AGN Physics and Cosmology with Large-Area X-Ray Surveys

Damien Coffey

München 2022

AGN Physics and Cosmology with Large-Area X-Ray Surveys

Damien Coffey

Dissertation
an der Fakultät Physik
der Ludwig–Maximilians–Universität
München

vorgelegt von
Damien Coffey
aus Kilkenny, Irland

München, den 17.10.2022

Erstgutachter: Prof. Kirpal Nandra
Zweitgutachter: Prof. Daniel Gruen
Tag der mündlichen Prüfung: 22.11.2022

Contents

List of Figures

List of Tables

Zusammenfassung

Der wissenschaftliche Schwerpunkt der vorliegenden Arbeit besteht in der Analyse der Wechselwirkung der physikalischen Prozesse von Akkretionsscheiben in aktiven Galaxien mit deren Umgebung. Im ersten Teil der Arbeit wird der Katalog der optischen spektralen Eigenschaften aller im Röntgenbereich detektierten SPIDERS (**SP**ectroscopic **ID**entification of ***eR****OSITA* **S**ources) Galaxien beschrieben, die Teil des Sloan Digital Sky Surveys (SDSS) DR14 Datenkatalogs sind. SPIDERS war ein SDSS-IV-Programm, welches optische Spektroskopie der Röntgenquellen des 2. ROSAT Datenkatalogs (2RXS) und der XMM-Newton "slew" Himmelsdurchmusterung im Rahmen des eBOSS-Projektes durchgeführt hat. Zum Zeitpunkt der Fertigstellung dieser Arbeit war SPIDERS DR14 die größte einheitlich ausgewählte Stichprobe von im Röntgenbereich selektierten AGN mit optischer spektroskopischer Nachbeobachtung. Der hier vorgestellte Katalog basiert auf einer Stichprobe von 9399 2RXS ($0 \leq z \leq 3.94$, $\bar{z} = 0.47$) und 1413 *XMM-Newton*-Durchmusterungsquellen ($0 \leq z \leq 3.94$, $\bar{z} = 0.35$) mit optischer Spektroskopie. Die optischen Spektren wurden einzeln validiert und liefern damit zuverlässige Messwerte für die Quellklassifikation und die Rotverschiebung der Objekte. Die Emissionslinienparameter der Hβ-und MgII-Emissionslinien wurden basierend auf einer genauen Kontinuumsmodelierung bestimmt um die bolometrischen Helligkeiten und die Massen der Schwarzen Löcher bis zu einer Rotverschiebung von $z \simeq 2.5$ abzuleiten. Es wurde ein Vergleich zwischen den optischen spektralen Eigenschaften und den aus Hβ und MgII abgeleiteten Masseschätzungen für Schwarze Löcher durchgeführt für Objekte die beide Linien in ihrem Spektrum abdecken. Diese Zusatzinformationen zum SDSS-IV DR14 Katalog wurden öffentlich zugänglich gemacht.

Im zweiten Teil dieser Arbeit wird untersucht wie Quasare als Standardkerzen für kosmologische Parameterableitungen verwendet werden können. Es ist bekannt, dass eine nichtlineare Beziehung zwischen der ultravioletten (UV) Emission von Akkretionsscheiben und der Röntgenemission der umgebenden Korona existiert. Diese Beziehung wurde genutzt, um Quasare als Standardkerzen in ein Hubble-Diagramm aufzunehmen, um die kosmologischen ΛCDM-Modellparameter Ω_M und Ω_Λ zu bestimmen (Risaliti & Lusso 2015). In dieser Arbeit wurde die Beziehung zwischen Röntgen- und UV-Emission in Quasaren anhand von zwei großen Stichproben in weiteren Wellenlängen untersucht, die von *ROSAT* und *XMM-Newton* entdeckt wurden. Es wurde untersucht ob die in Risaliti & Lusso (2015) erzielten Ergebenisse genauer bestimmt werden können, indem Ω_M und Ω_Λ unabhängig voneinander über das Hubble-Diagramm der Quasare gemessen wurden unter Verwendung von *ROSAT* und *XMM-Newton* Himmelsdurchusterungsdaten. Es wird

gezeigt, dass diese Datensätze inkonsistente Messungen von $\Omega_{\rm M}$ und Ω_Λ liefern.

Um gleichzeitig die Grenzen der *ROSAT*- und *XMM-Newton* Daten zur Bestimmung kosmologischer Parameter zu erforschen und um Vorhersagen für die mögliche Nutzung zukünftiger großer Datensätze in Himmelsdurchmusterungen zu erzielen, wurden Simulationen basierend auf dem *eROSITA*-4MOST AGN Datensatz entwickelt. *eROSITA* wird voraussichtlich $\sim 3{\times}10^6$ AGN bis zum Ende seiner Durchmusterung des gesamten Himmels detektieren. Etwa ein Drittel dieser Quellen wird mittels 4MOST optische spektroskopiert werden. Eine Methode zur Auswahl einer *eROSITA*-4MOST-Stichprobe mit einer Detektionsschwelle, die für die Verwendung im Quasar-Kosmologie-Experiment ausreicht, wurde skizziert. Mit dieser simulierten *eROSITA*-4MOST-Stichprobe wurden die kosmologischen Parameter $\Omega_{\rm M}$ und Ω_Λ mit $\Omega_{\rm M} = 0.253^{+0.079}_{-0.061}$ und $\Omega_\Lambda = 0.751^{+0.271}_{-0.393}$ bestimmt. Dieser simulierte Datensatz wurde dann verwendet, um viele der offenen Fragen im Zusammenhang mit dem Quasar-Kosmologie-Experiment zu klären, einschließlich der Auswirkungen der Probengröße, die Genauigkeit zur Bestimmung der Flussdichten, der intrinsischen Streuung in der Quasar-Röntgen-UV-Beziehung und der Auswirkungen zur Genauigkeit der Bestimmung der Regressionsparameter.

Abstract

The main scientific focus that permeates this work is the analysis of the interaction between the quasar accretion disk and the surrounding environment. The first part of this work presents the catalogue of optical spectral properties for all X-ray selected SPIDERS active galactic nuclei prepared as part of the fourteenth Sloan Digital Sky Survey data release (SDSS DR14). SPIDERS (**SP**ectroscopic **ID**entification of ***eR****OSITA* **S**ources) was an SDSS-IV programme that conducted optical spectroscopy of the counterparts to the X-ray selected sources detected in the *ROSAT* all-sky survey and the *XMM-Newton* slew survey in the footprint of the Extended Baryon Oscillation Spectroscopic Survey (eBOSS) project. At the time this work was completed, the SPIDERS DR14 sample was the largest uniformly selected sample of X-ray selected AGN with optical spectroscopic follow-up. The catalogue presented here is based on a clean sample of 9399 2RXS ($0 \leq z \leq 3.94$, $\bar{z} = 0.47$) and 1413 *XMM-Newton* slew survey sources ($0 \leq z \leq 3.94$, $\bar{z} = 0.35$) with optical spectra. Visual inspection results for each object in this sample are available from a combination of literature sources and the SPIDERS group, which provide both reliable redshifts and source classifications. The spectral regions around the Hβ and MgII emission lines have been fit in order to measure both line and continuum properties, estimate bolometric luminosities, and provide black hole mass estimates using the single-epoch (or photoionisation) method. The use of both Hβ and MgII allowed black hole masses to be estimated up to $z \simeq 2.5$. A comparison was made between the spectral properties and black hole mass estimates derived from Hβ and MgII using the subsample of objects which have coverage of both lines in their spectrum. These results have been made publicly available as an SDSS-IV DR14 value added catalogue.

The second part of this work presents an analysis of the potential of adopting quasars as standard candles. A non-linear relationship is known to exist between the ultraviolet (UV) emission from quasar accretion disks and the X-ray emission from the surrounding corona. This relationship has been used to include quasars as standard candles in a Hubble diagram in order to measure the ΛCDM cosmological model parameters $\Omega_{\rm M}$ and Ω_{Λ} (Risaliti & Lusso 2015). In this work, the relationship between X-ray and UV emission in quasars was investigated using two large multiwavelength quasar samples detected by *ROSAT* and *XMM-Newton*. An attempt was made to further the work presented in Risaliti & Lusso (2015) by independently measuring $\Omega_{\rm M}$ and Ω_{Λ} via the quasar Hubble diagram using survey-quality data from *ROSAT* and *XMM-Newton*. These datasets were found to produce inconsistent measurements of $\Omega_{\rm M}$ and Ω_{Λ}.

In order to simultaneously explore the limitations of the *ROSAT* and *XMM-Newton* datasets, and to provide forecasts for the potential use of future large survey-quality datasets in the quasar cosmology experiment, simulations of the *eROSITA*-4MOST AGN sample were developed. *eROSITA* is expected to detect $\sim 3 \times 10^6$ AGN by the end of its all-sky survey, and up to a third of these sources will be targets for follow-up optical spectroscopy by 4MOST. A method for selecting an *eROSITA*-4MOST subsample with flux density precision sufficient for use in the quasar cosmology experiment was outlined. Using this simulated higher-precision *eROSITA* - 4MOST subsample, the cosmological parameters $\Omega_{\rm M}$ and Ω_{Λ} were measured as $\Omega_{\rm M} = 0.253^{+0.079}_{-0.061}$ and $\Omega_{\Lambda} = 0.751^{+0.271}_{-0.393}$. This simulated dataset was then used to address many of the open issues relating to the quasar cosmology experiment including the effects of sample size, flux density precision, intrinsic scatter in the quasar X-ray - UV relationship, and the impact of regression dilution.

List of Acronyms

AGN	Active Galactic Nucleus
AIS	All-Sky Imaging Survey (GALEX)
APO	Apache Point Observatory
Art-XC	Astronomical Roentgen Telescope - X-ray Concentrator
BAO	Baryon Acoustic Oscillations
BALQSO	Broad Absorption Line Quasi-Stellar Object
BCES	Bivariate Correlated Errors and Intrinsic Scatter
BH	Black Hole
BIC	Bayesian Information Criterion
BLAGN	Broad Line Active Galactic Nucleus
BLR	Broad Line Region
BOSS	Baryon Oscillation Spectroscopic Survey
BXA	Bayesian X-ray Analysis
CCD	Charge-Coupled Device
CDFS	Chandra Deep Field South
CGS	Centimetre-Gram-Second
CMB	Cosmic Microwave Background
DIS	Deep Imaging Survey (GALEX)
DM	Distance Modulus
DOF	Degrees of Freedom

DR	Data Release
eBOSS	Extended Baryon Oscillation Spectroscopic Survey
eFEDS	eROSITA Final Equatorial Depth Survey
EPIC	European Photon Imaging Camera
eRASS	eROSITA All-Sky Survey
eROSITA	Extended ROentgen Survey with an Imaging Telescope Array
ESA	European Space Agency
FUV	Far-Ultraviolet
FWHM	Full Width at Half Maximum
GALEX	Galaxy Evolution Explorer
HRI	High Resolution Imager
IR	Infrared
LOS	Line of Sight
LHDF	Lockman Hole Deep Field
MCMC	Markov Chain Monte Carlo
MIS	Medium Imaging Survey (GALEX)
MJD	Modified Julian Date
MOS	Metal Oxide Semiconductor
MPE	Max-Planck Institute for Extraterrestrial Physics
MW	Milky Way
NASA	National Aeronautics and Space Administration
NLAGN	Narrow Line Active Galactic Nucleus
NLR	Narrow Line Region
NLS1	Narrow Line Seyfert 1
NUV	Near-Ultraviolet
OLS	Ordinary Least Squares

OM	Optical Monitor
PI	Principal Investigator
PSF	Point Spread Function
PSPC	Position Sensitive Proportional Counters
QSO	Quasi-Stellar Object
RASS	ROSAT All-Sky Survey
RGS	Reflection Grating Spectrometer
RM	Reverberation Mapping
ROSAT	ROentgen SATellite
SCP	Supernovae Cosmology Project
SDSS	Sloan Digital Sky Survey
SED	Spectral Energy Distribution
SEQUELS	Sloan Extended Quasar, Emission Line Galaxy, and Luminous Red Galaxy Survey
SFR	Star Formation Rate
SMBH	Supermassive Black Hole
SMC	Small Magellanic Cloud
SN	Supernovae
SNe 1a	Type Ia Supernovae
SPIDERS	SPectroscopic IDentification of eROSITA Sources
SRG	Spectrum-Roentgen-Gamma
TDSS	Time Domain Spectroscopic Survey
UV	Ultraviolet
VBC	Very Broad Component
VEILS	VISTA Extragalactic Infrared Legacy Survey
VISTA	Visible and Infrared Survey Telescope for Astronomy

WFC	Wide-Field Camera
WISE	Wide-field Infrared Survey Explorer
XMMSL1	First *XMM-Newton* Slew Survey
XRT	X-Ray Telescope
XSPEC	X-Ray Spectral Fitting Package
2RXS	Second ROSAT All-Sky Survey
4DE1	4-Dimensional Eigenvector 1
4MOST	4-Metre Multi-Object Spectroscopic Telescope
ΛCDM	Λ Cold Dark Matter

Chapter 1

Introduction

Parts of this chapter have also appeared in Coffey et al. (2019).

1.1 Observing AGN: Scientific Motivation

Of the rich variety of astrophysical phenomena discovered to date, observations of active galactic nuclei (AGN) rank among the most significant in terms of furthering our understanding of the evolution and distribution of matter in the Universe. They are the most energetic non-transient natural phenomenon known, often outshining their host galaxies by a factor of a thousand. The immediate vicinity of a supermassive black hole (SMBH) offers a remote laboratory for testing physical processes in one of the most extreme environments in nature. The number density and level of activity of AGN have evolved throughout cosmological time, which was one of the observations that led the scientific community to reject the steady-state model of the Universe. The detection of AGN above $z = 7$ (Mortlock et al. 2011) has made them valuable probes of the early Universe, and an anchor for theoretical models of physical conditions less than 1 Gyr after the Big Bang. Most massive galaxies are expected to host a SMBH, which suggests that they are an integral component of galactic structure and evolution (e.g. Gebhardt et al. 2000).

After decades of increasingly intense study, AGN continue to pose a number of questions which are the subject of ongoing research: From what seeds did the SMBHs currently observed originally form? To what extent do SMBHs have angular momentum? What is the physical nature of the accretion flow onto SMBHs? What mechanism triggers the onset of accretion onto SMBHs? What is the nature of the physical interaction between the accretion disk and the surrounding electron corona? How are AGN jets launched? Does AGN feedback regulate the star-formation of the host galaxy? From an empirical perspective, a detailed census of AGN observational properties is essential in order to facilitate progress in these open areas of inquiry.

A crucial requirement for the exploration of AGN demographics and evolution is the ability to select a dataset for analysis that is as complete and unbiased as possible. Wide-area optical spectroscopic surveys, such as the Sloan Digital Sky Survey (SDSS; York

et al. 2000; Gunn et al. 2006), have provided the opportunity to classify and analyse large numbers of objects. However, these catalogues contain their own intrinsic biases and incompletenesses, which may misrepresent the true AGN population. Carefully chosen selection criteria are therefore required in order to extract a uniform subsample. X-ray observations have frequently been used for AGN selection since the high-energy emission associated with mass accretion by a black hole distinguishes AGN from inactive galaxies and stars. Combining wide-area X-ray surveys with the ability to classify large numbers of objects spectroscopically via the SDSS provides a powerful resource for the study of AGN.

This work adopts an empirical approach to the study of the observational signatures of SMBH accretion. An overview of the main scientific goals of this work is presented in section 1.4. First, however, a brief history of AGN research will be presented along with an overview of the current understanding of the physical structure of AGN.

1.2 The Discovery of Quasars and Historical Context

While studying the then poorly understood "spiral nebulae[1]", Fath (1909) measured the first optical spectrum of an AGN; the Seyfert 2 galaxy NGC 1068. A second study of NGC 1068 was published a few years later by Slipher (1917), who identified a series of redshifted emission lines and suggested that the source was "receding from the Sun". The first systematic study of galaxies with nuclear emission lines was carried out by Seyfert (1943). Seyfert (1943) presented spectra that exhibited blue continuum emission and broad optical emission lines (FWHM[2] $> 8,500\,\mathrm{km\,s^{-1}}$), which they suggested were the result of Doppler motion. Radio surveys carried out throughout the 1950s (e.g. the 3C and 3CR surveys; Edge et al. 1959; Bennett 1962) revealed many point-like sources, prompting efforts to identify their optical counterparts. Baade & Minkowski (1954) were the first to present optical counterparts for the strong sources of radio emission Cassiopeia and Cygnus A. Interest in the objects studied by Seyfert (1943) then grew after Baade & Minkowski (1954) highlighted the similarity of the Seyfert galaxies' optical spectra to that of the Cygnus A radio source. The name "Seyfert" would eventually be given to galaxies whose spectra exhibit high-excitation nuclear emission lines.

An early insight into the nature of these objects came from Woltjer (1959) who studied the sample of Seyfert galaxies published by Baade & Minkowski (1954). Woltjer (1959) concluded that the observed nuclear emission originated from a compact, massive region, is persistent over long timescales ($> 10^8$ years), and may be a common occurrence in galaxies. Woltjer (1959) also suggested that the observed emission line broadening was caused by the emitting material's motion in the gravitational field of the nucleus. However, at the time, it was believed that the phenomena observed in galactic nuclei were the result of a dense central concentration of stars.

Many radio sources that lay out of the plane of the Milky Way were associated with galaxies that were resolved at optical wavelengths. However, some radio sources were found

[1] At the time, galaxies were referred to as "nebulae".

[2] Full width at half maximum.

to have point-like optical counterparts. The first of these was 3C 48 (Matthews & Sandage 1963), whose spectrum showed broad emission lines that were unidentified at the time. In 1963, Maarten Schmidt presented an optical counterpart to the radio source 3C 273, and for the first time recognised that the observed broad emission lines were MgII and the Hydrogen Balmer series, but strongly redshifted ($z = 0.158$, Schmidt 1963). This placed 3C 273 at a distance which was considered to be controversially large at the time. Soon after, groups began to discover objects with much higher redshifts, an example of which is the discovery of a quasar at $z = 2.223$ by Arp et al. (1967). These objects were originally referred to as quasi-stellar objects (QSO), or quasi-stellar radio sources, names which were later contracted to the term "quasar".

The extreme distances and luminosities associated with quasars implied an exotic and powerful source of energy. Hoyle & Fowler (1963a,b) were the first to suggest that quasars may be the result of the gravitational collapse of an extremely high mass star-like object. The following year, Salpeter (1964) and Zeldovich & Novikov (1964) independently proposed that the vast amounts of energy emitted by quasars may be produced by black holes accreting matter from their surroundings. This insight came just one year after the interpretation of the redshift of 3C 273 by Schmidt (1963), however, a large segment of the scientific community did not accept the black hole hypothesis at the time. Further support for a central black hole came from the work of Lynden-Bell (1969), who suggested that "collapsed bodies" (or black holes) lie at the centre of nearby galaxies that once were active quasars. This interpretation gradually gained favour since it accounted for many of the inferred and observed properties of quasars in addition to providing a mechanism for the release of the large amounts of energy associated with these objects.

1.3 AGN Structure and Emission Regions

1.3.1 AGN Unified Model

Initially, AGN were studied and classified at different wavelengths, which resulted in a slew of different classification schemes (e.g. Seyfert 1 and 2, radio quiet and loud, etc.). At first, however, the connection between these classes was unclear. Despite the myriad observed properties, AGN were found to exhibit three primary variable characteristics: the presence or absence of broad optical permitted emission lines, radio emission strength, and bolometric luminosity (Tadhunter 2008). Attempts at defining a model that would account for these variables began with two different approaches; the unification of radio properties and the unification of optical/X-ray properties.

Evidence for an obscuring medium in the inner regions of AGN was initially provided by X-ray observations which revealed that Seyfert 2 AGN exhibit obscuration by gas while Seyfert 1 AGN do not (Lawrence & Elvis 1982). This distinction between Seyfert 1 and 2 AGN led to the development of AGN unification models that were built on the concept of classification based on orientation. The first sketch of the inner obscuring structure of AGN was presented by Antonucci (1984) and was based on a study of the radio and

optical properties of a sample of radio galaxies. Antonucci (1984) found that Seyfert 1 and 2 AGN have polarisations that align parallel and perpendicular to the extended radio emission axis, respectively. The author also concluded that the Seyfert 2 perpendicular alignments may arise from the scattering of emission off an optically thick torus-shaped absorber. The following year, Antonucci & Miller (1985) presented polarisation spectra of the Seyfert 2 AGN NGC 1068 in which they discovered broad Balmer emission lines. This work suggested that Seyfert 1 and 2 AGN are intrinsically the same type of object, but that the broad line region (BLR) of Seyfert 2 AGN is obscured behind an optically thick medium.

The unification model of AGN (Antonucci 1993; Urry & Padovani 1995) postulates a black hole at the centre of the system, surrounded by an accretion disk emitting at optical and ultraviolet (UV) wavelengths (see figure 1.1). Above the black hole-accretion disk system lies a high-temperature, X-ray emitting, electron plasma known as the "corona". At larger distances, broad emission line clouds orbit the black hole in Keplerian motion. Beyond the BLR, an axisymmetric, optically-thick, obscuring structure (the torus) provides orientation-dependent obscuration of the interior, and re-emits radiation incident from the accretion disk at infrared (IR) wavelengths. An unobstructed view of the BLR, corona, and accretion disk is possibly only at a high viewing angle, where the torus does not intersect the line of sight. A source viewed at this angle would be classified as a Seyfert 1 AGN, or broad line quasar, as shown in figure 1.1. Conversely, Seyfert 2 AGN are observed at lower viewing angles where the inner structure is obscured by the torus. Further detail on the physical structure of AGN is presented in the following sections.

1.3.2 AGN Physical Structure

Accretion Disk

The highly-energetic phenomena associated with AGN are theorised to be powered by an accretion disk around the central black hole. At each radius the disk emits as a blackbody, with the disk temperature increasing towards smaller radii from the black hole. The superimposed blackbody spectra are emitted at optical and UV wavelengths, and are referred to as the "big blue bump".

Very high accretion rates can result in radiation pressure sufficient to repel material from the disk and halt accretion. The maximum accretion disk luminosity that can be expected is therefore reached when the radiation pressure on electrons exceeds the gravitation force exerted on protons by the black hole. This is known as the Eddington luminosity;

$$L_{Edd} = 4\pi c G M_{BH} m_p / \sigma_T$$

where c is the speed of light, G is the gravitational constant, M_{BH} is the black hole mass, m_p is the proton mass, and σ_T is the Thomson scattering cross-section.

X-Ray Corona

One of the defining characteristics of AGN is that they are powerful sources of persistent high-energy radiation, emitting a large portion ($\sim 10\%$, Peterson 1997) of their bolometric luminosity at X-ray wavelengths. This X-ray emission is generated when UV and optical photons from the accretion disk are increased in energy via inverse Compton scattering in a high-energy electron plasma referred to as the corona (Fabian et al. 1989; Haardt & Maraschi 1991; Zdziarski et al. 1994). Schnittman et al. (2013) have shown that the corona may form as a result of magnetohydrodynamic turbulence in the accretion disk.

The X-ray variability observed in AGN indicates that the corona is a relatively compact region, possibly only a few gravitational radii from the central black hole (e.g. Uttley et al. 2014). The process of Comptonisation in the corona produces a power law X-ray spectrum of the form

$$\mathrm{F(E)} = \mathrm{NE}^{-\Gamma}$$

which represents the flux, F, as a function of energy, E. Nandra & Pounds (1994) have found that for a sample of 27 low-redshift Seyfert galaxies the intrinsic X-ray continuum slope, Γ, is 1.95 ± 0.05. Inverse Compton scattering can take place when the seed photon energies are less than the thermal energy of the electrons in the corona. Photons that have energies approximately equal to the corona thermal energy do not undergo inverse Compton scattering. This results in the high energy cut-off of the X-ray power law (e.g. Sunyaev & Truemper 1979).

Broad and Narrow Line Regions

Superimposed on a blue optical - UV continuum, AGN exhibit broad emission lines, the most prominent of which are the hydrogen Balmer series (Hα [6, 563Å], Hβ [4, 861Å], etc.), the hydrogen Lyman α line (Lyα [1, 216Å]), the singly-ionised magnesium doublet (MgII [2, 796, 2, 803Å]), as well as triply-ionised carbon (CIV [1, 549Å]). To a first approximation, the BLR emission lines are assumed to be Doppler-broadened, which is a result of the Keplerian motion of the line emitting gas around the central black hole. This motion results in line widths of a few thousand $\mathrm{km\,s^{-1}}$, however, widths in excess of ten thousand $\mathrm{km\,s^{-1}}$ have also been observed. Electron densities in the BLR are expected to be high ($\sim 10^9\,\mathrm{cm^{-3}}$, or as high as $\sim 10^{11}\,\mathrm{cm^{-3}}$ for the Lyα-CIV emitting BLR, see Ferland et al. (1992)), which results in the complete collisional suppression of all forbidden transitions.

The BLR is expected to lie inside the obscuring torus, with radii that extend from light-months for lower luminosity Seyfert galaxies, up to light-years for high luminosity quasars, with a natural outer boundary set by the dust sublimation radius which itself is dependent upon the dust grain temperature (Barvainis 1987). Broad emission line variability has been observed to correlate with variations in the continuum emission, which suggests photoionisation as the line production mechanism. Since the BLR reprocesses the incident accretion disk emission, and its kinematics are dominated by the gravitational influence of the central black hole, spectral measurements of the broad emission lines are an important

tool for understanding the central engine of AGN. In section 1.5, the method for inferring black hole mass from measurements of AGN broad emission lines will be presented.

At much larger scales (hundreds to thousands of parsecs), AGN ionise the interstellar medium of their host galaxies. This produces emission lines from regions whose kinematics are not dominated by the gravitational influence of the central black hole, but instead by the gravitational potential of the entire galaxy. This results in lines with FWHM typically in the range of 350 - 400 $\mathrm{km\,s^{-1}}$ (Peterson 1997), and therefore this extended emission region is referred to as the "narrow line region" (NLR). Since the NLR electron densities are relatively low ($\sim 10^3$ - $10^5\,\mathrm{cm^{-3}}$, Beckmann & Shrader 2012), the time between collisions is longer than the forbidden transition decay time. Therefore, unlike the BLR, forbidden transitions are not collisionally suppressed in the NLR, resulting in a rich diversity of emission features which include low-ionisation permitted lines (e.g. Hβ [4, 861Å]) and high-ionisation forbidden lines (e.g. [OIII] [4, 959, 5, 007Å]).

Torus

As discussed in section 1.3.1, unification models propose that the observed differences between Seyfert 1 and 2 AGN is that the latter exhibit obscuration of the BLR by an axisymmetric, dust-filled, toroidal obscurer (Antonucci 1993; Urry & Padovani 1995). Conversely, Seyfert 1 AGN are orientated such that the line of sight to the BLR does not intersect the torus. The ratio of the numbers of observed Seyfert 1 and 2 AGN allows the opening angle of the torus to be estimated (found to be 48 deg by Schmitt et al. 2001). Long timescale X-ray monitoring of Seyfert galaxies has revealed that cloud transit events occur, suggesting that the torus is a clumpy agglomeration of clouds rather than a smooth homogeneous medium (e.g., Markowitz et al. 2014). Elitzur & Shlosman (2006) suggested that the torus may be formed by an accretion disk wind, which consists of clumpy, dusty, optically thick clouds. The dust sublimation radius is considered to be the inner boundary of the torus, however, the BLR and torus may form a continuous structure that changes from gas-dominated to dust-dominated at the sublimation radius. As in the case of the BLR, the torus is expected to be within the region dominated by the gravitational influence of the central black hole.

The torus absorbs part of the incident emission from the accretion disk, corona, and BLR, and reemits it across the $\sim$2 - 1,000 μm spectral region (Sanders et al. 1989). Some of the optical emission is scattered into the line of sight as evidenced by the detection of broad Balmer emission lines in the polarisation spectrum of the Seyfert 2 AGN NGC 1068 (Antonucci & Miller 1985). In addition, the torus Compton scatters incident X-ray emission from the corona, and the resulting Compton reflection spectrum can be used to infer the geometry of the obscuring region (e.g., Buchner et al. 2014).

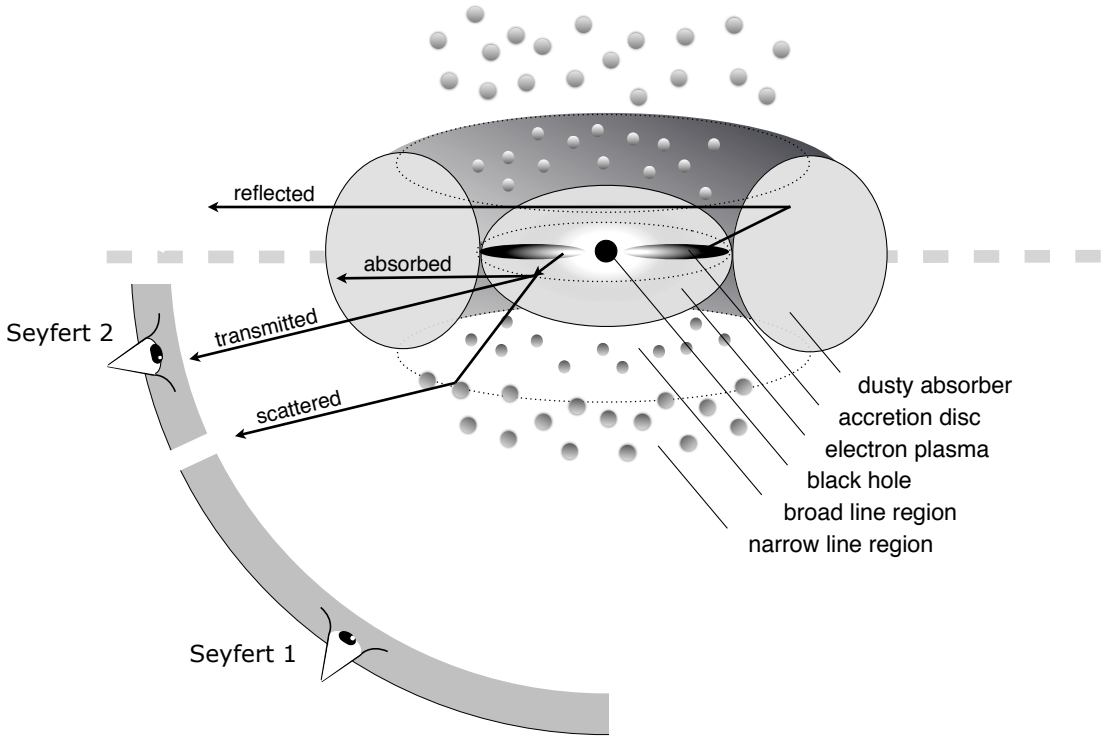


Figure 1.1: Schematic of the inner structure of AGN based on the AGN unified model. Adapted from figure 1.1 from Beckmann & Shrader (2012), illustrated by Marie-Luise Menzel.

1.4 Scientific Overview of this Work

The primary scientific goal that permeates this work is the measurement of the interaction between the central AGN accretion disk and the surrounding environment (the X-ray corona and the BLR). This work presents an empirical analysis of that relationship, and explores a number of scientific applications that stem from it: the photoionisation of the BLR gas by the central ionising source has allowed the black hole mass to be estimated (chapter 2). This in turn has been used to investigate the phenomenon of black hole mass downsizing (section 3.3). An introduction to the methods used for estimating black hole mass will be presented in the following section. A decomposition of the Hβ lines of a large number of AGN hints at the presence of an inner BLR dominated by inflow, possibly associated with the inner edge of the accretion disk (section 3.2.1).

The relationship between accretion disk emission and the resulting emission from the surrounding corona has been used by Risaliti & Lusso (2015, hereafter RL15) to calibrate AGN as standard candles (see section 1.6.2). Part of the work presented in this thesis aims to determine whether survey quality observations of AGN can be calibrated as cosmological distance indicators. The application of datasets that were available at the commencement of this work from *ROSAT* and *XMM-Newton* to the quasar Hubble diagram has been investigated, and is presented in chapter 4. The potential application of future *eROSITA* - 4MOST datasets to the quasar cosmology experiment has also been assessed, and is discussed in chapter 5.

As mentioned above, this work adopts an observational approach to the study of AGN using the catalogues of X-ray selected AGN available from the SPIDERS survey and simulations of the future *eROSITA* - 4MOST catalogues. These catalogues are based on X-ray observations from the *ROSAT*, *XMM-Newton*, and *eROSITA* X-ray observatories, which are discussed in section 1.7.1. SDSS and 4MOST, the programmes which provide optical spectroscopic follow-up for the X-ray detections, are discussed in section 1.7.2.

1.5 Measuring Black Hole Mass

An accurate measurement of the SMBH mass is necessary for the study of AGN and their coevolution with their host galaxies. Black hole mass has been found to scale with a number of host galaxy spheroid properties; stellar velocity dispersion (the M_{BH} - σ relation, e.g. Gebhardt et al. 2000; Merritt & Ferrarese 2001; Tremaine et al. 2002), stellar mass (e.g. Magorrian et al. 1998; Häring & Rix 2004), and luminosity (e.g. Kormendy & Richstone 1995). These correlations suggest a symbiotic evolution of SMBHs and their host galaxies.

1.5.1 Reverberation Mapping

Reverberation mapping (RM) has been used to measure the approximate radius of the BLR in AGN (e.g. Bahcall et al. 1972; Capriotti et al. 1982; Blandford & McKee 1982; Peterson 1993; Bentz & Katz 2015; Shen et al. 2015). This technique involves measuring

the time delay between variations in the continuum emission, which is expected to originate in the accretion disk, and the induced variations in the broad emission lines. It was found that different emission lines have different time delays, which is what would be expected for a stratified BLR, with emission lines of lower ionisation being produced further from the central ionising source (e.g. Gaskell & Sparke 1986). An example of this is the time delay of the high-ionisation line CIV, which has been found to be shorter than that of the lower-ionisation Hβ line (Peterson & Wandel 2000).

The RM effort has also revealed a tight relationship between the accretion disk continuum luminosity and the radius of the BLR (Kaspi et al. 2000; Bentz et al. 2006, 2009a). Therefore, by using the measured luminosity as a proxy for the BLR radius, and measuring the BLR line-of-sight velocity from the width of the broad emission lines, black hole masses can be estimated from a single spectrum (Vestergaard 2002; McLure & Jarvis 2002; Vestergaard & Peterson 2006; Assef et al. 2011; Shen & Liu 2012; Shen 2013). This approach is known as the single-epoch, or photoionisation, method.

Hβ is the most widely studied RM emission line to date, and it is also considered to be the most reliable line to use for single-epoch mass estimation. In addition, AGN Hβ emission lines typically exhibit a clear inflection point between the broad and narrow line components, making the virial FWHM measurement relatively straightforward (see section 2.2.4). The MgII line width correlates well with that of Hβ (see section 3.1.2), and therefore MgII has also been used for single-epoch mass estimation (e.g. McLure & Jarvis 2002). For SDSS spectra, either MgII or Hβ is visible in the redshift range $0 \leq z \lesssim 2.5$.

At higher redshifts, the broad, high-ionisation line CIVλ1549 is available. The CIV line width does not correlate strongly with that of low ionisation lines (e.g. Baskin & Laor 2005; Trakhtenbrot & Netzer 2012) and this, along with the presence of a large blueshifted component (e.g. Richards et al. 2002) makes it difficult to employ CIV for mass estimation. A number of calibrations have been developed which aim to improve the mass estimates derived from CIV (Denney 2012; Runnoe et al. 2013; Park et al. 2013; Coatman et al. 2017), however, whether CIV can provide reliable mass estimates when compared with low ionisations lines is still a subject of debate (see Mejía-Restrepo et al. 2018).

1.6 Quasars as Standard Candles

1.6.1 Quasar Cosmology: History and Rationale

Standard candles are sources for which the intrinsic luminosity can be estimated based on some physical property of the system. Phillips (1993) demonstrated that Type Ia supernovae (SNe Ia, thought to be the thermonuclear explosion of an accreting carbon/oxygen white dwarf that has reached the Chandrasekhar mass limit) could be used as standard candles. This led to the first direct evidence for the accelerating expansion of the Universe (Perlmutter et al. 1998, 1999; Riess et al. 1998; Schmidt et al. 1998; Garnavich et al. 1998). Since then, many groups have confirmed this result (Knop et al. 2003; Astier et al. 2006; Wood-Vasey et al. 2007; Kowalski et al. 2008; Kessler et al. 2009; Hicken et al. 2009; Guy

et al. 2010; Sullivan et al. 2011; Suzuki et al. 2012).

However, when considering what sources would be most suitable for use as standard candles, it is clear that there are a number of disadvantages associated with SNe Ia that should not affect quasars. The most distant SNe Ia discovered as of this writing is at a redshift of 2.26 (Rodney et al. 2015), while current SNe Ia cosmological constraints extend only to $z \simeq 1.7$ (Riess et al. 2007). Quasars, however, are observable over a wide range of redshifts, with the most distant source observed to date at $z = 7.085$ (Mortlock et al. 2011), and quasar number density per unit volume peaking between redshifts 2 and 3 (Richards et al. 2006b). In addition, SNe Ia are transient objects, which, apart from making them a challenge to discover in large numbers, results in many SNe Ia being detected after their peak brightness, making standard candle calibration less reliable. In contrast, quasars, though variable, are persistent sources which can be reobserved, with the interesting exception of "changing-look" objects (see LaMassa et al. 2015; Merloni et al. 2015).

The potential value of quasars as cosmological probes has led many groups to explore ways in which they can be implemented as standard candles. Baldwin (1977) discovered the correlation between continuum luminosity and emission line equivalent width, which subsequently became known as the "Baldwin Effect" (see Shields 2007). This correlation was suggested by Baldwin (1977) as a method of standardising quasar luminosity. More recently, Watson et al. (2011) have explored the possibility of using the observed luminosity-BLR size relation (Koratkar & Gaskell 1991; Kaspi et al. 2000; Peterson et al. 2004; Kaspi et al. 2005; Bentz et al. 2009a; Czerny & Hryniewicz 2011; Bentz et al. 2013) as a cosmological tool (Melia 2015). This method relies on using the line-continuum time lags and the observed flux to infer a luminosity distance that is independent of any specific cosmological model. Watson et al. (2011) show that their estimated distances are relatively accurate when compared with the distances derived from the *Wilkinson Microwave Anisotropy Probe* observations (Komatsu et al. 2011). Czerny et al. (2013) have investigated the accuracy of the distance moduli obtained via this method using simulated quasar light-curves, and have shown that the resulting accuracy in the distance moduli should be enough to provide useful cosmological constraints.

A physical (rather than empirical) method for estimating quasar distances was outlined by Yoshii et al. (2014). Their model was based on dust reverberation, and had the advantage of not relying on the empirical distance ladder, thereby giving an independent measurement of H_0 that was comparable in accuracy to that obtained using Cepheid variable stars (Freedman et al. 2001). Hönig et al. (2017) simulated the results of using hot-dust time lags to standardise quasar luminosity in the context of the **V**isible and **I**nfrared **S**urvey **T**elescope for **A**stronomy (VISTA, Sutherland et al. 2015) Extragalactic Infrared Legacy Survey (VEILS)[3]. This simulation suggested that such work could provide constraints on the dark energy density parameter, Ω_Λ, that could rival those obtained using supernovae.

[3]Additional information on the VEILS survey can be found at https://people.ast.cam.ac.uk/~mbanerji/VEILS/.

1.6.2 Quasar Cosmology: The Risaliti & Lusso Method

A new method of implementing quasars as standard candles has recently been introduced, and has been used to place constraints on the matter density parameter, $\Omega_{\rm M}$, and the dark energy density parameter, Ω_{Λ} (RL15). RL15 presented a method that can be used to estimate distance moduli for quasars in a way that does not require the assumption of a cosmological model based on the known relationship between their X-ray and UV emission. This non-linear correlation was discovered after early X-ray surveys were conducted (Tananbaum et al. 1979; Zamorani et al. 1981; Avni & Tananbaum 1986), and can be expressed as

$$\mathrm{Log_{10}}(\mathrm{L}_{\nu_{\rm e},\,2\,\rm keV}) = \beta + \gamma \mathrm{Log_{10}}(\mathrm{L}_{\nu_{\rm e},\,2{,}500\,\text{Å}}) \tag{1.1}$$

where the slope, γ, was found to lie in the 0.5 - 0.7 range (Vignali et al. 2003a; Strateva et al. 2005; Steffen et al. 2006; Just et al. 2007; Lusso et al. 2010; Young et al. 2010). $\mathrm{L}_{\nu_{\rm e},\,2\,\rm keV}$ and $\mathrm{L}_{\nu_{\rm e},\,2{,}500\,\text{Å}}$ are the monochromatic luminosities at rest-frame 2 keV and 2,500 Å in units of erg s^{-1} Hz^{-1}.

By selecting a sample of unobscured quasars with homogeneous spectral energy distributions (SEDs), Lusso & Risaliti (2016) demonstrated that the dispersion in this relation is ∼0.21 - 0.24 dex. This relatively tight relationship, which holds over four decades in luminosity, is interpreted as being the observational result of an intrinsic physical interaction between the UV emission from the accretion disk and the X-ray emission from the electron corona (Lusso & Risaliti 2017). Re-writing equation 1.1 in terms of the flux gives

$$\mathrm{Log_{10}}\left(\frac{4\pi \mathrm{D_L^2}}{1+\mathrm{z}}\,\mathrm{F}_{\nu_{\rm o},\,2\,\rm keV}\right) = \beta + \gamma \mathrm{Log_{10}}\left(\frac{4\pi \mathrm{D_L^2}}{1+\mathrm{z}}\,\mathrm{F}_{\nu_{\rm o},\,2{,}500\,\text{Å}}\right)$$

where $\mathrm{F}_{\nu_{\rm o},\,2\,\rm keV}$ is the flux density measured at the observed-frame frequency corresponding to the rest-frame 2 keV, $\mathrm{F}_{\nu_{\rm o},\,2{,}500\,\text{Å}}$ is the flux density measured at the observed-frame frequency corresponding to the rest-frame 2,500Å, and $\mathrm{D_L}$ is the luminosity distance. Rearranging gives

$$\mathrm{Log_{10}}\left(\frac{\mathrm{F}_{\nu_{\rm o},\,2\,\rm keV}}{1+\mathrm{z}}\right) = \gamma \mathrm{Log_{10}}\left(\frac{\mathrm{F}_{\nu_{\rm o},\,2{,}500\,\text{Å}}}{1+\mathrm{z}}\right) + \beta' + (\gamma-1)\mathrm{Log_{10}}(\mathrm{D_L^2}) \tag{1.2}$$

Where $\beta' = \beta + (\gamma - 1)\mathrm{Log_{10}}(4\pi)$. In a sufficiently narrow redshift bin (RL15 adopt $\Delta\mathrm{Log_{10}(z)} = 0.1$), the term containing $\mathrm{D_L}$ (highlighted in purple) will add an almost constant contribution to each flux value. Therefore, for a sufficiently narrow redshift bin, one can fit the following linear relation to recover the slope in that specific bin, γ_z,

$$\mathrm{Log_{10}}\left(\frac{\mathrm{F}_{\nu_{\rm o},\,2\,\rm keV}}{1+\mathrm{z}}\right) = \gamma_{\rm z} \mathrm{Log_{10}}\left(\frac{\mathrm{F}_{\nu_{\rm o},\,2{,}500\,\text{Å}}}{1+\mathrm{z}}\right) + \beta'_{\rm z} \tag{1.3}$$

Rearranging equation 1.2 allows the luminosity distance to be written as a function of the

X-ray and UV flux densities, redshift, γ, and $\beta^{'}$:

$$\mathrm{Log_{10}D_L} = \frac{1}{2(\gamma-1)}\left[\mathrm{Log_{10}}\left(\frac{\mathrm{F_{\nu_o,\,2\,keV}}}{1+\mathrm{z}}\right) - \gamma\mathrm{Log_{10}}\left(\frac{\mathrm{F_{\nu_o,\,2{,}500\,\mathring{A}}}}{1+\mathrm{z}}\right) - \beta^{'}\right]$$

The distance modulus is defined as

$$\mathrm{DM} = 5\mathrm{Log_{10}}\left(\frac{\mathrm{D_L}}{10\mathrm{pc}}\right)$$

which then can be written as

$$\mathrm{DM} = \frac{5}{2(\gamma-1)}\left[\mathrm{Log_{10}}\left(\frac{\mathrm{F_{\nu_o,\,2\,keV}}}{1+\mathrm{z}}\right) - \gamma\mathrm{Log_{10}}\left(\frac{\mathrm{F_{\nu_o,\,2{,}500\,\mathring{A}}}}{1+\mathrm{z}}\right) - \beta^{'}\right] \tag{1.4}$$

Since the value of $\beta^{'}$ is not known, absolute distance moduli cannot be estimated from equation 1.4, and therefore, following RL15, $\beta^{'}$ will be treated as an arbitrary scaling factor. Instead, relative, or uncalibrated, distance moduli can be estimated as follows

$$\mathrm{DM} = \frac{5}{2(\gamma-1)}\left[\mathrm{Log_{10}}\left(\frac{\mathrm{F_{\nu_o,\,2\,keV}}}{1+\mathrm{z}}\right) - \gamma\mathrm{Log_{10}}\left(\frac{\mathrm{F_{\nu_o,\,2{,}500\,\mathring{A}}}}{1+\mathrm{z}}\right)\right] \tag{1.5}$$

where the mean of the best-fit γ_z values measured from fitting equation 1.3 in narrow redshift bins can be adopted as the value of γ. This method of deriving distance modulus estimates assumes that the value of γ does not evolve with redshift.

1.6.3 Fitting the Distance Modulus - Redshift Relation

As discussed in section 1.6.2, RL15 have outlined a method for estimating distance moduli for quasars that is based on the empirically observed relationship between quasar X-ray and UV emission. The resulting distance modulus - redshift distribution can then be fit with a standard ΛCDM cosmological model in order to place constraints on the cosmological parameters Ω_M and Ω_Λ via the following method. For a fixed cosmological constant, Λ, the luminosity distance can be expressed as

$$D_L(z, \Omega_M, \Omega_\Lambda) = \frac{(1+z)}{\sqrt{|\Omega_K|}} \times \mathrm{sinn}\left(\int_0^z \frac{\sqrt{|\Omega_K|}}{\sqrt{\Omega_M(1+z')^3 + \Omega_K(1+z')^2 + \Omega_{de}(z')}} dz'\right) \quad (1.6)$$

where $\Omega_K = 1 - \Omega_M - \Omega_\Lambda$, and the dark energy density parameter is

$$\Omega_{de}(z) = \Omega_\Lambda \exp\left(3\int_0^z \frac{1+w(x)}{1+x} dx\right)$$

The equation above represents the general case where the dark energy equation of state, w(z), is a function of redshift. In this experiment however, the standard ΛCDM model will be assumed, where w is constant with redshift and equal to -1 (and $\Omega_{de} = \Omega_\Lambda$, see appendix A). Given this assumption, equation 1.6 becomes

$$D_L(z, \Omega_M, \Omega_\Lambda) = \frac{(1+z)}{\sqrt{|\Omega_K|}} \times \mathrm{sinn}\left(\int_0^z \frac{\sqrt{|\Omega_K|}}{\sqrt{\Omega_M(1+z')^3 + \Omega_K(1+z')^2 + \Omega_\Lambda}} dz'\right) \quad (1.7)$$

The function sinn is defined as follows

$$\mathrm{sinn}(x) = \begin{cases} x & \text{for } \Omega_K = 0 \quad \text{(Flat Universe)} \\ \sin(x) & \text{for } \Omega_K < 0 \quad \text{(Closed Universe)} \\ \sinh(x) & \text{for } \Omega_K > 0 \quad \text{(Open Universe)} \end{cases}$$

Note that throughout this work the distance modulus is shown in figures rather than the luminosity distance. The two quantities are related as follows:

$$DM(z, \Omega_M, \Omega_\Lambda) = 5\log_{10}\left(\frac{D_L(z, \Omega_M, \Omega_\Lambda)}{10\,\mathrm{pc}}\right) \quad (1.8)$$

The Markov chain Monte Carlo (MCMC) ensemble sampler "EMCEE" (Foreman-Mackey et al. 2013) was used to fit a likelihood function to the distance modulus - redshift relation

of the form

$$\begin{aligned} &\mathrm{Ln\,L}(\Omega_\mathrm{M}, \Omega_\Lambda, \beta, \delta \,|\, \mathrm{z}, \mathrm{DM}, \delta\mathrm{DM}) = \\ &-\frac{1}{2}\sum_{\mathrm{i}=1}^{\mathrm{n}} \left(\frac{[\mathrm{DM_i} - \mathrm{DM}(\mathrm{z_i}, \Omega_\mathrm{M}, \Omega_\Lambda, \beta)]^2}{\mathrm{s_i^2}} + \mathrm{Ln}\left(2\pi \mathrm{s_i^2}\right) \right) \end{aligned} \tag{1.9}$$

(see appendix B for a derivation of this likelihood function). The variance is defined as

$$\mathrm{s_i^2} = \delta\mathrm{DM_i^2} + \delta^2 \tag{1.10}$$

where $\mathrm{DM_i}$ are the distance moduli derived from flux density measurements (see equation 1.5) and $\mathrm{DM(z}, \Omega_\mathrm{M}, \Omega_\Lambda, \beta)$ are the values expected based on the measured redshift for each source using a standard ΛCDM cosmology (equation 1.8). Note that since the empirical distance modulus estimates are uncalibrated, when fitting equation 1.8 a scaling factor β is also included in the fit. $\delta\mathrm{DM_i}$ are the uncertainties on $\mathrm{DM_i}$, and these uncertainties are assumed to underestimate the variance by some amount δ^2, where δ is the intrinsic dispersion (in units of DM).

Following Bayes' Theorem, the probability distribution function can be written as

$$\mathrm{P}(\Omega_\mathrm{M}, \Omega_\Lambda, \beta, \delta \,|\, \mathrm{z}, \mathrm{DM}, \delta\mathrm{DM}) \propto \pi(\Omega_\mathrm{M}, \Omega_\Lambda, \delta)\mathrm{L}(\Omega_\mathrm{M}, \Omega_\Lambda, \beta, \delta \,|\, \mathrm{z}, \mathrm{DM}, \delta\mathrm{DM})$$

where π represents the prior on the values to be measured. Following RL15, the prior was set to

$$\begin{aligned} \pi(\Omega_\mathrm{M}, \Omega_\Lambda, \delta) = (0 < \Omega_\mathrm{M} < 1.19) \;\&\&\; (0 < \Omega_\Lambda < 1.49) \\ \&\&\; (\Omega_\Lambda < 0.99 + 1.8\Omega_\mathrm{M}) \;\&\&\; (0.37 < \delta < 2.72) \end{aligned} \tag{1.11}$$

Writing the probability distribution function in terms of the natural logarithm gives

$$\mathrm{LnP}(\Omega_\mathrm{M}, \Omega_\Lambda, \beta, \delta \,|\, \mathrm{z}, \mathrm{DM}, \delta\mathrm{DM}) \propto \mathrm{Ln}\pi(\Omega_\mathrm{M}, \Omega_\Lambda, \delta) + \mathrm{Ln\,L}(\Omega_\mathrm{M}, \Omega_\Lambda, \beta, \delta \,|\, \mathrm{z}, \mathrm{DM}, \delta\mathrm{DM})$$

This posterior probability function was then sampled using EMCEE.

The method described in section 1.6.2 was used by RL15 to derive distance modulus estimates for a sample of 808 quasars. The authors then fit the resulting distance modulus - redshift relation with a standard ΛCDM cosmological model in order to measure Ω_M and Ω_Λ. The resulting confidence contours in Ω_M and Ω_Λ parameter space are at an angle to those obtained from an analysis of SNe Ia (see figure 6, RL15). For this reason, quasars and SNe Ia can be considered to be complimentary methods of measuring Ω_M and Ω_Λ, as they exclude different regions of the parameter space.

One of the goals of this work was to attempt to extend the progress made by RL15 on the quasar cosmology experiment by using survey-quality data from *XMM-Newton* (the XMM-XXL North survey, Pierre et al. 2016; Menzel et al. 2016) and *ROSAT* (the 2RXS

survey, Boller et al. 2016). These datasets were selected since they were the largest X-ray selected quasar samples with follow-up optical spectroscopy available at the commencement of this work. The strengths and weaknesses of survey-quality data (rather than data from targeted observations) in the context of the quasar cosmology experiment were investigated.

A detailed understanding of the inclusion of survey-quality data in this experiment is especially relevant when considering the use of *eROSITA* survey data in the near future. The application of currently available *XMM-Newton* and *ROSAT* data to the quasar cosmology experiment is presented in chapter 4. This work also aimed to provide a forecast of the cosmological constraints that will be achieved via the quasar Hubble diagram when using the *eROSITA* - 4MOST quasar sample. Using the *eROSITA* mock catalogue produced by Comparat et al. (2019), the quasar X-ray - UV relation was simulated and the potential constraints on $\Omega_{\rm M}$ and Ω_{Λ} that can be derived using the RL15 method was investigated. This analysis of the potential of the *eROSITA* - 4MOST sample is presented in chapter 5.

1.7 Observatories Used in this Work

The observatories and survey programmes that have provided the data used in this work are discussed in the following sections. In addition, since chapter 5 presents simulations of the *eROSITA* - 4MOST dataset in the context of the quasar cosmology experiment, both the *eROSITA* mission and the 4MOST facility are discussed in sections 5.1 and 5.2, respectively.

1.7.1 X-Ray Observatories

XMM-Newton

The experiments presented in this work use data obtained by the X-ray Multi-Mirror (*XMM-Newton*) space observatory which was developed as part of the European Space Agency's (ESA) Horizons 2000 programme and placed into orbit on 10/12/1999 (Jansen et al. 2001). *XMM-Newton's* scientific payload includes the following:

- One European Photon Imaging Camera (EPIC) PN (Strüder et al. 2001) and two EPIC Metal Oxide Semiconductor (MOS) charge-coupled device (CCD) -based imaging spectrometers (Turner et al. 2001) which have a bandpass of 0.2 - 10 keV.
- Two Reflection Grating Spectrometers (RGS; den Herder et al. 2001), which are high-resolution ($E/\Delta E = 100 - 500$) spectrographs that cover the 0.3 - 2.1 keV range.
- One Optical Monitor (XMM-OM) which provides imaging in the 1,700 - 6,500 Å range (Mason et al. 2001).

These detectors are placed below three X-ray telescopes, each of which is composed of 58 Wolter I mirrors, with a focal length of 7.5 meters. *XMM-Newton* boasts the largest

effective area of a focusing X-ray telescope available at the commencement of this work, with each of the three telescopes reaching 1,550 cm^2 at 1.5 keV.

The datasets resulting from two surveys carried out by *XMM-Newton* were used in this work; the first *XMM-Newton* Slew survey (XMMSL1, Saxton et al. 2008), which is based on observations made by *XMM-Newton* while slewing between targets, and the equatorial XMM-XXL north survey (Pierre et al. 2016; Menzel et al. 2016). These surveys, and the resulting datasets, are discussed in sections 2.1 and 4.1.1, respectively.

ROSAT

ROSAT (Röntgensatellit, Truemper 1982) was the first instrument to perform an X-ray imaging survey of the entire sky (referred to as the *ROSAT* All-Sky Survey or RASS). *ROSAT* operated in the soft X-ray and extreme UV bands, and its scientific payload consisted of (Briel et al. 1996)

- The X-ray telescope (XRT), *ROSAT's* primary instrument, which observed in the 0.1 - 2.4 keV range. It had a focal-length of 2.4 m and consisted of four nested Wolter I mirrors. The XRT's focal plane contained two different types of detectors; two Position Sensitive Proportional Counters (PSPC), and the High Resolution Imager (HRI).

- The wide-field camera (WFC), which observed in the 0.05 - 0.21 keV range, was co-aligned with the XRT and had a focal-length of 0.525 m. The WFC consisted of three nested Wolter-Schwarzschild mirrors and a micro-channel plate detector.

ROSAT was launched on the 1st of July 1990, and operated in a low Earth orbit at $\sim$ 580 km with an orbital period of 96 min. The RASS was conducted using the PSPC and the WFC while the satellite scanned great circles that intersected the ecliptic poles. The survey progressed at a rate of $\sim$ 1 deg per day, and therefore was completed in $\sim$ 6 months. The remainder of the mission was mainly devoted to pointed observations, and the satellite was decommissioned on the 12th of February 1999. This work makes use of the second *ROSAT* all-sky survey (2RXS) catalogue (Boller et al. 2016), which is discussed in section 2.1.

1.7.2 Sloan Digital Sky Survey

The X-ray detected sources included in this work have been spectroscopically identified using data obtained by the SDSS (York et al. 2000). The SDSS has provided the most extensive photometric and spectroscopic survey of astronomical objects available as of the commencement of this work. Using the 2.5 m Sloan foundation telescope at Apache Point Observatory (APO), New Mexico, USA (Gunn et al. 2006), the SDSS has had four main survey phases thus far; SDSS-I (2000 - 2005, York et al. 2000), SDSS-II (2005 - 2008, York et al. 2000), SDSS-III (2008 - 2014, Eisenstein et al. 2011), and SDSS-IV (2014 - 2020, Blanton et al. 2017), each with a range of observational goals and sub-surveys.

Table 1.1: Technical details of the *ROSAT* and *XMM-Newton* satellites. The information presented below was taken from the "*XMM-Newton* Users Handbook", Issue 2.17, 2019 (ESA: *XMM-Newton* SOC) and https://heasarc.gsfc.nasa.gov/docs/rosat/pspc.html.

	ROSAT	*XMM-Newton*
Mission Duration	01/06/90 - 12/02/99	10/12/99 - currently operating
Energy Range [keV]	0.1 - 2.4	0.15 - 12
Total Effective Area at 1 keV [cm^2]	400	4,650
Field of View [deg]	∼2 (PSPC)	∼0.5
Focal Length [m]	2.4 (XRT)	7.5
Spectral Resolution at 1 keV [eV]	500	∼80 (EPIC pn)
Spatial Resolution at 1 keV [arcsec]	∼25	6
Time resolution [μs]	∼130	7 (EPIC pn burst mode)

In SDSS-I and -II, spectroscopic redshifts for almost one million galaxies were obtained (Abazajian et al. 2009). These observations were carried out using the SDSS 640-fiber multi-object spectrograph (Smee et al. 2013), which has a bandpass of 3,800 - 9,200 Å and a fiber diameter of 3 arcsec. By the end of SDSS-II, the survey had covered a photometric footprint of 11,663 deg^2 and a spectroscopic footprint of 9,380 deg^2 (Abazajian et al. 2009). SDSS-I and -II targeted optical counterparts to *ROSAT* X-ray detected sources (Anderson et al. 2007); a method of target selection that would be continued in both BOSS and eBOSS.

The Baryon Oscillation Spectroscopic Survey (BOSS, Dawson et al. 2013) was the largest sub-survey in SDSS-III (Eisenstein et al. 2011). The primary goal of BOSS was to measure the Baryon Acoustic Oscillation (BAO) scale which can be inferred from the clustering behaviour of objects in large cosmological volumes. To achieve this goal, BOSS measured ∼1.5 million galaxies out to $z < 0.7$, and 150,000 quasars in the redshift range $2.15 \leq z \leq 3.5$ in two sub-surveys that ran simultaneously. BOSS mainly observed high Galactic latitude objects (at equatorial declination $\delta > -3.75$ deg), and covered a spectroscopic footprint of ∼10,000 deg^2. An updated version of the SDSS spectrograph was used for BOSS, which utilised a larger number (1,000) of smaller fibers (2 arcsec in diameter) and covered a larger wavelength range (3,600 - 10,400 Å) than the original SDSS spectrograph (Smee et al. 2013). As part of the BOSS program, X-ray sources in the equatorial XMM-XXL north field were followed up with the BOSS spectrograph (Menzel et al. 2016). Data from the XMM-XXL north field is used in the quasar cosmology experiment presented in section 4, and further details on this survey can be found in section 4.1.1.

eBOSS/SPIDERS

The Extended Baryon Oscillation Spectroscopic Survey (eBOSS Dawson et al. 2016) began with SDSS-IV (Blanton et al. 2017) in July 2014, and was a continuation of BOSS. Within eBOSS, two sub-surveys were run in parallel; the TDSS (**T**ime **D**omain **S**pectroscopic **S**urvey; Principal Investigators (PIs) Green and Anderson, Morganson et al. 2015), which targeted variable objects, and SPIDERS (**SP**ectroscopic **ID**entification of ***eR****OSITA* **S**ources; PIs Merloni and Nandra). SPIDERS conducted optical spectroscopy of extragalactic X-ray detections in wide-area *ROSAT* and *XMM-Newton* surveys (Dwelly et al. 2017). Lying at the bright end of the X-ray source population, these sources will also be detected by *eROSITA* (Merloni et al. 2012; Predehl et al. 2016). The rich X-ray and optical datasets provided by SPIDERS were used throughout this work, and so further details on the SPIDERS programme are presented in section 2.1.1.

1.8 Conventions Adopted in this Work

Throughout this work, a standard ΛCDM cosmology was assumed, where $\Omega_{\rm M} = 0.3$, and $\Omega_{\Lambda} = 1 - \Omega_{\rm M}$ (Komatsu et al. 2009). It is noted here that the method for standardising quasar luminosities results in relative distance moduli estimates. These quasar distance estimates, like SNe Ia distance estimates, need to be calibrated to give absolute distances. Since the calibration for the quasar distance moduli is unknown, a Hubble constant value of $\rm H_0 = 70\,km\,s^{-1}\,Mpc^{-1}$ will be assumed in this work, except for chapter 5, where $\rm h = 0.6777$ is adopted in order to conform to the cosmological model used by Comparat et al. (2019). Magnitudes are expressed in the AB system. Following the tradition set by many research groups in the Astrophysics community, the CGS (centimetre-gram-second) unit system is used almost exclusively. The wavelengths of atomic transitions are given in the rest-frame of the source.

Chapter 2

SDSS-IV/SPIDERS: A Catalogue of X-Ray Selected AGN Properties.

Spectral Properties and Black Hole Mass Estimates for SPIDERS SDSS DR14 Type 1 AGN

This chapter discusses the spectroscopic analysis of the X-ray selected type 1 AGN in the SPIDERS SDSS DR14 sample, and the production of the corresponding SDSS DR14 Value Added Catalogue. This work is also presented in Coffey et al. (2019).

2.1 Preparing the Input Catalogue

The *ROSAT* sample used in this work is part of the second *ROSAT* all-sky survey (2RXS) catalogue (Boller et al. 2016), which has a limiting flux of $\sim 10^{-13}\,\mathrm{erg\,cm^{-2}\,s^{-1}}$ in the 0.1 - 2.4 keV band, corresponding to a luminosity of $\sim 10^{43}\,\mathrm{erg\,s^{-1}}$ at z = 0.5. This is an approximate limit since there is a range of different exposure times in the survey; for example, the minimum detectable flux is lower at the Ecliptic poles. Compared to the first *ROSAT* data release (Voges et al. 1999), the 2RXS catalogue is the result of an improved detection algorithm, which uses a more detailed background determination relative to the original *ROSAT* pipeline. A full visual inspection of the 2RXS catalogue has been performed, which provides a reliable estimate of its spurious source content (see Boller et al. 2016). The first *XMM-Newton* slew survey catalogue release 1.6 (XMMSL1; Saxton et al. 2008) was also used in this work. This catalogue includes observations made by the European Photon Imaging Camera (EPIC) pn detectors while slewing between targets, and has a limiting flux of $6 \times 10^{-13}\,\mathrm{erg\,cm^{-2}\,s^{-1}}$ in the 0.2 - 2 keV band, which corresponds to a luminosity of $5.8 \times 10^{44}\,\mathrm{erg\,s^{-1}}$ at z=0.5.

2.1.1 The SPIDERS Programme

The SPIDERS programme, which is part of SDSS-IV, has been providing spectroscopic observations of 2RXS and XMMSL1 sources[1] in the BOSS imaging footprints. Before the start of the eBOSS survey in 2014, the SPIDERS team compiled a sample of X-ray selected spectroscopic targets and submitted this sample for spectroscopic follow-up using the BOSS spectrograph (see Dwelly et al. 2017, for further details on the SPIDERS programme). As of the end of eBOSS in February 2019, the eBOSS/SPIDERS survey has covered a sky area of 5321 $\deg^2$. The SDSS DR14 SPIDERS sample presented in this work covers an area of $\sim 2200\,\deg^2$ ($\sim 40\%$ of the final eBOSS/SPIDERS area).

The spectroscopic completeness achieved by the SPIDERS survey as of SDSS DR14 in the eBOSS area is $\sim 53\%$ for the sample as a whole, $\sim 63\%$ considering only high-confidence X-ray detections (see section 2.1.4), and $\sim 87\%$ considering sources with high-confidence X-ray detections and optical counterparts with magnitudes within the nominal survey limits ($17 \leq m_{\mathrm{Fiber2,i}} \leq 22.5$). Outside the eBOSS area, the spectroscopic completeness of this sample is lower: $\sim 28\%$ for the sample as a whole, $\sim 39\%$ considering only high-confidence X-ray detections, and $\sim 57\%$ considering sources with high-confidence X-ray detections and optical counterparts with magnitudes within the nominal survey limits. The spectroscopic completeness of the SPIDERS SDSS DR16 sample inside and outside the eBOSS area is expected to be similar to that of the sample presented here.

In addition to those targeted during eBOSS/SPIDERS, a large number of 2RXS and XMMSL1 sources had their spectra taken during the SDSS-I/II (2000-2008; York et al. 2000) and the SDSS-III (Eisenstein et al. 2011) BOSS (2009-2014; Dawson et al. 2013) surveys (see figure 2.3). This work includes spectra obtained by eBOSS/SPIDERS up to DR14 (2014-2016) as well as spectra from SDSS-I/II/III. The resulting dataset is a powerful resource for the multiwavelength analysis of AGN. This work aims to capitalise on the wealth of information already available by providing detailed optical spectral measurements, as well as estimates of black hole masses and Eddington ratios.

2.1.2 Identifying IR Counterparts

To identify SPIDERS spectroscopic targets, the Bayesian cross-matching algorithm "NWAY" (Salvato et al. 2018) was used to select AllWISE (Cutri et al. 2013) infrared (IR) counterparts for the 2RXS and XMMSL1 X-ray selected sources in the BOSS footprint. The AllWISE catalogue consists of data obtained during the two main survey phases of the Wide-field Infrared Survey Explorer mission (WISE; Wright et al. 2010) which conducted an all-sky survey in the 3.4, 4.6, 12, and 22 μm bands (magnitudes in these bands are denoted [W1], [W2], [W3], and [W4] respectively). The matching process used the colour-magnitude priors [W2] and [W2-W1] (see Dwelly et al. 2017) which, at the depth of the 2RXS and XMMSL1 surveys, can distinguish between the correct counterparts and chance

[1]The SPIDERS programme targets both point-like and extended X-ray sources. This work focuses on the counterparts to point-like X-ray detections, which are predominantly AGN, and therefore, the samples discussed in the subsequent paragraphs are derived from the SPIDERS-AGN programme.

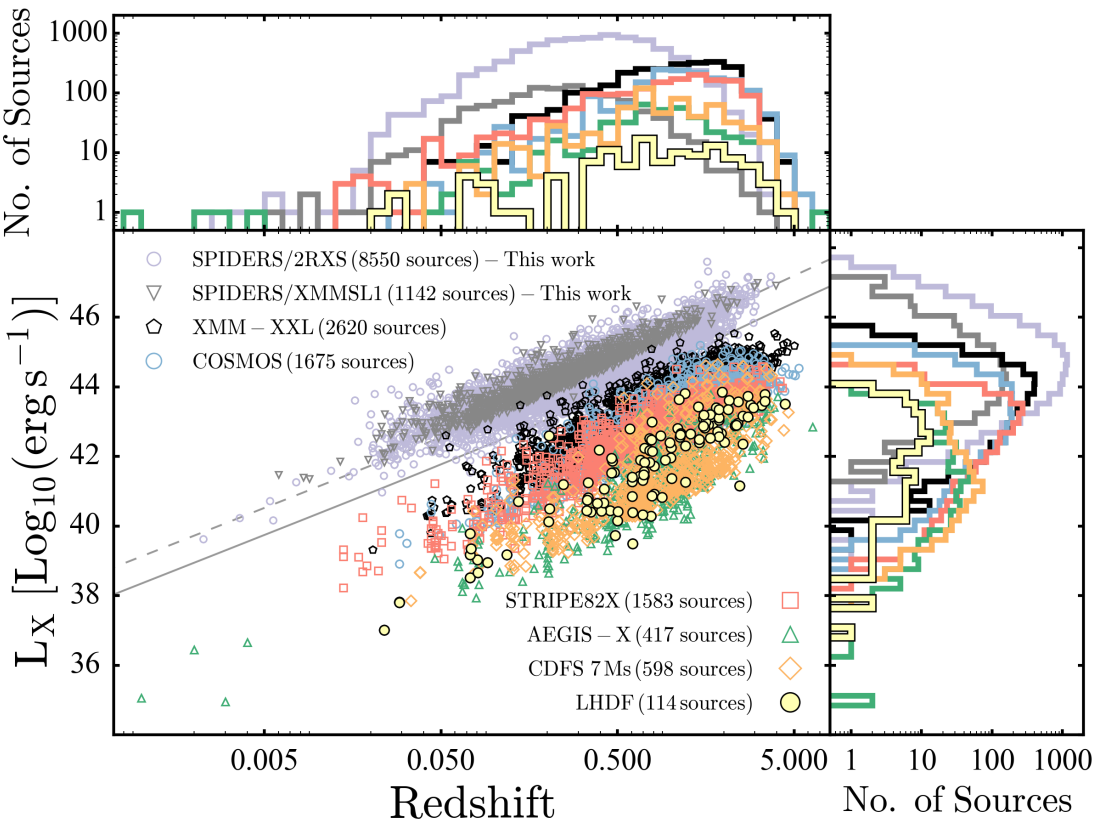


Figure 2.1: Soft X-ray luminosity versus spectroscopic redshift for the samples presented in this work and the following previously published X-ray selected samples; XMM-XXL (Menzel et al. 2016; Liu et al. 2016), CDFS (Luo et al. 2017), STRIPE82X (LaMassa et al. 2016), COSMOS (Marchesi et al. 2016a,b; Civano et al. 2016), AEGIS-X (Nandra et al. 2015), and the Lockman Hole deep field (LHDF) (Brunner et al. 2008; Fotopoulou et al. 2012). For each sample, the 0.5-2 keV luminosities are shown, except for the 2RXS sample, where the 0.1-2.4 keV luminosities are shown, and the XMMSL1 sample, where the 0.2-2 keV luminosities from Saxton et al. (2008) are shown. The approximate detection limit for the 2RXS and XMMSL1 samples are shown by the solid and dashed grey lines respectively. The X-ray luminosities for the 2RXS sample are derived from the classical flux estimates described in section 2.4.1, however it is noted here that some low count rate 2RXS sources do not have flux estimates. For sources that were detected in both 2RXS and XMMSL1, only the XMMSL1 luminosities are shown. Sources classified as stars have not been included in this figure.

associations. These colours would not be efficient if the 2RXS survey was much deeper (see Salvato et al. 2018, for a complete discussion). The resulting 2RXS and XMMSL1 catalogues with AllWISE counterparts contained 53455 and 4431 sources respectively. AllWISE positions were then matched to photometric counterparts, where available, in SDSS.

2.1.3 Comparison with Previous X-ray Surveys

Figure 2.1 displays the sources in the 2RXS and XMMSL1 samples which have spectroscopic redshifts and measurements of the soft X-ray flux. For comparison, a series of previously published X-ray selected samples that have optical spectroscopic redshifts are also shown. The large number of sources present in the 2RXS and XMMSL1 samples motivated the optical spectroscopic analysis discussed in the following sections.

2.1.4 Selecting a Reliable Subsample

The selection of SPIDERS spectroscopic targets was discussed in detail by Dwelly et al. (2017). Here, this selection procedure is summarised, along with the additional cuts made in this work to select a sample for spectral analysis. The sequence of selection criteria used and the resulting sample size are shown in figure 2.2.

2RXS sources with an X-ray detection likelihood (EXI_ML) ≤ 10 were excluded since these detections are considered highly uncertain with a spurious fraction $\geq 20\%$ (see Boller et al. 2016). XMMSL1 sources with an X-ray detection likelihood (XMMSL_DET_ML_B0) ≤ 10 were also excluded. Salvato et al. (2018, figure 1) show the distribution of flux with detection likelihood for both samples. These cuts returned 23245/53455 2RXS and 3803/4431 XMMSL1 sources.

The following cuts, which were described in detail in Dwelly et al. (2017), have also been applied to the sample:

- For each X-ray source, Salvato et al. (2018) give the probability, p_any, that a reliable counterpart exists among the possible AllWISE associations. Sources with p_any < 0.01 were removed. This returned 23046/23245 2RXS and 3558/3803 XMMSL1 sources.

- For each X-ray source, the most probable AllWISE counterpart was chosen by selecting sources with NWAY match_flag=1 (see Salvato et al. 2018). This returned 20585/23046 2RXS and 3321/3558 XMMSL1 sources.

- For each AllWISE counterpart, the brightest SDSS-DR13 photometric source within the AllWISE matching radius was selected using FLAG_SDSSv5b_best=1. This returned 19385/20585 2RXS and 3063/3321 XMMSL1 sources.

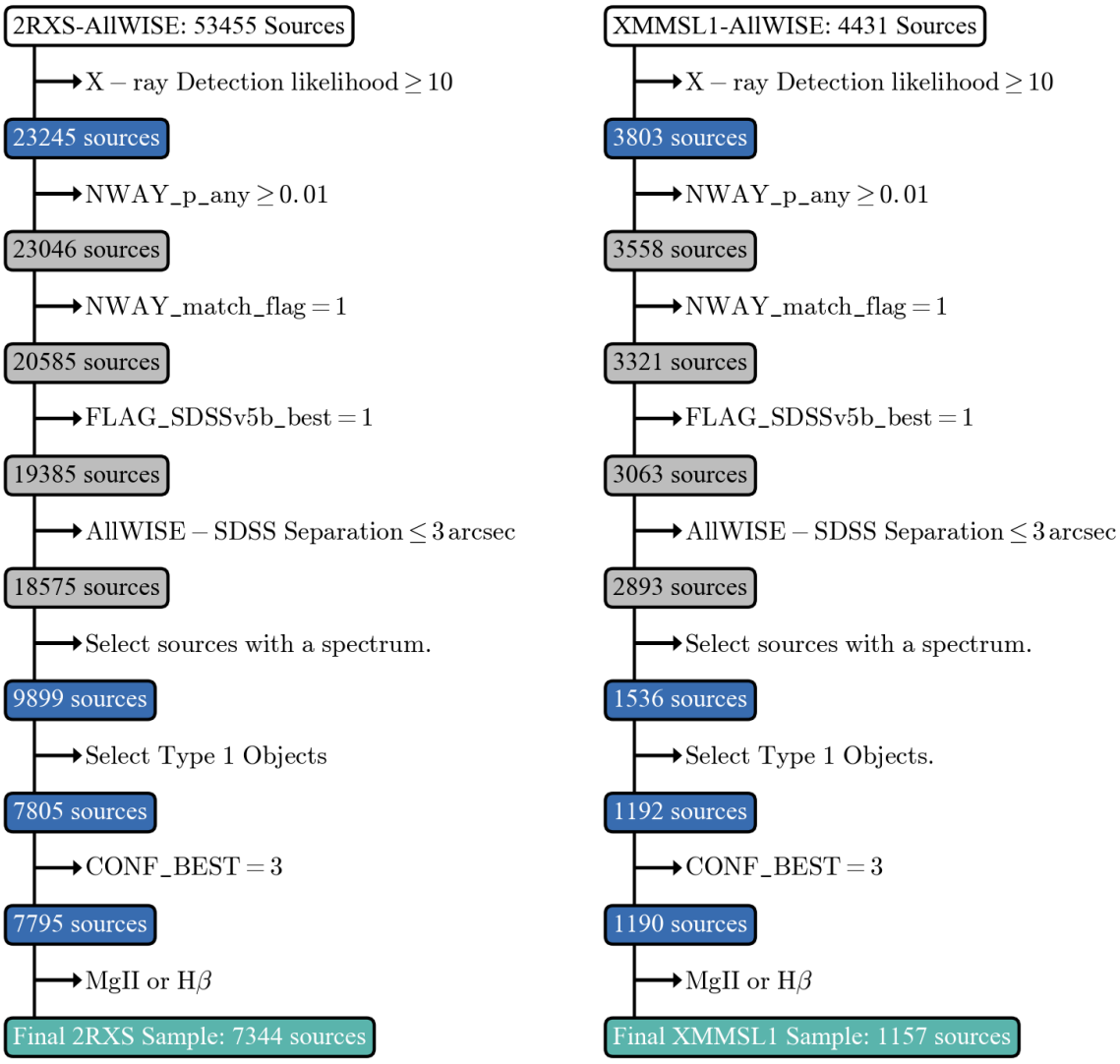


Figure 2.2: Sequence of quality cuts applied to the 2RXS and XMMSL1 samples to produce the subsample used for spectral analysis. The starting points (2RXS-AllWISE and XMMSL1-AllWISE) are the full samples of 2RXS and XMMSL1 selected sources with AllWISE IR counterparts in the BOSS footprint (see section 2.1.2). The steps in grey are those that have been discussed in Dwelly et al. (2017).

- Cases where the AllWISE-SDSS separation exceeded 3 arcsec were removed. This returned 18575/19385 2RXS and 2893/3063 XMMSL1 sources.

The results of these cuts are displayed in grey in figure 2.2. As shown above, sources with match_flag=1 were targeted; however, for 14% of the 2RXS sample and 10% of the

XMMSL1 sample, more than one counterpart was highly likely. This implies that either the counterpart association was not reliable, or that the X-ray detection was the result of emission from multiple sources. These sources were not included in the discussion of optical spectral properties as a function of X-ray properties in section 3.2.3. After selecting the brightest SDSS-DR13 (Albareti et al. 2017) photometric source, there were 15 cases where two unique 2RXS sources were matched to the same AllWISE/SDSS counterpart and 3 cases where two unique XMMSL1 sources were matched to the same AllWISE/SDSS counterpart. These sources were also removed.

Of the samples with reliable SDSS photometric counterparts, 8777 2RXS and 1315 XMMSL1 sources had their spectra taken during SDSS-I/II/III while 1122 2RXS and 221 XMMSL1 sources had their spectra taken during the SPIDERS programme (including SEQUELS), resulting in a sample of 9899 2RXS and 1536 XMMSL1 sources with spectra as of DR14 (see figure 2.3). The distribution of SDSS i band fiber2 magnitudes for this sample (showing the different spectroscopic programmes) is presented in figure 2.4. Due to targeting constraints (as discussed in section 2.1.1), the sample completeness is much lower outside of the nominal magnitude limits for the survey ($17 \leq m_{\mathrm{Fiber2,i}} \leq 22.5$ for eBOSS).

2.1.5 Source Classification

Visual inspection results for each object in this sample are available from a combination of literature sources (Anderson et al. 2007; Plotkin et al. 2010; Schneider et al. 2010; Pâris et al. 2017) and the SPIDERS group. The SPIDERS visual inspection (see Dwelly et al. 2017, for further details) provides a visual confirmation of the SDSS pipeline redshift and object classification. The results of this inspection include a flag indicating the confidence of the redshift, "CONF_BEST", which can take the values 3 (highly secure), 2 (uncertain), 1 (poor/unusable), 0 (insufficient data). Sources may be flagged as having low confidence redshifts ("CONF_BEST"<3) in cases where specific emission lines cannot be identified, typically due to a low spectral S/N. A confirmation of the source classification was also added during the visual inspection, which uses the categories QSO, broad absorption line QSO (BALQSO), blazar, galaxy, star, and none. Anderson et al. (2007) provide the broad line AGN (BLAGN) and narrow line AGN (NLAGN) classifications, which are defined based on the presence or absence of broad ($\mathrm{FWHM} > 1000\,\mathrm{km\,s^{-1}}$) permitted emission lines.

The main goal of this work is to analyse the type 1 AGN in the SPIDERS sample, and therefore only sources that have been classified via their optical spectra as either "BLAGN" or "QSO" were selected for spectroscopic analysis. This returned 7805/9899 2RXS and 1192/1536 XMMSL1 sources. Since the categories "BLAGN" and "QSO" are based on different classification criteria, there will be some overlap between the two sets of sources. Therefore, no distinction will be made between the two categories; instead, both sets of objects will be considered type 1 AGN in this work.

When selecting sources classified as "BLAGN" or "QSO", 1376 2RXS and 208 XMMSL1 sources that were optically classified as galaxies were removed. However, most of these

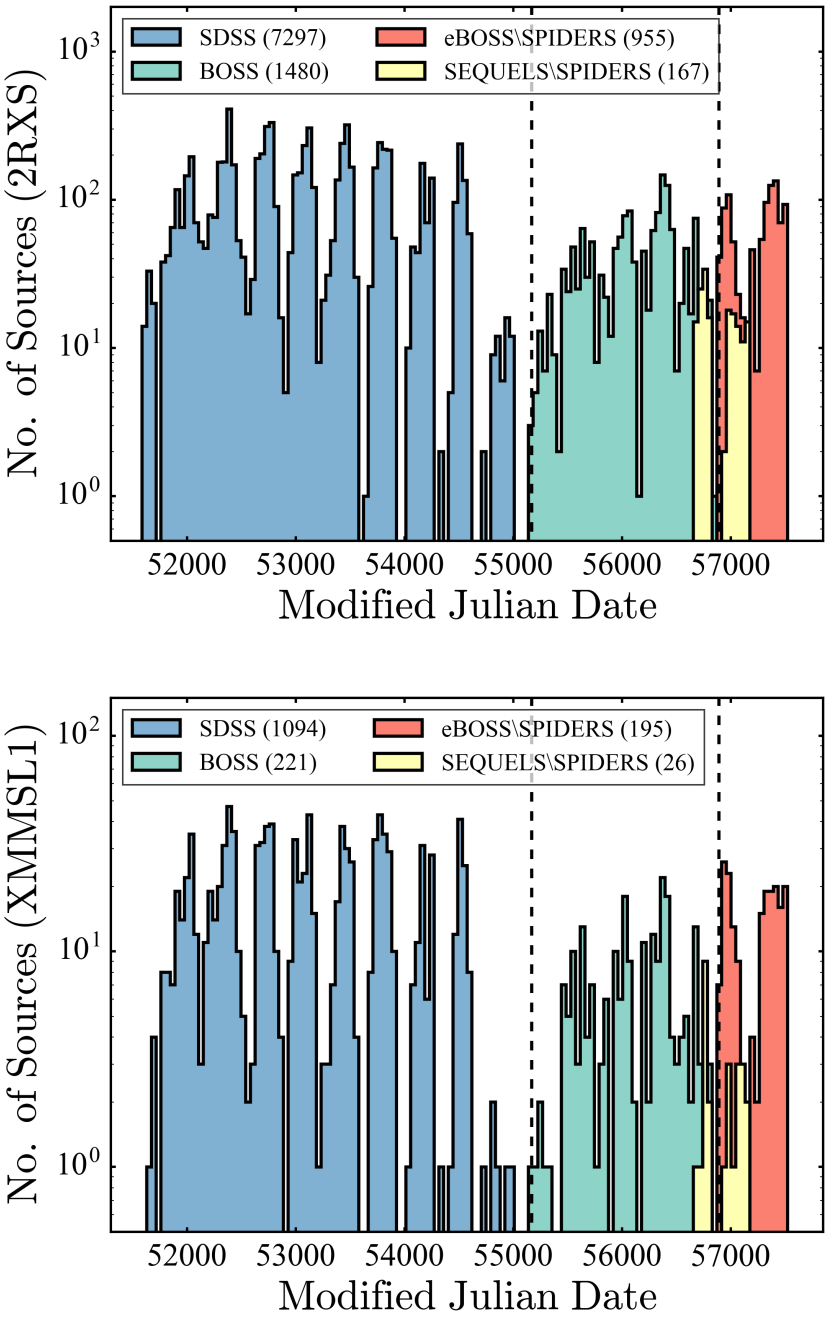


Figure 2.3: Distribution of MJDs on which the spectra were acquired for the sample of sources with spectra, colour-coded based on the survey from which the spectra were taken. The dashed lines indicate the beginning of the BOSS (MJD=55171) and eBOSS (MJD=56893) surveys.

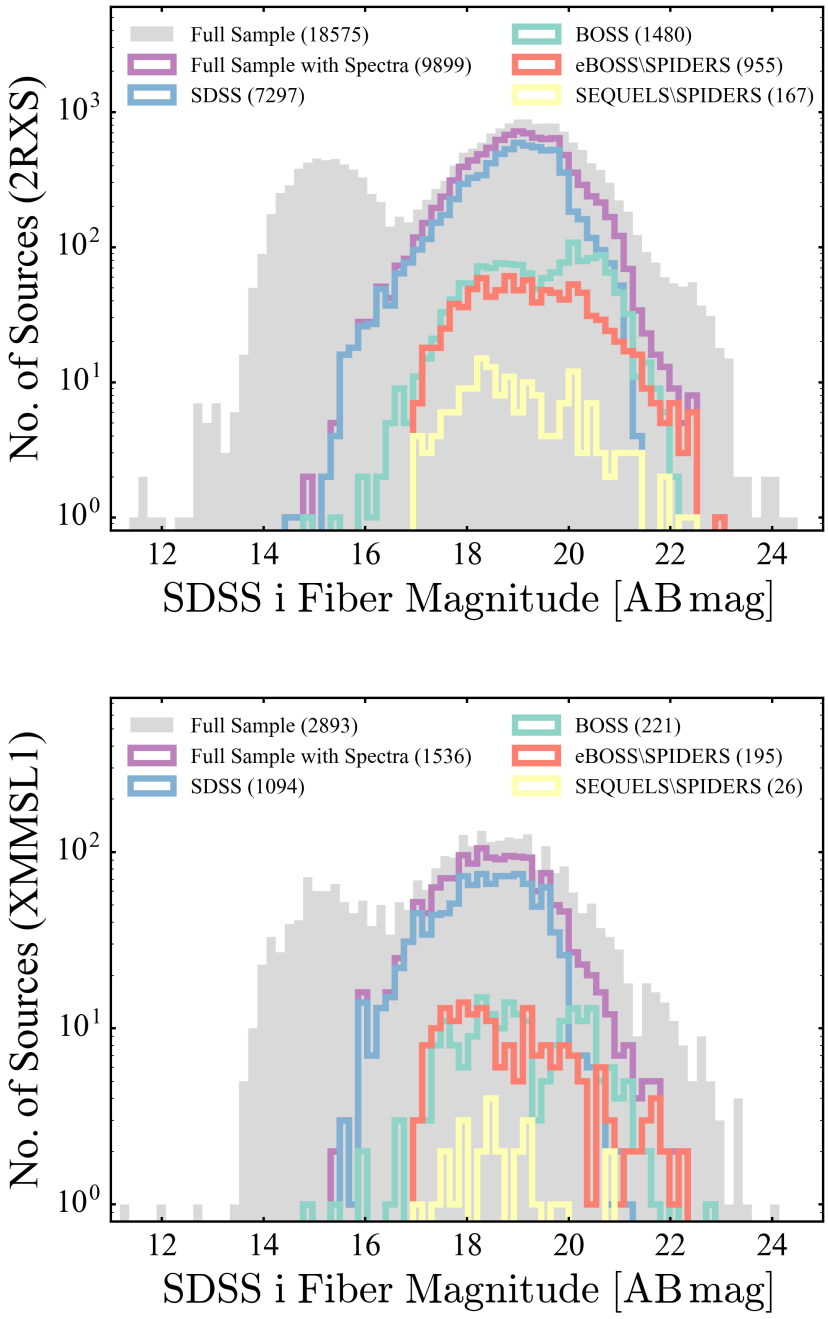


Figure 2.4: Distribution of i-band fiber magnitudes (fiber2Mag). The coloured curves represent all of the sources with spectra, and the survey from which the spectra were taken. The grey histogram displays the X-ray sources with a reliable SDSS photometric counterpart, including stars which cannot be targets for spectroscopy due to their brightness.

objects (90% and 98% for the 2RXS and XMMSL1 sources, respectively) had measured X-ray luminosities above or equal to $10^{42}\,\mathrm{erg\,s^{-1}}$. X-ray luminosities above this limit are expected to be primarily the result of AGN activity (see Aird et al. 2012; Lehmer et al. 2012; Menzel et al. 2016). These sources may be cases where the AGN emission is diluted by the host galaxy at optical wavelengths but is still detectable at X-ray wavelengths (e.g. Dwelly et al. 2017).

2.1.6 Contamination from Starforming Galaxies

Although the sample presented in this work probes luminosity ranges typically associated with AGN emission, starforming galaxies are also powerful X-ray sources and may be present as contaminants in this dataset. The X-ray emission from starforming galaxies is expected to originate from a number of energetic phenomena including supernova explosions and X-ray binaries (e.g. Persic & Rephaeli 2002). Therefore, the X-ray emission from starforming galaxies can be expected to be correlated with the star formation rate (SFR). Using their sample of luminous infrared galaxies, and a sample of nearby galaxies from Ranalli et al. (2003), Pereira-Santaella et al. (2011) found that the total SFR is related to the soft X-ray luminosity as follows:

$$\mathrm{SFR_{IR+UV}}\,(\mathrm{M_\odot\,yr^{-1}}) = 3.4 \times 10^{-40}\,\mathrm{L_{0.5-2keV}}\,(\mathrm{erg\,s^{-1}}) \tag{2.1}$$

Ilbert et al. (2015), figure 3, show the specific SFR for the COSMOS (Scoville et al. 2007) and GOODS (Giavalisco et al. 2004) surveys for a series of redshift bins in the range $0.2 < z < 1.4$. The peak of the redshift distribution of the 2RXS/XMMSL1 samples presented in this work is ~ 0.25. Therefore, assuming that the COSMOS/GOODS sample in the redshift bin 0.2-0.4 is representative of the 2RXS/XMMSL1 samples, the upper limit on the SFR that can be expected for galaxies in our sample is $\sim 50\,\mathrm{M_\odot\,yr^{-1}}$. According to equation 2.1, this corresponds to a soft X-ray luminosity of $\sim 10^{41}\,\mathrm{erg\,s^{-1}}$, which is below the lower range probed by our samples ($\sim 10^{42}\,\mathrm{erg\,s^{-1}}$, see figure 2.1).

2.1.7 Redshift Constraints

Using the "CONF_BEST" flag, sources with uncertain redshift or spectral classification (identified during the visual inspection of the sample) were also removed. This process resulted in a sample of 7795/7805 2RXS sources and 1190/1192 XMMSL1 sources. In the spectral fitting procedure (described in section 2.2), the Hβ and MgII lines were fit independently. Sources with Hβ and MgII present in their optical spectrum were selected using the following logic:

$$\begin{aligned} \mathrm{H}_\beta : (\mathrm{SN\_MEDIAN\_ALL} > 5)\ \&\& \\ (((\mathrm{INSTRUMENT} == \mathrm{SDSS})\ \&\&\ (0 < \mathrm{Z\_BEST} < 0.81))\ \|\| \\ ((\mathrm{INSTRUMENT} == \mathrm{BOSS})\ \&\&\ (0 < \mathrm{Z\_BEST} < 1.05))) \end{aligned}$$

$$\begin{aligned} \text{MgII} : (\text{SN\_MEDIAN\_ALL} > 5) \;\&\& \\ (((\text{INSTRUMENT} == \text{SDSS}) \;\&\& \; (0.45 < \text{Z\_BEST} < 2.1)) \;\| \\ ((\text{INSTRUMENT} == \text{BOSS}) \;\&\& \; (0.38 < \text{Z\_BEST} < 2.5))) \end{aligned}$$

Different redshift ranges have been used because the BOSS spectrograph has a larger wavelength coverage than the SDSS spectrograph. In some cases, parts of the fitting region will have been redshifted out of the SDSS/BOSS spectrograph range (Smee et al. 2013), and therefore will not be fit. However, the redshift limits where chosen so that both samples contain the broad lines used for estimating black hole mass. Sources with a median signal-to-noise ratio (S/N) less than or equal to five per resolution element were excluded from the spectral analysis since for these sources the broad line decomposition and resulting black hole mass estimates may be unreliable (see Denney et al. 2009; Shen et al. 2011).

Table 2.1 lists the numbers of sources with spectral coverage of either Hβ or MgII, while figure 2.5 shows the redshift distribution of these sources. There are 711 cases where the same optical counterpart was detected by both 2RXS and XMMSL1. The final combined (2RXS and XMMSL1) sample for spectral analysis contains 7790 unique type 1 sources. The S/N distribution for this sample of 7790 sources is shown in figure 2.6.

Table 2.1: The coverage of the Hβ and MgII emission lines in the two samples used in this work. There are 711 sources which were detected in both the 2RXS and XMMSL1 surveys. The "total" row lists the total number of unique sources obtained from combining the 2RXS and XMMSL1 samples.

	MgII	Hβ	Hβ and MgII	Hβ or MgII
2RXS	3310	6268	2234	7344
XMMSL1	314	1070	227	1157
Total	3473	6654	2337	7790

2.2 Spectral Analysis

A series of scripts have been written to perform spectral fits using the MPFIT least-squares curve fitting routine (Markwardt 2009). Each spectrum was corrected for Milky Way extinction using the extinction curve from Cardelli et al. (1989), and the dust map from Schlegel et al. (1998), with an R_V=3.1. No attempt has been made to estimate

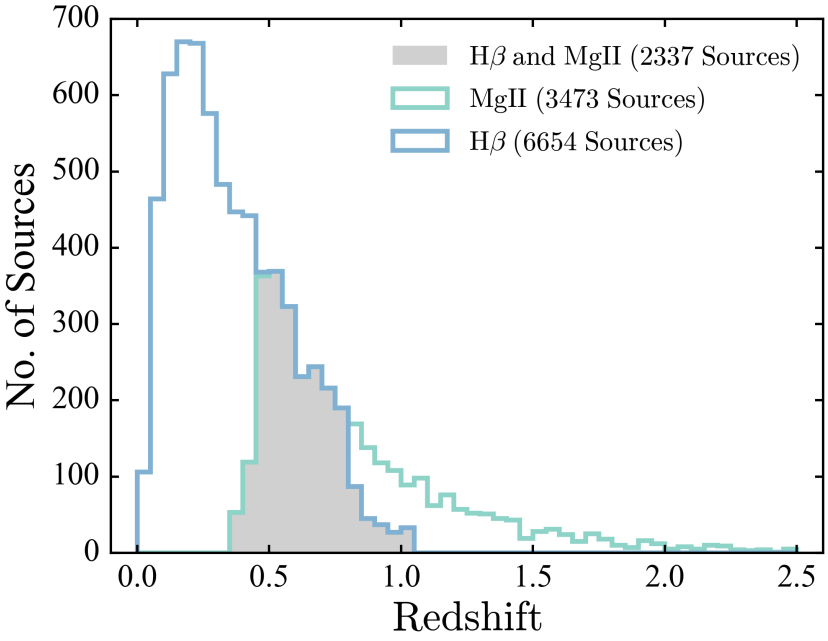


Figure 2.5: Redshift distribution of the sample of type 1 AGN with coverage of Hβ and/or MgII.

and correct for the intrinsic (host) extinction of each source[2]. Measured line widths were corrected for the resolution of the SDSS/BOSS spectrographs (see section 2.2.1). The Hβ and MgII emission line regions were fit independently using similar methods described in the following sections[3].

2.2.1 Correcting Line Widths for Spectrograph Resolution

The resolving power of a spectrograph, $R(\lambda_o)$, is defined as

$$R(\lambda_o) \equiv \frac{\lambda_0}{FWHM_R}$$

where the spectrograph resolution, $FWHM_R$, is the one-dimensional point spread function FWHM at a given wavelength λ_o, in wavelength units (Smee et al. 2013). The change in line width ($\sigma_R = FWHM_R/2\sqrt{2ln2}$) due to resolution is therefore

$$\sigma_R = \frac{\lambda_o}{R(\lambda_o)2\sqrt{2ln2}}$$

[2] Also note that extinction laws (e.g. Calzetti and Prevot) are based on samples of nearby SB and irregular type galaxies. Due to the lack of nearby passive galaxies, an extinction law for these galaxy types is not yet available.

[3] For each model parameter, the 1-sigma uncertainties from MPFIT were adopted.

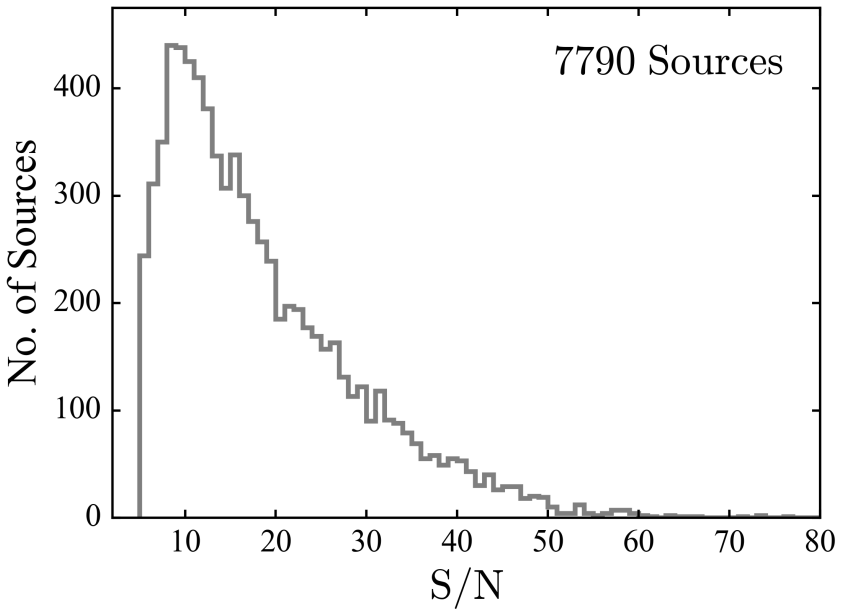


Figure 2.6: Distribution of the median signal-to-noise per resolution element for the sample of 7790 type 1 AGN discussed in section 2.1.7. A lower limit of S/N=5 was imposed in order to select sources for which the spectral fits can be expected to return reliable results.

λ_o is the observed frame wavelength, and therefore to compute R for a line with rest-frame peak wavelength λ_e, the equation above should be evaluated for $\lambda_o = (1+z)\lambda_e$ (e.g. Comparat et al. 2013)

$$\sigma_R(\lambda_o) = \frac{\lambda_o}{R(\lambda_o)2\sqrt{2\ln 2}} \tag{2.2}$$

The observed frame measured line width, $\sigma_{o,uncor}$, is a combination of the observed frame intrinsic (physical) line width, $\sigma_{o,cor}$, and the additional broadening induced by the resolution of the spectrograph, σ_R. $\sigma_{o,cor}$ can therefore be expressed as

$$\sigma_{o,cor} = \sqrt{\sigma^2_{o,uncor} - \sigma^2_R}$$

Peterson et al. (e.g. 2004). Inserting equation 2.2 gives

$$\sigma_{o,cor} = \sqrt{\sigma^2_{o,uncor} - \left(\frac{\lambda_o}{R(\lambda_o)2\sqrt{2\ln 2}}\right)^2}$$

For a source observed at redshift z, the observed frame and rest frame Gaussian widths are related as $\sigma_o = \sigma_e(1+z)$. Before the models are fit, the spectra are shifted to their rest-frame, and therefore the rest frame uncorrected Gaussian widths, $\sigma_{e,uncor}$, are measured,

and the rest frame intrinsic Gaussian width is calculated using

$$\sigma_{\mathrm{e,cor}} = \sqrt{\sigma^2_{\mathrm{e,uncor}} - \left(\frac{\lambda_{\mathrm{o}}}{\mathrm{R}(\lambda_{\mathrm{o}})(1+\mathrm{z})2\sqrt{2\ln 2}}\right)^2}$$

2.2.2 General Outline of the Model

The fitting routine first carries out a power law fit to the continuum region (excluding the emission line regions) of the form

$$\mathrm{f_{PL}}(\lambda) = \mathrm{N_{PL}}\lambda^{\alpha}$$

where $\mathrm{N_{PL}}$ is the power law normalisation in units of $10^{-17}\,\mathrm{erg\,cm^{-2}\,s^{-1}\,\AA^{-1}}$ and α is the slope. The best fit values for $\mathrm{N_{PL}}$ and α are then used as the initial guesses for the power law parameters in the second step in the fitting procedure. This second step fits the continuum region with the full continuum model, which consists of a power law, a galaxy template, and an iron (FeII) emission template. The full continuum model is

$$\mathrm{f_{cont}}(\lambda) = \mathrm{N_{PL}}\lambda^{\alpha} + \mathrm{N_{Gal}f_{Gal}}(\lambda) + \mathrm{N_{FeII}f_{FeII}}(\lambda) * \mathrm{g}(\sigma_{\mathrm{FeII}})$$

where $\mathrm{f_{Gal}}(\lambda)$ and $\mathrm{f_{FeII}}(\lambda)$ are the galaxy and FeII (see section 2.2.3) templates, $\mathrm{N_{Gal}}$ and $\mathrm{N_{FeII}}$ are the normalisations of the galaxy and FeII templates in units of $10^{-17}\,\mathrm{erg\,cm^{-2}\,s^{-1}\,\AA^{-1}}$, and g represents the Gaussian convolved with the FeII template. The best fit parameters for the continuum model are then passed to the final step, which involves fitting the spectral region with the full model, including both continuum and emission line components. Broad and narrow emission lines are fit with a series of Gaussian profiles of the form

$$\mathrm{f_{line}}(\lambda) = \mathrm{N_{line}}\mathrm{e}^{-\frac{(\lambda-\lambda_{\mathrm{c}})^2}{2\sigma^2}}$$

where $\mathrm{N_{line}}$ is the normalisation in units of $10^{-17}\,\mathrm{erg\,cm^{-2}\,s^{-1}\,\AA^{-1}}$, λ_{c} is the central wavelength of the line, and σ is the line width (standard deviation). The full model fit is iterated five times, and the best fit parameters obtained by one iteration are passed as initial estimates to the next iteration. This was done in order to aid the minimisation routine in finding the true best fit parameters.

2.2.3 Iron Emission Template

At UV and optical wavelengths, many AGN show a pseudocontinuum of blended emission features associated with FeII. The large number of individual FeII emission lines present in AGN spectra are blended together due to both their close wavelength spacing and the motion of the gas from which they are emitted. The magnitude of this broadening is observed to vary significantly from $300\,\mathrm{km\,s^{-1}}$ (Vestergaard & Wilkes 2001) up to $9000\,\mathrm{km\,s^{-1}}$ (Hu et al. 2008). Boroson & Green (1992) have reported that the $\mathrm{FWHM_{H\beta}}$ and $\mathrm{FWHM_{FeII}}$ are similar, and therefore have suggested that the FeII emission originates in the BLR

(see also Shields et al. 2010). However, other studies have found evidence that the FeII emission originates at larger distances than the Hβ emitting region (Kuehn et al. (2008); Hu et al. (2008) and also see Marziani et al. 2003). Baldwin et al. (2004) have shown that a photoionisation model alone cannot account for the observed UV FeII emission. Instead, they show that in order to reproduce the observed FeII emission, they require a model consisting of either photoionisation and microturbulence (with a microturbulent velocity parameter $> 100\,\mathrm{km\,s^{-1}}$) or a model where the FeII emission originates from a collisionally ionised region that is separate from the classical BLR. The presence of FeII emission in the optical and UV portions of the spectrum can be a significant complication when attempting to accurately measure broad emission line profiles. Therefore it is crucial that the model used to derive line widths for black hole mass measurements also accounts for the nearby FeII emission.

The spectra of NLS1 galaxies are especially useful for deriving FeII emission templates which can then be used in the spectral analysis of other classes of AGN. This is due to a number of reasons: Typically, NLS1 FeII emission is relatively strong, and therefore allows even weak FeII transitions to be detected. The "broad" emission lines are quite narrow (FWHM $< 2000\,\mathrm{km\,s^{-1}}$), and therefore FeII lines around the broad line component can be observed. In addition, since the virial broadening is low, the FeII lines are less blended than in broad line AGN, which allows FeII transitions with smaller wavelength separations to be resolved.

Figure 2.7 shows the two FeII templates used in this work; the Vestergaard & Wilkes (2001) and Boroson & Green (1992) templates used for the UV and optical regions of the spectrum, respectively. Both of these templates have been derived from spectra of the NLS1 galaxy I Zwicky 1, which has been selected due to its unusually narrow emission lines and its relatively strong FeII emission. Both groups have subtracted the AGN continuum emission, as well as any emission and/or absorption features not associated with FeII. As can be seen from figure 2.7, strong FeII emission is present in both the Hβ and MgII emission line regions. These template were incorporated into the model used in this work to fit the AGN continuum emission (see section 2.2.2). In order to model the observed blending of the FeII emission, the templates were convolved with a Gaussian whose width was included as a free parameter in the fitting procedure.

2.2.4 Hβ

The region from 4420-5500 Å was fit for each spectrum. The continuum model consisted of a power law, a galaxy template, and the Boroson & Green (1992) FeII emission template. The FeII template was convolved with a Gaussian while fitting, and the width of this Gaussian, along with the normalisation of the template were included as free parameters in the fit (see section 2.2.3). Previous analyses of AGN spectra have assumed an early-type galaxy component in the model (Calderone et al. 2017). Following this method, we use

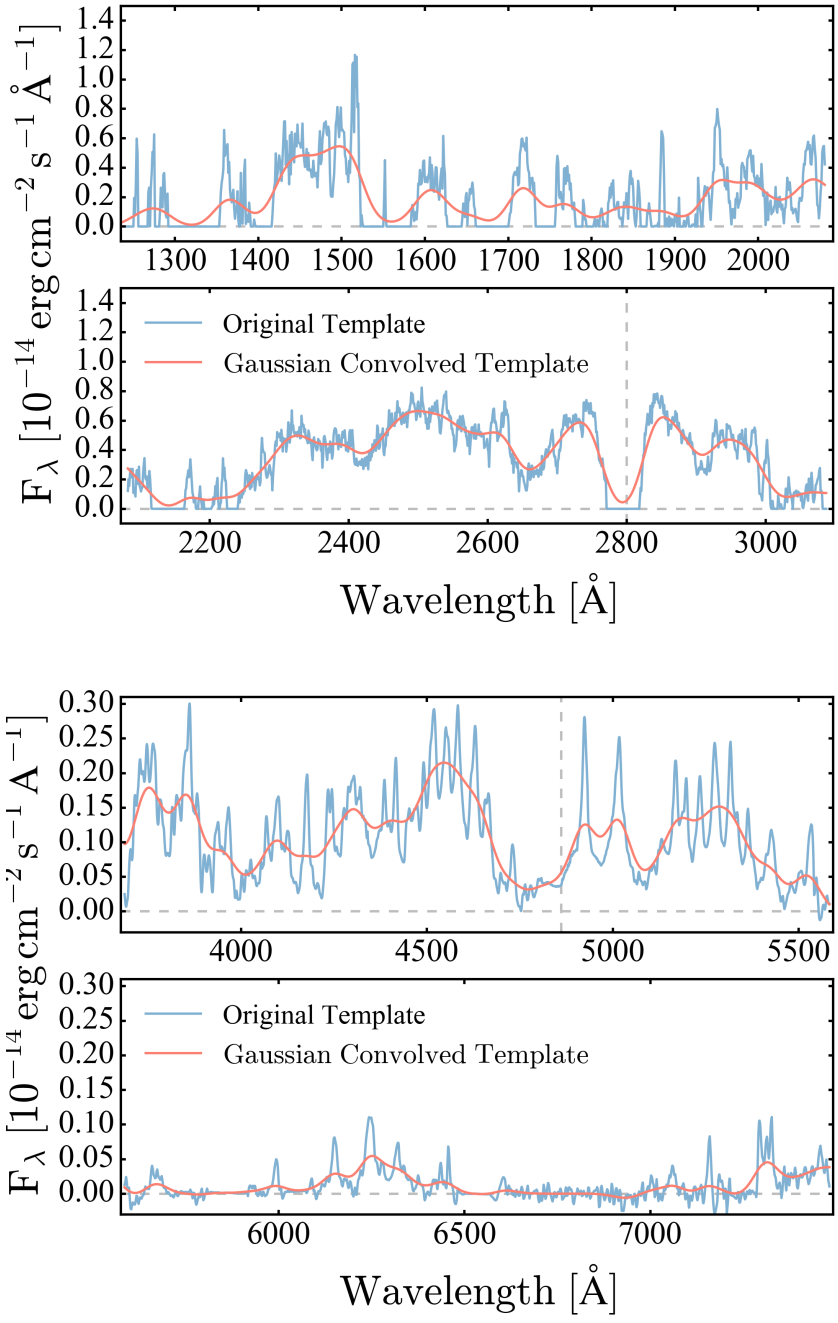


Figure 2.7: Upper panel: The Vestergaard & Wilkes (2001) FeII template which was used when fitting the MgII emission line region. Lower panel: The Boroson & Green (1992) FeII template which was used when fitting the Hβ emission line region. In both cases, the original template is shown (blue) along with the same template convolved with a Gaussian with FWHM $= 4000\,\mathrm{km\,s^{-1}}$ (red). The vertical dashed lines correspond to the position of MgII and Hβ at 2800Å and 4861Å.

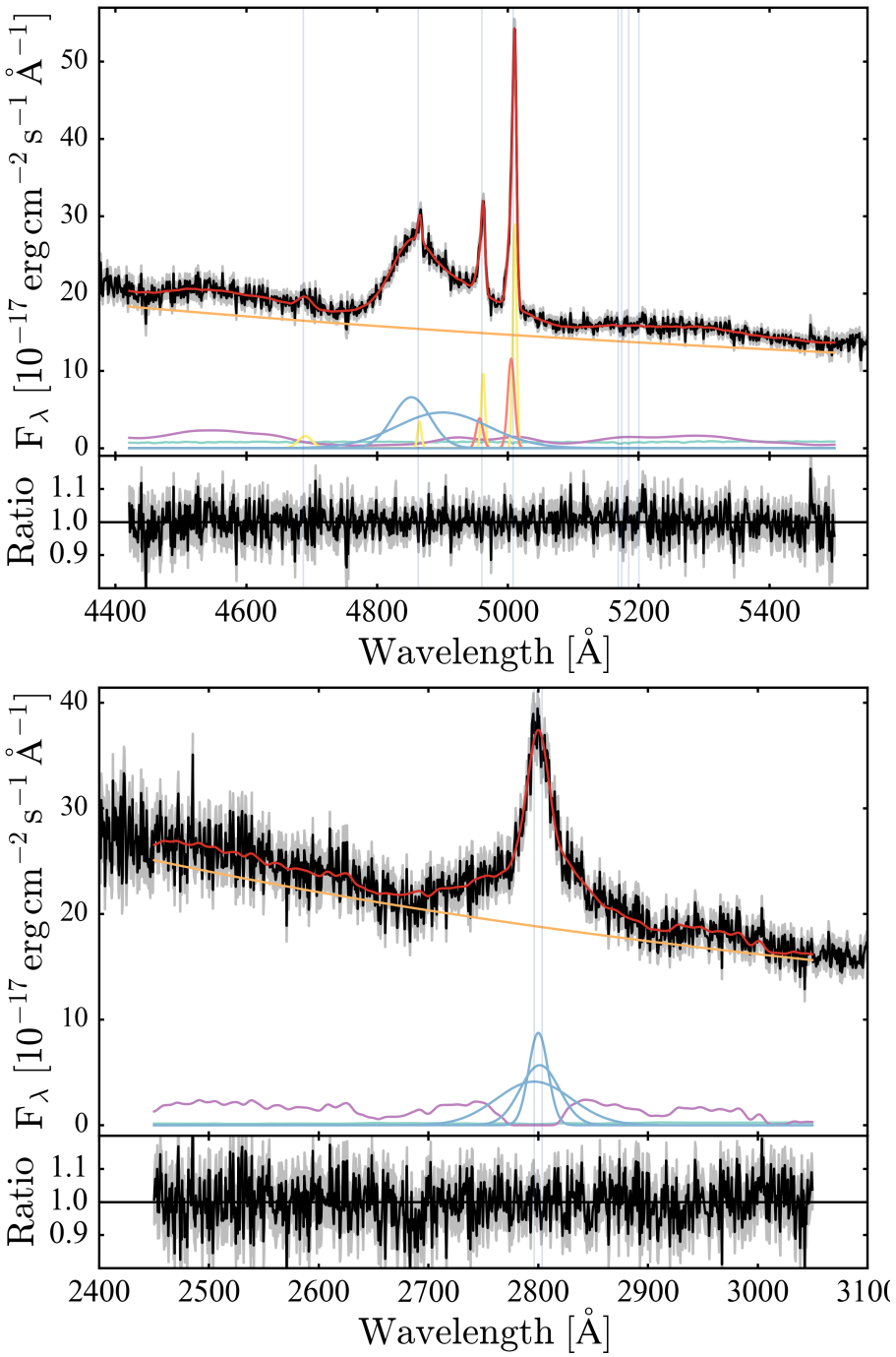


Figure 2.8: Examples of model fits to the Hβ (upper panel; plate=1159, MJD=52669, fiber=470) and MgII (lower panel; plate=423, MJD=51821, fiber=250) spectral regions. The model components are colour-coded as follows; power law (orange), iron emission (violet), broad lines (blue), narrow lines (yellow), [OIII] shifted wings (red), and the total model (red). The panels beneath the spectra show the data/model ratio.

an early-type SDSS galaxy template[4] in the fit, and the normalisation of this template as well as the normalisation and slope of the power law were also free parameters. The use of a single, early-type galaxy template is an approximation. It is considered to be justified however, since AGN are typically found to reside in bulge dominated galaxies (e.g. Grogin et al. 2005; Pierce et al. 2007), and the spectroscopic fiber collects emission mostly from the bulge (which is characterised by an old stellar population) and the active nucleus.

The [OIII]λ4959 and [OIII]λ5007 narrow lines were each fit with two Gaussians, one used to fit the narrow core, and an additional Gaussian to account for the presence of blue-shifted wings which are often detected (Boroson 2005). A single Gaussian was used to fit the HeIIλ4686 emission line. To avoid overfitting the Hβ line, the fitting process was run four times, with one, two, three, and four[5] Gaussian components used to fit the Hβ line. For each fit, the velocity width and peak wavelength of one of the Gaussian components was fixed to that of [OIII]λ4959 and [OIII]λ5007 in order to aid the identification of the narrow Hβ component. The normalisation ratio of the [OIII]λ4959 and [OIII]λ5007 lines was fixed to the expected value of 1:3 (e.g. Storey & Zeippen 2000). The best-fit model was then selected using the Bayesian information criterion (BIC; Schwarz 1978). BIC is a model selection criterion which discriminates between models using a penalty term which increases with the number of free parameters, and therefore avoids selecting a model that overfits the data. It can be written as

$$\mathrm{BIC} = \ln(\mathrm{n})\mathrm{k} + \chi^2$$

where n is the number of data points, k is the number of model parameters, and χ^2 is the chi-square of the fit. The preferred model is that with the lowest BIC. An example of a fit to the Hβ spectral region is shown in the upper panel of figure 2.8.

Broad Line Decomposition

The narrow Hβ and [OIII] components are required to have widths $\leq 800\,\mathrm{km\,s^{-1}}$. Any of the additional Gaussians used to fit MgII and Hβ with FWHM $> 800\,\mathrm{km\,s^{-1}}$ are considered "broad". This threshold of $800\,\mathrm{km\,s^{-1}}$ is taken from the approximate division between broad and narrow FWHM distributions in the lower panels of figure 3.5. The virial FWHM used for black hole mass estimation is the FWHM of the line profile defined by the sum of these broad Gaussian components (see figure 2.9). A major challenge with using the single-epoch method for estimating black hole mass is decomposing the broad and narrow components of the line in order to measure the virial FWHM. Figure 2.9 presents an example of the decomposition of a broad Hβ line. In this case, the narrow Hβ core can be easily distinguished and removed before measuring the virial FWHM. However, there are many cases where the broad and narrow components are blended, making it difficult

[4] template number 24 on http://classic.sdss.org/dr5/algorithms/spectemplates/

[5] Three broad Gaussians are used in addition to a single narrow component to account for the three distinct broad components that are expected to be present (see section 3.2.1) in at least some sources (Marziani et al. 2010).

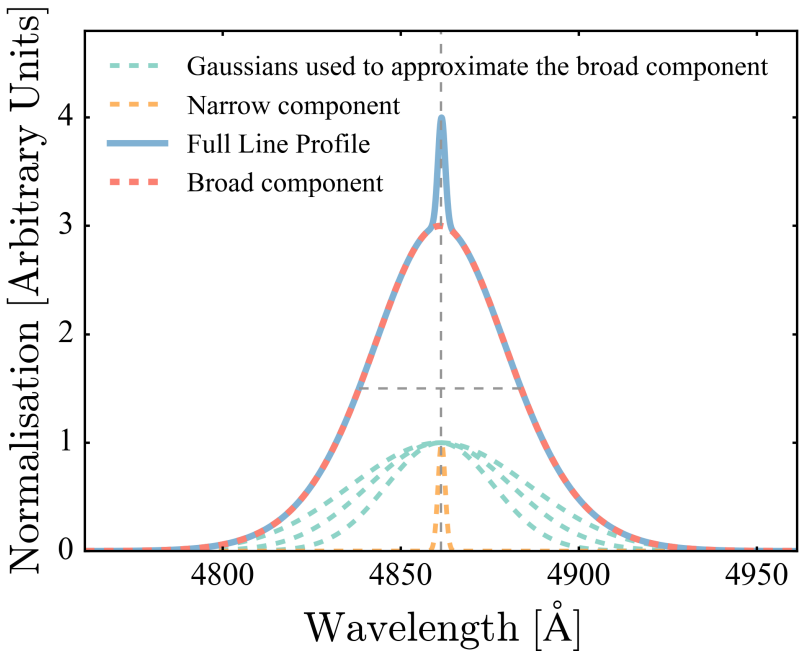


Figure 2.9: An example of the decomposition of a typical AGN Hβ line (plate=7276, MJD=57061, fiber=470). The horizontal dashed line represents the FWHM used for black hole mass estimation. The vertical dashed line indicates the rest-frame wavelength of Hβ. See section 2.2.4 for further details.

to successfully identify the appropriate virial FWHM. There are also cases where there is a clear distinction between two broad line components that are shifted in wavelength relative to each other (known as "double-peaked emitters"). How one should interpret the single-epoch black hole mass estimates for these unusual objects is uncertain (also see section 3.4.2).

2.2.5 MgII

The region from 2450-3050Å was fit for each spectrum. As in the case of the Hβ fits, a power law, an early-type galaxy template (5 Gyr old elliptical galaxy; Silva et al. 1998; Polletta et al. 2007), and the Vestergaard & Wilkes (2001) FeII emission template were used to fit the continuum. Again, the FeII template normalisation, and width of the Gaussian smoothing applied to the template, were included as free parameters in the fit. The MgII line is a doublet; however, due to the close spacing and virial broadening of the lines, it usually appears as a single broad component in AGN spectra. The narrow MgII line cores are usually not observed in AGN spectra, therefore the MgII profile was fit using three broad Gaussians. An example of a fit to the MgII spectral region is presented in the lower panel of figure 2.8.

Table 2.2: Black hole mass calibrations used in this work. A, B, and C are the calibration constants for single-epoch mass estimation (see equation 2.3).

	A	B	C	Reference
MgII, $L_{3000Å}$	1.816	0.584	1.712	Shen & Liu (2012)
H_β, $L_{5100Å}$	0.91	0.5	2	Vestergaard & Peterson (2006)
H_β, $L_{5100Å}$	0.895	0.52	2	Assef et al. (2011)

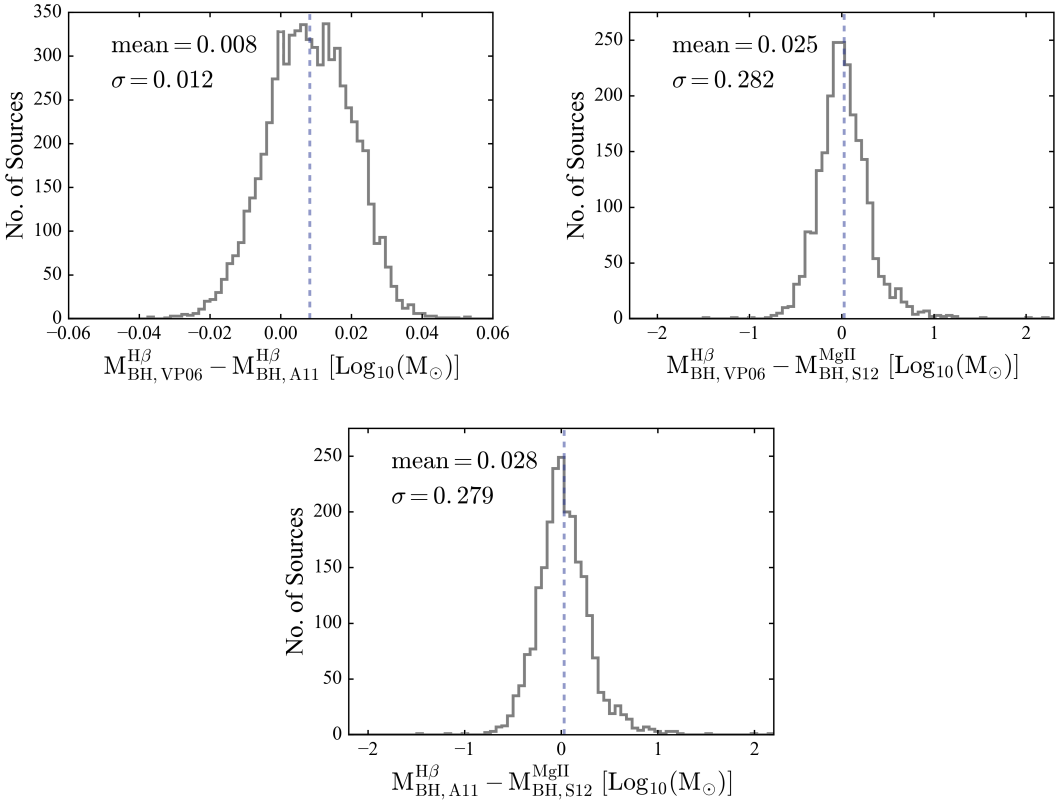


Figure 2.10: Differences between the BH mass calibrations used in this work (see section 2.3 for further details). The vertical blue lines indicate the mean value of each distribution. The standard deviation, σ, of each distribution is also shown.

2.3 Bolometric Luminosity and Black Hole Mass Estimation

Bolometric luminosities were estimated from the monochromatic luminosities using the bolometric corrections from Richards et al. (2006a); Shen et al. (2011):

$$L_{Bol} = 5.15\, L_{3000Å}$$
$$L_{Bol} = 9.26\, L_{5100Å}$$

These bolometric corrections have been derived using mean AGN SEDs; however, Richards et al. (2006a) note that using a bolometric correction resulting from a single mean SED can result in bolometric luminosities with inaccuracies up to 50%.

Under the assumption that the BLR gas is virialised, the single-epoch method can be used to estimate black hole mass as follows:

$$\log\left(\frac{M_{BH}}{M_{\odot}}\right) = A + B\log\left(\frac{\lambda L_{\lambda}}{10^{44}\,\mathrm{erg\,s^{-1}}}\right) + C\log\left(\frac{FWHM}{\mathrm{km\,s^{-1}}}\right) \tag{2.3}$$

where L_{λ} is the monochromatic luminosity at wavelength λ and FWHM is the full width at half maximum of the broad component of the emission line. A, B, and C are constants that are calibrated using RM results and vary depending on which line is used.

Over the years, many groups have provided calibrations of equation 2.3 for MgII and Hβ. In this work, the calibrations from Vestergaard & Peterson (2006) and Assef et al. (2011) are used for Hβ. Vestergaard & Peterson (2006) based their work on an updated study of the $R_{BH} - L$ relationship (Kaspi et al. 2005; Bentz et al. 2006) and a reanalysis of the RM mass estimates (Peterson et al. 2004) and therefore presented an improved mass calibration relative to previous studies. Assef et al. (2011) provide a mass calibration that is based on the $R_{BLR} - L$ relationship from Bentz et al. (2009b). Vestergaard & Peterson (2006) and Assef et al. (2011) both provide similar calibrations for single-epoch Hβ mass estimation, as can be seen from the left panel of figure 2.10.

The Shen & Liu (2012) calibration is used in this work for MgII. This calibration is based on a sample of 60 high-luminosity ($L_{5100Å} > 10^{45.4}\mathrm{erg\,s^{-1}}$) quasars in the redshift range 1.5 - 2.2. Shen & Liu (2012) use the Vestergaard & Peterson (2006) mass estimates as a reference when determining their MgII calibration. The centre and right panels of figure 2.10 show the comparison between the Shen & Liu (2012) MgII calibration and the Hβ calibrations from Vestergaard & Peterson (2006) and Assef et al. (2011). These calibrations agree reasonably well, with the standard deviation $\sigma \simeq 0.3$ in both cases, which is likely due to the fact that these BH mass estimates were derived using two different emission lines. A list of the three black hole mass calibrations used in this work is given in Table 2.2.

Black hole masses were computed for each of these calibrations and are included in the catalogue. Black hole masses were only estimated for sources with a detected broad line

component (see section 2.2.4). These black hole mass estimates were then used to estimate the Eddington luminosity and the Eddington ratio, defined as

$$L_{Edd} = 4\pi c G M_{BH} m_p / \sigma_T$$
$$\lambda_E = L_{Bol} / L_{Edd}$$

where c is the speed of light, G is the gravitational constant, M_{BH} is the black hole mass, m_p is the proton mass, and σ_T is the Thomson scattering cross-section.

2.4 X-ray Flux Estimates

Since X-ray detections are available for all objects in this sample, X-ray flux estimates have also been included in the catalogue. XMMSL1 fluxes in the 0.2-12 keV range from Saxton et al. (2008) are included. Saxton et al. (2008) convert the XMMSL1 count rates to fluxes using a spectral model consisting of an absorbed power law with a photon index of 1.7 and $N_H = 3 \times 10^{20} cm^{-2}$. The 2RXS fluxes were estimated using the method outlined below.

2.4.1 Estimating 2RXS X-Ray Fluxes

Many of the sources in the 2RXS sample have flux measurements close to the *ROSAT* flux limit ($\sim 10^{-13}$ erg cm^{-2} s^{-1}). Therefore, when estimating fluxes for this sample, it was necessary to correct for the Eddington bias. This was done by adopting a Bayesian method to derive a probability distribution of fluxes based on the known distribution of AGN as a function of flux. Following Kraft et al. (1991), Laird et al. (2009), and Georgakakis & Nandra (2011), the probability of a source having flux f_X, given an observed number of counts C, is

$$P(f_X, C) = \frac{T^C e^{-T}}{C!} \pi(f_X) \tag{2.4}$$

where C is the total number of observed source and background counts, T is the mean expected total counts in the detection cell for a given flux, and $\pi(f_X)$ is the prior. Source and background counts were taken from the 2RXS catalogue (Boller et al. 2016). The distribution of AGN per unit X-ray flux interval (Georgakakis et al. (2008), equation 1) is used as the prior. The inclusion of this term aims to correct for the Eddington bias. T in the equation above can be calculated as follows

$$T = f_x \times ECF \times EEF \times t + B \tag{2.5}$$

where ECF is the flux-count rate conversion factor, EEF is the energy encircled factor, t is the exposure time, and B is the number of background counts. The ECF was derived using `XSPEC` (Arnaud 1996) assuming a model consisting of a power law (with $\Gamma = 2.4$, following Dwelly et al. 2017) absorbed by the Milky Way column density (using the PHABS model).

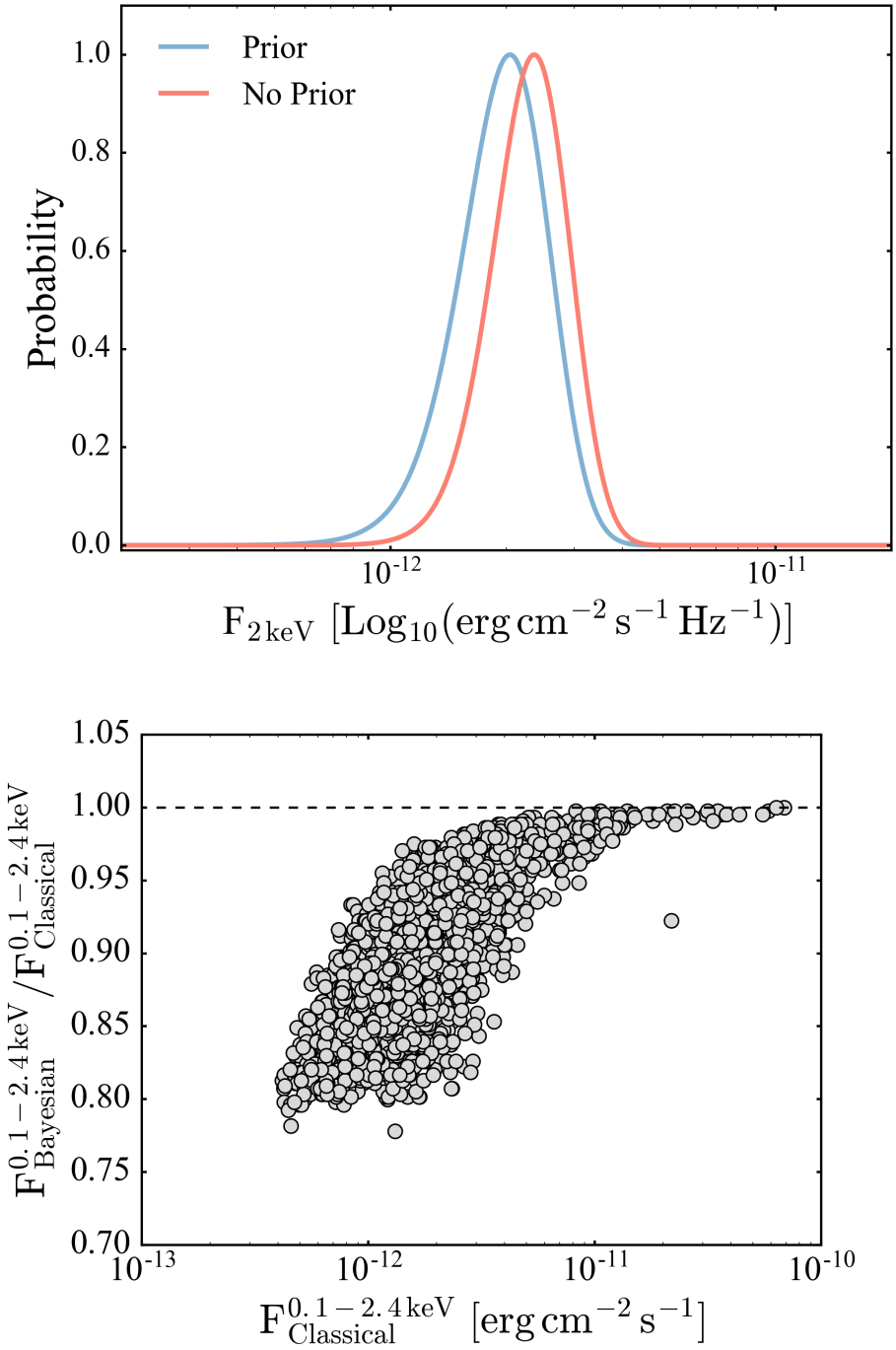


Figure 2.11: Upper panel: Example of a flux probability distribution with and without applying the prior. Lower panel: Comparison between the classical and Bayesian methods for estimating 2RXS fluxes. The deviation from a ratio of one at fainter fluxes results from the attempt to correct for the effect of the Eddington bias (see section 2.4.1 for further details).

This method was used to estimate the flux in the full ROSAT band (0.1-2.4 keV) as well the monochromatic flux at 2 keV.

The fluxes resulting from the method described above with and without applying the prior (termed "Bayesian" and "classical", respectively) are compared in the lower panel of figure 2.11. The disagreement between the two flux estimates increases with decreasing flux, which is expected since, without the prior, the classical method fails to account for the Eddington bias. Low count rate sources in this sample would be assigned unrealistically low Bayesian fluxes. To avoid this, the flux was left as undetermined when the Bayesian flux estimate was more than a factor of ten smaller than the classical flux estimate.

2.5 Comparison with the Shen et al. (2011) SDSS DR7 Catalogue

Shen et al. (2011) have published a catalogue of optical spectroscopic measurements for the sample of SDSS DR7 (Abazajian et al. 2009) quasars. Figure 2.12 shows the comparison between the monochromatic luminosities, broad emission line FWHM, and BH mass estimates presented in this work and in Shen et al. (2011). A sample of 2000 randomly selected sources with measurements from both the UV and optical portions of the spectrum was selected in order to allow a fair comparison of mean values and standard deviations.

The two upper panels of figure 2.12 show the difference between the UV and optical monochromatic luminosities from this work and Shen et al. (2011). In both cases, the monochromatic luminosities presented in this work are slightly larger, which is due to the fact they have been measured directly from the observed AGN continuum, while Shen et al. (2011) measure their luminosities from their best-fit power law. In this work, luminosities were measured from the AGN continuum directly in order to avoid the uncertainties associated with identifying the intrinsic AGN power law and removing the contributions from the FeII pseudo-continuum and the host galaxy contribution. Despite this difference in approach, the offset between the two sets of measurements is quite small compared to the standard deviation of the distributions, especially for $L_{5100\text{Å}}$.

The two middle panels of figure 2.12 show the difference in Hβ and MgII FWHM for both samples. The results presented in this work agree well with those presented in Shen et al. (2011), which is expected, since the methods used to decompose the emission line profile and measure the virial FWHM are very similar. The sources which show a large disagreement are likely cases where the amount of FeII emission subtracted from the region around the line was different. This is evident from the fact that the offset for MgII is larger than that of Hβ; the MgII spectral region typically exhibits much stronger FeII emission than the Hβ region. Finally, the two lower panels show the comparison of the BH mass estimates from this work and Shen et al. (2011), which agree reasonably well considering the large (0.4 dex) uncertainties associated with single-epoch BH mass estimates.

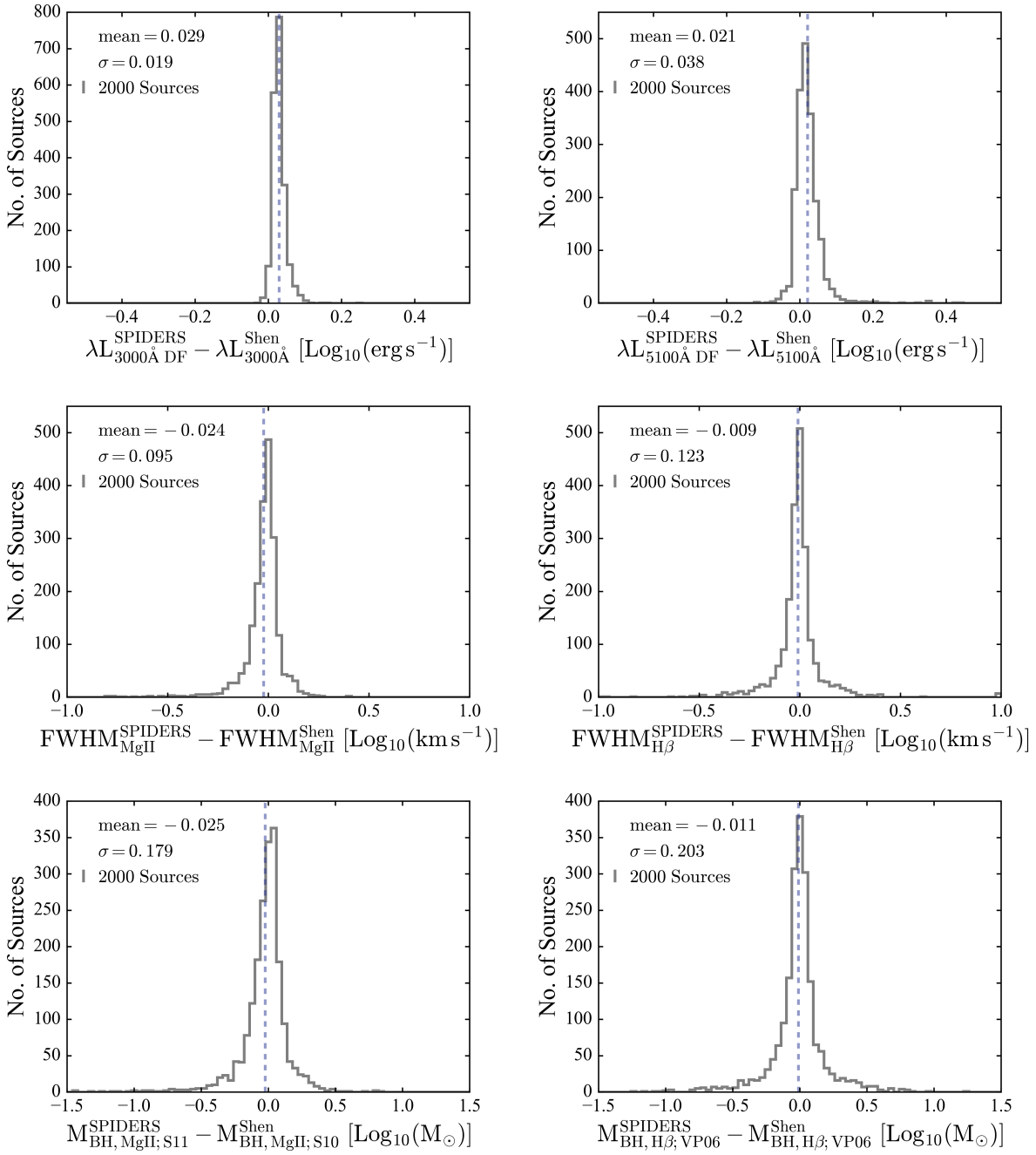


Figure 2.12: Comparison between the monochromatic luminosities (upper panels), broad emission line FWHM (middle panels), and BH mass estimates (lower panels) presented in this work and the Shen et al. (2011) SDSS DR7 sample. In each panel, the blue dashed line shows the mean value. The standard deviation of each distribution, σ, is also given.

2.6 Accessing the Data

The results from the spectral analysis discussed above, along with X-ray flux measurements and visual inspection results, have been made available in an SDSS DR14 value added catalogue which is available at http://www.sdss.org/dr14/data_access/value-added-catalogs/. Additionally, an extended version of the catalogue will be maintained at http://www.mpe.mpg.de/XraySurveys/SPIDERS/SPIDERS_AGN/

2.7 Overview and Scientific Applications

In this chapter, the catalogue of optical spectroscopic properties of SPIDERS SDSS DR14 X-ray selected type 1 AGN was presented. This dataset represents a well-defined, low-redshift, and high-luminosity sample of 7790 AGN. The redshift and spectroscopic classification of each source have been visually confirmed, resulting in a reliable set of optical counterparts to the X-ray detections. Detailed modelling has provided optical spectroscopic measurements of the spectral regions around the MgII and Hβ emission lines, along with derived quantities such as BH masses and Eddington ratios. A catalogue containing these results has been made available as part of a set of SDSS DR14 value added catalogues.

In the following chapter, the SPIDERS AGN catalogue is used to investigate a number of physical processes in AGN, the main focus of which is the interaction between the ionising source and the surrounding regions. In addition to the applications presented therein, the SPIDERS AGN catalogue is intended to provide the scientific community with a useful resource for the multiwavelength analysis of AGN. For example, this sample may be specifically useful for the study of the high-luminosity portion of the accreting BH population. In addition, the extensive coverage at low-redshifts may render the sample particularly relevant to the study of AGN host galaxy properties. The detailed decomposition of the low-ionisation broad emission lines MgII and Hβ can facilitate the exploration of the BLR geometry and kinematics (e.g., Wolf et al. 2020).

As noted previously, eROSITA will observe most of the X-ray sources present in the current SPIDERS catalogues. Therefore, SPIDERS provides optical spectroscopic counterparts for a large sample of sources that will be detected in the eROSITA all-sky survey. The sample presented in this work may therefore tentatively be expected to provide a useful tool for the study of AGN well into the era of eROSITA.

Chapter 3

SPIDERS: Accretion Driven Properties and BH Mass Evolution

This chapter presents the scientific application of the catalogue presented in chapter 2. Parts of this chapter were also published in Coffey et al. (2019).

3.1 Comparing the UV and Optical Spectral Measurements

The catalogue presented in chapter 2 includes a large subsample of sources that have spectral coverage of both the MgII and Hβ regions of the spectrum (see table 2.1). This subsample allowed the relationship between the UV and optical spectral properties to be investigated. In section 3.1.1, the 2500Å and 5100Å monochromatic luminosities, $L_{2500Å}$ and $L_{5100Å}$, are compared in order to explore the consistency of the independent modelling of both spectral regions. An empirically derived conversion between the 2500Å and 5100Å monochromatic luminosities was obtained, which can be used to include low-redshift sources in the α_{OX} analysis presented in section 3.2.3. Section 3.1.2 presents an analysis of the relationship between the MgII and Hβ FWHM measurements.

3.1.1 Relationship Between the UV and Optical Monochromatic Luminosities

A subsample of AGN with spectral coverage of the rest-frame wavelengths 2500Å and 5100Å was selected using the following criteria:

$$((\text{INSTRUMENT} = \text{SDSS}) \;\&\&\; (0.52 < \text{redshift} < 0.8)) \;||$$
$$((\text{INSTRUMENT} = \text{BOSS}) \;\&\&\; (0.46 < \text{redshift} < 1.04))$$

Of this sample, 1718 sources had reliable measurements of both $L_{2500Å}$ and $L_{5100Å}$. The $L_{2500Å} - L_{5100Å}$ relation was fit using the LINMIX (Kelly 2007) package. LINMIX is a Bayesian linear regression algorithm that accounts for uncertainties in both dependent and independent variables, as well as non-detections. The upper panel of figure 3.1 shows the $L_{2500Å} - L_{5100Å}$ distribution and the best-fit relation

$$Log_{10}(L_{5100Å}) = (0.841 \pm 0.007)Log_{10}(L_{2500Å}) - (6.0 \pm 0.3) \tag{3.1}$$

where the regression intrinsic scatter is 0.0151. Equation 3.1 can be used to estimate $L_{2500Å}$ from measurements of $L_{5100Å}$, which allows lower-redshift sources to be included in experiments that require UV luminosities (e.g., the analysis of the α_{OX} relation, discussed in section 3.2.3). The comparison between the bolometric luminosity estimates derived from $L_{2500Å}$ and $L_{5100Å}$ (see section 2.3) is shown in the lower panel of figure 3.1.

3.1.2 Relationship Between the MgII and Hβ FWHM Measurements

Previous analyses of AGN emission lines have found that the broad MgII and Hβ FWHM are very similar (e.g. McLure & Jarvis 2002; McLure & Dunlop 2004) which would suggest that both lines are emitted from the same location in the BLR. However, Salviander et al. (2007) have shown that though the $FWHM_{MgII}$ and $FWHM_{H\beta}$ show a strong correlation, above 4000 km s^{-1} the $FWHM_{H\beta}$ tends to be larger than $FWHM_{MgII}$. The authors suggest that in some cases this may be due to excess emission from the red Hβ wing. Corbett et al. (2003) arrived at the opposite conclusion and reported that $FWHM_{MgII}$ is typically broader than $FWHM_{H\beta}$. Dietrich & Hamann (2004) found that for their sample of seven quasars at $z \simeq 3.5$ the $FWHM_{H\beta}$ tends to be slightly broader than the $FWHM_{MgII}$. In order to test this relationship using the sample presented in this work, a subsample of AGN whose spectra cover the broad MgII and Hβ emission lines was selected using the following criteria

$$\begin{aligned}((\text{INSTRUMENT} = \text{SDSS}) \;\&\&\; (0.45 < \text{redshift} < 0.81)) \;||\\ ((\text{INSTRUMENT} = \text{BOSS}) \;\&\&\; (0.38 < \text{redshift} < 1.05))\end{aligned}$$

These criteria return 2337 sources with coverage of both MgII and Hβ. Of this sample, 2323/2337 sources had FWHM measurements for both Hβ and MgII. The upper panel of figure 3.2 shows the distribution of $Log_{10}(FWHM_{H\beta}/FWHM_{MgII})$ values for this sample of 2323 sources, which has a mean of 0.046 and a standard deviation (dispersion) of 0.123 dex. For their samples, McLure & Dunlop (2004) found a mean of -0.004 and a dispersion of 0.16 dex, while Shen et al. (2008) reported a mean of 0.0062 and a dispersion of 0.11 dex. The lower panel of figure 3.2 displays the comparison of the MgII and Hβ virial FWHM

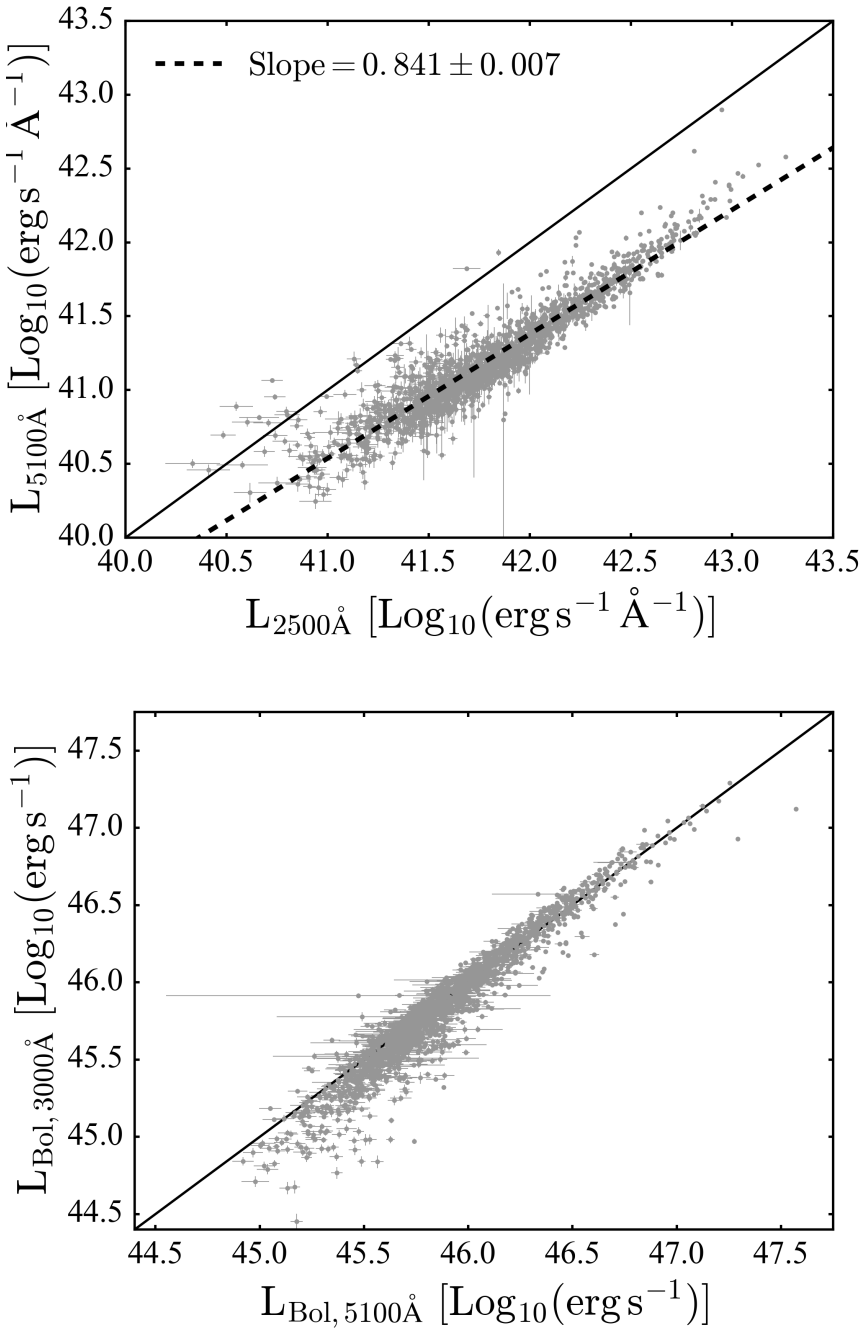


Figure 3.1: Upper panel: Comparison between the 2500Å and 5100Å monochromatic luminosities. The best-fit relation, illustrated by the black dashed line, is given in equation 3.1. Lower panel: Comparison between the bolometric luminosities estimated from the 5100Å and 3000Å monochromatic luminosities. In both panels, the solid black line is the unity line.

measurements for the 2323 sources in this sample. The resulting best-fit relation, fit using LINMIX, is

$$\mathrm{Log_{10}(FWHM_{MgII}) = (0.65 \pm 0.01)\,Log_{10}(FWHM_{H\beta}) + (1.21 \pm 0.04)}$$

with a regression intrinsic scatter of 0.0049. This deviation from the one-to-one relation has also been observed by Wang et al. (2009), who reported

$$\mathrm{Log_{10}(FWHM_{MgII}) = (0.79 \pm 0.03)\,Log_{10}(FWHM_{H\beta}) + (0.68 \pm 0.02)}$$

and Shen & Liu (2012), who found a slope of 0.57 ± 0.09. The results from this work strongly suggest that the $\mathrm{FWHM_{H\beta}}$ is broader than $\mathrm{FWHM_{MgII}}$. This conclusion is also supported by results from the SDSS reverberation mapping project, which has shown that the MgII emission line may respond more slowly to continuum variations than the Hβ emission line, suggesting that the MgII emission region is located at larger distances from the BH (Li et al. 2017). The single-epoch BH mass relations (equation 2.3) account for the $\mathrm{FWHM_{MgII}} - \mathrm{FWHM_{H\beta}}$ slope; when the correct BH mass calibration is used, the Hβ and MgII virial FWHM measurements yield BH masses that are in close agreement (see figures 2.10 and 3.3).

3.2 Sample Properties

Figure 3.4 presents the comparison between this sample and the full sample of SDSS DR7 AGN with optical spectral properties measured by Shen et al. (2011) in the bolometric luminosity-redshift and bolometric luminosity-black hole mass planes. As discussed in section 1, Hβ-derived black hole masses are used where available (shown in blue), while MgII-derived masses are used for the remaining higher-redshift sources (shown in green). The left panel of figure 3.4 shows that this sample populates the low-redshift, high-luminosity region of the parameter space, which is partially due to the high flux threshold of the X-ray selection.

3.2.1 Hβ Line Components

Section 2.2.4 described how the Hβ line profile was fit with either one, two, three, or four Gaussian components. Figure 3.5 displays the resulting distribution of Hβ FWHM measurements (where the panels are split based on the number of Gaussian components required to fit the line). There is a clear peak in the distribution at low FWHM associated with the narrow Hβ core typically measured at a few hundred $\mathrm{km\,s^{-1}}$. Above $\sim 1000\,\mathrm{km\,s^{-1}}$ the distribution is bimodal (in the two lower panels) with a large number of sources showing evidence for a component with a FWHM $\gtrsim 10000\,\mathrm{km\,s^{-1}}$, which is typically referred to as the "very broad component" (VBC).

The VBC is usually only observed in population B ($\mathrm{FWHM\,H_\beta^{BC}} \geq 4000\,\mathrm{km\,s^{-1}}$) AGN (see Marziani et al. 2009, 2010, 2015), and has been reported to show a significant velocity

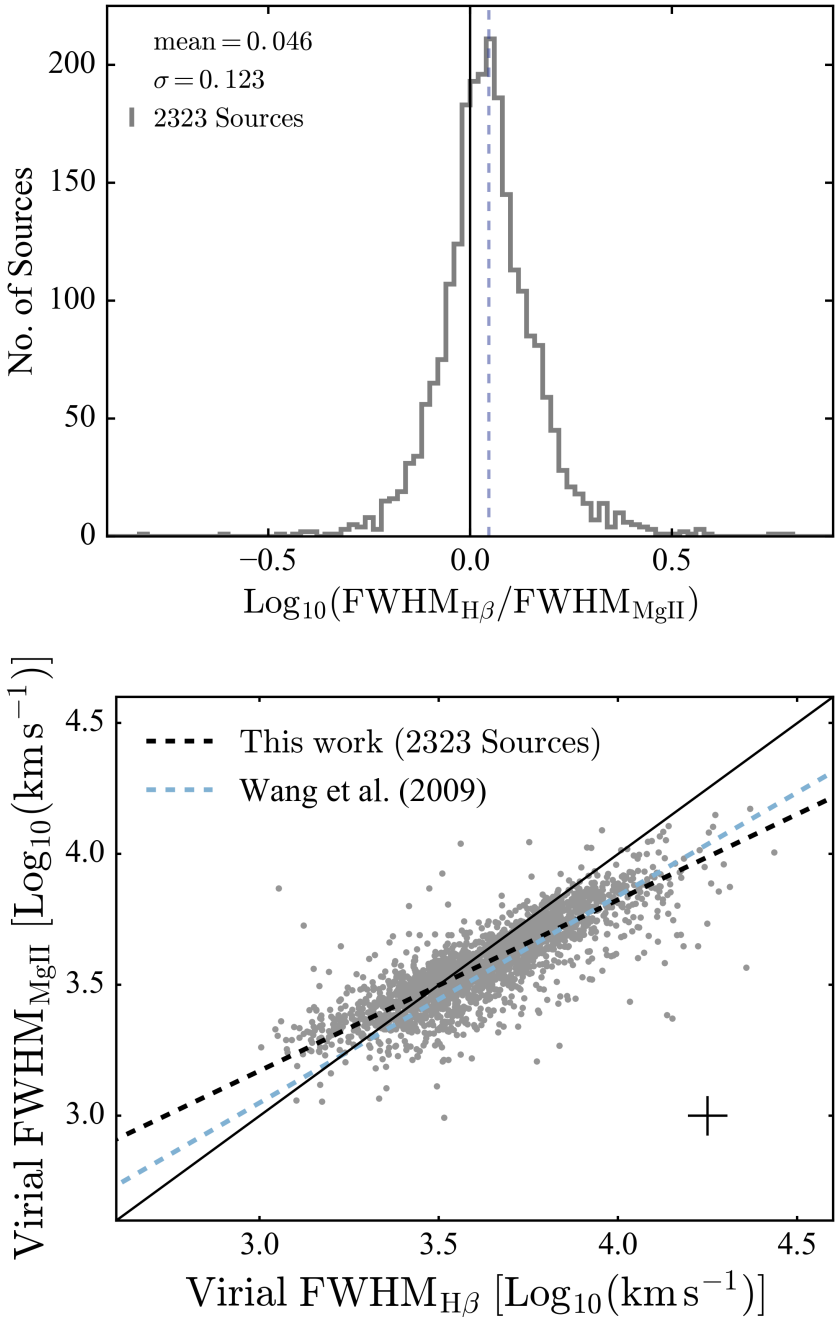


Figure 3.2: Upper panel: Ratio of the broad Hβ and MgII FWHM measurements for the subsample discussed in section 3.1.2. The mean value of the distribution (blue dashed line) and the standard deviation (σ) are both shown. Lower panel: Comparison of the FWHM measurements of Hβ and MgII for the same subsample. The best linear fit to the distribution is shown by the black dashed line. The best fit relation from Wang et al. (2009) (blue dashed line) is also shown for comparison. The interquartile mean of the uncertainties for both sets of measurements is shown, however, note that some sources have much larger uncertainties, possibly due to a lower spectral S/N. The solid line is the unity line.

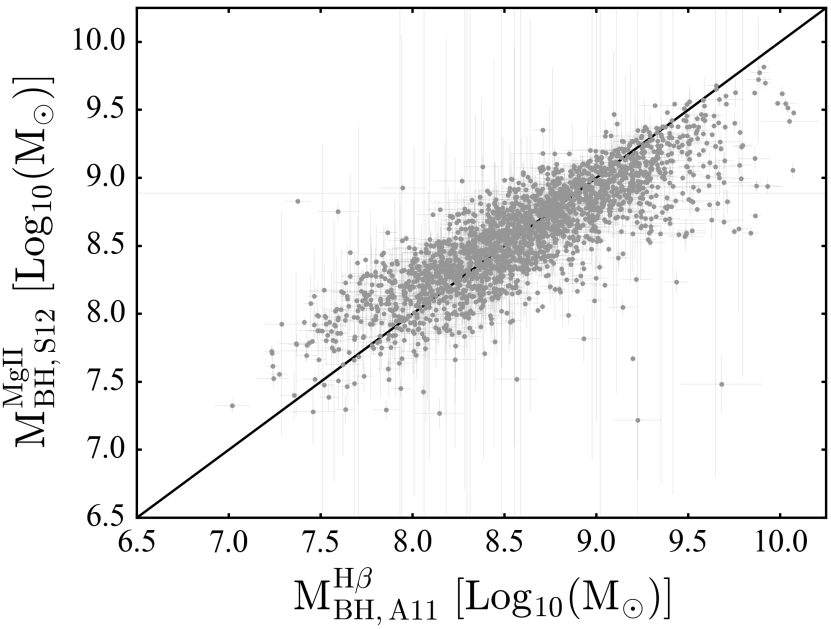


Figure 3.3: Comparison of the black hole mass estimates derived from Hβ (using the Assef et al. (2011) calibration) and MgII (using the Shen & Liu (2012) calibration). The solid black line is the unity line.

offset relative to the systemic redshift (e.g. Brotherton 1996; Sulentic et al. 2000a, 2002; Marziani et al. 2009; Zamfir et al. 2010; Marziani et al. 2010), which increases with the bolometric luminosity of the source (Corbin 1995). It has been suggested that the VBC is closer to the central ionising source than the classical BLR (e.g. Peterson & Ferland 1986; Wills et al. 1993; Corbin 1995; Corbin & Boroson 1996; Popović et al. 2004), and may have non-virialised motion (e.g. Marziani & Sulentic 2012). Bon et al. (2009) have proposed that the VBC is a result of line emission from the accretion disk.

If the VBC represents emission from the accretion disk, then a strong VBC may result in a bias towards a higher BH mass estimate, since the single-epoch method assumes a calibration that is based on the luminosity-BLR radius relation. However, since the kinematics and physical origin of the VBC remains uncertain, detected VBCs have not been excluded from the broad line profiles used to measure the virial FWHM in this analysis (as discussed in section 2.2.4).

3.2.2 This Sample in the 4D Eigenvector 1 Context

The 4D Eigenvector 1 (4DE1) system (Boroson & Green 1992; Sulentic et al. 2000b, 2011) aims to define a set of parameters that account for the diversity in the optical spectral properties of type 1 AGN. Two main 4DE1 parameters are the FWHM of the broad component of Hβ (FWHM H^{BC}_{β}) and the strength of the FeII emission relative to that of

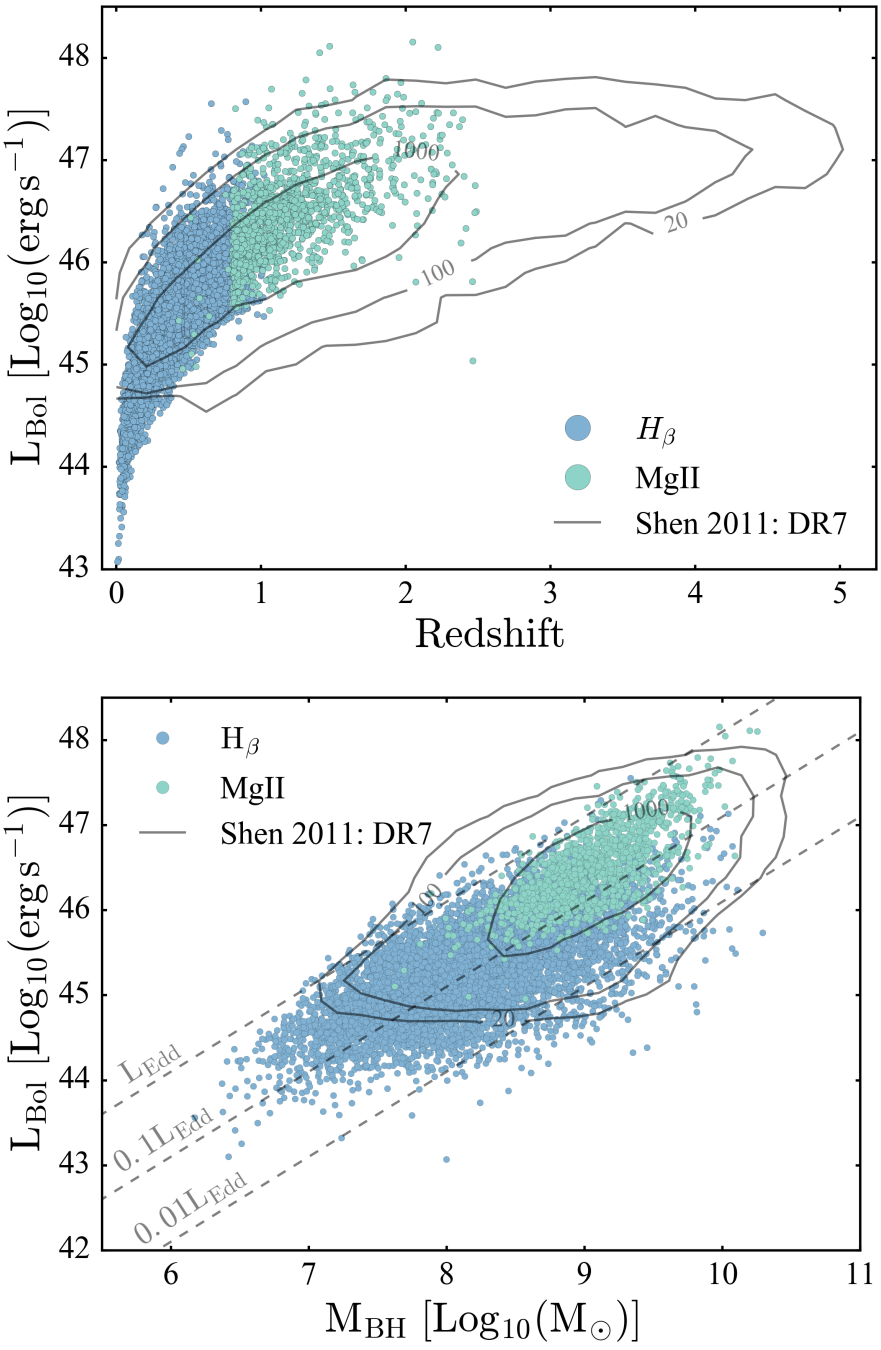


Figure 3.4: Bolometric luminosity versus redshift (upper panel) and bolometric luminosity versus BH mass (lower panel) for the sample presented in this work and the Shen et al. (2011) sample. Sources with Hβ-derived BH masses are shown in blue, and sources with MgII-derived BH masses are displayed in green.

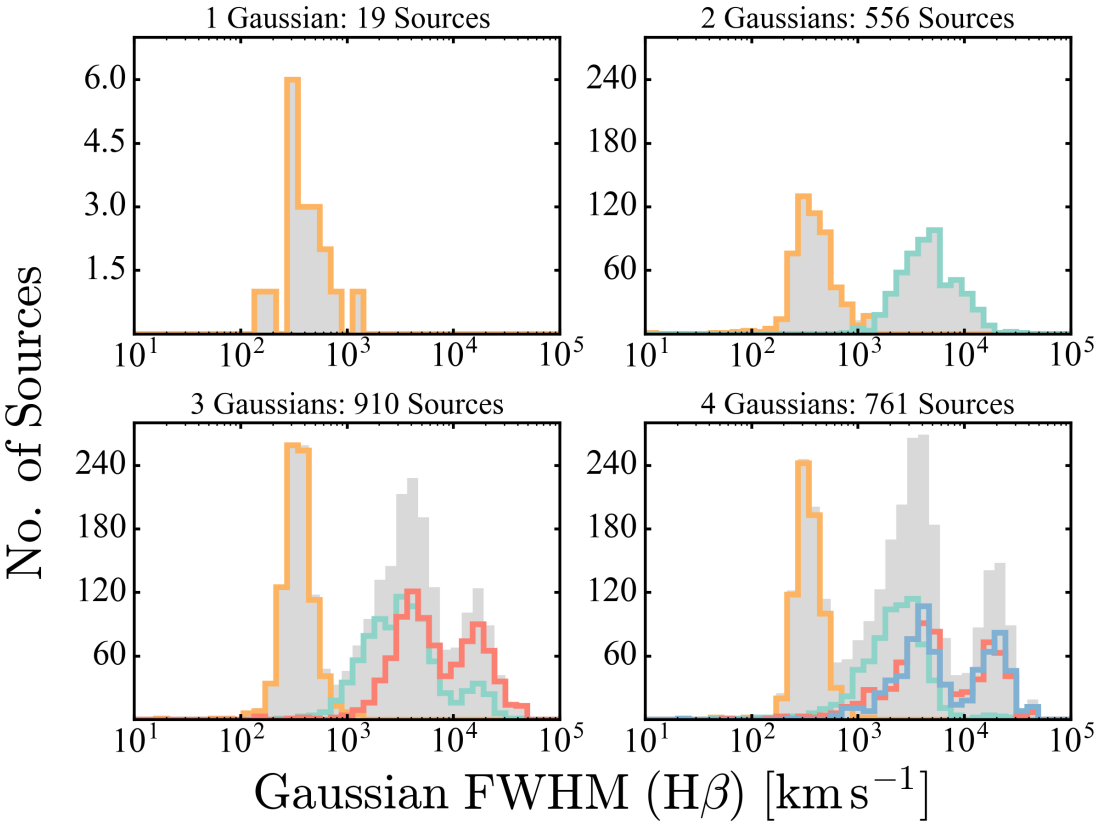


Figure 3.5: Distribution of Hβ Gaussian FWHM values. The panels are split based on the number of Gaussians required to fit the line. The coloured histograms each represent one of up to four possible Gaussians used to fit the Hβ line. The grey histograms represent the sum of the individual coloured histograms.

Hβ, defined as

$$R_{FeII} = F_{FeII}/F_{H\beta}$$

where F_{FeII} and $F_{H\beta}$ are the fluxes of the FeII emission in the 4434-4684Å range and broad Hβ line, respectively. A sample of 2098 sources with measurements of these parameters and reliable spectral fits ($0 \leq \chi^2_{\nu,H\beta} \leq 1.2$) was selected. The upper panel of figure 3.6 shows the distribution of this sample in the 4DE1 parameter space. It is expected that a reliable measurement of the FeII component will be difficult for many of the lower S/N sources (see Marziani et al. 2003). For this reason, the subset of sources in figure 3.6 with a median S/N greater than or equal to 20 per resolution element is also shown (blue). The lower panel of figure 3.6 presents the higher S/N sources, colour-coded as a function of Eddington ratio. The expected trend of increasing Eddington ratio towards smaller FWHM H^{BC}_{β} and larger R_{FeII} is observed for this sample of high-S/N sources. Typically, sources with both high R_{FeII} and high FWHM H^{BC}_{β} are not observed. If these sources exist, they may be difficult to detect since strong FeII emission might conceal a faint Hβ broad component. The potential

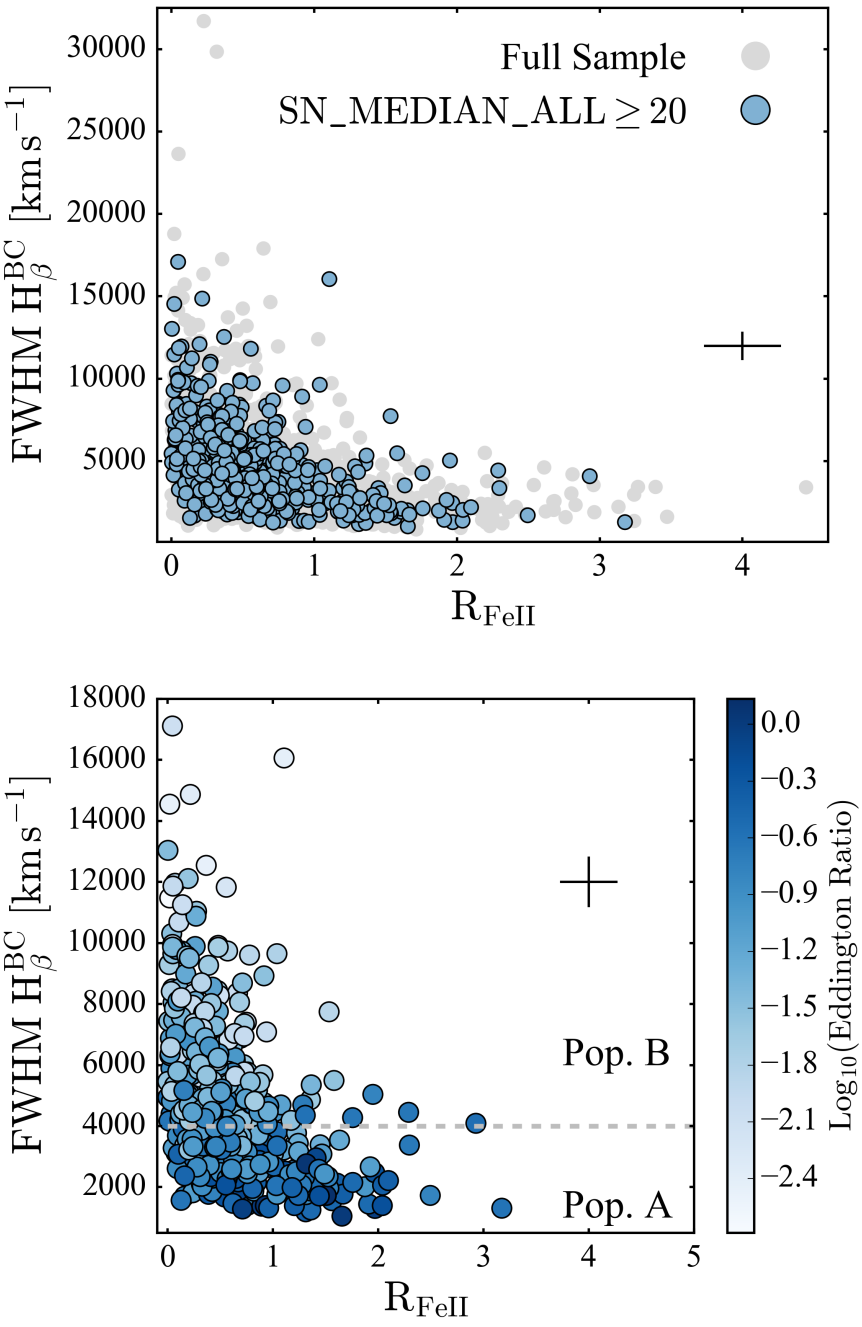


Figure 3.6: FWHM of the broad component of Hβ versus R_{FeII}. The upper panel displays the sample described in section 3.2.2 (grey), and sources with a median spectral S/N ≥ 20 (blue). The lower panel presents the same subsample of high S/N sources, colour coded to indicate the trend in Eddington ratio across the distribution. The grey dashed line marks the division between population A (FWHM $H_\beta^{BC} \leq 4000\,km\,s^{-1}$) and population B (FWHM $H_\beta^{BC} \geq 4000\,km\,s^{-1}$) sources. The size of the typical uncertainties, multiplied by a factor of five, in both variables for the high-S/N subsample is also shown.

bias in the 4DE1 plane source distribution due to model limitations and spectral S/N is discussed in section 3.4.1.

The grey dashed line in the lower panel of figure 3.6 indicates the division between population A (FWHM $H_\beta^{BC} \leq 4000\,km\,s^{-1}$) and population B (FWHM $H_\beta^{BC} \geq 4000\,km\,s^{-1}$) sources in the 4DE1 context (see Sulentic et al. 2011). Population A sources often possess Lorentzian broad line profiles, and it has been suggested that Gaussian fits to population A broad lines will result in an underestimation of the black hole mass (see Sulentic et al. 2014).

3.2.3 Relationship Between AGN X-ray and UV Emission

Quasars exhibit a non-linear relationship between their X-ray and UV emission, usually represented by the α_{OX} parameter

$$\alpha_{OX} = \frac{Log(L_{2\,keV}/L_{2500\,Å})}{Log(\nu_{2\,keV}/\nu_{2500\,Å})}$$

where $L_{2\,keV}$, $L_{2500\,Å}$, $\nu_{2\,keV}$, and $\nu_{2500\,Å}$ are the monochromatic luminosities and frequencies at 2 keV and 2500 Å, respectively (Vignali et al. 2003b; Strateva et al. 2005; Steffen et al. 2006; Just et al. 2007; Kelly et al. 2008; Green et al. 2009; Young et al. 2009; Lusso et al. 2010). The α_{OX} parameter is considered to be a proxy for the relative contribution of the UV accretion disk emission and the X-ray emission from the surrounding corona to the total luminosity. In order to study this relationship, a sample of sources with measurements of the 2keV, 2500Å, and 5100Å luminosities was selected. For lower redshift sources without spectral coverage of 2500Å, equation 3.1 was used to derive 2500Å luminosity estimates from 5100Å luminosity measurements. Extended sources were removed in order to prevent additional scatter in the relationship due to the contribution of the host galaxy. This was done by requiring that the SDSS g band "stellarity"[1] (defined as S(g) = cModelMag_g-psfMag_g) lies between ±0.1. This sample does not contain X-ray sources with more than one potential AllWISE counterpart and therefore avoids cases where the X-ray detection includes emission from more than one object. This selection process resulted in a sample of 4777 sources. Figure 3.7 shows the α_{OX} parameter versus $L_{2500\,Å}$ for the full sample of 4777 sources (upper panel), and for the subsample of sources with direct measurements of L_{UV}. The $\alpha_{OX} - L_{2500Å}$ relation was fit using LINMIX. The best fit relation for the sample with both direct measurements of $L_{2500Å}$ and estimates of $L_{2500Å}$ derived from $L_{5100Å}$ from equation 3.1 is

$$\alpha_{OX} = 2.39 \pm 0.16 - (0.124 \pm 0.005)Log(L_{2500Å})$$

with a regression intrinsic scatter of 0.0034. This slope is consistent with previous results from the literature (e.g. Kelly et al. 2008). The best fit relation for the subsample with

[1] For a description of how cModelMag_g and psfMag_g are measured see https://www.sdss.org/dr12/algorithms/magnitudes/

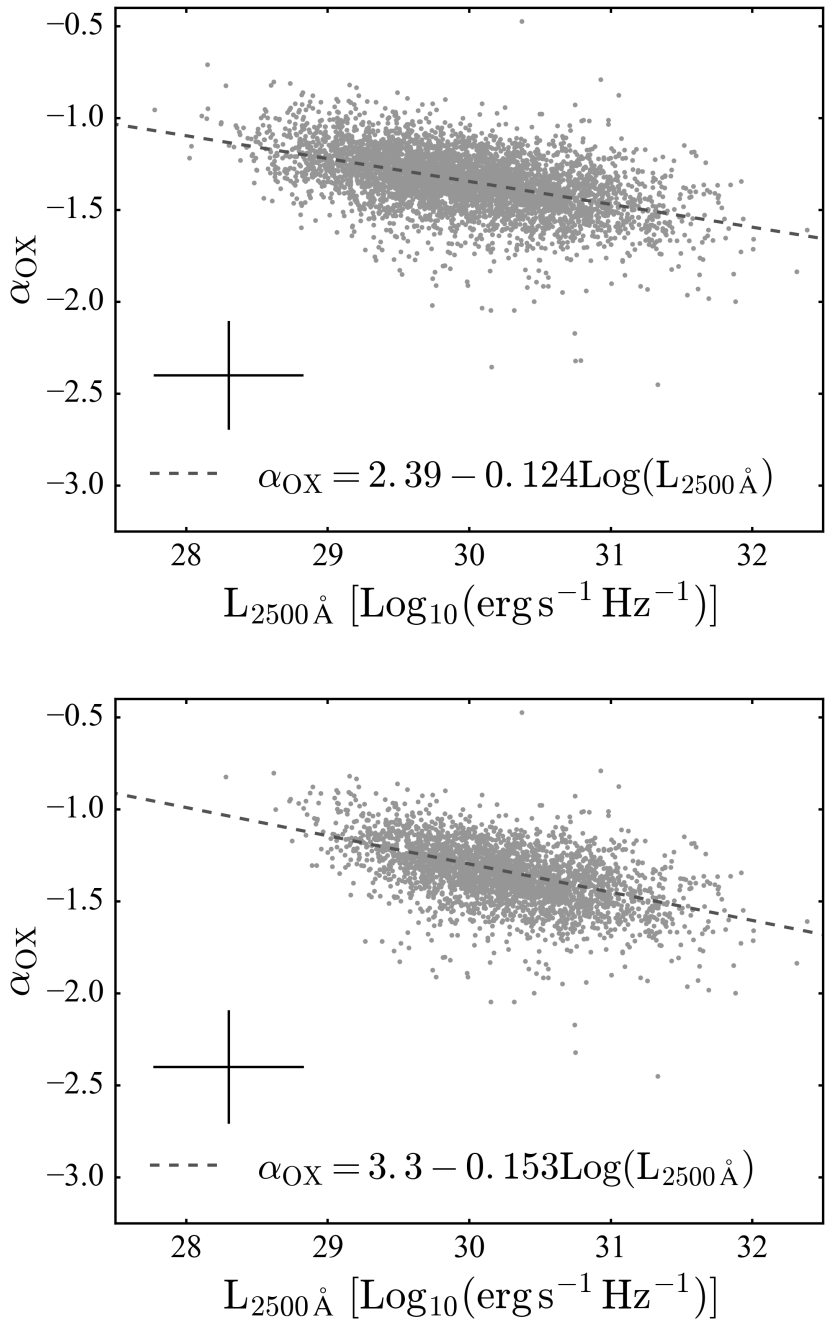


Figure 3.7: α_{OX} versus the 2500Å luminosity for the full sample of sources with direct measurements of $L_{2500Å}$ and estimates of $L_{2500Å}$ derived from $L_{5100Å}$ using equation 3.1 (upper panel), and the sample of sources with direct measurements of $L_{2500Å}$ (lower panel). The dashed line is the best linear fit to the distribution. The size of the typical uncertainties in both variables is also shown.

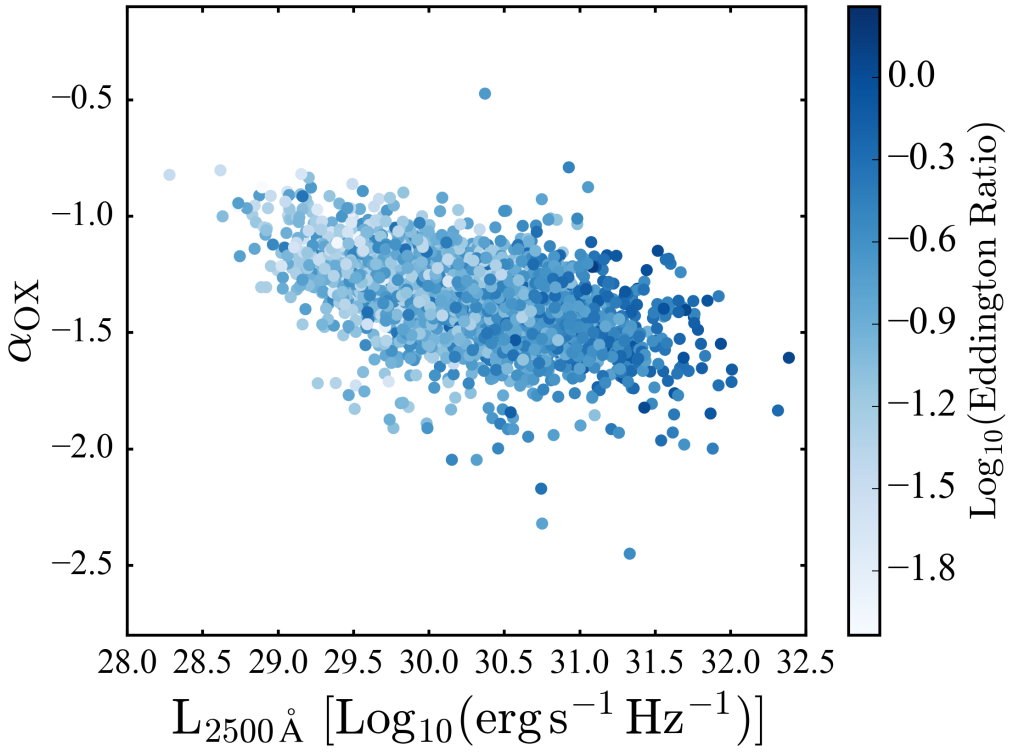


Figure 3.8: α_{OX} versus the 2500Å luminosity for the sample of sources with direct measurements of $L_{2500Å}$ and measurements of the Eddington ratio from the MgII derived BH mass estimates.

direct measurements of $L_{2500Å}$ is

$$\alpha_{OX} = 3.3 \pm 0.2 - (0.153 \pm 0.007)\mathrm{Log}(L_{2500Å})$$

with a regression intrinsic scatter of 0.0028. It has been suggested that the dispersion in the α_{OX} - $L_{2500Å}$ relationship may be in part due to the presence of sources with different Eddington ratios (Chiaraluce et al. 2018). Indeed, several studies have found that α_{OX} is anti-correlated with the Eddington ratio (Grupe et al. 2010; Lusso et al. 2010; Chiaraluce et al. 2018), with evidence for an offset between BLAGN and NLAGN in this parameter space (Grupe et al. 2010). Figure 3.8 shows the trend in the Eddington ratio across the α_{OX} - $L_{2500Å}$ plane for the sample presented in this work, which also shows an increase in the Eddington ratio with decreasing α_{OX}. However, as discussed in Chiaraluce et al. (2018), this trend may be a result of the Eddington ratio's dependence on $L_{2500Å}$.

3.3 Quasar Mass-Luminosity Plane

3.3.1 Quasar Mass Downsizing

Quasars are known to exhibit "downsizing", where the most luminous quasars at high redshift have higher luminosities than the most luminous quasars at low redshift (e.g. Schmidt 1968). Quasar masses have also been found to show this trend (e.g. Steinhardt & Elvis 2010). This effect is illustrated in the upper panel of figure 3.9, where BH mass is observed to increase as a function of redshift. The lower panel of figure 3.9 shows the distribution of BH masses for a series of redshift bins with width $\delta \log_{10} z = 0.25$. These figures clearly show that the high BH mass measurements that are present in the higher redshift bins are not present in the lower redshift bins. This suggests that a number of high mass quasars reach undetectably low luminosities over time, possibly due to a decrease in, or complete cessation of, the mass accretion rate.

As discussed by Steinhardt & Elvis (2010), the Malmquist bias introduced by the use of flux-limited surveys will affect the observed distribution of low-mass sources. The Malmquist bias will result in an increase in the average observed luminosity with redshift since lower-luminosity sources will gradually drop below the minimum detectable flux as redshift increases. Since more massive BHs tend to have higher nuclear luminosities, the lowest observable quasar mass increases with distance. Therefore, the samples presented in this work may not be used to study the evolution of the lower end of the mass distribution with redshift. However, it should be possible to detect high mass, high luminosity sources at both low and high redshifts, and therefore their absence at low redshifts cannot be attributed to selection effects.

3.3.2 A Sub-Eddington Limit for Quasar Accretion?

The quasar mass-luminosity diagram for the sample presented in this work is shown in the upper panel of figure 3.10. A physically motivated boundary on quasar luminosity as a function of BH mass is set by the Eddington limit. In their figure 6, Kollmeier et al. (2006) have shown that sources in their sample lie between $0.1 L_{Edd}$ and L_{Edd}, and typically do not significantly exceed the Eddington limit. For $\mathrm{Log}_{10}(M_{BH}) \gtrsim 7.5$, the sample presented in this work appears to be well bounded by the Eddington limit. However, following Steinhardt & Elvis (2010), this parameter space was examined in more detail by dividing the sample into a series of redshift bins with width $\mathrm{Log}_{10}(z) = 0.25$. The resulting distributions of BH mass and bolometric luminosity for each redshift bin are shown in the lower panel of figure 3.10.

From this figure, it is clear that the sources presented in this work do not reach the Eddington limit until $z \simeq 0.35$. Above this redshift, the sample is well bounded by the Eddington limit, except for the highest mass objects in each bin, which tend to emit below their Eddington luminosity. Steinhardt & Elvis (2010) have also shown that for each of their redshift bins, the higher mass quasars do not reach the Eddington limit. They refer to this as the "sub-Eddington Boundary". It appears that at each redshift, the highest

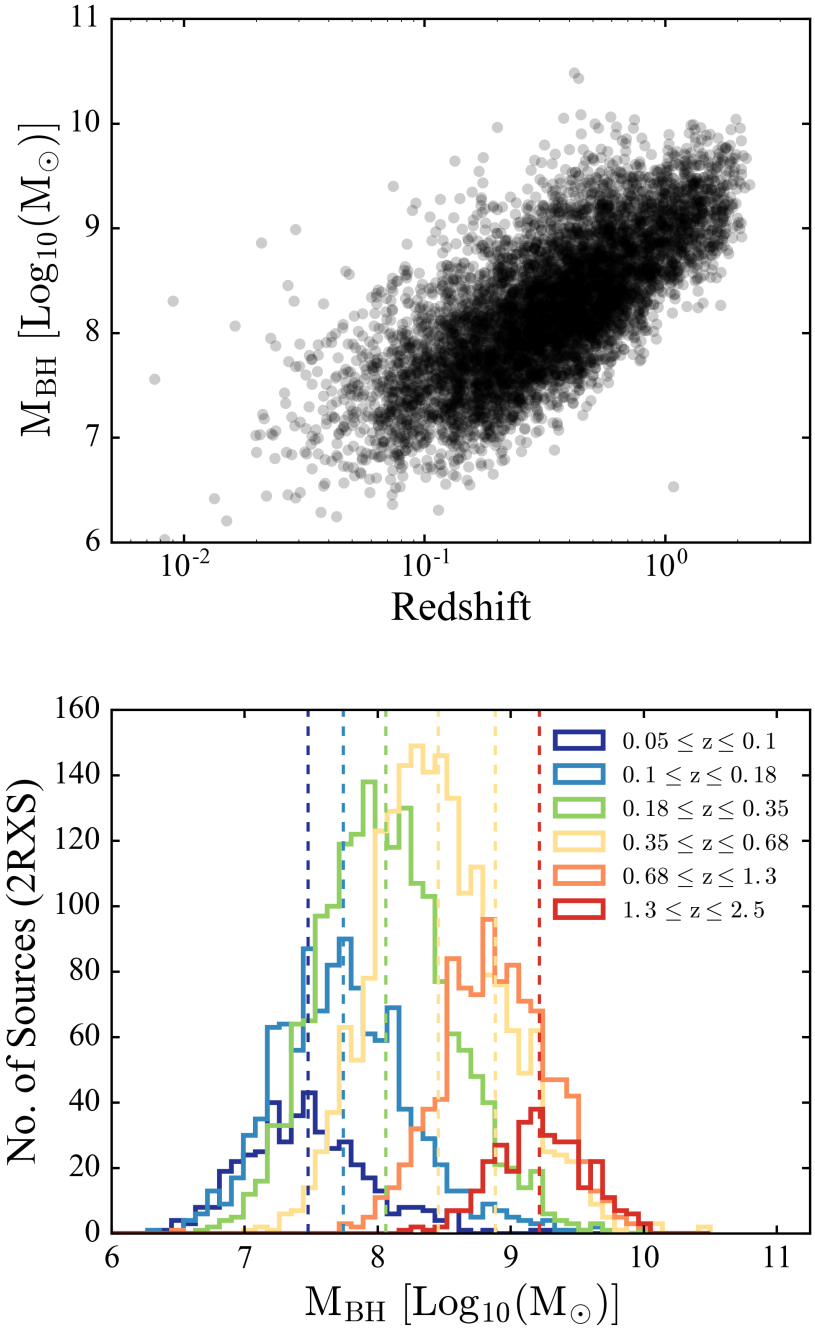


Figure 3.9: Upper panel: Black hole mass versus redshift. Lower panel: Distribution of black hole masses for a series of redshift bins with width $\delta \log_{10} z = 0.25$. The dashed lines show the mean mass for each bin.

mass AGN fail to reach their Eddington luminosity. These high-mass objects then seem to disappear from the sample at lower redshifts as discussed above. This observation is consistent with a gradual decrease in accretion rate at a certain critical mass before the AGN turns off completely. This critical upper limit on mass also appears to decrease towards lower redshifts.

3.4 Interpreting the Data and Limitations

In this section, the reliability and limitations of the sample will be discussed.

3.4.1 Measuring the FeII Emission

Distinguishing the FeII component from the continuum emission can be challenging when using low S/N spectra. In addition, for a given S/N, it may also be more difficult to detect FeII emission if the intrinsic broadening of the FeII lines is large, since broader, blended FeII emission lines are more likely to be fit by the model as continuum emission (see Marziani et al. 2003). Using simulated AGN spectra, Marziani et al. (2003) estimate the minimum detectable optical FeII emission as a function of Hβ width for different bins of S/N.

A poor fit to the FeII emission may affect the accuracy of the BH mass estimates, since FeII emission can influence measurements of both the broad line width (discussed in the following section) and the continuum luminosity. FeII emission may also conceal a broad Hβ component thus biasing a source's position in the 4DE1 plane (figure 3.6). These potential issues are tested in the following three sections.

Accuracy of the Broad Emission Line FWHM Measurements for Sources with FeII Continuum Emission

A poor fit to the FeII emission may affect the measurement of the broad emission line width. To quantify this effect, the Hβ fitting script (using four Gaussians to fit Hβ) was run with and without the FeII template on a sample of ~ 400 randomly selected sources. The fit without the FeII template represents the most extreme case where the FeII emission is completely ignored by the model. Therefore, the change in line widths measured by these two models should be the upper limit on what can be expected for cases where the FeII fit is inadequate. Figure 3.11 shows that the line width dispersion induced by ignoring the presence of FeII emission is $\simeq 640\,\mathrm{km\,s^{-1}}$. However, one must note that the severity of this effect is likely to be significantly lower for population B AGN which often show little to no FeII emission.

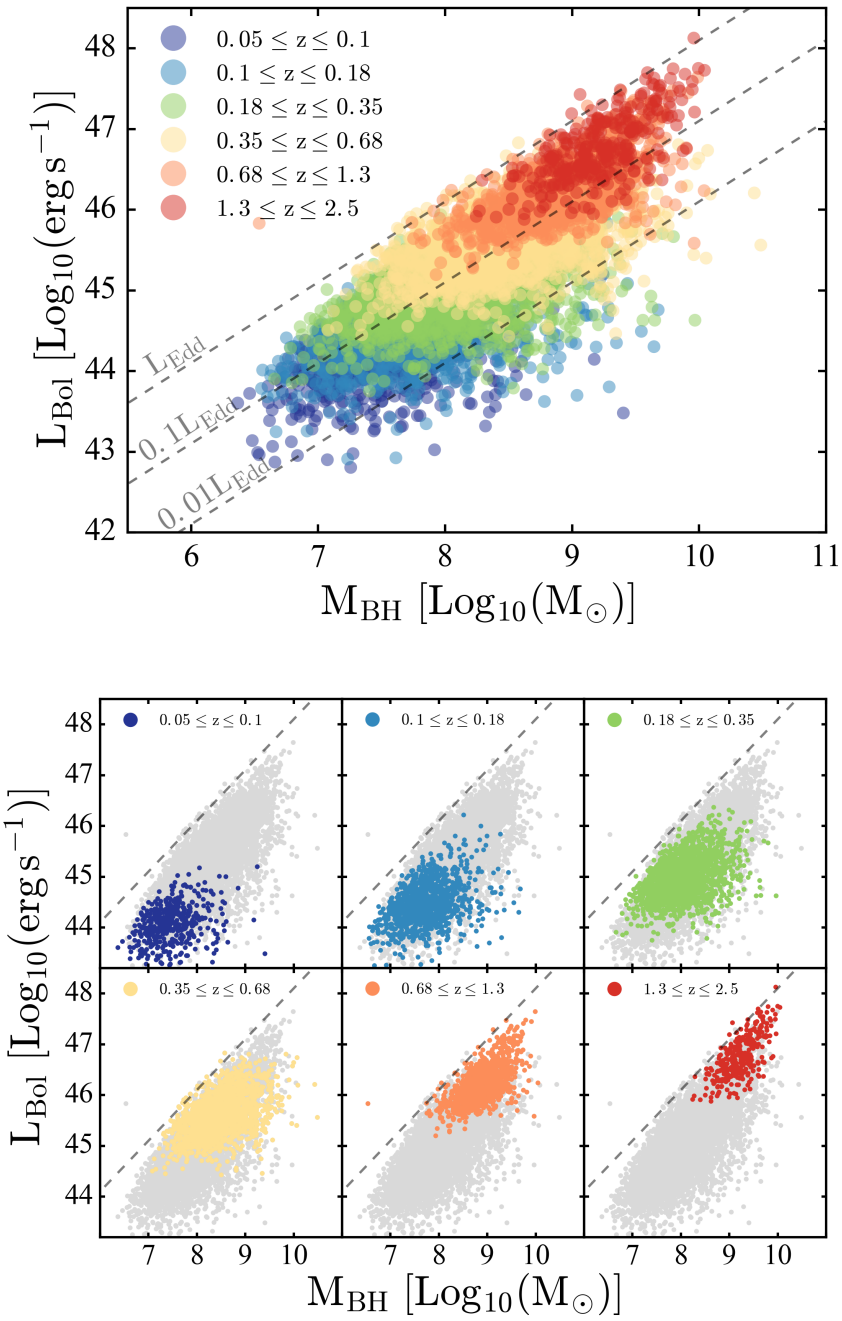


Figure 3.10: Upper panel: BH mass versus bolometric luminosity for the sample. Different colours represent different redshift bins separated by $\delta \log_{10} z = 0.25$. Diagonal lines indicate various fractions of the Eddington luminosity. Lower panel: BH mass versus bolometric luminosity, subdivided to show each redshift bin (with width $\delta \log_{10} z = 0.25$) in a separate panel.

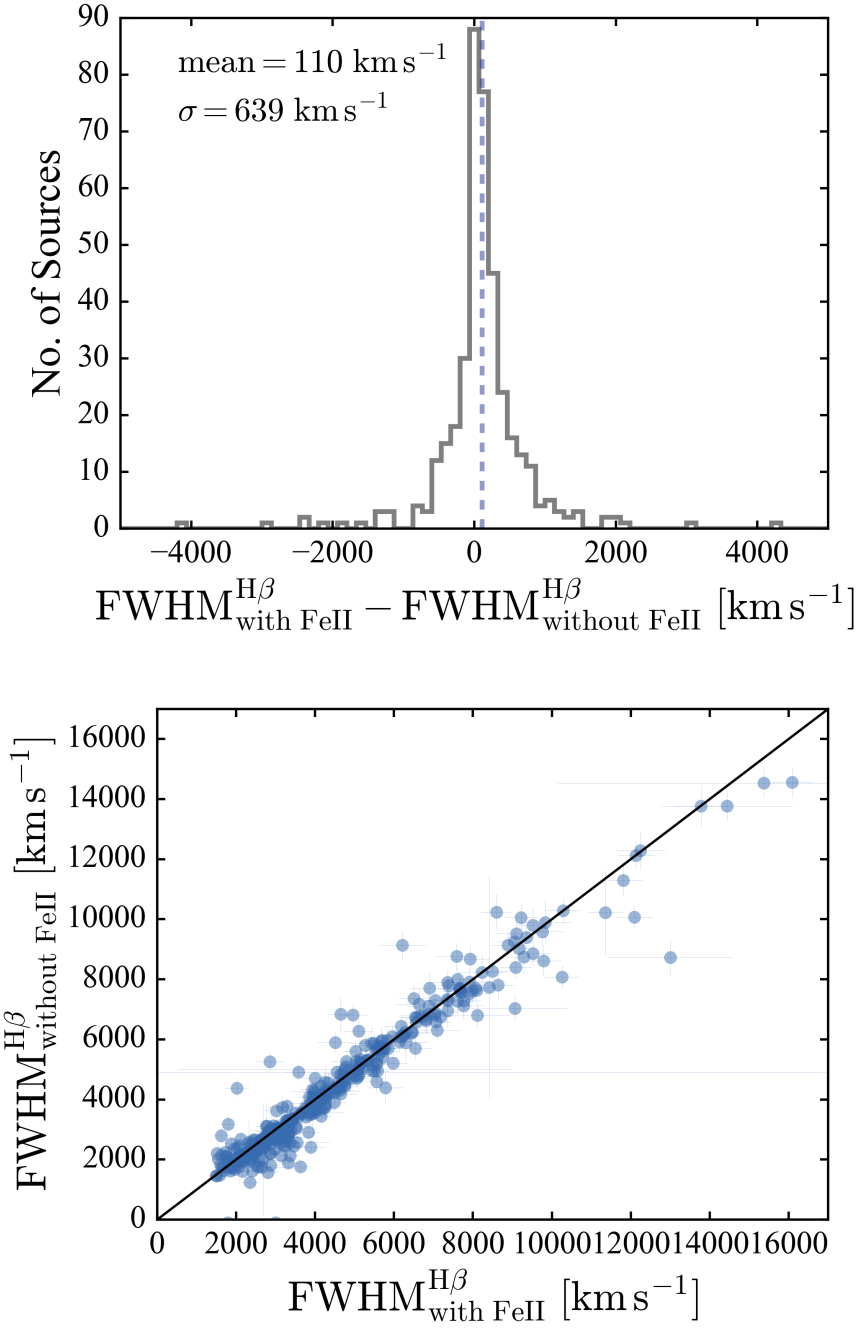


Figure 3.11: Upper panel: Difference between the Hβ FWHM measurements derived using a model with and without an FeII template. The mean value of the distribution (blue dashed line) and the standard deviation (σ) are both shown. Lower panel: Comparison of the Hβ FWHM measurements obtained with and without applying an FeII template. The solid black line is the unity line.

Model Limitations in Detecting Sources in the 4DE1 Plane

Sources with both large R_{FeII} and large FWHM H_β^{BC} are typically not observed, however, this absence may be due to model limitations; at high R_{FeII} the broad Hβ component may be concealed beneath the FeII emission, and therefore may not be detected. The experiment outlined in this section was carried out in order to determine whether the spectral fitting code used in this work would return accurate measurements for sources with high R_{FeII} and FWHM H_β^{BC} values.

A parameter space defined by $0.1 \leq R_{FeII} < 5$ and $1000\,km\,s^{-1} \leq$ FWHM $H_\beta^{BC} < 15000\,km\,s^{-1}$ was divided into a 12×12 grid. 10 S/N bins between 5 and 50 (a representative range for the samples presented in this work) were selected for each point on the grid, and 10 spectra were simulated for each R_{FeII} − FWHM H_β^{BC} − S/N combination, resulting in 14400 simulated spectra. For the parameters that were fixed in this experiment, the interquartile mean of the best-fit values for the type 1 AGN in this sample were used. The Hβ line profile was modelled with one narrow and one broad Gaussian. The wavelength range was set to 4420 − 5500Å (as in section 2.2.4) and the logarithmic wavelength spacing[2] was set to be equal to that of SDSS spectra;

$$Log_{10}\lambda_{i+1} - Log_{10}\lambda_i = 0.0001$$

These spectra were fit using a version of the Hβ fitting script which used one narrow and one broad Gaussian component to fit Hβ. The minimum S/N required for the fitting script to return the correct R_{FeII} and FWHM H_β^{BC} combinations was then determined. In order to consider an R_{FeII} and FWHM H_β^{BC} combination detectable at a given S/N, at least 7/10 spectra were required to have best fit R_{FeII} and FWHM H_β^{BC} values that agreed with the input values.

Figure 3.12 shows the detectable R_{FeII} and FWHM H_β^{BC} combinations for each point on the grid, along with the minimum S/N required to detect that combination. It is clear from figure 3.12 that at the S/N levels available in this sample, a large region of the 4DE1 parameter space would not be detected.

Bias in the R_{FeII} and FWHM H_β^{BC} Distribution due to Model Limitations

Figure 3.13 shows the comparison between the simulated and measured R_{FeII} and FWHM H_β^{BC} values for the highest and lowest S/N bins used in section 3.4.1. The upper panel of figure 3.13 shows that at low S/N the results are clearly biased against high R_{FeII} and FWHM H_β^{BC} values. At higher S/N (figure 3.13, lower panel), the accuracy of the lower left quadrant measurements is significantly improved. However, even at S/N=45.5, which is approximately the upper end of the S/N distribution of the samples presented in this work (see figure 2.6), the high R_{FeII} - FWHM H_β^{BC} measurements deviate significantly from the corresponding "true" values. This may suggest that the L-shaped distribution of sources in the 4DE1 plane (e.g. figure 3.6) is at least in part due to model limitations.

[2] https://www.sdss.org/dr12/spectro/spectro_basics/

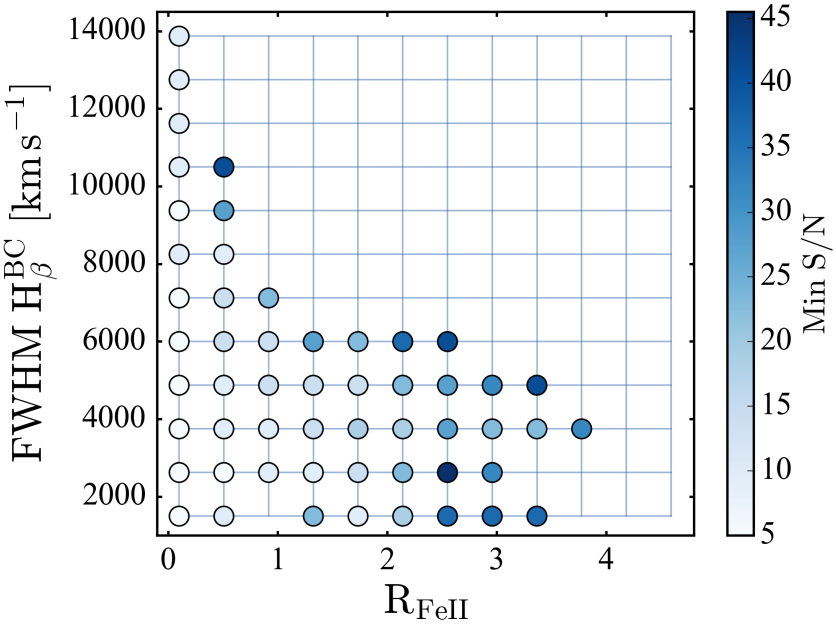


Figure 3.12: Minimum S/N required to detect a range of R_{FeII} and FWHM H_β^{BC} combinations. The blue grid indicates the range of the parameter space covered in the simulation described in section 3.4.1. Points on the grid that do not have a minimum S/N indicator represent R_{FeII} and FWHM H_β^{BC} combinations that are not detectable even at the highest S/N used in this experiment. Sources detected at these R_{FeII} and FWHM H_β^{BC} combinations are likely to be spurious (see figure 3.13).

3.4.2 Reliability of the Single-Epoch Method for Mass Estimation

Assuming that AGN broad emission lines are produced by gas whose motion is dominated by the gravitational potential of the central SMBH, the single-epoch method is expected to produce reliable mass estimates when compared to RM (see Vestergaard & Peterson 2006), with a systematic uncertainty of 0.3-0.4 dex. However, it is not clear how to measure the virial FWHM of lines that deviate from this norm.

The spectrum shown in the upper panel of figure 3.14 is an example of a source which exhibits a double-peaked Hβ line profile, where a clear inflection point is visible between two velocity-shifted broad line components. Double-peaked broad line profiles in AGN are expected to be the result of emission from the accretion disk (Perez et al. 1988; Chen et al. 1989; Eracleous & Halpern 1994; Strateva et al. 2003; Eracleous & Halpern 2003). Zhang et al. (2007) have found that single-epoch BH mass estimates obtained from double-peaked line profiles are significantly larger than BH mass estimates derived from stellar velocity dispersion measurements. Zhang et al. (2007) suggest that this discrepancy is the result of an overestimation of the BLR radius by the single-epoch mass calibrations for these

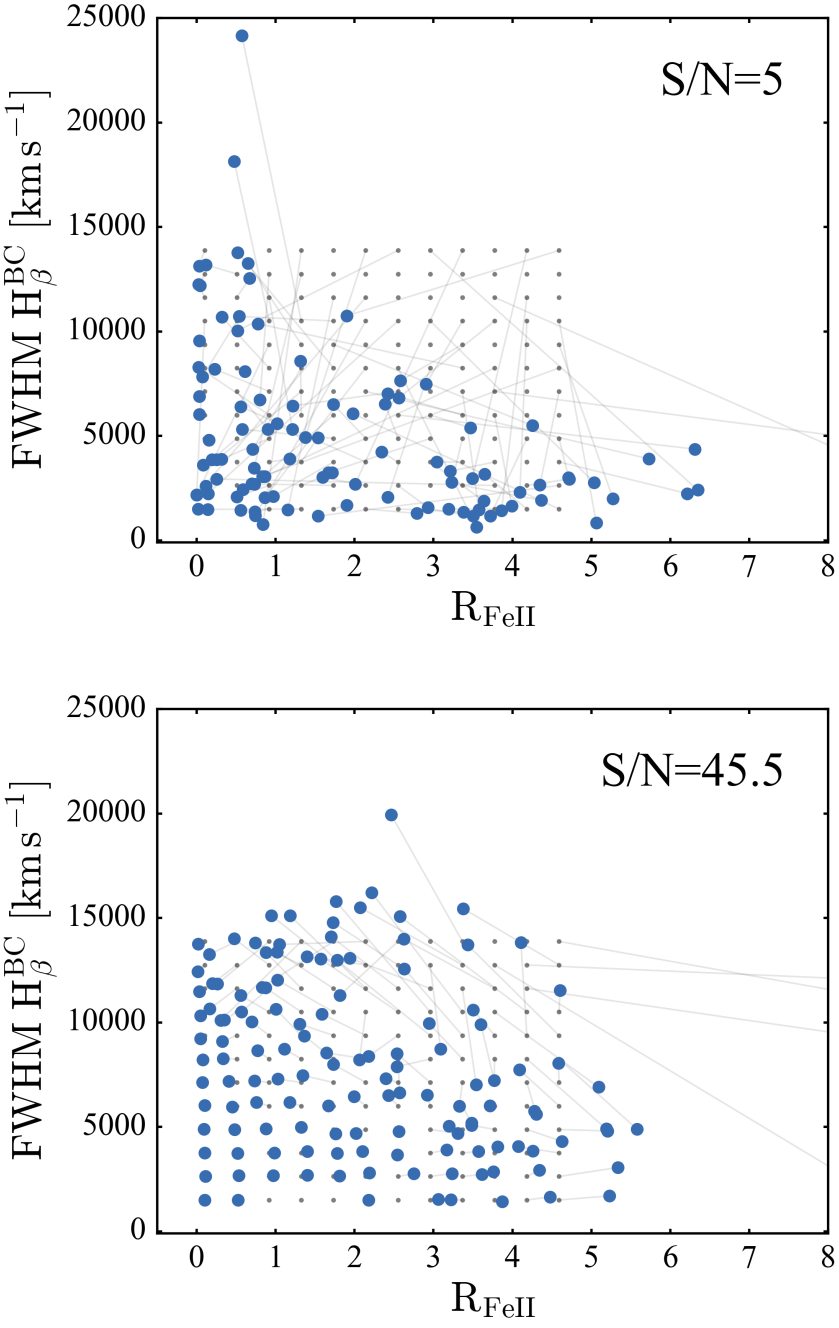


Figure 3.13: Comparison between the measured and simulated R_{FeII} and FWHM H_β^{BC} values for the highest and lowest S/N bins used in section 3.4.1. For clarity, each figure displays only one of the ten sets of spectra for that S/N. The values used to simulate the spectra (grey points) are connected to the corresponding best fit measurements (except for cases where either the broad Hβ or FeII components were not detected). For clarity, the figures do not show a small number of unphysically high R_{FeII} measurements.

objects. Therefore, the BH mass estimates provided in this work for sources which exhibit double-peaked broad emission lines should be treated with caution.

The lower panel of figure 3.14 shows an example of narrow absorption in the UV portion of the spectrum caused by intervening absorbing material along the line of sight to the AGN. Sources identified during the visual inspection as having narrow absorption lines have been flagged in the catalogue. These sources were fit using the model described in section 2.2.5 with the absorption line regions masked. However, in many cases, the absorption features distort the broad MgII line, and therefore the resulting BH mass estimates may not be reliable.

3.5 Conclusions

In this chapter, the catalogue presented in chapter 2 was used to investigate a number of phenomena associated with AGN accretion. The main results of this work can be summarised as follows:

- The virial FWHM of Hβ was found to be larger than that of MgII, suggesting that these lines are emitted from regions at different distances from the BH, confirming the conclusions drawn from previous spectral analyses of AGN (Wang et al. 2009; Shen & Liu 2012) as well as RM mapping (Li et al. 2017).

- Surveying a large number of sources, it was found that the Hβ line shows evidence for three distinct emission regions; the narrow component ($\sim 400\,\mathrm{km\,s^{-1}}$), the classical broad component ($\sim 4000\,\mathrm{km\,s^{-1}}$), as well as a third component with FWHM $> 10000\,\mathrm{km\,s^{-1}}$. This third component, referred to as the "very broad component", is expected to originate from the inner BLR, or perhaps the outer accretion disk.

- It was confirmed that the sources presented in this sample exhibit Eddington limited accretion above $z \simeq 0.35$. However, there is some evidence that the higher mass objects above this redshift fail to reach their Eddington limits. This phenomenon is referred to the sub-Eddington boundary (Steinhardt & Elvis 2010).

- It was found that the observed distribution of sources along the 4DE1 sequence may be influenced by the spectral S/N, and the inability of the spectral decomposition method to detect broad Hβ emission in sources with relatively strong FeII emission.

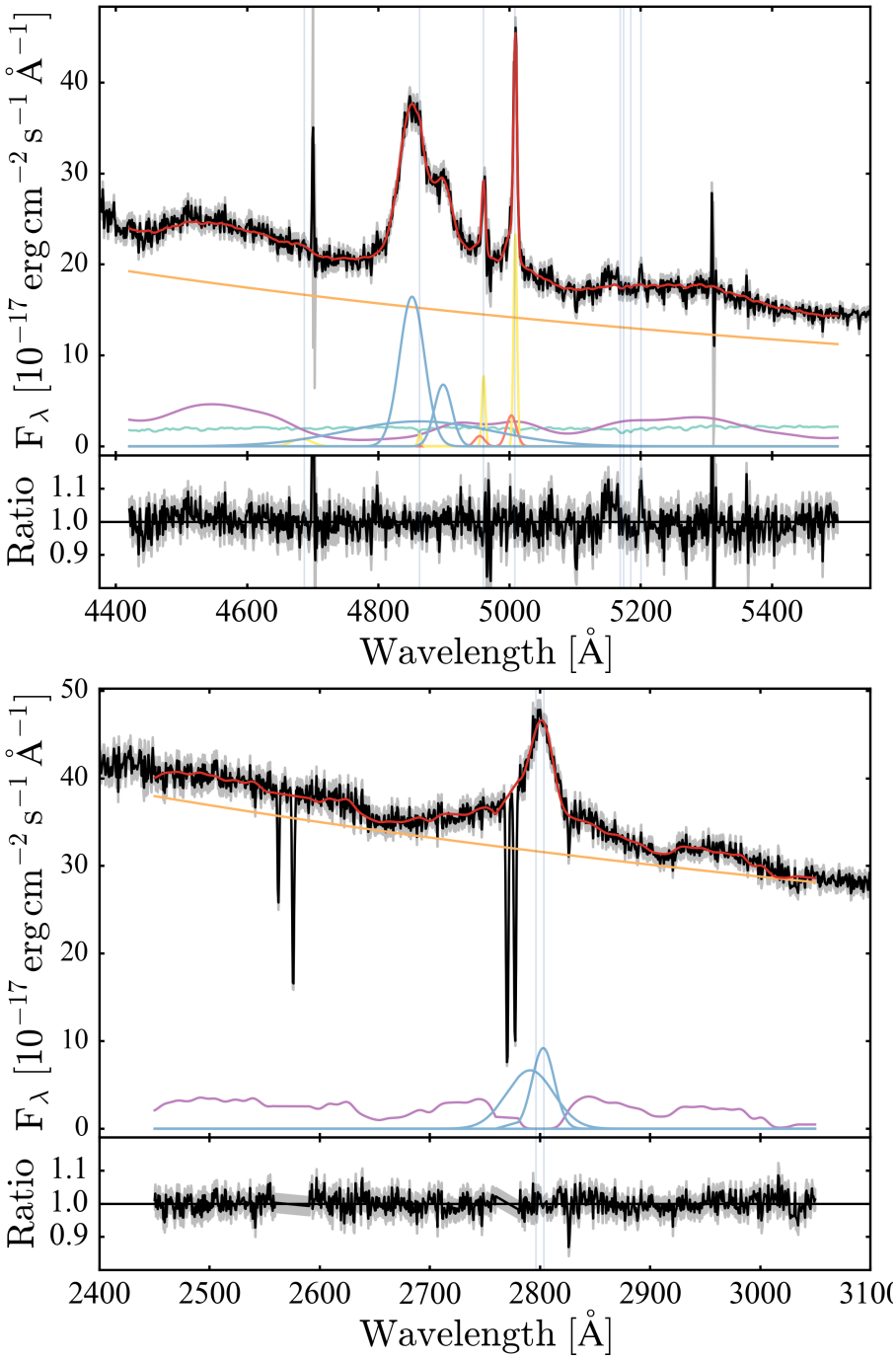


Figure 3.14: Upper panel: Example of a source with a double-peaked Hβ line profile (plate=7512, MJD=56777, fiber=321). Lower panel: Example of a source showing narrow UV absorption features which have been masked when fitting the model (plate=8188, MJD=57348, fiber=946). See section 3.4.2 for details.

Chapter 4

Quasars as Standard Candles

Using the relationship between X-ray and UV emission in quasars, RL15 have shown that a quasar Hubble diagram can be used to place constraints on the cosmological parameters Ω_M and Ω_Λ (see section 1.6.2). This chapter aims to investigate the application of the large survey-quality quasar datasets provided by ROSAT and XMM-Newton to the RL15 quasar cosmology experiment.

4.1 X-Ray Surveys

4.1.1 XMM−XXL North

This work has made use of the *XMM-Newton* XMM-XXL survey (PI M. Pierre; Pierre et al. 2016), which was designed to detect hundreds of galaxy clusters and tens of thousands of AGN. The survey was allotted an exposure time of 6.9 Ms and covers an area of 50 $\deg^2$, which was split into two 25 $\deg^2$ fields (Pierre et al. 2016). XMM-XXL observations reached an intermediate depth, with 10 ks per pointing.

Menzel et al. (2016) presented the optical spectroscopic follow-up of X-ray sources detected in the northern XMM-XXL field. The XMM-XXL north field is an extension of the 11 $\deg^2$ XMM-LSS survey (Pierre et al. 2004; Chiappetti et al. 2013; Clerc et al. 2014). The region with optical spectroscopic follow-up presented by Menzel et al. (2016) covers an area of $\sim 18\,\deg^2$ with a limiting flux of $1.27 \times 10^{-14}\,\mathrm{erg\,cm^{-2}\,s^{-1}}$ for 50% of this area. Menzel et al. (2016) matched the 8,445 point-like X-ray detected sources to SDSS DR8 optical (Aihara et al. 2011) and WISE IR (Wright et al. 2010) counterparts using the likelihood ratio method presented by Georgakakis & Nandra (2011). 3,042 of these sources were then targeted with the BOSS optical spectrograph. After visually inspecting the resulting spectra, 2,578/3,042 sources were found to have reliable redshifts. 1,787 sources were assigned the "BLAGN1" classification based on their optical emission line properties. The XMM-XXL north field is one of the largest samples of X-ray selected AGN in a contiguous and homogenous area with optical spectroscopic follow-up to date.

4.1.2 2RXS

This work also makes use of the sample of ROSAT X-ray sources with SDSS DR12[1] (Alam et al. 2015) optical spectroscopic follow up obtained as part of the SPIDERS programme (Dwelly et al. 2017). The Dwelly et al. (2017) SPIDERS catalogue contained the original *ROSAT* All-Sky Survey (RASS, Voges et al. 1999) X-ray detections. However, since this work was completed after the publication of the 2RXS catalogue (Boller et al. 2016), X-ray properties from the 2RXS catalogue were used instead. The Dwelly et al. (2017) catalogue was produced following a method similar to that discussed in section 2.1, which resulted in a sample of 11,913 sources. This catalogue was then cross-matched to the 2RXS catalogue which resulted in 9,924 sources. The 1,989 sources lost during this match are most likely cases where sources detected in 2RXS were not detected in RASS (and vice versa). The source classification made available by Dwelly et al. (2017) was then used to select objects classified as either "QSO" or "BLAGN", which resulted in a sample of 7,847 sources.

4.2 Obtaining UV Photometric Counterparts

In order to provide the UV photometric coverage which will be used in the SED fitting described in section 4.4, UV photometry was obtained from GAlaxy Evolution EXplorer (GALEX; Martin et al. 2005) observations. GALEX was a NASA Small Explorer Class mission that was launched on 28/04/03 and conducted the first space-based all-sky UV imaging survey. GALEX observed in the near-UV (NUV; 1771Å $\leq$ z $\leq$ 2831Å, $\lambda_{\rm Eff}$ = 2315.7Å) and far-UV (FUV; 1344Å $\leq$ z $\leq$ 1786Å, $\lambda_{\rm Eff}$ = 1538.6Å bands simultaneously (Morrissey et al. 2007), with an effective area of $\sim 50\,{\rm cm}^2$. The GALEX all-sky survey is divided into a number of sub-surveys at different depths; including the All-Sky Imaging Survey (AIS; typical exposure $\sim 100\,{\rm s}$), the Medium Imaging Survey (MIS; typical exposure $\sim 1500\,{\rm s}$), and the Deep Imaging Survey (DIS; typical exposure $\sim 30000\,{\rm s}$) (the limiting magnitudes of which are given in table 4.1). At $m_{\rm AB}$ ~23.5, the GALEX surveys become background limited (Martin et al. 2005).

The GALEX PSF ($\sim 4.5 - 6\,{\rm arcsec}$, see Bianchi et al. 2007) is large relative to that of SDSS[2] ($\sim 1.4\,{\rm arcsec}$), and therefore, to ensure that one GALEX detection always corresponds to one SDSS detection (rather than multiple sources blended together), SDSS sources in both catalogues with a neighbouring SDSS source within 8 arcsec were removed. This was done by cross matching the SDSS optical counterpart positions to the SDSS DR9 (Ahn et al. 2012) photometric catalogue which resulted in 5256 2RXS and 1366 XMM-XXL "isolated" sources that will be matched to GALEX counterparts. It is expected that each of these sources will have been resolved successfully by GALEX with few blending issues. The GALEX data can be accessed via the GalexView[3] tool.

Using the SDSS optical coordinates, all GALEX catalogues from GR6/GR7 were searched

[1] At the time of this work, SDSS DR12 was the most recent available data release.
[2] http://www.sdss.org/dr12/imaging/other_info/
[3] http://galex.stsci.edu/GalexView/

for counterparts within 4 arcsec of the SDSS position (see Budavári et al. 2009). For sources with multiple possible counterparts, the source closest to the SDSS position was selected. GALEX counterparts which had an FUV magnitude measurement, but did not have a corresponding NUV detection were removed, since these sources are expected to be spurious[4]. In addition, GALEX counterparts which had window reflection and dichroic reflection artefact flags were also removed since these sources may have problematic fluxes[5]. This resulted in 5233 2RXS and 868 XMM-XXL sources with GALEX counterparts.

Trammell et al. (2007) have pointed out that the GALEX source extraction method assumes a photometric uncertainty based on the Poisson uncertainty, and ignores any contribution from systematic errors (for example, variations in the detector background level). Using multi-epoch GALEX observations of stars, Trammell et al. (2007) have outlined a method to empirically determine the true GALEX photometric uncertainties. Following their method, the empirical GALEX photometric errors were estimated as follows.

$$\mathrm{Log}(t_{\mathrm{eff}} \leq 2.6) = \begin{cases} \mathrm{Log}(\delta\mathrm{FUV}) = 0.1336\,\mathrm{m} - 3.0578 \\ \mathrm{Log}(\delta\mathrm{NUV}) = 0.20525\,\mathrm{m} - 4.7302 \end{cases}$$

$$\mathrm{Log}(t_{\mathrm{eff}} > 2.6) = \begin{cases} \mathrm{Log}(\delta\mathrm{FUV}) = 0.24413\,\mathrm{m} - 5.8804 \\ \mathrm{Log}(\delta\mathrm{NUV}) = 0.23466\,\mathrm{m} - 5.7196 \end{cases}$$

where t_{eff} is the NUV effective exposure time and m is the NUV/FUV magnitude. GALEX photometric uncertainties measured in this way were used during the SED fitting discussed in section 4.4.

4.3 X-Ray Flux Estimation

4.3.1 XMM−XXL

A detailed X-ray spectral analysis has been carried out by Liu et al. (2016) on the 2512 XMM−XXL sources which received reliable spectroscopic observations from BOSS (see Menzel et al. 2016). Liu et al. (2016) use the Bayesian X-ray Analysis (BXA) software (Buchner et al. 2014) and provide the best fit model parameters in a catalogue accompanying their paper. Using these best fit model parameters, we use XSPEC (Arnaud 1996) to estimate the rest−frame 2 keV flux for each of the 1786 quasars in our sample. Note that five XMM-XXL sources have X-ray luminosities below that which would be expected from AGN emission ($L_{2-10\,\mathrm{keV}} < 10^{40}\,\mathrm{erg\,s^{-1}}$), and therefore these sources were excluded from this analysis presented below, leaving a sample of 1781 sources.

[4] https://asd.gsfc.nasa.gov/archive/galex/FAQ/#DI_1
[5] http://galex.stsci.edu/GR6/?page=ddfaq#6

4.3.2 ROSAT

The 2RXS fluxes used in this experiment were derived using the method outlined in section 2.4.1. This method (see Kraft et al. 1991; Laird et al. 2009; Georgakakis & Nandra 2011) aims to correct for the effects of the Eddington bias, which is expected to affect the 2RXS measurements due to the high flux limit of ROSAT. Following the method outlined in section 2.4.1, reliable bayesian 2 keV flux densities were computed for 2413/5256 sources in the 2RXS sample. The resulting flux correction is illustrated in figure 4.1 (see section 2.4.1 for further details).

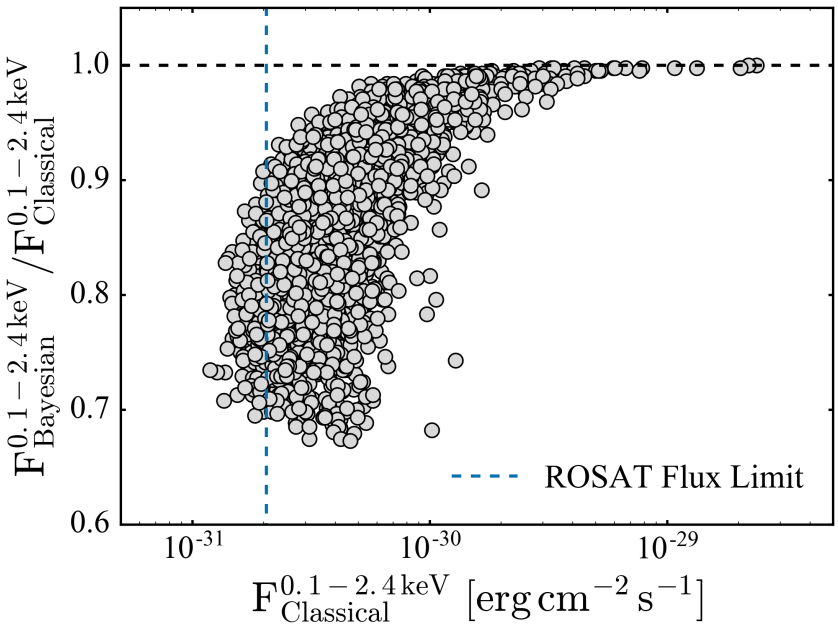


Figure 4.1: Comparison between the classical and Bayesian methods for estimating 2RXS 2 keV flux densities. The horizontal dashed line indicates the limiting ROSAT flux (see section 4.3.2).

4.4 SED Fitting and Extinction Correction

A significant challenge faced when attempting to measure the intrinsic UV luminosity of AGN is the effect of extinction by dust on the line of sight to the emitting region. The wavelength dependent extinction is defined as the magnitude difference caused by the presence of dust;

$$A_\lambda \equiv m_\lambda - m_{\lambda,0}$$

where m_λ is the observed magnitude at wavelength λ and $m_{\lambda,0}$ is the magnitude that would be observed in the absence of dust extinction. A_λ can also be expressed as

$$A_\lambda = k(\lambda)E_{B-V}$$

where $k(\lambda)$ is the extinction law, and E_{B-V} is the colour excess (or reddening) which is defined as the difference between the extinction in the B ($\lambda_{\text{eff}} = 4450$Å) and V ($\lambda_{\text{eff}} = 5510$Å) photometric bands (see Bradt 2004, pg. 308-311);

$$E_{B-V} \equiv A_B - A_V$$

The following sections discuss the SED fitting procedure, which aims to estimate the luminosity at 2500Å corrected for the effect of intrinsic (host galaxy) extinction.

4.4.1 Photometric Quality Cuts

Before carrying out the SED analysis discussed in section 4.4.2, it was necessary to define a clean sample of sources with reliable photometric measurements. The first step taken towards this goal was to select photometric measurements that can be expected to be sensitive to relatively low levels of dust extinction (reddening).

Figure 4.2 shows the per cent loss in flux at 2500Å as a function of the colour excess, for the three extinction laws used in the LePHARE fit discussed in section 4.4.2. It is clear from figure 4.2 that the loss in flux at 2500Å is very similar for each law. The rapid drop in flux with increasing colour excess illustrates the importance of having an accurate estimate of the extinction when measuring $F_{2500\text{Å}}$. For the SED fitting procedure described in section 4.4.2, only photometry which can discern a minimum flux loss (due to extinction) of 10 per cent will be used.

The upper panel of figure 4.3 shows the magnitude ratio of an SED model ("qso-1.25_84.sed") with no extinction applied, and with an extinction corresponding to a flux loss of 10 per cent, as a function of wavelength. Photometric uncertainties smaller than this magnitude ratio should be able to detect a flux loss due to extinction of at least 10 per cent. The lower panel of figure 4.3 shows the result of adding this level of extinction to a LePHARE AGN template ("qso-1.25_84.sed"). As can be seen from this figure, the difference between the two models increases towards UV wavelengths, while they become almost indistinguishable towards longer wavelengths. Bluer bands will therefore be much

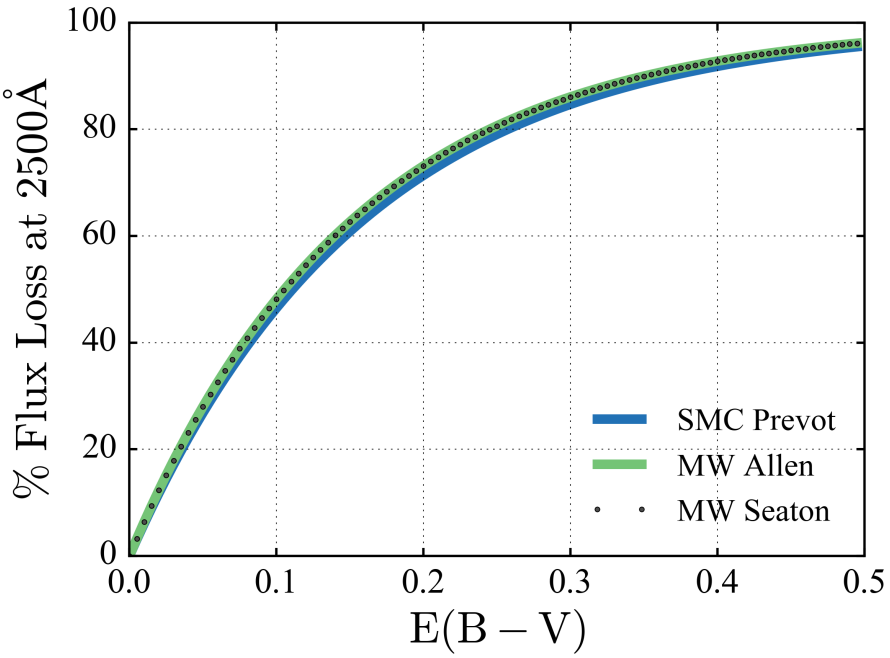


Figure 4.2: The per cent flux reduction relative to intrinsic flux at 2500Å is shown as a function of colour excess for the three extinction laws used in the LePHARE fit (see section 4.4.2). As can also be seen from figure 4.5, the three laws are very similar at 2500Å.

more likely to distinguish between these two cases than redder bands. Therefore it will be required that of the bands that lie blueward of the rest frame 2500Å, at least one of them must have photometric errors smaller than the magnitude difference shown by the upper panel of figure 4.3.

The SED templates used in this work cover the 600Å to 25000Å region, and therefore, as redshift increases, it was necessary to gradually remove bands from the fit which do not overlap with the template, starting with the FUV band, and progressing redward. It was also necessary to remove bands which lie above the magnitude limit for that band, or which have magnitude errors above 0.5 mag. Any GALEX magnitudes from the calibration imaging (CAI) dataset were also rejected, since these detections are of calibration white dwarf standards[6]. Magnitudes above the magnitude limit (see table 4.1) for that band were rejected. Finally, sources with less than two accepted bands were also rejected. The distribution of the number of accepted photometric bands is illustrated in figure 4.4.

[6]http://www.galex.caltech.edu/researcher/techdoc-ch2.html

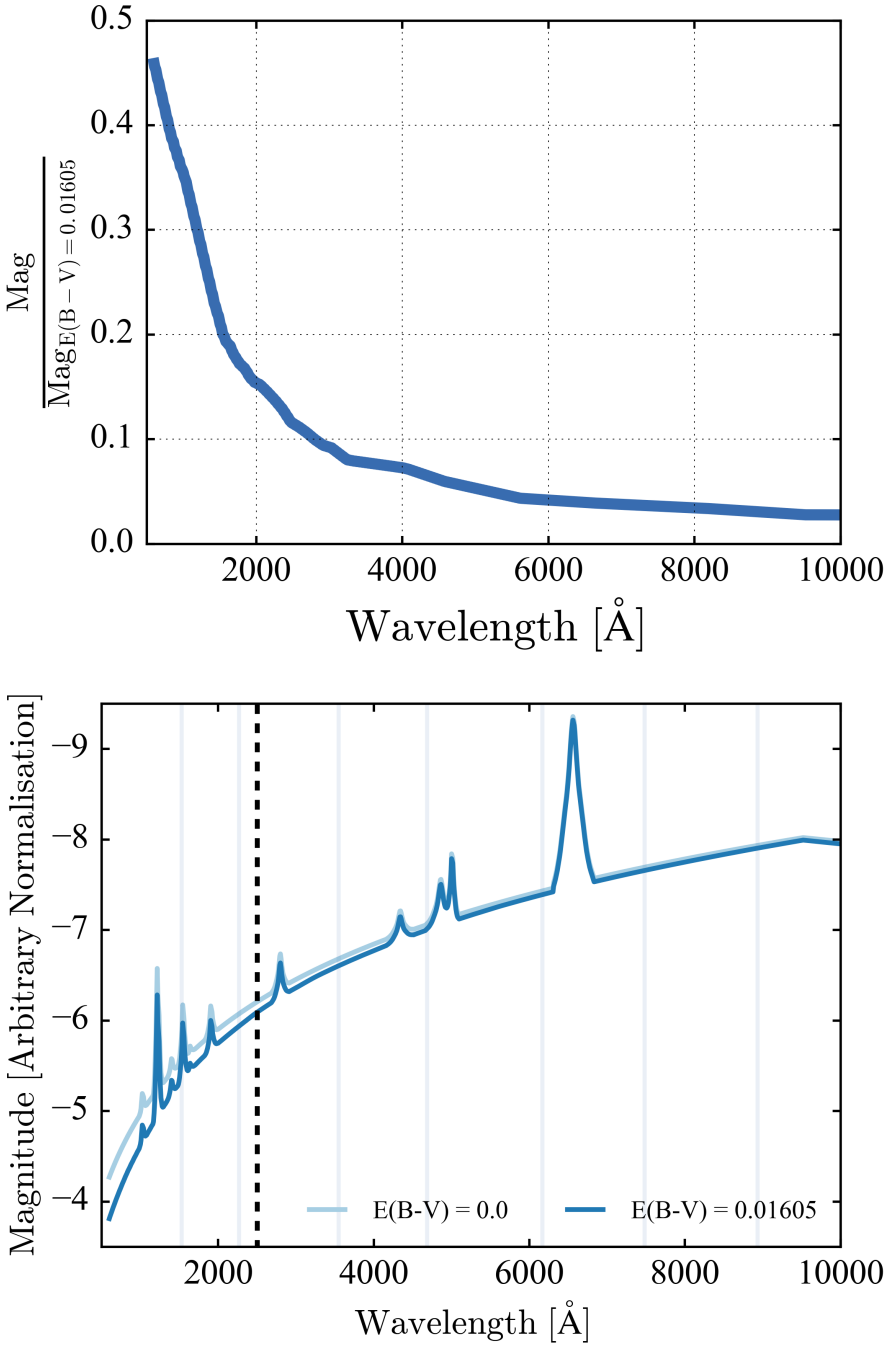


Figure 4.3: Upper panel: The magnitude ratio for two models; one with no extinction applied, and the other with an extinction corresponding to a flux loss of 10 per cent at 2500Å, as a function of wavelength. Lower panel: Comparison between a LePHARE template ("qso-1.25 84.sed") with no extinction applied (light blue) and the same template with a 10 per cent flux loss at 2500Å due to extinction (dark blue). The solid horizontal lines show the location of the SDSS and GALEX photometric bands. The dashed horizontal line marks the location of 2500Å.

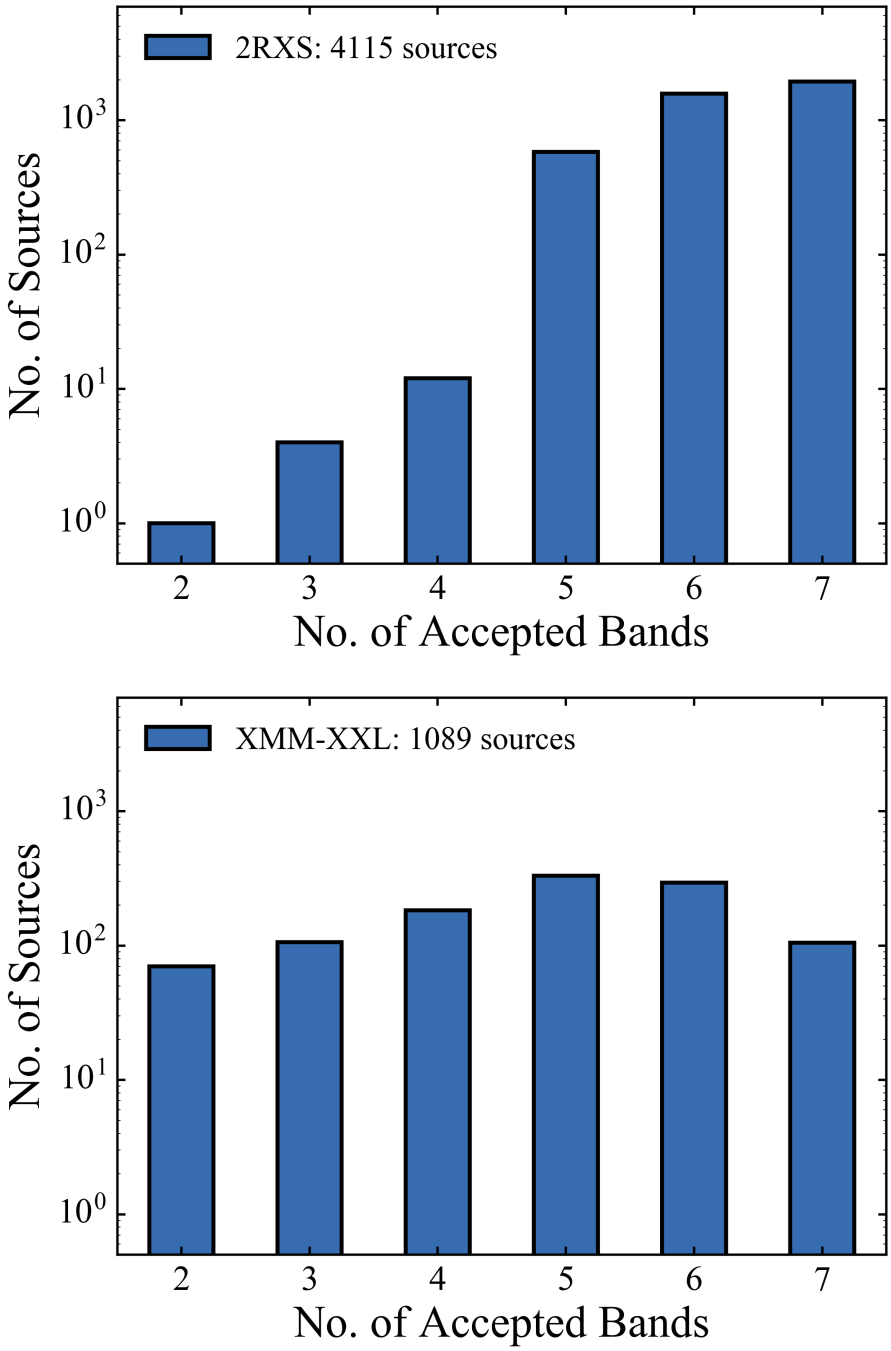


Figure 4.4: Distribution of the number of photometric bands accepted for SED fitting after the selection rules described in section 4.4.1 were imposed.

4.4.2 SED Fitting Procedure

The photometry available from SDSS and GALEX provides coverage of the optical and UV portion of the SED, and is summarised in table 4.1. SDSS "cmodel" magnitudes were used, which provide a good agreement with PSF magnitudes for point-like sources, but also are a good approximation of the Petrosian magnitudes for extended sources[7]. The SDSS u and z bands are known to deviate slightly from the AB system, and following the prescription on the SDSS flux calibration webpage[8] these bands were converted to the AB system using the conversion

$$m_u^{AB} = m_u^{SDSS} - 0.04$$
$$m_z^{AB} = m_z^{SDSS} - 0.02$$

The remaining SDSS and GALEX photometric bands are in the AB system. All photometric measurements were then corrected for Milky Way extinction using the Schlegel et al. (1998) dust maps and the Cardelli extinction law with R_v=3.1.

The optical-UV photometry was fit using the LePHARE[9] SED fitting software (Arnouts et al. 1999; Ilbert et al. 2006). LePHARE takes an input list of photometric data and performs a chi-square fit using a library of SED templates. The LePHARE "SYNTH" quasar model templates were used, which consist of a blackbody component, a broken power law, and a series of the main broad and narrow emission lines typically found in quasar spectra, with an intensity which is set to be proportional to the continuum. LePHARE also models the intrinsic extinction, and provides an estimate of the E_{B-V} value for each source. Each SED template is fit using a Milky Way (MW) (Allen 1976), MW (Seaton 1979), and Small Magellanic Cloud (SMC) (Prevot et al. 1984) extinction law (see figure 4.5), and the best fit law is selected for the extinction correction. The distributions of E_{B-V} values and extinction laws for both samples are shown in figure 4.6. As can be seen from the upper panels of figure 4.6, many sources were found to have low levels of extinction, typical for an X-ray selected, unobscured quasar sample, and in agreement with what was found in Salvato et al. (2009). An example of a LePHARE fit is shown in figure 4.7.

[7] https://www.sdss.org/dr12/algorithms/magnitudes/#cmodel
[8] https://www.sdss.org/dr12/algorithms/fluxcal/#SDSStoAB
[9] http://www.cfht.hawaii.edu/~arnouts/LEPHARE/lephare.html

Table 4.1: Summary of the photometric data used in the LePHARE SED fitting procedure discussed in section 4.4.2. Effective wavelengths, $\lambda_{\rm eff}$, and limiting magnitudes are taken from Morrissey et al. (2007) and Gunn et al. (1998). The different limiting magnitudes for GALEX correspond to surveys of different depths; the all-sky imaging survey (AIS), the medium imaging survey (MIS), and the deep imaging survey (DIS).

Instrument	Band	$\lambda_{\rm eff}$ [Å]	Limiting Magnitude [AB]
GALEX	FUV	1538.6	19.9 (AIS), 22.6 (MIS), 24.8 (DIS)
GALEX	NUV	2315.7	20.8 (AIS), 22.7 (MIS), 24.4 (DIS)
SDSS	u	3550	22.1
SDSS	g	4770	23.2
SDSS	r	6230	23.1
SDSS	i	7620	22.5
SDSS	z	9130	20.8

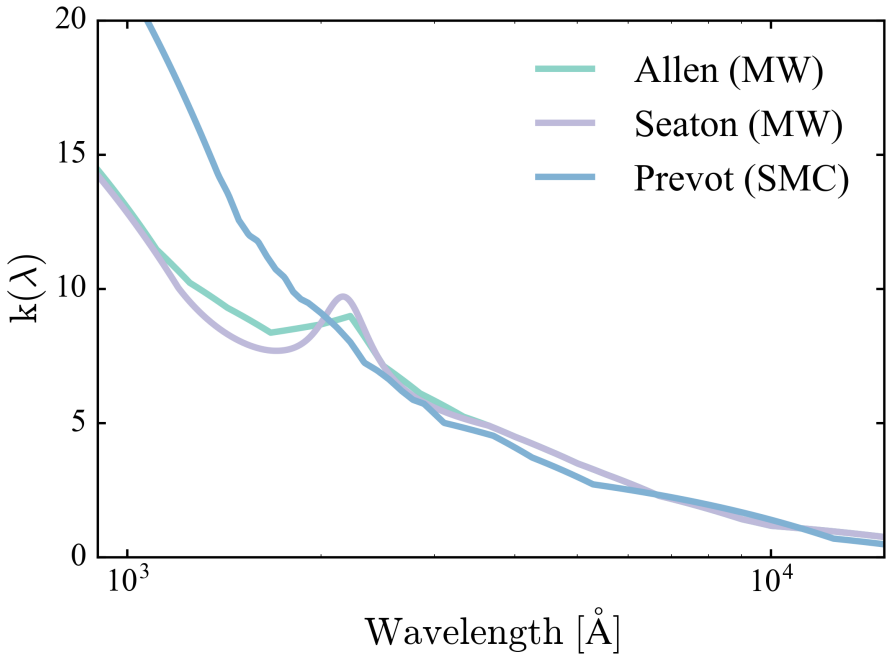


Figure 4.5: The three extinction laws used in the LePHARE SED fitting routine. See section 4.4.2 for further details.

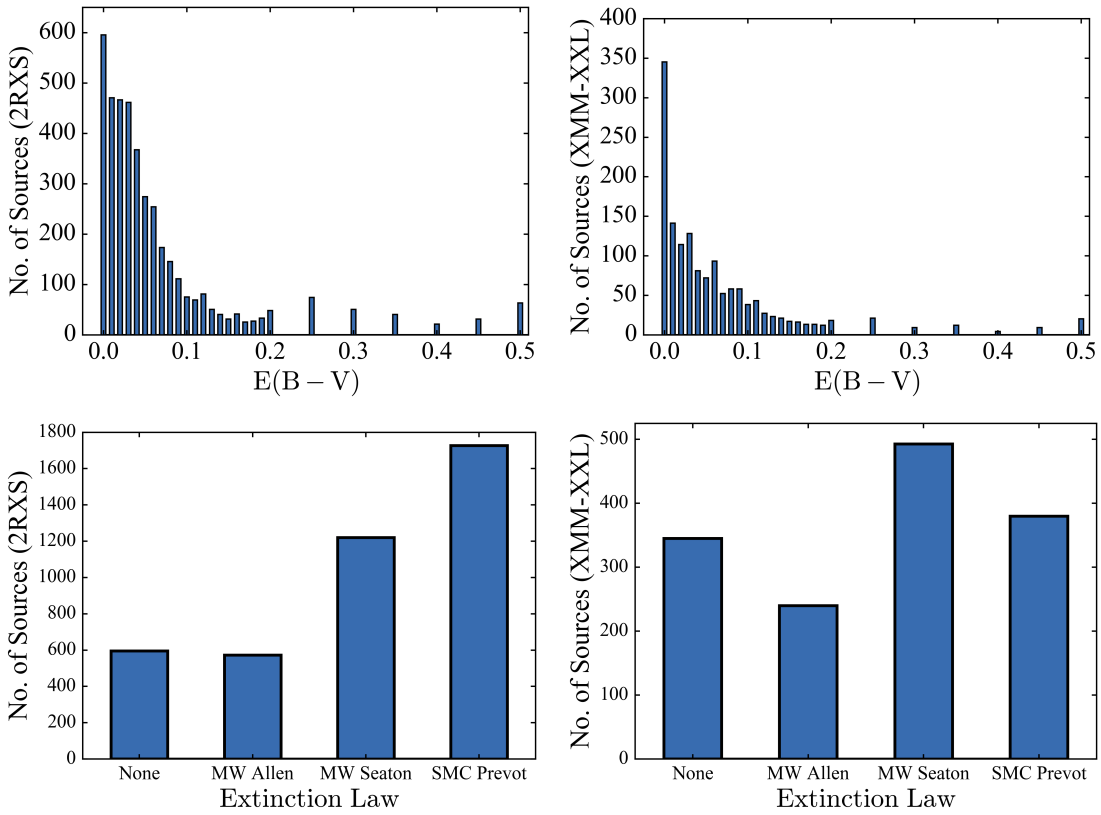


Figure 4.6: Upper panels: The distributions of the measured colour excess values for the 2RXS and XMM-XXL samples. Lower panels: The best-fit extinction laws used in the fit for the 2RXS and XMM-XXL samples. Note that for some sources, the application of extinction was not required in order to achieve a good fit.

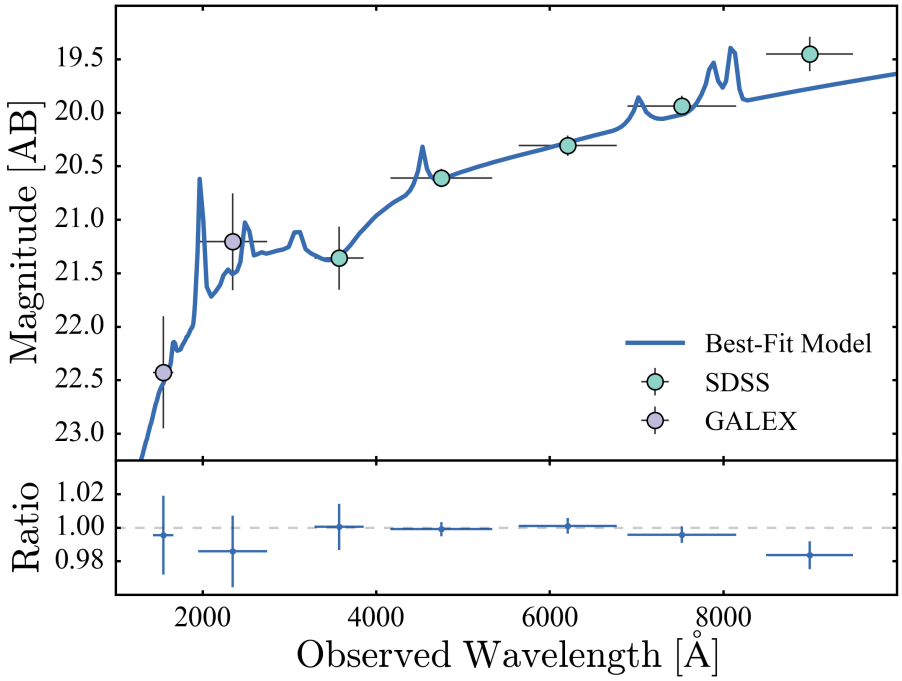


Figure 4.7: Example of an SED fit using LePHARE to GALEX and SDSS photometry.

4.5 Analysis of the F_X-F_{UV} Relation

Equation 1.5, which can be used to estimate distances modulus values for quasars, assumes that the quasar X-ray - UV luminosity relationship does not evolve with redshift. Redshift evolution would mean that any distance estimate obtained from equation 1.5 would be systematically under- or overestimated with redshift. Using their sample of 808 quasars, RL15 measured the slope of the $F_X - F_{UV}$ relationship in narrow redshift bins, γ_z, and found no significant evolution over the redshift range 0.5 - 6.5, with an average value of $\bar{\gamma}_z = 0.60 \pm 0.02$.

The evolution of the quasar $F_X - F_{UV}$ relationship was also measured using the two samples presented in this work. Following the method outlined in RL15, the sample was divided into bins with redshift width Log(z)$\sim$0.1. The X-ray-UV flux relation

$$\mathrm{Log}(F_X) = \gamma \mathrm{Log}(F_{UV}) + \beta' \tag{4.1}$$

was then fit in each bin. To perform the fit, the LINMIX (Kelly 2007) package was used (see section 3.1). LINMIX is a Bayesian linear regression algorithm that accounts for errors in both dependent and independent variables, as well as non-detections. In the following sections, the evolution of the best−fit slope, γ_z, as a function of redshift is discussed. A number of selection effects can be expected to influence the measurement of γ_z, and these include dust extinction, stellarity, and the X-ray detection likelihood. The rational behind these cuts, and their affect on the $F_X - F_{UV}$ relation, is also discussed in the following sections.

4.5.1 The Combined 2RXS - XMM-XXL Sample

The 2RXS sample includes 1872/3989 sources with measurements of the 2 keV flux. The redshift distribution of the two samples presented in this work is shown in figure 4.8. As can be seen from figure 4.8, the 2RXS sample provides a large number of sources at $z < 1$, while the majority of XMM-XXL sources lie at $z > 1$. Combined, the 2RXS and XMM-XXL samples provide coverage of the AGN population out to z=3.

The method for testing the cosmological evolution of the $\gamma - z$ relation discussed in section 4.5 requires a large range in UV and X-ray flux measurements. The UV and X-ray flux distributions as a function of redshift for the 2RXS and XMM-XXL samples are shown in the upper panels of figure 4.9. The 2RXS and XMM-XXL samples occupy different UV and X-ray flux ranges (see also the lower panels of figure 4.9), and, when combined, occupy a range similar to that of the RL15 sample, albeit with a larger dispersion, especially at low redshifts.

RL15 combined multiple quasar samples from the literature in order to achieve a wide range in redshift and luminosity. Following this method, the 2RXS and XMM-XXL samples presented in this work have been combined in order to study the $\gamma - z$ relation. Two sources which were detected by both the 2RXS and XMM-XXL surveys were removed from the 2RXS sample. This combined sample contained 2959 sources, and is hereafter be referred to as the "2RXS-XXL" sample. The mean SEDs for both samples are shown in figure 4.10.

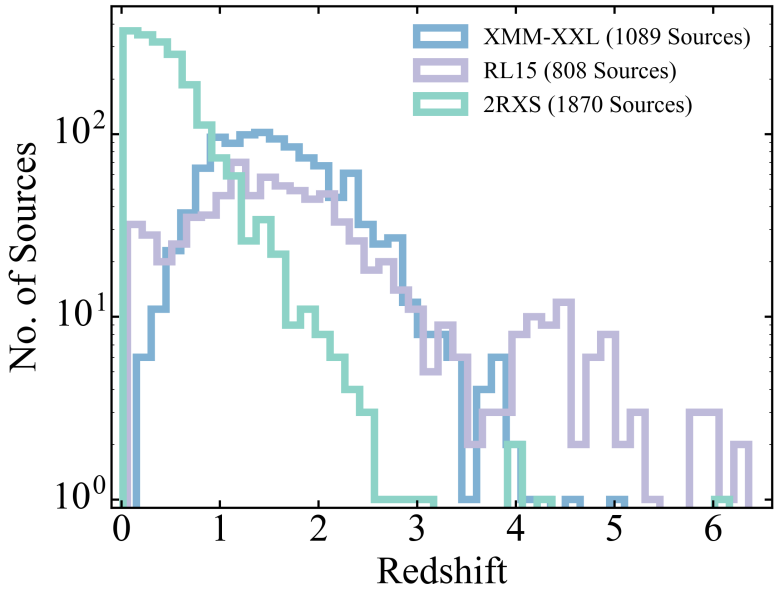


Figure 4.8: Redshift distribution of the 2RXS (green), XMM-XXL (blue), and RL15 (purple) samples.

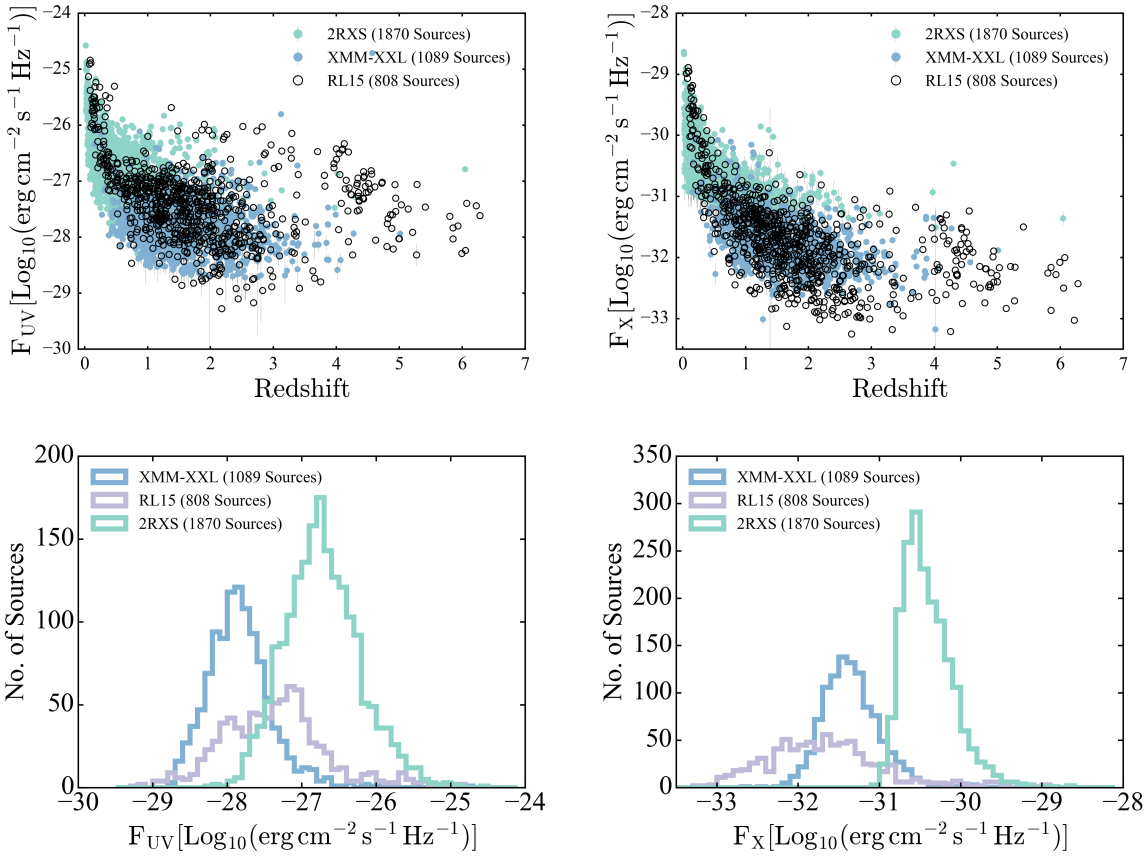


Figure 4.9: Upper panels: UV (left) and X-ray (right) fluxes versus redshift for the 2RXS (green) and XMM-XXL (blue) samples. The RL15 sample (black) is also shown for comparison. Lower panels: UV (left) and X-ray (right) flux distribution for the 2RXS (green), XMM-XXL (blue), and RL15 (violet) samples.

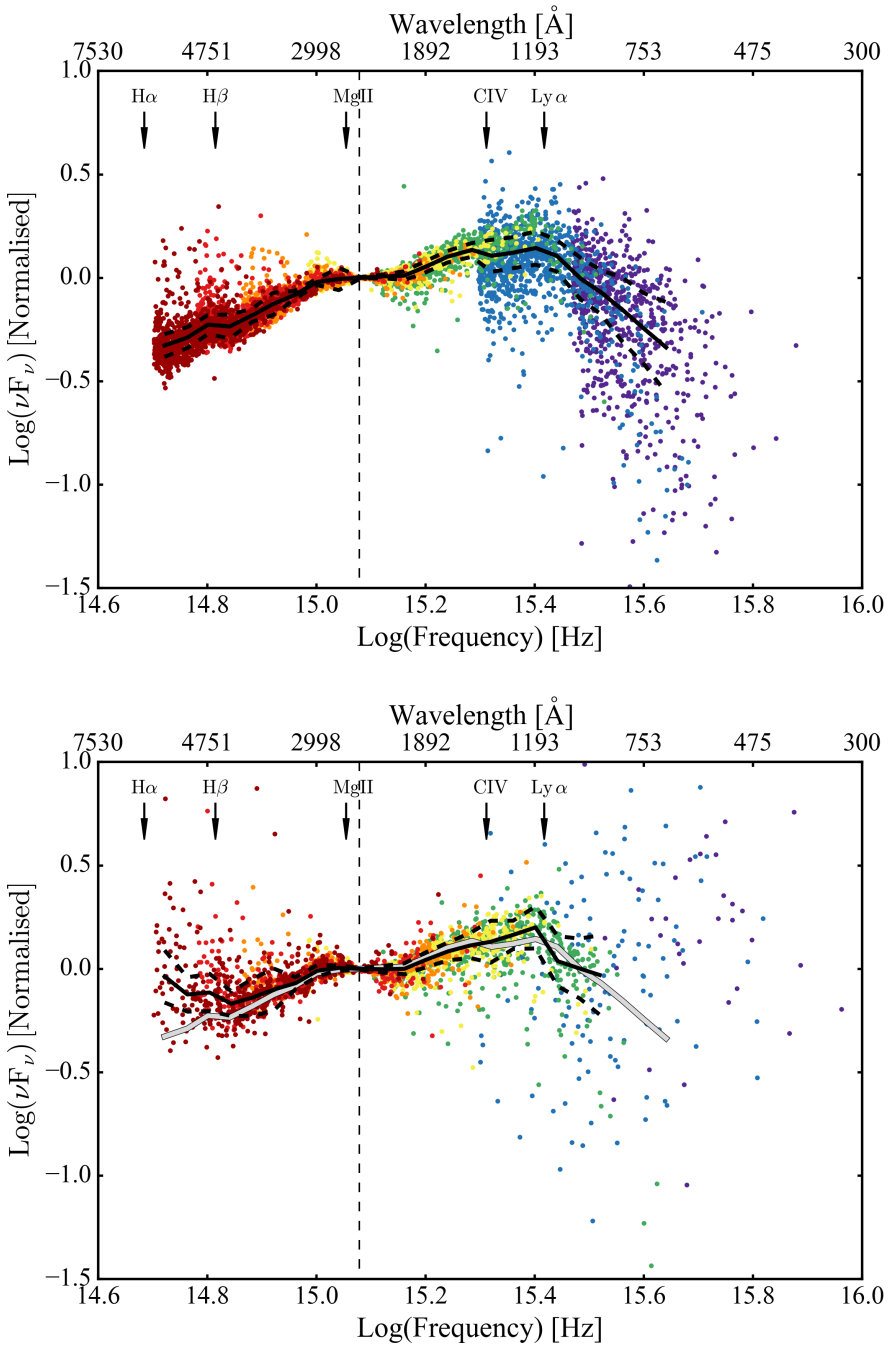


Figure 4.10: Rest-frame SEDs for the sources in the clean 2RXS subsample (upper panel), and XXL sample (lower panel) above a redshift of 0.5. The symbol colours (from red to purple), represent the SDSS and GALEX photometric bands (from z to FUV). Each SED is normalised to its flux at 2500Å (indicated by the dashed line). In both panels, the solid black line represents the median SEDs, and the black dashed lines represent the standard deviation of the fluxes in each bin. The white line in the lower panel shows the ROSAT median SED for comparison. The location of prominent quasar emission lines are indicated for reference.

4.5.2 Sample Quality Cuts and γ Measurement

A number of preliminary quality cuts were made to the "2RXS-XXL" sample of 2959 sources as follows:

- Remove 2RXS sources with uncertain redshift or classification (CONF_BEST = 3).
- Remove unreliable AllWISE counterparts to 2RXS detections (p_any < 0.01, see section 2.1.4).
- Remove 2RXS counterparts with SDSS spectral warning flags (ZWARNING==0).

In addition, two 2RXS sources were removed which did not have reliable redshifts (ID_2RXS=4121 and ID_2RXS=8491). This resulted in a clean sample of 2753 2RXS-XXL sources. This sample of 2753 2RXS-XXL sources was used to measure the X-ray - UV slope, γ, following the method outlined in section 4.5, the results of which are show in figure 4.11.

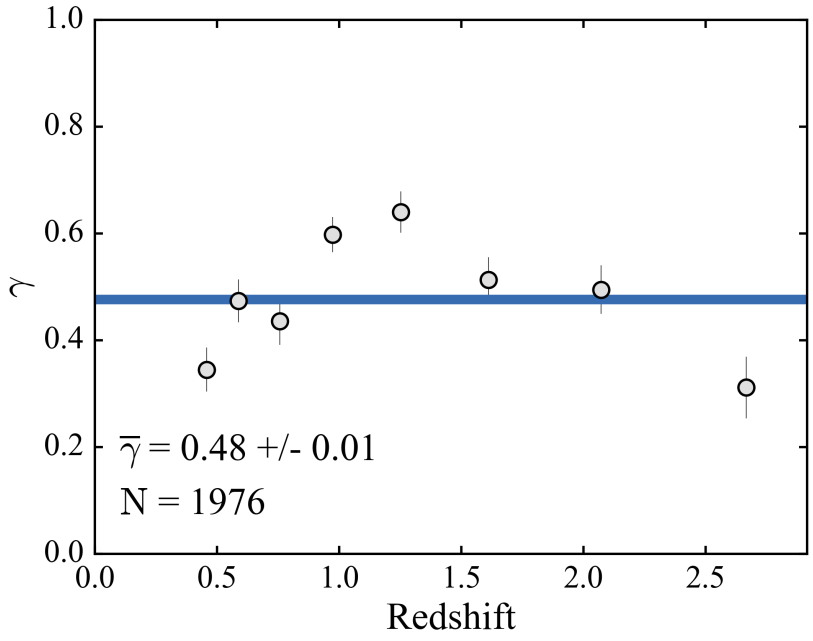


Figure 4.11: Best-fit X-ray-UV slope, γ_z, as a function of redshift obtained from fitting the 1976 sources in the 0.5 - 2.5 redshift range from the main 2RXS-XXL sample using LINMIX. The weighted mean of the γ_z measurements (illustrated by the horizontal blue line) and corresponding 1σ uncertainties are also shown.

The results presented in figure 4.11 show that the best-fit measurement of γ from the 2RXS-XXL sample is lower than the value of $\gamma = 0.6$ found by RL15. In addition, figure 4.11 illustrates how the best-fit γ measurement varies significantly across the redshift range

covered by this sample. The fact that γ is not found to remain constant with redshift indicates that this sample is not appropriate for use in the quasar cosmology experiment (as discussed in section 4.5).

4.6 Exploring the 2RXS-XXL Sample Properties

4.6.1 Stellarity

The SED fitting procedure described in section 4.4.2 does not include a host galaxy component in the SED model, and therefore cannot distinguish between the host and AGN contributions to the UV flux. Sources at lower redshifts are likely to have a significant host galaxy contribution to their UV flux, and therefore will bias the measurement of the $F_X - F_{UV}$ slope γ. Sources that are likely to have a significant host galaxy contribution to their UV flux can be identified by defining a u band "stellarity" index, as follows:

$$S(u) = \text{cModelMag\_u} - \text{psfMag\_u}$$

A description of how cModelMag_u and psfMag_u are measured is available on the SDSS "magnitudes" webpage[10]. The distribution of stellarity as a function of redshift for the 2RXS and XMM-XXL samples is presented in figure 4.12.

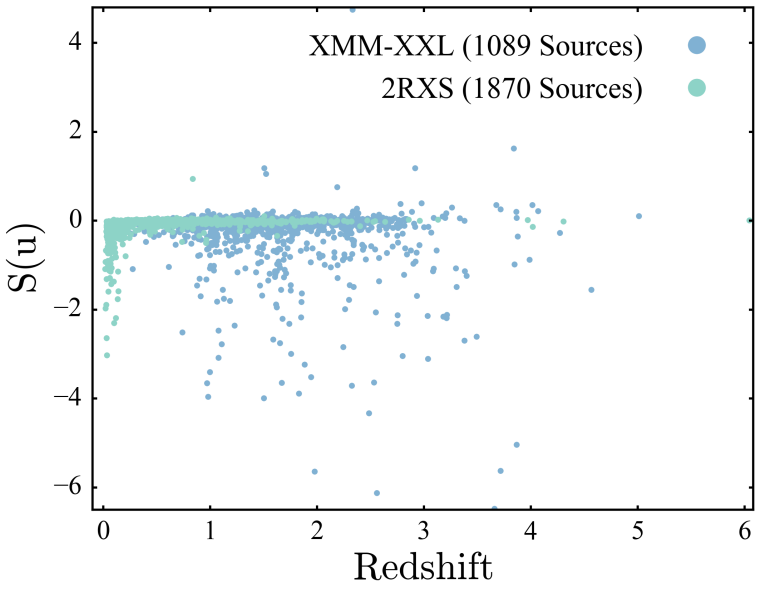


Figure 4.12: SDSS u band stellarity index versus redshift for the 2RXS and XMM-XXL samples.

[10] https://www.sdss.org/dr12/algorithms/magnitudes/

Extended sources can be seen as a sharp decrease in S(u), clearly visible, especially at low redshifts, for the 2RXS sample in figure 4.12. Sources with stellarity indices that deviate from zero can be expected to significantly reduce the accuracy of the UV flux density measurements derived from the SED fitting procedure used in this work. This will in turn bias the measurement of γ, thus rendering the sources inappropriate for use as standard candles in this experiment. Therefore, when selecting a sample for use in the quasar cosmology experiment, it may be necessary to either select a sample with a low stellarity index, or to implement a host galaxy - AGN decomposition in the SED fitting procedure.

4.6.2 Extinction

As discussed in section 4.4.2, the LePHARE SED fitting procedure provides an estimate of the intrinsic extinction for each fit. Both the 2RXS and XMM-XXL samples primarily contain sources with low levels of extinction (see figure 4.6, upper panels). However, as can be seen from figure 4.2, an E_{B-V} of 0.1 results in a flux correction of $\sim 50\%$ at 2500Å. This introduces a large source of uncertainty in the UV flux density measurements, since any unreliability in the extinction correction will dramatically affect the resulting extinction-corrected fluxes for any source that shows evidence for even modest levels of extinction.

When selecting a sample for use in the quasar cosmology experiment it is therefore necessary to define a sample for which the extinction correction is small, or even negligible. An alternative approach would be to ensure that the photometric coverage and SED fitting procedure can measure the E_{B-V} precisely for each source in order to apply a reliable extinction correction. Since the 2RXS sources are at lower redshifts, the photometry used when measuring E_{B-V} will typically not sample the far-UV portion of the SED where the effects of extinction are most prominent. The extinction measurements for the 2RXS sample are therefore not considered to be especially reliable.

4.6.3 Detection Likelihood

For each source, the 2RXS pipeline (Boller et al. 2016) calculates the detection likelihood, EXI_ML. Boller et al. (2016) carried out simulations of the ROSAT all-sky survey in order to estimate the number of spurious detections as a function of EXI_ML. The 2RXS catalogue produced by Boller et al. (2016) includes sources down to a minimum EXI_ML of 6.5, which is the limit applied to the ROSAT FSC (Voges et al. 2000). This limit of EXI_ML=6.5 therefore also applies to the sample presented in this work (see figure 4.13, right panel. The detection likelihood for the XMM-XXL sample is also shown for comparison). However, Boller et al. (2016), figure 1, indicates that sources close to the EXI_ML=6.5 limit can have a spurious fraction up to 34% (for fields close to the North Ecliptic Pole).

The spurious fraction of a given quasar sample may be expected to reduce both the precision and accuracy of the resulting γ measurement. For example, if the presence of spurious sources in a sample has the effect of increasing the dispersion in the $F_X - F_{UV}$ source distribution, the expectation is that this will decrease the precision of the resulting γ measurement. This prediction resulted from the analysis of the quasar cosmology experiment using simulations of the *eROSITA*-4MOST sample presented in section 5.7.2. Future work attempting to develop the use of survey quality data in the quasar cosmology experiment may be advised to explore the specific influence of the spurious fraction, and to determine a threshold above which a given sample would not be expected to return reliable results.

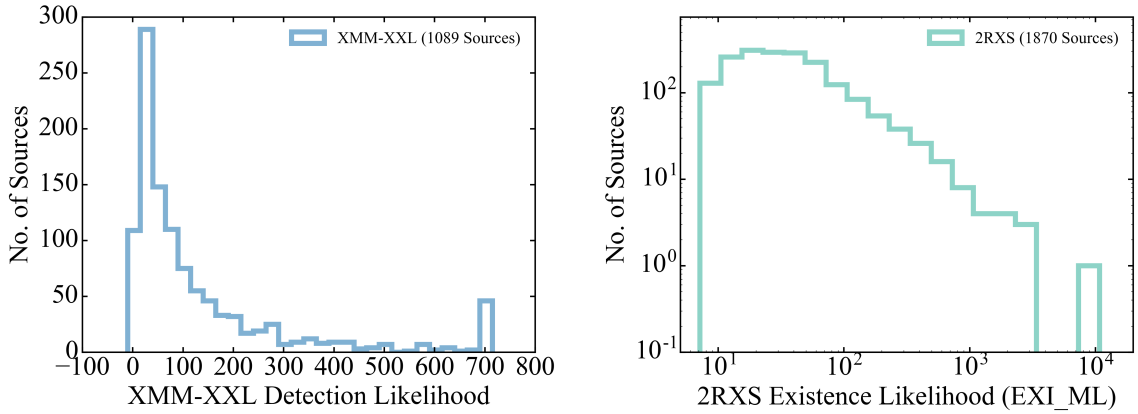


Figure 4.13: Left panel: XMM-XXL detection likelihood in the full band. Right panel: 2RXS detection likelihood.

4.7 Advantages and Disadvantages of Quasars as Standard Candles

A significant source of error not discussed so far in this work is the intrinsic quasar X-ray and UV variability. This would introduce scatter in the X-ray-UV flux relationship even in the ideal case where the observations were carried out simultaneously. This is due to the fact that emission at X-ray and UV wavelengths does not vary instantaneously. Instead, the UV emission often lags the X-ray emission (see McHardy et al. 2014). Using non-simultaneous observations will increase this intrinsic scatter. This is likely to be an important contribution to the dispersion in the Hubble diagram for the quasar samples presented here (which have X-ray-UV observations separated by months to years). As mentioned above, the analysis of the *eROSITA*-4MOST simulations presented in section 5.7.2 indicated that increasing the scatter in the X-ray-UV flux relationship decreases the precision of the resulting γ measurement.

Several groups have made efforts to quantify the intrinsic dispersion of the X-ray-UV relation in quasars. Using a sample of radio-quiet, non-BAL quasars with non-simultaneous observations, Gibson et al. (2008) have shown that observed X-ray flux varies by a factor of $\sim$2 relative to the X-ray flux that would be estimated from the UV flux. Selecting a sample of unobscured quasars with similar SEDs, and multiple X-ray observations, Lusso & Risaliti (2016) have shown that the measured dispersion in the L_X-L_{UV} relation is $\sim$0.23 dex, using the X-ray detection with the longest exposure time. When the X-ray luminosities are averaged, the dispersion falls to $\sim$0.21 dex.

Simm et al. (2015) carried out a detailed analysis of the influence of variability on the estimation of photometric redshifts, for a sample of X-ray selected AGN in the XMM-COSMOS field. It was found that when building an SED using photometric points from different epochs, the resulting photo-z accuracy was significantly reduced. This is due to the SED continuum shape changing between epochs, or to more extreme cases where bands at different epochs appear as false emission lines (see Simm et al. 2015, figure 8). Besides intrinsic source variability, the differences in quasar SEDs, known to be significant, will inevitably add to the scatter in the Hubble diagram. It is likely that a more detailed characterisation of quasar SEDs would allow for a more self-similar sample to be selected.

Despite these challenges, there are several obvious advantages that quasars have as standard candles. Quasars are both detectable in large numbers and out to high redshifts, properties that are not represented by current SNe Ia samples. The ability to preferentially select optically blue quasars can in principle provide sources for which the extinction correction is minimal. This, as discussed in section 4.6.2, may be critical in order to produce reliable measurements of the UV flux density, which may reduce one of the main sources of scatter in the resultant Hubble diagram. The fact that quasars are primarily persistent objects also provides the opportunity to compare multiple re-observations of a given sample. The average flux obtained from these campaigns could be used instead of a single epoch flux, which may also serve to reduce the $F_X - F_{UV}$ dispersion.

4.8 Summary and Conclusions

In this work, an attempt was made to test the application of survey quality data to the RL15 quasar cosmology experiment. The 2RXS and XMM-XXL X-ray selected samples with SDSS optical spectroscopic follow-up were prepared and X-ray and UV flux densities were measured for a reliable subsample of 2753 sources. It was found that this sample did not provide a measurement of the $F_X - F_{UV}$ slope γ that was stable with redshift, thus invalidating one of the key assumptions of the RL15 method. In addition, it was found that the average value of γ measured using this sample was significantly lower than that found by RL15. Therefore it was concluded that the survey quality data available at the commencement of this work was not of sufficient quality to provide a sample of reliable standard candles for use in the quasar cosmology experiment.

The analysis presented in this chapter raised the question: Could future large-scale X-ray-UV quasar datasets overcome the limitations of the data used in this work? To this end, simulations of the future *eROSITA*-4MOST sample were developed in order to provide a forecast of the cosmological constraints that that sample may provide. This work is presented in chapter 5. In chapter 5, the *eROSITA*-4MOST simulated data is also used to address many of the potential reasons why the 2RXS and XMM-XXL data used in this chapter did not result in a reliable sample of standard candles, which include the influence of flux density precision, intrinsic scatter, sample size, γ accuracy, and regression dilution.

Chapter 5

Quasar Cosmology in the Era of *eROSITA*

By the end of its all-sky survey, eROSITA will have detected $\sim 3 \times 10^6$ AGN, a third of which are expected to receive optical spectroscopic follow-up from 4MOST. This wealth of observational data will be of signal importance to many areas of quasar research, including the quasar cosmology experiment discussed in the previous chapter. The work presented in this chapter aims to provide a forecast of the cosmological constraints that may be achieved via the quasar Hubble diagram using a simulation of the eROSITA-4MOST sample.

5.1 The *eROSITA* Mission

The extended Roentgen Survey with an Imaging Telescope Array (*eROSITA*, Predehl et al. 2010) is the primary instrument on board the Russian-German Spectrum-Roentgen-Gamma (*SRG*, Pavlinsky et al. 2008) mission alongside the Astronomical Roentgen Telescope X-ray Concentrator (*Art-XC*, Pavlinsky et al. 2016). *SRG* was launched from Baikonur Cosmodrome in Kazakhstan on the 13$^{\text{th}}$ of July 2019 with a Proton-M launch vehicle. *eROSITA* will spend the first four years of operation conducting an all-sky survey in the 0.3 - 10 keV band. This will be the first all-sky imaging survey in the hard X-ray band (2 - 10 keV), while in the soft X-ray band (0.2 - 2 keV) *eROSITA* will have a sensitivity that is a factor of $\sim$10 deeper than *ROSAT* (see figure 5.1). *eROSITA's* all-sky survey will be followed by approximately three years of pointed observations. During this time, targets and observation campaigns will be selected based on proposals from the scientific community.

eROSITA was built by the Max-Planck Institute for Extraterrestrial Physics (MPE), Garching, Munich, and contains 7 identical X-ray telescope modules, each with 54 mirror shells, which have a focal length of 1.6 m (Merloni et al. 2012). *eROSITA* has an effective area of $\sim$1,500 cm^2 at 1.5 keV, an angular resolution (half energy width, on-axis) of 16 arcsec (Cappelluti et al. 2011), and an energy resolution of 138 eV at 6 keV. The 7 telescopes are co-aligned, and have the same 1 deg^2 field of view for each telescope (Merloni et al. 2012).

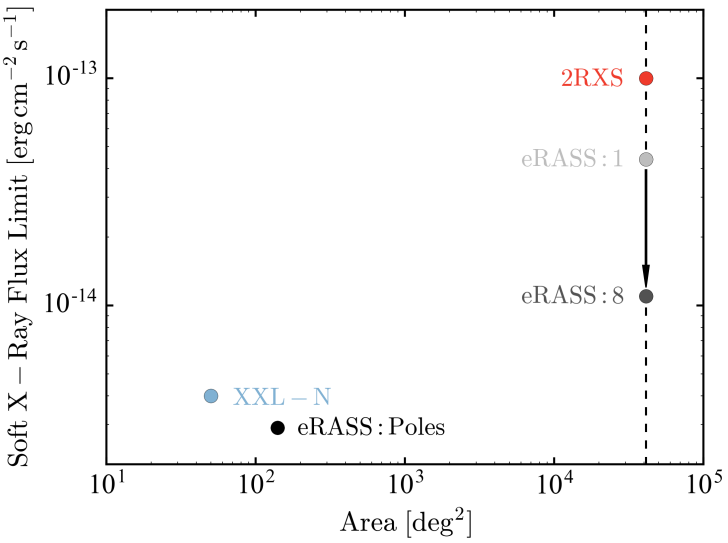


Figure 5.1: Soft X-ray flux limit versus survey area for the *eROSITA*, 2RXS (Boller et al. 2016), and XMM-XXL North surveys. The projected *eROSITA* survey depths represent the 0.5 - 2 keV point-source flux limit averaged over the full sky after the first 6 months (eRASS:1), after the final 4 year survey (eRASS:8), and after the final 4 year survey for a region of 140 deg^2 around the ecliptic poles (eRASS:Poles, Merloni et al. 2012). The XMM-XXL north limit shown is the 90% completeness limit in the 0.5 - 2 keV band (Pierre et al. 2016). The 2RXS flux limit is an approximate limit taken from Boller et al. (2016). The vertical dashed line represents the area of the full sky.

eROSITA orbits the second Lagrange point which allows for continuous observation free from occultation by the Earth. *eROSITA* scans the full sky in one six-month orbit, and will complete eight full scans during the four-year survey phase. The final *eROSITA* catalogue will include detections from a combination of the eight individual scans (Merloni et al. 2012).

eROSITA will provide an unprecedented level of observational detail across the entire X-ray sky, and will observe a diverse range of objects from Galactic stars to high-redshift quasars. *eROSITA* has been designed to detect a large sample ($\sim 10^5$) of galaxy clusters out to $z \simeq 1$. Galaxy clusters trace the large-scale distribution of matter in the Universe, and their number density and distribution can be used to constrain cosmological models (e.g. Rosati et al. 2002) and the amount of Dark Energy and Dark Matter. The sample that *eROSITA* will provide is expected to significantly improve upon the current constraints on a number of cosmological parameters, as well as provide a large dataset for the study of cluster structure and thermodynamics (see Merloni et al. 2012). In addition to galaxy clusters, *eROSITA* is expected to detect $\sim 3 \times 10^6$ AGN (Merloni et al. 2012), which will allow a detailed investigation into their accretion history (through the study of the AGN X-ray Luminosity Function), as well as the link between the evolution of BHs and large

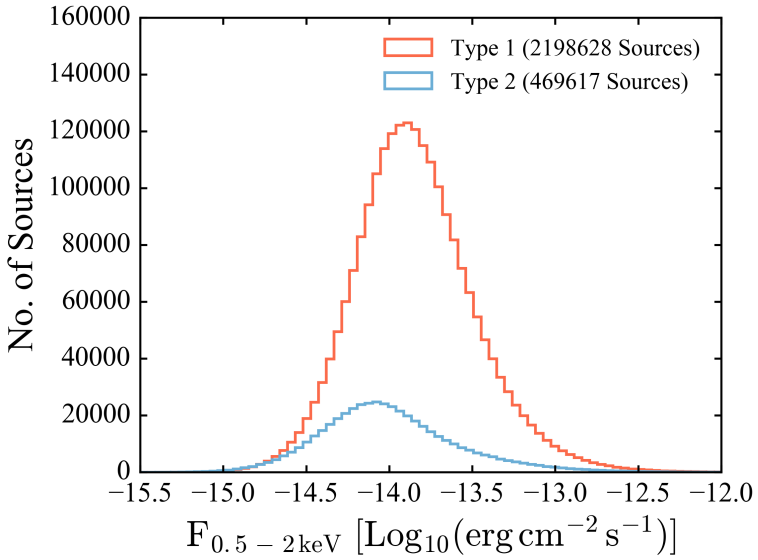


Figure 5.2: Soft X-ray flux distribution for optically classified type 1 and 2 AGN in the Comparat et al. (2019) *eROSITA* mock catalogue.

scale structure.

Being the first instrument to carry out an all-sky survey in both the hard and soft X-ray bands, *eROSITA* is also expected to detect a large sample of obscured AGN (see figure 5.2). This sample will provide a more complete view of the true AGN population than is currently available and will aid the understanding of AGN and BH growth across cosmic time. This sample of obscured AGN will also permit a detailed study of the relationship between AGN and the properties of their host galaxies (Merloni et al. 2012). Along with galaxy clusters and AGN, *eROSITA* is expected to detect tens of thousands of compact objects (stellar mass black holes, neutron stars, and white dwarfs), several hundred thousand stars, map the Milky Way interstellar medium, observe many nearby galaxies, and detect a variety of variable X-ray sources (Merloni et al. 2012).

5.2 Multiwavelength Observations of Quasars From *eROSITA* - 4MOST

As in the case of *ROSAT* and *XMM-Newton* (see section 1.7.1), to maximise the scientific potential of the large X-ray dataset that will be provided by *eROSITA*, a dedicated programme of optical spectroscopic follow-up is required. Optical spectroscopic campaigns targeting *eROSITA* sources will be conducted by both SDSS-V (Black Hole Mapper, Kollmeier et al. 2017) and the 4-metre Multi-Object Spectroscopic Telescope (4MOST, de Jong et al. 2014, 2019). 4MOST is a wide-field, multi-object spectrograph which will be installed at the 4 m European Southern Observatory Visible and Infrared Survey Telescope

for Astronomy (VISTA, Sutherland et al. 2015) at Paranal, Chile. 4MOST has a total of 2,436 1.45 arcsec spectroscopic fibres, 1,624 of which are connected to two lower-resolution spectrographs ($\lambda/\Delta\lambda \sim 6{,}500$) and 812 of which are connected to one higher-resolution spectrograph ($\lambda/\Delta\lambda \sim 20{,}000$). 4MOST has a 4.2 deg^2 field-of-view, and throughout the planned five-year survey is expected to cover a sky area of 17,000 deg^2 twice. During the planned two-hour observation period, 4MOST is expected to obtain redshifts for objects down to an r magnitude of 22.5 (de Jong et al. 2019).

eROSITA will have completed multiple sky scans by the time 4MOST commences science operations in 2022. Following the method employed for the 2RXS survey (see section 2.1.2), the Bayesian cross-matching algorithm NWAY (Salvato et al. 2018) will be used to identify optical counterparts and provide targets for 4MOST spectroscopy using a prior optimised for the depth of the *eROSITA* survey. 4MOST will target both the point-like and extended optical counterparts to the X-ray sources detected by *eROSITA* (Merloni et al. 2019). The 4MOST follow-up of *eROSITA* sources is expected to reach a spectroscopic completeness of 80 - 90 per cent for X-ray sources in the accessible area of the sky which have $F_{0.5-2\,\mathrm{keV}} > 10^{-14}\,\mathrm{erg\,cm^{-2}\,s^{-1}}$. 4MOST is expected to target up to 10^6 *eROSITA*-detected AGN over 10,000 deg^2 out to $z = 6$. This will increase the number of AGN with optical spectroscopic confirmation by a factor of a hundred relative to the samples that were available at the time the analysis carried out in chapter 4 was commenced.

5.3 The *eROSITA* Mock Catalogue

Using N-body simulations (Klypin et al. 2016), Comparat et al. (2019) have produced a mock catalogue of *eROSITA* AGN. Comparat et al. (2019) began by simulating *eROSITA* AGN from $0 < z < 6$ over the full sky using a model which is intended to follow the observed AGN X-ray luminosity function from Aird et al. (2015). An empirical obscuration model was also implemented, which provided both X-ray and optical AGN classification. In addition to object classifications, UV and optical magnitudes and optical spectra have also been included in the simulation. Comparat et al. (2019) adopted a flat ΛCDM cosmology with $\Omega_M = 0.307115$, $h = 0.6777$, and $\sigma_8 = 0.8228$ (see Klypin et al. 2016).

This simulated *eROSITA* catalogue contained 2,668,245 AGN over the full sky. Sources from this simulated sample that will be in the German[1] *eROSITA*-4MOST region of the sky were selected using the following criteria:

- |galactic latitude| > 10 deg → deselects the sky area occupied by the Milky Way
- galactic longitude > 180 deg → selects the German half of the *eROSITA* sky
- equatorial declination < 5 deg & equatorial declination > −80 deg
  → selects the 4MOST maximum sky coverage

[1]For a fixed period of time the all-sky survey data will be available only to the *eROSITA* collaboration members, and will be divided equally between the German and Russian consortia.

This returned a sample of 965,636 sources. When assigning optical and X-ray classifications to sources in their mock catalogue, Comparat et al. (2019) adopted the classification scheme defined by Merloni et al. (2014). Sources classified as type 1 at both X-ray and optical wavelengths, identified in the mock catalogue via the designation "AGN_type==11", were selected, which returned 752,985 sources. Throughout the remainder of this chapter, the term "quasar" will be used to refer to type 1 AGN and quasars collectively, without making a distinction based on source luminosity. Finally, the analysis presented here was restricted to *eROSITA* quasars with an SDSS r magnitude $\leq$ 20.5, which 4MOST can be expected to detect at a signal-to-noise ratio that is approximately ≥ 10. This returned a subsample of 317,391 sources, the X-ray luminosity - redshift distribution of which is shown in figure 5.3.

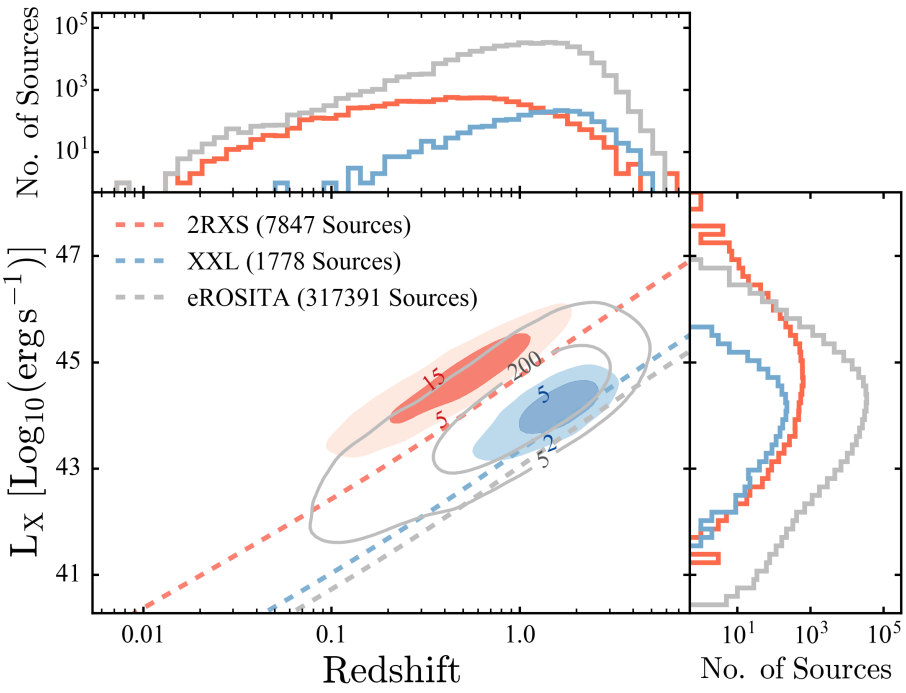


Figure 5.3: Soft X-ray luminosity versus redshift for the *eROSITA*-4MOST simulated sample (grey contours) from Comparat et al. (2019) (see section 5.3) along with the 2RXS (red contours) and XMM-XXL (blue contours) samples used in the quasar cosmology experiment discussed in chapter 4. For both XMM-XXL and 2RXS, the full sample of quasars with both X-ray detections and spectroscopic redshifts are shown. The dashed lines represent the flux limits for each survey; 1×10^{-13} erg cm^{-2} s^{-1} (approximate 2RXS flux limit, Boller et al. 2016), 4×10^{-14} erg cm^{-2} s^{-1} (XMM-XXL 90 per cent soft X-ray completeness limit, Pierre et al. 2016), and 2×10^{-15} erg cm^{-2} s^{-1} (*eROSITA* soft X-ray flux limit at the ecliptic poles, Predehl et al. 2016).

5.4 Simulating *eROSITA*-4MOST Flux Density Measurements

The method used to implement quasars as standard candles discussed in section 1.6.2 assumes that a relationship exists between the quasar X-ray and UV luminosities that does not evolve with redshift. In order to use the Comparat et al. (2019) mock catalogue to test the RL15 method, it was first necessary to simulate a sample of sources that follow the quasar $L_X - L_{UV}$ relation. The *eROSITA*-4MOST mock catalogue produced by Comparat et al. (2019) contains rest-frame 0.5 - 2 keV luminosities, $L_{0.5-2\,keV}$, for each source. The rest-frame 2 keV monochromatic luminosity was estimated from this as follows:

$$L_{E_e,\,2\,keV} = \frac{L_{0.5-2\,keV}}{(2-0.5\,keV)} \qquad [erg\,s^{-1}\,keV^{-1}]$$

where $L_{E_e,\,2\,keV}$ is a monochromatic luminosity that is at a rest-frame energy approximately equal to the middle of the 0.5 - 2 keV band. This monochromatic luminosity was then converted to the desired units;

$$L_{\nu_e,\,2\,keV}\ [erg\,s^{-1}\,Hz^{-1}] = h \times L_{E_e,\,2\,keV} \qquad [erg\,s^{-1}\,keV^{-1}]$$

where h is the Planck constant ($4.135667662 \times 10^{-18}$ keV s). The quasar $L_X - L_{UV}$ relation can be modelled as

$$Log_{10}(L_{\nu_e,\,2\,keV}) = \beta + \gamma Log_{10}(L_{\nu_e,\,2{,}500\,\text{Å}}) + \epsilon \tag{5.1}$$

where ϵ is the intrinsic scatter, which is Gaussian distributed with a mean of zero and a standard deviation of σ; $\epsilon \sim \mathcal{N}(0, \sigma)$.

Equation 5.1 was used to simulate the 2,500 Å monochromatic luminosity for each source assuming $\gamma = 0.6$ and $\beta = 8$. A σ of 0.3 was adopted which corresponds to the mean dispersion of the RL15 sample (see their figure 4). Using the cosmological parameters adopted by Comparat et al. (2019) to produce the simulation ($\Omega_M = 0.307115$ and $\Omega_\Lambda = 0.692885$) the corresponding observed-frame flux densities were then calculated as follows:

$$F_{\nu_o,\,2\,keV} = \frac{L_{\nu_e,\,2\,keV}(1+z)}{4\pi D_L^2} \qquad\qquad F_{\nu_o,\,2{,}500\,\text{Å}} = \frac{L_{\nu_e,\,2{,}500\,\text{Å}}(1+z)}{4\pi D_L^2}$$

Simulated measurements of these flux densities, $\widehat{F_{\nu_o,\,2\,keV}}$ and $\widehat{F_{\nu_o,\,2{,}500\,\text{Å}}}$, were computed assuming a Gaussian distribution with means equal to $F_{\nu_o,\,2\,keV}$ and $F_{\nu_o,\,2{,}500\,\text{Å}}$, and standard deviations equal to $\delta F_{\nu_o,\,2\,keV}$ and $\delta F_{\nu_o,\,2{,}500\,\text{Å}}$, respectively;

$$\begin{aligned} \widehat{F_{\nu_o,\,2\,keV}} &\sim \mathcal{N}(F_{\nu_o,\,2\,keV}, \delta F_{\nu_o,\,2\,keV}) \\ \widehat{F_{\nu_o,\,2{,}500\,\text{Å}}} &\sim \mathcal{N}(F_{\nu_o,\,2{,}500\,\text{Å}}, \delta F_{\nu_o,\,2{,}500\,\text{Å}}) \end{aligned} \tag{5.2}$$

The *eROSITA* Final Equatorial Depth Survey (eFEDS, Brunner et al. 2022) is a $\sim 140\,\mathrm{deg}^2$ survey that was designed to reach the depth that is expected from the final *eROSITA* all-sky survey (eRASS:8). The flux density precision and range covered by eFEDS observations are also expected to be representative of the final *eROSITA* sample. In this work, the X-ray flux density uncertainty model will be based on the eFEDS observations. To derive a flux density uncertainty model, the eFEDS "ML_FLUX_2" was used, which is the observed flux in the 0.6 - 2.3 keV band. In order to select a subsample of reliable eFEDS observations, sources with a detection likelihood < 8, sources with anomalously large fractional flux uncertainties ($\log_{10}$(ML_FLUX_ERR_2/ML_FLUX_2) ≥ 1.0), and sources with anomalously large measured extensions (EXT ≥ 10), were removed. X-ray flux densities were then estimated from the eFEDS 0.6 - 2.3 keV flux as follows:

$$\mathrm{F_{E_o}} = \frac{\mathrm{F_{0.6-2.3\,keV}}}{(2.3\ -\ 0.6\,\mathrm{keV})} \qquad [\mathrm{erg\,cm^{-2}\,s^{-1}\,keV^{-1}}]$$

where $\mathrm{F_{E_o}}$ is a flux density that is at an observed-frame energy approximately equal to the middle of the 0.6 - 2.3 keV band. This flux density, expressed in units of $[\mathrm{erg\,cm^{-2}\,s^{-1}\,Hz^{-1}}]$, appears to follow a relation of the form

$$\mathrm{Log_{10}}\left(\frac{\delta \mathrm{F_{\nu_o}}}{\mathrm{F_{\nu_o}}}\right) = -0.53\mathrm{Log_{10}}(\mathrm{F_{\nu_o}}) - 17.27 \tag{5.3}$$

which is shown by the red line in figure 5.4.

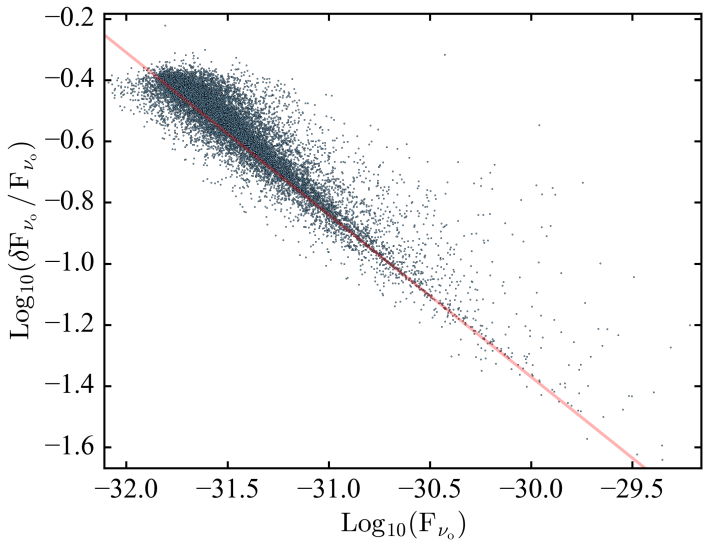


Figure 5.4: Fractional X-ray flux density uncertainty versus X-ray flux density for the eFEDS sample. The red line, which does not represent a fit to the data, is defined by equation 5.3.

This relation has not been fit to the data but merely represents the general trend of decreasing fractional flux density uncertainty with increasing flux density. Note that $\mathrm{F_{\nu_o}}$

is not the same as $\widehat{F_{\nu_o\,2\,keV}}$. The main difference is that the frequency at which $\widehat{F_{\nu_o,\,2\,keV}}$ is measured changes with redshift, whereas F_{ν_o} always corresponds to the same observed frequency. However, in this work, it is assumed that equation 5.3 can be used to compute reasonable approximations of the uncertainties on $\widehat{F_{\nu_o,\,2\,keV}}$. Therefore, for each value of $F_{\nu_o,\,2\,keV}$, equation 5.3 was used to compute a corresponding uncertainty, $\delta F_{\nu_o,\,2\,keV}$, which, via equation 5.2, was then used to simulate measured X-ray flux densities ($\widehat{F_{\nu_o,\,2\,keV}}$). The fractional UV flux density uncertainty was set to be the same as the corresponding fractional X-ray flux density uncertainty. That is, for a given source, the UV flux density uncertainty is defined as

$$\delta F_{2{,}500\,\text{Å}} = \frac{\delta F_{2\,keV}}{F_{2\,keV}} \times F_{2{,}500\,\text{Å}}$$

It is noted here that while the X-ray flux density uncertainty model is expected to be a realistic prediction of the final *eROSITA* sample, the method used to simulate UV flux density uncertainties is considered a simplifying assumption, given that it is more difficult to predict in advance the UV flux density uncertainties.

The datasets simulated throughout this work can be said to exhibit heteroscedastic uncertainties due to the fact that the uncertainty model (equation 5.3, and subsequent variants) sets the fractional uncertainty to increase with decreasing flux density. Heteroscedasticity refers to the scenario where the standard deviation of a dependent variable is not constant across a range of values of a corresponding independent variable. Heteroscedastic uncertainties normally invalidate the assumptions built into conventional fitting routines. The LINMIX documentation does not appear to refer to the issue of heteroscedastic uncertainties directly, however, Kelly (2007) claim that the method they have developed allows for heteroscedastic uncertainties. It is therefore assumed that the LINMIX fitting routine is the most appropriate choice for analysing heteroscedastic datasets such as the one simulated in this work.

5.5 Measuring Ω_M and Ω_Λ Using the Quasar Hubble Diagram

Following the method discussed in section 1.6.2, the quasar $L_X - L_{UV}$ relationship, represented by equation 5.1, can be written in terms of flux densities as follows:

$$\text{Log}_{10}\left(\frac{4\pi D_L^2}{1+z}\,F_{\nu_o,\,2\,keV}\right) = \beta + \gamma \text{Log}_{10}\left(\frac{4\pi D_L^2}{1+z}\,F_{\nu_o,\,2{,}500\,\text{Å}}\right)$$

$$\hookrightarrow \text{Log}_{10}\left(\frac{F_{\nu_o,\,2\,keV}}{1+z}\right) = \beta' + \gamma \text{Log}_{10}\left(\frac{F_{\nu_o,\,2{,}500\,\text{Å}}}{1+z}\right) \tag{5.4}$$

The RL15 method of adopting quasars as standard candles requires a measurement of γ in order to compute distance modulus estimates. Throughout this work, γ was measured by

selecting sources from a given sample in the 0.5 - 2.5 redshift range and dividing them into redshift bins with width, $\Delta \mathrm{Log}_{10}(\mathrm{z})$, selected based on the sample size. The LINMIX fitting package (Kelly 2007) was used to fit equation 5.4 to the $\widehat{\mathrm{F}_{\nu_{\mathrm{o}},\,2\,\mathrm{keV}}} - \widehat{\mathrm{F}_{\nu_{\mathrm{o}},\,2{,}500\,\text{Å}}}$ relation in each bin using 10,000 iterations for each fit.

The inverse-variance weighted mean of the resulting best-fit γ_{z} values, $\overline{\gamma_{\mathrm{z}}}$, was then used to compute distance modulus estimates using

$$\mathrm{DM} = \frac{5}{2(\overline{\gamma_{\mathrm{z}}} - 1)} \left[\mathrm{Log}_{10} \left(\frac{\widehat{\mathrm{F}_{\nu_{\mathrm{o}},\,2\,\mathrm{keV}}}}{1+\mathrm{z}} \right) - \overline{\gamma_{\mathrm{z}}}\, \mathrm{Log}_{10} \left(\frac{\widehat{\mathrm{F}_{\nu_{\mathrm{o}},\,2{,}500\,\text{Å}}}}{1+\mathrm{z}} \right) \right] \tag{5.5}$$

Corresponding distance modulus uncertainties were estimated assuming that the uncertainties on $\widehat{\mathrm{F}_{\nu_{\mathrm{o}},\,2\,\mathrm{keV}}}$, $\widehat{\mathrm{F}_{\nu_{\mathrm{o}},\,2{,}500\,\text{Å}}}$, and $\overline{\gamma_z}$ are random and independent. Following the method adopted in section 1.6.3, the Markov chain Monte Carlo ensemble sampler EMCEE (Foreman-Mackey et al. 2013) was used to fit a standard flat ΛCDM cosmological model to the distance modulus-redshift relation in order to measure Ω_{M} and Ω_{Λ}. 10,000 steps were used for each EMCEE run with 100 walkers and a burn-in of 5,000 steps.

5.6 Forecast of the *eROSITA*-4MOST Quasar Cosmological Results

5.6.1 Attempting To Recover γ Using the eFEDs Uncertainty Model

In this section, the ability to measure γ using the simulated *eROSITA*-4MOST sample will be explored. Section 5.3 described the criteria used to select the subsample of *eROSITA* sources from the Comparat et al. (2019) mock catalogue that are in the German half of the *eROSITA* sky, and that can be expected to be detected by 4MOST at a signal-to-noise ratio that is approximately ≥ 10. This sample contained 317,391 sources. Following the method outlined in section 5.4, measurements of the X-ray and UV flux densities were simulated for this sample assuming the eFEDS uncertainty model (see equation 5.3). When simulating measured values for the flux densities using equation 5.2, a number of sources were assigned values of $\widehat{\mathrm{F}_{\nu}}$ that were less than, or equal to, zero. In addition, some sources had $\mathrm{F}_{\nu} \leq \delta \mathrm{F}_{\nu}$, which represent measurements that only provide upper limits. These sources were excluded from the analysis, which resulted in a sample of 315,627 sources, 269,258 of which are in the 0.5 - 2.5 redshift range. Following the method outlined in section 5.5, γ was measured for this sample (with $\Delta \mathrm{Log}_{10}(\mathrm{z}) = 0.007$), the results of which are shown in figure 5.5.

The resulting measurement of $\overline{\gamma_{\mathrm{z}}} = 0.582 \pm 0.0011$ is clearly inconsistent with the value of γ (0.6) used to simulate the sample. However, increasing the precision of the uncertainty model used to simulated the sample by varying the normalisation parameter, N, in a model

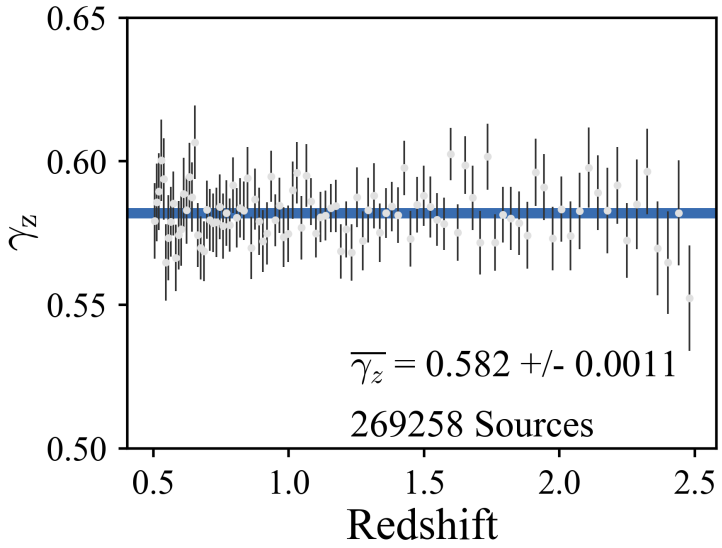


Figure 5.5: Best-fit X-ray-UV slope, γ_z (see equation 5.4), as a function of redshift obtained from fitting sources in the 0.5 - 2.5 redshift range from the main simulated *eROSITA*-4MOST sample using LINMIX. These results were obtained assuming a flux density uncertainties model representative of eFEDS observations. The weighted mean of the γ_z measurements (illustrated by the horizontal blue line) and corresponding 1σ uncertainties are also shown.

of the form

$$\mathrm{Log}_{10}\left(\frac{\delta \mathrm{F}_{\nu_o}}{\mathrm{F}_{\nu_o}}\right) = -0.53\mathrm{Log}_{10}(\mathrm{F}_{\nu_o}) - \mathrm{N} \tag{5.6}$$

was found to result in an increasingly accurate measurement of γ (see table 5.1). This biasing of the best-fit slope is discussed in further detail in section 5.7.6. Note that the number of sources included in the fit increases with increasing N, since the precision of the uncertainty model (set by N) determines how many simulated sources will have $\widehat{\mathrm{F}_\nu} \leq 0$ and/or $\mathrm{F}_\nu \leq \delta \mathrm{F}_\nu$, and therefore be excluded from the fit.

The eFEDS uncertainty model is expected to be representative of the precision of the complete *eROSITA* all-sky survey (eRASS:8). However, due to the survey strategy, the survey sensitivity will be deeper towards the ecliptic poles. These regions of higher sensitivity will provide sources with more precise flux density measurements than those from the rest of the sky. From the results presented in table 5.1, it appears that the following flux density uncertainty model

$$\mathrm{Log}_{10}\left(\frac{\delta \mathrm{F}_{\nu_o}}{\mathrm{F}_{\nu_o}}\right) = -0.53\mathrm{Log}_{10}(\mathrm{F}_{\nu_o}) - 17.85 \tag{5.7}$$

represents the minimum required precision in order to recover a measurement of γ consistent with the input value of 0.6 using this sample. A subsample of sources with a

Table 5.1: Weighted mean best-fit measurements of γ_z from fitting the main simulated *eROSITA*-4MOST sample using a series of different values of the normalisation of the uncertainty model, N (see equation 5.6).

N	$\overline{\gamma_z}$	No. of Sources
17.27	0.5820±0.0011	269,258
17.5	0.5951±0.0011	270,605
17.6	0.5971±0.0011	270,664
17.7	0.5982±0.0011	270,674
17.85	0.5990±0.0011	270,677

flux density precision characterised by equation 5.7 can be expected to return an accurate measurement of γ and therefore be used to accurately measure Ω_M and Ω_Λ. The following sections explore how to identify this higher-precision subsample and provide a forecast of the cosmological constraints that it can be expected to provide.

5.6.2 Determining the Relationship Between Exposure Time and Flux Density Precision

It was necessary to determine what percentage of the main simulated *eROSITA*-4MOST sample can be expected to have flux density measurements with the level of precision characterised by equation 5.7. To achieve this, the eFEDS catalogue was used to determine the relationship between the exposure time of an observation and the resulting X-ray flux density precision. The eFEDS catalogue contained the exposure time in the 0.6 - 2.3 keV band ("ML_EXP_2"). "ML_EXP_2" is the exposure time of the band from which the flux density estimates were derived and will be referred to as T_{exp}.

Figure 5.6 shows that the fractional X-ray flux density uncertainty distribution shifts with T_{exp}. Therefore, assuming all else being equal, a higher exposure time will result in more precise flux density measurements, which will in turn result in a more accurate measurement of γ. The eFEDS sample was divided into 12 T_{exp} bins between 300 s and 3,500 s with a linear bin width of 250 s (in the manner shown by figure 5.6).

The sources in each T_{exp} bin were fit with equation 5.6 in order to determine the normalisation, N, of this model for each T_{exp} bin (see figure 5.6). The best-fit value for N was then plotted against the mean T_{exp} for the corresponding bin ($\overline{T}_{exp}$, see figure 5.7). The $N - \overline{T}_{exp}$ distribution was then fit with the following model:

$$N = \overline{T}_{exp}^{0.1} + 15.14 \tag{5.8}$$

which is shown by the grey line in figure 5.7. The $N-\overline{T}_{exp}$ distribution was also extrapolated to higher exposure times (shown by the blue line in figure 5.7). Equation 5.7 represents the

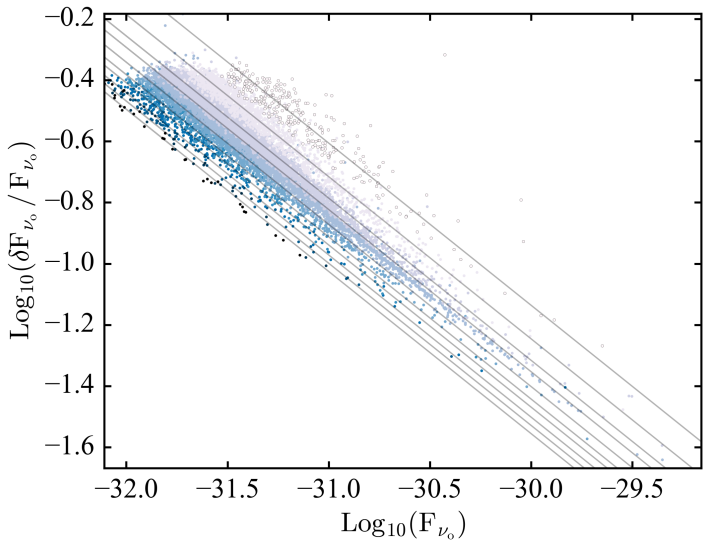


Figure 5.6: Fractional flux density uncertainty as a function of flux density for the eFEDS sample. The colours represent increasing $\mathrm{T}_{\mathrm{exp}}$, light to dark colours representing low to high $\mathrm{T}_{\mathrm{exp}}$, respectively. The grey lines represent the best-fit models (equation 5.6) to each exposure time bin.

minimum precision required in order to recover the input value of γ, and requires N=17.85. It was determined using equation 5.8 that $\mathrm{N} = 17.85$ is achieved at $\mathrm{T}_{\mathrm{exp}} = 15{,}288\,\mathrm{s}$. The linear extrapolation however suggests that $\mathrm{N} = 17.85$ is achieved at $\mathrm{T}_{\mathrm{exp}} = 8{,}470\,\mathrm{s}$. Since few sources are expected to have eRASS:8 exposure times above 15 ks, the remainder of this analysis will explore the scenario where the required flux density precision is reached at $\mathrm{T}_{\mathrm{exp}} = 8{,}470\,\mathrm{s}$.

It was then necessary to determine the fraction of the sky that can be expected to receive eRASS:8 observations with exposure times $\geq 8{,}470\,\mathrm{s}$. The eRASS:1 exposure map[2] was downloaded from the MPE DATool[3]. The *eROSITA*-4MOST area was selected using the criteria outlined in section 5.3. The eRASS:1 exposure times were then multiplied by a factor of eight in order to provide an approximation of the eRASS:8 exposure map (see figure 5.8). Using this map, it was determined that ∼0.8 per cent of the eRASS:8 sky can be expected to have exposure times $\geq 8{,}470\,\mathrm{s}$.

[2] Map_eRASS1_exp_band022_946.fits
[3] https://erosita.mpe.mpg.de/DATool/

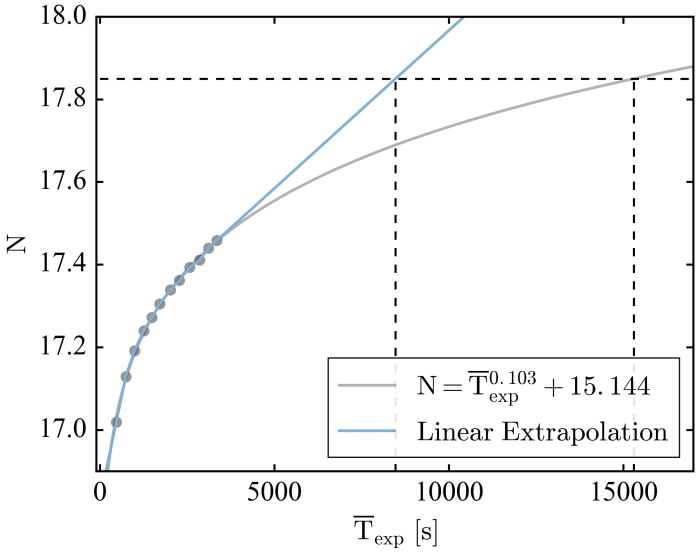


Figure 5.7: Normalisation of the fractional X-ray flux density uncertainty model, N, versus mean exposure time per bin, $\overline{T}_{exp}$, for the eFEDS sample. The grey line shows the best-fit model represented by equation 5.8. The blue line shows the linear extrapolation of the distribution. The horizontal line shows the value of N required to recover the input value of γ (see equation 5.7). The vertical lines show the values of $\overline{T}_{exp}$ at which both models predict $N = 17.85$ will be achieved.

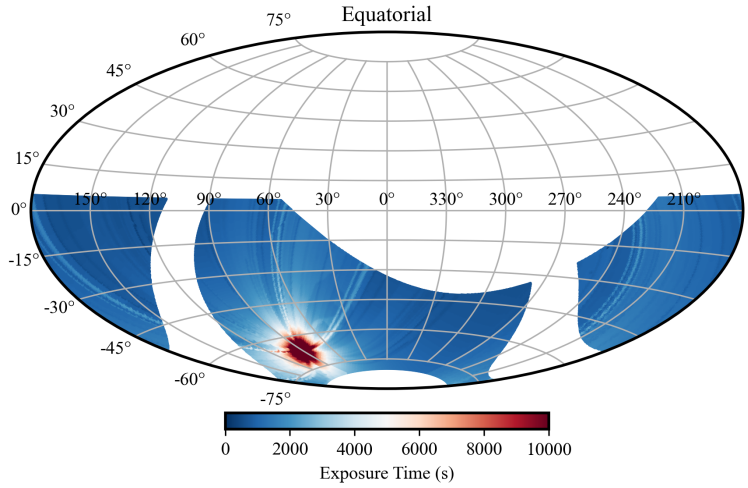


Figure 5.8: Estimated exposure map for the eRASS:8 survey over the *eROSITA*-4MOST sky area (equatorial coordinates, Aitoff projection). The exposure times shown are for the 0.6 - 2.3 keV band.

5.6.3 Results Using the Higher-Precision Subsample

From the analysis presented in the previous section approximately 0.8 per cent of the *eROSITA*-4MOST area is expected to have flux densities with precision sufficient for quasar cosmology by the end of eRASS:8. From the main simulated *eROSITA*-4MOST sample of 317,391 sources, a single, unique subsample of 2,539 sources ($\sim$0.8 per cent of the main sample) was randomly selected. The sources in this subsample were not forced to be distributed uniformly around the parameter space occupied by the parent sample. Therefore, measurements of γ, Ω_M, and Ω_Λ can be expected to vary between different randomly selected subsamples (see section 5.7.4).

This subsample of 2,539 sources was then used to measure γ following the method outlined in section 5.5 with $\Delta\mathrm{Log}_{10}(z) = 0.03$, the results of which are shown in the left panel of figure 5.9. Using the weighted mean of the best-fit γ_z measurements, distance modulus estimates were derived and the distance modulus-redshift relation was then fit using EMCEE, with the resulting Ω_M and Ω_Λ confidence contours shown in the right panel of figure 5.9. Note that the confidence contours used throughout this work represent the 68 and 95 per cent confidence levels, rather than the 1 and 2 σ levels, in order to facilitate a direct comparison of these results with the previously published literature on this subject.

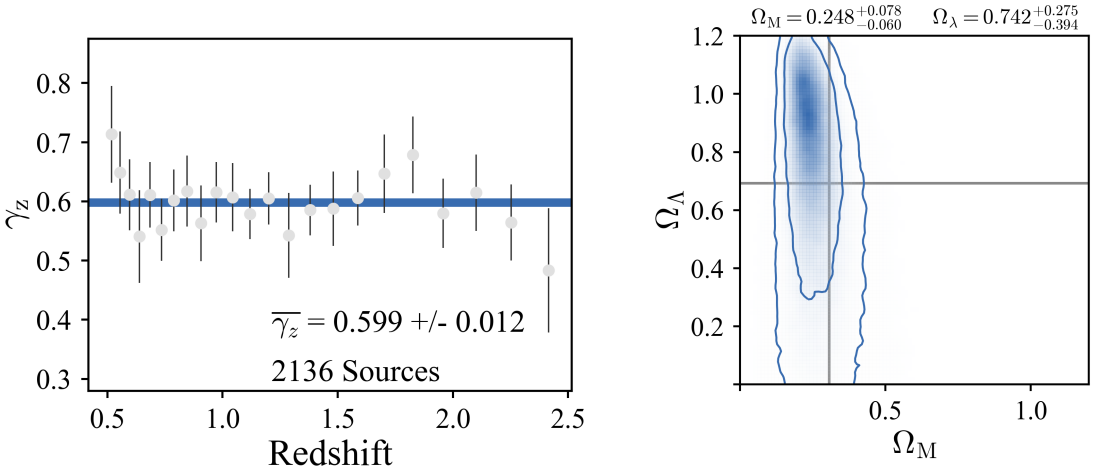


Figure 5.9: Results obtained from the simulated subsample of *eROSITA*-4MOST sources with high-precision flux density measurements described in section 5.6.3. Left panel: The best-fit X-ray-UV slope, γ_z (see equation 5.4), as a function of redshift obtained from fitting sources in the 0.5 - 2.5 redshift range using LINMIX. The weighted mean of the γ_z measurements (illustrated by the horizontal blue line) and corresponding 1σ uncertainties are also shown. Right panel: The constraints on Ω_M and Ω_Λ obtained from fitting the distance modulus-redshift relation using EMCEE. The contours show the 68 and 95 per cent confidence levels. The grey crosshairs show the values of Ω_M and Ω_Λ used to create the mock ($\Omega_M = 0.307115$ and $\Omega_\Lambda = 0.692885$). The best-fit results and corresponding 1σ uncertainties are also shown.

The cosmological constraints derived using this simulated higher-precision *eROSITA*-4MOST subsample ($\Omega_M = 0.253^{+0.079}_{-0.061}$ and $\Omega_\Lambda = 0.751^{+0.271}_{-0.393}$) are consistent with the values used to produce this simulation ($\Omega_M = 0.307115$ and $\Omega_\Lambda = 0.692885$), and provide a level of precision comparable to that presented by RL15 (see table 5.2). While the forecasts provided here for this *eROSITA*-4MOST subsample naturally fall short of the precision of the combined EE+lowE power spectra, and combined *Planck* cosmic microwave background (CMB) power spectra, CMB lensing, and baryon acoustic oscillations (BAO) constraints (Planck Collaboration et al. 2018) shown in table 5.2, it is clear that such a sample would provide a relevant and complementary set of standard candles.

Table 5.2: Comparison of the constraints on Ω_M and Ω_Λ from different datasets. The *eROSITA*-4MOST constraints are derived from the simulated subsample of *eROSITA*-4MOST sources with high-precision flux density measurements described in section 5.6.3. The RL15 results, the *Planck* collaboration's 2018 results using the *Planck* CMB power spectra, and the results from the combination of the *Planck* CMB power spectra, CMB lensing, and BAO are also shown for comparison (Planck Collaboration et al. 2018).

Dataset	Ω_M	Ω_Λ	Reference
RL15	$0.22^{+0.10}_{-0.08}$	$0.92^{+0.18}_{-0.30}$	Risaliti & Lusso (2015)
eROSITA-4MOST	$0.250^{+0.078}_{-0.060}$	$0.758^{+0.262}_{-0.387}$	this work
EE, lowE	$0.289^{+0.026}_{-0.033}$	$0.711^{+0.033}_{-0.026}$	Planck Collab. (2018)
TT, TE, EE, lowE, lensing, BAO	$0.3111^{+0.0056}_{-0.0056}$	$0.6889^{+0.0056}_{-0.0056}$	Planck Collab. (2018)

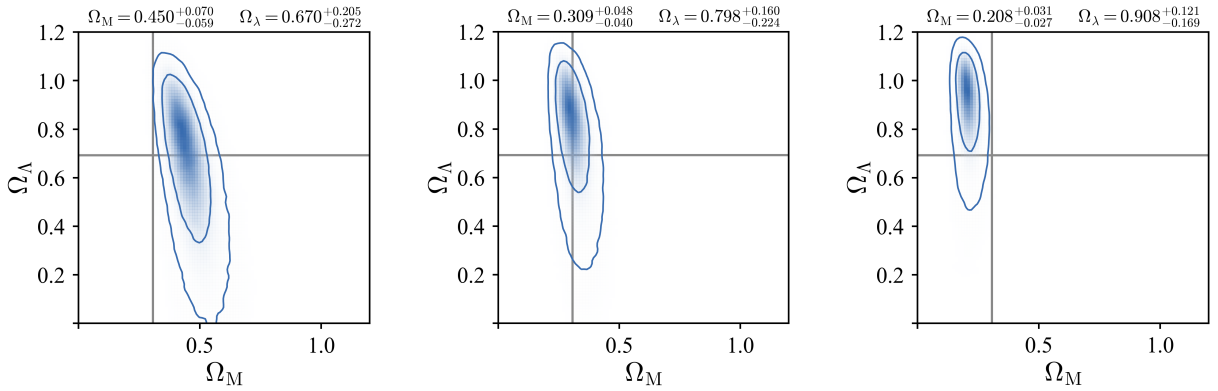
5.7 Exploring the Quasar Cosmology Method Using the *eROSITA*-4MOST Simulated Data

The simulated *eROSITA*-4MOST quasar data presented in this work offers an opportunity to examine the effect that the properties of a given dataset have on the resulting cosmological constraints. In the following sections, the simulation developed in this work will be used to explore a number of aspects of the quasar cosmology method in more detail, including the necessity of accurately measuring γ (section 5.7.1), the results obtained considering different levels of intrinsic scatter in the $\mathrm{Log}_{10}(L_{\nu_e,2{,}500\,\text{Å}}) - \mathrm{Log}_{10}(L_{\nu_e,2\,\mathrm{keV}})$ relationship (section 5.7.2), the influence of sample size (section 5.7.3), the effect of using different subsamples (section 5.7.4), the stability of the results using different fitting methods (section 5.7.5), and the effects of regression dilution (section 5.7.6).

5.7.1 The Influence of γ Accuracy on the Measurements of Ω_M and Ω_Λ

From the results presented in table 5.2, it is clear that the ability to recover an accurate measurement of γ is strongly influenced by the flux density precision. In turn, the accuracy of the γ measurement used to compute the distance modulus values (using equation 5.5) directly determines the accuracy of the resulting cosmological constraints. To explore this effect, 10,000 sources randomly selected from the main sample were simulated following the method outlined in section 5.4, assuming a standard deviation of the $\mathrm{Log}_{10}(\mathrm{L}_{\nu_e,\,2500\,\text{Å}}) - \mathrm{Log}_{10}(\mathrm{L}_{\nu_e,\,2\,\mathrm{keV}})$ relation intrinsic scatter, σ, of 0.3, and an uncertainty model defined by equation 5.7. Three sets of distance modulus estimates were extracted from this sample with γ set to be equal to 0.59, 0.6, and 0.61, and γ uncertainties set to a precision that is representative of what would be obtained using a sample of this size.

Confidence contours derived using the reliable subsample of 2,539 sources (see section 5.6.3) for $\gamma = 0.59$ have significant overlap with the boundary of the parameter space defined by the prior (equation 1.11). A sample containing 10,000 sources was therefore chosen for this experiment instead since, for this sample size, the overlap is minimal which facilitates a clearer comparison of the confidence intervals for each value of γ. The distance modulus-redshift relation was then fit for each of these samples following the method outlined in section 5.5, and the resulting Ω_M and Ω_Λ confidence contours are shown in figure 5.10.



(a) $\gamma = 0.590 \pm 0.004$
10,000 sources

(b) $\gamma = 0.6 \pm 0.004$
10,000 sources

(c) $\gamma = 0.61 \pm 0.004$
10,000 sources

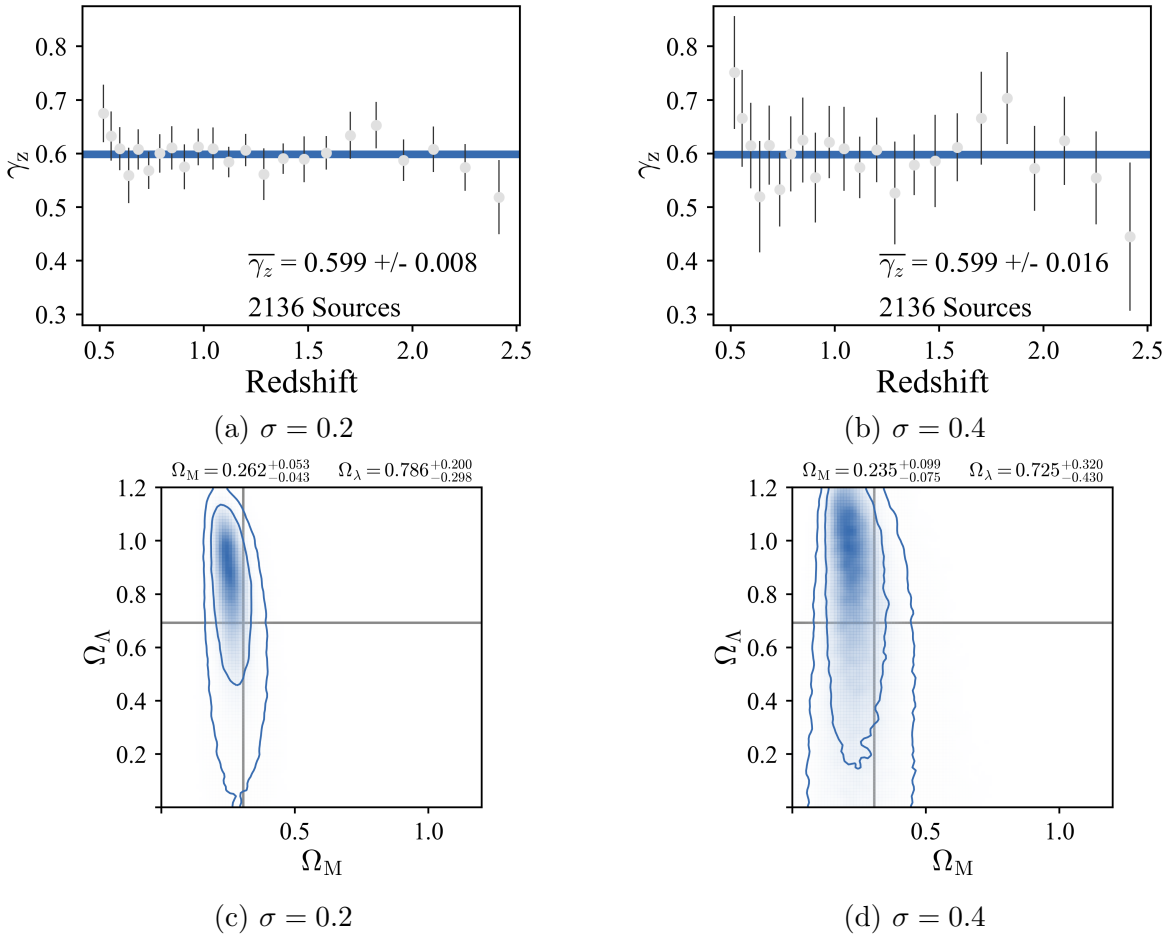
Figure 5.10: Constraints on Ω_M and Ω_Λ obtained from fitting the distance modulus-redshift relation of 10,000 sources randomly selected from the main simulated *eROSITA*-4MOST sample using EMCEE. Each panel represents a different set of distance modulus estimates derived using different values of γ. The contours show the 68 and 95 per cent confidence levels. The grey crosshairs show the values of Ω_M and Ω_Λ used to create the mock ($\Omega_M = 0.307115$ and $\Omega_\Lambda = 0.692885$). The best-fit results and corresponding 1σ uncertainties are shown above each panel.

From figure 5.10, it is clear that using an inconsistent measurement of γ to compute distance modulus estimates will result in inconsistent Ω_{M} and Ω_{Λ} measurements, but when the true value of γ is used, consistent measurements of Ω_{M} and Ω_{Λ} are recovered. Recent work has suggested that the cosmological constraints provided by quasar standard candles are inconsistent with the ΛCDM cosmological model (Risaliti & Lusso (2019), see also Lusso et al. (2019); Yang et al. (2020)). The effect illustrated by figure 5.10 suggests that this deviation may be caused by an inconsistent measurement of γ.

5.7.2 The Intrinsic Scatter in the $\mathrm{Log_{10}(L_{UV})} - \mathrm{Log_{10}(L_X)}$ Relationship

The intrinsic scatter in the quasar $\mathrm{Log_{10}(L_{\nu_e,\,2{,}500\,\text{Å}})} - \mathrm{Log_{10}(L_{\nu_e,\,2\,keV})}$ relationship is not known precisely (however see Lusso & Risaliti 2016) and can be expected to influence the measurement of γ, Ω_{M}, and Ω_{Λ}. In order to quantify this, two iterations of the reliable subsample of 2,539 sources were produced following the method outlined in section 5.4 using two additional values of the scatter in the intrinsic $\mathrm{Log_{10}(L_{UV})} - \mathrm{Log_{10}(L_X)}$ relation ($\sigma = 0.2$ and $\sigma = 0.4$). Following the method outlined in section 5.5, γ_z was measured for both samples with $\Delta\mathrm{Log_{10}(z)} = 0.03$ and the resulting distance modulus-redshift relation was then fit using EMCEE, the results of which are shown in figure 5.11.

Considering both figures 5.11 and 5.9, it is apparent that increasing σ reduces the precision of $\overline{\gamma_z}$, but does not affect its accuracy. Of significant note is that even in the case of an extreme level of dispersion in the $\mathrm{Log_{10}(L_{\nu_e,\,2{,}500\,\text{Å}})} - \mathrm{Log_{10}(L_{\nu_e,\,2\,keV})}$ relation ($\sigma = 0.4$), the relatively precise flux density measurements of this subsample allow a consistent measurement of γ to be made. Varying σ had the same effect on the measurements of Ω_{M} and Ω_{Λ}, where the precision of those measurements decreased as σ increased, and consistent measurements of Ω_{M} and Ω_{Λ} were recovered even in the case of large values of σ.



(a) $\sigma = 0.2$ (b) $\sigma = 0.4$

(c) $\sigma = 0.2$ (d) $\sigma = 0.4$

Figure 5.11: Results obtained using different iterations of the reliable subsample simulated using two values for the standard deviation of the $\mathrm{Log_{10}(L_{\nu_e,\,2500\,\text{Å}})} - \mathrm{Log_{10}(L_{\nu_e,\,2\,keV})}$ relation intrinsic scatter ($\sigma = 0.2$ and 0.4). Upper panels: Best-fit X-ray-UV slope, γ_z (see equation 5.4), as a function of redshift obtained from fitting sources in the 0.5 - 2.5 redshift range from both samples using LINMIX. The weighted mean of the γ_z measurements (illustrated by the horizontal blue line) and corresponding 1σ uncertainties are also shown. Lower panels: Constraints on Ω_M and Ω_Λ obtained from fitting the distance modulus-redshift relation of both samples using EMCEE. The contours show the 68 and 95 per cent confidence levels. The grey crosshairs show the values of Ω_M and Ω_Λ used to create the mock ($\Omega_M = 0.307115$ and $\Omega_\Lambda = 0.692885$). The best-fit results and corresponding 1σ uncertainties are also shown.

5.7.3 The Influence of Sample Size on the Measurement of Ω_M and Ω_Λ

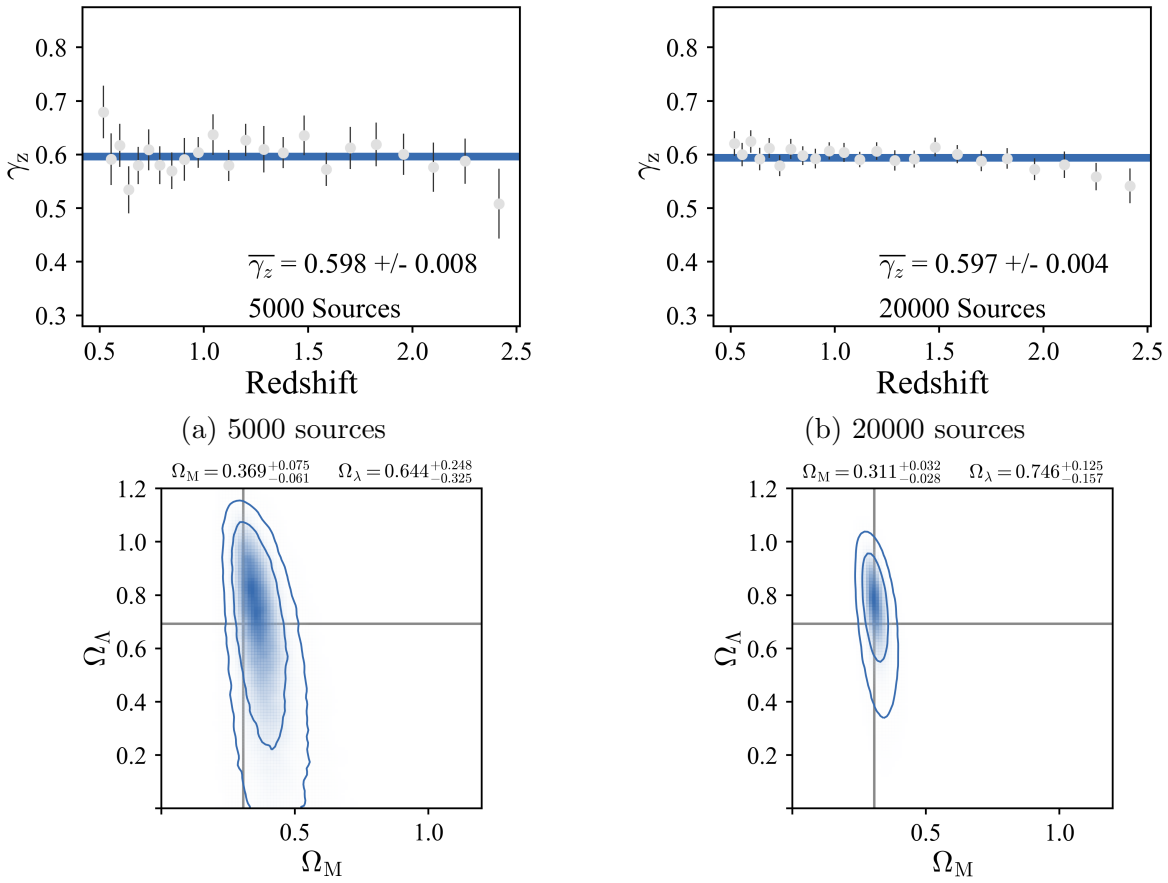
The results presented in section 5.6.3 aim to provide a forecast of the cosmological constraints that may be achieved using the *eROSITA*-4MOST quasar sample that will be available at the end of eRASS:8. However, it may be the case that the reliable eRASS:8 sample is larger than what was predicted in section 5.6.3. For example, the minimum 4MOST signal-to-noise ratio required in order to achieve reliable UV flux density measurements may be lower than the value of 10 assumed in this work. This would result in a larger sample size than that adopted in section 5.6.3.

It is also possible that the flux density precision of the final eRASS:8 sample will be higher than what was assumed in section 5.6.3, which would also increase the size of the reliable subsample. Another consideration is that additional *eROSITA*-4MOST sources may become available from future *eROSITA* observing campaigns. Given that these future datasets are expected to be made available, the simulations developed in this work were used to quantify the corresponding improvement in the Ω_M and Ω_Λ constraints in order to determine whether future work incorporating these samples into the quasar Hubble diagram is justified. In light of these considerations, in this section the improvement in the constraints on Ω_M and Ω_Λ with increasing sample size will be explored.

Following the method outlined in section 5.4, two quasar samples were simulated assuming a standard deviation of the $\mathrm{Log_{10}}(\mathrm{L}_{\nu_e,\,2500\,\text{Å}}) - \mathrm{Log_{10}}(\mathrm{L}_{\nu_e,\,2\,\mathrm{keV}})$ relation intrinsic scatter of $\sigma = 0.3$, an uncertainty model defined by equation 5.7, and with 5,000 and 20,000 sources in the 0.5 - 2.5 redshift range[4]. Following the method outlined in section 5.5, γ was measured for both samples with $\Delta\mathrm{Log_{10}(z)} = 0.03$ and the resulting distance modulus-redshift relation was then fit using EMCEE, the results of which are shown in figure 5.12.

Comparing figures 5.12 and 5.9 it can be seen that there is a consistent increase in the precision of the γ, Ω_M, and Ω_Λ measurements with increasing sample size. There are also differences in the accuracy of the Ω_M and Ω_Λ measurements between the three samples. This, however, is to be expected; when comparing any number of different samples, there will be differences in the absolute best-fit measurements of Ω_M and Ω_Λ that will not necessarily be the result of differing sample size (see section 5.7.4).

[4] γ was measured throughout this work in the 0.5 - 2.5 redshift range. Note that when measuring Ω_M and Ω_Λ, sources in the sample are included in the fit regardless of their redshift.



(a) 5000 sources

(b) 20000 sources

(c) 5871 sources

(d) 23404 sources

Figure 5.12: Results obtained from two simulated *eROSITA*-4MOST samples illustrating the effect of increasing sample size. Upper panels: Best-fit X-ray-UV slope, γ_z (see equation 5.4), as a function of redshift obtained from fitting sources in the 0.5 - 2.5 redshift range from both samples using LINMIX. The weighted mean of the γ_z measurements (illustrated by the horizontal blue line) and corresponding 1σ uncertainties are also shown. Lower panels: Constraints on Ω_M and Ω_Λ obtained from fitting the distance modulus-redshift relation of both samples using EMCEE. The contours show the 68 and 95 per cent confidence levels. The grey crosshairs show the values of Ω_M and Ω_Λ used to create the mock ($\Omega_M = 0.307115$ and $\Omega_\Lambda = 0.692885$). The best-fit results and corresponding 1σ uncertainties are also shown.

5.7.4 Comparing the Results From Different Subsamples

The main *eROSITA*-4MOST sample contains 317,391 sources (see section 5.3). When selecting subsamples for analysis throughout this work, sources are randomly drawn from this main sample and are not forced to be distributed uniformly around the parameter space occupied by the main sample. Best-fit measurements of Ω_{M} and Ω_{Λ} can therefore be expected to vary between these subsamples. In order to illustrate this, four different subsamples, each with a sample size that is $\sim$0.8 per cent that of the main sample, were selected. This sample size corresponds to that of the high-precision subsample discussed in section 5.6.3. Following the method outlined in section 5.4, $\mathrm{L}_{\nu_{\mathrm{e}},\,2500\,\text{Å}}$ and $\mathrm{L}_{\nu_{\mathrm{e}},\,2\,\mathrm{keV}}$ values were simulated for each sample assuming a standard deviation of the $\mathrm{Log}_{10}(\mathrm{L}_{\nu_{\mathrm{e}},\,2500\,\text{Å}}) - \mathrm{Log}_{10}(\mathrm{L}_{\nu_{\mathrm{e}},\,2\,\mathrm{keV}})$ relation intrinsic scatter of $\sigma = 0.3$ and an uncertainty model defined by equation 5.7.

In order to determine the effect specifically due to the differences in source distribution between the samples, distance modulus estimates were computed for each sample assuming $\gamma = 0.6 \pm 0.01$. Ω_{M} and Ω_{Λ} were measured for each sample following the method outlined in section 5.5, and the resulting $\Omega_{\mathrm{M}} - \Omega_{\Lambda}$ confidence contours are shown in figure 5.13. From figure 5.13, it can be seen that different randomly selected subsamples result in different best-fit results for Ω_{M} and Ω_{Λ}, some of which are not consistent with the input values.

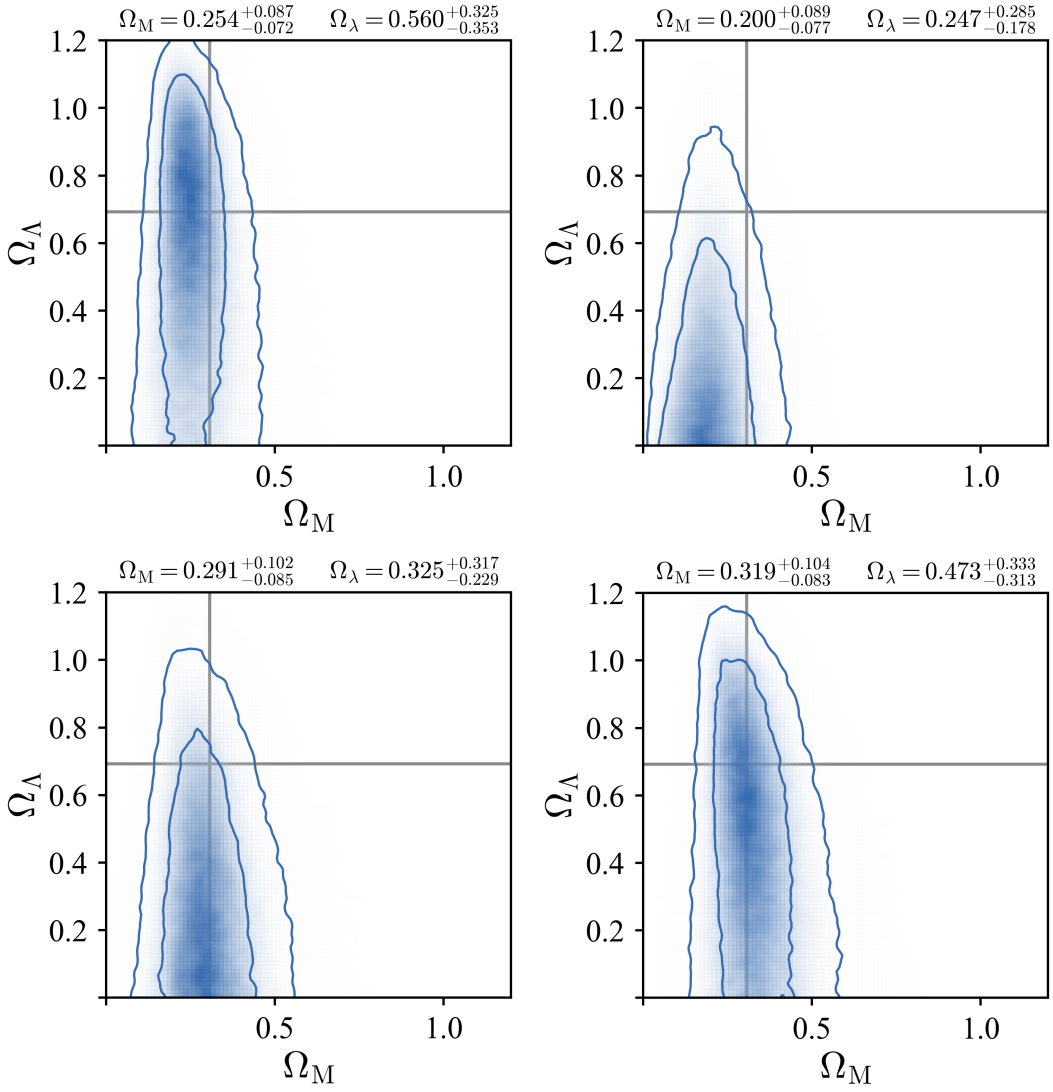


Figure 5.13: Constraints on Ω_M and Ω_Λ obtained from fitting the distance modulus-redshift relation using EMCEE. The different panels represent results obtained from a set of unique, randomly selected, simulated subsamples, each of which are 0.8 per cent the size of the main sample. The contours show the 68 and 95 per cent confidence levels. The grey crosshairs show the values of Ω_M and Ω_Λ used to create the mock ($\Omega_M = 0.307115$ and $\Omega_\Lambda = 0.692885$). The best-fit results and corresponding 1σ uncertainties are shown above each panel.

5.7.5 Are the Measurements of γ Made Using Different Regression Methods Consistent?

From a number of possible options, LINMIX (Kelly 2007) was selected as the most appropriate regression package for the measurement of the X-ray-UV slope, γ, throughout this work. This was due to the fact that LINMIX allows for heteroscedastic uncertainties and accounts for uncertainties in both the dependent and independent variables. However, it is important to determine whether the ability to recover an accurate measurement of γ using the reliable subsample described in section 5.6.3 is independent of the fitting method used.

In order to test this, γ was measured for the reliable subsample of 2,136 simulated *eROSITA*-4MOST sources in the 0.5 - 2.5 redshift range (see section 5.6.3) using the Markov chain Monte Carlo ensemble sampler EMCEE fitting package (Foreman-Mackey et al. 2013). This analysis followed a method similar to that used when fitting with LINMIX; the sample was divided into a series of redshift bins with a bin width $\Delta\mathrm{Log}_{10}(z) = 0.03$, EMCEE was then used to fit a linear model to the $\widehat{F_{\nu_o,\,2\,\mathrm{keV}}} - \widehat{F_{\nu_o,\,2{,}500\,\text{Å}}}$ distribution for each of these redshift bins with 20,000 steps, 200 walkers, and a burn-in fraction of 0.5. The resulting best-fit measurements of γ as a function of redshift are presented in figure 5.14. A comparison of figures 5.9 and 5.14 demonstrates that LINMIX and EMCEE return best-fit measurements of γ that are both consistent with each other, and consistent with the input value of 0.6.

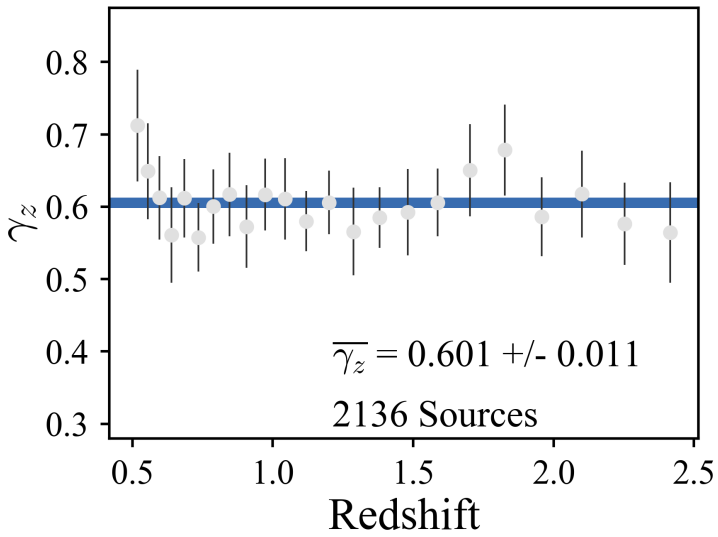


Figure 5.14: Best-fit X-ray-UV slope, γ_z (see equation 5.4), as a function of redshift obtained from fitting sources in the 0.5 - 2.5 redshift range from the simulated reliable subsample described in section 5.6.3 using EMCEE. The weighted mean of the γ_z measurements (illustrated by the horizontal blue line) and corresponding 1σ uncertainties are also shown.

5.7.6 What Dataset Properties Influence Regression Dilution?

The results presented in table 5.1 illustrate how the measurement of the X-ray - UV slope, γ, becomes increasingly biased towards zero as the uncertainty precision is decreased. This behaviour appears to be the result of an effect known as regression dilution. Regression dilution is typically understood as the scenario where the best-fit slope estimate derived using a linear regression routine is biased towards zero due to the presence of uncertainties in the independent variable (e.g. Frost & Thompson 2000). The issue of slope biasing under certain conditions affects linear regression in general, for example, the Akritas & Bershady (1996) BCES routine (Graham 2016) and the OLS routine (Kelly 2007). Therefore, it is perhaps to be expected that the LINMIX fitting routine used in this work will be similarly affected.

For a given linear regression routine an expression can be derived for the degree to which the best-fit slope will be biased compared to the true value (for example, the OLS case is presented in section 3 of Kelly (2007)). This expression could then be used to correct the best-fit slope for the effect of regression dilution. However, due to the complexity of the LINMIX fitting routine, the derivation of an analogous expression for LINMIX as part of this work was not considered feasible. Though a corresponding expression is not available for LINMIX, the example of the OLS expression for slope biasing offers an indication of the nature of the issue that may affect LINMIX; namely, that the magnitude of the slope bias is related to specific properties of a given dataset; for example, the uncertainty variance and covariance.

A biased measurement of γ can in turn significantly reduce the accuracy of the cosmological constraints that a given sample can provide (as was demonstrated in section 5.7.1). Therefore, the simulation presented in this work was used to empirically explore the issue of regression dilution in LINMIX. This involved examining how adjusting the uncertainty model used to simulate the data affects the LINMIX measurement of γ. Apart from the changes made to the uncertainty model, the results discussed in this section were derived from data that was simulated and fit following the method outlined in sections 5.4 and 5.5, with 100 iterations used for each LINMIX fit. The absence of a mathematical model for regression dilution in LINMIX means that it is not known in advance what properties of a dataset influence the biasing of the best-fit slope. Therefore, using the OLS slope biasing expression (Kelly 2007) as a guide, a set of parameters describing various dataset properties was chosen as potentially being related to regression dilution in LINMIX. For each of the tests discussed below, these parameters were calculated and presented along with the best-fit measurement of γ.

In the first of these experiments the original X-ray uncertainty model was retained, while the UV flux density uncertainties were assigned values of zero. The second experiment was the inverse of the first, with the original UV uncertainty model being retained and the X-ray uncertainties set to zero. The results of these two experiments are shown in sections 1 and 2 of table 5.3. It can be seen from section 1 of table 5.3 that the biasing of $\overline{\gamma_z}$ with decreasing uncertainty precision (traced by the parameter N) is still present even after UV flux density uncertainties are set to zero. However, section 2 of table 5.3 shows that setting

the X-ray flux density uncertainties to zero largely negates the effect of regression dilution.

In the course of this investigation into regression dilution, two additional tests were carried out, the first of which involved retaining the original X-ray uncertainty model while setting the UV flux density uncertainties to be a constant fractional value of $\delta F_{\nu_o,\,2500\,\text{Å}} = 0.1 \times F_{\nu_o,\,2500\,\text{Å}}$. The second test was the inverse, where the original UV uncertainty model was retained and the X-ray flux density uncertainties were set to a constant fractional value of $\delta F_{\nu_o,\,2\,\text{keV}} = 0.1 \times F_{\nu_o,\,2\,\text{keV}}$. The results of these tests are shown in sections 3 and 4 of table 5.3. Once again, it is clear that while changing the UV flux density uncertainties did not remove the effect of regression dilution, setting the X-ray flux density uncertainties to a constant fractional value resulted in largely unbiased measurements of γ. A final test was carried out in which the parameter λ, representing the slope of the X-ray flux density uncertainty model in the following equation:

$$\text{Log}_{10}\left(\frac{\delta F_{\nu_o,\,2\,\text{keV}}}{F_{\nu_o,\,2\,\text{keV}}}\right) = -\lambda \text{Log}_{10}(F_{\nu_o,\,2\,\text{keV}}) - 17.5, \tag{5.9}$$

was varied, the results of which are shown in table 5.4.

The parameters describing the dataset properties listed in tables 5.3 and 5.4 were compared in order to identify whether any one parameter predicts the degree of slope biasing. From this comparison it was noted that LINMIX appears to behave in a way that is similar to, but not the same as, OLS, with the slope bias appearing to be proportional to the variance in the dependent variable uncertainties ($\overline{\sigma}^2_{\delta F_X}$). In most of the tests discussed here, the best-fit slope decreases as $\overline{\sigma}^2_{\delta F_X}$ increases. There are only two exceptions to this; the case of N=17.27 in sections 2 and 4 of table 5.3, where, at the lowest uncertainty precision, it appears that an additional unknown effect is in play. This work does not present a mathematical demonstration of which parameters determine the magnitude of regression dilution in LINMIX. However, the tests discussed here can be considered an empirical demonstration that the LINMIX fitting routine suffers from regression dilution that is proportional to the variance in the dependent variable uncertainties.

5.8 Quasar Cosmology: An Epilogue

5.8.1 Conclusions

The work presented in this chapter explored the application of the forthcoming *eROSITA*-4MOST quasar sample to the RL15 method for implementing quasars as standard candles. Using the *eROSITA* mock presented by Comparat et al. (2019) as a starting point, simulated measurements of quasar X-ray and UV flux densities following the known quasar $L_X - L_{UV}$ relationship were produced. From this simulated *eROSITA*-4MOST dataset the subsample of sources that are expected to have eRASS:8 flux density measurements at a quality sufficient for use in the quasar cosmology experiment was identified. A method that may be used to select this subsample using the observation exposure time was presented. This simulated subsample was then used to forecast the measurements of the cosmological

Table 5.3: Results from fitting the X-ray - UV relation for a series of different realisations of the main simulated *eROSITA - 4MOST* sample. Section 1 presents results obtained from a sample simulated with the UV flux density uncertainties set to zero. Section 2 presents results from a sample simulated with the X-ray flux density uncertainties set to zero. Section 3 presents results from a sample simulated with UV flux density uncertainties set to a constant fractional value. Section 4 presents results from a sample simulated with X-ray flux density uncertainties set to a constant fractional value. For each section, a series of values of the normalisation of the uncertainty model, N, (see equation 5.6) were used. For each fit, a redshift bin width of $\Delta \mathrm{Log}_{10}(\mathrm{z}) = 0.007$ was adopted.

N	$\overline{\gamma_{\mathrm{z}}}$	$\overline{\sigma}^2_{\delta \mathrm{F_{UV}}}$	$\overline{\sigma}^2_{\delta \mathrm{F_X}}$	$\mathrm{Var(F_{UV})}$	$\sigma_{\delta \mathrm{F_{UV}}, \delta \mathrm{F_X}}$	$\mathrm{Cov(F_{UV}, F_X)}$	No. of Sources
			(1) $\delta \mathrm{F}_{\nu_{\mathrm{o}},\, 2500\,\text{Å}} = 0$				
17.27	0.5837±0.001	0	0.002	0.292	0	0.175	269,949
17.5	0.5956±0.0011	0	0.0005	0.294	0	0.177	270,636
17.6	0.5973±0.0011	0	0.0003	0.294	0	0.177	270,668
17.7	0.5981±0.0011	0	0.0002	0.294	0	0.176	270,676
17.85	0.5992±0.0011	0	0.00009	0.294	0	0.176	270,677
			(2) $\delta \mathrm{F}_{\nu_{\mathrm{o}},\, 2\,\mathrm{keV}} = 0$				
17.27	0.5961±0.0011	0.002	0	0.300	0	0.175	269,900
17.5	0.599±0.0011	0.0005	0	0.297	0	0.177	270,640
17.6	0.5992±0.0011	0.0003	0	0.296	0	0.177	270,673
17.7	0.5996±0.0011	0.0002	0	0.295	0	0.177	270,675
17.85	0.5996±0.0011	0.00009	0	0.294	0	0.176	270,677
			(3) $\delta \mathrm{F}_{\nu_{\mathrm{o}},\, 2500\,\text{Å}} = 0.1 \times \mathrm{F}_{\nu_{\mathrm{o}},\, 2500\,\text{Å}}$				
17.27	0.5841±0.0011	0.00002	0.002	0.294	0.0000003	0.175	269,949
17.5	0.5955±0.0011	0.00002	0.0005	0.296	0.00000006	0.177	270,636
17.6	0.5974±0.0011	0.00002	0.0003	0.296	0.00000007	0.177	270,668
17.7	0.5983±0.0011	0.00002	0.0002	0.296	0.00000007	0.177	270,676
17.85	0.5991±0.0011	0.00002	0.00009	0.296	0.00000005	0.176	270,677
			(4) $\delta \mathrm{F}_{\nu_{\mathrm{o}},\, 2\,\mathrm{keV}} = 0.1 \times \mathrm{F}_{\nu_{\mathrm{o}},\, 2\,\mathrm{keV}}$				
17.27	0.5961±0.0011	0.002	0.00002	0.3	-0.00000002	0.175	269,900
17.5	0.5993±0.0011	0.0005	0.00002	0.297	-0.00000005	0.177	270,640
17.6	0.5997±0.0011	0.0003	0.00002	0.296	0.000000003	0.177	270,673
17.7	0.5999±0.0011	0.0002	0.00002	0.295	-0.000000007	0.177	270,675
17.85	0.5996±0.0011	0.00009	0.00002	0.294	-0.00000001	0.176	270,677

N: Normalisation of the uncertainty model slope (see equation 5.6).

$\overline{\gamma_{\mathrm{z}}}$: Inverse-variance weighted mean of the best-fit measurements of γ in each redshift bin.

$\overline{\sigma}^2_{\delta \mathrm{F_{UV}}}$: Variance of the UV flux density uncertainties.

$\overline{\sigma}^2_{\delta \mathrm{F_X}}$: Variance of the X-ray flux density uncertainties.

$\mathrm{Var(F_{UV})}$: Sample variance of the UV flux densities.

$\sigma_{\delta \mathrm{F_{UV}}, \delta \mathrm{F_X}}$: Covariance of the UV and X-ray flux density uncertainties.

$\mathrm{Cov(F_{UV}, F_X)}$: Sample covariance of the UV and X-ray flux densities.

Table 5.4: Results from fitting the X-ray - UV relation for a series of realisations of the main simulated *eROSITA - 4MOST* sample using different values for the uncertainty model slope λ (see equation 5.9). A redshift bin width of $\Delta\mathrm{Log}_{10}(z) = 0.007$ was adopted.

λ	$\overline{\gamma_z}$	$\overline{\sigma}^2_{\delta F_{UV}}$	$\overline{\sigma}^2_{\delta F_X}$	$\mathrm{Var}(F_{UV})$	$\sigma_{\delta F_{UV},\delta F_X}$	$\mathrm{Cov}(F_{UV}, F_X)$	No. of Sources
0.4	0.5997±0.0011	0.000000000002	0.000000002	0.294	0.000000000002	0.176	270,677
0.45	0.5997±0.0011	0.000000003	0.000000003	0.294	0.000000003	0.176	270,677
0.5	0.5996±0.0011	0.000005	0.000005	0.294	0.000005	0.176	270,677
0.53	0.5949±0.0011	0.0005	0.0005	0.297	0.0005	0.178	270,605
0.55	0.5103±0.0012	0.006	0.006	0.281	0.003	0.138	234,965

λ: Uncertainty model slope (see equation 5.9).

$\overline{\gamma_z}$: Inverse-variance weighted mean of the best-fit measurements of γ in each redshift bin.

$\overline{\sigma}^2_{\delta F_{UV}}$: Variance of the Log_{10} UV flux density uncertainties.

$\overline{\sigma}^2_{\delta F_X}$: Variance of the Log_{10} X-ray flux density uncertainties.

$\mathrm{Var}(F_{UV})$: Sample variance of the Log_{10} UV flux densities.

$\sigma_{\delta F_{UV},\delta F_X}$: Covariance of the Log_{10} UV and Log_{10} X-ray flux density uncertainties.

$\mathrm{Cov}(F_{UV}, F_X)$: Sample covariance of the Log_{10} UV and Log_{10} X-ray flux densities.

parameters Ω_M and Ω_Λ that may be achieved after the eRASS:8 *eROSITA*-4MOST sample becomes available. In addition, the simulated dataset was used to explore how the properties of a given quasar sample affect the quality of the cosmological constraints it can provide.

It was found that the majority of quasars in the simulated eRASS:8 *eROSITA*-4MOST sample have flux density measurements that are insufficiently precise to provide a consistent measurement of the $\mathrm{Log}_{10}(L_{\nu_e,\,2{,}500\,\text{Å}}) - \mathrm{Log}_{10}(L_{\nu_e,\,2\,\mathrm{keV}})$ slope γ (see section 5.6.1). As is to be expected, an accurate measurement, or theoretical prediction, of γ is in general required in order to achieve consistent measurements of Ω_M and Ω_Λ (see section 5.7.1). In section 5.6.1, it was demonstrated that the accuracy of the measurement of γ increases as the flux density measurement precision increases and that a relatively high flux density precision is required in order to recover the input value of $\gamma = 0.6$. This biasing of the measurement of γ appears to be due to regression dilution, the effects of which are discussed in section 5.7.6.

It was determined in section 5.6.2 that a sample with sufficient flux density precision could be identified by selecting sources above a threshold observation exposure time. Using the projected eRASS:8 exposure map, it was found that 2,539 sources can be expected to have flux densities with sufficiently high precision for use as reliable standard candles. Fitting the distance modulus-redshift relation using this reliable subsample returned $\Omega_M = 0.250^{+0.078}_{-0.060}$ and $\Omega_\Lambda = 0.758^{+0.262}_{-0.387}$ (see section 5.6.3), which are consistent with the values used to create the simulation ($\Omega_M = 0.307115$ and $\Omega_\Lambda = 0.692885$). In addition to verifying the efficacy of the selection criteria, this result demonstrates that even when working with only the 0.8 per cent highest-precision measurements in the sample, consistent cosmological constraints can be achieved.

Of significant note is that even in the case of an extreme level of dispersion in the

$L_{\nu_e,\,2\,keV} - L_{\nu_e,\,2{,}500\,Å}$ relation, using a sample with sufficiently precise flux density measurements will allow a consistent measurement of γ, Ω_M, and Ω_Λ to be made (section 5.7.2). As was discussed in section 5.7.3 it may be the case that the number of *eROSITA*-4MOST sources with sufficiently precise flux density measurements is larger than what is assumed in section 5.6.3, which would increase the precision of the Ω_M and Ω_Λ measurements as illustrated by figure 5.12. Finally, it is also apparent that the specific sample used, and how it populates the distance modulus - redshift parameter space, can dramatically influence whether reliable measurements of Ω_M and Ω_Λ are recovered even when a reliable measurement of γ is available (see section 5.7.4).

5.8.2 With Hindsight

The work presented in chapter 4 was an attempt to extend the RL15 method for using quasars as standard candles using larger sample sizes derived from the archival survey datasets that were available at the time. It was found that these archival datasets failed to deliver measurements of the cosmological model parameters Ω_M and Ω_Λ that were accurate compared to fiducial literature results. This remained the case even after experimenting with many different strategies for selecting a reliable subsample. It must be acknowledged that during the course of the work presented in chapter 4, the archival survey datasets were blindly applied to the quasar cosmology experiment in a heuristic manner since, at the time, significantly less was known about what properties of a given dataset would permit accurate measurements of Ω_M and Ω_Λ.

In addition to providing forecasts for the results of potential future work, the simulated *eROSITA*-4MOST dataset presented in this chapter was used as a testbed for exploring many of the challenges associated with the quasar cosmology experiment. Specifically, it was used to identify the main systematic effects that reduce the reliability of the Ω_M and Ω_Λ measurements (section 5.7). Before this additional work was completed, it was not possible to arrive at a firm conclusion as to why the archival survey data failed to return reliable results. A significant issue that demands careful consideration was found to be the precision of the flux density measurements; the precision requirements for this experiment, when considering survey-quality data, lies in the territory of *eROSITA* data and beyond. The lessons learned from working with the simulated data indicate that the archival data failed to provide reliable measurements of Ω_M and Ω_Λ due to the fact that these datasets contained sources with insufficiently precise flux density measurements.

The analysis presented in this chapter also provides a framework for identifying a dataset that is appropriate for use in the quasar cosmology experiment. For example, it was demonstrated that selecting a sample above a threshold observation exposure time is required in order to evade the effects of regression dilution (section 5.6.1). This will allow future datasets to be assessed in advance of carrying out the full analysis, thus preventing resources being spent working on datasets that are unlikely to provide reliable results. If these insights had been available prior to commencing the work presented in chapter 4 it is highly likely that those archival datasets would have been identified as being inappropriate for use in the quasar cosmology experiment.

5.8.3 Future Prospects

In addition to providing a larger sample size and increased redshift and luminosity coverage, *eROSITA* observations will provide information on the quasar X-ray continuum. This will allow source properties such as the intrinsic absorption column density, n_H, to be measured for at least the higher-count sources. For sources with reliable measurements of n_H, the flux densities can be corrected for intrinsic absorption (which, due primarily to the limited energy resolution, was not possible when using the 2RXS sample), while for the remaining sources hardness ratios may be used to identify and exclude from the analysis sources that show evidence of X-ray obscuration. The ability to correct for the effect of intrinsic absorption in this way can be expected to decrease the dispersion in the distance modulus-redshift relation, and therefore increase the precision of the resulting cosmological constraints. For higher-count sources reliable measurements of the X-ray continuum slope, Γ_X, may also be possible. This would allow sources with unusual spectral slopes to be excluded from the analysis, which can also be expected to reduce the dispersion in the distance modulus-redshift relation.

Despite continuing work (e.g. Lusso & Risaliti 2017, and references therein), a complete physical model describing the quasar accretion disk-corona interaction has yet to be conclusively determined. This relationship, if understood, may provide a theoretical prediction of the true value of γ, as well as whether γ can be expected to evolve with redshift. Such a model may also conclusively determine whether the quasar $\mathrm{Log}_{10}(L_{\nu_e,\,2{,}500\,\text{Å}}) - \mathrm{Log}_{10}(L_{\nu_e,\,2\,\mathrm{keV}})$ relation depends on BH mass (as suggested by Lusso & Risaliti (2017)), or the source's Eddington ratio, or whether additional parameters (such as Γ_X) can be expected to influence the dispersion in the observed relation.

Appendix A

Dark Energy Density Parameter in ΛCDM

In the general case where the dark energy equation of state is a function of redshift, the dark energy density parameter is given by

$$\Omega_{de}(z) = \Omega_\Lambda \exp\left(3\int_0^z \frac{1+w(x)}{1+x}dx\right) \tag{A.1}$$

For the standard ΛCDM case, where w is constant with redshift and equal to -1, the integral can be solved as follows

$$\int_0^z \frac{1+w}{1+x}dx$$

Substitution:

$$u = x + 1 \qquad \frac{du}{dx} = 1$$

which gives

$$(w+1)\int_0^z \frac{1}{u}du$$

$\int \frac{1}{u}du$ is a standard integral, and equal to ln(u), giving

$$(w+1)\ln(u)|_0^z$$

Undo substitution

$$(w+1)\ln(x+1)|_0^z$$

$$(w+1)\ln(x+1) - (w+1)\ln(0+1)$$

$$(\mathrm{w}+1)\ln(\mathrm{x}+1)$$

Substitute back into equation A.1

$$\Omega_{\mathrm{de}}(\mathrm{z}) = \Omega_{\Lambda} \exp\left(3(\mathrm{w}+1)\ln(\mathrm{x}+1)\right)$$

Which, for w = -1

$$\Omega_{\mathrm{de}} = \Omega_{\Lambda}$$

Appendix B

Likelihood Function for the Distance Modulus - Redshift Fit.

The Gaussian probability density function for the fit can be written as

$$\mathrm{p}(\mathrm{z}, \mathrm{DM}, \delta\mathrm{DM} \,|\, \Omega_\mathrm{M}, \Omega_\Lambda, \beta, \delta) = \frac{1}{\sqrt{2\pi \mathrm{s}^2}} \mathrm{e}^{-\frac{(\mathrm{DM} - \mathrm{DM}(\Omega_\mathrm{M}, \Omega_\Lambda, \mathrm{z}, \beta))^2}{2\mathrm{s}^2}}$$

where

$$\mathrm{s}^2 = \delta\mathrm{DM}^2 + \delta^2$$

Switching to logs, this can be written as

$$\mathrm{Ln}\, \mathrm{p}(\mathrm{z}, \mathrm{DM}, \delta\mathrm{DM} \,|\, \Omega_\mathrm{M}, \Omega_\Lambda, \beta, \delta) = -\frac{1}{2} \left(\mathrm{Ln}(2\pi \mathrm{s}^2) + \frac{(\mathrm{DM} - \mathrm{DM}(\Omega_\mathrm{M}, \Omega_\Lambda, \mathrm{z}, \beta))^2}{\mathrm{s}^2} \right) \quad \text{(B.1)}$$

The joint density, or likelihood function, is the product of the individual densities.

$$\mathrm{L}(\theta \,|\, \mathrm{y}) = \prod_{\mathrm{i}=1}^{\mathrm{n}} \mathrm{p}(\mathrm{y_i} \,|\, \theta)$$

which, written in terms of the natural logarithm, is

$$\mathrm{Ln}\, \mathrm{L}(\theta \,|\, \mathrm{y}) = \sum_{\mathrm{i}=1}^{\mathrm{n}} \mathrm{Ln}\, \mathrm{p}(\mathrm{y_i} \,|\, \theta) \quad \text{(B.2)}$$

Since $\mathrm{Ln}(\prod_i x_i = \sum_i (x_i))$. From equations B.1 and B.2 the likelihood function used in the distance modulus - redshift fits discussed in section 1.6.3 can be written as

$$\mathrm{Ln}\, \mathrm{L}(\Omega_\mathrm{M}, \Omega_\Lambda, \beta, \delta \,|\, \mathrm{z}, \mathrm{DM}, \delta\mathrm{DM}) =$$
$$-\frac{1}{2} \sum_{\mathrm{i}=1}^{\mathrm{n}} \left(\frac{[\mathrm{DM_i} - \mathrm{DM}(\mathrm{z_i}, \Omega_\mathrm{M}, \Omega_\Lambda, \beta)]^2}{\mathrm{s_i^2}} + \mathrm{Ln}(2\pi \mathrm{s_i^2}) \right)$$

Appendix C

Extinction Correction

The extinction corrected flux can be expressed as

$$F_{int} = F_{obs}\, 10^{0.4 A_\lambda} \tag{C.1}$$

where F_{obs} is the observed flux. The wavelength-dependent extinction is

$$A_\lambda = k(\lambda)\, E(B-V) = \frac{k(\lambda)\, A_V}{R_V} \tag{C.2}$$

where E(B-V) is the colour excess and $k(\lambda)$ is the reddening curve. Note that

$$R_V = \begin{cases} 3.1 & \text{in general} \\ 2.72 \pm 0.21 & \text{for the Small Magellanic Cloud} \\ 4.05 \pm 0.8 & \text{for Calzetti's law} \end{cases}$$

Appendix D

$\Omega_M - \Omega_\Lambda$ Constraints from Risaliti & Lusso (2015)

As discussed in section 1.6.2, RL15 presented constraints on the cosmological parameters Ω_M and Ω_Λ based on distance modulus estimates derived from 808 X-ray selected quasars. The left panel of figure D.1 shows the Hubble diagram resulting from their work. Their dataset was derived by combining a series of previously published quasar samples (Steffen et al. 2006; Just et al. 2007; Shemmer et al. 2006; Lusso et al. 2010; Young et al. 2010), and the authors noted that the combination of multiple samples was required in order to achieve adequate coverage of the $F_X - z$ and $F_{UV} - z$ parameter spaces. The Ω_M and Ω_Λ constraints derived from the RL15 sample are shown in the right panel of figure D.1.

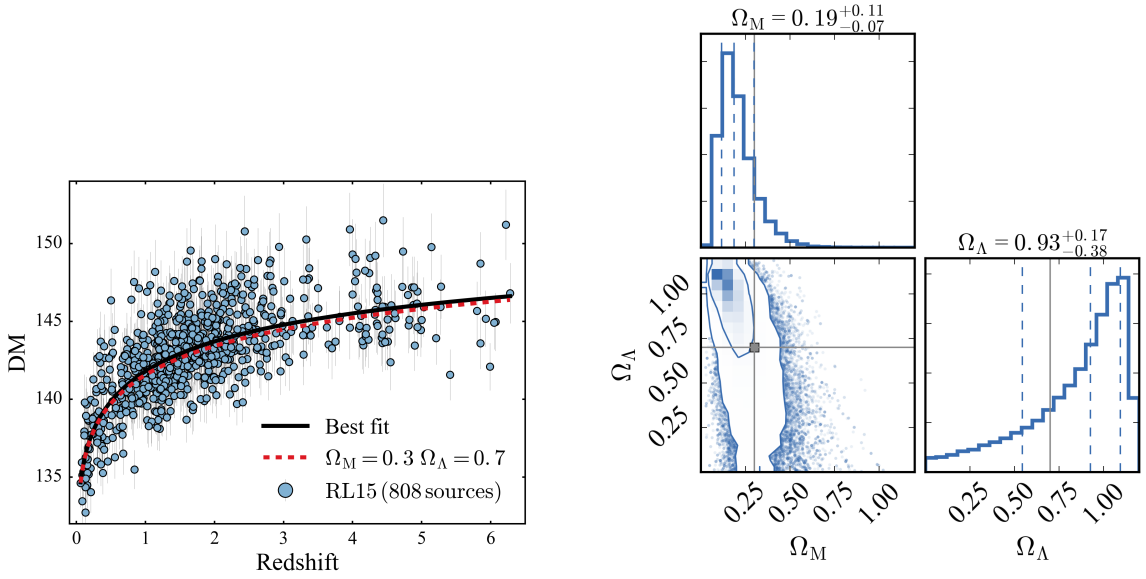


Figure D.1: Left panel: Distance Modulus versus redshift for the sample presented by RL15. Right panel: Constraints on Ω_M and Ω_λ obtained from fitting the RL15 sample. The 68 and 95 per cent confidence regions are shown, along with the 1 sigma projection of the posterior probability distributions for both parameters. The grey crosshairs indicate the initial values for the MCMC walkers, which were set to the fiducial values of $\Omega_M = 0.3$ and $\Omega_\Lambda = 0.7$. These figures are a reproduction of RL15 figures 5 and 6.

Bibliography

Abazajian, K. N., Adelman-McCarthy, J. K., Agüeros, M. A., et al. 2009, ApJS, 182, 543

Ahn, C. P., Alexandroff, R., Allende Prieto, C., et al. 2012, ApJS, 203, 21

Aihara, H., Allende Prieto, C., An, D., et al. 2011, ApJS, 193, 29

Aird, J., Coil, A. L., Georgakakis, A., et al. 2015, MNRAS, 451, 1892

Aird, J., Coil, A. L., Moustakas, J., et al. 2012, ApJ, 746, 90

Akritas, M. G. & Bershady, M. A. 1996, ApJ, 470, 706

Alam, S., Albareti, F. D., Allende Prieto, C., et al. 2015, ApJS, 219, 12

Albareti, F. D., Allende Prieto, C., Almeida, A., et al. 2017, ApJS, 233, 25

Allen, C. W. 1976, Astrophysical Quantities

Anderson, S. F., Margon, B., Voges, W., et al. 2007, AJ, 133, 313

Antonucci, R. 1993, Annu Rev Astron Astr, 31, 473

Antonucci, R. R. J. 1984, ApJ, 278, 499

Antonucci, R. R. J. & Miller, J. S. 1985, ApJ, 297, 621

Arnaud, K. A. 1996, in Astronomical Society of the Pacific Conference Series, Vol. 101, Astronomical Data Analysis Software and Systems V, ed. G. H. Jacoby & J. Barnes, 17

Arnouts, S., Cristiani, S., Moscardini, L., et al. 1999, MNRAS, 310, 540

Arp, H. C., Bolton, J. G., & Kinman, T. D. 1967, ApJ, 147, 840

Assef, R. J., Denney, K. D., Kochanek, C. S., et al. 2011, ApJ, 742, 93

Astier, P., Guy, J., Regnault, N., et al. 2006, A&A, 447, 31

Avni, Y. & Tananbaum, H. 1986, ApJ, 305, 83

Baade, W. & Minkowski, R. 1954, ApJ, 119, 206

Bahcall, J. N., Kozlovsky, B.-Z., & Salpeter, E. E. 1972, ApJ, 171, 467

Baldwin, J. A. 1977, ApJ, 214, 679

Baldwin, J. A., Ferland, G. J., Korista, K. T., Hamann, F., & LaCluyzé, A. 2004, ApJ, 615, 610

Barvainis, R. 1987, ApJ, 320, 537

Baskin, A. & Laor, A. 2005, MNRAS, 356, 1029

Beckmann, V. & Shrader, C. R. 2012, Active Galactic Nuclei

Bennett, A. S. 1962, MmRAS, 68, 163

Bentz, M. C., Denney, K. D., Grier, C. J., et al. 2013, ApJ, 767, 149

Bentz, M. C. & Katz, S. 2015, PASP, 127, 67

Bentz, M. C., Peterson, B. M., Netzer, H., Pogge, R. W., & Vestergaard, M. 2009a, ApJ, 697, 160

Bentz, M. C., Peterson, B. M., Pogge, R. W., Vestergaard, M., & Onken, C. A. 2006, ApJ, 644, 133

Bentz, M. C., Walsh, J. L., Barth, A. J., et al. 2009b, ApJ, 705, 199

Bianchi, L., Rodriguez-Merino, L., Viton, M., et al. 2007, ApJS, 173, 659

Blandford, R. D. & McKee, C. F. 1982, ApJ, 255, 419

Blanton, M. R., Bershady, M. A., Abolfathi, B., et al. 2017, AJ, 154, 28

Boller, T., Freyberg, M. J., Trümper, J., et al. 2016, A&A, 588, A103

Bon, E., Popović, L. Č., Gavrilović, N., La Mura, G., & Mediavilla, E. 2009, MNRAS, 400, 924

Boroson, T. 2005, AJ, 130, 381

Boroson, T. A. & Green, R. F. 1992, ApJS, 80, 109

Bradt, H. 2004, Astronomy Methods: A Physical Approach to Astronomical Observations

Briel, U. G., Aschenbach, B., Hasinger, G., et al. 1996, The ROSAT Users' Handbook

Brotherton, M. S. 1996, ApJS, 102, 1

Brunner, H., Cappelluti, N., Hasinger, G., et al. 2008, A&A, 479, 283

Brunner, H., Liu, T., Lamer, G., et al. 2022, A&A, 661, A1

Buchner, J., Georgakakis, A., Nandra, K., et al. 2014, A&A, 564, A125

Budavári, T., Heinis, S., Szalay, A. S., et al. 2009, ApJ, 694, 1281

Calderone, G., Nicastro, L., Ghisellini, G., et al. 2017, MNRAS, 472, 4051

Cappelluti, N., Predehl, P., Böhringer, H., et al. 2011, Memorie della Societa Astronomica Italiana Supplementi, 17, 159

Capriotti, E. R., Foltz, C. B., & Peterson, B. M. 1982, ApJ, 261, 35

Cardelli, J. A., Clayton, G. C., & Mathis, J. S. 1989, ApJ, 345, 245

Chen, K., Halpern, J. P., & Filippenko, A. V. 1989, ApJ, 339, 742

Chiappetti, L., Clerc, N., Pacaud, F., et al. 2013, MNRAS, 429, 1652

Chiaraluce, E., Vagnetti, F., Tombesi, F., & Paolillo, M. 2018, A&A, 619, A95

Civano, F., Marchesi, S., Comastri, A., et al. 2016, ApJ, 819, 62

Clerc, N., Adami, C., Lieu, M., et al. 2014, MNRAS, 444, 2723

Coatman, L., Hewett, P. C., Banerji, M., et al. 2017, MNRAS, 465, 2120

Coffey, D., Salvato, M., Merloni, A., et al. 2019, A&A, 625, A123

Comparat, J., Kneib, J.-P., Bacon, R., et al. 2013, A&A, 559, A18

Comparat, J., Merloni, A., Salvato, M., et al. 2019, MNRAS, 487, 2005

Corbett, E. A., Croom, S. M., Boyle, B. J., et al. 2003, MNRAS, 343, 705

Corbin, M. R. 1995, ApJ, 447, 496

Corbin, M. R. & Boroson, T. A. 1996, ApJS, 107, 69

Cutri et al., R. M. e. 2013, VizieR Online Data Catalog, 2328

Czerny, B. & Hryniewicz, K. 2011, A&A, 525, L8

Czerny, B., Hryniewicz, K., Maity, I., et al. 2013, A&A, 556, A97

Dawson, K. S., Kneib, J.-P., Percival, W. J., et al. 2016, AJ, 151, 44

Dawson, K. S., Schlegel, D. J., Ahn, C. P., et al. 2013, AJ, 145, 10

de Jong, R. S., Agertz, O., Berbel, A. A., et al. 2019, The Messenger, 175, 3

de Jong, R. S., Barden, S., Bellido-Tirado, O., et al. 2014, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 9147, Proc. SPIE, 91470M

den Herder, J. W., Brinkman, A. C., Kahn, S. M., et al. 2001, A&A, 365, L7

Denney, K. D. 2012, ApJ, 759, 44

Denney, K. D., Peterson, B. M., Dietrich, M., Vestergaard, M., & Bentz, M. C. 2009, ApJ, 692, 246

Dietrich, M. & Hamann, F. 2004, ApJ, 611, 761

Dwelly, T., Salvato, M., Merloni, A., et al. 2017, MNRAS, 469, 1065

Edge, D. O., Shakeshaft, J. R., McAdam, W. B., Baldwin, J. E., & Archer, S. 1959, MmRAS, 68, 37

Eisenstein, D. J., Weinberg, D. H., Agol, E., et al. 2011, AJ, 142, 72

Elitzur, M. & Shlosman, I. 2006, ApJ, 648, L101

Eracleous, M. & Halpern, J. P. 1994, ApJS, 90, 1

Eracleous, M. & Halpern, J. P. 2003, ApJ, 599, 886

Fabian, A. C., Rees, M. J., Stella, L., & White, N. E. 1989, MNRAS, 238, 729

Fath, E. A. 1909, Lick Observatory Bulletin, 149, 71

Ferland, G. J., Peterson, B. M., Horne, K., Welsh, W. F., & Nahar, S. N. 1992, ApJ, 387, 95

Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. 2013, PASP, 125, 306

Fotopoulou, S., Salvato, M., Hasinger, G., et al. 2012, ApJS, 198, 1

Freedman, W. L., Madore, B. F., Gibson, B. K., et al. 2001, ApJ, 553, 47

Frost, C. & Thompson, S. G. 2000, Journal of the Royal Statistical Society: Series A (Statistics in Society), 163, 173

Garnavich, P. M., Jha, S., Challis, P., et al. 1998, ApJ, 509, 74

Gaskell, C. M. & Sparke, L. S. 1986, ApJ, 305, 175

Gebhardt, K., Bender, R., Bower, G., et al. 2000, ApJL, 539, L13

Georgakakis, A. & Nandra, K. 2011, MNRAS, 414, 992

Georgakakis, A., Nandra, K., Laird, E. S., Aird, J., & Trichas, M. 2008, MNRAS, 388, 1205

Giavalisco, M., Ferguson, H. C., Koekemoer, A. M., et al. 2004, ApJ, 600, L93

Gibson, R. R., Brandt, W. N., & Schneider, D. P. 2008, ApJ, 685, 773

Graham, A. W. 2016, in Astrophysics and Space Science Library, Vol. 418, Galactic Bulges, ed. E. Laurikainen, R. Peletier, & D. Gadotti, 263

Green, P. J., Aldcroft, T. L., Richards, G. T., et al. 2009, ApJ, 690, 644

Grogin, N. A., Conselice, C. J., Chatzichristou, E., et al. 2005, ApJL, 627, L97

Grupe, D., Komossa, S., Leighly, K. M., & Page, K. L. 2010, ApJS, 187, 64

Gunn, J. E., Carr, M., Rockosi, C., et al. 1998, AJ, 116, 3040

Gunn, J. E., Siegmund, W. A., Mannery, E. J., et al. 2006, AJ, 131, 2332

Guy, J., Sullivan, M., Conley, A., et al. 2010, A&A, 523, A7

Haardt, F. & Maraschi, L. 1991, ApJ, 380, L51

Häring, N. & Rix, H.-W. 2004, ApJ, 604, L89

Hicken, M., Wood-Vasey, W. M., Blondin, S., et al. 2009, ApJ, 700, 1097

Hönig, S. F., Watson, D., Kishimoto, M., et al. 2017, MNRAS, 464, 1693

Hoyle, F. & Fowler, W. A. 1963a, Nature, 197, 533

Hoyle, F. & Fowler, W. A. 1963b, MNRAS, 125, 169

Hu, C., Wang, J.-M., Ho, L. C., et al. 2008, ApJ, 687, 78

Ilbert, O., Arnouts, S., Le Floc'h, E., et al. 2015, A&A, 579, A2

Ilbert, O., Arnouts, S., McCracken, H. J., et al. 2006, A&A, 457, 841

Jansen, F., Lumb, D., Altieri, B., et al. 2001, A&A, 365, L1

Just, D. W., Brandt, W. N., Shemmer, O., et al. 2007, ApJ, 665, 1004

Kaspi, S., Maoz, D., Netzer, H., et al. 2005, ApJ, 629, 61

Kaspi, S., Smith, P. S., Netzer, H., et al. 2000, ApJ, 533, 631

Kelly, B. C. 2007, ApJ, 665, 1489

Kelly, B. C., Bechtold, J., Trump, J. R., Vestergaard, M., & Siemiginowska, A. 2008, ApJS, 176, 355

Kessler, R., Becker, A. C., Cinabro, D., et al. 2009, ApJS, 185, 32

Klypin, A., Yepes, G., Gottlöber, S., Prada, F., & Heß, S. 2016, MNRAS, 457, 4340

Knop, R. A., Aldering, G., Amanullah, R., et al. 2003, ApJ, 598, 102

Kollmeier, J. A., Onken, C. A., Kochanek, C. S., et al. 2006, ApJ, 648, 128

Kollmeier, J. A., Zasowski, G., Rix, H.-W., et al. 2017, arXiv e-prints, arXiv:1711.03234

Komatsu, E., Dunkley, J., Nolta, M. R., et al. 2009, ApJS, 180, 330

Komatsu, E., Smith, K. M., Dunkley, J., et al. 2011, ApJS, 192, 18

Koratkar, A. P. & Gaskell, C. M. 1991, ApJl, 370, L61

Kormendy, J. & Richstone, D. 1995, Annu Rev Astron Astr, 33, 581

Kowalski, M., Rubin, D., Aldering, G., et al. 2008, ApJ, 686, 749

Kraft, R. P., Burrows, D. N., & Nousek, J. A. 1991, ApJ, 374, 344

Kuehn, C. A., Baldwin, J. A., Peterson, B. M., & Korista, K. T. 2008, ApJ, 673, 69

Laird, E. S., Nandra, K., Georgakakis, A., et al. 2009, ApJS, 180, 102

LaMassa, S. M., Cales, S., Moran, E. C., et al. 2015, ApJ, 800, 144

LaMassa, S. M., Urry, C. M., Cappelluti, N., et al. 2016, ApJ, 817, 172

Lawrence, A. & Elvis, M. 1982, ApJ, 256, 410

Lehmer, B. D., Xue, Y. Q., Brandt, W. N., et al. 2012, ApJ, 752, 46

Li, J., Shen, Y., Horne, K., et al. 2017, ApJ, 846, 79

Liu, Z., Merloni, A., Georgakakis, A., et al. 2016, MNRAS, 459, 1602

Luo, B., Brandt, W. N., Xue, Y. Q., et al. 2017, ApJS, 228, 2

Lusso, E., Comastri, A., Vignali, C., et al. 2010, A&A, 512, A34

Lusso, E., Piedipalumbo, E., Risaliti, G., et al. 2019, A&A, 628, L4

Lusso, E. & Risaliti, G. 2016, ApJ, 819, 154

Lusso, E. & Risaliti, G. 2017, A&A, 602, A79

Lynden-Bell, D. 1969, Nature, 223, 690

Magorrian, J., Tremaine, S., Richstone, D., et al. 1998, AJ, 115, 2285

Marchesi, S., Civano, F., Elvis, M., et al. 2016a, ApJ, 817, 34

Marchesi, S., Lanzuisi, G., Civano, F., et al. 2016b, ApJ, 830, 100

Markowitz, A. G., Krumpe, M., & Nikutta, R. 2014, MNRAS, 439, 1403

Markwardt, C. B. 2009, in Astronomical Society of the Pacific Conference Series, Vol. 411, Astronomical Data Analysis Software and Systems XVIII, ed. D. A. Bohlender, D. Durand, & P. Dowler, 251

Martin, D. C., Fanson, J., Schiminovich, D., et al. 2005, ApJ, 619, L1

Marziani, P. & Sulentic, J. W. 2012, NewAR, 56, 49

Marziani, P., Sulentic, J. W., Negrete, C. A., et al. 2015, Ap&SS, 356, 339

Marziani, P., Sulentic, J. W., Negrete, C. A., et al. 2010, MNRAS, 409, 1033

Marziani, P., Sulentic, J. W., Stirpe, G. M., Zamfir, S., & Calvani, M. 2009, A&A, 495, 83

Marziani, P., Sulentic, J. W., Zamanov, R., et al. 2003, ApJS, 145, 199

Mason, K. O., Breeveld, A., Much, R., et al. 2001, A&A, 365, L36

Matthews, T. A. & Sandage, A. R. 1963, ApJ, 138, 30

McHardy, I. M., Cameron, D. T., Dwelly, T., et al. 2014, MNRAS, 444, 1469

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

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

Mejía-Restrepo, J. E., Trakhtenbrot, B., Lira, P., & Netzer, H. 2018, MNRAS, 478, 1929

Melia, F. 2015, Ap&SS, 359, 34

Menzel, M.-L., Merloni, A., Georgakakis, A., et al. 2016, MNRAS, 457, 110

Merloni, A., Alexander, D. A., Banerji, M., et al. 2019, The Messenger, 175, 42

Merloni, A., Bongiorno, A., Brusa, M., et al. 2014, MNRAS, 437, 3550

Merloni, A., Dwelly, T., Salvato, M., et al. 2015, MNRAS, 452, 69

Merloni, A., Predehl, P., Becker, W., et al. 2012, ArXiv e-prints

Merritt, D. & Ferrarese, L. 2001, ApJ, 547, 140

Morganson, E., Green, P. J., Anderson, S. F., et al. 2015, ApJ, 806, 244

Morrissey, P., Conrow, T., Barlow, T. A., et al. 2007, ApJS, 173, 682

Mortlock, D. J., Warren, S. J., Venemans, B. P., et al. 2011, Nature, 474, 616

Nandra, K., Laird, E. S., Aird, J. A., et al. 2015, ApJS, 220, 10

Nandra, K. & Pounds, K. A. 1994, MNRAS, 268, 405

Pâris, I., Petitjean, P., Ross, N. P., et al. 2017, A&A, 597, A79

Park, D., Woo, J.-H., Denney, K. D., & Shin, J. 2013, ApJ, 770, 87

Pavlinsky, M., Akimov, V., Levin, V., et al. 2016, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 9905, Proc. SPIE, 99051J

Pavlinsky, M., Sunyaev, R., Churazov, E., et al. 2008, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 7011, Proc. SPIE, 70110H

Pereira-Santaella, M., Alonso-Herrero, A., Santos-Lleo, M., et al. 2011, A&A, 535, A93

Perez, E., Penston, M. V., Tadhunter, C., Mediavilla, E., & Moles, M. 1988, MNRAS, 230, 353

Perlmutter, S., Aldering, G., della Valle, M., et al. 1998, Nature, 391, 51

Perlmutter, S., Aldering, G., Goldhaber, G., et al. 1999, ApJ, 517, 565

Persic, M. & Rephaeli, Y. 2002, A&A, 382, 843

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

Peterson, B. M. 1997, An Introduction to Active Galactic Nuclei

Peterson, B. M. & Ferland, G. J. 1986, Nature, 324, 345

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

Peterson, B. M. & Wandel, A. 2000, ApJ, 540, L13

Phillips, M. M. 1993, ApJl, 413, L105

Pierce, C. M., Lotz, J. M., Laird, E. S., et al. 2007, ApJL, 660, L19

Pierre, M., Pacaud, F., Adami, C., et al. 2016, A&A, 592, A1

Pierre, M., Valtchanov, I., Altieri, B., et al. 2004, JCAP, 9, 011

Planck Collaboration, Aghanim, N., Akrami, Y., et al. 2018, arXiv e-prints, arXiv:1807.06209

Plotkin, R. M., Anderson, S. F., Brandt, W. N., et al. 2010, AJ, 139, 390

Polletta, M., Tajer, M., Maraschi, L., et al. 2007, ApJ, 663, 81

Popović, L. Č., Mediavilla, E., Bon, E., & Ilić, D. 2004, A&A, 423, 909

Predehl, P., Andritschke, R., Babyshkin, V., et al. 2016, in Proc. SPIE, Vol. 9905, Space Telescopes and Instrumentation 2016: Ultraviolet to Gamma Ray, 99051K

Predehl, P., Andritschke, R., Böhringer, H., et al. 2010, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 7732, Proc. SPIE, 77320U

Prevot, M. L., Lequeux, J., Prevot, L., Maurice, E., & Rocca-Volmerange, B. 1984, A&A, 132, 389

Ranalli, P., Comastri, A., & Setti, G. 2003, A&A, 399, 39

Richards, G. T., Lacy, M., Storrie-Lombardi, L. J., et al. 2006a, ApJS, 166, 470

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

Richards, G. T., Vanden Berk, D. E., Reichard, T. A., et al. 2002, AJ, 124, 1

Riess, A. G., Filippenko, A. V., Challis, P., et al. 1998, AJ, 116, 1009

Riess, A. G., Strolger, L.-G., Casertano, S., et al. 2007, ApJ, 659, 98

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

Risaliti, G. & Lusso, E. 2019, Nature Astronomy, 3, 272

Rodney, S. A., Riess, A. G., Scolnic, D. M., et al. 2015, AJ, 150, 156

Rosati, P., Borgani, S., & Norman, C. 2002, ARA&A, 40, 539

Runnoe, J. C., Brotherton, M. S., Shang, Z., & DiPompeo, M. A. 2013, MNRAS, 434, 848

Salpeter, E. E. 1964, ApJ, 140, 796

Salvato, M., Buchner, J., Budavári, T., et al. 2018, MNRAS, 473, 4937

Salvato, M., Hasinger, G., Ilbert, O., et al. 2009, ApJ, 690, 1250

Salviander, S., Shields, G. A., Gebhardt, K., & Bonning, E. W. 2007, ApJ, 662, 131

Sanders, D. B., Phinney, E. S., Neugebauer, G., Soifer, B. T., & Matthews, K. 1989, ApJ, 347, 29

Saxton, R. D., Read, A. M., Esquej, P., et al. 2008, A&A, 480, 611

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

Schmidt, B. P., Suntzeff, N. B., Phillips, M. M., et al. 1998, ApJ, 507, 46

Schmidt, M. 1963, Nature, 197, 1040

Schmidt, M. 1968, ApJ, 151, 393

Schmitt, H. R., Antonucci, R. R. J., Ulvestad, J. S., et al. 2001, ApJ, 555, 663

Schneider, D. P., Richards, G. T., Hall, P. B., et al. 2010, AJ, 139, 2360

Schnittman, J. D., Krolik, J. H., & Noble, S. C. 2013, ApJ, 769, 156

Schwarz, G. 1978, Ann. Stat., 6, 461

Scoville, N., Aussel, H., Brusa, M., et al. 2007, ApJS, 172, 1

Seaton, M. J. 1979, MNRAS, 187, 73P

Seyfert, C. K. 1943, ApJ, 97, 28

Shemmer, O., Brandt, W. N., Schneider, D. P., et al. 2006, ApJ, 644, 86

Shen, Y. 2013, Bulletin of the Astronomical Society of India, 41, 61

Shen, Y., Brandt, W. N., Dawson, K. S., et al. 2015, ApJS, 216, 4

Shen, Y., Greene, J. E., Strauss, M. A., Richards, G. T., & Schneider, D. P. 2008, ApJ, 680, 169

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

Shen, Y., Richards, G. T., Strauss, M. A., et al. 2011, ApJS, 194, 45

Shields, G. A., Ludwig, R. R., & Salviander, S. 2010, ApJ, 721, 1835

Shields, J. C. 2007, in Astronomical Society of the Pacific Conference Series, Vol. 373, The Central Engine of Active Galactic Nuclei, ed. L. C. Ho & J.-W. Wang, 355

Silva, L., Granato, G. L., Bressan, A., & Danese, L. 1998, ApJ, 509, 103

Simm, T., Saglia, R., Salvato, M., et al. 2015, A&A, 584, A106

Slipher, V. M. 1917, Lowell Observatory Bulletin, 3, 59

Smee, S. A., Gunn, J. E., Uomoto, A., et al. 2013, AJ, 146, 32

Steffen, A. T., Strateva, I., Brandt, W. N., et al. 2006, AJ, 131, 2826

Steinhardt, C. L. & Elvis, M. 2010, MNRAS, 402, 2637

Storey, P. J. & Zeippen, C. J. 2000, MNRAS, 312, 813

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

Strateva, I. V., Strauss, M. A., Hao, L., et al. 2003, AJ, 126, 1720

Strüder, L., Briel, U., Dennerl, K., et al. 2001, A&A, 365, L18

Sulentic, J., Marziani, P., & Zamfir, S. 2011, Baltic Astronomy, 20, 427

Sulentic, J. W., Marziani, P., Olmo, A. d., & Plauchu-Frayn, I. 2014, Advances in Space Research, 54, 1406

Sulentic, J. W., Marziani, P., Zamanov, R., et al. 2002, ApJL, 566, L71

Sulentic, J. W., Marziani, P., Zwitter, T., Dultzin-Hacyan, D., & Calvani, M. 2000a, ApJL, 545, L15

Sulentic, J. W., Zwitter, T., Marziani, P., & Dultzin-Hacyan, D. 2000b, ApJ, 536, L5

Sullivan, M., Guy, J., Conley, A., et al. 2011, ApJ, 737, 102

Sunyaev, R. A. & Truemper, J. 1979, Nature, 279, 506

Sutherland, W., Emerson, J., Dalton, G., et al. 2015, A&A, 575, A25

Suzuki, N., Rubin, D., Lidman, C., et al. 2012, ApJ, 746, 85

Tadhunter, C. 2008, NewAR, 52, 227

Tananbaum, H., Avni, Y., Branduardi, G., et al. 1979, ApJl, 234, L9

Taylor, M. B. 2005, in Astronomical Society of the Pacific Conference Series, Vol. 347, Astronomical Data Analysis Software and Systems XIV, ed. P. Shopbell, M. Britton, & R. Ebert, 29

Trakhtenbrot, B. & Netzer, H. 2012, MNRAS, 427, 3081

Trammell, G. B., Vanden Berk, D. E., Schneider, D. P., et al. 2007, AJ, 133, 1780

Tremaine, S., Gebhardt, K., Bender, R., et al. 2002, ApJ, 574, 740

Truemper, J. 1982, Advances in Space Research, 2, 241

Turner, M. J. L., Abbey, A., Arnaud, M., et al. 2001, A&A, 365, L27

Urry, C. M. & Padovani, P. 1995, PASP, 107, 803

Uttley, P., Cackett, E. M., Fabian, A. C., Kara, E., & Wilkins, D. R. 2014, A&A Rev, 22, 72

Vestergaard, M. 2002, ApJ, 571, 733

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

Vestergaard, M. & Wilkes, B. J. 2001, ApJS, 134, 1

Vignali, C., Brandt, W. N., & Schneider, D. P. 2003a, AJ, 125, 433

Vignali, C., Brandt, W. N., Schneider, D. P., et al. 2003b, AJ, 125, 2876

Voges, W., Aschenbach, B., Boller, T., et al. 1999, A&A, 349, 389

Voges, W., Aschenbach, B., Boller, T., et al. 2000, IAU Circ., 7432

Wang, J.-G., Dong, X.-B., Wang, T.-G., et al. 2009, ApJ, 707, 1334

Watson, D., Denney, K. D., Vestergaard, M., & Davis, T. M. 2011, ApJl, 740, L49

Wills, B. J., Brotherton, M. S., Fang, D., Steidel, C. C., & Sargent, W. L. W. 1993, ApJ, 415, 563

Wolf, J., Salvato, M., Coffey, D., et al. 2020, MNRAS, 492, 3580

Woltjer, L. 1959, ApJ, 130, 38

Wood-Vasey, W. M., Miknaitis, G., Stubbs, C. W., et al. 2007, ApJ, 666, 694

Wright, E. L., Eisenhardt, P. R. M., Mainzer, A. K., et al. 2010, AJ, 140, 1868

Yang, T., Banerjee, A., & Ó Colgáin, E. 2020, Phys. Rev. D, 102, 123532

York, D. G., Adelman, J., Anderson, Jr., J. E., et al. 2000, AJ, 120, 1579

Yoshii, Y., Kobayashi, Y., Minezaki, T., Koshida, S., & Peterson, B. A. 2014, ApJl, 784, L11

Young, M., Elvis, M., & Risaliti, G. 2009, ApJS, 183, 17

Young, M., Elvis, M., & Risaliti, G. 2010, ApJ, 708, 1388

Zamfir, S., Sulentic, J. W., Marziani, P., & Dultzin, D. 2010, MNRAS, 403, 1759

Zamorani, G., Henry, J. P., Maccacaro, T., et al. 1981, ApJ, 245, 357

Zdziarski, A. A., Fabian, A. C., Nandra, K., et al. 1994, MNRAS, 269, L55

Zeldovich, Y. B. & Novikov, I. D. 1964, Sov. Phys. Dokl., 158, 811

Zhang, X.-G., Dultzin-Hacyan, D., & Wang, T.-G. 2007, MNRAS, 376, 1335

Acknowledgements

Personal

My time at MPE has been enormously enriching for both my professional and personal life. I have been fortunate to have worked under the supervision of Paul Nandra, Mara Salvato, and Thomas Boller. Being able to learn from the example set by such successful and esteemed members of the scientific community has been of tremendous value. They have given me the opportunity to experience life at one of the top research institutions in the world, and for that I will always be grateful. There are many others to whom I owe a great debt of gratitude to, too many to list here, but in particular I would like to mention:

- Ricky, Linda, Vlas, Adam, Jacob, Yru, Julien, and the many other good friends that I have made during my time at MPE.
- Andrea Merloni, Johan Comparat, Johannes Buchner, Tom Dwelly, Guido Risaliti, and the other colleagues with whom I have had many productive conversations throughout the years; their input has been of great value to me and has without doubt improved the quality of my work.
- The High-Energy Group at MPE, which provided a stimulating, competitive, and rewarding working environment. From group retreats at Ringberg to the launch of *eROSITA*, I'll carry some great memories with me from my time spent among them.
- Christian Elsner and Harald Baumgartner who provided much appreciated advice on the many IT-related issues I've brought to them.
- Birgit Boller, Annette Hilbert, and the MPE administrative staff, who were always prepared to help me with logistical and bureaucratic tasks.
- My family back home who were always there to provide encouragement and support. Christmases spent with them back in Tipperary and their visits to Munich have been highlights for me during these busy years, and have meant that I have never felt too far from home.

Finally, I am extremely lucky to have had my fiancé Hanna by my side; throughout my time at MPE and during all our years together she has been a constant source of love, support, and encouragement.

Software and Data

Funding for the Sloan Digital Sky Survey IV has been provided by the Alfred P. Sloan Foundation, the U.S. Department of Energy Office of Science, and the Participating Institutions. SDSS-IV acknowledges support and resources from the Center for High-Performance Computing at the University of Utah. The SDSS web site is www.sdss.org.

SDSS-IV is managed by the Astrophysical Research Consortium for the Participating Institutions of the SDSS Collaboration including the Brazilian Participation Group, the Carnegie Institution for Science, Carnegie Mellon University, the Chilean Participation Group, the French Participation Group, Harvard-Smithsonian Center for Astrophysics, Instituto de Astrofísica de Canarias, The Johns Hopkins University, Kavli Institute for the Physics and Mathematics of the Universe (IPMU) / University of Tokyo, Lawrence Berkeley National Laboratory, Leibniz Institut für Astrophysik Potsdam (AIP), Max-Planck-Institut für Astronomie (MPIA Heidelberg), Max-Planck-Institut für Astrophysik (MPA Garching), Max-Planck-Institut für Extraterrestrische Physik (MPE), National Astronomical Observatories of China, New Mexico State University, New York University, University of Notre Dame, Observatário Nacional / MCTI, The Ohio State University, Pennsylvania State University, Shanghai Astronomical Observatory, United Kingdom Participation Group, Universidad Nacional Autónoma de México, University of Arizona, University of Colorado Boulder, University of Oxford, University of Portsmouth, University of Utah, University of Virginia, University of Washington, University of Wisconsin, Vanderbilt University, and Yale University.

Funding for SDSS-III has been provided by the Alfred P. Sloan Foundation, the Participating Institutions, the National Science Foundation, and the U.S. Department of Energy Office of Science. The SDSS-III web site is http://www.sdss3.org/.

SDSS-III is managed by the Astrophysical Research Consortium for the Participating Institutions of the SDSS-III Collaboration including the University of Arizona, the Brazilian Participation Group, Brookhaven National Laboratory, Carnegie Mellon University, University of Florida, the French Participation Group, the German Participation Group, Harvard University, the Instituto de Astrofísica de Canarias, the Michigan State/Notre Dame/JINA Participation Group, Johns Hopkins University, Lawrence Berkeley National Laboratory, Max Planck Institute for Astrophysics, Max Planck Institute for Extraterrestrial Physics, New Mexico State University, New York University, Ohio State University, Pennsylvania State University, University of Portsmouth, Princeton University, the Spanish Participation Group, University of Tokyo, University of Utah, Vanderbilt University, University of Virginia, University of Washington, and Yale University.

Funding for the SDSS and SDSS-II has been provided by the Alfred P. Sloan Foundation, the Participating Institutions, the National Science Foundation, the U.S. Department of Energy, the National Aeronautics and Space Administration, the Japanese Monbukagakusho, the Max Planck Society, and the Higher Education Funding Council for England. The SDSS Web Site is http://www.sdss.org/.

The SDSS is managed by the Astrophysical Research Consortium for the Participating Institutions. The Participating Institutions are the American Museum of Natural His-

tory, Astrophysical Institute Potsdam, University of Basel, University of Cambridge, Case Western Reserve University, University of Chicago, Drexel University, Fermilab, the Institute for Advanced Study, the Japan Participation Group, Johns Hopkins University, the Joint Institute for Nuclear Astrophysics, the Kavli Institute for Particle Astrophysics and Cosmology, the Korean Scientist Group, the Chinese Academy of Sciences (LAMOST), Los Alamos National Laboratory, the Max-Planck-Institute for Astronomy (MPIA), the Max-Planck-Institute for Astrophysics (MPA), New Mexico State University, Ohio State University, University of Pittsburgh, University of Portsmouth, Princeton University, the United States Naval Observatory, and the University of Washington.

This publication makes use of data products from the Wide-field Infrared Survey Explorer, which is a joint project of the University of California, Los Angeles, and the Jet Propulsion Laboratory/California Institute of Technology, and NEOWISE, which is a project of the Jet Propulsion Laboratory/California Institute of Technology. WISE and NEOWISE are funded by the National Aeronautics and Space Administration.

Plot colours were in part based on www.ColorBrewer.org, by Cynthia A. Brewer, Penn State. The TOPCAT tool (Taylor 2005) was also used during this work.