Characterising the X-ray selected AGN Content of the Nearby Universe

Supervisors:
Prof. Mike Watson, Dr James Aird & Dr Rhaana Starling

Thesis submitted for the degree of

Doctor of Philosophy

at the University of Leicester

by

Keir Birchall

Department of Physics & Astronomy

July 2021

Characterising the X-ray selected AGN Content of the Nearby Universe

Keir Birchall

Abstract

Large-scale surveys of the Active Galactic Nuclei (AGN) population performed over the past few decades have highlighted the existence of strong correlations between the central supermassive black hole and their host galaxy. Whilst such samples are incredibly useful, they tend to be incomplete. This skews our understanding of the AGN population as samples consist of largely brighter, more actively accreting objects.

In this thesis, I combined SDSS galaxy samples with XMM-Newton serendipitous sources to measure the incidence of AGN as a function of galaxy properties in the nearby ($z < 0.35$) Universe, carefully accounting for the incompleteness of the X-ray imaging. Regardless of how the galaxy population were split, the probability of hosting an AGN was well described by a power law and highlighted significant amounts of activity at low X-ray luminosities and accretion rates.

I identified 61 AGN in dwarf galaxies, of which 40 have not been previously identified. The probability of hosting an AGN in this regime was unaffected by either the stellar mass or redshift of the host galaxy. Then I studied the wider local galaxy population. As with dwarf galaxies, the probability of hosting an AGN is not affected by stellar mass but it does increases with redshift. When comparing my results to higher redshift measurements, I identified a stellar-mass-dependent distribution of black hole growth rates. Finally, I investigated the effect of star-forming activity. I found that the probability of a galaxy hosting an AGN increases with the star formation rate.

Throughout, I also compared the effectiveness of AGN selection at X-ray and optical wavelengths. I found that the BPT diagnostic is not as effective as X-ray selection. The optical diagnostic incorrectly classified a significant proportion of the X-ray selected AGN sample as star-forming, and was unable to identify weak AGN emission.

Acknowledgements

Pursuing a PhD is challenging but rewarding. When done in the middle of a global pandemic, these experiences are heightened. Thus, I am eternally grateful for the support of so many people, throughout my PhD, but especially in the past 18 months.

My first thanks go to my supervisors: Mike, James and Rhaana. Without your guidance and belief in my ability, none of this research would have been possible. Thanks to your support, I finish this thesis as a confident scientist and look forward to the new discoveries that I could make in the future.

Next, I would like to thank my wonderful housemates, Joe and Nadhya. Whether it was a movie night, board games, or just dancing around the front room, you two were the perfect companions to indefinitely remain indoors with. Thanks also to Dean and Joe for joining us every week to mess up our fictional kingdom with own poor decisions. You know what you did...

Thanks to my fellow PhD students, especially to my PhD siblings, Manika, Jack and Cassandra. Whether it was in the office, or over Zoom, you have always provided interesting discussions and made me feel welcome.

It would be remiss of me not to mention the support of the queer community in Nottingham. Particularly, the spaces created by the fabulous Nana Arthole and Marliyn Sane. They have allowed me to stay sane during a very intense period of my life. I am counting the days until it is safe to go back.

Last but by no means least, I want to thank my parents and my partner, Ross. Your unconditional love and support makes me feel like I can do anything.

Publications

The results presented in chapter 3 have been published in the Monthly Notices of the Royal Astronomical Society:

Birchall, K. L., M. G. Watson, and J. Aird (2020). "X-ray detected AGN in SDSS dwarf galaxies".MNRAS 492.2, pp. 2268–2284

The results presented in chapter 4 have been submitted to the Monthly Notices of the Royal Astronomical Society:

Birchall K. L., M. G. Watson, J. Aird, and R. L. C Starling (Submitted), "The Incidence of X-ray selected AGN in Nearby Galaxies"

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Active galactic nuclei (AGN) have been studied for over a century. In that time the field has grown from a handful of emission line and radio jet observations to the understanding that every galaxy, at some point in its life, will host AGN activity. These advances in our knowledge have been made through careful observation and classification of the AGN population. Despite these advances, however, there are still many critical questions left unanswered. This thesis is focused on trying to understand the relationship between the growth of galaxies and the AGN activity they host. This activity is driven by the accretion of gas and dust onto a supermassive black hole (SMBH) that resides at the galaxy centre, causing the SMBH to grow. Some models argue that the growth of the galaxy through processes like star formation could provide the appropriate conditions for fuelling concurrent SMBH growth. On the other hand, higher mass galaxies have greater quantities of gas and dust, so this could increase the chances of accretion onto the SMBH. Crucially, the precise origin of this material, and what increases its chances of accretion are unclear. By adopting and adapting a novel completeness-corrected identification method discussed later in this chapter, I aim to perform my own careful observations of AGN in the local ($z < 0.35$) Universe and uncover the galactic conditions under which AGN activity is most prevalent. In this chapter I provide an account of some of the key studies and theory required to put the results of my observations into context.

1.1 The First Black Holes

From small perturbations in the initial conditions of the Universe, structures began to form, this is generally referred to as the ΛCDM model (White and Rees 1978). Dark matter coalesced into halos which then gravitationally attracted baryonic gas clouds. These halos then merged together to form proto-galaxies. Black holes also formed rapidly in the early Universe. Observational evidence suggests

that at around $z \approx 7$ there were already quasars of $\log_{10} M_{BH}/M_{\odot} = 9 - 10$ (Bañados et al. 2018; Mortlock et al. 2011). There are several types of possible formation mechanisms that could grow black holes to such high masses so quickly: formation from Population III stars or the direct collapse of proto-galactic gas clouds. These mechanisms are summarised in figure 1.1.

Black hole formation from stellar collapse is possible at all redshifts but Population III stars, the first stellar generation, can only be created in the low metallicity gas of the early Universe. They are thought to be composed from exclusively H & He (Ostriker and Gnedin 1996) and have masses around 100 - 1000 $M_{\odot}$ (Nakamura and Umemura 2001). Thus, their collapse would create a larger black hole seed than if formed from a later generation star (Madau and Rees 2001). However, figure 1.1 shows that to reach the observed masses in this short period, these Population III seeds would have to undergo a series of above-Eddington accretion episodes (Madau et al. 2014). Such high rates of accretion could be avoided if the Population III stars were part of a cluster. In this scenario, the resultant seeds would merge to form more massive black holes (Miller and Davies 2012).

Black holes are also thought to form from the direct collapse of a proto-galactic gas cloud. Low metallicity gas is believed to cool inefficiently, so instead of fragmenting and forming stars, the material collapses into a higher mass black hole seed (Begelman et al. 2006). Crucially, however the low angular momentum gas required for this method of formation only existed in sufficient quantities in the early Universe. As the Universe evolved and disk-like structures began to form, the gas within would gain angular momentum and become less likely to collapse and form a black hole. For a more detailed discussion of these formation mechanisms, see Latif and Ferrara (2016).

To determine which formation mechanism dominated in the early Universe, we can look for their observational consequences in the present day dwarf galaxy population. Dwarf galaxies are well suited to this task because, unlike their higher mass counterparts, they have experienced a much calmer merger history (Reines and Comastri 2016), making the observational signatures clearer in this mass regime (e.g. Bellovary et al. 2011). For example, Volonteri et al. (2008) shows that at higher velocity dispersions, and thus stellar masses, the fraction of galaxies hosting black holes is the same regardless of formation mechanism. These high mass structures grew via numerous merging and accretion episodes which would remove any direct observational correlation between the black hole and how it formed. However, some early galaxies will have evolved without being merged into a larger counterpart.

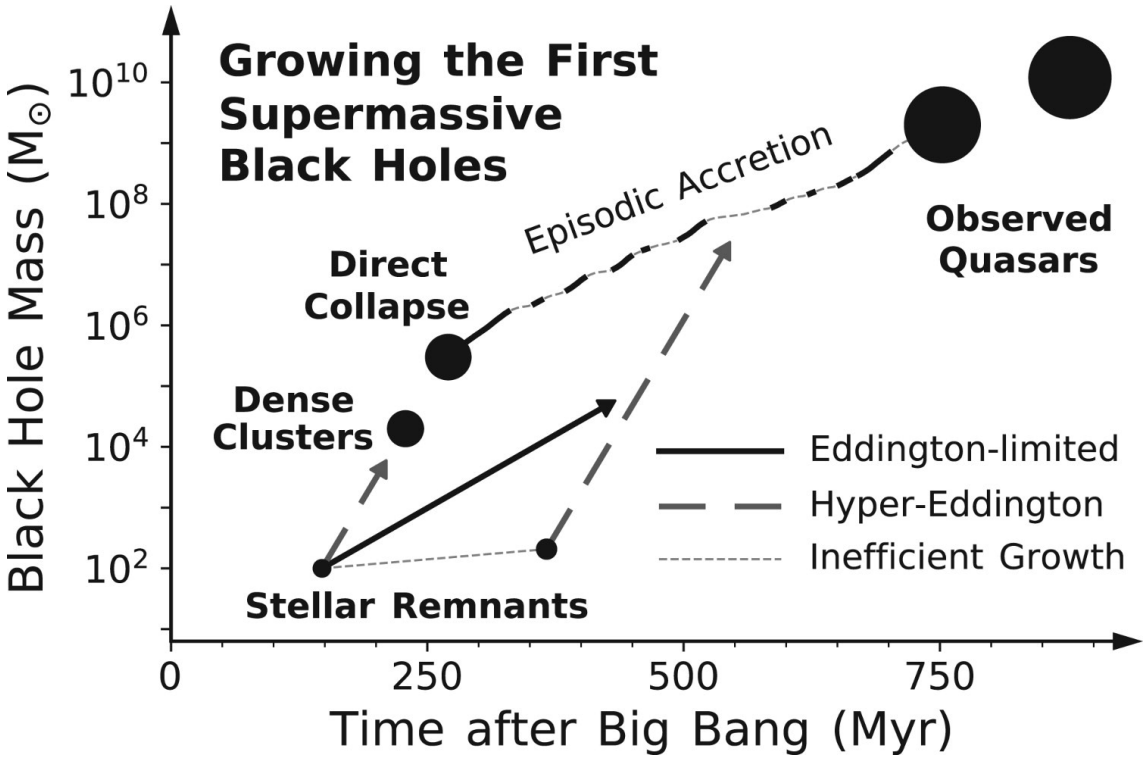

Figure 1.1: Illustration of the relative amounts of growth required by each proposed black hole formation mechanism to form $10^9 - 10^{10}$ $M_\odot$ quasars at $z \approx 6 - 7$. Source: *Smith and Bromm (2019)*

Thus, present day low mass galaxies will retain some signature of the mechanism by which its black hole was formed. By calculating the black hole occupation fraction in this low mass regime, we can constrain the dominant formation mechanism. Since formation by direct collapse requires low angular momentum gas, the redshift regime within which it could dominate would be limited. If this were the dominant seeding mechanism then the fraction of dwarf galaxies hosting black holes is expected to be much lower than if the Population III stellar seeding mechanism was dominant (Greene 2012). For a more in-depth review of the observational evidence for black hole seeds in dwarf galaxies, see Reines and Comastri (2016).

1.2 AGN Structure

Identifying black holes in the early Universe, low mass galaxies or any other environment is difficult as, by definition, black holes cannot be directly observed. Instead we have to infer their existence from AGN emission or the black hole's effect on the host galaxy. To understand the origin of this emission we must first consider the structure of an AGN.

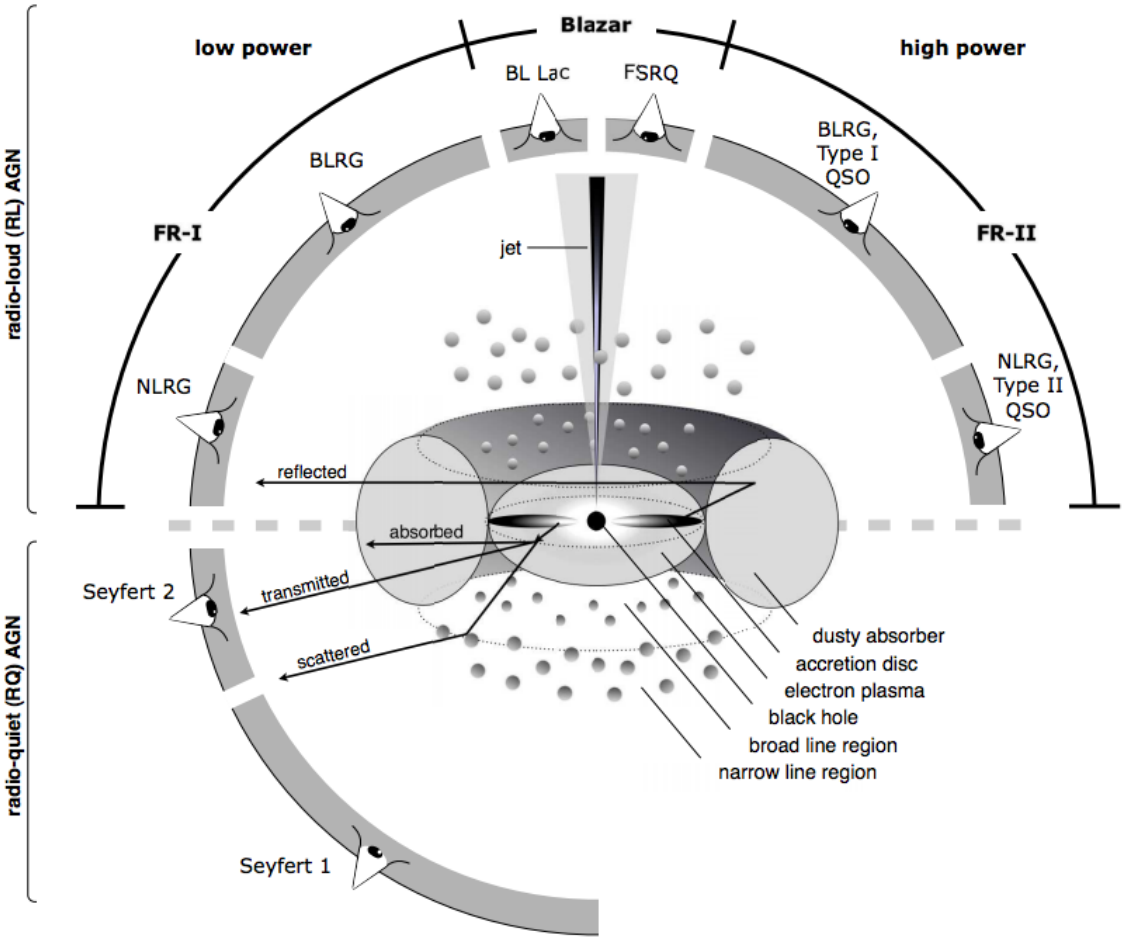

Figure 1.2: The unified model of an AGN (Urry and Padovani 1995) with the expected AGN classification highlighted at the appropriate viewing angle. Source: *Beckmann and Shrader (2012)*

In figure 1.2 we can see an illustration of the 'unified model' (Urry and Padovani 1995) for an AGN. The model aims to unify the full range of AGN observations into a single structure. It argues that the full range of AGN activity can be explained by the presence or absence of radio activity, and the observer's viewing angle. In this scheme, AGN are classified as type-1 if the central region is within the observer's line-of-sight. Otherwise, it is classified as type-2.

1.2.1 Central SMBH

At the centre of this structure lies the SMBH whose gravitational potential energy is the primary driver of AGN emission. Material is accreted onto the SMBH, transforming the gravitational potential energy into electromagnetic energy. This process, and the resulting mechanisms, produces a unique spectral energy distribution (SED), shown in figure 1.3. Soltan (1982) outlined the relation between AGN

luminosity, L, and the mass accretion rate, $\dot{m}$,

$$L = \nu \dot{m} c^2 \tag{1.1}$$

where ν is the accretion efficiency and c is the speed of light. By fitting a radiatively-efficient accretion disk model to optical quasar observations, Davis and Laor (2011) suggest that SMBH rotation has a strong effect on the conversion efficiency. They find non-rotating SMBHs have $\approx 6\%$ conversion efficiency, whereas for maximally rotating SMBHs this rises to $\approx 29\%$.

The resultant electromagnetic emission exerts a radiation pressure on the incoming material. Accretion onto the SMBH can continue as long as the gravitational force is greater than outward force due to radiation pressure. The maximum luminosity at which accretion can occur is the Eddington luminosity, L_{Edd}, and is given by,

$$L_{Edd} = \frac{4\pi G M m_p c}{\sigma_T} \approx 1.25 \times 10^{38} (\frac{M}{M_\odot}) \text{ erg s}^{-1} \tag{1.2}$$

where G, m_p and σ_T are the gravitational constant, the proton mass and the Thompson cross section respectively. This limit assumes a spherically symmetric accretion of a fully ionised gas.

1.2.2 Accretion Disk

One of the main factors preventing accretion onto the SMBH is the gas' angular momentum. To conserve this quantity, the incoming gas orbits the SMBH forming an accretion disk. Gas in the accretion disk can lose angular momentum through viscous or turbulent processes. As this occurs, the gas moves inwards, angular momentum moves outwards and the temperature of the disk increases.

Emission from a typical optically thick, geometrically thin accretion disk (Shakura and Sunyaev 1973), is described by a black body spectrum. This is thermal emission that arises from a perfect emitter and depends only on the temperature of the body. The black body spectral energy distributions can be described by Planck's law, a continuous frequency spectrum with the peak of emission directly related to the body's temperature. Assuming all the energy released when material moves from radius r to $r + dr$ in an disk accreting at a rate, $\dot{m}$ and orbiting a black hole of mass M is radiated as a black body then,

$$\frac{GM\dot{m}dr}{r^2} \approx \sigma \; 2\pi r dr \; T^4 \tag{1.3}$$

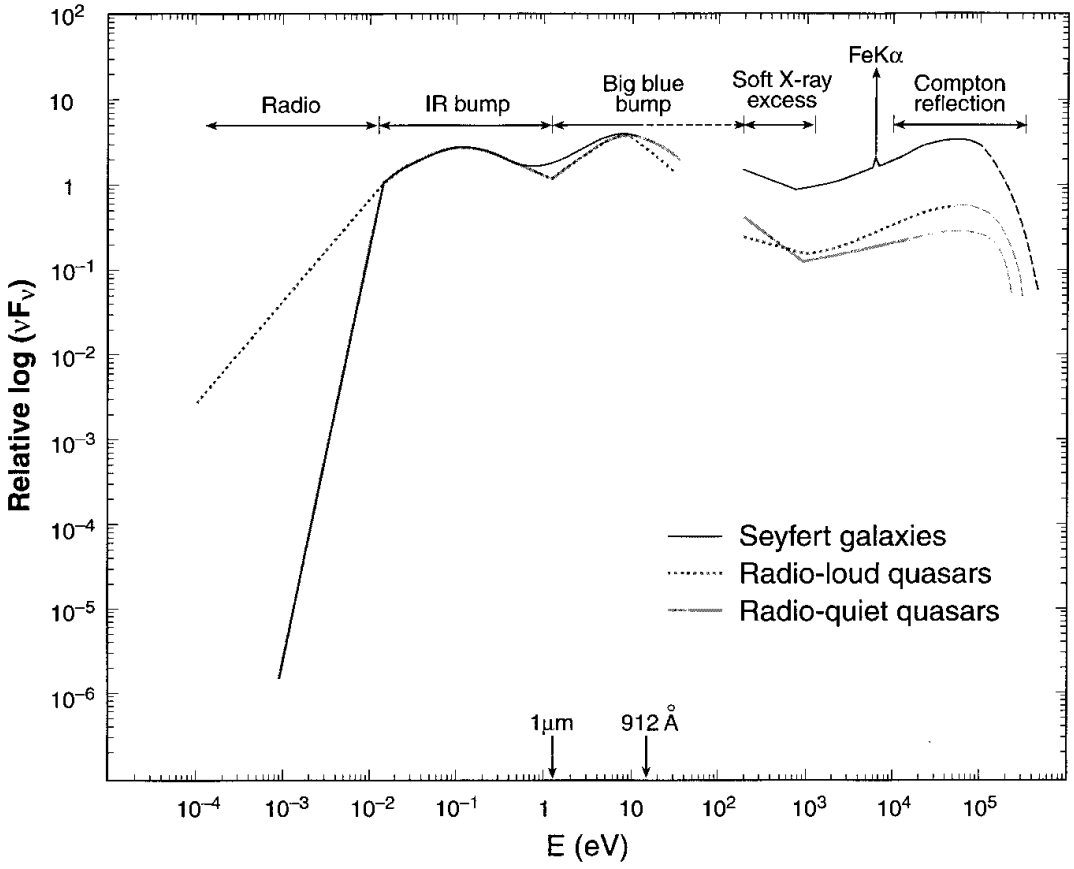

Figure 1.3: The spectral energy distribution (SED) for different types of AGN. Source: *Koratkar and Blaes (1999)*

where σ is the Stefan-Boltzmann constant. Thus the temperature, T in an annular black body follows $T \propto r^{-3/4}$. To obtain the full emission spectrum, we integrate across the full disk radius. The total emission spectrum can be interpreted as a multi-colour black body emission, with the highest temperature found at the innermost stable circular orbit. For a SMBH of stellar mass 10^8 $M_\odot$, the temperature of the innermost disk region is $\approx 10^5$K. Thus accretion disk emission dominates in the UV region and is responsible for the "big blue bump" in the AGN SED, shown in figure 1.3.

1.2.3 X-ray Corona

The X-ray corona is composed of a hot cloud of electrons very close to the centre of the accretion disk. As can be seen in figure 1.4, emission from the corona produces a unique X-ray spectrum. Its underlying form is a power law. Lower energy photons emitted by the accretion disk interact with the relativistic electrons that make up the corona. This inverse Compton scattering boosts the interacting photons to higher energies and produces this continuum of hard X-ray emission. As we reach a few hundred keV, however, the power law is subject to a sharp cut-off. Significantly fewer photons are boosted up to these energies as the electrons performing the scattering are only as energetic as the overall temperature of the corona (Haardt and Maraschi 1991).

There are additional features overlaid on the underlying power law form. Above 6 keV, are the iron Kα line and Compton reflection hump. Both these features are caused by corona and accretion disk radiation reflecting off other material in the AGN. For the iron Kα line, the radiation causes iron in the broad line region (BLR) to fluoresce. The Compton reflection hump occurs at 30-40 keV and is thought to be caused by the superposition of reflection from material in both the inner part of the accretion flow, and material further away in the BLR or torus (Nandra et al. 1989; Pounds et al. 1989, 1990).

At lower energies, the AGN X-ray spectrum is dominated by the "Soft Excess". Its origin is uncertain. It could be caused by blurred reflection in the inner parts of the accretion disk, or a second cooler corona scattering to lower energies (Crummy et al. 2006; Done et al. 2012; Petrucci et al. 2018).

Figure 1.2 shows the "electron plasma" of the X-ray corona as a sphere in the central region, however, the exact shape of this sea of energetic electrons is unclear. There are several proposed models: a hot sphere surrounding the SMBH (Zdziarski et al.

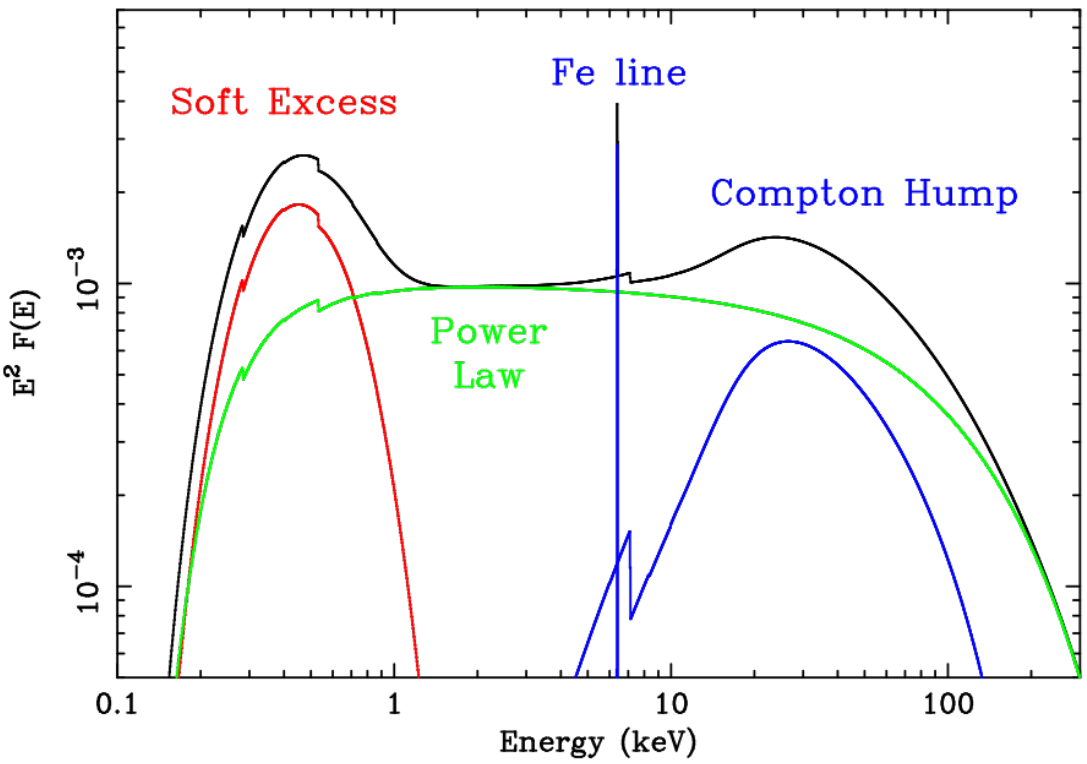

Figure 1.4: The main components of the X–ray spectrum of unobscured AGN are shown: the soft excess whose origin is uncertain (red); power law from Comptonization of photons from the accretion disc (green); reflection continuum and narrow Iron (Fe) line due to reflection of the hard X–ray emission from dense gas (blue). Source: *Fabian and Miniutti (2005)*

1994); parallel planes covering the accretion disk (Haardt and Maraschi 1991; Haardt et al. 1994); a hot sphere with an inner warm disc (Matt et al. 1991; Miniutti and Fabian 2004); or an extended sphere and plane structure (Petrucci et al. 2013).

1.2.4 Emission Line Regions

Further out from the SMBH are the emission line regions. Photons from the accretion disk photo-ionise the material in these regions. Soon after, the emitted electrons re-combine to produce photons at characteristic wavelengths. The aggregate effect of this ionisation and re-combination produces the prominent optical emission lines that characterise these regions. However, there are distinct changes in the form and combination of emission line spectra with increasing distance from the central SMBH.

Closer to the SMBH is the broad-line region (BLR). It is filled with high-density ($> 10^9$ cm^{-3}) gas predominantly emitting lines in the Hydrogen Balmer and Lyman series. By observing the delay between UV variations (originating in the accretion

disk) and optical variations in the same AGN's spectrum, a process called Reverberation Mapping, we can estimate that the emission line region tends to be a few to hundreds of light days in radius. At these distances, the speed of the gas' rotation around the central SMBH causes Doppler broadening of BLR emission lines up to thousands of km/s.

Further out, the gas in the narrow line region (NLR) slows down significantly, with gas velocities of hundreds of km/s, hence this region lacks broad emission lines in its spectrum. The gas in the NLR is also significantly less dense ($\approx 10^4$ cm^{-3}). Such low densities mean very few atomic collisions occur which allows the existence of forbidden emission lines, alongside Hydrogen Balmer lines, in the NLR spectrum (Netzer 2013).

1.2.5 Obscuring Torus

Along some lines of sight, shown in figure 1.2, the optical and UV radiation is obscured by an optically thick torus. As noted earlier, this is the crucial structure required to divide the AGN population between type-1 and type-2. If viewed along the obscured lines of sight then the emission from the central region is absorbed by the toroidal material and emitted at IR wavelengths. This forms an "IR bump" in the AGN SED, shown in figure 1.3, peaking between 10 - 30 μm.

Initial studies of this structure considered a simplified torus model with a homogeneous dust distribution. However, subsequent work has found the picture to be much more complex. From both IR and X-ray observations, the gas in the torus has been shown to be clumpy (Markowitz et al. 2014; Ramos Almeida et al. 2009). Detailed ALMA observations have also shown evidence to suggest that the torus funnels gas into the central engine, and back out to the host galaxy's circumnuclear disk (Garcıa-Burillo et al. 2016; Müller Sánchez et al. 2009). An increasing number of AGN also show evidence of extended polar gas flows (López-Gonzaga et al. 2016) which, when included alongside a nuclear dust disk, could form the shape of the obscuring structure.

1.2.6 Radio Jets

As outlined earlier, the other major classification parameter is whether the AGN exhibits strong, non-thermal radio emission. Approximately 10% of optically-selected AGN are radio-loud (Ivezić et al. 2002). The jets that characterise these radio-loud AGN are believed to be produced by the SMBH's spin. As the highly conducting material of the accretion disk rotates around the SMBH, a magnetic field is generated.

This changing magnetic field produces a strong electric field capable of accelerating the plasma produced by the accretion disk. These electrons spin around the dynamic magnetic field lines, are accelerated to relativistic velocities and produce jets that can extend out to kiloparsec scales (Blandford and Znajek 1977).

1.3 Extragalactic Sources of X-ray Emission

Whilst AGN are capable of producing emission across the electromagnetic spectrum, this thesis focuses on identifying AGN using their X-ray emission. Studying in this wavelength presents a number of advantages to identifying AGN activity, but there are other significant populations of X-ray emitting objects that have to be taken into account before a detection can be confirmed.

1.3.1 Non-AGN X-ray Emission

There are several non-AGN phenomena that produce significant amounts of X-ray emission. It is important to understand their origin so their effect on the galactic X-ray luminosity function (XLF) can be modelled and accounted for.

Galaxies are known to have significant amounts of hot-ionised gas at sub-keV temperatures. This is a source of copious amounts of X-ray emission (e.g. Grimes et al. 2005; Li and Wang 2013). The morphology of this gas suggests that it is associated with galactic outflows from supernovae and winds from massive stars (e.g. Chevalier and Clegg 1985). For example, Tyler et al. (2004) traced mid-IR and Hα emission across a sample of spiral arms and found strong correlations between these sites of star-formation and the diffuse X-ray emission. Thus, the total gas luminosity has been found to correlate strongly with the host galaxy's SFR (e.g. Mineo et al. 2012b).

The X-ray output of a typical galaxy is dominated by the emission of its X-ray binary (XRB) population. They consist of a compact object - typically a neutron star or stellar mass black hole - accreting matter from a donor star. X-ray emission is produced through accretion processes similar to those occurring in an AGN. XRBs can be split into low- and high-mass categories, depending on the mass of the donor star. Low-mass-XRBs (LMXBs) are fairly long-lived as they accrete from their donor star on longer timescales, thus these objects are typically found in older stellar populations. High-mass-XRBs (HMXBs), on the other hand, accrete from their donor stars much more rapidly so are typically found in younger stellar populations (Tauris and van den Heuvel 2006). Given these differing accretion timescales, the contributions that HMXBs and LMXBs each make to the galactic XLF are cor-

related with different galaxy properties. For example, Lehmer et al. (2016) correlate the LMXB contribution with the galaxy's stellar mass, the HMXB contribution with the SFR and also a redshift component to account for the changing metallicity of the galaxy. From this model, they match with typical XRB population luminosities of between $10^{37} - 10^{39}$ erg s^{-1} (Fabbiano 2006).

The brightest non-AGN source of X-ray emission is the ultraluminous X-ray source (ULX). They are typically off-nuclear X-ray point sources with observed X-ray luminosities generally exceeding 10^{39} erg s^{-1}. ULXs are thought to be distinct from XRBs because these luminosities are too bright to be associated with accretion onto neutron stars or stellar mass black holes (Miller et al. 2004). Assuming that the radiation is emitted isotropically, ULX emission could be evidence for the existence of intermediate mass black holes (Fabbiano et al. 2001). However, it could also be explained by beamed radiation from a stellar mass black hole (King et al. 2001). ULXs are thought to be relatively rare. Where they are found, no more than one ULX, on average, was hosted in that galaxy. Thus, their contribution the host's XLF can be folded into the high luminosity end of the XRB's XLF (Swartz et al. 2004).

1.3.2 Identifying AGN Activity

X-ray surveys are particularly suited to identifying AGN activity for a number of reasons. First, X-ray emission appears to be nearly universal in luminous AGN. Even when identified at other wavelengths they almost always show signs of X-ray AGN emission as well (Avni and Tananbaum 1986; Gibson et al. 2008; Mushotzky 2004). Second, X-ray emission can easily penetrate the substantial amounts of absorbing material that surround the AGN. This emission is believed to travel through hydrogen column densities up to $N_H \approx 10^{24}$ cm^{-2}, above which the AGN are classified as Compton-thick. This ability to penetrate absorbing material is crucial as a significant portion of AGN are believed to be obscured. The column density distribution for the obscured AGN population has a log-normal shape, peaking around 10^{23} cm^{-2} (Tozzi et al. 2006). Third, X-ray emission from AGN activity tends to dominate over the stellar processes previously discussed. Finally, the distinct shape of the AGN's X-ray spectrum can aid their detection, and help isolate a purer sample of active galaxies. Given these advantages, it has been argued that this wavelength produces the most complete samples which are best suited to studying the AGN population (e.g. Brandt and Alexander 2015; Mushotzky 2004). As a result, there are numerous ways in which X-rays have been used to identify AGN emission. Brandt and

Alexander (2015) summarised the most popular as follows:

- AGN emit strongly in the X-ray band, so sources with X-ray luminosity greater than 3×10^{42} ergs^{-1} are good candidates as the star-formation rates necessary to reach such high X-ray luminosities exceed 200 $M_{\odot}$ yr^{-1} (Ranalli et al. 2003)
- Numerous studies have established relations between X-ray luminosity and SFR for galaxies without AGN (e.g. Lehmer et al. 2016; Mineo et al. 2014). X-ray sources that lie at a significant distance above these thresholds can be considered AGN
- X-ray emission from an AGN experiences little dilution from starlight so sources with high X-ray:Optical/NIR flux ratios are strong AGN candidates
- Given the unique structure of the AGN's spectrum, there are features that can be used to identify AGN activity: for example, a flat effective power law or high hardness ratios
- Rapid X-ray variability of a significant magnitude is commonly seen in AGN where emission from regions close to the SMBH is directly observed
- AGN are also, typically (but not exclusively (Reines et al. 2020)), coincident with the apparent nucleus of the host galaxy

Whilst these techniques have been used to identify large and secure AGN samples, they can still be biased towards SMBHs that dominate over galaxy emission. This presents an issue if we want to understand AGN activity in lower mass galaxies, as low mass black holes that are detected are the much rarer, more actively accreting ones. Combining some of these techniques, however, has facilitated a huge increase in the number of AGN detected in dwarf galaxies. For example, X-ray selected AGN in dwarf galaxies can be identified through isolating centrally located emission that cannot be explained by other sources, including X-ray binaries and hot gas (e.g. Baldassare et al. 2017; Lemons et al. 2015; Mezcua et al. 2018; Pardo et al. 2016). Generally, the most secure detections are those that meet several of the above criteria, or successfully match the X-ray signal to a multi-wavelength galactic counterpart.

1.4 The Connection between AGN and their Host Galaxies

It is in part due to the application of these selection techniques that we now believe SMBHs are ubiquitous at the centres of the most massive galaxies in the Universe. This ubiquity drove the idea that the SMBH and host galaxy might influence one another as they evolve (Kormendy and Ho 2013). For example, galaxies grow by forming stars, a process driven by the availability of cold gas. This material is also thought to fuel the growth of SMBHs (Alexander and Hickox 2012). However, a lot of uncertainty remains around the nature of this co-evolution. Large scale surveys of AGN activity across the electromagnetic spectrum have been employed to shed light on this relationship. By exploring how the incidence of AGN changes with the SFR of its host galaxy we have uncovered several key pieces of evidence that suggest the growth of SMBHs and their host galaxies are connected.

One of the most well-constrained relationships of this nature is the correlation between a SMBH's mass and the luminosity of the host galaxy's classical bulge. The first signs of this correlation appeared in small scale studies (e.g. Dressler and Richstone 1988; Kormendy and Richstone 1992). Magorrian et al. (1998) was the first study to observe this correlation using a larger sample; a mixture of ground and space-based observations of a sample of 32 galaxies with their black hole masses measured using dynamical techniques. Soon after, similarly strong correlations between the related bulge quantities like mass and velocity dispersion (e.g. Ferrarese and Merritt 2000; Gebhardt et al. 2000) were found. A sample of kinematically detected SMBHs displaying these correlations can be seen in figure 1.5. Observations such as these have been used to argue for the existence of historic, correlated SMBH-galaxy growth.

Another strong piece of evidence for the existence of the correlation between SMBH and galaxy growth can be seen in figure 1.6. It shows several well-known determinations of black hole accretion rate density (BHARD) distributions with redshift, and compares them to the best-fit star formation history (SFH; solid, black line) (Madau and Dickinson 2014). Using Soltan's argument outlined in equation (1.1), the observed AGN luminosity can be used to estimate the amount of accretion onto a SMBH. Aird et al. (2010) derived the BHARD distribution from measurements of the 2-10 keV X-ray luminosity function across a large redshift range (green region). This distribution appears to peak later and decline more rapidly than the best-fit SFH. Bolometric AGN luminosity functions from Shankar et al. (2009) (red) and Delvecchio et al. (2014) (blue) are also shown in figure 1.6. These studies appear to

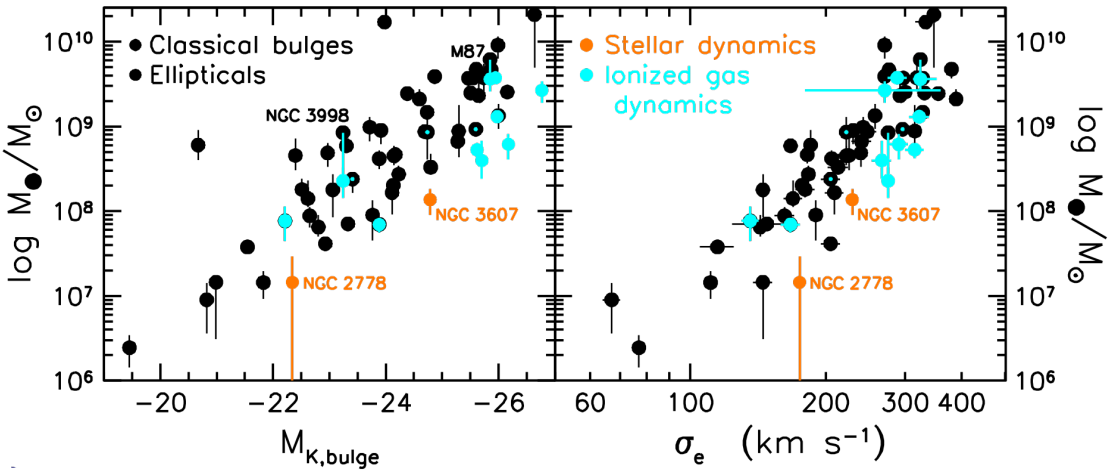

Figure 1.5: A sample of kinematically-detected SMBHs displaying strong correlations between black hole mass, $M_\bullet$, K-band bulge magnitude, $M_{K,bulge}$, and velocity dispersion σ_e. Source: *Kormendy and Ho (2013)*

track the SFH more closely. Both the BHARD distribution and SFH increase from $z = 6$, peak at around $z = 2$, and then decline to about a tenth of their peak in the present-day Universe.

When directly attempting to connect the AGN activity with the host galaxy's SFR, however, the results are less clear. On the one hand, numerous studies found that the average SFR of AGN-hosting galaxies increased out to at least $z \approx 3$ (Harrison et al. 2012; Mullaney et al. 2012b; Rosario et al. 2012, 2013) consistent with the results in figure 1.6, and that the SFR is found to tightly correlate with the average AGN luminosity (e.g. Chen et al. 2013; Mullaney et al. 2012a). On the other hand, it was found that AGN at a fixed X-ray luminosity can have a broad range of SFRs (e.g. Alexander et al. 2005; Mullaney et al. 2010) - in some cases covering up to 5 orders of magnitude (Rafferty et al. 2011). Thus, it became increasingly clear that there exists a strong connection between star formation and SMBH growth but not a correlation between the SFR of individual galaxies and their AGN. Hickox et al. (2014) tried to reconcile these contradictory observations using the relative stability of these processes. AGN activity varies on shorter timescales, typically on Myr timescales, whereas star formation typically varies over much longer periods, $\geq$ 100Myr. To test this hypothesis, they constructed a model that assumed the long-term SMBH growth rate is exactly proportional to the SFR, but allowed the observed X-ray luminosity to vary by orders of magnitude on shorter timescales. Their model successfully reproduced the proportional relationship between SFR and average AGN luminosity, showing that short-term variations can disguise an underlying long-term correlation, such as that between AGN activity and the growth of the host galaxy.

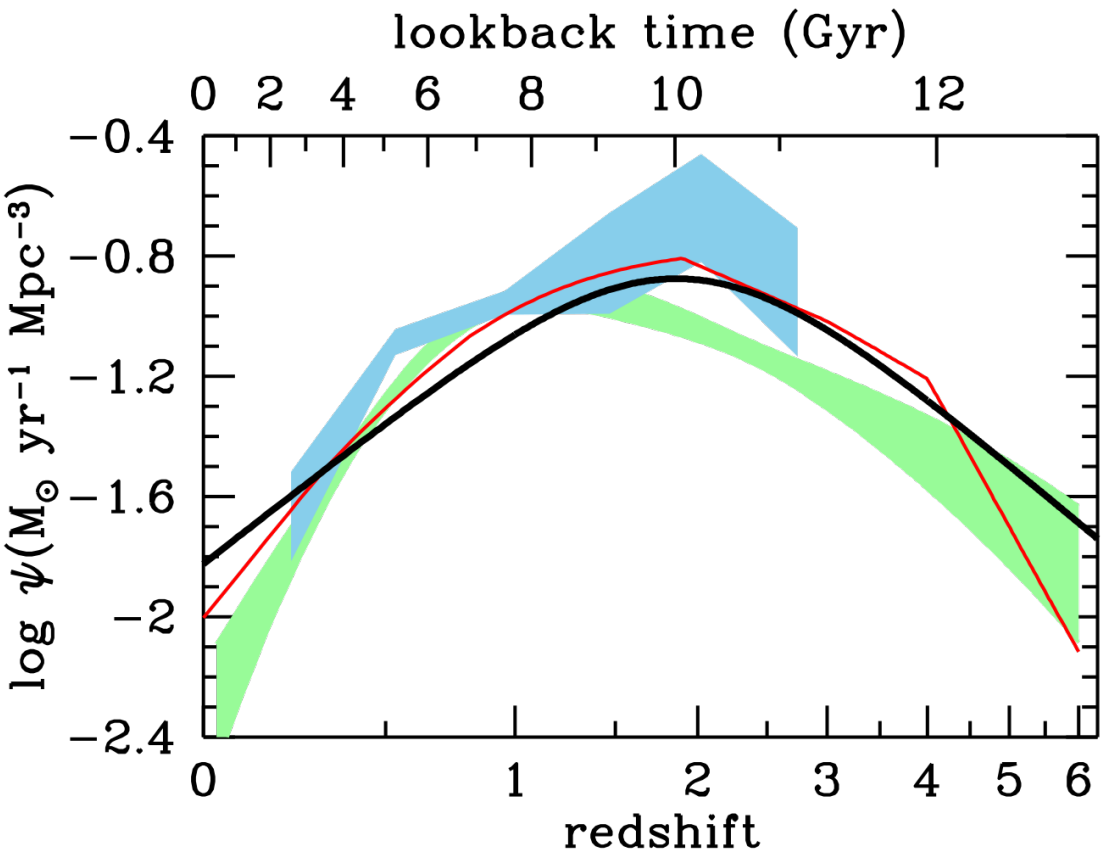

Figure 1.6: Comparison of best-fit star formation history (solid black line) with the black hole accretion rate density distributions calculated from X-ray (Shankar et al. (2009) - red line; Aird et al. (2010) - green region) and IR (Delvecchio et al. (2014) - blue region) data. Source: *Madau and Dickinson (2014)*

1.5 Completeness-corrected AGN Samples

Taking large samples of galaxies and identifying the AGN activity therein has been very effective at highlighting the relationship between the SMBH and host galaxy. If the sample is sufficiently big, then it is possible to smooth out stochastic differences and try to identify trends (Aird et al. 2013; Delvecchio et al. 2020; Hickox et al. 2014). However, studies that take this approach often lack sufficient correction for the limitations of the survey data used. To better understand this relationship, it is important to model and correct for any sample incompleteness.

Aird et al. (2012) was one of the first AGN population studies to attempt to correct for these observational limitations. They started with a sample of galaxies out to $z = 1$, from the Prism Multi-object Survey (Coil et al. 2011; Cool et al. 2013) and identified AGN activity using X-ray data from Chandra and XMM-Newton. In this investigation, they aimed to determine the probability of finding AGN activity as a function of host galaxy colour, stellar mass and redshift. However, the varying sensitivity of their X-ray data prevented them from detecting low luminosity or high redshift AGN activity. Since the observed distribution is not reflective of the underlying AGN population they created a method which corrects for this imperfect detection process. They estimated the probability of detecting AGN activity of luminosity, $L_{X,i}$ hosted in a galaxy at redshift, z_i, - $p_{\rm det}(L_{X,i}, z_i)$ - by calculating the fraction of the survey sky area that is sensitive to the resulting flux, $f_X(L_{X,i}, z_i)$. It has the form,

$$p_{\rm det}(L_{X,i}, z_i) = \frac{A(f_X(L_{X,i}, z_i))}{A_{\rm total}} \tag{1.4}$$

where $A(f_X)$ is the survey sky area sensitive to a given flux, and $A_{\rm total}$ is the total survey sky area. These probabilities were calculated for each galaxy in a given mass and redshift bin, and their sum used to correct the observed AGN count in that bin.

Aird et al. (2012) showed that the probability of hosting an AGN can be described by a power-law distribution of specific black hole accretion rates (sBHAR). This quantity has the form,

$$\lambda_{\rm sBHAR} = \frac{25 L_{2-10{\rm keV}}}{1.26 \times 10^{38} \times 0.002 M_*} \approx \frac{L_{\rm bol}}{L_{\rm Edd}} \tag{1.5}$$

They considered $\lambda_{\rm sBHAR}$ as a tracer of the rate of black hole growth relative to its stellar mass, not an Eddington ratio. In doing so, they could present an Eddington-

ratio-scaled accretion rate but disregard any uncertainties in the assumption of black hole mass.

Figure 1.7 shows the key results from Aird et al. (2012). Overall, they find that the AGN population is well described by a power law with most accretion onto the black hole occurring at low rates. These distributions were found to be consistent over a broad range of stellar masses ($9.5 \leq \log_{10} M_*/M_\odot \leq 12$) but with a normalisation that evolves strongly with redshift. This shows us that the probability of hosting an AGN with a given sBHAR is independent of stellar mass but drops rapidly between $z \approx 1$ and the present day.
The effect of the completeness corrections can be clearly seen in the bottom-right panel of the lower plot in figure 1.7. Aird et al. (2012) contrast the AGN probability distribution (dashed black line) for the full stellar mass range with a histogram describing the observed AGN distribution (to which no completeness corrections have been applied). It is clear that the observed distribution misses a significant amount of low-level AGN activity. Thus, the application of completeness corrections, as pioneered in Aird et al. (2012), is crucial to understanding the relationship between SMBHs and their host galaxies.

Similar approaches to Aird et al. (2012) have been adopted by subsequent AGN population studies and adapted to suit the surveys used. These studies tend to agree on the overall shape of the AGN distribution, with activity increasing towards lower sBHARs. However, this distribution has been modelled using a simple power law (e.g. Aird et al. 2012; Bongiorno et al. 2012; Wang et al. 2017), a broken power law (e.g. Bongiorno et al. 2016; Yang et al. 2017, 2018) and with complex non-parametric models (e.g. Aird et al. 2017, 2018, 2019; Georgakakis et al. 2017).

There is also agreement around the general decrease in AGN activity towards lower redshifts (e.g. Aird et al. 2018; Bongiorno et al. 2016), and that the incidence of AGN increases in star-forming galaxies (e.g. Aird et al. 2019; Azadi et al. 2015; Georgakakis et al. 2014). The effect of stellar mass, however, is unclear. Bongiorno et al. (2012) recreated the mass-independent distributions from Aird et al. (2012) at higher redshifts, whereas Georgakakis et al. (2017), Yang et al. (2018) and Aird et al. (2018) find evidence of differential stellar mass growth rates when probing a wider range of masses and redshifts. Furthermore, since stellar mass, redshift and SFR are correlated quantities, it can be difficult to disentangle these effects and uncover the precise nature of the relationship between the AGN and its host galaxy. This completeness-corrected approach has produced some promising results, but more work is needed, with bigger samples to effectively disentangle the correlation

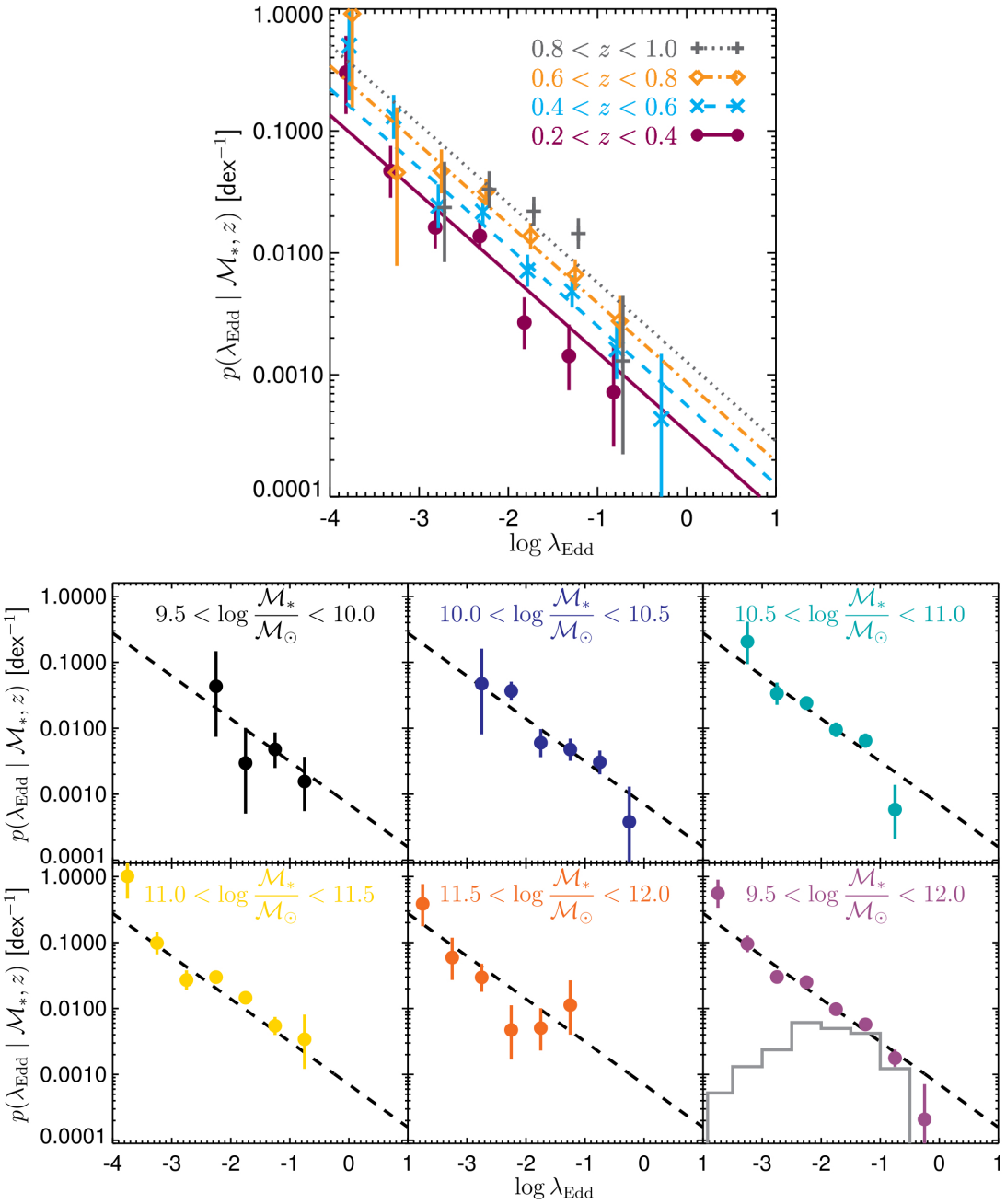

Figure 1.7: Results from Aird et al. (2012) showing the probability of finding AGN activity, $p(\lambda_{\rm Edd}|M_*, z)$, as a function of Eddington ratio, $\lambda_{\rm Edd}$. Above, the sample was split into redshift bins, and evaluated at the centre of that bin. Below, the sample has been split into stellar mass bins, and evaluated at the median sample redshift, $z = 0.6$. The bottom-right panel fits the calculated AGN probability (coloured points) for the full sample mass range with a power law (dashed black line) and compares it to the histogram describing the observed AGN distribution (to which no completeness corrections have been applied).

between these properties.

1.6 This Thesis

Robust measurements of the AGN X-ray luminosity and Eddington ratio distributions at the lowest redshifts would provide a useful benchmark to understand how this population evolved through cosmic time. Unfortunately, none of the papers outlined in section 1.5 focus on this region. There are a few studies that attempt to characterise the AGN population at low redshifts (e.g. Kauffmann and Heckman 2009; Schulze et al. 2009) however they are subject to issues which have been discussed previously. Firstly, they use optically-selected AGN samples which are not as complete as X-ray-selected samples. Secondly, these studies of the nearby AGN population make little attempt to correct for their incompleteness. This would bias the results towards the brightest, most actively accreting AGN.

Thus in this thesis I perform a robust quantification of the X-ray selected AGN population in the nearby Universe ($z \leq 0.35$). I combine galaxy samples from the Sloan Digital Sky Survey data release 8 (SDSS-DR8) with XMM-Newton serendipitous sources (3XMM-DR7) to measure the incidence of AGN as a function of galaxy properties in the nearby ($z < 0.35$) Universe, carefully accounting for the incompleteness of the X-ray imaging.

Chapter 2 provides a discussion of the instruments used, how they detect objects and outline the data I took from them to perform this work. Chapter 3 focuses on the AGN content found in nearby dwarf galaxies. I calculate the fraction of galaxies that host an AGN as a function of stellar mass and redshift. Chapter 4 extends the work of the previous chapter by calculating these quantities for the whole nearby AGN population. Chapter 5 focuses on the effect star formation in the host galaxy has on the nearby AGN population. Finally, chapter 6 summarises the results of this thesis and discusses possible future avenues of study.

Throughout this thesis, I assume Friedmann-Robertson-Walker cosmology: $\Omega = 0.3$, $\Lambda = 0.7$ and $H_0 = 70$ km s^{-1}Mpc^{-1}.

Chapter 2

Observational Data & Instruments

Performing a robust survey of the AGN content of the local Universe requires high quality data appropriate for the task. For each study presented in this thesis I draw on the galaxy surveys produced by the SDSS and the serendipitous source catalogues from XMM-Newton. In this chapter I will explain why these surveys were chosen, outline how they collect data and describe the data sets they produced.

2.1 Sloan Digital Sky Survey

The Sloan Digital Sky Survey (SDSS) is an astronomical project which operates several telescopes located across both hemispheres. The primary telescope in this network is the Sloan Foundation 2.5m telescope located at Apache Point Observatory in New Mexico. It has been observing for for over 20 years, having seen first light in May 1998.

Data began being recorded in 2000. Over that period there have been many different observational priorities. SDSS-I, the first phase of observations between 2000 - 2005, focused primarily on taking images in five bandpasses over > 8000 deg^2 of sky. Around $5,700$ deg^2 of this region also contains galactic spectra. SDSS-II (2005 - 2008) expanded on this work and completed imaging of half the northern sky. It also completed two primarily spectroscopic projects: a study of supernovae at the celestial equator, and a kinematic and stellar population study of the Milky Way (SEGUE). After phase II the imaging camera was retired. Thus SDSS-III, IV and V all focus on spectroscopic studies. Some of the targets included the Milky Way (e.g. APOGEE), nearby galaxies (e.g. MaNGA, BOSS) and exoplanets (MARVELS). This thesis makes use of the Legacy survey, a statistically complete, magnitude-limited galaxy survey produced in SDSS-I and II. Whilst the imaging camera was retired and the spectrograph updated several times, what follows is a description of the instruments used to make these observations, and how these data

were subsequently analysed.

2.1.1 The Sloan Foundation Telescope

The Sloan Foundation telescope uses two hyperbolic mirrors: a 2.5-metre wide primary and a 1.08-metre secondary in the Ritchey-Chrétien arrangement. Light enters the telescope, is reflected off the primary mirror, into the secondary which then directs it through the central 1.17 metre hole onto the focal plane. By adding two corrector lenses to this construction, the Sloan Foundation telescope achieves a 3° distortion-free field of view (FoV).

The imaging camera collects photometric data using an array of thirty 2048×2048 pixels CCDs. As can be seen in figure 2.1, they are arranged in 6 columns, each containing 5 CCDs. Each row corresponds to an SDSS filter, ordered as follows: *r, i, u, z* and *g*. Table 2.1 has more information about the imaging camera properties. An additional 24 CCDs are placed before and after the photometric arrays to collect astrometric data by focusing on bright reference stars. This camera operates on a drift scan mode. Data is collected as the camera slowly reads the CCDs. At the same time, the telescope moves along great circles in the sky so that the objects move across each CCD row. An object takes 54 seconds to move from the beginning of a CCD to the end. However, given the gaps between CCDs, it takes an image 71.7 seconds to move from the beginning of one row to the next. Thus, whilst the effective exposure time is 54 seconds, there are 71.7 seconds between observations on each row. This approach allows the camera to produce 5 images of the same object in each filter. However, there are also gaps between the CCD columns, so two passes along a great circle are required to produce a solid image area, with the second pass being slightly offset from the first.

Spectroscopic observations are made with the imaging camera by mounting a fibre plug plate onto the focal plane. These aluminium plates have been individually drilled for each field and hold around 640 optical fibre cables that collect incident light from target objects. Each fibre has a projected fibre diameter of about 3" on the sky. Once collected, the light from the object fibres is directed into two spectrographs. Each of these spectrographs record the spectra on two 2048×2048 CCDs, one covering bluer light (3800Å – 6100Å), the other covering redder light (5900Å–9100Å). Galaxies were selected for spectroscopic observation from extended objects with strong ($> 5\sigma$) r band detections in the imaging survey. Exposures typically consist of three exposures of 15 minutes each, so the spectra can reach the required signal-to-noise. Alongside the object fibres are several fibre bundles which

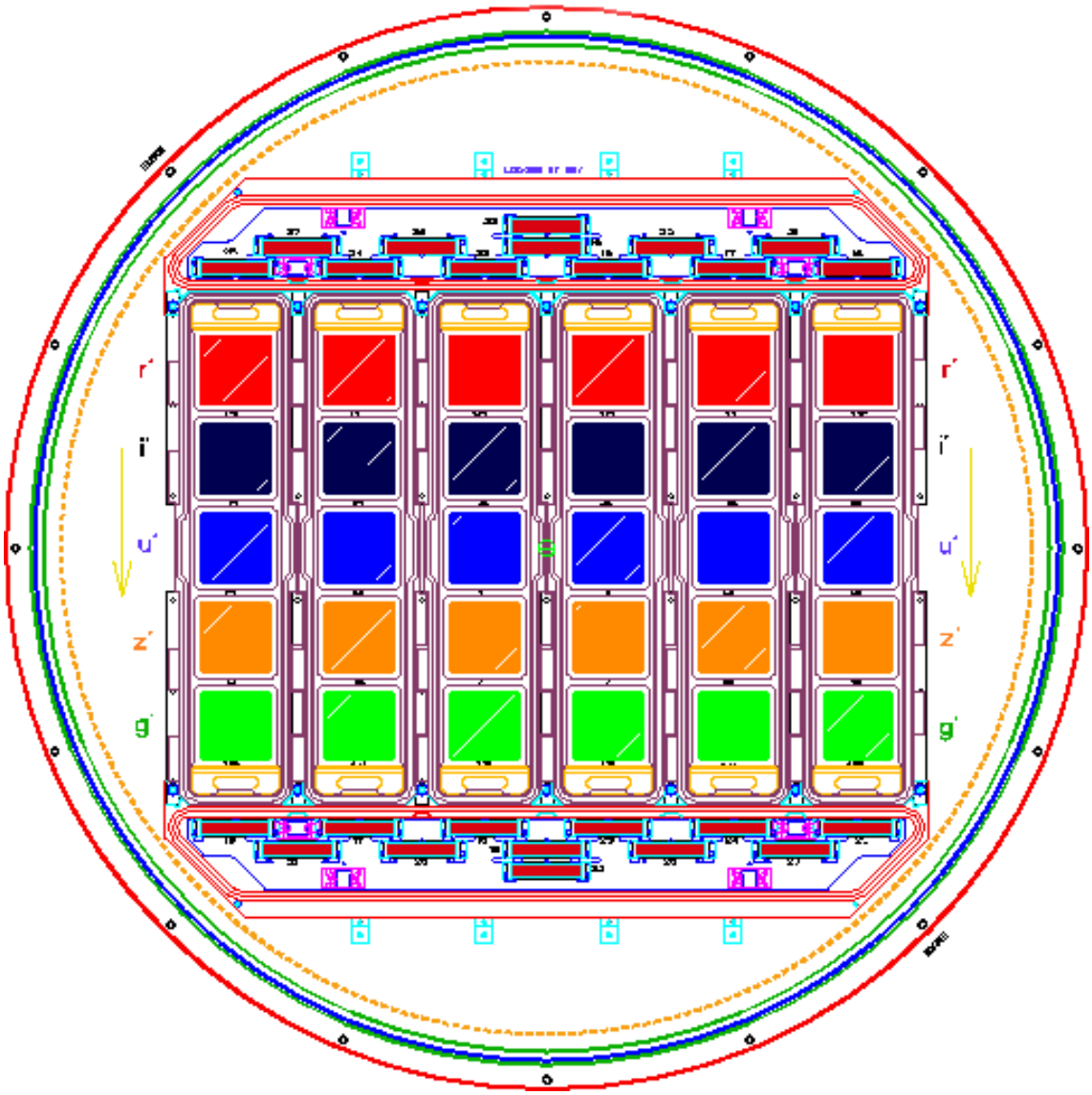

Figure 2.1: Diagram of the CCD arrangement in the SDSS camera. Astrometric CCDs lie above and below the imaging CCD columns. Imaging CCDs are coloured and labelled with the relevant filter. The imaging direction is highlighted with arrows. Credit: *SDSS*

are aimed at guide stars to correctly align the observation. A further large fibre bundle, covering about 30", is used to measure sky brightness. After allowing for slewing, calibration and CCD read times, the total exposure time for a given field is 63 minutes. See table 2.1 for more details about the spectrograph parameters (Gunn et al. 2006).

Imaging Camera					
Filter bands	u'	g	r	i	z'
Central Wavelength (Å)	3551	4686	6166	7480	8932
Magnitude Limits	22	22.2	22.2	21.3	20.5
PSF Width	1.4" median in r				

Spectrograph	
Wavelength Coverage	3800 – 9200Å
Spectral Resolution	2.5 – 3.6Å
Magnitude Limit	$r < 17.77$ for Galaxies
Signal-to-noise	> 4 per pixel for $g = 20.2$

Table 2.1: Summary of SDSS photometric & spectroscopic properties

2.1.2 MPA-JHU Catalogue

Optical photometry and spectroscopy covering 9274 deg^2 of the sky can be found in the Sloan Digital Sky Survey Data Release 8 (SDSS DR8). The MPA-JHU catalogue provides estimates of galaxy properties such as stellar mass, SFR and emission line fluxes, for 1,472,583 objects observed as part of the Legacy survey in SDSS DR8.

The MPA-JHU catalogue is formally deprecated by the SDSS. Alongside MPA-JHU, three alternative catalogues are presented whose analysis is, formally, favoured: Wisconsin, Portsmouth and Granada[1]. However, after assessing the catalogues I found that they were insufficient for these purposes. The Wisconsin catalogue did not calculate any SFR values, nor did they present any spectral quantity from which one could be derived. The Portsmouth catalogue does calculate the appropriate quantities but 64% of their SFRs have a zero value across a wide range of magnitudes. In addition, the non-zero SFRs are gridded, implying a lack of precision at lower values. The Granada catalogue has the most promising quantities, having a greater proportion of non-zero SFRs. However, there is also evidence of gridding at lower values. Furthermore, there appears to be a strict limit applied to the specific SFR. It is unclear why this limit has been applied and it has the effect of structures like

[1]For more information about these catalogues, visit the SDSS galaxy properties page

the galactic main sequence of star formation (e.g. Elbaz et al. 2011; Noeske et al. 2007). MPA-JHU is not subject to any of the above issues. It is a widely used catalogue with robust stellar mass and SFR values. Thus, despite its deprecated status, MPA-JHU is the best catalogue for this work and its quantities are used throughout.

Stellar masses are provided by the MPA-JHU catalogue whose calculation method is based on that described in Kauffmann et al. (2003a). They use template spectra made from a linear combination of single stellar population models generated using the Bruzual and Charlot (2003) code. These models consider 10 possible ages - from 0.005 to 10 Gyr - and 4 possible metallicites - from 0.25 to 2.4 $Z_{\odot}$. They model galaxies as a single metallicity population with the chosen model being the one that yields the minimum χ^2. They then subtract this from the observed spectrum and the remaining emission lines are modelled as Gaussians. For each observed and calculated property they employ a Monte-Carlo fitting technique. This produces a probability distribution from which they extract the most likely value, the distribution median. Given the size of the SDSS fibre aperture, the spectral measurements required by Kauffmann et al. (2003a) would be dominated by light from the galactic centre if based on SDSS spectroscopy. Instead, MPA-JHU use the ugriz photometry from the full extent of the galaxy, and calculate the total stellar mass by fitting to model magnitudes. They assume a Kroupa (2001) initial mass function. MPA-JHU stellar masses were found to be largely consistent with those from the Galex-SDSS-WISE Legacy Catalogue (Salim et al. 2016). The typical difference ranged from $0.03 - 0.13$ dex, consistent with the errors on the stellar mass measurement.Thus throughout this thesis, the median of the mass probability distribution is used as the mass value.

As with the mass, SFR is measured using fits to the SED. Photometry from the full extent of the galaxy is taken from the SDSS and GALEX (UV; Martin et al. (2005)). MPA-JHU SFRs are calculated using the method outlined in Salim et al. (2007). They construct stellar populations from the Bruzual and Charlot (2003) population synthesis models and a Chabrier (2003) IMF. Star formation histories were not single stellar populations, but the combination of an exponentially declining continuous star formation $\propto e^{-\gamma t}$, with $0 \leq \gamma \leq 1$ Gyr^{-1} uniformly distributed across this range, and with random starbursts superimposed. These bursts were constructed such that the occurrence of a single event over the past 2 Gyr is 50% and had a duration uniformly distributed in the 30-300 Myr range. Once constructed, each model is subjected to the Charlot and Fall (2000) dust attenuation model. The model SEDs at the redshift closest to the galaxy in question are, in turn, compared and their χ^2 values evaluated. From this a probability distribution corresponding

to a range of possible SFRs is produced. Salim et al. (2016) found that the MPA-JHU SFRs showed no evidence of bias when compared with two independent SFR measures, including mid-IR SFRs. Thus throughout this thesis, the median of the SFR probability distribution is used as the SFR value.

2.2 XMM-Newton

The X-ray Multi-mirror Mission (XMM-Newton) is an X-ray space observatory launched in December 1999. After launch, it was placed into a 48-hour elliptical orbit, at an inclination of 40°, moving between its perigee of 7,000 and apogee of 114,000 km from Earth. XMM-Newton was the second cornerstone mission of the European Space Agency's Horizon 2000 program which had, initially, been planned to operate for 10 years, but it has now been successfully observing for over 20 years.

XMM-Newton hosts three Wolter type-1 X-ray telescopes which are 250cm long and 90cm in diameter, shown in figure 2.2. Due to their high energy X-rays can be difficult to focus, so the design of the telescope array is different to other wavelength observations. Incident X-ray photons hit mirrors on the inside of the telescope at a shallow angle and are strongly reflected towards the focal plane. To facilitate this process of glancing reflection, each telescope consists of 58 nested mirror shells. There are two sets of mirrors: the outer paraboloid mirrors, followed by the inner hyperboloid shells. The Wolter type-1 design is inset on the upper panel of figure 2.2. Stray light, singly reflected by the hyperboloid mirrors, can also reach the focal plane. However, each telescope includes baffles made of concentric annular apertures, to diminish stray X-ray light.

At the primary focus of each telescope are the instruments that make up the European Photon Imaging Camera (EPIC). It consists of two types of CCD: the Metal Oxide Semi-conductor (MOS; Turner et al. (2001)) and PN arrays (Strüder et al. 2001). These instruments are designed for X-ray imaging and moderate resolution spectroscopy. The data used in this thesis is taken largely from the EPIC cameras, so their construction will discussed further in section 2.2.1.

At the secondary focus of two X-ray telescopes are the Reflection Grating Spectrometers (RGSs; den Herder et al. (2001)). These telescopes contain Reflection Grating Arrays which direct about 40% of the X-ray light away from the EPIC camera and onto the RGS Focal Plane Camera. Once incident on the camera, the RGS will provide high resolution spectra for bright sources in the 0.35 - 2.5 keV energy range.

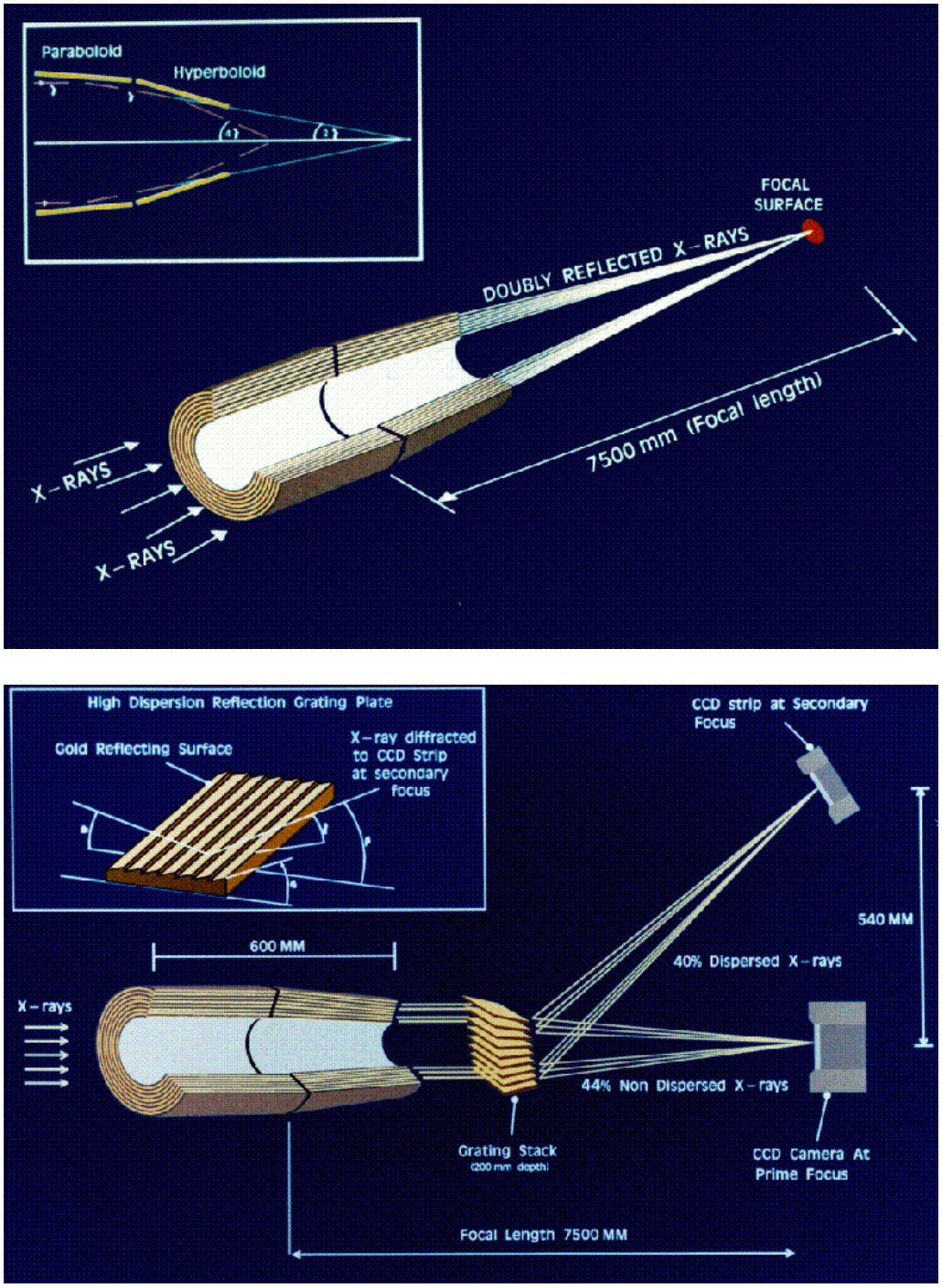

Figure 2.2: Diagrams showing the construction of the XMM-Newton telescopes. Above is the construction of the PN telescope. Inset into this figure is a schematic of the mirror construction for each telescope. Below is the construction of the MOS telescope. Inset into this figure is a diagram of the Reflection Grating Array. Source: *XMM-Newton User's Handbook*

An Optical Monitor (OM; Mason et al. (2001)) is mounted on the mirror support platform, situated alongside the three X-ray telescopes. The 30cm Ritchey-Chrétien telescope provides optical and UV imaging and spectroscopy in the 170 to 650 nm wavelength range. It has a 17 square arcminute FoV centred so as to align with the X-ray telescope's observation. This alignment with the X-ray image's FoV allows for the simultaneous optical and UV data to collected about the target object.

XMM-Newton's construction gives it several advantages over other X-ray observatories. Firstly, it has good angular resolution. The on-axis point-spread function (PSF) is narrow and varies little over a wide energy range. Each telescope has its own PSF: the PN detector has a full width at half maximum (FWHM) of ~ 6" and half energy width (HEW) of ~ 16"; the MOS detectors have a FWHM of ~ 5" and HEW of ~ 15". Secondly, it has a large effective mirror area, with an effective area of ~ 1550 cm^2 at 1.5 keV. Since the telescopes have their FoVs aligned to within 1-2" this combines to produce an total area of ~ 4650 cm^2. And finally, the EPIC cameras have a moderate spectral resolution ($\frac{E}{\Delta E} \approx 20 - 50$). These properties make XMM-Newton useful for sensitive X-ray imaging over large areas on the sky, and thus well suited to this work.

2.2.1 EPIC Cameras

For this thesis work, I used data captured on the EPIC cameras. They are capable of performing sensitive imaging over a 30' FoV in the 0.2-12 keV energy range with moderate spectral and angular resolution. At the primary focus of each X-ray telescope lie the CCD arrays that make up the EPIC camera system.

MOS CCD arrays share their telescopes with the two RGS spectrometers, shown in the lower panel of figure 2.2. For this reason, only about 44% on the incoming light is incident on the MOS arrays. Figure 2.3 shows the circular arrangement of seven identical silicon CCD chips that compose the MOS detectors: one lies at the centre of the array, with the others surrounding it. These 600×600 pixel CCDs are arranged as such to follow the curvature of the focal plane, however this causes the surrounding chips to overlap the central one, creating unusable areas. To account for this, both MOS detectors are rotated at 90° with respect to each other. The MOS detector is front-illuminated meaning its quantum efficiency is limited at higher energies.

At the focal plane of the third telescope is the PN Array, shown in the upper panel of figure 2.2. This array consists a single silicon wafer with tweleve 200×64 inte-

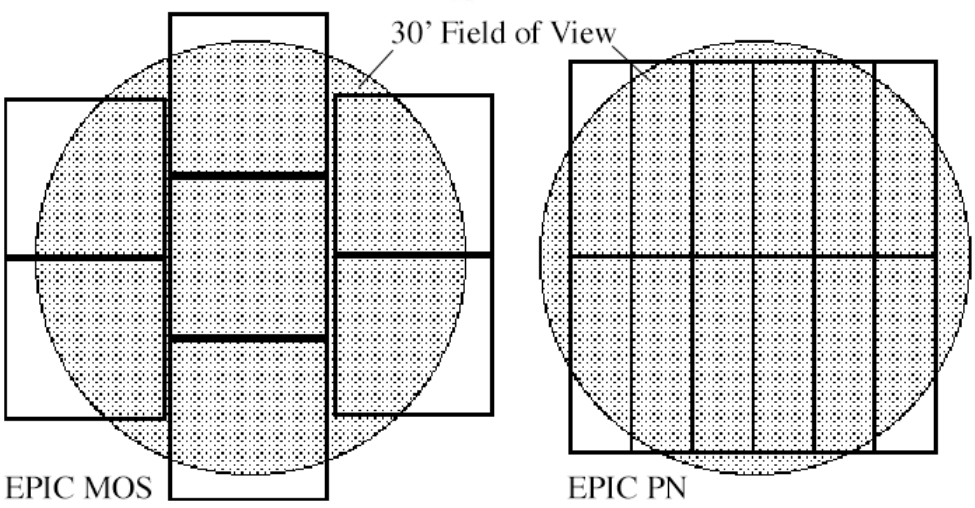

Figure 2.3: CCD arrangements for the MOS (left) and PN (right) arrays on the XMM-Newton 30' field of view. Source: *XMM-Newton Users' Handbook*

grated CCD chips arranged in parallel, shown in figure 2.3. These chips are slightly offset from the optical axis so the observing target object does not fall along a chip boundary. Unlike MOS, the PN CCDs are back-illuminated meaning they have high quantum efficiency across the full XMM energy range. In addition, the readout of the PN chips is much faster than MOS because each pixel column has its own read-out node.

2.2.2 X-ray Upper Limits

XMM-Newton's ability to detect an X-ray source is limited by the sensitivity of the EPIC cameras, the strength of the X-ray background and exposure time. Sensitivity is represented as the flux of the most sensitive possible detection in any given region of an EPIC pointing. The success of the studies outlined in section 1.5 depends on detailed knowledge of the relevant survey's sensitivity limits, thus, access to accurate upper limits is crucial for this thesis. These limits allowed me to assess and account for any incompleteness in my X-ray sample.

A constant lower limit on detection sensitivity is provided by the X-ray background. This is due to high energy particles interacting with the body of the EPIC detectors and soft photons scattered onto the X-ray focal plane (Hasinger et al. 2001). The X-ray background also experiences strong fluctuations, which can commonly increase by factors of > 10 (Watson et al. 2001). Furthermore, detections made at high off-axis angles will have fewer photons reaching the focal plane, a process

known as vignetting. In fact, at large offset angles, the fraction of signal reaching the EPIC camera can be as low as 20% of the incident flux.

To assess the sensitivity of XMM-Newton's detections, I used the Flux Limits for Images in XMM-Newton using DR7 data (Flix) service[2]. This uses data from the Serendipitous Source Catalogue, 3XMM-DR7 (see section 2.2.3), to estimate the upper limit of the X-ray flux sensitivity at a given point in an image. Flix follows an algorithm outlined in Carrera et al. (2007) to calculate this upper limit. It builds on the sensitivity calculations from image processing pipeline (see section 2.2.3). They identified empirical relations between the observed EPIC PN count rates and the pure Poisson count rate calculated from the background map. From this analysis they found a series of coefficients for each XMM band to modify the upper limit calculation. Flix assesses the upper limit at any given XMM pointing by, first, using the observed background and Poisson fluctuations to find the number of counts which exceeds a chosen detection threshold. Then the predicted count is divided by the product of the average exposure time at that location and the relevant coefficient for that band. This process can be repeated for any set of co-ordinates and threshold values for any region of the sky observed by XMM-Newton.

2.2.3 Serendipitous Source Catalogues

Given XMM-Newton's large effective area and good spatial resolution, pointed observations with the XMM-Newton EPIC cameras can reach very faint flux limits. At these fluxes each EPIC field contains substantial numbers of "serendipitous" X-ray sources. For any given XMM pointing, it has been estimated that between 50 - 100 serendipitous detections are made (Watson et al. 2009). As XMM-Newton makes around 600 observations per year, covering a sky area of around 100 deg^2, the number of unique serendipitous sources grows by, on average, 30,000 per year (after allowing for overlaps in sky coverage and areas where EPIC was not used). XMM-Newton thus provides a sensitive, large area sky survey which is well suited to this thesis work.

Source detection in EPIC fields is performed as part of the image processing pipeline. Since July 2013, the XMM-Newton Science Operation Centre has been responsible for the bulk processing of observation data files with this pipeline. The source detection process is multi-stepped, with each step being performed simultaneously across all bands and cameras. First, a detection mask is made for each camera. This defines the suitable area for source detection: only those CCDs where the unvignetted

[2]Found at ledas.ac.uk/flix/flix.html

exposure map values are at least 50% of the maximum.

Next, prominent X-ray sources are found using "local mode box detection". This moves a 20" × 20" box across the image region within the detection mask. A local background to this source is also established in the 8" region around this box. Using a maximum likelihood approach, measurements from both the search region and background for each band were used to calculate the probability and corresponding likelihood that the observed count rate was due to fluctuations in the background. These band-specific likelihoods are then summed and used to calculate the full-band fluctuation probability, P_F. A source is only included in the initial source list if the detection likelihood, calculated from $1 - P_F$, exceeds a given threshold.

After this, a more sensitive detection pass is carried out. New background maps are created for each camera and energy band with areas of the image where sources were previously detected, blanked out. The cut-out radius depends on the source brightness, increasing until the source count per unit area fell below 0.002 counts arcsec^{-2}. A 12×12 node spline surface was fitted to the source-free image so as to create the new smooth background map. The box detection method described above was then repeated on the new background map.

Maximum likelihood PSF fits are then made to the count distributions in each source list. The spatial distribution of an input source is compared to the PSF model and a likelihood value that both distributions are the same is calculated by varying parameters like position, extent and counts. Values for the total source count rate, hardness ratios and detection and source extent likelihoods are derived. If the detection likelihood after this fitting process exceeded a given threshold then it was included in the final source list.

Over the course of XMM-Newton's operation, these serendipitous detections have formed the basis of the Serendipitous Source Catalogues. 1XMM, the first Serendipitous Source Catalogue, was released by the XMM-Newton Science Survey Centre (SSC) in September 2003. To date the SSC have been responsible for 10 further catalogues released.

The X-ray data used in this study comes from the 3XMM DR7 catalogue released in June 2017 (Rosen et al. 2016) and was the most recent release at the time this thesis work began. Whilst slightly bigger catalogues were subsequently released, 3XMM-DR7 remained the most up-to-date version available with comprehensive upper limits from Flix. It is based on 9,710 pointed observations with the XMM-

Newton EPIC cameras. DR7 contains ~ 400,000 unique X-ray sources based on 727,790 individual detections. Typical position errors for DR7 are ≈ 1.5" (1σ) and extend down to a flux limit of $\approx 10^{-15}$ erg cm^{-2} s^{-1}. For this work I use the unique source list rather than the individual detections. The results are thus averaged over several individual observations for a significant number of sources. From this sample I summed fluxes in the 2 - 4.5 keV and 4.5 - 12 keV bands and converted them to luminosities in the 2 - 12 keV energy range using the MPA-JHU redshifts. Since these objects are at such a low redshift, no rest-frame correction was applied.

Chapter 3

The Incidence of AGN in Dwarf Galaxies

3.1 Introduction

Identifying AGN within dwarf galaxies ($M_* \leq 3 \times 10^9$ $M_\odot$) is of increasing interest having been catalysed by a combined X-ray and radio detection in Henize 2-10 almost 10 years ago (Reines et al. 2011). Cataloguing these objects can provide insights into numerous aspects of galaxy evolution. Since some dwarf galaxies have much lower metallicity relative to their high mass counterparts and have not tidally interacted with their neighbours, they can be considered an analogue for galaxies in the high redshift Universe (Bellovary et al. 2011). Thus this quantification of AGN in dwarf galaxies can be used to provide insights into the possible mechanisms that seed SMBHs in the very early Universe.

There are two possible mechanisms by which black holes may have formed in the early Universe: formation from Population III stars (Madau and Rees 2001) or by direct collapse (Begelman et al. 2006). Black hole formation from stellar collapse is possible at all redshifts but the early Universe contained low metallicity gas in abundance which facilitated the growth of more massive Population III stars. After collapsing into stellar mass black hole seeds (Madau and Rees 2001) they could have undergone a series of intermittent super-Eddington accretion episodes (Madau et al. 2014) or merged with other seeds to form more massive black holes (Miller and Davies 2012). Black holes could also form from the direct collapse of a proto-galactic gas cloud if it is subject to inefficient gas cooling (Begelman et al. 2006). Crucially, however the low angular momentum gas required for this method of formation only existed in sufficient quantities in the early Universe. As disk-like structures began to form, the gas within would gain angular momentum and become less likely to

collapse and form a black hole. This restriction means that if direct collapse were the dominant seeding mechanism then the fraction of dwarf galaxies hosting black holes (and thus hosting AGN) is expected to be much lower than if the Population III stellar seeding mechanism was dominant (Greene 2012). For a more in-depth review of black hole formation mechanisms see the review by Latif and Ferrara (2016).

Before the applicability of these models can be discussed, however, the black holes need to be found. Within massive galaxies, black holes can be identified using dynamical methods, however current technological limits mean these methods are difficult to apply at lower galactic masses. Hence the field focuses on AGN. There is increasing evidence across a range of wavelengths to show that at least a small number of AGN do exist within the dwarf galaxy mass range (see review by Reines and Comastri 2016). Some of the first large-scale studies into this area focused on optical emission (Reines et al. 2013) finding black holes through a combination of identifying broad line emission and measuring line strength. However, the effectiveness of this method can be limited by obscuration from dust or star formation signatures.

AGN emission in the X-ray band generally dominates over other sources and is less easily obscured. A number of studies have had success in using this emission to identify the presence of AGN in dwarf galaxies (e.g. Lemons et al. 2015; Mezcua et al. 2018, 2016; Paggi et al. 2016; Pardo et al. 2016; Reines et al. 2011). However, the techniques employed are not free from problems. Lower mass galaxies, like the ones being studied, tend to host lower mass central black holes hence less luminous emission is expected. This difficulty can also be compounded if the AGN being studied are very weakly accreting. The biggest challenge for studies like these is to rule out emission from other potential X-ray sources. In this chapter, I quantify the X-ray selected AGN population in local dwarf galaxies.

3.2 Identifying AGN in Dwarf Galaxies

As discussed in chapter 2, the parent sample is constructed using data from the SDSS value-added catalogue, MPA-JHU, and 3XMM DR7. In this section I outline the matching and verification process applied to these catalogues to construct the dwarf galaxy ($M_* \leq 3\times10^9$ $M_\odot$) and comparison high mass galaxy ($M_* \geq 10^{10}$ $M_\odot$) samples. This is followed by a discussion of how AGN were selected from the resulting sample of X-ray emitting galaxies.

3.2.1 Position Matching

To get a sense of how many AGN-hosting galaxy candidates are present, I first searched for any MPA-JHU galaxies within a 10" search radius of every 3XMM-DR7 X-ray object; this yielded 3,440 matches. From within this sample of X-ray emitting galaxies, possible AGN-hosting dwarf galaxies could be found. To achieve this I first established a set of more robust matching criteria to improve the confidence of association between the optical and X-ray objects. For each dwarf galaxy/X-ray pair I calculated the position-error-normalised separation, x, as follows,

$$x = \frac{d_{\mathrm{O,X}}}{\Delta_{\mathrm{X}}} \tag{3.1}$$

where $d_{\mathrm{O,X}}$ is the separation between the X-ray and optical signals, and Δ_{X} is the error in the X-ray position. A dwarf galaxy was considered robustly matched only if $x < 3.5$. This gave a sample completeness of 99.8%. I also limited the extent of the X-ray source to less than 10" to ensure point-like emission consistent with an AGN (Rosen et al. 2016). This approach yielded 101 possible AGN hosts. Applying the same matching criteria to the high mass sample yields 2,237 matches.

3.2.2 Verifying the Matches

I performed a number of checks on the matches produced by the position-error-normalised separation criterion in equation (3.1). By imposing the dwarf galaxy mass limit, I could have removed higher mass objects with a smaller separation to the X-ray source, resulting in a poorer dwarf galaxy match being included. To check if any better matches to excluded objects exist, I uploaded the 101 X-ray co-ordinates to the SDSS SciServer and compared them to the full SDSS DR8 to find their nearest neighbour. 11 X-ray sources were found to have smaller separations to a higher mass galaxy but had been removed by imposing the dwarf galaxy mass limit during the matching process. These sources, and their dwarf galaxy matches, were removed from the X-ray emitting galaxy sample.

I then performed a visual assessment of the dwarf galaxy sample. The optical images from the SDSS Finding Charts identified 3 sources that were extremely off-nuclear. If the photometry was constructed correctly for these objects then the refinement process outlined above should have removed them, however this was not the case. In certain instances a galaxy that should not have passed the mass criterion was being broken into smaller sections and only had the MPA-JHU mass calculations

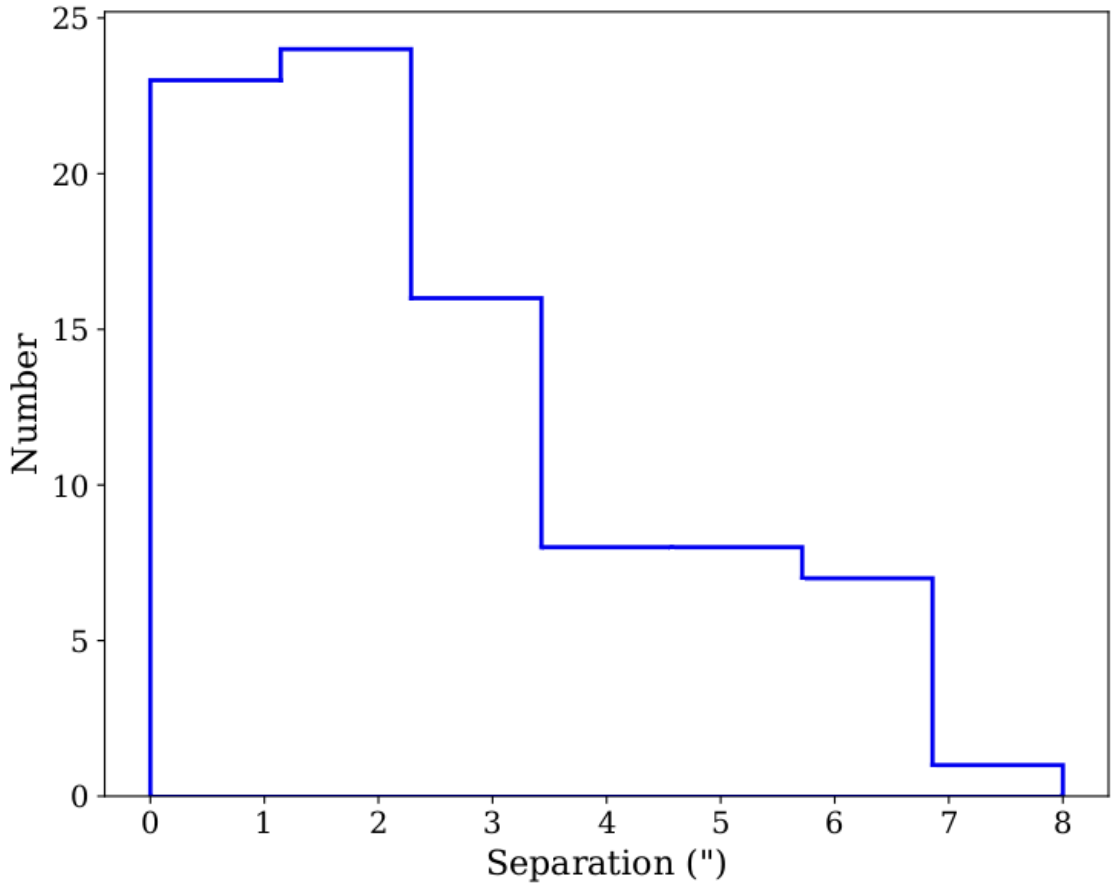

Figure 3.1: Separation in arcseconds between the X-ray and optical signals for the 86 dwarf galaxies that meet the matching criteria outlined in section 3.2. 84% of the signals in this sample are matched within 5" and the median matching radius is 2.2".

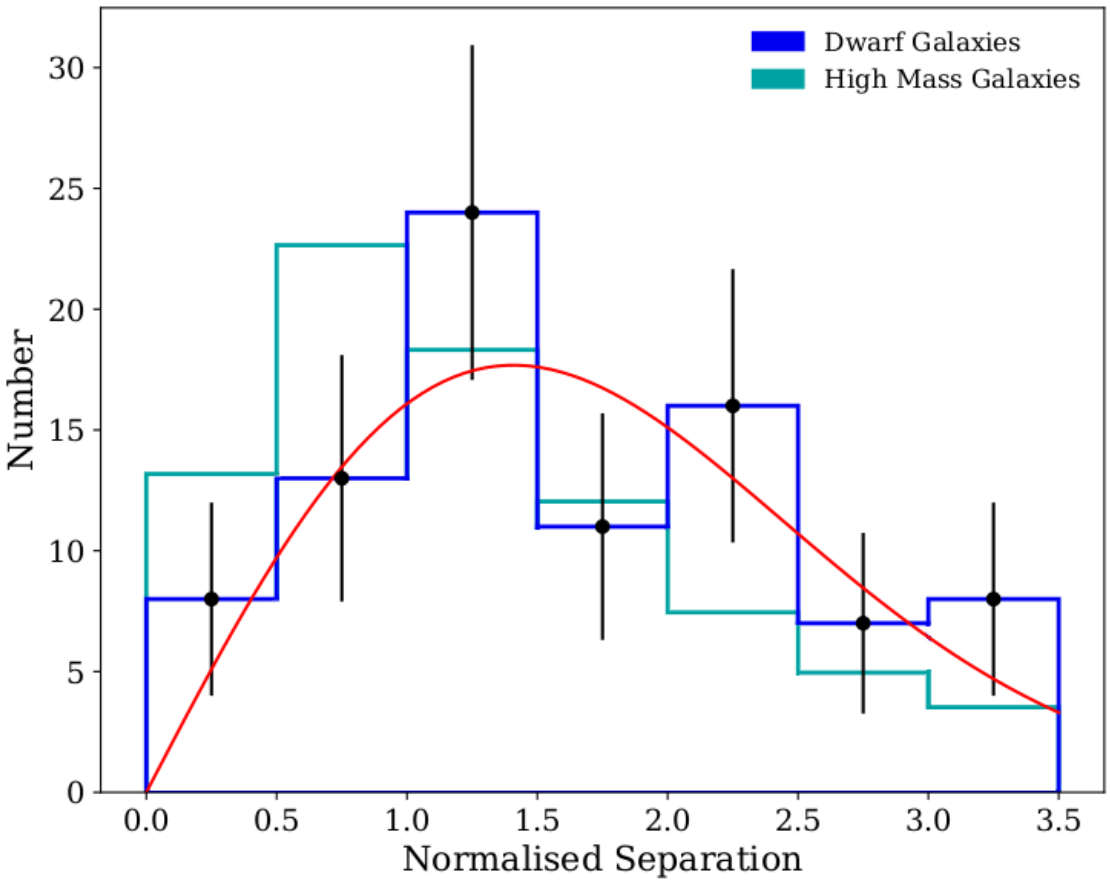

Figure 3.2: Position-error-normalised separation distribution (dark blue) for the 86 dwarf galaxies that met the matching criteria described in section 3.2. Poisson errors are assumed in the histogram bins. The Rayleigh distribution fit (red) has a $\sigma = 1.4$ implying an underestimation in the XMM positional errors, a result that has been noted previously (Watson et al. 2009). A sample of high mass galaxies (light blue) has undergone the same matching process and had its normalised separation scaled using the normalisation found from the Rayleigh fit.

applied to this small section. This was then matched to a nearby X-ray source, fooling my identification criteria into thinking it was a valid target. These 3 sources were removed from the sample leaving 87 dwarf galaxies with nuclear X-ray activity. This photometric fragmentation process has been observed in other, similar studies such as Sartori et al. (2015).

A final assessment of the data found an object with extremely high redshift error. This is indicative of a poor fitting process, leading me to doubt the accuracy of this object's data. Removing this object left me with a final sample of 86 dwarf galaxies. Their distribution of separations between the X-ray and optical signals is shown in figure 3.1: 84% of these dwarf galaxies match an X-ray object within 5", and the median matching radius is 2.2". A sample of dwarf galaxy images, with their X-ray sources and position error overlaid, can be found in appendix A.

Despite all these checks, there remains the possibility that some matches are due to chance alignment. To understand whether this might affect the validity of my sample, I estimated the expected number of false matches. Assuming a uniform distribution of the 65,461 dwarf galaxies in the SDSS area (9274 deg^2), the dwarf galaxy sky density would be 7.05 deg^{-2}. Also within the SDSS sky area, I found 122,834 point-like X-ray sources. Assuming I then perform a search, of radius $3.5 \times \Delta_X$, around each of these sources it would cover a sky area of 0.91 deg^2 within which I would expect to falsely match to 6 dwarf galaxies. However, given a lot of these X-ray sources will have well-defined counterparts, this value represents a conservative upper limit on the false match number.

Once all of the matches had been checked, I performed one final test to assess the accuracy of my position-error-normalised separation criterion. Figure 3.2 shows distributions of this quantity for both the dwarf galaxy and high mass samples. I performed a 1-D KS-test to compare this distribution to the expected distribution of XMM errors, described by the Rayleigh distribution. The KS-test returned a p-value of 10^{-4}, indicating the observed distribution is not consistent with the expected Rayleigh distribution, given the fiducial positional uncertainties provide in the 3XMM catalogue. To investigate whether the observed distribution was better described by a skewed Rayleigh function, I fit it to this equation,

$$f(x) = N\frac{x}{\sigma^2}e^{\frac{-x^2}{(2\sigma^2)}} \tag{3.2}$$

where x is the position-error-normalised separation. The normalisation constant, N and σ value were both free and I assumed Poisson errors for the number of objects

per bin. Figure 3.2 shows the results of this fitting: it was found that the skewed Rayleigh distribution (red) had a $\sigma = 1.4$, with a reduced χ^2 of 1.90, implying that the X-ray errors were underestimated. It is also clear that both the dwarf galaxy and high mass distributions are consistent with this error underestimation. The underestimation of errors is a known issue in 3XMM, as noted in Watson et al. (2009). Given these positional uncertainties, I was confident the position-error-normalised separation criterion functioned correctly, thus I expect the dwarf galaxy X-ray sources to be nuclear.

3.2.3 X-ray Emitting Galaxies' Properties

Figure 3.3 summarises some of the properties for both the dwarf galaxy and high mass samples. The top panel shows the distribution of both samples of objects in mass and redshift. The dwarf galaxies, in dark blue, and high mass sample, in light blue, are plotted over the full sample of 3XMM X-ray detections within the SDSS area, in grey. Dwarf galaxies are found out to $z = 0.25$ and span a mass range of 3.08×10^6 $M_\odot$ to 2.92×10^9 $M_\odot$. Higher mass objects span a mass range 1.01×10^{10} $M_\odot$ to 2.84×10^{12} $M_\odot$ and can be detected at redshifts up to $z = 0.33$.

The bottom panel of figure 3.3 shows the distribution of X-ray luminosities for both samples. I summed fluxes in the 2 - 4.5 keV and 4.5 - 12 keV bands and converted them to luminosities in the 2 - 12 keV energy range using the MPA-JHU redshifts. Since these objects are at such a low redshift, no rest-frame correction was applied. The dwarf galaxies, in dark blue, have observed X-ray luminosities of between 2.43×10^{36} erg s^{-1} and 5.35×10^{42} erg s^{-1}; the modal group is between 5×10^{39} erg s^{-1} and 1×10^{40} erg s^{-1}. My dwarf galaxies are significantly less luminous than their high mass counterparts, shown in the arbitrarily-scaled light blue distribution, as expected.

3.2.4 Identifying AGN

Given most of the dwarf galaxy X-ray detections do not meet the suggested 3×10^{42} erg s^{-1} threshold, other aspects of the emission need to be assessed in order to identify any AGN activity. One possible source of X-ray contamination is the combined emission of unresolved X-ray binary (XRBs) populations within the galaxy. To estimate the contribution these objects make to a galaxy's X-ray emission, I used the relationship provided in Lehmer et al. (2016). It takes into account a galaxy's stellar mass (M_*), SFR and redshift (z) and takes the following form,

$$L_{\mathrm{XRB}} = \alpha_0(1+z)^\gamma M_* + \beta_0(1+z)^\delta \mathrm{SFR} \tag{3.3}$$

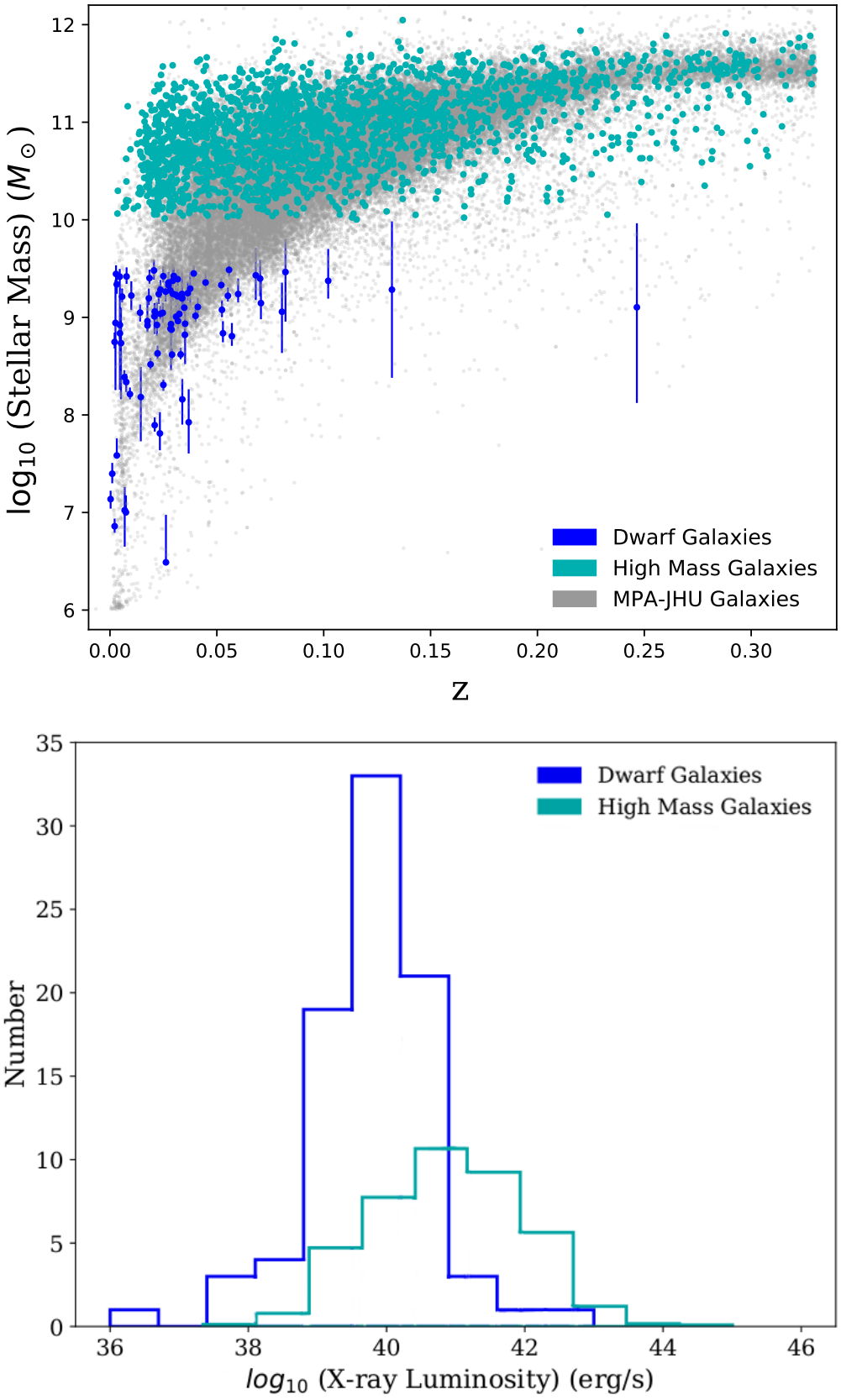

Figure 3.3: A summary of the observed properties of the dwarf galaxy (dark blue) and high mass (light blue) samples. In the top panel is stellar mass against redshift (contrasted with the underlying MPA-JHU sample in grey). In the bottom panel is the X-ray luminosity distribution (the high mass distribution has been arbitrarily scaled to allow for comparison). Section 3.2.3 breaks down the X-ray emission properties of these samples.

where $\log_{10}(\alpha_0) = 29.37 \pm 0.15$, $\gamma = 2.03 \pm 0.60$, $\log_{10}(\beta_0) = 39.28 \pm 0.03$ and $\delta = 1.31 \pm 0.13$ for 2 - 10 keV.

The low-mass X-ray binary (LMXB) contribution is correlated to the stellar mass, the high mass X-ray binary (HMXB) contribution to the SFR and the redshift dependence accounts for changes in metallicity and evolution of the XRB population. About 15% of the galaxy sample used to fit the Lehmer et al. (2016) relationship falls within the dwarf galaxy range defined in this chapter. Thus equation (3.3) was appropriate to calculate the expected emission from every dwarf galaxy's binary population, L_{XRB}. A total of 76 galaxies were found to have an observed luminosity that is greater than L_{XRB}. Thus, there is a significant sample of X-ray emitting dwarf galaxies whose emission cannot be accounted for solely by their XRB population.

Hot gas in the interstellar medium can also produce X-rays, which could also potentially account for some of the emission. Its contribution can be estimated using the Mineo et al. (2012b) relation,

$$L_{\mathrm{Gas}} = (8.3 \pm 0.1) \times 10^{38} \ \mathrm{SFR} \ (\mathrm{M}_{\odot} \ \mathrm{yr}^{-1}) \tag{3.4}$$

About a third of the sample used to derive this relationship fell within my dwarf galaxy definition. Thus equation (3.4) was appropriate to calculate the expected emission from the hot gas, L_{Gas}. I then added this to each galaxy's L_{XRB} to determine the expected non-AGN contribution to the X-ray emission. Despite its relatively low magnitude, L_{Gas} was still useful to calculate as all significant alternative X-ray sources needed to be considered. However, as expected all 76 objects which already have observed emission exceeding L_{XRB} also exceed the sum of L_{XRB} and L_{Gas}.

Before I accepted this sample of 76 objects as AGN hosts, a level of significance needed to be applied to allow for potential uncertainties in the observed values of the relationships used. For this reason, dwarf galaxies which met or exceeded the following X-ray excess criterion,

$$\frac{L_{\mathrm{X,Obs}}}{L_{\mathrm{XRB}} + L_{\mathrm{Gas}}} \geq 3 \tag{3.5}$$

where $L_{\mathrm{X,Obs}}$ is the observed X-ray luminosity, were considered to have sufficiently excessive X-ray emission to potentially host an AGN. Figure 3.4 shows the results of these calculations. A total of 61 objects highlighted in dark blue and red both

have sufficiently excess emission to pass this criterion and host an AGN. However, they have their SFR measured in different ways. The point colour indicates the method by which that galaxy's SFR was calculated: the 56 galaxies indicated by the dark blue points had SFRs calculated from their SED fits by MPA-JHU; the 5 red points were flagged as having poor SFR fits so instead I used the Kennicutt and Evans (2012) formalism and the Hα emission line flux to give an SFR. Given the Hα emission line can be contaminated by AGN light, this SFR is a conservative upper limit. X-ray detections that do not meet the criterion given by my excess X-ray criterion are plotted in grey. As a comparison, I also performed the same analysis to the high mass sample. It identified 1,316 objects with emission likely originating from an AGN.

To confirm if these dwarf galaxies with AGN had been previously identified, I entered their RA and Dec co-ordinates into both SIMBAD and NED. Of the 61 AGN I have identified, 40 have not been found previously. In appendix B I present a breakdown of the observed and calculated properties for this sample of 61 AGN alongside the references of any previous identifications.

3.3 Spectral Properties

In this section I analyse the X-ray and optical spectra to see how AGN in dwarf galaxies compare to other samples.

3.3.1 Hardness Ratio

Hardness Ratios (HRs) are defined between -1 and 1, with more positive values indicating harder X-ray emission, likely from an AGN, and more negative values indicating softer emission, likely from stellar processes. It is calculated using the counts from two different energy bands, A and B,

$$\mathrm{HR} = \frac{\mathrm{Band\ A} - \mathrm{Band\ B}}{\mathrm{Band\ A} + \mathrm{Band\ B}} \tag{3.6}$$

I compare two HR bands: the soft band is the mean HR of the 0.5 - 1.0 keV and 1.0 to 2.0 keV bands; the hard band is the mean HR of the 1.0 to 2.0 keV and 2.0 - 4.5 keV bands. Analysing this quantity can help indicate whether the X-ray emission from our dwarf galaxy sample is coming from an AGN, rather than stellar processes.

Figure 3.5 shows the error-weighted distribution of hardness ratios for both my X-ray detected dwarf galaxy sample (upper panel) and high mass comparison sample

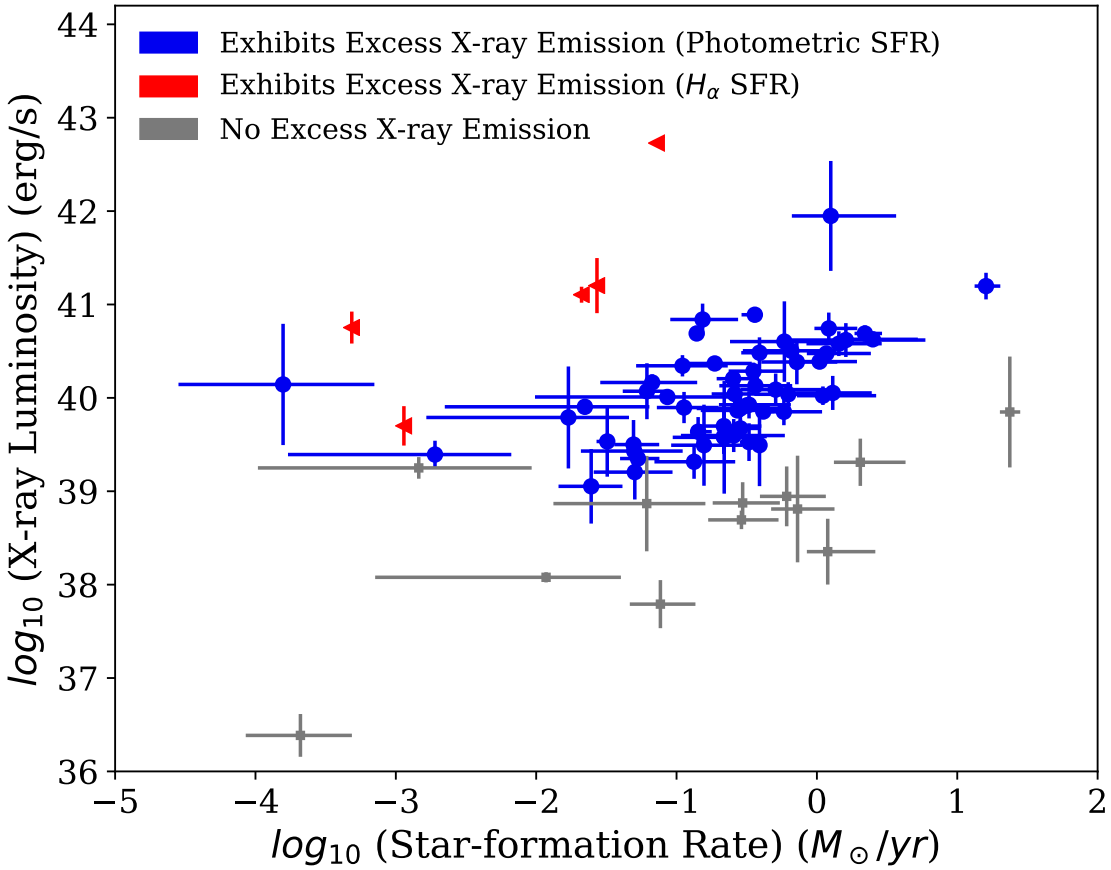

Figure 3.4: X-ray luminosity as a function of SFR for the 86 X-ray emitting dwarf galaxies. A total of 56 objects (blue) have observed X-ray emission which exceeds the excess X-ray criterion: possessing X-ray emission three times greater than the sum of $L_{\rm XRB}$, predicted X-ray emission from X-ray binary stars, and $L_{\rm Gas}$, predicted X-ray emission from hot gas. 5 objects (red) are those which meet or exceed the excess X-ray criterion but have their SFR calculated using the Kennicutt and Evans (2012) formalism and $H\alpha$ emission line. The grey points are those that did not meet the X-ray excess criterion.

(lower panel) before (dotted line) and after (solid line) applying the X-ray excess criterion. By overlaying the distributions in this way it is clear the criterion is working as intended: the emission in both distributions becomes, on average, harder. The contour lines can be seen to shift towards more positive values as objects with softer emission are removed. This effect is particularly pronounced in the high mass distribution, where a distinct group of harder emission objects had been isolated from the full high mass sample. However, less strict attention was given to matching this sample so it does appear more diffuse than that of the dwarf galaxies.

Higher mass galaxies generally have a larger number of confirmed AGN so this distribution will reflect some of their spectral properties. To determine whether the thresholded dwarf galaxy distribution shared any similarities with the high mass distribution and thus with AGN spectra I performed two KS tests. Firstly, the thresholded dwarf galaxy distribution was compared with the full high mass distribution but they were found to be inconsistent, with a 2D 2 sample KS test producing a p-value of 10^{-3}. When both the thresholded dwarf galaxy and thresholded high mass galaxies were compared, however, a 2D 2 sample KS test showed they are consistent at a 3σ confidence. Thus the emission from my thresholded dwarf galaxy sample shows characteristics of coming from AGN.

These X-ray spectra alone are not definitive in determining the source of emission. XRBs can also produce hard spectra, similar to an AGN. Plotting these spectral distributions next to each other and performing the KS tests, however, highlights similarities between the dwarf galaxy and high mass distributions which suggest the dwarf galaxy emission is dominated by AGN.

3.3.2 BPT Classification

AGN can also produce signatures detectable in the visible part of the spectrum. The central accreting black hole ionises the surrounding gas causing various emission lines to come to prominence in the galaxy's spectrum. Other processes such as star formation can also ionise a galaxy's gas and dust but the associated radiation is much softer than from an AGN. The BPT diagram plots a pair of optical emission line ratios - $\frac{\text{[OIII] } \lambda 5007}{\text{H}_\beta}$ and $\frac{\text{[NII] } \lambda 6583}{\text{H}_\alpha}$ - against each other to try and distinguish the source of this ionising radiation (Baldwin et al. 1981). Given the BPT diagnostic's popularity in the field, I investigate whether the 61 AGN hosts identified using X-ray selection techniques would also be found by the BPT diagram. Of the 61 AGN hosts in my sample, 53 had significant detections ($\frac{\text{Line Flux}}{\text{Line Flux Error}} > 3$) in each of the

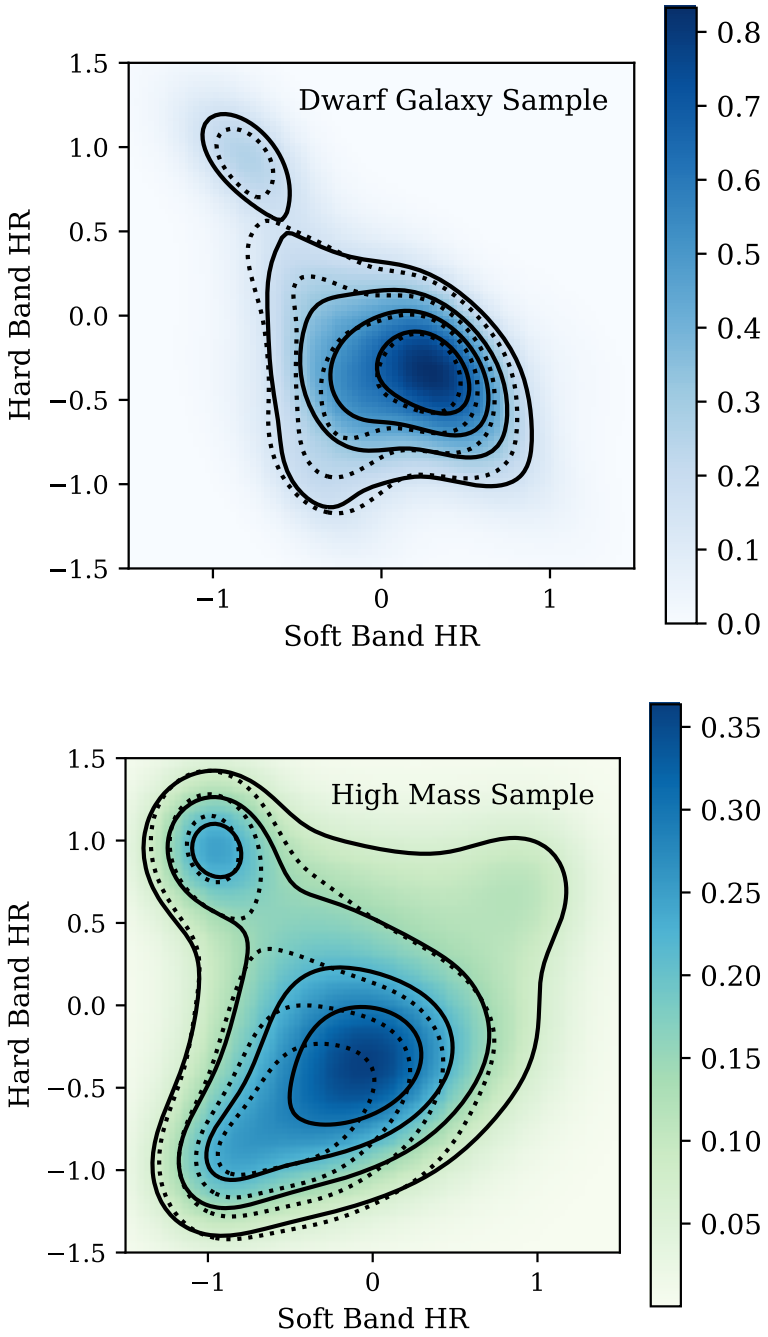

Figure 3.5: Error-weighted hardness ratio (HR) distributions for dwarf galaxy and high mass samples before (dotted line) and after (solid line) the application of the X-ray excess criterion. The soft band is the mean hardness ratio of the 0.5 - 1.0 keV and 1.0 to 2.0 keV bands; the hard band is the mean hardness ratio of the 1.0 to 2.0 keV band and 2.0 - 4.5 keV bands. Applying the X-ray excess criterion shows a positive shift in HRs, indicative of possible AGN activity. The plots have been smoothed using a Gaussian kernel, whose width corresponds to the sample size and dimensions of the data: ~ 0.5 for the dwarf galaxies and ~ 0.3 for the high mass galaxies. In addition, the points are weighted by their errors. The contours encompass 80%, 60%, 40% and 20% of the samples in each plot.

required emission lines so these were used in this analysis.

The results of the BPT analysis are shown in figure 3.6 with AGN hosts plotted as large, coloured points in the foreground. Black lines separate the AGN hosts into different classifications: objects with ionisation signatures predominately from AGN lie in the top-right, those dominated by star formation in the bottom-left, and those that have a composite spectra are in the central region (Kauffmann et al. 2003a; Kewley et al. 2001). Underneath these points, the BPT classification for the 62,703 galaxies in MPA-JHU with $M_* \leq 3 \times 10^9$ $M_\odot$ and $z \leq 0.25$ are plotted in light grey. They are dominated by star-forming objects. The dark grey points show local galaxies in MPA-JHU with an X-ray counterpart within 10". These points are more concentrated in the upper-right region as they likely contain significant numbers of high mass AGN-hosting galaxies.

To investigate whether star formation dominates optical emission in my AGN hosts I attempted to isolate emission from AGN processes at a single wavelength and compare it to the total observed optical emission. I chose to investigate it at 3550Å, ~ the central wavelength of the SDSS u-band as this UV/Optical region will contain light from active star-formation.

First, I consider the AGN emission prediction. To ensure little contamination from stellar processes I used the X-ray emission to calculate the AGN's contribution to the observed optical light. Lusso and Risaliti (2016) published the α_{OX} relation which has the form,

$$\log_{10}\left(L_{\nu,\ 2500\text{Å}}\right) = \frac{1}{0.6}\left(\log_{10}\left(L_{\nu,\ 2\ \mathrm{keV}}\right) - 7\right) \tag{3.7}$$

It relates the luminosity density at 2 keV, $L_{\nu,\ 2\ \mathrm{keV}}$ to that at 2500Å, $L_{\nu,\ 2500\text{Å}}$, from a sample of SDSS quasars. Assuming my dwarf galaxies follow the same relation it can provide a useful starting point in my efforts to estimate the effects of the AGN on optical observations.
To find the luminosity density at 2 keV, I calculated the geometric means of 3XMM bands 2 & 3 and bands 4 & 5, giving the luminosity densities at 1 keV and 5 keV respectively. Using linear interpolation between these values I calculated $L_{\nu,\ 2\ \mathrm{keV}}$ and used it in equation (3.7) to find $L_{\nu,\ 2500\text{Å}}$. Translating $L_{\nu,\ 2500\text{Å}}$ to the required emission at 3550Å required the composite UV-optical quasar spectrum from Vanden Berk et al. (2001). They model this region as a power law spectrum, $f_\nu = c\nu^{-\alpha}$, where $\alpha = 0.44$. This required $L_{\nu,\ 2500\text{Å}}$ to be converted to flux density, $f_{\nu,\ 2500\text{Å}}$ which was used, first, to scale the power for each AGN host. From this an average spectrum

was constructed and the flux density at 3550Å extracted.
The total observed optical emission was more straightforward to find, I simply converted the SDSS u-band magnitude to a flux density at 3550Å. I then divided the predicted AGN emission at 3550Å by the observed optical emission at 3550$\dot{A}$, $L_{\nu,3550\dot{A}}(\mathrm{AGN}_{\mathrm{predicted}}/(\mathrm{AGN}+\mathrm{Galaxy})_{\mathrm{observed}})$, and split each object into groups based on this value. These dictate the colours used for the AGN host points in figure 3.6.

A total of 50 AGN hosts have predicted optical AGN contributions less than the total observed optical emission - 44 have AGN contributions < 25% of the total - the majority of which lie in the star-forming region. This helps confirm the idea that the AGN emission in this part of the spectrum is being hidden by star formation processes, causing the BPT diagnostic to mis-classify them. Of the 2 remaining AGN hosts with a predicted optical AGN excess, only one lies in the AGN region. This host has the largest predicted optical AGN excess in the sample, over 10 times that of the galaxy emission. The other lies within the star-forming region because it has a much smaller excess, only 1.08 times that of the galaxy. This is likely insufficient to produce emission lines of the appropriate proportions to move it into the AGN region. This further suggests that the BPT diagnostic is biased against identifying the low-luminosity AGN expected in dwarf galaxies.

Some studies have shown that the Lusso and Risaliti (2016) relation used to predict the relative optical AGN contribution has an increased dispersion in the regime of dwarf galaxies. Furthermore, this relation assumes that the galaxies are un-obscured. However, due to the MPA-JHU selection criteria, the galaxies in my sample are all narrow-line objects. This produces an overestimate of the AGN contribution making $L_{\nu,3550}(\mathrm{AGN}_{\mathrm{predicted}}/(\mathrm{AGN}+\mathrm{Galaxy})_{\mathrm{observed}})$ appear larger (Baldassare et al. 2017; Plotkin et al. 2016). This can be seen in the two cases where the predicted AGN contribution is larger than the total optical emission. Despite this, I have still identified a significant number of X-ray selected AGN mis-identified as star-forming galaxies. This finding is consistent with a increasing body of work suggesting that optical spectroscopic measurements are insufficient to identify AGN in these environments (Agostino and Salim 2019; Cann et al. 2019).

3.4 Specific Black Hole Accretion Rate

To investigate the activity of the central black holes powering my AGN I calculate their growth rates in terms of the specific black hole accretion rate (sBHAR), λ_{sBHAR}. This compares the bolometric AGN luminosity of the galaxy with an estimate of the black hole's Eddington luminosity to give an indication of how efficiently the black

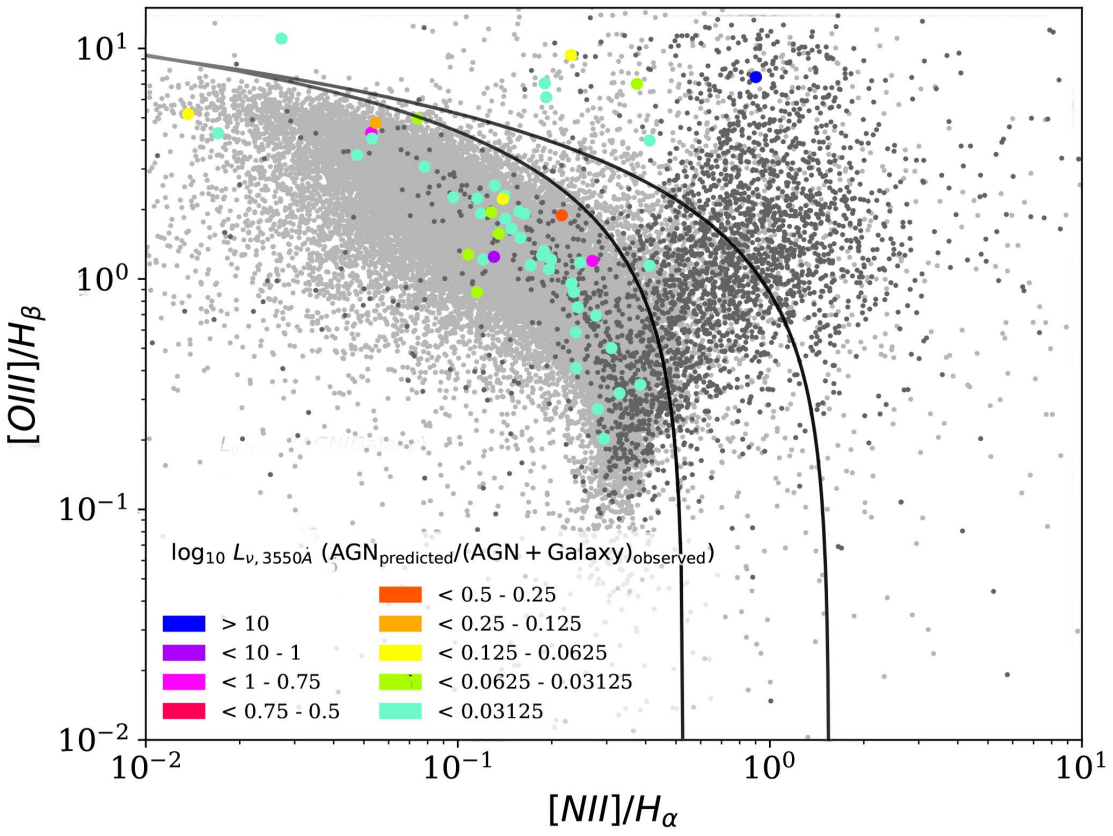

Figure 3.6: BPT diagram for various subsets of the MPA-JHU. The large, coloured points highlight the 53 dwarf galaxies identified as likely to contain an AGN with significant emission line detections. Their colour indicates the size of predicted optical AGN luminosity compared to total optical emission, both at 3550Å - see section 3.3.2 for details on how this was calculated. The smaller, background points are defined as follows: the light grey points show all the low mass objects ($M_* \leq 3 \times 10^9$ $M_\odot$) found within the MPA-JHU highlighting a large distribution of points along the star-forming track; the dark grey points are the objects in the MPA-JHU that have an X-ray counterpart, largely dominated by the high mass AGN-like objects.

hole is accreting. It is found using,

$$\lambda_{\mathrm{sBHAR}} = \frac{25 L_{2-10\mathrm{keV}}}{1.26 \times 10^{38} \times 0.002 M_{*}} \approx \frac{L_{\mathrm{bol}}}{L_{\mathrm{Edd}}} \tag{3.8}$$

and is taken from Aird et al. (2012). I assume that the black hole and stellar masses scale in the same way as their higher mass counterparts to get a sense of the black hole growth relative to the total galaxy mass. In order to clarify the validity of this assumption, I would need to compare this scaling relation to black hole masses but these are very difficult to accurately ascertain. Instead I looked at the morphologies of my AGN hosts to see if they are bulge dominated. Of the 36 AGN hosts found in the Galaxy Zoo DR1 (Lintott et al. 2011), 33 have uncertain morphology and the remaining 3 are likely spirals. Despite this, I chose to cautiously continue using the relationship and present the results in figure 3.7. It plots the mass of a particular host against its observed X-ray luminosity, the colour gives an indication of the sBHAR. I have also plotted lines of constant sBHAR to give an idea of the typical mass and luminosity expected from these galaxies.
The most common object in my sample is a host of mass $\approx 10^{9-9.5}$ $\mathrm{M}_{\odot}$. None of the SMBHs are accreting very efficiently; most of this sample have an sBHAR of less than 1% of their Eddington luminosity with the most frequent accretion rate being around 0.1%, with only one rising above 10%. In the high mass region I see a wide range of sBHARs which is restricted further down the mass scale. This effect is likely due to the fact that objects with lower sBHAR are missing at low masses as less efficiently accreting hosts will be more difficult to observe (due to their extremely low X-ray luminosities), and likely won't appear on the plot because of the threshold applied in the excess X-ray criterion.

3.5 Completeness-corrected Luminosity and Accretion Rate Distributions

Currently the sample is subject to significant observational biases as I preferentially identify the most luminous and actively accreting AGN, this shows a skewed picture of the distribution of AGN in the wider galaxy population. In this section I attempt to correct these biases to try and understand how the probability of a galaxy hosting an AGN as a function of luminosity and accretion rate varies across the full dwarf galaxy population.

Firstly, I made sure that my sample of AGN was consistent with the bulk of the underlying galaxy population. Figure 3.3 shows how the shape of the underlying

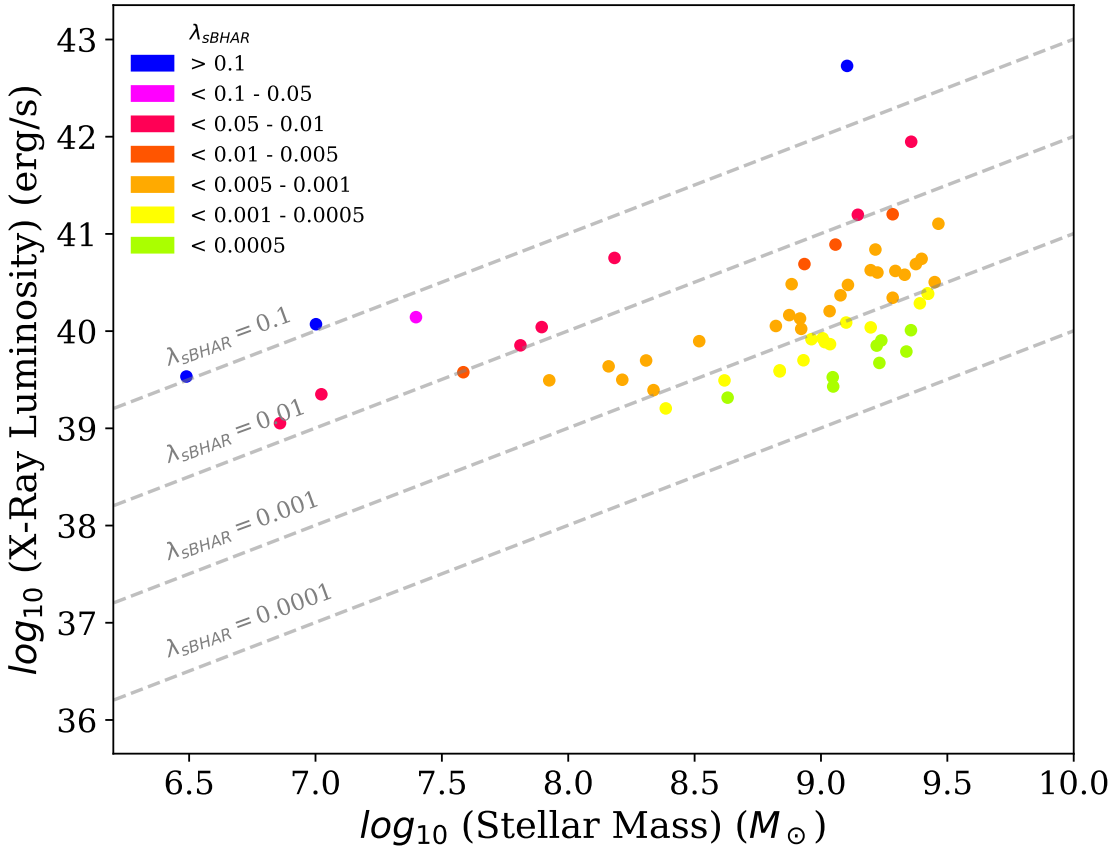

Figure 3.7: Figure showing the X-ray luminosity against galaxy mass for the 61 AGN hosts identified in section 3.2.4. The colour gives an indication of the accretion rate for the central black hole described in section 3.4

.

galaxy distribution changes with increasing mass and that some of my observed AGN lie at a comparatively high redshift. To correct this I took galaxies of all masses in narrow bands of redshifts from MPA-JHU and worked out the mass which contained ~ 90% of the galaxies. Through this process I defined a mass completeness limit as a function of redshift. The observed AGN were split into 3 mass bins: $9 - \log_{10}(M_*/M_\odot) < 9.5$; $8 - \log_{10}(M_*/M_\odot) < 9$; $6.8 - \log_{10}(M_*/M_\odot) < 8$. Each had the completeness function applied to it and any AGN that fell below it were removed. As a result, the sample was also limited to $z \leq 0.06$.

In addition, my sample needs to be constrained so it only contains AGN with a statistically significant detection in the energy bands used in the X-ray completeness analysis. This analysis needs to be done to account for AGN that may have been missed due to the varying sensitivity of 3XMM. To do this I used Flix (Carrera et al. 2007), 3XMM's upper limits service. It provides an upper limit, flux estimate and associated error broken down by band and instrument for the whole of 3XMM.

I use the upper limits and observed fluxes from band 8 in the PN camera; not only does this band cover the entire energy range of 3XMM but it also has the greatest number of AGN hosts that meet or exceed the 3XMM's detection threshold. Thus I restricted my observed AGN sample to only those with a detection likelihood > 6 in this band. Taking this and my mass and redshift corrections into account leaves me with 28 AGN hosts. I will refer to these AGN as the statistical sample which form the basis of the probability distributions. These AGN are marked with an asterisk in the summary table found in appendix B.

To account for the varying sensitivity, I need to characterise the distribution of X-ray detection upper limits in the region of MPA-JHU with coverage from 3XMM. There are 6,447 dwarf galaxies in MPA-JHU within 15' of a 3XMM source. The co-ordinates of these galaxies were uploaded to Flix. 4,331 dwarf galaxies fell within the 3XMM sky so Flix was able to produce X-ray flux limits at their co-ordinates. Thus these 4,331 dwarf galaxies are considered as the parent sample. Using the dwarf galaxy's redshift, this flux limit could be turned into a luminosity upper limit. Once all the luminosity upper limits were recorded, they were converted into a cumulative histogram as a function of X-ray luminosity and normalised by the size of the parent sample. This gave me a distribution of the fraction of galaxies where an AGN could have been detected above a given X-ray luminosity, a distribution which will be referred to as the luminosity sensitivity function. I can use this information to correct the observed distributions of luminosities and account for the varying sensitivity of 3XMM, allowing the recovery of estimates of the true underlying distribution of

luminosities within my samples of dwarf galaxies.

To produce the probability distribution, the observed AGN in each mass interval were binned as a function of the observed X-ray luminosity. For each luminosity bin, I use the sensitivity function to determine the number of galaxies within the parent sample where the 3XMM data is sufficiently sensitive to detect an AGN of this luminosity. I then divided the total number of X-ray detections by the expected number of galaxies to provide an estimate of the true probability of finding an AGN with such luminosities. This process produced the probability distributions seen in the left-hand column of figure 3.8. They show the probability of finding an AGN within nearby dwarf galaxies, in each mass and redshift interval, as a function of the observed X-ray luminosity.

Correcting the statistical sample of AGN with its corresponding luminosity sensitivity function has removed some of the observational bias described earlier and can be seen in figure 3.8. In contrast to the observed distribution shown in figure 3.3, the probability of finding an AGN generally increases towards lower observed X-ray luminosities across all mass bins.

Up until this point I have assumed all these objects are AGN, based on the low-mass and high mass X-ray binary checks applied to my sample. Whilst the models used do consider the integrated X-ray emission from the galaxy up to around 10^{39}erg s^{-1}, an ultra luminous X-ray object (ULX) located in the galactic centre and emitting significantly more than the rest of the galaxy could potentially have been included in this sample. To check whether or not these objects had been included in my sample, I used the Mineo et al. (2012a) X-ray luminosity function (XLF). It models the XRB populations, including ULXs, as a two-part power law normalised by the host galaxy's SFR. In each mass interval, the galaxies' XLFs were calculated and averaged to show how the average number of ULXs compared to the average number of AGN as a function of observed X-ray luminosity. As can be seen on the plots in the left-hand column of figure 3.8, the vast majority of the data points do not overlap with the ULX XLFs.

Whilst this large gap does exist, it is important to quantify how many ULXs are expected in this sample. Stochastic star formation could produce a single, very luminous ULX which may account for some of the overlap seen in the lowest mass probability distributions. To calculate the number of expected ULXs, I folded the ULX luminosity function through the correction fractions extracted from the 3XMM sensitivity curve. This calculation suggests there are 1.26 individual ULX detections

within my dwarf galaxy sample. Thus, the probability that 1 or more of my X-ray detections is in fact a ULX, rather than an AGN, is ~72% whilst the probability of 2 or more contaminants is ~36%, assuming a Poisson distribution.

It has also been noted that galaxies with lower than solar metallicities, like the dwarf galaxies being studied, have an enhanced HMXB population (Brorby et al. 2014). Thus the Mineo et al. (2012a) XLF could be underestimating the probability of finding a HMXB in this population of galaxies. Lehmer et al. (2019) observes this enhancement in 4 dwarf galaxies of similar masses and metallicities to those in my sample. I follow the same procedure outlined earlier but instead use the normalisations calculated in Lehmer et al. (2019) and observe some increase in the probability of finding a HMXB at any given X-ray luminosity. I have not shown this relationship on figure 3.8 because the small sample size means the relationship will have large and undefined uncertainties and I believe plotting the relationship would appear overly definitive. When the Lehmer et al. (2019) relationship is plotted on the lowest galaxy mass panel it is clear that the increase does overlap with the lower luminosity half of the fit line which suggests that some of these detections may not be AGN. However, when plotted in the higher host galaxy mass bins the increase is shown to be insufficient to overlap with the probability distributions. Thus, I am confident that the observed sample residing in the higher mass dwarf galaxies are AGN.

After confirming my sample were AGN I could confidently construct probability distributions for the sBHAR, as first considered in section 3.4. Observed X-ray luminosity can be affected by a number of host galaxy properties so by doing this I can reduce the observational bias whereby a black hole growing at a given accretion rate in a low mass galaxy produces a similar observable X-ray luminosity when compared to a black hole with a lower accretion rate in a higher mass host galaxy. I can also confirm whether the results shown in figure 3.7 are consistent with the underlying population of AGN in dwarf galaxies. I repeated the upper limits correction process described earlier but instead binned the observed and upper limits data as a function of sBHAR. The results, plotted in the right-hand column of figure 3.8, show a similar dynamic range to the observed X-ray luminosities that varies significantly depending on the stellar mass bin. In addition as we move down the mass scale, the average sBHAR increases. This is consistent with the results in figure 3.7 as lower mass galaxies require a larger sBHAR to be observed.

A power law of the following form is fit to each of these plots,

$$\log_{10}(p(X)) = A + k(\log_{10}(X) - x') \tag{3.9}$$

where $p(X)$ is the probability of observing an AGN with corresponding X-axis quantity, X. Each equation is centred at $x' = 40.5$ for luminosity and -2.5 for sBHAR, this is the median value for each quantity in the full statistical sample. The power laws are shown as a dashed red line allowing me to more clearly identify how the probability of finding an AGN in a dwarf galaxy, within a given mass and redshift, changes as a function of observed X-ray luminosity and sBHAR.

These power law fits reinforce the effect of the upper limits correction, giving me insight into the true extent of black hole activity across the dwarf galaxy population. I can confidently say that, in the higher mass plots (top and middle rows), the average number of AGN in dwarf galaxies as a function of both luminosity and sBHAR are well described by a power law. The luminosity power law in both these mass bins is distinct from the Mineo et al. (2012a) XLF in both normalisation and index, highlighting the fact I have identified a distinct sample of AGN. Moreover, the luminosity power laws describing the high and middle mass samples are identical within the error regions. The sBHAR power laws do not have the same degree of similarity. However, they are very likely derived from a clean sample of AGN and their error regions are of a similar size to the luminosity power laws which suggests that sBHAR, like luminosity, could be a fundamental property of an AGN. This has been previously observed in Aird et al. (2012) for a higher redshift and higher mass sample.

A slight downward trend exists in the lowest mass bins but it is much more uncertain given only 4 AGN were identified in this region (compared to the 15 and 10 AGN found in the high and middle mass bins respectively). Thus each point is based on a single observation and leads to comparatively large errors.

Overall, AGN in dwarf galaxies are detected emitting at a range of luminosities, driven by a correspondingly large range of accretion rates. For each mass bin, the average number of AGN increases with decreasing X-ray luminosity and sBHAR. This is consistent with other AGN population studies (e.g. Aird et al. 2018; Georgakakis et al. 2017), highlighting the expectation that AGN with lower luminosities and sBHARs are much more numerous, despite our current inability to detect them.

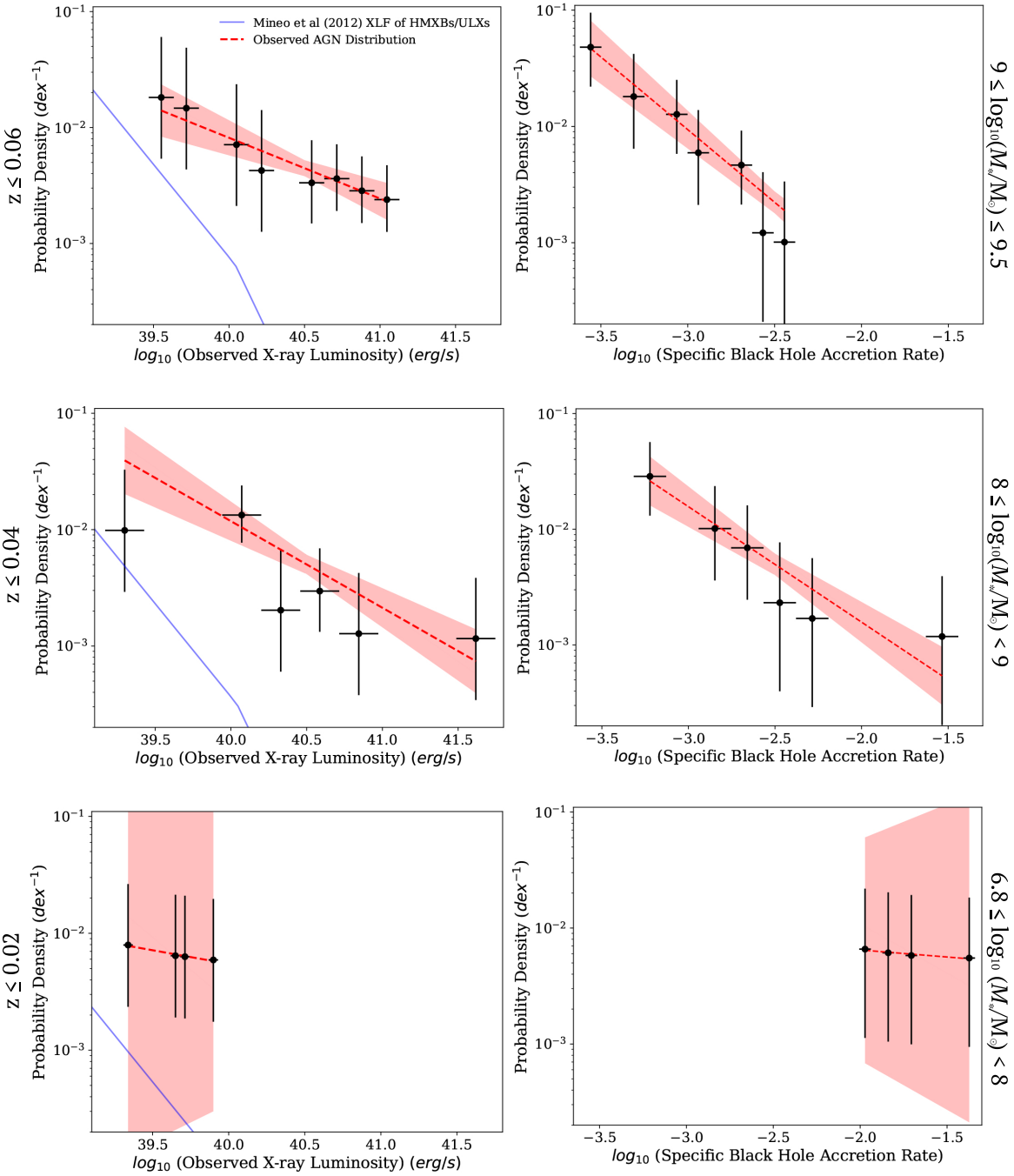

Figure 3.8: Distributions showing the probability of finding an AGN within the completeness corrected dwarf galaxy sample as a function of both observed X-ray luminosity (left-hand column) and specific black hole accretion rate (right-hand column). Power laws (dashed red line) were fit to each distribution, their uncertainty is also shown (pale red region). The redshift and mass limits applied to each row are shown. The blue line shows the expected average contribution from discrete, stellar-origin sources (HMXBs, ULXs) within my galaxy sample, based on the Mineo et al. (2012a) XLF scaled by the average SFR of the galaxies in my sample. It is plotted on each of the left-hand luminosity plots to rule out the possibility of a central ULX or group of ULXs being responsible for some of the low luminosity emission. See section 3.5 for more details on how these plots were constructed.

3.6 AGN Fraction as a Function of Host Galaxy Mass and Redshift

The fraction of galaxies that host an AGN within given mass and redshift regimes can be easily derived from the probability distributions shown in figure 3.8. In this section I use the luminosity probability distributions shown in the left-hand column of figure 3.8 to calculate robust AGN fractions above fixed luminosity limits which provide a direct comparison with previous work.

First, I consider how AGN fraction varies as a function of host galaxy mass. For each plot in the left-hand column of figure 3.8 the probabilities were summed, converted into fractions, with the following equation,

$$f(L_{\rm X} > L_{\rm min}) = \sum_{L_{\rm min}}^{42} p(\log_{10}(L_{\rm X})) \times \Delta \log_{10}(L_{\rm X}) \tag{3.10}$$

where $f(L_{\rm X} > L_{\rm lim})$ is the AGN fraction with X-ray luminosity $> L_{\rm min}$. The AGN fractions for the full range of X-ray luminosity ($L_{\rm X} > 10^{39}$ erg s^{-1}) are plotted in green on figure 3.9 at the median mass for that sample. The higher luminosity half of the observed AGN hosts ($L_{\rm X} > 10^{40.5}$ erg s^{-1}) are plotted as purple squares on the same figure. Overall there appears to be an increase in AGN fraction with host galaxy mass. This is consistent with the expectation that black holes in higher mass galaxies are increasingly common, thus as the incidence of black holes increases so too does the the incidence of AGN. As highlighted in section 3.5, AGN with low accretion rates are more common, thus the likelihood that an AGN would produce detectable emission rises with host galaxy mass. However, the $L_{\rm X} > 10^{39}$ erg s^{-1} fractions are also consistent with no change across the dwarf galaxy mass range.
At a fixed host galaxy mass, an increased luminosity threshold highlights the higher accretion rate end of the AGN probability distributions. As expected from the downward slope of the power laws in figure 3.8, there is a clear overall drop in $L_{\rm X} > 10^{40.5}$ erg s^{-1} AGN fraction. The increase with stellar mass for these fractions appears more statistically significant. It could suggest increasing host galaxy mass has some effect on increasing the incidence of AGN activity in highly accreting AGN. However, I cannot confirm this result for the lowest stellar masses in this regime as there were no AGN observed with these properties. The upper limit, presented as a triangle, is based on the assumption that a single AGN does actually exist; a black hole in a host of this mass would have to be accreting at an exceptionally high rate to produce detectable luminosities.

The errors on the AGN fraction as a function of host galaxy mass, $f(L_{\rm X} > L_{\rm lim})$,

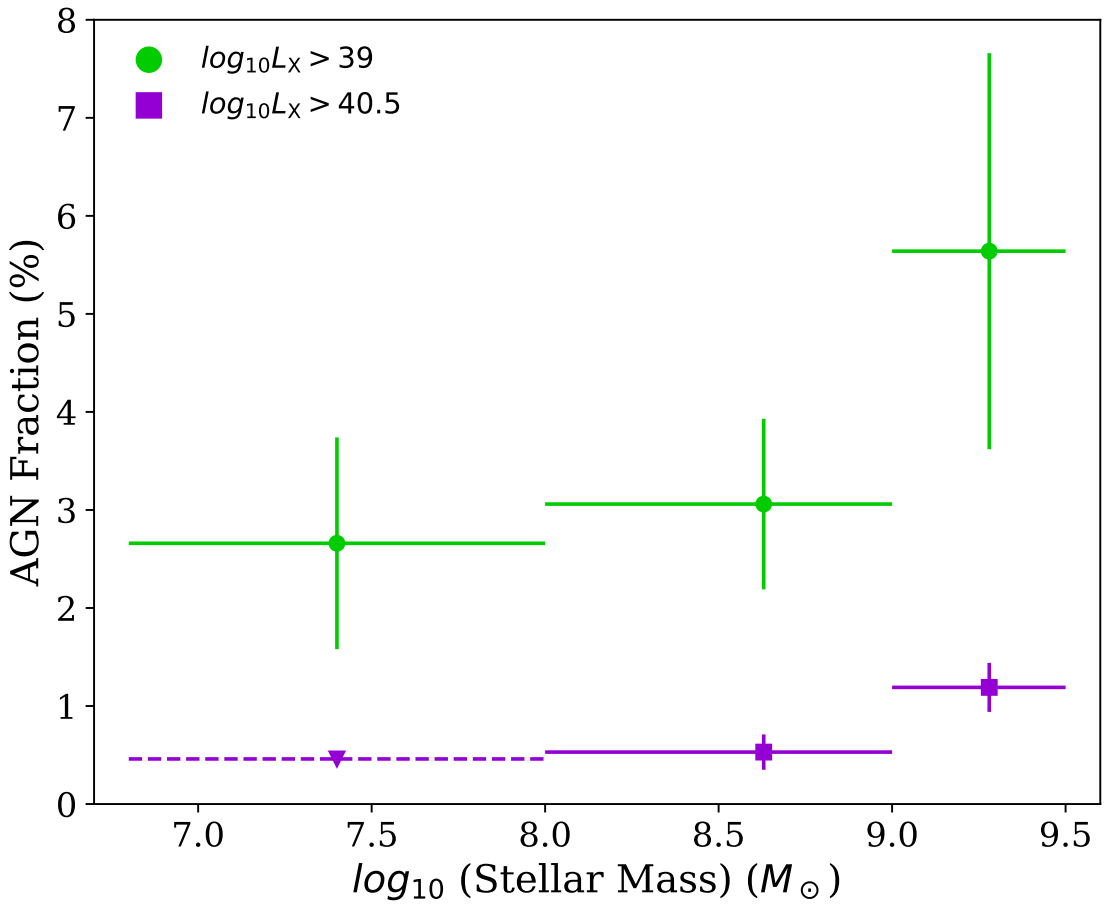

Figure 3.9: AGN fraction ($100 \times f(L_X > L_{lim})$) as a function of host galaxy mass for the full (green circles) and high luminosity (purple squares) samples as a function of host galaxy mass. AGN fraction increases with stellar mass in both luminosity regimes, however, the increase is only statistically significant for AGN with $\log_{10}(L_X) > 40.5$.

are found by summing the errors on each data point in quadrature. For the full luminosity bin, $\log_{10} L_X > 39$, the errors are consistent with a constant fraction up to $M_* \sim 10^{9.5}$ $M_\odot$. Perhaps host galaxy mass is not the only property that affect the incidence of AGN for the bulk of the population.

I also consider how the AGN fraction varies as a function of the host galaxy's redshift and compare it to Mezcua et al. (2018). They looked at dwarf galaxies ($10^7 < M_*/M_\odot < 3 \times 10^9$) in the Chandra COSMOS-Legacy survey out to $z \sim 2.4$. Using similar techniques to those outlined in section 3.2.4, they identified a sample of 40 AGN which they use to study the evolution of the AGN fraction as a function of stellar mass, X-ray luminosity and redshift. Their sample allows them to measure the AGN fraction out to $z \sim 0.7$ and $\log_{10} L_X \sim 42.4$; there is only one observed AGN in my sample that reaches luminosities in that range.

By extrapolating the power law fit made in section 3.5 I can estimate the AGN fraction at higher luminosities that are not directly probed by my study. I used the power law from the high mass bin ($9 < \log_{10}(M_*/M_\odot) < 9.5$) and re-ran the X-ray upper limits analysis for the $7 \leq \log_{10}(M_*/M_\odot) < 9$ bin used in Mezcua et al. (2018). I calculated the predicted AGN fraction as follows,

$$f_{\rm extrap}(L_{\rm X}) = \int_{L_{\rm min}}^{L_{\rm max}} A + k(\log_{10}(L_{\rm X}) - 40.5)\ d\log_{10}(L_{\rm X}) \tag{3.11}$$

where $f_{\rm extrap}(L_{\rm X})$ is the predicted AGN fraction between the mass and luminosity limits shown in figure 3.10. This figure shows my results plotted as circular points at the median redshift of each luminosity and mass bin. Shown alongside this is the Mezcua et al. (2018) data (square points), and others studies as lines, regions and points. The errors in each of my AGN fractions is found by accounting for the 1σ uncertainty in the power law fit parameters. Due to the limited redshift range of MPA-JHU, my points cover only the low redshift end of the axis, however in both panels they are consistent, within the errors, with the results from prior studies.

For the lower luminosity AGN, in the upper panel, my AGN fractions are consistent in both stellar mass bins. Within the uncertainties, my data is also consistent with Mezcua et al. (2018). Thus I don't find any evidence for a mass dependence in these lower luminosity AGN.

Generally lower fractions, in the bottom panel of figure 3.10, show that higher luminosity AGN are less abundant than their dimmer counterparts. This is consistent with earlier findings suggesting that AGN are more numerous at lower observed X-ray luminosity. Mezcua et al. (2018) suggested that the AGN fraction may decline with increasing redshift, although given the large uncertainties, their measurements are consistent with a constant AGN fraction out to $z = 0.7$. My result, in this mass and redshift bin, provides an estimate at low redshift, consistent with the Mezcua results at higher redshifts. My results, therefore, indicate that the AGN fraction in dwarf galaxies is constant with increasing redshift.

3.7 Conclusion

In this chapter, I have rigorously matched a sample of dwarf galaxies to their central X-ray counterpart. I then predicted the contribution that XRBs and hot gas may make to this X-ray emission based on each galaxy's mass and SFR. Any galaxies that had observed emission three times greater than what was predicted were considered

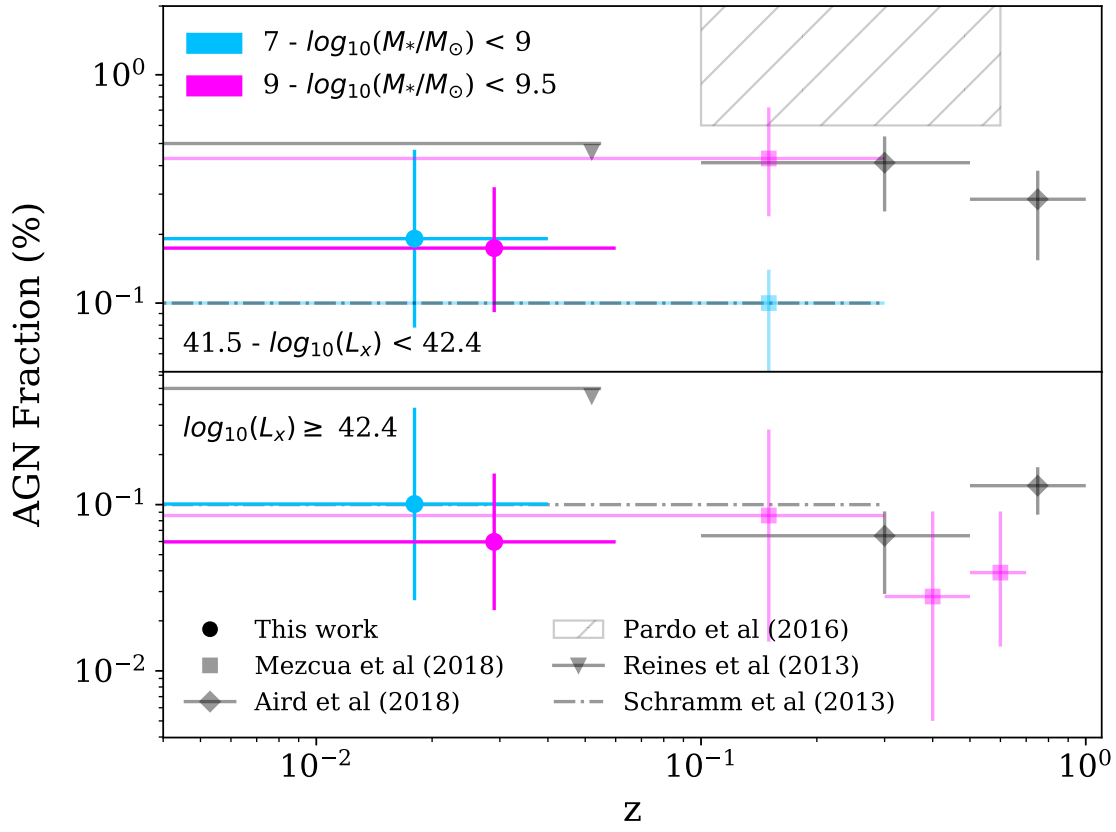

Figure 3.10: AGN fraction ($100 \times f_{\text{extrap}}$) as a function of redshift for different observed X-ray luminosity ranges, adapted from Mezcua et al. (2018). The circular points were derived from power law fits, within highlighted mass and corresponding 90% complete redshift regimes discussed in section 3.5.

likely to be host an AGN. I showed that placing this threshold likely isolated the hardest X-ray emission in this sample. Off-nuclear emission cannot be completely ruled out given the resolution of 3XMM, however recent simulations by Bellovary et al. (2019) and observations in Reines et al. (2020) suggest that such a detection could still be an AGN. Thus, I can confidently say I have identified 61 dwarf galaxy AGN hosts. 40 of these AGN are brand new detections (see appendix B for details).

I then performed BPT analysis on the 53 AGN hosts with significant detections in the required emission lines to see if their optical characteristics matched their X-ray classification. I found that the vast majority of these galaxies were classified as star-forming. My result adds to a growing body of evidence suggesting that optical selection methods may miss AGN, particularly those residing in dwarf galaxies. To investigate if star formation dominated the optical component of AGN emission, I took the X-ray emission - dominated by the AGN - and translated it down to the SDSS u-band, to find out the AGN's contribution to the host galaxy's optical emission which is compared to the observed u-band flux. All but one of the X-ray selected AGN classified as star-forming had their optical emission dominated by the galaxy. This finding shows that star formation can confuse the results from BPT diagnostics and hide the signatures of AGN.

Next I investigated the activity of the central SMBHs powering our AGN by calculating their accretion rates. I found that the SMBHs span a wide range of accretion rates but that none are accreting at high rates. The most common environment in my sample is a host of mass $\sim 10^{9-9.5}$ $M_\odot$ with an SMBH accreting at about 0.1% of its Eddington luminosity. The most active SMBH has an accretion rate of $\sim$ 10% of its Eddington luminosity.

Finally, I attempted to correct my sample to account for AGN that could have been missed due to the varying sensitivity of 3XMM. To do this, the observed AGN sample needed to be reduced so it matched the underlying galaxy distribution and had significant detections in the appropriate 3XMM band. Thus the observed sample was reduced from 61 to a statistically robust sample of 29 AGN, which were then split into 3 mass and redshift bins. To correct the observed distribution, I determined the upper X-ray flux limit at the positions of the 4,331 dwarf galaxies in MPA-JHU that lie within the 3XMM footprint. I used these flux upper limits to determine a sensitivity function that allows me to correct the observed distributions of AGN luminosities for incompleteness and recover the true probability distribution functions of AGN luminosities and specific accretion rates within the dwarf galaxy population. As a final check, the luminosity probability distributions

were compared to the Mineo et al. (2012a) XLF of HMXBs and ULXs. They were found to lie above the XLF indicating that I am identifying a distinct population of AGN within dwarf galaxies and are not significantly contaminated by the detection of individual, bright ULXs within the galaxy. The probability distributions show that AGN in dwarf galaxies have a wide range of activity, with the probability of identifying an AGN being well described by a power law. AGN are more numerous at lower X-ray luminosities and sBHARs.

I used my robust measurements of the probability of hosting an AGN as a function of X-ray luminosity to determine how the incidence of AGN varies as a function of other galaxy properties. I found evidence that the fraction of galaxies with an AGN above a luminosity limit of $L_{\rm X} > 10^{39}$ erg s^{-1} increases as a function of stellar mass, rising from $\sim 2.7\%$ at $M_* \sim 10^{7.4}$ M$_\odot$ to $\sim 6\%$ at $M_* \sim 10^{9.5}$ M$_\odot$. However, the fractions are consistent with showing no change across the stellar mass range. I also extrapolated my measurements to higher luminosity thresholds and compared to higher redshift measurements from Mezcua et al. (2018), finding no evidence for any evolution in the AGN fraction in dwarf galaxies out to $z \sim 0.7$.

Chapter 4

The Incidence of AGN in the Nearby Universe

4.1 Introduction

The degree to which an SMBH and its host galaxy interact and affect each other's evolution is an important and contested question. Among the well-studied and constrained examples of co-evolution are the relations between an SMBH's mass and the velocity dispersion, mass and luminosity of the host galaxy's classical bulge. (Ferrarese and Merritt 2000; Gebhardt et al. 2000; Magorrian et al. 1998). Such relationships imply that the SMBH and bulge co-evolve by regulating each others growth (Kormendy and Ho 2013). SMBH growth can also be identified through a range of electromagnetic signatures released when they undergo periods of accretion. Tracing the features of AGN activity across cosmic time has highlighted more evidence of co-evolution. For example, total AGN accretion rate density is highly correlated with changes in star-formation rate (SFR) density (Aird et al. 2015; Delvecchio et al. 2014; Madau and Dickinson 2014); the effect of AGN on their host galaxies in the form of feedback is associated with the quenching of star formation (Di Matteo et al. 2005; Fabian 2012; Greene et al. 2020); and feedback could also explain the observed bi-modality in colour-magnitude and colour-mass space (Baldry et al. 2004; Martin et al. 2007; Schawinski et al. 2014).

It is still unclear what mechanism is the main driver of gas and dust moving down into the centre (Alexander and Hickox 2012). As discussed in a number of studies, taking large samples of galaxies is a useful way of constraining the AGN fuelling mechanism as it allows us to smooth out stochastic differences and try to identify trends (Aird et al. 2013; Delvecchio et al. 2020; Hickox et al. 2014). The most complete AGN samples are typically found by using X-ray surveys as they can iden-

tify AGN across large redshift ranges and down to relatively low luminosities where non-AGN emission may dominate (Brandt and Alexander 2015). Studies that have adopted this approach suggest that AGN are much more likely to be found in higher mass galaxies, and in galaxies at higher redshifts (e.g. Haggard et al. 2010; Hernán-Caballero et al. 2014; Mendez et al. 2013; Williams and Röttgering 2015; Xue et al. 2010).

Whilst these surveys employ large samples of AGN, they are subject to several observational biases. Firstly, most X-ray detection techniques define AGN as objects that exceed absolute luminosity thresholds, or more generally dominate galaxy emission (Brandt and Alexander 2015), causing them to miss the lower luminosity black holes, accreting at lower rates, typically found in dwarf galaxies. Any low mass black holes that are detected are the much rarer, more actively accreting ones. However, over the past decade there has been a huge increase in the number of AGN detected in dwarf galaxies aided, in part, by adapting these techniques. X-ray selected AGN in dwarf galaxies are identified through isolating a centrally located excess emission that cannot be explained by other sources, including X-ray binaries and hot gas (Baldassare et al. 2017; Birchall et al. 2020; Lemons et al. 2015; Mezcua et al. 2018; Paggi et al. 2016; Pardo et al. 2016). Such a change in detection method allows us to overcome the bias towards actively accreting, higher mass black holes associated with the absolute X-ray luminosity thresholds and helps challenge the assumption that AGN are less likely in lower mass galaxies.

Secondly, a lot of these studies make little attempt to understand and correct for the limitations of the survey data used. Aird et al. (2012) was one of the first AGN studies to attempt to overcome these observational limitations and apply completeness corrections to their sample. By applying these corrections to a sample of galaxies out to $z \approx 1$, from the Prism Multi-object Survey (Coil et al. 2011; Cool et al. 2013), they were able to determine the probability of finding an AGN as a function of various host galaxy properties. Aird et al. (2012) showed that the probability of hosting an AGN can be described as a power law of X-ray luminosity and specific black hole accretion rate (sBHAR). Both X-ray luminosity and sBHAR distributions were found to be consistent over a wide range of stellar masses but with normalisations that dropped rapidly with decreasing redshift.

Similar approaches to Aird et al. (2012) have been adopted and adapted by subsequent AGN population studies (Aird et al. 2017, 2018; Birchall et al. 2020; Bongiorno et al. 2016; Georgakakis et al. 2017). In chapter 3, I found power law distributions describing the AGN incidence in dwarf galaxies as a function of X-ray luminosity

and sBHAR. From this I identified a redshift-independent fraction out to $z \approx 0.7$. I also found a weak increase in AGN fraction with stellar mass, however it was also consistent with being flat for the range probed.

In this chapter, I extend the work done in chapter 3, remove the mass threshold and investigate the X-ray luminosity and sBHAR distributions for this expanded sample of local AGN. With these distributions I aim to extend the ranges probed previously and identify whether there is any connection between AGN incidence and stellar mass or redshift.

4.2 Data & Sample Selection

As this chapter builds upon the work done in chapter 3, I continue to use the SDSS value-added catalogue, MPA-JHU, and 3XMM DR7. See chapter 2 for more information on these catalogues. In this section, I will outline the new position-matching process needed to efficiently match my expanded sample.

4.2.1 Position Matching

In chapter 3, the dwarf galaxies in MPA-JHU catalogue were matched to 3XMM with a simple but robust statistical technique. The small sample size allowed for individual inspection of degenerate matches i.e. galaxies matched to multiple X-ray signals, and vice versa. By removing the mass threshold, the potential AGN sample expands massively. This opened up the possibility of thousands of potential degenerate matches between MPA-JHU and 3XMM. An automated method that assessed the strength of these degenerate matches would be needed to feasibly process such a large catalogue.

To overcome this challenge I turned to the ARCHES cross-correlation tool, `xmatch` (Pineau et al. 2017). It is an astronomical matching tool able to identify the counterparts of one catalogue to multiple others, whilst also computing probabilities of associations using background sources and positional errors. In consultation with its creator, I wrote a script to match MPA-JHU and 3XMM DR7. It uses a standard approach adopted when working with XMM, where the data and matching process is broken down into individual XMM fields. Every 3XMM and SDSS object within 15' of the field's central co-ordinates was found. Once complete, a level of quality control was applied, to isolate the most secure X-ray detections - those entries whose absolute position error was no greater than 4" and whose physical extent was less

than 10". Crucially, `xmatch` calculates the Bayesian probability of association and non-association for each remaining pair of X-ray and galaxy co-ordinates within the current 3XMM field. After repeating this process for all 3XMM fields, `xmatch` identified 110,274 potential matches.

Unfortunately, by focusing on individual fields the matching process did not account for any overlap that might occur between them. Before I assessed the match strengths, I removed repeating galaxy or X-ray matches produced from overlapping fields. Once done, I isolated the SDSS objects with an entry in MPA-JHU. All the SDSS co-ordinates were matched to MPA-JHU co-ordinates within 1", leaving 3,357 galaxies with mass and SFR estimates. Next, I identified the most likely X-ray & galaxy pairs. I chose to set a 90% probability of association as the matching threshold which produced a well-matched sample of 1,559 X-ray emitting galaxies.

4.2.2 Clarifying the Effect of High Optical Emission on the SFRs & Stellar Masses

As outlined in section 2.1.2, I am confident that these values are robust and well-suited to our study. However, in this sample I expect to find high luminosity AGN. Thus I want to check that strong central optical emission, likely originating from an AGN, is not biasing these quantities within my sample of candidate AGN-hosting galaxies. To determine what proportion of the total observed optical emission originates from the AGN, I drew upon the techniques used in section 3.3.2. This technique aimed to calculate and compare the luminosity densities at $3550\dot{A}$ originating from the AGN and compare to it the overall optical emission ($L_{\nu,3550\dot{A}}(\mathrm{AGN}_{\mathrm{predicted}}/(\mathrm{AGN}+\mathrm{Galaxy})_{\mathrm{observed}})$). Thus, for the sake of this calculation, I make the assumption that all 1,559 X-ray emitting galaxies host an AGN.

Figure 4.1 shows the stellar mass and SFR distribution of the 1,559 X-ray emitting galaxies as black contours. Overlaid on these contours are the galaxies with high predicted AGN contributions, shown as coloured points. As discussed previously, the Lusso and Risaliti 2016 relation assumes the AGN are un-obscured. Since MPA-JHU selects only narrow-line objects this predicted AGN luminosity will be an upper limit on its contribution. Thus some of these AGN have predictions that are greater than the total. However, this constitutes less than 20% of the galaxy sample. This drops to ~ 7% at $L_{\nu,3550\dot{A}}(\mathrm{AGN}_{\mathrm{predicted}}/(\mathrm{AGN}+\mathrm{Galaxy})_{\mathrm{observed}}) > 10$ and ~ 1% at $L_{\nu,3550\dot{A}}(\mathrm{AGN}_{\mathrm{predicted}}/(\mathrm{AGN}+\mathrm{Galaxy})_{\mathrm{observed}}) > 100$. These are the AGN with strong optical emission that I am interested in. Figure 4.1 clearly shows that even with a strong predicted AGN contribution, the stellar mass and SFR values

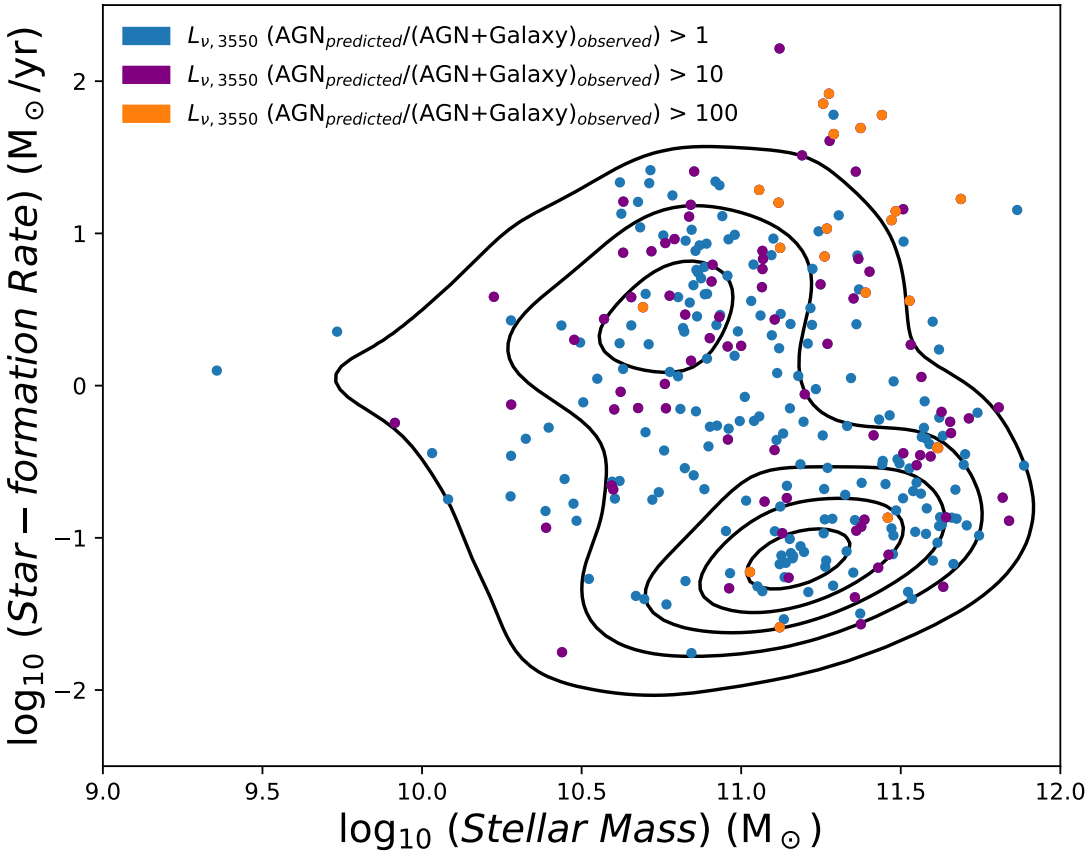

Figure 4.1: Comparing the stellar mass and SFR of galaxies with a high, predicted relative optical AGN contribution ($L_{\nu,3550\dot{A}}(\mathrm{AGN_{predicted}}/(\mathrm{AGN}+\mathrm{Galaxy})_{\mathrm{observed}})$; coloured points) against the the full X-ray emitting galaxy sample (black contours). The contours encompass 90%, 70%, 50%, 30% and 10% of the galaxy sample. See section 4.2.2 for more information on how $L_{\nu,3550\dot{A}}(\mathrm{AGN_{predicted}}/(\mathrm{AGN}+\mathrm{Galaxy})_{\mathrm{observed}})$ was calculated.
.

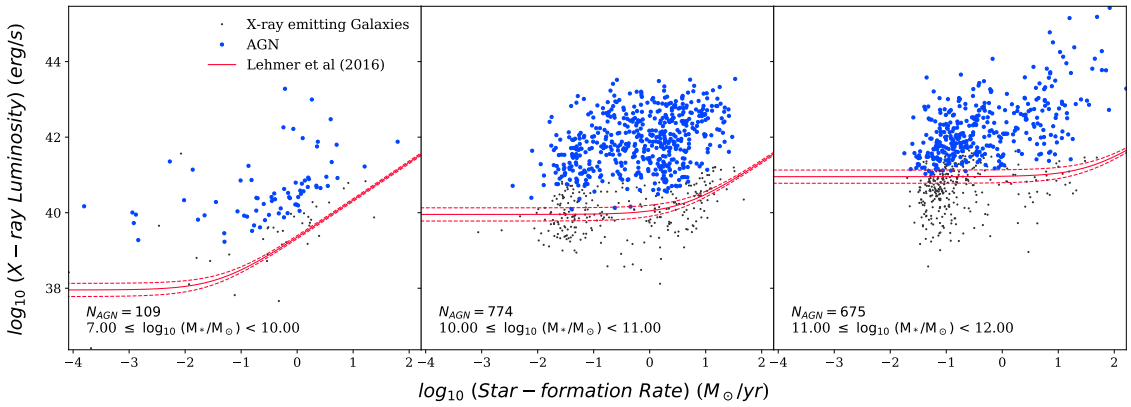

Figure 4.2: Observed X-ray luminosity against SFR for the 1,559 X-ray emitting galaxies which are candidates to host AGN. For clarity, these candidates were split up into three panels of different host galaxy masses (highlighted in text on each panel). Objects that met or exceeded the threshold are plotted in blue, those that did not are grey. The red line is the Lehmer et al. (2016) equation, the dominant source of X-ray emission prediction, described in section 4.2.3. Variation due to uncertainty in coefficients is highlighted by the dashed lines. By comparing it across plots we can clearly see the variation in mass and SFR dependencies.

calculated are consistent with the wider X-ray emitting galaxy distribution. Thus I can continue to confidently use these values.

4.2.3 Identifying AGN

To identify emission from an AGN, I modelled the combined emission coming from other X-ray emitting sources - X-ray binary stars and hot gas emission - and compared it to the observed X-ray luminosity, following the process used in chapter 3.

Based on this criterion, 949 X-ray emitting galaxies were classified as AGN. Figure 4.2 shows a breakdown of this classification process, comparing the observed X-ray luminosity with the SFR for every X-ray emitting galaxy and highlighting whether it has been classified as an AGN. The sample was split into several mass bins for clarity but it also highlights the changing nature of the Lehmer et al. (2016) equation. Each illustrated line was calculated using the mean stellar mass, indicated on the panel, and mean redshift of the bin. So each line should be considered an illustrative threshold. As the stellar mass increases, so too does the normalisation of equation (3.3) at low SFR. So for the highest mass galaxies, this relationship behaves almost like a luminosity cut of $> 10^{41}$ erg s^{-1}, lower than the typical threshold of 10^{42} erg s^{-1} (Brandt and Alexander 2015). At such low redshifts and SFRs, however, I trust that the selection criterion will continue to identify AGN activity. This increase in normalisation is believed to originate from a growing population of long-lived low mass X-ray binaries that occurs in more massive galaxies. Variation

due to uncertainty in coefficients is highlighted by the dashed lines.

To measure an accurate AGN fraction for this sample I had to ensure it is a statistically complete sample of galaxies above a given stellar mass limit. Thus, I took galaxies of all masses in narrow bands of redshifts from MPA-JHU and determined the mass range which contained ~ 90% of the galaxies. This process allowed me to create a 90% mass completeness function. Figure 4.3 shows the distribution of AGN on the redshift and stellar mass plane, with the completeness function highlighted as a red line. I have split this into 6 redshift intervals between 0 and 0.35 - 5 of width 0.05 and the final being 0.1 in width. These bins are highlighted by their colour in the figure. Only those AGN which lie above the completeness function were binned, resulting in the removal of 32 objects.

4.3 BPT Classification

AGN activity can impact the host galaxy's emission across the electromagnetic spectrum. The BPT diagnostic (Baldwin et al. 1981) is a commonly used technique to identify the primary source of ionising radiation in the optical part of the spectrum. By comparing the ratios of various emission lines, it is possible to gain insight into whether star formation, AGN or a composite of both processes dominate in any given galaxy. In chapter 3, I found that this diagnostic would have missed around 85% of my X-ray selected AGN in dwarf galaxies. Now with this larger sample I am able to investigate how the accuracy of this diagnostic changes with stellar mass.

Of the 917 AGN hosts I identified using X-ray selection techniques, 658 had significant detections ($\frac{\text{Line Flux}}{\text{Line Flux Error}} > 3$) in each of the required emission lines so these were used in my analysis. In the upper panel of figure 4.4, I present the results of the BPT analysis with my AGN sample plotted as large points, coloured to indicate their stellar mass. Black lines separate the AGN hosts into different classifications: objects with ionisation signatures predominately from AGN lie in the top-right, those dominated by star formation in the bottom-left, and those that have composite spectra are in the central region (Kauffmann et al. 2003b; Kewley et al. 2001). Underneath these points, in grey, are a subset of the MPA-JHU galaxies used to illustrate the underlying BPT distribution. Encouragingly, the vast majority of my AGN appear to lie within the AGN region. However, the colours clearly show that most of the objects found here are higher mass galaxies, with the lower mass points favouring the star-forming and composite regions. The behaviour exhibited by these low mass galaxies mirrors what I found in chapter 3 and is consistent with a growing body of

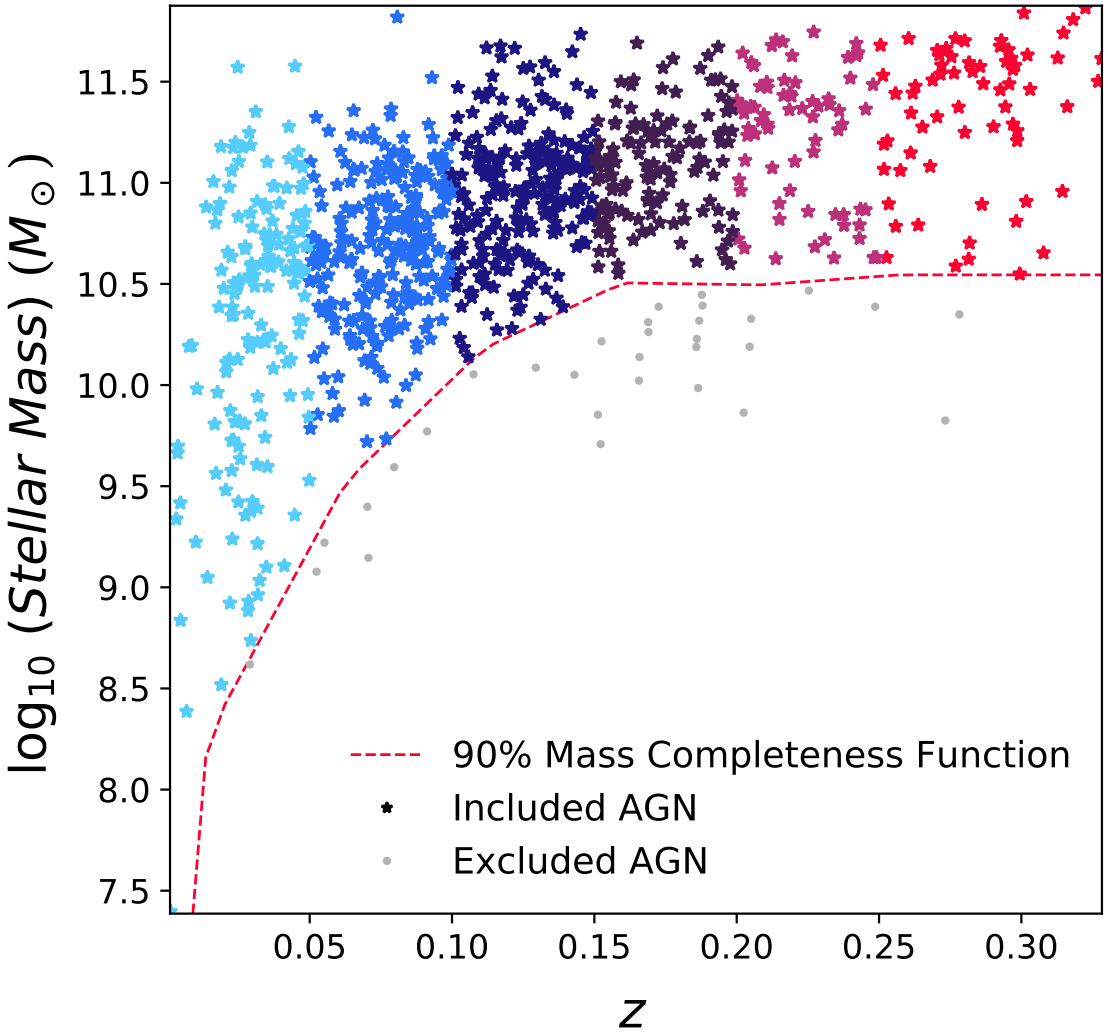

Figure 4.3: Redshift against stellar mass for the 949 AGN identified with equation (3.5). This sample has been limited by the 90% mass completeness function (red, dashed line), resulting in the removal of 32 AGN (grey points). The remaining 917 AGN have been split into 6 redshift bins outlined in section 4.2.3, and separated by colour for clarity.

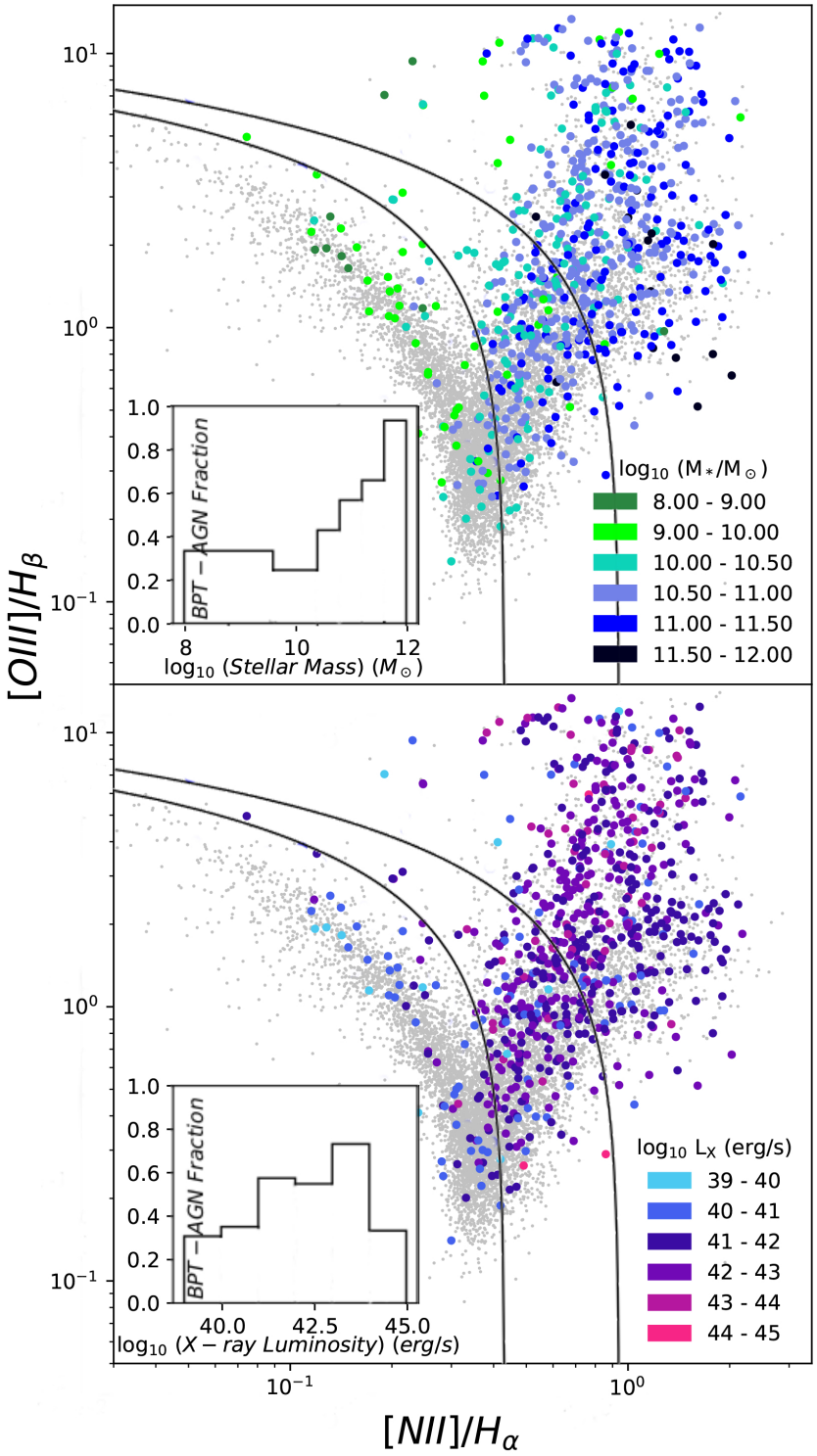

Figure 4.4: BPT diagnostics for the 658 X-ray selected AGN with significant detections in all the emission lines shown. In the upper panel, the colour of each AGN point represents the mass of the host galaxy, outlined in the legend. In the lower panel, the colour of each AGN point represents the observed X-ray luminosity of the AGN. Inset into each panel, we see the fraction of X-ray selected AGN with the same classification from the BPT diagnostic, as a function of the relevant property.

work suggesting that these optical spectroscopic measurements are less effective at identifying AGN in lower mass galaxies (Agostino and Salim 2019; Cann et al. 2019).

To further interrogate this trend, I calculated the proportion of X-ray-selected AGN with the same BPT classification as a function of stellar mass. Inset into the upper panel of figure 4.4, we can see that the effectiveness of the BPT diagnostic in reproducing my X-ray based classification increases with stellar mass. At the low mass end we can see similar behaviour to that found in my previous work, with the diagnostic missing around 75% of my dwarf galaxy AGN. However, the accuracy rises steadily until it identifies 93% of my AGN in the highest mass host galaxies.

Chapter 3 concluded that the AGNs' relatively low luminosity caused their stark mis-classification. In the lower panel of figure 4.4 I investigated how the changing X-ray luminosity affects the BPT classification. Inset in this panel is a distribution showing the changing BPT accuracy with X-ray luminosity. Whilst there is an increase in BPT accuracy, it clearly peaks before reaching the highest luminosity AGN: rising from 36% accurate, in the lowest mass bin, to 74% in the most populous. Clearly then, increasing X-ray luminosity is not the only variable that affects mass dependence of the BPT accuracy.

Behroozi et al. (2019) collated the quenched fraction of galaxies in the nearby universe (based on Bauer et al. 2013; Moustakas et al. 2013; Muzzin et al. 2013) and showed it increases strongly with stellar mass. This means higher mass galaxies are much less likely to have their emission lines driven by star formation activity. As we move towards higher mass galaxies, the typically higher luminosity AGN combined with this drop in star formation would see AGN-driven signatures dominate and produce the concurrent increase in BPT accuracy.

4.4 Specific Black Hole Accretion Rate

AGN are powered by mass accretion onto a galaxy's central SMBH. However we gain little insight into this accretion process by only looking at the observed X-ray luminosity. For example, one black hole growing at a higher accretion rate in a lower mass galaxy could produce a similar X-ray luminosity to another black hole with a lower accretion rate in a high mass galaxy. To break this degeneracy and understand more about how the black hole activity is distributed across my observed AGN sample, I looked at the specific black hole accretion rate (sBHAR), λ_{sBHAR}. This quantity compares the bolometric AGN luminosity of the galaxy with an estimate of the black hole's Eddington luminosity to give an indication of how

efficiently the black hole is accreting. It is found using,

$$\lambda_{\mathrm{sBHAR}} = \frac{25 L_{2-10\mathrm{keV}}}{1.26 \times 10^{38} \times 0.002 M_{*}} \approx \frac{L_{\mathrm{bol}}}{L_{\mathrm{Edd}}} \tag{4.1}$$

and is taken from Aird et al. (2012). Figure 4.5 shows how my observed AGN sample are distributed on the stellar mass and X-ray luminosity plane, with the data point colour giving an indication of the sBHAR of each galaxy. I have also highlighted lines of constant sBHAR to give a sense of the luminosities expected from a host galaxy of a given mass at those accretion rates. This figure further highlights that the vast majority of the observed AGN sample lies above stellar masses of 10^{10} $\mathrm{M}_{\odot}$. There is also a clear preference for black holes with relatively low accretion rates. Most of them have accretion rates that are less than 0.5% of their Eddington luminosity and only a handful of the most massive galaxies venturing above 10%.

4.5 Completeness-corrected Probability Distributions

Despite using AGN selection criteria that modify the X-ray luminosity threshold depending on the mass, SFR and redshift of the host galaxy, figure 4.5 clearly shows that my sample as it is currently constructed preferentially identifies AGN found in the highest mass galaxies. In this section I will extend the method described in chapter 3 to the full mass and redshift range. In doing so, I will create a series of probability distributions comparing different configurations of host galaxy properties with the effects of observational bias removed. From these distributions I aim to understand more about the underlying distribution of AGN in the nearby Universe.

4.5.1 Calculating Completeness Corrections

Using the 3XMM serendipitous source survey as my X-ray catalogue introduced a significant amount of variation in the detection sensitivity. Such variation brings with it the possibility that lower luminosity AGN within the overlapping region of 3XMM and MPA-JHU could have been missed as the flux limit was insufficient to detect it. To overcome this, I established the fraction of the parent galaxy sample that lie within 3XMM fields with sufficient sensitivity to detect an AGN above a given luminosity threshold.

To perform this analysis, I made use of Flix (Carrera et al. 2007), 3XMM's upper limits service. It provides upper limits broken down by band and instrument for the whole of 3XMM. From this data I chose the upper limits and observed fluxes from

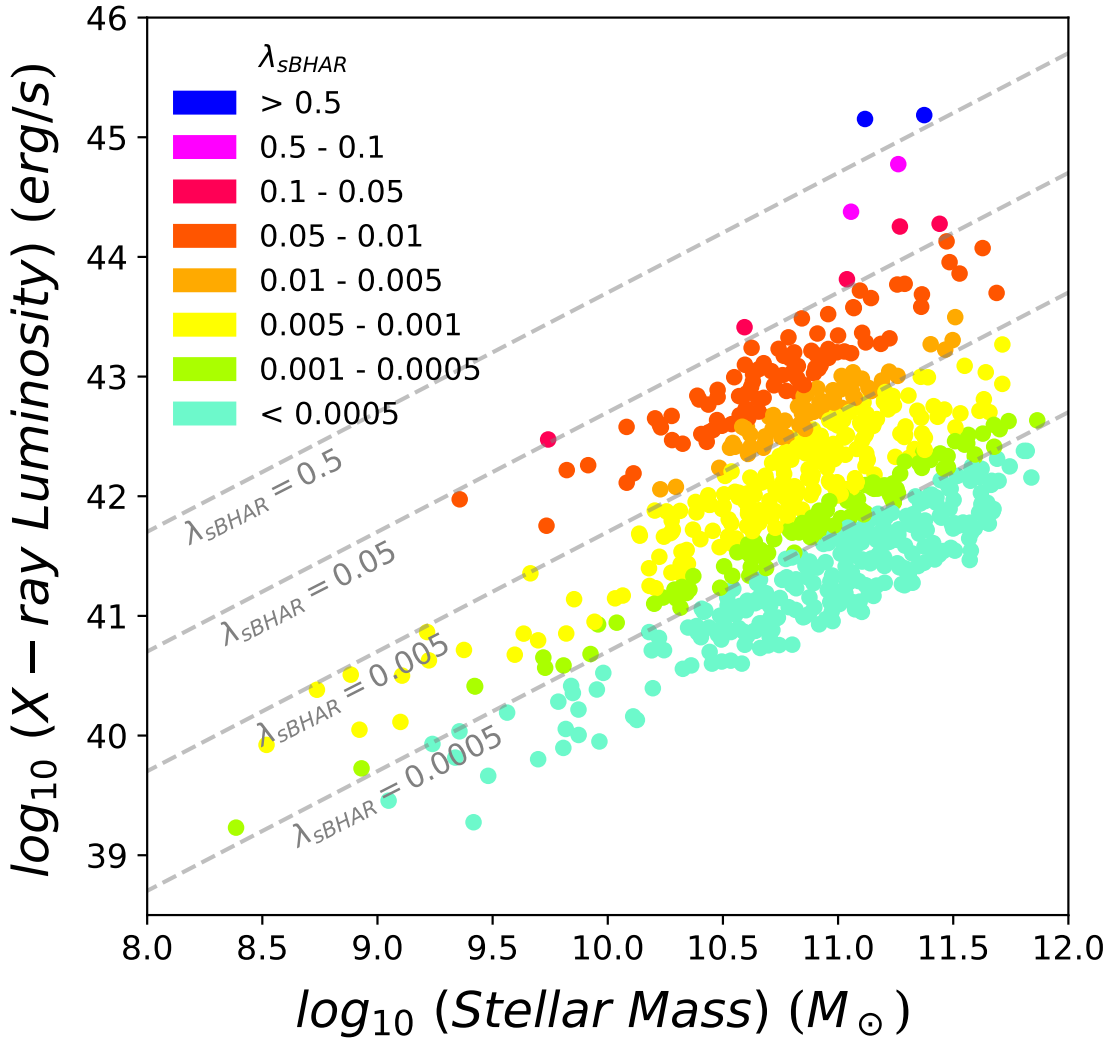

Figure 4.5: Stellar mass against the observed X-ray luminosity for the 917 AGN detected in section 4.2.3. Each point has been assigned a colour to provide an indication of the accretion rate (λ_{sBHAR}) of the galaxy's central SMBH, calculated in section 4.4 . Several grey lines of constant sBHAR have also been plotted for reference.

band 8 in the PN camera as it covers the entire energy range of 3XMM. Because of this restriction, I also limited the observed AGN to only those with detections in that same band. Not all objects will meet this detection threshold and some XMM fields will not have been observed by the PN camera so my observed AGN sample was reduced from 917 to 739.

To characterise this sensitivity variation, I need to find the X-ray detection upper limits of all the galaxies in the region of MPA-JHU that have coverage from 3XMM. There are 28,545 MPA-JHU galaxies found within 3XMM whose co-ordinates were uploaded to Flix. It returned the X-ray flux upper limits at the co-ordinates of 25,949 galaxies – my parent sample. I then extracted the flux upper limits for a given range of stellar masses and redshifts. These fluxes were then converted into luminosities using the redshift of the galaxy associated with each upper limit. Finally, these upper limits were used to construct a cumulative histogram function normalised by the total number of galaxies in the current mass and redshift range. This luminosity sensitivity function allowed me to determine the fraction of galaxies where an AGN could have been detected above a given X-ray luminosity within a range of masses or redshifts.

4.5.2 Creating the Probability Distributions

I am interested in calculating how the probabilities of finding AGN vary in the nearby Universe as a function of stellar mass and redshift. To illustrate the process of constructing these probability distributions, I will focus on how the probability varies with X-ray luminosity and stellar mass.

For this configuration of host galaxy properties, I split up the parent sample into a series of stellar mass bins. In each of these stellar mass bins, the data was further broken down as a function of X-ray luminosity, producing an observed AGN count distribution. In addition, I constructed a bespoke luminosity sensitivity function, as outlined in section 4.5.1. Each bin's count distribution was divided by the correction fractions extracted from this sensitivity function to recover the expected number of galaxies sensitive enough to detect an AGN as a function of luminosity. Finally, I divided these corrected AGN counts by the parent sample size in this bin to produce a series of probability distributions. The results of this process are shown in figure 4.6. It shows the probability of finding an AGN as a function of luminosity in bins of increasing mass. Some useful reference information is printed on each panel including the sizes of the AGN and parent samples, and the stellar mass range of the galaxies included therein.

By applying corrections extracted from the luminosity sensitivity function to the observed AGN counts I am able to provide robust measurements of the true incidence of AGN within the nearby galaxy population. In figure 4.6 we see that there is an abundance of AGN across stellar mass despite the clear favouring of higher mass AGN seen in figure 4.5. Of particular significance are the continuation of the findings from chapter 3: AGN populations are well described by power law distributions, with AGN being found predominantly at lower X-ray luminosities (see also Aird et al. 2012).

To calculate the errors in each probability data point I used the confidence limits equations presented in Gehrels (1986) meaning that the size of the error is determined based on the number of detected AGN in a given bin.

One final check is performed on the AGN populations shown in figure 4.6. Section 4.2.3 highlights how I have considered contamination from X-ray binaries and hot gas, but there remains the possibility that this emission could originate from a ULX, particularly in lower mass galaxies. To check whether a ULX could account for this emission, I used the Mineo et al. (2012a) XLF. XRB populations, including ULXs, are modelled using a two-part power law normalised by the host galaxy's SFR. In each stellar mass bin, XLFs associated with each individual galaxy were calculated and averaged to show how the predicted number of ULXs compared to my AGN observations. The blue lines in each panel of figure 4.6 show that the vast majority of the data points do not overlap with the ULX XLFs. The only point at which the two distributions do meet is at the low luminosity end of the lowest stellar mass bin.

Since the overwhelming majority of my sample are now all confirmed AGN I could confidently fit power laws to each of these panels. They have the following form,

$$p(X) = A\left(\frac{X}{x'}\right)^{k} d\log_{10} X \tag{4.2}$$

where $p(X)$ is the probability of observing an AGN with corresponding X-axis quantity, X and centred on a value x'. For this probability distribution configuration, each power law is centred on the median luminosity of the sample, $\log_{10} x' = 42.1$ (and $\log_{10} x' = -2.55$ for sBHAR). The power laws are shown as dashed red lines in each panel and more clearly identify how the probability of finding an AGN, within the highlighted mass range, changes as a function of X-ray luminosity. The pale red regions surrounding each power law function give an indication of the error in each

fit. This was calculated by performing a χ^2 fit with equation (4.2) to the black data points in each bin and their associated errors. Fit parameter errors were estimated by taking the square-root of the covariance matrix's diagonal. With this I could outline the extent of the uncertainty in each fit. Encouragingly, in nearly all of the bins the power law fits are appropriate.

4.5.3 Probability Distribution Comparison

Using the method outlined in section 4.5.2, I created a number of other probability distributions to investigate how different configurations might shed light on the AGN population in the nearby Universe. As can be seen in figure 4.7, I have added redshift and sBHAR to my properties of interest. By including redshift in my range of host galaxy properties, we can gain some insight into how the changing conditions of the Universe might affect the AGN population. Each configuration has been split up into 6 colour-coded bins of the corresponding property – either redshift (top row) or stellar mass (bottom row) – and had a probability distribution calculated from the data points in that region. For ease of comparison, each set of distributions has been placed in the appropriate column - sBHAR on the left, and X-ray luminosity on the right. The results from figure 4.6 have been placed in the bottom-right panel of this figure. For clarity, I have only included the fits and their error region in each configuration panel. However the full set of data points for each configuration can be found in appendix C. Table D.1 outlines the best-fit coefficients, and associated errors, used in equation (4.2) to create these probability distributions. Overall, this figure shows me that regardless of the host galaxy property configuration, the average number of AGN in the nearby Universe are well described by a power law. However, there are key differences between these configurations which I will outline in this section.

Slope

In each panel of figure 4.7, there is some evidence of the slope changing with the respective property. Along the top row, we can see a series of slopes steadily steepening with redshift. This steepening effect is most pronounced in the right-hand X-ray luminosity column. However, we see in the top-left panel that AGN at higher redshifts favour more moderate accretion rates. This drives the increase in the incidence of relatively low luminosity AGN emission for the corresponding redshift bin in the top-right panel. A similar steepening can be seen in the bottom-right stellar mass, X-ray luminosity panel, however the trend is much less consistent. It shares the steepening seen in the redshift, X-ray luminosity panel but there is a lot

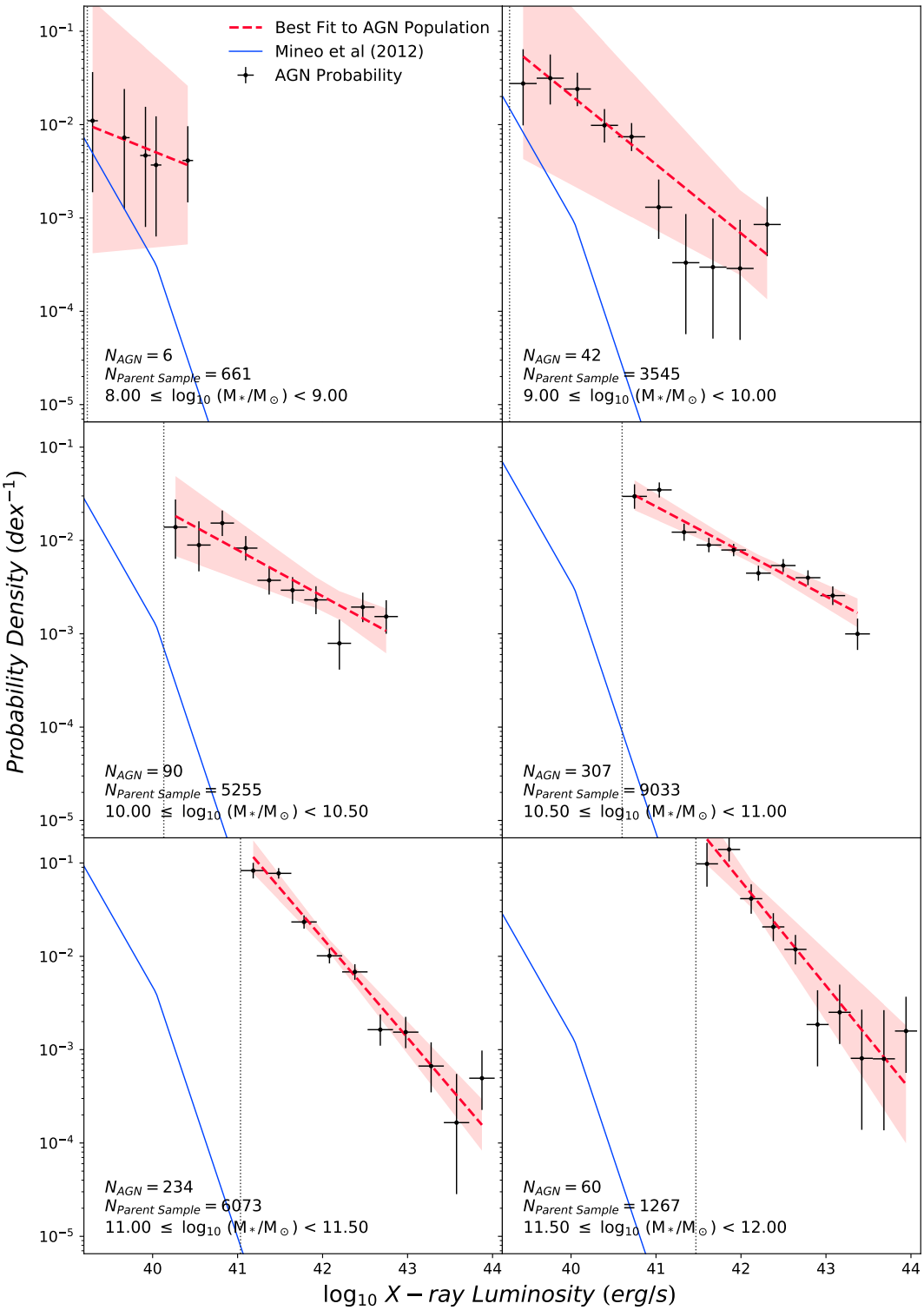

Figure 4.6: Probability of finding an AGN in the nearby Universe, using the completeness-corrected AGN samples, as a function of observed X-ray luminosity and stellar mass. Power laws (dashed red lines) have been fit to the data in each panel and displayed alongside their 1σ uncertainty (pale red region). Information about the number of PN-8-detected AGN and parent galaxies in the each mass range are printed on the panels. The blue line shows the average number of high luminosity stellar sources like ULXs. The form of these lines are based on the Mineo et al. (2012a) XLF, scaled by the average SFR of the galaxies in each bin. Dashed grey lines indicate the AGN detection limit in each bin. See section 4.5.2 for more details on how these plots were constructed.

of variation between these two extremes.

Normalisation

As outlined in section 4.4, studying sBHAR allows me to break observational degeneracies associated with X-ray luminosity so we can see how mass accretion is distributed across the AGN population. What is most striking about these distributions is that the distributions in the left-hand column of figure 4.7 all appear to be within the same limits. Regardless of redshift or stellar mass, there is little change in the normalisation, implying that the average amount of material a black hole accretes remains relatively consistent in the nearby Universe.

This trend is in stark contrast to that seen in the luminosity plots, in the right-hand column of figure 4.7. Both redshift and stellar mass distributions show consistently increasing normalisations. There are a number of processes that could be feeding into this shift: the magnitude limitation of the galaxy sample removing high redshift, low luminosity AGN; and the stellar mass and redshift dependencies of the Lehmer et al. (2016) XLF in equation (3.3). However, these effects only explain why the lower limits increase and don't account for full shifting of these luminosity-dependent distributions. This can be understood if we consider the lines of constant sBHAR in figure 4.5: for a fixed accretion rate, stellar mass determines the range of observable luminosities. At higher stellar masses the much more common, lower accretion rate sources have higher observed luminosities. When we consider that these mass and redshift bins have the same accretion limits, it is clear that changing stellar mass is the primary driver of this shift in normalisation, whether caused by explicit binning or the effect of magnitude limitation.

4.6 AGN Fractions

Using the fits produced in section 4.5.3, I can calculate how the AGN fraction varies with stellar mass and redshift in the nearby Universe. The results of these calculations are shown in figure 4.8. In this section I will outline this calculation process and the significance of the results.

AGN fractions were calculated by integrating under the probability distributions shown in figure 4.7. Each distribution produces a single point in the corresponding position, colour and panel of figure 4.8. Upper and lower limits on each AGN fraction were calculated by integrating both edges of the error region. The limits of these integrations are shown in the bottom-right-hand corner of the stellar mass

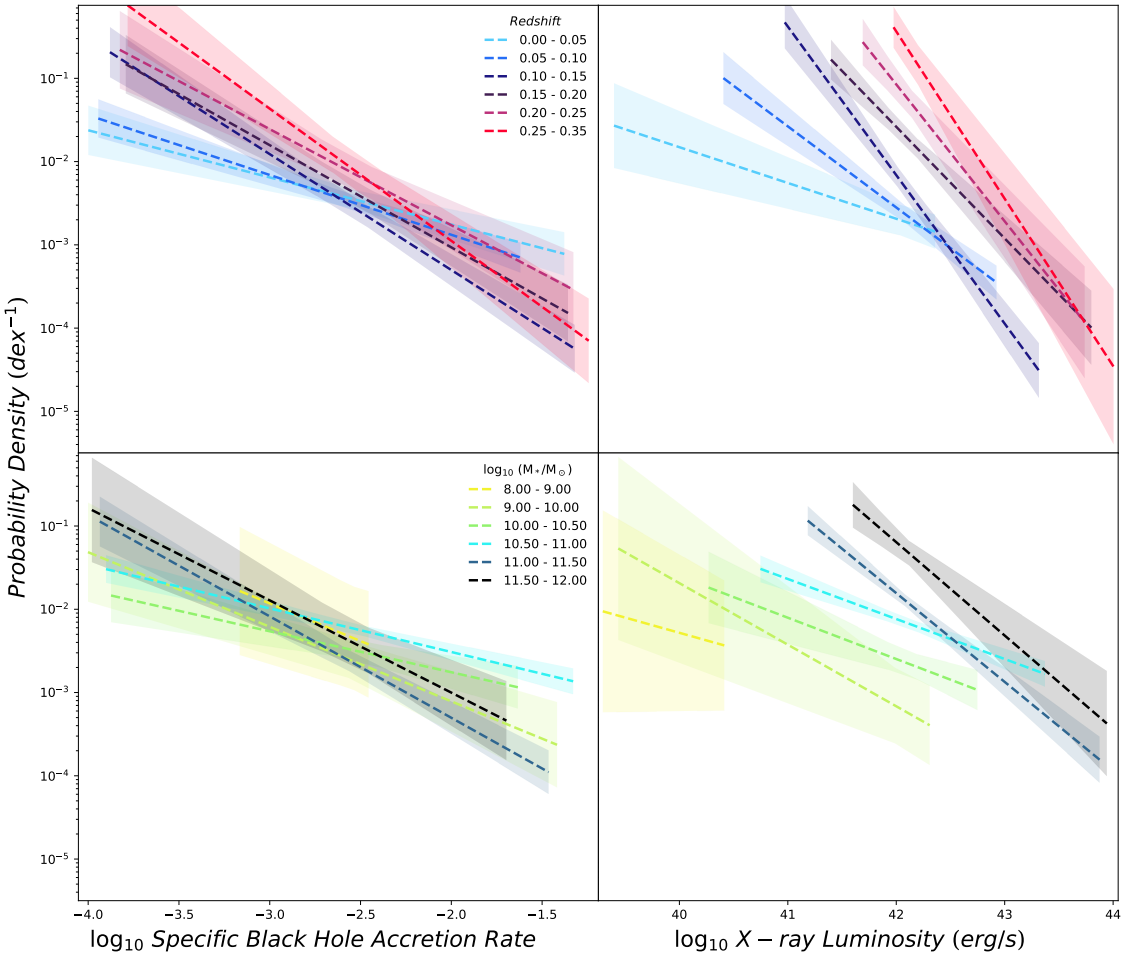

Figure 4.7: A comparison of the probability distributions and associated error regions calculated using different configurations of host galaxy properties. The left-hand column looks at how the probability varies with sBHAR and the right-hand column shows this with X-ray luminosity. The top row bins the AGN in redshift intervals, the bottom row does this in stellar mass. The fits for the bottom-right hand plot were taken from figure 4.6, the rest are presented in appendix C.

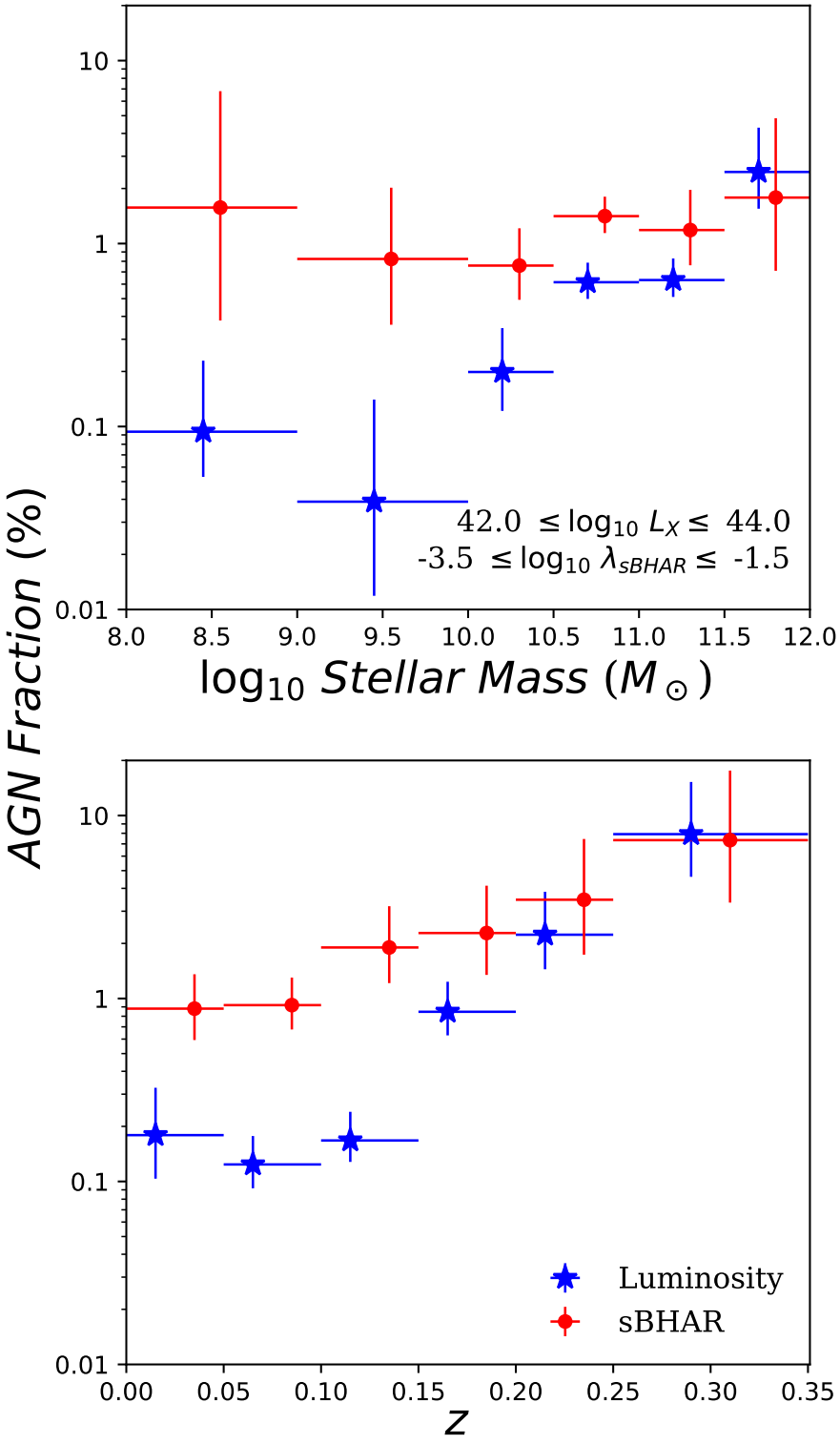

Figure 4.8: AGN fractions as a function of stellar mass (upper panel) and redshift (lower panel) calculated from the probability distributions in figure 4.7. Each probability distribution was integrated within the relevant limit shown in the bottom-right of the upper panel. There is a clear difference in AGN fraction depending on whether the luminosity (blue stars) or sBHAR (red circles) distributions were integrated.

panel. Consistent limits were chosen to encapsulate the region within which the majority of the distributions are defined and to make sure each point is directly comparable.

This approach leaves me with two sets of AGN fractions in each galaxy property panel: luminosity-derived fractions (blue stars) and sBHAR-derived fractions (red dots). In effect, I have created two separate AGN definitions to compare. The luminosity-derived points create an observational-style definition whereby an AGN is an object emitting at an X-ray luminosity $\geq 10^{42}$ erg s^{-1}. Whereas the sBHAR-derived points define an AGN as a black hole accreting at rates $\lambda_{\mathrm{sBHAR}} \geq 10^{-3.5}$. As is evident from figure 4.8, these differing definitions have implications for the predicted distribution of AGN.

In the upper panel, we can see how the two AGN fraction definitions vary with stellar mass. The sBHAR-derived fractions predict a constant occupation of just over 1% across the stellar mass range. Whereas the luminosity-derived fractions predict a much lower AGN fraction at low stellar mass before rising to agree with sBHAR.
In the lower panel, we can see how both AGN fraction definitions vary with redshift. The sBHAR-derived fractions show a steady increase rising from 1% to 10%. The luminosity-derived fractions describe an even steeper increase from about 0.1% to 10%, agreeing with the sBHAR-derived data at the highest redshifts.

Clearly, these two AGN fraction definitions follow different trends. Since the luminosity-derived fractions only consider higher-luminosity AGN, the sample used to calculate these will be restricted and so a drop in AGN fraction is likely. As discussed previously, all the AGN in my sample appear to be accreting over a similar range of sBHAR, so stellar mass is the main driver of changes in observed X-ray luminosity. Therefore, we would expect that lower mass AGN, which are much less likely to meet such high luminosities, are the objects most likely being excluded. In addition, figure 4.3 shows that most of the lowest mass AGN are found at the lowest redshifts. So this would also concentrate this deficit to the lowest redshift regimes. To pinpoint where this deficit occurs, I first split the AGN by stellar mass - those above and below the median stellar mass of $10^{10.89}$ M$_\odot$ - and recalculated the AGN fractions as a function of redshift. The top row of figure 4.9 confirms my expectations by showing that this luminosity-derived fraction deficit very clearly occurs in the lowest stellar mass and redshift galaxies. The right-hand, high-mass panel shows that when both limits encapsulate the majority of the AGN sample, they predict very similar trends. We can see that the luminosity-derived fractions are still subject

to observational biases despite being calculated from completeness-corrected data. Other studies, with different samples, have observed this effect as well e.g. Aird et al. (2012), Bongiorno et al. (2016), and Weigel et al. (2017).

Whilst the sBHAR-derived points are not free from these biases, selecting AGN based on an Eddington ratio limit has been shown to yield a wider range of AGN and host galaxy properties compared to a luminosity limit and so better represents the underlying AGN population (Jones et al. 2017). It is for these reasons that I focus on the sBHAR-derived points when drawing conclusions about the underlying AGN population in the nearby Universe. The upper panel of figure 4.8 showed us that the sBHAR-derived AGN fraction was constant across the stellar mass range. Furthermore, the bottom row of figure 4.9 shows that when I split the AGN by redshift - those above and below the median redshift of 0.11 - and recalculate the AGN fractions as a function of stellar mass, that this flat distribution remains. Thus, the sBHAR-derived AGN fraction does not depend on stellar mass in the nearby universe.

In figure 4.8, we saw that the sBHAR-derived AGN fraction increased from 1% to 10% with increasing redshift. We see similar behaviour in both panels of the top row of figure 4.9, where the sample is split by stellar mass. Furthermore, when I compared the low and high redshift bins in the bottom row, we see evidence of a systematic increase in AGN fraction for both the sBHAR- and luminosity-derived values. My results show that there is an intrinsic increase in the sBHAR-derived AGN fraction with redshifts out to $z \sim 0.35$.

Figure 4.10 compares my sBHAR-derived AGN fractions with other predictions derived from Eddington ratio-based probability distributions: Aird et al. (2018), Aird et al. (2012), Birchall et al. (2020), Bongiorno et al. (2012), Wang et al. (2017), and Yang et al. (2018). I have increased the limits of integration compared to figure 4.8 to incorporate the generally higher threshold of AGN activity used in these other studies. By shifting the lower limit from $\log_{10} \lambda_{\mathrm{sBHAR}} = -3.5$ to -3, I have halved the average AGN fraction predicted from my data to 0.5%. This shows there is a significant amount of low level accretion in the nearby Universe.

My results provide one of the first, robust measurements of the incidence of AGN within the nearby galaxy population. To highlight this I presented the median redshift of each paper's sample in the legend. Overall we see that as redshift increases so too does the AGN fraction. It is thought that the increased availability of cold gas at earlier times (Popping et al. 2012; Vito et al. 2014) would drive this increase

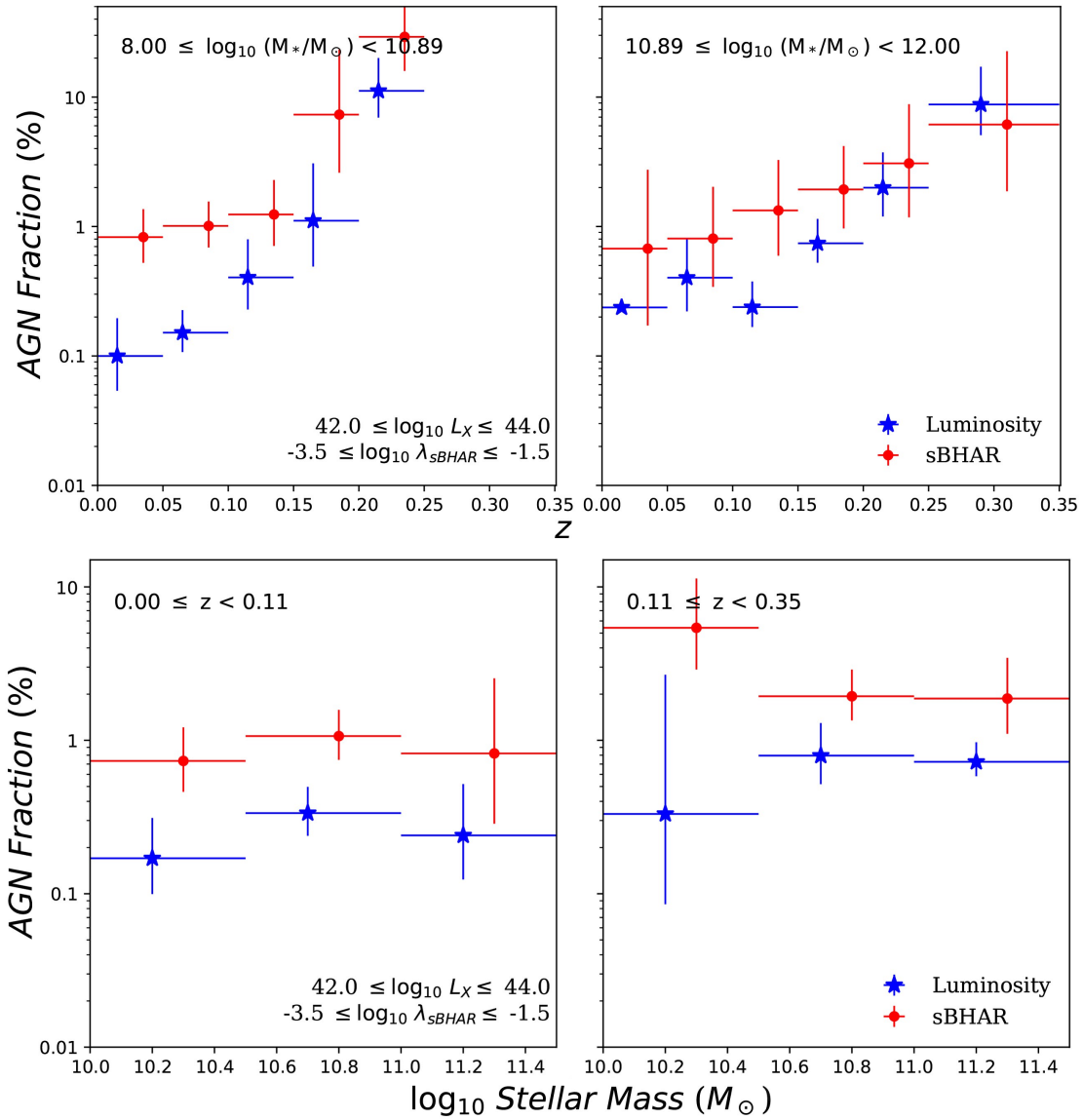

Figure 4.9: The top row depicts AGN fractions as a function of redshift split up into stellar mass bins - those above and below the median mass. Below are the AGN fractions as a function of stellar mass split up into redshift bins - those above and below the median redshift. The limits between which the full AGN sample has been split are shown in the top-left of each panel. As with figure 4.8, the luminosity-derived fraction are blue stars and the sBHAR-derived fractions are red circles.

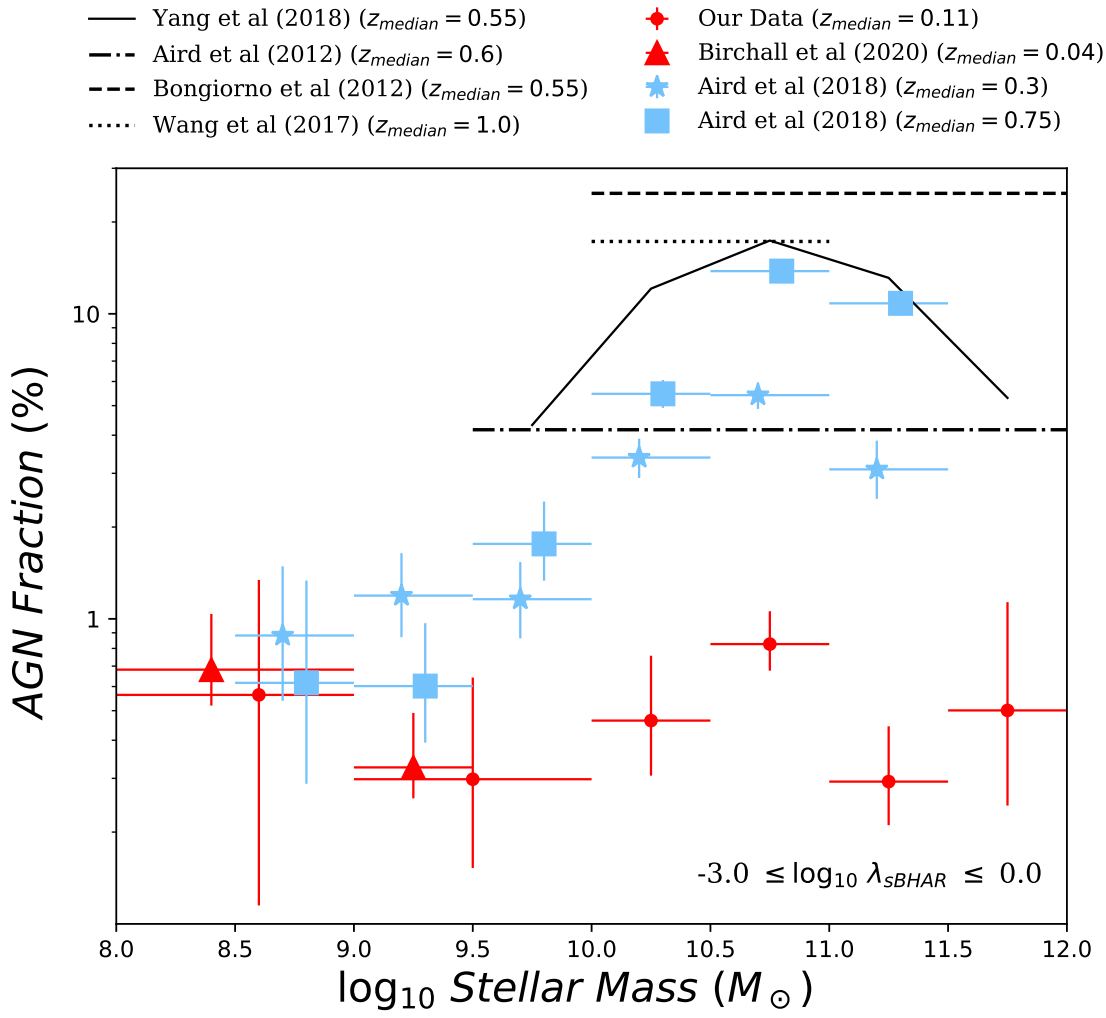

Figure 4.10: A comparison of how AGN fraction changes as a function of stellar mass at different redshifts. My data is calculated from the sBHAR-dependent fits. Alongside this are numerous other AGN fractions calculated from Eddington ratio based probability distributions within the limits shown in the bottom-right of the plot. The median sample redshift is highlighted in the legend. My results agree with other studies that probe the lowest stellar masses. However, it is clear there is strong redshift evolution of the AGN fraction at higher stellar masses.

in activity. Other AGN samples taken from a range of redshifts and wavelengths appear to show similar patterns of activity (Aird et al. 2018; Delvecchio et al. 2018; Yang et al. 2018). To check whether my results are consistent with this evolution, I re-integrated equation (12) of Aird et al. (2012) between the same sBHAR limits shown in figure 4.10. However, this time I modified the redshift-dependent normalisation term by evaluating it at $z = 0.11$, the median redshift of my work. This produced a new fraction of 1.1%, consistent with most of my fractions.

However, figure 4.10 shows that this evolution is suppressed in lower mass galaxies. We can see significant alignment between points accounting for AGN samples measured out to $z = 1$ in the dwarf galaxy mass range ($\log_{10} M_*/M_\odot \leq 9.5$); consisting of my results, and those from Birchall et al. (2020) and Aird et al. (2018). It is thought that since the potential wells for lower mass galaxies are shallower, gas is less likely to fall into the galactic centre and fuel the SMBH (Bellovary et al. 2013). A similar drop in AGN activity at lower stellar masses has also been observed in different AGN samples (Aird et al. 2018; Yang et al. 2017).

4.7 Summary & Conclusions

In this chapter I have rigorously matched a sample of local galaxies from MPA-JHU to their X-ray counterparts in 3XMM using `xmatch`. Before identifying AGN, I confirmed that the nuclear emission from my matched sample did not affect the stellar mass and star-formation rate measurements. With this confirmation I could use these quantities to confidently predict the X-ray luminosity due to emission from X-ray binaries and hot gas. Any galaxies that had observed X-ray emission at least three times larger than the sum of these predictions was considered an AGN host. Finally, any AGN that fell outside of the 90% mass completeness function were removed from the sample. This left me with 917 AGN.

I then performed BPT analysis on the 658 AGN with significant detections in all relevant emission lines. I have demonstrated that there is a strong stellar mass dependence in the BPT diagnostic. Higher mass X-ray selected AGN are much more likely to be assigned the same classification by the BPT diagnostic. In contrast, I found little dependence in the BPT-selected fraction of AGN on X-ray luminosity. Since the quenched fraction of galaxies is believed to increase with stellar mass, then the influence of star-formation on these emission lines would drop significantly and produce the increase in BPT accuracy.

Next I investigated the activity of the black holes powering my AGN by calculating

their specific accretion rates. I found that there is a clear preference for black holes with relatively low accretion rates. Most of them have accretion rates that are less than 0.5% of their Eddington luminosity (assuming a nominal scaling between black hole mass and galaxy stellar mass), with only a handful of the most massive galaxies venturing above 10%. Only two AGN have an accretion rate that is greater than 50% of their Eddington luminosity.

I then corrected for observational bias in my AGN sample. Drawing on the method used in chapter 3, I used Flix to determine the flux limits of the 25,949 MPA-JHU galaxies found within 3XMM - my parent sample. These upper limits form the basis from which a sensitivity function can be created and used to correct my observed distributions of AGN. From this I created a series of probability distributions splitting the AGN population by stellar mass and redshift, and looking at trends with observed X-ray luminosity and specific black hole accretion rate. These distributions are well-described by power laws but when compared show distinct trends. As redshift increases, the power laws steepen suggesting more moderate accretion rates and thus luminosities, are favoured at earlier times. A similar steepening trend can be seen with stellar mass, however, the progression is less certain. I also find my galaxies contain AGN accreting over approximately the same, broad range of specific accretion rates ($-3.5 \lesssim \log \lambda_{\mathrm{sBHAR}} \lesssim -1.5$) regardless of stellar mass or redshift.

Finally, integrating under each of these distributions allows me to calculate robust AGN fractions and determine how these vary with host galaxy stellar mass and redshift. Since I have investigated the AGN distribution as a function of both X-ray luminosity and specific black hole accretion rate, I adopt two definitions of an AGN when determining the fraction. The first is pseudo-observational, defining an AGN an object with X-ray luminosity $\geq 10^{42}$ erg s^{-1}; the second defines an AGN as a black hole accreting at rates $\lambda_{\mathrm{sBHAR}} \geq 10^{-3.5}$. I find that the pseudo-observational luminosity-limited definition reproduces observational biases inherent in previous AGN studies so I focus on the accretion rate derived fractions. I find no evidence that my accretion-rate-derived AGN fractions depend on the stellar mass, finding a constant fraction of just over 1% for $8 \lesssim \log_{10}(\mathrm{M}_*/\mathrm{M}_\odot) \lesssim 12$. In addition, I find that AGN fraction increases with redshift, rising from 1% to 10% at $z \sim 0.35$.
I then derived AGN fractions from sBHAR distributions produced by other studies and compared them to my results. My low redshift results help confirm the previously identified, strong redshift evolution of the AGN fraction in galaxies with $M_* \gtrsim 10^{10}$ M$_\odot$ brought about by an increase in the availability of cold gas. In contrast, my results confirm a lack of evolution at low stellar masses, indicating a constant AGN fraction out to $z = 1$ (see also Birchall et al. 2020) The suppression

in this region is thought to be driven by the shallower potential wells in lower mass galaxies that are unable to capture as much gas as their higher mass counterparts.

Chapter 5

The Relationship between AGN & Star-forming Activity

5.1 Introduction

Numerous studies have linked star-forming activity with an increase in the incidence of AGN (e.g. Aird et al. 2017; Aird et al. 2012; Rosario et al. 2013; Wang et al. 2017) through the common fuelling mechanism of cold gas accretion. One of the most compelling pieces of evidence supporting this relationship is the strong correlation between the SFR and BHAR density distributions. Both increase from $z = 6$, peak at around $z = 1-3$, and then decline to about a tenth of their peak in the present-day Universe (e.g. Aird et al. 2015; Boyle and Terlevich 1998; Delvecchio et al. 2014; Madau and Dickinson 2014). Concurrently there is a significant change in the star-forming composition of the overall galaxy population. Since $z = 2$, there has been a significant build up of the quiescent galaxy population, particularly in higher mass galaxies (Barro et al. 2017; Brammer et al. 2011; Tomczak et al. 2014), which highlights a significant amount of star-formation quenching in this period. This decline in the density of star-forming activity across recent cosmic time is thought to be driven by a decreasing density of molecular gas (Maeda et al. 2017; Popping et al. 2012).

An observed excess of AGN in star-forming galaxies would imply that quiescent hosts are less effective at fuelling their central black holes. Kauffmann and Heckman (2009) explained this disparity in activity through their "feast and famine" fuelling model. By analysing the Eddington ratio distributions of a sample of optically-selected SDSS AGN they were able to demonstrate the existence of two distinct populations of AGN implying there were different regimes of black holes growth. The "feast" mode is associated with galaxies containing significant amounts of star

formation in their central regions. The large amounts of cold gas required for this star-forming activity fuels black hole growth. The "famine" mode is associated with galaxies hosting older stellar populations. In this case, black hole growth is regulated by the rate at which stars lose their mass.

Whilst the optical selection method used in Kauffmann and Heckman (2009) produced an incomplete sample (Jones et al. 2016), this dual fuelling model has been observed in more recent studies with more complete samples. For example, Aird et al. (2019) explores the effect of star-forming activity on a sample of AGN from the CANDELS survey ($0.1 \leq z \leq 4$), selected using Chandra X-ray data. After applying observational corrections to this sample, they also find evidence of an SFR-dependent fuelling mechanism, reflective of the model proposed in Kauffmann and Heckman (2009). Given the incomplete nature of the Kauffmann and Heckman (2009) sample, performing a detailed analysis of the star-forming properties of AGN activity in this redshift regime would be invaluable in confirming the nature of this relationship.

In this chapter, I will explore how the star-forming properties of a host galaxy affect the likelihood of finding a local ($z \leq 0.35$) X-ray selected AGN. This analysis makes use of the AGN sample constructed in the previous chapter, see section 4.2 for more information about how these AGN were identified.

5.2 Star-forming Classification

The galaxy population in the nearby Universe is strongly bimodal: there are star-forming galaxies that lie on the main sequence (e.g. Elbaz et al. 2011; Noeske et al. 2007) and quiescent galaxies that lie below. However, to accurately split these two populations I cannot just use absolute SFRs. As previously discussed, the level of star formation has changed dramatically through recent cosmic time. In addition, Salim et al. (2007) found that for a large sample of star-forming SDSS galaxies, stellar mass and SFR are related by a power law, $\mathrm{SFR} \propto \mathrm{Mass}^{0.65}$. To isolate any effect that star-forming activity may have on the incidence of AGN in the nearby Universe, I need to account for the effect that both stellar mass and redshift have on SFR.

5.2.1 Star-forming or Quiescent

My method for splitting the sample into different star-forming classes is based on Moustakas et al. (2013). They calculate a quantity referred to as the "rotated SFR", SFR_{rot}, which attempts to account for the stellar mass effect on SFR. Thus SFR_{rot} has the form,

$$\log_{10}(SFR_{rot}) = \log_{10}(SFR) - 0.65(\log_{10} Mass - 10) \tag{5.1}$$

where SFR is in units of $M_\odot$ yr^{-1} and mass is in units of $M_\odot$. I plotted histograms of SFR_{rot} binned by redshift to produce distributions of this mass-independent SFR for the underlying galaxy population. This then allowed me to identify the SFR_{rot} corresponding to the local minimum between the star-forming and quiescent peaks in each redshift bin. Fitting a straight line to the change of SFR_{rot} minima with redshift produced an appropriately normalised equation which could be used to split the whole galaxy population into star-forming and quiescent, $SFR_{SF/Q}$. It took the form,

$$\log_{10}(SFR_{SF/Q}) = 0.65(\log_{10} Mass - 10) + \frac{z - 0.75}{0.79} \tag{5.2}$$

Figure 5.1 shows the results of this analysis. My AGN sample is plotted as large, coloured points on the mass and SFR plane and contrasted with the underlying MPA-JHU galaxy population. Each panel contains a green line, described by equation (5.2), which is used to split these AGN into star-forming (blue stars) and quiescent (red circles). Any objects that lie above the line are classified as star-forming, those below it are quiescent.

5.2.2 SFR relative to the Main Sequence

Dividing the sample into star-forming and quiescent galaxies is useful to compare the general effect of star-formation on AGN activity. However, to understand this effect in greater detail I extended the above analysis to further divide the galaxy sample by its changing level of star-formation. For this quantity I shifted equation (5.2) up to the galactic main sequence of star formation at a given redshift. It has the form,

$$\log_{10}(SFR_{MS}) = 0.65(\log_{10} Mass - 10) + \frac{z - 0.05}{0.79} \tag{5.3}$$

With this equation I established 5 bins of $\log_{10}(SFR/SFR_{MS})$ to track the changing level of star-formation in the sample:

- Starburst: Star-forming galaxies with excess star-formation relative to the

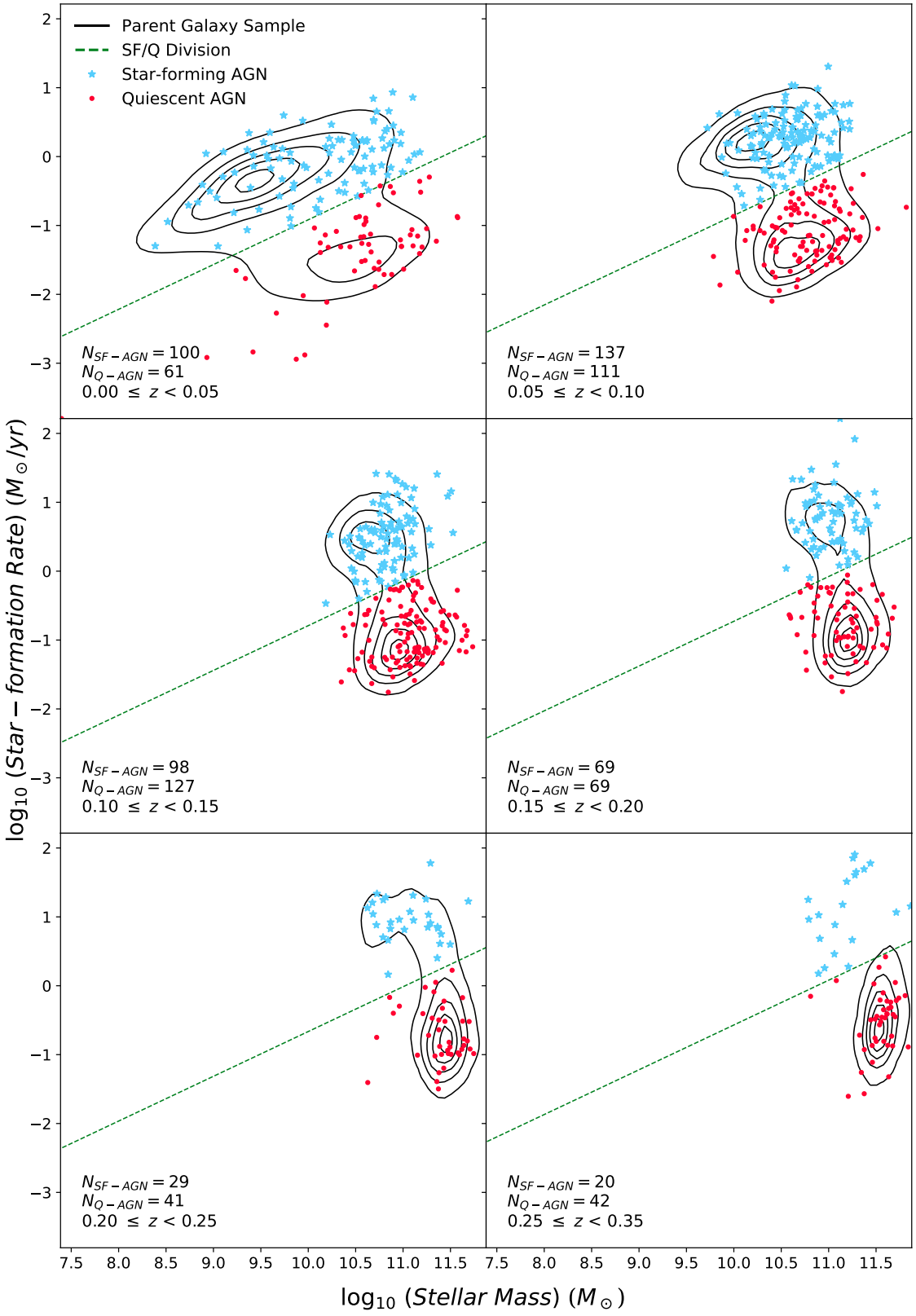

Figure 5.1: The AGN sample is contrasted with the underlying galaxy population (black contours, containing 90%, 70%, 50%, 30%, and 10% of the population). The green dashed line describes the definition of the main sequence of star-formation for this galaxy population and is used to determine whether an AGN is "star-forming" (blue, stars) or "quiescent" (red, circular points). See section 5.2.1 for more information on how this was calculated.

main sequence ($\log_{10}(\mathrm{SFR}/\mathrm{SFR_{MS}}) > 0.3$)

- Main Sequence: Star-forming galaxies with SFRs consistent with the main sequence ($-0.2 \leq \log_{10}(\mathrm{SFR}/\mathrm{SFR_{MS}}) \leq 0.3$) corresponding to 50% of the total star-forming galaxy population
- Sub-Main Sequence: Galaxies with SFRs lower than the bulk of the main sequence ($-1.3 \leq \log_{10}(\mathrm{SFR}/\mathrm{SFR_{MS}}) < -0.2$) consisting of weak star-forming galaxies and relatively strong quiescent galaxies.
- Quiescent: Quiescent galaxies with moderate SFRs ($-1.8 \leq \log_{10}(\mathrm{SFR}/\mathrm{SFR_{MS}}) < -1.3$)
- Quiescent (Low): Quiescent galaxies with SFRs in weakest 50% of that population ($\log_{10}(\mathrm{SFR}/\mathrm{SFR_{MS}}) < -1.8$)

Figure 5.2 shows how each bin maps onto the AGN sample. By taking this approach I have ensured there are sufficient numbers of observed AGN in each bin and that, for a fixed redshift, an increase in the $\log_{10}(\mathrm{SFR}/\mathrm{SFR_{MS}})$ tracks only the effects of SFR.

5.3 BPT Classification

In chapters 3 and 4, I recreated a known fault with the BPT diagram whereby star-formation signatures dilute the strength of AGN indicators, causing typically lower mass AGN to be missed by this diagnostic (see also Agostino and Salim 2019; Cann et al. 2019). Having established techniques to separate my AGN sample by star-forming activity, in this section I will outline how the star-forming classification maps onto the BPT diagnostic.

Of the total sample of 917 X-ray selected AGN, 658 have significant detections ($\frac{\text{Line Flux}}{\text{Line Flux Error}} > 3$) in all four emission lines. These AGN are shown on the BPT diagnostic in figure 5.3. Blue, star-forming AGN can clearly be seen throughout the BPT diagnostic, whereas red, quiescent AGN are only found in the AGN region. Given quiescent AGN are defined by their lack of star-forming activity, this concentration in the AGN region is expected as there are no significant stellar emission lines to dilute the optical AGN indicators. However, over half of the X-ray selected quiescent AGN sample are missing from the BPT diagram, compared with only a handful of star-forming AGN. These objects do not appear because they lack sufficiently significant detections in all the emission lines, with the most frequently weak emission line being Hβ. Cid Fernandes et al. (2010) also found a significant

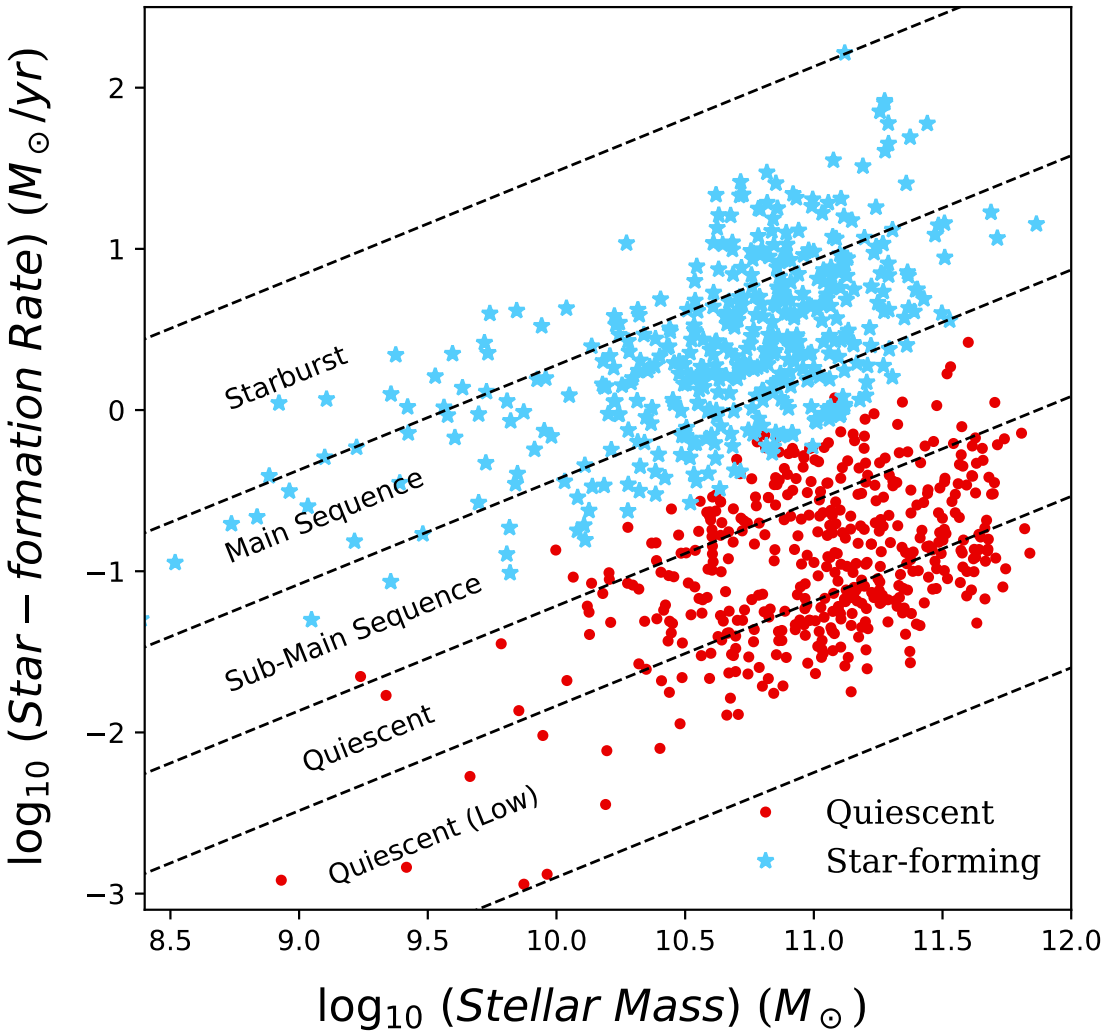

Figure 5.2: The distribution of stellar masses and SFRs for my AGN sample. Overlaid are black, dashed lines describing bins of $\log_{10}(\mathrm{SFR}/\mathrm{SFR}_{\mathrm{MS}})$ calculated based on the underlying galaxy population. Star-forming AGN are shown as blue stars, and quiescent AGN are red circles. Section 5.2.2 provides an explanation of how these limits were chosen. This figure was evaluated at the median sample redshift, $z = 0.11$.

population of SDSS galaxies were missed out from the BPT diagram due to lack of significant Hβ line detections. This predominantly affected galaxies that would have been classified as AGN, should the significance criterion not have been required.

To investigate why a large proportion of quiescent AGN are missing, I looked at how the fraction of BPT-detected AGN changes with stellar mass and star-forming classification. The results are inset in figure 5.3. The blue star-forming distribution show high proportions of BPT-detected AGN, reaching 95% in the $10.5 \leq \log_{10}\ (M_*/M_\odot) \leq 11$ bin and never dropping below 67%. However, the red quiescent distribution peaks at 76% in the $10 \leq \log_{10}\ (M_*/M_\odot) \leq 11$, before rapidly dropping to 17% in the highest mass bin. There was only one quiescent AGN detected in the lowest mass bin which was missed, hence the 0% detection rate.

By definition, quiescent AGN are not forming as many new stars as their star-forming counterparts so stellar absorption lines will dominate their spectra. These absorption lines will be most pronounced in the spectra of the highest mass galaxies as there is a larger amount of stellar material. Whilst low level AGN activity in high mass hosts can be clearly detected by robust X-ray detection, prominent stellar absorption lines could hide weak optical emission lines in the galaxy's spectrum. This would produce the significant drop in the fraction of high mass, quiescent AGN appearing on the BPT diagram seen in figure 5.3. This is further evidence to suggest that the BPT diagram is not effective at identifying weakly accreting AGN.

5.4 Completeness-corrected Probability Distributions

In this section, I build on the probability distribution analysis performed in section 4.5 by using the star-forming classifications to highlight how they affect the AGN population in the nearby Universe.

Figure 5.4 shows sBHAR probability distributions for each property of interest, split up into quiescent (right-hand column) and star-forming (left-hand column) AGN. Each panel contains numerous probability distributions coloured to indicate the respective stellar mass (top row), redshift (middle row) or $\log_{10}(\mathrm{SFR}/\mathrm{SFR}_{\mathrm{MS}})$ (bottom row) bin. Based on the AGN fraction analysis in section 4.6 I only consider sBHAR-based probability distributions. See section 4.4 for more information about sBHAR.

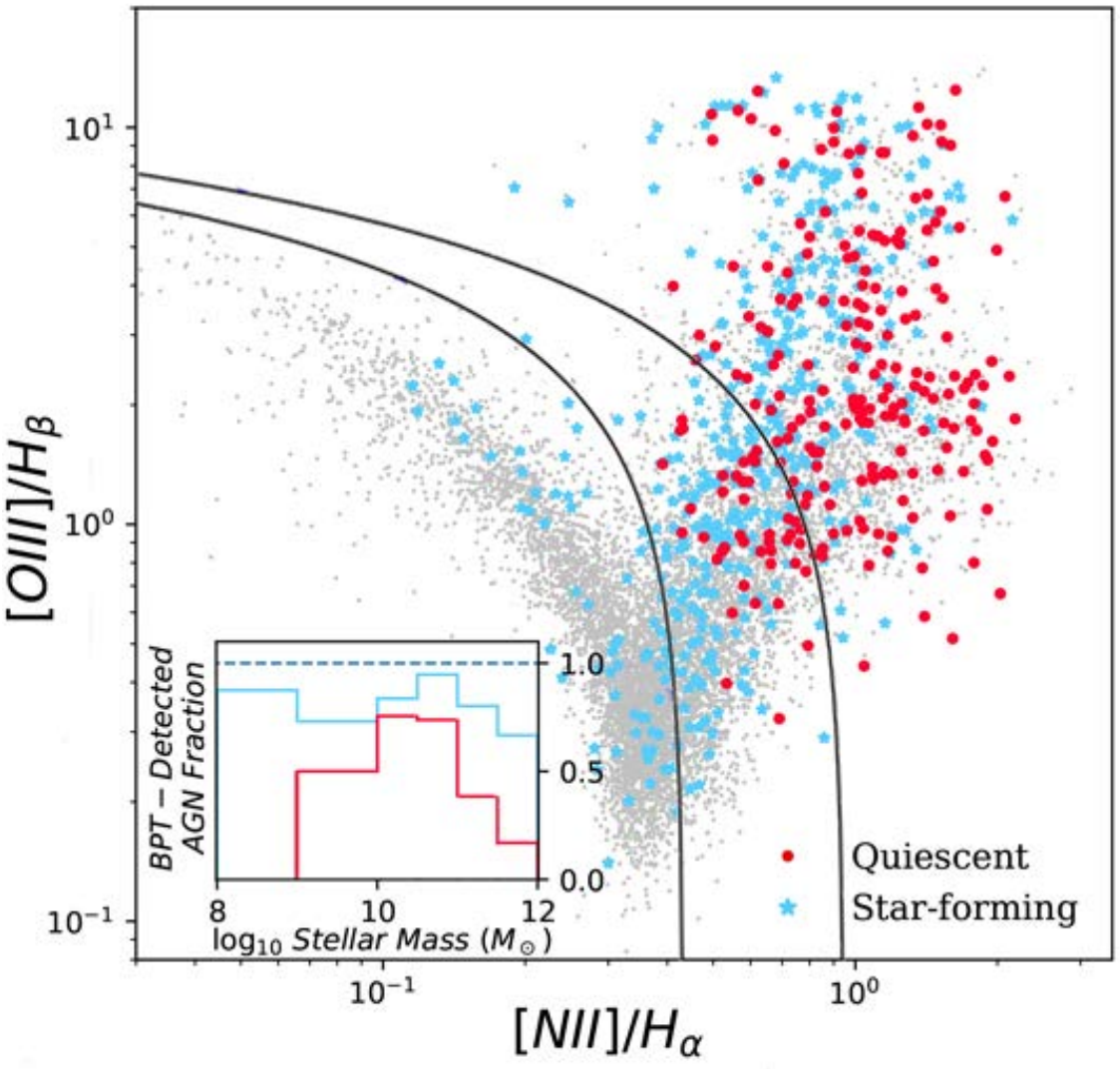

Figure 5.3: BPT diagram for the 658 AGN with significant detections in all four emission lines overlaid onto the underlying galaxy population in grey. Star-forming AGN are shown as blue stars, and quiescent AGN are red circles. The inset figure shows the fraction of AGN that have sufficiently strong emission lines to be detected by the BPT diagnostic, split up by star-forming category and as a function of stellar mass.

5.4.1 Normalisation

Only considering sBHAR-dependent distributions results in few changes in normalisation. This is consistent with the distributions in section 4.5. The only significant shifts occur in the extreme ends of the stellar mass bins. The lowest mass, quiescent AGN and highest mass, star-forming AGN both appear restricted when compared to the other bins. This is due to the limited number of observations in these bins. There is no indication of any significant enhancement or deficit relative to the rest of the distributions. However, it does highlight the changing mix of classifications with stellar mass: star-forming galaxies dominate at lower stellar masses and quiescent at higher masses.

5.4.2 Slope

The clearest changes in slope can be seen in the redshift panels of figure 5.4. The quiescent distributions are much steeper than those composed of star-forming AGN. There is also a distinct steepening between the quiescent distributions, showing that lower accretion rates are favoured with increasing redshift. The star-forming distributions, however, do not show this trend. In fact, the highest redshift star-forming distributions appear significantly flatter than their quiescent counterparts, suggesting that star-forming activity facilitates higher accretion rates. This is a clear example of star-forming activity having a direct effect on the distribution of AGN in the nearby Universe.

There is also evidence for some change in slope in the $\log_{10}(\mathrm{SFR}/\mathrm{SFR}_{\mathrm{MS}})$ panels. Given the nature of this quantity the quiescent/star-forming division occurs along the $\log_{10}(\mathrm{SFR}/\mathrm{SFR}_{\mathrm{MS}})$ axis, there are only three bins in each panel. However, it is clear that there are three distinct groups: the bottom two quiescent, the middle three main-sequence-adjacent and the highest $\log_{10}(\mathrm{SFR}/\mathrm{SFR}_{\mathrm{MS}})$ distributions share similar forms. Increasing $\log_{10}(\mathrm{SFR}/\mathrm{SFR}_{\mathrm{MS}})$ appears to bring a slightly flatter distribution further highlighting that star-forming activity fuels higher rates of accretion, similar to the effect seen in redshift.

There does not appear to be any consistent trend within the stellar mass panels, nor any clear effect when comparing the quiescent and star-forming populations. Whilst the highest mass, star-forming bin has a distinctly flat gradient, it is poorly constrained and is consistent with the lower mass bin within the errors.

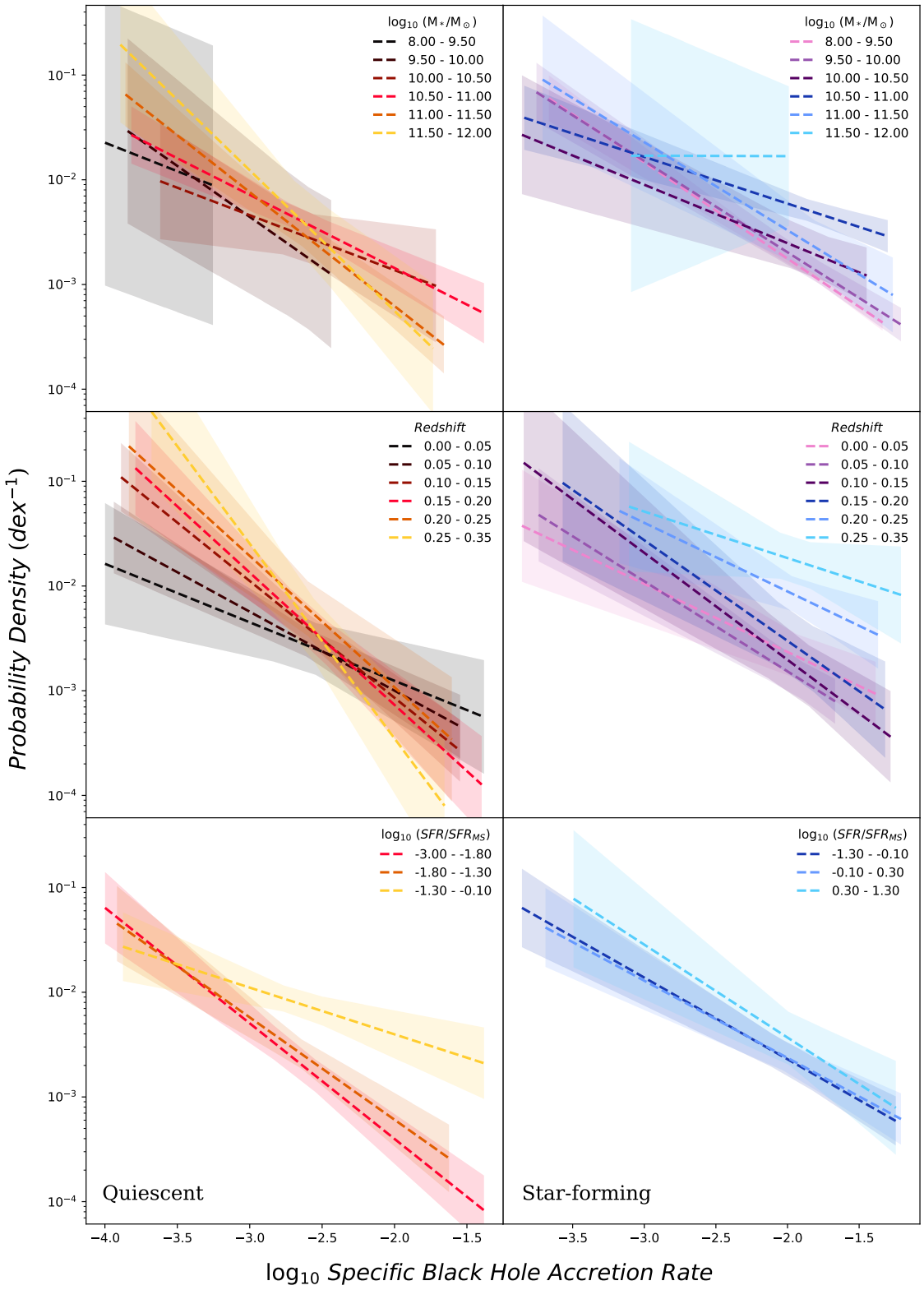

Figure 5.4: Comparison of the full range of sBHAR-dependent probability distributions. The left-hand column plots the distributions of quiescent AGN, the right-hand column plots the distributions of the star-forming AGN. The distributions are binned as a function of stellar mass (top row), redshift (middle row) and $\log_{10}(\mathrm{SFR}/\mathrm{SFR}_{\mathrm{MS}})$ (bottom row).

5.5 AGN Fractions

Integrating under the probability distributions shown in figure 5.4, allowed me to further analyse how star-forming activity affects the AGN population in the nearby Universe. In this section, I will outline the results of this calculation and explain their significance.

Figure 5.5 shows the AGN fraction as function of each property, split up into star-forming and quiescent populations. As before I am only considering the sBHAR-derived fractions; the corresponding integration limits are shown in the bottom-left corner of each panel. It is encouraging to see that the top and middle panels, exploring the AGN fraction with stellar mass and redshift respectively, highlight similar trends shown in section 4.6. Previously I found little change in AGN fraction with stellar mass, averaging around 1%. In the top panel of figure 5.5 there is a similarly flat AGN fraction with stellar mass, averaging about 1% for the quiescent galaxies, and 2% for star-forming galaxies. Section 4.6 also shows that I found AGN fraction with redshift increased from around 1% to 10%. In the middle panel of figure 5.5 there is also a clear increase in AGN fraction with redshift. Between $z = 0$ and 0.35, AGN fraction rises from 0.5% to 4.5% for quiescent galaxies, and from 1.5% to 7% for star-forming objects.

Splitting the AGN sample by star-forming classification shows that star-forming galaxies have slightly enhanced AGN fractions. However, this enhancement does not appear statistically significant. To check its significance, I calculated the overall fraction of AGN found in star-forming galaxies and compared it to the fraction in quiescent galaxies. The star-forming AGN fraction was found to be enhanced by a factor of 2 at a $> 3.5\sigma$ significance. Thus there does appear to be a real increase in the incidence of AGN in star-forming galaxies. Azadi et al. (2015) observed a similarly sized star-forming-driven enhancement in a different X-ray selected AGN sample out to $z \approx 1.2$.

This enhancement of AGN fraction in star-forming galaxies is also reflected in the bottom panel of figure 5.5. The AGN fraction rises from 0.6% to 3.6% with increasing $\log_{10}(\mathrm{SFR}/\mathrm{SFR}_{\mathrm{MS}})$. However, this increase is not linear. As with the probability distributions in figure 5.4, the fractions appear to be grouped into similar levels of star-forming activity: groups dominated by quiescent galaxies, groups centred around the main sequence, and starburst galaxies. Each group was observed to have a slightly higher AGN fraction than the previous. A similar positive correlation has been observed between average black hole accretion rate and SFR in

samples at higher redshifts (e.g. Chen et al. 2013; Yang et al. 2017).

Both redshift and SFR/SFR_{MS} appear to drive similarly strong increases in AGN fraction. To disentangle these effects, I divided the galaxy and AGN populations and re-plotted the AGN fractions as a function of $\log_{10}(SFR/SFR_{MS})$ for these new samples. It is clear from figure 5.6, that AGN fraction increases with SFR throughout the sample. Whether split at the median mass ($\log_{10}(M_*/M_\odot) = 10.89$) in the top row or redshift ($z = 0.11$) in the bottom row, both trends show a systematic increase in fraction and a steepening gradient between the low and high value bins.

The systematic increase between the low and high redshift bins is expected given the observed trend in figure 5.5 but for this same reason the systematic increase between lower and higher mass galaxies is not. By splitting the sample into lower and higher mass galaxies, however, there would also be a change to the average sample redshift. So this systematic increase in AGN fraction between low and high mass galaxies is also likely due to redshift effects. To confirm this, I consolidated each set of fractions from the low and high mass bins into quiescent and star-forming classifications at each panel's median redshift (0.08 for lower mass galaxies, 0.15 for higher mass galaxies). I found that the increase between these consolidated fractions was consistent with the overall AGN fraction increase with redshift.
The steepening gradient between low and high value bins will largely be due to changing combinations of star-forming and quiescent hosts. As I outlined previously, SFR is known to increase with both stellar mass and redshift. Thus as the SFR increases in the higher value bins, the proportion of AGN in star-forming galaxies will increase and those in quiescent galaxies will decrease, producing a steeper increase.

5.6 AGN Fractions in Different Accretion Regimes

The AGN fraction definition I used to create figure 5.5 covers a significant range of accretion rates. With this definition I have highlighted that star-forming activity could enhance the AGN fraction by a factor of 2. However, the redshift row of figure 5.4 shows that quiescent distributions at higher redshifts favour lower accretion rates, whereas star-forming galaxies favour higher accretion rates. So by integrating across such a broad range of accretion rates I could be obscuring the real role star formation plays in fuelling AGN. In this section, I change my definition of AGN fraction: integrating under smaller areas of the probability distributions so I can further interrogate how star-forming activity affects the incidence of AGN fraction

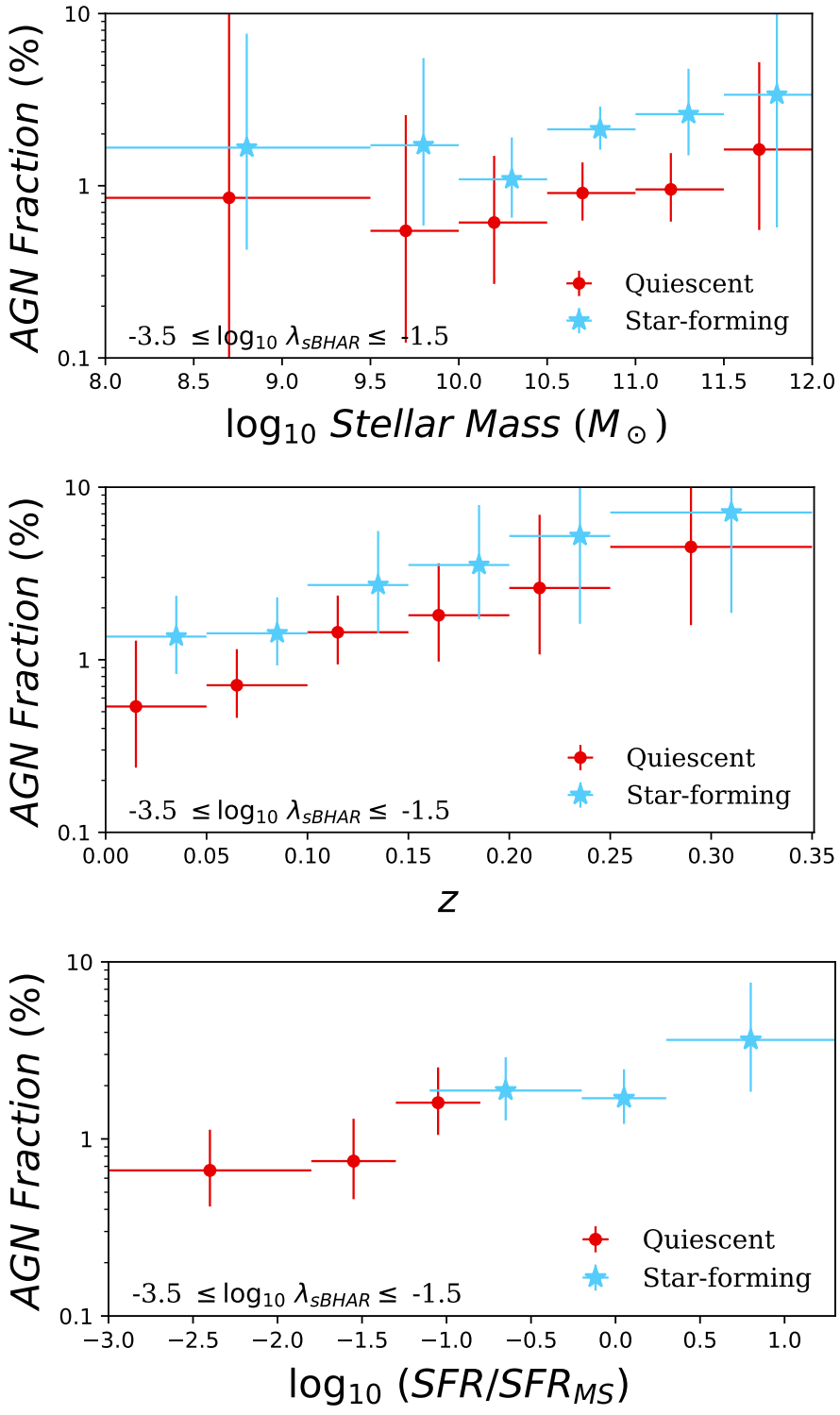

Figure 5.5: AGN fraction as a function of stellar mass (top), redshift (middle) and $\log_{10}(\mathrm{SFR}/\mathrm{SFR}_{\mathrm{MS}})$ (bottom). In each panel star-forming AGN fractions are shown as blue stars, and quiescent AGN fractions are red circles. The sBHAR (λ_{sBHAR}) integration limits are also displayed in the bottom-left corner.

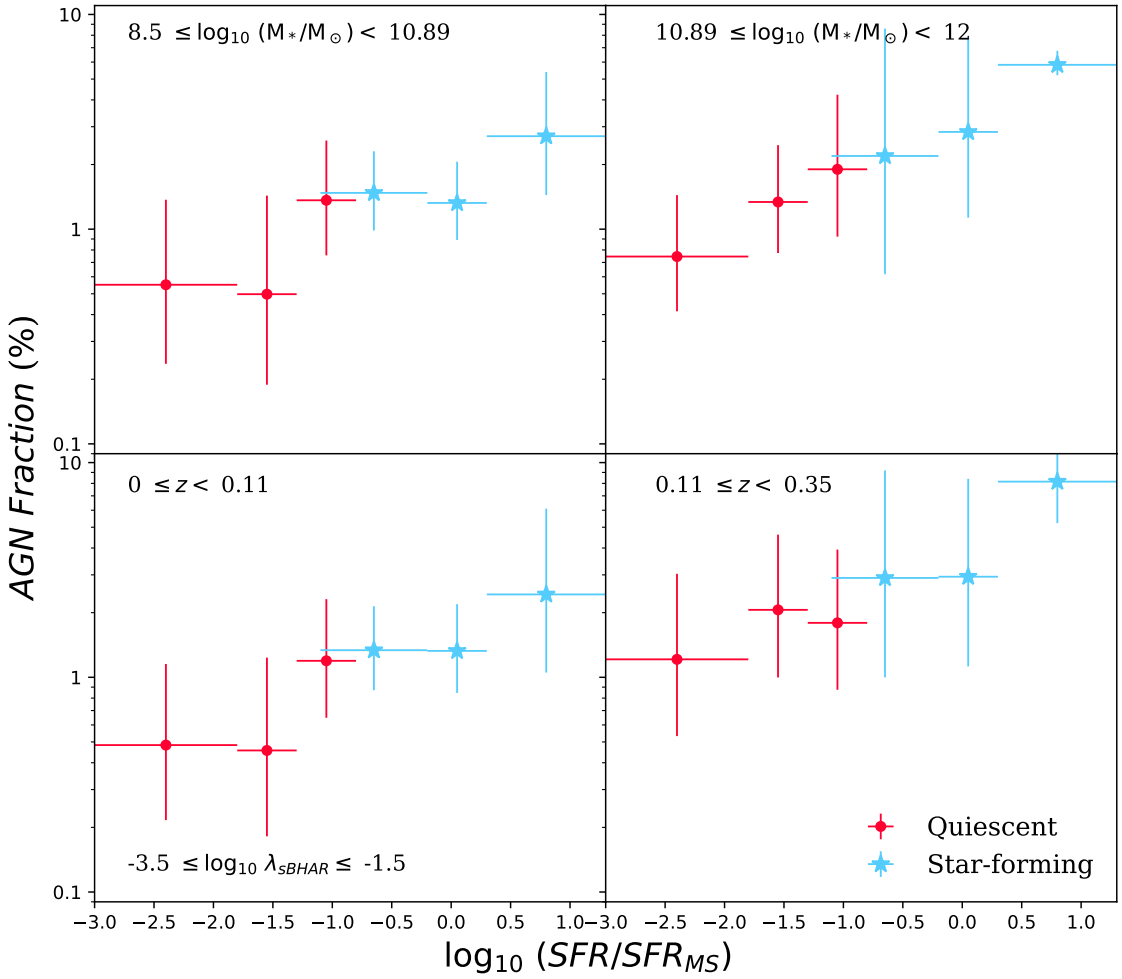

Figure 5.6: AGN fraction as a function of $\log_{10}(\mathrm{SFR}/\mathrm{SFR_{MS}})$ split at the median stellar mass (top row) and redshift (bottom row) of the sample. In each panel star-forming AGN fractions are shown as blue stars, and quiescent AGN fractions are red circles. The sBHAR (λ_{sBHAR}) integration limits are also displayed in the bottom-left corner.

in different accretion rate regimes.

Figure 5.7 shows the results of this analysis. The accretion rate limits are shown in the bottom-left corner and increases from $-3.5 \leq \log_{10} \lambda_{\mathrm{sBHAR}} < -2.5$ in the top panel to $-1.5 \leq \log_{10} \lambda_{\mathrm{sBHAR}} < -0.5$ in the bottom panel. The majority of my sample is found within the low accretion rate limits of figure 5.7's top panel. For this reason it is unsurprising to see a recreation of the AGN fraction increase with redshift seen in figure 5.5. It rises from around about 1% to just under 10% across the sample's redshift range, with a slight enhancement in star-forming galaxies. For moderate accretion levels, in the middle panel, there is an overall drop in AGN fraction, compared to low-level accretion, to around 0.1%. As before, both populations maintain similar levels of AGN incidence. However, at $z = 0.15$ they diverge, where the star-forming AGN fraction becomes increasingly enhanced compared to the constant quiescent AGN fraction. Highly accreting AGN, in the bottom panel, experience a further drop in AGN incidence, to around 0.05%. Both quiescent and star-forming galaxies, again, have similar AGN fractions at low redshift, however this time both fractions are decreasing. There is also a divergence at $z = 0.15$ which appears more significant: as the quiescent AGN fraction continues to decrease, the star-forming AGN fraction increases.

Using the data in figure 5.7, I would argue that there are two different mechanisms that fuel AGN accretion in the nearby Universe. There is agreement between the star-forming and quiescent AGN fractions for the full redshift range for low accretion rate AGN, and up to $z = 0.15$ for more actively accreting AGN. This strong correlation in these regions implies that these AGN are fuelled at the same rate, and thus by a common mechanism. Above $z = 0.15$, the quiescent AGN fraction continues to plateau or drop in the moderate and high accretion rate panels respectively. Whereas, the star-forming AGN fractions begin to increase. Therefore this second mechanism dominates in star-forming galaxies, above this redshift threshold. The presence of these two distinct evolutionary behaviours is reflective of the 'feast' and 'famine' fuelling modes theorised in Kauffmann and Heckman (2009). However, the point at which one mechanism dominates over the other depends on more than just star-forming classification.

Stellar mass loss is thought to be able to provide the necessary supply of gas to sustain low levels of AGN accretion (Aird et al. 2018, 2019). This is likely the only source of fuel in quiescent galaxies and, in the absence of significant amounts of cold gas, will also fuel AGN in nearby star-forming galaxies (Ciotti and Ostriker 2007; Kauffmann and Heckman 2009; Wang et al. 2017). This fuelling mecha-

nism dominates at the lowest accretion rates and helps facilitate an increasing AGN fraction with redshift. Stellar mass loss continues to fuel a decreasing amount of higher rate AGN accretion at the lowest redshifts in both star-forming and quiescent galaxies. Above z = 0.15 in the higher accretion rate panels, AGN activity in star-forming galaxies becomes increasingly enhanced compared to the quiescent population. With increasing redshift comes a greater density of cold gas which fuels more star-formation and black hole accretion (Madau and Dickinson 2014). Figure 5.7 provides an estimate of the redshift at which a critical density of cold gas becomes available for gas-driven accretion to become the dominant fuelling mechanism. The upward trend of the star-forming fractions suggests that an ever greater amount of AGN activity at higher redshifts is fuelled by cold gas accretion. By definition, this star formation does not occur in the quiescent galaxies which serves as further evidence that two distinct mechanisms fuel AGN at higher redshifts. The plateauing and declining quiescent AGN fractions highlight that stellar mass loss alone is insufficient at fuelling highly accreting AGN.

Other AGN studies tend to focus on higher redshift regimes, thus their studies are less sensitive to lower level X-ray emission. For this reason, they tend to adopt Eddington ratio thresholds of 10^{-2} or an equivalent X-ray luminosity threshold of 10^{42} erg s^{-1} (for 10^{10} M$_\odot$ galaxies). When compared to figure 5.7, it is clear that the star-forming AGN fraction enhancement would dominate their results (e.g. Rosario et al. 2013; Wang et al. 2017; Xue et al. 2010; Yang et al. 2018). Thus they observe the regime where gas-driven accretion dominates and largely miss the AGN fuelled by stellar mass loss. However, by including AGN at lower redshifts and accretion rates I have further highlighted the existence of a second fuelling mechanism that dominates in these regimes.

5.7 Conclusions

In this chapter I have extended the analysis from chapter 4 to investigate how star-forming activity might influence the distribution of AGN in the nearby Universe. For this investigation I created two sets of definitions designed to track changing star-forming activity whilst accounting for redshift and mass driven enhancements. The first definition split the sample into star-forming and quiescent populations; the second, split the sample into even more refined bins of star-forming activity relative to the galactic main sequence.

I then applied these star-forming classifications to the BPT-selected AGN sample.

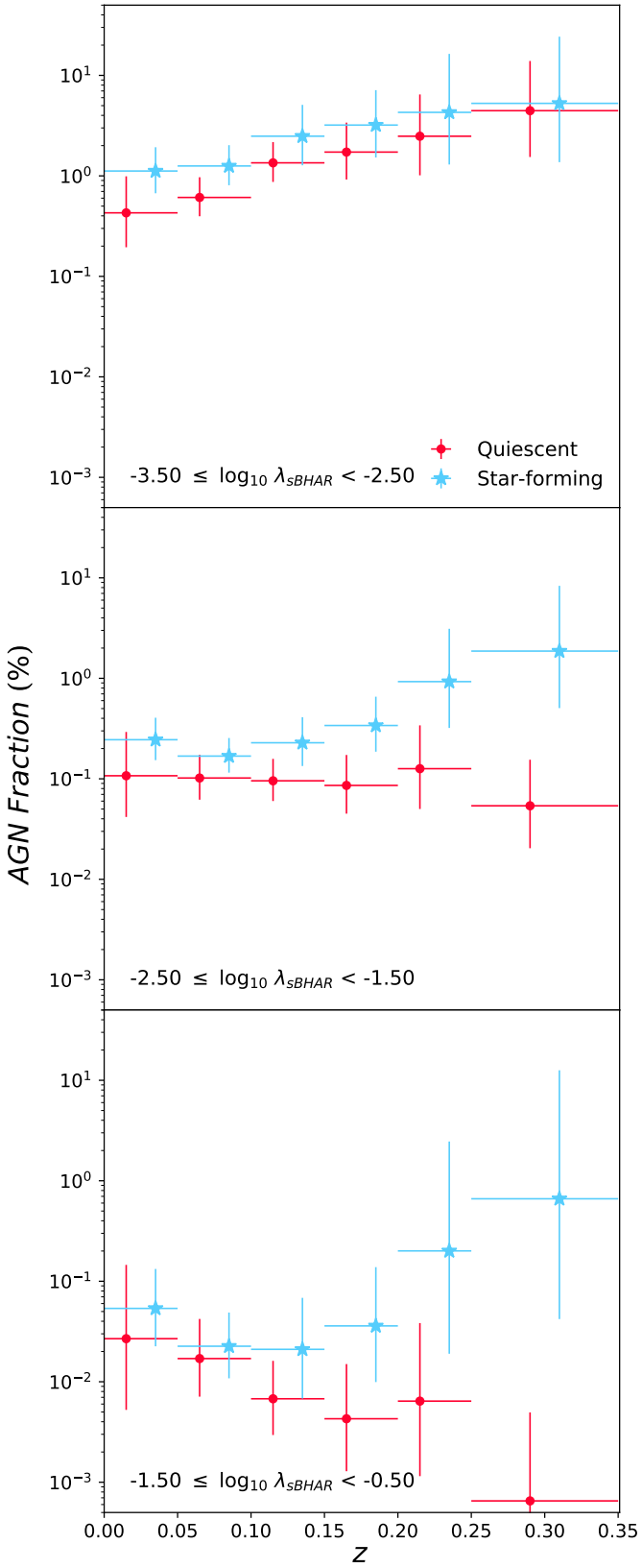

Figure 5.7: AGN fraction as a function of redshift, split up by specific black hole accretion rate, λ_{sBHAR}, increasing from the least-actively accreting AGN in the top row, to the most active in the bottom row. In each panel star-forming AGN fractions are shown as blue stars, and quiescent AGN fractions are red circles.

The vast majority of the star-forming X-ray selected AGN had sufficiently strong emission lines to be placed across the full extent of the BPT diagnostic. About half of the quiescent X-ray selected AGN were identified by the BPT diagnostic but were limited to the AGN region. The other half of the quiescent sample was made up of higher mass galaxies which did not have sufficiently significant emission line activity to make it onto the BPT diagnostic. I believe this is due to stellar absorption lines hiding the weaker optical emission lines from low levels of AGN activity. This is further evidence to suggest that the BPT diagram is not effective at identifying weakly accreting AGN.

I built on the probability distribution analysis performed in section 4.5 by applying the star-forming classifications to highlight how they affect the AGN population in the nearby Universe. The strongest star-formation-driven changes were seen in the redshift-binned distributions. Quiescent AGN showed a significant steepening with redshift, appearing to favour much lower accretion rate activity. Star-forming galaxies appeared to show the opposite trend with flatter distributions at higher redshifts. This implies star formation facilitates more active accretion at higher redshifts.

By integrating under these distributions I could calculate robust AGN fractions and determine how star-forming activity affects the AGN population in the nearby Universe. Reassuringly, both AGN fraction trends with stellar mass and redshift seen in section 4.6 are recreated when split by star-forming classification. There is little change in AGN fraction with stellar mass, and a noticeable increase with redshift. I found that star-forming activity increases the incidence of AGN by a factor of 2 at a $> 3.5\sigma$ significance. This enhancement is also seen when binning the AGN fraction as a function of $\log_{10}(\mathrm{SFR}/\mathrm{SFR}_{\mathrm{MS}})$.
Finally, I re-integrated the distributions within reduced accretion rate limits so I could investigate how the AGN fraction changes in different accretion regimes. This highlighted evidence for two distinct fuelling mechanisms. Both the star-forming and quiescent fractions show strong agreement at the lowest accretion rates, and at very low redshifts for more actively accreting black holes. Stellar mass loss is thought able to sustain the weak levels of AGN accretion for quiescent galaxies, and also in star-forming galaxies when there is a lack of cold gas. However, above $z = 0.15$ there is a strong enhancement of star-forming AGN activity. This is likely the threshold at which cold gas forms a critical density to allow gas-driven AGN activity to become the dominant fuelling mechanism for more active black holes.

Chapter 6

Summary & Conclusions

In this thesis I have performed a robust characterisation of the AGN population in the nearby Universe. To perform this analysis, I have adopted a similar approach to studies outlined in section 1.5. I combined SDSS galaxy samples with XMM-Newton serendipitous sources (both discussed in chapter 2) to measure the incidence of AGN as a function of galaxy properties in the nearby ($z < 0.35$) Universe, carefully accounting for the incompleteness of the X-ray imaging. With this novel approach I was able to perform a focused study of the X-ray selected AGN population in dwarf galaxies (chapter 3) and then expand that analysis to include the wider local galaxy population (chapter 4). In addition, I studied how star-forming activity might affect the probability of finding an AGN in these nearby galaxies (chapter 5). In this chapter I aim to summarise the key findings from this work and outline potential avenues for future research.

6.1 Identifying AGN Activity

My AGN identification technique involved, first, robustly matching X-ray signals to a galaxy's optical nucleus. Then I modelled the expected X-ray luminosity due to X-ray binaries and hot gas from a galaxy of a given stellar mass, SFR and redshift to further clarify the nature of the matched X-ray emission. If the observed X-ray luminosity was at least three times larger than the prediction I attributed that emission to AGN activity. Whilst this approach has been successful, no AGN selection technique is perfect. This work's sole reliance on X-ray selection will have missed AGN that would have been identified with other techniques. Throughout this thesis I have compared my X-ray selected AGN sample with the BPT diagnostic. This technique is primarily used to identify AGN activity in the local Universe as some of the required optical emission lines required are redshifted out of the observed band above $z \sim 0.4$. Given its popularity within the field, it would be revealing to see how effective the diagnostic is compared to X-ray selection.

Overall, I found that the BPT diagnostic is not as effective as X-ray selection at identifying a wide range of AGN activity. Around 85% of the dwarf galaxy AGN sample identified in chapter 3 had sufficiently significant emission lines to be placed on the BPT diagnostic, this dropped to around 72% for the full AGN sample in chapter 4. In chapter 5, I analysed the properties that increase the probability of a galaxy not making it onto the BPT diagram. Over half of the AGN I classified as quiescent hosts did not possess sufficiently strong emission lines, compared with only a handful of AGN hosts I defined as star-forming. Hβ was the most frequently non-significant emission line. Lack of strength in this emission line also caused large numbers of AGN to be missed by the BPT diagnostic in Cid Fernandes et al. (2010). At all stellar masses, quiescent AGN made up a smaller proportion of my BPT-selected X-ray AGN sample, peaking at 75% of the X-ray selected sample in the $10 \leq \log_{10}(M_*/M_\odot) < 11$ bin, before dropping rapidly to 17% at the highest masses. I argued that this is due to strong stellar absorption lines potentially hiding weak optical emission lines from the central AGN.

Even when X-ray selected AGN made it onto the BPT diagnostic, their classifications were not always the same. Chapter 3 showed that X-ray selected AGN in dwarf galaxies primarily lay in the star-forming region of the BPT diagnostic. This implies star-forming signatures dominated over AGN activity in the optical part of these galaxies' spectra. Using the α_{OX} relation (Lusso and Risaliti 2016), I attempted to calculate the contribution made by the AGN to the optical emission, calculated from the observed X-ray luminosity. This analysis confirmed that emission from the galaxy, and thus from star-forming processes, dominated over optical AGN emission. Chapter 4 helped to confirm this analysis. I found that a greater proportion of X-ray selected AGN have their classifications agree at higher stellar masses. In the dwarf galaxy mass regime, approximately 25% of galaxies are classified as AGN based both on X-ray emission and the BPT diagnostic. This rises steadily, reaching 93% agreement at the highest stellar masses. This does not appear to be driven by increased AGN activity as no correlation was found between AGN classification agreement and observed X-ray luminosity. Behroozi et al. (2019) collated results that suggested the fraction of galaxies with ongoing star formation decreases with stellar mass. Given the low rate of classification agreement at lower stellar masses, I argued that star-forming activity can cloud the detection of AGN signatures by the BPT diagnostic and result in misclassification. This result adds to a growing body of work suggesting that the BPT diagram does not produce accurate classification in blue, star-forming galaxies (e.g. Agostino and Salim 2019; Cann et al. 2019; Moran et al. 2002).

6.2 The Nearby AGN Population

Once a sample of AGN was identified, I constructed sensitivity curves comprised of 3XMM upper limits within the relevant stellar mass and redshift ranges. These curves were used to correct for incompleteness in the 3XMM survey. The observed AGN distributions were divided by the correction fractions taken from these sensitivity curves to recover the expected number of galaxies where X-ray surveys were sensitive enough to detect an AGN. From these corrected AGN count distributions, I could measure distributions that described the probability of finding an AGN in a galaxy as a function of X-ray luminosity and specific black hole accretion rate. Throughout, these distributions were well described by a power law. In particular, they highlight significant amounts of activity at low X-ray luminosities and low accretion rates.

These probability distributions were also useful in understanding the relationship between AGN activity and the host galaxy. Each probability distribution was binned as a function of different properties so that possible changes could be observed. To help clarify the extent of these changes, I integrated these distributions between certain X-ray luminosity ($42 \leq \log_{10}\ \mathrm{L_X}\ (\mathrm{erg\ s^{-1}}) \leq 44$) and specific black hole accretion rate ($-3.5 \leq \log_{10} \lambda_{\mathrm{sBHAR}} \leq -1.5$) limits to calculate the fraction of galaxies expected to host AGN. This approach left me with two sets of AGN fractions. The luminosity-derived points created an observational-style definition whereby an AGN is an object emitting at an X-ray luminosity $\geq 10^{42}$ erg s^{-1}. Whereas the sBHAR-derived points defined an AGN as a black hole accreting at rates $\geq 10^{-3.5}$. However, the luminosity-derived points reproduced the observational biases inherent in previous AGN studies, despite the application of completeness corrections. Thus, these fractions were disregarded in favour of the sBHAR-derived fractions.

6.2.1 Stellar Mass & Redshift

In each chapter I analysed changes in the AGN fraction as a function of stellar mass and redshift. Regardless of how the galaxy population was divided, there appears to be no change in the AGN fraction with stellar mass. Thus it is unlikely that possessing a larger amount of gas and dust increases the probability of a galaxy fuelling AGN activity. Each chapter also found that an increase in redshift increased the probability that a given galaxy hosts AGN activity. This is believed to be due to the increased availability of cold gas at higher redshifts (e.g. Popping et al. 2012; Vito

et al. 2014). This behaviour is also consistent with the increases seen in Madau and Dickinson (2014) suggesting an increase in the average level of accretion towards $z = 2 - 3$.

However, these single property analyses hide a more complex picture. Chapter 4 compared probability distributions split by stellar mass with those same distributions from chapter 3 and results derived from comparable Eddington ratio distributions out to $z \sim 1$. These were used to calculate AGN fractions within reduced $\lambda_{\rm sBHAR}$ limits which reflected the generally higher threshold of AGN activity used in other studies. This comparison highlighted the existence of differential stellar mass growth rates. Higher mass galaxies (typically $M_* > 10^{10} M_\odot$) show evidence of a strong redshift-driven increase out to $z \sim 1$. As before, this is thought to be caused by the increasing availability of cold gas. Aird et al. (2018), Georgakakis et al. (2017), and Yang et al. (2018) all observed similar behaviour in different AGN samples taken from a range of redshifts and wavelengths. Lower mass galaxies, however, have consistent AGN fractions out to $z \sim 1$. It is thought that their shallower potential wells are less likely to attract gas into the galactic centre and fuel the SMBH (Bellovary et al. 2013). A similar drop in AGN activity in these mass regimes has also been observed in different AGN samples (Aird et al. 2018; Yang et al. 2017). Thus, regardless of formation mechanism, black hole seeds appear difficult to grow.

6.2.2 SFR

There is strong observational evidence to suggest a correlation between the volume-averaged black hole accretion activity and star formation history going back to $z \sim 6$. Thus, in chapter 5, I investigated whether the incidence of AGN in the nearby Universe correlates with star-forming activity. Using the "rotated SFR" quantity from Moustakas et al. (2013), I established a measure of star formation that takes into account the enhancing effects that stellar mass and redshift have on the SFR of a given galaxy. I normalised this quantity to my galaxy sample and then split it into star-forming and quiescent populations. With this definition, I calculated the AGN fraction as a function of stellar mass and redshift for both the star-forming and quiescent populations. I found evidence to suggest a factor of two increase in AGN fraction in star-forming galaxies at a 3.5σ significance. A similar level of star-forming-driven enhancement was observed by Azadi et al. (2015) in a different X-ray selected AGN sample out to $z \sim 1.2$. In addition, I further divided the galaxy sample into five finer bins of rotated SFR. I found that AGN fraction rises with increasing values of rotated SFR. Similar positive correlations have been observed between AGN activity and SFR in samples at higher redshifts (e.g. Chen et al. 2013; Yang

et al. 2017). However, it is difficult to determine with the data available whether this positive correlation is brought about because of the increase in the availability of cold gas which drives star-formation, the motion of that gas within the galaxy, or a combination of both.

6.2.3 Fuelling Mechanisms

The AGN fraction definition was chosen to encompass the full range of activity in my observed sample. However, observations of the probability distributions in chapter 5 suggested that there were differences in the AGN population that would be hidden by such a wide definition. There were very distinct differences in the redshift-driven progression of quiescent and star-forming AGN populations. Quiescent galaxies at higher redshifts appeared much more likely to host AGN with lower levels of accretion. However, star-forming galaxies at these same redshifts were more likely to host AGN with higher levels of accretion. To investigate this phenomenon further, I split the integration range into three smaller bands: low ($-3.5 \leq \log_{10}\lambda_{\mathrm{sBHAR}} < -2.5$), moderate ($-2.5 \leq \log_{10}\lambda_{\mathrm{sBHAR}} < -1.5$) and high ($-1.5 \leq \log_{10}\lambda_{\mathrm{sBHAR}} < -0.5$) accretion rates. I analysed how these new AGN fraction definitions changed as a function of redshift. At low accretion rates, and up to $z = 0.15$ for more actively accreting AGN, there is strong agreement in AGN fraction. Above this redshift in the moderate and higher accretion rate regimes, however, there is an enhancement in the AGN fraction for star-forming galaxies. The presence of these two distinct evolutionary behaviours is reflective of the "feast and famine" fuelling modes theorised in Kauffmann and Heckman (2009). Stellar mass loss is thought able to sustain the weak levels of AGN accretion for quiescent galaxies and star-forming galaxies lacking in cold gas. And the strong enhancement of star-forming AGN activity above $z = 0.15$ suggests that gas-driven AGN activity becomes the dominant fuelling mechanism for these more active black holes.

6.3 Final Remarks & Future Work

In this thesis I have performed a robust characterisation of the X-ray selected AGN population in the nearby Universe. I identified AGN across the nearby galaxy population, including dozens of new dwarf galaxy AGN. I have highlighted the complex role that host galaxy stellar mass and redshift play in changing the probability of hosting an AGN. Furthermore, I have shown that star-forming activity enhances the overall AGN fraction, particularly in black holes accreting at higher accretion rates and in higher redshift hosts. These AGN probability distributions and fractions will

provide useful constraints for simulations and observational studies of black hole evolution.

This work has been able to effectively explore these relationships because of the large AGN samples and comprehensive X-ray upper limits which can correct for incompleteness. However, to further explore some of these complex relationships, a larger AGN sample is required. Such an increase would help disentangle the nature of the the differential stellar mass growth rates, by allowing the identification of changing redshift dependencies in different stellar mass ranges. Furthermore, it would allow for a more detailed study into the existence of multiple black hole fuelling mechanisms and where precisely the change occurs. eRosita (Merloni et al. 2012) is a new X-ray instrument, launched as part of the Spectr-Roentgen-Gamma mission, that would be well suited to constructing this expanded sample. It is currently over one year into its three-and-a-half year plan to perform a $0.2-12$ keV all-sky survey, with ambitions to make around 3 million AGN detections. Samples of this size and distribution would be invaluable in constructing representative populations of AGN needed to further study these questions.

Appendix A

Images illustrating Dwarf Galaxy Cross-Matching Procedure

Figure A.1 provides examples from my dwarf galaxy AGN sample and illustrates results of the cross-matching process. The images are ordered by increasing sky separation between the optical centre of the galaxy (black cross) and the X-ray source (magenta cross). In the top left panel, the X-ray source is 0.3" away from the optical centre of the galaxy, and it sits well within the X-ray position error (solid circle). In contrast, the optical centre of the galaxy in the bottom right panel sits at the edge of the 3.5 × position error (dashed circle) as it lies 6.8" away from the X-ray source.

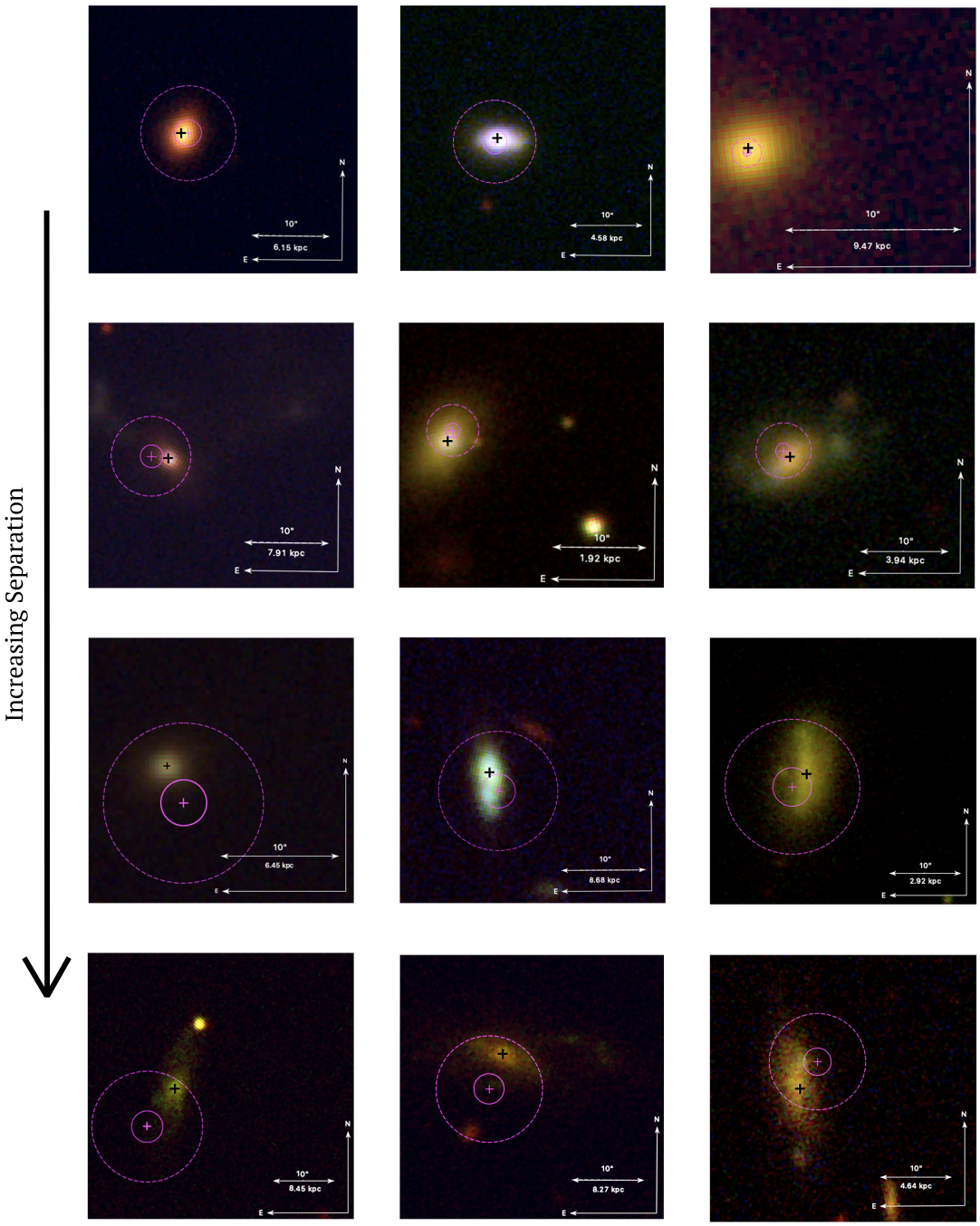

Figure A.1: A sample of 12 fields showing the positions of the galaxy (black cross) and X-ray source (magenta cross) surrounded associated position error (solid circle) and 3.5 × position error (dashed circle). These images are ordered by increasing sky separation between these signals.

Appendix B

List of Dwarf Galaxy AGN Candidates

Table B.1 contains data for the 61 dwarf galaxies that I identified as hosting X-ray AGN. The table is split into two sections: first are the AGN with a valid photometric SFR, followed by the AGN whose SFR was calculated from their H_α luminosity, using the Kennicutt and Evans (2012) method. Each galaxy's co-ordinates have also been uploaded to SIMBAD and NED to check if these AGN have been previously detected. The columns are defined as follows:

- (1) Common name of galaxy taken from SIMBAD and NED. Galaxies marked with an asterisk are those which made up the statistical sample (see section 3.5 for more details).
- (2) A letter in this column indicates that the AGN has been identified previously. Here follows a list of the studies which have previously identified my AGN, the letters used to represent them and the wavelength range within which their study focuses:
  - a: Véron-Cetty, M.-P. and Véron, P. (2010) (Optical, X-ray & Radio)
  - b: Reines et al. (2013) (Optical)
  - c: Lemons et al. (2015) (X-ray)
  - d: Sartori et al. (2015) (Optical & IR)
  - e: Sun and Shen (2015) (Optical)
  - f: Baldassare et al. (2017) (X-ray & UV)
  - g: Kawasaki et al. (2017) (Optical)
  - h: Marleau et al. (2017) (IR)
  - i: Nucita et al. (2017) (X-ray)

- (3) & (4) Galaxy co-ordinates taken from MPA-JHU (based on SDSS DR8).
- (5) & (6) X-ray co-ordinates taken from 3XMM DR7.
- (7) Separation between the optical and X-ray co-ordinates in arcseconds.
- (8) Total stellar mass of galaxy and associated errors in units of $M_{\odot}$. The median error for the upper/lower error on the mass is 0.1/0.08 dex.
- (9) SFR and associated errors in units of $M_{\odot}$ yr^{-1}. The median error for the upper/lower error on the SFR is 0.27/0.21 dex.
- (10) Observed 2 - 12 keV X-ray luminosity and associated errors, taken from 3XMM, in units of erg s^{-1}.
- (11) Predicted 2 - 12 keV X-ray luminosity ($L_{\rm XRB} + L_{\rm Gas}$; see chapter 3 for more information) in units of erg s^{-1}.

Name	Previously found?	Optical (^{o}) RA	Optical (^{o}) Dec	X-ray (^{o}) RA	X-ray (^{o}) Dec	Separation (")	Mass ($M_\odot$)	SFR ($M_\odot\ yr^{-1}$)	$L_{\rm Obs}$ ($erg\ s^{-1}$)	$L_{\rm Pred}$ ($erg\ s^{-1}$)
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
SDSS J011523.96+003808.7 *		18.850199	0.635465	18.849929	0.634991	1.964	$8.60\times10^{8^{+5.9E8}_{-8.2E8}}$	$1.39\times10^{-1^{+4.0E-2}_{-3.5E-2}}$	$4.90\times10^{40}\pm5.1E40$	$5.47\times10^{38}\pm1.6E38$
SDSS J014529.26+001036.0	e	26.372410	0.177643	26.371791	0.176987	3.245	$1.14\times10^{9^{+7.8E8}_{-1.1E9}}$	$3.61\times10^{-1^{+4.1E-2}_{-9.5E-2}}$	$7.77\times10^{40}\pm6.4E40$	$1.21\times10^{39}\pm3.4E38$
SDSS J024117.09-001352.3 *		40.321250	-0.231196	40.320863	-0.232070	3.444	$1.70\times10^{9^{+3.7E8}_{-2.9E8}}$	$2.85\times10^{-1^{+1.4E-1}_{-1.1E-1}}$	$4.71\times10^{39}\pm7.7E39$	$1.10\times10^{39}\pm3.6E38$
2MASX J02564580+0603173		44.189983	6.055122	44.190282	6.054252	3.308	$3.09\times10^{6^{+3.5E6}_{-1.1E5}}$	$3.21\times10^{-2^{+5.2E-2}_{-7.6E-2}}$	$3.41\times10^{39}\pm1.9E40$	$7.62\times10^{37}\pm2.0E37$
LEDA 2402319 *		123.634210	51.884872	123.635037	51.883489	5.309	$2.82\times10^{9^{+5.4E8}_{-4.0E8}}$	$6.57\times10^{-1^{+3.2E-1}_{-3.5E-1}}$	$3.19\times10^{40}\pm4.1E40$	$2.28\times10^{39}\pm2.2E39$
2MASX J08193880+2103521 *	i	124.911720	21.064728	124.912299	21.064209	2.694	$1.12\times10^{9^{+2.2E8}_{-2.4E8}}$	$5.00\times10^{-2^{+3.5E-1}_{-3.8E-1}}$	$2.69\times10^{39}\pm4.8E39$	$3.86\times10^{38}\pm1.8E38$
SDSS J082228.93+034551.7		125.620580	3.764374	125.619971	3.764018	2.543	$6.63\times10^{8^{+1.3E8}_{-1.2E8}}$	$1.30\times10^{0^{+2.8E-1}_{-2.6E-1}}$	$1.13\times10^{40}\pm4.9E39$	$3.24\times10^{39}\pm3.5E39$
2MASS J08320053+1912058	g	128.002140	19.201637	128.002743	19.201694	2.074	$1.97\times10^{9^{+7.8E8}_{-8.0E8}}$	$1.61\times10^{0^{+5.7E-1}_{-4.1E-1}}$	$4.16\times10^{40}\pm1.8E40$	$4.32\times10^{39}\pm7.8E39$
SDSS J085629.97+380456.1		134.124920	38.082253	134.126541	38.080536	7.693	$1.04\times10^{9^{+2.0E8}_{-1.5E8}}$	$2.85\times10^{-1^{+3.5E-1}_{-3.1E-1}}$	$7.70\times10^{39}\pm1.4E40$	$9.42\times10^{38}\pm9.5E38$
SDSS J090335.40+151142.0		135.897540	15.195035	135.897672	15.195871	3.045	$7.49\times10^{8^{+1.7E8}_{-1.5E8}}$	$6.70\times10^{-2^{+3.2E-1}_{-3.7E-1}}$	$1.46\times10^{40}\pm9.8E39$	$3.44\times10^{38}\pm2.3E38$
SDSSCGB 15.2		139.676320	16.479506	139.677069	16.479499	2.602	$1.57\times10^{9^{+6.5E8}_{-3.8E8}}$	$2.50\times10^{0^{+3.2E-1}_{-2.5E-1}}$	$4.24\times10^{40}\pm2.9E40$	$6.33\times10^{39}\pm7.2E39$
2XMMi J092720.4+362407 *	i	141.835160	36.401897	141.835438	36.402101	1.090	$3.30\times10^{8^{+5.0E7}_{-4.0E7}}$	$1.13\times10^{-1^{+2.5E-1}_{-1.9E-1}}$	$7.88\times10^{39}\pm3.7E39$	$3.44\times10^{38}\pm2.4E38$
PWC2011 J100805.1+125650		152.021360	12.947362	152.021634	12.947564	1.199	$2.46\times10^{9^{+5.3E8}_{-4.1E8}}$	$3.53\times10^{-1^{+2.2E-1}_{-1.8E-1}}$	$1.93\times10^{40}\pm1.3E40$	$1.45\times10^{39}\pm7.0E38$
SDSS J102526.59+124540.3		156.360800	12.761314	156.359322	12.760766	5.533	$1.02\times10^{9^{+2.8E8}_{-1.8E8}}$	$3.28\times10^{-1^{+3.0E-1}_{-2.1E-1}}$	$8.44\times10^{39}\pm1.7E40$	$1.03\times10^{39}\pm8.4E38$
LEDA 30866		157.254560	29.635242	157.254775	29.635034	1.003	$8.41\times10^{7^{+6.6E7}_{-6.2E7}}$	$3.90\times10^{-1^{+5.3E-2}_{-8.9E-2}}$	$3.12\times10^{39}\pm2.4E40$	$9.49\times10^{38}\pm2.8E38$
Mrk 1434 *	c	158.542300	58.063630	158.542303	58.063435	0.698	$1.00\times10^{7^{+4.0E6}_{-4.4E5}}$	$6.15\times10^{-2^{+1.3E-1}_{-1.7E-1}}$	$1.18\times10^{40}\pm3.0E39$	$1.44\times10^{38}\pm9.1E37$
SDSS J103844.88+533005.2 *		159.686900	53.501450	159.687208	53.501374	0.703	$3.84\times10^{7^{+1.6E7}_{-1.6E6}}$	$2.19\times10^{-1^{+1.9E-1}_{-3.7E-1}}$	$3.78\times10^{39}\pm2.4E38$	$5.10\times10^{38}\pm5.9E38$
LEDA 2116718		166.425690	38.056496	166.424690	38.056521	2.835	$4.15\times10^{8^{+5.8E7}_{-5.0E7}}$	$1.57\times10^{-1^{+2.7E-1}_{-2.3E-1}}$	$3.11\times10^{39}\pm2.3E40$	$4.73\times10^{38}\pm3.9E38$
UGC 6192 *	i	167.301650	61.396324	167.301427	61.396035	1.110	$2.43\times10^{8^{+4.1E7}_{-3.0E7}}$	$5.04\times10^{-2^{+2.7E-1}_{-2.9E-1}}$	$1.60\times10^{39}\pm4.2E38$	$1.74\times10^{38}\pm1.4E38$
SDSS J112830.77+583342.9	h	172.128300	58.561844	172.128389	58.561797	0.242	$1.67\times10^{9^{+5.7E8}_{-5.7E8}}$	$5.87\times10^{-1^{+4.8E-1}_{-3.9E-1}}$	$4.01\times10^{40}\pm5.5E39$	$1.76\times10^{39}\pm2.5E39$
SDSS J112910.56+582309.0 *	i	172.294000	58.385834	172.293456	58.385212	2.468	$1.28\times10^{9^{+2.8E8}_{-2.0E8}}$	$1.17\times10^{0^{+3.2E-1}_{-1.4E-1}}$	$2.99\times10^{40}\pm3.4E40$	$3.12\times10^{39}\pm2.7E39$
Mrk 1303 *	i	175.055150	-0.411672	175.055270	-0.411804	0.656	$8.35\times10^{8^{+4.5E8}_{-2.2E8}}$	$1.10\times10^{0^{+3.8E-1}_{-1.9E-1}}$	$1.06\times10^{40}\pm6.7E39$	$2.79\times10^{39}\pm3.1E39$
2XMM J114501.7+194549 *		176.257550	19.763748	176.257412	19.763534	0.907	$2.27\times10^{9^{+5.8E8}_{-5.2E8}}$	$8.60\times10^{-2^{+4.8E-1}_{-9.4E-1}}$	$1.02\times10^{40}\pm7.2E39$	$7.65\times10^{38}\pm6.2E38$
SDSS J115558.40+232730.7	i	178.993360	23.458563	178.993608	23.459134	2.210	$2.14\times10^{9^{+3.5E8}_{-2.5E8}}$	$1.43\times10^{0^{+3.1E-1}_{-2.3E-1}}$	$3.80\times10^{40}\pm2.1E40$	$4.02\times10^{39}\pm3.9E39$
NGC 4117 *	i	181.942140	43.126354	181.942229	43.126437	0.380	$2.18\times10^{9^{+4.5E8}_{-4.8E8}}$	$1.70\times10^{-2^{+4.3E-1}_{-1.0E0}}$	$6.17\times10^{39}\pm5.0E38$	$5.53\times10^{38}\pm1.5E38$
ECO 11516 *		182.253720	42.475260	182.253462	42.474715	2.089	$1.92\times10^{9^{+4.4E8}_{-3.8E8}}$	$1.10\times10^{-1^{+3.2E-1}_{-3.3E-1}}$	$2.21\times10^{40}\pm1.3E40$	$7.31\times10^{38}\pm3.6E38$
SDSS J121352.97+141312.5		183.470730	14.220132	183.469156	14.220695	5.867	$1.11\times10^{9^{+2.6E8}_{-2.5E8}}$	$3.28\times10^{-1^{+7.9E-2}_{-6.5E-2}}$	$3.36\times10^{39}\pm8.5E39$	$1.05\times10^{39}\pm2.4E38$
SDSS J121707.89+034056.3 *		184.282880	3.682264	184.283938	3.680705	6.775	$1.05\times10^{7^{+5.8E6}_{-9.1E6}}$	$5.30\times10^{-2^{+1.5E-1}_{-1.3E-1}}$	$2.24\times10^{39}\pm1.4E39$	$1.25\times10^{38}\pm7.2E37$
LEDA 39539		184.632190	5.849806	184.632586	5.851398	5.909	$2.17\times10^{8^{+4.2E7}_{-5.1E7}}$	$1.90\times10^{-3^{+5.4E-1}_{-1.0E0}}$	$2.47\times10^{39}\pm1.3E39$	$5.62\times10^{37}\pm1.7E37$
NGC 4395 *	b; c; f; h; i	186.453610	33.546870	186.453591	33.546854	0.091	$2.50\times10^{7^{+6.4E6}_{-5.6E6}}$	$1.57\times10^{-4^{+6.5E-1}_{-7.5E-1}}$	$1.39\times10^{40}\pm7.0E38$	$6.23\times10^{36}\pm1.6E36$
2XMM J123519.9+393110	i	188.833530	39.519196	188.832976	39.519672	2.298	$7.85\times10^{7^{+1.5E7}_{-1.2E7}}$	$2.60\times10^{-1^{+1.7E-1}_{-1.6E-1}}$	$1.10\times10^{40}\pm4.4E39$	$6.26\times10^{38}\pm4.3E38$
NVSS J123542-001252		188.927000	-0.215192	188.926617	-0.214698	2.255	$6.47\times10^{7^{+3.2E7}_{-2.6E7}}$	$4.15\times10^{-1^{+7.8E-2}_{-5.3E-2}}$	$7.14\times10^{39}\pm9.3E39$	$9.89\times10^{38}\pm2.7E38$
LEDA 44693 *		195.004040	27.945436	195.004703	27.945354	2.118	$1.73\times10^{9^{+3.6E8}_{-4.1E8}}$	$2.22\times10^{-2^{+4.6E-1}_{-1.0E0}}$	$8.02\times10^{39}\pm6.0E39$	$4.77\times10^{38}\pm1.8E38$
7W 1258+27W06 *	i	195.140300	27.637766	195.140594	27.637073	2.659	$2.64\times10^{9^{+6.1E8}_{-3.9E8}}$	$1.04\times10^{0^{+2.7E-1}_{-1.8E-1}}$	$2.45\times10^{40}\pm2.3E40$	$3.10\times10^{39}\pm2.3E39$
2MASX J13070847+5357446 *	i	196.785110	53.962387	196.784717	53.962394	0.835	$2.37\times10^{9^{+9.2E8}_{-8.1E8}}$	$2.20\times10^{0^{+1.2E-1}_{-7.4E-2}}$	$4.89\times10^{40}\pm3.3E40$	$5.78\times10^{39}\pm2.1E39$
SDSS J130821.42+113055.0		197.089300	11.515293	197.087516	11.515225	6.279	$2.03\times10^{8^{+2.5E7}_{-2.9E7}}$	$2.17\times10^{-1^{+2.7E-1}_{-1.8E-1}}$	$4.99\times10^{39}\pm1.5E40$	$5.60\times10^{38}\pm4.8E38$
SDSS J131930.27+552146.0		199.876170	55.362810	199.877721	55.362280	3.691	$1.09\times10^{9^{+2.0E8}_{-1.4E8}}$	$2.72\times10^{-1^{+2.8E-1}_{-2.6E-1}}$	$7.34\times10^{39}\pm6.0E39$	$9.04\times10^{38}\pm7.2E38$
2XMM J134107.9+263047	a	205.283140	26.513401	205.282967	26.513346	0.596	$2.50\times10^{9^{+1.1E9}_{-1.0E9}}$	$1.22\times10^{0^{+2.0E-1}_{-1.0E-1}}$	$5.54\times10^{40}\pm1.2E41$	$3.67\times10^{39}\pm1.9E39$
2XMM J134427.6+560130	d	206.114010	56.024930	206.114470	56.025227	1.410	$1.40\times10^{9^{+7.9E8}_{-5.3E8}}$	$1.60\times10^{1^{+1.0E-1}_{-8.1E-2}}$	$1.57\times10^{41}\pm8.2E40$	$3.99\times10^{40}\pm1.5E40$
2XMM J134719.1+581437 *		206.830260	58.243744	206.829865	58.243869	0.876	$1.26\times10^{9^{+2.6E8}_{-1.7E8}}$	$5.08\times10^{-1^{+3.2E-1}_{-2.5E-1}}$	$1.23\times10^{40}\pm5.5E39$	$1.52\times10^{39}\pm1.4E39$

Table B.1: Data for all 61 dwarf galaxies believed to be strong candidates for hosting an AGN. continued on next page...

Name	Previously found?	Optical (^{o}) RA	Dec	X-ray (^{o}) RA	Dec	Separation (")	Mass ($M_\odot$)	SFR ($M_\odot\ yr^{-1}$)	L_{Obs} ($erg\ s^{-1}$)	L_{Pred} ($erg\ s^{-1}$)
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
2XMM J134736.4+173404	d	206.901690	17.567960	206.901721	17.567890	0.278	$2.28 \times 10^{9^{+6.4E8}_{-3.9E8}}$	$1.26 \times 10^{0^{+4.7E-1}_{-2.8E-1}}$	$8.88 \times 10^{41} \pm 5.9E40$	$3.59 \times 10^{39} \pm 4.7E39$
UGC 9215		215.862980	1.726289	215.863058	1.726305	0.299	$6.87 \times 10^{8^{+1.4E8}_{-9.3E7}}$	$2.17 \times 10^{-1^{+3.1E-1}_{-2.6E-1}}$	$3.85 \times 10^{39} \pm 1.6E39$	$6.61 \times 10^{38} \pm 5.9E38$
SDSS J143102.57+281625.9		217.760760	28.273851	217.760817	28.273226	2.258	$9.17 \times 10^{8^{+2.5E8}_{-1.7E8}}$	$3.14 \times 10^{-1^{+1.7E-1}_{-1.0E-1}}$	$8.26 \times 10^{39} \pm 9.3E39$	$9.71 \times 10^{38} \pm 4.2E38$
2MASX J14401271+0247441 *	b; c; d; f; h; i	220.052920	2.795424	220.052756	2.795508	0.657	$2.66 \times 10^{9^{+5.6E8}_{-3.9E8}}$	$7.20 \times 10^{-1^{+2.9E-1}_{-2.5E-1}}$	$2.42 \times 10^{40} \pm 8.1E39$	$2.36 \times 10^{39} \pm 1.9E39$
2XMM J144056.3+033145		220.236020	3.528175	220.235329	3.527134	4.501	$1.44 \times 10^{8^{+6.9E7}_{-8.6E7}}$	$1.42 \times 10^{-1^{+9.7E-2}_{-5.1E-2}}$	$4.34 \times 10^{39} \pm 8.9E39$	$3.74 \times 10^{38} \pm 1.1E38$
ECO 2050 *		228.550740	13.809550	228.549996	13.810569	4.487	$4.27 \times 10^{8^{+7.8E7}_{-5.6E7}}$	$1.33 \times 10^{-1^{+2.9E-1}_{-2.8E-1}}$	$2.07 \times 10^{39} \pm 4.8E39$	$4.17 \times 10^{38} \pm 3.8E38$
SDSS J153704.18+551550.5 *		234.267400	55.264060	234.267702	55.263194	3.186	$7.24 \times 10^{6^{+1.3E6}_{-1.1E6}}$	$2.46 \times 10^{-2^{+2.2E-1}_{-2.3E-1}}$	$1.13 \times 10^{39} \pm 1.8E38$	$5.82 \times 10^{37} \pm 5.4E37$
SDSS J154818.94+350741.2		237.078920	35.128174	237.079120	35.128482	1.259	$1.66 \times 10^{9^{+3.3E8}_{-2.2E8}}$	$5.82 \times 10^{-1^{+2.7E-1}_{-2.0E-1}}$	$7.10 \times 10^{39} \pm 1.4E40$	$1.85 \times 10^{39} \pm 1.4E39$
2XMM J160531.8+174825 *	b; f; i	241.382700	17.807276	241.382749	17.807222	0.245	$1.64 \times 10^{9^{+3.5E8}_{-3.5E8}}$	$1.53 \times 10^{-1^{+2.5E-1}_{-2.3E-1}}$	$6.91 \times 10^{40} \pm 3.2E40$	$7.73 \times 10^{38} \pm 3.8E38$
AGC 262533 *	i	241.712550	8.157967	241.712419	8.158277	1.218	$1.64 \times 10^{8^{+2.5E7}_{-2.0E7}}$	$4.94 \times 10^{-2^{+1.8E-1}_{-1.4E-1}}$	$3.16 \times 10^{39} \pm 9.4E38$	$1.53 \times 10^{38} \pm 7.7E37$
SDSS J161321.26+510534.8		243.338610	51.093000	243.340961	51.094039	6.507	$1.57 \times 10^{9^{+2.2E8}_{-1.8E8}}$	$6.25 \times 10^{-1^{+2.8E-1}_{-2.4E-1}}$	$1.09 \times 10^{40} \pm 2.0E40$	$1.88 \times 10^{39} \pm 1.6E39$
SDSS J162642.49+390842.8 *		246.677060	39.145226	246.676418	39.145578	2.203	$7.65 \times 10^{8^{+1.7E8}_{-1.0E8}}$	$3.90 \times 10^{-1^{+2.3E-1}_{-1.3E-1}}$	$3.04 \times 10^{40} \pm 1.4E40$	$1.11 \times 10^{39} \pm 6.9E38$
SDSS J162729.77+385455.1		246.874050	38.915330	246.873416	38.915903	2.733	$1.08 \times 10^{9^{+2.4E8}_{-1.7E8}}$	$2.53 \times 10^{-1^{+1.8E-1}_{-1.2E-1}}$	$1.60 \times 10^{40} \pm 1.8E40$	$8.69 \times 10^{38} \pm 3.8E38$
SDSS J213732.54+002800.1		324.385620	0.466721	324.386684	0.466883	3.874	$6.87 \times 10^{8^{+2.0E8}_{-1.5E8}}$	$2.56 \times 10^{-1^{+3.6E-1}_{-3.8E-1}}$	$3.95 \times 10^{39} \pm 8.9E39$	$7.97 \times 10^{38} \pm 9.6E38$
SDSS J213743.69+003125.5		324.432130	0.523779	324.432454	0.523861	1.206	$1.19 \times 10^{9^{+2.7E8}_{-2.1E8}}$	$1.87 \times 10^{-1^{+2.6E-1}_{-2.2E-1}}$	$2.34 \times 10^{40} \pm 2.2E40$	$7.63 \times 10^{38} \pm 4.7E38$
6dFGS gJ233225.3-005049 *		353.105260	-0.847034	353.106172	-0.848164	5.240	$8.26 \times 10^{8^{+1.1E8}_{-8.4E7}}$	$3.65 \times 10^{-1^{+2.6E-1}_{-2.6E-1}}$	$1.35 \times 10^{40} \pm 9.8E39$	$1.05 \times 10^{39} \pm 9.2E38$
SDSS J011421.73+001335.6		18.590588	0.226537	18.590958	0.228235	6.258	$1.92 \times 10^{9^{+3.1E9}_{-4.0E9}}$	$2.71 \times 10^{-2^{+3.6E-3}_{-3.6E-3}}$	$1.59 \times 10^{41} \pm 6.2E41$	$6.63 \times 10^{38} \pm 9.1E38$
SDSS J012325.32-002921.4 *		20.855507	-0.489282	20.854883	-0.488693	3.088	$8.53 \times 10^{8^{+6.3E8}_{-9.3E8}}$	$1.14 \times 10^{-3^{+9.4E-5}_{-9.4E-5}}$	$5.02 \times 10^{39} \pm 1.3E40$	$2.15 \times 10^{38} \pm 1.6E38$
SDSS J030446.14-011208.1 *		46.192287	-1.202276	46.192499	-1.204069	6.501	$1.52 \times 10^{8^{+1.1E8}_{-1.6E8}}$	$4.85 \times 10^{-4^{+3.8E-5}_{-3.8E-5}}$	$5.66 \times 10^{40} \pm 2.6E40$	$3.81 \times 10^{37} \pm 2.8E37$
SDSS J220558.60-003049.3		331.494140	-0.513736	331.495163	-0.513464	3.809	$2.92 \times 10^{9^{+2.3E9}_{-3.4E9}}$	$2.11 \times 10^{-2^{+1.2E-3}_{-1.2E-3}}$	$1.27 \times 10^{41} \pm 8.6E40$	$8.64 \times 10^{38} \pm 6.7E38$
SDSS J234759.26+010344.2		356.996920	1.062305	356.996524	1.061061	4.697	$1.27 \times 10^{9^{+2.5E9}_{-2.9E9}}$	$7.23 \times 10^{-2^{+2.1E-2}_{-2.1E-2}}$	$5.35 \times 10^{42} \pm 5.2E42$	$7.09 \times 10^{38} \pm 8.4E38$

Table B.1: Data for all 61 dwarf galaxies believed to be strong candidates for hosting an AGN.

Appendix C

Other Probability Distributions

Figure C.1 shows the probability distributions split by stellar mass and redshift, and analysing the remaining trends with X-ray luminosity and sBHAR. They have the same form as figure 4.6 and are included for the sake of transparency, to show the strength of our power law fits to the data.

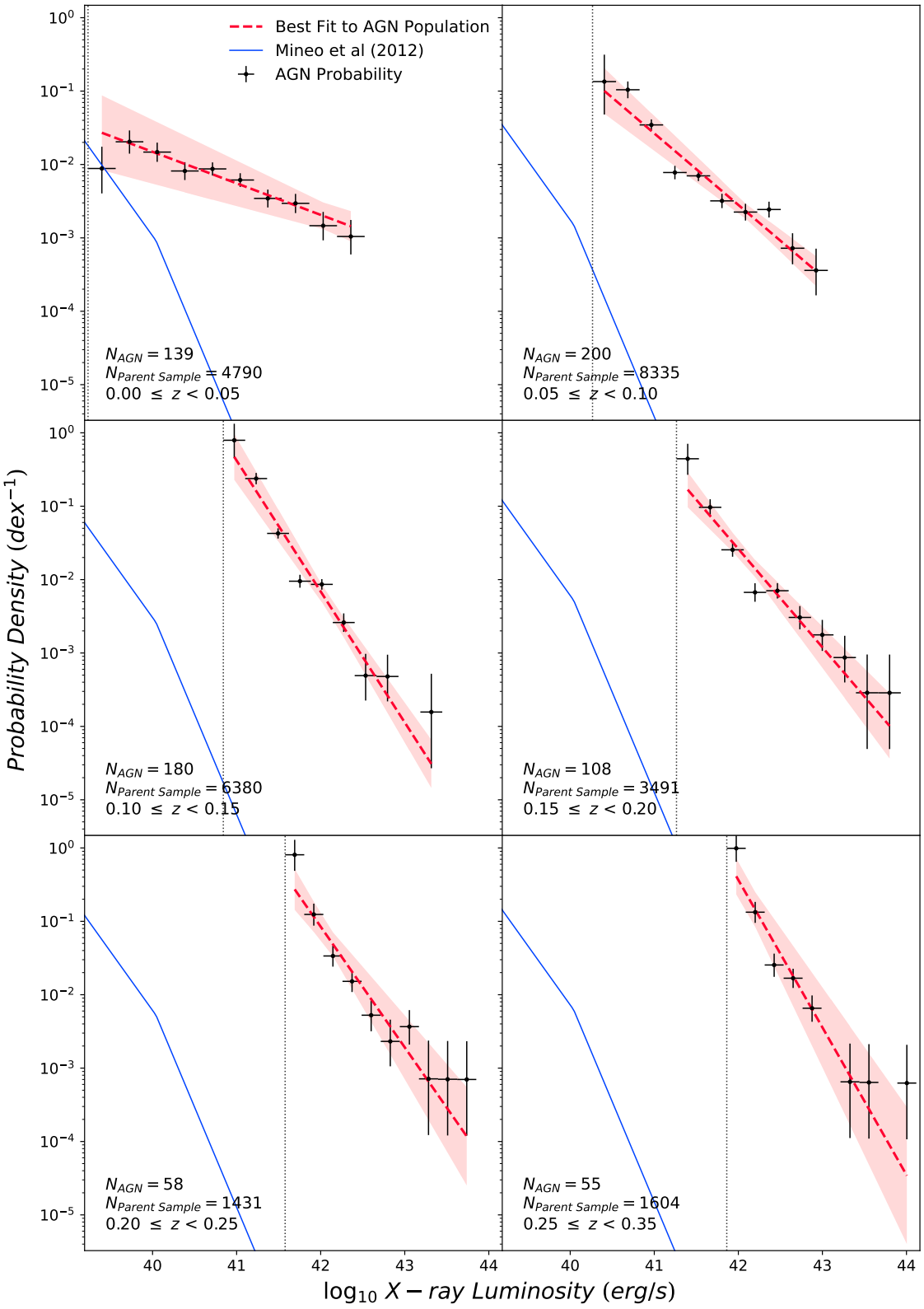

Figure C.1: Additional probability distribution used to calculate fits and associated error regions for the other host galaxy properties. This figure shows the probability of a galaxy hosting an AGN as a function of X-ray luminosity and split into bins of redshift. As with figure 4.6, power laws (dashed red lines) have been fit to the data in each panel and displayed alongside their 1σ uncertainty (pale red region). These plots were constructed using the method outlined in section 4.5

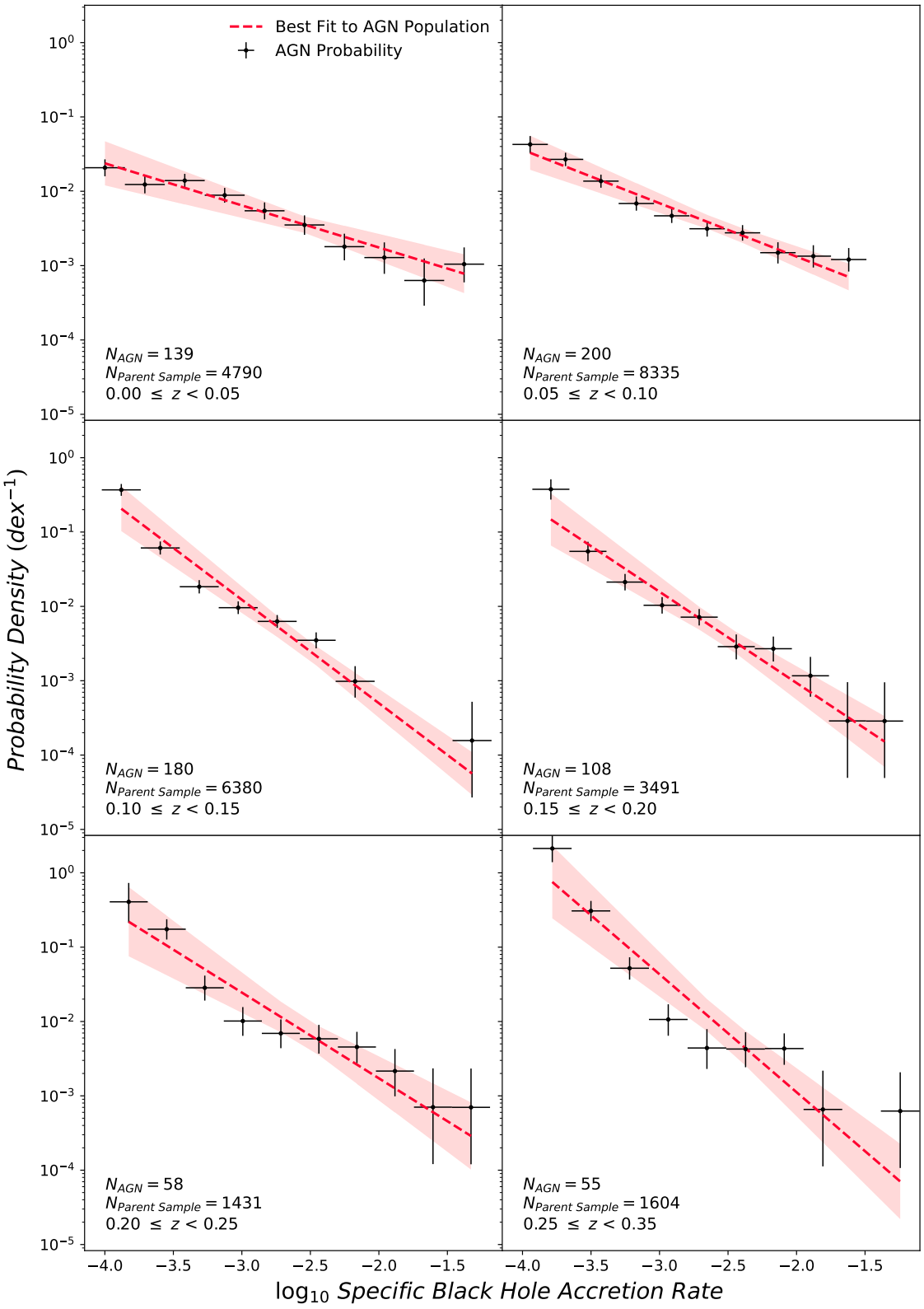

Figure C.1: Additional probability distribution used to calculate fits and associated error regions for the other host galaxy properties. This figure shows the probability of a galaxy hosting an AGN as a function of λ_{sBHAR} and split into bins of redshift. As with figure 4.6, power laws (dashed red lines) have been fit to the data in each panel and displayed alongside their 1σ uncertainty (pale red region). These plots were constructed using the method outlined in section 4.5

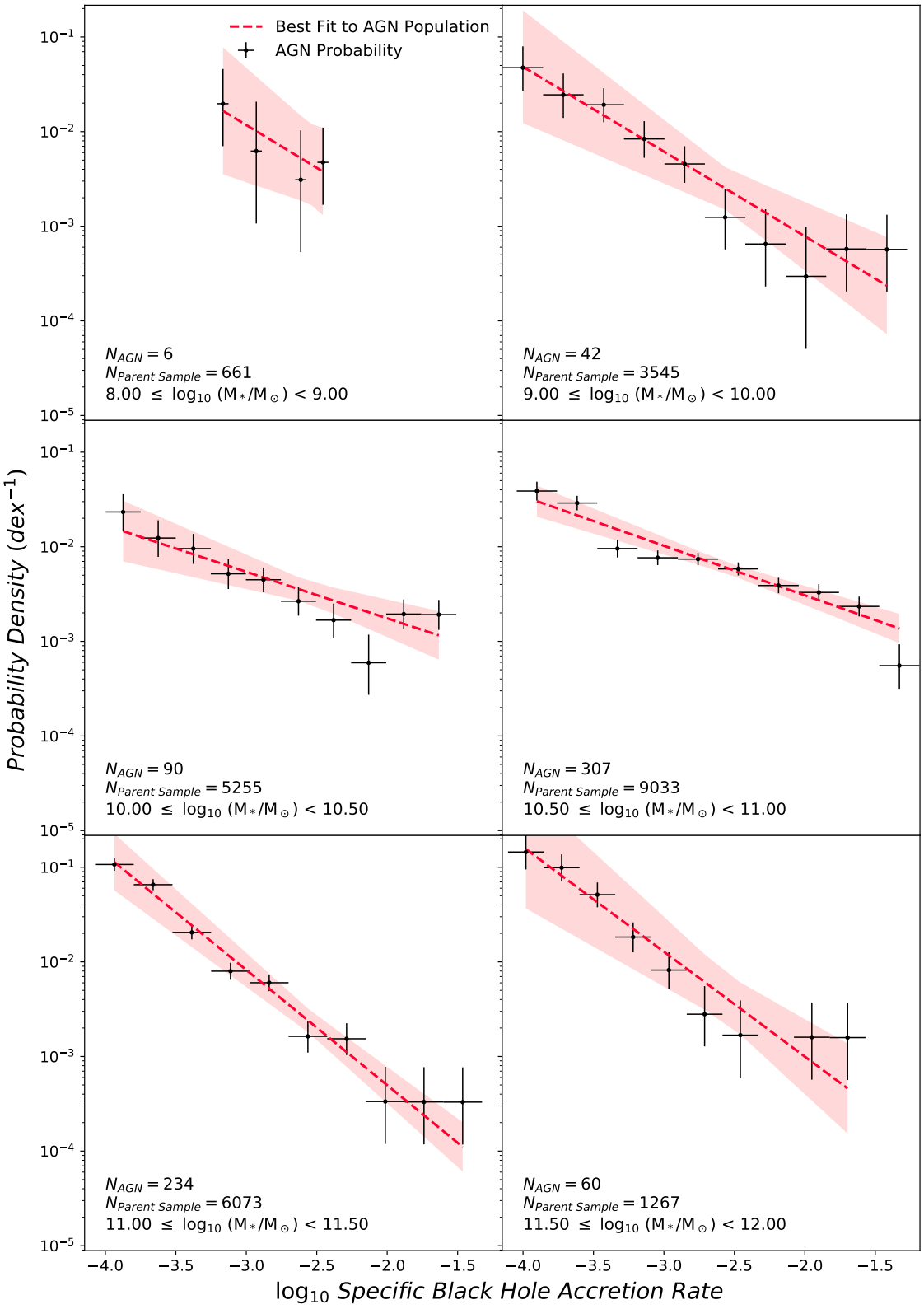

Figure C.1: Additional probability distribution used to calculate fits and associated error regions for the other host galaxy properties. This figure shows the probability of a galaxy hosting an AGN as a function of λ_{sBHAR} and split into bins of stellar mass. As with figure 4.6, power laws (dashed red lines) have been fit to the data in each panel and displayed alongside their 1σ uncertainty (pale red region). These plots were constructed using the method outlined in section 4.5

Appendix D

Probability Distribution Fit Coefficients

Table D.1 outlines the best-fit coefficients, and associated errors, for equation (4.2) to create every probability distribution shown in figure 4.7.

	sBHAR ($\log_{10}\ x' = -2.55$)		X-ray Luminosity (erg s^{-1}) ($\log_{10}\ x' = 42.1$)	
z	$\log_{10}\ A$	k	$\log_{10}\ A$	k
0.00 - 0.05	-2.44 ± 0.11	-0.57 ± 0.13	-2.73 ± 0.13	-0.43 ± 0.12
0.05 - 0.10	-2.48 ± 0.08	-0.72 ± 0.11	-2.65 ± 0.10	-0.97 ± 0.12
0.10 - 0.15	-2.54 ± 0.12	-1.39 ± 0.14	-2.34 ± 0.10	-1.78 ± 0.19
0.15 - 0.20	-2.36 ± 0.13	-1.23 ± 0.18	-1.72 ± 0.10	-1.34 ± 0.20
0.20 - 0.25	-2.13 ± 0.17	-1.15 ± 0.23	-1.24 ± 0.15	-1.64 ± 0.32
0.25 - 0.35	-2.08 ± 0.19	-1.59 ± 0.24	-0.64 ± 0.20	-2.01 ± 0.39
$\log_{10}$ Stellar Mass ($M_\odot$)	$\log_{10}\ A$	k	$\log_{10}\ A$	k
8.00 - 9.00	-2.24 ± 0.61	-0.90 ± 0.01	-3.05 ± 0.00	-0.37 ± 0.43
9.00 - 10.00	-2.61 ± 0.22	-0.90 ± 0.26	-3.24 ± 0.43	-0.74 ± 0.25
10.00 - 10.50	-2.49 ± 0.11	-0.49 ± 0.16	-2.65 ± 0.14	-0.50 ± 0.16
10.50 - 11.00	-2.23 ± 0.06	-0.52 ± 0.08	-2.16 ± 0.05	-0.48 ± 0.08
11.00 - 11.50	-2.63 ± 0.12	-1.22 ± 0.13	-1.91 ± 0.07	-1.07 ± 0.12
11.50 - 12.00	-2.39 ± 0.25	-1.11 ± 0.26	-1.31 ± 0.14	-1.12 ± 0.27

Table D.1: Best-fit coefficients used to fit equation (4.2) to all probability distribution configurations.

Bibliography

Agostino, C. J. and S. Salim (2019). "Crossing the Line: Active Galactic Nuclei in the Star-forming Region of the BPT Diagram". *ApJ* 876.1, 12, p. 12.

Aird, J., A. L. Coil, and A. Georgakakis (2017). "X-rays across the galaxy population - I. Tracing the main sequence of star formation". *MNRAS* 465.3, pp. 3390–3415.

— (2018). "X-rays across the galaxy population - II. The distribution of AGN accretion rates as a function of stellar mass and redshift." *MNRAS* 474, pp. 1225–1249.

— (2019). "X-rays across the galaxy population - III. The incidence of AGN as a function of star formation rate". *MNRAS* 484.3, pp. 4360–4378.

Aird, J. et al. (2010). "The evolution of the hard X-ray luminosity function of AGN". *MNRAS* 401.4, pp. 2531–2551.

Aird, J. et al. (2015). "The evolution of the X-ray luminosity functions of unabsorbed and absorbed AGNs out to z~ 5". *MNRAS* 451.2, pp. 1892–1927.

Aird, J. et al. (2012). "PRIMUS: The Dependence of AGN Accretion on Host Stellar Mass and Color". *ApJ* 746, 90, p. 90.

Aird, J. et al. (2013). "PRIMUS: An Observationally Motivated Model to Connect the Evolution of the Active Galactic Nucleus and Galaxy Populations out to z ~1". *ApJ* 775.1, 41, p. 41.

Alexander, D. M. and R. C. Hickox (2012). "What drives the growth of black holes?" *NewAR* 56.4, pp. 93–121.

Alexander, D. M. et al. (2005). "Rapid growth of black holes in massive star-forming galaxies". *Nature* 434.7034, pp. 738–740.

Avni, Y. and H. Tananbaum (1986). "X-Ray Properties of Optically Selected QSOs". *ApJ* 305, p. 83.

Azadi, M. et al. (2015). "PRIMUS: The Relationship between Star Formation and AGN Accretion". *ApJ* 806.2, 187, p. 187.

Baldassare, V. F. et al. (2017). "X-ray and Ultraviolet Properties of AGNs in Nearby Dwarf Galaxies". *ApJ* 836.1, 20, p. 20.

Baldry, I. K. et al. (2004). "Quantifying the Bimodal Color-Magnitude Distribution of Galaxies". *ApJ* 600.2, pp. 681–694.

Baldwin, J. A., M. M. Phillips, and R. Terlevich (1981). "Classification parameters for the emission-line spectra of extragalactic objects". *PASP* 93, pp. 5–19.

Bañados, E. et al. (2018). "An 800-million-solar-mass black hole in a significantly neutral Universe at a redshift of 7.5". *Nature* 553.7689, pp. 473–476.

Barro, G. et al. (2017). "Structural and Star-forming Relations since z ˜ 3: Connecting Compact Star-forming and Quiescent Galaxies". *ApJ* 840.1, 47, p. 47.

Bauer, A. E. et al. (2013). "Galaxy And Mass Assembly (GAMA): linking star formation histories and stellar mass growth". *MNRAS* 434.1, pp. 209–221.

Beckmann, V. and C. R. Shrader (2012). *Active Galactic Nuclei*.

Begelman, M. C., M. Volonteri, and M. J. Rees (2006). "Formation of supermassive black holes by direct collapse in pre-galactic haloes". *MNRAS* 370.1, pp. 289–298.

Behroozi, P. et al. (2019). "UNIVERSEMACHINE: The correlation between galaxy growth and dark matter halo assembly from z = 0-10". *MNRAS* 488.3, pp. 3143–3194.

Bellovary, J. et al. (2011). "The First Massive Black Hole Seeds and Their Hosts". *ApJ* 742.

Bellovary, J. et al. (2013). "The Relative Role of Galaxy Mergers and Cosmic Flows in Feeding Black Holes". *ApJ* 779.2, 136, p. 136.

Bellovary, J. M. et al. (2019). "Multimessenger signatures of massive black holes in dwarf galaxies". *MNRAS* 482.3, pp. 2913–2923.

Birchall, K. L., M. G. Watson, and J. Aird (2020). "X-ray detected AGN in SDSS dwarf galaxies". *MNRAS* 492.2, pp. 2268–2284.

Blandford, R. D. and R. L. Znajek (1977). "Electromagnetic extraction of energy from Kerr black holes." *MNRAS* 179, pp. 433–456.

Bongiorno, A. et al. (2012). "Accreting supermassive black holes in the COSMOS field and the connection to their host galaxies". *MNRAS* 427.4, pp. 3103–3133.

Bongiorno, A. et al. (2016). "AGN host galaxy mass function in COSMOS. Is AGN feedback responsible for the mass-quenching of galaxies?" *A&A* 588, A78, A78.

Boyle, B. J. and R. J. Terlevich (1998). "The cosmological evolution of the QSO luminosity density and of the star formation rate". *MNRAS* 293.2, pp. L49–L51.

Brammer, G. B. et al. (2011). "The Number Density and Mass Density of Star-forming and Quiescent Galaxies at 0.4 lt;= z lt;= 2.2". *ApJ* 739.1, 24, p. 24.

Brandt, W. N. and D. M. Alexander (2015). "Cosmic X-ray surveys of distant active galaxies. The demographics, physics, and ecology of growing supermassive black holes". *A&ARv* 23.

Brorby, M., P. Kaaret, and A. Prestwich (2014). "X-ray binary formation in low-metallicity blue compact dwarf galaxies". *MNRAS* 441.3, pp. 2346–2353.

Bruzual, G. and S. Charlot (2003). "Stellar population synthesis at the resolution of 2003". *MNRAS* 344, pp. 1000–1028.

Cann, J. M. et al. (2019). "The Limitations of Optical Spectroscopic Diagnostics in Identifying Active Galactic Nuclei in the Low-mass Regime". *ApJ* 870, L2, p. L2.

Carrera, F. J. et al. (2007). "The XMM-Newton serendipitous survey. III. The AXIS X-ray source counts and angular clustering". *A&A* 469.1, pp. 27–46.

Chabrier, G. (2003). "Galactic Stellar and Substellar Initial Mass Function". *PASP* 115.809, pp. 763–795.

Charlot, S. and S. M. Fall (2000). "A Simple Model for the Absorption of Starlight by Dust in Galaxies". *ApJ* 539.2, pp. 718–731.

Chen, C.-T. J. et al. (2013). "A Correlation between Star Formation Rate and Average Black Hole Accretion in Star-forming Galaxies". *ApJ* 773.1, 3, p. 3.

Chevalier, R. A. and A. W. Clegg (1985). "Wind from a starburst galaxy nucleus". *Nature* 317.6032, pp. 44–45.

Cid Fernandes, R. et al. (2010). "Alternative diagnostic diagrams and the 'forgotten' population of weak line galaxies in the SDSS". *MNRAS* 403.2, pp. 1036–1053.

Ciotti, L. and J. P. Ostriker (2007). "Radiative Feedback from Massive Black Holes in Elliptical Galaxies: AGN Flaring and Central Starburst Fueled by Recycled Gas". *ApJ* 665.2, pp. 1038–1056.

Coil, A. L. et al. (2011). "The PRIsm MUlti-object Survey (PRIMUS). I. Survey Overview and Characteristics". *ApJ* 741.1, 8, p. 8.

Cool, R. J. et al. (2013). "The PRIsm MUlti-object Survey (PRIMUS). II. Data Reduction and Redshift Fitting". *ApJ* 767.2, 118, p. 118.

Crummy, J. et al. (2006). "An explanation for the soft X-ray excess in active galactic nuclei". *MNRAS* 365.4, pp. 1067–1081.

Davis, S. W. and A. Laor (2011). "The Radiative Efficiency of Accretion Flows in Individual Active Galactic Nuclei". *ApJ* 728.2, 98, p. 98.

Delvecchio, I. et al. (2014). "Tracing the cosmic growth of supermassive black holes to z ~ 3 with Herschel". *MNRAS* 439.3, pp. 2736–2754.

Delvecchio, I. et al. (2018). "SMBH accretion properties of radio-selected AGN out to z ~ 4". *MNRAS* 481.4, pp. 4971–4983.

Delvecchio, I. et al. (2020). "The Evolving AGN Duty Cycle in Galaxies Since z ~ 3 as Encoded in the X-Ray Luminosity Function". *ApJ* 892.1, 17, p. 17.

den Herder, J. W. et al. (2001). "The Reflection Grating Spectrometer on board XMM-Newton". *A&A* 365, pp. L7–L17.

Di Matteo, T., V. Springel, and L. Hernquist (2005). "Energy input from quasars regulates the growth and activity of black holes and their host galaxies". *Nature* 433.7026, pp. 604–607.

Done, C. et al. (2012). "Intrinsic disc emission and the soft X-ray excess in active galactic nuclei". *MNRAS* 420.3, pp. 1848–1860.

Dressler, A. and D. O. Richstone (1988). "Stellar Dynamics in the Nuclei of M31 and M32: Evidence for Massive Black Holes". *ApJ* 324, p. 701.

Elbaz, D. et al. (2011). "GOODS-Herschel: an infrared main sequence for star-forming galaxies". *A&A* 533, A119, A119.

Fabbiano, G. (2006). "Populations of X-Ray Sources in Galaxies". *ARA&A* 44.1, pp. 323–366.

Fabbiano, G., A. Zezas, and S. S. Murray (2001). "Chandra Observations of "The Antennae" Galaxies (NGC 4038/9)". *ApJ* 554.2, pp. 1035–1043.

Fabian, A. C. (2012). "Observational Evidence of Active Galactic Nuclei Feedback". *ARA&A* 50, pp. 455–489.

Fabian, A. C. and G. Miniutti (2005). "The X-ray spectra of accreting Kerr black holes". *arXiv e-prints*, astro-ph/0507409, astro–ph/0507409.

Ferrarese, L. and D. Merritt (2000). "A Fundamental Relation between Supermassive Black Holes and Their Host Galaxies". *ApJL* 539.1, pp. L9–L12.

Garcıa-Burillo, S. et al. (2016). "ALMA Resolves the Torus of NGC 1068: Continuum and Molecular Line Emission". *ApJL* 823.1, L12, p. L12.

Gebhardt, K. et al. (2000). "A Relationship between Nuclear Black Hole Mass and Galaxy Velocity Dispersion". *ApJL* 539.1, pp. L13–L16.

Gehrels, N. (1986). "Confidence Limits for Small Numbers of Events in Astrophysical Data". *ApJ* 303, p. 336.

Georgakakis, A. et al. (2014). "Investigating evidence for different black hole accretion modes since redshift z ˜ 1". *MNRAS* 440.1, pp. 339–352.

Georgakakis, A. et al. (2017). "Observational constraints on the specific accretion-rate distribution of X-ray-selected AGNs". *MNRAS* 471.2, pp. 1976–2001.

Gibson, R. R., W. N. Brandt, and D. P. Schneider (2008). "Are Optically Selected Quasars Universally X-Ray Luminous? X-Ray-UV Relations in Sloan Digital Sky Survey Quasars". *ApJ* 685.2, pp. 773–786.

Greene, J. E. (2012). "Low-mass black holes as the remnants of primordial black hole formation". *Nature Communications* 3, 1304, p. 1304.

Greene, J. E. et al. (2020). "The Role of Active Galactic Nuclei in the Quenching of Massive Galaxies in the SQuIGG ěc{L} E Survey". *ApJL* 899.1, L9, p. L9.

Grimes, J. P. et al. (2005). "A Chandra X-Ray Investigation of the Violent Interstellar Medium: From Dwarf Starbursts to Ultraluminous Infrared Galaxies". *ApJ* 628.1, pp. 187–204.

Gunn, J. E. et al. (2006). "The 2.5 m Telescope of the Sloan Digital Sky Survey". *AJ* 131.4, pp. 2332–2359.

Haardt, F. and L. Maraschi (1991). "A Two-Phase Model for the X-Ray Emission from Seyfert Galaxies". *ApJL* 380, p. L51.

Haardt, F., L. Maraschi, and G. Ghisellini (1994). "A Model for the X-Ray and Ultraviolet Emission from Seyfert Galaxies and Galactic Black Holes". *ApJL* 432, p. L95.

Haggard, D. et al. (2010). "The Field X-ray AGN Fraction to z = 0.7 from the Chandra Multiwavelength Project and the Sloan Digital Sky Survey". *ApJ* 723.2, pp. 1447–1468.

Harrison, C. M. et al. (2012). "No Clear Submillimeter Signature of Suppressed Star Formation among X-Ray Luminous Active Galactic Nuclei". *ApJL* 760.1, L15, p. L15.

Hasinger, G. et al. (2001). "XMM-Newton observation of the Lockman Hole. I. The X-ray data". *A&A* 365, pp. L45–L50.

Hernán-Caballero, A. et al. (2014). "Higher prevalence of X-ray selected AGN in intermediate-age galaxies up to z ~ 1". *MNRAS* 443.4, pp. 3538–3549.

Hickox, R. C. et al. (2014). "Black Hole Variability and the Star Formation-Active Galactic Nucleus Connection: Do All Star-forming Galaxies Host an Active Galactic Nucleus?" *ApJ* 782.1, 9, p. 9.

Ivezić, Ž. et al. (2002). "Optical and Radio Properties of Extragalactic Sources Observed by the FIRST Survey and the Sloan Digital Sky Survey". *AJ* 124.5, pp. 2364–2400.

Jones, M. L. et al. (2016). "The Intrinsic Eddington Ratio Distribution of Active Galactic Nuclei in Star-forming Galaxies from the Sloan Digital Sky Survey". *ApJ* 826.1, 12, p. 12.

Jones, M. L. et al. (2017). "Do You See What I See? Exploring the Consequences of Luminosity Limits in Black Hole-Galaxy Evolution Studies". *ApJ* 843.2, 125, p. 125.

Kauffmann, G. et al. (2003a). "Stellar masses and star formation histories for 10^5 galaxies from the Sloan Digital Sky Survey". *MNRAS* 341, pp. 33–53.

Kauffmann, G. et al. (2003b). "The host galaxies of active galactic nuclei". *MNRAS* 346, pp. 1055–1077.

Kauffmann, G. and T. M. Heckman (2009). "Feast and Famine: regulation of black hole growth in low-redshift galaxies". *MNRAS* 397.1, pp. 135–147.

Kawasaki, K. et al. (2017). "Active Galactic Nuclei with a Low-metallicity Narrow-line Region". *ApJ* 842.1, 44, p. 44.

Kennicutt, R. C. and N. J. Evans (2012). "Star Formation in the Milky Way and Nearby Galaxies". *Annual Review of Astronomy and Astrophysics* 50, pp. 531–608.

Kewley, L. J. et al. (2001). "Theoretical Modeling of Starburst Galaxies". *ApJ* 556, pp. 121–140.

King, A. R. et al. (2001). "Ultraluminous X-Ray Sources in External Galaxies". *ApJL* 552.2, pp. L109–L112.

Koratkar, A. and O. Blaes (1999). "The Ultraviolet and Optical Continuum Emission in Active Galactic Nuclei: The Status of Accretion Disks". *PASP* 111.755, pp. 1–30.

Kormendy, J. and L. C. Ho (2013). "Coevolution (Or Not) of Supermassive Black Holes and Host Galaxies". *Annual Review of Astronomy and Astrophysics* 51, pp. 511–653.

Kormendy, J. and D. Richstone (1992). "Evidence for a Supermassive Black Hole in NGC 3115". *ApJ* 393, p. 559.

Kroupa, P. (2001). "The Local Stellar Initial Mass Function". In: *Dynamics of Star Clusters and the Milky Way*. Ed. by S. Deiters et al. Vol. 228. Astronomical Society of the Pacific Conference Series, p. 187.

Latif, M. A. and A. Ferrara (2016). "Formation of Supermassive Black Hole Seeds". *Publications of the Astronomical Society of Australia* 33, e051.

Lehmer, B. D. et al. (2016). "The Evolution of Normal Galaxy X-Ray Emission through Cosmic History: Constraints from the 6 MS Chandra Deep Field-South". *ApJ* 825, 7, p. 7.

Lehmer, B. D. et al. (2019). "X-Ray Binary Luminosity Function Scaling Relations for Local Galaxies Based on Subgalactic Modeling". *ApJS* 243.1, 3, p. 3.

Lemons, S. M. et al. (2015). "An X-Ray Selected Sample of Candidate Black Holes in Dwarf Galaxies". *ApJ* 805.

Li, J.-T. and Q. D. Wang (2013). "Chandra survey of nearby highly inclined disc galaxies - I. X-ray measurements of galactic coronae". *MNRAS* 428.3, pp. 2085–2108.

Lintott, C. et al. (2011). "Galaxy Zoo 1: data release of morphological classifications for nearly 900 000 galaxies". *MNRAS* 410, pp. 166–178.

López-Gonzaga, N. et al. (2016). "Mid-infrared interferometry of 23 AGN tori: On the significance of polar-elongated emission". *A&A* 591, A47, A47.

Lusso, E. and G. Risaliti (2016). "The Tight Relation between X-Ray and Ultraviolet Luminosity of Quasars". *ApJ* 819, 154, p. 154.

Madau, P. and M. Dickinson (2014). "Cosmic Star-Formation History". *ARA&A* 52, pp. 415–486.

Madau, P., F. Haardt, and M. Dotti (2014). "Super-critical Growth of Massive Black Holes from Stellar-mass Seeds". *ApJ* 784.

Madau, P. and M. J. Rees (2001). "Massive Black Holes as Population III Remnants". *ApJ* 551, pp. L27–L30.

Maeda, F., K. Ohta, and A. Seko (2017). "Evolution of Cosmic Molecular Gas Mass Density from z ~ 0 to z = 1-1.5". *ApJ* 835.2, 120, p. 120.

Magorrian, J. et al. (1998). "The Demography of Massive Dark Objects in Galaxy Centers". *AJ* 115.6, pp. 2285–2305.

Markowitz, A. G., M. Krumpe, and R. Nikutta (2014). "First X-ray-based statistical tests for clumpy-torus models: eclipse events from 230 years of monitoring of Seyfert AGN". *MNRAS* 439.2, pp. 1403–1458.

Marleau, F. R. et al. (2017). "Infrared signature of active massive black holes in nearby dwarf galaxies". *A&A* 602, A28, A28.

Martin, D. C. et al. (2005). "The Galaxy Evolution Explorer: A Space Ultraviolet Survey Mission". *ApJL* 619.1, pp. L1–L6.

Martin, D. C. et al. (2007). "The UV-Optical Galaxy Color-Magnitude Diagram. III. Constraints on Evolution from the Blue to the Red Sequence". *ApJS* 173.2, pp. 342–356.

Mason, K. O. et al. (2001). "The XMM-Newton optical/UV monitor telescope". *A&A* 365, pp. L36–L44.

Matt, G., G. C. Perola, and L. Piro (1991). "The iron line and high energy bump as X-ray signatures of cold matter in Seyfert 1 galaxies." *A&A* 247, p. 25.

Mendez, A. J. et al. (2013). "PRIMUS: Infrared and X-Ray AGN Selection Techniques at 0.2 lt; z lt; 1.2". *ApJ* 770.1, 40, p. 40.

Merloni, A. et al. (2012). "eROSITA Science Book: Mapping the Structure of the Energetic Universe". *arXiv e-prints*, arXiv:1209.3114, arXiv:1209.3114.

Mezcua, M. et al. (2018). "Intermediate-mass black holes in dwarf galaxies out to redshift âˆ¼2.4 in the Chandra COSMOS-Legacy Survey". *MNRAS* 478.2, pp. 2576–2591. ISSN: 0035-8711.

Mezcua, M. et al. (2016). "A Population of Intermediate-mass Black Holes in Dwarf Starburst Galaxies Up to Redshift=1.5". *ApJ* 817.

Miller, J. M., A. C. Fabian, and M. C. Miller (2004). "A Comparison of Intermediate-Mass Black Hole Candidate Ultraluminous X-Ray Sources and Stellar-Mass Black Holes". *ApJL* 614.2, pp. L117–L120.

Miller, M. C. and M. B. Davies (2012). "An Upper Limit to the Velocity Dispersion of Relaxed Stellar Systems without Massive Black Holes". *ApJ* 755.1, 81, p. 81.

Mineo, S., M. Gilfanov, and R. Sunyaev (2012a). "X-ray emission from star-forming galaxies - I. High-mass X-ray binaries". *MNRAS* 419.3, pp. 2095–2115.

— (2012b). "X-ray emission from star-forming galaxies - II. Hot interstellarmedium". *MNRAS* 426, pp. 1870–1883.

Mineo, S. et al. (2014). "X-ray emission from star-forming galaxies - III. Calibration of the L_X-SFR relation up to redshift z ≈ 1.3". *MNRAS* 437.2, pp. 1698–1707.

Miniutti, G. and A. C. Fabian (2004). "A light bending model for the X-ray temporal and spectral properties of accreting black holes". *MNRAS* 349.4, pp. 1435–1448.

Moran, E. C., A. V. Filippenko, and R. Chornock (2002). ""Hidden" Seyfert 2 Galaxies and the X-Ray Background". *ApJL* 579.2, pp. L71–L74.

Mortlock, D. J. et al. (2011). "A luminous quasar at a redshift of z = 7.085". *Nature* 474.7353, pp. 616–619.

Moustakas, J. et al. (2013). "PRIMUS: Constraints on Star Formation Quenching and Galaxy Merging, and the Evolution of the Stellar Mass Function from z = 0-1". *ApJ* 767.1, 50, p. 50.

Mullaney, J. R. et al. (2010). "Characterizing the far-infrared properties of distant X-ray detected AGNs: evidence for evolution in the infrared-X-ray luminosity ratio". *MNRAS* 401.2, pp. 995–1012.

Mullaney, J. R. et al. (2012a). "GOODS-Herschel: the far-infrared view of star formation in active galactic nucleus host galaxies since z ≈ 3". *MNRAS* 419.1, pp. 95–115.

Mullaney, J. R. et al. (2012b). "The Hidden "AGN Main Sequence": Evidence for a Universal Black Hole Accretion to Star Formation Rate Ratio since z ~2 Producing an M $_{BH}$-M $_{*}$ Relation". *ApJL* 753.2, L30, p. L30.

Müller Sánchez, F. et al. (2009). "Molecular Gas Streamers Feeding and Obscuring the Active Nucleus of NGC 1068". *ApJ* 691.1, pp. 749–759.

Mushotzky, R. (2004). "How are AGN Found?" In: *Supermassive Black Holes in the Distant Universe*. Ed. by A. J. Barger. Vol. 308, p. 53.

Muzzin, A. et al. (2013). "The Evolution of the Stellar Mass Functions of Star-forming and Quiescent Galaxies to z = 4 from the COSMOS/UltraVISTA Survey". *ApJ* 777.1, 18, p. 18.

Nakamura, F. and M. Umemura (2001). "On the Initial Mass Function of Population III Stars". *ApJ* 548.1, pp. 19–32.

Nandra, K. et al. (1989). "Detection of iron features in the X-ray spectrum of the Seyfert I galaxy MCG -6-30-15." *MNRAS* 236, 39P–46.

Netzer, H. (2013). "The Physics and Evolution of Active Galactic Nuclei". In: chap. Main Components of AGNs.

Noeske, K. G. et al. (2007). "Star Formation in AEGIS Field Galaxies since z=1.1: The Dominance of Gradually Declining Star Formation, and the Main Sequence of Star-forming Galaxies". *ApJL* 660.1, pp. L43–L46.

Nucita, A. A. et al. (2017). "A Catalog Sample of Low-mass Galaxies Observed in X-Rays with Central Candidate Black Holes". *ApJ* 837.1, 66, p. 66.

Ostriker, J. P. and N. Y. Gnedin (1996). "Reheating of the Universe and Population III". *ApJL* 472, p. L63.

Paggi, A. et al. (2016). "Hidden Active Galactic Nuclei in Early-type Galaxies". *ApJ* 823.

Pardo, K. et al. (2016). "X-Ray Detected Active Galactic Nuclei in Dwarf Galaxies at 0 lt; z lt; 1". *ApJ* 831.

Petrucci, P. .-O. et al. (2013). "Multiwavelength campaign on Mrk 509. XII. Broad band spectral analysis". *A&A* 549, A73, A73.

Petrucci, P. .-O. et al. (2018). "Testing warm Comptonization models for the origin of the soft X-ray excess in AGNs". *A&A* 611, A59, A59.

Pineau, F. .-X. et al. (2017). "Probabilistic multi-catalogue positional cross-match". *A&A* 597, A89, A89.

Plotkin, R. M. et al. (2016). "The X-Ray Properties of Million Solar Mass Black Holes". *ApJ* 825.2, 139, p. 139.

Popping, G. et al. (2012). "An indirect measurement of gas evolution in galaxies at 0.5 lt; z lt; 2.0". *MNRAS* 425.3, pp. 2386–2400.

Pounds, K. A. et al. (1989). "Iron features in the X-ray spectra of three Seyfert galaxies." *MNRAS* 240, pp. 769–783.

Pounds, K. A. et al. (1990). "X-ray reflection from cold matter in the nuclei of active galaxies". *Nature* 344.6262, pp. 132–133.

Rafferty, D. A. et al. (2011). "Supermassive Black Hole Growth in Starburst Galaxies over Cosmic Time: Constraints from the Deepest Chandra Fields". *ApJ* 742.1, 3, p. 3.

Ramos Almeida, C. et al. (2009). "The Infrared Nuclear Emission of Seyfert Galaxies on Parsec Scales: Testing the Clumpy Torus Models". *ApJ* 702.2, pp. 1127–1149.

Ranalli, P., A. Comastri, and G. Setti (2003). "The 2-10 keV luminosity as a Star Formation Rate indicator". *A&A* 399, pp. 39–50.

Reines, A. E. and A. Comastri (2016). "Observational Signatures of High-Redshift Quasars and Local Relics of Black Hole Seeds". *Publications of the Astronomical Society of Australia* 33.

Reines, A. E., J. E. Greene, and M. Geha (2013). "Dwarf Galaxies with Optical Signatures of Active Massive Black Holes". *ApJ* 775.

Reines, A. E. et al. (2011). "An actively accreting massive black hole in the dwarf starburst galaxy Henize2-10". *Nature* 470, pp. 66–68.

Reines, A. E. et al. (2020). "A New Sample of (Wandering) Massive Black Holes in Dwarf Galaxies from High-resolution Radio Observations". *ApJ* 888.1, 36, p. 36.

Rosario, D. J. et al. (2012). "The mean star formation rate of X-ray selected active galaxies and its evolution from z ~2.5: results from PEP-Herschel". *A&A* 545, A45, A45.

Rosario, D. J. et al. (2013). "Nuclear Activity is More Prevalent in Star-forming Galaxies". *ApJ* 771.1, 63, p. 63.

Rosen, S. R. et al. (2016). "The XMM-Newton serendipitous survey. VII. The third XMM-Newton serendipitous source catalogue". *A&A* 590, A1, A1.

Salim, S. et al. (2007). "UV Star Formation Rates in the Local Universe". *ApJS* 173, pp. 267–292.

Salim, S. et al. (2016). "GALEX-SDSS-WISE Legacy Catalog (GSWLC): Star Formation Rates, Stellar Masses, and Dust Attenuations of 700,000 Low-redshift Galaxies". *ApJS* 227.1, 2, p. 2.

Sartori, L. F. et al. (2015). "The search for active black holes in nearby low-mass galaxies using optical and mid-IR data". *MNRAS* 454.4, pp. 3722–3742.

Schawinski, K. et al. (2014). "The green valley is a red herring: Galaxy Zoo reveals two evolutionary pathways towards quenching of star formation in early- and late-type galaxies". *MNRAS* 440.1, pp. 889–907.

Schulze, A., L. Wisotzki, and B. Husemann (2009). "Low redshift AGN in the Hamburg/ESO survey. I. The local AGN luminosity function". *A&A* 507.2, pp. 781–793.

Shakura, N. I. and R. A. Sunyaev (1973). "Reprint of 1973A&A....24..337S. Black holes in binary systems. Observational appearance." *A&A* 500, pp. 33–51.

Shankar, F., D. H. Weinberg, and J. Miralda-Escudé (2009). "Self-Consistent Models of the AGN and Black Hole Populations: Duty Cycles, Accretion Rates, and the Mean Radiative Efficiency". *ApJ* 690.1, pp. 20–41.

Smith, A. and V. Bromm (2019). "Supermassive black holes in the early universe". *Contemporary Physics* 60.2, pp. 111–126.

Soltan, A. (1982). "Masses of quasars." *MNRAS* 200, pp. 115–122.

Strüder, L. et al. (2001). "The European Photon Imaging Camera on XMM-Newton: The pn-CCD camera". *A&A* 365, pp. L18–L26.

Sun, J. and Y. Shen (2015). "Dissecting the Quasar Main Sequence: Insight from Host Galaxy Properties". *ApJL* 804.1, L15, p. L15.

Swartz, D. A. et al. (2004). "The Ultraluminous X-Ray Source Population from the Chandra Archive of Galaxies". *ApJS* 154.2, pp. 519–539.

Tauris, T. M. and E. P. J. van den Heuvel (2006). "Formation and evolution of compact stellar X-ray sources". In: *Compact stellar X-ray sources*. Vol. 39, pp. 623–665.

Tomczak, A. R. et al. (2014). "Galaxy Stellar Mass Functions from ZFOURGE/CANDELS: An Excess of Low-mass Galaxies since z = 2 and the Rapid Buildup of Quiescent Galaxies". *ApJ* 783.2, 85, p. 85.

Tozzi, P. et al. (2006). "X-ray spectral properties of active galactic nuclei in the Chandra Deep Field South". *A&A* 451.2, pp. 457–474.

Turner, M. J. L. et al. (2001). "The European Photon Imaging Camera on XMM-Newton: The MOS cameras". *A&A* 365, pp. L27–L35.

Tyler, K. et al. (2004). "Diffuse X-Ray Emission in Spiral Galaxies". *ApJ* 610.1, pp. 213–225.

Urry, C. M. and P. Padovani (1995). "Unified Schemes for Radio-Loud Active Galactic Nuclei". *PASP* 107, p. 803.

Vanden Berk, D. E. et al. (2001). "Composite Quasar Spectra from the Sloan Digital Sky Survey". *AJ* 122, pp. 549–564.

Véron-Cetty, M.-P. and Véron, P. (2010). "A catalogue of quasars and active nuclei: 13th edition*". *A&A* 518, A10.

Vito, F. et al. (2014). "Black hole accretion preferentially occurs in gas-rich galaxies*". *MNRAS* 441.2, pp. 1059–1065.

Volonteri, M., G. Lodato, and P. Natarajan (2008). "The evolution of massive black hole seeds". *MNRAS* 383.3, pp. 1079–1088.

Wang, T. et al. (2017). "AGN-host connection at 0.5 lt; z lt; 2.5: A rapid evolution of AGN fraction in red galaxies during the last 10 Gyr". *A&A* 601, A63, A63.

Watson, M. G. et al. (2001). "The XMM-Newton Serendipitous Survey. I. The role of XMM-Newton Survey Science Centre". *A&A* 365, pp. L51–L59.

Watson, M. G. et al. (2009). "The XMM-Newton serendipitous survey. V. The Second XMM-Newton serendipitous source catalogue". *A&A* 493, pp. 339–373.

Weigel, A. K. et al. (2017). "AGNs and Their Host Galaxies in the Local Universe: Two Mass-independent Eddington Ratio Distribution Functions Characterize Black Hole Growth". *ApJ* 845.2, 134, p. 134.

White, S. D. M. and M. J. Rees (1978). "Core condensation in heavy halos: a two-stage theory for galaxy formation and clustering." *MNRAS* 183, pp. 341–358.

Williams, W. L. and H. J. A. Röttgering (2015). "Radio-AGN feedback: when the little ones were monsters". *MNRAS* 450.2, pp. 1538–1545.

Xue, Y. Q. et al. (2010). "Color-Magnitude Relations of Active and Non-active Galaxies in the Chandra Deep Fields: High-redshift Constraints and Stellar-mass Selection Effects". *ApJ* 720.1, pp. 368–391.

Yang, G. et al. (2017). "Black Hole Growth Is Mainly Linked to Host-galaxy Stellar Mass Rather Than Star Formation Rate". *ApJ* 842.2, 72, p. 72.

Yang, G. et al. (2018). "Linking black hole growth with host galaxies: the accretion-stellar mass relation and its cosmic evolution". *MNRAS* 475.2, pp. 1887–1911.

Zdziarski, A. A. et al. (1994). "Physical processes in the X-ray/gamma-ray source of IC 4329A." *MNRAS* 269, pp. L55–L60.