The SOUX AGN sample: SDSS–*XMM-Newton* optical, ultraviolet, and X-ray selected active galactic nuclei spanning a wide range of parameter space – sample definition

Daniel Kynoch [1,2,3]★ Jake A. J. Mitchell [3] Martin J. Ward,[3] Chris Done,[3] Elisabeta Lusso [4,5] and Hermine Landt[3]

[1]*School of Physics and Astronomy, University of Southampton, University Road, Southampton SO17 1BJ, UK*
[2]*Astronomical Institute, Czech Academy of Sciences, Boční II 1401, CZ-141 00 Prague, Czech Republic*
[3]*Centre for Extragalactic Astronomy, Department of Physics, Durham University, South Road, Durham DH1 3LE, UK*
[4]*Dipartimento di Fisica e Astronomia, Universita di Firenze, via G. Sansone 1, Sesto Fiorentino, I-50019 Firenze, Italy*
[5]*INAF – Osservatorio Astrofisico di Arcetri, L.go Enrico Fermi 5, I-50125 Firenze, Italy*

ABSTRACT
We assemble a sample of 696 type 1 active galactic nuclei (AGN) up to a redshift of $z = 2.5$, all of which have an SDSS spectrum containing at least one broad emission line (H α, H β, or Mg II) and an *XMM-Newton* X-ray spectrum containing at least 250 counts in addition to simultaneous optical/ultraviolet photometry from the *XMM* Optical Monitor. Our sample includes quasars and narrow-line Seyfert 1s: thus our AGN span a wide range in luminosity, black hole mass, and accretion rate. We determine single-epoch black hole mass relations for the three emission lines and find that they provide broadly consistent mass estimates whether the continuum or emission line luminosity is used as the proxy for the broad emission line region radius. We explore variations of the UV/X-ray energy index $\alpha_{\rm ox}$ with the UV continuum luminosity and with black hole mass and accretion rate, and make comparisons to the physical quasar spectral energy distribution model QSOSED. The majority of the AGN in our sample lie in a region of parameter space with $0.02 < L/L_{\rm Edd} < 2$ as defined by this model, with narrow-line type 1 AGN offset to lower masses and higher accretion rates than typical broad-line quasars. We find differences in the dependence of $\alpha_{\rm ox}$ on UV luminosity between both narrow/broad-line and radio-loud/quiet subsets of AGN: $\alpha_{\rm ox}$ has a slightly weaker dependence on UV luminosity for broad-line AGN and radio-loud AGN have systematically harder $\alpha_{\rm ox}$.

Key words: accretion, accretion discs – black hole physics – galaxies: active – galaxies: high-redshift – quasars: emission lines – quasars: supermassive black holes.

1 INTRODUCTION

Active galactic nuclei (AGN) are powered by mass accretion onto the supermassive black holes at the centres of galaxies. Most of the gravitational potential energy of the infalling material is liberated as radiation across the full electromagnetic spectrum. For typical, luminous AGN, the majority of their power emerges in the ultraviolet and optical bands as thermal radiation from a viscous accretion disc (Novikov & Thorne 1973; Shakura & Sunyaev 1973). The non-thermal, hard X-ray emission emerges via the Compton upscattering of soft seed photons from the disc in a hot ($T_e \sim 100$ keV) plasma above the disc (the 'corona' e.g. Haardt & Maraschi 1991). A strong relationship between the ultraviolet and X-ray emission of AGN was found in early studies (e.g. Tananbaum et al. 1979) and its nature has been explored further in more recent investigations (e.g. Lusso & Risaliti 2017; Salvestrini et al. 2019). The non-linear and remarkably tight relation between the ultraviolet and X-ray luminosities enables luminous AGN to be used as 'standardizable candles' to test cosmological models (e.g. Sacchi et al. 2022).

Clearly the disc and the corona are intimately related, but a detailed description of how has yet to be formulated.

For the majority of AGN, the peak of their energy output occurs in the far-UV/soft X-ray band which is unobservable because of Galactic extinction. This complicates the determination of the energetics and the disc–corona relationship. One means of circumventing this problem is to make studies of high- z AGN for which the spectral peak is redshifted to observable optical-UV wavelengths, as was done by Collinson et al. (2015). Making detailed studies of the optical and X-ray spectra of 11 sources, Collinson et al. (2015) were able to accurately determine the bolometric luminosities and accretion rates of the AGN, demonstrating estimates of the bolometric luminosity made by scaling from a single wavelength or narrow-band luminosity (e.g. at 2500 Å or 2–10 keV) are generally inaccurate (see also Netzer 2019). However, obtaining high-quality multiwavelength spectra for high-z AGN necessarily restricts the sample to a small number of bright quasars, thereby limiting the parameter space which can be explored. Another approach is to sample the far-UV and soft X-ray emission on either side of the peak and use a physical model (e.g. Done et al. 2012) to 'bridge the gap' and recover the intrinsic spectral energy distribution (SED). This method was adopted by Jin et al. (2012a), who explored the optical and X-ray spectroscopic

★ E-mail: d.kynoch@soton.ac.uk

properties of 51 AGN in a series of works (Jin et al. 2012a; Jin, Ward & Done 2012b,c). They showed that the broad-band SEDs of AGN could generally be fit with three components: a standard accretion disc, a hot corona, and an intermediate 'warm corona' responsible for the observed excess of soft X-ray emission (e.g. Magdziarz et al. 1998; Gierliński & Done 2004). The findings of Jin et al. (2012a,b), Jin, Ward & Done (2012c) have enabled the development and refinement of new, physical models of AGN SEDs (Kubota & Done 2018, 2019). These models have been constructed with reference to a small number of representative AGN SEDs. In the decade since the work of Jin et al. (2012a,b,c) the continuing SDSS (Blanton et al. 2017; Pâris et al. 2018) and *XMM-Newton* (Webb et al. 2020) surveys have greatly increased the number of AGN with quality optical/UV and X-ray spectra. We now have the opportunity to test the predictions made by current physical models against the observed properties of a large and diverse sample of AGN.

Achieving a more rigorous understanding of the accretion flow and disc–corona relationship will also enable us to better address another outstanding problem in astrophysics: the origin of relativistic jets. How AGN launch and power relativistic jets of outflowing matter is an active area of research. Progress towards a fuller understanding of the nature of jets and their relation to the accretion flow may be made via comparisons to accreting stellar-mass black holes (X-ray binaries: XRBs, e.g. Fender, Belloni & Gallo 2004) in which the disc–corona system likely regulates the launching of jets. Many studies have explicitly investigated the analogies between AGN and XRBs in terms of the accretion flow properties (e.g. Noda & Done 2018; Ruan et al. 2019; Arcodia et al. 2020) and the disc–jet coupling (e.g. Merloni, Heinz & di Matteo 2003; Körding, Jester & Fender 2006; Svoboda, Guainazzi & Merloni 2017; Fernández-Ontiveros & Muñoz-Darias 2021; Moravec et al. 2022). Generally these studies are supportive of a unified scheme of mass accretion and ejection across the mass scale, although many open questions have yet to be resolved. Studies of high accretion rate AGN in the local Universe (narrow-line Seyfert 1s: NLS1s) have found a relative deficit of radio-loud sources: only ≈5 per cent of NLS1s are radio-loud (e.g. Komossa et al. 2006; Rakshit et al. 2017) compared with ≈15–20 per cent of quasars (e.g. Kellermann et al. 1989), again suggesting a relationship between the accretion flow (the disc–corona configuration) and the presence of a radio jet.

Previous studies have mostly either investigated very large and diverse AGN samples without a detailed exploration of their spectroscopic properties (e.g. Svoboda et al. 2017) or have presented a thorough analysis of much smaller, focused samples of AGN with high-quality spectroscopic data (e.g. Jin et al. 2012a,b,c and Collinson et al. 2015, 2017). In this first paper in a series, we expand on the work of Jin et al. (2012a,b,c) and Collinson et al. (2015, 2017) and make use of recent optical, UV, and X-ray catalogues to compile a large sample of AGN (696 unique sources) with both optical and X-ray spectra in addition to broad multiwavelength coverage. This new sample has the advantage of being large enough to investigate population statistics whilst still having good quality multiwavelength data. It has been selected to probe a diverse range of AGN properties; as we will demonstrate, the sample spans several orders of magnitude in black hole mass and luminosity (accretion rate). The sample is a well-suited selection with which to perform detailed investigations of the evolution of the disc–corona system with black hole mass and accretion rate (Mitchell et al., submitted). In this paper we make key measurements of AGN properties and take a preliminary look at the evolution of the spectral energy distribution with both mass and luminosity. In Section 2 we describe the assembly of the sample; in Section 3 we obtain multiwavelength measurements for

the selected AGN; in Section 4 we determine black hole masses from optical/UV data; we investigate the UV/X-ray energy index α_{ox} and make comparisons with a physical model in Section 5; finally in Sections 6 and 7 we discuss our results and summarize our conclusions. To convert fluxes to luminosities we have assumed a flat ΛCDM cosmology with $H_0 = 70$ km s^{-1} Mpc^{-1}, $\Omega_m = 0.3$, and $\Omega_\Lambda = 0.7$.

2 SAMPLE SELECTION

2.1 The parent catalogues

2.1.1 Optical spectra: SDSS-DR14Q

The Quasar Catalogue of the Fourteenth SDSS Data Release (SDSS-DR14Q; Pâris et al. 2018) contains 526 356 quasars. Each quasar has been the subject of at least one optical spectroscopic observation recorded on or before 2016 December 5. The catalogue also contains SDSS photometric data from imaging observations. Multiwavelength data from *ROSAT*, *XMM-Newton*, *GALEX*, 2MASS, *WISE*, UKIDSS, and FIRST are included, where available. The *XMM-Newton* data are taken from the 3XMM-DR7 catalogue (an earlier version of the serendipitous survey catalogue: see below). 14 736 of the SDSS-DR14Q quasars (≈3 per cent) have an X-ray source detection within a 5 arcsec matching radius. Although the catalogue contains some X-ray data obtained from an *XMM-Newton* catalogue (3XMM-DR7), it does not contain any optical/UV photometry recorded with the *XMM-Newton* Optical Monitor (OM).

2.1.2 X-ray data: 4XMM-DR9

4XMM-DR9 is the first iteration of the fourth source catalogue compiled from the *XMM-Newton* Serendipitous Survey; this catalogue, released in 2019 December, is the ninth data release from the survey overall (Webb et al. 2020). The catalogue contains 810 795 X-ray detections pertaining to 550 124 unique sources from observations made between 2000 February 3 and 2019 February 26.

2.1.3 Optical and ultraviolet photometry: XMM-SUSS4.1

We take *XMM-Newton* optical/UV photometric data from the fourth *XMM-Newton* Serendipitous Ultraviolet Source Survey (SUSS: Page et al. 2012) catalogue, XMM-SUSS4.1. This catalogue, released in 2018 January, compiles data from observations made up to 2017 July. It contains 8 176 156 entries from 5 503 765 sources.

2.2 Initial source selection criteria

2.2.1 Quasars

We first establish which optical quasars have an X-ray source detection in the latest *XMM-Newton* data release. To do this, we cross-match the optical catalogue with the slimline version of the X-ray catalogue. The slimline 4XMM-DR9 catalogue contains just one row per unique X-ray source (rather than one row per detection as in the full catalogue). There are 550 050 X-ray sources in the catalogue, excluding 74 which are flagged as 'CONFUSED' (i.e. with a non-zero probability of being associated with more than one distinct source). Following Pâris et al. (2018), we choose a match radius of 5 arcsec and retain all of the best (nearest) matching X-ray sources. Fig. 1 shows the distribution of angular separations between the optical quasars and X-ray sources. Our cross-match returns 17 336
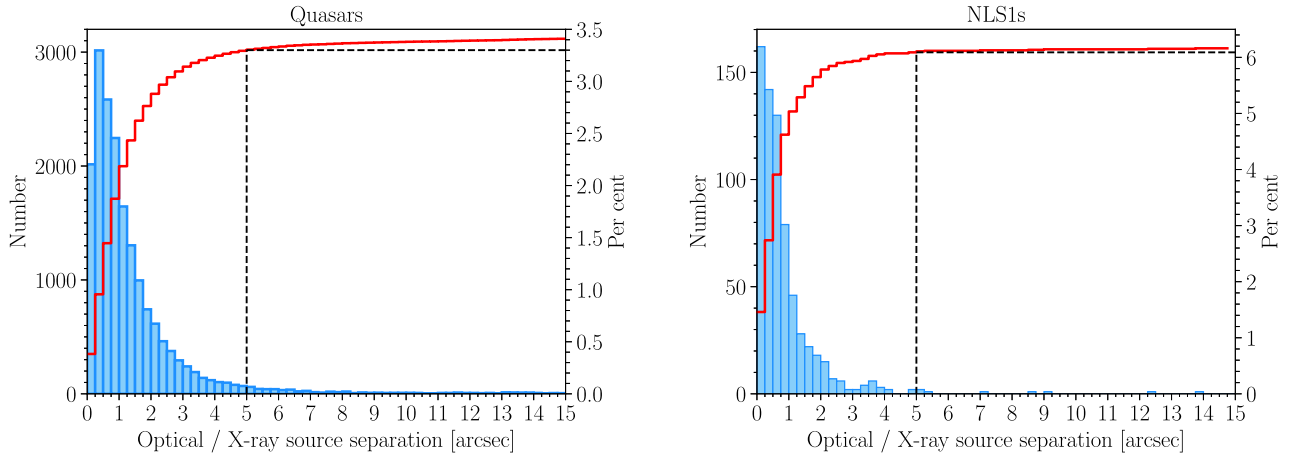
Figure 1. The left-hand ordinate axis indicates the number of matched quasars. The right-hand ordinate axis gives the percentage of all 526 356 quasars, with the cumulative distribution shown in red. To select our sample, we chose a matching radius of 5 arcsec which returns 17 336 quasars, approximately 3.3 per cent of all SDSS-DR14Q quasars. Right-hand panel: The same, for Rakshit et al. (2017) NLS1s. Of 11 101 SDSS NLS1s, 676 (6.1 per cent) have a clean X-ray detection within 5 arcsec.
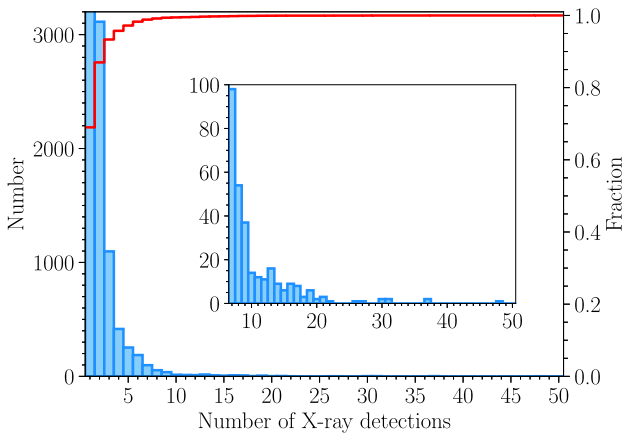
Figure 2. The number of clean X-ray detections for 17 299 X-ray detected optical quasars. 11 935 quasars (69 per cent) have only 1 clean X-ray detection. Only 109 have 10 or more detections. The red line shows the cumulative fraction, with the fraction indicated on the right-hand ordinate axis.

X-ray detected optical quasars (approximately 3.3 per cent of those in SDSS-DR14Q). We record the unique X-ray source ID ('SRCID') of each quasar. We do not remove broad absorption line (BAL) quasars from the optical catalogue initially, but will weed these out of our sample by visually inspecting the optical spectra.

Before performing a cross-match to the full X-ray catalogue, we clean it by excluding the sources which:

(i) have a poor source summary flag (SUM_FLAG ≥3);
(ii) are flagged as 'CONFUSED';
(iii) were observed with a high background.

The clean catalogue then contains 88 per cent of the total number of records, with 713 866 X-ray detections from 484 726 unique sources.

Using the X-ray source IDs, we find all clean observations of the optical quasars. We match all records in the cleaned, full 4XMM-DR9 catalogue to our pre-selected optical quasars. We find that 17 299 of the pre-selected 17 336 quasars have a clean X-ray detection. Of these, 5364 (31 per cent) were observed by *XMM-Newton* more than once; Fig. 2 shows the number of clean X-ray observations per optical quasar.

We then match with sources in the OM catalogue. This matching requires that the OM source is within 1 arcsec of the optical quasar coordinates and that the observation ID of the OM record matches that of the X-ray observation (i.e. our OM and EPIC data were recorded simultaneously). 4499 quasars meet these criteria of which 868 have more than one simultaneous OM-EPIC observation.

At $z \gtrsim 0.4$, the OM photometry may be compromised by the strong Ly α λ1216 UV emission line and Ly α forest absorption bluewards of 912 Å (rest frame). To determine which OM filters would be free of Ly α emission and absorption, we calculated the quasar redshifts for which the red wing of Ly α would fall just outside of the effective bandpass of each filter. Then, for each redshift bin we require an OM detection in one of the following:

(i) $z < 0.40$: any filter;
(ii) $0.40 \leq z < 0.55$: *V, B, U, UVW1,* or *UVM2* filters;
(iii) $0.55 \leq z < 0.90$: *V, B, U,* or *UVW1* filters;
(iv) $0.90 \leq z < 1.35$: *V, B,* or *U* filters;
(v) $1.35 \leq z < 1.95$: *V* or *B* filters;
(vi) $1.95 \leq z < 2.50$: *V* filter only.

1768 quasars have at least one useful OM photometry point.

We wish to obtain reasonable constraints on the X-ray spectral parameters of these quasars, for which we require relatively high-quality spectra. We therefore make a quality cut on the number of X-ray counts recorded by an EPIC detector in the 0.2–12 keV band. The number of sources as a function of X-ray counts is shown in Fig. 3. It can be seen in Fig. 3 that the counts distribution peaks $\gtrsim 200$ counts and that only 17 per cent of detections are made with 1000 counts or more. We chose to make a cut at 250 counts, which includes 46 per cent of the clean detections. For quasars with more than one X-ray observation, we retain the one with the greatest number of counts in any detector (so that we now have only one simultaneous EPIC and OM observation of each quasar). After applying this cut, we are left with 782 quasars.

To make an estimate of the black hole mass, we require spectral coverage of at least one of the emission lines H α, H β, or Mg II. H α or H β will be visible in the spectra of low-redshift sources and the Mg II emission line will be visible in the wavelength range of the BOSS spectrograph up to $z \lesssim 2.5$ (see Fig. 4). We therefore consider
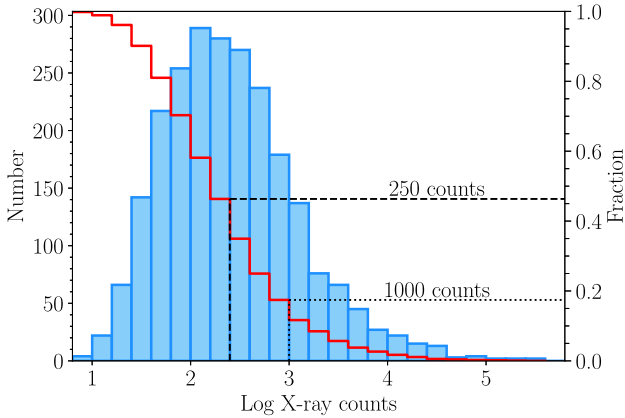
Figure 3. The X-ray counts distribution for 2374 'clean' X-ray observations with simultaneous *XMM*-OM photometry redwards of Ly α. For each observation, we take X-ray counts to be the greatest number of counts recorded by any of the three EPIC detectors (pn, MOS1, or MOS2). The left-hand ordinate axis gives the total number of observations and the right gives the fraction with a minimum number of counts (shown by the red line).
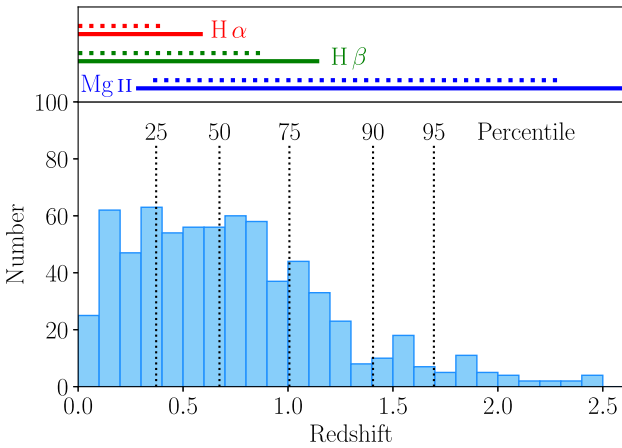
Figure 4. The redshift distribution of the 696 AGN in our sample. The black dotted lines indicate the 25th, 50th, 75th, 90th, and 95th percentiles of the distribution. The top panel shows the redshift range for which the centroids of prominent broad emission lines (H α, H β, and Mg II) are just visible in the wavelength coverage of the SDSS (dotted lines) and BOSS (solid lines) spectrographs.

only the SDSS quasars with $z \leq 2.5$. Following this redshift cut, our sample contains 768 quasars.

2.2.2 Narrow-line Seyfert 1s

Narrow-line Seyfert 1s (NLS1s) are a subset of unobscured AGN with relatively narrow broad lines. They are generally defined as optical type 1 AGN with FWHM(H β) $\lesssim$ 2000 km s^{-1}, weak [O III]λ5007 relative to H β and strong Fe II emission (e.g. Osterbrock & Pogge 1985; Goodrich 1989). NLS1s are typically found in spiral galaxies in the local Universe. Because Seyfert AGN are less luminous than quasars, many NLS1s are absent from the SDSS quasar catalogues. For example, of the 12 NLS1s in the sample of Jin et al. (2012a), only two are present in our selection of quasars. Nevertheless, NLS1s represent a unique and important region of parameter space to explore, and enable contrasts to be made with typical broad-line AGN and quasars (e.g. Jin et al. 2012a,b,c).

Rakshit et al. (2017) (hereafter R17) assembled a sample of 11101 NLS1s from the SDSS-DR12 spectroscopic data base, by analysing the spectra of 'QSO'-type sources and retaining those with FWHM(H β) $\leqslant$ 2200 km s^{-1} and flux ratio [O III]λ5007/H β < 3. We identify NLS1s with *XMM-Newton* OM and EPIC data by matching the R17 catalogue to the X-ray and OM serendipitous catalogues, following the method described in Section 2.2.1. Our criteria return 147 NLS1s: 63 new sources and 84 that were already included from the quasar catalogue.

2.3 Optical spectral fitting

Rakshit, Stalin & Kotilainen (2020) (hereafter R20) performed fits to the optical spectra of all SDSS DR14Q quasars using the publicly available PYTHON code PYQSOFIT (Guo, Shen & Wang 2018; Guo et al. 2019; Shen et al. 2019). For the quasars in our sample, we can take many emission line and continuum measurements from R20. However, the 63 NLS1s not present in the quasar catalogue have only optical spectral measurements from R17 who used a private, custom spectral fitting routine to obtain their measurements of NLS1s. These two routines differ in several respects. The broad emission lines were often modelled with a single Lorentzian profile in R17; Lorentzian profiles are not currently an option in PYQSOFIT and instead one or more Gaussians may be used to fit a line. R20 chose to use three Gaussians to model broad H β and H α, for example. The host galaxy subtraction, which is often required for the low-luminosity NLS1s, is also approached in different ways. R17 removed the host galaxy contribution by performing simple stellar population modelling whereas principal component analysis (PCA: e.g. Yip et al. 2004a,b) is employed in PYQSOFIT. Naturally these different methods will yield somewhat different results for the continuum and emission line parameters. Derived quantities such as the virial black hole mass and Eddington ratio will then also differ.

For consistency of the optical spectral measurements of our sample, we adapted PYQSOFIT to replicate the spectral fitting procedure of R20 and refit the spectra of the sources drawn from R17. We briefly summarize the adopted PYQSOFIT spectral fitting routine here. In this setup, the Schlegel, Finkbeiner & Davis (1998) dust map is queried to determine the line-of-sight extinction and the spectrum is dereddened using the extinction curve of Fitzpatrick (1999). The spectrum is then transformed to the rest-frame. The PCA method is used to decompose stellar and AGN emission with 5 and 20 eigenspectra used to construct the host galaxy and quasar contributions, respectively. The host galaxy subtracted continuum is then modelled with a power law plus Fe II templates and a functional Balmer continuum. Emission line complexes are then modelled separately in several broad windows:

(i) 6400–6800 Å: H α, the [N II] doublet ($\lambda = 6549, 6585$ Å), and the [S II] doublet ($\lambda = 6718, 6732$ Å);
(ii) 4640–5100 Å: H β, the [O III] doublet ($\lambda = 4959, 5007$ Å), and He II;
(iii) 2700–2900 Å: Mg II.

Following the fitting procedure of R20, the broad components of H α and H β and He II are each modelled with three Gaussian components; broad Mg II is modelled with two Gaussians. The broad Gaussians have a minimum full-width at half-maximum (FWHM) of 900 km s^{-1} and maximum velocity offset $\pm$3000 km s^{-1} from their centre. Narrow Gaussians are used to model forbidden lines and the narrow components of the permitted lines; these are allowed a width in the range $100 \leq$ FWHM ≤ 900 km s^{-1}. [O III]$\lambda\lambda$4959, 5007 additionally has Gaussian wings that may be broader than the narrow

cores (in the range $400 \leq$ FWHM ≤ 2000 km s^{-1}). The velocity offsets of the narrow lines are not allowed to exceed ± 1000 km s^{-1}. The flux ratios of the [O III] and [N II] doublets are fixed to their theoretical value 1:3. In each emission line complex the widths and offsets of the narrow lines are tied together. Uncertainties on the emission line and continuum parameters are estimated using the Monte Carlo method, resampling each spectrum 50 times. Following the procedures of R20 and Calderone et al. (2017), we added a quality flag for each measured emission line and continuum point; the flags are listed in Appendix C.

PYQSOFIT uses optical galaxy eigenspectra to perform the host galaxy subtraction; these eigenspectra do not extend into the UV and therefore the fitted spectrum is truncated to match the wavelength range of the eigenspectra (≈ 3500–9000 Å). Consequently, there are no measurements of the UV continuum and emission lines for sources in which host galaxy subtraction was performed in the optical. To obtain the UV spectral measurements for these AGN, we checked for spectra that contained a Mg II line that was not initially measured and then fit the short-wavelength part of these spectra ($\lambda_{rest} = 1000$–4000 Å) without performing host-galaxy subtraction. The measurements for Mg II and the UV continuum were added to our catalogue, and we have added a flag ('MGII_SEP') to indicate the AGN for which the Mg II region was fitted separately from the optical region.

2.3.1 Identification of broad- and narrow-line type 1 AGN

To ease the comparison between AGN with relatively narrow and broad permitted emission lines, we assigned each AGN a linewidth flag ('B' for broad-line type 1 AGN; 'N' for narrow-line type 1 AGN). The determination is based on the fitted width of H β where it is available, Mg II for the higher redshift sources without H β, and H α only if neither H β or Mg II were measured. (Spectroscopic coverage of these emission lines as a function of redshift is shown in Fig. 4.) We set the division between the subsets at the usual 2000 km s^{-1}. The differences in fitting routine and the division between narrow- and broad-line AGN mean that our classifications differ from those of R17 (those authors measured lines by fitting a Lorentzian profile and chose a limit of FWHM[H β] ≤ 2200 km s^{-1}). Consequently, not all AGN drawn from R17 are flagged 'N' in our sample. Typically, the broad line FWHMs measured by PYQSOFIT are slightly broader than those measured by R17. However, the differences in linewidth measured by the two routines are relatively minor: the median difference in FWHM(H β) is $\approx +100$ km s^{-1} and for approximately two-thirds of the AGN from the NLS1 catalogue in our sample, FWHM(H β) as measured by PYQSOFIT is within ± 350 km s^{-1} of the width measured by R17 (350 km s^{-1} is the median uncertainty on FWHM[H β] from PYQSOFIT). We note that the narrow-line type 1 AGN in our sample are not necessarily NLS1s by the standard definition described in Section 2.2.2: we have not always used broad H β as the characteristic line width and we have made no assessment of the strengths of [O III] $\lambda5007$ or the optical [Fe II] emission.

2.4 Quality checks and selection of the final SOUX sample

We performed a visual inspection of all 831 optical spectra and removed any sources which had low signal to noise or any absorption features particularly affecting the line regions from which black hole mass measurements were derived. In addition any objects with a diffuse emission flag in their X-ray spectra were removed from the sample. Following this quality control, our final sample consists of 696 AGN. [1] Of these, 636 are drawn from SDSS-DR14Q, 60 unique sources are added from R17, and 72 are common to both catalogues.

3 MULTIWAVELENGTH MEASUREMENTS

We will present a detailed study of the broad-band SEDs of our AGN in a future paper (Mitchell et al., submitted). However, the SED shape can be characterized by two common parameters: the radio loudness and UV-X-ray energy index. The radio-loudness parameter indicates the power of AGN outflows relative to the accretion flow, with radio-loud AGN exhibiting powerful jets (Section 3.3). The optical luminosity at 4400 Å and radio luminosity at 5 GHz are required to calculate the radio-loudness. The UV-X-ray energy index (α_{ox}: Section 5) indicates the relative power of the UV-emitting accretion disc to the X-ray emitting corona. To calculate α_{ox} we require a measurement of the UV luminosity at 2500 Å and the X-ray luminosity at 2 keV. In this section, we describe how the measurements of these quantities were obtained.

3.1 UV and optical luminosities

For sources at $z \gtrsim 0.5$ the rest-frame monochromatic luminosity at 2500 Å is available directly from the SDSS optical spectrum. For nearer sources, this wavelength falls outside the SDSS spectrograph range, but photometry covering it may be available from either *XMM* OM or SDSS. For AGN at lacking spectroscopic coverage of 2500 Å we first attempt obtain an estimate from the *XMM* OM photometry. We determine which OM bands sample 2500 Å in the rest frame and from these we calculate the luminosity from the flux in the band with an effective wavelength closest to rest frame 2500 Å. If there is no *XMM* OM coverage of 2500 Å we take the PSF flux from the appropriate Sloan filter (either the *u* or *g* band). We deredden the photometry, taking the line-of-sight colour excess from the Schlegel et al. (1998) dust maps as was done for the spectral analysis and convert the fluxes to rest-frame luminosities. Finally, L_{2500} is obtained from the optical spectrum of 421 AGN, from *XMM* OM photometry of 237 AGN and from SDSS photometry of 27 AGN; overall, we have a measurement of L_{2500} for 685 AGN (98 per cent of the full sample). The AGN lacking a L_{2500} measurement are all at the low-redshift end of our sample ($z < 0.21$).

The optical continuum at rest-frame 4400 Å was not measured by R20 or R17 but, similarly to L_{2500}, L_{4400} can be measured for the majority of our AGN either from the SDSS spectrum, or from the Sloan or *XMM* OM photometry. When making a measurement from the optical spectrum, we avoid contamination of the continuum flux measurement by the blended H γ, [O III], and Fe II emission lines by performing a linear interpolation between line-free continuum windows at 4200 and 5100 Å, where possible. In spectra where these two windows are not present (or are too near the end of the spectrum to be reliable), we measure the 4400 Å flux directly. For AGN lacking optical spectral coverage of 4400 Å we took the flux from the photometric band containing rest-frame 4400 Å, corrected for Galactic reddening and converted the flux to a rest-frame luminosity, as was done for the L_{2500} measurements. In our final sample of 696 AGN, 482 have L_{4400} measured by interpolating the spectral continuum; 111 from a direct measurement of the 4400 Å spectral

---

[1]In this work, we consider only one optical spectrum per source: that listed as the 'science primary' spectrum at the time of the catalogue compilation. However, 116 AGN have multiple spectra recorded on different nights; we describe these in Appendix B.

flux; 38 estimates from SDSS photometry; and 65 have no 4400 Å coverage in either the spectrum or photometry. None of the 65 sources lacking L_{4400} from the SDSS spectrum or photometry had coverage in the *XMM* OM data because they were higher-z AGN and the OM filters do not cover long enough wavelengths to contain the rest-frame 4400 Å. The AGN lacking a L_{4400} measurement are all at the high-redshift end of our sample ($z > 1$).

The mixture of spectroscopically and photometrically derived UV and optical luminosities is not ideal. Whilst the *XMM* OM photometry is contemporaneous with the X-ray data, broad-band photometry is susceptible to contamination from non-continuum emission, particularly at lower redshifts. The SDSS and *XMM* data are not contemporaneous, so AGN variability will introduce some scatter between spectroscopic and photometric fluxes. Additionally, aperture effects may also introduce systematic differences. In Appendix A we assess the effects of mixing different measures of these luminosities. We find that on average the differences between spectroscopic and photometric luminosities are small, with $\langle \Delta \log (L_{4400}) \rangle = -0.03 \pm 0.13$ and $\langle \Delta \log (L_{2500}) \rangle = -0.02 \pm 0.16$. As Fig. A1 shows, the spectroscopic and photometric luminosity measures are approximately equal across several orders of magnitude in luminosity, although there is substantial scatter. We discuss the potential impact on our results in more detail in Appendix A and Section 6.3.
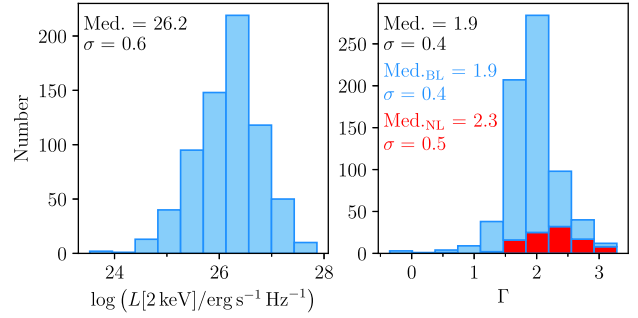
3.2 X-ray luminosities

A detailed analysis of the X-ray spectra of our sources is beyond the scope of this paper but will be presented in future work. However, it is possible to estimate the X-ray luminosities from data provided in the DR9 catalogue. Our sources span a wide range of redshifts and therefore the EPIC fluxes reported for various (observed) energy bands are not directly comparable since they sample different parts of the emitted spectrum.

To calculate the monochromatic X-ray flux at rest-frame 2 keV, we follow a similar approach to that of Lusso & Risaliti (2016) (see section 2.1 of that paper). We determine the EPIC flux in soft (0.5–2 keV) and hard (2–12 keV) energy ranges. An approximate correction for Galactic photoelectric absorption is applied to the X-ray fluxes.[2] From these band fluxes, we estimate the monochromatic fluxes at 1 and 3.5 keV, assuming a photon index $\Gamma = 1.7$ in each energy band. We then translate these fluxes at observed energies to luminosities at rest-frame frequencies in $\log(\nu)$–$\log(\nu L_\nu)$ form. Using these two points, the rest-frame monochromatic 2 keV luminosity is calculated by linear interpolation (for sources with $z < 1$) or extrapolation (for sources with $z > 1$). An estimate of the X-ray photon index Γ is also made by computing the gradient between the absorption-corrected monochromatic fluxes at 1 and 3.5 keV.[3]

In Fig. 5 we show the distributions of $\log(L_{2\,\mathrm{keV}})$ and Γ. Our sample spans approximately 4 dex in X-ray luminosity, with median $\log(L_{2\,\mathrm{keV}}) = 26.2$ and $\sigma = 0.6$. The median estimated photon index is $\Gamma = 1.9$ with $\sigma = 0.4$. 673 of the AGN (97 per cent) have 1
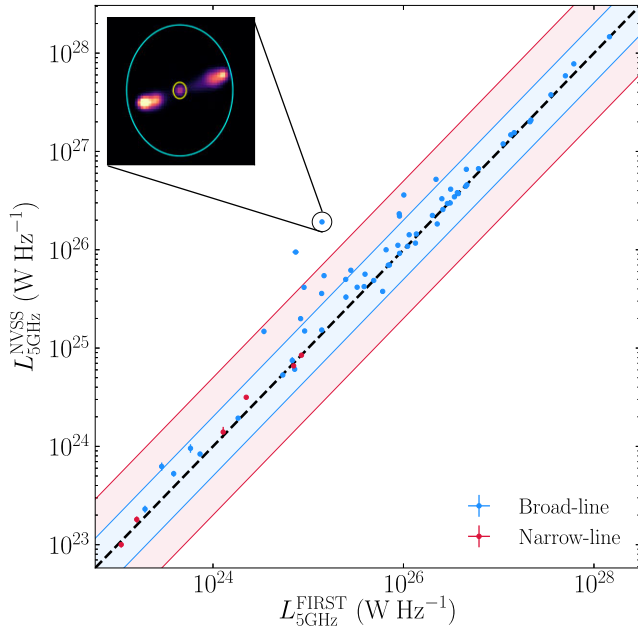

Figure 5. Distributions of the monochromatic luminosity at 2 keV in the rest frame (left-hand panel) and the estimated X-ray photon index Γ (right-hand panel). In the right-hand panel the subsample of 101 AGN with relatively narrow permitted lines ('NL') are shown in red, and the 595 broad-line sources ('BL') in blue.

$\leq \Gamma \leq 3$. Less than 2 per cent of sources have $\Gamma < 1$, perhaps indicating absorption of soft X-rays in the AGN or host galaxy. The Γ distribution of the 101 narrow-line AGN is different from that of the 595 broad-line AGN. The narrow-line AGN have a median $\Gamma = 2.3$ and $\sigma = 0.5$, indicating that their X-ray spectra are generally softer than typical quasars. Kolmogorov–Smirnov and Anderson–Darling tests both indicate that the photon indices of the narrow-line and broad-line AGN are drawn from different distributions. For a null hypothesis that the two samples are drawn from the same distribution we determine a KS statistic 0.41 and $p = 8 \times 10^{-14}$ and AD statistic 44 and $p < 0.001$.

3.3 Radio properties

Our sample of 696 AGN was cross-matched with both FIRST (Becker, White & Helfand 1995) and NVSS (Condon et al. 1998), using a matching radius of 10 arcsec that has been shown to have a false association rate of only 0.2 per cent with FIRST (Lu et al. 2007). Both FIRST and NVSS sample the sky at 1.4 GHz, with beam-widths of 5.6 and 45 arcsec, respectively. 679 of our AGN were within the FIRST footprint, of which we found a matching radio source for 107. 73 AGN also have an NVSS radio detection (all of which were FIRST-detected). We converted the 1.4 GHz fluxes to rest-frame 5 GHz luminosities using the method of Alexander et al. (2003), assuming a common radio spectral index (α_R) of 0.6. A comparison of the NVSS and FIRST luminosities for the 73 of objects matched with both catalogues can be seen in Fig. 6. It is clear that for the majority of sources the FIRST and NVSS luminosities match to within a factor of 2. The population is skewed towards higher luminosities in NVSS due to its larger beam width; this is most pronounced in objects which possess extended morphologies as can be seen by the radio map of the FR-II source SDSS J151443.07+365050.4 (Fig. 6, inset).

For the 101 radio-detected AGN with a L_{4400} measurement, we calculated the radio loudness according to the common definition of $L_{5\,\mathrm{GHz}}/L_{4400}$ (Kellermann et al. 1989). Fig. 7 displays the distribution of the 5 GHz luminosities of the radio-detected AGN. The distribution has been separated into distinct populations: the left-hand panel shows the radio-loud and radio-quiet subsets and the right-hand panel shows the narrow-line and broad-line subsets (see Section 2.3.1). For many undetected radio sources, we can still estimate whether they are likely to be radio-quiet by considering the FIRST detection threshold as an upper limit to their radio flux. Assuming a FIRST survey flux density threshold of 1 mJy (Becker et al. 1995), we calculate a curve

---

[2]The line-of-sight Galactic neutral hydrogen column density is obtained from the maps of the Leiden/Argentine/Bonn (LAB) Galactic H I Survey (Kalberla et al. 2005) using the PYTHON tool GDPYC (https://pypi.org/project/gdpyc/).

[3]We note that our results do not depend very strongly on the assumption of a standard $\Gamma = 1.7$ in the soft and hard bands to estimate the 1 and 3.5 keV fluxes. If we compare the derived distributions of Γ and $\log(L_{2\,\mathrm{keV}})$ that result from instead assuming $\Gamma = 1.5$ and 1.9, we find that the mean Γ differ by ≈ 0.1 and the mean $\log(L_{2\,\mathrm{keV}})$ differ by only ≈ 0.01 dex.

Figure 6. A comparison of the rest-frame luminosities at 5 GHz derived from FIRST and NVSS fluxes at 1.4 GHz, as described in the text. The dashed line marks the one-to-one relation, the shaded blue region being a factor of 2 deviation, and the pink shaded region a factor of 5. The top left-hand inset figure shows the radio map for the FR-II source SDSS J151443.07+365050.4, which has a much higher luminosity in NVSS than FIRST, as a result of the larger beam size of the former survey. The 45 arcsec beamwidth of NVSS is in blue, and the 5.6 arcsec beamwidth of FIRST is in yellow.
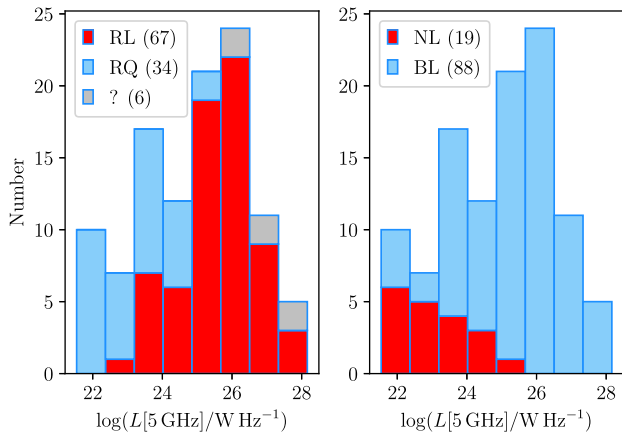
Figure 7. For the 107 FIRST radio detected AGN in our sample we show the distribution of 5 GHz rest-frame radio luminosities derived from the observed 1.4 GHz fluxes, as described in the text. In the left-hand panel, the distribution is divided into radio-loud and quiet ('RL' and 'RQ') subsamples in addition to six sources lacking an optical L_{4400} measurement with unknown radio loudness ('?'). In the right-hand panel the distribution is divided into narrow-line and broad-line ('NL' and 'BL') subsamples on the basis of the relative widths of the permitted optical emission lines.

corresponding to the lowest rest-frame 5 GHz luminosity we would expect for a source to be detected; we show this curve in Fig. 8. For radio undetected sources we plot the radio luminosity corresponding to $R = L_{5\,\mathrm{GHz}}/L_{4400} = 10$ (the luminosity above which they would be considered radio-loud). Radio-undetected AGN with luminosities more than 3σ above the curve (206 AGN) are considered to be radio-quiet, since we would expect them to have been radio-detected if
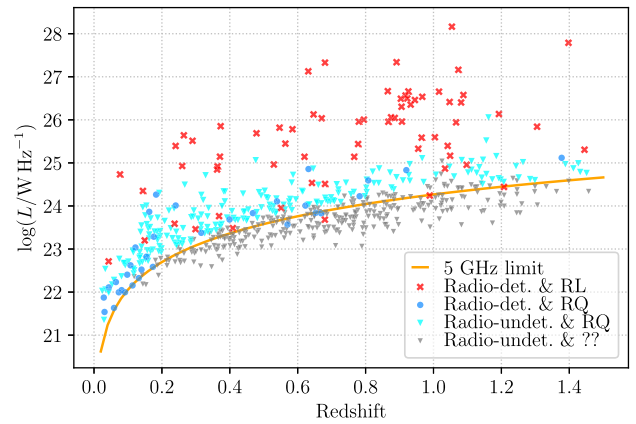
Figure 8. The orange curve shows the limiting rest-frame 5 GHz luminosity for a source to be detected by FIRST at 1.4 GHz, assuming a survey flux density threshold of 1 mJy and radio spectral index $\alpha = 0.6$. Radio-detected and radio-loud (radio-quiet) sources are shown with red crosses (blue circles). For radio-undetected sources, we plot the luminosity corresponding to $R = L_{5\,\mathrm{GHz}}/L_{4400} = 10$. Undetected sources more than 3σ above the curve (cyan triangles) are considered radio-quiet. Undetected sources within 3σ of the curve and below (grey triangles) could be either radio-quiet or radio-loud.

they were radio-loud. However, as Fig. 8 illustrates, there remain 304 AGN below the curve which are too faint for us to determine whether they are either radio-loud or radio-quiet. Following this exercise, we determine that our sample contains 68 radio-loud AGN, 239 radio-quiet AGN, and a total of 389 AGN for which we are unable to determine the radio-loudness. The radio-loudness classifications are summarized in Table 1 and illustrated in Fig. 9.

Overall, ∼10 per cent of our sample with both FIRST and 4400 Å coverage is found to be radio-loud, similar to previous estimates for bright quasars (Kellermann et al. 1989; Miller, Peacock & Mead 1990). However, we caution that this radio-loud fraction is a lower limit because of the large number of sources in our sample without any radio-loudness determination. We tentatively find differences in the radio-loud fractions of broad-line and narrow-line type 1 AGN. Whilst narrow-line AGN make up ≈15 per cent of the sample overall, they constitute only ≈9 per cent of the radio-loud sources. This is consistent with previous findings that the radio-loud fraction of NLS1s is lower than that of typical broad-line AGN and quasars (Komossa et al. 2006; Zhou et al. 2006; Rakshit et al. 2017). Again, selection biases and incomplete data mean it is not possible to draw firm conclusions from our sample.

4 BLACK HOLE MASSES

The widths and luminosities of the broad permitted emission lines, as well as the continuum luminosities can be used to derive the masses of the black holes in our AGN. In this Section, we assess the broad emission line properties and calculate the black hole masses using new scaling relations.

4.1 Broad emission line properties

We compare the broad line widths for sources with spectra containing multiple lines. The line width comparisons are shown in Fig. 10. We find that H β is slightly broader than both H α and Mg II. 207 spectra contain both H α and H β and from the median of the logarithmic distribution we find that H β is on average ≈10 per cent broader than H α. There is greater scatter between the widths of H β and Mg II,

**Table 1.** Radio-loudness of the sample.

Subset	N	Per cent
No radio coverage (Outside FIRST footprint)	20	2.8 per cent of the full sample
No 4400 Å coverage	65	9.3 per cent of the full sample
Radio-detected and radio-loud	68	11.1 per cent of AGN with radio and 4400 Å coverage
Radio-detected and radio-quiet	33	5.4 per cent of AGN with radio and 4400 Å coverage
Radio-undetected and radio-quiet	206	33.7 per cent of AGN with radio and 4400 Å coverage
Radio-undetected, undetermined radio-loudness	304	49.8 per cent of AGN with radio and 4400 Å coverage
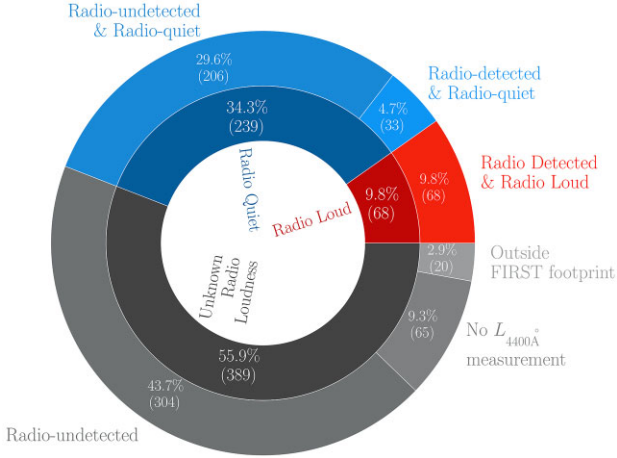
Figure 9. Radio detection and radio loudness fractions for the AGN in our sample.
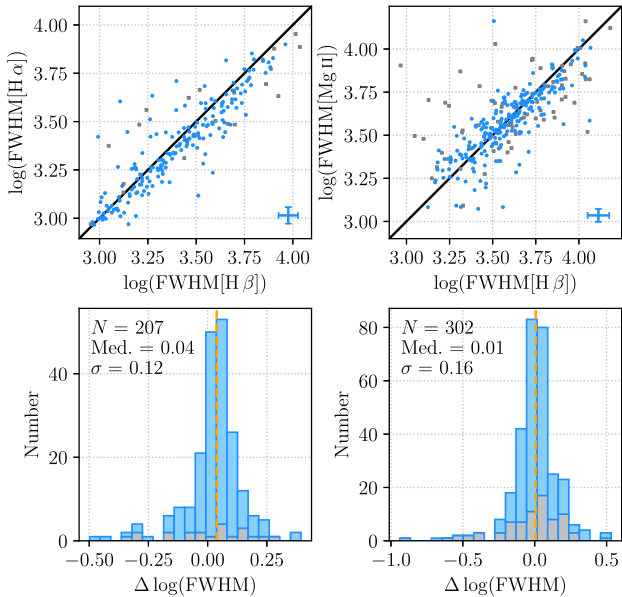
Figure 10. A comparison of the full widths at half maxima of H α, H β, and Mg II. Upper panels show the values for sources with at least two lines in their spectra. Representative errorbars are shown in the lower right-hand corners of the plots. The lower panels show the logarithm of the line width ratios.

and we do not find strong evidence of a systematic difference in the line widths (H β is on average only ≈ 2 per cent broader than Mg II) as measured in 302 spectra. Qualitatively, these results somewhat are similar to those of previous studies, e.g. Osterbrock & Shuder (1982) found H β to be ≈ 16 per cent broader than H α in Seyfert galaxies;
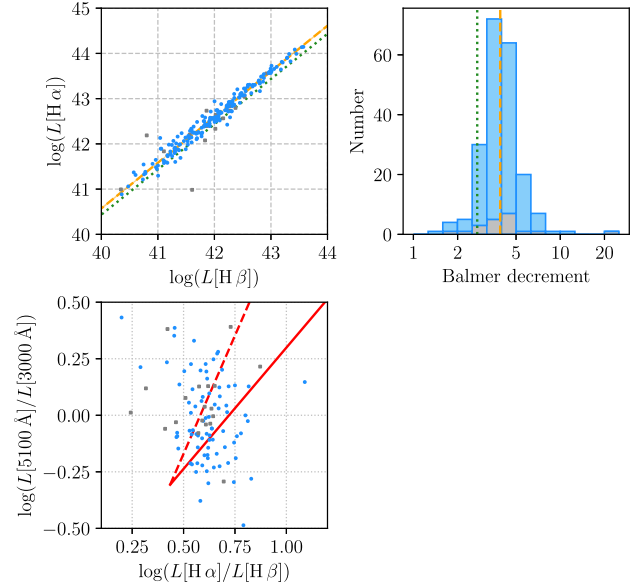
Figure 11. Top left-hand panel: A comparison of the H α and H β broad line luminosities for low-z sources with spectra containing both lines. Top right-hand panel: The distribution of the Balmer decrements of these sources. The orange dashed lines indicate the best-fitting linear regression, corresponding to $L_{H\alpha}/L_{H\beta} = 3.9$. The green dotted lines correspond to $L_{H\alpha}/L_{H\beta} = 2.72$ (Gaskell 2017). Bottom: Balmer decrements versus ratios of the monochromatic continuum luminosities at 5100 and 3000 Å. The solid red line shows the expected trend from the Gaskell et al. (2004) AGN reddening curve; the predicted trend for the Milky Way reddening curve of Cardelli, Clayton & Mathis (1989) is shown by the dashed line, for comparison. Data with good quality measurements of all quantities are shown as blue circles; the grey squares represent data in which at least one measurement has a quality flag raised. All luminosities have been derived from the SDSS optical spectrum and are therefore contemporaneous.

Mejía-Restrepo et al. (2016) (hereafter MR16) also found H β to be broader than both H α and Mg II.

In Fig. 11 we show the ratio of the H α to H β luminosities (the Balmer decrement). For the logarithmic mean value, we find $\langle H\alpha/H\beta \rangle = 3.9$ which is higher than the Case B value 2.74 (Osterbrock & Ferland 2006) and the intrinsic value 2.72 ± 0.04 suggested by Gaskell (2017), based on observations of blue AGN. This Balmer decrement is also greater than the average found in previous studies of type 1 AGN: $\langle H\alpha/H\beta \rangle = 3.16$ (Lu et al. 2019); $\langle H\alpha/H\beta \rangle = 3.06$ (Dong et al. 2008); $\langle H\alpha/H\beta \rangle = 3.45$ (La Mura et al. 2007). From the Shen et al. (2011) Catalogue of SDSS DR7 Quasar Properties, we determine $\langle H\alpha/H\beta \rangle = 3.55$ with $\sigma = 0.14$ dex. Taken as a reddening indicator, this large average Balmer decrement implies a high degree of intrinsic dust extinction in our sample of AGN: on average the colour excess $E(B - V) \approx 0.4$ mag. However, if a large Balmer decrement is due to reddening, we would expect the
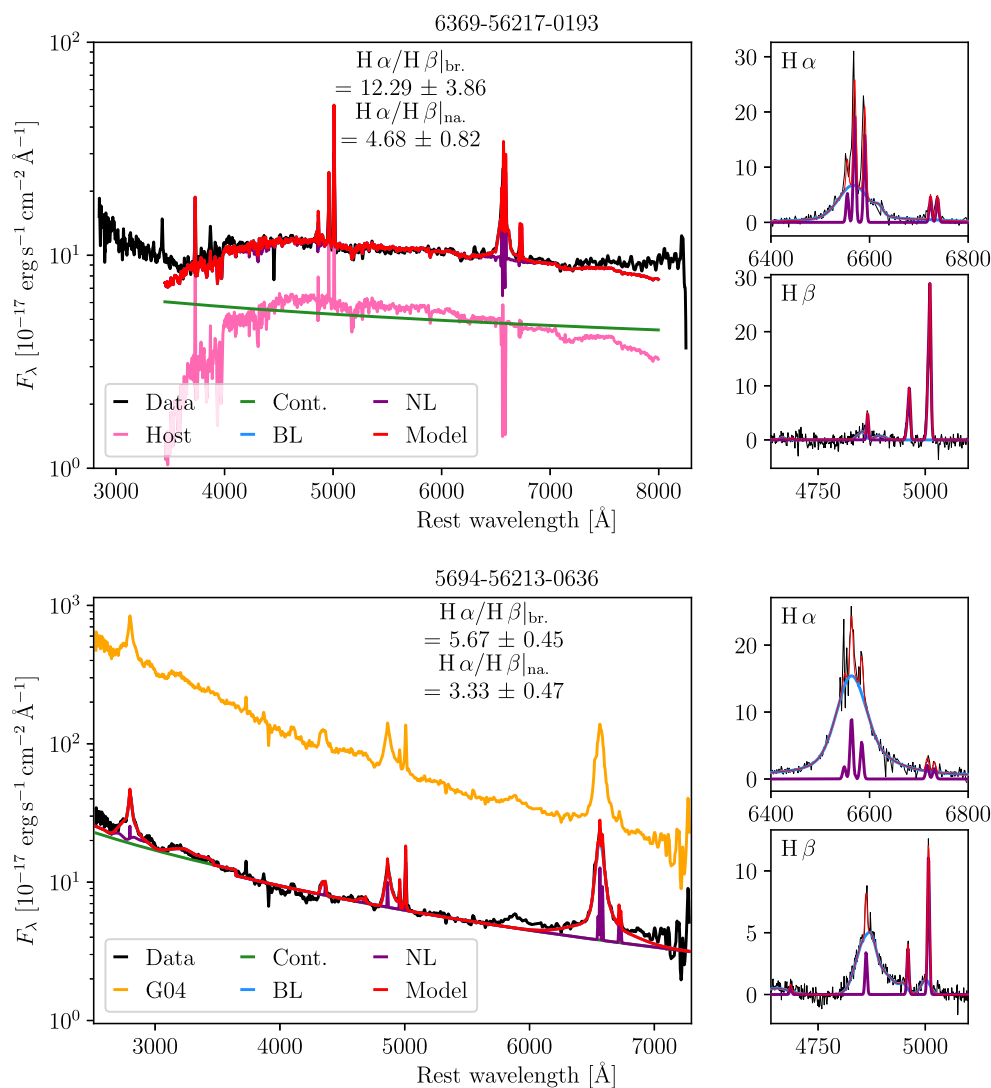
Figure 12. Example spectra with high broad Balmer line decrements. The $H\alpha/H\beta$ flux ratios of both broad and narrow lines are given in the inset text. Each spectrum (shown in black) has been corrected for Galactic reddening and transformed to the rest frame. The inset legend indicates the components of the total PYQSOFIT model (red): the host galaxy emission (pink); the AGN continuum (green) and selected, modelled broad and narrow emission lines ('BL' and 'NL', coloured blue and purple, respectively). In the lower panel, we also show in orange the spectrum 5694-56213-0636 additionally dereddened with the Gaskell et al. (2004) curve assuming an $A_v = 5.3$ derived from the broad Balmer line decrement.

optical-UV continuum to be similarly affected such that AGN with a higher Balmer decrement will have a correspondingly greater value of L_{5100}/L_{3000}. Assuming an intrinsic (Case B) broad line Balmer decrement of 2.72 (Gaskell 2017) and an AGN continuum slope $\alpha = 1/3$, we have used the AGN reddening curve of Gaskell et al. (2004) to calculate Balmer decrements and the corresponding continuum luminosity ratios for increasing dust extinction: this trend line is shown in the lower panel of Fig. 11. The Gaskell et al. (2004) AGN curve is much flatter in the UV than Milky Way curves; for comparison, Fig. 11 also shows the expected luminosity ratio trend derived from the Cardelli et al. (1989) Galactic reddening curve. Each data point in the figure has been obtained from a single spectrum; it can be seen, that these data do not follow either trend.

We further investigated whether spectra with large Balmer decrements may be reddened by visually inspecting those with ($H\alpha/H\beta > 5$): two examples are shown in Fig. 12. Whilst reddening is a plausible explanation for the large Balmer decrement in the spectrum 6369-

56217-0193, visual inspection of the spectrum 5694-56213-0636 does not indicate any obvious reddening. Since the UV continuum is not highly extinguished, if the AGN is reddened its reddening curve must flatten towards shorter wavelengths. Gaskell et al. (2004) presented such a curve for AGN, and we have used this curve to modify this spectrum using an $A_v = 5.3$ (estimated from the broad line Balmer decrement assuming the intrinsic line ratio is 2.72). The modified spectrum is clearly much brighter than the one used in our analysis, but is still quasar-like so we cannot definitively rule out intrinsic reddening. We discuss this point further in Section 6.1.

We have computed the best-fitting linear regressions between various luminosities using the package LINMIX (Kelly 2007). The results are shown in Fig. 13 and the values are quoted in Table 2. The Balmer lines show an almost 1:1 correspondence with the 5100 Å continuum luminosity, whereas the relation with Mg II is slightly shallower with a gradient $m = 0.87 \pm 0.03$. The relationships of both $H\beta$ and Mg II luminosities with the 3000 Å continuum luminosity are
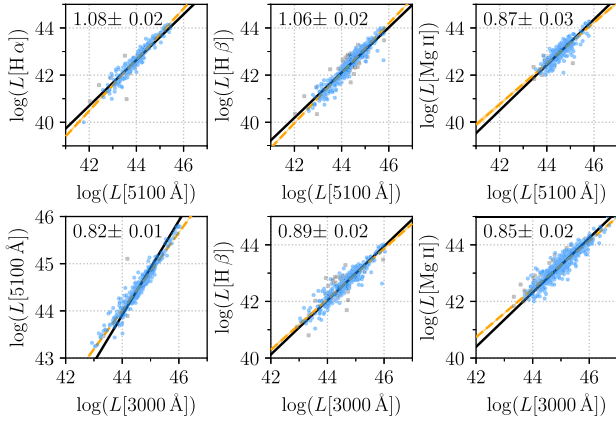
Figure 13. Correlations between emission line and continuum luminosities. The dashed orange lines show the best-fitting linear regression to the data, and the value of the slope (*m*) is given in the inset text. Relations with a gradient *m* = 1 are shown for comparison (solid black lines). Data with good quality measurements of all quantities are shown as blue circles; the grey squares represent data in which at least one measurement has a quality flag raised. All luminosities have been derived from the SDSS optical spectrum and are therefore contemporaneous.

Table 2. Correlations between emission line and continuum luminosities. We give the independent and dependent variables (*x* and *y*) and the slope and intercept (*m* and *c*) of the best-fitting linear regression. σ is the dispersion of the data about the linear regression.

x	y	m	c	σ
$\log L_{5100}$	$\log L_{H\alpha}$	1.08 ± 0.02	-5.0 ± 1.0	0.20
$\log L_{5100}$	$\log L_{H\beta}$	1.06 ± 0.02	-4.4 ± 0.7	0.18
$\log L_{5100}$	$\log L_{Mg\,II}$	0.87 ± 0.03	4.4 ± 1.3	0.21
$\log L_{3000}$	$\log L_{5100}$	0.82 ± 0.01	7.8 ± 0.5	0.11
$\log L_{3000}$	$\log L_{H\beta}$	0.89 ± 0.02	2.8 ± 0.9	0.20
$\log L_{3000}$	$\log L_{Mg\,II}$	0.85 ± 0.02	5.1 ± 0.7	0.19

shallower compared with those calculated for the 5100 Å luminosity. The strongest correlation we find is between the two continuum luminosities: the best-fitting linear regression has a slope $m = 0.82 \pm 0.01$ and a dispersion of only 0.11 dex (around half that of the emission lines).

4.2 Black hole mass estimates from broad emission lines

Assuming that the gas in the BLR is virialized, an estimate of the mass of the black hole (M_{BH}) can be made by applying the virial theorem:

$$M_{BH} = f \frac{R_{BLR} V^2}{G}, \tag{1}$$

where R_{BLR} is the radius of the BLR, V is the velocity of the gas, and G is the gravitational constant. R_{BLR} is often determined from radius–luminosity relations derived for reverberation-mapped AGN (e.g. Peterson et al. 2004; Bentz et al. 2006; Landt et al. 2011a). The virial factor f accounts for the geometry and inclination to the line of sight of the BLR and cannot be easily determined for individual sources. Where the broad-line FWHM is used as a proxy for V then $f \sim 1$.

We first make estimates of the black hole masses of the AGN in our sample using recently calibrated relations for virial black hole masses based on the broad Hα, Hβ, and Mg II emission lines. The

Table 3. The virial black hole mass relations from literature used.

| FWHM	L	Relation ref.	N_{spectra}	$N_{\mathrm{spectra}}^{\mathrm{good}}$
Hα	5100 Å	MR16	219	189 (86 per cent)
Hα	Hα	MR16	220	194 (88 per cent)
Hβ	5100 Å	MR16	481	401 (83 per cent)
Hβ	Hβ	G10	482	406 (84 per cent)
Mg II	3000 Å	MR16	517	479 (93 per cent)
Mg II	Mg II	W18	521	483 (93 per cent)

Note. We list the emission-line FWHM and luminosity L used to calculate the black hole masses. N_{spectra} is the number of spectra in which the relevant quantities were measured. $N_{\mathrm{spectra}}^{\mathrm{good}}$ is the subset of N_{spectra} (and the percentage) for which no quality flags were raised on the relevant emission line or luminosity measurements. *References*: MR16 = Mejía-Restrepo et al. (2016); G10 = Greene et al. (2010); W18 = Woo et al. (2018).

relations used are listed in Table 3. It is common to use the AGN continuum luminosity as a proxy for the BLR radius, however MR16, Greene et al. (2010) and Woo et al. (2018) provide relations using instead the broad emission line luminosity (of Hα, Hβ, and Mg II, respectively). Both MR16 and Woo et al. (2018) provide relations to estimate the black hole mass directly from a luminosity and broad emission line FWHM, assuming $f = 1$ and $f = 1.12$, respectively. For the MR16 relations we use the values calibrated from the global fit to the continuum (the middle column of table 7 in that paper); these are the most appropriate relations for our sample since PYQSOFIT performs a global fit to the broad-band AGN continuum, rather than a local fit under each separate emission line. The Woo et al. (2018) relation we adopt for Mg II was derived with free β and γ, which has least scatter with respect to fiducial black hole masses estimated using FWHM(Hβ) and L_{5100}. Greene et al. (2010) provide a formula to calculate R_{BLR} from the Hβ emission line luminosity; with an estimate of R_{BLR} we then calculate the black hole mass using equation (1) assuming $f = 1$.

4.3 Comparisons of black hole mass estimates

In Fig. 14 we compare the mass estimates obtained using the relationships given in Table 3. Generally, we find good agreement between the different estimates. For estimates using the continuum luminosity, we see excellent agreement between Hα and Hβ masses; on average the difference between the mass estimates are negligible ($\Delta M_{BH} \approx 5$ per cent) and the scatter is $\sigma = 0.22$ dex. There is slightly greater scatter between the Hβ and Mg II estimates ($\sigma = 0.26$), as has been noted in previous studies (e.g. MR16). The average agreement is less good, also, with the Mg II masses being systematically higher than the Hβ masses by ≈ 60 per cent. The greatest discrepancy we see is between the two Hα relations: masses calculated using the broad Hα line luminosity are systematically lower than those made with the continuum luminosity by almost a factor of two. Mass estimates using the broad Hβ line luminosity are also systematically lower than those involving the continuum luminosity, but the discrepancy is less than between the Hα estimates (≈ 60 per cent). In the case of Mg II, there is a very minor systematic in the opposite sense with estimates using the continuum luminosity being ≈ 20 per cent smaller than those using the line luminosity.

To obtain better consistency between the various mass estimates, we have calculated new virial mass relations. We have assumed that the relation using the FWHM of broad Hβ and the 5100 Å luminosity provides the best and most reliable estimate of the mass.
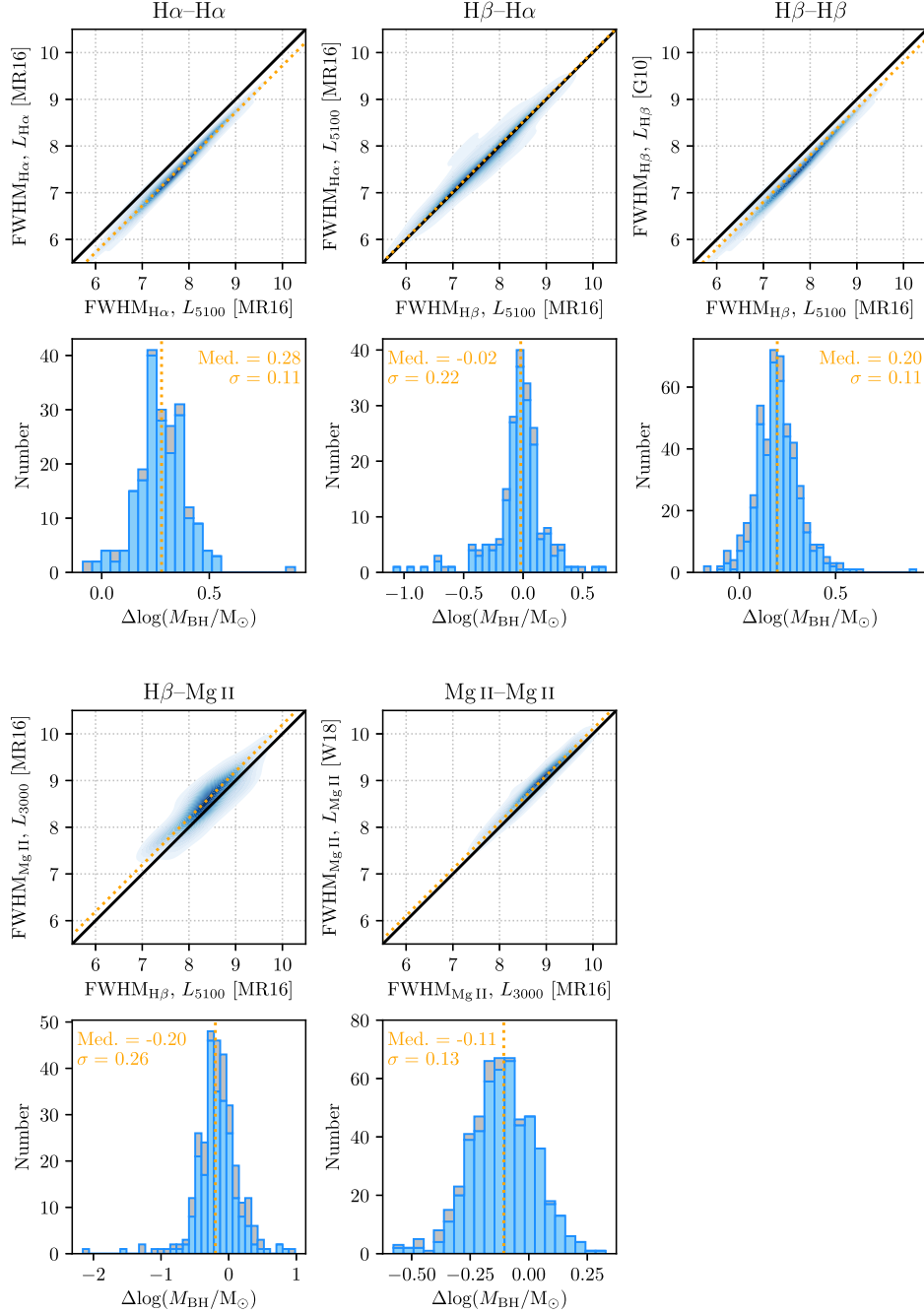
Figure 14. Comparisons of black hole mass estimates made using single-epoch relations taken from literature. The contour plots in the upper panels show $\log(M_{BH}/M_\odot)$, calculated using the broad emission line FWHM and the luminosity indicated on the axis label. The literature reference for the relation used is also given in square parentheses (see Table 3 in the text). The lower panels show the difference in $\log(M_{BH}/M_\odot)$. Reliable mass differences (with no quality flags raised on any of the quantities used to calculate the two masses) are shown in blue and the quality-flagged values are shown in grey. The median and standard deviations of the distributions for the reliable differences are given in the inset text. The solid black lines in the upper panels show the 1:1 relation and the orange dotted lines mark the median in both lower and upper panels.

Then, assuming a virial mass relation of the form

$$\log\left(\frac{M_{BH}}{M_\odot}\right) = \log(K) + \alpha \log\left(\frac{L}{10^{44} \text{ erg s}^{-1}}\right)$$
$$+ 2\log\left(\frac{\text{FWHM}}{1000 \text{ km s}^{-1}}\right), \qquad (2)$$

we determine the quantities K and α which reduce the systematic offset and scatter in the relations with respect to the Hβ-5100 Å

mass estimates. These values are given in Table 4 along with the median $\Delta\log(M_{BH})$ with respect to the Hβ-5100 Å mass estimates, and the dispersion σ. Here, and in Fig. 15, it can be seen that the systematic offsets between different relations are greatly reduced.

For additional quality assurance, we visually inspected all of the optical spectra of the AGN in our sample to identify problematic line profiles (e.g. very noisy or low-contrast profiles; profiles affected by absorption features or data gaps, etc.) and select the best broad

Table 4. Recalibrated virial black hole mass relations.

FWHM	L	$\log(K)$	α	Median $\Delta\log(M_{BH})$	σ	N_{spec}
Hα	Hα	7.56 ± 0.04	0.47 ± 0.02	0.00	0.21	182
Hα	5100 Å	6.91 ± 0.02	0.55 ± 0.03	0.02	0.22	182
Hβ	Hβ	7.83 ± 0.01	0.494 ± 0.007	-0.01	0.10	401
Mg II	Mg II	7.70 ± 0.05	0.46 ± 0.04	-0.03	0.29	224
Mg II	3000 Å	6.86 ± 0.03	0.51 ± 0.03	-0.02	0.26	224

Note. The relations are of the form $\log(M_{BH}) = \log(K) + \alpha\log(L) + 2\log(FWHM)$. Values of K and α have been determined which minimize the systematic offset and scatter with respect to masses estimated from the relation involving the FWHM of broad Hβ and the 5100 Å luminosity. The relations are calculated with the luminosity L in units 10^{44} erg s^{-1} and FWHM in units 1000 km s^{-1}. For each relation we have calculated we give the resulting systematic offset and dispersion with respect to the Hβ-5100 Å mass estimates. N_{spec} is the number of spectra used to establish the relation.
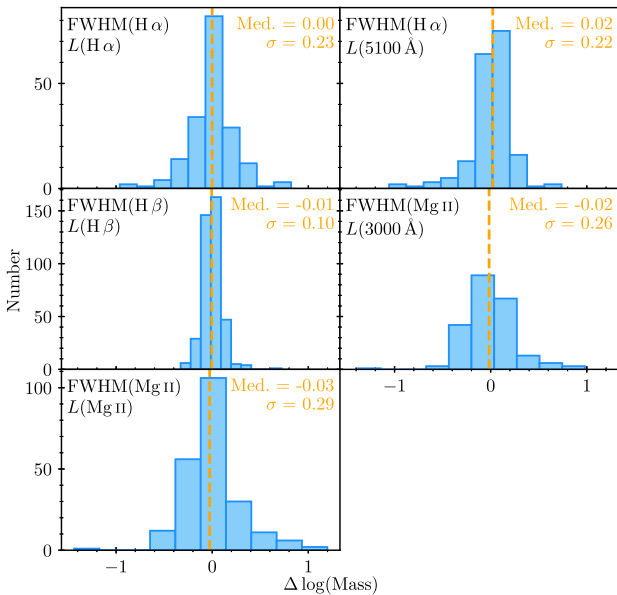
Figure 15. Comparisons of BH mass estimates from broad emission lines. $\Delta\log(\text{Mass}) = \log\left(\text{Mass}|^{\text{FWHM}(H\beta)}_{L(5100\,\text{Å})}\right) - \log\left(\text{Mass}|^{\text{FWHM}(x)}_{L(y)}\right)$.

line from which to estimate the black hole mass, considering the line profile shape and S/N in the vicinity of the line. For each AGN in our sample, we list the 'preferred' mass estimate indicating our choice of broad emission line and luminosity scaling relation for that source. We have adopted, wherever possible, the black hole mass from estimates using FWHM(Hβ) and L_{5100}, since Hβ is the broad emission line most commonly used to establish the BLR radius–luminosity relationship in optical reverberation mapping campaigns and this relationship forms the basis of single-epoch mass estimates such as those used here. As noted in the literature, other line–luminosity relations are often associated with a greater overall uncertainty on the black hole mass and typically relations involving other lines are calibrated to that of Hβ (e.g. Netzer et al. 2007; Shen et al. 2011; Mejía-Restrepo et al. 2016; Woo et al. 2018). In cases where the Hβ profile was unsuitable, we have chosen another broad line and nearby continuum luminosity (usually Hα-L_{5100} for low-redshift sources and Mg II-L_{3000} Å for high-redshift sources). Spectra lacking any broad emission line from which a black hole mass could be reliably estimated were previously removed in our initial quality inspection (Section 2.4).

5 α_{OX}: THE UV/X-RAY RELATIONSHIP

We calculate the UV/optical-X-ray energy index

$$\alpha_{ox} = -0.384 \log\left(\frac{L_{2\,keV}}{L_{2500}}\right) \tag{3}$$

for the AGN in our sample. An α_{ox} determination is possible for 685 AGN in our sample (98 per cent), with the remaining AGN having no spectroscopic or photometric coverage of 2500 Å.

We find a mean $\alpha_{ox} = 1.40$ with standard deviation $\sigma = 0.16$. These values are in good agreement with those found by Lusso et al. (2010) for a sample of quasars in the *XMM*–COSMOS field (mean 1.37 and dispersion 0.18). Lusso & Risaliti (2016) also show that there is a correlation of α_{ox} with L_{2500}. Fig. 16 (upper, left-hand panel) shows this for the 685 AGN in our sample. Plainly there is a good correlation of these quantities in our data. We use LINMIX, the PYTHON version of the package LINMIX_ERR (Kelly 2007), to perform a linear regression analysis. We find a statistically significant correlation with

$$\alpha_{ox} = -(2.07 \pm 0.19) + (0.117 \pm 0.007) \times \log L_{2500}. \tag{4}$$

The dispersion about this line is 0.12 dex, showing that including the luminosity dependence does indeed reduce the scatter, as required if this is to be used as a cosmological probe (Lusso & Risaliti 2016). By contrast, there is much larger scatter of α_{ox} with the monochromatic X-ray luminosity, $L_{2\,keV}$. For these quantities we find

$$\alpha_{ox} = (0.94 \pm 0.25) + (0.018 \pm 0.010) \times \log L_{2\,keV}; \tag{5}$$

unlike the $\alpha_{ox}-L_{2500}$ relation, this relation is not statistically significant at the 2σ level.

5.1 α_{ox} related to emission line width and radio-loudness

In addition to values for the full sample, we have also calculated median α_{ox} and UV/X-ray luminosities for the narrow- and broad-line and radio-loud and radio-quiet subsets. We have also determined the luminosity–α_{ox} relations for these subpopulations. The resultant values are listed in Table 5. It is clear that the narrow-line AGN are typically lower in luminosity (both UV and X-ray) than the broad-line AGN. A slightly greater difference is seen in the UV luminosities (median $\Delta L_{2500} \approx 0.9$ dex) than the X-ray luminosities (median $\Delta L_{2\,keV} \approx 0.8$ dex) between the narrow- and broad-line AGN. We therefore see a small difference in the α_{ox} distributions of the subsets, with the narrow-line AGN having, on average, slightly lower α_{ox}. Fig. 17 shows that this is a consequence of the narrow-line sources having lower UV luminosities with $\alpha_{ox} \propto L_{2500}$. If the
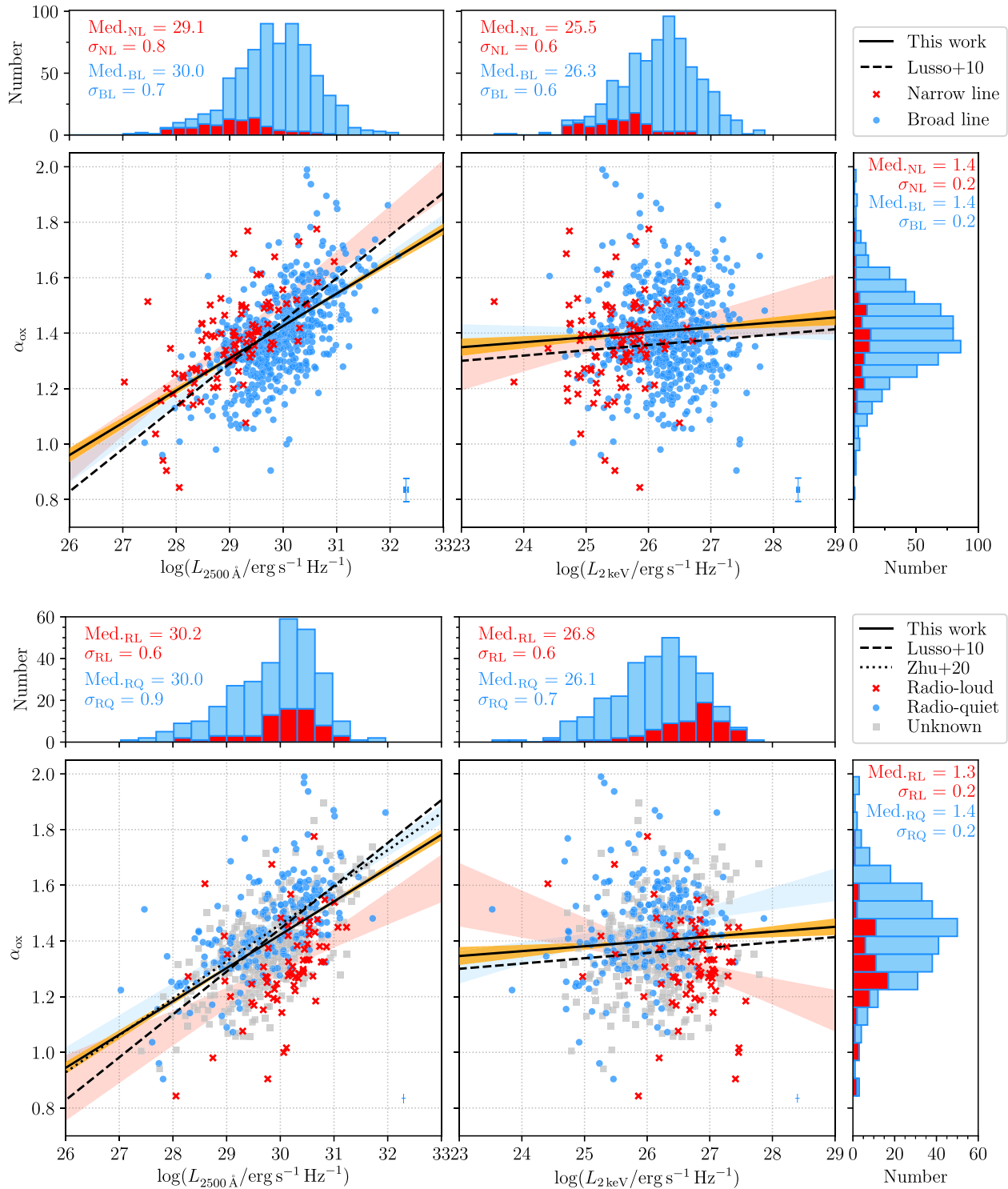
Figure 16. The relationship between α_{ox} and L_{2500} (left-hand panel) and between α_{ox} and $L_{2\,keV}$ (right-hand panel). We show the best-fitting relation found in this work for all 673 AGN with an α_{ox} determination (solid black line) and its 1σ confidence region (in orange). For comparison we show the best-fitting relations found by Lusso et al. (2010) (dashed black line) and Zhu et al. (2020) (dotted black line). Stacked histograms show the distributions of the three quantities. In the top figure, narrow-line (FWHM < 2000 km s^{-1}) and broad-line sources are coloured red and blue, respectively; the pale red and blue areas in the scatter plots are the linear regression 1σ confidence regions for the respective subsets. In the bottom figure data pertaining to the radio-loud and radio-quiet sources are similarly coloured red and blue, respectively. The median 1σ uncertainty on each quantity is shown by the error bars in the lower right-hand corner of each scatter plot.

AGN are binned by L_{2500} then it can be seen that the α_{ox} values of the narrow-line AGN are systematically *higher* than those of the broad-line AGN across the luminosity bins, and progressively so with increasing L_{2500}.

The differences in luminosities between the RL and RQ populations are less marked than those seen in the narrow- and broad-line subsets. The RL and RQ AGN have similar UV luminosities, although the RL subset is on average more luminous in X-rays

Table 5. Median, standard deviation, and linear regression values from α_{ox} analysis. For the whole sample of AGN with an α_{ox} determination (673 AGN) we list the median (M) and standard deviation (σ) of α_{ox}, the logarithms of the 2500 Å UV luminosity and 2 keV X-ray luminosity as well as the slope (m) and intercept (c) of the best-fitting linear regression of the luminosities with respect to the dependent variable α_{ox} and the dispersion (σ) of the data about the best fit. These quantities are also given for the narrow-line ('NL'), broad-line ('BL'), radio-loud ('RL'), and radio-quiet ('RQ') subsets of the sample. N is the number of AGN in each set. Luminosities are in units erg s^{-1} Hz^{-1}.

Set	N	$M_{\alpha_{ox}}$	$\sigma_{\alpha_{ox}}$	$M_{\log L_{2500}}$	$\sigma_{\log L_{2500}}$	$M_{\log L_{2\,keV}}$	$\sigma_{\log L_{2\,keV}}$	$m(\log L_{2500})$	$c(\log L_{2500})$	$\sigma(\log L_{2500})$	$m(\log L_{2\,keV})$	$c(\log L_{2\,keV})$	$\sigma(\log L_{2\,keV})$
All	685	1.40	0.16	29.85	0.77	26.23	0.63	0.117 ± 0.007	-2.07 ± 0.19	0.12	0.018 ± 0.010	0.94 ± 0.25	0.15
NL	96	1.37	0.17	29.09	0.76	25.51	0.57	0.147 ± 0.017	-2.89 ± 0.50	0.12	0.038 ± 0.031	0.40 ± 0.79	0.17
BL	589	1.40	0.16	30.00	0.70	26.31	0.56	0.129 ± 0.008	-2.45 ± 0.23	0.12	0.004 ± 0.012	1.31 ± 0.30	0.15
RL	68	1.29	0.16	30.19	0.65	26.81	0.60	0.101 ± 0.030	-1.72 ± 0.92	0.15	-0.074 ± 0.033	3.28 ± 0.87	0.16
RQ	229	1.45	0.16	29.99	0.85	26.14	0.66	0.126 ± 0.010	-2.29 ± 0.29	0.12	0.049 ± 0.017	0.17 ± 0.43	0.16
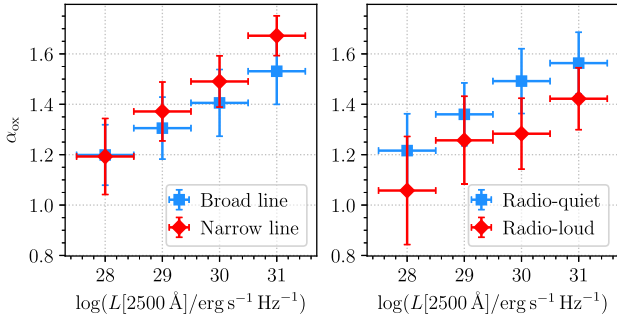
Figure 17. α_{ox} versus UV luminosity for (left-hand panel) narrow-line / broad-line and (right-hand panel) radio-loud / radio-quiet subsamples. In each luminosity bin, narrow-line sources have higher α_{ox} values (softer SEDs) than broad-line sources whereas radio-loud sources have lower α_{ox} values (harder SEDs) than radio-quiet sources.

(median $\Delta L_{2\,keV} \approx 0.7$ dex). The radio-loud AGN have, on average, lower values of α_{ox} than the radio-quiet AGN. In the lower left-hand panel of Fig. 16 we show the L_{2500}–α_{ox} relation derived by Zhu et al. (2020) from a sample of radio-quiet quasars. We see that our relation for the RQ subset agrees very well with theirs, and that our RL sources have systematically lower α_{ox}.

## 5.2 α_{ox} related to black hole mass and accretion rate

We also test the correlation between α_{ox} and black hole mass. There is substantial scatter but LINMIX determines a significant relation with mass:

$$\alpha_{ox} = (0.83 \pm 0.08) + (0.069 \pm 0.009) \times \log(M_{BH}/M_\odot), \quad (6)$$

which is shown in Fig. 18. Whilst the correlation between α_{ox} and M_{BH} is not particularly strong (the Pearson correlation coefficient is ≈ 0.3), we can exclude the absence of a correlation with high confidence: the gradient 0.069 ± 0.009 is $>7\sigma$ inconsistent with zero and the Pearson correlation has a p-value $\ll 0.05$ for a null hypothesis of the distributions being uncorrelated.

We now investigate how much of the scatter in α_{ox}–L_{2500} can be explained by a two-parameter dependence on both mass and mass accretion rate (or equivalently, mass and L/L_{Edd}). The left-hand panel of Fig. 19 shows the range of $\log L_{2500}$ (a tracer of mass accretion rate) and black hole mass spanned by our sample, binned onto a grid with two squares per decade. Within each grid square, we calculate the average α_{ox}, where this should be a more reliable estimate as we are using only AGN with approximately the same mass and mass accretion rate. We remove squares containing fewer than three AGN, so the dispersion is dominated by the intrinsic scatter. This shows a
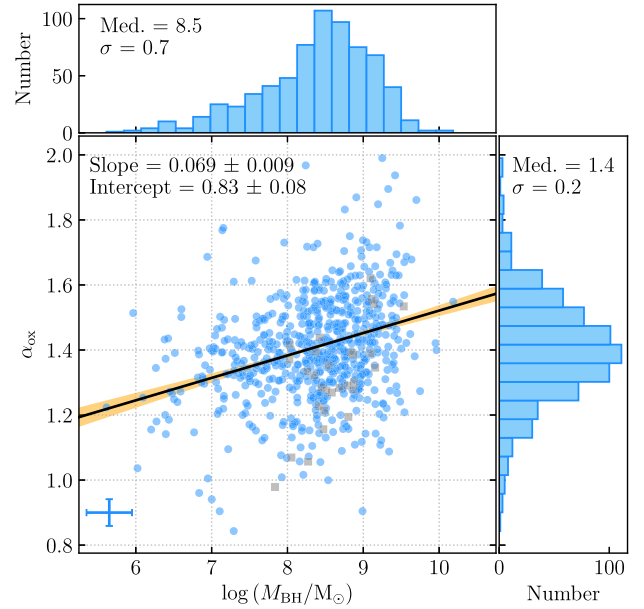
Figure 18. The relationship between α_{ox} and the logarithm of black hole mass. We show the best-fitting relation found (solid black line) and its one sigma confidence region (in orange). Histograms show the distributions of the two quantities. A representative error bar is shown in the bottom left.

subtle but systematic trend in the data, where α_{ox} depends on both mass and L_{2500}, rather than L_{2500} alone.

For comparison, the right-hand panel of Fig. 19 shows the predicted α_{ox} over this parameter space from the quasar SED model QSOSED (Kubota & Done 2018). This model was built from extrapolating trends seen in a much smaller sample of AGN than used here (Jin et al. 2012a,c). These earlier studies showed that optical-X-ray AGN SEDs are typically fit by three components: an outer standard disc, emitting thermal blackbody down to some radius, where it transitions to incomplete thermalization (warm Comptonization). This optically thick structure truncates below the radius R_{hot} so that the gravitational power within some $R_{hot} > R_{isco}$ (where R_{isco} is the radius of the innermost stable circular orbit) is dissipated instead in heating optically thin plasma which produces the power-law tail by Comptonization. The QSOSED model assumes that the total X-ray luminosity is fixed at $L_X = 0.02L_{Edd}$, i.e. the maximal ADAF limit. This captures the main trends seen in the data: at low mass accretion rates, only just above $L_{bol} = 0.02L_{Edd}$, then almost all of the accretion power is required to power the X-rays, and R_{hot} is large. Conversely, at high mass accretion rates, $L_{bol} \approx L_{Edd}$ then almost all of the accretion power is dissipated in the disc, and the X-rays at $L_X = 0.02L_{Edd}$ are weak in comparison and $R_{hot} \rightarrow R_{isco}$.
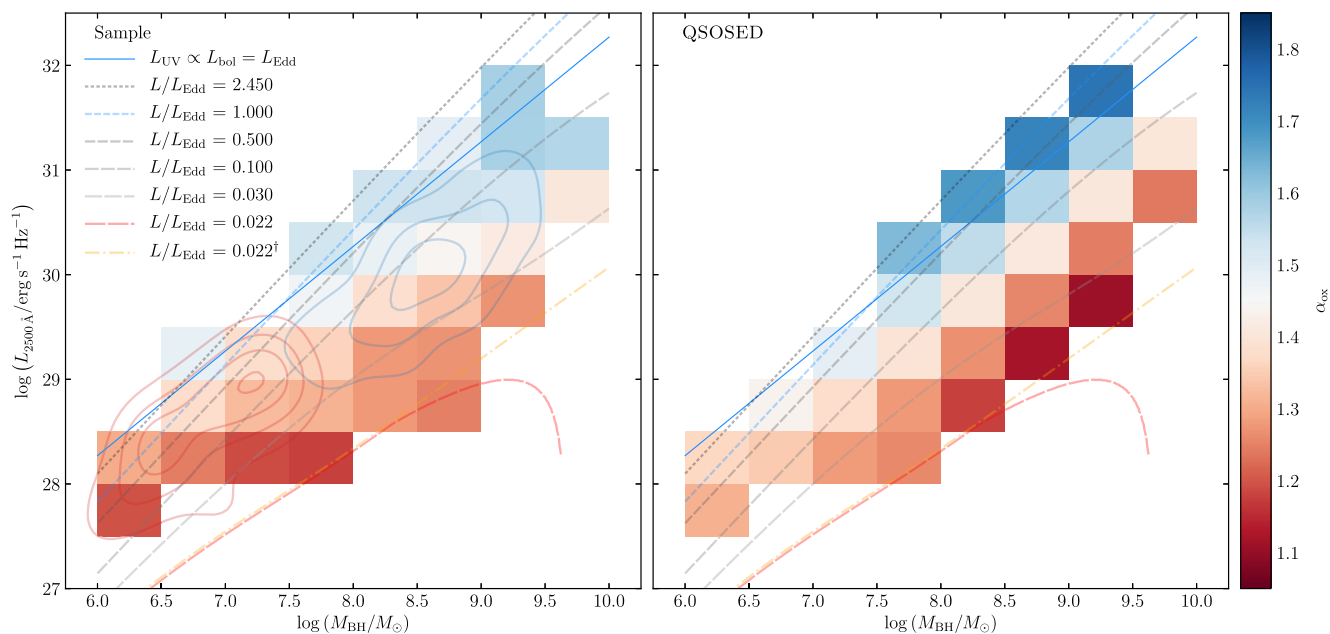
Figure 19. The relationship between L_{2500}, black hole mass, and $\alpha_{\rm ox}$. The left-hand panel shows the average values of $\alpha_{\rm ox}$ for sources in our sample binned by mass and UV luminosity (only bins containing three or more AGN are drawn). The right-hand panel shows the value of $\alpha_{\rm ox}$ taken from an SED generated by the Kubota & Done (2018) model QSOSED, for mass and UV luminosity at the centre of each grid point. The dashed lines correspond to curves of constant $L/L_{\rm Edd}$ as determined by QSOSED. †The dot-dashed orange line is computed for a model in which the outer disc radius $\log (R_{\rm out}/R_{\rm g}) = 5$, whereas in all other models the outer radius is the disc self-gravity radius (cf. the red dashed line at the same $L/L_{\rm Edd}$). The solid blue line shows a curve of constant $L/L_{\rm Edd} = 1$ assuming that $L_{2500} \propto L_{\rm bol} \propto M_{\rm BH}$; clearly this relation is shallower than that predicted by QSOSED (the dashed blue line) in which $L_{2500} \propto M_{\rm BH}^{4/3}$. The blue and red contours in the left-hand panel indicate the parameter space spanned by AGN drawn from the SDSS quasar catalogue and the supplementary sources added from the R17 NLS1 catalogue, respectively.

The changing size of $R_{\rm hot}$ and changing luminosity of the disc also sets the spectral slope of the power-law tail self-consistently, so that the model predicts the entire SED (with some additional assumptions about the warm Compton region).

We use this model, with spin fixed at 0 and the inclination at $30°$ to generate a set of SEDs from black hole masses in the range $\log (M_{\rm BH}/{\rm M_\odot}) = 5.75$–9.75 and adjust the mass-normalized accretion rate between $\dot{m} = 0.022 - 2.5$ (the limits of the QSOSED model) to match $\log L_{2500}$ to the midpoint of each grid square. We extract $\alpha_{\rm ox}$ from these model SEDs, and colour code each grid point using the same colour scale as from the real data. We also overplot lines of constant $L/L_{\rm Edd}$ on both panels. At high Eddington ratio, the UV is always dominated by the standard disc over this mass range, so the monochromatic flux $L_{2500} \propto (M_{\rm BH}\dot{M})^{2/3}$, giving rise to the straight lines for high $L/L_{\rm Edd}$ on the $\log (M_{\rm BH}/{\rm M_\odot}) - \log L_{2500}$ plot. Importantly this is not a linear relationship even for a standard disc. $L_{2500} \propto \dot{M}^{2/3} \propto L_{\rm bol}^{2/3}$ rather than being directly proportional to $L_{\rm bol}$. Thus using monochromatic flux (or flux in a narrow band) with a constant K-correction to convert to bolometric flux and hence derive Eddington ratio is subtly incorrect even for standard disc models. The solid blue line in Fig. 19 shows the Eddington luminosity assuming a bolometric correction of $L_{\rm bol} = 2.75 \times L_{2500}$ (Krawczyk et al. 2013).

The QSOSED model assumes the disc extends outwards only to the self-gravity radius, which may be only a few hundred $R_{\rm g}$ at the lowest accretion rates for the highest masses (see Mitchell et al., submitted). This causes the very sharp drop in L_{2500} for these objects. If instead we set the outer disc radius to $10^5 R_{\rm g}$ (as is usually done in the near-infrared spectroscopic modelling of AGN, e.g. Landt et al. 2011b), the $0.022L_{\rm Edd}$ curve more closely resembles those of higher Eddington ratios (the orange dot-dashed line in Fig. 19).

It is very clear that the data sit between the Eddington limits hardwired into the QSOSED code. The model assumes that below $L/L_{\rm Edd} \sim 0.02$, the flow is completely dominated by the optically thin, geometrically thick, hot plasma (ADAF/RIAF flow). This transition is seen directly in the 'changing-look' AGN (Noda & Done 2018; Ruan et al. 2019). Above $L/L_{\rm Edd} \sim 2.5$, the disc photosphere becomes brighter than its local Eddington flux (Kubota & Done 2019), so its structure is probably modified by strong winds and/or radial advection of the excess radiation. Both these probably lead to vertical structure so the radiation escaping preferentially only in a narrow funnel, making these objects rare (as well as probably short lived).

In the left-hand panel of Fig. 19 we see that the contours indicating the location of our data do not extend above the $L/L_{\rm Edd} = 0.5$ line (taken from the QSOSED model for the same mass and UV luminosity) at the high mass end; this implies that high Eddington fraction flows are rare at the highest black hole masses in our sample. These objects are very rare in the local Universe, and are only likely to be present at the peak of the quasar epoch at $z \sim 2$–4, beyond our sample redshift limit. Conversely, we only see the lowest mass black holes ($M \sim 10^6 M_\odot$) at the highest Eddington ratios as these are otherwise diluted by their host galaxy (e.g. Done et al. 2012).

In Fig. 20 we show the difference between $\alpha_{\rm ox}$ as predicted by QSOSED and as determined in our sample. This shows that there is generally reasonable agreement between the model predictions and $\alpha_{\rm ox}$ estimated from our data, in the majority of bins (97 per cent) $|\Delta\alpha_{\rm ox}| < 0.2$ and in 52 per cent of the bins $|\Delta\alpha_{\rm ox}| < 0.1$. A detailed exploration of the differences between the data and the model predictions as a function of black hole mass, luminosity, and accretion rate will be performed in future work (Mitchell et al., submitted).
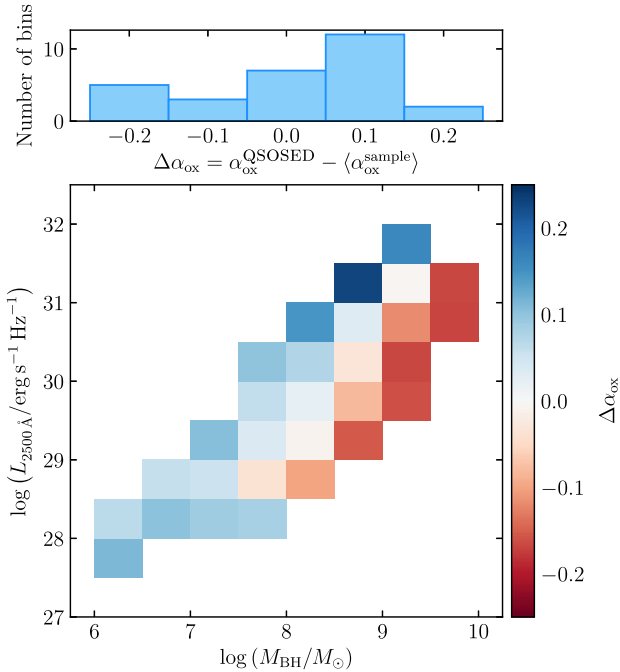
Figure 20. The difference between $\alpha_{\rm ox}$ as predicted by the QSOSED model and as found from the data in our sample. The bottom panel shows the difference in each mass–luminosity bin, the top panel shows the distribution of $\Delta\alpha_{\rm ox}$ values.

6 DISCUSSION

The SOUX sample is a medium-sized selection of ≈ 700 AGN up to $z = 2.5$ with quality multiwavelength data. Although it is not a complete or representative AGN sample, it has been assembled with the intention of exploring a wide range of parameter space. This first study has focused on making key measurements from the data (the broad optical emission line properties, multiwavelength luminosities etc.) and on determining a consistent set of black hole masses from which we will be able to determine Eddington ratios. We have taken a preliminary look at changes in the shape of the AGN SED (characterized by the parameter $\alpha_{\rm ox}$) as a function of black hole mass and AGN luminosity. To first order, $\alpha_{\rm ox}$ is found to vary with both mass and luminosity as broadly expected from a theoretical model. However, the systematic differences seen at second order will require a deeper exploration. Our main findings are discussed below.

6.1 Broad emission line properties

In Section 4.1, we found that the luminosities of the broad emission lines $H\alpha$ and $H\beta$ were very well-correlated, with the mean luminosity ratio (the Balmer decrement) being $\langle H\alpha/H\beta\rangle = 3.9$ (Fig. 11). The analysed spectra have all been corrected for line-of-sight Galactic reddening, therefore if (as proposed by Gaskell 2017) the intrinsic broad line ratio is close to the Case B value (≈ 2.7) then the typically higher values that we find imply that a large fraction of the AGN in our sample are reddened by dust in their host galaxies or galactic nucleus. In Fig. 12 we show the spectra of two sources with high Balmer decrements. One spectrum (6369-56217-0193) exhibits both high broad line and narrow line Balmer decrements and lacks a dominant blue AGN continuum. (Note however that PYQSOFIT performs a principal component analysis using galaxy and

AGN eigenspectra to obtain a best estimate of the AGN continuum level.) This AGN may plausibly be reddened by dust in its host galaxy affecting both the BLR and NLR; however the presence of reddening in many other spectra with high Balmer decrements is more ambiguous. For example, the spectrum 5694-56213-0636 (also shown in Fig. 12) has a high broad-line Balmer decrement but a narrow line decrement consistent with Case B and a very strong, blue continuum. The modified spectrum, dereddened using the Gaskell et al. (2004) AGN curve and an $A_{\rm v}$ estimated from the Balmer decrement, is shown in orange in Fig. 12. This spectrum is ~ 1.25 dex brighter than the spectrum from which we have measured emission line and continuum properties, but it is still quasar-like. Whilst we cannot definitively rule out the intrinsic reddening of such individual sources, Fig. 11 shows that across our sample Balmer decrements and the ratio of 5100 and 3000 Å continuum luminosities are uncorrelated: we would expect there to be a correlation in the reddening scenario. Alternatively, it is possible that the intrinsic Balmer decrements in AGN BLRs are not always close to Case B. This was a conclusion of Schnorr-Müller et al. (2016), who found that the intrinsic Balmer decrements of type 1 AGN spanned a wide range (2.5–6.6), derived from photoionization modelling of their data; the vast majority of the AGN in our sample fall within this range. A range of intrinsic broad-line Balmer decrements would partly explain why we do not find a correlation between Balmer decrements and continuum luminosity ratios in our sample.

We showed in Section 4.1 that the luminosities of the three broad emission lines $H\alpha$, $H\beta$, and Mg II are strongly correlated with the AGN continuum luminosities. Of all emission line-continuum relations we tested, the least scatter is found between $L_{H\beta}$ and L_{5100}, although the relationship between the two continuum luminosities L_{5100} and L_{3000} Å had least scatter overall. Whilst the luminosities of the two Balmer lines are approximately linearly proportional to L_{5100}, the $\log(L_{\rm Mg\,II})$–$\log(L_{5100})$ relation has a gradient noticeably lower than 1 ($m = 0.87 \pm 0.03$). These results are similar to those reported by Shen & Liu (2012), who also found least scatter between L_{5100} and L_{3000} Å. Shen & Liu (2012) report a $\log(L_{H\alpha})$–$\log(L_{5100})$ relation with a gradient consistent with 1 and a $\log(L_{\rm Mg\,II})$–$\log(L_{5100})$ proportionality consistent with the one we find here (0.86 ± 0.07). However, in their work they find a much steeper relation between $\log(L_{H\beta})$–$\log(L_{5100})$ (1.25 ± 0.07). Collinson et al. (2017) also find a shallower gradient between $\log(L_{\rm Mg\,II})$ and $\log(L_{5100})$ than between the Balmer lines and $\log(L_{5100})$. In their work, the gradients of all three emission line luminosities with respect to $\log(L_{5100})$ were shallower than 1, although they were working with a much smaller sample (11 AGN) and much more limited range in luminosities (~ 1 dex). It has been shown both in observations of reverberation-mapped and 'changing-look' AGN (e.g. Sun et al. 2015; Zhu, Sun & Wang 2017; Kynoch et al. 2019a) and photoionization models (e.g. Guo et al. 2020) that Mg II responds more weakly to changes in the continuum than does $H\beta$. Here, we appear to see a similar trend in a statistical sense across several hundred snapshots of individual AGN. Overall, the strong correlations between the emission line and continuum luminosities demonstrate that the former are also reasonable proxies for the BLR radius, which we discuss below.

6.2 Relations for virial black hole masses

In Section 4, we computed the black hole masses of the AGN in our sample from the broad emission line widths and either a nearby continuum luminosity or the broad emission line luminosity. We used several single-epoch virial black hole mass relations taken from recent literature (MR16; Greene et al. 2010; Woo et al. 2018).

Mass estimates from the FWHM of the Balmer lines and L_{5100} are generally in very good agreement, with a minor systematic difference $|\Delta \log (M_{BH})| = 0.2$ (i.e. a 5 per cent difference in M_{BH}) and $\sigma = 0.22$ dex.

We noticed that estimates made using the emission line luminosities were systematically lower than those made using the continuum luminosities. The masses were ≈ 60 per cent lower in the case of Hβ and a factor two lower in the case of Hα. Discrepancies of a similar magnitude have been observed in previous studies calculating the black hole mass of single objects (e.g. Landt et al. 2017; Kynoch et al. 2019b) and also in large samples (e.g. Shen & Liu 2012). Shen & Liu (2012) compared masses calculated with FWHM(Hα)–$L_{H\alpha}$ (Shen et al. 2011) and FWHM(Hβ)–$L_{H\beta}$ (Greene & Ho 2005) with fiducial masses calculated with FWHM(Hβ)–L_{5100} (Vestergaard & Peterson 2006) and again estimates using the line luminosities were lower, although to a lesser extent than we find here (Shen & Liu 2012 report $\langle \Delta \log [M_{BH}] \rangle = -0.08$ and -0.04 dex for Hα and Hβ, respectively). The cause of these systematic discrepancies is not apparent. Since, as we noted above, there is generally very good correspondence between the emission line and continuum luminosities, in principle it should be possible to determine a reliable mass estimate using either luminosity as a proxy for the BLR radius. It is not simply the case that we have taken scaling relations from different papers in which the relations were derived for different samples of AGN: both relations involving Hα are taken from MR16. Such large systematic difference in mass are difficult to explain since by equation (1) $M_{BH} \propto L^{1/2}$, a factor four difference in luminosity is required to reconcile a factor two difference in mass. It is very unlikely that the 5100 Å luminosity is contaminated by stellar light from the AGN host galaxies, resulting in masses determined using L_{5100} to be overestimates. PYQSOFIT performs a subtraction of the host galaxy spectrum where necessary, so in principle the stellar contamination has been removed from our measurements. Also, we see a systematic offset even for the highest mass sources: highly luminous quasars with optical luminosities far in excess of their host galaxies. It is possible that the procedure we used to fit the broad emission lines to determine their width and luminosity could differ from those used to determine the M_{BH} relations (MR16 used only two Gaussians for the Balmer lines, for example). As described in Section 2.3, in this work the broad emission lines were fit with three Gaussians. In many cases, the total line profile was decomposed into broad, very broad, and intermediate Gaussians. The shallow but very broad component (of Hα in particular) which was often present can add a substantial amount of flux to the line without greatly increasing the FWHM. However, this would have the effect of increasing the emission line luminosities, whereas what we find is that the Balmer lines are apparently *underluminous* compared to the continuum for a given FWHM, resulting in the lower mass estimates. Differences in the assumed virial factor f will also result in equivalently larger or smaller estimates of the mass. However, the relations we test here use $f \approx 1$ so the mass differences ought to be very minor.

To avoid problems introduced by systematic differences in black hole masses, we have calculated new mass scaling relations tailored to our sample. As described in Section 4.2 these minimize the discrepancy with respect to estimates made from FWHM(Hβ) and L_{5100} (which we judge is the most reliable estimate). The new scaling relations are presented in Table 4. We note here that all of our relations have a slope α consistent with 0.5, the expected value for an inverse-square scaling of radius with luminosity. The majority of the AGN in our sample (52 per cent) have their adopted masses derived using the FWHM(Hβ)–L_{5100} relation from MR16.

45 per cent of the sample have their preferred mass estimate taken from our FWHM(Mg II) and L_{3000} Å relation; these were mostly the high-z AGN in our sample, lacking optical spectral coverage of Hβ. Our newly derived FWHM(Mg II)–L_{3000} Å mass relation has a very minor systematic offset with respect to the fiducial masses calculated via the FWHM(Hβ)–L_{5100} relation of MR16: the median difference in mass is negligible, although there is substantial scatter ($\sigma = 0.26$ dex). Following these recalibrations, we do not expect that using different mass relations for different AGN will have an adverse impact on our results.

6.3 α_{ox}

In Section 5, we explored the relationship between the UV-X-ray energy index α_{ox} with the UV and X-ray luminosities and with black hole masses and accretion rates. As mentioned in Section 3.1, we have used a mixture of spectroscopic and photometric measurements of L_{2500} (from which we derive α_{ox}) and there are subtle differences between the two luminosity estimates. In Appendix A, we have performed a detailed investigation of the potential impact of mixing different luminosity measures on our results, in particular on the $\log (L_{2500})$–α_{ox} correlation. To summarize, the average difference between spectroscopic and photometric luminosities is small ($\langle \Delta \log (L_{2500}) \rangle = -0.02$ dex) and we derive very similar distributions of α_{ox} from both the spectroscopic and photometric estimates (the α_{ox} distributions are statistically indistinguishable). We have recalculated $\log (L_{2500})$–α_{ox} relations for the spectroscopically and photometrically derived quantities separately. Each subset shows a statistically significant positive correlation between $\log (L_{2500})$ and α_{ox}, with similar dispersions (≈ 0.1 dex) so it does not appear that photometric measurements of L_{2500} are substantially less reliable than the spectroscopic ones. The photometric $\log (L_{2500})$–α_{ox} relation differs most from the spectroscopic one at the highest luminosities (Fig. A2). However, this difference will have little effect on our analysis in Section 5. The high-luminosity sources are mostly at high redshift ($z > 0.5$) and for these we have adopted $\log (L_{2500})$ from spectroscopy rather than photometry. Very few sources lie in the region of parameter space in which the photometric $\log (L_{2500})$–α_{ox} relation is substantially different from the spectroscopic one *and* for which $\log (L_{2500})$ and α_{ox} were estimated from photometry. We conclude that the mixing of spectroscopically and photometrically derived UV luminosities does not greatly affect our results and note that several of our relations are in good agreement with those of previous studies.

Our analysis of the final sample in Section 5 demonstrates a highly significant correlation of α_{ox} with the monochromatic UV luminosity $\log (L_{2500})$ for all AGN as well as the four subsets (narrow- and broad-line; radio-loud and quiet) taken separately. No highly significant $\log(L_{2\,keV})$–α_{ox} correlation was found for any individual subset, or for the full sample, although a marginally significant correlation ($>3\sigma$) was found for the RQ subset only (see Table 5 and Fig. 16). In a previous study of quasars, Lusso et al. (2010) found no significant relation between α_{ox} and $\log(L_{2\,keV})$, consistent with our results from a sample containing both quasars and NLS1s.

Figs 16 and 17 illustrate differences in the α_{ox} distributions, and the relationship between α_{ox} and UV luminosity, between different subsets of our sample. Whilst *on average* narrow-line AGN have lower α_{ox} than broad-line AGN they have slightly higher α_{ox} when compared with broad-line AGN in the same L_{2500} bin, likely as a result of the stronger soft X-ray excess (warm corona) component in the former. Figs 16 and 17 show a subtle difference in the dependence of α_{ox} on the UV luminosity between the NL and BL

subsets, with the best-fitting $\alpha_{\rm ox}$–L_{2500} relation being slightly steeper for NL AGN. Starker differences are apparent between our radio-loud and radio-quiet AGN: the radio-loud AGN are on average more X-ray luminous than the radio-quiet sources although both sets have similar UV luminosities (Fig. 16). We can see in Fig. 16 that radio-loud AGN are systematically offset to lower $\alpha_{\rm ox}$ with respect to the radio-quiet sources, and very few radio-loud AGN are found to have higher $\alpha_{\rm ox}$ than the average for radio-quiet AGN at a given $\log L_{2500}$. Even when binning by L_{2500}, radio-loud AGN have systematically lower $\alpha_{\rm ox}$ (harder SEDs) than the radio-quiet AGN. This trend has been noted in previous studies. Lawther et al. (2017) found that radio-loud objects in their sample of $z \approx 2$ AGN were offset from the best-fitting L_{2500}–$\alpha_{\rm ox}$ relation determined by Strateva et al. (2005) for radio-quiet quasars. More recently, Zhu et al. (2020) found that their large sample of radio-loud quasars were offset from a comparison radio-quiet sample, also in the same sense that we find here. In this work we show that the same trend extends to lower luminosities and redshifts than probed by Lawther et al. (2017), consistent with the findings of Zhu et al. (2020). Often this systematic offset between RL and RQ AGN is attributed to the radio jet contributing to the hard X-ray band, resulting in an increased X-ray flux and lower UV/X-ray ratio. However, Zhu et al. (2020) argue that for steep radio spectrum sources the excess X-ray luminosity of radio-loud AGN is likely *not* due to contamination of the X-ray emission by the jet, but rather some evolution of the disc–corona configuration. Zhu et al. (2020) proposed that the best-fitting relation between $\alpha_{\rm ox}$ and $\log L_{2500}$ for radio-quiet quasars (which we show in Fig. 16) represented an approximate 'jet line' for AGN, in analogy to the jet line of accreting stellar mass black holes (e.g. Fender et al. 2004): jets are quenched for AGN on the soft-SED side of the line and the most powerful jets are found in sources with the hardest SEDs for a given L_{2500}. The best-fitting $\alpha_{\rm ox}$–$\log L_{2500}$ relation for radio-quiet quasars found by Zhu et al. (2020) is entirely consistent with our relation for radio-quiet AGN (Table 5 and Fig. 16).

In Section 5.2, we calculated the mean $\alpha_{\rm ox}$ for 31 subsets of AGN, binned in both $\log(M_{\rm BH})$ and $\log(L_{2500})$, ranging over four orders of magnitude in each quantity (Fig. 19). Fig. 19 illustrates the diversity of our sample – blue and red contours show the location in parameter space of AGN drawn from the quasar and NLS1 catalogues, respectively. The inclusion of AGN from the NLS1 catalogue extends our sample to, on average, lower black hole masses and higher accretion rates than probed by the quasars. This wide coverage of parameter space allows us to track changes in the spectral shape (characterized by $\alpha_{\rm ox}$) with both mass and accretion rate. We compared the mean $\alpha_{\rm ox}$ for each bin to the prediction made from the AGN SED model QSOSED. The correspondence between the model predictions and our sample data is remarkably good: nearly two thirds of bins have $|\Delta\alpha_{\rm ox}| < 0.1$ and in all but one bin $|\Delta\alpha_{\rm ox}| < 0.2$. Fig. 20 shows that the largest differences between the data and the model are at the extremes of the parameter space: for high-mass/high-luminosity AGN, the model predicts SEDs which are too red and for low-mass/low-luminosity AGN the predictions are too blue. Of course, if there is a systematic error in the calculation of the black hole masses then the QSOSED predictions are not being compared with the appropriate AGN (however, as described previously, we have made efforts to ensure our mass estimates are reliable). Alternatively, whilst QSOSED makes good predictions for moderate masses and luminosities, it is possible that the actual disc–corona coupling is more complex than assumed by the model. We will explore these issues in more detail by modelling the full broad-band SEDs of our sample (Mitchell et al., submitted).

7 SUMMARY AND CONCLUSIONS

We have assembled a sample of 696 AGN with SDSS optical spectra which also have X-ray spectra and simultaneous UV photometry recorded by *XMM-Newton*. Although this is not a complete or representative sample of AGN, the objects in our sample cover a wide range in black hole mass, accretion rate, spectral hardness, and redshift. The majority of sources in our sample are luminous AGN found in the SDSS-DR14 quasar catalogue; these are supplemented with a selection of lower luminosity, narrow-line type 1 AGN in the local Universe (taken from R17).

We have performed new optical spectral fits so that we have comparable measurements of the AGN continuum and at least one broad emission line (Hα, Hβ, or Mg II) for the narrow-line AGN (and broad-line AGN that lacked a measurement of the Mg II spectral region). We computed single-epoch virial black hole masses for the AGN in our sample using these broad emission lines with both the local continuum luminosity and the emission line luminosity. However, we found some large systematic differences between mass estimated made using different relations (masses calculated with FWHM[Hα] and $L_{\rm H\alpha}$ are on average a factor two lower than those made using L_{5100}). We have calculated new virial black hole mass relations which minimize the systematic offsets with respect to fiducial black hole masses calculated using FWHM(Hβ) and L_{5100}.

From the X-ray catalogue fluxes, we have estimated rest-frame 2 keV luminosities and X-ray photon indices for the full sample. Combining these with our UV measurements, we have calculated the energy index $\alpha_{\rm ox}$, an indicator of the broad-band SED shape. We found that narrow-line sources have steeper X-ray spectra (higher Γ) and, when controlling for $L_{\rm UV}$, have softer broad-band SEDs (higher $\alpha_{\rm ox}$) than typical broad-line AGN. Differences were also found with respect to radio properties with radio-loud AGN being on average more X-ray luminous and having harder SEDs (lower $\alpha_{\rm ox}$) than the radio-quiet subset. Fig. 19 illustrates that the majority of the AGN in our sample lie between $0.02 < L/L_{\rm Edd} < 2$, with narrow-line type 1 offset to lower masses and higher accretion rates compared with the broad-line quasars. We have shown that, to first order, the model QSOSED makes very good predictions of $\alpha_{\rm ox}$ for a given black hole mass and UV luminosity (assumed to be a proxy for the accretion rate). Importantly however, there are systematic differences between the data and model predictions at the extremes of the mass-accretion rate parameter space. In a subsequent paper (Mitchell et al., submitted), we will model the broad-band SEDs, testing the physical assumptions made in the QSOSED model and further explore the relationships between the UV/optical and X-ray spectra.

Space Science Institute (ISSI), Bern, Switzerland, and we thank ISSI for their hospitality. Many thanks also go to Suv Rakshit, Vicky Fawcett, and Hengxiao Guo for their help and advice on using PYQSOFIT.

DATA AVAILABILITY

The data underlying this work are available from the catalogues and archives described in Section 2. The details of the AGN in the SOUX sample, relevant measures, and derived quantities are presented in the table available online which is described in Appendix D.

[4] http://www.astropy.org
[5] https://healpix.sourceforge.io.

REFERENCES

Alexander D. M. et al., 2003, AJ, 125, 383
Arcodia R., Ponti G., Merloni A., Nandra K., 2020, A&A, 638, A100
Astropy Collaboration, 2013, A&A, 558, A33
Astropy Collaboration, 2018, AJ, 156, 123
Becker R. H., White R. L., Helfand D. J., 1995, ApJ, 450, 559
Bentz M. C., Peterson B. M., Pogge R. W., Vestergaard M., Onken C. A., 2006, ApJ, 644, 133
Blanton M. R. et al., 2017, AJ, 154, 28
Calderone G., Nicastro L., Ghisellini G., Dotti M., Sbarrato T., Shankar F., Colpi M., 2017, MNRAS, 472, 4051
Cardelli J. A., Clayton G. C., Mathis J. S., 1989, ApJ, 345, 245
Collinson J. S., Ward M. J., Done C., Landt H., Elvis M., McDowell J. C., 2015, MNRAS, 449, 2174
Collinson J. S., Ward M. J., Landt H., Done C., Elvis M., McDowell J. C., 2017, MNRAS, 465, 358
Condon J. J., Cotton W. D., Greisen E. W., Yin Q. F., Perley R. A., Taylor G. B., Broderick J. J., 1998, AJ, 115, 1693
Done C., Davis S. W., Jin C., Blaes O., Ward M., 2012, MNRAS, 420, 1848
Dong X., Wang T., Wang J., Yuan W., Zhou H., Dai H., Zhang K., 2008, MNRAS, 383, 581
Fender R. P., Belloni T. M., Gallo E., 2004, MNRAS, 355, 1105
Fernández-Ontiveros J. A., Muñoz-Darias T., 2021, MNRAS, 504, 5726
Fitzpatrick E. L., 1999, PASP, 111, 63
Gaskell C. M., 2017, MNRAS, 467, 226
Gaskell C. M., Goosmann R. W., Antonucci R. R. J., Whysong D. H., 2004, ApJ, 616, 147
Gierliński M., Done C., 2004, MNRAS, 349, L7
Ginsburg A. et al., 2019, AJ, 157, 98
Goodrich R. W., 1989, ApJ, 342, 224
Górski K. M., Hivon E., Banday A. J., Wandelt B. D., Hansen F. K., Reinecke M., Bartelmann M., 2005, ApJ, 622, 759
Greene J. E., Ho L. C., 2005, ApJ, 630, 122
Greene J. E. et al., 2010, ApJ, 723, 409
Guo H., Shen Y., Wang S., 2018, Astrophysics Source Code Library, record ascl:1809.008
Guo H., Liu X., Shen Y., Loeb A., Monroe T., Prochaska J. X., 2019, MNRAS, 482, 3288
Guo H. et al., 2020, ApJ, 888, 58
Haardt F., Maraschi L., 1991, ApJ, 380, L51
Jin C., Ward M., Done C., Gelbord J., 2012a, MNRAS, 420, 1825
Jin C., Ward M., Done C., 2012b, MNRAS, 422, 3268
Jin C., Ward M., Done C., 2012c, MNRAS, 425, 907
Kalberla P. M. W., Burton W. B., Hartmann D., Arnal E. M., Bajaja E., Morras R., Pöppel W. G. L., 2005, A&A, 440, 775
Kellermann K. I., Sramek R., Schmidt M., Shaffer D. B., Green R., 1989, AJ, 98, 1195
Kelly B. C., 2007, ApJ, 665, 1489
Komossa S., Voges W., Xu D., Mathur S., Adorf H.-M., Lemson G., Duschl W. J., Grupe D., 2006, AJ, 132, 531
Körding E. G., Jester S., Fender R., 2006, MNRAS, 372, 1366
Krawczyk C. M., Richards G. T., Mehta S. S., Vogeley M. S., Gallagher S. C., Leighly K. M., Ross N. P., Schneider D. P., 2013, ApJS, 206, 4
Kubota A., Done C., 2018, MNRAS, 480, 1247
Kubota A., Done C., 2019, MNRAS, 489, 524
Kynoch D., Ward M. J., Lawrence A., Bruce A. G., Landt H., MacLeod C. L., 2019a, MNRAS, 485, 2573
Kynoch D. et al., 2019b, MNRAS, 487, 181
La Mura G., Popović L. Č., Ciroi S., Rafanelli P., Ilić D., 2007, ApJ, 671, 104
Landt H., Bentz M. C., Peterson B. M., Elvis M., Ward M. J., Korista K. T., Karovska M., 2011a, MNRAS, 413, L106
Landt H., Elvis M., Ward M. J., Bentz M. C., Korista K. T., Karovska M., 2011b, MNRAS, 414, 218
Landt H. et al., 2017, MNRAS, 464, 2565
Lawther D., Vestergaard M., Raimundo S., Grupe D., 2017, MNRAS, 467, 4674
Lu Y., Wang T., Zhou H., Wu J., 2007, AJ, 133, 1615
Lu K.-X., Zhao Y., Bai J.-M., Fan X.-L., 2019, MNRAS, 483, 1722
Lusso E., Risaliti G., 2016, ApJ, 819, 154
Lusso E., Risaliti G., 2017, A&A, 602, A79
Lusso E. et al., 2010, A&A, 512, A34
Magdziarz P., Blaes O. M., Zdziarski A. A., Johnson W. N., Smith D. A., 1998, MNRAS, 301, 179
Mejía-Restrepo J. E., Trakhtenbrot B., Lira P., Netzer H., Capellupo D. M., 2016, MNRAS, 460, 187
Merloni A., Heinz S., di Matteo T., 2003, MNRAS, 345, 1057
Miller L., Peacock J. A., Mead A. R. G., 1990, MNRAS, 244, 207
Moravec E., Svoboda J., Borkar A., Boorman P., Kynoch D., Panessa F., Mingo B., Guainazzi M., 2022, A&A, 662, A28

Netzer H., 2019, MNRAS, 488, 5185

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

Noda H., Done C., 2018, MNRAS, 480, 3898

Novikov I. D., Thorne K. S., 1973, in DeWitt C., DeWitt B. S., eds, Black Holes (Les Astres Occlus). Gordon and Breach, New York, p. 343

Osterbrock D. E., Ferland G. J., 2006, Astrophysics of Gaseous Nebulae and Active Galactic Nuclei. University Science Books, New York

Osterbrock D. E., Pogge R. W., 1985, ApJ, 297, 166

Osterbrock D. E., Shuder J. M., 1982, ApJS, 49, 149

Page M. J. et al., 2012, MNRAS, 426, 903

Pâris I. et al., 2018, A&A, 613, A51

Peterson B. M. et al., 2004, ApJ, 613, 682

Rakshit S., Stalin C. S., Chand H., Zhang X.-G., 2017, ApJS, 229, 39

Rakshit S., Stalin C. S., Kotilainen J., 2020, ApJS, 249, 17

Ruan J. J., Anderson S. F., Eracleous M., Green P. J., Haggard D., MacLeod C. L., Runnoe J. C., Sobolewska M. A., 2019, ApJ, 883, 76

Sacchi A. et al., 2022, A&A, 663, L7

Salvestrini F., Risaliti G., Bisogni S., Lusso E., Vignali C., 2019, A&A, 631, A120

Schlegel D. J., Finkbeiner D. P., Davis M., 1998, ApJ, 500, 525

Schnorr-Müller A. et al., 2016, MNRAS, 462, 3570

Shakura N. I., Sunyaev R. A., 1973, A&A, 24, 337

Shen Y., Liu X., 2012, ApJ, 753, 125

Shen Y. et al., 2011, ApJS, 194, 45

Shen Y. et al., 2019, ApJS, 241, 34

Strateva I. V., Brandt W. N., Schneider D. P., Vanden Berk D. G., Vignali C., 2005, AJ, 130, 387

Sun M. et al., 2015, ApJ, 811, 42

Svoboda J., Guainazzi M., Merloni A., 2017, A&A, 603, A127

Tananbaum H. et al., 1979, ApJ, 234, L9

Vestergaard M., Peterson B. M., 2006, ApJ, 641, 689

Webb N. A. et al., 2020, A&A, 641, A136

Woo J.-H., Le H. A. N., Karouzos M., Park D., Park D., Malkan M. A., Treu T., Bennert V. N., 2018, ApJ, 859, 138

Yip C. W. et al., 2004a, AJ, 128, 585

Yip C. W. et al., 2004b, AJ, 128, 2603

Zhou H., Wang T., Yuan W., Lu H., Dong X., Wang J., Lu Y., 2006, ApJS, 166, 128

Zhu D., Sun M., Wang T., 2017, ApJ, 843, 30

Zhu S. F., Brandt W. N., Luo B., Wu J., Xue Y. Q., Yang G., 2020, MNRAS, 496, 245

SUPPORTING INFORMATION

Supplementary data are available at *MNRAS* online.

SOUX_Sample_Properties.dat

Please note: Oxford University Press is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

APPENDIX A: UV AND OPTICAL LUMINOSITIES

In Section 3.1, we estimated the UV luminosities at rest-frame 2500 Å from either the optical spectrum of a source, or its *XMM-Newton* OM or SDSS photometry and in Section 5 we used these to calculate $\alpha_{\rm ox}$. Similarly, in Section 3.1 we estimated the optical luminosities at rest-frame 4400 Å from either the optical spectrum of a source or its SDSS photometry and used these to calculate the radio loudness parameter R in Section 3.3. This mixture of luminosity measures from different sources is not ideal and raises the following issues:

(i) Variability: only the OM photometry was recorded simultaneously with the X-ray data but the SDSS spectra and photometry were not.

(ii) Aperture effects: the *XMM* OM photometric aperture (12–35 arcsec diameter) is larger than the SDSS spectral fibre diameter (3 arcsec for the SDSS spectrograph and 2 arcsec for the BOSS spectrograph).

(iii) Flux sampling: the luminosities are arguably best determined from the spectra, from which the flux is measured in a narrow (100 Å wide) window centred on the relevant rest-frame wavelength. In contrast the OM and SDSS photometric bands sample flux from a larger wavelength window: the effective widths of the bands are several hundred Å and therefore they possibly contain some broad-line emission. Whilst we took photometric fluxes from a band containing either 2500 or 4400 Å, the bands will not generally be centred on these wavelengths. The Galactic dust dereddening we have applied to the fluxes will be more accurate for the spectroscopic measures than the photometric ones (although Galactic extinction is very low in the SDSS survey area, and the effect of reddening at 4400 Å is relatively minor anyway).

Here we assess the possible scale of the discrepancies that may be introduced, and their effect on $\alpha_{\rm ox}$ and R across the sample.

For L_{2500} 162 sources have a measurement derived from both the SDSS spectrum and *XMM* OM photometry; 414 sources have a measurement derived from both the spectrum and SDSS photometry. For L_{4400} 588 sources have a measurement derived from both the SDSS spectrum and SDSS photometry (note that we did not use any *XMM* OM photometry to determine L_{4400}). In Fig. A1, we compare the spectroscopic and photometric measures of the luminosities at 2500 and 4400 Å; we have used LINMIX to determine the best-fitting linear regressions and present the values in Table A1. Panels (d) in Fig. A1 show that the relationship between the spectroscopic and photometric luminosities is approximately one-to-one across several orders of magnitude in luminosity, however there is substantial scatter ($\sigma = 0.14$ dex for L_{2500} and 0.09 dex for L_{4400}). We have also calculated the luminosity differences $\Delta \log (L) = \log (L^{\rm spec.}) - \log (L^{\rm phot.})$ and do not find, on average, strong evidence for a systematic offset between the spectroscopic and photometric luminosities with $\langle \Delta \log (L_{2500}) \rangle = -0.02 \pm 0.16$ and $\langle \Delta \log (L_{4400}) \rangle = -0.03 \pm 0.13$. This is clear from panels (g) of Fig. A1 which show the distributions of $\Delta \log (L)$. Panels (e) show $\Delta \log (L)$ as a function of $\log (L^{\rm spec.})$; in these panels we see a weak tendency for photometric measurements to be higher than spectroscopic ones at low spectroscopic luminosities and vice versa. Both $\Delta \log (L_{2500})$ and $\Delta \log (L_{4400})$ are weakly but significantly correlated with their corresponding spectroscopic luminosities.[6]

As panels (f) in Fig. A1 show, we can directly compare spectroscopic and photometric measures of L_{4400}, but not L_{2500}, for AGN at $z \lesssim 0.5$. We would expect to see greater discrepancies at 4400 Å for these low redshifts because (i) aperture effects will be greatest for AGN in resolved host galaxies and (ii) galaxies will be relatively brighter in the optical than the UV. Panel (f) in the L_{4400} plot shows that there is no significant correlation between $\Delta \log (L_{4400})$ and z: the Pearson correlation coefficient is 0.05 with p-value 0.20. This provides some reassurance that the photometric UV fluxes will not be strongly enhanced either. The extremes of the discrepancies along the best-fitting linear regressions are $\Delta \log (L_{2500}) \lesssim \pm 0.1$ dex and $\Delta \log (L_{4400}) \lesssim \pm 0.05$ dex which are small when compared

---

[6]The Pearson correlation coefficient of $\Delta \log (L_{2500})$ with $\log(L_{2500}^{\rm spec.})$ is 0.21 with a p-value $\ll 0.05$; for the correlation of $\Delta \log (L_{4400})$ with $\log(L_{4400}^{\rm spec.})$ we determine a coefficient 0.13 and p-value 0.002.
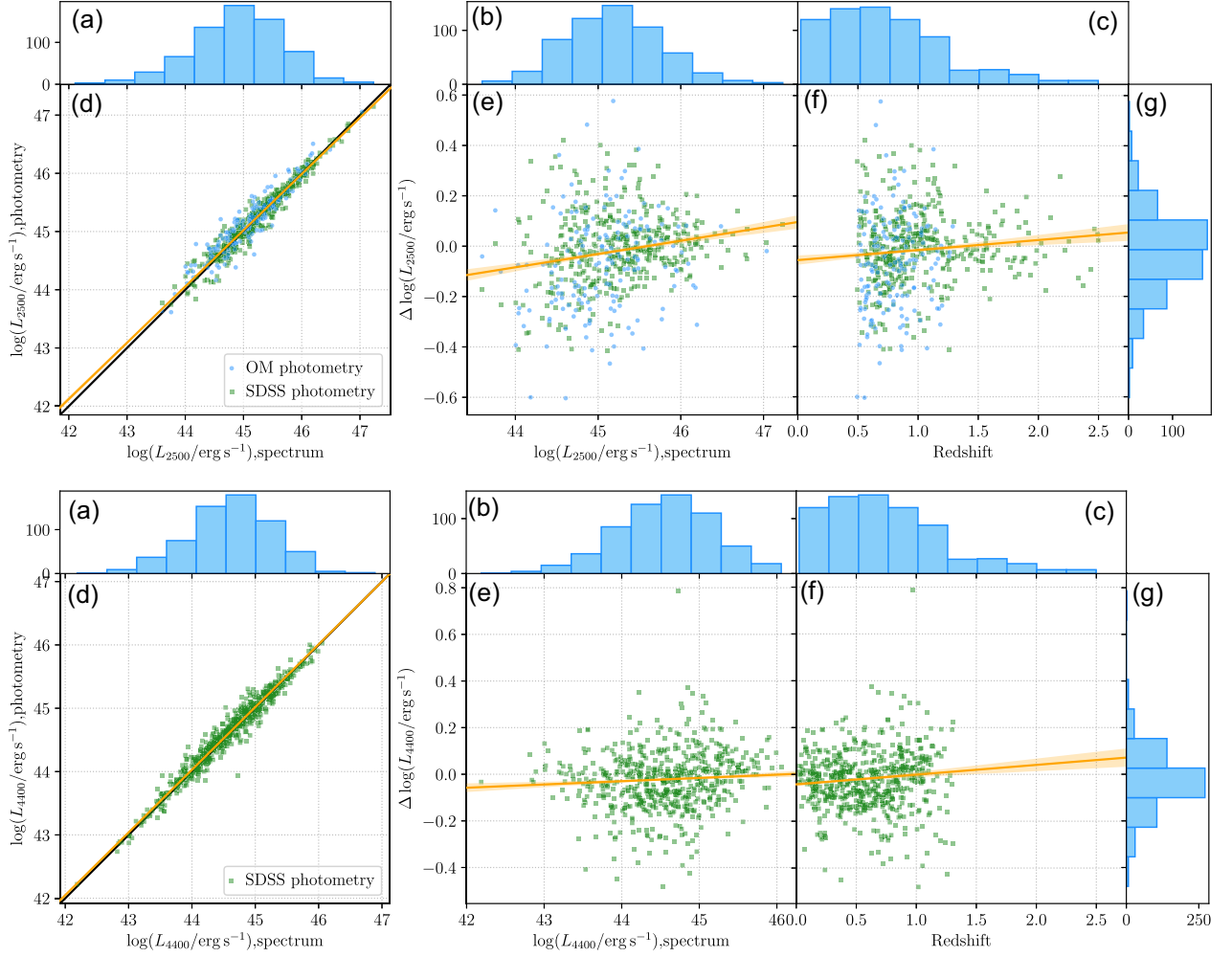
Figure A1. A comparison of spectroscopic and photometric measures of the luminosities at 2500 Å (top) and 4400 Å (bottom). Panels (d) show a scatter plot of the spectroscopic and photometric luminosities; the black line is the one-to-one relation and the orange line in panels (d), (e), and (f) shows the best-fitting linear regression, with the 1σ confidence regions shaded orange. Panels (e) and (f) are scatter plots of the luminosity difference $\Delta\log(L) = \log(L^{\text{spec.}}) - \log(L^{\text{phot.}})$ as a function of $\log(L^{\text{spec.}})$ and redshift, respectively. Histograms of $\Delta\log(L)$ are shown in panels (g). Panels (a) show the distributions of $\log(L)$ for all possible sources. Panels (b) show the distributions of the spectroscopic measures of $\log(L)$ for sources plotted in panels (e). Panels (c) show the redshift distribution of all 696 AGN. Note that we used both *XMM* OM and SDSS photometry to determine L_{2500} but only SDSS photometry to determine L_{4400}.

Table A1. Linear regression parameters for the comparisons of L_{2500} and L_{4400} measurements from spectra and photometry.

| Indep. var.	Dep. var.	Slope	Intercept	σ
$\log\left(L_{2500}^{\text{spec.}}\right)$	$\log\left(L_{2500}^{\text{phot.}}\right)$	0.966 ± 0.012	1.55 ± 0.54	0.14
$\log\left(L_{2500}^{\text{spec.}}\right)$	$\Delta\log\left(L_{2500}\right)$	0.053 ± 0.012	-2.40 ± 0.53	0.14
Redshift	$\Delta\log\left(L_{2500}\right)$	0.040 ± 0.018	-1.06 ± 0.02	0.14
$\log\left(L_{4400}^{\text{spec.}}\right)$	$\log\left(L_{4400}^{\text{phot.}}\right)$	0.990 ± 0.007	0.47 ± 0.33	0.09
$\log\left(L_{4400}^{\text{spec.}}\right)$	$\Delta\log\left(L_{4400}\right)$	0.014 ± 0.007	-0.65 ± 0.33	0.09
Redshift	$\Delta\log\left(L_{4400}\right)$	0.042 ± 0.017	-0.042 ± 0.009	0.09

Note. We define $\Delta\log(L_{2500}) = \log(L_{2500}^{\text{spec.}}) - \log(L_{4400}^{\text{phot.}})$.

with the scatter between the luminosities. We therefore judge that any systematic differences between spectroscopic and photometric luminosities have a negligible effect on our results.

We examined the impact on the radio loudness categorization by determining the number of sources which would change category if their photometric L_{4400} were 'wrong' on the scale of the scatter between spectroscopic and photometric measurements. Only 35 AGN with radio information (6 per cent) have L_{4400} derived from photometry, compared with 593 which have L_{4400} derived from their optical spectrum. We have checked these 35 sources and only three have $|[\log(L_{5\text{GHz}}) - 1] - \log(L_{4400})| > 0.09$ dex (where 0.09 dex is the scatter between measurements determined above) meaning their radio loudness classification would change if the optical luminosities change by ± 0.09 dex. All three were undetected in FIRST, so the radio loudness classifications are only estimates based on the FIRST flux limit (as described in Section 3.3). One source, which we have classified as radio-quiet, would have undetermined radio loudness if it were 0.09 dex optically fainter. The other two have undetermined radio loudness but would be classified as radio-quiet if they were 0.09 dex optically brighter. A change in the radio loudness category of three AGN out of 611 will not have a noticeable effect on our results.

There was a greater mixture of L_{2500} measurements: as stated in Section 3.1, for 421 AGN we have used L_{2500} from the optical spectrum and for 264 AGN (39 per cent) we have used photometry. We therefore test for possible discrepancies in α_{ox} that may result from this mixture. Optical spectroscopic coverage of rest-frame
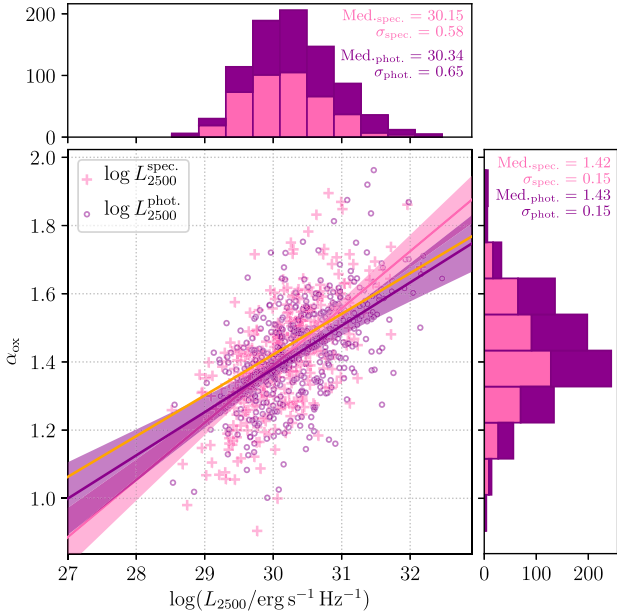
Figure A2. Relationships of $\alpha_{\rm ox}$ and $\log(L_{2500})$ for $\log(L_{2500})$ derived from the optical spectra (pink pluses) and from photometry (purple circles). All 416 sources shown have a measure of $\log(L_{2500})$ from *both* the spectrum and photometry, so they cover the same range in redshift. The pink and purple solid lines show the best-fitting linear regressions for $\log(L_{2500}^{\rm spec.})$–$\alpha_{\rm ox}$ and $\alpha_{\rm ox}$–$\log(L_{2500}^{\rm phot.})$, respectively, with 3σ uncertainty regions shaded in the same colours. The orange line shows the best-fitting linear regression for the full sample (as in Fig. 16). The top and right-hand panels show the distributions of the two quantities for each subset, and we quote the median value and standard deviation.

2500 Å is only possible for sources at $z \gtrsim 0.5$, whereas the wide wavelength range spanned by the UV-optical-infrared photometric filters mean that a L_{2500} determination is possible across the full range of redshifts covered by our sample. To avoid biases introduced by selecting different subsets of AGN in our parameter space, we test only the set of 416 AGN which have *both* a spectroscopic and photometric value of L_{2500}. For each of these AGN, we have two values of L_{2500} and two corresponding values of $\alpha_{\rm ox}$; these are shown in Fig. A2. The plot shows that the distributions of L_{2500} and $\alpha_{\rm ox}$ are slightly different for the spectroscopic and photometric subsets. The median $\log(L_{2500}/\mathrm{erg\,s^{-1}\,Hz^{-1}}) = 30.15$ for the spectroscopic measurements and is slightly higher for the photometric measurements with a median of 30.34. KS and Anderson–Darling tests both indicate that the $\log(L_{2500})$ measurements are drawn from different distributions (the p values of the tests are both <0.001). However, whilst the $\alpha_{\rm ox}$ distributions are not identical, the same statistical tests cannot rule out that the spectroscopically and photometrically derived quantities are drawn from the same distribution (the p values are 0.44 and 0.25, respectively). So we see consistency in the derived $\alpha_{\rm ox}$ values despite a slight difference in $\log(L_{2500})$.

We have recomputed $\log(L_{2500})$–$\alpha_{\rm ox}$ relations for the spectroscopic and photometric quantities separately. The best-fitting linear regressions for the two subsets both show a statistically significant positive correlation between $\alpha_{\rm ox}$ and $\log(L_{2500})$, as was found for the full sample in Section 5 (see Table A2). In Fig. A2 we show these best-fitting relations and their 3σ uncertainty regions. The relation for the spectroscopic subset is slightly steeper (slope 0.168 ± 0.011) than that of the photometric subset (slope 0.127 ± 0.009). In the plot we see that the two relations agree to within 3σ up

Table A2. Linear regression parameters for $\log(L_{2500})$–$\alpha_{\rm ox}$ relations when L_{2500} is derived from either spectra or photometry.

| Subset	N	Slope	Intercept	σ
$\log\left(L_{2500}^{\rm spec.}\right)$	416	0.168 ± 0.011	-3.64 ± 0.32	0.11
$\log\left(L_{2500}^{\rm phot.}\right)$	416	0.127 ± 0.009	-2.43 ± 0.28	0.12
All with $\alpha_{\rm ox}$	685	0.120 ± 0.006	-2.16 ± 0.19	0.13

Note. For the $\log(L_{2500}^{\rm spec.})$ subset $\log(L_{2500})$ has been derived from the optical spectrum and this value is used to determine $\alpha_{\rm ox}$. The $\log(L_{2500}^{\rm phot.})$ subset contains the same 416 sources, but $\alpha_{\rm ox}$ has been determined from the photometrically derived $\log(L_{2500})$. For comparison we also give values for the full set of 685 AGN with $\alpha_{\rm ox}$, which contains a mixture of spectroscopically (421) and photometrically (264) derived quantities.
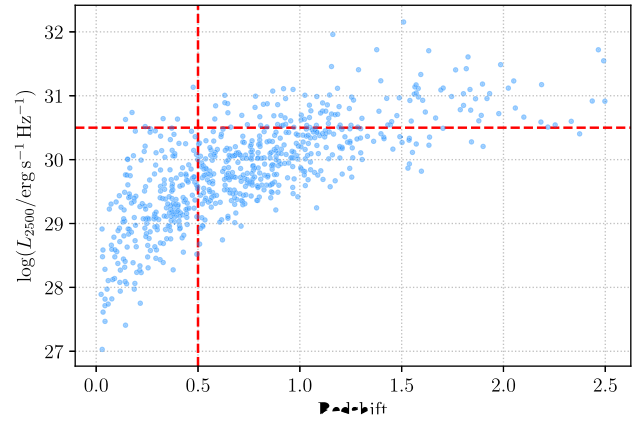
Figure A3. The UV luminosity at 2500 Å versus source redshift. The vertical red dashed line indicates the approximate redshift above which rest-frame 2500 Å is contained within an SDSS spectrum. The horizontal red dashed line indicates the luminosity above which we see a substantial discrepancy in the $\log(L_{2500})$–$\alpha_{\rm ox}$ relations for spectroscopically and photometrically derived quantities (see Fig. A2).

to $\log(L_{2500}/\mathrm{erg\,s^{-1}\,Hz^{-1}}) \approx 30.5$, above which we see a greater discrepancy. Again, we judge that this will have only a minor effect on our results. As stated in Section 3.1, we have taken L_{2500} from the optical spectrum where it is available (which it is for sources at $z \gtrsim 0.5$) and used this value to determine $\alpha_{\rm ox}$. Only where L_{2500} is not available in an optical spectrum will we have estimated L_{2500} and $\alpha_{\rm ox}$ from photometry. Fig. A3 demonstrates that there are very few sources in the region of parameter space in which L_{2500} can only be determined from photometry and the spectroscopic and photometric $\log(L_{2500})$–$\alpha_{\rm ox}$ relations are substantially different. Since there is very good agreement between the spectroscopic and photometric $\log(L_{2500})$–$\alpha_{\rm ox}$ relations in the region of parameter space that contains the majority of our sources, we do not believe that mixing the two types of measurement will have a substantial impact on our results and conclusions. This is corroborated by the fact that we see good agreement between our $\log(L_{2500})$–$\alpha_{\rm ox}$ relations and those of previous studies (e.g. Strateva et al. 2005; Lusso et al. 2010; Zhu et al. 2020).

Based on the findings of previous work, we do not expect the scale of aperture effects to be very substantial in the SOUX sample. In a sample of 51 $z < 0.4$ AGN, Jin et al. (2012a) noticed a discrepancy between flux levels of the SDSS spectra and *XMM* photometry for 17 sources, all with $z \lesssim 0.24$. The photometric points were above the extrapolation from the optical spectra, so it was proposed that the discrepancies were due to aperture effects, as described above.
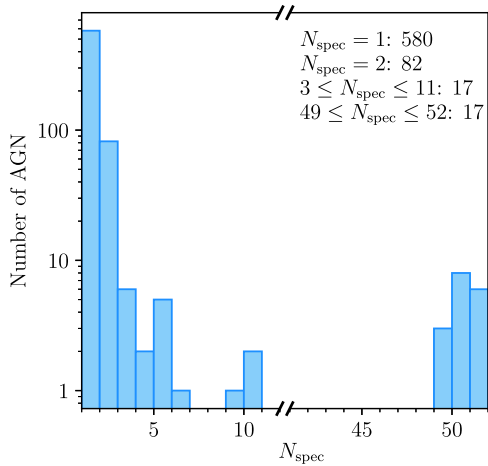
Figure B1. The number of SDSS optical spectra ($N_{\rm spec}$) available for the 696 AGN in our sample, up to and including DR14. The majority of our sample (83 per cent) have only one spectrum, but a small number have $\approx$50 spectra.

To obtain better agreement, Jin et al. (2012a) reanalysed the optical-UV images for affected sources and remeasured the photometry using a smaller aperture, excluding more of the extended host galaxy flux. The optical-UV photometric fluxes were generally reduced by $\sim$50 per cent when using a smaller aperture (see fig. 1 of Jin et al. 2012a). A bespoke reanalysis of the OM photometry for AGN in the SOUX sample is not feasible since it contains hundreds of sources and our investigation here is based on the analysis of standard catalogue and pipeline data, rather than tailored data reductions. Unlike the low-redshift AGN sample of Jin et al. (2012a), the SOUX sample consists of AGN drawn from NLS1 and quasar catalogues, so our AGN are generally UV-bright and not strongly contaminated by their host galaxy flux in the UV.

APPENDIX B: AGN WITH MULTIPLE OPTICAL SPECTRA

Whilst in this study, we consider only one optical spectrum per AGN, many sources have been observed multiple times. Here we show quantitatively the number of spectra for the AGN in the SOUX sample. We use ASTROQUERY (Ginsburg et al. 2019) to search for all optical spectra of our 696 AGN in SDSS DR14. The query performs a 2 arcsec-radius search around the input coordinates (taken from the Pâris et al. 2018 and R17 catalogues) for sources with spectroscopy. We checked that all spectra returned by the query relate to the same object (i.e. they have the same `ObjID`) and that the specific spectrum analysed in this work is one of those. No results were returned by ASTROQUERY for seven AGN; for these we manually checked for all spectra in the SDSS Science Archive Server.[7] We show the

[7] https://dr14.sdss.org/optical/spectrum/search

distribution of the number of spectra in Fig. B1. Whilst 83 per cent of the AGN in our sample have only one optical spectrum, 17 have many (>49). The wealth of optical data for these AGN may be explored further in future studies.

APPENDIX C: QUALITY FLAGS

We assess the quality of measurements determined from the spectral fits by assigning a quality flag. If no issues are raised the quality flag value is 0 and the measurement may be considered to be reliable. The quality flags are integer numbers which are the sum $2^{\rm Bit\,0} + 2^{\rm Bit\,1}... + 2^{\rm Bit\,n}$ for set bits. The bit numbers and the conditions for setting them are listed in Tables C1 and C2. For example, if the fractional uncertainty of the FWHM of a measured broad Hβ line is greater than 2 and the uncertainty on the velocity offset of the line is greater than 1000 km s^{-1} then bits 5 and 8 are set and the Hβ quality flag has the value $2^5 + 2^8 = 32+256 = 288$.

Table C1. Quality flags for the fitted continuum.

Bit	Condition
–	no quality flags raised
0	the luminosity or its uncertainty are zero, NaN, or infinite
1	fractional uncertainty of the luminosity >1.5
2	the continuum slope or its uncertainty are zero, NaN, or infinite
3	the continuum slope has reached the lower or upper limit in the fit
4	the uncertainty on the continuum slope >0.3
5	the continuum fit $\chi_\nu^2 > 50$

Table C2. Quality flags for the fitted emission lines.

Bit	Condition
–	no quality flags raised
0	S/N in the continuum around the line < 3
1	the luminosity or its uncertainty are zero, NaN, or infinite
2	fractional uncertainty of the luminosity >1.5
3	the FWHM or its uncertainty are zero, NaN, or infinite
4	the FWHM ≤ 910 km s^{-1}
5	the fractional uncertainty of the FWHM > 2
6	the velocity offset or its uncertainty are zero, NaN, or infinite
7	the velocity offset has reached the lower or upper limit in the fit
8	the uncertainty on the velocity offset >1000 km s^{-1}

APPENDIX D: TABLE INFORMATION

We have made available online a table containing the details (RA, Dec., z, etc.), measurements (emission line widths, optical, and X-ray continuum luminosities etc.) and derived quantities (black hole masses, $\alpha_{\rm ox}$ etc.) of the 696 AGN in the SOUX sample. Table D1 gives an overview of the columns of that table.

Table D1. The SOUX AGN sample – column information for the table containing the details, measurements, and derived quantities for the 696 AGN.

#	Column name	Units	Description
	Source and SDSS optical spectrum details:		
1	SDSS_ID		Unique SDSS spectrum identifier (PLATE-MJD-FIBER)
2	RA	Degrees	Right Ascension (J2000)
3	DEC	Degrees	Declination (J2000)
4	Z		Spectroscopic redshift
5	DR14Q		0: source is not in the SDSS DR14 quasar catalogue; 1: source is in the SDSS DR14 quasar catalogue
6	R17		0: source is not in the R17 NLS1 catalogue; 1: source is in the R17 NLS1 catalogue
7	SN_RATIO_CONT		Median S/N sampled in emission-line free windows of the spectrum
8	MIN_WAVE	Å	Shortest rest-frame wavelength covered in the optical spectrum
9	MAX_WAVE	Å	Longest rest-frame wavelength covered in the optical spectrum
	UV-optical continuum measurements:		
10	LOG_L1350	erg s^{-1}	Logarithm of the continuum luminosity at 1350 Å (rest-frame) from the SDSS spectrum
11	LOG_L1350_ERR	erg s^{-1}	Uncertainty on LOG_L1350_ERR
12	QUALITY_L1350		Quality flag on LOG_L1350 (see Table C1)
13	LOG_L2500_SPEC	erg s^{-1}	Logarithm of the continuum luminosity at 2500 Å (rest-frame) from the SDSS spectrum
14	LOG_L2500_SPEC_ERR	erg s^{-1}	Uncertainty on LOG_L2500_SPEC
15	LOG_L2500_PHOT_OM	erg s^{-1}	Logarithm of the continuum luminosity at 2500 Å (rest-frame) from *XMM-Newton* OM photometry
16	LOG_L2500_PHOT_OM_ERR	erg s^{-1}	Uncertainty on LOG_L2500_PHOT_OM
17	LOG_L2500_PHOT_SDSS	erg s^{-1}	Logarithm of the continuum luminosity at 2500 Å (rest-frame) from SDSS photometry
18	LOG_L2500_PHOT_SDSS_ERR	erg s^{-1}	Uncertainty on LOG_L2500_PHOT_SDSS
19	LOG_L2500_ALPHA_OX		Which L_{2500} measurement was used to calculate ALPHA_OX (#86)
20	LOG_L3000	erg s^{-1}	Logarithm of the continuum luminosity at 3000 Å (rest-frame) from the SDSS spectrum
21	LOG_L3000_ERR	erg s^{-1}	Uncertainty on LOG_L3000_ERR
22	QUALITY_L3000		Quality flag on LOG_L3000 (see Table C1)
23	LOG_L4400_SPEC	erg s^{-1}	Logarithm of the continuum luminosity at 4400 Å (rest-frame) from the SDSS spectrum
24	LOG_L4400_SPEC_ERR	erg s^{-1}	Uncertainty on LOG_L4400_SPEC
25	LOG_L4400_PHOT_OM	erg s^{-1}	Logarithm of the continuum luminosity at 4400 Å (rest-frame) from *XMM-Newton* OM photometry
26	LOG_L4400_PHOT_OM_ERR	erg s^{-1}	Uncertainty on LOG_L4400_PHOT_OM
27	LOG_L4400_PHOT_SDSS	erg s^{-1}	Logarithm of the continuum luminosity at 4400 Å (rest-frame) from SDSS photometry
28	LOG_L4400_PHOT_SDSS_ERR	erg s^{-1}	Uncertainty on LOG_L4400_PHOT_SDSS
29	LOG_L4400_R_LOUD		Which L_{4400} measurement was used to calculate R_LOUD (#90)
30	LOG_L5100	erg s^{-1}	Logarithm of the continuum luminosity at 5100 Å (rest-frame) from the SDSS spectrum
31	LOG_L5100_ERR	erg s^{-1}	Uncertainty on LOG_L5100_ERR
32	QUALITY_L5100		Quality flag on LOG_L5100 (see Table C1)
	Broad emission line measurements:		
33	FWHM_HA_BR	km s^{-1}	Full-width at half-maximum of the broad Hα emission line
34	FWHM_HA_BR_ERR	km s^{-1}	Uncertainty on FWHM_HA_BR
35	PEAK_HA_BR	Å	Peak wavelength of the broad Hα emission line
36	PEAK_HA_BR_ERR	Å	Uncertainty on PEAK_HA_BR
37	EW_HA_BR	Å	Equivalent width of the broad Hα emission line
38	EW_HA_BR_ERR	Å	Uncertainty on EW_HA_BR
39	LOGL_HA_BR	erg s^{-1}	Logarithm of the luminosity of the broad Hα emission line
40	LOGL_HA_BR_ERR	erg s^{-1}	Uncertainty on LOGL_HA_BR
41	QUALITY_HA		Quality flag on the broad Hα emission line (see Table C2)
42	FWHM_HB_BR	km s^{-1}	Full-width at half-maximum of the broad Hβ emission line
43	FWHM_HB_BR_ERR	km s^{-1}	Uncertainty on FWHM_HB_BR
44	PEAK_HB_BR	Å	Peak wavelength of the broad Hβ emission line
45	PEAK_HB_BR_ERR	Å	Uncertainty on PEAK_HB_BR
46	EW_HB_BR	Å	Equivalent width of the broad Hβ emission line
47	EW_HB_BR_ERR	Å	Uncertainty on EW_HB_BR
48	LOGL_HB_BR	erg s^{-1}	Logarithm of the luminosity of the broad Hβ emission line
49	LOGL_HB_BR_ERR	erg s^{-1}	Uncertainty on LOGL_HB_BR
50	QUALITY_HB		Quality flag on the broad Hβ emission line (see Table C2)
51	BALMER_SOURCE		K23: Balmer lines measurements from this work; R20: Balmer lines measured by R20
52	FWHM_MGII_BR	km s^{-1}	Full-width at half-maximum of the broad Mg II emission line
53	FWHM_MGII_BR_ERR	km s^{-1}	Uncertainty on FWHM_MGII_BR
54	PEAK_MGII_BR	Å	Peak wavelength of the broad Mg II emission line
55	PEAK_MGII_BR_ERR	Å	Uncertainty on PEAK_MGII_BR
56	EW_MGII_BR	Å	Equivalent width of the broad Mg II emission line
57	EW_MGII_BR_ERR	Å	Uncertainty on EW_MGII_BR

Table D1 – *continued*

#	Column name	Units	Description
58	LOGL_MGII_BR	erg s^{-1}	Logarithm of the luminosity of the broad Mg II emission line
59	LOGL_MGII_BR_ERR	erg s^{-1}	Uncertainty on LOGL_MGII_BR
60	QUALITY_MGII		Quality flag on the broad Mg II emission line (see Table C2)
61	MGII_SOURCE		K23: Mg II measurements from this work; R20: Mg II measured by R20
62	MGII_SEP		0: Mg II region modelled in global fit to full spectrum; 1: Mg II region modelled separately to Balmer line regions (see Section 2.3)
63	LINEWIDTH		N: narrow-line type 1 AGN; B: broad-line type 1 AGN (see Section 2.3.1)
	Black hole mass estimates:		
64	MASS_HA_5100_MR16	M$_\odot$	BH mass estimate from FWHM_HA_BR and LOG_L5100 using the relation of MR16
65	MASS_HA_HA_MR16	M$_\odot$	BH mass estimate from FWHM_HA_BR and LOGL_HA_BR using the relation of MR16
66	MASS_HB_5100_MR16	M$_\odot$	BH mass estimate from FWHM_HB_BR and LOG_L5100 using the relation of MR16
67	MASS_HB_HB_G10	M$_\odot$	BH mass estimate from FWHM_HB_BR and LOGL_HB_BR using the relation of Greene et al. (2010)
68	MASS_MGII_3000_MR16	M$_\odot$	BH mass estimate from FWHM_MGII_BR and LOG_L3000 using the relation of MR16
69	MASS_MGII_MGII_W18	M$_\odot$	BH mass estimate from FWHM_MGII_BR and LOGL_MGII_BR using the relation of Woo et al. (2018)
70	MASS_HA_5100_K23	M$_\odot$	BH mass estimate from FWHM_HA_BR and LOG_L5100 using the relation found in this work
71	MASS_HA_HA_K23	M$_\odot$	BH mass estimate from FWHM_HA_BR and LOGL_HA_BR using the relation found in this work
72	MASS_HB_HB_K23	M$_\odot$	BH mass estimate from FWHM_HB_BR and LOGL_HB_BR using the relation found in this work
73	MASS_MGII_3000_K23	M$_\odot$	BH mass estimate from FWHM_MGII_BR and LOG_L3000 using the relation found in this work
74	MASS_MGII_MGII_K23	M$_\odot$	BH mass estimate from FWHM_MGII_BR and LOGL_MGII_BR using the relation found in this work
75	Q_MASS_HA_5100		Quality flag on Hα-5100 Å mass: =1 unless QUALITY_HA and QUALITY_L5100 are both 0
76	Q_MASS_HB_5100		Quality flag on Hβ-5100 Å mass: =1 unless QUALITY_HB and QUALITY_L5100 are both 0
77	Q_MASS_MGII_3000		Quality flag on Mg II-3000 Å mass: =1 unless QUALITY_MGII and QUALITY_L3000 are both 0
78	LOG_MASS_PREF_VALUE	M$_\odot$	Logarithm of our preferred BH mass estimate
79	LOG_MASS_PREF_SOURCE		Which BH mass estimate is the preferred one
	Multiwavelength measurements:		
80	XMM_OBSID		The *XMM-Newton* Observation ID for the X-ray and OM photometric data
81	XMM_SRCID		The unique X-ray source number taken from the *XMM-Newton* catalogue
82	LOG_L2KEV	erg s^{-1} Hz^{-1}	Logarithm of L_ν(2 keV rest-frame) estimated from the *XMM-Newton* band fluxes (see Section 3.2)
83	LOG_L2KEV_ERR	erg s^{-1} Hz^{-1}	Uncertainty on LOG_L2KEV
84	GAMMA_X		Estimated X-ray photon index (see Section 3.2)
85	GAMMA_X_ERR		Uncertainty on GAMMA_X
86	ALPHA_OX		Optical-X-ray energy index $\alpha_{\rm ox}$ calculated from LOG_L2500_ALPHA_OX and LOG_L2KEV
87	ALPHA_OX_ERR		Uncertainty on ALPHA_OX
88	LOG_L5GHZ	W Hz^{-1}	Logarithm of L_ν(5 GHz rest-frame) estimated from the FIRST radio flux (see Section 3.3)
89	LOG_L5GHZ_ERR	W Hz^{-1}	Uncertainty on L5GHZ
90	R_LOUD		Radio loudness calculated from LOG_L4400_R_LOUD and LOG_L5GHZ
91	R_LOUD_FLAG		RL: radio-loud; RQ: radio-detected and radio-quiet; RQu: radio-undetected and radio-quiet; RU: radio-undetected and undetermined radio-loudness; −999: outside FIRST footprint or no L_{4400}

This paper has been typeset from a TeX/LaTeX file prepared by the author.