DEVELOPING A PHOTOMETRIC ESTIMATE OF QUASAR REDSHIFTS:

An independent measurement of the distances to the most powerful objects in the Universe

by

Jeremy Moss

A thesis
submitted to the Victoria University of Wellington
in fulfilment of the requirements for the degree of
Master of Science

Victoria University of Wellington

2019

Contents

List of Figures

List of Tables

Abstract

While spectroscopy is the standard method of measuring the redshift of luminous objects, it is a time-intensive technique, requiring, in some cases, hours of telescope time for a single source. Additionally, spectroscopy favours brighter objects, and therefore introduces an intrinsic bias towards luminous or closer sources. A simple method of estimating the redshift through photometry would prove invaluable to forthcoming surveys on the next generation of large radio telescopes, as well as alleviating the inherent bias towards the most optically bright sources. While there is a well-established correlation between the near-infrared K-band magnitude and redshift for galaxies, we find that the K-z relation breaks down for samples dominated by quasi-stellar objects (QSOs).

Current methods of estimating photometric redshift rely either on template spectra, which requires a high number of infrared photometry points, or computationally intensive machine learning methods.

Using photometric data from the Sloan Digital Sky Survey (SDSS) we investigate the relationship between combinations of magnitudes of a group of quasars, and their redshift. We find a high correlation between the colour relation (I-W2)/(W3-U) and redshift for a group of broad-line emission sources from the SDSS, and we conclude that this could be a robust estimator of the redshift.

1 Introduction

1.1 Quasi-Stellar Objects and Quasars

1.1.1 Discovery

Karl Jansky is credited with the first detection of radio waves from an astronomical object, when in 1932, he observed radiation coming from the Milky Way, which was particularly strong in the direction of the constellation of Sagittarius. By the 1960s, rapid advances had been made in the field of radio astronomy, and Hazard et al. (1963) performed an in-depth study of a star-like object labelled 3C 273. Comparing the spectrum of 3C 273 with the Balmer lines of hydrogen showed that the absorption lines in its spectrum were highly redshifted, by a factor of 1.16, with reference to the laboratory frame (Figure 1.1). Additionally, while stars can exhibit proper motion against the sky, typically on the order of ~ 0.1 arcsec/year, 3C 273 had a proper motion of less than the detection limit at the time (~ 0.01 arcsec/year). This suggested that the source was at a far greater distance than any star yet observed.

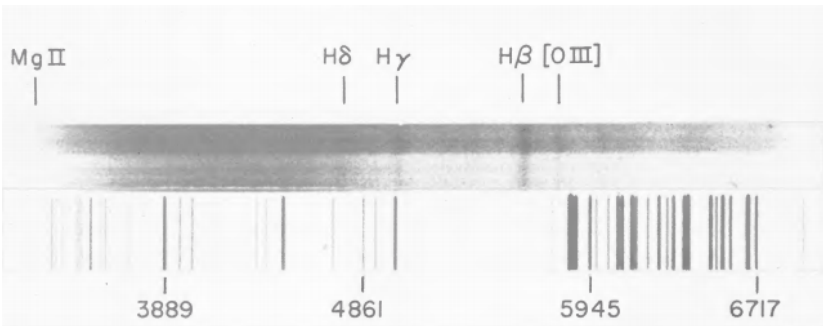

Figure 1.1: The original spectrum, taken by Greenstein and Schmidt (1964), of the quasar 3C 273. Comparison spectrum is H + He + Ne. Redshifted emission lines of H and OIII are indicated.

Correcting for cosmological expansion yielded a redshift for 3C 273 of $z = 0.154$ (for the then-current value of $H_0 \approx 100$ km.s^{-1} Mpc^{-1}). Although consistent with the observed lack of any proper motion against the sky, such a high redshift was unheard of, and Greenstein and Schmidt (1964) published a detailed analysis of 3C 273, in which they considered explanations of the redshift by rapid motion of objects in or near the Milky Way, and gravitational and cosmological redshifts. The lack of any observed proper motion of the source, an absolute magnitude closer to that of galaxies than stars, and the width of the emission lines (see, for example, Section 1.1.3) all pointed to cosmological expansion as being the most likely cause of the redshift (Gunn, 1971). Furthermore, since 3C 273 varies in flux by about 30% in a year, it could not be larger than $\sim$1 pc across, due to the minimum time it

would take a light signal to cross the diameter. It was these arguments that established the source to be a highly compact, but extremely powerful source of radio light – the source of this power is now understood to be a supermassive black hole, with a mass on the order of $\sim 10^9$ $M_\odot$.

A star-like object with a very high redshift (and thus a large distance from Earth) is now called a *quasi-stellar object* (QSO). Some QSOs also have a high luminosity in the radio spectrum, in which case it is referred to as a *quasar*. The terms QSO and quasar are generally grouped under the more general name *Active Galactic Nuclei* (AGNs).

1.1.2 The Nature and Phenomenology of AGN

Quasars, being very luminous AGN, are visible across the Universe; the most distant quasar known is almost 29 billion light-years away (in terms of co-moving distance); the most luminous quasars can have a core luminosity of $\sim 10^{49}$ erg s^{-1}. By comparison, the Milky Way Galaxy's (MW) luminosity is $\sim 10^{44}$ erg s^{-1}. AGN are now understood to be associated with an actively feeding supermassive black hole (SMBH) at the core of a galaxy, and are believed to be powered by accretion of matter onto the central black hole (BH). In fact, it is thought that almost every spiral galaxy harbours an SMBH in its nucleus, and thus probably undergoes at least one AGN phase throughout its lifetime. Indeed, Su et al. (2010) report on two large gamma-ray bubbles, extending as much as 50° above and below the plane of the MW, imaged by the Fermi Gamma-Ray Large Area Telescope, *Fermi-LAT*. It is thought that these bubbles are the result of some large episode of energy injection into the SMBH at the Galaxy's core.

As shown in Figure 1.2, the orientation-based obscuration model of Urry and Padovani (1995) places a large, hydrostatically equilibrated torus around the accretion disk and the BLR. Observers for whom the central SMBH and BLR are obscured see either a radio galaxy, or a Seyfert 2 galaxy (a type of QSO with a low radio luminosity), depending on the total radio flux being produced by the AGN. This torus acts to collimate the NLR, leading to jet-like plumes of line-emitting gas being ejected axisymmetrically from the SMBH. The NLR is generally axially aligned with polar jets of X-ray emitting relativistic electrons, and an AGN whose jet happens to be aligned with our viewpoint looks like an extremely bright X-ray source called a blazar. Figure 1.3 is a schematic representation of the unified model, illustrating the approximate scale and spectral contributions of the main regions of an AGN.

As matter falls onto the black hole, it emits energy as electromagnetic radiation of varying wavelengths. Most of the radiation emitted by the accretion disk is due to accretion flow onto the central BH, as well as frictional heating as a consequence of differential rotation within the disk. In order to produce the observed AGN luminosities of $L \sim 10^{44} - 10^{49}$ erg s^{-1}, matter must be accreted at a rate of ~ 2 M$_\odot$ yr^{-1}.

1.1.3 AGN spectra

The continuous spectrum of an AGN is unlike that of stars. A stellar spectrum has a blackbody curve, having a peak at some wavelength $\lambda_{\rm max}$ near the optical bands. The hottest stars emit a significant fraction of their light at ultraviolet wavelengths, while cooler stars tend to emit mostly in the red to infrared bands. This gives stars, and therefore quiescent galaxies, whose light is dominated by the stellar population, the characteristic thermal spectrum typified by the blackbody curve. AGN, on the other hand, are luminous in the X-ray, ultraviolet, visible, infrared and, in the case of quasars, radio bands, and they have similar power at all wavelengths, and are thus not a single blackbody object. The radio spectrum is the result of emission from synchrotron radiation produced by charged particles spiralling around magnetic field lines at relativistic speeds, while the AGN contribution depends on whether the central SMBH is obscured by a dusty circumnuclear medium. Since the medium may not fully obscure the central source, some of the high-frequency radiation escapes, and leads to the production of emission line contributions, notably from Mg IIλ2799, the semi-forbidden C III] λ1908, (see Figure 1.9) O IIλ3727, O IIIλ1666 and H$\alpha\lambda$6565 leading to a skew towards higher magnitudes at certain redshifts. In certain AGN, the optical spectrum is dominated by non-thermal emission. This is especially true of QSOs, for which the continuum radiation from the central source outshines the stellar light of the host galaxy (Kauffmann et al., 2003).

Observationally, AGN can vary in their appearance, in terms of the width of their emission lines, spectroscopic features, their variability, the relative position of high and low surface brightness (called the Fanaroff-Riley class), and other parameters. These many different, and sometimes overlapping, criteria have meant that the history of AGN classification has been rife with confusing classes and groups, such as Seyfert Galaxies, Broad Line Radio Galaxies, Optically Violent Variable sources, and more. It is now broadly accepted that what appear to be different types of sources can, in fact, be explained as an orientation-based unified model, which satisfactorily explains the variation in

luminosity among the different classes of AGN (Figure 1.2).

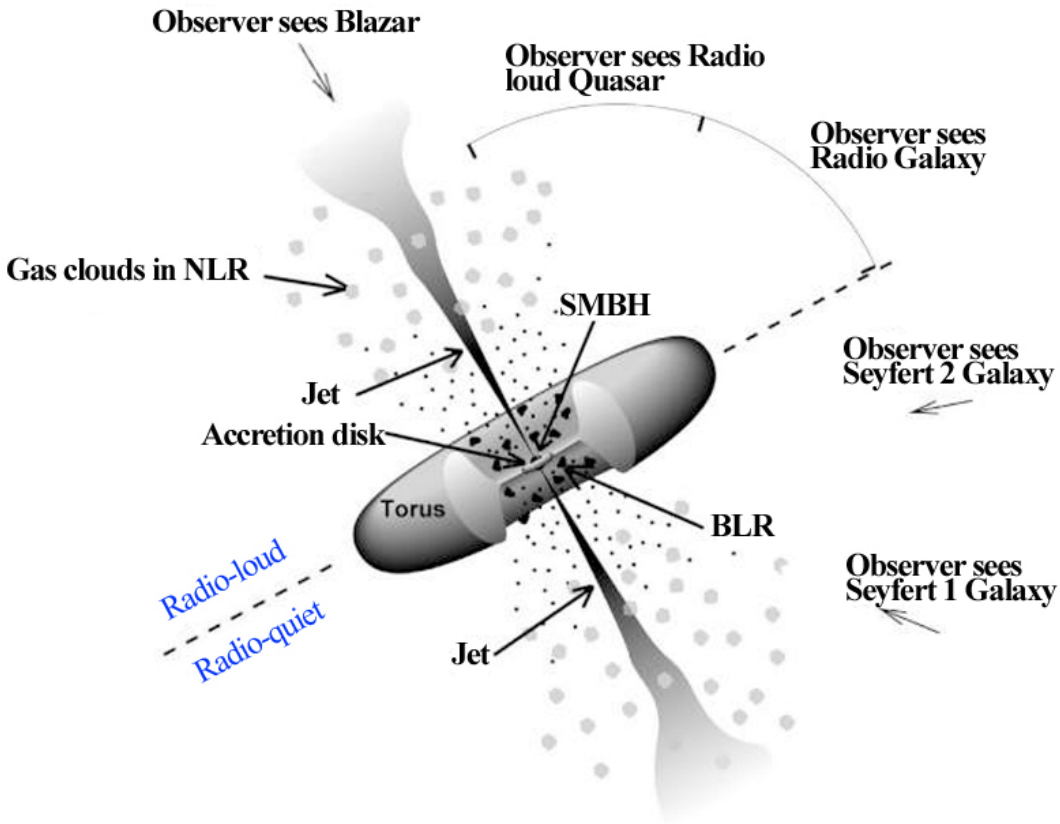

Figure 1.2: The orientation-based AGN unified model, showing the Broad Line Region (BLR) and Narrow Line Region (NLR), as well as different types based on the observational angle. Based on a schematic diagram by Urry and Padovani (1995).

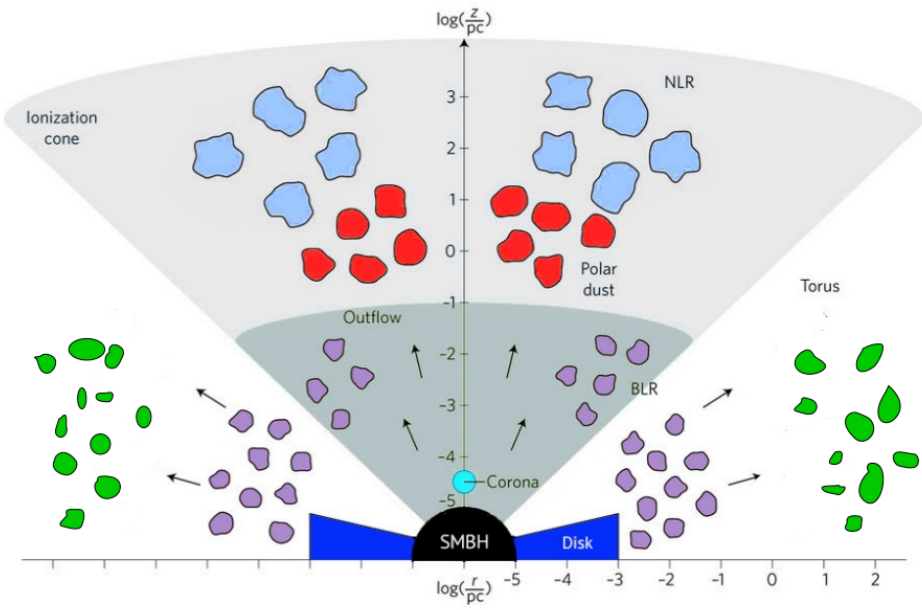

Figure 1.3: Schematic representation of the AGN physical model, illustrating the approximate scales of the main obscuring regions. The colours indicate the regions responsible for the main spectral components of an AGN. Green: torus (IR), dark blue: accretion disk (IR-optical), red: polar dust (IR), purple: BLR, light blue: NLR, dark blue: accretion disk. Arrows indicate outflow of mass and energy. Adapted from Ramos Almeida and Ricci (2017).

1.2 Motivation – Photometric redshifts

1.2.1 Redshift

Redshift is given by

$$z = \frac{\lambda' - \lambda_0}{\lambda_0} = \frac{\lambda'}{\lambda_0} - 1 \ , \tag{1.1}$$

where λ' is the observed wavelength of a line in the object's spectrum and λ_0 is the wavelength of the same spectral line observed in the laboratory frame. Furthermore, for two spectral lines with the same redshift, and with rest wavelengths λ_1 and λ_2, we have the relation

$$\frac{\lambda_1}{\lambda_2} = \frac{\lambda_1'}{\lambda_2'} \tag{1.2}$$

Thus, the amount of Doppler shift for two corresponding lines in a laboratory frame and in a redshifted frame, is not dependent on the difference in wavelengths of the two lines. The standard method of measuring redshift is by spectroscopy. Spectroscopy requires a bright source to produce an optical spectrum, while photometric observation can identify much fainter objects than can be found by spectroscopic observation at the same exposure time.

Continuum surveys of large parts of the sky are expected to add huge amounts of data to our already substantial database of faint sky objects, for which an estimate of the redshift will prove invaluable in extending the understanding of the sources and their distribution in space. The *Evolutionary Map of the Universe* (EMU, Norris et al. (2011) on the *Australian Square Kilometer Array Pathfinder* (ASKAP) will make a deep radio continuum survey of the entire Southern sky, cataloging around 70 million galaxies (Allison et al., 2016), including AGN with distributions extending to the edge of the visible Universe. Additionally, photometric techniques are useful because they expand the volume of "distance-luminosity" space where redshifts can be measured. Expanding this space is an important step in determining not only the type of templates used to estimate the photometric redshift, but also the best coverage for the templates (see Section 3.4). The ability to reliably determine the redshift of hundreds of objects by photometry alone will greatly increase the value of such surveys in determining the population and structure of the Universe.

The absorption lines in the spectrum of an AGN can be due to gas associated with the host galaxy, in the intergalactic medium (IGM), or within a galaxy lying along the line of sight to a

background continuum source. Even surveys where wide-band radio spectroscopy is already available, for example, the Boolardy Engineering Test Array (BETA) of the ASKAP (McConnell, 2016), often require follow-up observations in order to ascertain the source of the absorbing gas (Allison et al., 2015). This is also important in determining the populations of active and quiescent sources in the distant Universe, and will provide a valuable addition to machine learning methods. For very large surveys, such as the First Large Absoprtion Survey in HI (FLASH), which is expected to yield spectra for 150 000 radio sources, optical spectroscopy is not a viable option.

Furthermore, knowing the redshift is necessary in order to know the tuning frequency for radio telescopes looking for redshifted HI transition lines. However, since the amount of absorption is correlated with the redness of the source (Webster et al., 1995; Carilli et al., 1998; Curran et al., 2006) and this reddening is likely due to dust grains shielding the neutral gas from the local UV field, this produces a bias in the surveys against dust-rich systems. Sources which are sufficiently bright to be visible at large redshifts, also produce enough photons to ionise all of the neutral gas within the host galaxy (Curran et al., 2008; Curran and Whiting, 2012). Hence, even the SKA will not be able to detect star-forming regions in high redshift radio sources. Therefore, there is a need for some other method to estimate the redshift for fainter objects. De Breuck et al. (2002) found a correlation

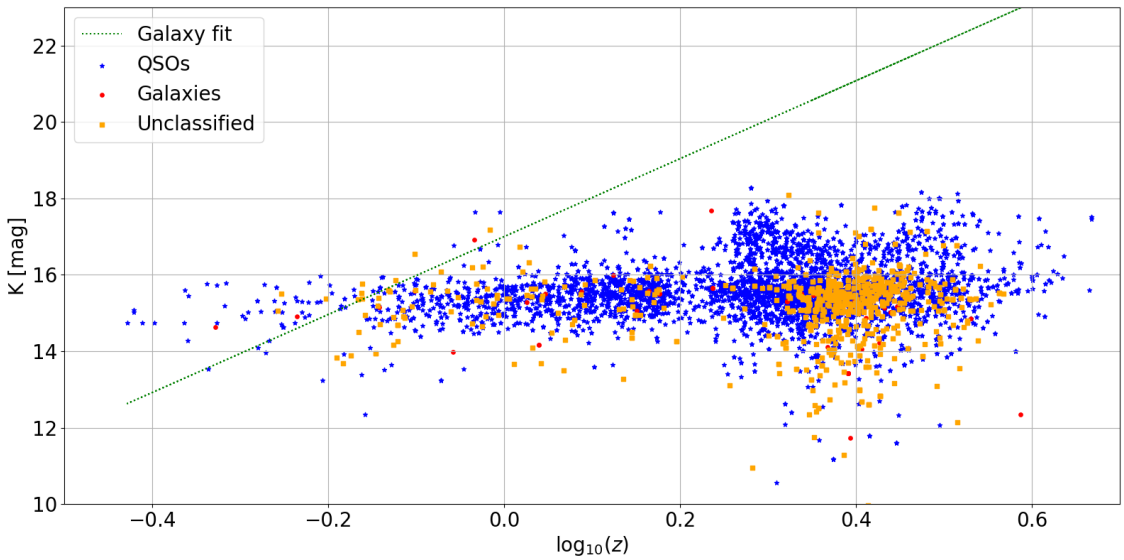

Figure 1.4: A plot of the relationship between the K-magnitude and redshift ($K - z$) for a large sample of QSOs. The broken line shows the fit for galaxies ($K = 4.43 \log_{10} z + 17$), De Breuck et al. (2002).

between the near-infrared K magnitude at $\lambda = 2.2$ μm and the redshift of galaxies. Even though

more distant objects will be fainter, the narrow spread of this correlation is remarkable, given that each source will have its own intrinsic luminosity. When applied to QSOs, however, this relationship is lost (Figure 1.4). It can be seen that the K-magnitude underestimates the redshift in the case of QSOs, and thus, at best, the K magnitude can provide only a lower limit to the redshift.

Since the $K-z$ relation for quasars is scattered to the left of the fit in Figure 1.4, we suspect there to be a range of AGN contributions. Because a quasar is a result of the direct line of sight observation of the central engine of an AGN (Netzer, 2013), even a partially obscured SMBH accretion disk would contribute a significant portion of the K-band magnitude from the quasar. However, because of its small angular diameter, this contribution would appear unresolved, and although the images were not of a high enough quality to directly detect an AGN contribution, this may contribute to the K-band emission, and thus been indirectly detected.

Glowacki et al. (2017) used the The Large Area Radio Galaxy Evolution Spectroscopic Survey (LARGESS) catalogue to investigate a similarly simple method, to find a correlation between the W1 ($\lambda = 3.4\mu$m) and W2 ($\lambda = 4.6\mu$m) Widefield Infrared Survey Explorer (WISE) magnitudes (Wright et al., 2010), including broad line sources such as quasars, and redshift. However, this correlation has a low regression coefficient of just $r = 0.56$ for the W1 band and $r = 0.36$ for W2. From our own photometry matching by NED name, we find $r = 0.77$ and 0.65, respectively (Figure 1.5), albeit with much lower numbers of sources in the sample. Thus, it is hoped that there is a combination of magnitudes sensitive to the AGN contribution, which can yield a tighter relationship with the redshift. Furthermore, we favour a simple linear fit, since a polynomial will introduce degeneracies into the predictions, with a photometric redshift matching more than one spectroscopic redshift.

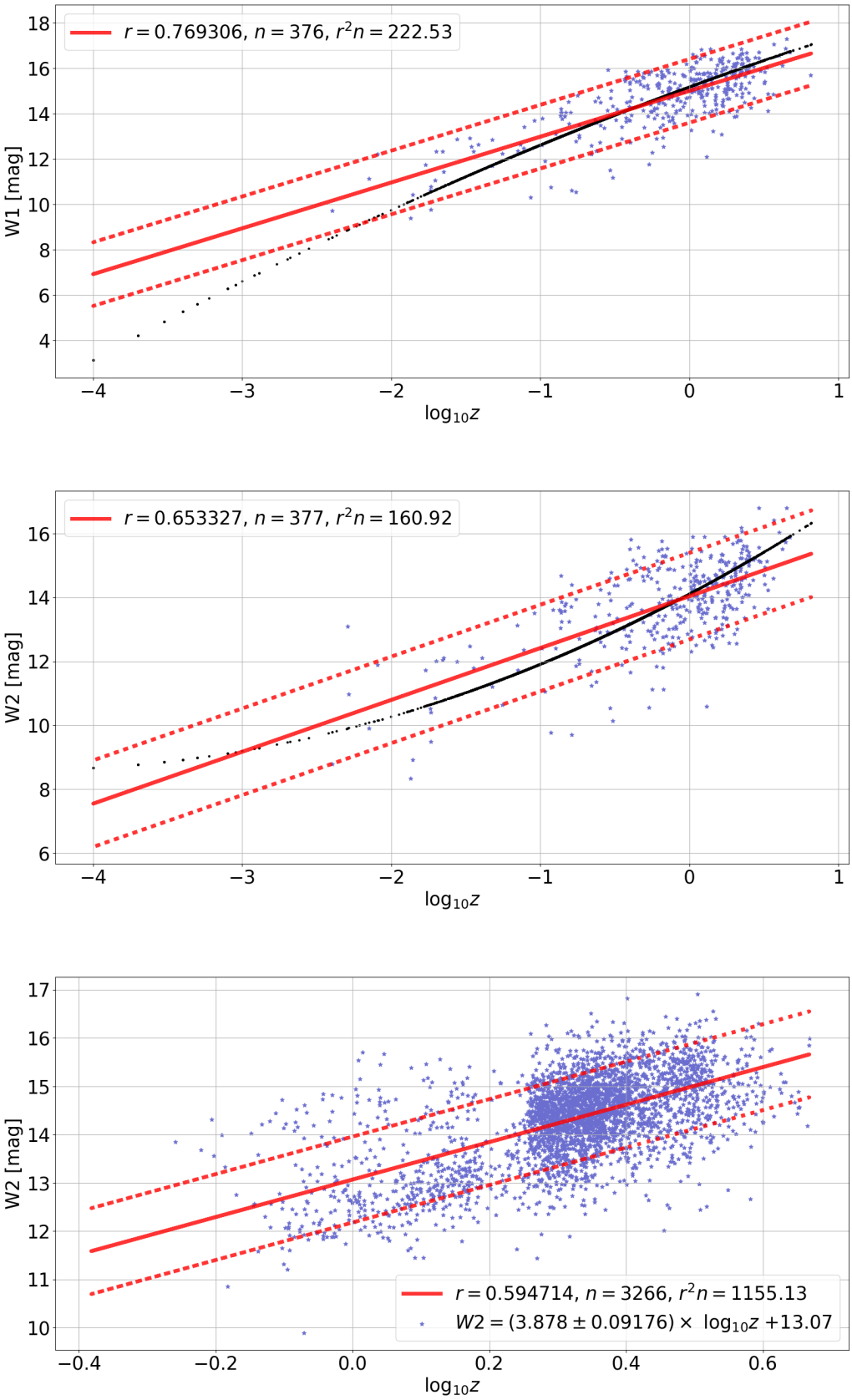

Figure 1.5: The Hubble W1 (top) and W2 (bottom) distributions from the LARGESS sample. The solid line shows the least-squares best fit with the regression coefficient r. The dotted curves show the fits from Glowacki et al. (2017) (their Table 2) to the QSOs ($-0.15\log_{10} z^2 + 2.41\log_{10} z + 15.19$ for W1 and $0.28\log_{10} z^2 + 2.48\log_{10} z + 14.12$ for W2). The bottom panel shows as above, but for our initial dataset (the MgII sample). The red lines show best fit $\pm 1\sigma$.

It is the aim of this investigation to find a tighter statistical correlation between the photometric magnitude, or a combination of magnitudes of quasars, and their redshift, without the need for a template fit or a full optical spectrum (Section 1.2.3).

1.2.2 Redshift distributions

The number of detected AGNs per unit area on the sky depends on the wavelength band and survey depth (Netzer, 2013). The redshift distribution of the Sloan Digital Sky Survey (SDSS) sample is shown in Figure 1.6. The sample is dominated by QSOs, particularly from $z \approx 1$, but there is a marked drop off in numbers with redshift, as is expected from the Malmquist bias.

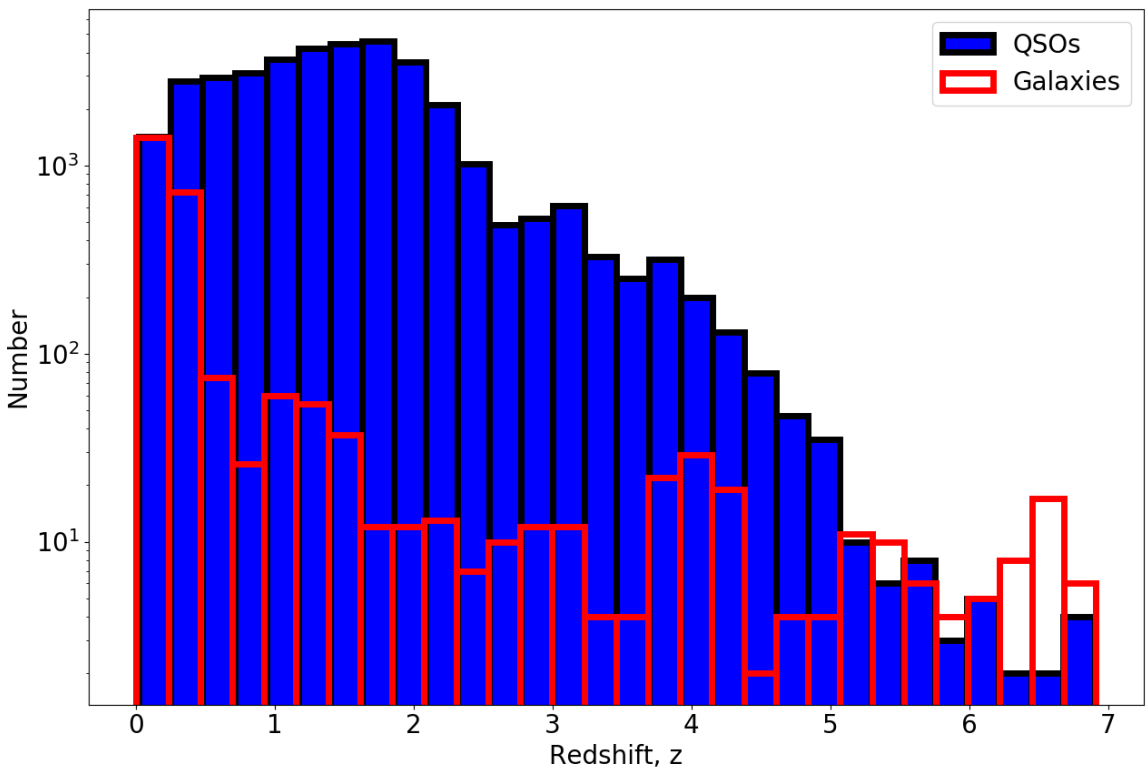

Figure 1.6: Distribution of QSOs and galaxies with redshift from the SDSS dataset.

Since AGNs have been detected all the way to $z \sim 7$, we can use them as tracers of the distribution of matter in the early Universe. AGNs are valuable tracers of the intergalactic medium for a number of reasons (Marziani et al., 2019):

- AGNs are plentiful, and easily recognizable by their spectra, luminosity and positions on a colour-colour diagram,

- having such a high luminosity means that they are visible across the Universe,
- they are observed over a wide range of redshifts, and
- unlike Type Ia supernovae and other transient phenomena that are employed as cosmological yardsticks, they are very stable.

A fundamental first step in this understanding, is determining the redshift of the AGN.

1.2.3 Spectral and photometric sensitivities

For historical reasons, brightness measurements are normally given in magnitudes (an inverse logarithmic measure of the flux). When collecting spectral data for a particular source or sample, it is necessary to take an exposure for long enough (have a long enough integration time) that the signal-to-noise ratio rises above a particular threshold, governed in part by the magnitude of the object. The integration time t is (Prasad, 2007)

$$t = \frac{\left[\sigma_{\rm rn}^2 \left(R_\star + n_{\rm pix} R_{\rm sky}\right] + \left[\left(\sigma_{\rm rn}^4 \left(R_\star + n_{\rm pix} R_{\rm sky}\right)^2 + 4R_\star^2 \sigma_{\rm rn}^2 n_{\rm pix} \rho^2\right]^{\frac{1}{2}}\right.}{2R_\star^2} \tag{1.3}$$

where $\sigma_{\rm rn}$ = signal/noise ratio, $R_\star$ = count rate from source * in e^- s^{-1}, $n_{\rm pix}$ = number of pixels in the aperture, $R_{\rm sky}$ = count rate from the sky in e^- s^{-1} $pixel^{-1}$, and ρ = read noise †. For bright sources, whose count rate is $> 3\times$ read noise, the noise from the source dominates the $\sigma_{\rm rn}$ contribution in Equation 1.3, while the signal to noise ratio is

$$\sigma_{\rm rn} \simeq \sqrt{R_\star t} \propto t^{\frac{1}{2}} \tag{1.4}$$

and in the sky-limited regime (in which the noise from the sky is greater than $3\times$ read noise):

$$\sigma_{\rm rn} \simeq \frac{R_\star t}{\sqrt{n_{\rm pix} R_{\rm sky} t}}, \tag{1.5}$$

*The number of photoelectrons emitted per second by the photomultiplier
†The static sources of noise that are independent of the signal

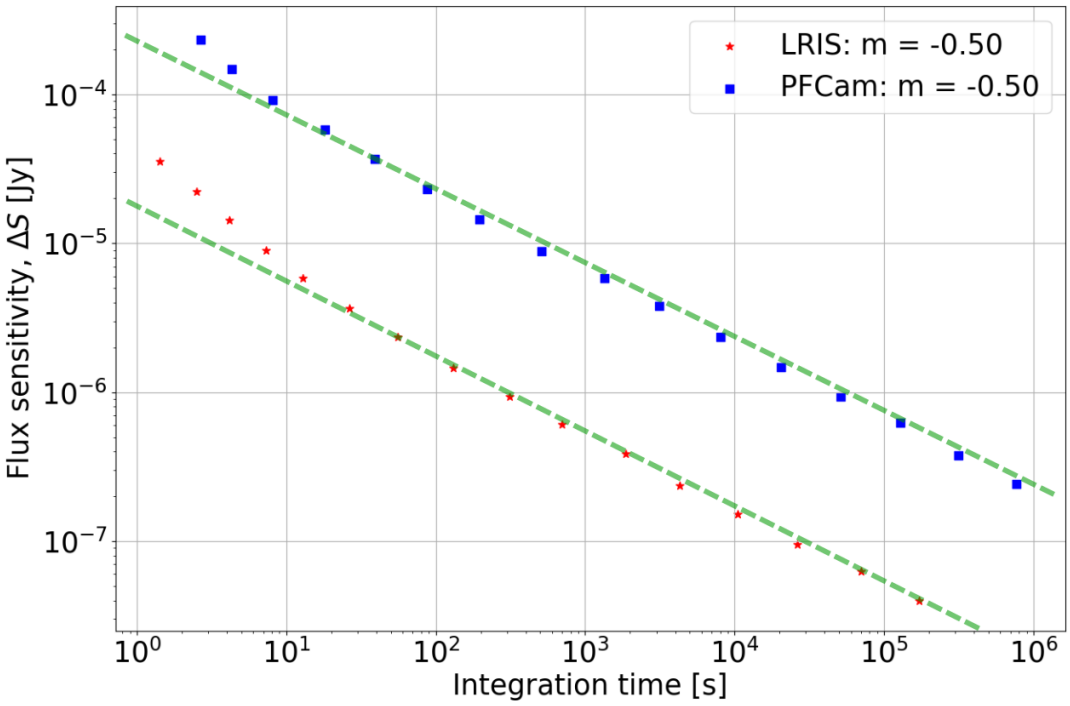

Figure 1.7: The V-band flux density versus integration time for the Low Resolution Imaging Spectrometer (LRIS) on the Keck Telescope (Oke et al., 1995) and the Prime Focus Camera (PFCam) on the Shane 3-meter telescope, located at Lick Observatory (Allen and Clarke, 2000).

which can be written in terms of the radiometer equation (Wilson et al., 2000) as

$$\Delta S = \frac{S_{\rm sys}}{\sqrt{\Delta\nu t}} \tag{1.6}$$

where ΔS is the sensitivity, $S_{\rm sys}$ is the system flux density, both in Janskys, and $\Delta\nu$ is the spectral resolution.

The system flux density is related to the system temperature $T_{\rm sys}$, in Kelvin, by the objective area and efficiency of the telescope, so that

$$\sigma_{\rm rn} = \frac{T_{\rm sys}}{T_\star\sqrt{\Delta\nu t}} \tag{1.7}$$

where $T_\star$ is the power received from the source, and $T_{\rm sys} = T_{\rm sky} + T_{\rm rx}$ describes the power received due to both the sky $T_{\rm sky}$, and the thermal noise from the receiver electronics $T_{\rm rx}$.

From Equation 1.6, the sensitivity of a particular observation depends on the spectral resolution $\Delta\nu$. Thus, since spectral resolution for obtaining an optical spectrum is necessarily much finer than

that for photometry, $\Delta S_{\rm spec} \gg \Delta S_{\rm photo}$. Emission lines in the optical bands of AGN spectra can have widths of $\sim 10^3$ km s^{-1} (Netzer, 2013), although the redshift accuracy will be determined by the spectral resolution

$$\Delta z = (z+1)\frac{\Delta v}{c} \tag{1.8}$$

where Δv is the spectral resolution in km s^{-1} and c is the speed of light. Therefore, for instance, a spectral resolution of $\Delta v = 100$ km s^{-1} at $z = 2$ gives $\Delta z = 0.001$. Although the integration time required for a spectrum depends not only on the required signal-to-noise ratio (such as $\Delta S \gtrsim 10\sigma_{\rm rms}$, where $\sigma_{\rm rms}$ is the r.m.s noise) but also the strength of the emission line, we can compare the integration time of a photometric observation with that of a spectral observation by considering the effect of a spectrometer on the instruments in Figure 1.7. For example, the best fit to the LRIS data is the power law $\log_{10}\Delta S = -0.565\log_{10} t - 4.54$. However, this is skewed by a region of non-linearity at $t \lesssim 10$ s, possibly due to the saturation of pixels at high flux, and so forcing the expected slope of -0.50, we obtain

$$\log_{10}\Delta S = -0.50\log_{10} t - 4.71 \Rightarrow \Delta S = \frac{10^{-4.71}}{\sqrt{t}} \tag{1.9}$$

If the photometric measurement is taken over the full V band with a wavelength of 551 ± 44 nm, then $\Delta\nu = 8.75 \times 10^{13}$ Hz. Hence, from Equation 1.6,

$$10^{-4.71} = \frac{S_{\rm sys}}{\sqrt{8.75\times10^{13}}} \Rightarrow S_{\rm sys} = 10^{-4.71}\sqrt{8.75\times10^{13}} = 182.4 \Rightarrow \Delta S = \frac{182.4}{\sqrt{\Delta\nu t}} \tag{1.10}$$

for the Low Resolution Imaging Spectrometer (LRIS) on the Keck Telescope.

Figure 1.8 shows the required integration times required to reach a given sensitivity at different spectral resolutions, using, in the non-relativistic regime,

$$\Delta v = c\frac{\Delta\nu}{\nu_{\rm obs}} \tag{1.11}$$

where, for the V band, $\nu_{\rm obs} = 5.44\times10^{14}$ Hz (551 nm). From this, it can be seen that for $\Delta v = 1000$ km s^{-1}, ($\Delta z = 0.008$ at $z = 1.4$), which is barely sufficient to resolve the MgII emission line in Figure 1.9, requires ~ 50 times the integration time to reach the same sensitivity as the photometric observation.

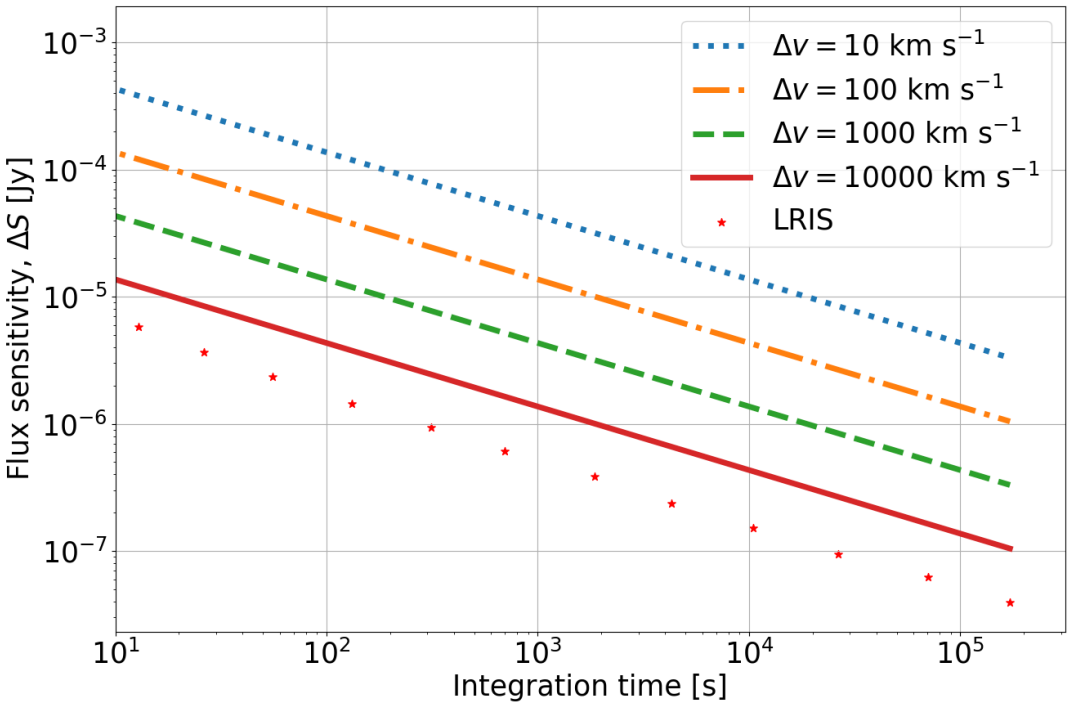

Figure 1.8: The sensitivity per spectral channel for the LRIS at four different spectral resolutions (from Equation 1.6, ranging from $\Delta v = 10$ km s^{-1} ($\Delta\nu = 1.8 \times 10^{10}$ Hz) to $\Delta v = 10\,000$ km s^{-1} ($\Delta\nu = 1.8 \times 10^{13}$ Hz). Stars show one channel over the full V-band for the LRIS ($\Delta\nu = 8.75 \times 10^{13}$ Hz).

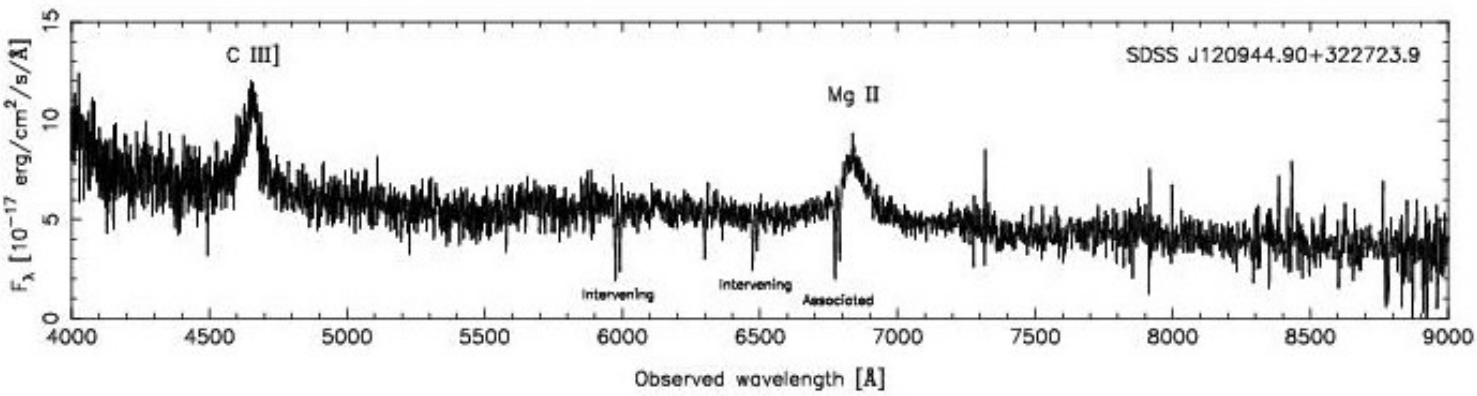

Figure 1.9: Example of an SDSS spectrum exhibiting both associated and intervening MgII absorption lines (labelled). The broad MgII and the semi-forbidden CIII] emission lines are also indicated. The full-width half maximum (FWHM) of the MgII line is $\Delta\lambda \sim 100$ Å, giving $\Delta \approx 4000$ km s^{-1} at $\lambda \approx 6830$ Å. This gives the QSO a redshift of $z = 1.443$. Curran et al (in prep).

1.2.4 Spectral lines

Another issue is the spectral line being redshifted out of the observing band. From Figure 1.8, we see that the SDSS band spans 4000 - 9000 Å, although the spectrum is significantly noisier within the edge $\approx$ 1000Å. In addition to the CIII] and MgII lines, there are several other QSO emission lines detected. The redshift ranges over which these lines can be detected within $5000 - 8000$ Å is shown in Figure 1.10.

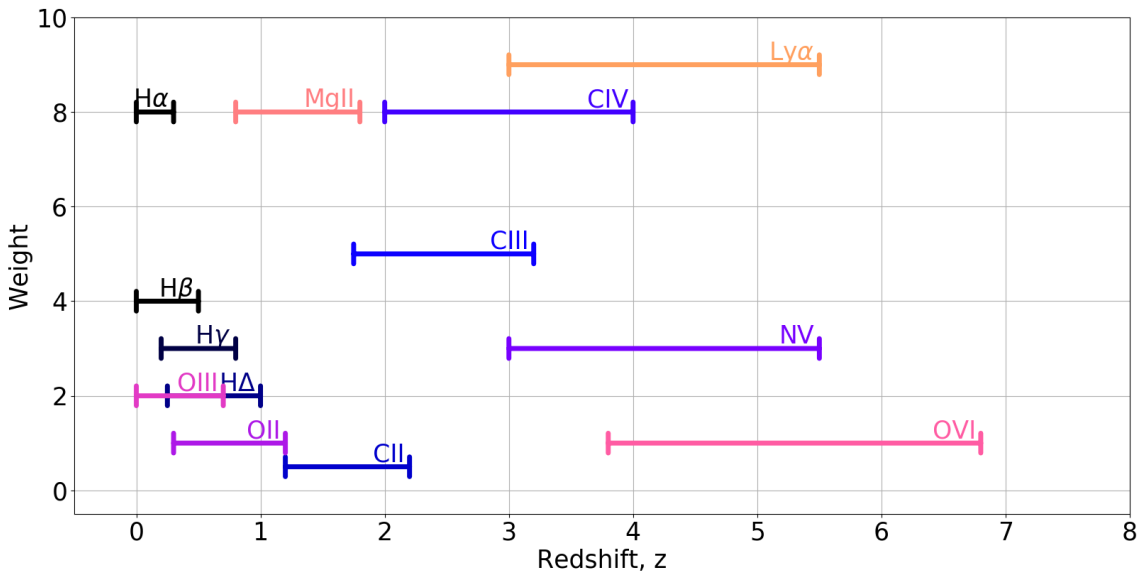

Figure 1.10: The redshift ranges of the QSO emission lines used by the SDSS over $5000 - 8000$ Å. Weight is a measure of the relative strength of the emission line in a spectrum (Vanden Berk et al., 2001).

From this, of the highly weighted lines in terms of line strength, we see that the highest redshift probed is $z = 5.583$ (Lyα, at $\lambda = 1215.24$ Å). One of the strongest lines in the emission spectrum is that of MgII, which is limited to redshifts of $z = 0.786$ to $z = 1.858$ (or $z = 0.429$ to $z = 2.215$ for $4000 - 9000$ Å). Figure 1.6 shows a steep drop-off in QSO numbers at $z \approx 3.5$. The only highly-weighted lines for the distribution are the Lyα and CIV emission lines, hence the sensitivity of the survey is decreased at higher redshifts. In addition to this, higher magnitudes (dimmer sources) require higher integration times to reach a given flux sensitivity, also contributing to the marked decrease at higher redshift (Figure 1.6).

2 Initial testing and analysis

Preliminary testing used the Mg II catalogue (Zhu and Ménard, 2013) in the Sloan Digital Sky Survey (SSDS) Data Release 12 (DR12), containing Mg II absorbing sources (mostly QSOs) up to $z = 5.5$ ($\log_{10} z = 0.74$ in Figure 1.4). Added to this were the WISE $W1$, $W2$, $W3$ and $W4$ bands. Values of wavelength λ and F_0 (the flux, in Janskys, at $m = 0$, where 1 Jy $= 10^{-26}$ W Hz^{-1} m^{-2}) for each of the bands used are given in Table 2.1.

Band	Wavelength λ [μm]	F_0 [Jy]
FUV	0.152	3630
NUV	0.227	3630
U	0.365	1810
B	0.445	4260
V	0.551	3640
R	0.658	3080
I	0.806	2550
J	1.220	1600
H	1.630	1080
K	2.190	670
$W1$	3.4	309.54
$W2$	4.6	171.787
$SPIT8$	8	64.13
$W3$	12	37.674
$W4$	22	8.363

Table 2.1: Observed wavelengths and fluxes for the wavelength bands used. Wavelengths given are the effective wavelength midpoint for standard filters. F_0 is the flux at a magnitude of 0. Values of F for bands U to K are from Harvard-Smithsonian Center for Astrophysics; for the *Wide-field Infrared Survey* (WISE) bands from Explanatory Supplement to the WISE All-Sky Data Release Products; and for the SPITZER 8μm band ($SPIT8$) from SPITZER, 2MASS, optical filtersets/bandpasses.

Figure 2.1 shows some example spectral energy distributions (SEDs) in terms of the photometric magnitudes for some typical samples from the dataset.

For each source, we matched the coordinates to the closest source within a 6 arcsecond search radius in the NASA/IPAC Extragalactic Database (NED), from which we obtained the specific flux densities. The NED names were also used to query the WISE, Spitzer Space Telescope (Capak et al., 2013), the Two Micron All-Sky Survey (2MASS) (Skrutskie et al., 2006) and the Galaxy Evolution Explorer (GALEX data release GR6/7)[‡] databases. In order to obtain a uniform measure of magnitude, the measurements were added if they fell within $\Delta \log_{10} \nu = \pm 0.05$ of the central frequency (Figure 2.1). For more than one point in the band, the fluxes were averaged before being

[‡]http://galex.stsci.edu/GR6/#mission

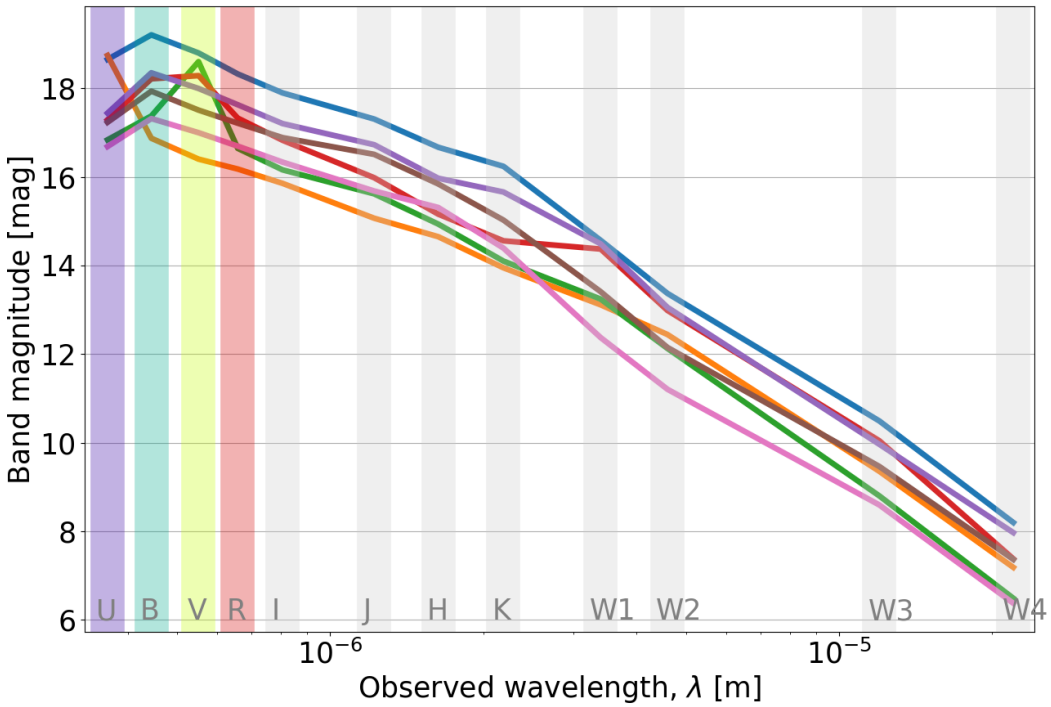

Figure 2.1: Examples of the Spectral Energy Distribution, in terms of magnitudes, for several of the sources in the MgII dataset. Canonical band frequencies are shown as vertical bands, with a thickness of $\Delta \log_{10} \nu = \pm 0.05$.

converted to magnitude. The apparent magnitude is related to the flux density, F, via

$$m = -2.5 \log_{10} \left(\frac{F}{F_0} \right) \ , \tag{2.1}$$

where F is the observed flux, in Janskys, at the frequency of the wavelength band, and F_0 is the flux at $m = 0$ (Table 2.1).

2.1 Single magnitudes

As a preliminary test, each magnitude was tested against z to find the values of r. Table 2.2 summarises the results.

Magnitude	$\lvert r \rvert > 0.5$	n
U	0.647813	5998
$W2$	0.594714	3266
NUV	0.591015	12 420

Table 2.2: Best fits for single magnitudes versus $\log_{10} z$, for which $|r| > 0.5$.

While these all have high correlations with $\log_{10} z$, their spreads are too wide to provide a good estimate of photometric redshift (Figure 2.2).

2.2 Single magnitude combinations

A colour index is a difference in two magnitudes ($U - B$, $U - V$, ...), each at a different wavelength (Bessell, 2005). As such, it is effectively the ratio between two fluxes, giving a distance-independent measurement of a fundamental quantity of the source. Comparison of these magnitudes can allow us to, for instance, separate populations of stars and galaxies based on the relative flux at different wavelengths (Strait, 2015). This ratio does not change with redshift, even taking into account the evolution of rest-frame wavelength with redshift. Therefore, since the photometric colours provide a distance-independent measure of a fundamental property of the quasars (see Section 1.2.1), their relationships to $\log_{10} z$ were investigated first.

Once the colours were computed, their Ordinary Least Squares regression (linear best fit – see Section 6) against $\log_{10} z$ was calculated. At each stage of the computation, those combinations with

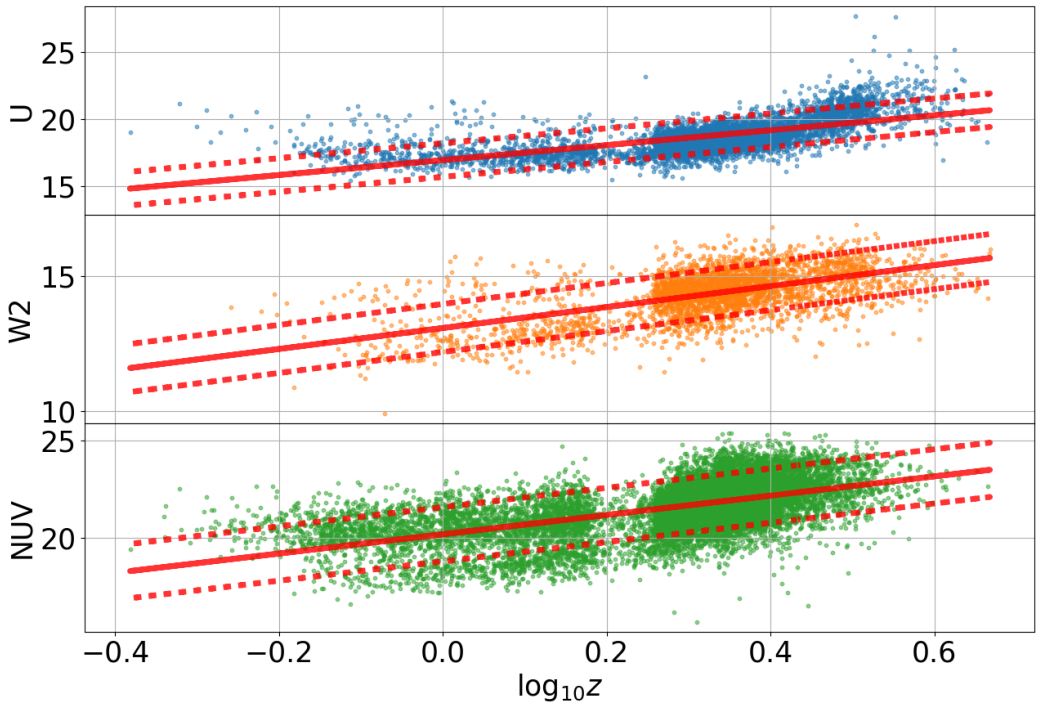

Figure 2.2: The distributions for the three best-fitting magnitudes versus redshift for the MgII dataset over all redshift ranges. Top: $r = 0.647813, r^2n = 2517.13$. Middle: $r = 0.594714, r^2n = 1155.13$. Bottom: $r = 0.591015, r^2n = 4338.30$. Red lines show the best fit $\pm 1\sigma$.

the highest values of r were identified and plotted explicitly.

Initial testing using the method employed here yielded, for example, 182 single magnitude subtractive combinations (*i.e.* photometric colours), of which the one with the highest r value is $(NUV - V)$ ($r = 0.702604$, $r = 52$). However, since this combination has only 52 corresponding points, it would be a poor predictor of redshift for large datasets. Thus, since a very low number of observations can give a high r value, the value of r^2n was also calculated for each combination, as a way of taking into account both the r-value and the number of observations, both of which varied for each combination. Visual inspection of the combinations shows the optimum value of r^2n to be ~ 1000, and for this reason, only combinations with values of $r^2n > 1000$ will be considered, as a way of selecting both high r and high n combinations. Applying the constraint of $|r| > 0.5$ and $r^2n > 1000$ reduces the yield to 27 subtractive combinations, of which the one with the highest r value is $(NUV - R)$ ($r = 0.676618$, $r = 3851$, $r^2n = 1763.04$). However, Figure 2.3 shows that the residuals for this combination, when plotted against redshift, are not homogeneously spread throughout the redshift range, indicating that this combination would have a high number of outliers, and therefore would be a limited usefulness as a predictor of photometric redshift.

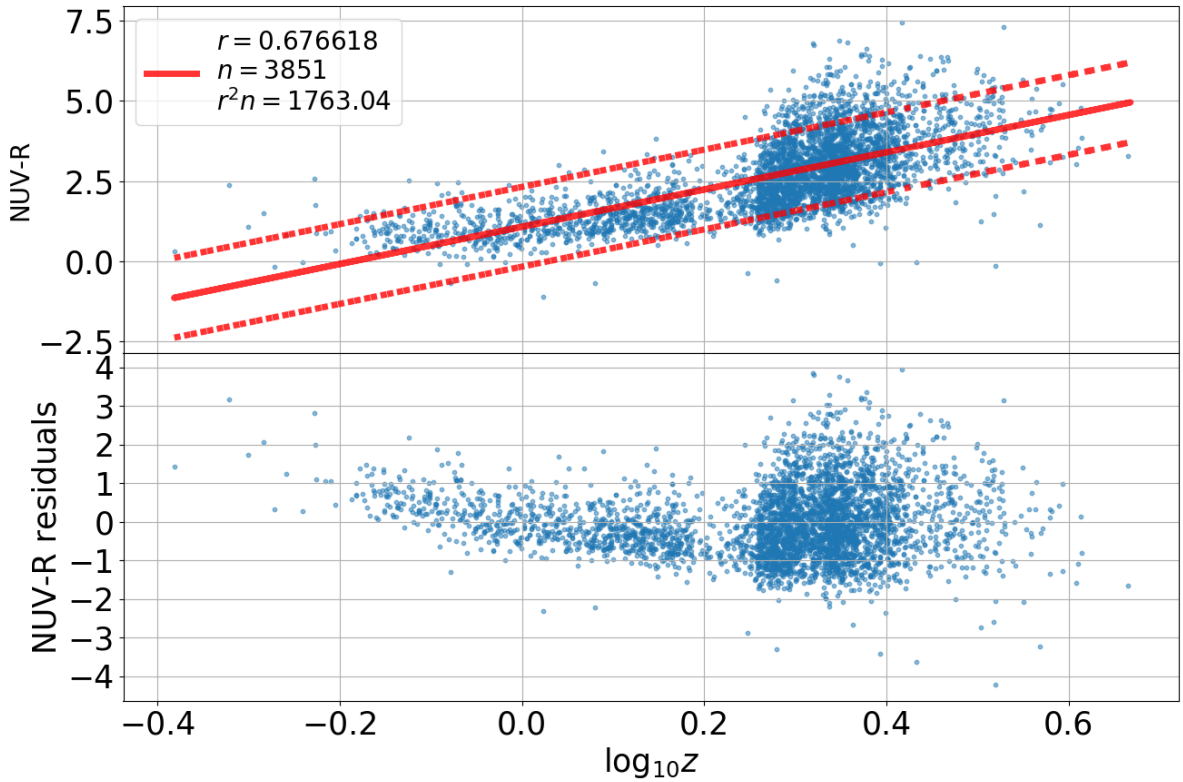

Figure 2.3: The $NUV - R$ versus redshift for the MgII dataset over all redshift ranges. The red lines show best fit $\pm 1\sigma$.

2.3 Double magnitude combinations

The rationale behind finding double combinations, is that including more colour indices would approach the same amount of information as a low dispersion spectrum, thus giving more information than pairs of magnitudes. A colour-colour diagram, for example, is a means of comparing the apparent magnitudes of sources at different wavelengths, using one or a pair of colour indices. Since a colour-colour diagram plots two photometric colours against each other, each colour was divided into the others, giving the combinations $(U - B)/(U - V)$, $(U - B)/(U - R)$, $(U - B)/(U - K)$ ... Using this as a starting point, the subtractive combinations already found were operated upon, giving the following combinations:

$$
\begin{array}{cccc}
(U-B)/(U-V) & (U-B)\times(U-V) & (U-B)+(U-V) & (U-B)-(U-V) \\
(U-B)/(U-R) & (U-B)\times(U-R) & (U-B)+(U-R) & (U-B)-(U-R) \\
(U-B)/(U-K) & (U-B)\times(U-K) & (U-B)+(U-K) & (U-B)-(U-K) \\
\vdots & \vdots & \vdots & \vdots
\end{array}
$$

The other arithmetic operations were also explored:

$$
\begin{array}{cccc}
(U+B)/(U+V) & (U+B)\times(U+V) & (U+B)+(U+V) & (U+B)-(U+V) \\
(U+B)/(U+R) & (U+B)\times(U+R) & (U+B)+(U+R) & (U+B)-(U+R) \\
(U+B)/(U+K) & (U+B)\times(U+K) & (U+B)+(U+K) & (U+B)-(U+K) \\
\vdots & \vdots & \vdots & \vdots \\
(U/B)/(U/V) & (U/B)\times(U/V) & (U/B)+(U/V) & (U/B)-(U/V) \\
(U/B)/(U/R) & (U/B)\times(U/R) & (U/B)+(U/R) & (U/B)-(U/R) \\
(U/B)/(U/K) & (U/B)\times(U/K) & (U/B)+(U/K) & (U/B)-(U/K) \\
\vdots & \vdots & \vdots & \vdots \\
(U\times B)/(U\times V) & (U\times B)\times(U\times V) & (U\times B)+(U\times V) & (U\times B)-(U\times V) \\
(U\times B)/(U\times R) & (U\times B)\times(U\times R) & (U\times B)+(U\times R) & (U\times B)-(U\times R) \\
(U\times B)/(U\times K) & (U\times B)\times(U\times K) & (U\times B)+(U\times K) & (U\times B)-(U\times K) \\
\vdots & \vdots & \vdots & \vdots
\end{array}
$$

The first, and simplest combinations explored in this investigation were subtractive combinations of the wavelength bands, such as $U-B$, $U-V$,..., as given in Table 3.2. As an example, the $B-V$ colour index is:

$$
\begin{aligned}
B-V &= -2.5\log_{10} f_B + 2.5\log_{10} f_V \\
&= -2.5\log_{10}\left(\frac{f_B}{f_V}\right)
\end{aligned}
\tag{2.2}
$$

while

$$
\begin{aligned}
B+V &= -2.5\log_{10} f_B - 2.5\log_{10} f_V \\
&= -2.5\log_{10}\left(f_B\cdot f_V\right)
\end{aligned}
\tag{2.3}
$$

However, taking, for example, the product of two magnitudes would be:

$$
\begin{aligned}
B\times V &= (-2.5\log_{10} f_B)(-2.5\log_{10} f_V) \\
&= 6.25\log_{10} f_B^{-2.5\log_{10} f_V}
\end{aligned}
\tag{2.4}
$$

while

$$
\begin{aligned}
B/V &= \left(\frac{-2.5\log_{10} f_B}{-2.5\log_{10} f_V}\right) \\
&= \log_{f_V} f_B
\end{aligned}
\tag{2.5}
$$

Like B and V individually, $B \times V$ and B/V would be distance-dependent, and therefore have no relationship to any intrinsic property to the source. Furthermore, a combination like B/V would be ill-behaved near $V = 0$, and so the algorithm generated for this investigation deliberately filters out these invalid values. Beyond the photometric colours, since it is not known *a priori* what relationships may yield corrections to the $K - z$ relationship, all possible combinations of wavelength bands from the MgII dataset (*e.g.* $U + B$, $U + V$, ..., $U \times B$, $U \times V$, ..., U/B, U/V, ...) were calculated in a brute-force manner in order to provide some insight on how best to proceed. Many of these combinations produced r values in excess of $r \sim 0.7$, so only those with $r > 0.7$ and $r^2 n > 1000$ are considered.

Most of the magnitude combinations have complex physical interpretations (for example, the multiplication of magnitudes), and may not have any physical significance in terms of the individual quasars. However, the combinations explored in this study show high values of regression coefficient when plotted against $\log_{10} z$, and so, from a purely statistical standpoint, may be useful in estimating a redshift based on photometric magnitude, despite not containing any information about the individual sources themselves.

3 Results

3.1 Single magnitude combinations

A total of 84 magnitude combinations with $r > 0.5, r^2n > 1000$ were found. Those with the highest values are given in Table 3.1. To further test whether each combination would be a good predictor of photometric redshift, other statistical tests were employed, as outlined in Section 6. These other goodness-of-fit test statistics are also shown in the tables. Skew, optimally ~ 0, indicates a leftward (negative) or rightward (positive) tail in the distribution, thus giving an indication of outliers. Kurtosis, optimally $|k| \sim 3$, (Moors, 1986) is a measure of the sharpness of the peak in the probability distribution, and, similarly to skew, gives an indication of how much any outliers will affect the distribution (Section 6).

Combination	$\|r\| > 0.5$	n	r^2n	Skew	Kurtosis
$U + W2$	0.733229	2054	1104.28	−0.151	3.135
$U \times W2$	0.725858	2054	1082.19	−0.172	2.968
$I/W2$	−0.707549	2086	1044.31	0.815	3.870
$W2/I$	0.698607	2086	1018.08	−0.650	3.594
$U - R$	0.596446	5971	2124.17	1.966	12.472

Table 3.1: Best fit models of the single magnitude combinations, for which $|r| > 0.74$ and $r^2n > 1000$. $U - R$ is recognizable as a photometric colour. It is compared with the results for the next two best-fit colours in Table 3.2.

Table 3.1 shows the single magnitude combinations with the highest values of r and r^2n. While there are some high-r combinations, only $U - R$ is recognisable as a photometric colour. This is compared with the next two highest-scoring colours in Table 3.2.

Even given the high values of r and r^2n, these other test statistics indicate that, as a predictor of z, it may produce too many outliers to be useful. Figure 3.1 plots the photometric colour $U - R$, the best of these models, against $\log_{10} z$, along with the equation of the linear fit, the r value and r^2n.

Colour	$\|r\| > 0.5$	n	$r^2n > 1000$	Skew	Kurtosis
U-R	0.596446	5971	2124.17	1.966	12.472
U-I	0.580567	5605	1889.21	2.074	11.952
U-B	0.567813	5997	1933.5	1.737	16.913

Table 3.2: The photometric colours with the highest values of r and r^2n when plotted against $\log_{10} z$. $U - I$, having a high positive skew, indicates that it would produce outliers at high redshift. Each combination has high values of kurtosis, indicating the distribution has a sharp peak, and therefore is not normally distributed (Section 6).

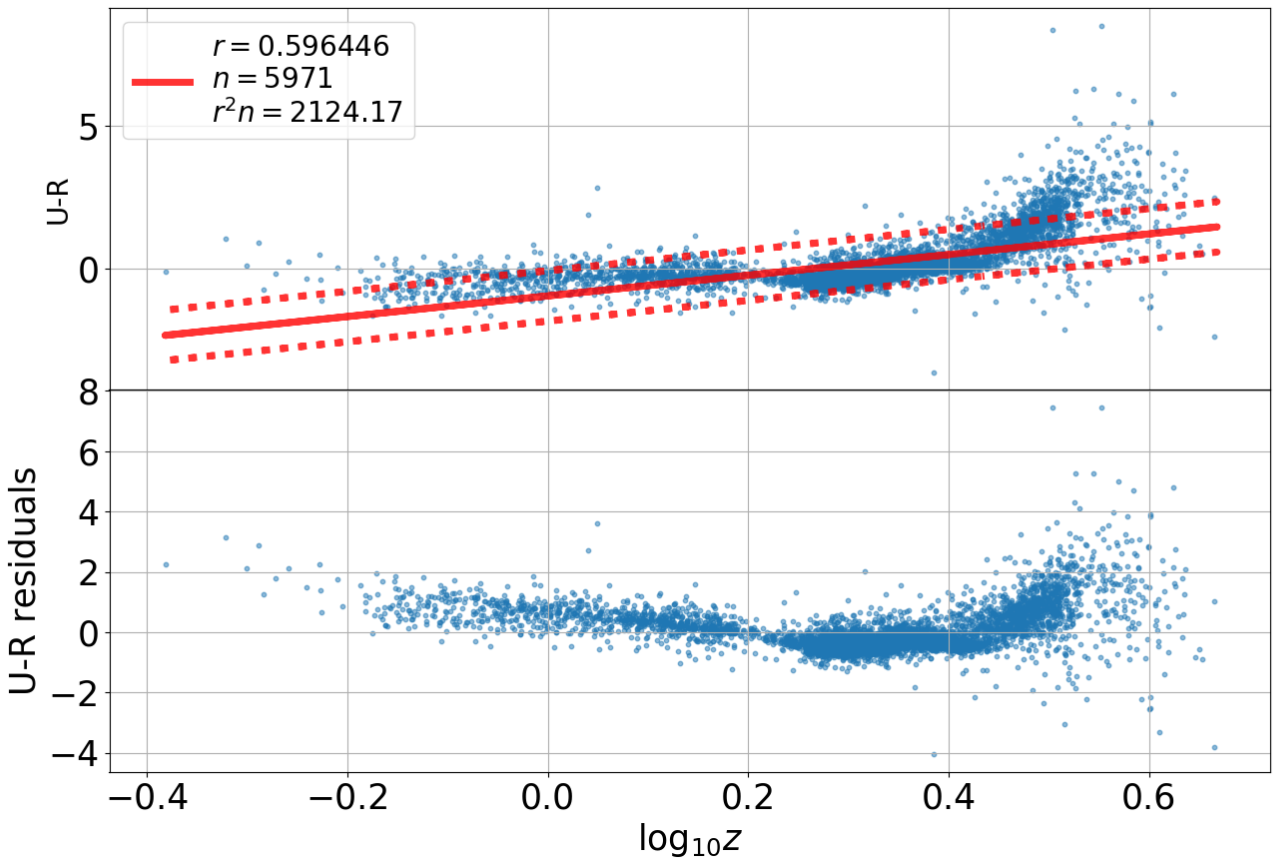

Figure 3.1: Top: Plot of $U - R$ colour against $\log_{10} z$. Among the photometric colours, this has the highest value of r. Bottom: Residuals plot for $U - R$ when plotted against $\log_{10} z$ as a test for the goodness of fit. The red lines show best fit $\pm 1\sigma$.

The bottom plot in Figure 3.1 shows the residuals for $U - R$; that is, the distance in colour-space of each point from the line of best fit. The fact that the residuals for $U - R$ are less spread at low z than at high z, indicates that a power law may not be the best fit for this set. It was noted that photometric colours involving the higher frequencies (*e.g.* $U - NUV$) tended to produce distributions

that were both less correlated, and had more heterogeneous spread, than colours involving lower frequencies (*e.g.* $I - W2$). This may be due to the fact that the spectral emission of quasars is dominated by thermal dust emission from the near-infrared through to submillimeter wavelengths (Sanders, 1999). For example, Figure 3.2 compares the plots of the comparatively short ultraviolet wavelengths $U - NUV$ (producing a poor fit, with much higher spread at high z), and the longer infrared wavelengths $I - W2$.

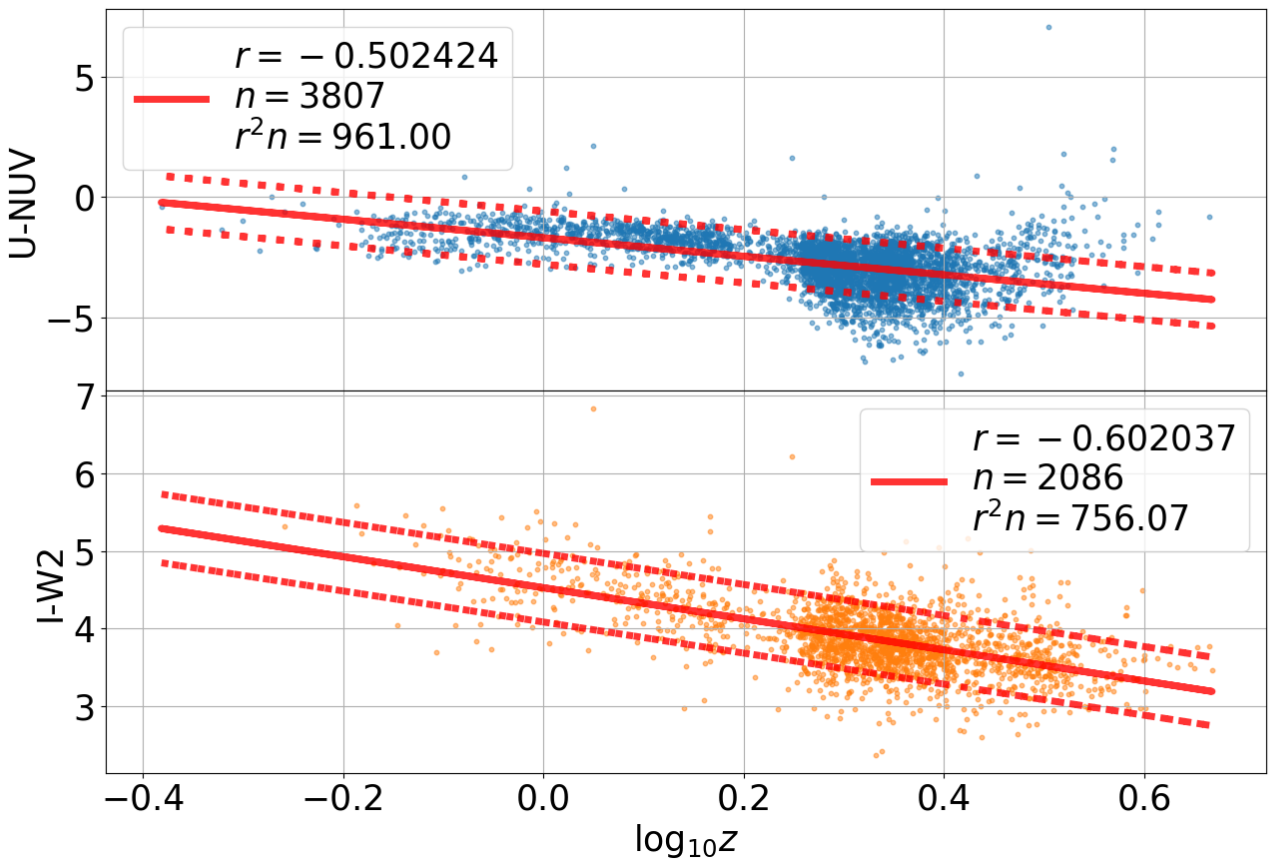

Figure 3.2: Comparing the distributions of a short wavelength combination $U - NUV$ to that of a long wavelength combination $I - W2$. The red lines show best fit $\pm 1\sigma$.

3.2 Double magnitude combinations

When the photometric colours were divided into each other, the correlation coefficient increased by around 50% compared to using just a single colour (*e.g.* $r \sim 0.5$ *c.f.* $r \sim 0.75$), with a smaller drop in n for a particular combination of colours. For example, the colour with the highest values of r and r^2n is $U - R$ ($r = 0.596446$, $r^2n = 2124.17$), while the best-fitting colour combination is

$(I - W2)/(W3 - U)$ ($r = 0.828545$, $r^2n = 1328.35$). These changes were balanced out by only a slight drop in the value of r^2n. Similar results were obtained when applying other operations to the photometric magnitudes. A total of 10 208 colour combinations with $r > 0.5, r^2n > 1000$ were found. Those with the highest values are given in Table 3.3.

Combination	$\|r\| > 0.5$	n	r^2n	Skew	Kurtosis
$(I/U) + (I/W2)$	-0.840933	2039	1441.92	0.004	3.917
$(I/U)/(W2/I)$	-0.838606	2039	1433.95	0.049	4.179
$(B \times I)/(U \times W2)$	-0.832803	2039	1414.17	0.403	4.587
$(U/I) - (I/W2)$	0.830791	2039	1407.35	0.379	5.048
$(I - W2)/(W3 - U)$	0.828545	1935	1328.35	-0.374	3.560
$(B + I)/(U + W2)$	-0.818597	2039	1366.33	0.191	4.834
$(U \times W2) - (I \times W3)$	0.753815	1935	1099.54	0.457	4.281
$(U - I) + (W2 - FUV)$	0.758322	1997	1148.38	0.135	5.082

Table 3.3: Best fit models of the double magnitude combinations, for comparison against each other. Most of these imply a complex physical relationship between the magnitudes, and of these, only $(I - W2)/(W3 - U)$, with $r = 0.828545$ is recognisable as a ratio of photometric colours.

Despite the fact that the values of r, n and r^2n are not quite as high for $(I - W2)/(W3 - U)$ as they are for some of the other combinations, such as $(I/U) + (I/W2)$ ($r = -0.840933$), the latter implies a complex physical relationship, which is difficult to interpret and justify in terms of what quantity the combination represents. Hence, $(I - W2)/(W3 - U)$ is likely to be the best combination in this set, being the ratio of two photometric colours. Therefore, this is the colour combination that the remainder of this report will focus on. It is compared with other colour combinations in Table 3.4. It is not clear why the $(I - W2)/(W3 - U)$ colour ratio should trace the redshift with such high correlation, but it should be borne in mind that $z \gg 0$, these would have been emitted at significantly shorter wavelengths (Figure 3.8).

Combination	$\|r\| > 0.5$	n	r^2n	Skew	Kurtosis
$(I-W2)/(W3-U)$	0.828545	1935	1328.35	−0.374	3.560
$(I-W2)/(W4-U)$	0.810052	2018	1324.18	−0.107	3.032
$(R-W2)/(W3-U)$	0.798750	1942	1239.00	−0.721	4.159
$(B-W2)/(W3-U)$	0.766935	1944	1143.44	−0.889	4.595
$(R-W2)/(W4-U)$	0.765999	2025	1188.18	−0.546	3.482
$(U-W3)/(I-W2)$	0.752859	1935	1096.75	0.609	4.352
$(U-W4)/(I-W2)$	0.745342	2018	1121.07	0.674	3.948
$(U-W3)/(R-W2)$	0.741070	1942	1066.52	0.261	4.288

Table 3.4: Best fit models of the colour combinations, for comparison against each other.

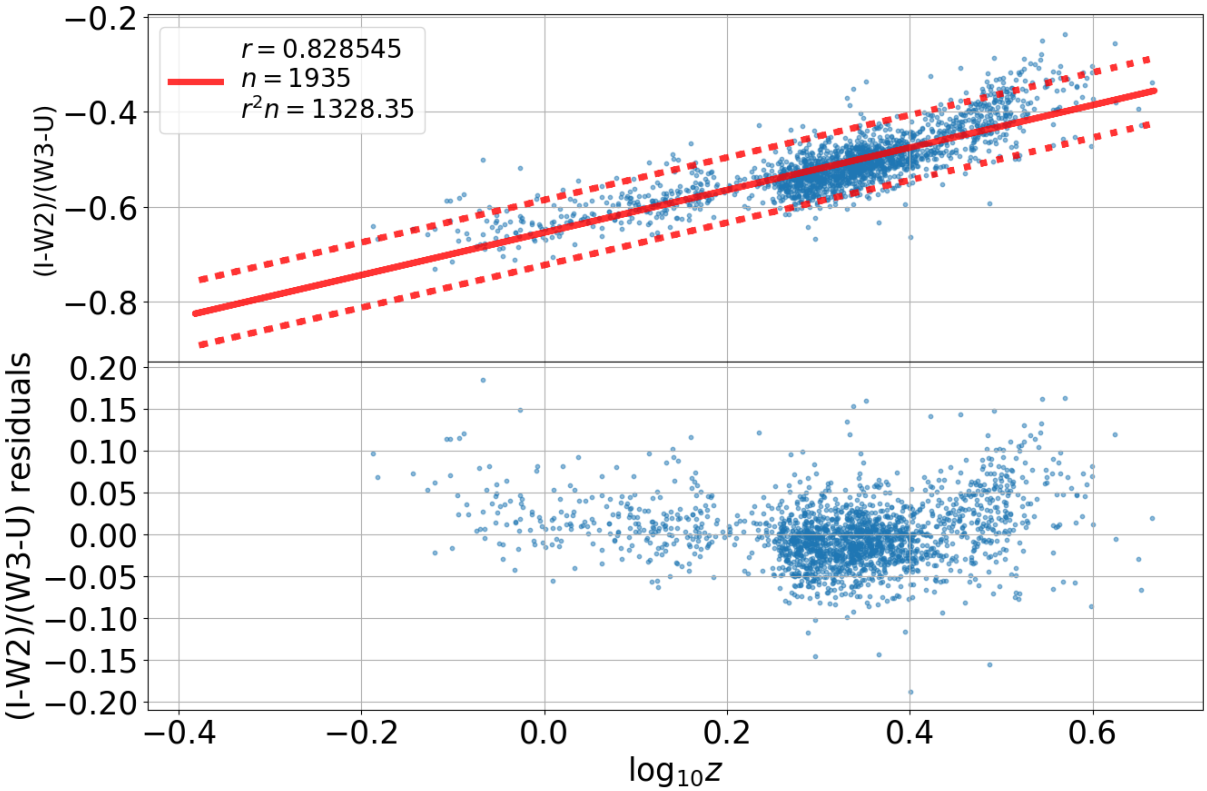

Figure 3.3: The distribution for the colour ratio $(I-W2)/(W3-U)$ against $\log_{10} z$. Residuals are shown on the bottom plot. The red lines show best fit $\pm 1\sigma$.

One issue is the relatively small number of sources which have measurements in all four wavelength bands, as discussed in Section 4. However, given the independence of relative redshift with distance,

it was hoped that the fit should be comparable for any other four bands with similar separations. Figure 3.4 shows the distributions for the wavelengths one band redward, *i.e.* $(R-W1)/(W2-U)$, and one band blueward, $(J-W3)/(W4-B)$, of the best fit model. As can be seen, the fits for each of these are not as good as the best fit model of $(I-W2)/(W3-U)$.

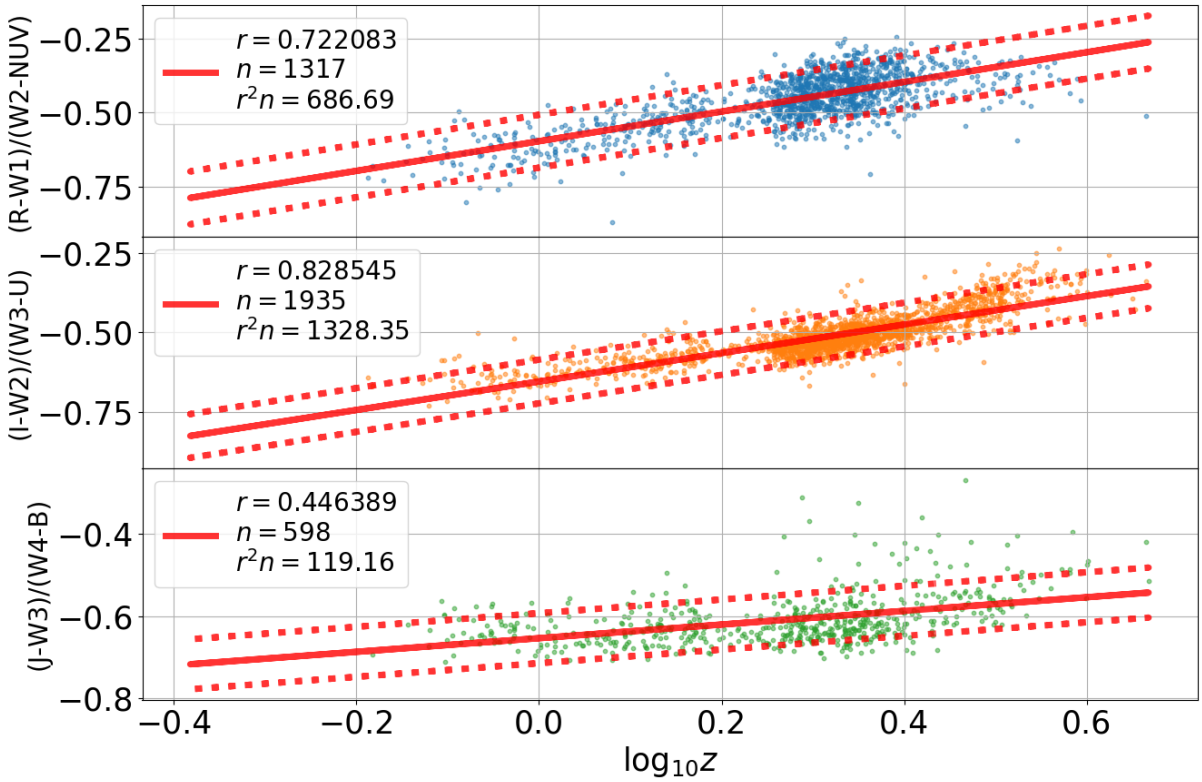

Figure 3.4: The distributions for the wavelengths one band blueward (top), and one band redward (bottom), of the best fit model $(I-W2)/(W3-U)$. The red lines show best fit $\pm 1\sigma$.

3.3 Photometric redshift tests

3.3.1 Single combinations

In order to assess the value of each of the best combinations as a predictor of redshift, tests of the prediction models were conducted on our sample. The best-fit magnitude combinations were each self-tested against the original dataset. These were $U-R$ ($r = 0.596446$), $U-B$ ($r = 0.567813$) and $U-I$ ($r = 0.580567$). Figure 3.5 shows the distribution of $z_{\rm spec} - z_{\rm photo}$ for the best colour, $U-R$, from the SDSS dataset. The strong upward curve and high standard deviation ($\sigma = 0.2749$) mean this colour would make a very poor predictor of redshift.

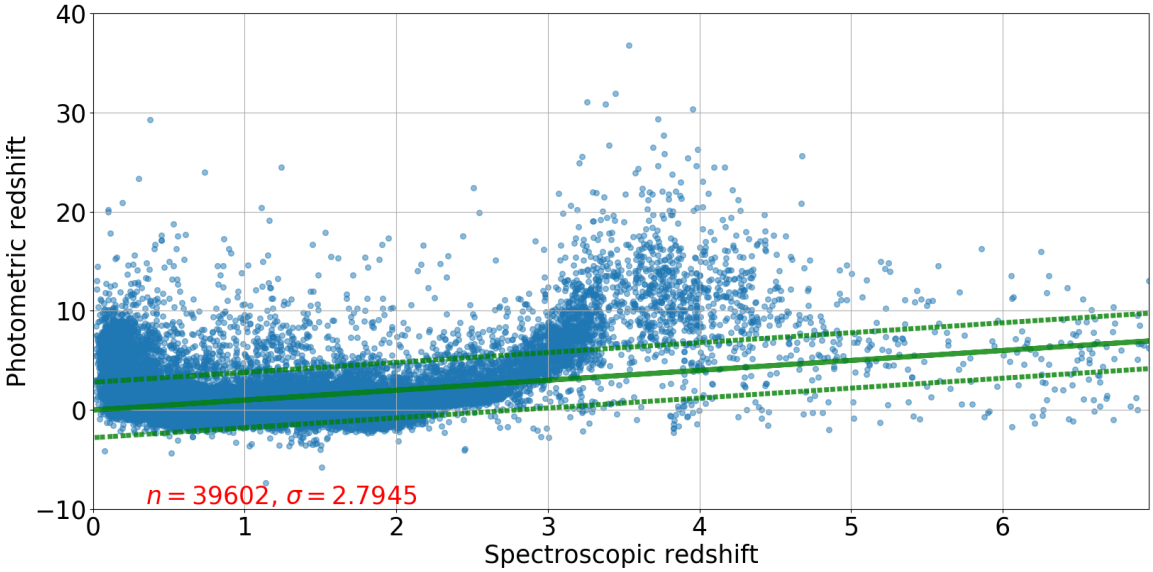

Figure 3.5: Distribution of $z_{\text{spec}} - z_{\text{photo}}$ for $U - R$ over all redshift ranges with no source filtering, and over all redshift ranges, from the SDSS dataset. The green lines show $z_{\text{photo}} = z_{\text{spec}} \pm 1\sigma$.

3.3.2 Redshift predictions for best fit model

The best fit model $(I - W2)/(W3 - U)$ from the MgII dataset was tested against the SDSS dataset. The distribution is shown in Figure 3.6. The best fit model produces good estimates of the photometric redshift for this datase, with a tight correlation and most points lying within $\pm 1\ \sigma\ =\ 0.939$ of $z_{\text{photo}} = z_{\text{spec}}$. This compares favourably to the single colour prediction (Figure 3.5).

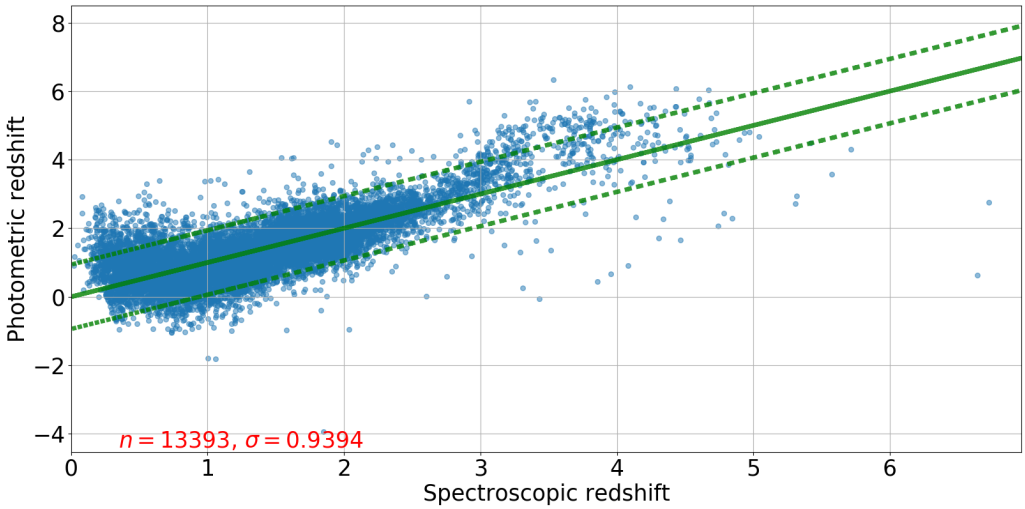

Figure 3.6: The distribution of $z_{\text{spec}} - z_{\text{photo}}$ for $(I-W2)/(W3-U)$ versus z over all redshift ranges with no source filtering, tested against the SDSS dataset. The green lines show $z_{\text{photo}} = z_{\text{spec}} \pm 1\sigma$.

It can be seen in Figure 3.6 that the best fit prediction is not as good in the ranges $z < 1$ and $z > 3$, which is also evident in the increased scatter in the rightmost panel of Figure 3.7.

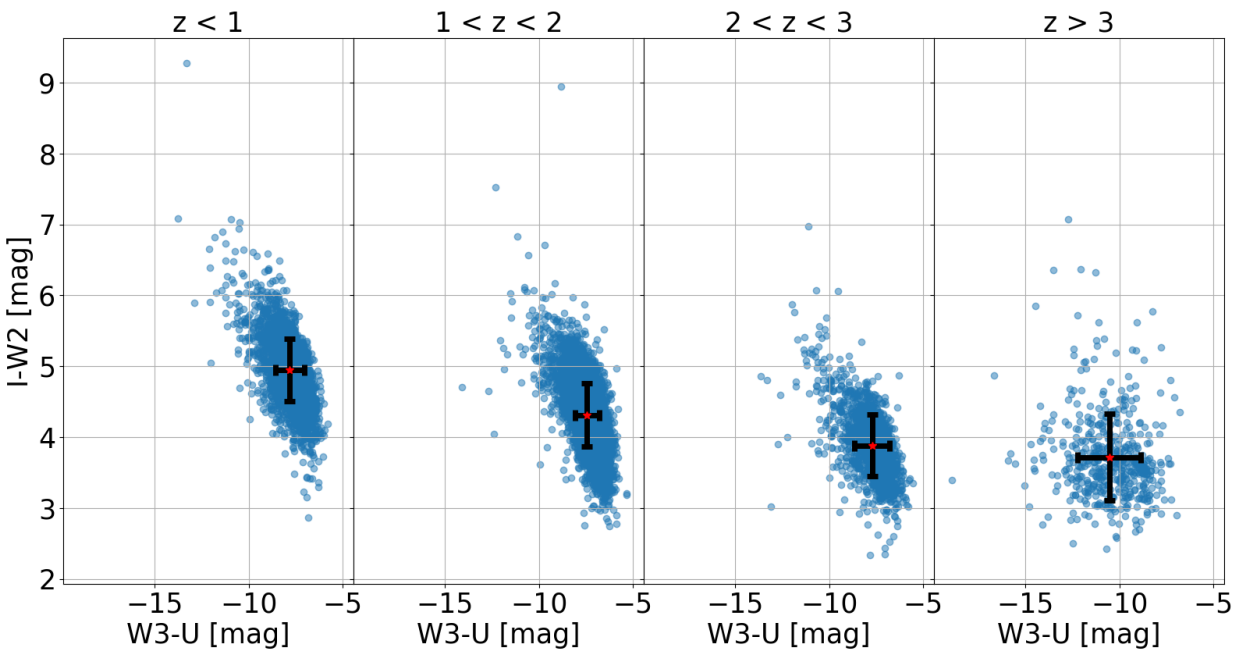

Figure 3.7: Redshift evolution of $I - W2$ versus $W3 - U$ colour-colour diagram. Error bars show $\pm 1\sigma$ about the mean.

In an attempt to find a fit for the sources at either end of the redshift range, further combinations of magnitudes were tested for $z < 1$ and $z > 3$. Due to the redshift, the rest-frame wavelengths have been shifted to longer wavelengths; for example, at a redshift of ~ 1, $W2 \sim U$ (see Figure 3.8).

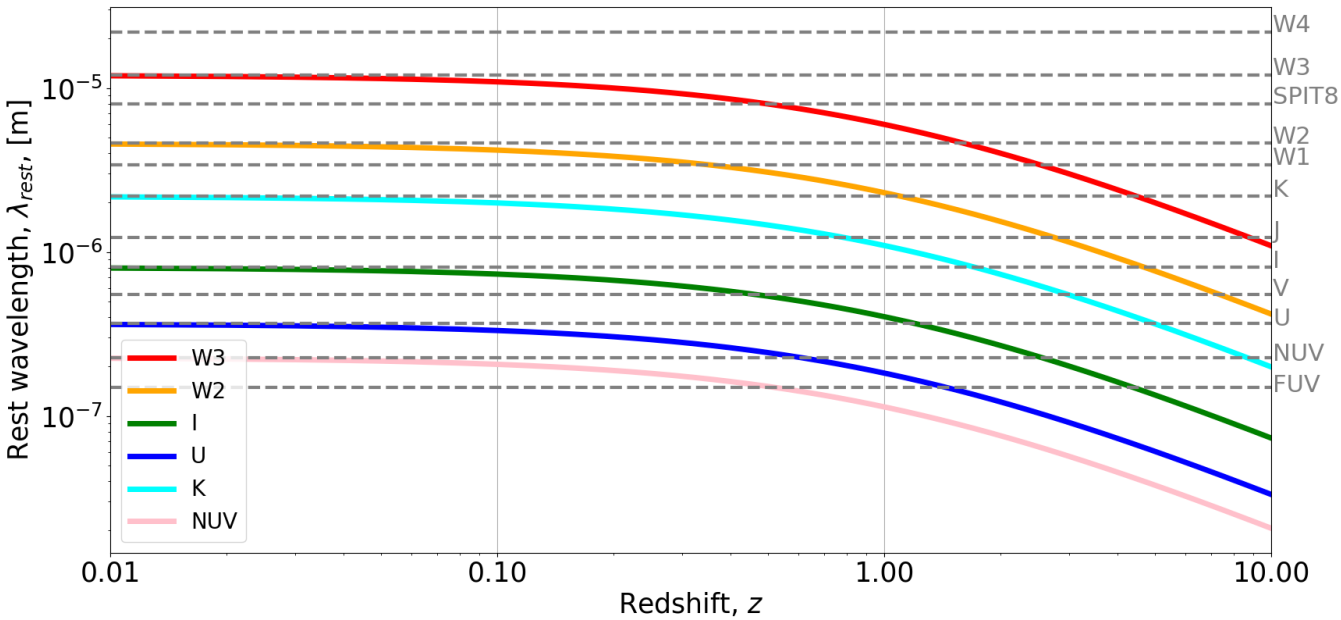

Figure 3.8: The evolution of rest-frame wavelength with redshift for W3, W2, I and U. Dotted horizontal lines indicate the observed-frame near-infrared, optical and GALEX near-ultraviolet (NUV) and far-ultraviolet (FUV) wavelengths.

Thus, for sources of redshift $z \gg 0.1$, the $W3$ and $W2$ bands trace the IR/near-IR wavelengths, while I and U have been shifted to extreme blue/near UV. At $z < 1$, for example, $I \sim U$, $W2 \sim K$, $W3 \sim W2$ and $U \sim NUV$, and thus our best fit model becomes

$$\frac{I - W2}{W3 - U} \longrightarrow \frac{U - K}{W2 - NUV} \text{ for } z < 1 \tag{3.1}$$

which we confirm (Figure 3.9). The $(U - K)/(W2 - NUV)$ colour ratio is the best fit for the range $z < 1$, as expected, with a correlation coefficient of $r = 0.53194$. At the higher redshift $z > 3$, the rest-frame combination $(U - K)/(W2 - NUV)$ will be approximately $(J - SPIT8)/(W4 - V)$. Possibly due to the limited data, the best fit for $z > 3$ is actually $(I - SPIT8)/(W4 - R)$, although the numbers remain small. Table 3.5 gives a summary of the best fits for each range, and Figure 3.9 shows the fit for each of these redshift regimes.

z range	Best fit	r	n	r^2n
$z < 1$	$(U-K)/(W2-NUV)$	0.531940	1242	351.44
$1 < z < 3$	$(I-W2)/(W3-U)$	0.724005	9023	4729.71
$z > 3$	$(I-SPIT8)/(W4-R)$	0.730038	49	26.11

Table 3.5: The best fits for each of the three redshift ranges. Figure 3.9 shows the distributions for each combination in their respective redshift ranges.

The redshift predictions for these combinations were then tested for the SDSS dataset, as shown in Figure 3.10. It can be seen that each of the best fits yields reasonably tight correlations, and most points lie within $z_{\text{photo}} = z_{\text{spec}} \pm 1\sigma$. However, for this method to provide accurate estimates of photometric redshift, we would need to have prior knowledge of the redshift to determine which fit to apply.

Given that a multicomponent fit requires *a priori* knowledge of the redshift, another method to find the switch points was tried. The bottom panel of Figure 1.5 shows that at $z = 1$ ($\log_{10} z = 0.0$), the $W2$ magnitude is ~ 12.5, and at $z = 3$ ($\log_{10} z = 0.47$), $W2 \sim 15$. Since we would not know the redshift before applying the model, we used the $W2$ magnitude to estimate where the switch should occur. Thus, the switch from $(U-K)/(W2-NUV)$ to $(I-W2)/(W3-U)$ should occur at a magnitude of $W2 \approx 12.5$, and from $(I-W2)/(W3-U)$ to $(I-SPIT8)/(W4-R)$ at $W2 \approx 15$. These fits are shown in Figure 3.11, and the z_{photo} predictions in Figure 3.12.

From Figure 3.11 we can see that the equations of the lines of best fit for each redshift regime, using the $W2$ magnitude as a guide to the switching points, are:

z range	Best fit	Equation
$z < 1$	$(U-K)/(W2-NUV)$	$\frac{U-K}{W2-NUV} = 0.2285 \times \log_{10} z - 0.3061$
$1 < z < 3$	$(I-W2)/(W3-U)$	$\frac{I-W2}{W3-U} = 0.1637 \times \log_{10} z - 0.6025$
$z > 3$	$(I-SPIT8)/(W4-R)$	$\frac{U-SPIT8}{W4-R} = 0.1525 \times \log_{10} z - 0.6381$

Table 3.6: The best fits and their equations for each of the three redshift ranges (Figure 3.11).

Hence, this method of estimating the switching points gives the following relations:

$$\log_{10} z = \begin{cases} \frac{1}{0.2285}\left(\frac{U-K}{W2-FUV} + 0.3061\right) & \text{if } W2 \leq 12.5 \\ \frac{1}{0.1637}\left(\frac{I-W2}{W3-U} + 0.6025\right) & \text{if } 12.5 < W2 \leq 15 \\ \frac{1}{0.1525}\left(\frac{I-SPIT8}{W4-R} + 0.6381\right) & \text{if } W2 > 15 \end{cases}$$

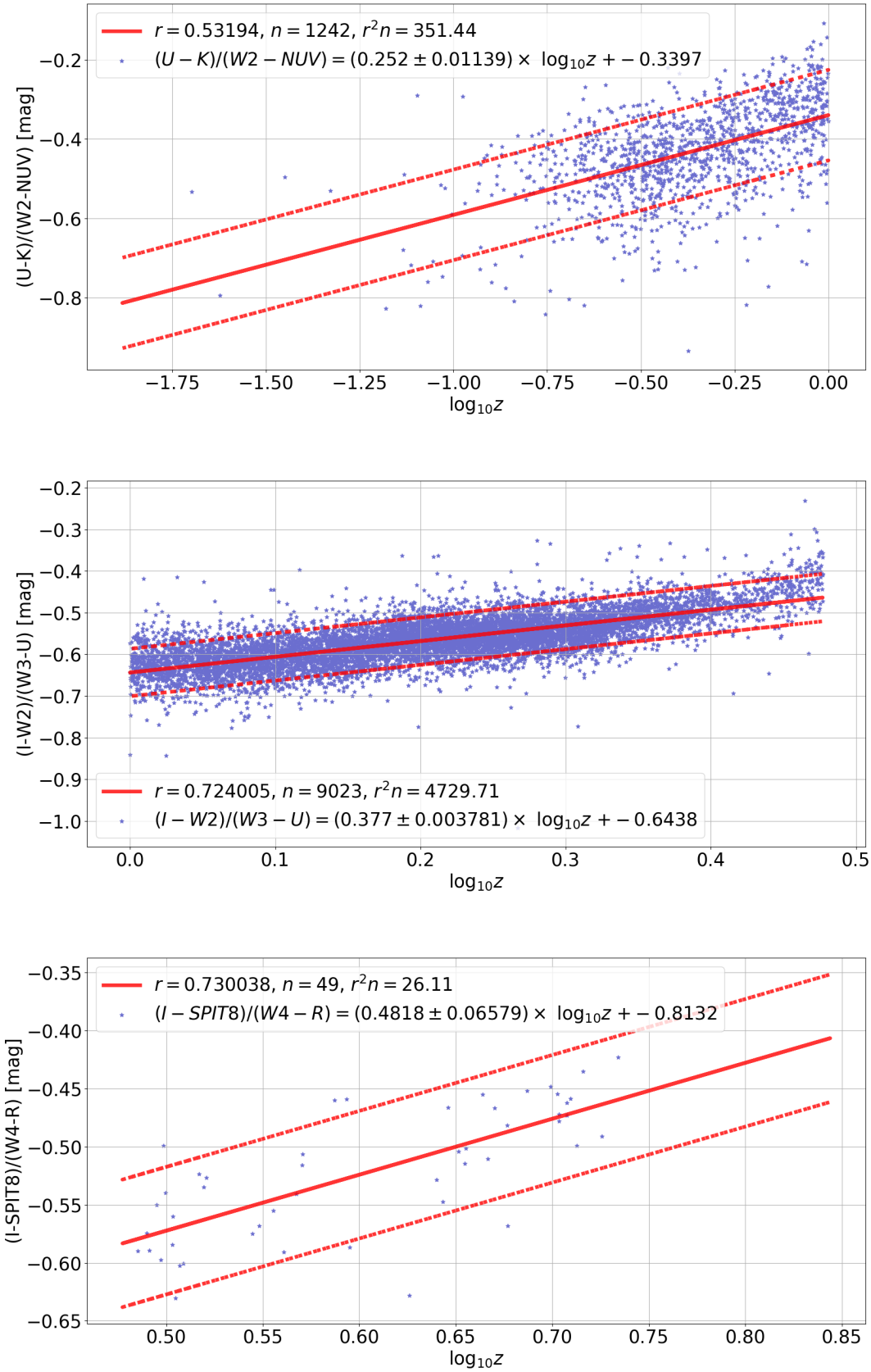

Figure 3.9: The best fits for each redshift range in the SDSS dataset, $(U-K)/(W2-NUV)$ for $z < 1$ (top), $(I-W2)/(W3-U)$ for $1 < z < 3$ (middle) and $(I-SPIT8)/(W4-R)$ for $z > 3$ (bottom). The red lines show best fit $\pm 1\sigma$.

From Figure 3.11, there is some scatter at $z < 0.5$ ($\log_{10} z \sim -0.30$, middle panel), possibly due to an imperfect split at $W2 = 12.5$, as well as a low number of data points. However, the multicomponent fit based on the $W2$ magnitude switch does give accurate predictions of the photometric redshift.

3.3.3 Testing on other databases

Since the best fit model presented in this study is derived from an optically selected sample, we hope to maximise the robustness of our model by testing it on disparate and heterogeneous datasets, so that it may be applied to future datasets for which information may be limited.

The LARGESS catalogue comprises 19 179 radio sources matched with SDSS counterparts, providing redshifts for 10 883 sources (Ching et al., 2017). After removing duplicate lines of sight, the constraints $U \cap K \cap W2 \cap NUV$, $U \cap I \cap W2 \cap W3$, $I \cap SPIT8 \cap W4 \cap B$ required to cover all redshift ranges yielded only 263 sources. When the best fit model is applied to the LARGESS dataset (Figure 3.13), the fit is good, but with a wide spread. At smaller redshifts, galaxies contribute a large number of sources to the sample, and $|z_{\text{photo}} - z_{\text{spec}}| \gg 0$. Figure 1.6 shows that, at $z < 1$, galaxies contribute a large number of sources to the sample. Hence, the bottom panel of Figure 3.13 shows the distribution of Δz for the LARGESS sample filtered for broad line sources (mostly QSOs). Filtering the LARGESS sample for broad line sources reduces the total number of samples from $n = 287$ to $n = 253$, but it eliminates the outliers at $z \approx 0$, and lowers the standard deviation from $\sigma = 1.6$ to $\sigma = 1.1$.

The Second Realization of the International Celestial Reference Frame (ICRF2) by Very Long Baseline Interferometry (Ma et al., 2009), comprises strong flat-spectrum radio sources, of which 1 682 have known redshifts. Upon removal of duplicate lines of sight and unreliable redshifts, 123 sources satisfy the magnitude constraints. However, the standard deviation ($\sigma = 1.05$) is comparable to that of the LARGESS sample ($\sigma = 1.11$). Figure 3.14 shows the distribution of z_{photo} against z_{spec} for the ICRF2 sample.

3.4 Comparison with other studies

Other studies have also used releases of the SDSS data to determine accurate photometric redshifts for quasars, with narrower distributions of Δz, typically with a spread of $\sigma_{\Delta z} \approx 0.1$ (Richards et al., 2001; Weinstein et al., 2004; Maddox et al., 2012). The Hubble Deep Field (HDF) has also been

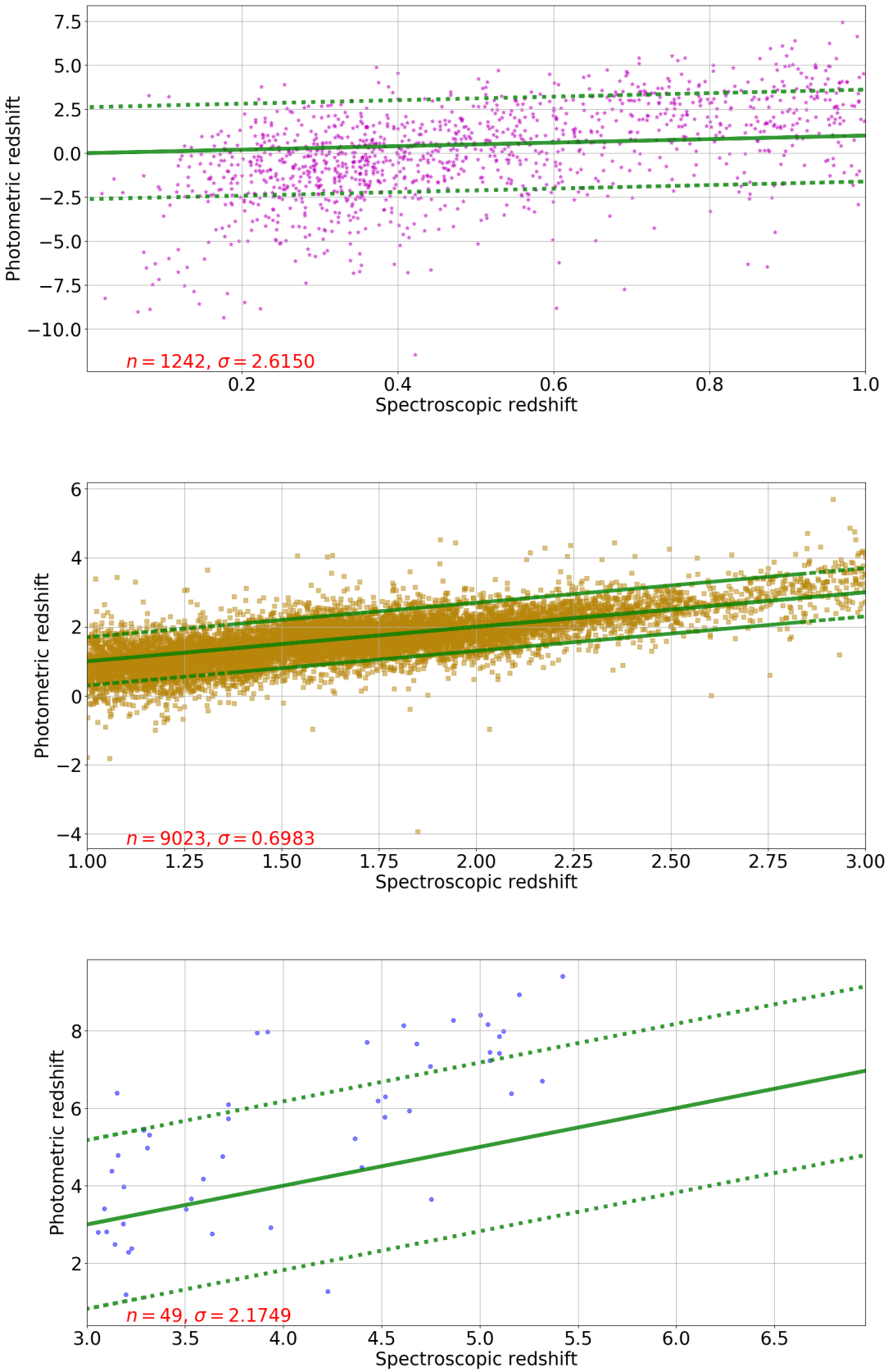

Figure 3.10: Comparison of $z_{\text{spec}} - z_{\text{photo}}$ for $(U-K)/(W2-NUV), z < 1$ (top), $(I-W2)/(W3-U), 1 < z < 3$ (middle), and $(I-SPIT8)/(W4-R), z > 3$ (bottom) against z, applied to the SDSS dataset. The green lines show $z_{\text{photo}} = z_{\text{spec}} \pm 1\sigma$.

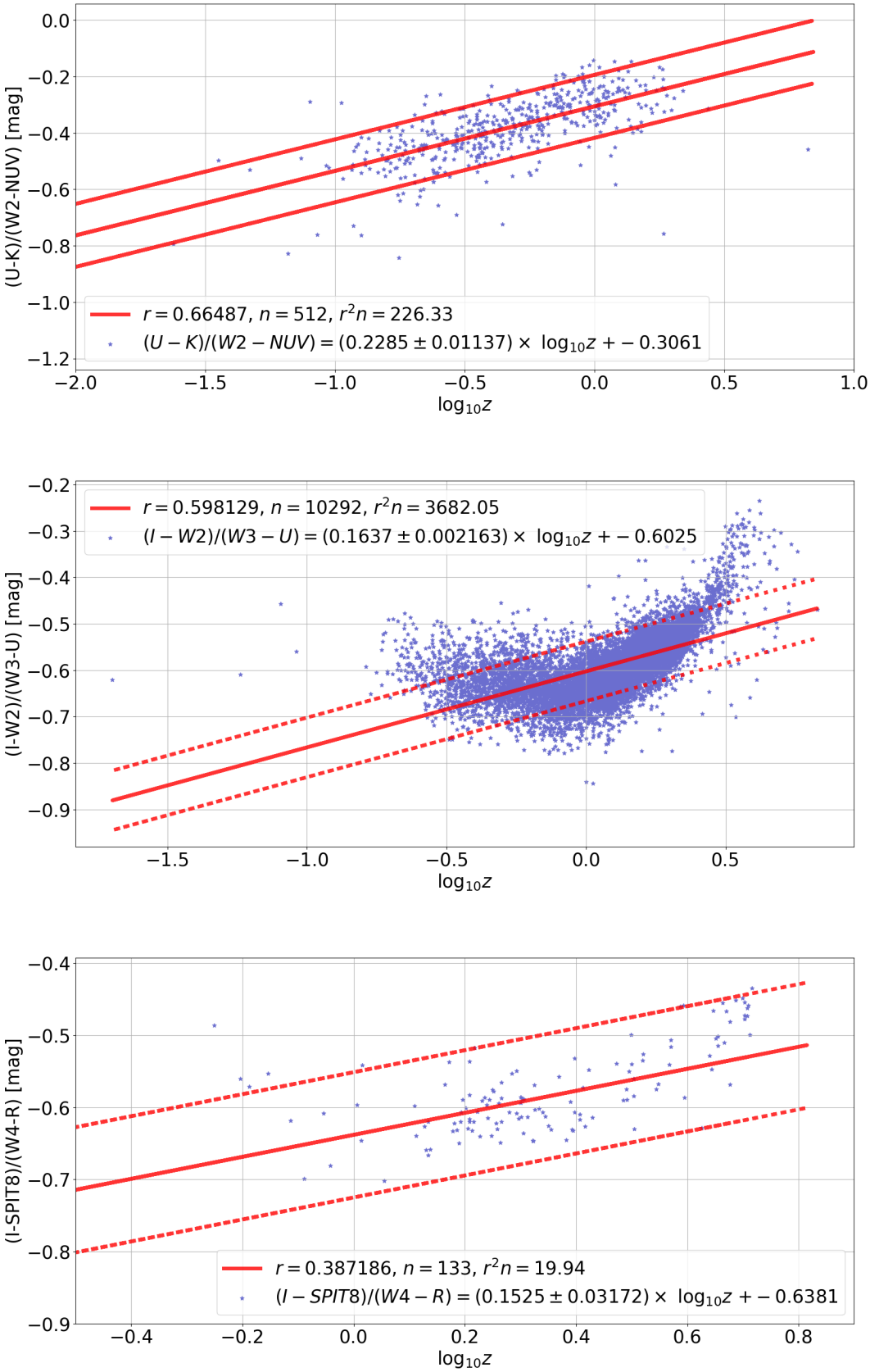

Figure 3.11: As for Figure 3.9, but with the split based on the magnitudes in the bottom panel of Figure 1.5. The red lines show best fit $\pm 1\sigma$. Standard deviations are σ =0.1116 (top panel), σ =0.0641 (middle panel), σ =0.0868 (bottom panel).

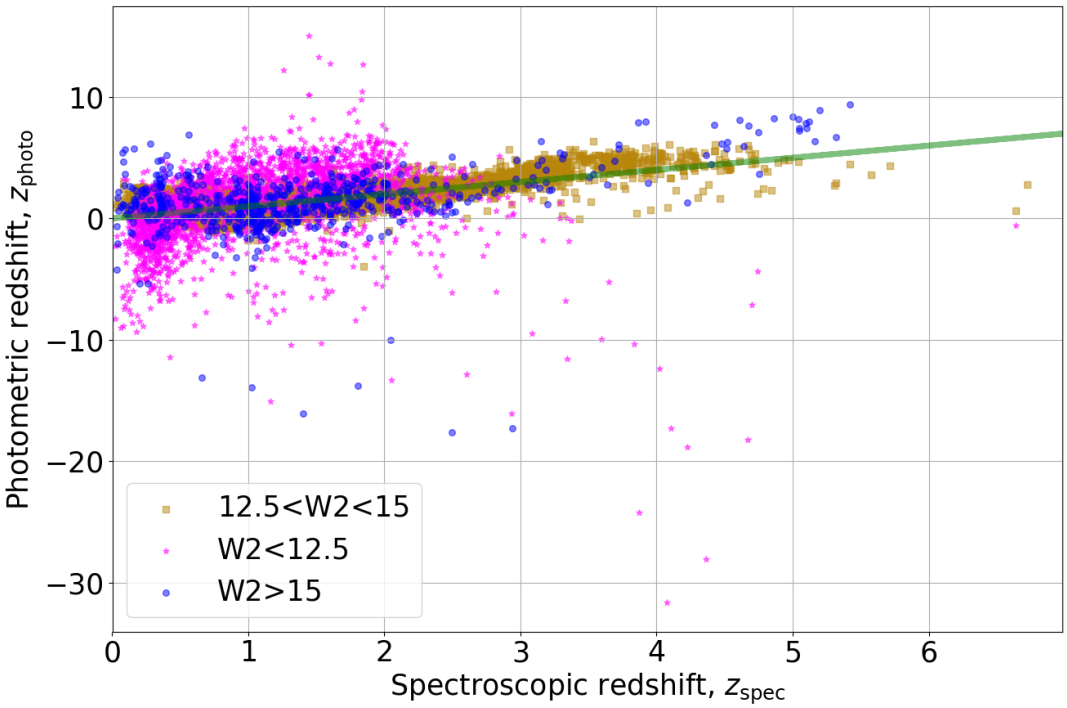

Figure 3.12: Predictions of $z_{\rm photo}$ from the SDSS dataset, with the split based on the $W2$ magnitudes. Symbols as per Figure 3.10 to show switch points at each $W2$ magnitude.

extensively surveyed for photometric estimates of redshifts (Gwyn and Hartwick, 1996; Massarotti et al., 2001; Fernández-Soto et al., 2002). All of these produce estimates with $\sigma_{\Delta z} < 0.1$, but they either rely on template spectra, or methods considerably more complex than our own, in that they employ sophisticated machine learning algorithms, or complex techniques to break degeneracies. Section 6 gives an outline of these.

Our method is much simpler than those of Han et al. (2016), which used the computationally intensive *kNN* machine learning algorithm; Richards et al. (2001) which used a conflagration of four photometric colours compared to the two used for our method; and those of Massarotti et al. (2001) and Weinstein et al. (2004), both of which used a multicomponent fit for every source in their surveys. Our sample set is much larger than those of Polsterer et al. (2013) (1106 sources) and Maddox et al. (2012) (324 sources). Unlike the studies which used χ^2 (Gwyn and Hartwick, 1996; Richards et al., 2001; Fernández-Soto et al., 2002; Wu et al., 2004)) to determine redshift, our method is not subject to degeneracies; this is also a drawback of the probability method employed by Richards et al. (2001), which also had high uncertainties for $z < 0.3$.

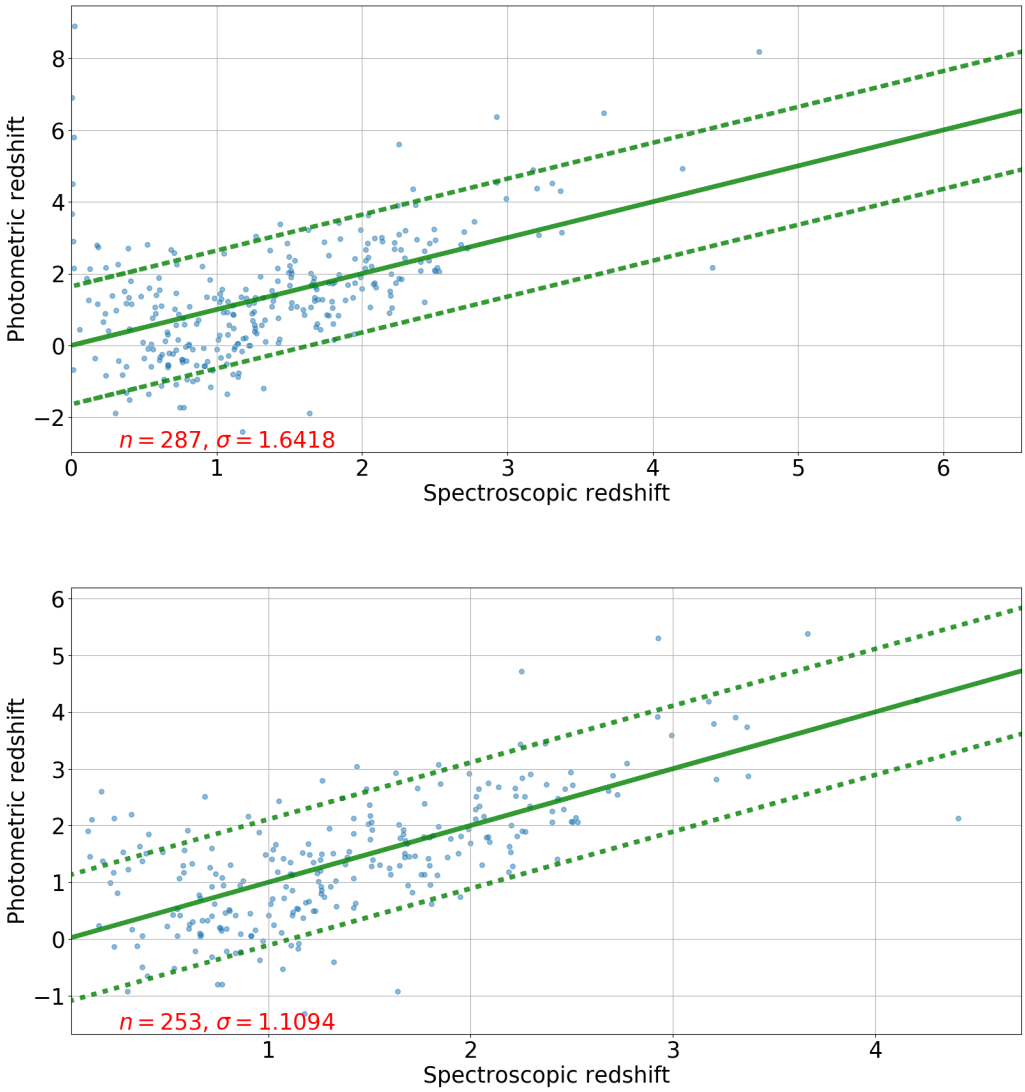

Figure 3.13: Comparison of $z_{\rm spec} - z_{\rm photo}$ for $(I-W2)/(W3-U)$ against z, applied to the LARGESS dataset with no source filtering (top), and filtered for broad line sources (bottom). The green lines show $z_{\rm photo} = z_{\rm spec} \pm 1\sigma$.

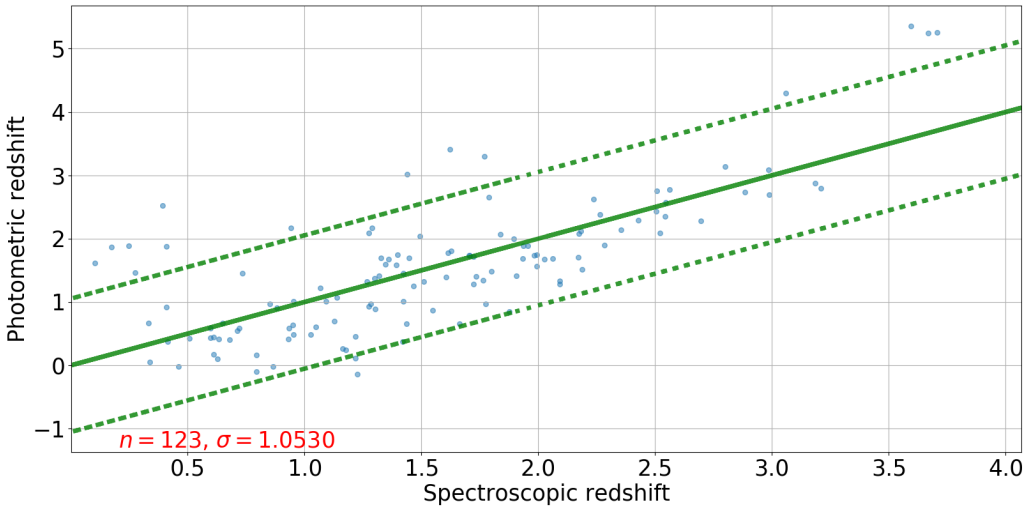

Figure 3.14: As for Figure 3.13 bottom panel, but for the ICRF2 dataset.

4 Discussion

4.1 Physical interpretation

The near-infrared emission in QSOs is believed to be due to torus dust around the nucleus being heated by the AGN (Hatziminaoglou, 2010), while the UV emission gives rise to an excess blue colour of the QSO (Shields, 1978). Because of this, an increasing AGN contribution may be apparent as a decrease in $U-K$. The emitted $U-K$ is redshifted to the observed $I-W2$, which does indeed decrease with redshift (Figure 3.7), as expected. At $z < 1$, we see no correlation between $W2 - NUV$, which traces a wider wavelength range than $U - K$, and redshift. The fact that $U - K$ is correlated with redshift while $W2 - NUV$ is not, suggests that the $U - K$ colour may be sufficient for estimating the photometric redshift. However, the top panel of Figure 4.1 shows that there is a degeneracy between the $U - K$ colour and redshift, where the colour index matches more than one spectroscopic redshift, and the $W2 - NUV$ colour is required to break this degeneracy and improve the fit. Hence, this suggests that the rest-frame $U - K$ colour traces the excess flux due to the AGN, and thus offers a measure of the redshift. Figure 4.1 shows the $U - K$ colour for $z > 1$ (top panel), and $W2 - NUV$ for $z < 1$ (bottom panel). Figure 3.7 shows that the $I - W2$ and $W3 - U$ colours both decrease, and become more scattered, with redshift. It is not clear why these colours should be particularly sensitive to the redshift, although we offer the following speculation.

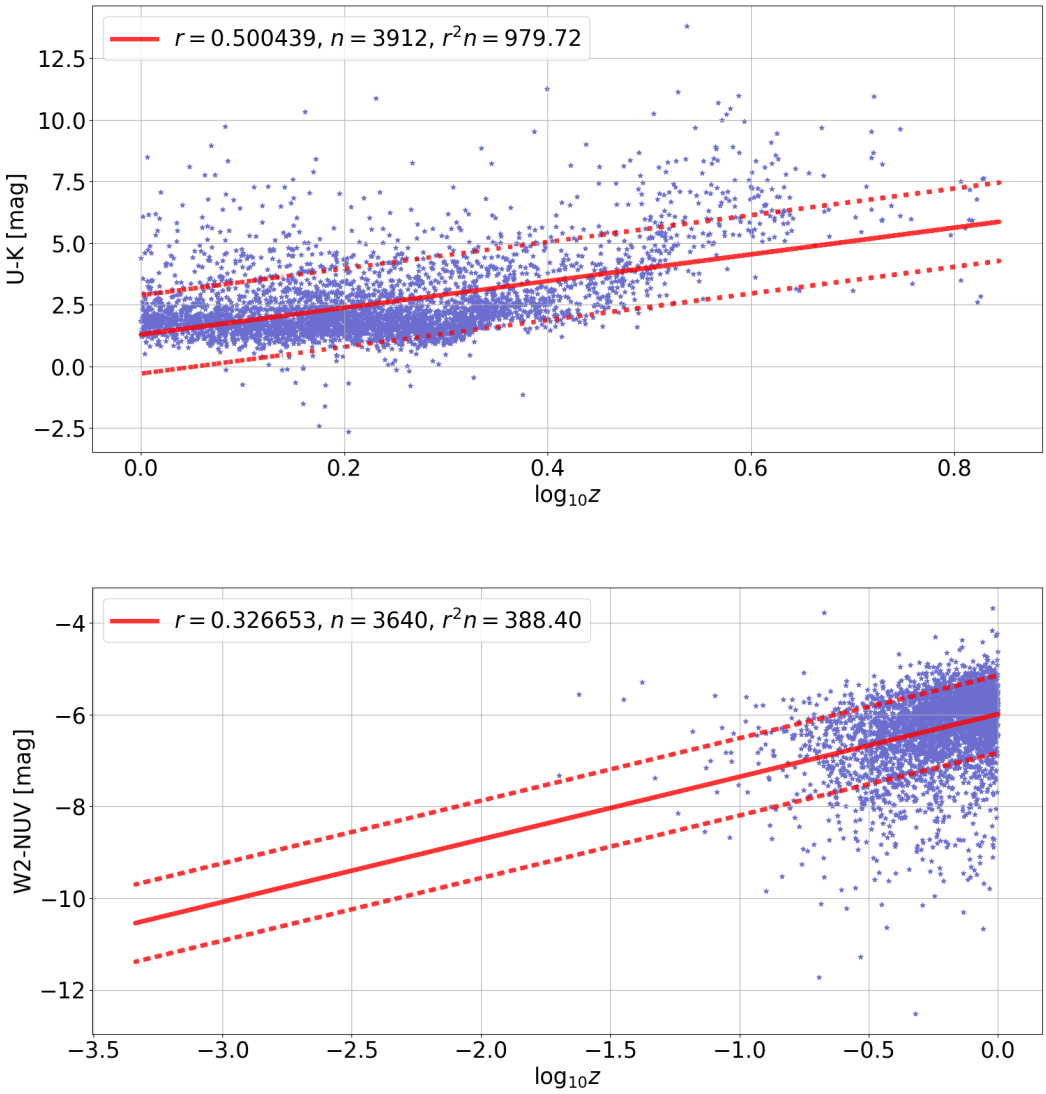

Figure 4.1: The $U - K$ and $W2 - NUV$ colours versus $\log_{10} z$ for the SDSS sample. The top plot has $r = 0.500439$, $n = 3912$, $r^2n = 979.72$ for the range $z > 1$. The bottom plot has $r = 0.326653$, $n = 3640$, $r^2n = 338.40$ for the range $z < 1$. The red lines show best fit $\pm 1\sigma$.

The hot dusty torus of an AGN emits mostly in the infrared[§], with a substantial fraction of their bolometric luminosity being emitted at wavelengths $\sim 8 - 1000\ \mu m$ (Sanders, 1999), and, in the case of an obscured AGN, we can expect most of the UV emission to be blocked by the optically thick disk. This dusty torus can be considered to cool the outer regions of the accretion disk by absorbing the ultraviolet emission and isotropically re-emitting it at infrared wavelengths (Goulding et al., 2012). The axisymmetric, but anisotropic geometry of the torus means that for some lines of sight, the emission from the accretion disk is obscured, while for other, narrower lines of sight, the central accretion disk is directly observable. Obscuration of the accretion disk can also be due to the host galaxy, in the form of dust-obscured star forming regions and dust lanes. This can be expected to have some impact on the demography of the observed sample of AGNs. Different sized tori will have a different covering factor (the fraction of the sky covered by the obscuring material, from the point of view of the SMBH), and this is one of the main elements affecting the intensity of the reprocessed radiation (Ramos Almeida and Ricci, 2017). This, in addition to the small angular size of the accretion disk, means that the obscuration is likely to be more significant for inclined and edge-on galaxies, since, on average, the typical optical depth at any given viewing angle will be higher than for a face-on view. Higher obscuration can be expected to be correlated with higher infrared emission, due to the re-emission of shorter wavelengths by the dusty torus. The best-fit model with the widest range of redshifts, $(I-W2)/(W3-U)$, traces mostly infrared emission (the I, $W2$ and $W3$ bands), and therefore likely reflects the higher proportion of obscured AGN in the sample. Figure 1.3 shows the main components responsible for the obscuration of an AGN, colour-coded to indicate their impact on the observed spectrum.

4.2 Limitations

The best results are achieved when using colours, rather than magnitudes. This may be because the effects of differences in the intrinsic characteristics of the objects are minimized by the dimension reduction from filter band to colour space, giving a normalisation effect. Additionally, it suggests that our $(I-W2)/(W3-U)$ colour ratio gives an accurate measure of the AGN activity (Section 4). The use of both $I - W2$ and $W3 - U$ colours from the sample dataset requires the sources in the survey to have readings at $I \cap W2 \cap W3 \cap U$. This requirement significantly limits the number of

[§]Although stars also heat interstellar dust. It would be difficult to distinguish the source of the heating.

sources for which the best fit model can be used. From Table 4.1, we see that the limiting factor is the measurements of the WISE wavebands. However, the inclusion of the $W2$ and $W3$ bands is necessary to provide reliable redshift predictions.

Sample	Total	$W3$	$W2$	I	U
MgII	16 580	2911	3316	6150	6032
SDSS	50 000	17 230	17 717	49 205	48 522
LARGESS	10 931	326	377	10 846	9746

Table 4.1: The total number of measurements in each of the magnitude bands for the sample datasets.

The multicomponent fit shown in Section 3.3.2 (Figures 3.9 and 3.10) provide good photometric redshifts, but the requirement of measurements in $U \cap K \cap W2 \cap NUV$, $U \cap I \cap W2 \cap W3$, $I \cap SPIT8 \cap W4 \cap B$ to cover all redshift ranges further limits the number of sources.

5 Conclusion

Given that a simple, accurate and reliable photometric estimate of redshift will be invaluable for upcoming large extragalactic radio surveys, we have tested the feasibility of a statistical estimate of redshift from near-infrared and visible magnitudes. This builds upon the work of De Breuck et al. (2002), who find a tight correlation between the K-band magnitude of galaxies, and their redshift. When applied to our sample, dominated by illuminated MgII absorption systems (mostly QSOs), the fit of De Breuck et al. (2002) provides only a lower limit to the redshift, underestimating the redshift of QSOs in the sample.

Glowacki et al. (2017) estimated photometric redshifts from the WISE W1 and W2 bands in the LARGESS catalogue. However, with low correlation coefficients of $r = 0.56$ and $r = 0.36$, the correlation is not tight enough to provide an accurate estimate of redshift, with poor predictions when applied to other datasets. We therefore test the correlations of numerous combinations of magnitudes with spectroscopic redshift in the MgII sample, with 17 285 sources. We find that the ratio of $(I - W2)$ to $(W3 - U)$ colours yields a regression coefficient of $r = 0.83$ for the 1 975 sources for which all four magnitudes are available. Upon application of this ratio to the 50 000-source SDSS DR12 sample, the fit fails to provide accurate estimates at redshifts of $z \leq 1$ and $z \geq 3$, and we further test two different combinations over these redshift ranges. We find that the ratio

$(U-K)/(W2-NUV)$, which is essentially the observed $(I-W2)$ to $(W3-U)$ ratio in the rest-frame of the source, provides the best fit for $z \leq 1$, and $(I-SPIT8)/(W4-R)$ provides the best fit at $z \geq 3$, where $SPIT8$ is the Spitzer 8.0 μm wavelength. However, given that we have no *a priori* knowledge of the redshift, we attempt to estimate this with the weak $W2-z$ relation, where we find $W2 > 12.5$ at $z > 1$ and $W2 > 15$ at $z > 3$. This gives the following relations:

$$\log_{10} z = \begin{cases} \frac{1}{0.2285}\left(\frac{U-K}{W2-FUV} + 0.3061\right) & \text{if } W2 \leq 12.5 \\ \frac{1}{0.1637}\left(\frac{I-W2}{W3-U} + 0.6025\right) & \text{if } 12.5 < W2 \leq 15 \\ \frac{1}{0.1525}\left(\frac{I-SPIT8}{W4-R} + 0.6381\right) & \text{if } W2 > 15 \end{cases}$$

Even though the standard deviations are larger than those of other studies, with $\sigma_{\Delta z} \sim 0.1$, these often have long Δz wings extending past the Gaussian, and apply only over a limited range of redshifts. Furthermore, the methods are often significantly more complex, requiring either machine learning algorithms, or complex techniques to break the degeneracies which these methods introduce. We also apply our model to radio samples for which we have redshifts. For these sources, the Δz distribution is very similar to that of the SDSS dataset, which indicates that the model may also be applicable to other, future surveys, which may lack spectroscopic redshift measurements.

It is, in fact, the rest-frame $U-K$ colour which is anticorrelated with redshift, which is to be expected if the ultraviolet emission traces the AGN activity, while the far-infrared is dominated by stellar activity. Thus, the relation provides an analogue of the $K-z$ relation for galaxies. The major drawback with our method, is the requirement of having measurements for nine individual magnitudes ($U \cap K \cap W2 \cap NUV$, $U \cap I \cap W2 \cap W3$, $I \cap SPIT8 \cap W4 \cap B$). Relying on only two magnitudes, as for the $W1-z$ and $W2-z$ of Glowacki et al. (2017) would increase the applicability of this method as an estimator of photometric redshift. However, the WISE bands are required to measure K at $z \geq 1$, and $W2$ is required to estimate the redshift range for the application of the multicomponent fit. Nevertheless, even a 2% yield will allow us to obtain photometric redshifts for over one million of the sources expected to be detected with the Evolutionary Map of the Universe.

6 Appendices

Statistical tests

Since the regression coefficient r is the most important statistical parameter in this investigation, this was the main focus of the goodness of fit for the correlations investigated. However, applying a linear fit for the data comes with some underlying assumptions. Violations of these assumptions can affect the conclusions reached, and therefore, the following statistical tests were applied to ensure that the assumptions are met, and hence that the linear fit is valid.

Ordinary Least Squares

In order to ensure that a linear fit is valid for a certain set of data, the following assumptions must be satisfied:

1. **The expected value of y (in this case, magnitude, or combination of magnitudes) is a linear function of x ($\log_{10} z$)**; that is, the slope of the line does not depend on the values of other variables. In the case of this investigation, since we are not working with a time series, non-linearity is best diagnosed by taking a plot of residuals (*i.e.* the difference between the plotted value and the best fit line) versus $\log_{10} z$. For a good linear fit, the residuals should be distributed symmetrically about zero.

2. **The unexplained variations of y are independent random variables.** This can be tested for using the *Durbin-Watson statistic*, which should be close to 2.

3. **All of the data points have the same variance** ("homoscedasticity"). Deviation from homoscedasticity (called heteroscedasticity) implies that either:

   (a) a linear fit is not appropriate – in which case a polynomial fit may be tried, or

   (b) there are other variables which affect the dependent variable.

   Homoscedasticity can also be judged by plotting the residuals against the independent variable.

4. **All of the data points are normally distributed.** This was tested for by using the Jarque-Bera test (Jarque and Bera, 1980), which is a goodness-of-fit test based on the skew

(which should be close to zero) and kurtosis (which should have an absolute value of < 3, as a rule of thumb) of a sample.

Kendall's Rank Coefficient

Kendall's τ can be formulated as a special case of r, and provides a way to measure the ordinal association between two quantities. A value of $\tau = 1$ indicates that the two quantities have identical rank (relative position label of the observations within the variables), while $\tau \sim 0$ indicates that there is little to no correlation between the relative positions of the two variables.

Previous studies of photometric estimates of quasar redshift

Methods employing template spectra

In an effort to investigate the redshift distribution and luminosity functions of objects in the Hubble Deep Field, Gwyn and Hartwick (1996) converted the magnitude in each bandpass to a flux using the same method as described from Equation 2.1 onwards, and then, by plotting the flux against wavelength, to a low-resolution spectral energy distribution (SED). Using a set of template spectra of all Hubble types, the SED of each object was matched with the closest spectrum, determined by minimizing χ^2, defined as

$$\chi^2 = \sum_{i=1}^{N_f} \frac{(F_i - \alpha T_i)^2}{\sigma F i^2} \tag{6.1}$$

for each source, where F_i and σF_i are the flux and the uncertainty in the flux, respectively in each bandpass of the observed galaxy, T_i is the flux in each bandpass of the template, α is a normalisation constant, and the sum is over the number of filters N_f in the set. The photometric redshift was then tested against simulations and photometric observations of galaxies with known spectroscopic redshifts. They used this method to show that the redshift distribution of HDF objects contains two peaks: one at $z \approx 0.6$ and another at $z \approx 2.2$. This indicates that larger galaxies form stars early in the Universe's history, at $z \sim 3$, while in the dwarf galaxies star formation is delayed, until around $z \sim 1$.

Wu et al. (2004) used the χ^2 minimization technique on data from the SDSS to estimate the

photometric redshifts on a colour-colour diagram of a large sample of quasars, by comparing theoretical photometric colours with the observed colours. They show that identifying the redshift corresponding to the smallest value of χ^2 for a given quasar can give a reasonable estimate (to within $|\Delta z| \leq 0.2$ in most cases) of its redshift. However, this method can lead to incorrect results. If the spectrum of the quasar is very different from the composite spectrum, or if there are large uncertainties in the photometric magnitudes, the calculated χ^2 can reach a minimum at the wrong redshift. More composite spectra for different types of quasars can alleviate this problem. In addition, their selection criteria can miss many quasars, especially those with lower redshifts, as they may be located in the same colour-space as stars. The experimental results from this template-fitting method have shown that the accuracy of their estimates rely on templates constructed by either simulation or by observational data.

Machine Learning and Statistical methods

In addition, and complementary to, the template-fitting methods, machine learning is a fundamental component of current photometric estimation methods. In this approach, a computerised algorithm learns the relationship between colour and redshift, based on a training sample of either purely photometrically derived redshifts, or a full template of both photometry and reliable spectroscopically determined redshifts, for the training sample (Salvato et al., 2019).

Han et al. (2016) used the data-mining method k-nearest neighbour (kNN) to estimate the redshift. The kNN algorithm compares the Euclidean distance between a test sample point and its k nearest neighbours in a feature space, such as colour or luminosity, and assigns a weighted combination of the redshifts of those nearest neighbors to the test object in order to sort each source into a group. They then maximised the separation of the groups, thereby minimizing the classification error, to help mitigate the catastrophic failure inherent in other photometric redshift approaches. The kNN algorithm, however, is computationally intensive, and for a sample of, for example $\sim 10^5$ samples, can be prohibitively difficult. However, it is possible to apply specific spatial data structures, such as k-dimensional trees, which can be used to pre-organise the data points into a reduced space, thus facilitating its implementation on large datasets (Norris et al., 2019).

Also using the kNN method, Polsterer et al. (2013) found that the best results in terms of the separation of the groups in the algorithm are obtained using colours, rather than magnitudes. They found that their method performed well for $z > 4.8$, and as the redshifts in the reduced reference sample were homogeneously distributed, they avoided catastrophic outliers at these high z values, despite the sample size being decreased by 98.5% (*i.e.* from 77096 to 1106 samples).

Fernández-Soto et al. (2002) describe a technique of assigning confidence limits to redshift measurements, along with their associated probability functions. In order to mitigate the uncertainty in redshift associated with uncertainty in the photometric measurements, they calculated the likelihood function,

$$L\left(z, \quad T\right) = \prod_{i=1}^{N} exp\left\{-\frac{1}{2}\left[\frac{f_i - AF_i\left(z,T\right)}{\sigma i}\right]^2\right\} \tag{6.2}$$

where the product extends to the number of filters used, A is a normalization constant, f_i and σ_i are the flux and associated error of the source measured in the ith band, and $F_i\left(z,T\right)$ are the model fluxes for a galaxy of type T at redshift z in the ith band. Using a template model to produce the fluxes cannot reproduce the SEDs of all galaxies, particularly for very bright sources. High-quality photometry will amplify any differences between the model and the observations, producing a very bad χ^2 fit for a very bright source.

Massarotti et al. (2001) used the Hubble Deep Field North to minimize the residuals between photometric $z_{\rm phot}$ and spectroscopic $z_{\rm spec}$ redshifts. Photometric redshifts were obtained by comparing the observed broad-band colours of galaxies with the model fluxes provided by Buzzoni (1989, 1995); Buzzoni (2002). For each source in their sample they used a multi-component fit to the galaxy SED, finding the best template that minimized the χ^2 value. They then analysed the residuals by defining two statistical moments

$$\overline{\Delta z} = \sum^{N_g} \frac{\left(z_{\rm spec} - z_{\rm phot}\right)}{N_g} \tag{6.3}$$

and

$$\sigma_z^2 = \sum^{N_g} \frac{\left[(z_{\rm spec} - z_{\rm phot}) - \overline{\Delta z}\right]^2}{N_g - 1} \tag{6.4}$$

for the number of galaxies N_g in each sample. Catastrophic outliers were discarded, and the photometric redshifts were compared to the spectroscopic values across different library templates. They found that all of their templates found very good agreement between $z_{\rm phot}$ and $z_{\rm spec}$ for $z_{\rm spec} < 1.5$, while for $z_{\rm spec} > 2$ they had to account for the absorption effects of the interstellar and intergalactic media (ISM and IGM), which scatter and absorb light, reddening the SED.

Weinstein et al. (2004) constructed an empirical colour-redshift relation (CZR) using a set of quasars for which both the spectroscopic redshift $z_{\rm spec}$ and photometric redshift $z_{\rm phot}$ are known. The most heavily reddened quasars in the $u - g$ and $g - r$ colours were removed to prevent the CZR being heavily skewed toward the red. They then computed the mean colour vector M_i (in four dimensions, one for each colour), and the 4×4 colour covariance V_i (a sixteen-element matrix) for all Q quasars in each of N redshift bins (a total of 3814 quasars) in the study using the $u - g$, $g - r$, $r - i$, and $i - z$ colours, for which

$$M_i^j = \frac{1}{Q_i} \sum_{q=1}^{Q_i} x_{j,q} \quad (j = 1, 2, 3, 4) \tag{6.5}$$

and

$$V_i^{jk} = \frac{1}{Q_i - 1} \sum_{q=1}^{Q_i} \left(x_{j,q} - M_i^j\right) \left(x_{k,q} - M_i^k\right) \quad (j, k = 1, 2, 3, 4) \tag{6.6}$$

where j and k are integers representing the colours and $x_{1,q}$, $x_{2,q}$, $x_{3,q}$, and $x_{4,q}$ are the four photometric colours of the qth nonreddened quasar in the ith bin. They tested their algorithm first on the entire training set, and then on a subset of the same training set, finding similar results in both tests. They were able to predict the $z_{\rm phot}$ to within 0.3 for up to 80.8% of the quasars with known $z_{\rm spec}$ in the subset test.

Bovy et al. (2012) use a similar method to Weinstein et al. (2004), in that they compute the joint probability of an object's fluxes, its redshift, and the proposition that it is a quasar, which they write as

$$p\,(\text{fluxes, z, quasar}) = p\,(\text{fluxes}|\text{z, quasar}) \times p\,(\text{z}|\text{quasar})\,P\,(\text{quasar}) \tag{6.7}$$

$$= p\,(\text{fluxes}|\text{z, quasar})\,P\,(\text{quasar}) \tag{6.8}$$

They used the extreme deconvolution method of Bovy et al. (2011) to optimize the uncertainty distribution function for noisy, incomplete, or heteroscedastic datasets, which is often the case in astronomy. Their predictions (Figure 15 in Bovy et al. (2012)) show high levels of uncertainty for $z < 3$, and the *ugriz*-only photometric data they use suffers from inherent degeneracies, which are propagated through their algorithm (see their Figure 12).

Richards et al. (2001) used a sample of 2625 quasars from the SDSS to show that it is possible to determine accurate and precise photometric redshifts for quasars. They had reasonably good photometric redshift estimates using a combination of four colours to determine the CZR for each of three redshift bins, and then minimizing the χ^2 value between all four observed colours and the median colour for each redshift bin. Their χ^2 value is computed as

$$\chi^2_Z = \frac{\left[(u' - g') - (u' - g')_Z\right]^2}{\sigma^2_{u'-g'} + \sigma^2_{(u'-g')_Z}} + C_{gr} + C_{ri} + C_{iz} \tag{6.9}$$

where $(u' - g')$ is the measured colour for a quasar, $(u' - g')_Z$ is the median CZR at a given redshift, $\sigma_{(u'-g')}$ is the error in the $u' - g'$ colour, and $\sigma_{(u'-g')_Z}$ is the 1σ error width of the median CZR with respect to redshift. C_{gr}, C_{ri} and C_{iz} have the same form as the first term, but for the $g' - r'$, $r' - i'$, $i' - z'$ colours respectively. As with other techniques involving minimizing χ^2, this method can give multiple possible redshifts, leading to degeneracies in the estimate. After testing the method on a subset of the training set, they also applied their algorithm to a set of 642 quasars from the NASA/IPAC Extragalactic Database (NED) to obtain similar results.

Maddox et al. (2012) attempted to exploit the K-band excess shown by all quasars to construct

a more complete quasar catalogue for $1 \leq z \leq 3.5$. They used a combination of optical and near infrared (NIR) colours to take advantage of the K-band flux excess of quasars, and discriminate between stars and quasars at all redshifts. Their data are taken from the SDSS Data Release 7 and the United Kingdom Infrared Telescope Infrared Deep Sky Survey (UKIDSS) Data Release 4, with the measured redshifts being obtained from the catalogue published by Hewett and Wild (2010). From their initial dataset, objects with SDSS spectra were removed, as were SDSS quasar objects with no spectroscopic data. This was to facilitate follow-up spectroscopic observations. They further restricted their samples to objects with $z_{\rm phot} \geq 1.0$ to ensure that the remaining objects have long path lengths for the occurrence of intervening absorbers. For their entire survey region, which spans $-5 < \text{dec} < 15$ and $0 < \text{R.A.} < 360$, the magnitudes were limited to

$$[K, i] = \begin{cases} K \leq 16.5, i \leq 19.7 \text{ for R.A.} < 210 \\ K \leq 16.6, i \leq 22.0 \text{ for R.A.} > 210 \end{cases}$$

The final step in cleaning up their data was the removal by visual inspection of objects with poor photometric environment, such as nearby bright stars. The effective area surveyed was 567.0 deg^2, with a total of 324 objects observed. After being reduced using standard software made available by the Image Reduction and Analysis Facility at the National Optical Astronomy Observatory, the spectra were calibrated from observations of standard stars, resulting in an offset calibration being required, and then individually classified by hand into groups based on the type of object. This method allowed them to compile a quasar sample with fewer biases than optical selection.

Acknowledgements

I would first like to thank my thesis advisor, Dr Steven Curran, Senior Lecturer in Physics, School of Chemical and Physical Sciences, Victoria University of Wellington. Dr Curran provided guidance and help at each step of the process, and he was always forthcoming with comments and constructive criticism. He consistently allowed this paper to be my own work, while steering me in the right direction.

This research has made use of the NASA/IPAC Extragalactic Database (NED) which is operated by the Jet Propulsion Laboratory, California Institute of Technology, under contract with the National

Aeronautics and Space Administration. This research has also made use of NASA's Astrophysics Data System Bibliographic Services.

References

Allen, S. L. and Clarke, D. A. (2000). Three Instruments in Four Years: The UCO/Lick Data-Driven Toolkit. In Manset, N., Veillet, C., and Crabtree, D., editors, *Astronomical Data Analysis Software and Systems IX*, volume 216 of *Astronomical Society of the Pacific Conference Series*, page 339.

Allison, J. R., Sadler, E. M., Moss, V. A., Harvey-Smith, L., Heywood, I., Indermuehle, B. T., McConnell, D., Sault, R. J., and Whiting, M. T. (2016). Tracing the neutral gas environments of young radio AGN with ASKAP. *Astronomische Nachrichten*, 337:175.

Allison, J. R., Sadler, E. M., Moss, V. A., Whiting, M. T., Hunstead, R. W., Pracy, M. B., Curran, S. J., Croom, S. M., Glowacki, M., and Morganti, R. (2015). Discovery of H I gas in a young radio galaxy at z = 0.44 using the Australian Square Kilometre Array Pathfinder. *Monthly Notices of the Royal Astronomical Society*, 453(2):1249–1267.

Bessell, M. S. (2005). Standard photometric systems. *Annual Review of Astronomy and Astrophysics*, 43(1):293–336.

Bovy, J., Hogg, D. W., and Roweis, S. T. (2011). Extreme deconvolution: Inferring complete distribution functions from noisy, heterogeneous and incomplete observations. *Ann. Appl. Stat.*, 5(2B):1657–1677.

Bovy, J., Myers, A. D., Hennawi, J. F., Hogg, D. W., McMahon, R. G., Schiminovich, D., Sheldon, E. S., Brinkmann, J., Schneider, D. P., and Weaver, B. A. (2012). Photometric Redshifts and Quasar Probabilities from a Single, Data-driven Generative Model. *The Astrophysical Journal*, 749:41.

Buzzoni, A. (1989). Evolutionary Population Synthesis in Stellar Systems. I - A global approach. *Astrophysical Journal Supplement*, 71:817–869.

Buzzoni, A. (1995). Evolutionary Population Synthesis in Stellar Systems. II. Early-Type Galaxies. *Astrophysical Journal Supplement*, 98:69.

Buzzoni, A. (2002). Ultraviolet properties of primeval galaxies: Theoretical models from stellar population synthesis. *The Astronomical Journal*, 123(3):1188–1199.

Capak, P. L., Teplitz, H. I., Brooke, T. Y., Laher, R., and Science Center, S. (2013). Science Quality Mosaics and Source List for the Spitzer Heritage Archive. In *American Astronomical Society Meeting Abstracts #221*, volume 221 of *American Astronomical Society Meeting Abstracts*, page 340.06.

Carilli, C. L., Menten, K. M., Reid, M. J., Rupen, M. P., and Yun, M. S. (1998). Redshifted Neutral Hydrogen 21 Centimeter Absorption toward Red Quasars. *The Astrophysical Journal*, 494(1):175–182.

Ching, J. H. Y., Sadler, E. M., Croom, S. M., Johnston, H. M., Pracy, M. B., Couch, W. J., Hopkins, A. M., Jurek, R. J., and Pimbblet, K. A. (2017). The Large Area Radio Galaxy Evolution Spectroscopic Survey (LARGESS): survey design, data catalogue and GAMA/WiggleZ spectroscopy. *Monthly Notices of the Royal Astronomical Society*, 464:1306–1332.

Curran, S. J. and Whiting, M. T. (2012). Complete ionization of the neutral gas: why there are so few detections of 21 cm hydrogen in high-redshift radio galaxies and quasars. *The Astrophysical Journal*, 759(2):117.

Curran, S. J., Whiting, M. T., Murphy, M. T., Webb, J. K., Longmore, S. N., Pihlström, Y. M., Athreya, R., and Blake, C. (2006). A survey for redshifted molecular and atomic absorption lines - I. The Parkes half-Jansky flat-spectrum red quasar sample. *Monthly Notices of the Royal Astronomical Society*, 371:431–443.

Curran, S. J., Whiting, M. T., Wiklind, T., Webb, J. K., Murphy, M. T., and Purcell, C. R. (2008). A survey for redshifted molecular and atomic absorption lines - II. Associated HI, OH and millimetre lines in the $z > 3$ Parkes quarter-Jansky flat-spectrum sample. *Monthly Notices of the Royal Astronomical Society*, 391:765–784.

De Breuck, C., van Breugel, W., Stanford, S. A., Röttgering, H., Miley, G., and Stern, D. (2002). Optical and Near-Infrared Imaging of Ultra-Steep-Spectrum Radio Sources: The K-z Diagram of Radio-selected and Optically Selected Galaxies. *Astrophysical Journal*, 123:637–677.

Fernández-Soto, A., Lanzetta, K. M., Chen, H.-W., Levine, B., and Yahata, N. (2002). Error analysis of the photometric redshift technique. *Monthly Notices of the Royal Astronomical Society*, 330:889–894.

Glowacki, M., Allison, J. R., Sadler, E. M., Moss, V. A., and Jarrett, T. H. (2017). WISE data as a photometric redshift indicator for radio AGN. *Monthly Notices of the Royal Astronomical Society*. Submitted (arXiv:1709.08634).

Goulding, A. D., Alexander, D. M., Bauer, F. E., Forman, W. R., Hickox, R. C., Jones, C., Mullaney, J. R., and Trichas, M. (2012). 755(1):5.

Greenstein, J. L. and Schmidt, M. (1964). The Quasi-Stellar Radio Sources 3C 48 and 3C 273. *The Astrophysical Journal*, 140:1.

Gunn, J. E. (1971). On the Distances of the Quasi-Stellar Objects. *Astrophysical Journal Letters*, 164:L113.

Gwyn, S. D. J. and Hartwick, F. D. A. (1996). The Redshift Distribution and Luminosity Functions of Galaxies in the Hubble Deep Field. *Astrophysical Journal Letters*, 468:L77.

Han, B., Ding, H.-P., Yan-Xia, Z., and Zhao, Y.-H. (2016). Photometric redshift estimation for quasars by integration of knn and svm. *Research in Astronomy and Astrophysics*, 16(5):074.

Hatziminaoglou, E. e. (2010). HerMES: Far infrared properties of known AGN in the HerMES fields. *Astronomy and Astrophysics*, 518:L33.

Hazard, C., Mackey, M. B., and Shimmins, A. J. (1963). Investigation of the Radio Source 3C 273 By The Method of Lunar Occultations. *Nature*, 197:1037–1039.

Hewett, P. C. and Wild, V. (2010). Improved redshifts for SDSS quasar spectra. *Monthly Notices of the Royal Astronomical Society*, 405:2302–2316.

Jarque, C. M. and Bera, A. K. (1980). Efficient tests for normality, homoscedasticity and serial independence of regression residuals. pages 255–259.

Kauffmann, G., Heckman, T. M., Tremonti, C., Brinchmann, J., Charlot, S., White, S. D. M., Ridgway, S. E., Brinkmann, J., Fukugita, M., Hall, P. B., Ivezić, Ž., Richards, G. T., and Schneider,

D. P. (2003). The host galaxies of active galactic nuclei. *Monthly Notices of the Royal Astronomical Society*, 346:1055–1077.

Ma, C., Arias, E. F., Bianco, G., Boboltz, D. A., Bolotin, S. L., Charlot, P., Engelhardt, G., Fey, A. L., Gaume, R. A., Gontier, A.-M., Heinkelmann, R., Jacobs, C. S., Kurdubov, S., Lambert, S. B., Malkin, Z. M., Nothnagel, A., Petrov, L., Skurikhina, E., Sokolova, J. R., Souchay, J., Sovers, O. J., Tesmer, V., Titov, O. A., Wang, G., Zharov, V. E., Barache, C., Boeckmann, S., Collioud, A., Gipson, J. M., Gordon, D., Lytvyn, S. O., MacMillan, D. S., and Ojha, R. (2009). The Second Realization of the International Celestial Reference Frame by Very Long Baseline Interferometry. *IERS Technical Note*, 35.

Maddox, N., Hewett, P. C., Péroux, C., Nestor, D. B., and Wisotzki, L. (2012). The large area KX quasar catalogue - I. Analysis of the photometric redshift selection and the complete quasar catalogue. *Monthly Notices of the Royal Astronomical Society*, 424:2876–2895.

Marziani, P., Bon, E., Bon, N., del Olmo, A., Martínez-Aldama, M., D'Onofrio, M., Dultzin, D., Negrete, C., and Stirpe, G. (2019). Quasars: From the Physics of Line Formation to Cosmology. *Atoms*, 7:18.

Massarotti, M., Iovino, A., Buzzoni, A., and Valls-Gabaud, D. (2001). New insights on the accuracy of photometric redshift measurements. *Astronomy and Astrophysics*, 380:425–434.

McConnell, D. e. (2016). The Australian Square Kilometre Array Pathfinder: Performance of the Boolardy Engineering Test Array. *Publications of the Astronomical Society of Australia*, 33:e042.

Moors, J. J. A. (1986). The meaning of kurtosis: Darlington reexamined. *The American Statistician*, 40(4):283–284.

Netzer, H. (2013). *Preface*, pages xiii–xiv. Cambridge University Press.

Netzer, H. (2013). *The Physics and Evolution of Active Galactic Nuclei*.

Norris, R. P., Hopkins, A. M., Afonso, J., Brown, S., Condon, J. J., Dunne, L., Feain, I., Hollow, R., Jarvis, M., Johnston-Hollitt, M., and et al. (2011). Emu: Evolutionary map of the universe. *Publications of the Astronomical Society of Australia*, 28(3):215–248.

Norris, R. P., Salvato, M., Longo, G., Brescia, M., Budavari, T., Carliles, S., Cavuoti, S., Farrah, D., Geach, J., Luken, K., Musaeva, A., Polsterer, K., Riccio, G., Seymour, N., Smolčić, V., Vaccari, M., and Zinn, P. (2019). A Comparison of Photometric Redshift Techniques for Large Radio Surveys. *Astronomical Society of the Pacific, Publications*. Submitted (arXiv:1902.05188).

Oke, J. B., Cohen, J. G., Carr, M., Cromer, J., Dingizian, A., Harris, F. H., Labrecque, S., Lucinio, R., Schaal, W., Epps, H., and Miller, J. (1995). The keck low-resolution imaging spectrometer. *Publications of the Astronomical Society of the Pacific*, 107:375.

Polsterer, K. L., Zinn, P.-C., and Gieseke, F. (2013). Finding new high-redshift quasars by asking the neighbours. *Monthly Notices of the Royal Astronomical Society*, 428:226–235.

Prasad, B. K. (2007). S/n in optical astronomy. In Käufl, H. U. and Siebenmorgen, R., editors, *Modern Observational Techniques*, pages 7–8, Berlin. Springer.

Ramos Almeida, C. and Ricci, C. (2017). Nuclear obscuration in active galactic nuclei. *Nature Astronomy*, 1:679–689.

Richards, G. T., Weinstein, M. A., Schneider, D. P., Fan, X., Strauss, M. A., Berk, D. E. V., Annis, J., Burles, S., Laubacher, E. M., York, D. G., Frieman, J. A., Johnston, D., Scranton, R., Gunn, J. E., Ivezić, Ž., Nichol, R. C., Budavári, T., Csabai, I., Szalay, A. S., Connolly, A. J., Szokoly, G. P., Bahcall, N. A., Benítez, N., Brinkmann, J., Brunner, R., Fukugita, M., Hall, P. B., Hennessy, G. S., Knapp, G. R., Kunszt, P. Z., Lamb, D. Q., Munn, J. A., Newberg, H. J., and Stoughton, C. (2001). Photometric redshifts of quasars. *The Astronomical Journal*, 122(3):1151–1162.

Salvato, M., Ilbert, O., and Hoyle, B. (2019). The many flavours of photometric redshifts. *Nature Astronomy*, 3:212–222.

Sanders, D. B. (1999). Infrared emission from AGN. In Terzian, Y., Khachikian, E., and Weedman, D., editors, *Activity in Galaxies and Related Phenomena*, volume 194 of *IAU Symposium*, page 25.

Shields, G. A. (1978). Thermal continuum from accretion disks in quasars. *Nature*, 272:706–708.

Skrutskie, M. F., Cutri, R. M., Stiening, R., Weinberg, M. D., Schneider, S., Carpenter, J. M., Beichman, C., Capps, R., Chester, T., Elias, J., Huchra, J., Liebert, J., Lonsdale, C., Monet,

D. G., Price, S., Seitzer, P., Jarrett, T., Kirkpatrick, J. D., Gizis, J. E., Howard, E., Evans, T., Fowler, J., Fullmer, L., Hurt, R., Light, R., Kopan, E. L., Marsh, K. A., McCallon, H. L., Tam, R., Van Dyk, S., and Wheelock, S. (2006). The Two Micron All Sky Survey (2MASS). *Astrophysical Journal*, 131:1163–1183.

Strait, V. (2015). Separating Stars and Galaxies Probabilistically Based on Color. In *American Astronomical Society Meeting Abstracts #225*, volume 225 of *American Astronomical Society Meeting Abstracts*, page 336.51.

Su, M., Slatyer, T. R., and Finkbeiner, D. P. (2010). Giant Gamma-ray Bubbles from Fermi-LAT: Active Galactic Nucleus Activity or Bipolar Galactic Wind? *The Astrophysical Journal*, 724:1044–1082.

Urry, C. M. and Padovani, P. (1995). Unified Schemes for Radio-Loud Active Galactic Nuclei. *Astronomical Society of the Pacific, Publications*, 107:803.

Vanden Berk, D. E., Richards, G. T., Bauer, A., Strauss, M. A., Schneider, D. P., Heckman, T. M., York, D. G., Hall, P. B., Fan, X., Knapp, G. R., Anderson, S. F., Annis, J., Bahcall, N. A., Bernardi, M., Briggs, J. W., Brinkmann, J., Brunner, R., Burles, S., Carey, L., Castander, F. J., Connolly, A. J., Crocker, J. H., Csabai, I., Doi, M., Finkbeiner, D., Friedman, S., Frieman, J. A., Fukugita, M., Gunn, J. E., Hennessy, G. S., Ivezić, Ž., Kent, S., Kunszt, P. Z., Lamb, D. Q., Leger, R. F., Long, D. C., Loveday, J., Lupton, R. H., Meiksin, A., Merelli, A., Munn, J. A., Newberg, H. J., Newcomb, M., Nichol, R. C., Owen, R., Pier, J. R., Pope, A., Rockosi, C. M., Schlegel, D. J., Siegmund, W. A., Smee, S., Snir, Y., Stoughton, C., Stubbs, C., SubbaRao, M., Szalay, A. S., Szokoly, G. P., Tremonti, C., Uomoto, A., Waddell, P., Yanny, B., and Zheng, W. (2001). Composite Quasar Spectra from the Sloan Digital Sky Survey. *Astrophysical Journal*, 122:549–564.

Webster, R. L., Francis, P. J., Petersont, B. A., Drinkwater, M. J., and Masci, F. J. (1995). Evidence for a large undetected population of dust-reddened quasars. *Nature*, 375:469–471.

Weinstein, M. A., Richards, G. T., Schneider, D. P., Younger, J. D., Strauss, M. A., Hall, P. B., Budavari, T., Gunn, J. E., York, D. G., and Brinkmann, J. (2004). An empirical algorithm for

broadband photometric redshifts of quasars from the sloan digital sky survey. *The Astrophysical Journal Supplement Series*, 155(2):243–256.

Wilson, T., Rohlfs, K., and Huettemeister, S. (2000). Springer-Verlag Berlin Heidelberg.

Wright, E. L., Eisenhardt, P. R. M., Mainzer, A. K., Ressler, M. E., Cutri, R. M., Jarrett, T., Kirkpatrick, J. D., Padgett, D., McMillan, R. S., Skrutskie, M., Stanford, S. A., Cohen, M., Walker, R. G., Mather, J. C., Leisawitz, D., Gautier, III, T. N., McLean, I., Benford, D., Lonsdale, C. J., Blain, A., Mendez, B., Irace, W. R., Duval, V., Liu, F., Royer, D., Heinrichsen, I., Howard, J., Shannon, M., Kendall, M., Walsh, A. L., Larsen, M., Cardon, J. G., Schick, S., Schwalm, M., Abid, M., Fabinsky, B., Naes, L., and Tsai, C.-W. (2010). The Wide-field Infrared Survey Explorer (WISE): Mission Description and Initial On-orbit Performance. *Astrophysical Journal*, 140:1868–1881.

Wu, X. B., Zhang, W., and Zhou, X. (2004). Color-redshift relations and photometric redshift estimations of quasars in large sky surveys. *Chinese Journal of Astronomy & Astrophysics*, 4:17–27.

Zhu, G. and Ménard, B. (2013). The JHU-SDSS Metal Absorption Line Catalog: Redshift Evolution and Properties of Mg II Absorbers. *The Astrophysical Journal*, 770:130.