UNIVERSITY OF SOUTHAMPTON
FACULTY OF ENGINEERING, SCIENCE & MATHEMATICS
SCHOOL OF PHYSICS & ASTRONOMY

SEARCHING FOR THE GHOST OF LYMAN ALPHA

A RESONANCE LINE DRIVING SIGNATURE IN THE ABSORPTION TROUGHS OF BROAD ABSORPTION LINE QUASARS

Matthew W. M. North

Submitted for the degree of Master of Philosophy

November, 2004

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF ENGINEERING, SCIENCE & MATHEMATICS
SCHOOL OF PHYSICS & ASTRONOMY

Master of Philosophy

SEARCHING FOR THE GHOST OF LYMAN ALPHA

By Matthew W. M. North

Within this thesis I have searched the broad absorption line quasar (BALQSO) sample presented by Reichard and collaborators for objects exhibiting the so-called "ghost of Lyman alpha". This ghost manifests as a hump near -5900 km/s in the troughs of broad absorption lines and provides strong evidence for the importance of line-driving in powering the outflows from BALQSOs.

Of the 224 BALQSOs, selected by Reichard and collaborators, from the Sloan Digital Sky Survey (SDSS) Early Data Release (EDR), 198 satisfy my redshift constraints, and 58 show clear evidence of multiple trough (MT) structure in their CIV λ1550Å broad absorption line (BAL). A composite spectrum constructed from this MT sample already shows evidence for a ghost feature. Narrowing this classification scheme further, I define a set of 36 objects that individually show evidence of a ghost feature and then apply additional cuts to arrive at a final sample that contains my 7 strongest ghost candidates. Despite its limited size, this final sample triples the number of known BALQSOs with clear ghost signatures and should make an excellent basis for detailed follow-up studies.

As a consequence of these initial results, I have developed an algorithm for the purpose of identifying BALQSOs within the latest SDSS Data Release 3 (DR3). This algorithm is designed to improve upon the SDSS determined redshifts and so enable the quick and accurate selection of BALQSOs exhibiting Ghost signatures. With the SDSS DR3 currently containing ~50,000 quasar spectra, this routine will likely uncover ~5,000-10,000 BALQSOs of which ~200-400 objects may well display the Ghost signature.

This thesis details the algorithm and presents results from its testing phase. As an indication of its capabilities, I include herein the spectra of 5 new BALQSOs from the SDSS EDR, previously overlooked by Reichard.

List of Figures

ACKNOWLEDGEMENTS

I wish to acknowledge and thank Dr Christian Knigge for his expert tutelage, patience and support. His enthusiasm and scientific excellence have been an education in their own right. I am forever in his debt for the opportunity he has provided me and for his guidance throughout this work.

I wish to thank my beautiful and loving wife, without her constant belief, encouragement and support, this thesis would not exist.

I thank the Sloan Digital Sky Survey for their freely available spectroscopic data, without which this study would not have been possible.

I would also like to thank Dr Mike Goad for his invaluable assistance throughout this past year.

Finally, a warm 'thank you' to my office colleagues, Jose Galache, Vanessa McBride and Carolyn Brinkworth who have aided my progress and kept me going through to the completion of this thesis.

1.0 Introduction

Approximately 10%-20% of quasi-stellar objects (QSOs) exhibit broad, blue-shifted absorption troughs associated with strong ultraviolet (UV) emission lines, such as NV λ1240 Å, SiIV λ1400 Å and CIV λ1550 Å. These troughs have long been regarded as signs of large-scale outflows or winds, whose velocities (as inferred from the widths of the troughs) can reach 0.1-0.2 c (Korista [1992], see also § 1.2). These outflows remove mass, energy and momentum (both linear and angular) from the QSO and deposit them in the host galaxy. As a result, they can significantly affect the evolution of the QSO and the metal enrichment of its host (e.g. Artymowicz [1993]; and more recently Gupta, Srianand, Petijean & Ledou [2003]). Furthermore, there is increasing evidence (e.g. Reichard 2003b), that Broad Absorption Line QSOs (BALQSOs) are fundamentally the same objects as Broad Emission Line (BEL) QSOs, except that we are viewing the central engine through these powerful winds (see § 1.1). This, in turn, suggests that these outflows may be ubiquitous throughout the QSO community. If so, then the low incidence of BALQSOs among QSOs implies a relatively narrow wind opening-angle. Indeed, in the context of unified models, orientation is the key difference between all types of Active Galactic Nuclei (AGN) and QSOs. Consequently, analyses of the physics, as well as the overall statistics, of BALQSOs are also important to studies of QSOs and AGN more generally.

An observational feature that has promised to provide much insight into the physics of outflows from BALQSOs is the so-called "ghost of Lyman alpha" (Lyα) (Arav & Li 1994). This term refers to a hump near -5900 km/s, the Lyα - NV velocity separation, seen in the troughs of the broad absorption lines of some BALQSOs. This local maximum can be explained naturally if the outflow is radiatively accelerated via resonance line scattering (see § 1.4). However, despite the potential significance of the ghost of Lyα for our understanding of the (BAL)QSO population, the set of 4 objects discussed by Arav (1996) is currently still the only observational sample of BALQSOs exhibiting clear ghost signatures. It is thus the goal of this thesis to expand upon this sample by selecting new 'ghost candidates' from QSO spectra obtained from the Sloan Digital Sky Survey (SDSS) and to present a new tool for selecting BALQSOs (and, ultimately, even more ghost candidates) from large samples of QSOs.

In the remainder of this introduction, I briefly introduce AGN and BALQSOs (§1.1). I then discuss the key observational signature of BALQSOs – the P-Cygni profile – and explain its significance as a tracer of mass loss from these objects (§ 1.2). I go on to introduce the Balnicity Index (BI) (§ 1.3), which is a tool used for the formal classification of BALQSOs. I then describe the so-called "ghost of Lyα" itself and explain why it provides direct evidence that BALQSO winds are driven by radiation pressure mediated by spectral lines (§ 1.4). I finally conclude by giving an overview of the rest of this thesis (§ 1.5).

1.1 AGN, QSOs and BALQSOs

Active Galactic Nuclei (AGN) are amongst the most luminous objects known. Their extreme luminosities (10^{39}-10^{47} erg/s), coupled with observed short timescale variations precludes nuclear burning as the origin of this emission. Instead, the generally accepted paradigm for energy production in these sources is the release of gravitational potential energy from material falling into the deep potential well of a super-massive (10^{6}-10^{9} M$_\odot$) black hole located at their centre.

The AGN family encompasses a wide variety of objects with sometimes quite different observational characteristics (e.g. radio-loud vs radio-quiet sources and sources exhibiting broad [type I] vs narrow [type II] emission lines). However, it is now thought that all of these objects are fundamentally similar, with the observational differences reflecting mainly differences in viewing angle, luminosity and the presence or absence of powerful jets (for a good introduction to AGN unification models, see, for example, Peterson 1999).

This thesis deals with Broad Emission Line (BEL) and Broad Absorption Line (BAL) Quasars (QSOs). These objects represent the high luminosity tail of the AGN distribution, and are so intrinsically bright that they outshine the integrated light of their host galaxies by a large factor[1]

[1] This explains the origin of their name: in optical imaging surveys, QSOs usually appear point-like and hence resemble stars. The terms "quasar"/ "QSO" are thus a contraction/acronym of "quasi-stellar object".

The broad emission lines seen in BEL and BAL QSOs are thought to be formed deep within the potential well of the AGN, in a region containing high-density gas moving at high velocities. The main line formation mechanism is thought to be thermal emission, with the lines being broadened by Doppler shifting.

In the rest-frame ultraviolet (UV) and optical wavebands, BELQSOs are generally well described by a reddened power law continuum, superposed on which are the BELs associated mainly with permitted transitions of various ionic species. BALQSOs appear basically the same, but additionally exhibit broad absorption troughs located blueward of the strong UV resonance lines (most notably C IV, N V, Si IV but also C III] λ1909Å and others).

Weymann et al (1991; Hereafter WMFH91) and more recently Reichard et al (2003; Hereafter R2003b), have shown that BALQSOs and BELQSOs appear to be drawn from the same parent population. This fits nicely with the idea that all AGN are intrinsically similar, with most of the observational differences being due to different viewing angles. In the context of this unified framework, it should not come as a surprise that the dividing line between BELQSOs and BALQSOs can be blurry, and that distinguishing the two classes is not always a simple task.

One particular difficulty is differentiating between the narrow absorption lines (NALs) seen in many (BEL)QSOs and the broad absorption troughs that are the defining characteristic of BALQSOs. In practice, the dividing line between the two is usually set somewhat arbitrarily to 2000 km/s. Whereas BALs have long been recognized as signatures of large-scale outflows, NALs have traditionally been thought of as being formed in orbiting clouds of ionised gas located along our line of sight to the continuum source, but perhaps more associated with the host galaxy than the AGN itself. However, this view may be changing. For example, in the quasar unification model presented by Elvis (2000), the NALs are formed in essentially the same outflow that gives rise to BALs, with the observational difference again being due mainly to viewing angle: if the bright central engine is viewed across the flow, we see NALs; if it is viewed through the flow, we see BALs (see Fig 1). This type of unified model can therefore accommodate all three types of (type II) QSOs – BELQSOs with and without NALs and also BALQSOs – and even explains the relative numbers of the different types as a result of the outflow geometry/opening angle.

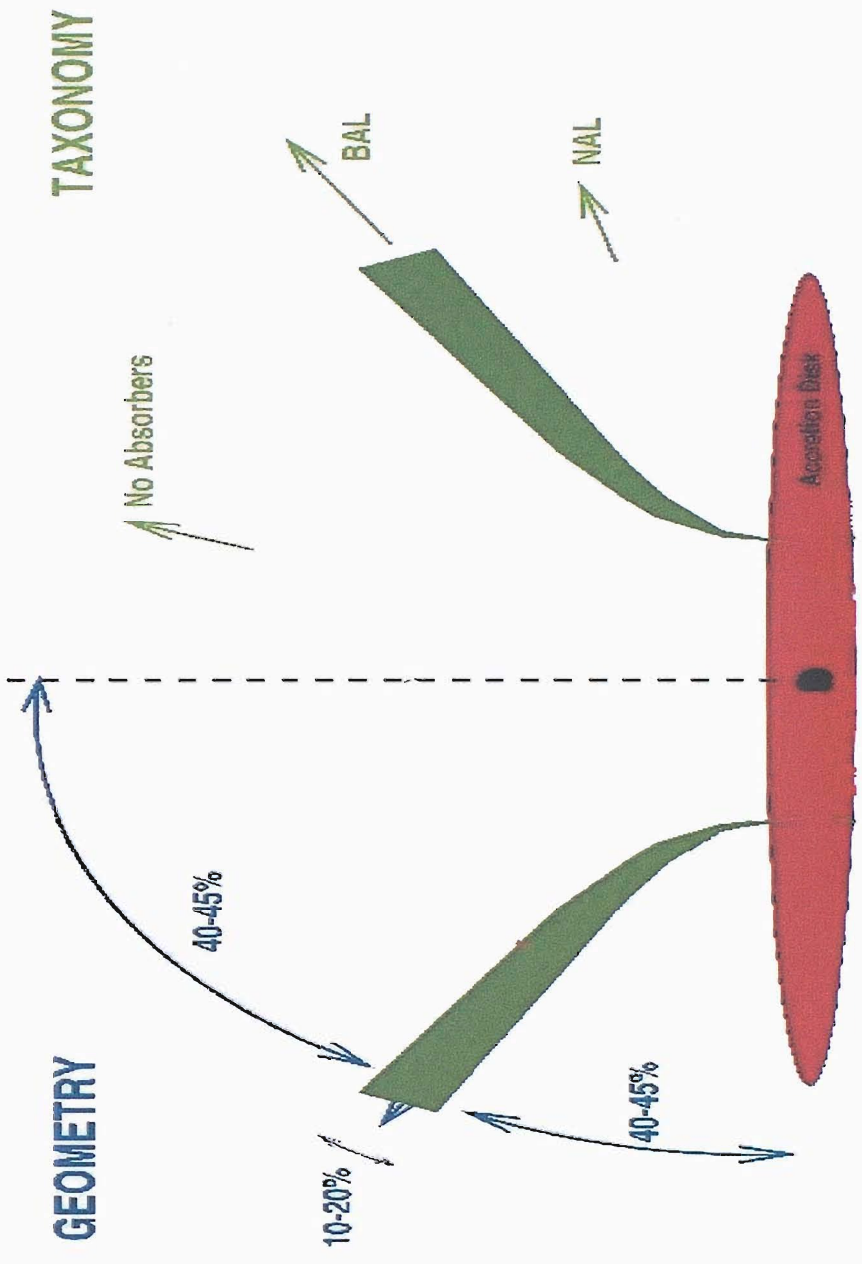

Figure 1 - Schematic of a simple unifying model for BELQSOs (with and without NALs) and BALQSOs. The green regions represent the outflow. The red ellipse represents the accretion disk with the main continuum source at its centre. Based on Fig.1 from Elvis (2000).

1.2 Outflows and P-Cygni Profiles

In the previous section, I have introduced the subset of AGN known as BALQSOs. I now outline the physical mechanisms that give rise to the observational signature that defines this class of objects. This signature is the broad, blue-shifted absorption seen particularly in the strongest UV resonance lines (such as N V, Si IV, C IV). Line profiles characterized by broad, blue-shifted absorption and associated red-shifted emission are often referred to as "P Cygni profiles", after the massive hot star in which they were first observed.

The presence of P Cygni profiles in the spectrum of an astronomical source provides strong evidence for an outflow from this source. This is most easily understood by considering the specific case of line formation taking place in a spherical, non-rotating outflow from a spherical star that emits a pure continuum spectrum. The outflow velocity field varies smoothly from zero at the stellar surface to v_{MAX} at infinity. We only consider a simple, strong resonance transition, and treat line formation as a pure, isotropic scattering process. The geometry of the situation is illustrated in figure 2, an extract from Knigge, 1995. The narrow cylinder of wind material 'in front' of the star (as seen by a distant observer) scatters stellar continuum photons emitted towards the observer out of his or her line of sight. Since all of this material is moving towards the observer according to the wind velocity law, the resulting flux deficit ("underlying absorption") in the observed spectrum is seen blueward of the rest wavelength, λ_T, out to

$$\lambda_{MIN} = [1 - (v_{MAX}/c)]\lambda_T \ .$$

Equation 1

Here, λ_T represents the wavelength of our strong resonance transition.

All other parts of the outflow (except the occulted region "behind" the star) can only scatter continuum photons into the observer's line of sight. By symmetry (see figure 3, also an extract from Knigge, 1995), the resulting added flux ("emission") must be symmetric about λ_T.

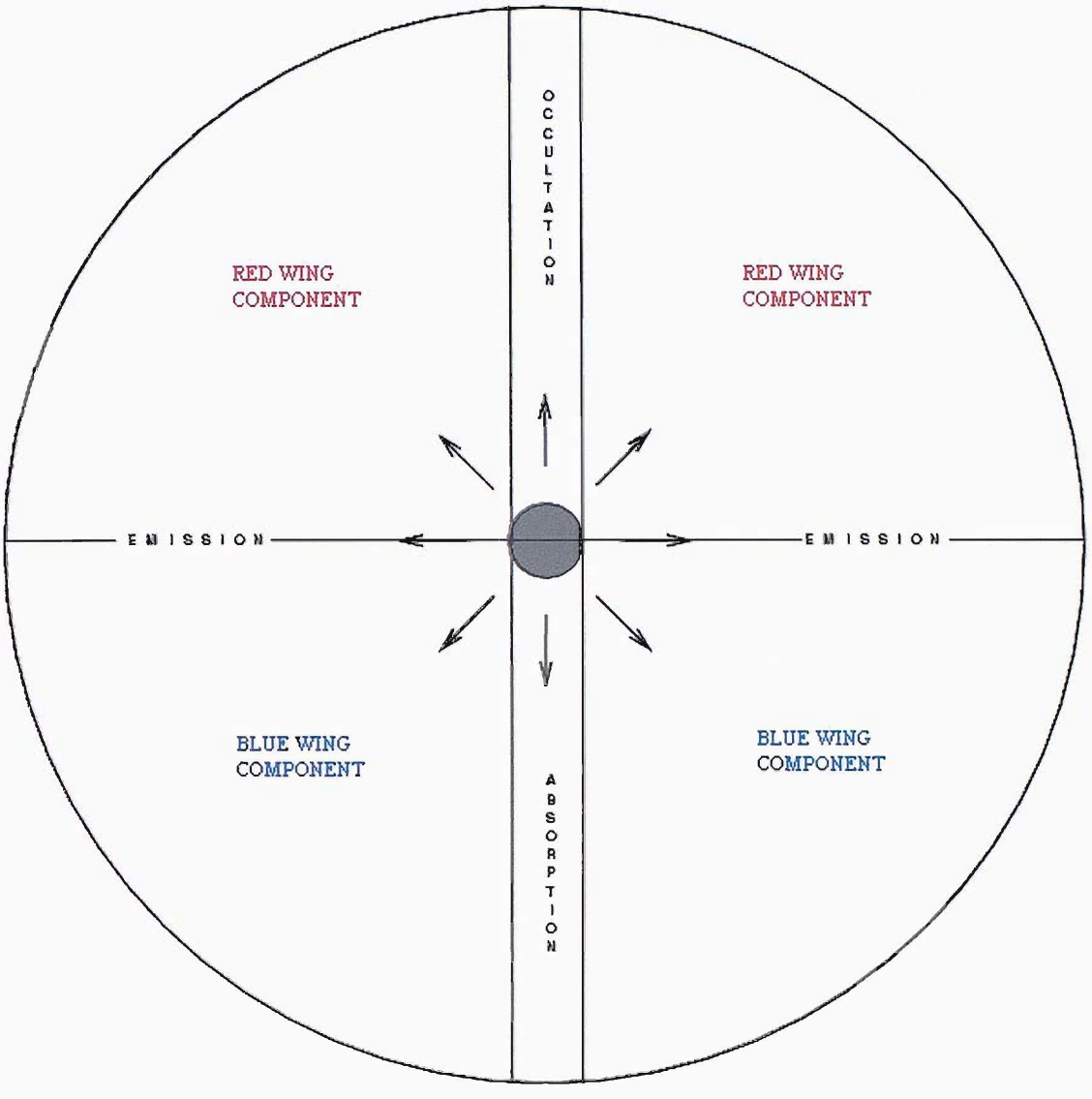

Figure 2 - Diagram of a spherical outflow with the different regions labelled according to their contribution towards a P-Cygni profile. The direction towards the observer is down.

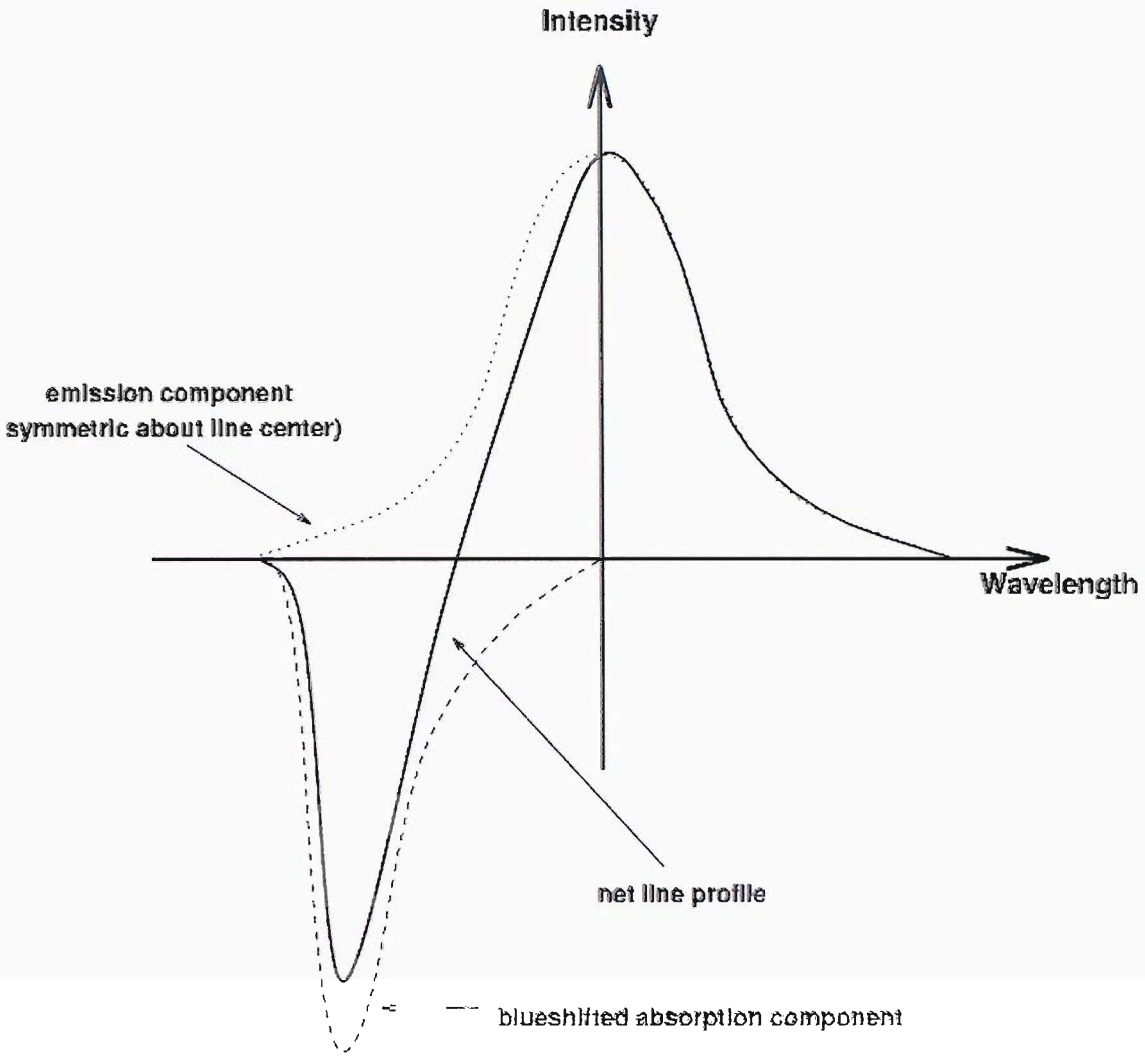

Figure 3 - Sketch of a typical P-Cygni profile and its decomposition into the underlying emission and absorption components[2].

[2] Note that the shape of the profile displayed in figure 3 isn't quite correct, since the net line flux is not zero. This is a reflection of the crudity of the diagram and by no means should be considered as reflective of a 'different dynamic'. The net line flux, within the scope of the simple scenario described in the text, will always be zero, since all viewing angles are equivalent and the wind does not create or destroy photons.

The observed line profile is then a superposition of these underlying absorption and emission components, i.e. a classic P Cygni profile.

Before concluding this section, it is important to note that the actual appearance of a P Cygni profile depends strongly on the mass loss rate, geometry and kinematics (velocity law and rotation) of the outflow, and also on the geometry of the continuum source (e.g. spherical or disk-like). In all of these respects, BALQSOs and their outflows differ strongly from the simple, spherical stellar wind I used to illustrate the underlying physics of P Cygni line formation. In fact, given the similarity between the BELs seen in BELQSOs and the red emission components associated with BALs in BALQSOs, it is likely that the emission components in the latter are produced by thermal emission, rather than scattering. However, the invariant aspect of the line formation mechanism described above is that, broad, blue-shifted absorption is produced when the bright continuum source is viewed directly through the outflow. Moreover, the shape of this absorption trough contains information about the velocity field of the outflow. This final statement is central to the explanation for the 'ghost of Lyα' presented in section 1.5.

1.3 The Balnicity Index

It has previously been noted that it is often difficult to distinguish whether a QSO is a BELQSO or BALQSO. To this end, WMFH91 devised a simple method to remove much of the subjectivity involved in taking this decision and assigned a "balnicity index" to each object. This index is determined as follows:

- Define a continuum, as sensibly as possible, between the rest wavelengths of the CIV and SiIV BELs.
- Define the systemic rest frame as accurately as possible by: (a) Centering the MgII BEL to its rest frequency wherever possible[3]; (b) If MgII is not available then CIII] should be used[3]; If CIII] is not available then CIV can be used, but only provided there are no significant absorption features in the BEL itself.

[3] Using MgII (or indeed CIII]) for the systemic rest frame stems, in the main, from research of observational data. I review this research and its implications in §4.1 where I discuss emission line shifts and redshift determination in more detail

- The BI is then calculated for CIV, and is a modified equivalent width of strong absorption features (expressed in km/s) and given by:

$$BI = -\int_{25000}^{3000} [1 - f(\mathrm{v})/0.9] C d\mathrm{v}$$

Equation 2

Here, WMFH91 define f(v) to be the normalised flux as a function of velocity displacement from the line centre. The value of C is initially zero, but is set to 1.0 whenever the quantity in brackets has been continuously positive over an interval of 2000 km/s. However, C is reset to zero whenever the quantity in brackets becomes negative.

1.4 The Ghost of Lyman Alpha

In a series of publications between 1994 and 1996, Nahum Arav and co-workers presented a dynamical model describing a signature that could arise if much of the primary energy used in generating BAL wind is provided by resonant line scattering. Specifically, Arav modelled the scattering of Lyα photons by NV ions present in the outflowing material and showed that, provided that certain criteria (listed below) are fulfilled, a hump of increased flux will be seen in the broad absorption trough of a BALQSO (see figure 4).

This feature can be understood in the context of the simple model for the formation of blue-shifted absorption troughs described in section 1.2. More specifically, the optical depth in a particular wavelength (= velocity) range is proportional to the column of material moving with velocities in this range. Now this column is proportional to the size of the region of the flow moving with this velocity, and this in turn is inversely proportional to the velocity gradient, dv/dr, in this region (my notation here retains the spherical outflow approximation, but this is not essential to the argument). Thus

$$\tau_{\mathrm{v}} \propto \left[\frac{d\mathrm{v}}{dr}\right]_{\mathrm{v}}^{-1}$$

Equation 3

Now suppose that the outflow itself is actually driven by the transfer of momentum associated with the resonance line scatterings that produce the BAL troughs. Moreover, consider an incident spectrum that is not a pure continuum but also contains a strong BEL due to Lyα. Now Lyman alpha is separated from the NV λ1240Å rest wavelength by only -5900 km/s in velocity space, so NV ions moving outwards at this velocity will see the BEL red-shifted into their rest wavelengths. They are thus able to scatter these Lyman alpha photons efficiently, leading to an enhanced momentum transfer and increased velocity gradient in this region of the flow. Equation 3 thus predicts a decrease in the optical depth around -5900 km/s (hereafter v_{GH}) in this model. This manifests itself as a 'hump' in all the BAL troughs created by the flow. The shape of this feature should directly reflect the profile of the Lyα BEL, and thus the resultant feature is appropriately named "the ghost of Lyα". It is important to realise that this scenario only holds for flows driven by radiation pressure mediated by spectral lines. Thus the presence of the ghost in the spectra of some systems provides direct evidence that this driving mechanism is the key to the outflows in these, and perhaps all, BALQSOs.

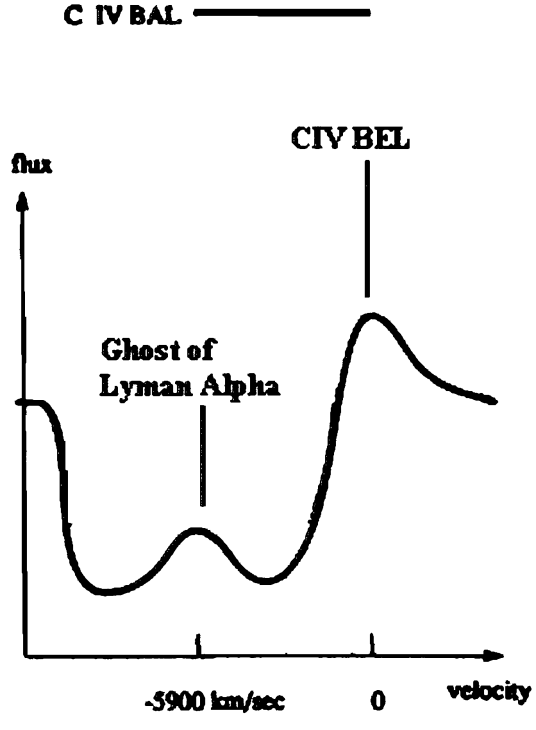

Figure 4 - Illustration of a Ghost signature in the CIV BAL.

Building on this theoretical framework, Arav (1996) listed a set of 5 characteristics that a BALQSO should possess in order to produce a strong ghost signature. These are:

1. A significant absorption trough in the BAL of CIV between –3000 and –9000 km/s must be present.

2. A strong Lyα emission line must be present in the intrinsic spectrum of the QSO.

3. A relatively narrow Lyα emission line should be present, as this provides a steep increase in radiation pressure at and around v_{GH} (this in turn will help to provide a clear ghost feature).

4. A visible BAL associated with NV should be observed as these ions must be present in the flow for this model to apply.

5. The Lyman forest, between 200 Å – 1000 Å, should be relatively low. This condition arises as a result of the large number of resonant transitions present throughout this region of the spectrum. If there is a lot of energy output in this zone then one should reasonably expect that a significant amount of acceleration might arise from this part of the continuum and consequently contaminate any ghost feature.

Arav suggests that this last criterion can be tested without actually viewing this part of the spectrum. He bases this idea upon the publications by MacAlpine (1981) and Mathews & Ferland (1987, [MF81]) who successfully correlated the equivalent width (EW) of HeII (1640Å) to the flux at the HeII ionisation energy at 228Å.

In Arav's most recent paper on this matter (Arav 1996), he presents 3 BALQSOs that show weak, but significant, signatures in their CIV broad absorption troughs (Arav 1996, figures: 2; 4; 7; 8; 9 &10). He also presents one additional object in which a ghost appears to be present in the OVI BAL (Arav 1996, figure 9). To date, these are still the only individual objects that have been reported to exhibit the ghost signature. Given the potential significance of the features in shedding light on the outflows from BALQSOs, it is clearly desirable to increase the size of this sample.

1.5 Structure of this Thesis

In Chapter 2, I describe the data at the core of this research, obtained from the SDSS.

In Chapter 3, I present my search through the SDSSs Early Data Release (EDR) BALQSO catalogue for any ghost signatures present. I discuss the methodology employed in § 3.1 and present the results in § 3.2 and § 3.3.

In chapter 4, I return to the issue of BALQSO classification, introduce my ongoing work, and detail the development of an improved tool for the quick and accurate identification of these objects. This is motivated by the recognition that the EDR BALQSO catalogue was formed from just a small fraction of the total data set currently available from the SDSS. The aim is thus to allow a continuation of the search for Arav's ghost signature amongst the ever-expanding quasar data sets available. I include within this section the main issues that require consideration when analysing these objects (§ 4.1 & § 4.2) and subsequently detail the development of this tool (§ 4.3). I conclude this chapter (§ 4.4) by presenting the results of test runs on the EDR sample and comparison to the results of Reichard et al. (2003a).
Chapter 5 presents my general thoughts on both aspects of this research (i.e. ghosts in the EDR and the development of the BALQSO identification tool).

Chapter 6 presents my conclusions and suggests promising follow-up strategies.

2.0 The Data

The current statistics on individual QSOs exhibiting ghost signatures are poor (i.e. Arav 1996, QSO 1314+0116, QSO 1336+1335, QSO 1120+0019, QSO 2212-1759). In addition the few spectra that show this feature are of low quality. Thus the Sloan Digital Sky Survey (SDSS), an optical imaging and spectroscopic survey, which began observations in April 2001, offers an excellent opportunity to improve upon these limitations.

The SDSS (York et al 2000) was commissioned to provide the astronomical community with an unprecedented number of images and optical spectra. One of their key targets is to deliver spectra for some 100,000 QSOs. In June 2001 they made available their Early Data Release (EDR) which was subsequently followed up by their data release 1 (DR1) in 2003. In 2004 they have delivered both their data releases 2 and 3 (DR2, DR3) with the DR3 catalogue containing ~51000 quasar spectra.

As described in brief by Menou et al (2001), all the SDSS spectra are observed with a dedicated 2.5m telescope at Apache Point Observatory, New Mexico. An array of CCD's (Gunn et al, 1998) are employed with all observations carried out in 5 wavebands (u', g', r', i' and z') which cover the full wavelength range of the CCD's longward of the atmospheric cut-off at short wavelengths (Fukugita et al 1996). A 20" telescope at the same site provides the photometric calibrations. This spectroscopic survey uses two fibre-fed double spectrographs designed to cover the wavelength range of 3800 Å – 9200 Å and achieve a resolution of ~1800 across the entire range. Quasar candidates are flagged from their point-source optical morphology and optical colours such that they lie outside of the stellar locus in colour-colour space.

The SDSSs algorithms only flag spectra as likely quasars, or QSOs, based upon their observed spectral characteristics. Thus any attempt to focus upon any subset of QSO, requires additional independent classifications to be carried out.

Each SDSS quasar spectrum has been wavelength calibrated, sky subtracted and corrected for galactic extinction. A redshift determination, based on fitting 'à trous' wavelets (Starck, Siebenmorgen & Gredel 1997) to each continuum-subtracted spectrum, is also supplied. This 'wavelet transform' algorithm identifies peaks in the

spectrum that are then marked as emission lines. The algorithm then compares the approximate locations of these emission lines to a template quasar spectrum (specifically: Vanden Berk et al 2001) and consequently assigns a 'trial' line identification and redshift. Each emission line is also assigned a non-zero weighting. The algorithm then fits a Gaussian to each emission line peak, returning a central wavelength value for each. It then, for each line independently, searches the spectrum for other 'expected' common emission lines at their appropriate wavelengths (as defined by the template spectrum) for that trial redshift. It thus records the related weighting for each expected line it finds, sums these 'returned' weightings and then divides through by the total weighting expected. In this manor, the algorithm returns a weighted score, or 'confidence level', for each designated 'significant' emission lines trial redshift and finally selects that redshift associated with the highest confidence level. They note that the confidence level for each line is penalised if the emission line centres do not quite match and, in addition, they have "tweaked" this process to best match their empirical success rate in assigning correct emission line redshifts, based on the manual inspection of a large number of spectra from the EDR.

2.1 The SDSS Early Data Release, BALQSO Catalogue

In their catalogue of BALQSOs from the SDSS EDR, Reichard et al employed an automated method for the selection of BALQSOs. From their sample of EDR QSOs they identified 224 BALQSOs of which 198 fall into the redshift window required for the study described in chapter 3. They developed an algorithm based on the assumption that the continuum slope of all QSO spectra are basically the same, but subject to a difference in their power law index, α, and host galaxy extinction, E(B-V). This algorithm ascertains the values of these two physical parameters (i.e. α[4] and E(B-V)), via χ^2 minimisation, with respect to a composite spectrum produced from the full 3814 EDR spectra. Reichard et al also fitted a reddened power law to the individual spectra that they analysed. From these two continuum-fitting processes, they produced a BI for the CIV BAL region. However, for the composite fitting routine, they altered Weymann's BI definition to account for the potential breadth of the CIV BEL and any absorption therein. They achieved this by scaling the BEL to the peak of the flux of the

[4] Within this thesis, the use of the power law index, α is always taken to be the wavelength index, α_λ, and is defined via $F_\lambda \propto \lambda^\alpha$.

input spectrum and calculating the BI relative to the continuum plus BEL, rather than just the continuum.

Their final step was to inspect by eye any spectrum that was not well fitted by the continuum, determined by χ^2, and rejected the source as a BALQSO if necessary.

Reichard et al's EDR BALQSO catalogue provided immediate access to a large, bone fide, set of BALQSOs amongst which I can search for the ghost of Lyα.

3.0 The Ghost of Lyman Alpha in the SDSS EDR

With the recent data releases from the Sloan Digital Sky Survey (SDSS), there is current access to an unprecedented number of spectra of QSOs, including BALQSOs. As a first step, here I take as my parent sample the BALQSO catalogue presented by Reichard et al (2003), which is selected from the SDSS Early Data Release (EDR).

For the purpose of this analysis, I have chosen to select only objects whose spectra fully cover the CIV BEL and its associated BAL. The CIV BAL displays a particularly deep, well-defined trough and is thus the most likely BAL to exhibit a clear ghost feature. Given the optical wavelength coverage of the SDSS 1-D spectra, a suitable red-shift window of $1.66<z<4.94$ was identified.[5] Using, Reichard et al.'s (2003) sample this redshift restriction yields 198 BALQSOs.

3.1 Methodology

The approach taken within this study is to divide my BALQSO sample into sensible sub-categories, at each step eliminating those objects that are unlikely to, or do not, show convincing ghost signatures. For each sub-sample, I produce a composite spectrum to highlight structure that is common to objects across the sample (§ 3.2).

Figure 5 illustrates my process of sub-division of the full EDR BALQSO sample into manageable and more revealing data sets. Starting from the complete set of BALQSOs contained within our red-shift window, I first create two subsets, namely, the high-ionisation BALQSOs (HiBALs) and low-ionisation BALQSOs (LoBALs). HiBALs are defined by the presence of broad absorption troughs associated with one or more high-ionisation lines. LoBALs define BALQSOs that display broad absorption troughs associated with one or more low-ionisation lines also. I adopt Reichard et al.'s "by eye" classifications for this purpose. The primary reason for this step is that it is more common to see highly structured BALs in LoBALs. One might therefore expect it to be more difficult to find clear ghost signatures amongst LoBALs.

[5] In this chapter I adopt the redshifts given by Reichard et al. (2003) unless otherwise noted. See chapter 4 for my alternative redshift determinations.

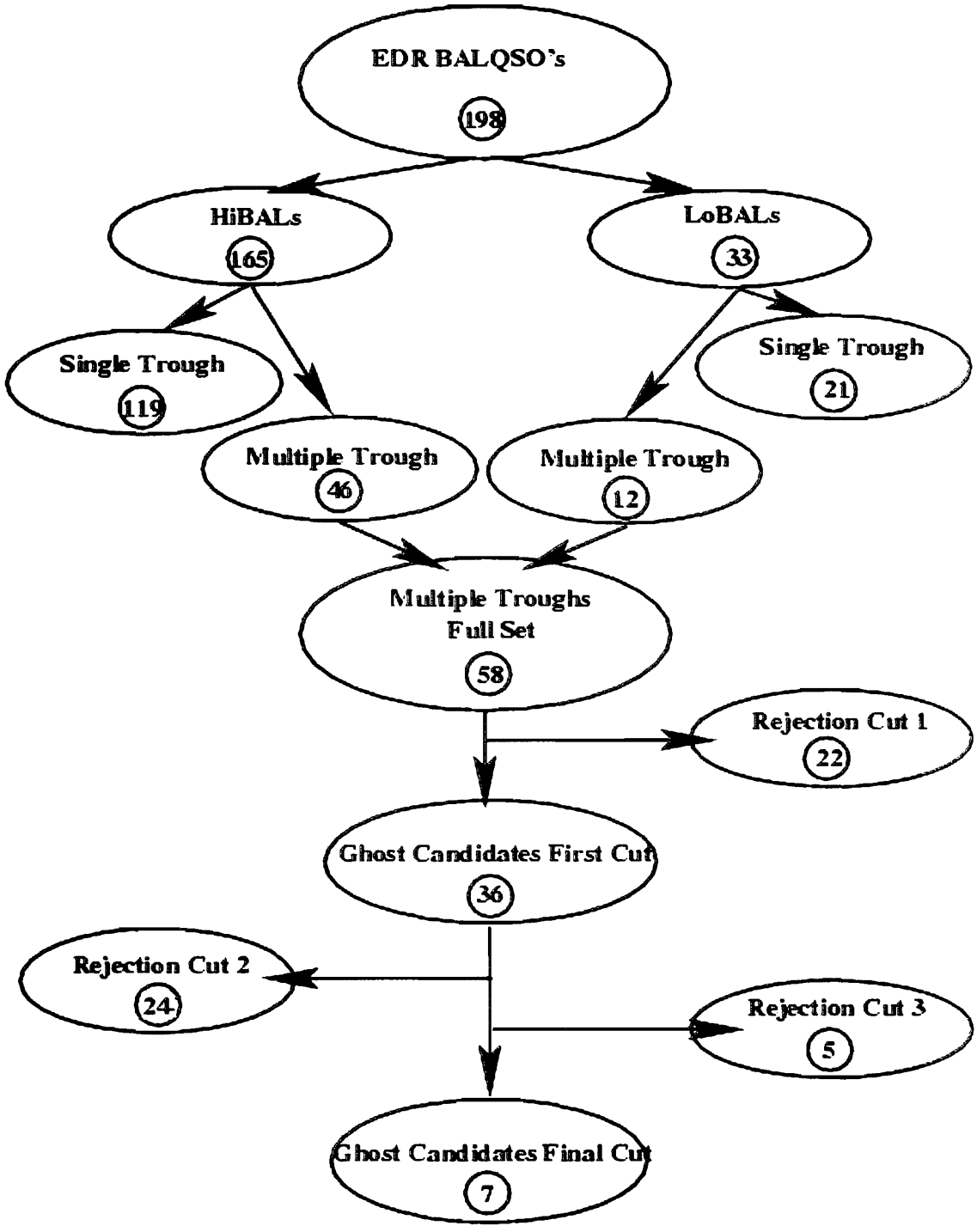

Figure 5 - This displays a simple flow diagram illustrating the method of sub-division employed in this chapter.

The next sub-division splits these samples into groups exhibiting single trough (ST) and multiple trough (MT) BALs. My working definition of a MT BAL is simply that it should exhibit more than one clear minimum in its absorption trough. This classification process was done by eye. The justification for this division is that all ghost candidates must, by definition, exhibit MTs. As expected, the proportion of MTs is somewhat higher amongst the LoBALs (36%) than the HiBALs (28%).

Even though one may expect it to be more difficult to find convincing ghost candidates amongst the LoBAL MT set, I feel it is nevertheless important to inspect all BALQSOs exhibiting MTs for this feature. I therefore re-merge the MT HiBAL and LoBAL sets before making the final rejection cuts.

In order to help distinguish likely ghost features from other BAL structure, I introduce the concept of a "ghost zone" (GZ). This I define as the region in velocity space within which one would expect to see the peak of any ghost signature. More specifically, I define the limits of the GZ by the most extreme combinations of the doublet structures that give rise to the signature itself. Thus I first calculate the rest-frame velocity differences ($V_{Gh,blue}$ and $V_{Gh,red}$) between Lyα and each of the two NV doublet transitions. I then locate where in the spectra these velocities occur for each of the doublet pairs in the BEL/BAL of interest (CIV [always] and SiIV [where available]). The maximum and minimum values of these locations fundamentally define the edges of the GZs for SiIV and CIV.

Finally, I slightly expand the GZs to allow for redshift errors. These errors are larger for BALQSOs than "ordinary" QSOs, with typical values (statistical + systematic) around $\Delta z = 0.01$ (Donald Schneider, private communication; see also Schneider et al. 2002). This error approximation is also born out by the results I present in § 4.4.3. However, for the purpose of providing a GZ, which is reflective of these errors, I use the median redshift to generate a percentage error in redshift. This is a more reflective, and at the lower redshifts more conservative, representation of errors arising in individual spectra (i.e. using a percentage error will ensure that a systematic error causing a shift of 1000km/s, in velocity space, will remain a shift of 1000km/s, irrespective of redshift).

The median redshift of the full BALQSO sample is $z_{med} = 2.11$, so we expand the GZs by multiplying the limiting wavelengths by a factor:

$$[1 \pm \Delta z/(1+z_{MED})] = 1 \pm 0.0032.$$

Here, the positive and negative signs refer to the red [blue] edges, respectively.

I then refine the MT sample by carrying out three 'rejection cut' iterations. In the first cut, I remove what I consider obvious non-ghosts and weak candidates. Thus in this iteration I reject objects exhibiting bumps well away from v_{Gh}, sources with particularly low S/N spectra, and sources with highly structured BALs. At the end of this iteration, I am left with the 'Rejection cut 1' (RC1) and 'Ghost candidate First Cut' samples.

The final rejection cuts are more subjective and are specifically designed to leave a final sample that contains only sources with clear, strong, local maxima in their BALs that appear well within their GZs ('Ghost Candidates Final Cut'). These final rejection cuts, although in reality are achieved in a single step, generate the two sub-sets of 'Rejection Cut 2' and 'Rejection Cut 3' (RC2 & RC3 respectively). Essentially the criterion applied here is the same but, with review of the spectra rejected at this stage, a natural division is apparent. I thus create the two sub-sets of RC2 and RC3, and consequently display all the members of RC3 in figure 7, as these are clearly worthy of consideration as ghost candidates in their own right.

Since the final 'rejection steps' are the most subjective aspects of this selection process, I show, in Figure 6, selected spectra from the RC1 & RC2 samples (as well as the complete RC3 sample Figure 7, as previously mentioned) and briefly discuss the reasoning for their rejection in the accompanying text below the figures.

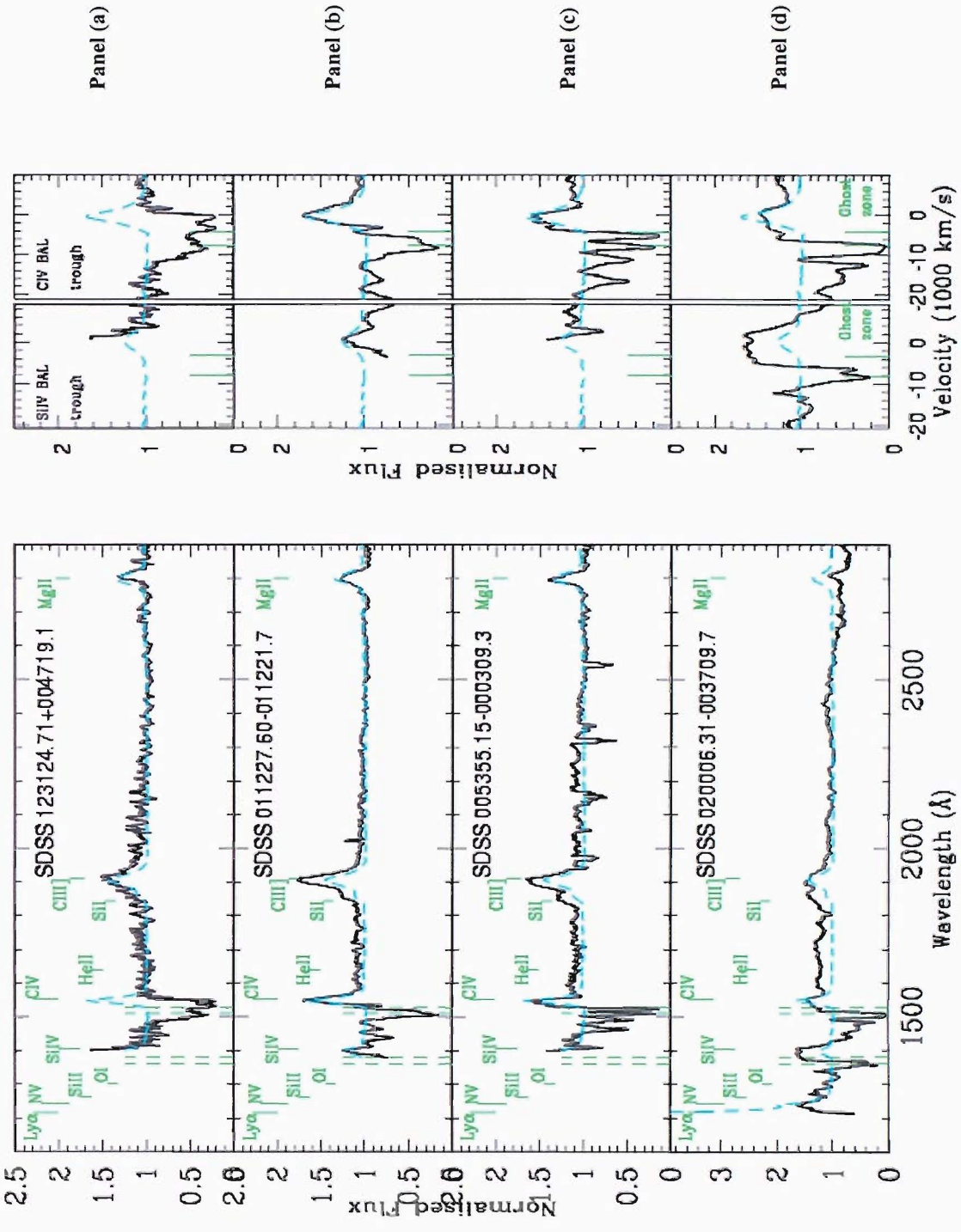

Figure 6 - Examples of MTBALQSOs rejected at cuts 1 & 2. The black lines represents the normalised spectra. The blue dashed lines displays My EDR QSO sample composite. The panels on the right (top of this page) display the CIV and SiIV BALs in more detail. See text for details.

With reference to figure 6, I summarise the spectra plotted and and outline the reasons for their rejection:

- Panels (a) & (b) are both from RC1.
    - SDSS 123124.71+004719.1 displays a feature in the CIV GZ but the S/N ratio is sufficiently low for this not to be considered. In addition, the presence of a second peak casts further doubt.
    - SDSS 011227.60-011221.7 is clearly a MT BALQSO again there is more than one local maximum, and the strongest feature in not located in or near to the GZ of CIV.
- Panels (c) & (d) are both from RC2.
    - SDSS 005355.15-000309.3 displays a significant feature centred in the GZ of CIV. However, the other two maxima immediately blueward of this potential Ghost give a strong appearance of multiple narrow absorbers thus this is rejected.
    - SDSS 020006.31-003709.7 has a clear feature but despite its appearance it is clearly outside the CIV GZ. However, the flattened appearance of the BELs makes the redshift determination uncertain. Hence this is an RC2 spectrum.

With reference to figure 7, I summarise the spectra plotted and detail the criterion applied for their respective rejections:

- SDSS 143022.47-002045.2, SDSS 145045-004400.3 & SDSS 171330.98+610707.8 all display a deep, broad, trough with significant features right on the blue edge of their GZs. This factor ultimately led to their rejection but clearly all are good candidates in their own right.
- SDSS 113544.33+001118.6 has a well-positioned feature but is sufficiently noisy to be rejected.
- SDSS 110736.67+000329.4 similarly has a well-positioned feature but this feature is no different in size to that of the 2 other features located at and to the red of the SiIV BEL. However, as is evident with all the spectra displayed here, the feature seen in the CIV BAL is sufficient clear enough to be distinguished from the other rejected spectra.

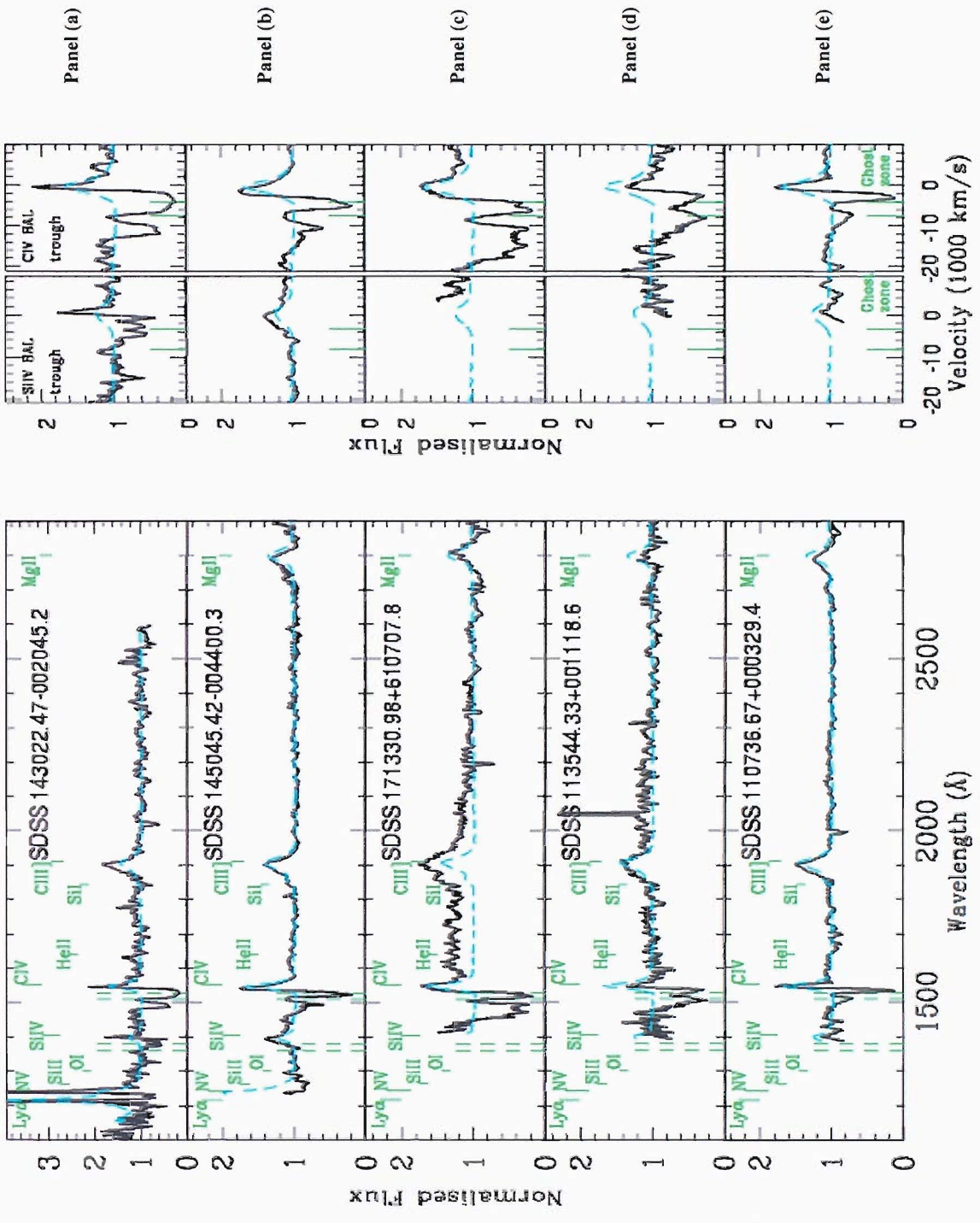

Figure 7 - Individual plots of the Rejection Cut 3 sample. The black lines represent the normalised spectra. The blue dashed lines display my EDR QSO sample composite. The panels on the right display the CIV and SiIV BALs in more detail. The vertical dashed green lines mark out the GZs in both the CIV and SiIV BALs.

3.2 The Composite Spectra

In this section, I present and discuss the composite spectra produced for the various sub-samples. Each composite is constructed as the arithmetic mean of the normalized BALQSO spectra within a given sample. The normalization is achieved by fitting a composite spectrum with a reddened power law to selected continuum windows for each source in the sample and dividing the spectrum by the fit (see § 4.3.1). All of the composites are displayed across the spectral range between λ1150Å - λ2900Å with particular focus on SiIV and CIV, as these lines typically display the strongest BAL signatures.

I begin by displaying the full set of 198 BALQSOs (Fig.8, panel [a]) within my redshift window to provide a sanity check of the composite construction routine. My EDRQSO composite is a good match to this composite spectrum except within the BAL regions themselves. This is expected as the EDRQSO sample is dominated by BELQSOs with only a small fraction (~20%) of the sample comprising of BALQSOs. This validates my composite construction methodology. It is also interesting to note that the RMS spectrum shows good agreement throughout the spectra (<20% deviation) with only the blue wings of the BELs of the high-ionisation transitions (Lyα, NV, SiIV and CIV) showing significant deviations (~60-80%). Again this is no surprise when one considers that we are looking at BALQSOs classified according to the appearance / properties of their absorption troughs which begin blueward of BEL centre.

Figure 8 panel (b) displays the EDR multi-trough HiBALs. Again we see very good agreement with the EDRQSO composite throughout the spectrum with the exception of the BALs. However, unlike the EDR-BALQSO full sample we already see a feature starting to appear within the ghost zone of CIV. We also note here the increase in depth of the primary trough located immediately to the blue of the CIV BEL. Inspection of the RMS spectra reveals a subtle, but clear, drop in amplitude (~40%) corresponding to the feature we see appearing in the Ghost zone. We also see a clear deepening of the trough between Lyα and NV in the composite spectrum, although we see nothing significant in SiIV apart from a small deepening of the BAL.

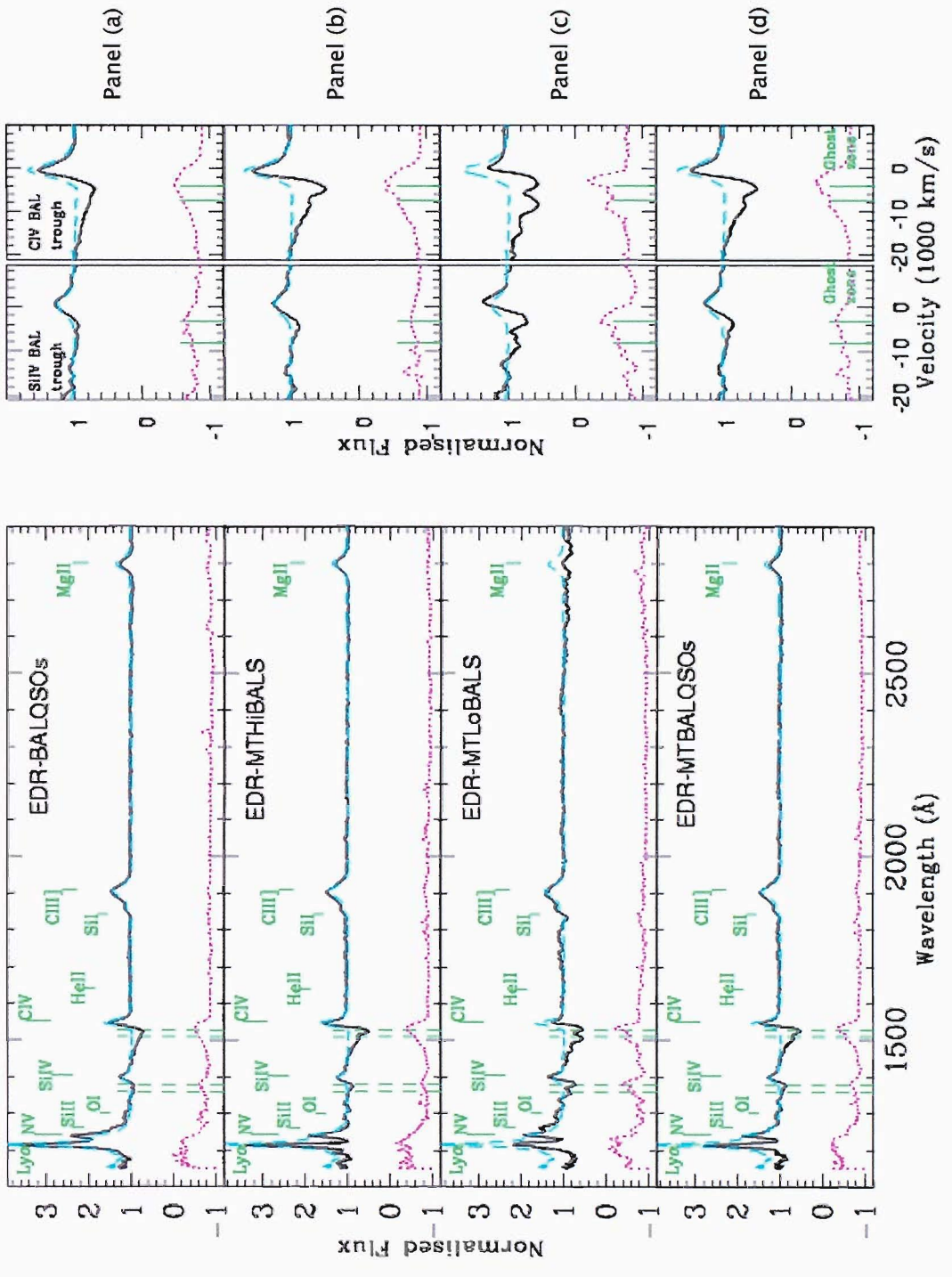

Figure 8 - Composite Spectra from the EDR. Here, the black lines display the composite spectra. The blue dashed lines display my EDR sample composite. The red dotted lines display each composites respective RMS spectrum. As before, the vertical dashed lines trace out the GZs in both the CIV & SiIV BALs.

Figure 8 panel (c) displays the LoBAL multi-trough set. This composite displays a significant deepening of both the CIV and SiIV BALs, relative to the full EDR QSO composite in my redshift window, with clear features in both ghost zones. This is not a complete surprise as a similar observation was made by Reichard et al (R2003b). In addition, the feature within the SiIV BAL displays double peaks, separated by ~2000km/s which is consistent with the separation of the SiIV doublet. This is precisely what one would hope to see if we are indeed looking at a Lyα-NV line-locked system. Again, we also see, an obvious drop in the variability of the RMS spectra corresponding to the feature discussed.

The final plot in Figure 8, panel (d), is the merged sample of the MT HiBALs and MT LoBALs, which form the complete multi-trough set. Here we see a marginally more pronounced feature in the CIV Ghost zone than that of the HiBAL set alone, but the SiIV feature observed in the LoBAL sample has all but disappeared. Since the HiBALs out-number the LoBALs by about 4:1, it is no surprise that this composite looks quite similar to that of the HiBALs alone.

The fifth composite displayed, Figure 9 panel (a), is the set generated by the rejection of spectra at RC1, discussed in §3.1. This is included as it demonstrates the success of the selection criteria employed. Although we see the slightest hint of a feature in the CIV BAL, the 22 objects that make up this sample are not the most convincing ghost candidates. It is interesting though that there is still a drop in amplitude of the RMS spectrum corresponding to the 'hinted' feature. However, it is only very slight and overall this plot looks much like the original EDR-BALQSO composite with just a slight deepening of both the CIV and NV BALs.

The next plot displayed, figure 9 panel (b), is the composite of the RC2 set. We see a clearer hint of a feature on the blue edge of the CIV GZ and there is again a corresponding drop in amplitude in the RMS spectrum and a more pronounced NV BAL.

Figure 9, panel (c), displays the composite of the Rejection Cut 3 set. This contains all the BALQSOs that have distinct ghost-like features but suffer from bad red-shift determination / Wrong location but debateable Z / low signal to noise or a combination of these factors (see §3.2). This shows the clear emergence of a feature in the CIV BAL but it is located just outside the blue edge of the GZ. However, this is actually expected if one considers the asymmetry of the BAL profile. Specifically, by constructing a composite from two spectra that display this asymmetric profile (e.g. Figure 10, SDSS 170056.85+602639.8), and ensuring that the two spectra have a relative offset of, e.g., 1000km/s from their correct systemic locations, then the result will display a ghost feature which is significantly shifted blueward of its correct 'central' location. In other words, errors in redshifts (symmetrically distributed about systemic) combining with the asymmetric profile of BALs, results in composite ghost features shifting bluewards of the correct systemic location.

The final composite displayed comprises of the complete set of BALQSOs from the SDSS EDR exhibiting the Ghost of Lyα (EDR-MTBALQSOs-Ghosts-Final-Cut, Fig.9.). Unsurprisingly, in light of the selection process that led to this sample, we see a clear Ghost signature in the GZ of CIV. Nevertheless, the strength of the ghost signature in the composite is encouraging, and it is particularly promising that the signature also appears in the SiIV BAL of the composite. Indeed, we have two strong features from 5 objects contributing to the SiIV BAL in the composite spectrum. Significantly here, the RMS spectrum almost returns to its continuum value at the corresponding position. This point is important to understand and worth stressing: The ghost of Lyα is a reduction in the absorption of the continuum flux. Hence, any strong ghost can, at best, return the flux (at and around the ghost zone) to its continuum level. Thus, if we have a set of objects with strong ghost signatures, the RMS spectra of any composite produced will, at the location of the ghost, tend to the continuum level of the RMS spectra, whilst the surrounding BAL is essentially unique to each object, and will thus produce an increase in the amplitude of deviation in the RMS spectrum.

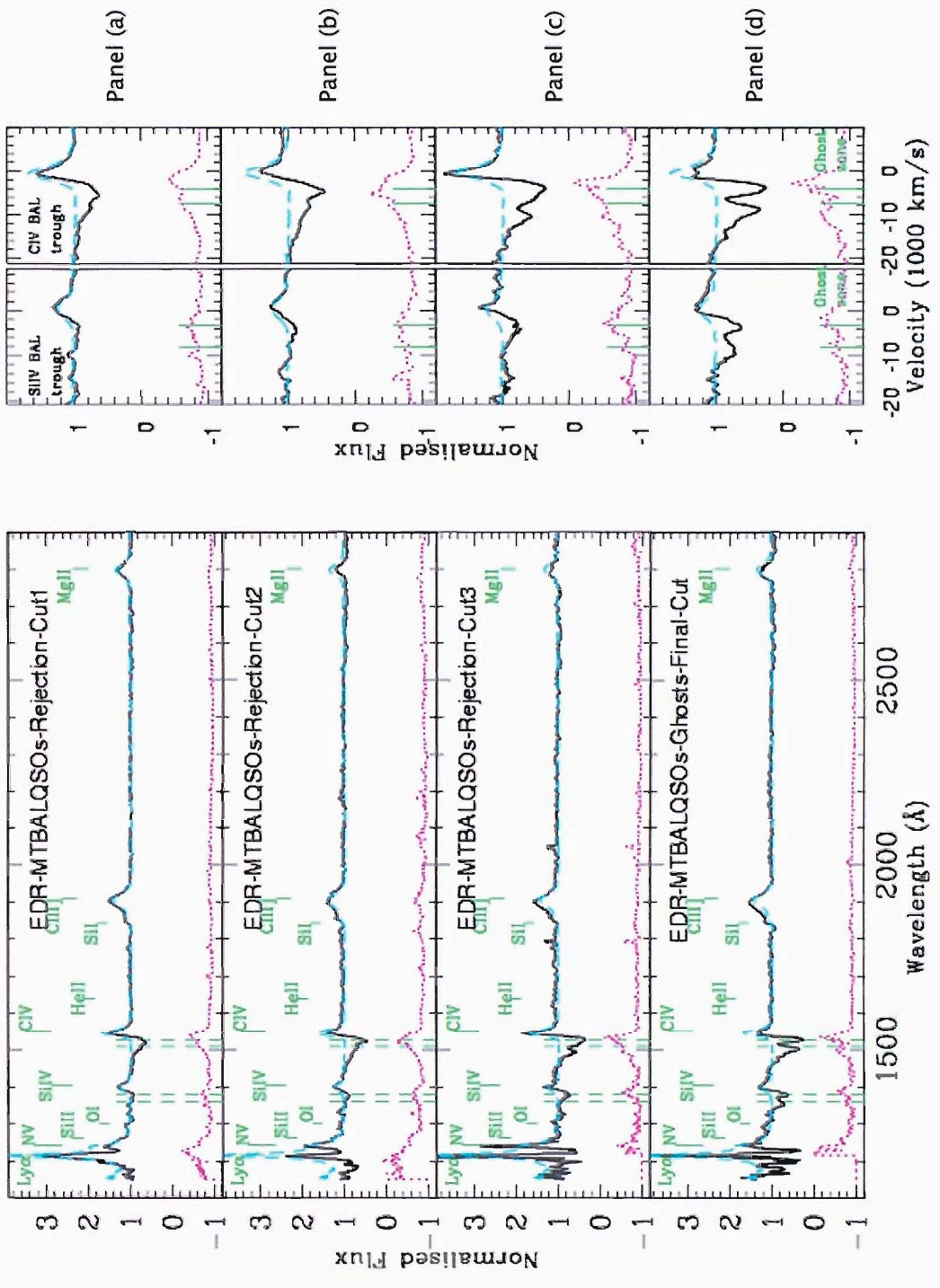

Figure 9 - Composite spectra from the EDR. Here, the black lines display the composite spectra. The blue dashed lines display My EDR sample composite. The red dotted lines display each composites respective RMS spectrum. And, as before, the vertical dashed lines trace out the GZs in Both the CIV & SiIV BALs.

3.3 The Best Ghost Candidates

In figure 10 below, I show the 7 best candidates of individual spectra displaying strong ghost signatures. Each spectrum has been normalised via the composite fitting routine (described in § 4.3.1). An arithmetic mean composite spectrum of the full QSO sample from the EDR, in my redshift window, is over-laid to aid in the recognition of various BELs, redshift determination and to highlight the BALs present.

SDSS 132304.58-003856.5 (figure 10 [a]): I first note the lack of Lyα coverage. Despite this, this object clearly can be regarded as a strong candidate as the depth of the BAL either side of the ghost feature is particularly pronounced. It is noted that the ghost feature itself sits toward the blue edge of the GZ and inspection of the CIII] and MgII BELs suggests that the given redshift could be improved. In fact, improvement of the redshift by centering these two BELs does actually centre this Ghost feature in the GZ (see §4.4.3). It also worth of noting that the HeII line is not obvious here and this fits well with Arav's selection criterion. The SiIV BEL shows the beginnings of a shallow absorption trough, but the spectrum cuts off too early to search for a ghost in this line.

SDSS 142050.34-002553.1 (figure 10 [b]): This again shows a strong CIV BAL and a clear feature centered in the ghost zone. However, I note that the MgII and CIII] BELs are again misaligned sitting blueward of their rest wavelengths. In this instance centering on these lines shifts the peak of the ghost feature to the red edge of the GZ. However, observation of the SiIV BAL shows a deep trough either side of the Ghost feature. Shifting the spectrum to centre the MgII line actually serves to center this feature in the middle of the GZ. Furthermore, the SiIV feature also shows a double peak that one may reasonably expect to see due to the relatively broad separation of the SiIV (1394Å & 1403Å) doublet. Again, within this spectrum, there appears to be little or no trace of a HeII BEL, in line with Arav's prediction. I again note the lack of coverage of Lyα, but emphasise the similarities in profile between the ghost features in both BALs (SiIV & CIV) and their respective BELs. This correlation is suggested by Arav[6].

[6] This idea stems from an observational fact that the BELs of Lyα and CIV tend to be very similar. Thus in the absence of a Lyα BEL, Arav suggests the CIV BEL can be used as a guide.

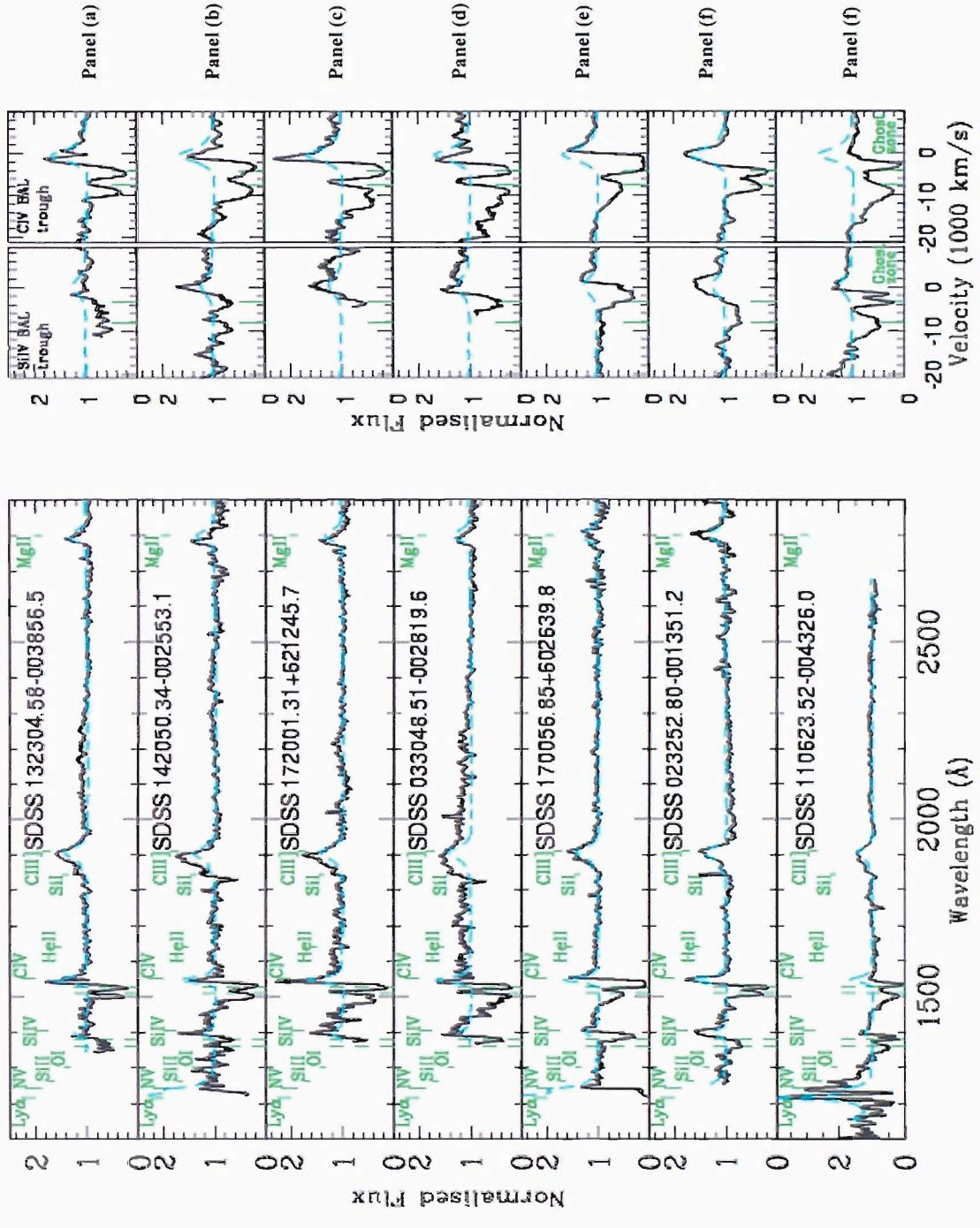

Figure 10 - The 'Ghost Candidates'. The black lines represents the normalised spectra. The blue dashed lines displays My EDR QSO sample composite. The panels on the right display the CIV and SiIV BALs in more detail. The vertical dashed green lines mark out the GZs in both the CIV and SiIV BALs.

SDSS 172001.31+621245.7 (Figure 10 [c]): Again there is a lack of coverage of Lyα. This spectrum actually cuts off just as the SiIV BAL begins. Also, I note that as above, the CIV ghost feature sits just on the blue edge of the GZ but, also as above [panel (a)] centering of the MgII line centers the ghost feature also. This feature does actually mimic the CIV BEL quite closely and again there is no obvious HeII emission. This feature also sits clearly in what is otherwise a very broad, deep, trough.

SDSS 033048.51-002819.6 (Figure 10 [d]): This spectrum, similarly, suffers from a slightly dubious redshift and, again, the MgII and CIII] BELs are slightly blue of their respective rest frequencies. Again though, centring these BELs also centres the ghost feature. In addition, this feature also appears to share a similar profile to that of the CIV BEL, it also sits in a particularly deep and broad trough, and once again there is little sign of HeII emission. The spectrum cut off just short of the SiIV BEL.

SDSS 170056.85+602639.8 (Figure 10 [e]): This source displays a spectrum that cuts off just into the NV BEL, and the SiIV BAL terminates longward of the blue edge of the GZ. However, despite the noise, or absorption, present in the MgII line this spectrum appears to be correctly redshifted, as determined by the CIII] BEL. It also displays a well centred ghost feature and although the blue side of the BAL tails back to the continuum, and is not as deep as the previous examples, it is particularly smooth in profile and very suggestive of a BAL that, in the absence of a ghost feature, would continue down to the very deep trough immediately to the blue of CIV rest. As before, this feature is highly reminiscent of the CIV BEL and again there is no obvious HeII emission.

SDSS 023252.80-001351.2 (Figure 10 [f]): This spectrum appears to be correctly redshifted, any minor adjustments applied here do not significantly alter the position of the ghost feature. Again there is a lack of Lyα coverage but, as before, there is an obvious parallel in the profiles of the ghost and CIV BEL. In fact, a closer look shows a peculiar asymmetry in both the ghost feature and the CIV BEL that makes this a compelling case for a ghost signature. This spectrum has a relatively narrow BAL, and thus it is no surprise that a ghost feature cannot be identified in the SiIV BAL.

SDSS 110623.52-004326.0 (Figure 10 [g]): The seventh and final clear ghost candidate does actually have coverage of the Lyα BEL. This BEL is appropriately broad and

strong with the 'breadth' clearly repeated in the ghost features seen in both the CIV and SiIV BAL. In addition, there is a clear and respectively deep absorption trough associated with the NV BEL. However, I note that the double troughed feature seen between the SiIV BEL and its respective ghost feature is reminiscent of a Narrow Absorption feature. Significantly, the wavelength separation of this double trough does not correlate with the wavelength separation of the SiIV doublet whereas the peaks of the Ghost feature do. Equally significant here is that no NAL feature is observed in the CIII] (low ionisation) BEL either, this would be expected for a NAL. The redshift appears well determined as inferred by the location of the CIII] BEL, and any slight shifting would still maintain the presence of these features in their respective GZs.

3.4 Chapter Summary

I have set out an empirically based methodology to investigate and identify the presence of Arav's 'Ghost of Lyman Alpha' radiation driving signature in BALQSO spectra.

Through the creation of sub-samples of BALQSOs in the SDSS EDR I have provided compelling evidence of the presence of this signature in several objects. In particular, via a combination of composite construction and inspection of individual spectra, I have demonstrated that the ghost of Lyα appears to be a real phenomenon. The composites not only serve as a sanity check on the selection process, but also provide additional insight into the consistent presence of a feature at, and around, v_{GH}. Specifically, they show that all ***but*** the 'humps' within, and close to, the GZ of CIV are washed out in multi-object composites. This is most clearly demonstrated in the LoBAL and Rejection Cut composites, as both contain relatively large numbers of spectra, both contain MT features and yet both only display residual 'humps' within (and on the blue edge) of the GZ. Furthermore, inspection of the individual spectra in both Figure 7 and 10 show clear signatures around the GZ of CIV, as well as on occasion displaying strong signatures in the SiIV GZs also. If we take both sets to be Ghost of Lyα candidates then this constitutes a quadrupling of the known BALQSOs displaying this signature. However, even by remaining cautious and only accepting the 'Ghost candidates, final cut' sample, this still provides the largest known sample of BALQSOs displaying Arav's Ghost of Lyα radiation-driving signature.

The success of this initial study suggests that much could be gained from the development of an algorithm that can identify BALQSOs and improve the determination of their redshifts, and thus enable a continuation of this research. This is particularly pertinent when one considers the relatively low numbers of QSOs in the EDR compared to that of the SDSSs DR3 (i.e. ~3,800 to ~50,000). I thus reserve further comment on the importance of these results until I present my conclusions and recommendations at the end of this thesis (Chapters 5 & 6). I consequently return my attention to the issue of BALQSOs generally and take the reader through the development of a code for the identification of BALQSOs.

4.0 Automated Classification and Parameterisation of QSOs

The work described in §3 suggests two things: The need for larger samples and the importance of establishing accurate redshifts for BALQSOs, to correctly identify ghost signatures.

Establishing an accurate redshift, as has been repeatedly stressed, is critical to revealing the presence of Lyα ghost features. The SDSS provide a redshift estimate in the header files of the individual spectra and also quote a statistical error for this value. The typical error values quoted are in the order of $\Delta z \approx \pm$ 0.001 - 0.002. However, upon close scrutiny of the spectra it is evident (particularly in §3.3) that these errors are, more often than not, ~ 5 - 10 times, too small. In fact, Schneider et al (2001), a key contributor to the SDSS program, quoted typical redshift errors in their BALQSOs to be on the order of: $\Delta z \approx \pm$ 0.01 - 0.03, which is a far more realistic error estimation. He clarified this in a private communication (12/10/2004) where he explained that the errors in the header files are calculated purely by a 'fitting' algorithm and take no account of such issues as emission line shifts (see: § 4.1) amongst other contaminating factors (e.g. sky lines etc.). As a result, I have begun a process, within the development of the algorithm herein, to independently ascertain the redshifts for each spectrum.

Consequently, this chapter not only deals with the development of a composite fitting routine for the purpose of identifying BALQSOs but it also attempts to determine a more reliable redshift than those quoted by the SDSS.
In attempting to achieve these goals, I begin by addressing two issues that must be considered before any routine is developed.

4.1 Emission Line Shifts & Redshift Determination

The identification of accurate redshifts for my sample is wholly dependent upon the presence of strong emission lines in spectra. Essentially, by dividing the observed wavelength of the emission line centres by the laboratory (rest-frame) values, a redshift is determined. However, as has been shown in several papers (e.g. Tytler & Fan, 1992; Richards et al, 2002), it is common to observe relative shifts between different emission line species present in a single spectrum. How then do we determine the 'correct'

systemic redshift? To attempt to answer this question, closer scrutiny of the current literature on QSO emission line shifts is required.

I begin with one of the earliest papers that dealt with this issue, which was written by C.M. Gaskell (1982). Although, at that time, emission line shifts in the whole array of lines observed in quasars (AGN), were well documented, Gaskell was the first to address this issue explicitly. He noted that, in the most extreme cases, particular lines (e.g. Balmer) were seen to shift by up to 2600 km/s relative to others (e.g. Semi-forbidden). Clearly such shifts, in the context of this research, will cause major misclassifications of ghost signatures, as well as BIs.

Gaskells study dealt systematically with all of the major UV resonance lines, both high & low ionisation, and for each he put forward all the possible contaminants that could give rise to the various shifts observed. He then proceeded, where possible, to rule out these various factors to ultimately arrive at the best redshift estimator for these objects. Without discussing his analysis in full, his results can be summarised as follows: For MgII (λ2798Å) he finds that this doublet does suffer from blending with FeII emission. However, later research has revealed that MgII sits in an FeII minimum, this can affect the FWHM but does not disturb the centroid of this line. For the CIII] (λ1909Å) BEL he notes that there are 3 possible contaminants. Specifically, there is probable SiIII], but also AlII and FeII emission that will likely have the effect of yielding a slightly shorter apparent wavelength, (i.e. a higher redshift estimate). For the OI triplet (λ1304Å), he notes that this is generally uncontaminated, but at the optical depths expected in the BLR, the mean wavelength will be lengthened slightly (from λ1303.5Å to λ1304.36Å). On the scales important for my study here, this would have negligible impact, but this line is generally far weaker than the other UV lines. The possible exception here is OI contamination by the SiII triplet (λ1304Å, λ1307Å & λ1309Å) in heavily reddened (dusty host galaxy) QSOs. The Lyα and NV lines are well known to be heavily blended. This fact, together with the, often visible, blue wing absorption of both of these BELs, in general, renders these two lines as unreliable redshift indicators. The SiIV and CIV BELs suffer from blue wing absorption, as well as heavy blending from other ionic species, they are thus also unreliable as redshift indicators. Finally, OIV], OVI, HeII and [OIII] are also heavily blended and so should not be used for accurate redshift determinations.

In essence then, MgII and OI are the most reliable and accurate redshift indicators. But the relative weakness of OI implies that this line can only be used when inspecting spectra with high signal to noise ratios. CIII] may also provide a reasonable estimate for the redshift in the absence of MgII (and OI), but should be viewed with some caution. Gaskell notes that "on average" the narrow forbidden lines may also provide a good redshift determination, but he also observes that some spectra can show up to 2000 km/s disagreement between the narrow lines and the OI & MgII BELs. Thus within the scope of my research, this latter redshift estimator is not considered. In conclusion, I draw the reader's attention to several later papers (e.g. Wilkes, 1984; Espey et al, 1989), which basically confirm Gaskell's findings.

4.2 Host Galaxy Extinction

Dust extinction is a ubiquitous phenomenon that affects the observed spectra of astrophysical objects. Extinction, or dust reddening, laws essentially aim to account for the amount of flux lost at all wavelengths of interest. What is generally observed is that the spectra of objects can often appear redder than we believe them to be. It is commonly accepted that dust grains in our local galaxy, as well as in very distant galaxies, preferentially scatter wavelengths with increasing intensity at shorter wavelengths. The net result is that we will observe more flux from the red end of a spectrum than the blue end. Hence we have the term 'reddening'. Although for a spectrum with sufficient S/N this phenomenon is unlikely to alter any small-scale features, it can certainly alter the slope of a spectrum, and in some cases quite significantly. It is thus a phenomenon that one cannot ignore if one is attempting to determine the true underlying continuum slope of a distant object. Defining a continuum, albeit locally, is a prerequisite for determining the Balnicity of an object. Thus I am required to identify a method for correcting for host galaxy extinction.

It is commonly accepted that high red-shift objects have an intrinsic dust reddening process that is more uniform, across a wavelength range, than the Galactic extinction seen within the Milky Way. Pei (1992) examined three different reddening laws. In summary, he modelled extinction for the Milky Way, Large Magellanic Cloud and Small Magellanic Cloud (SMC). Subsequently, Pei et al's values, given for the SMC extinction law, have been repeatedly used (notably, Reichard et al, R2003a) to describe 'host galaxy reddening' of objects in the high redshift. Pei's SMC dust reddening law

with colour excess, E(B-V), is thus included as an additional parameter within the continuum-fitting process adopted here.

Pei's empirical extinction curve is expressed as follows:

$$\xi(\lambda) = [(E_{\lambda - V} / E_{B - V}) + R_V]/(1 + R_V)$$

Equation 2 – The Pei empirical extinction curve.

Here, $E_{\lambda - V} = A_\lambda - A_V$ is the colour excess in magnitudes; $R_V = A_V / E_{B - V}$ is the ratio of total-to-selective extinction, with the subscript V indicating the visual photometric band. For the SMC, $R_V \approx 2.93$; and $(E_{\lambda - V} / E_{B - V})$ is tabulated in Pei (1992) against λ^{-1}.

4.3 Continuum Fitting

In the fitting process presented here I assume that the continuum of all BAL/BEL QSOs are essentially the same, differing only in their extinction ratio, power law index and arbitrary constant of proportionality[7] (§ 4.3.1). This assumption formed the basis for R2003a's composite fitting method also.

I note here that Reichard et al did not allow for the constant of proportionality to be a free parameter in his fitting process. Instead, before applying a χ^2 function to return best-fit values of α and E only, "The adjusted composite spectrum is normalised to match the average flux of the input spectrum in a small wavelength range". I believe this may have inhibited his fitting process slightly, as the constant of proportionality should be included as a free parameter. Hence this immediately provides one avenue for potential improvement. Furthermore, Reichard et al note that in producing their χ^2 fits, in view of the degrees of freedom in their model, they should normally require a reduced χ^2 value of "at most slightly greater than unity" for the fits to be deemed acceptable. None of their fits met this standard. They go on to reflect that to produce a BI indicative of the QSO under analysis, one need only generate a continuum that is accurate locally over the CIV broad absorption trough and consequently they do not claim "physicality for the parameters". Although fitting accurately over the CIV BAL is clearly an important property of any BI-calculating routine, I believe that failing to provide "physicality" here, hints toward further avenues of improvement. In other

[7] The constant of proportionality referred to here is a result of the different intensities of flux received (e.g. identical objects at different distances etc.).

words, the fact that one has based a fitting-methodology on the assertion that: "the spectral index, reddening parameter and constant of proportionality are the only substantial differences between spectra" then one should also demand that the values returned should indeed be physically sensible. Thus throughout the iterations undertaken to develop Reichard et al's method further, I have been regularly guided by how 'sensible' these aforementioned values are. I consequently display plots of the distributions of these values throughout the analysed sample and discuss their implications in the results presented at the end of this chapter § 4.4.

4.3.1 The Composite Fitting Method

To generate a composite fit to a given set of continuum bands, I express the spectrum to be fitted (the 'input spectrum') in terms of the composite spectrum and allow for changes in the continuum slope by solving for a 'relative' power law index, a constant of proportionality and the intrinsic reddening.

The wavelength dependent continuum flux, $F_{cont}(\lambda; \alpha, E)$, for all quasar spectra can then be expressed as follows:

$$F_{cont}(\lambda; \alpha, E) = K \cdot \lambda^{\alpha} 10^{-aE \cdot \xi(\lambda)}$$

Equation 3 - Reddened power law

Where: $a=0.4(1 + R_V)$; $\xi(\lambda)$ is the Pei SMC extinction curve; λ is the rest wavelength; E is the host galaxy extinction E(B-V); K is a constant of proportionality; and α is the spectral index.

The QSO composite spectrum used for this fit is that created by Vanden Berk et al (2001). This was used as it is a geometric mean composite and hence is characterised by the mean of the spectral indices and intrinsic extinction coefficients of the spectra from which it is comprised (re: Francis et al [1991], but also Vanden Berk et al [2001]). This is critical for the method applied here. Arithmetic mean composites do not retain these properties and are thus inadequate for the purpose of this composite fitting method as (discussed in §4.3) I require that 'physically sensible' values for both α and E are returned by the fit.

By expressing both the input spectrum and the composite spectrum in terms of equation 2, subtracting one from the other allows us to solve for the constants in the left hand side of equation 3.

Hence I calculate the continuum by solving for the linear least squares fit of:

$$\mathrm{Log}[f_i(\lambda;\, \alpha, E)] - \mathrm{Log}[f_c(\lambda;\, \alpha, E)] = \mathrm{Log}\ b + \alpha^*\mathrm{Log}(\lambda) - aE^*\xi(\lambda)$$

Equation 4 - The composite fit

Where $\alpha^* = \alpha_i - \alpha_c$. $E^* = E_i - E_c$ and b is a constant of proportionality (strictly $\log(K_i) - \mathrm{Log}(K_c)$. Obviously in this case α_i, E_i and K_i are the spectral index, colour excess and constant for the input spectrum respectively.

Each individual spectrum is fitted to carefully selected continuum windows. The selection of these windows was very much a 'trial and error' process and is critical to the optimisation of the fitting process. This is discussed further in section (§ 4.3.4 and § 4.4.1).

To solve for the three unknowns in equation 3, I have employed the SM function LINFIT that produces a least squares fit to arbitrary numbers of parameters. As per R2003a, I allow for absorption in the blue wing of the CIV BEL and hence the emission lines in this fitting routine are re-scaled accordingly (as they can potentially form part of the BI calculation). This is done as follows:

- Having established the constants α, K & E, these values are applied to the full composite spectrum to generate an adjusted composite fit that includes un-scaled emission lines.
- This new composite is then divided-up into 'new' continuum regions by excluding BEL ***scaling*** regions[8] only.
- From these new continuum regions, the SM function INTERP2[9] is used to create a spectrum over the full, common, wavelength range of the composite spectrum. This in effect provides us with two partially fitted continuums, one is inclusive of the ***un-scaled*** BELs and the other ***excludes*** all designated BELs.

[8] Once again, as was the case for the continuum window selection, this process proved sensitive to the regions selected (see § 4.4.1 for details).

[9] INTERP2 is a simple linear interpolation function.

- The latter spectrum is then subtracted from both the input spectrum and the former, composite spectrum with ***un-scaled*** BELs. Hence this provides two spectra of BELs only, with the composite fit BELs tending precisely to zero outside of the BELs themselves.
- I then apply a further set of BEL windows to both spectra by defining a smaller 'NARROW BEL' set of zones which straddle the emission line peaks[8] (rest wavelength).
- Each pair of redefined 'narrow' emission lines are divided, composite into object, and the maximum value, for each emission line in turn, delivers the scale factor required to ensure that all emission lines of the composite can now be matched in height to the input spectrum. Hence the composite emission lines are multiplied by their scale factors and then added back to the continuum that excludes the BELs.

It is noted that although taking the narrow BEL zones allows in some sense for any relative BEL shifts, it does not actually shift the composite BELs to match those of the object. Nor does it cater for significantly broader, or narrower, lines. This point was also mentioned by R2003a and is discussed further in chapter 5 (§ 5.2). However, the composite used herein is made from some 2200 BEL(BAL) QSOs from the SDSS and is thus considered representative of the spectra in my sample.

The remainder of the code is essentially related to the input, plotting and output of this program, except for the redshift calculation, which is discussed in the following section.

4.3.2 The Redshift Adjustment

As has been highlighted repeatedly throughout this thesis, a highly accurate redshift is required for the correct identification of Arav's ghost signature. It was thus an additional aim to try to improve upon the redshifts determined by the SDSS.

The method detailed here was employed in conjunction with my continuum fitting routine and is based upon a simple approach. In essence, I allow for a change in the redshifts from the SDSS values by $\Delta z \leq \pm 0.03$. Thus, by offsetting the given value by $z = z_{SDSS} + 0.03$ to start with, the routine then, via forty iterations, steps through to $z = z_{SDSS} - 0.03$. At each step, the routine, having recalculated the continuum and BEL

fits, calculates the RMS for the red wings of the major BELs relative to the composite spectrum[10]. Given this set of RMS values for each BEL, associated with each redshift increment, a weighting[11] is applied for each emission line, and hence a single 'weighted' RMS value calculated for each of the redshift steps. The program then selects the redshift associated with the 'best fit' RMS.

4.3.3 Producing the BIs

Having established a composite fitting spectrum, Reichard et al's (2003a) 'emission-line inclusive' BI calculation is determined. This 'adapted' BI is calculated in accordance with Weymann's procedure (§ 1.3), but differs in its allowance for absorption relative to the BEL itself. Specifically, by providing for scaled BELs from a BELQSO composite, we calculate the BI relative to the flux in the blue wing of this BEL, rather than relative to an underlying continuum. Consequently, as a sanity check, I also calculated BIs with the stricter Weymann definition by utilising the continuum generated in the BEL scaling process (i.e. the continuum which excludes the BELs).

4.3.4 Optimisation

Having created a code for the identification of BALQSOs via the production of a BI, a process of optimisation is still required. Specifically, the selection of continuum regions by which the composite spectrum is fitted to the input spectrum proves particularly sensitive in providing sensible results (R2003a). Also, in the case of redshift determination, the choice of both the BEL regions and their respective 'narrow-BEL' regions is critical to the algorithm's success.

To establish the 'fitting-regions' that provide the overall 'best' results a process of fitting-region selection and output was iterated. With every iteration, slight alterations are made to the fitting-regions with the subsequent outputs then compared and used as indicators to select new regions. In all, I repeated this process approximately 40 times.

---

[10] As was the case for both the continuum fitting and the BEL scaling, this was also sensitive to the exact wavelength 'windows' used for calculating the best RMS fits. Thus I again neglect to detail the 'windows' here and consequently discuss this in detail in the results section (§4.4.2).

[11] The weightings applied are designed to be reflective of the relative importance of the major emission line for the determination of systemic redshift values (re: §4.1).

To begin with, I used the EDR composite (Vanden Berk et al [2001]) to select emission line free zones and also to select sensible regions to cover the strong BELs of Lyα, NV, SiIV, CIV, CIII] and MgII. My initial instinct was also to not include the BAL region of CIV.

The 'outputs' used to indicate the relative success, are as follows:

1. A comparison of all resulting numbers of BALQSOs to those present in the EDR BALQSO catalogue, overlapping with my sample (i.e. 198[12]).
2. A comparison of the resulting numbers of BALQSOs flagged by the more classical Weymann definition to those of EDR BALQSO catalogue.
3. A plot of the fractional RMS distribution vs the fraction of the overall sample of QSOs processed.
4. A histogram plot of the frequency distribution of the sample within suitable bins of the resultant power law index, α.
5. A histogram plot of the frequency distribution of the sample within suitable bins of the resultant extinction E(B-V).
6. ***For the Redshift calculation only:*** A histogram plot of the frequency distribution of the redshift differences of those generated by my routine vs the SDSS redshifts.

The idea here is to return sensible values for all of the fitted parameters whilst also returning the maximum number of bone fide BALQSOs from Reichard's EDR BALQSO catalogue and minimising the number of non-BALQSOs found. Furthermore, I also try to maintain a high proportion of the full input sample with good fits, as determined by the fractional RMS.

By first proceeding on a 'trial & error' basis, I hoped to hone in on a set of fitting-regions that optimised the results (i.e. returned the least number of spurious BALQSOs and gave improved estimates for the redshifts).

[12] It is worthy of note here that Reichard's algorithm only found 162 out of his final published sample of 198 within my redshift window and that they do not give the number of non-BALs found by their automated routine.

4.4 Results

In view of the large number of iterations applied in arriving at my optimised fitting-region selection, I have chosen to display, in the following three sections, only those results pertaining to my optimised fitting-regions. However I do discuss any significant factors, influencing, the optimisation process detailed here.

I have divided the results of this chapter into three sections. The first section presents the results obtained by the composite fitting algorithm when supplied with the SDSS determined redshifts (§ 4.4.1). The second section presents the results of the algorithm when allowed to determine its own redshifts for each QSO (§ 4.4.2). The third and final section in this chapter subsequently uses the 'new' redshifts to revisit the sample of ghost candidates found in the EDR BALQSO Catalogue (Chapter 3) and discusses the impact of the new redshifts (§ 4.4.3).

4.4.1 Comparisons to Reichard's BIs (SDSS Redshifts)

The final fitting regions are as follows:

- The Continuum bands used for fitting the parameters α, E(B-V) and C are: [λ1274Å-λ1350Å]; [λ1410Å-λ1500Å]; [λ1650Å-λ1800Å]; [λ2100Å-λ2750Å]; [λ2900Å-λ3200Å].
- The BEL 'broad-bands' are: [λ1180Å-λ1230Å]; [λ1230Å-λ1280Å]; [λ1350Å-λ1420Å]; [λ1525Å-λ1585Å]; [λ1810Å-λ1960Å]; [λ2745Å-λ2840Å].
- The BEL 'narrow-bands' are: [λ1210Å-λ1220Å]; [λ1235Å-λ1245Å]; [λ1393Å-λ1403Å]; [λ1540Å-λ1560Å]; [λ1900Å-λ1915Å]; [λ2772Å-λ2822Å].

The trial and error process of defining appropriate continuum bands revealed the requirement that the CIV BAL region ***must*** be included to return sensible results for the fits. This can be understood by recognising that there is a lack of emission line free regions shortward of CIV, as was also noted in R2003a. Without a window in this zone, the fitting routine will always produce a poor fit blueward of CIV. However, over-emphasising this region results in too low a continuum level across the CIV BAL trough and thus many BALQSOs are not flagged. R2003a dealt with this by attributing a relative weight to the CIV BAL region of 50%. Within my routine the weightings actually come from the relative number of pixels in each fitting region.

One other important point, which became clear through the optimisation process, is the ability of a composite fitting method to fit to regions that include emission/absorption lines. Specifically, if there exists a region of QSO spectra whose variations from one spectrum to the next is relatively small, then any composite spectrum, for the region in point, will be wholly representative of every spectrum. Thus allowing for the power law, extinction and systematic offset, we can fit to such a region just as well as to those regions that generally lack emission/absorption. The upshot of this enabled me to select a reasonably large region covering the OI emission line. As noted previously (§ 4.1) the OI line doesn't generally suffer from shifts or blending. Consequently all spectra, with this region present, are very well fitted.

However, driven by the desire to produce as large a data set as possible, many spectra stopped short of the OI line and thus the CIV BAL region was required. I suspect that this led to an increased number of iterations taken in establishing the 'optimising-regions'. I consequently suggest that any future runs on large datasets apply a redshift window to include this zone.

I summarise the basic statistics of the run of the algorithm, utilising the afore mentioned fitting-regions as follows:

- Applying the R2003a BI definition, the code found 348 BALQSOs from the 1412 QSOs within my redshift window. Of these, 189 are present in the 198 BALQSOs in the EDR BALQSO Catalogue. Furthermore, from the 159 objects also classified as BALQSOs by this routine (not present in the EDR BALQSO Catalogue) approximately half, when inspected by eye were verified as BALQSOs. I have consequently included a plot of five of these 'NEW' BALQSOs at the end of this section in figure 14. As is evident from inspection, these are clearly correctly classified.
- Applying Weymann's definition for a BI yielded 343 BALQSOs and also found the same 189 objects present in the EDR BALQSO catalogue.

Comparison of the BIs calculated by each method (Reichard's & Weymann's) proved important throughout most of the early iterations. These two methods initially provided quite different numbers of BALQSOs. However, this particular run displayed the most

significant degree of convergence of these two methods, adding weight to the claim that I have indeed found a well-optimised set of fitting regions.

In figures 11, 12 and 13, I display all the plots described in § 4.3.4. These were also subject to a lot of variation throughout every iteration, but as is clear from inspection all have returned good results.

Specifically, review of figure 11 shows that 90% of the sample have a fractional RMS value less than 0.0725 and that 80% of the sample has a fractional RMS of less than 0.014 (i.e. the majority of the sample are exceptionally well fitted). Review of figure 12 shows that the calculated power law indices for almost all of the spectra (>80%) fall in the range: $-1 \leq \alpha \leq 1$, relative to the composite spectrum. To interpret this result one must first recognise that the power law index of the composite spectrum is, $\alpha_\lambda \approx -1.56$ (Vanden Berk et al [2001]). Then noting that we expect the UV continuum of QSOs to be produced by a hot accretion disk, then we can expect the power law index to fall within the range: $-4 \leq \alpha \leq 0$ (In general, one can reasonably expect a power index, $\alpha \sim -2.3$; Frank, King and Raine [2003]). Consequently, relative to this composite spectrum, one may expect to see a fairly symmetric distribution about zero with most spectra retaining a relative power index in the range: $-2.44 \leq \alpha^* \leq 1.56$ with perhaps a slight bias immediately negative of zero (if $\alpha \sim -2.3$ is correct). This is precisely what is seen here, in fact over 93% of the results fall within this range with approximately 80% falling within the range: $-1 \leq \alpha^* \leq 1$.

In figure 13 I display the extinction for each spectrum calculated relative to the composite. In a similar fashion to above, we would again expect that the distribution is symmetric about zero, as we are calculating the host galaxy extinction relative to a composite that is supposed to be representative of our quasar sample. In the case of these extinction values, as we assumed that the Pei-curve is indeed indicative of QSOs generally, we would expect a very narrow distribution about zero also. These results are remarkably good and perhaps go some way to verifying that the assumption that the SMC reddening law is indeed a good reflection of host galaxy extinction for quasar UV spectra.

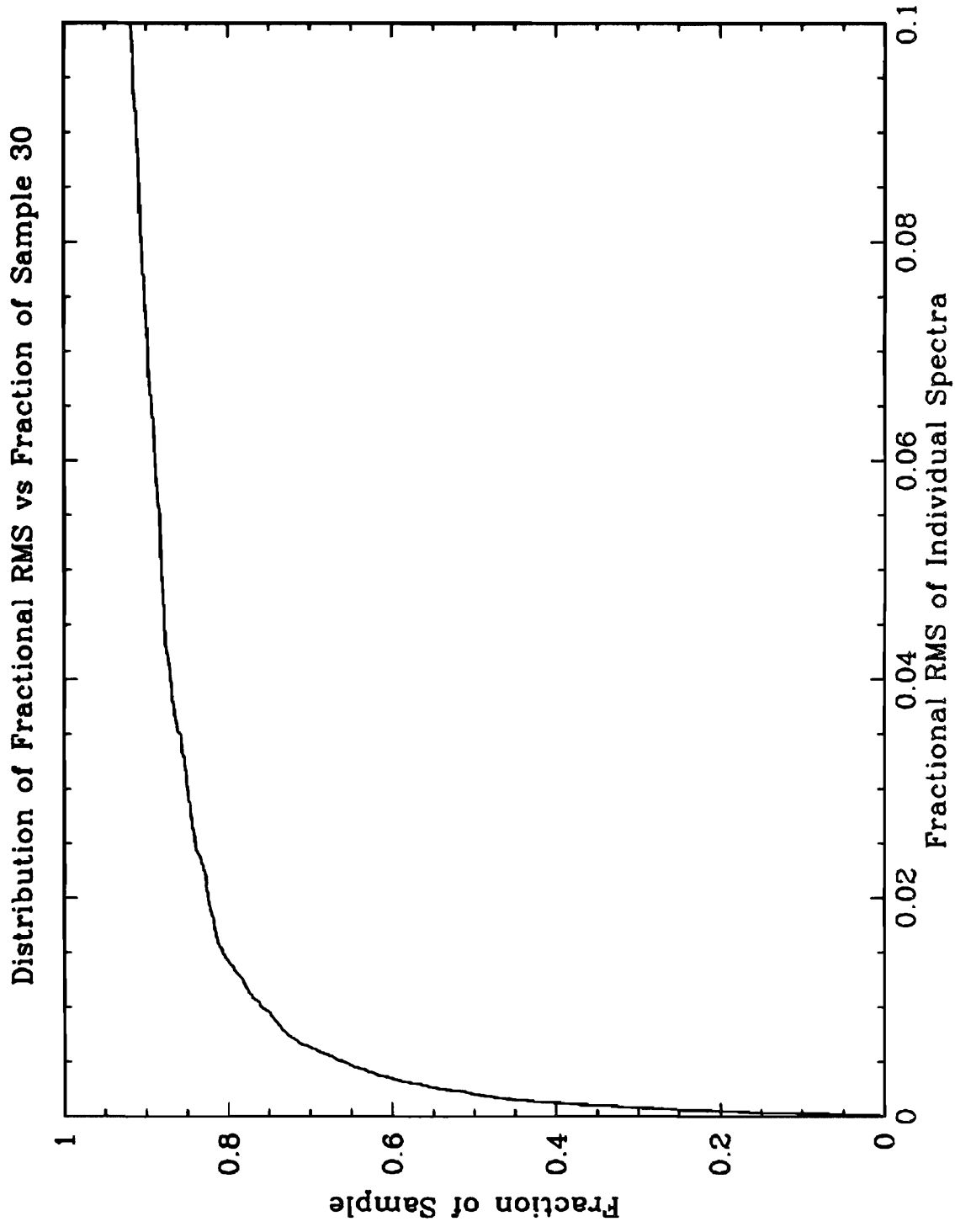

Figure 11 - Fractional RMS Distribution curve for the 'optimised' run.

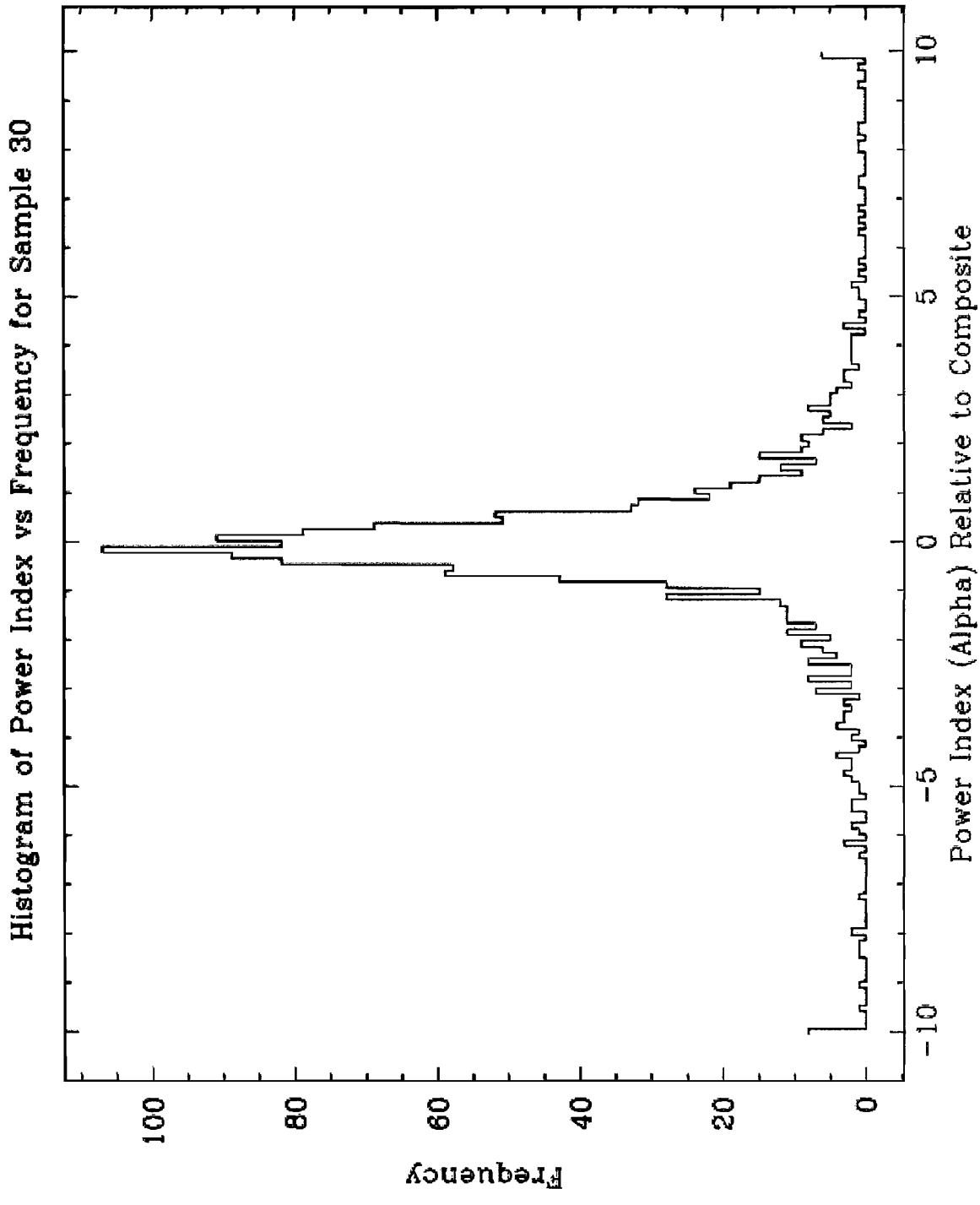

Figure 12 - Histogram plot of the frequency distribution per [$\alpha = 0.12$] Power Law Index bin.

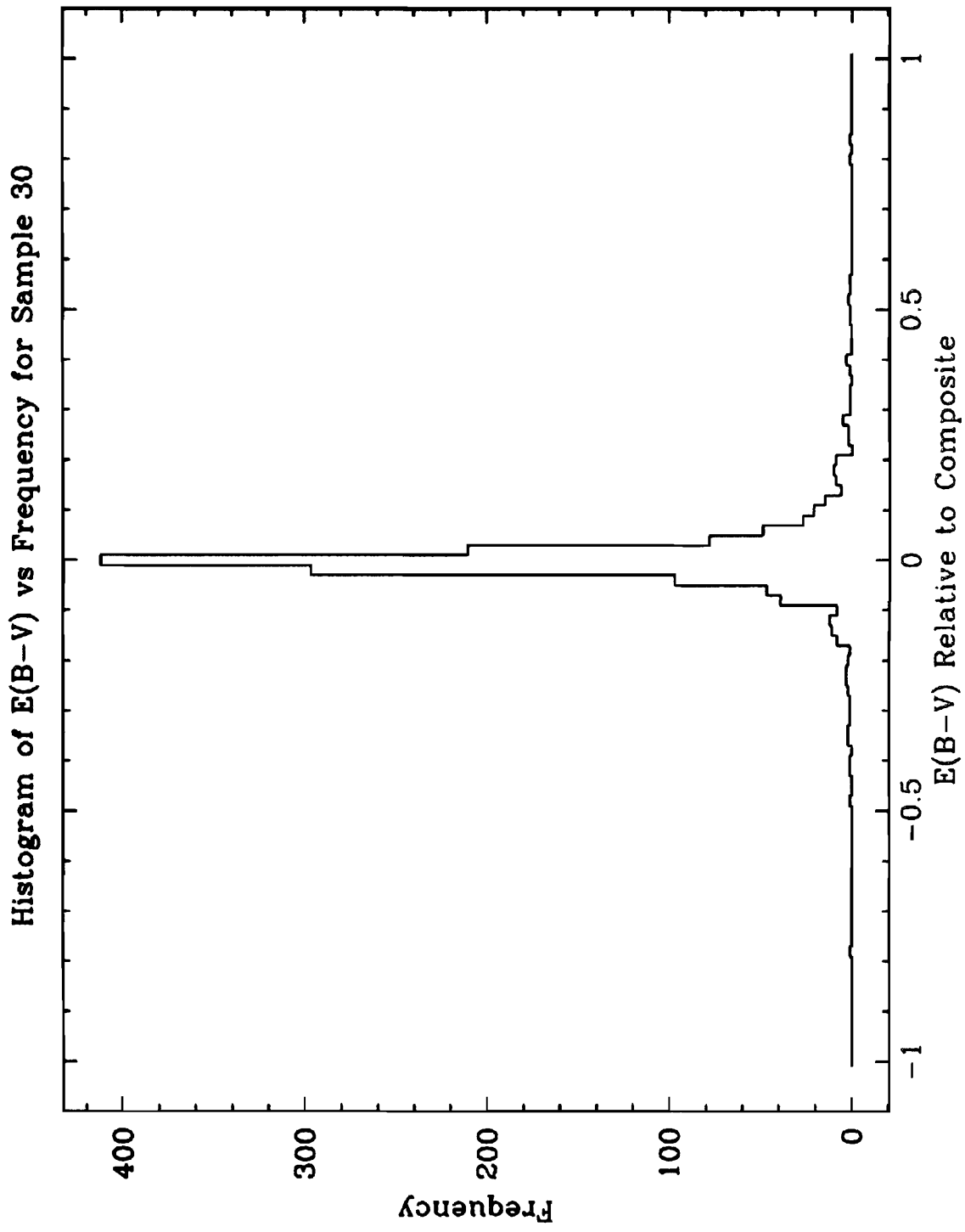

Figure 13 - Histogram plot of the frequency distribution per [E(B-V) = 0.02] Extinction bin.

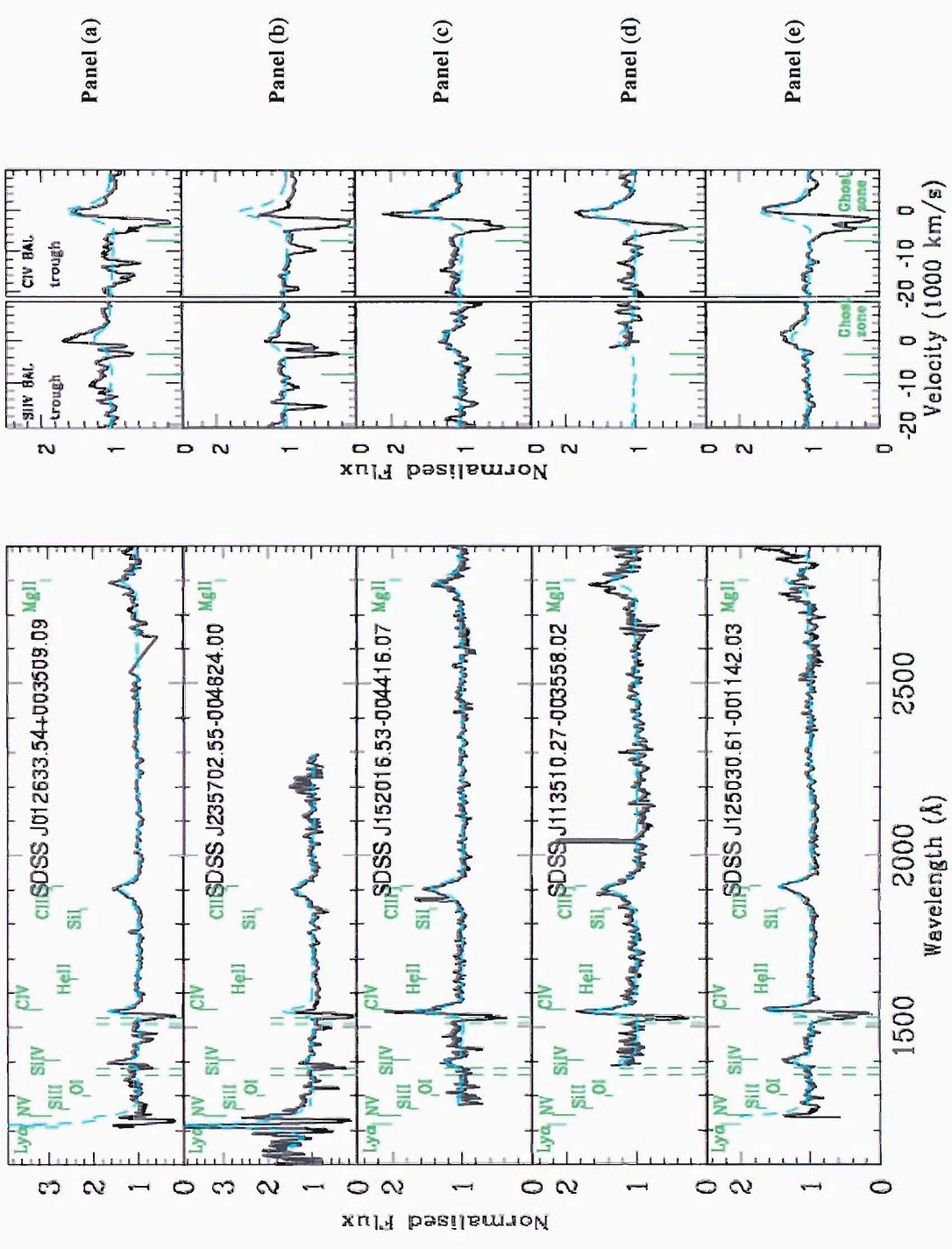

Figure 14 - Extract of 5 spectra from my new sample of BALQSOs in the EDR. The black lines represent the normalised spectra. The blue dashed lines display my EDR QSO sample composite. The panels on the right display the CIV and SiIV BALs in more detail. The vertical dashed green lines mark out the GZs in both the CIV and SiIV BALs.

4.4.2 Comparisons to the SDSS Redshifts

Having discussed the significance of the outputs generated by the fitting process in the context of optimising my algorithms 'fitting-regions' I will not cover the same ground again. I thus display all the same plots here, list any statistics quoted in § 4.4.1 for comparison, and discuss only the issues relating to the success of the redshift fitting process specifically.

Before discussing the results in detail, I first appraise the reader of the additional fitting parameters used to generate these improved redshifts. In particular, as mentioned earlier, the redshift routine applied here is producing an RMS value for the composite fit to the red wing of the main BELs (i.e. Lyα, NV, SiIV, CIV, CIII], and MgII). To achieve this, my first attempt just specified the BEL centre wavelengths and covered up to the red edges defined in the scaling fit. However, it became evident that this often resulted in the BEL peaks of all lines shifting too far to the blue. I consequently, for the CIII] and MgII lines only, specified a 'new' blue edge to encourage the algorithm to fit (shift) to the peaks of these lines. Thus subsequent 'best-fit' windows for the BEL 'red-wing' fits were determined: Lyα (λ1215.67Å-λ1230Å); NV (λ1240.15Å-λ1280Å); SiIV (λ1396.75Å-λ1420Å); CIV (λ1549.06Å-λ1570Å); CIII] (λ1900Å-λ1960Å); MgII (λ2789Å-λ2840Å).

The first plot displayed (fig.15), as detailed in § 4.3.4, is the frequency distribution of the redshift differences between the algorithms determination of redshift and the SDSS quoted redshift ($\Delta z = z_{SDSS} - z_{ALGORITHM}$). The points to note here are that approximately 50%, of the quasar sample fall within the region: $-0.005 \leq \Delta z \leq 0.005$. The remaining 50% are distributed fairly symmetrically about zero with a peak of nearly 50 quasars located at the most negative location ($\Delta z \sim -0.03$). Inspection of this sub-sample revealed that roughly half were correctly classified. In these instances the SDSS algorithm appeared to have centred on residual skylines still present in the spectra. Of the remaining fraction, approximately half again were slightly misaligned as a result of low S/N, or heavy absorption features, in the MgII BEL. Where this was the case, the SDSS redshifts were equally unreliable.

The remaining 25% of these 49 objects had badly determined redshifts by my algorithm due to two factors. Specifically, this was caused by a particularly high S/N with a very

narrow MgII BEL. This was initially confusing, but after due consideration it became clear that the following 'break' in logic had occurred:

- Given the high S/N, the continuum fitting process was matching the spectrum nearly perfectly.
- Given the narrow emission line of MgII (which carries a significantly higher weighting than all of the other lines) the initial redshift adjustment had shifted the MgII BEL totally out of the BEL 'scaling' region.
- The algorithm consequently computed a scaling factor of zero for the MgII BEL.
- Hence, given that the continuum fitting was nearly perfect, the subsequent RMS value returned was approximately zero (i.e. 'a perfect fit').

This issue may perhaps be useful in pointing a way towards an improvement to the method currently employed by the algorithm.

Figure 16 displays the Fractional RMS curve relative to the fraction of the sample. Here, 80% of the sample returned a fractional RMS less than 0.014. This is identical to the value returned when using the SDSS redshifts. However, 90% of the sample returned a fractional RMS of less than 0.056. This constitutes an improvement on the RMS values returned using the SDSS redshifts.

Figures 17 and 18 although displaying marginally different distributions are remarkably similar. Thus clearly shows that the success of the fitting process is still intact.

From the results discussed so far, any substantial success has yet to be revealed since the whole purpose of redefining the redshifts is to provide a level of accuracy beyond that of the SDSS and so to enable the accurate identification of Lyα-Ghost signatures. With this in mind, I draw the readers' attention to the final section in this chapter, § 4.4.3, where I revisit the RC3 sample and EDR Ghosts from chapter 3 and display them in Figures 19 and 20 with their new defined redshifts.

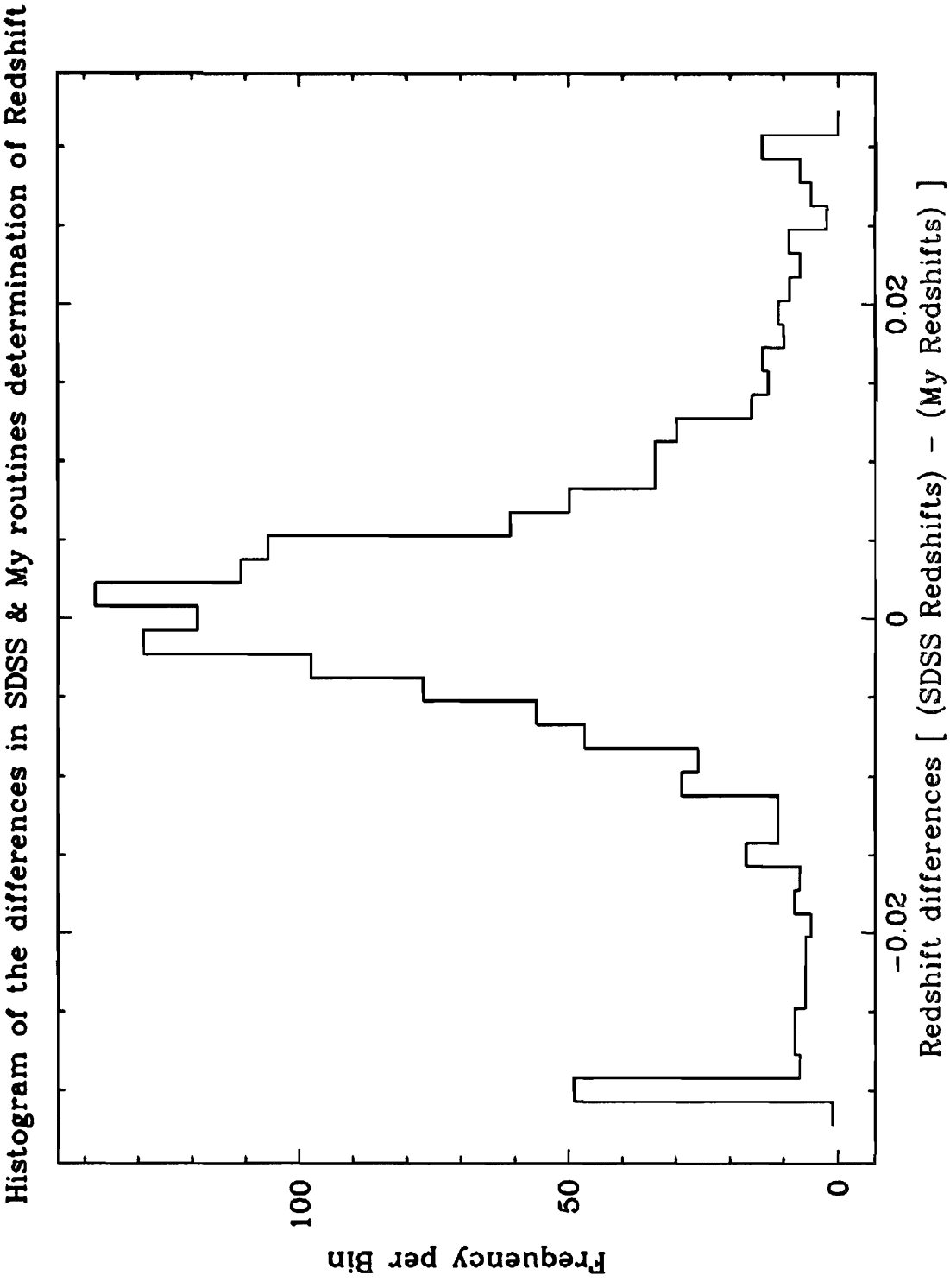

Figure 15 - Histogram plot of the frequency of QSOs per Redshift-difference [$\Delta z = 0.0014$] bin.

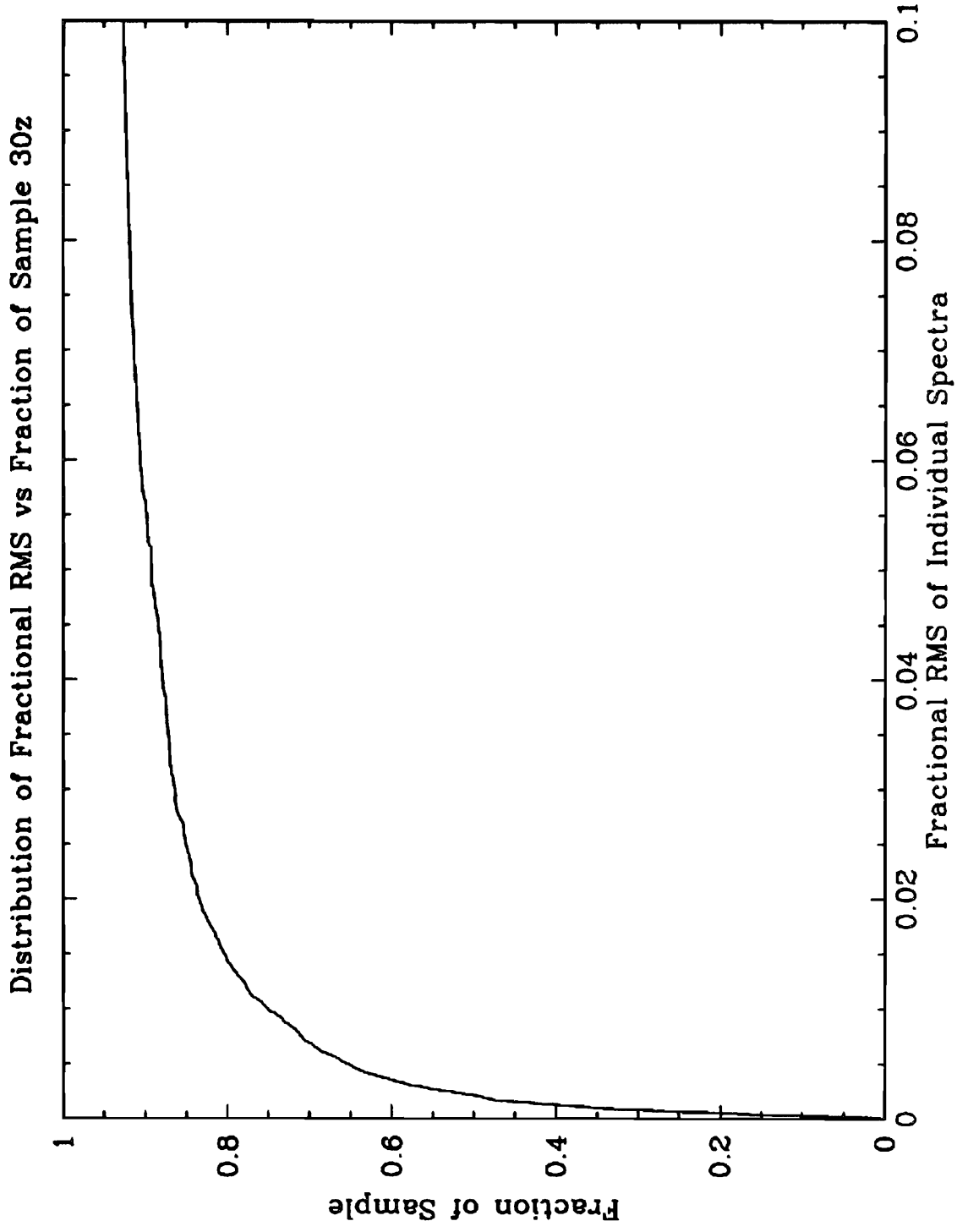

Figure 16 - Fractional RMS Distribution curve from the Redshift determination run with optimised fitting-regions.

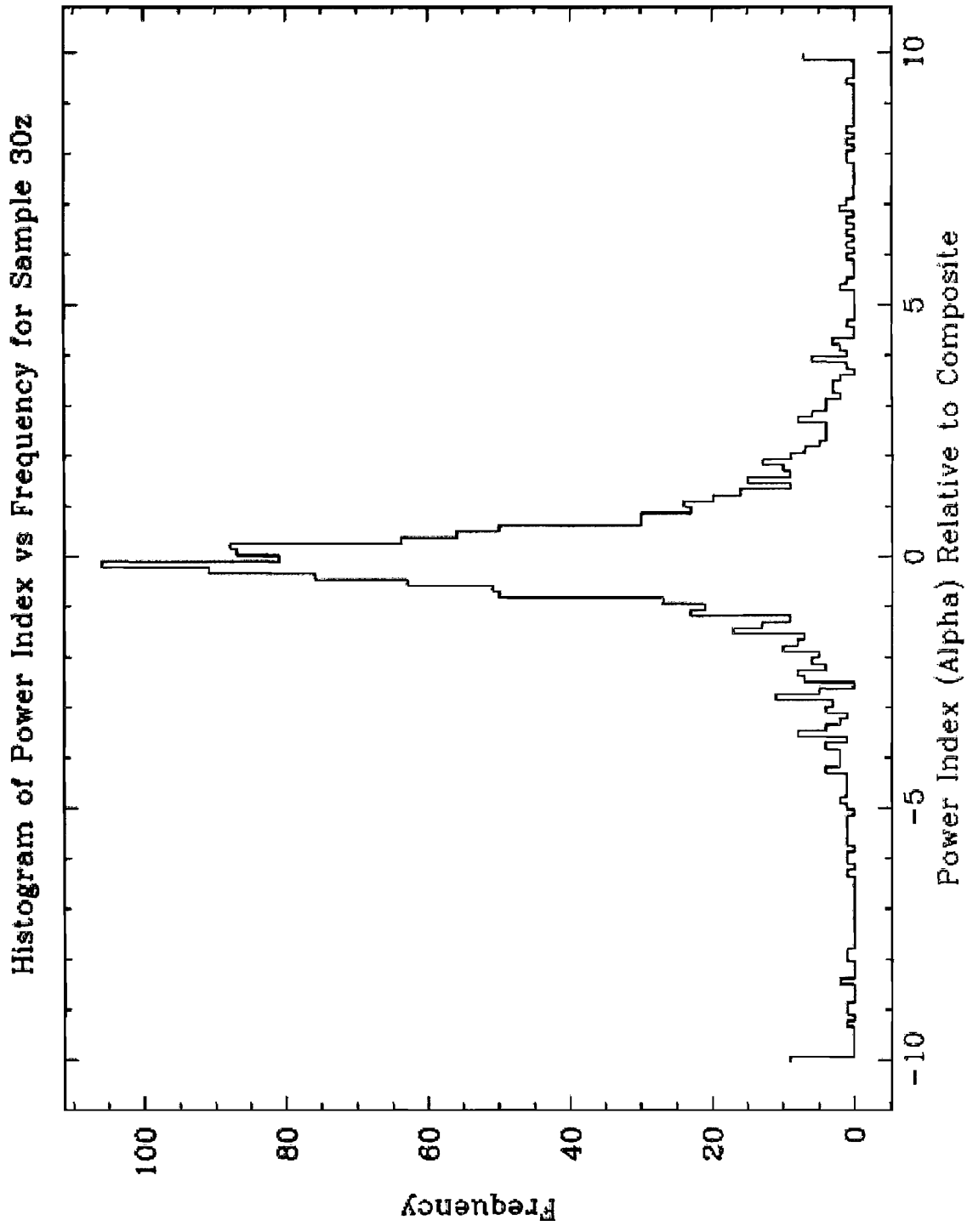

Figure 17 - Histogram plot of the frequency distribution per [a = 0.12] Power Law Index bin.

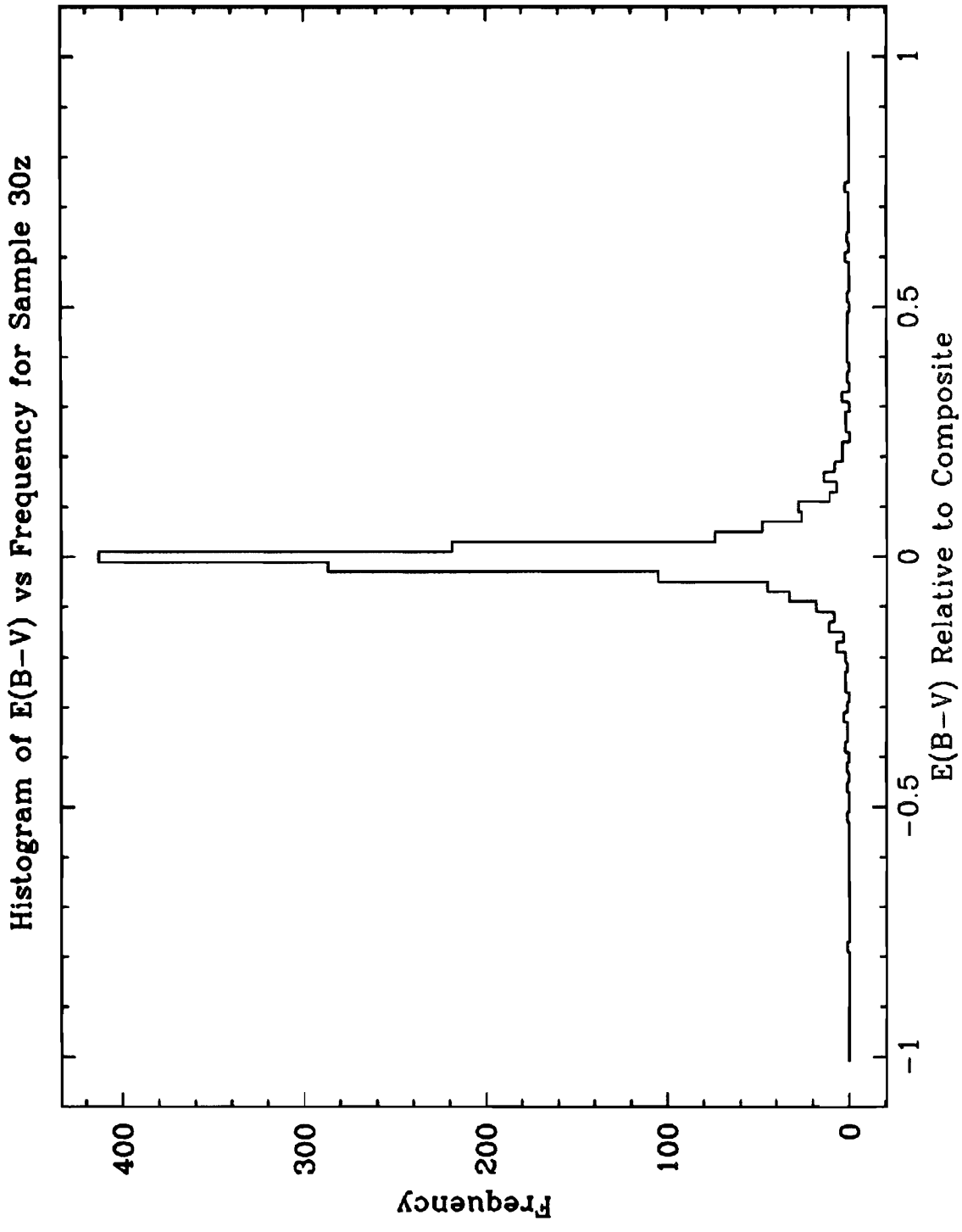

Figure 18 - Histogram plot of the frequency distribution per [E(B-V) = 0.02] Extinction bin.

4.4.3 Revisiting the EDR Ghosts

Displayed in figures 19 and 20 are the RC3 samples and EDR Ghost candidates, presented in chapter 3, with redefined redshifts. I summarise the results as follows:

With reference to figure 19 and the RC3 sample:

- SDSS 143022.47-002045.2, (figure 19 [a]), has been very slightly shifted toward the red. In the absence of a MgII BEL the CIII] BEL has provided the major influence here. The ghost feature has been shifted by about 1000km/s, into the ghost zone. Although the feature still clearly sits blue of centre its location is evidently improved.
- SDSS 145045.42-004400.3, (figure 19 [b]), has a well-defined MgII BEL and consequently the routine has relocated the redshift so as to centre this emission line. It is also evident from closer inspection that the CIII] BEL is now also better placed. The consequence of this is to improve the location of the ghost feature within the GZ and provides for another good candidate.
- SDSS 171330.98+610707.8, (figure 19 [c]), via inspection of the MgII BEL it is obvious that the redshift, as determined by the fit to this line, has been improved. However, the resulting shift has moved the ghost feature about 2000km/s to the blue of its original location and thus shifts this feature further away from the GZ. Consequently this has to be considered an unlikely ghost candidate. It is interesting to note however that, despite the clarity of the feature itself, the profile is clearly not mimicking that of the CIV BEL. This is not the case for any of the other 4 RC3 candidates displayed here.
- SDSS 113544.33+001118.6, (figure 19 [d]), still has a well-positioned feature despite being subject to a minor shift to the red. The shift, again, is well justified as the MgII BEL is well centred on the composite emission line. It also appears that the CIV BEL is also better located here.
- SDSS 110736.67+000329.4, (figure 19 [e]), similarly remains a well-positioned feature despite also shifting slightly to the red. However, as before the shift seen here has clearly improved the locations of all the visible emission lines relative to the EDR composite.

With reference to figure 20 and the EDR Ghost set:

- SDSS 132304.58-003856.5 (figure 20 [a]), SDSS 172001.31+621245.7 (Figure 20 [c]), SDSS 033048.51-002819.6 (Figure 20 [d]), SDSS 023252.80-001351.2 (Figure 20 [f]), SDSS 110623.52-004326.0 (Figure 20 [g]) have all had their respective ghost features well centred within their GZs.
    - SDSS 132304.58-003856.5, SDSS 172001.31+621245.7, SDSS 033048.51-002819.6 and SDSS 023252.80-001351.2 all have MgII present in their spectra, and the redshifts that have been returned are clearly an excellent fit.
    - Only SDSS 110623.52-004326.0 lacks MgII but clearly the fit to the CIII here cannot be improved with this spectrum.
- SDSS 142050.34-002553.1 (Figure 20 [b]) and SDSS 170056.85+602639.8 (Figure 20 [e]) have both returned a fit that has put their respective ghost features on the edges of their GZs.
    - SDSS 142050.34-002553.1 appears to suffer from substantial absorption features in both the red and blue wings of its MgII BEL, which may well have generated an unreliable redshift determination. However, one cannot condemn the algorithm for this alignment as clearly it has fitted the given feature exceptionally well.
    - SDSS 170056.85+602639.8, despite covering the wavelength range of MgII, obviously lacks a significant MgII BEL. Again there is evidence here that it has been absorbed away. Irrespective though, review of the CIII BEL clearly indicates that a marginal shift of the spectrum toward the red is perhaps necessary. This would serve to re-centre the ghost feature in its zone.

Review of figures 9 and 10 in chapter 3 does in fact show a clear improvement in all of the spectra displayed here apart from SDSS 170056.85+602639.8 and SDSS 110623.52-004326.0. In the case of SDSS 110623.52-004326.0 there is no improvement because there is no perceivable change in redshift. Thus it is only in the case of SDSS 170056.85+602639.8 where a marginally incorrect redshift determination was obtained.

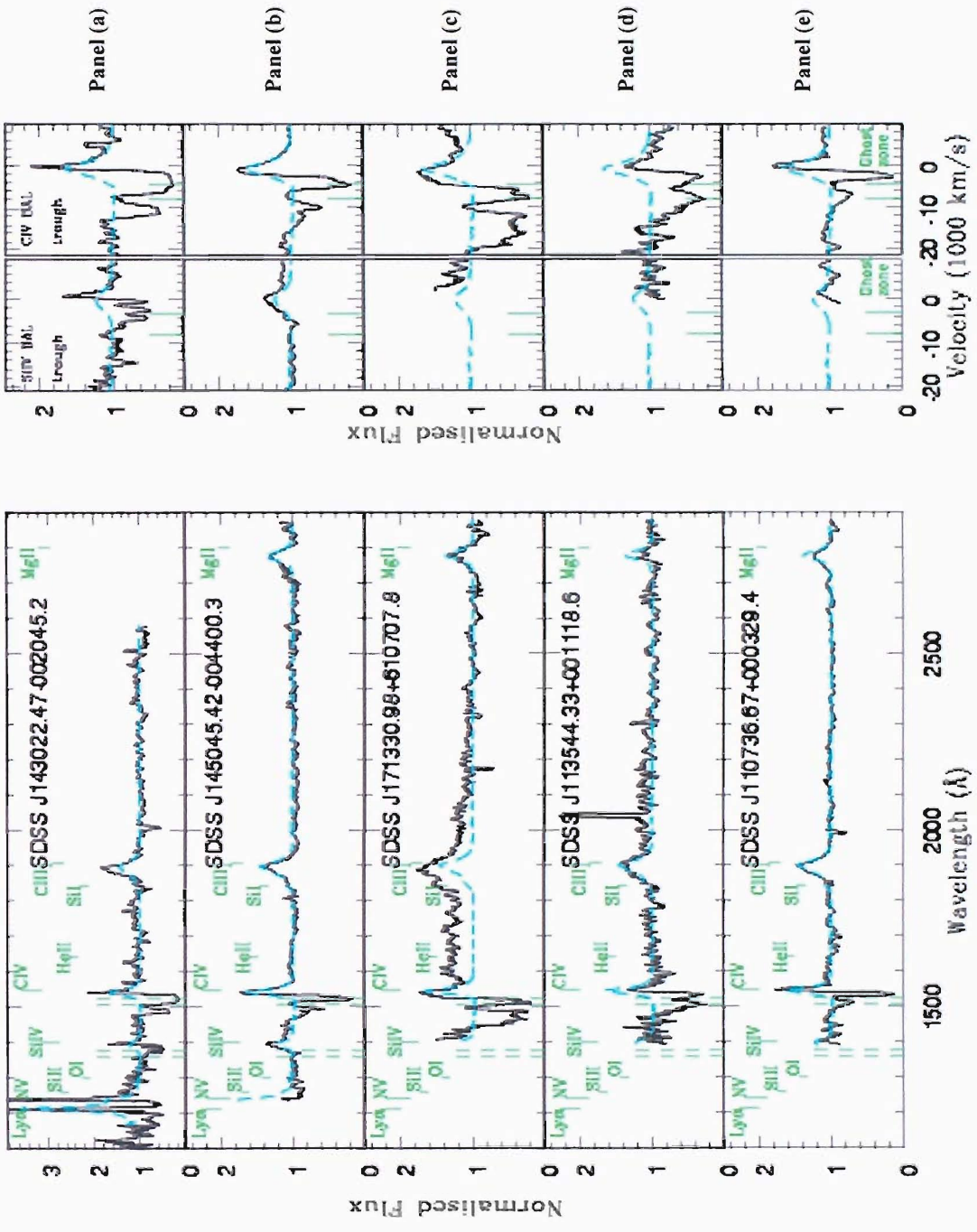

Figure 19 - The Rejection Cut 3 sample, re-plotted with the redshifts generated by my algorithm. The black lines represent the normalised spectra. The blue dashed lines display my EDR QSO sample composite. The panels on the right display the CIV and SiIV BALs in more detail. The vertical dashed green lines mark out the GZs in both the CIV and SiIV BALs.

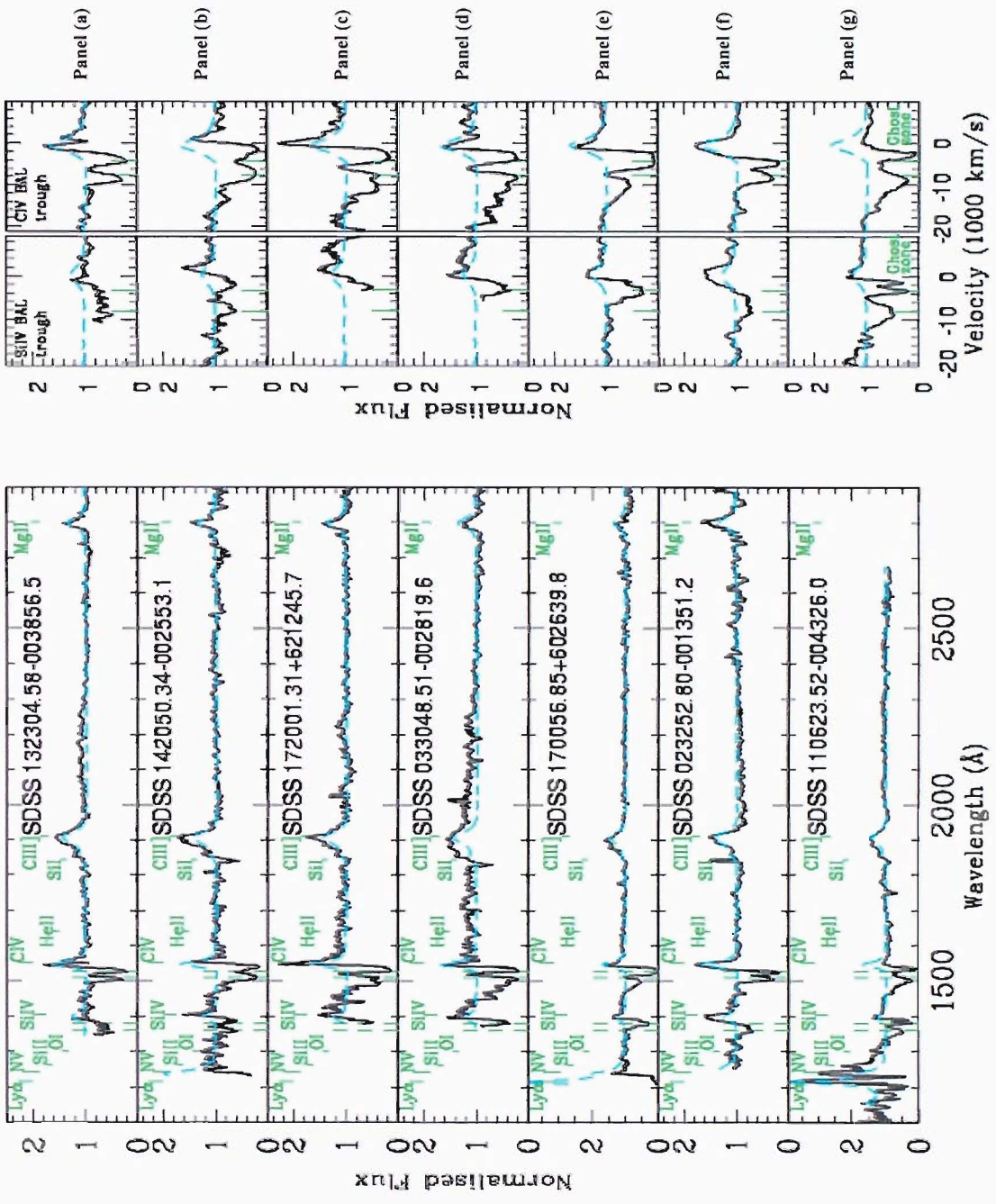

Figure 20 - The EDR Ghost set, re-plotted with the redshifts generated by my algorithm. The black lines represent the normalised spectra. The blue dashed lines display my EDR QSO sample composite. The panels on the right display the CIV and SiIV BALs in more detail. The vertical dashed green lines mark out the GZs in both the CIV and SiIV BALs.

4.5 Chapter Summary

Within this chapter I have presented a new algorithm for the identification of BALQSOs. The results from the initial testing of this algorithm clearly indicate that this development has been successful. In particular, I have displayed five new BALQSOs from the SDSS EDR, which were not found by Reichard et al's algorithm. In addition, as previously discussed in § 4.3, R2003a decline from inferring any 'physicality' from the parameters (α & E) fitted. However, I have no such reservations here as, I believe that I have illustrated that my routine generally does. Furthermore, the algorithm developed herein has successfully identified 189 BALQSOs from the 198 published in the EDR BALQSO catalogue, which also compares favourably with the 162 found by R2003a's algorithm.

Finally, although I have identified still more room for improvement, the redshift determination has also been shown to provide an increased level of accuracy over and above that of the SDSS's algorithm. In doing so, it has broadly borne out the communication from Donald Schneider, where he asserted that, based on inspections 'by-eye' a redshift error of 0.01 is a fairer reflection of the errors in the redshift determinations, than the formal redshift errors provided by the SDSS.

5.0 Discussion

The focus of this thesis has been twofold:

[1.] Identification of ghost signatures in the SDSS EDR BALQSO Catalogue and the construction of a new sample of ghost candidates.

[2.] Determining Balnicity Indices for QSOs to provide further sets of BALQSOs that may display ghost signatures and redshift improvement.

Consequently, this discussion is divided accordingly and I address part each in turn.

5.1 Ghosts in the EDR BALQSO Catalogue

With respect to the composites produced in chapter 3, the interesting point to note is that by purely selecting BALQSOs that have MT structure, we immediately see a feature appearing within the GZ. Whether this could just be a random occurrence was the focus of a paper by Korista et al (1993). Korista analysed a set of quasars, identified by Weymann, displaying a double trough structure and questioned whether a Monte Carlo simulation could recreate the same statistical results. His conclusion tended toward this not being random. Furthermore, on considering the consistent structure seen in the RMS spectra as well, (see §3.2) the conclusion that Arav's theory is indeed at work in BALQSOs becomes quite convincing. This view is obviously further compounded by inspection of the Ghost candidates I present from the SDSS EDR BALQSO catalogue.

With respect to the individual ghost candidates themselves, there is a disappointing lack of coverage of the Lyα BEL (hence also the Lyman forest). Consequently a comprehensive testing of Arav's model has not been possible. It is interesting to note that Arav's suggestion that the CIV BEL is similar in shape to that of the Lyα BEL and hence will also mimic the profile of the ghost, does in fact appear to be the case here.

On a slightly different tack, it is worth considering that even if Arav's model for the construction of a ghost feature is correct. It is not clear that his suggestion that the features will correlate well with the Lyα BEL is correct. Specifically, the model is based

upon the strong, preferential, scattering of Lyα photons by the 'resident' NV ions, thus is it not necessarily reasonable to assume that the resulting redistribution of Lyα photons will remain in our view. In other words, in any BALQSO whose viewing angle is such, and dynamics are such, that we see a strong ghost signature, then dependent upon the geometry of the QSO, one may expect that the Lyα emission would be severely reduced. Obviously, this model requires a strong emission for the phenomenon to occur but it is not necessarily reasonable to assume that both will be observed simultaneously.

Again, changing tack slightly, there may be an interesting by-product of this dynamic. If the various emission line shifts, are indeed, caused by red / blue wing absorption (see § 4.2) of the various BELs and is not actually a reflection of the 'real' emission from these objects then we can safely assume that the rest-wavelength calculation of where the ghost feature should lie is actually where it is, with respect to systemic. Hence, given the ability to correlate certain properties of the ghost with other features in a spectrum (as yet still to be achieved) then, knowing that we have found a ghost might enable a highly accurate systemic redshift to be determined. Currently, establishing an accurate redshift is a difficult issue with BALQSOs, as well as QSOs in general.

5.2 Automated Continuum Fitting and the Production of BIs

It is evident from the plots presented throughout this thesis that the composite fitting method detailed herein is particularly successful at providing high quality fits to quasar spectra. This automated routine correctly provides BIs for the majority of the QSOs analysed.

Despite these successes though, there are still some improvements to be made. In particular it would be ideal to have an algorithm that could discover all, and only, bone fide BALQSOs with a highly reliable redshift.

As projects like the SDSS produce ever-larger data sets for analysis, automated methods of this type are becoming increasingly important. Clearly, the results herein justify the use of such approaches and moreover, even without any further improvements, the algorithm I have constructed is evidently fully operational.

In summary of this discussion here, one generic conclusion I consequently draw is that there is certainly more work to be done in this field. I have a tool for the selection of BALQSOs, despite certain inherent difficulties, from which more ghost features will surely be found. And, given coverage of the spectrum shortward of Lyα, a comprehensive analysis of this phenomenon and its related properties will undoubtedly yield a greater understanding of BALQSO outflows.

6.0 Conclusions

The primary purpose of the research undertaken here was to attempt to expand the current sample of BALQSOs displaying the radiation driving signature known as the Ghost of Lyα. With access to the SDSS EDR BALQSO catalogue, this goal was achieved. The immediate consequence of this was to promote the development of a tool for the continued identification of BALQSOs to expand the ghost-set still further.

The results of this study will hopefully encourage a greater level of scrutiny to be given to the outflows of QSOs and understanding the physical mechanisms that drive them. This will hopefully provide an increased level of precision in determining parameters such as the covering fraction of these outflows. It may also indirectly imply what role, if any, magneto-hydro-dynamical processes play in driving winds from the centres of AGN.

I would further hope, particularly in the light of the lack of Lyα coverage, that spectroscopic follow ups will be carried out on at least the top 7 candidates identified in § 3.3. This could enable a full testing of Arav's theory to be undertaken (by providing coverage of Lyα) as well as providing insight into the variability of the features and their associated outflows. On this point, WHT service time has already been provided for this purpose, but a full analysis of the spectra obtained has not been possible within the time constraints of this research. However, only three targets, present in table 1.0, were obtained as part of this proposal. I have indicated which three in table 1.0 along with the basic attributes of the MT BALQSOs within the EDR BALQSO catalogue. I include in this table the magnitude and S/N in the SDSS g' band. This table also includes their redshifts and BIs as calculated by my routine (Chapter 4) as well as Reichard's composite-fit calculated BIs. I have included table 1.0 to provide the reader with the relevant information to assist in target selection of any of these BALQSOs for further spectroscopic study.

Finally, the algorithm developed herein has been extensively tested, and the initial results are clearly excellent. The algorithm's ability to return realistic values for the power law index, α, and extinction, E(B-V), are compelling. Coupled with the general quality of each fit, illustrated throughout all of the plots presented here, this routine has proved its potential. With the SDSS DR3 there is an immediate opportunity to put this

routine to work. Broadly speaking from ~50,000 QSOs I expect to find ~5,000 – 10,000 BALQSOs. This could well lead to the discovery of ~200 – 400 BALQSOs exhibiting ghost of Lyα signatures. Clearly these statistics will radically alter the current data sets available, which may in turn provide a platform from which to further enhance our understanding of QSOs, and perhaps AGN more generally.

Table 1.0

Object ID	My BI SDSS Z	My BI Z Adjusted	Reichards BI	My Z	SDSS Z	Magnitude (in g′)	S/N (in g′)	Category[13]
J110623.52-004325.10	12,130	11,580	4,034	2.444	2.45	19.88	10.14	GC_FC
J023252.80-001351.02[14]	4,968	5,106	2,092	2.028	2.025	19.19	10.16	GC_FC
J170056.86+602639.08	6,000	5,862	1,400	2.132	2.125	19.24	9.427	GC_FC
J033048.51-002819.06	11,710	11,230	5,548	1.77	1.779	20.06	6.593	GC_FC
J172001.31+621245.08	5,447	4,550	3,290	1.75	1.762	19.79	5.781	GC_FC
J142050.34-002553.01	6,761	7,310	3,442	2.085	2.103	20.28	4.345	GC_FC
J132304.58-003856.07	2,554	2,416	287	1.822	1.828	18.78	8.491	GC_FC
J170056.86+602639.08	6,000	5,862	1,400	2.132	2.125	19.24	9.427	PG_RC3
J033048.51-002819.06	11,710	11,230	5,548	1.77	1.779	20.06	6.593	PG_RC3
J172001.31+621245.08	5,447	4,550	3,290	1.75	1.762	19.79	5.781	PG_RC3
J142050.34-002553.01	6,761	7,310	3,442	2.085	2.103	20.28	4.345	PG_RC3
J132304.58-003856.07	2,554	2,416	287	1.822	1.828	18.78	8.491	PG_RC3
J005355.15-000309.03	3,509	3,647	1,088	1.713	1.715	18.72	13.23	PG_RC2
J010616.06+001524.00	8,754	8,823	2,520	3.038	3.05	20.5	2.649	PG_RC2
J010612.22+001920.03[14]	10,620	11,720	2,453	2.321	2.3	19.42	7.406	PG_RC2
J020006.31-003709.09[14]	11,900	12,670	9,550	2.111	2.136	19.03	14.02	PG_RC2
J025042.45+003536.07	8,070	9,034	3,544	2.393	2.38	19.73	7.418	PG_RC2
J100809.64-000209.08	276	830	56	2.551	2.561	19.81	5.162	PG_RC2
J104109.86+001051.08	5,519	5,450	1,913	2.253	2.25	19.66	12.09	PG_RC2
J104233.86+010206.03	1,519	139	401	2.102	2.123	19.18	14.02	PG_RC2
J104841.03+000042.08	3,450	3,381	1,176	2.023	2.022	19.48	12.73	PG_RC2
J120657.01-002537.08	1,795	2,554	110	2.014	2.005	19.38	6.306	PG_RC2
J123947.61+002516.02	9,310	10,760	7,299	1.861	1.869	20.82	2.972	PG_RC2
J130035.29-003928.03	4,054	4,123	853	3.636	3.63	20.47	2.994	PG_RC2
J134544.55+002810.09	5,096	2,817	1,510	2.544	2.516	18.75	11.07	PG_RC2
J134808.80+003723.02	3,175	0	1,309	3.65	3.62	20.66	2.697	PG_RC2
J143054.04-003627.04	6,810	6,810	9,064	3.712	3.71	22.16	0.3717	PG_RC2
J145913.72+000215.09	1,795	1,588	356	1.909	1.91	19.05	12.99	PG_RC2
J151636.78+002940.05	8,582	8,582	4,035	2.24	2.24	18.8	12.65	PG_RC2
J171944.76+554408.05	1,450	553	205	3.863	3.886	21.28	1.23	PG_RC2
J171949.92+532132.09	6,692	6,071	4,903	1.771	1.777	18.56	16.94	PG_RC2
J173911.53+565551.00	1,243	1,657	919	1.777	1.772	19.53	8.888	PG_RC2
J234506.34+010135.07	7,032	6,894	2,488	1.792	1.794	19.61	7.732	PG_RC2
J134145.13-003631.02	3,589	5,589	870	2.231	2.205	19.8	8.652	PG_RC2
J000056.91-010409.06	5,520	4,554	1,560	2.099	2.111	20.62	3.437	PG_RC2
J143022.47-002045.02	4,139	3,311	1,957	2.533	2.544	20.7	2.473	PG_RC2
J003551.99+005726.06	3,727	3,727	1,731	1.905	1.905	19.44	8.533	PG_RC1
J004041.39-005537.03	138	138	0	2.082	2.092	18.39	17.24	PG_RC1
J011227.61-011221.07	5,981	5,981	3,033	1.755	1.755	18.5	16.32	PG_RC1
J012913.72+011428.00	1,863	1,793	345	1.776	1.782	19.72	6.105	PG_RC1
J015048.83+004126.03	1,381	0	105	3.675	3.703	19.75	5.479	PG_RC1
J024221.86+004912.07	1,577	2,198	229	2.078	2.071	18.66	16.48	PG_RC1
J031227.13-003446.03	2,206	4,481	0	1.759	1.772	19.59	7.315	PG_RC1

[13] The categories here are: GC – 'Ghost Candidate'; PG – 'Possible Ghost'; FC – 'Final Cut' with RC1, RC2 & RC3 indicating the 'Rejection Cut' samples of Chapter 3.
[14] WHT spectroscopy has been obtained.

J104152.62-001102.02	3,380	4,347	1,588	1.694	1.703	19.45	12.58	PG_RC1
J110041.20+003631.10	11,230	11,030	4,687	2.02	2.017	18.83	15.03	PG_RC1
J121803.28+001236.10	2,267	2,129	269	2.011	2.01	19.06	8.181	PG_RC1
J122228.40-011010.10	2,829	2,621	678	2.305	2.284	19.91	5.29	PG_RC1
J123124.72+004719.02	6,829	7,794	3,134	1.729	1.72	19.69	5.513	PG_RC1
J123824.91+001834.06	1,726	1,795	220	2.148	2.154	19.67	6.652	PG_RC1
J130348.94+002010.06	8,070	7,932	1,425	3.632	3.655	21.02	1.525	PG_RC1
J170903.06+594530.07	6,964	6,964	4,936	1.707	1.708	19.32	10.25	PG_RC1
J170930.100+630357.01	11,790	10,820	0	2.39	2.402	18.75	13.09	PG_RC1
J170951.03+570313.07	3,796	3,037	528	2.538	2.547	21.07	2.505	PG_RC1
J172012.39+545601.02	2,070	2,070	1,249	2.1	2.099	18.8	21.92	PG_RC1
J232205.46+004551.01	967	346	222	1.814	1.82	20.29	3.032	PG_RC1
J032246.82-005148.10	1,726	1,726	0	1.679	1.68	19.62	7.193	PG_RC1
J130208.27-003731.06	5,451	5,658	0	1.674	1.672	18.71	13.51	PG_RC1
J125241.55-002040.06	2,671	2,671	2,524	2.896	2.898	18.97	10.23	PG_RC1

Bibliography

Arav, N., Li, Z. 1994, ApJ, 427, 700.

Arav, N., Li, Z., Begelman, M.C. 1994, ApJ, 432, 62.

Arav, N., Begelman, M.C. 1994, ApJ, 434, 479.

Arav, N., Kotista, K.T., Barlow, T.A., Begelman, M.C. 1995, Nature, 376, 576.

Arav, N. 1996, ApJ, 465, 617.

Artymowicz, P. 1993, PASP, 105, 1032.

Elvis, M. 2000, ApJ, 545, 63.

Espey, B.R., Carswell, R.F., Bailey, J.A., Smith, M.G., Ward, M.J. 1989, ApJ, 342, 666.

Francis, P.J., Hewett, P.C., Foltz, C.B., Chaffee, F.H., Weymann, R.J., Morris S.L. 1991, ApJ, 373, 465.

Frank, J., King, A., Raine, D. 2002 ... 'Accretion power in astrophysics.' (Third Edition), Cambridge University Press.

Fukugita, M., Ichikawa, T., Gunn, J.E., Doi, M., Shiasaka, K., Schnieder, D.P. 1996, AJ, 111, 1748.

Gaskell, C.M. 1982, ApJ, 263, 79.

Gunn, J.E., et al. 1998, AJ, 116, 3040.

Gupta, N.,Srianand, R., Petijean, P., Ledoux, C. 2003, A&A, 406, 65.

Knigge, C. 1995, Oxford University, PhD Thesis.

Korista, K.T., et al. 1992, ApJ, 401, 529.

Korista, K.T., Voit, G.M., Morris, S.L., Weymann, R.J. 1993, ApJS, 88, 357.

MacAlpine, G.M. 1981, ApJ, 251, 465.

Mathews, W.G., Ferland, G.J. 1987, ApJ, 323, 456.

Menou, K., et al. 2001, ApJ, 561, 645.

Pei, Y.C. 1992, ApJ, 395, 130.

Peterson, B.M. 1997 ... 'An introduction to active galactic nuclei.' Cam. Uni. Press.

Reichard, T.A., et al. 2003, AJ, 125, 1711 ..... (R2003a)

Reichard, T.A., et al. 2003, AJ, 126, 2594 ..... (R2003b)

Richards, G.T., et al. 2002, AJ, 124, 1.

Schnieder, D.P., et al, 2002, AJ, 123, 567.

Starck, J., Siebenmorgen, R., Gredel, R. 1997, ApJ, 482, 1011.

Tytler, D., Fan, X. 1992, ApJS, 79, 1.

Vanden Berk, D.E., et al. 2001, AJ, 122, 549.

Weymann, R.J., Morris, S.L., Foltz, C.B., Hewett, P.C. 1991, ApJ, 373, 23.

Wilkes, B.J. 1986, MNRAS, 218, 331.

York, D.G., et al. 2000, AJ, 120, 1579.