. 2010, AJ, 139, 2360
Seyfert, C. K. 1943, ApJ, 97, 28
Shakura, N. I., & Sunyaev, R. A. 1973, A&A, 24, 337
Shang, Z., Wills, B. J., Wills, D., & Brotherton, M. S. 2007, AJ, 134, 294
Shang, Z., Brotherton, M. S., Green, R. F., et al. 2005, ApJ, 619, 41
Shang, Z., Brotherton, M. S., Wills, B. J., et al. 2011, ApJS, 196, 2
Shen, Y. 2013, Bulletin of the Astronomical Society of India, 41, 61
Shen, Y., & Ho, L. C. 2014, Nature, 513, 210
Shen, Y., & Liu, X. 2012, ApJ, 753, 125
Shen, Y., Richards, G. T., Strauss, M. A., et al. 2011, ApJS, 194, 45
Silk, J. 2013, ApJ, 772, 112
Stalevski, M., Fritz, J., Baes, M., Nakos, T., & Popović, L. Č. 2012, MNRAS, 420, 2756
Sulentic, J., & Marziani, P. 2015, Frontiers in Astronomy and Space Sciences, 2, 6
Sulentic, J. W., Bachev, R., Marziani, P., Negrete, C. A., & Dultzin, D. 2007, ApJ, 666, 757
Tarchi, A. 2012, in IAU Symposium, Vol. 287, Cosmic Masers - from OH to H0, ed. R. S. Booth, W. H. T. Vlemmings, & E. M. L. Humphreys, 323–332
Tibshirani, R. 1996, Journal of the Royal Statistical Society (Series B), 58, 267
Trakhtenbrot, B., & Netzer, H. 2012, MNRAS, 427, 3081
Tran, H. D., Cohen, M. H., & Goodrich, R. W. 1995, AJ, 110, 2597
Trevese, D., Perna, M., Vagnetti, F., Saturni, F. G., & Dadina, M. 2014, ApJ, 795, 164
Urry, C. M., & Padovani, P. 1995, PASP, 107, 803
Vacca, W. D., Cushing, M. C., & Rayner, J. T. 2003, PASP, 115, 389
van der Marel, R. P., & Franx, M. 1993, ApJ, 407, 525
van Dokkum, P. G. 2001, PASP, 113, 1420
Van Gorkom, K. J., Wardle, J. F. C., Rauch, A. P., & Gobeille, D. B. 2015, MNRAS, 450, 4240
Vanden Berk, D. E., Richards, G. T., Bauer, A., et al. 2001, AJ, 122, 549
Vanden Berk, D. E., Shen, J., Yip, C.-W., et al. 2006, AJ, 131, 84
Verner, E. M., Verner, D. A., Korista, K. T., et al. 1999, ApJS, 120, 101
Vernet, J., Dekker, H., D'Odorico, S., et al. 2011, A&A, 536, A105
Véron-Cetty, M.-P., Joly, M., & Véron, P. 2004, A&A, 417, 515
Vestergaard, M., & Osmer, P. S. 2009, ApJ, 699, 800
Vestergaard, M., & Peterson, B. M. 2006, ApJ, 641, 689
Wang, J.-G., Dong, X.-B., Wang, T.-G., et al. 2009, ApJ, 707, 1334
Whittle, M. 1985, MNRAS, 213, 1
Williams, M. J., Bureau, M., & Cappellari, M. 2010, MNRAS, 409, 1330
Wills, B. J., Brotherton, M. S., Fang, D., Steidel, C. C., & Sargent, W. L. W. 1993, ApJ, 415, 563
Wills, B. J., & Browne, I. W. A. 1986, ApJ, 302, 56
Wills, B. J., Netzer, H., & Wills, D. 1985, ApJ, 288, 94
Wills, B. J., Thompson, K. L., Han, M., et al. 1995, ApJ, 447, 139
Wright, E. L., Eisenhardt, P. R. M., Mainzer, A. K., et al. 2010, AJ, 140, 1868
Yip, C., Connolly, A., vanden Berk, D., et al. 2004a, in Astronomical Society of the Pacific Conference Series, Vol. 311, AGN Physics with the Sloan Digital Sky Survey, ed. G. T. Richards & P. B. Hall, 17
Yip, C. W., Connolly, A. J., Vanden Berk, D. E., et al. 2004b, AJ, 128, 2603
York, D. G., Adelman, J., Anderson, Jr., J. E., et al. 2000, AJ, 120, 1579
Zhu, L., Zhang, S. N., & Tang, S. 2009, ApJ, 700, 1173

Zou, H., & Hastie, T. 2005, Journal of the Royal Statistical Society, Series B, 67, 301