Spectral Energy Distribution Modelling of X-ray Selected AGNs and Their Host Galaxies

Adam Lee Marshall

Institute of Astronomy
University of Cambridge

This dissertation is submitted for the degree of
Doctor of Philosophy

Magdalene College February 2023

Declaration

I hereby declare that except where specific reference is made to the work of others, the contents of this dissertation are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other University. This dissertation is my own work and contains nothing which is the outcome of work done in collaboration with others, except as specified in the text and Acknowledgements. This dissertation contains fewer than 60,000 words, including the abstract, tables, footnotes and appendices.

- Chapters 2 and 3 are based on work published within the journal *Monthly Notices of the Royal Astronomical Society*, Marshall et al. (2022). I performed the analysis and writing of the work included within this paper.

Adam Lee Marshall
February 2023

Abstract

Spectral Energy Distribution Modelling of X-ray Selected AGNs and Their Host Galaxies: Adam Lee Marshall

The nature of the relation between active galactic nuclei (AGN), and their host galaxies have been observed in detail throughout the Universe. Such work has found an intrinsic link between central supermassive black hole (SMBH) masses, and host galaxy properties such as the velocity dispersion of stars, and bulge mass. However, the difference in scale between SMBH and their host galaxies has led to debate on how this relation might form, and develop over time. In order to aid in understanding the relation between AGN and their host galaxies, the work throughout this thesis has therefore focused on the development and implementation of a new SED fitting code, using an up-to-date AGN SED to accurately infer both AGN and host galaxy properties. To this end, we explore the intricacies involved in producing useful property inferences using a Bayesian MCMC fitting method, whilst working to avoid common issues such as bimodality and lack of convergence.

We then perform SED fitting using our methods to 711 luminous X-ray AGN at $0.7 < z < 4.5$ using 10-bands of optical and infra-red photometric data for objects within XMM-SERVS. Using these fits, we study the relation between AGN X-ray luminosity and host galaxy stellar mass, along with our ability to predict emission line strength and morphology from photometry alone. In order to further understand the intricacies of SED fitting, we also provide a case study into the effect of AGN SED choice on host galaxy and AGN property inferences, by comparing our AGN SED to another commonly used template. In this work, we show that it is important to consider host galaxy contamination when trying to produce a pure AGN template, and the effect that this contamination can have on AGN and host galaxy property inferences. We also find that the use of lower resolution SEDs can lead to repercussions on property inferences such as host galaxy stellar mass, which may provide incorrect assumptions on the relation between AGN and their host galaxies.

I would like to dedicate this thesis to my Nan Margaret, who always wanted to see me in a floppy hat.

Acknowledgements

I would like to thank my supervisors Manda Banerji, Matthew Auger-Williams, and Roberto Maiolino for constant help and support throughout my PhD. I'd also like to thank my house-mates throughout my PhD. Jake Bennett, Adam Langeveld, Stephen Thorp, Andy Everall and Matthew Nixon. I won't say you helped keep me sane throughout lockdown, but at least we all went crazy together.

Finally, I'd like to thank my family, for the constant travelling back and forth to Cambridge. Now, whenever you ask me what it is that I do, I can just hand you this thesis.

Table of contents

List of figures **xv**

1 Introduction **1**

1.1 Quasars . . . 1

1.1.1 The Unified AGN model . . . 1

1.1.2 AGN Emission lines . . . 4

1.1.3 The intrinsic link between AGN and host galaxy properties . . . 5

1.1.4 Star formation regulation mechanisms . . . 8

1.2 Multi-wavelength methods of AGN identification . . . 9

1.2.1 Infrared selection . . . 10

1.2.2 Optical selection . . . 11

1.2.3 X-ray selection . . . 12

1.2.4 Spectroscopic selection . . . 12

1.2.5 Future directions and multi-wavelength selection . . . 13

1.2.6 The AGN SED . . . 14

1.3 Introduction to SED fitting codes . . . 15

1.4 Thesis structure . . . 17

2 Development of a new SED fitting code **19**

2.1 Introduction . . . 19

2.2 Method . . . 20

2.2.1 Galaxy Template . . . 20

2.2.2 AGN Template . . . 20

2.2.3 Template combination and magnitude conversion . . . 24

2.3 Overcoming sampling difficulties . . . 24

2.3.1 MCMC Fitting Algorithm . . . 24

2.3.2 Quantifying bimodality and lack of convergence . . . 28

2.3.3 Selecting the best model family 29
2.4 Summary 30

3 Fitting with the XMM-SERVS AGN Sample 33
3.1 Introduction 33
3.2 The XMM-SERVS data sample 35
3.3 XMM-SERVS inferences 38
3.3.1 CFHT *u*-band exclusion 44
3.3.2 AGN Luminosity & Obscuration 45
3.3.3 The stellar mass-AGN luminosity relation 48
3.3.4 AGN emission line properties 50
3.4 Conclusions 55

4 A case study on the effect of AGN template on host galaxy stellar mass inferences 57
4.1 Introduction 57
4.1.1 The R06 template 59
4.1.2 Comparison of the AGN Templates 60
4.2 Comparison of results from the T21 and R06 SED fits 63
4.2.1 Quality of the fits 63
4.2.2 Discrepancies in AGN and host galaxy inferences 65
4.3 Reasons for changes in AGN and host galaxy property inferences 72
4.3.1 AGN emission line effects on AGN and host galaxy properties . . . 73
4.3.2 AGN continuum effects on AGN and host galaxy properties 76
4.3.3 Combined emission line and continuum effects on AGN and Galaxy properties 78
4.4 Comparison of median spectra 79
4.5 Conclusions 81

5 Summary and Future work 85
5.1 Thesis summary 85
5.2 Future work 87
5.2.1 Extension into the UV/MIR 87
5.2.2 Photometric redshift estimates 88
5.2.3 Target selection for future spectroscopic surveys 89
5.2.4 Machine learning 90

References 93

Appendix A Determining mean stellar masses for different populations from MCMC samples 105

List of figures

1.1 A cartoon depiction of the Unified AGN model, showing how the line-of-sight to the AGN can affect the regions that can be observed.
Image credit: http://fermi.gsfc.nasa.gov/science/eteu/agn/ 3

1.2 Correlation between supermassive black hole mass and host galaxy bulge mass, suggesting a co-evolution of AGN and host galaxy properties. The shown relation was found by Magorrian et al. (1998) based on the kinematics of 32 nearby galaxies using the Hubble space telescope photometry and ground-based kinematics. 6

1.3 The luminosity function for galaxies within the Universe from Croton et al. (2006). The black lines show simulated functions both with (solid-line), and without (dashed-line) the effects of AGN feedback included within the simulation. The blue points represent observational data. Feedback from the AGN is therefore required to model the brighter end of the luminosity function as observed within the Universe. 7

1.4 An example AGN SED from Hickox & Alexander (2018) with the individual contributions separated by colour. Here we can see the maximum in AGN flux occurring within the optical-UV region of the spectrum. For comparison, a star-forming galaxy SED is also shown in grey. A maximum in host galaxy flux can be seen at $\sim 1\mu m$. 14

2.1 The range of emission lines properties available in the Temple et al. (2021) AGN template spectrum, as parameterised by the emline_type. Negative values correspond to weaker, more highly blueshifted emission lines, whereas positive values correspond to stronger, more symmetric emission lines. An emline_type of 0 corresponds to the average emission line properties for an SDSS AGN at $z = 2$ with an absolute magnitude $M_i = -27$. 23

2.2 The posterior probability of walkers exploring parameter space for an example run of the MCMC fitting algorithm. Here we can see that around half of the walkers have found a global maxima corresponding to a solution with a average $\log P(\rho|data) \approx -23$. However, the remaining walkers are stuck in a significantly lower $\log P(\rho|data)$ solution. Parallel-tempering optimisation is therefore utilised to allow walkers to escape local maxima, and start closer to the global maximum within our actual MCMC run. 27

2.3 The posterior probability of walkers exploring parameter space for an example run of the MCMC fitting algorithm. Despite an initial parallel-tempering optimisation, two stray walkers have remained unconverged, and not found the global maximum at $\log P(\rho|data) \approx -17.5$. We therefore explore methods of quantifying $\log P(\rho|data)$ distributions with stray walkers, in order to identify unconverged fits from large samples of objects. 27

2.4 Histograms of the samples from two SED fits. The object in blue has converged to a single solution based on its associated corner plot. The orange object has a small number of walkers within a lower $\log P(\rho|data)$ solution, as can be seen by the increase in samples occurring around $\log P(\rho|data) \approx -24$. Whilst both objects have similar peaks in $\log P(\rho|data)$, the distribution of the converged fit samples is wider, despite the presence of unconverged walkers within the orange fit. 28

3.1 The coverage of the XMM-SERVS survey, provided by Chen et al. (2018). Each point represents the position of an X-ray selected object within the sample. Objects for which reliable multi-wavelength counterparts could not be found by Chen+18 are shown in red, and represent 7% of the total sample. These objects were not included within our SED fitting. The multi-wavelength survey footprints of matched optical and infrared data are also shown. 35

3.2 The filter transmission curves including atmospheric absorption and detector quantum efficiencies for the 10 bands of optical and infrared photometry used for SED fitting within this work. These originate from three surveys: HSC (Aihara et al., 2018), VISTA VIDEO (Jarvis et al., 2013) and Spitzer SERVS (Mauduit et al., 2012), each of which has been matched to an X-ray source observed by XMM-Newton (Chen et al., 2018). 37

3.3 Flowchart showing the cuts made to the initial XMM-SERVS parent sample (Chen et al., 2018). The figure is colour-coded to show the objects removed before fitting in red. The green boxes represent the seperation of the remaining objects into our model families, based on a comparison of their AIC values after fitting. The details of these model families, and the AIC designation method used is described in further detail within Sections 2.3.1 and 2.3.3 respectively. 39
3.4 An example of the marginalised one- and two dimensional posterior distributions for an unobscured (HR = -0.67) AGN at $z = 0.93$. In this example, we can see the degeneracies between a young, low stellar mass, star forming galaxy and an older, redder galaxy with a higher luminosity AGN. These degeneracies can be seen in the bimodalities of the host galaxy stellar mass and age, and AGN optical luminosity and extinction. 41
3.5 Example fits for the four model families. The top figures show examples of fits with both the galaxy and AGN components, one with a free emission line (left), and the other with emission line properties fixed to the average seen in an SDSS quasar at $z = 2$ AGN, with an average absolute magnitude $M_i = -27$. The bottom figures show fits where either a AGN-only (left) or Galaxy-only (right) fit is preferred. In each case, the total maximum likelihood spectrum is shown in green. The blue, orange and red spectra represent the AGN, galaxy and AGN hot dust components respectively. The observed photometric data are shown as red points. In each corner is the HSC DR-2 *gri* colour image of the object. 43
3.6 Comparison of the SED fits produced both with (left) and without (right) the inclusion of CFHT u-band band (Veillet, 2007). The blue, orange and red spectra represent the AGN, galaxy and AGN hot dust components respectively. The observed photometric data is shown as red points. The inclusion of the u-band data severely decreases the quality of the fit, as shown by the increase in reduced χ^2, despite the addition of a data point. 44

3.7 XMM-Newton X-ray luminosity vs AGN optical luminosity at 3000Å with the relation derived from Marconi et al. (2004) (left) and the 3000Å optical luminosity vs hot dust luminosity at 3μm with the relation from Jun & Im (2013) (right). The blue data points show the median solutions and uncertainties for these objects, and the contours show the combined inferences from our MCMC analysis. For clarity, uncertainties are limited to a third of the points shown. In the case where X-ray luminosity is plotted, objects where $\log_{10}(L_{2-10keV})$ is an upper limit are shown in grey. 45

3.8 Histograms showing the distribution of E(B-V) (left) and host galaxy stellar mass (right) MCMC inferences, separated by the measured X-ray hardness ratio. This ratio acts as a proxy for the AGN obscuration type. We have assumed values above and below HR = -0.2 to represent obscured and unobscured AGN respectively. The obscured AGN show a tail to higher extinction values when compared to the unobscured sample. 48

3.9 Host galaxy stellar mass MCMC inferences vs the measured X-ray luminosity from XMM-Newton. The blue points correspond to the median solutions for each object. For clarity, uncertainties limited to a third of the objects are shown. For objects where $\log_{10}(L_{2-10keV})$ is an upper limit, median solutions are shown in grey. Contours show all of the MCMC inferences for these objects. 49

3.10 SDSS spectra (in black) compared to the highest likelihood SED, including the host galaxy, AGN and hot dust components (in green) produced using SED fitting. The SDSS spectrum are normalised to the best fit SED to have the same *i*-band flux. The figures show examples of an object that preferred weaker, more blue shifted lines on the left, and stronger, more symmetrical lines on the right, compared to the emission lines seen in AGN at $z = 2$ and an average absolute magnitude of $M_i = -27$. In the left figure, whilst the emission line strength in the best fit SED appears to show good agreement between the width and strength of the lines observed, either AGN variability, or an offset in spectra calibration appears to change the continuum emission between the SDSS spectrum and best fit SED. 51

3.11 The X-ray to UV power law slope, α_{ox}, vs the inferred AGN emission line properties. More negative values of emission line properties are indicative of line driven disc winds resulting in weaker, more blueshifted lines, whereas more positive values are stronger and more symmetric. The sample shown does not include objects that preferred a fixed emission line within their fit, based on our AIC designation (such that the emission line properties value is fixed to 0). 54

4.1 Comparison of the R06 and T21 AGN templates used within this work. Both templates have been normalised to have the same flux at 3000Å. In order to directly compare the R06 and T21 templates, the fixed dust included within the R06 template (shown as a dashed line in orange) has been removed, and a separate template of a 1236K blackbody is re-added. 61

4.2 The change in g-r colour as a function of redshift for a range of smoothing of the Temple et al. (2021) AGN template, via a Gaussian filter. The smoothing is compared to the lower resolution AGN template from Richards et al. (2006), shown in brown. 62

4.3 Comparison of the χ^2_{red} using the R06 and T21 AGN templates, along with the 1-D histograms for each fit. Generally, the T21 template is shown to provide better quality fits for 83% of the sample. 64

4.4 Comparison of the median host galaxy stellar mass produced using the R06 and T21 AGN templates. There is a clear difference in stellar mass inferences, with the R06 template predicting lower stellar masses for 39% of objects. For clarity, uncertainties are limited to a third of the points shown. 66

4.5 Comparison of the median E(B-V) (top), AGN optical luminosity at 3000Å (bottom left) and hot dust luminosity at 3μm (bottom right) inferences for the R06 vs T21 AGN templates, colour-coded by the difference in the stellar mass inferences of the two fits. 67

4.6 2-10keV X-ray luminosity vs AGN optical luminosity at 3000Å with the relation derived from Marconi et al. (2004) (top) and the 3000Å optical luminosity vs hot dust luminosity at 3μm with the relation from Jun & Im (2013) (bottom). The data points show the median solutions and uncertainties for these objects, with the R06 and T21 AGN template fits shown in orange and blue respectively. The contours show all of the inferences from our MCMC analysis. For clarity, uncertainties are limited to a third of the points shown. 68
4.7 The upper and lower uncertainties associated with the AGN optical luminosity at 3000Å for the R06 and T21 template fits. Here we see that the uncertainties are generally higher for the R06. 69
4.8 Comparison of the stellar mass inferences vs observed X-ray luminosity. The data points show the median solutions and uncertainties for these objects, with the R06 and T21 AGN template fits shown in orange and blue respectively. The contours show the total inferences from our MCMC analysis. For clarity, uncertainties are limited to a third of the points shown. The lower mass inferences for the R06 template differ from the positive X-ray luminosity-host galaxy mass relation seen in the T21 template. 70
4.9 Comparison of the distributions of stellar mass inferences for the R06 (orange) and T21 (blue) AGN templates produced using the SED fitting code created for this thesis. Our stellar mass inferences are compared to previous work by Bongiorno et al. (2012) (green) of objects within the COSMOS field, also created via SED fitting using the R06 template. 71
4.10 Comparisons of the T21 (blue) and R06 (orange) stellar mass estimates and associated uncertainties. The x-axis gives the fits produced from a run using a higher minimum uncertainty limit of 0.1 placed on each of the 10-bands of photometry. The y-axis gives the same run with a lower uncertainty limit corresponding to 5% of the flux from the 3.0″ images. The inferred stellar mass estimates are consistent despite the increase in uncertainty. For clarity, uncertainties are limited to a third of the points shown. 72

4.11 The maximum likelihood SED fits produced using the T21 (left) and R06 (right) AGN templates. Here we can see a switch in the contribution from the AGN and host galaxy. Clear emission lines are not seen within the R06 AGN template due to its low resolution. Within the R06 fit, we therefore see the host galaxy change to attempt to compensate for the emission lines, and recreate the J-band photometry. The younger host galaxy inferences are seen to decrease the galaxy stellar mass inferences in the R06 fit. 74
4.12 Marginalised one- and two-dimensional posterior distributions for the SEDs fits shown in Fig. 4.11. The distributions for the T21 and R06 AGN templates are shown in black and red respectively. 75
4.13 The maximum likelihood SED fits produced using the T21 (top left), R06 (top right) AGN+GAL and R06 (bottom-left) AGN-only templates. The disagreement between the fits could be attributed to the difference in continuum for the two AGN templates, specifically affecting the IR bands. The more gradual dropoff at longer wavelengths in the R06 template results in the separate host galaxy contribution not being required for the R06 best fit. 77
4.14 The maximum likelihood SED fits produced using the T21 (left) and R06 (right) AGN templates. For the majority of the sample of discrepant stellar mass fits, the inferred AGN properties for the T21 and R06 fits are consistent, with the galaxy component changing instead to compensate for the lack of significant emission lines within the R06 template. 79
4.15 The median rest frame spectra of the galaxy (orange) and AGN (blue) and combined (green) contributions to the R06 (dashed lines) and T21 (solid lines) SED fits. In the case of both AGN templates, the host dust component is not included in the plot. The median R06 galaxy SED is younger than the T21 template, with higher rates of star formation resulting in prominent emission lines compensating for the lack of emission lines in the R06 AGN SED. 80

5.1 The three regions in which spectra will be collected the using MOONS (Maiolino et al., 2020a), in comparison with the transmissions curves, including atmospheric absorption and detector quantum efficiencies, of the bands used throughout this thesis (HSC, VISTA VIDEO and Spitzer SERVS), along with the CFHT *u*-band 90

Chapter 1

Introduction

1.1 Quasars

Observations of galaxies within our Universe suggest that most contain supermassive black-holes (SMBHs) at their centres, with masses ranging from $10^6 - 10^{10} M_\odot$. The first of these to be discovered was Cygnus A, as a radio bright object observed by Reber (1944). For bright radio objects such as Cygnus A, the emission was initially believed to originate from relatively nearby stars. However, followup spectroscopic observations of one such radio source, designated 3C 273, by Schmidt (1963), revealed the presence of emission lines consistent with the Hydrogen Balmer series at a redshift of $z = 0.158$. Such a redshift implied that the object could not be a star, but a new type of highly luminous ($> 10^{12} L_\odot$) distant object, designated as a quasi-stellar radio object, or Quasar. Our understanding of these objects was further developed by Lynden-Bell (1969), who suggested these luminous, compact 'quasar' objects could occur due to the accretion of mass onto supermasssive black holes at the centres of galaxies. Since these initial discoveries, modern surveys such as the Sloan Digital Sky Survey (SDSS; Lyke et al. (2020)) have continued to identify and increase the sample size of quasars up to $\approx 750,000$, with spectroscopic redshift measurements now recorded out to around $z = 7.6$ (Wang et al., 2021). These observations have allowed us to gain a more detailed understanding of the nature of quasars.

1.1.1 The Unified AGN model

Observations of Quasi-stellar objects (QSOs) since the discovery of the first radio-bright sources have revealed a variety in emission line and continuum flux properties (Padovani et al., 2017). Specifically, for some QSO spectra, emission lines are seen to be stronger, wider, or more asymmetrical. Spectral differences have led to a number of distinct categories of objects, collectively known as active galactic nuclei (AGN). For example, distinctions can be made on the strength of the AGN output in the radio, giving 'radio-loud' and 'radio-quiet' quasars. Examples of objects with extreme variability and radio emission are known as blazars (Falomo et al., 2014). AGN can also be characterised by the properties of emission lines within the UV/optical region of their spectrum, typically known as type-1 and type-2 AGN. With type-2 AGN, we only observe narrow emission lines, whereas in type-1 AGN we may see both narrow- and broad emission lines. In Type-2 AGN spectra, we also typically see clear signatures of a galaxy, which is often obscured by a bright continuum within type-1 AGN.

In order to understand the origins of type-1 and type-2 AGN, observers have tried to theorise what could cause the physical differences between AGN spectra. Such studies have led to the development and investigation of the unified model. The unified model states that the structure of all AGN in the Universe are similar, with the differences in AGN categorisations being a result of the angle at which we view the object (Antonucci, 1993; Urry & Padovani, 1995).

A cartoon depiction of the unified AGN model can be seen in Fig 1.1, which outlines their basic structure. At the centre of this model, we see a supermassive black hole surrounded by gas falling inwards. Due to the angular momentum stored within this gas, a fast moving accretion disc is formed around the black hole. The material within the accretion disc is turbulent and viscous, and as such, will lose energy and thus slowly spiral inwards. As this in-spiralling occurs, the gas will heat up and release gravitational potential energy. The accretion disc is formed as optically thick, but geometrically thin, thus, whilst some of the energy is used to increase the kinetic energy of the material as it decreases in orbit radius, the rest is emitted as a blackbody. Blackbody emission therefore occurs across the range of radii within the accretion disc, leading to an emission that represents the combination of blackbodies at multiple temperatures, with the temperature increasing as the orbital radius decreases. Material within the accretion disc will continue to fall inwards and release energy until the last stable orbit is reached, at which point it will then fall into the black hole. In general, this process is extremely efficient in the release of the energy when compared to alternative methods such as nuclear fusion. In practice, black hole accretion can therefore lead to an energy output of around $0.1\mathrm{M}c^2$ of the rest frame mass.

Due to the reliance on a constant influx of in-falling material in order to produce further energy, we find that there is a upper limit on the luminosity output for any given AGN. If the outwards pressure produced from accreting material is too strong, this will overpower

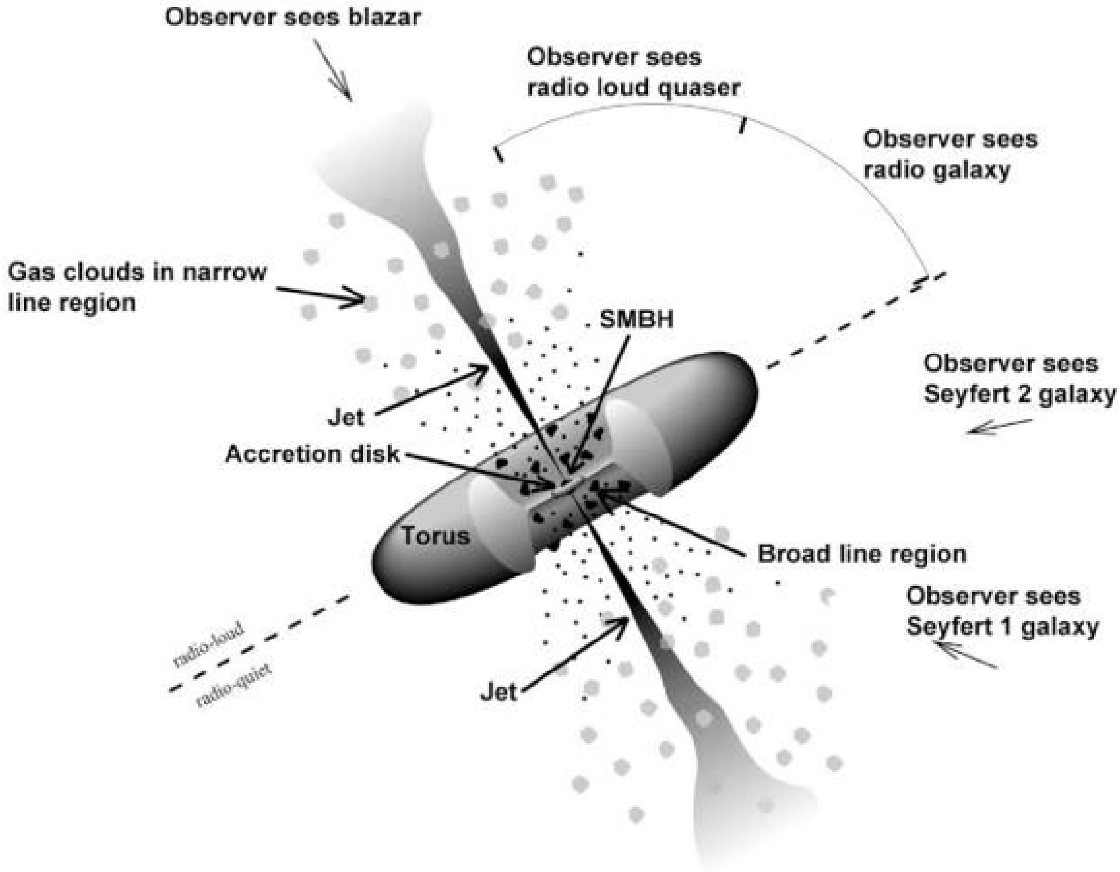

Fig. 1.1 A cartoon depiction of the Unified AGN model, showing how the line-of-sight to the AGN can affect the regions that can be observed.
Image credit: http://fermi.gsfc.nasa.gov/science/eteu/agn/

the force due to gravity, and thus prevent any further material to fall inwards, removing the source of fuel. An upper limit therefore occurs at the point where the outwards radiation pressure is equal to the gravitational force. Such a case can be estimated by assuming that the in-falling, spherically symmetric gas is comprised entirely of ionised hydrogen. The outwards pressure is produced via Thompson scattering by electrons, with a cross section of σ_T, from photons carrying a luminosity L. At the point where this outwards force balances against the gravitational force, we find that:

$$\frac{GM(m_e+m_p)}{r^2}=\frac{\sigma_T L}{4\pi r^2 c} \tag{1.1}$$

As $m_p >> m_e$ we can ignore the m_e as negligible. From this, we can then rearrange to find the Eddington Luminosity, the largest luminosity at which the outwards radiation pressure is equal to the gravitational potential as:

$$L_E=\frac{4\pi GMm_pc}{\sigma_T}\approx 30000\frac{M}{M_\odot}L_\odot \tag{1.2}$$

This equation can provide a good indication of the central supermassive black hole masses based on the typical observed luminosities of AGN in the Universe. For example, for an object with a luminosity of L = $10^{12}L_\odot$, we can predict a black hole mass of around $10^8 M_\odot$. Beyond the accretion disc in Fig. 1.1, we can see that a thick, dusty torus is formed, obscuring the inner material from view when observed edge-on. An additional effect of the inflow of material towards the black hole is the production of large magnetic fields. As these fields increase in strength, they can become strong enough to channel twin jets of high energy outflows. For an observer looking straight down on of these jets, the object AGN is viewed as a high energy blazar. Finally, at a greater distances (50-100 pc scales) we see clumps of cooler gas clouds.

1.1.2 AGN Emission lines

In terms of observed spectra, we are able to observe broad emission lines when there is a line of sight to the central region of the AGN, known as the broad line region (BLR). Here we see rapidly circling material close to the supermassive black hole, with velocities of the order of 10,000kms^{-1} (Peterson, 1997). Due to the presence of orbiting material moving away, and towards, the observer, there is both red- and blue-shifting of the emission lines within the heated gas of the BLR. The movement of the heated gas therefore has the effect of broadening the observed emission lines, leading to the type-1 spectra we observe in AGN. However, alternate models for type-1 spectra have also been suggested. For example, Matthews et al. (2020) suggest that the observed emission lines are a result of disc winds, in which material is blown out due to radiation pressure from the central SMBH, producing similar broadening of emission lines that would be expected from rapidly circling clouds within the BLR.

In the case of type-2 objects, it is believed that a dusty torus obscures the inner accretion disc. However, the exact morphology of the surrounding material has been questioned, with some work also suggesting evidence for polar dusty winds within the AGN structure (Hönig, 2019). In either case, an observer of a type-2 object is able to see cooler, slower moving gas at radii greater than the BLR, produced within what is therefore known as the narrow line region (NLR) (Padovani et al., 2017). This NLR is also within the line of sight of an observer for type-1 AGN. We therefore see narrow emission lines in both type-1 and type-2 AGN. However, the obscuration of the BLR by the dusty torus means that no broad lines are observed within the spectra of type-2 AGN.

In other cases, such as in the observations of blazars, radio loud AGN originate from electrons spiralling within strong magnetic fields produced within the disc of the accreting material. These electrons release concentrated radio jets via synchrotron emission, which are kept narrow by the strong magnetic fields, and thus only observed within specific line-of-sight orientations.

1.1.3 The intrinsic link between AGN and host galaxy properties

Despite the relatively small size of the central supermassive black hole (SMBH) compared to the host galaxy, its presence has been noted to play a significant role in the evolution of the combined system (Kormendy & Ho, 2013; Magorrian et al., 1998; Merritt & Ferrarese, 2001). Observations of nearby inactive galaxies show an intrinsic link between the host galaxy stellar bulge mass, and the mass of a SMBH (Gültekin et al., 2009; Häring & Rix, 2004; Magorrian et al., 1998; Marconi & Hunt, 2003). Such a link appears to suggest a coupling between the galaxy and SMBH developing at some point in the history of the system, and can be seen in Fig. 1.2. Work by Madau et al. (1998); Richards et al. (2006) have also shown an overlap in the time between peaks in star formation and AGN accretion rates within the Universe at z $\sim$ 2-3, also suggesting that these two phenomena are intertwined. However, the processes by which host galaxy and SMBH properties become linked is still not fully understood, and is an active area of study within the field of extra galactic astronomy.

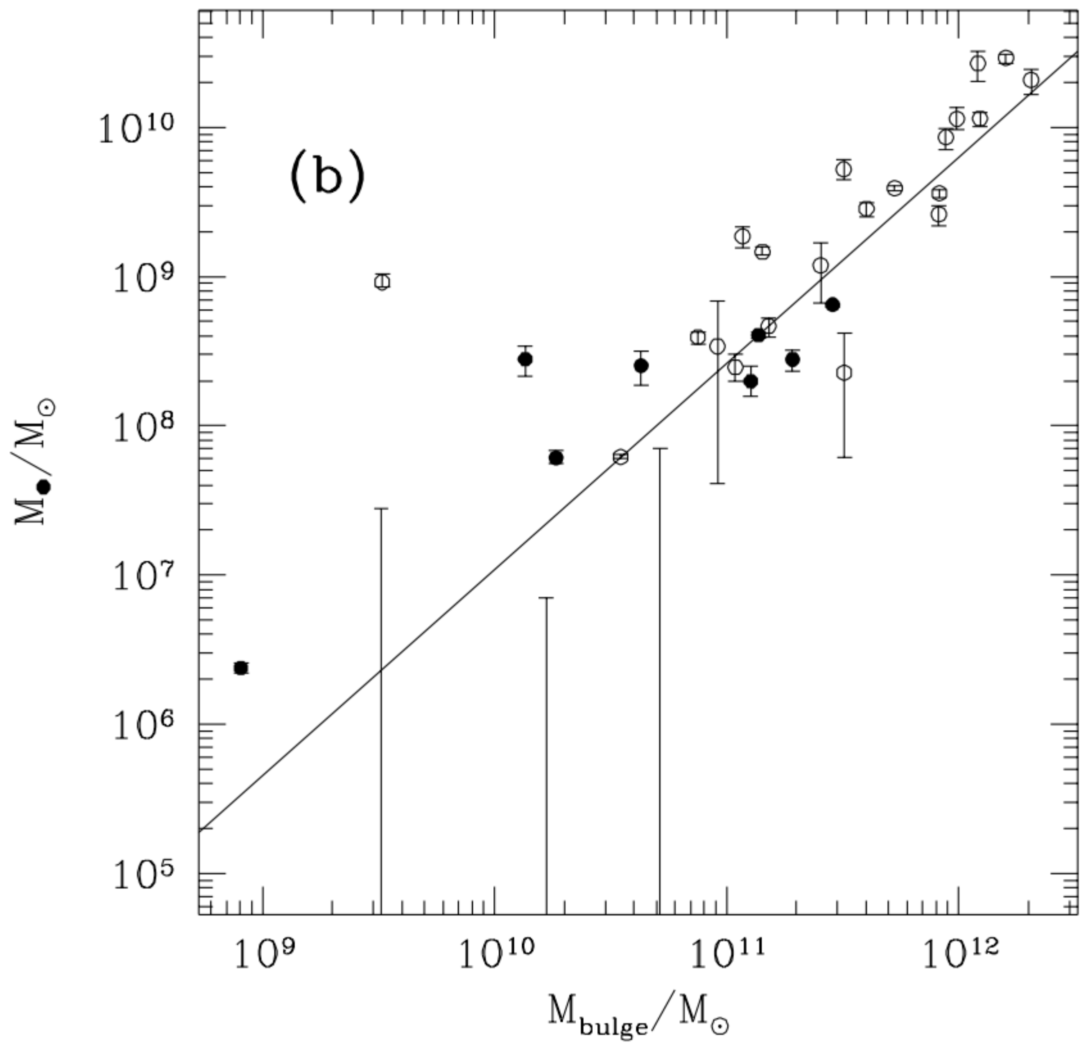

Fig. 1.2 Correlation between supermassive black hole mass and host galaxy bulge mass, suggesting a co-evolution of AGN and host galaxy properties. The shown relation was found by Magorrian et al. (1998) based on the kinematics of 32 nearby galaxies using the Hubble space telescope photometry and ground-based kinematics.

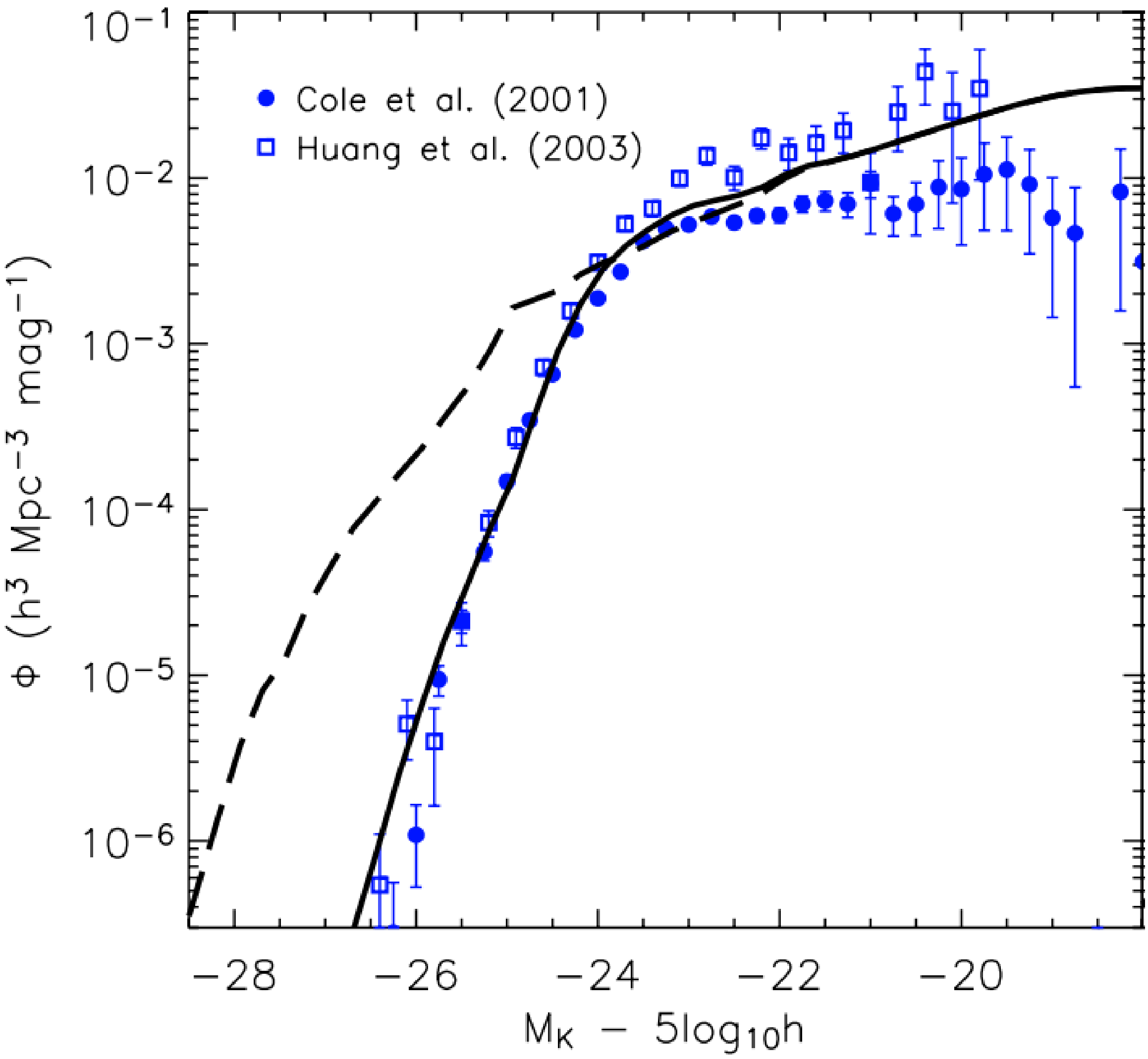

Fig. 1.3 The luminosity function for galaxies within the Universe from Croton et al. (2006). The black lines show simulated functions both with (solid-line), and without (dashed-line) the effects of AGN feedback included within the simulation. The blue points represent observational data. Feedback from the AGN is therefore required to model the brighter end of the luminosity function as observed within the Universe.

1.1.4 Star formation regulation mechanisms

A number of theories aimed at explaining the intrinsic link between SMBH and host galaxy properties have been proposed. Many simulations have shown that the 'feedback' produced by the AGN is required to accurately model observations. Simulations by Croton et al. (2006), have aimed to accurately model the luminosity function of galaxies in the Universe. Versions of the simulation both with and without the effects AGN feedback on the luminosity function were modelled. Comparisons to observations, as shown in Fig. 1.3, showed that the presence of this feedback in the simulation was key to recreate the observations of the bright end of luminosity function seen in the Universe.

Similarly, in both hydrodynamical (Bennett & Sijacki, 2022; Dubois et al., 2016; Habouzit et al., 2022; Scannapieco et al., 2012; Sijacki et al., 2015) and semi-analytical (Bower et al., 2006; Granato et al., 2004) models, AGN feedback is identified as key feature required to model galaxy evolution. Work in explaining AGN feedback mechanisms has suggested two contributing methods can occur depending on the 'mode' of AGN. These modes have been categorised as the 'transformative mode' (or 'quasar-mode') and 'maintenance mode' (or 'radio mode' Bower et al. (2006); Croton et al. (2006)) with highly luminous AGN typically displaying properties associated with the former category. In the case of both modes, the presence of the AGN can have the effect of quenching further star formation within the host galaxy, in what is known as negative AGN feedback (Fabian, 2012; Morganti, 2017).

In the case of a transformative mode AGN, high rates of accretion give rise to high-velocity winds that can hinder star formation by removing the surrounding material within the inter-stellar medium (Veilleux et al., 2005). The effects of these winds have been observed within galaxies hosting AGN through the presence of blue- and red-shifted wings on ionised emission lines (Cicone et al., 2014; Lyke et al., 2020; Müller-Sánchez et al., 2011). Removal of surrounding material by the AGN winds has the additional effect of removing the source of material for continuing accretion onto the SMBH. The outcome of an AGN in transformative mode would therefore be to halt star formation, fixing the host galaxy stellar mass, whilst simultaneously slowing further SMBH growth due to lack of material for accretion, thus explaining the origin of a link between these properties.

The maintenance mode AGN, which are often radio-loud, alternatively have the effect of heating any remaining material, such that gas within the galaxy is not able to cool sufficiently enough to form stars (Barišić et al., 2017; Best et al., 2005; McNamara & Nulsen, 2007). Maintenance mode AGN may therefore be seen in the aftermath of an AGN that has already gone through a transformative mode, and, as the name suggests, continue to enforce the host galaxy bulge-SMBH mass correlation.

The effect of AGN feedback on host galaxies is however still debated. For example, Ishibashi & Fabian (2012) suggests that instead of AGN winds shutting off star formation, they can trigger higher star formation rates at larger galactic radii (Shin et al., 2019). Initially, a central core would be produced for such a galaxy, this would be followed by a build-up of outer regions as radiation pressure from the accreting SMBH pushes on the surrounding gas and dust. The increased density within the outflowing regions would then lead to increased star formation rates via positive feedback (Bicknell et al., 2000; Dey et al., 1997).

It is also important to consider phenomena other than AGN feedback that may have a significant effect on host galaxy properties, including star formation rates. One such example is that of major and minor mergers (Jahnke & Macciò, 2011; Moreno et al., 2019; Peng et al., 2006; Sanders et al., 1988). Mergers can occur both between two initially compact galaxies of equal sizes, or through the accretion of a number of smaller satellites onto a larger central galaxy. The interactions between these galaxies will produce regions with higher densities of gas and dust, restarting active star formation. Further studies have also suggested a link between mergers and AGN activity (Davies et al., 2022; Ellison et al., 2013; Gao et al., 2020) due to the increase in dust and gas leading to greater accretion onto the SMBH. In addition to both mergers and AGN being contributing factors to galaxy evolution, the two phenomena may also therefore be intrinsically linked, with the star formation rates affected by the merger sharing a common cause to that of the active nature of the SMBH (Barnes & Hernquist, 1992; Capelo et al., 2015; Sanders et al., 1988).

1.2 Multi-wavelength methods of AGN identification

In order to develop our understanding of the intrinsic link between AGN and their host galaxies, it is important to build a full picture of AGN within the Universe. A more complete sample of AGN would help identify how their relation with their host galaxies developed over time, and how the galaxy-bulge-SMBH mass relation became fixed in local galaxies (Magorrian et al., 1998). As such, we require a full exploration of the various components of an AGN, such as those shown in Fig. 1.1. We also need a more complete sample without significant contamination from other astronomical objects, such as stars.

A benefit of data collection from high luminosity AGN is that optical, infrared, X-ray and radio observations can be used to probe different regions and mechanisms occurring simultaneously within the AGN. However, each method of identification can be subject to different sample biases. As such, a combination of catalogues using identification methods from a range of wavelengths is required to build a complete picture of AGN.

1.2.1 Infrared selection

Infrared (IR) based observations in the $\sim 1.2-22\mu m$ region of the spectrum are typically dominated by the light absorbed from the accretion disc and re-emitted in the IR by the dusty torus. The wavelength range in which IR selection is a useful tool is heavily dependant on host galaxy and stellar emission. For significant old stellar populations, the ratio of host galaxy emission to AGN emission reaches a maximum at around 1μm (Hickox & Alexander, 2018; Richards et al., 2006; Temple et al., 2021), and thus tends to dominate over the dusty torus emission. Within the far infrared, the dusty torus emission can be outshone by lower temperature dust within star forming galaxies ($\leq$ 40K compared to $\leq$1280K for the dusty torus) (Magnelli et al., 2012). Selection based on data from both the far and near infrared is therefore problematic, and thus IR AGN selection focuses on the mid infrared range.

The exact nature of the distribution of material within the dusty torus is still debated, with some arguing that the material is smoothly distributed within the region (Dullemond & van Bemmel, 2005; Fritz et al., 2006a; Pier & Krolik, 1992), or formed in clumps (Krolik & Begelman, 1988; Tristram et al., 2007). Further observations of IR spectra may therefore aid in our understanding of this specific region of the AGN. In the case of the most heavily obscured AGN, where the column density of the gas surrounding the accretion disc exceeds $1.5 \times 10^{24} cm^{-2}$, even highly ionising X-rays are unable to escape, and thus are undetectable within this wavelength region. However, as the IR specifically probes the emission of the surrounding gas and dust, such Compton-thick objects are still visible, and it is via infrared observations that we are able to provide insight into this subset of objects. Studies have shown that the number of Compton thick AGN within the Universe may be significant, with upper limits on the Compton thick AGN fraction at $z > 0.5$ being as high as 40% (Laloux et al., 2022), thus, including such objects is key to building a full picture of AGN in the Universe.

Despite these benefits however, selection via infrared can be subject to contamination from other sources with similar colours, specifically star forming galaxies with polycyclic aromatic hydrocarbon emission (Stern et al., 2005). As infrared observations are specifically probing hot dust emission, they are also not a useful identifier for dust poor AGN that may be outshone in the IR by host galaxy emission.

Within the infrared, absorption within the Earth's atmosphere makes ground-based observation difficult, thus the majority of IR data are collected from space-based satellites. The main sources of IR data are the *Spitzer space telescope* (Werner et al., 2004), and the Wide-field infrared Survey explorer (*WISE*) (Wright et al., 2010). Selection of AGN from these observations is based on colour-colour plots, in which AGN separate out from other MIR sources. IR observations will benefit from the recently launched *James Webb Space Telescope*, with instruments including NIRSpec, NIRCam and MIRI providing both IR spectra and photometric data spanning from 0.6-27μm (Gardner et al., 2006).

1.2.2 Optical selection

Within the optical region of the spectrum, we are able to probe material within the accretion disc of the AGN, assuming that the line-of-sight to the AGN is not entirely blocked by the dusty torus (see Fig. 1.1). AGN identification in the optical makes use of *ugriz* colour-colour diagrams (Richards et al., 2001), which have been used prominently in large scale surveys such as SDSS (Schneider et al., 2003). The shape of the continuum within the optical varies between AGN and stars, allowing for the identification and removal of a stellar locus to select large numbers of luminous and unobscured AGN.

However, the removal of non-AGN objects via colour-colour cuts is not exact. If too tight a cut is made to remove objects around the stellar locus, it is possible to also lose AGN from the sample. Conversely, too loose a cut might still include contamination from stellar objects (Padovani et al., 2017). Additionally, due to the highly dusty nature of type-2 AGN, large amounts of flux from the blue region of the spectrum is absorbed and remitted within the IR. Thus, the optical flux is severely reduced, and can be completely outshone by host galaxy light within this wavelength region. This therefore places a bias in optical selection against type-2 AGN.

Despite these issues, there are many advantages of optical AGN identification. Optical data can be collected from ground based observations. For example, the Sloan Digital Sky Survey (SDSS) (Schneider et al., 2003) features the collection of both multi-filter CCD imagining and multi-object spectra from the ground based 2.5m telescope located at the Apache Point Observatory in New Mexico. This is not the case for both the X-ray (Chen et al., 2018; Truemper, 1982) and IR (Mauduit et al., 2012), which require space-based observatories for useful data collection due to the effects of atmospheric absorption.

Additionally, when compared to other wavelength regions, deep optical photometry of the whole sky can be collected relatively inexpensively in terms of exposure time. Whilst whole sky surveys have been performed in other wavelength regions, such as *ROSAT* within the X-ray (Truemper, 1982), gaining the same depth requires much longer exposure times, and thus many non-optical surveys tend to focus instead on observing more limited regions of sky for longer periods (Chen et al., 2018).

Deeper optical observations allow for greater spatial resolution than is possible for IR, X-ray and Radio selection, which allows cuts to be made based on morphology. Extended galaxies can therefore be easily removed from target selection. Such a cut leaves only point source AGN and stars within the sample, which can be separated further via colour-colour plots. Within other wavelength regions, the shallower depth of observations mean that galaxies appear as point sources, further contaminating the sample.

1.2.3 X-ray selection

X-ray observations are useful in providing information on the processes related to the inner accretion disc (Mushotzky et al., 1993). X-rays are also typically a beneficial method for AGN identification due to their ubiquity in almost all AGN, and their penetrating nature allowing them to still be observed through the dust and gas in the surrounding torus. The strengths at which emitted X-rays are observed ($L_X > 10^{42} ergs^{-1}$) are not typically seen in other astronomical objects, such as star forming galaxies. As such, X-ray selection is useful in avoiding contamination within the selected sample. Identification using only X-rays can become difficult in cases of extremely low luminosity or heavily obscured, Compton-thick AGN. In these cases, the AGN may go undetected without the addition of IR data (Laloux et al., 2022), as discussed above.

Alternatively, what is believed to be a low luminosity AGN could be an alternate source. For example, X-ray binaries, degenerate stars which release energy due to accretion from their non-degenerate companion, can produce similar energies to low luminosity X-ray AGN ($L_X \approx 10^{41} ergs^{-1}$) (Padovani et al., 2017), therefore providing contamination to the sample.

Additionally, unlike observations within the optical and infrared, the extent of sky coverage is relatively limited for current X-ray observations. Whilst all-sky X-ray surveys have been performed before by *ROSAT* (Truemper, 1982), the most current X-ray surveys such as *XMM-Newton* (Chen et al., 2018) have small fields of view, and as such are limited in scale. However, modern surveys, including the data collected by the *eROSITA* telescope (Predehl et al., 2021), greatly improve both the X-ray coverage of *XMM-Newton*, providing an all sky-survey, and the sensitivity of *ROSAT*, with a soft X-ray band (0.2-2.3keV) 25 times more sensitive.

1.2.4 Spectroscopic selection

Spectroscopic observations focusing on emission lines are also used to identify AGN that might be missed from photometric selection methods. Observations of a multitude of broad emission lines, such as the Hydrogen Balmer series, can give an indication of the presence of an AGN even in the case where a host galaxy outshines the output of the AGN. Similarly, the use of the ratio of emission lines Log$\frac{[O_{III}]}{H_\beta}$ vs Log$\frac{[N_{II}]}{H_\alpha}$, typically known as the BPT diagram (Baldwin et al., 1981; Veilleux & Osterbrock, 1987) can be used to separate out inactive galaxies from type-2, low luminosity AGN in large scale spectroscopic surveys such as SDSS (Best et al., 2005; Juneau et al., 2014; Yuan et al., 2016). Such objects would not be observed using optical photometric selection.

Integral field unit spectroscopy using data from surveys such as MaNGA (Mapping Nearby Galaxies at APO) have also been useful in the identification of AGN within SDSS that previously have not been seen due to the dominance of galaxy emission. Through the use of spatially resolved spectroscopy, the spectra of the central region of the host galaxy can be observed separately, and thus signatures of the presence of an AGN can be found (Wylezalek et al., 2018).

1.2.5 Future directions and multi-wavelength selection

From looking at both the pros and cons of AGN selection in a number of wavelength regimes, it is clear that the use of multiple wavelength observations in combination is useful in the identification of AGN. Surveys such as SDSS have been able to identify large numbers of AGN using optical (Aihara et al., 2018), IR (Wright et al., 2010), UV (Veillet, 2007), X-ray and radio data, making use of the combined advantages of each wavelength range to optimise the sample for followup spectroscopic observations (Lyke et al., 2020).

Multi-wavelength identification also allows for multiple selection criteria to be applied simultaneously. This has the benefit that each individual wavelength selection cut can be more liberal, reducing the likelihood of missing AGN, whilst the combination of the multi-wavelength selection still reduces contamination from other astronomical sources.

Spectral properties more commonly associated with AGN than other astronomical objects also provide a useful method of identification. The power law continuum of an AGN between the X-ray and optical/UV region gives a distinct shape when compared to inactive galaxies, related to the central accretion disc (Hubeny et al., 2001; Laor & Netzer, 1989). Thus, data within these two regions can aid in removing contaminates from the sample. Additionally, one common property of AGN is variability (Angione, 1973; Ulrich et al., 1997) that has been used with instruments such as the Zwicky Transient Facility (López-Navas et al., 2023) to act as a reliable indicator to separate out AGN from other point sources, such as nearby stars (Graham et al., 2014; Pouliasis et al., 2019; Trevese et al., 2008). Future work will build

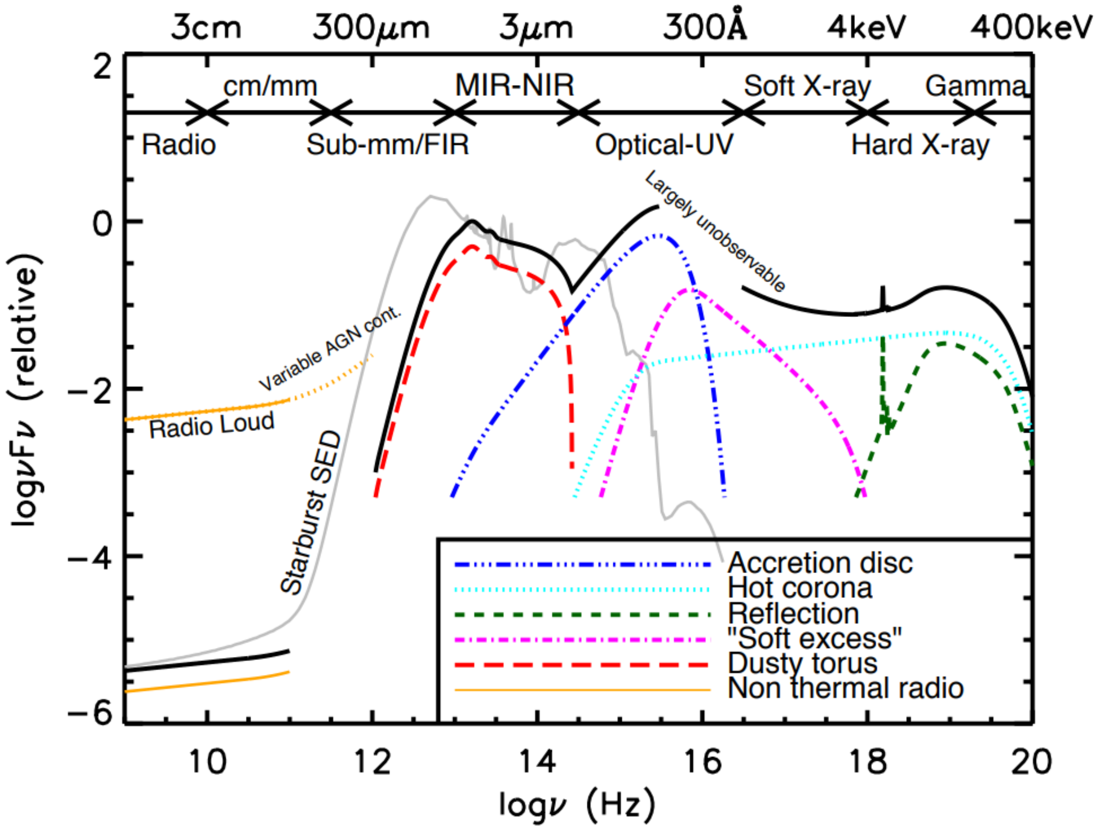

Fig. 1.4 An example AGN SED from Hickox & Alexander (2018) with the individual contributions separated by colour. Here we can see the maximum in AGN flux occurring within the optical-UV region of the spectrum. For comparison, a star-forming galaxy SED is also shown in grey. A maximum in host galaxy flux can be seen at $\sim 1\mu m$.

upon AGN identification from variability, using data from upcoming instruments such as the Large Synoptic Survey Telescope (LSST) (Ivezić et al., 2019).

1.2.6 The AGN SED

The previous section focused on the AGN selection methods used across numerous wavelength regions. In many cases however, contamination from host galaxies can hinder these methods, making it difficult to accurately disentangle AGN and host galaxy contributions. Looking at a typical AGN spectral energy distribution (SED) can therefore help to explain the observational advantages and biases discussed in Section 1.2.

A typical AGN SED can be seen in Fig. 1.4 from Hickox & Alexander (2018). An example host galaxy is also shown for comparison in grey. It is important to note that in the case of the galaxy SED, we are focusing on the specific case in which the star formation rate is particularly high as opposed to, for example, a quiescent elliptical galaxy. In such a case, we may expect the MIR/FIR of the galaxy to be significantly lower, due to the lack of stars heating surrounding gas during formation.

In the case of radio-loud AGN, we get non-thermal radio emission shown in yellow. The dusty torus then contributes in the infrared via re-emission of flux produced within the accretion disc, which itself mainly emits within the optical and UV region of the spectrum. When compared to the starbust galaxy SED shown in grey, there is a significant drop in the relative contribution from the AGN within the far-infrared region of the spectrum. As discussed in Section 1.2, this is due to the higher temperature of the dusty torus compared to the IR flux emitted by dust heated by star formation (Padovani et al., 2017).

Fig 1.4 also shows the maximum of AGN flux occurring within the optical and UV region of the spectrum, due to the accretion disc. The output from the accretion disc, shown in blue, represents the combination of multiple blackbodies at a range of temperatures, which are emitted as material slowly spirals inwards towards the central black hole and discussed in further detail within Section 1.1.1. The galaxy maximum compared to the AGN can also been seen at approximately 1μm. More generally, an important factor to take into account in relation to Fig. 1.4 is that it shows us when both the host galaxy and AGN contributions are comparatively similar in flux. We see a rapid drop-off in host galaxy contributions in both the radio and X-ray regions of the spectrum, and a dominance from the galaxy within the FIR region. Thus, whilst these spectral regions are useful for the identification of AGN (often in conjunction with identification from other regions), the data from these regions alone will not provide information on the host galaxy. As we are trying to further understand the apparent link between host galaxy and AGN properties, it is therefore useful to look within the infrared and optical region of the spectrum. Here we may be able to ascertain information on both AGN and host galaxy properties.

1.3 Introduction to SED fitting codes

Both host galaxy and black hole properties can potentially be understood through analysis of spectra for galaxies containing AGN. However, since the advent of large multi-wavelength galaxy surveys, SED-fitting using photometry has been routinely employed in order to infer photometric redshifts (e.g. Salvato et al. 2019) as well as galaxy and AGN physical properties for statistical samples (e.g. see Johnson et al. 2021; Rosario 2019; Thorne et al. 2022; Walcher et al. 2011 for a review of commonly used methods). In essence, SED fitting methods involve the comparison of observational data to models of AGN and host galaxies, in order to infer the AGN and host galaxy properties of the observed object.

At rest-frame ultraviolet (UV) and optical wavelengths much of the focus in SED-fitting has been on improving stellar population synthesis models to more accurately represent the emission from stars in galaxies (e.g. Conroy et al. 2013; Kriek & Conroy 2013 and references therein). Bayesian techniques are also increasingly being employed to constrain SED properties (e.g. Calistro Rivera et al. 2016) with greater awareness of some of the potential pitfalls of interpreting simple maximum likelihood estimates (Mountrichas et al., 2021). Within the galaxy SED fitting community, there is a recognition that SED-fit parameters can have complex degeneracies in the multi-dimensional fitting space (e.g. Lower et al. 2020), and underestimating the real inherent uncertainties in these fits has potential consequences for what we can conclude from them about galaxy formation and evolution (e.g. Curtis-Lake et al. 2021).

For high-redshift galaxies, the realisation that emission lines can contribute significant flux in some passbands and therefore influence the best-fit SED model has revolutionised our understanding of the results from SED-fitting (e.g. De Barros et al. 2013; Schaerer & de Barros 2009; Smit et al. 2014. Contemporaneously to these advances, there have also been notable developments in SED-fitting techniques that use a self-consistent approach to simultaneously model the ultraviolet through far infrared emission from galaxies (e.g. CIGALE Boquien et al. 2019, MAGPHYS Da Cunha et al. 2011). Large survey datasets that extend into the infrared, such as surveys conducted with the *Herschel Space Telescope* (Pilbratt et al., 2010), have driven these improvements to be able to model the cool dust emission from galaxies. AGN components are more commonly incorporated into SED modelling codes that cover an extensive wavelength range primarily because the multi-wavelength data can help break some of the degeneracies between the AGN and host galaxy parameters (Calistro Rivera et al., 2016).

There are two broad types of AGN templates commonly employed in SED-fitting codes: (i) empirically derived templates (e.g. Polletta et al. 2007; Richards et al. 2006) based on observations of known AGN and (ii) theoretical templates produced using radiative transfer models (e.g. Fritz et al. 2006b; Stalevski et al. 2016). The empirical templates, while providing a relatively simple parametrisation of the AGN emission, may not be representative of all AGN. Theoretical SEDs offer more flexibility to model diverse AGN emission but at the expense of a very large number of free parameters, many of which are difficult to constrain using broadband photometric data alone. Moreover, none of these templates have, as yet, assessed critically the effect of emission and absorption features to the broadband SED fitting, in a way analogous to what has been done for high-redshift galaxies.

The rapid advances in precision imaging datasets in the optical and near infrared — e.g. the Dark Energy Survey (DES; Abbott et al. 2021), Hyper Suprime-Cam (HSC; Aihara et al. 2018), the upcoming Vera C. Rubin Observatory Legacy Survey of Space Time (LSST; Ivezić et al. 2019) and *Euclid* (Percival et al., 2019) — means optical and infrared surveys are already far surpassing the flux limits achievable over a wider wavelength range. In the context of jointly studying AGN and host galaxy emission, this necessitates the development of parallel SED-fitting approaches that attempt to model both the AGN and host galaxy over a more limited wavelength range, with a relatively small number of free parameters and to the resolution required to match current and future large sky surveys. New wide-field spectroscopic surveys such as 4MOST (Merloni et al., 2019) and VLT-MOONS (Maiolino et al., 2020b) will also use SED-fitting to broadband photometry as the basis for their AGN target selection. Thus, in light of these new surveys, it is timely to critically assess how well current SED-fitting methods are able to jointly constrain galaxy and AGN properties.

With this in mind, we choose to produce our own SED fitting code for use within this thesis. We combine an empirically derived AGN SED model from Temple et al. (2021) with galaxy templates from Conroy & Gunn (2010); Conroy et al. (2009) to fit the observed optical to infrared SEDs of a sample of X-ray selected AGN. We make use of a Bayesian SED-fitting technique and Markov Chain Monte Carlo (MCMC) sampling to fully explore the AGN+host galaxy parameter space. Our focus is on relatively luminous and distant ($z > 0.7$) AGN using optical and near infrared photometry. As discussed in Section 1.2.6, the chosen wavelength region includes the point in which the AGN/galaxy flux ratio reaches a minimum. Observations in this region thereby enhance the contrast between the galaxy stellar population and the AGN (Bongiorno et al., 2012; Merloni et al., 2010), such that the constraint of both AGN and host galaxy properties may be possible.

1.4 Thesis structure

This thesis is structured as follows: First we describe the development of our SED fitting method in Chapter 2. Here we discuss the components included within the fitting code, and the AGN and host galaxy properties that we are able to infer from its application. We also address the nuances of Bayesian SED-fitting, and the methods used to combat bimodality and lack of convergence that can emerge due to the complexities within our parameter space.

The dataset to which the SED fitting code is applied, XMM-SERVS (Chen et al., 2018), is described in Chapter 3. We also describe the main results from our fits, including the produced inferences on AGN luminosities in the optical, infrared and X-ray. We then discuss how these luminosities relate to the stellar mass of the AGN host galaxy. Comparison is then made between the inferred emission line properties for our objects, to actual emission lines seen in SDSS spectra. Finally, we look at how our inferred emission lines relate to the X-ray-UV slope (α_{ox}), as a proxy for the Baldwin relation (Baldwin, 1977).

Chapter 4 presents a detailed comparison of how the choice of AGN template used in SED fitting can affect host galaxy stellar mass estimates. A frequently used AGN template developed by Richards et al. (2001) is compared to the Temple et al. (2021) template that has been used throughout this thesis. Both AGN templates are used on the same sample of X-ray selected objects from XMM-SERVS (Chen et al., 2018), using the SED fitting code described in Chapter 2.

Finally, Chapter 5 details how future work could build on the current findings presented in this thesis. Specifically, we discuss how the inclusion of additional bands of photometry, and photometric redshift as a free parameter could be used to extend our work to samples without spectroscopic redshift measurements. We also consider the complexities involved with further increasing the parameters within our MCMC fitting code, and how best to avoid bimodality and lack of convergence based on what was learned from the work in Chapter 2. Finally, we discuss how updates to our SED fitting code, including the use of machine learning, will greatly decrease the time it takes to make inferences, and therefore will vastly increase the number of objects we can investigate. Such work would provide an excellent target selection sample for upcoming spectroscopic surveys, such as MOONS and 4MOST.

Throughout the thesis, all magnitudes used are on the AB system. We adopt a ΛCDM cosmology with Ω_m=0.3, Ω_Λ=0.7, $H_0 = 70 kms^{-1}Mpc^{-1}$.

Chapter 2

Development of a new SED fitting code

The content of this Chapter includes work published within the journal *Monthly Notices of the Royal Astronomical Society*, Marshall et al. (2022). I performed the analysis and writing of the work included within this paper.

2.1 Introduction

Spectroscopic data are very useful in the identification of AGN. Properties such as as the presence of broad emission lines due to the rapid rotation of material around the SMBH, the evidence of ionising radiation from narrow line ratios, and the X-ray to optical/UV power law continuum (Hubeny et al., 2001; Laor & Netzer, 1989) can act as clear indicators of the presence of AGN. However, spectroscopic data is limited in scale due to factors such as the long exposure times, making identification of large samples of AGN via spectroscopic data alone difficult. Conversely, photometric data in a number of bands is already available through a number of large-area, multi-purpose surveys. Such photometric data can also be used in the identification of AGN, through the use of Spectral Energy Distribution (SED) fitting.

In order to gain information on a large number of AGN, it is therefore important to understand what can be inferred about AGN and their host galaxies specifically from photometry, along with the limitations that might be associated with this method. With up-to-date precision imaging datasets, such as the Dark Energy Survey (DES; Abbott et al. (2021)), and Hyper Supreme-Cam (HSC; (Aihara et al., 2018)), optical and infrared data is surpassing the flux limits of data available within other wavelength regions such as the X-ray and far infrared. We therefore also want a reliable SED fitting code that focuses on the optical and infrared wavelength range. To this end, this chapter will focus on the development of a new SED fitting code using a Bayesian MCMC analysis of AGN and their host galaxies, specifically to fit optical and infrared photometry.

From our overview of a typical AGN SED shown in Fig. 1.4, we are able to determine the components necessary to accurately model AGN optical and infrared photometry. We therefore incorporate three components which are combined to create a total model SED. These components are the AGN accretion disc and broad line emission, the hot dust emission from the AGN, and the AGN host galaxy (including stellar light and nebular emission). Due to the chosen wavelength range of the data, we do not include the contribution of cooler dust, which may provide significant flux at longer wavelengths. We describe the galaxy and AGN templates below, along with the development and testing of the SED code. Due to regions of high flux output from both AGN and their host galaxies (as shown in Fig. 1.4), there is a significant potential for degeneracies between the host galaxy and AGN contributions to the total SED. In this chapter, we therefore also discuss the methods used to combat bimodality and lack of convergence, where possible, through the use of parallel-tempering ensemble MCMC.

2.2 Method

2.2.1 Galaxy Template

The galaxy templates used were produced using the Flexible Stellar Population Synthesis (FSPS) code (Conroy & Gunn, 2010; Conroy et al., 2009), including nebular emission lines. Composite stellar populations (CSPs) were produced assuming an initial mass function from Chabrier (2003) and solar metallicity. We assume exponentially declining star-formation histories with a range of e-folding times and range of ages as detailed in Table 2.1. Dust extinction is applied to these templates, assuming the Calzetti et al. (1994) attenuation curve, for a range of optical depths, producing a total of $26 \times 20 \times 94 = 48880$ separate galaxy spectra. Finally, the normalisation of the galaxy template provides an additional free parameter corresponding to the stellar mass of the galaxy. In total, this provides a total of four free parameters associated with the AGN host galaxy. The range of these parameters are outlined within Table 2.1.

2.2.2 AGN Template

To model AGN emission, we make use of a new empirical AGN SED from Temple et al. (2021). This AGN model includes contributions from the accretion disk, broad and narrow

Table 2.1 Model parameters for our galaxy and AGN SED templates. All host galaxy, AGN and hot dust priors are flat in log space. The final column provides the step sizes, and number of templates used for each property. A scaling factor note within this column indicates that any value between the limits for each individual property could be produced using a scaling factor applied to the total SED.

Parameter	Range	Step sizes
Galaxy properties		
Star formation rate e-folding time (τ)	$0.08 \leq \tau \leq 25$ Gyr	26 templates in log(0.1) steps
Effective v-band Optical depth τV	$10^{-4} \leq \tau\mathrm{V} \leq 2$	20 templates where $\tau V = 2 \times \left(\frac{n}{19}\right)^2 + 10^{-4}$ for n = 0 - 19
Age (Years)	$3.2 \times 10^5 \leq \mathrm{Age} \leq 1.4 \times 10^{10}$	94 templates in log(0.05) steps
Stellar Mass ($\log_{10}(M_*/M_\odot)$	$10^7 \leq M_* \leq 10^{13} M\odot$	Scaling factor
AGN properties		
AGN Luminosity at 3000Å	$10^{40} \leq L_{3000} \leq 10^{50}$ ergs^{-1}	Scaling factor
AGN Reddening E(B-V)	-0.2 $\leq$ E(B-V) $\leq$ 2	21 templates in 0.11 steps
Hot dust Luminosity at 3μm	$10^{30} \leq L_{3\mu m} \leq 10^{49}$ ergs^{-1}	Scaling factor
Emission line strength	-2 $\leq$ emline type $\leq$ 3	2 templates (emline = -2, 3)

emission lines, and the hottest component of the dusty torus at close to the sublimation temperature. The model is empirically calibrated using colours of quasars in the Sloan Digital Sky Survey.

The Temple et al. (2021) model differs from previous AGN SED models (e.g. Richards et al. (2006)) in its aim to accurately reproduce the average colours of unobscured AGN over a more limited wavelength range (rest frame 912Å- 3μm) using only a small number of free parameters. Notable improvements in the Temple et al. (2021) model relative to previous work include a more accurate determination of the contamination from the host galaxy to the AGN continuum, as well as a thorough treatment of the effect of broad emission lines on the broadband AGN colours. The former improvement allows a more robust determination of the 'pure AGN' emission, consequently providing more reliable host galaxy properties. Even in the case of bright AGN at $z > 2$, host galaxy contribution was found to account for $> 5\%$ of the flux of the total SED (Temple et al., 2021). By accurately removing host galaxy flux from the AGN model, we are ensuring that this flux is attributed to our galaxy templates instead. The Temple et al. (2021) model also encompasses a range of possible emission line properties seen in AGN spectra, from weak, highly blueshifted lines through to high equivalent width, symmetrical lines. The AGN SED model is described in further detail in Temple et al. (2021).

The free parameters within the AGN model include the AGN reddening, E(B-V), assuming an empirically derived extinction curve (Temple et al., 2021). The AGN template was reddened following the relation shown below. This provides a reddened flux, f_{red}, given the emitted flux, f_{em}:

$$f_{red}(\lambda) = f_{em}(\lambda) \times 10^{-0.4A_\lambda} \tag{2.1}$$

The value of A_λ is the dust extinction at a given wavelength, λ, defined as:

$$A_\lambda = E(B-V)(k_\lambda + 3.1) \tag{2.2}$$

where E(B-V) is the reddening parameter and k_λ is the reddening law. For this work, the extinction curve was produced via a comparison of two similar AGN populations, with the exception that one was highly reddened, whilst the other was unobscured. k_λ is therefore a function that represents the way in which these spectra differ across the wavelength range of the AGN template, specifically due to this obscuration. This reddening law is discussed further within Section 2.6 of Temple et al. (2021).

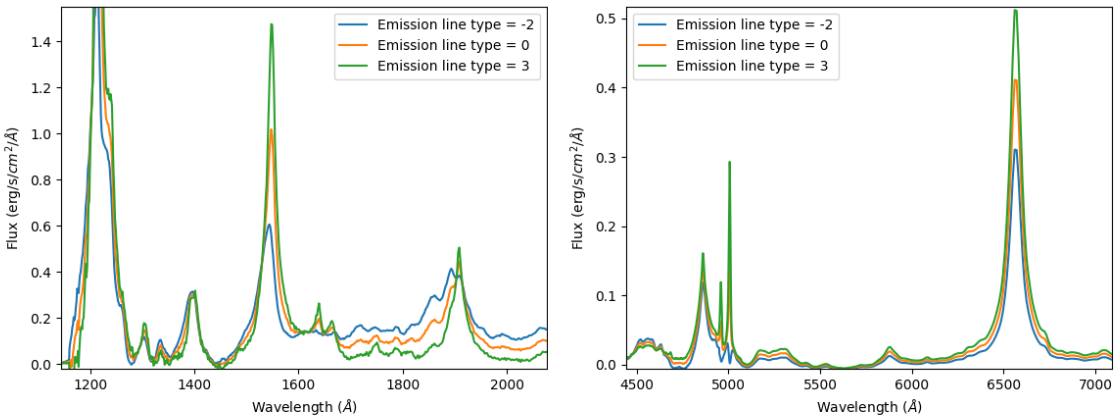

Fig. 2.1 The range of emission lines properties available in the Temple et al. (2021) AGN template spectrum, as parameterised by the emline_type. Negative values correspond to weaker, more highly blueshifted emission lines, whereas positive values correspond to stronger, more symmetric emission lines. An emline_type of 0 corresponds to the average emission line properties for an SDSS AGN at $z = 2$ with an absolute magnitude $M_i = -27$.

Temple et al. (2021) found that emission lines could affect AGN photometric colours by 0.1 magnitudes or more in some cases, a difference that could readily be measured within modern photometric surveys with typical uncertainties of ∼0.05 mag. Therefore, in the case of one run of our SED fitting code, the emission line properties are also allowed to vary as shown in Fig. 2.1, using an interpolation between the two extremes of AGN templates produced within appendix B of Temple et al. (2021). We are able to gain information on the relative blueshift of AGN emission lines from photometry due to the intrinsic link of the emission line blueshift to emission line strength (Temple et al., 2021). The emission line property values are analogous to the emission line range shown in the top of Fig. 3 in Richards et al. (2021). This figure shows CIV equivalent width vs CIV blueshift from a sample of 438 AGN. Negative values of the emission line type correspond to highly blueshifted, weak emission lines similar to objects in the bottom right of Richards et al. (2021) Fig. 3, whereas positive values of the emission line property correspond to stronger, more symmetrical emission lines, similar to objects in the top left of that figure.

The normalisation of the rest-frame UV to optical SED, applied at 3000Å before reddening, also produces a free parameter that provides a measure of the AGN luminosity. Finally, the hot dust component is modelled as a blackbody with a fixed temperature of 1236K, whose normalisation represents the hot dust luminosity at 3μm. Cooler components of the dust do not affect the observed-frame colours at λ <4.5μm given our redshift cut of $z > 0.7$, and thus are not considered in our models.

2.2.3 Template combination and magnitude conversion

In order to compare the galaxy and AGN templates to photometric data, they first require combination and conversion to the corresponding magnitudes that would be observed through each filter. The large number of dimensions that would be needed for a grid containing spectra for all possible parameter combinations makes producing this computationally inviable. Instead, three magnitude grids were produced over the physical parameters for the galaxy, AGN, and hot dust. To calculate the relevant magnitudes, each spectrum can be multiplied by the transmission function for the photometric bands used within the fit. For this thesis, these filter transmission functions were provided by the Spanish Virtual Observatory (Rodrigo et al., 2017). For each filter, the flux was then integrated over and converted to the AB photometric system, the same form in which the photometric data are provided, according to the relation:

$$F_s = \frac{\int \lambda F_\lambda S(\lambda) d\lambda}{\int \lambda S(\lambda) d\lambda} \tag{2.3}$$

where $S(\lambda)$ represents the relevant filter transmission function, and F_λ is the input flux. All template data were therefore placed in the same format as the observational data. Before use within the SED fitting code, the 3-dimensional galaxy, 2-dimensional AGN and 1-dimensional hot dust cubes were interpolated via a spline interpolation, allowing the MCMC code to call any parameter value between the assigned limits. Each cube would then provide the relevant magnitude values associated with the called parameters. Additional scaling factors relating to the AGN luminosity and the host galaxy stellar mass would then be applied, providing two additional free parameters for a total of eight. Finally, AGN (including the hot dust), and host galaxy components are combined to give the total template magnitudes for comparison to the observed data.

2.3 Overcoming sampling difficulties

2.3.1 MCMC Fitting Algorithm

After the production of our AGN, hot dust, and host galaxy templates, we then constructed a Bayesian MCMC fitting code to sample the parameter space shown in Table 2.1. Given the diversity of AGN, our code was developed in order to produce four different SED fits, hereafter referred to as model families. The best model was then chosen for each object, as detailed in Section 2.3.3. The four model families correspond to:

1. AGN and host galaxy template components are included, with the AGN emission properties fixed to the average seen in SDSS AGN at $z = 2$ and an average absolute magnitude $M_i = -27$ (Temple et al., 2021);

2. AGN and host galaxy template components are included, with variable emission line properties in the AGN component. Emission lines are allowed to vary between the limits shown in Fig. 2.1;

3. Only galaxy templates are included;

4. Only AGN templates are included, with the same variable emission line properties as used in model family 2.

The latter two model families allow for the fitting of objects where the optical to infrared emission is entirely dominated by either the galaxy or the AGN component. We also include the free, and fixed emission line model families in order to determine if, for a subset of objects, we are able to infer information on emission lines from photometry alone.

For each model family, the MCMC algorithm allows for the calculation of a posterior probability, P(ρ|data), for a specified set of model parameters, ρ, via the Bayesian relation:

$$P(\rho|data) = \frac{P(data|\rho)P(\rho)}{P(data)} \tag{2.4}$$

where *P(data)* is a normalisation term, and *P(ρ)* is a function containing information associated with any prior knowledge of the expected value of each free parameter. For each SED model, the prior information, *P(ρ)*, for all parameters of the model is a flat distribution. In the cases of τ, τ_v, age and mass for the galaxy templates, the AGN luminosity at 3000Å , and the hot dust luminosity at 3μm, the prior distribution is flat in logarithmic space (see Table 2.1). *P(data|ρ)* is the likelihood of observing the data given the model parameters ρ. For a given band *i*, we compute the model magnitude $m_{model_i}(\rho)$ and calculate the likelihood using the observed photometric magnitudes, m_i, and uncertainties, σ_i, assuming a Gaussian likelihood:

$$P(data|\rho) \propto \prod_i^n exp\left(- \frac{[m_{obs_i} - m_{model_i}(\rho)]^2}{2\sigma_i^2} \right). \tag{2.5}$$

In order to explore the posterior space, MCMC fitting makes use of 'walkers'. Walkers move through the parameter space by proposing a step (known as a Monte Carlo step) either accepting or rejecting this step based on the posterior probability of the new position compared to the current position. Each step will eventually lead to convergence towards the highest probability region within the parameter space. Walkers are not fully independent, and each will aim to move around the space to form a single chain. This aids in increasing the efficiency of the parameter space exploration, by decreasing the correlation scale of the samples in order to provide faster sampling.

We anticipate that the posterior will often be multi-modal, with some of the modes having negligible likelihood. Such modes can lead some walkers to remain in lower probability solutions, an example of which can be seen in Fig. 2.2. We therefore utilise parallel-tempering ensemble MCMC from the package EMCEE[1] (Foreman-Mackey et al., 2013). Instead of a single chain, this will produce multiple replicas of the system, each separated out to a number of different 'temperatures', T, to provide different probability distributions, D(ρ), according to the relation:

$$D_{new}(\rho) = exp\left(- \frac{1}{T} logD(\rho) \right) \tag{2.6}$$

For a large temperature, $D_{new}(\rho)$ will decrease, thus producing an overall 'flatter' distribution of probabilities. The chance of a walker moving out of a local maximum is increased within higher T systems, as the flatter distribution increases the likelihood that a proposed step will be accepted. More of the parameter space will therefore be explored, such that the chains are more likely to converge to the global maximum. The parallel tempering ensemble allows separate temperature walkers to exchange configurations. Thus, through the use of multiple temperature systems, we are able to both sample larger regions of parameter space due to higher temperature runs, and retain information on the peaks in the likelihood distribution due to lower temperature runs.

Based on this, we choose to run our SED fitting code as outlined below. We initially randomise the starting position of our walkers within the prior space. A parallel-tempering optimisation is performed, using eight temperatures. The choice to use eight temperatures was made as this value was found to provide a good balance between aiding walkers in escaping local maxima, without greatly increasing the code run time. The initial parallel-tempering run is expected to place the walkers close to the global maxima. The bulk of the posterior can then be fully explored using an MCMC run with walkers starting at the final positions of the parallel-tempering optimisation. We then take our inferences on AGN and host galaxy properties from the samples of this MCMC run.

[1]Made using emcee Version 2.2.1

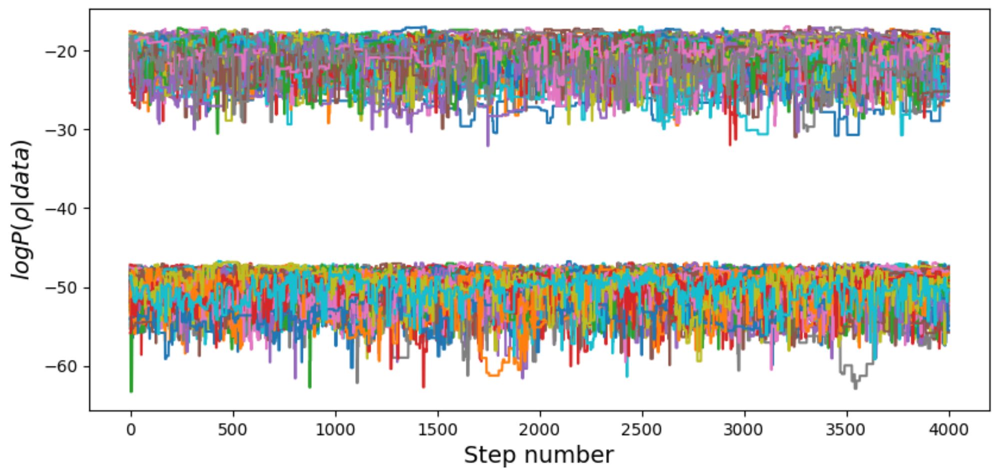

Fig. 2.2 The posterior probability of walkers exploring parameter space for an example run of the MCMC fitting algorithm. Here we can see that around half of the walkers have found a global maxima corresponding to a solution with a average $\log P(\rho|data) \approx -23$. However, the remaining walkers are stuck in a significantly lower $\log P(\rho|data)$ solution. Parallel-tempering optimisation is therefore utilised to allow walkers to escape local maxima, and start closer to the global maximum within our actual MCMC run.

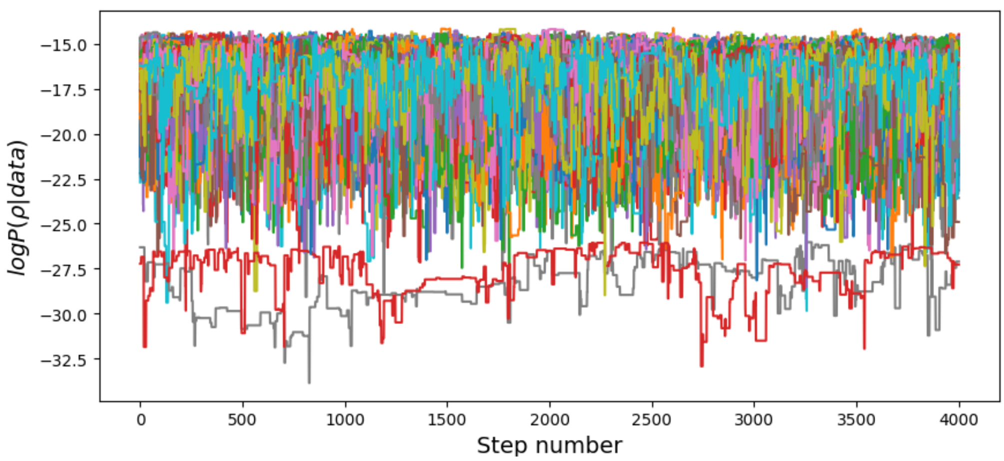

Fig. 2.3 The posterior probability of walkers exploring parameter space for an example run of the MCMC fitting algorithm. Despite an initial parallel-tempering optimisation, two stray walkers have remained unconverged, and not found the global maximum at $\log P(\rho|data) \approx -17.5$. We therefore explore methods of quantifying $\log P(\rho|data)$ distributions with stray walkers, in order to identify unconverged fits from large samples of objects.

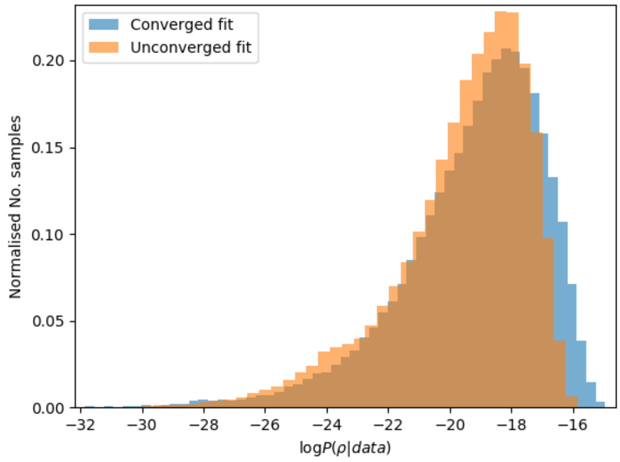

Fig. 2.4 Histograms of the samples from two SED fits. The object in blue has converged to a single solution based on its associated corner plot. The orange object has a small number of walkers within a lower $\log P(\rho|data)$ solution, as can be seen by the increase in samples occurring around $\log P(\rho|data) \approx -24$. Whilst both objects have similar peaks in $\log P(\rho|data)$, the distribution of the converged fit samples is wider, despite the presence of unconverged walkers within the orange fit.

2.3.2 Quantifying bimodality and lack of convergence

Whilst parallel-tempering is useful for removing walkers from lower posterior solutions, the nature of MCMC fitting can lead to a number of other sampling complexities that can persist within our MCMC run. Due to the intricacies of our chosen parameter space, we might expect some objects to produce truly bimodal distributions. In the case of AGN SED fitting, we often see degeneracies between bright, unobscured AGN fits with an old, quiescent galaxy, compared to a young, star forming galaxy with a highly obscured AGN. (An example of an object showing this bimodality can be seen in Fig. 3.4. In such cases, multiple solutions will have similar $\log P(\rho|data)$ values, and thus a single solution is not preferred over another. In other cases, despite the initial optimisation, stray walkers may remain unconverged, such as the example shown in Fig. 2.3. As we aim to apply our SED fitting code to statistically large samples of objects, we therefore attempted to find a diagnostic that is able to characterise these bimodal and unconverged fits.

To this end, we looked at the results from SED fitting for a small sample of 10 test objects, chosen at random from the XMM-SERVS catalogue (Chen et al., 2018). For each object, we looked at two diagnostics. The first was the upper and lower 1σ uncertainties on the sample inferences. The purpose of this test was to look for examples where the posterior had a broad distribution that may point towards a bimodal solution. For the case of the truly bimodal fits, whilst the posterior probability of each solution may be similar, we would still expect to observe a broadening of the property inferences due to the multiple peaks associated with each solution. The second diagnostic was to look at the $\log P(\rho|data)$ distribution. In the case of lack of convergence due to stray walkers, we might expect an increase in the width of the distribution, and a significant tail to lower $\log P(\rho|data)$ values, compared to a converged fit.

However, issues were identified in the use of these diagnostics as methods to determine a bimodal or unconverged fit. Both methods rely on providing a cut-off value above which either the $P(\rho|data)$ distribution or 1σ sample uncertainties are classed as unconverged or bimodal. From our example fits, we identified that a reasonable cut-off value varies on an object by object basis. This can see seen in Fig. 2.4, which shows the $\log P(\rho|data)$ for the samples from two test objects, both of which have similar $\log P(\rho|data)$ maxima. The orange object in this case shows an example of an unconverged fit. We can see evidence for this at around $\log P(\rho|data) \approx -24$, where there is an increase in samples due to the presence of walkers stuck in a local maxima. In the converged blue distribution, we don't see evidence of lack of convergence within our sample distributions. We can see that the distribution is wider for our converged object, with a standard deviation of 2.33 compared to 2.22 for the unconverged object. Any cut-off value based on the $\log P(\rho|data)$ distribution to remove the unconverged object would therefore also exclude the converged object.

Similarly for the 1σ sample uncertainties, there can be a number of complexities associated with each parameter which could lead to larger distributions. For example, an SED fit to a dust poor AGN might have poor constraints on the hot dust luminosity. These poor constraints could lead to a relatively wide distribution in $L_{3\mu m}$, but does not guarantee that the fit is bimodal.

Based on the issued outlined above, we chose not to attempt to make a specific cut that would remove bimodal objects, or objects with some unconverged walkers. Instead, the decision was made to include both median solutions, along with the total sample distributions within figures and in the analysis of property inferences, without attempting to separate out bimodal objects.

2.3.3 Selecting the best model family

As discussed in Section 2.3.1 we consider four model families for each object. These models include the two AGN+GAL fits, one with emission line properties as a free parameter and the other fixed to the average emission line properties for SDSS AGN derived by Temple et al. (2021), as well as the AGN- and galaxy-only fits. For each model family we calculate the Akaike Information Criterion (AIC) using the maximum likelihood, L, and the number of free parameters, k, as:

$$AIC = 2k - 2log(L) \tag{2.7}$$

where the model family with the lowest AIC is considered our best-fit. The benefit of using the AIC to determine which model family best represents the data for each object is that increasingly complex models with larger numbers of free parameters are penalised in the AIC. We can therefore, for example, determine if allowing the emission line properties to vary is truly improving the quality of the fit significantly enough to justify the increased model complexity. The AIC value is also explicitly tied to the reduced χ^2, which is commonly used as a determination of the quality of SED fits, and can be calculated by removing the prior distributions from our posterior. As outlined within Table 2.1, this work uses flat priors between the stated limits for each property. The prior will be different for each of our model families, due to the change in the number of prior distributions from each of the free parameters included. However, in general, the reduced χ^2 can be calculated from the AIC as:

$$\chi^2_{red} = \frac{AIC - log(P(\rho))}{n-k} \tag{2.8}$$

Where n is the total number of data points, k is the number of free parameters (which is dependant on the model family). $log(P(\rho))$ is the contribution to the posterior probability from the priors placed on our proprieties, the values of which are provided by the python package *PyMC* (Salvatier et al., 2016). As the reduced χ^2 is commonly in previous SED fitting work (e.g. Bongiorno et al. (2012); Merloni et al. (2010)), this value will be used throughout the thesis as a measure of the quality of our best fits for each of our model families.

2.4 Summary

In this chapter, we have outlined the development of a new SED fitting code, which will be used to infer both AGN and host galaxy properties in Chapter 3:

- SED fitting software was developed to fit AGN and their host galaxies within the optical and infrared. This fitting code included the use of three separate SED components that were combined to produce an optical and infrared spectrum. The SED components specifically allowed for contributions from the AGN, including the hot dust emission from the dusty torus, and host galaxy. These three components provided a total of eight free parameters to be included within our fits. Importantly, the AGN component produced by Temple et al. (2021) included the ability to change emission lines from strong and symmetrical through to weak and highly blueshifted, via the inclusion of a single additional free parameter. The inclusion of this emission line parameter allows for testing of our ability to infer emission line properties from photometry alone, and is explored further in Chapter 3.

- After the initial development of the SED fitting code was complete, a number of tests were performed in order to ensure that the code was working as intended. One of the most common issues with MCMC fitting with a large number of free parameters is that of bimodality, or lack of convergence. In order to combat these issues, we employed the use of parallel-tempering MCMC for initial optimisation. Further studies were conducted in order to understand how best to quantify the percentage of bimodal objects within a large sample. However, it was found that values such as the $\log P(\rho|data)$ standard deviation varied on an object by object basis, and could be larger for some converged objects when compared to unconverged examples. Such a determination was not possible, and it was therefore decided to include all of the samples from each fit within further analysis.

Chapter 3

Fitting with the XMM-SERVS AGN Sample

The content of this Chapter is based on work published within the journal *Monthly Notices of the Royal Astronomical Society*, Marshall et al. (2022). I performed the analysis and writing of the work included within this paper.

3.1 Introduction

One of the main aims of this thesis is to investigate the relationship between supermassive black holes, and their host galaxies. In Chapter 2 we introduced an SED modelling code, which included three separate components. These component templates allow us to model AGN, including their hot dust emission, and their host galaxies. From this, we are therefore able to explore their properties, and understand the relation between SMBH and their host galaxies in further detail. In this Chapter, we aim to apply our code to real observational data. First, we must therefore choose a sample of AGNs with appropriate photometric data. For this purpose, we first look at AGN selection.

There are many advantages of selecting AGN based on their X-ray properties. For example, high luminosity X-ray emission is uncommon from other astronomical objects, and thus X-ray selection is useful in providing an AGN sample free from contaminants. Additionally, the ubiquity and penetrating nature of light within the X-ray region also means that both type-1 and type-2 AGN can be identified, thus providing a more complete sample of AGN in the Universe.

Whilst X-ray observations are a useful tool for AGN selection, emission from host galaxies within this wavelength region is negligible (see Fig. 1.4). Thus, from X-ray observations alone, we are not able to simultaneously gain insight into both AGN and host galaxy properties. For the application of our SED modelling code, we therefore choose to use data from the optical and infrared region of the spectrum. As can be seen in Fig. 1.4, data at these wavelengths can incorporate regions of both high AGN and host galaxy flux, therefore allowing us to infer the properties of both simultaneously.

At the bluer end of the optical spectrum, emission is typically dominated by material within the accretion disc of the AGN, and thus is modelled in our SED fitting code by the Temple et al. (2021) AGN template (described in Section 2.2.2). Using this template, we are also able to take into account the effects of emission line strengths and morphology, which are included as a free parameter within our SED fitting code. Temple et al. (2021) note within their analysis that the presence of broad emission lines can be seen to have a significant effect on AGN photometric colours by 0.1 magnitudes or more in some cases, a difference that could easily be measured within modern photometric surveys. Thus the inclusion of varying emission lines may allow for inferences on emission line morphology and strength to be made through our SED fits.

The dusty torus surrounding the central region absorbs and re-emits the light from the accretion disc within the infrared region of the spectrum. The material in the torus has a sublimation temperature of $\approx 1280K$ (Magnelli et al., 2012), and thus is represented by the hot dust blackbody included within our modelling code. The output of flux from the dusty torus is key in understanding the output of type-2 AGN, in which line of sight to the accretion disc is blocked by the torus. This can lead to low AGN flux in the optical region of the spectrum (Padovani et al., 2017).

In between the torus and accretion disc emission, at around $1\mu m$, we see a maximum in host galaxy emission (Hickox & Alexander, 2018; Richards et al., 2006; Temple et al., 2021). In the production of the AGN template by Temple et al. (2021), care was taken to ensure that the flux around $1\mu m$ was purely AGN emission, with host galaxy contamination removed. Thus from the inclusion of the galaxy templates within our fitting code, we are able to infer host galaxy stellar mass.

Throughout this chapter, we aim to make use of the advantages of several wavelength regions to provide accurate inferences on both AGN and host galaxy properties. We choose an X-ray selection to allow us to produce a sample including type-2 AGN. In order to infer our AGN and host galaxy properties, the chosen X-ray sample is matched to a number of optical and infrared surveys that can be used within our SED fitting code. Using both optical and infra-red data, we are able to model the accretion disc, the surrounding hot dusty torus, and the host galaxy. Using the inferences from our SED modelling, we will look both at

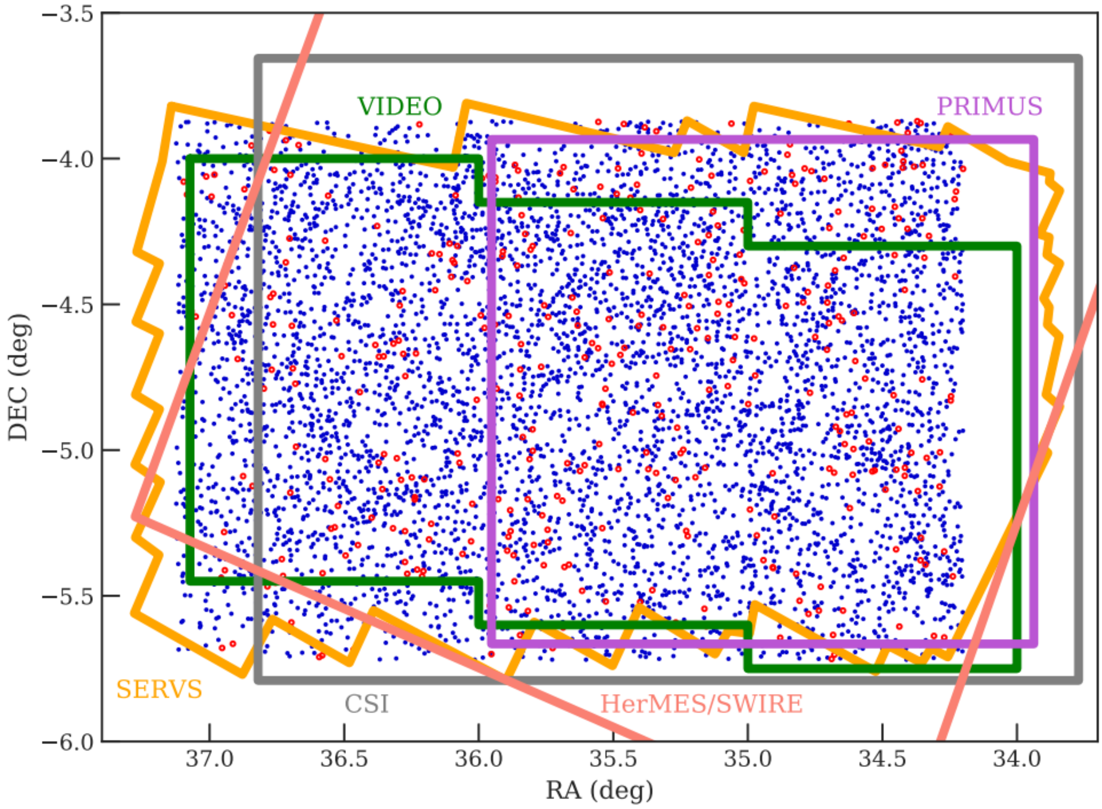

Fig. 3.1 The coverage of the XMM-SERVS survey, provided by Chen et al. (2018). Each point represents the position of an X-ray selected object within the sample. Objects for which reliable multi-wavelength counterparts could not be found by Chen+18 are shown in red, and represent 7% of the total sample. These objects were not included within our SED fitting. The multi-wavelength survey footprints of matched optical and infrared data are also shown.

trends between AGN properties and host galaxy stellar mass estimates, and compare our results with previous work by Jun & Im (2013), Marconi et al. (2004). We start by describing our chosen X-ray selected sample, and the matched optical and infrared photometry that will form the basis for our SED fits.

3.2 The XMM-SERVS data sample

The parent sample of X-ray selected AGN considered in this work originates from the XMM-SERVS survey (Chen et al., 2018), which provides 5242 AGN candidates over 5.3 deg^2 of the XMM-Large Scale Structure (XMM-LSS) survey region (shown in Fig. 3.1) with an X-ray survey integration time of $\approx$ 50ks. The X-ray data from this survey originates from European Space Agency's X-ray Multi-Mirror Mission, *XMM-Newton*, a space-based observatory

Table 3.1 The surveys and filters used within this work.

Survey name	Bands	Wavelength Range (Å)	Reference
Hyper Suprime-Cam Deep Survey (HSC)	`grizy`	3940-10931	Aihara et al. (2018)
VISTA DEEP Extragalactic Observations Survey (VIDEO)	`JHKs`	12524-23674	Jarvis et al. (2013)
Spitzer Extragalatic Representative Volumes Survey (SERVS)	`3.6 4.5` (μm)	31296-50561	Mauduit et al. (2012)

launched in 1999 and currently scheduled to operate until the end of 2026. Its instruments feature a set of three X-ray CCD camera, forming the European Photon Imaging Camera (EPIC) and providing observations across a detection range of 0.1-15keV. The advantage of X-ray selection is the ability of hard X-ray photons to penetrate significant dust and gas columns, thereby making identification of heavily obscured AGN possible. X-ray hardness ratios can also be used to separate AGN based on their levels of obscuration, as will be discussed further in Section 3.3.2.

The *XMM-Newton* survey region overlaps with several other optical and infrared multi-wavelength datasets including the *grizy* bands of HyperSuprimeCam Deep (Aihara et al., 2018), *ZYJHKs* bands from VISTA VIDEO (Jarvis et al., 2013), and the 3.6μm and 4.5μm bands of the Spitzer Extragalactic Representative Volume Survey (SERVS; Mauduit et al. 2012). Of these surveys, HSC and VISTA VIDEO data are collected from ground-based observatories. HyperSuprimeCam is a 900-megapixel wide field camera for the Subaru 8.2m telescope, developed by the National Astronomical Observatory of Japan and located at the Mauna Kea Observatory on Hawaii. The Visible and Infrared Survey Telescope for Astronomy (VISTA) is a 4.1 telescope based at the Paranal Observatory in Chile. The VISTA Deep Extragalactic Observations (VIDEO) Survey is a 12 deg^2 survey using this telescope, specifically designed to study large scale structure and galaxy evolution including the effects of AGN (Jarvis et al., 2013). Conversely, data from the Spitzer Extragalactic Representative Volume Survey (Mauduit et al., 2012) was collected from the Spitzer Space Telescope, launched in 2003. Whilst Spitzer carries three separate instruments, we will focus on the data from the Infrared Array Camera (IRAC). This camera provided four wavelength 256x256 pixel images across the mid-infrared.

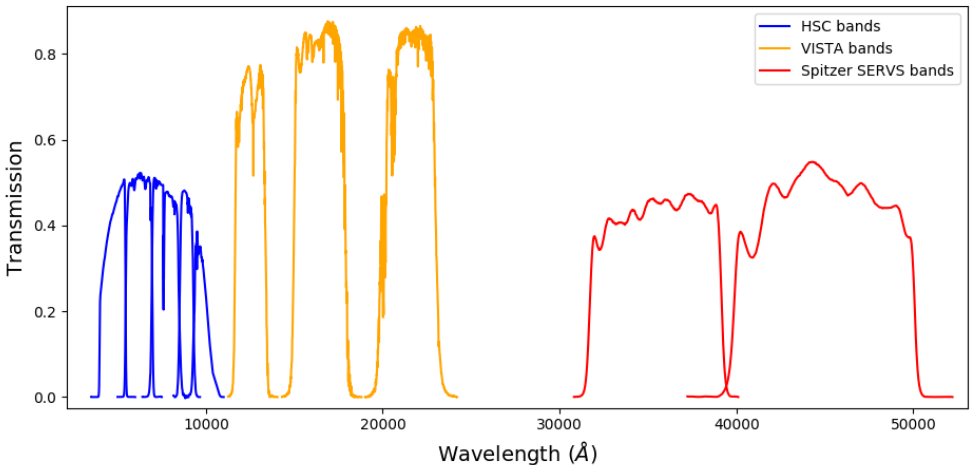

Fig. 3.2 The filter transmission curves including atmospheric absorption and detector quantum efficiencies for the 10 bands of optical and infrared photometry used for SED fitting within this work. These originate from three surveys: HSC (Aihara et al., 2018), VISTA VIDEO (Jarvis et al., 2013) and Spitzer SERVS (Mauduit et al., 2012), each of which has been matched to an X-ray source observed by XMM-Newton (Chen et al., 2018).

The total wavelength range and filter transmission curves for these surveys, including atmospheric absorption and detector quantum efficiencies, can be seen in Table 3.1 and Fig. 3.2 respectively. Chen et al. (2018) have identified optical and infrared counterparts to XMM-Newton X-ray sources within these surveys. Through the use of a likelihood ratio method, 93% of the X-ray sources were found to have reliable counterparts in HSC and VIDEO. In addition, 82% of the X-ray sources have reliable mid-infrared counterparts in either the 3.6μm or 4.5μm imaging from SERVS. Whilst VISTA VIDEO also provides Z- and Y-band photometry, we limit our analysis to only using the HSC z-and y-bands, which are deeper, and more complete within the survey region. Our multi-wavelength data therefore constitutes 10 bands of photometry ranging from the optical g-band through to 4.5μm.

Also providing coverage in the XMM-SERVS survey region is the Canada-France-Hawaii telescope (CFHT u-band). We tested the inclusion of this band within the SED fitting. However, comparison of fits with and without this u-band showed a significant increase in the χ^2 when the u-band was included. We therefore choose to exclude the u-band from our fits. The reason for this discrepancy could be due to the lower spatial resolution of the CFHT data compared to HSC and VISTA. The effect of the inclusion of the u-band is discussed at the end of Section 3.3.

We update the optical and near infrared photometry presented in Chen et al. (2018) by considering the latest data releases from both the HSC PDR2 (Aihara et al., 2019) and VISTA VIDEO surveys (Bowler et al., 2020). We use aperture photometry with a 3.0$''$ diameter aperture and an aperture correction. To avoid over-fitting due to unrealistically small uncertainties, a minimum uncertainty limit of 5% of the 3.0$''$ aperture flux was placed on each band. We further restrict our sample to only those sources with reliable spectroscopic redshifts – 1314 in total. We choose to limit to objects with spectroscopic redshift, as the inclusion of photometric redshift as an additional free parameter would introduce further degeneracies within our SED fits. Additionally, visual inspection of the HSC *gri* images suggests that a large number of the lowest redshift sources are galaxies hosting low-luminosity AGN. As the primary goal of this work is to investigate the spectral energy distributions of high-redshift, high luminosity AGN, we therefore place a redshift cut of $z > 0.7$ on our sample. In total, this cut provides 774 spectroscopically confirmed X-ray selected AGN. Finally, to avoid object blends affecting the optical and infrared photometry, we further remove any AGN with a neighbour in the full HSC DR2 catalogue that is $<2''$ from the AGN itself. Our nearby neighbour cut leads to a final sample of 711 spectroscopically confirmed, X-ray selected AGN, whose spectral energy distributions are studied in detail in this work. A breakdown of the cuts made to our initial XMM-SERVS sample can also be seen in Fig. 3.3.

3.3 XMM-SERVS inferences

Our sample of 711 objects was fit using the newly developed SED fitting code detailed in Chapter 2. Each object was modelled four times, once for each model family. Objects were then categorised base on the model family that provided the smallest Akaike Information Criterion value, as outlined in Section 2.3.3. Our four model families correspond to (i) a AGN+GAL SED with the AGN emission properties fixed to the average seen in AGN at $z = 2$ and an average absolute magnitude $M_i = -27$ (Temple et al., 2021); (ii) a AGN+GAL SED with variable emission lines properties in the AGN component, between the limits shown in Fig. 2.1; (iii) a galaxy-only SED; and (iv) a AGN-only SED with the same variable emission line properties as used in model family (ii). In order to confirm the use of AIC as a method of model family selection, we looked at the objects where the AIC was similar for different model families. Only 13 objects were found to have a galaxy-only and AGN+GAL AIC difference < 1. Additionally, no AGN-only fits met this criteria. Whilst it may be the case that, for some of our galaxy-only fit objects, further information on AGN properties

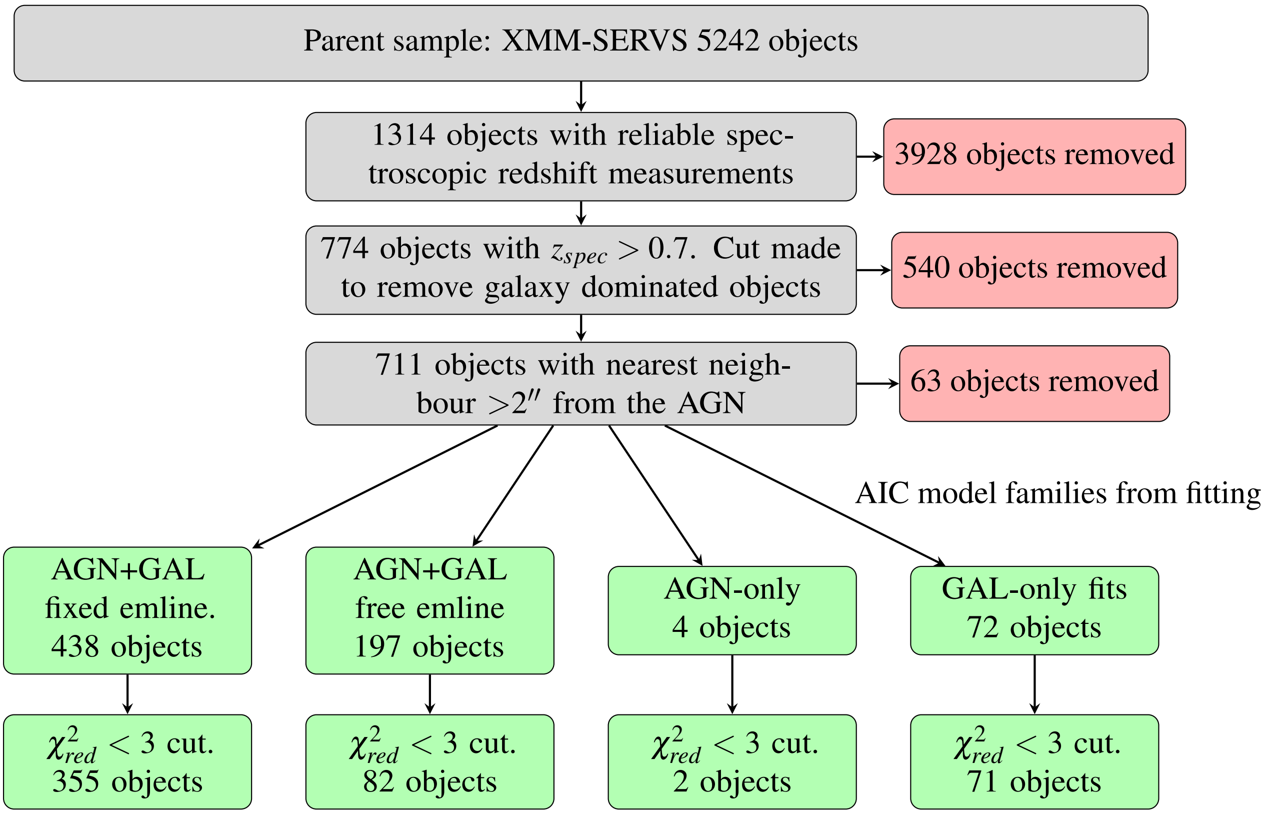

Fig. 3.3 Flowchart showing the cuts made to the initial XMM-SERVS parent sample (Chen et al., 2018). The figure is colour-coded to show the objects removed before fitting in red. The green boxes represent the seperation of the remaining objects into our model families, based on a comparison of their AIC values after fitting. The details of these model families, and the AIC designation method used is described in further detail within Sections 2.3.1 and 2.3.3 respectively.

could have been inferred, the sample of objects is small, and their inclusion does not have a significant effect on the distributions shown in the results of this work.

The flow of objects fit within each model family, from the initial parent sample from XMM-SERVS Chen et al. (2018), is outlined within Fig. 3.3. Of the 711 objects, 438 (61.6%) were best fit with the AGN+GAL model fixed to the average emission line properties in Temple et al. (2021), 197 (27.7%) preferred the AGN+GAL model family with varying emission line properties (see also Section 3.3.4), 72 (10.1%) preferred the GAL only model family and are the most obscured/low-luminosity AGN where the optical to infrared emission is completely dominated by the host galaxy. Finally, only 4 (0.6%) preferred the AGN only model family. Within some sections of our analysis, we compare our inferred AGN and host galaxy properties. Therefore, to maintain consistency in the sample used throughout our analysis, we do not include inferences from the small number of objects that preferred galaxy-only and AGN-only model families.

After the MCMC fitting had been run for all 711 objects, and each object had been assigned a preferred model family, a visual inspection of the marginalised one- and two-dimensional posteriors for each run was performed. An example of these posteriors is shown in Fig. 3.4 where the age, stellar mass, and E(B-V) are bimodal. We find that the age is in general poorly constrained for the majority of our sources, as is often the case when fitting spectral energy distribution models to broadband photometry without the use of UV bands (Ciesla et al., 2015). Whilst the example shown in Fig. 3.4 is more bimodal than the majority of our objects, the degeneracies visible highlight a typical issue for composite host galaxy and AGN SED fitting: a dust-reddened quasar residing in a young star-forming galaxy is often degenerate with an unobscured AGN in an older galaxy. The posterior distributions shown in E(B-V) of Fig. 3.4 are also often non-Gaussian with visible tails. For completeness, when comparing AGN and host galaxy properties, we include within our figures both the distribution of all of the MCMC inferences from each object, along with the corresponding median solutions inferred for each property. Both our median and all MCMC inferences show similar distributions for the AGN and host galaxy properties that we investigate in this thesis.

Before summarising the main results from our work, we make one final cut to the sample, to remove objects where the highest likelihood solution still provides a reduced $\chi^2 > 3$, which we calculate from the AIC using equation 2.8. This cut was chosen to be fairly liberal to account for the possible under-prediction of uncertainties in the photometric data, whilst still removing objects with poor fits based on a visual inspection. Our sample after these final cuts contains a total of 510 objects, which is comparable in size to previous work by e.g. Lanzuisi et al. (2017) in the COSMOS field. Of these 510 objects, 437 preferred a AGN+GAL model family, either with emission line properties that are allowed to vary within the fitting run (18.8%), or fixed to the average seen in AGN at $z = 2$ and an average absolute magnitude $M_i = -27$ (81.2%). Of the remaining objects that had a reduced $\chi^2 < 3$, 71 instead preferred a galaxy-only model family, and 2 an AGN-only model family.

In order to study the difference between obscured and unobscured AGN within our sample, we also separated our 437 AGN+GAL model family objects using their measured X-ray hardness ratio provided by Chen et al. (2018). A hardness ratio of -0.2 was selected as the threshold, with hardness ratio values below and above this value corresponding to unobscured and obscured objects respectively (Hasinger, 2008). This hardness ratio cut provided 316 unobscured AGN (72% of the AGN+GAL sample), and 121 obscured AGN

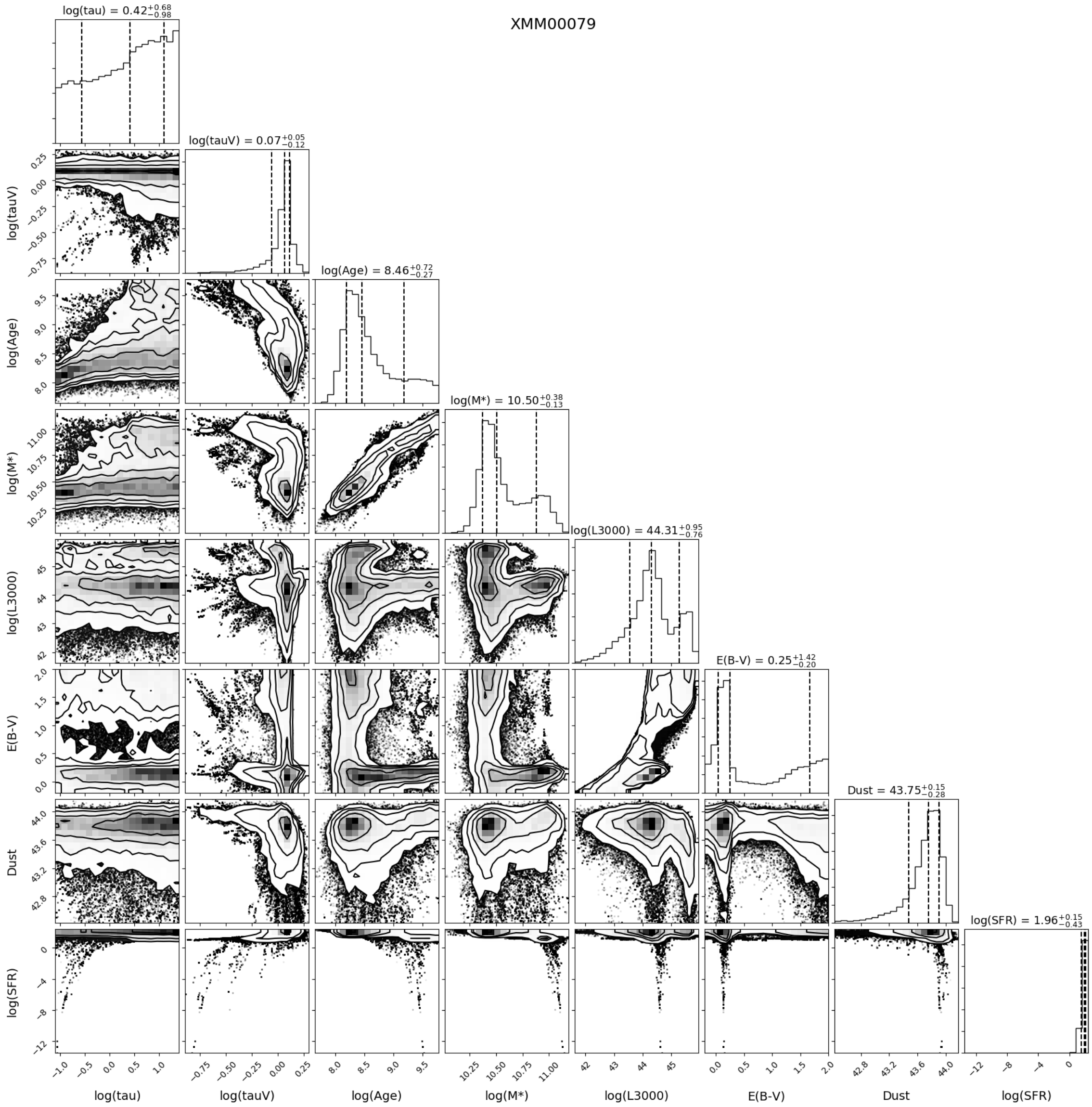

Fig. 3.4 An example of the marginalised one- and two dimensional posterior distributions for an unobscured (HR = -0.67) AGN at $z = 0.93$. In this example, we can see the degeneracies between a young, low stellar mass, star forming galaxy and an older, redder galaxy with a higher luminosity AGN. These degeneracies can be seen in the bimodalities of the host galaxy stellar mass and age, and AGN optical luminosity and extinction.

(28% of the AGN+GAL sample). For the 71 objects found to be dominated by the galaxy component, all but 18 only had upper limits on either the hard- or soft X-ray detections, such that a hardness ratio could not be calculated. Of the remaining 18, 13/18 (72%) were found to have a hardness ratio $>$ -0.2. We might expect our objects dominated by the galaxy in the optical and infrared region to also be more heavily obscured in the X-ray, which appears to be the case, especially when we consider that the majority of our galaxy sample appears to be so obscured that a definitive measurement was not possible within one of the X-ray bands.

In Fig. 3.5 we show example SED fits from the four model families of objects: AGN dominated (bottom-left), galaxy dominated (bottom-right), AGN+GAL with the AGN emission lines fixed to the average emission line template from Temple+21 (top-right), and AGN+GAL with variable emission line properties (top-left). Each example also meets our reduced χ^2 < 3 cut. We find that the majority of SEDs that are best fit by a AGN+GAL model family have a relatively unobscured AGN dominating the near-UV with old stellar populations from the host galaxy providing flux at longer wavelengths, as shown in the upper two panels of Fig. 3.5. The hot dust emission from the AGN only contributes beyond a rest-frame wavelength of $\sim 1\mu$m. Our inferences from the SED fits are broadly consistent with the HSC *gri* colour composite images shown in the inset panels. For example, in the top-right source XMM00076, the HSC image shows a blue point source surrounded by extended red emission. This is consistent with the bluer flux being dominated by the quasar and the redder flux having more significant contributions from the extended host galaxy. The upper left hand panel shows the best fit for XMM01134 using non-standard emission lines, preferring stronger, more symmetric lines compared to the average SDSS quasar. These lines appear to contribute significantly to the broadband flux in the *y*- and *J* bands. The subset of AGN fit with atypical emission lines will be discussed further in Section 3.3.4.

The bottom-right panel of Fig. 3.5 shows an example of one of the 71 objects that preferred a galaxy-only fit based on our AIC selection. This subset of objects is still actively accreting based on the detection of significant X-ray emission, but the AGN contribution is obscured across the entire optical and infrared region of the spectrum. The average redshift of the galaxy-dominated objects is $z = 0.89$, significantly lower than the AGN+GAL class of objects, which have an average redshift of $z = 1.53$. We consider the difference between the HSC *g*-band cModel and PSF magnitudes provided by HSC DR2 (Aihara et al., 2019) as a simple measure of extendedness at the bluer wavelengths and find that the AGN+GAL class of objects have an average gcModel-gPSF of -0.12 compared to -0.63 for the galaxy dominated class of sources. This supports the hypothesis that the galaxy-dominated sources are indeed more extended in the HSC images. Our galaxy-dominated sample therefore

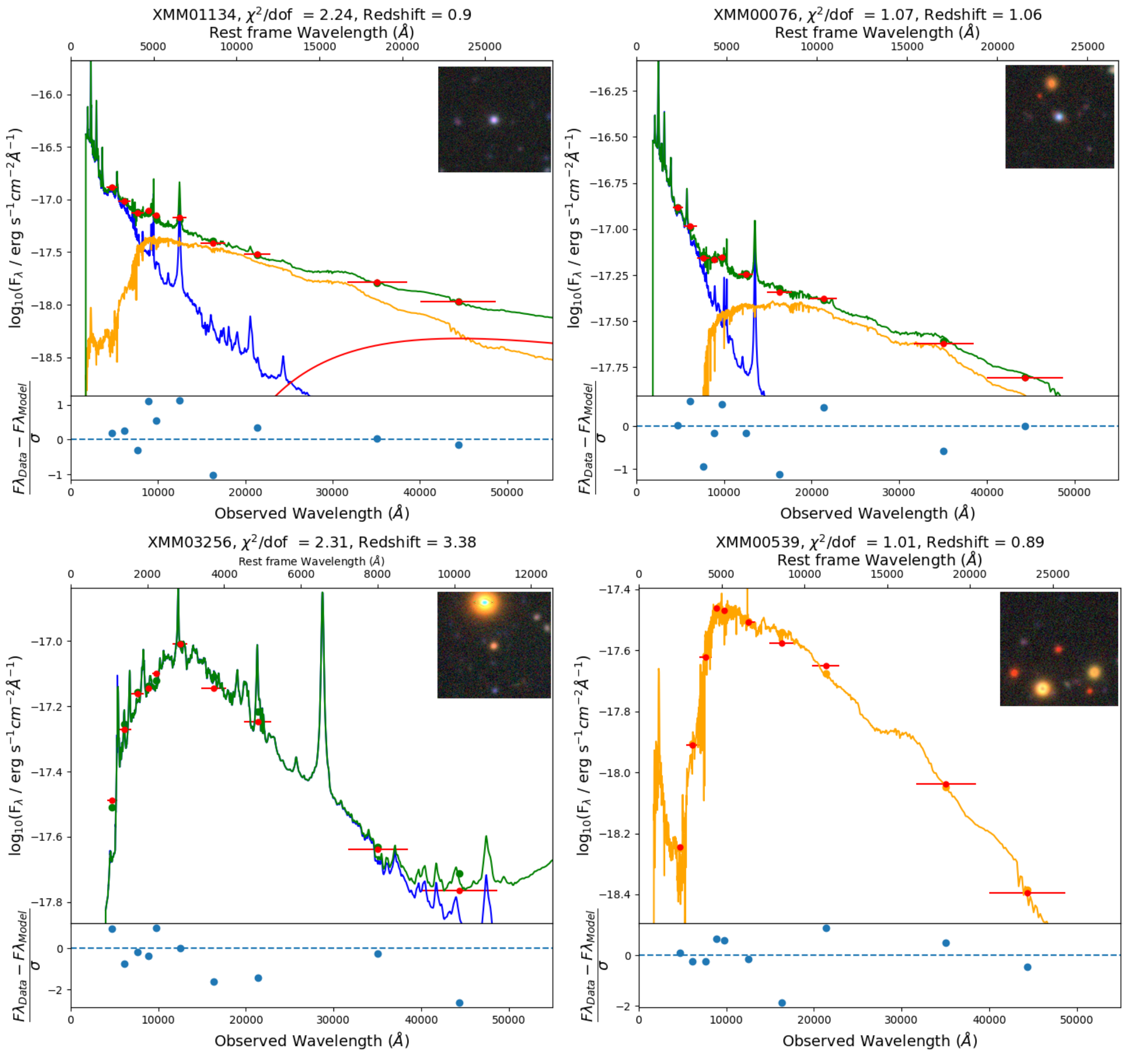

Fig. 3.5 Example fits for the four model families. The top figures show examples of fits with both the galaxy and AGN components, one with a free emission line (left), and the other with emission line properties fixed to the average seen in an SDSS quasar at $z = 2$ AGN, with an average absolute magnitude $M_i = -27$. The bottom figures show fits where either a AGN-only (left) or Galaxy-only (right) fit is preferred. In each case, the total maximum likelihood spectrum is shown in green. The blue, orange and red spectra represent the AGN, galaxy and AGN hot dust components respectively. The observed photometric data are shown as red points. In each corner is the HSC DR-2 *gri* colour image of the object.

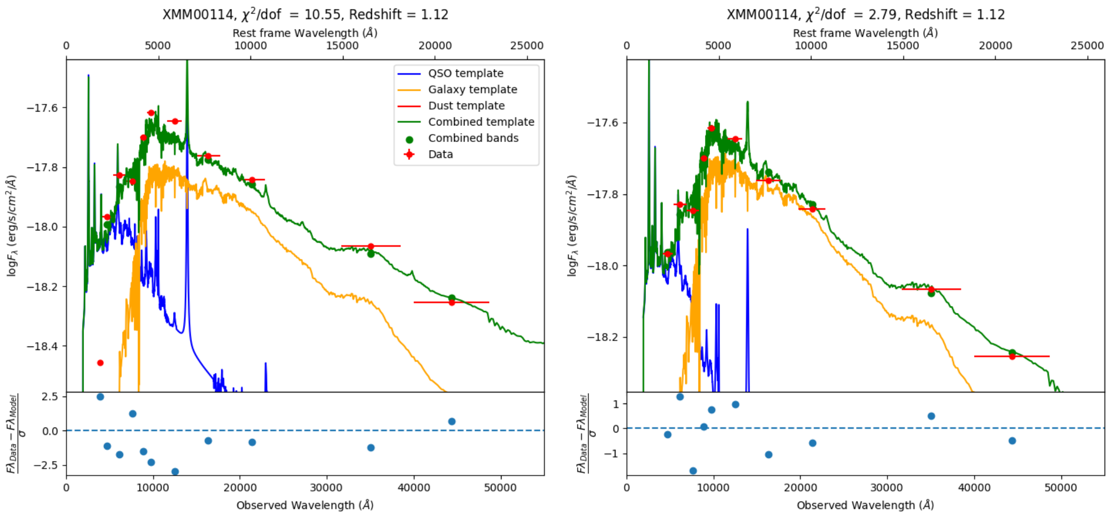

Fig. 3.6 Comparison of the SED fits produced both with (left) and without (right) the inclusion of CFHT *u*-band band (Veillet, 2007). The blue, orange and red spectra represent the AGN, galaxy and AGN hot dust components respectively. The observed photometric data is shown as red points. The inclusion of the *u*-band data severely decreases the quality of the fit, as shown by the increase in reduced χ^2, despite the addition of a data point.

accounts for a subset of AGN with low X-ray luminosities that are only observed within this sample due to being relatively nearby. This results in domination from the host galaxy across the observed wavelength range.

3.3.1 CFHT *u*-band exclusion

As previously mentioned, despite access to CFHT *u*-band data (Veillet, 2007) within the XMM-SERVS field, the decision was made not to include this band within SED fits. Fig. 3.6 shows an example of the inclusion of the *u*-band within an SED fit. The significant drop in *u*-band flux when compared to the neighbouring HSC *g*-band greatly reduces the quality of the fit, despite providing an additional data point that increases the number of degrees of freedom. With the combination of the host galaxy and AGN SEDs included within the SED fitting code, it is unclear how such a *g*-band to *u*-band drop could occur, suggesting a possible issue with the *u*-band aperture photometry.

We specifically focus on the CFHT *u*-band as the cause of the fitting issue due to the fact that, as we have found that it is possible to gain reasonable ($\chi^2 < 3$) fits from using the rest of our available XMM-SERVS matched photometry (Chen et al., 2018) without this band. The reason for the observed *u*-band discrepancies may be related to the uncertainties

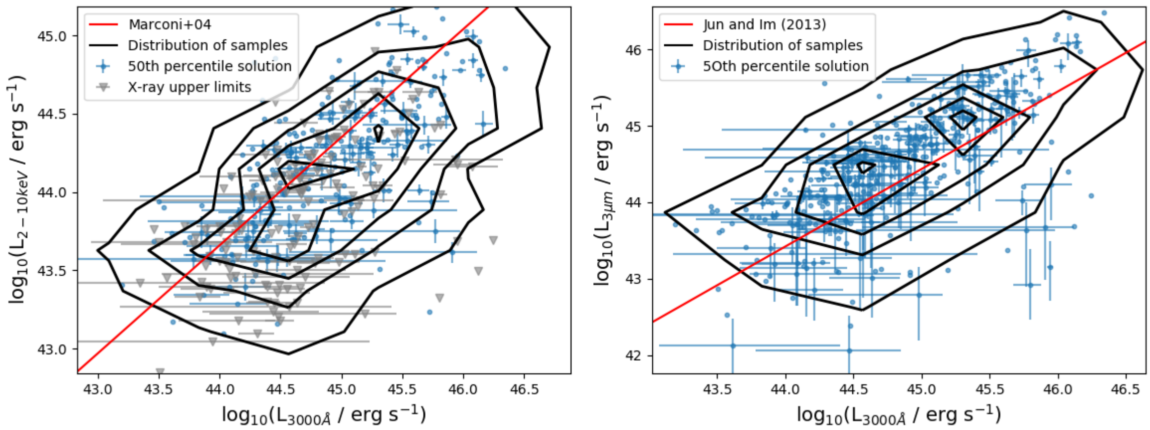

Fig. 3.7 XMM-Newton X-ray luminosity vs AGN optical luminosity at 3000Å with the relation derived from Marconi et al. (2004) (left) and the 3000Å optical luminosity vs hot dust luminosity at 3μm with the relation from Jun & Im (2013) (right). The blue data points show the median solutions and uncertainties for these objects, and the contours show the combined inferences from our MCMC analysis. For clarity, uncertainties are limited to a third of the points shown. In the case where X-ray luminosity is plotted, objects where $\log_{10}(L_{2-10keV})$ is an upper limit are shown in grey.

associated with the CFHT u-band photometry. It is possible that uncertainties in this case are under-predicted, leading to overly constrained u-band values that therefore provide poor reduced χ^2. Such an issue may be related to the calculation of u-band flux used to calculate the 3.0″ magnitudes from the CFHT images. Additionally, when compared to optical or near-IR bands from HSC, VISTA and Spitzer, we might expect additional reddening to have a greater effect on the observed flux within the shorter wavelength u-band, leading to the larger discrepancy in CFHT data. However, the case seen in the example in Fig. 3.6, the observed drop does appear large an non-physical when compared to our model SED, even taking into account greater u-band reddening sensitivity. For these reasons, SED fits within this thesis will not include u-band photometry within the analysis of our SED fits, focusing instead on the 10 bands from HSC, VISTA and Spitzer SERVS (Aihara et al., 2018; Jarvis et al., 2013; Mauduit et al., 2012).

3.3.2 AGN Luminosity & Obscuration

We now use our SED fits, without the use of u-band data, to further explore the inferred AGN luminosities and obscuration for the XMM-SERVS spectroscopic sample. As stated in Section 3.3, due to the non-Gaussian nature of many of the 1-D marginalised probability distributions, we choose to show both the median solutions with 68% uncertainties for each object, along with the contours corresponding to every MCMC inference in the posterior distribution for the 437 well-fit AGN+GAL model family objects. The left side of Fig. 3.7 shows the 2-10keV X-ray luminosity, $\log_{10}(L_{2-10keV})$, calculated from the X-ray flux in the catalogue from Chen et al. (2018), as a function of the extinction-corrected 3000Å luminosity, $\log_{10}(L_{3000})$, inferred from the AGN contribution to the SED. The contours represent the density of MCMC inferences that probe the full posterior distributions of the 3000Å luminosity. The straight line is the relation derived from Eq. 21 in Marconi et al. (2004):

$$log_{10}(L_{2-10keV}) = 0.69log_{10}(L_{3000}) + 13.3. \tag{3.1}$$

We see that the distribution of 3000Å luminosities is consistent with the relation, with a 1σ scatter of ± 0.4 dex around the Marconi et al. (2004) line for all of the sources in the sample. As the X-ray luminosities are independent of our SED fits, agreement of the AGN optical luminosity with the predictions from the Marconi et al. (2004) relation serves as a useful validation of our method.

To further understand the relationships between AGN luminosities, we also look at our inferences on the luminosity of the hot dust surrounding the SMBH. The right panel of Fig. 3.7 shows the distribution of optical luminosity at 3000Å with the AGN hot dust luminosity at $3\mu m$, $\log_{10}(L_{3\mu m})$, for the same sample of X-ray AGN. We also show the empirically derived relation from Jun & Im (2013):

$$log_{10}(L_{2.3\mu m}) = (1.014 \pm 0.002)\ log_{10}(L_{0.51\mu m}) - (0.655 \pm 0.076). \tag{3.2}$$

The Jun & Im (2013) relation compares the hot dust luminosity at $2.3\mu m$ and the AGN luminosity at 5100Å. We therefore convert these values to the optical luminosity at 3000Å and the hot dust luminosity at 3μm. As we are not changing the shape of the unreddened Temple+21 AGN template and dust blackbody, we can convert the luminosities as:

$$log_{10}(L_{2.3\mu m}) = log_{10}(L_{3.0\mu m}) - 0.058 \tag{3.3}$$

$$log_{10}(L_{5100}) = log_{10}(L_{3000}) - 0.592. \tag{3.4}$$

Even after making this conversion, our points are offset by -0.23 dex from the Jun & Im (2013) relation, with our SED fits inferring higher $3\mu m$ luminosities for a given 3000Å luminosity. A key difference between our analysis and that of Jun & Im (2013) is the incorporation of the new quasar template from Temple et al. (2021), which is believed to provide a more accurate representation of the intrinsic quasar SED without contamination from the host galaxy emission. The work by Jun & Im (2013) modelled the AGN SED as a power law continuum and hot dust blackbody emission, but did not include the contribution of emission lines. It is therefore possible that the observed discrepancy is due to the intrinsic difference in the AGN SED templates used. The effect of the inclusion of emission lines within AGN templates during SED fitting requires further study, and will form the basis of Chapter 4 of this thesis. The work by Jun & Im (2013) does also extend out further into the infra-red, making use of 3.5μm and 9μm data where available. In order to account for these additional data within the SED fits, further hot dust blackbodies at 500 and 200K were also added. In the cases where these additional blackbodies are included, it may be the case that some of the high temperature flux is attributed to a lower temperature blackbody. If Jun & Im (2013) only used their 1250K blackbody to calculate their hot dust luminosity, this could lead to an under-prediction that might cause an offset similar to that seen in Fig. 3.7. However, the use of further infrared data would not be expected to have a significant effect on the the specific optical and hot dust luminosities included within Fig. 3.7, which focus on a hot dust output at 3μm. Additionally, as this only applies to a subset of the total Jun & Im (2013) sample of AGN, it is unlikely that such a explanation would lead to the observed offset. Instead, it is possible that the removal of host galaxy contamination from the Temple et al. (2021) template might also relate to the Jun & Im (2013) offset. Re-adding a galaxy contribution to the Temple et al. (2021) template has the effect of shifting the Jun & Im (2013) line upwards by changing the conversions shown in equations 3.3 and 3.4. However, the re-added galaxy contribution within our new AGN template is dependant on redshift. Therefore, the extent of the shift in the Jun & Im (2013) relation is dependant on the redshift of the galaxy template. As the redshift differs for each object within our sample, understanding this effect would require further analysis beyond the scope of this thesis.

Separating out AGN by hardness ratio we found that the $L_{2-10keV}$ vs L_{3000} and $L_{3\mu m}$ vs L_{3000} distributions of unobscured (HR < -0.2) and obscured (HR > -0.2) objects appear similar, with the main difference being a larger spread on the distribution of inferences for the obscured sample of AGN. The 1σ scatter compared to the optical and X-ray luminosity relation from Marconi et al. (2004) increases from $\pm$0.4 dex to $\pm$0.5 dex from the X-ray unobscured to the obscured AGN. Whilst the AGN luminosity distributions are relatively similar, differences in the measured optical extinction between X-ray unobscured and obscured AGN are more significant, as can be seen on the left of Fig. 3.8. The histogram of obscured E(B-V) inferences has a tail extending to higher E(B-V) values, whereas the inferences from

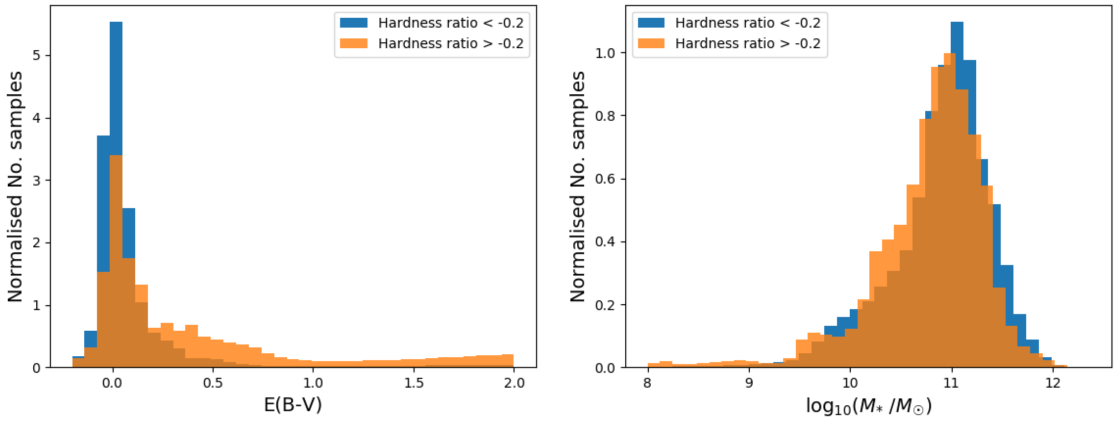

Fig. 3.8 Histograms showing the distribution of E(B-V) (left) and host galaxy stellar mass (right) MCMC inferences, separated by the measured X-ray hardness ratio. This ratio acts as a proxy for the AGN obscuration type. We have assumed values above and below HR = -0.2 to represent obscured and unobscured AGN respectively. The obscured AGN show a tail to higher extinction values when compared to the unobscured sample.

the sample of unobscured objects are tightly peaked at E(B-V) = 0.02. This difference is expected, as, assuming the unified theory of AGN, flux from obscured AGN passes through a larger amount of dust, leading to greater extinction. If we assume that obscured AGN generally have higher extinctions, AGN in older host galaxies will provide a majority of their flux in the same wavelength region as their host galaxies. Balancing of the contribution of flux from these two components therefore leads to more complex degeneracies between the AGN luminosity, AGN extinction and host galaxy stellar mass and may therefore extend the distribution of AGN optical luminosity estimates. Such an effect can be seen in the 1-D posterior distributions of these properties shown in Fig. 3.4.

3.3.3 The stellar mass-AGN luminosity relation

To understand how host galaxy stellar mass inferences relate to AGN properties, we first look at the total distribution of derived host galaxy stellar masses, separated by the hardness ratio proxy for AGN type. These are shown in the right of Fig. 3.8. Our findings show consistent host galaxy masses for unobscured and obscured AGN, having median mass values[1] of $\log_{10}(M_*/M_\odot) = 10.88 \pm 0.09 M_\odot$ and $\log_{10}(M_*/M_\odot) = 10.8 \pm 0.1 M_\odot$ for unobscured and obscured AGN respectively. Previous work by Zou et al. (2019) and Suh et al. (2019) have suggested a link between host galaxy stellar mass and AGN type. Their results varied, with

[1]The calculation of these values is discussed further in Appendix A

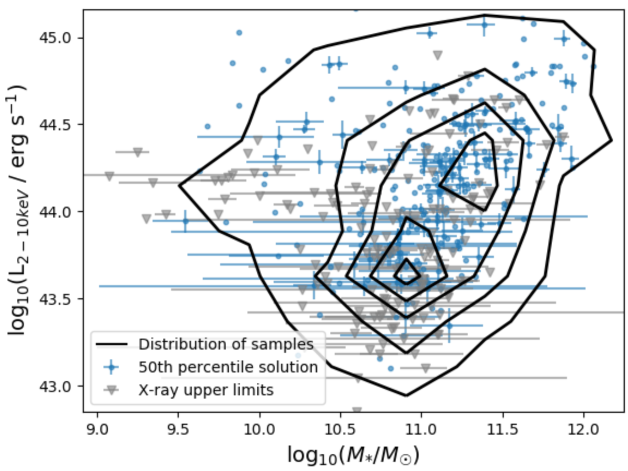

Fig. 3.9 Host galaxy stellar mass MCMC inferences vs the measured X-ray luminosity from XMM-Newton. The blue points correspond to the median solutions for each object. For clarity, uncertainties limited to a third of the objects are shown. For objects where $\log_{10}(L_{2-10keV})$ is an upper limit, median solutions are shown in grey. Contours show all of the MCMC inferences for these objects.

the former finding unobscured AGN are typically found inhabiting less massive host galaxies than obscured AGN, and the latter finding the opposite. Both Zou et al. (2019) and Suh et al. (2019) do however differ in their methods of AGN selection when compared to this work, instead using the presence of broad emission lines in the observed spectra to define type 1 AGN.

In order to further understand the underlying link between AGN and their host galaxies, we also looked at the inferred correlations between our AGN dust, optical, and X-ray luminosities with host galaxy stellar mass. We found that the strongest correlation of the AGN luminosities with stellar mass is with the X-ray luminosity, with a correlation coefficient of 0.15 ± 0.03. The uncertainty on this value was calculated using a Monte Carlo method using MCMC samples similar to the calculation of property uncertainties as described in Appendix A, and shows a statistically significant positive correlation between AGN and host galaxy properties.

This is slightly larger than the stellar mass-AGN optical luminosity, and stellar mass-hot dust luminosity correlations, which are 0.13 and 0.14 respectively. Focusing on the X-ray luminosity-stellar mass correlation, Fig. 3.9 shows the contours produced from our MCMC inferences comparing the measured X-ray luminosity with the stellar mass estimates, along with points corresponding to our median solutions. As in our previous analysis, both unobscured and obscured objects were found to show similar correlations, and as such are shown as a single distribution within Fig. 3.9. A positive correlation between X-ray luminosity and host galaxy stellar mass has been also been found in previous work, such as Magliocchetti et al. (2020). Whilst statistically significant, the stated correlation coefficient is still relatively weak. As can be seen in Fig. 3.9, there is a relatively large scatter on the uncertainties for our stellar mass estimates, with a tail in the sample contours that extends to lower mass inferences. It is therefore possible that the combination of a small number of low mass samples from each fit is masking a strong relation between host galaxy stellar mass and AGN X-ray luminosity. However, it is also possible that we might generally expect the relation between host galaxy and AGN properties properties to be weak, due to the difference in timescales on which such properties are observed to occur. The stellar mass of a host galaxy is expected to grow as additional stars are formed on million year timescales. On the other hand, AGN X-ray emission is a result of the accretion of material onto the central black hole. As the density of the accreting material can vary, the X-ray output can also be highly variable on much shorter timescales of the order of weeks. Whilst we therefore might generally expect a positive correlation between host galaxy stellar masses and AGN luminosity, the snapshots provided by X-ray measurements might lead to a relatively weak correlation, due to these timescale differences.

In terms of evolution with redshift, the median values of stellar mass MCMC inferences show no clear difference between galaxies at redshift < 1 compared to those at $z > 2$, with these values from our samples changing from $\log_{10}(M_*/M_\odot) = 10.83 \pm^{0.09}_{0.09} M_\odot$ to $\log_{10}(M_*/M_\odot) = 10.9 \pm^{0.1}_{0.1} M_\odot$. We see the uncertainty on stellar mass estimates increases with redshift. The uncertainty increase could be due to the fact that whilst the stellar mass values do not decrease, high redshift objects are likely observed due to the presence of a high luminosity AGN, which can lead to more complex degeneracies that inflate the stellar mass uncertainties.

3.3.4 AGN emission line properties

A novel feature of the Temple et al. (2021) quasar SED model is the incorporation of quasar emission line templates that reflect the full diversity of emission line strengths and morphologies seen in high-redshift quasar spectra. The details of these emission lines are further described in Section 2.2.2. A key question is whether these different emission line properties materially impact the broadband colours of quasars in such a way that the quasar emission line properties can be inferred from SED fitting. In order to conduct this test, the

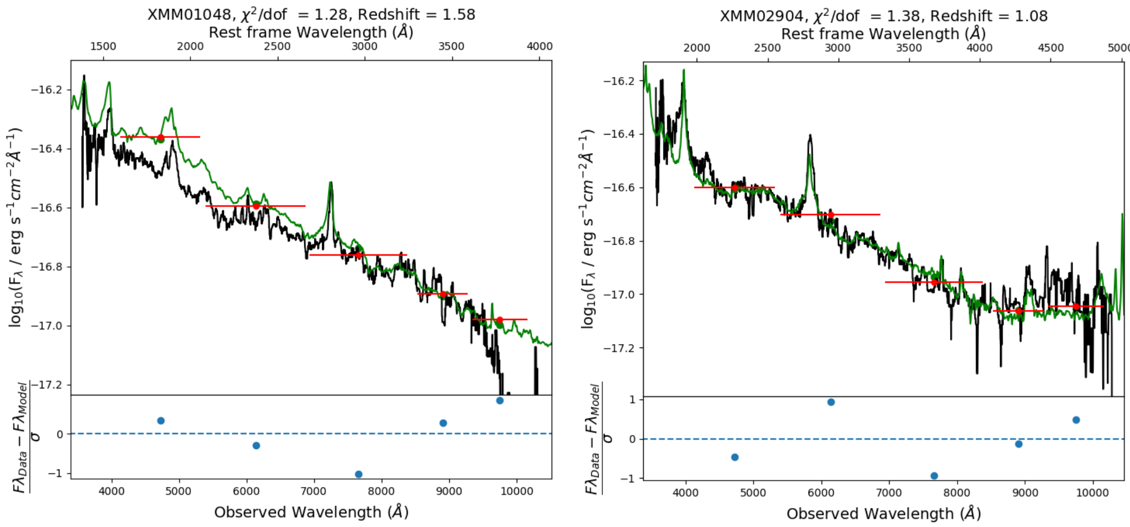

Fig. 3.10 SDSS spectra (in black) compared to the highest likelihood SED, including the host galaxy, AGN and hot dust components (in green) produced using SED fitting. The SDSS spectrum are normalised to the best fit SED to have the same *i*-band flux. The figures show examples of an object that preferred weaker, more blue shifted lines on the left, and stronger, more symmetrical lines on the right, compared to the emission lines seen in AGN at $z = 2$ and an average absolute magnitude of $M_i = -27$. In the left figure, whilst the emission line strength in the best fit SED appears to show good agreement between the width and strength of the lines observed, either AGN variability, or an offset in spectra calibration appears to change the continuum emission between the SDSS spectrum and best fit SED.

SDSS Data Release 16 quasar catalogue (Lyke et al., 2020) was matched to the XMM-SERVS sample of X-ray AGN to provide a sub-set of 408 AGN where emission line properties can directly be inferred from the spectra. As detailed in Section 2.3.1, we fit SED models with the quasar emission line properties fixed to the default value, which represents the average emission line properties for luminous, high-redshift SDSS quasars (Temple et al., 2021), as well as SED models where the emission line properties are a free parameter. The model with the lowest AIC is chosen as the best-fit model from all SED fits.

We find that 61 of the 289 XMM-SERVS-SDSS matched catalogue AGN, with a best fit reduced $\chi^2 < 3$, prefer non-standard emission line properties. The average redshift of this sample has a higher mean redshift of $z = 1.86$, compared to $z = 1.46$ for the standard emission line sample. For the fixed emission model family fits, we find that 64% have $z < 1.6$, which is only the case for 39% of the free emission line fits. This difference is expected, as for higher redshift objects, the strong emission line C IV present within the AGN SED is redshifted into the HSC g-band. For the lower redshift objects, this emission line is outside of the range fit by the SED, and thus information on its nature cannot be inferred for this sample.

The right-hand plot in Fig. 3.10 shows an example of one non-standard emission line object, fit using stronger, more symmetric lines (emline_type = 2.92) when compared to the emline_type = 0 emission lines that correspond to the average SDSS AGN at $z = 2$ and average absolute magnitude M_i = -27. The SDSS spectrum for the quasar is over-plotted and demonstrates excellent agreement with the emission line strength independently inferred from the photometry. Visual comparison of all of the best-fit SEDs for the 61 AGN with SDSS spectra that prefer non-standard emission line properties showed that 66% of the AGN show excellent agreement between the spectra and the spectral line strengths inferred from the photometry.

For a further 25% of AGN, the emission line strengths inferred from our SED fits are in reasonable agreement with the SDSS spectra but small differences in the continuum emission between the spectra and the best-fit SED are observed. These could arise for example due to quasar variability, or due to the typical uncertainties in the absolute flux calibration of the SDSS spectra. An example of such an object can be seen in the left-hand panel of Fig. 3.10, which also features an AGN fit with weaker, more highly blueshifted emission lines when compared to the average emission lines seen in AGN at $z = 2$ and an average absolute magnitude M_i = -27. Only 9% of the AGN with non-standard emission line properties were inconsistent with the SDSS spectra. In these cases the best-fit SEDs were often composites of young star-forming galaxies with an obscured quasar, whereas the SDSS spectrum confirms the presence of a relatively unobscured AGN with broad emission lines. The final 6% of objects had very low signal-to-noise SDSS spectra, precluding any firm conclusions regarding their nature.

Having confirmed that emission line properties can be effectively inferred from broadband photometry, we now consider the multi-wavelength properties of these sources in more detail. From the total sample of 61 AGN with non-standard emission line properties, 61% preferred stronger, more symmetric lines, with the remaining 36% preferring weaker, more highly blueshifted lines relative to the Temple et al. (2021) average quasar SED. The AGN with stronger, symmetric lines have an average 3000Å luminosity of $\log_{10}(L_{3000}/\mathrm{erg\ s^{-1}})$ = $44.9\pm^{0.2}_{0.2}$, similar to the average AGN luminosity for the sample best fit with the default emission line template in the Temple et al. (2021) model of $\log_{10}(L_{3000}/\mathrm{erg\ s^{-1}})$ = $44.7\pm^{0.1}_{0.1}$. On the other hand, the AGN that prefer weaker, more highly blue-shifted emission lines, indicative of line driven winds, have a higher average luminosity of $\log_{10}(L_{3000}/\mathrm{ergs^{-1}})$ = $45.2\pm^{0.2}_{0.2}$. These results confirm the well-known Baldwin effect – namely that the equivalent widths of strong emission lines in quasar spectra are anti-correlated with the AGN luminosity (Baldwin, 1977). Confirmation of this result from broadband SED-fitting however, gives us an additional confidence that our SED fits are indeed able to correctly infer emission line properties.

We also consider the stellar mass of the AGN host galaxies as a function of their emission line properties. AGN with stronger, more symmetric lines have an average $\log_{10}(M_*/M_\odot)$ = $10.9\pm0.1M_\odot$, which is consistent with the average for AGN best-fit by the default emission line template $\log_{10}(M_*/M_\odot)$ = $10.86\pm0.08M_\odot$. AGN with weaker, more highly blueshifted lines tend to have lower host galaxy stellar masses, but are still consistent within the associated uncertainties, with an average mass of of $\log_{10}(M_*/M_\odot)$ = $10.7\pm0.3M_\odot$.

The ratio of the X-ray luminosity at 2keV and the UV luminosity at 2500Å is often used to represent the hardness of the AGN ionising SED and can provide insight into the connection between the X-ray corona and the accretion disk of the AGN. We calculate $\log_{10}(L_{2500})$ using the same approach as described in Section 3.3.2. To find $\log_{10}(L_{2keV})$, we use the hard (2-10keV), and soft (0.5-2keV) X-ray bands from XMM-Newton. For each pair of robust X-ray measurements, we calculate a power law relationship for the flux density, and use this to k-correct these data to 2keV. This provides a distribution of photon indices, with an average value of $\gamma = 1.5$. This was used as the photon index for the objects in the sample where X-ray measurements only provided an upper limit in either the hard or soft X-ray bands, and as such individual photon indices could not be calculated. These values were used to calculate α_{ox}. α_{ox} is defined as the slope of the power law between the X-ray and UV luminosities using the relation:

$$\alpha_{ox} = 0.384Log_{10}(L_{2keV}/L_{2500}) \tag{3.5}$$

Our inclusion of emission line properties within the SED fits, and confirmation of their validity, allows us to compare our inferred power law slope to the AGN emission line properties. Previous work, such as Timlin et al. (2020), has found a significant correlation between α_{ox} and both the equivalent width and blueshift of the CIV emission line. Fig. 3.11 shows the distribution of MCMC inferences for the sub-selection of 82 AGN+GAL model family objects that preferred non-standard emission line properties. It is important to note the bias in this sample, in that it only contains objects with a measurable difference in broadband flux as a result of the varying emission line properties when compared to the average $z = 2$, absolute magnitude M_i = -27 AGN. In Fig. 3.11 we see a positive correlation between α_{ox} and emission line properties for our inferences, with a correlation coefficient of 0.58. This suggests that the weaker, blueshifted emission line objects, with smaller emission

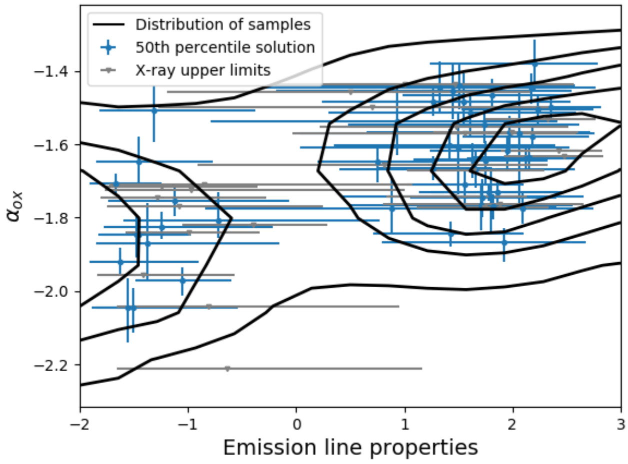

Fig. 3.11 The X-ray to UV power law slope, α_{ox}, vs the inferred AGN emission line properties. More negative values of emission line properties are indicative of line driven disc winds resulting in weaker, more blueshifted lines, whereas more positive values are stronger and more symmetric. The sample shown does not include objects that preferred a fixed emission line within their fit, based on our AIC designation (such that the emission line properties value is fixed to 0).

line equivalent widths, typically have softer AGN SEDs when compared to the stronger, more symmetric emission line objects. Previous studies (Richards et al., 2011; Timlin et al., 2020) have found similar results using SDSS spectra. Timlin et al. (2021) for example used X-ray data from the Chandra X-ray Observatory to gain $alpha_{ox}$ estimates for their sample of 2106 AGN. Figure 5 of the Timlin et al. (2020) paper is analogous to the results shown in Fig. 3.11, where the rest frame equivalent width of the CIV emission line is shown instead of our emission line inferences. Timlin et al. (2020) provide an average uncertainty on their α_{ox} values of around 0.05, along with a Spearman rank-order test statistic value of 0.470 for the scatter associated with this fit. These values align well with the results shown in Fig. 3.11, which provides an average α_{ox} uncertainty of 0.55 and Spearman value of 0.466, showing a similar scatter on this relation. However, our work shows it is possible to see such a trend using broadband photometry and SED fitting, compared to Timlin et al. (2020), which required spectra for their CIV equivalent width values.

3.4 Conclusions

We performed SED-fitting to a spectroscopically confirmed sample of 711 high-redshift ($z > 0.7$) X-ray selected AGN in the XMM-SERVS survey using optical and near infrared photometric data from the HSC Deep/UltraDeep, VISTA VIDEO and *Spitzer* SERVS surveys. 510 of these AGN were found to have reliable fits across all 10 filters and a reduced $\chi^2 < 3$.

We used the Aikake Information Criterion to classify these 510 AGN, and found 71 objects are galaxy-dominated, and 2 objects can be fit by a pure AGN template across the chosen wavelength range. 437 objects require both AGN and host galaxy components in the SED, thereby allowing the link between AGN and host galaxy properties to be investigated.

We used a newly developed AGN SED model from Temple et al. (2021) to characterise the AGN properties. The model incorporates a concise parametrisation of the variation in emission line properties across unobscured AGN, thereby allowing us to potentially infer AGN emission line properties from broadband photometry. In order to study the effects of bimodal or solutions on our inferred properties, we investigate both the median solutions and full posterior distribution of MCMC inferences. These two methods are shown to provide consistent results for the properties investigated in this work.

- We find that the AGN X-ray luminosity is correlated with the 3000Å luminosity inferred from SED-fitting in good agreement with previously known relations (e.g. Marconi et al. 2004). These results demonstrate that robust AGN luminosities can be inferred from the use of our SED fitting code.

- Comparison with previous work found an offset of -0.23 dex between the AGN optical luminosity and hot dust luminosity at $3\mu m$ from the relation found by Jun & Im (2013). This offset was possibly caused by the differences in the AGN templates used in this work and theirs.

- We used the X-ray hardness ratio (HR) to split the AGN into X-ray obscured (HR > -0.2) and X-ray unobscured (HR < -0.2) AGN. This showed a consistent distribution for both obscured and unobscured samples in terms of the extinction-corrected AGN and hot dust luminosites at 3000Å and $3\mu m$ respectively. As might be expected for obscured AGN, an extended tail to larger E(B-V) was seen in the MCMC inferences, suggesting a generally higher extinction for obscured AGN. We also find similar stellar masses for the sample of unobscured AGN ($\log_{10}(M_*/M_\odot)$ = 10.88 $\pm 0.09 M_\odot$ and $\log_{10}(M_*/M_\odot)$ = 10.8 $\pm 0.1 M_\odot$ for HR < -0.2 and HR > -0.2 AGN respectively).

- The ability to vary emission line properties in the Temple+21 AGN template used in SED fitting allowed us to determine that for 18.8% of the reduced $\chi^2 < 3$ sample, non-standard emission line properties were preferred over the average $z = 2$, absolute magnitude $M_i = -27$, AGN emission lines. By comparing to the SDSS spectra available for a subset of 82 AGN, we found that the emission line strengths inferred via SED-fitting to broadband photometry were broadly consistent with the results from spectroscopy for $\sim$91% of the sample. These results highlight that the current generation of precision photometric data-sets are able to infer emission line properties of broad-line AGN from photometry alone for a subset of AGN.

- We calculated α_{ox} based on the measured X-ray luminosity and the rest-frame UV luminosity inferred from our SED fits. We found a correlation between α_{ox} and the emission line properties inferred from photometry. This correlation showed that weaker, more blueshifted emission lines, indicative of line driven winds, were found to occur with softer α_{ox} slopes. In contrast, stronger, more symmetric emission lines preferred harder α_{ox} slopes. This is consistent with previous works (Richards et al., 2011), but hasn't previously been found using photometry alone.

The insights gained from our SED modelling could be used to aid in target selection for upcoming spectroscopic surveys such as 4MOST (Merloni et al., 2019) and VLT-MOONS (Maiolino et al., 2020b). This will be discussed further in Chapter 5.

Chapter 4

A case study on the effect of AGN template on host galaxy stellar mass inferences

4.1 Introduction

Due to the vastly different scales of the host galaxy and its central SMBH, the exact origin of the link between AGN and host galaxy properties is still not fully understood. Observations such as those by Marconi et al. (2004) have suggested the majority of SMBH mass growth occurs during ‘active‘ stages of mass accretion from surrounding gas and dust. Active SMBH accretion then produces winds- and large amounts of energy, both of which can have the effect of halting further star formation within the host galaxy, by either removal (Cicone et al., 2014; Lyke et al., 2020; Müller-Sánchez et al., 2011) or heating (Barišić et al., 2017; Best et al., 2005; McNamara & Nulsen, 2007) of the surrounding gas. Such events produce a link between the SMBH and host galaxy properties, and are typically referred to as active galactic nuclei (AGN) feedback. By studying properties such as host galaxy stellar mass, along with AGN properties such as luminosity, we are therefore able to further understand if AGN feedback has led to the galaxy-SMBH mass relations we see in the local Universe.

Understanding how SMBH-host galaxy relations develop over time requires observations of both local and distant SMBH and host galaxies for comparison. Such observations assist in understanding the timescale and processes by which the SMBH-host galaxy coupling occurs. However, inferences on the properties of SMBHs via stellar or gas dynamics are not possible for objects at high redshifts due to the available spatial resolution. In order to study high redshift SMBH, we therefore rely on the subset of highly luminous ($\approx 10^{45} ergs^{-1}$) AGN, which are visible within distant galaxies. In the case of bright AGN however, their flux is so great as to dominate over the host galaxy contribution, making it hard to infer the host galaxy properties needed to understand co-evolution. Within the optical, the AGN dominates in the bluest regions, whereas the hot dust surrounding the central black hole dominates in the infrared.

There is, however, a spectral region around 1μm, where the host galaxy contribution to the flux is at its maximum, as can be seen in Fig. 1.4. It is therefore possible to measure AGN, hot dust and host galaxy properties simultaneously, allowing for comparisons of observations with simulations. SED fitting can be used to gain information on AGN and host galaxy properties using data in the optical and infrared region. However, significant degeneracies can occur within SED fits (as seen in Fig. 3.4). In the case of empirically derived AGN templates, these degeneracies can make disentangling non-AGN emission, in order to create a pure AGN template, difficult. Additional residual host galaxy contamination could significantly affect our estimations of AGN and host galaxy properties.

To this end, this chapter will act as a case study on how the chosen AGN SED template can affect AGN and host galaxy inferences. We will compare our newly developed SED fitting code, which uses the AGN SED from Temple et al. (2021) (hereafter referred to as the T21 template), to a separate AGN SED developed by Richards et al. (2006) (hereafter referred to as the R06 template). The R06 template was chosen for comparison due to its common usage in SED fitting work (e.g. Bongiorno et al. 2012; Merloni et al. 2010). Both the R06 and T21 templates will be applied to the same sample of objects, using data collected within the optical and infrared region of the spectrum. The sample of AGN that will be used for comparison of our two AGN templates originates from the XMM-SERVS survey, specifically X-ray selected objects in a redshift range of $0.7 < z < 4.5$. Based on the hardness ratio provided by XMM-Newton (Chen et al., 2018), the majority of our sample (70%) are type-1 AGN, and as such are relatively unobscured by dust. A more complete outline of the XMM-SERVS catalogue, and the optical and infrared data used within our SED fits is described in Chapter 3.

The outline of this chapter is as follows. In Section 4.1.1, we describe the features of the R06 AGN template, and outline previous work in which it has been used for SED fitting. In Section 4.1.2 we discuss the main differences between the R06 and T21 template (described in Section 2.2.2). Section 4.2.1 gives an assessment of the quality of the SED fitting performed using the R06 and T21 templates, based on a comparison of the reduced χ^2 values. In Section 4.2.2, we compare our SED fits using the R06 and T21 templates to look for discrepancies in host galaxy stellar mass and AGN luminosity inferences. In Section 4.3 we look at the causes behind the discrepancies shown in Section 4.2.2. Finally, in Section 4.4, we look at how the choice of AGN template can affect the median AGN and host galaxy spectra of the sample, calculated from the highest likelihood solutions from each object.

4.1.1 The R06 template

The R06 template is an empirically derived AGN SED developed using 19 bands of photometry collected for a sample of 259 quasars. The observed photometry spans a wavelength range from 2keV to $\sim 24\mu$m, although not all objects have photometry in all bands. In cases where photometry is missing, a 'gap repair' method was utilised to estimate the missing photometry using neighbouring bands and the SED produced by Elvis et al. (1994). Rest-frame luminosities in bins with spacing of 0.02log(ν) were interpolated from the observed photometry of each object, and the resulting SED is the geometric mean of the measurements in each bin.

R06 attempt to account for the presence of host galaxy contamination within their AGN template. For each object in the sample, R06 predicted its r-band host galaxy luminosity using the relation proposed by Vanden Berk et al. (2006), assuming each object is accreting material close to its Eddington limit. The galaxy contribution to the total luminosity is then removed across the entire wavelength range by assuming that the host follows the elliptical galaxy SED produced by Fioc & Rocca-Volmerange (1997).

The R06 AGN template is well-established, and has been used frequently since its production. Assef et al. (2010) and Brown et al. (2019) produced their own AGN SEDs and compared their results to the R06 template. Assef et al. (2010) note the effect of galaxy contamination within the R06 template, which they were motivated to remove within their own AGN template. Merloni et al. (2010) and Bongiorno et al. (2012) have used the R06 template for SED fitting. In the case of Merloni et al. (2010), the simultaneous fitting of a host galaxy provided stellar mass estimates. In a comparison with the trend shown by Häring & Rix (2004) of the SMBH to host galaxy stellar mass ratio of local objects, the authors found a positive offset for the Merloni et al. (2010) SED fits, which predicted higher host galaxy stellar mass estimates. They provide a number of possible explanations for the offset, including incorrect estimations of stellar mass values based on uncertainties in the host galaxy IMF, and the AGN template choice. For the median spectrum calculated from a sub-sample of highly luminous ($L_{bol} > 10^{45.5} ergs^{-1}$) AGN collected from COSMOS, Merloni et al. (2010) also noted a smaller dip occurring around 1μm compared to the flux seen in the R06 template. Merloni et al. (2010) thus suggested that the use of the R06 template in SED fitting may have the effect of maximising stellar mass estimates. The Merloni et al. (2010) findings differ from other work such as Assef et al. (2010); Krawczyk et al. (2013), who claim that the R06 template still contains contamination from the host galaxy. If the R06 template does still contain host galaxy contamination, this would ascribe less flux to any additional galaxy SED included within the fit. Host galaxy contamination within the AGN template would therefore decrease the host galaxy stellar mass rather than maximise it as suggested by Merloni et al. (2010). Assef et al. (2010) highlighted this issue as a motivation to create their own AGN template, in order to address the R06 galaxy contamination.

For an independent comparison, our R06 and T21 fits will be compared to another SED fitting code. The work by Bongiorno et al. (2012) was chosen for this purpose. The SED fitting performed by Bongiorno et al. (2012) uses an X-ray selected sample of objects from XMM-Newton, as is also the case for the data within this thesis. Bongiorno et al. (2012) also make use the R06 AGN template, along with a range of galaxy templates (Bruzual & Charlot, 2003) with exponentially declining star formation histories (e-folding times of 0.1 - 30 Gyr) and galaxy extinctions ($0 \leq$ E(B - V)$\leq$ 0.5). Unlike the SED fitting code developed for this thesis, the Bongiorno et al. (2012) SFH and extinction values are discretized. Bongiorno et al. (2012) also makes use of a larger wavelength range of photometry, spanning 3800Å from the Canada–France–Hawaii Telescope (CFHT) u-band through to 24μm from the Multiband Imaging Photometer for Spitzer (MIPS).

Overall, the results from Bongiorno et al. (2012) provide a useful comparison for this thesis due to the similarity in both sample and method. The smaller wavelength range of our SED fits also allows us to test if the additional inclusion of UV and mid-infrared data with the Bongiorno et al. (2012) fits has a significant effect on AGN and host galaxy inferences. The results from Bongiorno et al. (2012), and the effects of SED fitting differences on the host galaxy inferences will be discussed further in Section 4.2.2.

4.1.2 Comparison of the AGN Templates

As discussed in Section 4.1, the work within this chapter focuses on the comparison of SED fitting using the well-established R06 template to the T21 AGN template. Using these SEDs, we will study how the choice of template can affect inferences of host galaxy properties. The R06 template differs from the T21 AGN template in a number of key properties, which can be seen in Fig. 4.1. Here, the templates have been normalised to have the same flux at 3000Å.

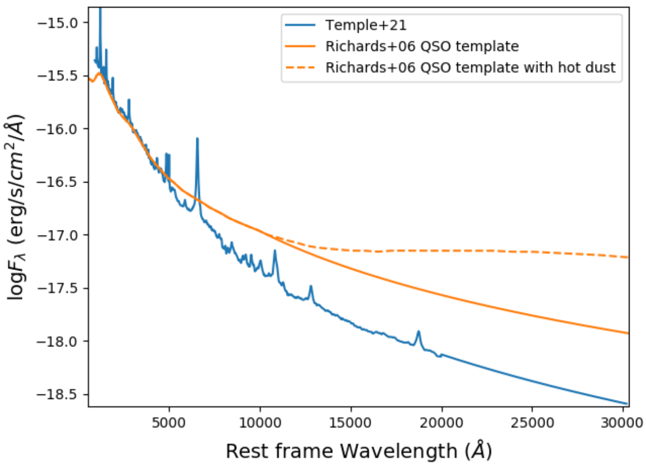

Fig. 4.1 Comparison of the R06 and T21 AGN templates used within this work. Both templates have been normalised to have the same flux at 3000Å. In order to directly compare the R06 and T21 templates, the fixed dust included within the R06 template (shown as a dashed line in orange) has been removed, and a separate template of a 1236K blackbody is re-added.

Whilst both templates are empirical in nature, the R06 template is formed using 19 bands of photometry from AGN selected in the infrared. The T21 template is instead a parametric model created via a combination of emission lines and power laws. The T21 template was matched to the broadband photometry of stacked SDSS quasars in order to accurately recreate observed AGN colours. Due to its origins from photometry, the R06 SED is lower resolution, but with a larger wavelength range than the T21 template. Whilst the R06 template does include emission lines, the lower resolution smooths the flux associated with these lines across a greater wavelength range, such that individual lines cannot be seen within the spectrum.

The motivation behind the comparison of the T21 and R06 templates is to assess how the differences in resolution and continuum affects inferences of host galaxy and AGN properties. If such differences are significant, the choice of AGN SED used within fits could lead to a different understanding of the evolution of galaxies in the Universe. The lower resolution of the R06 template has the effect of smoothing the contribution of emission lines across a wider wavelength range than that seen in the T21 template. By focusing on the effect that lower resolution templates can have on model photometry, we are able to understand how this may affect the ability to accurately match to observed photometry. In order to do this, we look at how smoothing our higher resolution T21 template affects g-r colour.

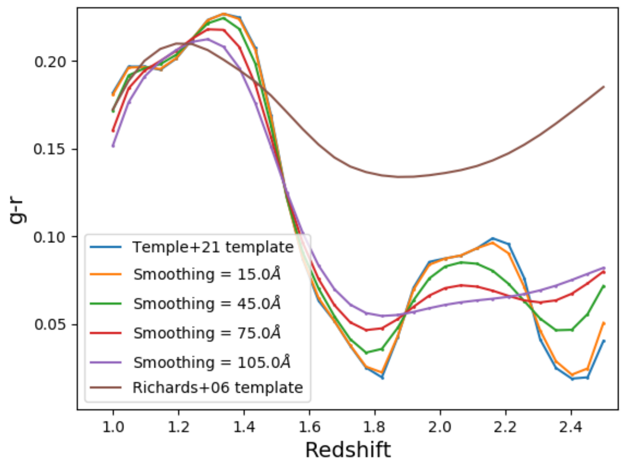

Fig. 4.2 The change in g-r colour as a function of redshift for a range of smoothing of the Temple et al. (2021) AGN template, via a Gaussian filter. The smoothing is compared to the lower resolution AGN template from Richards et al. (2006), shown in brown.

We initially took the T21 template, and applied increasing levels of smoothing via a Gaussian filter. Fig. 4.2 shows the change in g-r colour as a function of redshift for these AGN SEDs, along with the R06 template. Looking at the unsmoothed T21 template shown in blue, we can see a significant feature in the g-r colours between the redshifts of 1.8-2.4. This is due to the presence of a number of emission lines, most significantly Lyα and CIV, moving into, and out of the *g*- and *r*-bands. However, by smoothing range of around 105Å, this feature is no longer present within the g-r colours.

The effect of smoothing causes a maximum change in the g-r colour of around 0.04 within this redshift region, when compared to the original T21 template. Such a change is within the measured uncertainties that would be associated with currently available photometry, and thus would have a noticeable effect on the measured magnitudes. By smoothing a chosen AGN SED beyond 100Å, information on specific emission lines would be lost within the SED fitting. This would mean that the methods used in Section 3.3.4 could not be recreated with the smoothed T21 template. In the production of any AGN SED in the future, it is important to acknowledge that there is a trade-off between a long wavelength range, low resolution SED, and a short wavelength range, high resolution SED. Thus, depending on the level of smoothing utilised, the high resolution emission lines within the T21 template can have a large effect on the observed colours within the infrared and optical region of the spectrum. We therefore might expect smoothing to have a significant effect on the modelling of the observed photometry within our SED fits. By comparing the high resolution T21

SED to a smoothed, lower resolution T21 SED, we are therefore able to test the effect of the presence of emission line smearing on the inferences of AGN and host galaxy properties.

The R06 template already includes the contribution of hot dust within the template (as can be seen in the orange dashed-line SED in Fig 4.1). Within the T21 template, we add a 1236K blackbody into our fitting code as a free parameter, in order to infer the hot dust luminosity at $3\mu m$. To produce a fair comparison of our two template fits, the hot dust contribution within the R06 template was therefore removed, and replaced with a power law following the flux between 10000-10300Å. R06 template with the hot dust contribution removed is shown as the dashed orange SED in Fig. 4.1. Whilst the continuum of these two templates are similar up to around 5000Å, Fig. 4.1 also shows that the T21 template is steeper at longer wavelengths even with the hot dust component of the R06 template removed. The discrepancy in the continuum, along with the difference in emission lines due to smoothing, mark the two largest differences between the R06 and T21 AGN SEDs.

4.2 Comparison of results from the T21 and R06 SED fits

4.2.1 Quality of the fits

With the hot dust emission removed from the R06 AGN template, SED fitting was performed on the XMM-SERVS sample. We used the same SED fitting code and free parameters as used in Chapters 2 and 3, replacing the T21 AGN with the R06 template. The galaxy model used within the SED fitting code originates from the Flexible Stellar Population Synthesis (FSPS) code (Conroy & Gunn, 2010; Conroy et al., 2009), including nebular emission lines and composite stellar populations (CSPs), which are discussed further in Section 2.2.1.

As the focus of this work is to study how the choice of AGN SED can affect both AGN and host galaxy stellar mass inferences, further analysis focused only on the subset of objects deemed to require both galaxy and AGN components by the Akaike information criterion (AIC). In order to test the reliability of the R06 stellar mass inferences, SED fitting was also performed using only the R06 AGN template and hot dust SED, without the galaxy component. The AIC of the R06 fits both with, and without the galaxy component were then compared to ensure that the additional complexities associated with the galaxy SED did actually improve the quality of the fit, and as such allow for meaningful measurements of host galaxy properties. The calculation and full use of the AIC follows the same method discussed in Section 2.3.3. The determination of the AIC showed that, in comparison to the

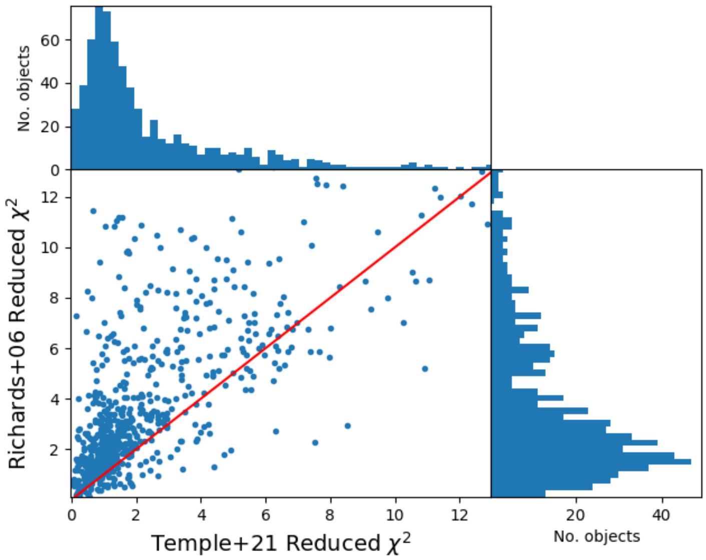

Fig. 4.3 Comparison of the χ^2_{red} using the R06 and T21 AGN templates, along with the 1-D histograms for each fit. Generally, the T21 template is shown to provide better quality fits for 83% of the sample.

T21 fits, a greater percentage — 80 (11%) objects compared to 4 objects for the T21 fit — preferred an AGN only fit with the R06 template.

The additional number of objects preferring the AGN only fit for the R06 template might be due to galaxy contamination, as previously noted by Assef et al. (2010). Galaxy flux remaining within the R06 template could mean that a separate galaxy template is not required to provide a reasonable fit, thus an 'AGN-only' fit is preferred. For reasonable comparison of both host galaxy and AGN properties, we reduce the sample to only include objects where both the R06 and T21 fits require AGN+GAL components based on the AIC designation. The AGN+GAL cut provided a sample of 559 objects for which stellar mass estimates are available for both template fits.

An initial comparison between the quality of the T21 and R06 AGN template fits was made using the reduced χ^2 (χ^2_{red}), and can be seen within Fig. 4.3, along with the 1-D distribution of each fit. Fits using the T21 AGN SED were found to peak at a lower χ^2_{red}, around 1.2. The χ^2_{red} distribution for the R06 AGN fit is generally broader, with a poorer χ^2_{red} peak at around 1.8. Within previous work using the T21 template (see Chapter 3), further analysis of the inferences of AGN and host galaxy properties was only performed on objects with a $\chi^2_{red} < 3$. 385 (69%) of the AGN+GAL model family objects were found to meet this criterion. Applying this same criterion to the R06 template fit, this value was found to reduce to 272 (49%) of the sample, suggesting that the T21 fit was able to produce combinations of AGN and host galaxy SEDs better able to match the observed photometry. Whilst the T21 template may generally provide better quality fits, it is also useful to test if the sub-sample of objects that meet the $\chi^2_{red} < 3$ criteria in the T21 fits are the same objects that meet the criterion in the R06 fits. If this is not the case, it would suggest that there are is a specific sub-sample of objects for which the R06 does appear to outperform the T21 fit. Of the 272 R06 fits that meet the $\chi^2_{red} < 3$ criterion, 255 also meet the criterion within the T21 fit, leaving only 17 objects (3% of the total sample) with a lower χ^2_{red} for the R06 fit. In general, this therefore suggests that the sample of objects fit well by the R06 template are fit equally well or better by the T21 template.

4.2.2 Discrepancies in AGN and host galaxy inferences

Our investigation of the quality of fits suggests that the T21 AGN template is more accurately able to match observed photometry. However, it is also important to note how the AGN template choice can affect inferences on host galaxy and AGN properties. We chose to limit our sample to only objects in which both the R06 and T21 fits provide reasonable inferences, therefore, we only include objects for which $\chi^2_{red} < 3$ for both fits.

Fig. 4.4 shows a comparison of the host galaxy stellar mass inferences produced using the T21 and R06 AGN templates for our $\chi^2_{red} < 3$ sample. The most prominent feature of this figure is the larger number of objects for which the R06 fit predicts a lower stellar mass than the T21 fit. Of our 255 objects, 171 (67%) are consistent within the uncertainties for these two fits, with all of the discrepant objects having lower stellar masses for the R06 fits. The average difference between T21 and R06 stellar mass inferences for the entire sample is 0.51 dex. In addition to our sample of discrepant stellar mass objects, there are also 27 objects for which the AIC has switched to a preference of an AGN-only model family fit for the R06 template. The AGN-only fits for these 27 objects do still meet the $\chi^2_{red} < 3$ criteria. In total, for 39% of reliable fits, there is therefore a significant discrepancy between stellar mass estimates when using the R06 template compared to the T21 fit.

A comparison of the inferences on the R06 and T21 fit AGN properties can be seen in Fig. 4.5, in each case colour-coded by the difference in host galaxy stellar mass. The bottom left-hand corner of this figure shows the difference in dust corrected AGN $logL_{3000}$. Comparison of the two template fits shows a systematic offset, in which the fit produced using the R06 template provides on average a higher L_{3000} by 0.07 dex. In the case of the AGN Luminosity at 3000Å, and hot dust luminosity at $3\mu m$, there is a trend in which the largest

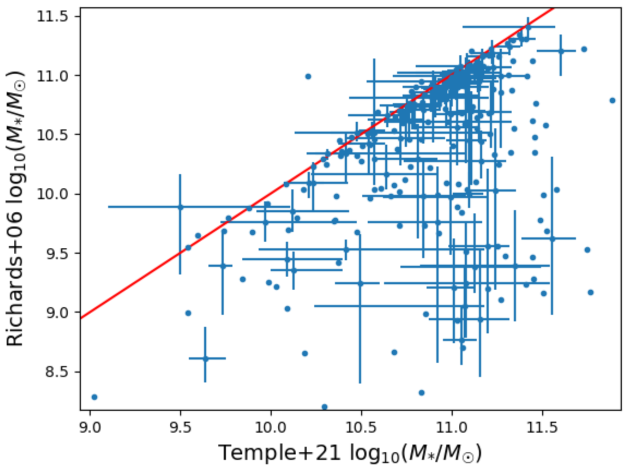

Fig. 4.4 Comparison of the median host galaxy stellar mass produced using the R06 and T21 AGN templates. There is a clear difference in stellar mass inferences, with the R06 template predicting lower stellar masses for 39% of objects. For clarity, uncertainties are limited to a third of the points shown.

stellar mass differences occur at the highest AGN luminosities. Such a trend in stellar mass differences might be expected, as higher AGN luminosities suggest a greater contribution to the SED from the AGN templates compared to the host galaxy SED. Consequently, the galaxy contribution to the flux becomes less dominant, and inferences on the stellar masses become more uncertain. Such a change may therefore be expected to lead to greater discrepancies between the two fits.

In the case of these high AGN luminosity, stellar mass discrepant objects, it may be the case that the R06 fits do not allow us to accurately infer galaxy properties due to the AGN contribution dominating the fit. Such a dominance leads to lower stellar mass estimates for R06 when compared to the T21 fits. Within our $\chi^2_{red} < 3$ sample, the total number of objects found to be inconsistent in $L_{3\mu m}$ between the R06 and T21 fits was small, accounting for 7% of objects based on their respective uncertainties. The most significant difference occurs in L_{3000}, of which 11% of the sub-sample were found to be discrepant. Discrepancies can also be seen in AGN E(B-V) between the T21 and R06 fits in the top panel of Fig. 4.5. In total, we find that for the subset of $\chi^2_{red} < 3$ objects, around 7% are inconsistent within the measured uncertainties for the R06 and T21 E(B-V) inferences. The cause of these measured inconsistencies between our host galaxy and AGN properties will be discussed further in Section 4.3.3.

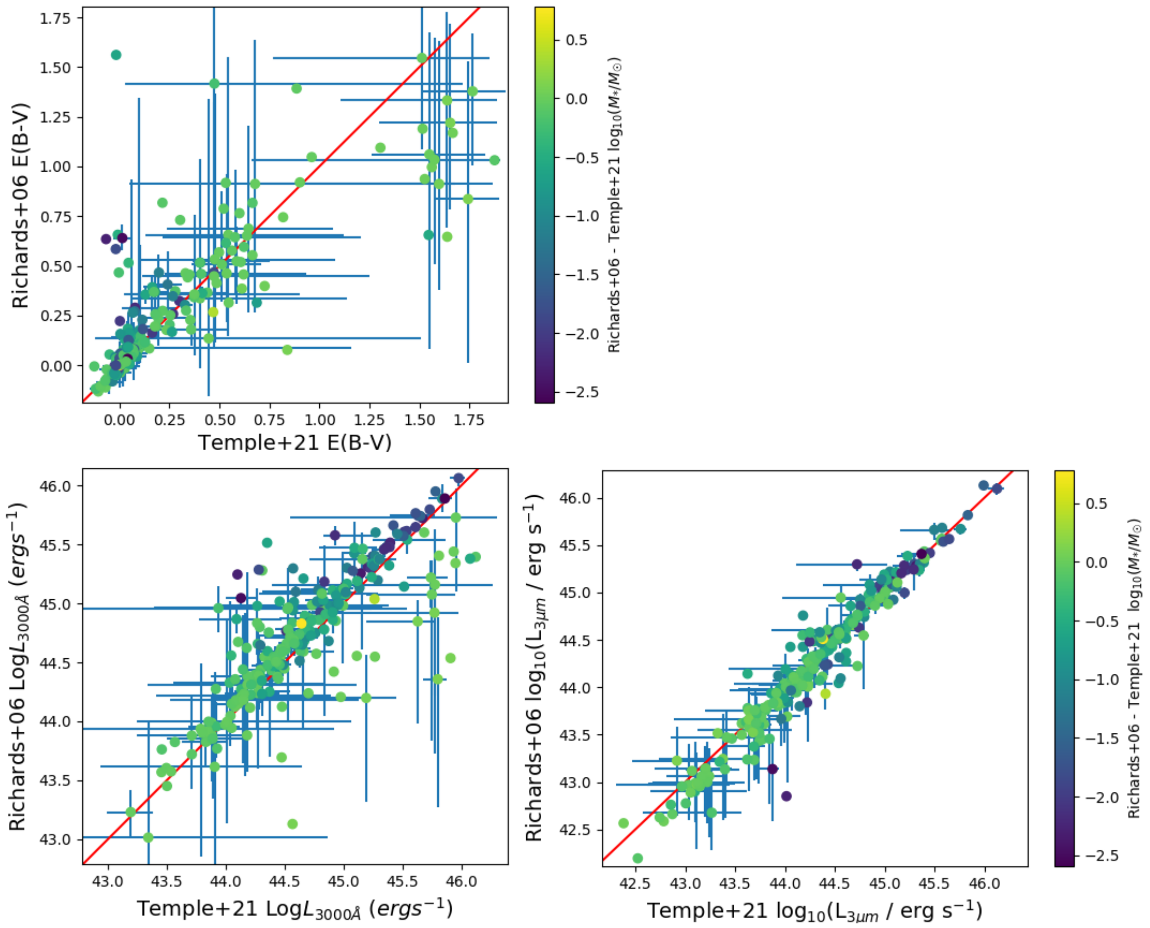

Fig. 4.5 Comparison of the median E(B-V) (top), AGN optical luminosity at 3000Å (bottom left) and hot dust luminosity at 3μm (bottom right) inferences for the R06 vs T21 AGN templates, colour-coded by the difference in the stellar mass inferences of the two fits.

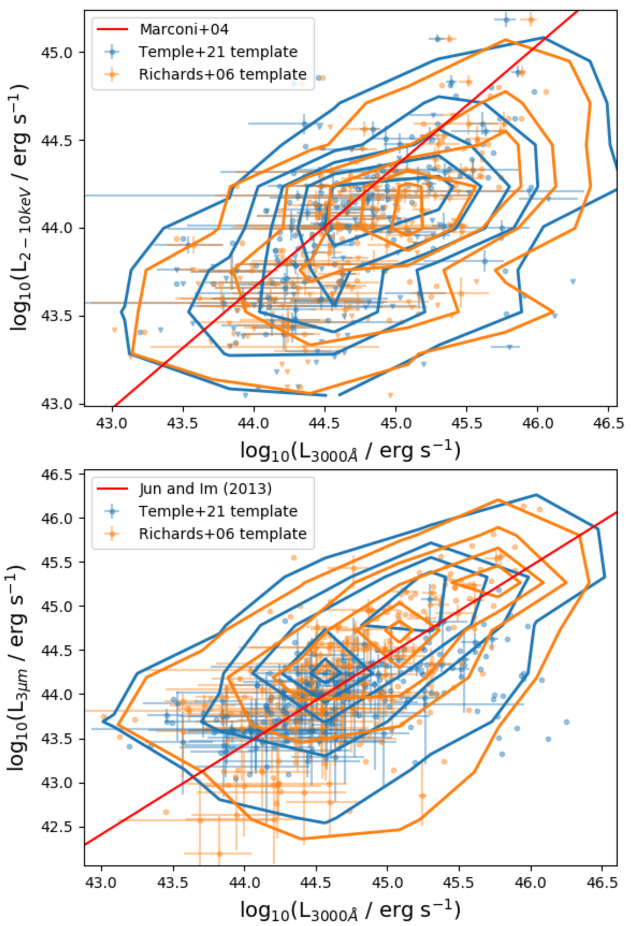

Fig. 4.6 2-10keV X-ray luminosity vs AGN optical luminosity at 3000Å with the relation derived from Marconi et al. (2004) (top) and the 3000Å optical luminosity vs hot dust luminosity at 3μm with the relation from Jun & Im (2013) (bottom). The data points show the median solutions and uncertainties for these objects, with the R06 and T21 AGN template fits shown in orange and blue respectively. The contours show all of the inferences from our MCMC analysis. For clarity, uncertainties are limited to a third of the points shown.

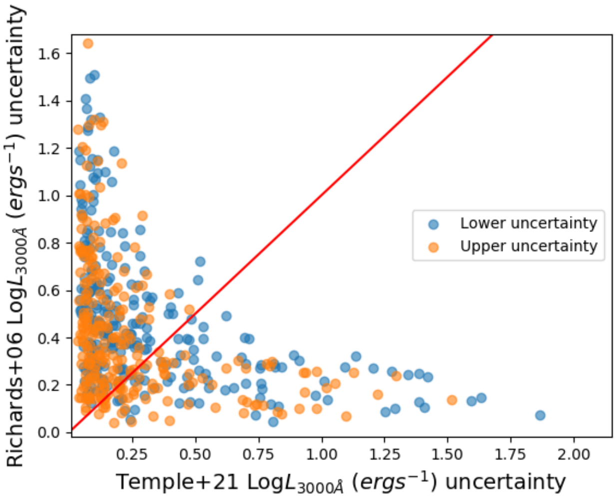

Fig. 4.7 The upper and lower uncertainties associated with the AGN optical luminosity at 3000Å for the R06 and T21 template fits. Here we see that the uncertainties are generally higher for the R06.

Fig. 4.6 shows $L_{2-10keV}$ vs L_{3000} and $L_{3\mu m}$ vs L_{3000} for both the T21 and R06 template fits. The relevant luminosity relations from Marconi et al. (2004) and Jun & Im (2013), as previously discussed in Section 3.3, are also included. In both cases, the shift in AGN inferences has little impact on agreement with previous work, due to the relatively large scatter. Additionally, looking at the total distribution of uncertainties for the AGN optical luminosities as an example (Fig. 4.7), we see that the uncertainties associated with the R06 fit are greater than those of the T21 fit for 66% of the sample. This could be associated with the R06 template providing a poorer fit to the data, as shown in the χ^2_{red} comparison in Fig. 4.3. In such a case, a flatter posterior distribution might be expected, which would therefore lead to greater uncertainties on each inference.

Fig. 4.8 compares the stellar mass inferences to the observed X-ray luminosities for the two AGN templates. Whereas a positive correlation can be seen between these two properties in the T21 template case, the low stellar mass inferences in the R06 template has a significant effect on this correlation. For the sub-sample of T21 and R06 fits with $\chi^2_{red} <$ 3, the correlation between the stellar mass and X-ray luminosity reverses from 0.16 for the T21 fit to an anti-correlation of -0.31 for the R06 fit. The relation is therefore completely different between the AGN and host galaxy for R06, and does not agree with previous work (e.g. Magliocchetti et al. 2020). Magliocchetti et al. (2020) also looked at comparing host galaxy stellar masses to AGN X-ray properties, using X-ray data from the Chandra X-ray observatory (Luo et al., 2016). Host galaxy mass estimates were provided by Galametz et al.

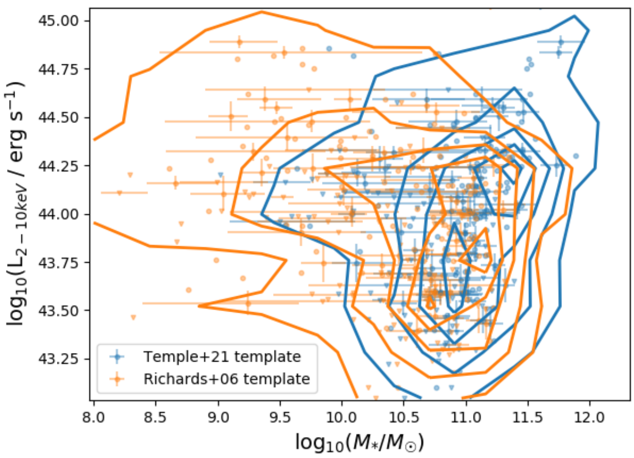

Fig. 4.8 Comparison of the stellar mass inferences vs observed X-ray luminosity. The data points show the median solutions and uncertainties for these objects, with the R06 and T21 AGN template fits shown in orange and blue respectively. The contours show the total inferences from our MCMC analysis. For clarity, uncertainties are limited to a third of the points shown. The lower mass inferences for the R06 template differ from the positive X-ray luminosity-host galaxy mass relation seen in the T21 template.

(2013) and Guo et al. (2013). These estimates were found using SED fitting of 19 bands from a near infrared selected sample of AGN ranging from 3860Å (from CFHT u-band) to 8μm (from Spitzer/IRAC). Magliocchetti et al. (2020) found that large host galaxy stellar masses appeared to favour AGN activity at X-ray wavelengths, with a non-negligible tail of lower stellar mass objects within their sample. Such a result is therefore similar to our results produced using T21 template, which can be seen in Fig. 3.9. The extension of the contours to the top left of Fig. 4.8 for the R06 fits also shows that the majority of the low stellar mass inferences for our SED fits are associated with the higher X-ray luminosity objects. Based on the expected positive relation between the X-ray and optical luminosity of AGN, this may be a result of poor constraint on host galaxy properties due to a dominance of a highly luminous AGN across the optical and infrared region. This will be discussed further in Section 4.3.2.

A comparison to previous work by Bongiorno et al. (2012), in which stellar mass inferences have been produced using the R06 AGN template, is shown in Fig. 4.9. Bongiorno et al. (2012) was chosen for comparison due to its similarity to the XMM-SERVS sample used in this work in terms of X-ray luminosity. The stellar mass inferences of 582 AGN from the work by Bongiorno et al. (2012) are shown in green. Cuts to the sample have been made to only include objects with a spectroscopic redshift > 0.7 as reported by Lusso et al. (2010, 2011). Such a redshift cut ensures that the sample of objects has a similar redshift

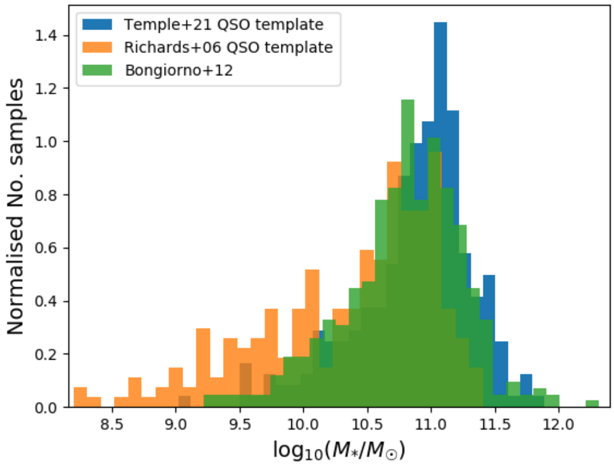

Fig. 4.9 Comparison of the distributions of stellar mass inferences for the R06 (orange) and T21 (blue) AGN templates produced using the SED fitting code created for this thesis. Our stellar mass inferences are compared to previous work by Bongiorno et al. (2012) (green) of objects within the COSMOS field, also created via SED fitting using the R06 template.

distribution as those from XMM-SERVS. The Bongiorno et al. (2012) sample also has a similar X-ray luminosity distribution as XMM-SERVS. Whilst the R06 fits produced using our SED fitting code have a longer tail to lower stellar mass inferences, both this, and the inferences from Bongiorno et al. (2012), peak at similar values, around $\log_{10}(M_*/M_\odot) = 10.9\ M_\odot$. Conversely, the stellar mass inferences from the T21 fit peak at a higher value of $\log_{10}(M_*/M_\odot) = 11.1\ M_\odot$. The differences in stellar mass inferences between the R06 and Bongiorno et al. (2012) fits might be explained by the difference in the photometry used. As has been previously stated by work such as Ciesla et al. (2015), properties such as star formation are typically poorly constrained using only optical and near infrared photometry. The Bongiorno et al. (2012) SED fitting differs from that performed in this work, as it also makes use of CFHT u-band data, extending the fitting further towards the UV than the HSC g-band. Similarly, previous work has highlighted the inclusion of longer wavelength bands as a useful tool in determining star formation rates of host galaxies (Coleman et al., 2022). Thus, the use of the 24μm band within the Bongiorno et al. (2012) SED fit may also be related to the difference in stellar mass inferences.

It is interesting to note however that the T21 template appears to provide higher stellar mass values of Bongiorno et al. (2012). The remaining offset between the Bongiorno et al. (2012) and T21 fits could therefore be attributed to a greater reduction of galaxy

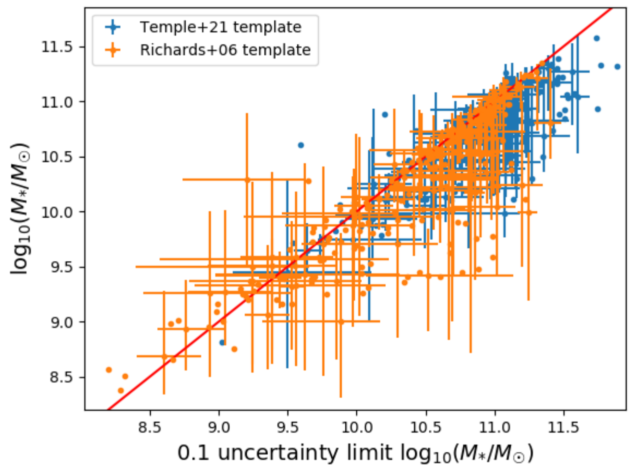

Fig. 4.10 Comparisons of the T21 (blue) and R06 (orange) stellar mass estimates and associated uncertainties. The x-axis gives the fits produced from a run using a higher minimum uncertainty limit of 0.1 placed on each of the 10-bands of photometry. The y-axis gives the same run with a lower uncertainty limit corresponding to 5% of the flux from the 3.0″ images. The inferred stellar mass estimates are consistent despite the increase in uncertainty. For clarity, uncertainties are limited to a third of the points shown.

contamination in the production of the T21 template, leading to an average higher stellar mass for the total sample.

4.3 Reasons for changes in AGN and host galaxy property inferences

Initially, it was noted that the uncertainties placed on our photometric data might be too restrictive, which could have a significant effect on the inferences for both of our fits. The uncertainties associated with the photometry used throughout this work are based on the measurements calculated from the 3.0″ images of each photometric survey. However, it is likely that in some cases our uncertainties over-predict the precision of the photometry. Whilst a lower uncertainty limit was enforced within the fitting code, set a 5% of the total flux for each band, there may be cases in which this value might be too restrictive on the SED fits as a whole. Conversely, there are examples of fits within the total XMM-SERVS sample which have a χ^2_{red} « 1, suggesting an under-prediction of uncertainties.

One concern of the choice of uncertainty limit was that in some cases, the overly restrictive nature of small uncertainties may have the effect of forcing fits into specific regions of parameter space, when in reality a larger span of properties should be explored. This could possibly lead to the difference in R06 and T21 AGN template fits. In order to test the effects of uncertainty on the AGN and host galaxy inferences, the SED fitting code was re-run using both the R06, and T21 AGN SEDs. In both cases, the minimum uncertainty for each band was limited to 0.1 mags. For the majority of the sample (98%) this set the uncertainty to 0.1 for all 10 bands used in the SED fit. To test how this uncertainty increase affected our inferences, we looked at the stellar mass values for each of our runs. In Fig. 4.10 we show a comparison of the stellar mass inferences for the two uncertainty runs using both our T21 and R06 templates. Here we see that the number of low stellar mass inferences for the R06 fit is comparable to the lower uncertainty run. Generally for all host galaxy and AGN properties, there is little change in the inferences between the two uncertainty runs for both the R06 and T21 fits. For the stellar mass, 99% and 98% of objects provide consistent inferences between the two uncertainty fits for the T21 and R06 templates respectively. Such a result suggests that the difference in stellar mass inferences is therefore unlikely to be a result of under-predicted uncertainties. Instead, discrepancies between the R06 and T21 fits are likely due to the differences between the two AGN templates.

Thus, in order to further understand the discrepancies between the T21 and R06 AGN template fits, we compared the SEDs of objects for which there is a significant change in host galaxy stellar mass inferences. The purpose of this analysis was to identify noticeable trends in the quality of fits that may be explained by the ability of the chosen AGN template to match the observed photometry. Based on the findings outlined in Section 4.2.2, we looked both at objects in which the quality of the T21 fit was deemed reasonable, but the R06 fit was not, as well as the objects included within Fig. 4.5 where both fits were reasonable, but gave significantly different stellar mass inferences. The discrepancies in stellar mass inferences were observed to relate to a combination of the two main differences in the R06 and T21 AGN templates as discussed in Section 4.1.2, specifically, the continuum and emission lines.

4.3.1 AGN emission line effects on AGN and host galaxy properties

Looking at the total sample of objects, around 7% were found to be highly discrepant in AGN E(B-V) inferences, for which the difference in E(B-V) was larger than the combined uncertainties for the T21 and R06 template fits. A typical example of one such object can be seen in Fig. 4.11, where the T21 fit has a lower χ^2_{red} than the R06 with values of 2.12 and 3.42 respectively. In this case, the most significant change is the wavelength region in which the galaxy and AGN contributions are most dominant.

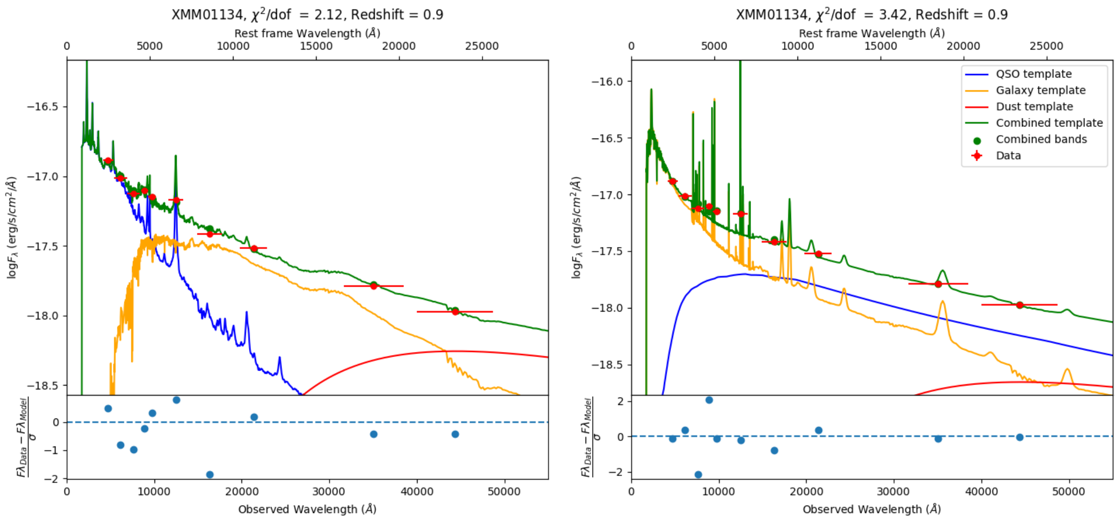

Fig. 4.11 The maximum likelihood SED fits produced using the T21 (left) and R06 (right) AGN templates. Here we can see a switch in the contribution from the AGN and host galaxy. Clear emission lines are not seen within the R06 AGN template due to its low resolution. Within the R06 fit, we therefore see the host galaxy change to attempt to compensate for the emission lines, and recreate the J-band photometry. The younger host galaxy inferences are seen to decrease the galaxy stellar mass inferences in the R06 fit.

In the T21 fit case (left-hand side of Fig. 4.11), a bright, unobscured AGN can be seen providing the bluer flux, whilst an older, quiescent galaxy, and hot dust contributes to the redder optical and IR bands. In the R06 fit, however, instead we see a young, star forming galaxy with strong emission lines providing the bluer flux, whilst a highly reddened AGN represents the redder flux. The flux provided by the reddened AGN has a shallower slope into the infrared than the host galaxy in the T21 fit. The additional AGN infrared contribution therefore leads to a decrease in the flux provided by the hot dust.

A switch in the region of flux dominance between the AGN and host galaxy was shown previously in Chapter 3 of this thesis. In that case, the switch was attributed to the bimodality attributed to the complex degeneracies associated with the high number of free parameters included within the SED fits. However, degeneracy does not seem to be the explanation in the case shown in Fig. 4.11. Looking at the marginalised one- and two dimensional posterior distributions for both SED fits (Fig. 4.12), no bimodality is seen in either the R06 or T21 parameter distributions. The lack of bimodality suggests that both fits provide different inferences for our host galaxy and AGN properties, and are not displaying two similar likelihood solutions. The different solutions are therefore likely to be due to the differences in the AGN templates.

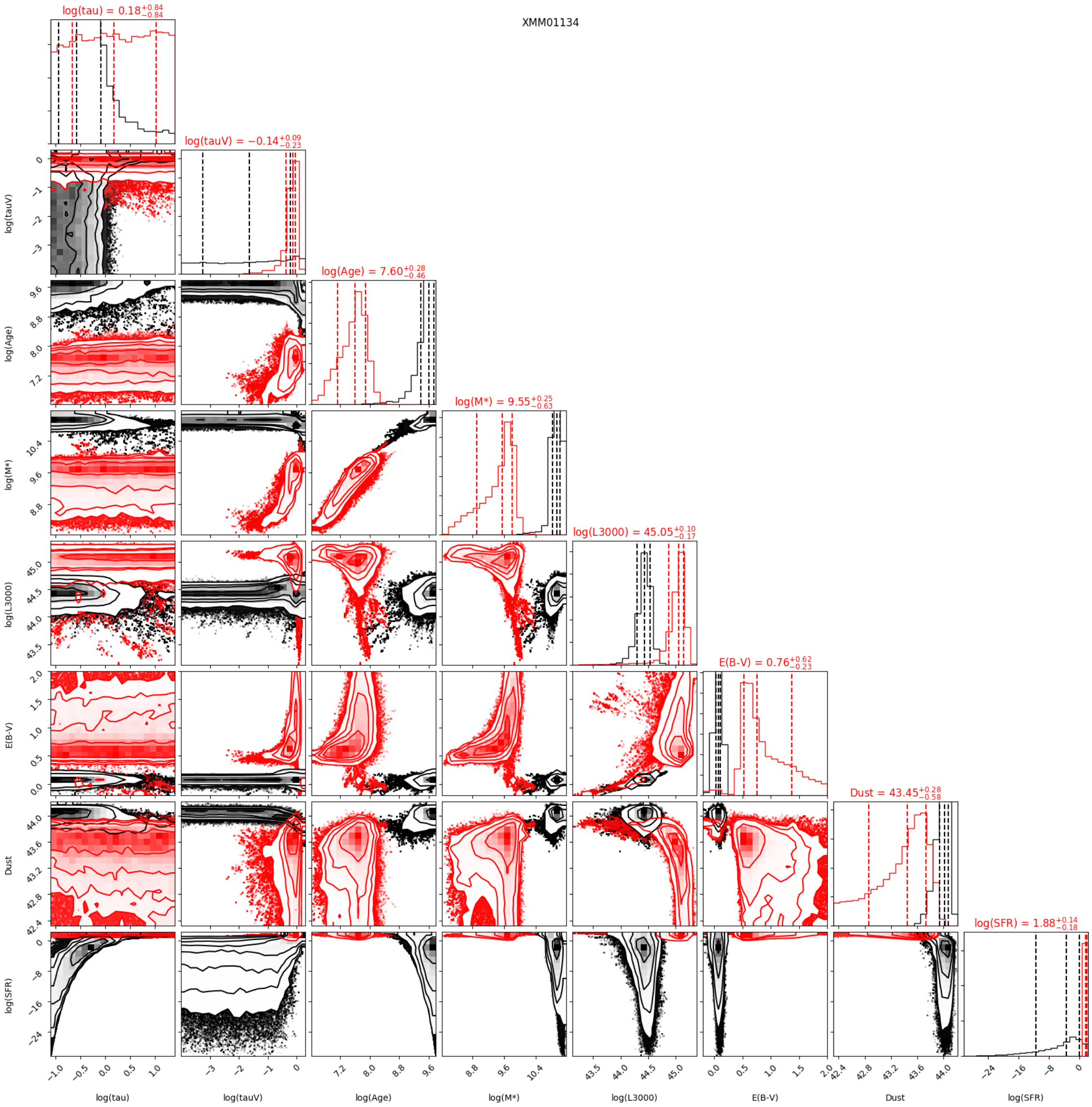

Fig. 4.12 Marginalised one- and two-dimensional posterior distributions for the SEDs fits shown in Fig. 4.11. The distributions for the T21 and R06 AGN templates are shown in black and red respectively.

In the case shown in Fig. 4.11, the difference in SED fits may be tied to the emission lines within the two AGN SED templates. One major difference can be prominently seen in the attempt to fit the Hα emission line. For the object shown in Fig 4.11, which is at redshift of 0.9, Hα appears within the VISTA J-band. The presence of Hα appears to significantly raise the J-band flux away from the continuum, when compared to the neighbouring y- and H-bands. Due to the low resolution of the R06 AGN template, strong features that have a large effect on broadband magnitudes are smoothed out, and thus have less significant influence on the AGN model magnitudes. In order to replicate the increase in flux seen in the observed photometry, the AGN emission lines are compensated for by a young, star forming galaxy. This young galaxy provides its own emission lines in order to model the peaks in the photometry. The galaxy emission lines are not a perfect match to the AGN lines, thus resulting in a worse fit to the photometry. Additionally, a more obscured AGN is then required to provide the redder flux not given by the young galaxy SED. We therefore infer a younger galaxy for the R06 template fits, along with the highly discrepant E(B-V) inferences for the AGN when compared to the T21 fit.

4.3.2 AGN continuum effects on AGN and host galaxy properties

Contrary to the case seen above, the difference between T21 and R06 fits cannot always be attributed to a specific emission line. For the majority of objects, we do not see a complete switch in galaxy and AGN contributions in the SED fits. In the example fits shown in Fig. 4.13, we see that for both our T21 and R06 fits, the AGN contribution is dominant over the galaxy. In the infrared, the hot dust provides a significant contribution to the *Spitzer* bands. However, in the case shown in Fig. 4.13, whilst the AIC of the T21 fit suggests the inclusion of the galaxy template, the R06 fit gives a lower AIC for its AGN-only fit. To understand the change in the R06 fit, both the AGN+GAL and AGN-only model families are included within this figure. Comparing the R06 and T21 AGN+GAL fits, a significant decrease can be seen in the stellar mass inferences for the R06 fit, with corresponding median values of $\log_{10}(M_*/M_\odot) = 10.36 \pm^{0.24}_{0.29} M_\odot$ and $\log_{10}(M_*/M_\odot) = 10.87 \pm^{0.18}_{0.21} M_\odot$ respectively. In cases such as Fig 4.13, the AGN is dominant across the wavelength region of study, and thus a match between the shape of the chosen AGN continuum and the observed photometry is key to provide a good fit. The T21 fit is better able to reproduce this continuum, leading to much lower χ^2_{red} of 1.2, compared to 5.42 for the R06 fit. The steeper drop-off towards redder wavelengths of the T21 AGN SED requires a larger contribution from old stellar populations in the host galaxy at longer wavelengths. The change in host galaxy contribution is large

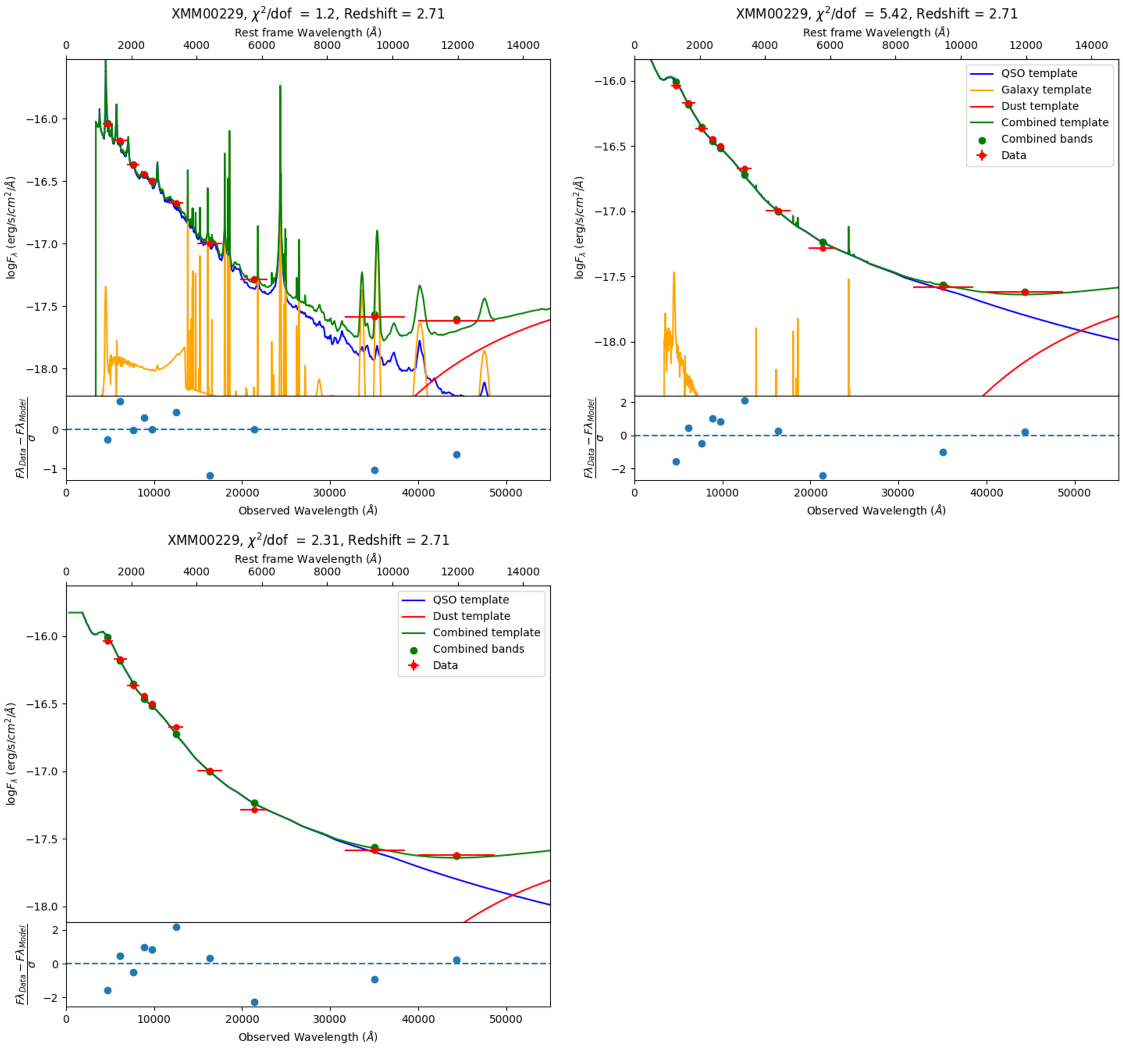

Fig. 4.13 The maximum likelihood SED fits produced using the T21 (top left), R06 (top right) AGN+GAL and R06 (bottom-left) AGN-only templates. The disagreement between the fits could be attributed to the difference in continuum for the two AGN templates, specifically affecting the IR bands. The more gradual dropoff at longer wavelengths in the R06 template results in the separate host galaxy contribution not being required for the R06 best fit.

enough that the R06 is better fit without the galaxy template and its additional complexities. There is thus a decrease in the number of AGN+GAL objects in the R06 sample, as discussed in Section 4.2.1. However, more generally, the change in continuum has the effect of lowering the stellar mass contribution. The larger contribution from old stellar populations in the host galaxy at longer wavelengths also has the additional effect of lowering the contribution of the hot dust in the case of the T21 fit, which accounts for the small sample of dust luminosity discrepant objects discussed in Section 4.2.2.

4.3.3 Combined emission line and continuum effects on AGN and Galaxy properties

The example SEDs shown in Figs. 4.11 and 4.13 were chosen as clear cases in which the continuum or emission line differences were responsible for a change in inferred properties. For the majority of the stellar mass discrepant sample (80% of the mass discrepant objects), the reason for T21 and R06 discrepancies is a combination of both effects. If the presence of Hα within specific bands was the sole reason for a significant difference on the host galaxy mass in the R06 fit, a trend might be expected between stellar mass discrepancy and redshift. We would see greater discrepancies in stellar mass estimates when Hα moves into one of the observed bands of photometry. However, the additional effects associated with the continuum mask this, and discrepancies between the R06 and T21 fits were not found to show any specific trend with redshift.

Fig. 4.14 shows an example of an object in which discrepancies between the R06 and T21 fits can be attributed to both AGN template emission line and continuum differences. The SEDs within Fig. 4.14 show an example of the most typical effect of changing between the R06 and T21 SED, and is representative of 80% of the stellar mass discrepant objects. In this case, there isn't a clear switch in the galaxy and AGN contributions as is the case for fits such as Fig. 4.11. We do however see an increase in flux in the J-band containing Hα that is compensated for by galaxy emission lines. Thus for the R06 fit, a younger, star forming host galaxy SED is required, similar to the fits in Section 4.3.1.

Similarities are also seen with the example object shown in Section 4.3.2. In both cases, the inferences on L_{3000} are consistent between the R06 and T21 fits. However, unlike the example SED given in Fig. 4.13, Fig. 4.14 has significant contributions from both the AGN and host galaxy. Despite the difference in the continuum shape at longer wavelengths, (as seen in Fig. 4.1) the shape of the R06 and T21 templates are similar for the bluer HSC bands. Within this region, the AGN is dominant over the galaxy contribution, thus both fits require

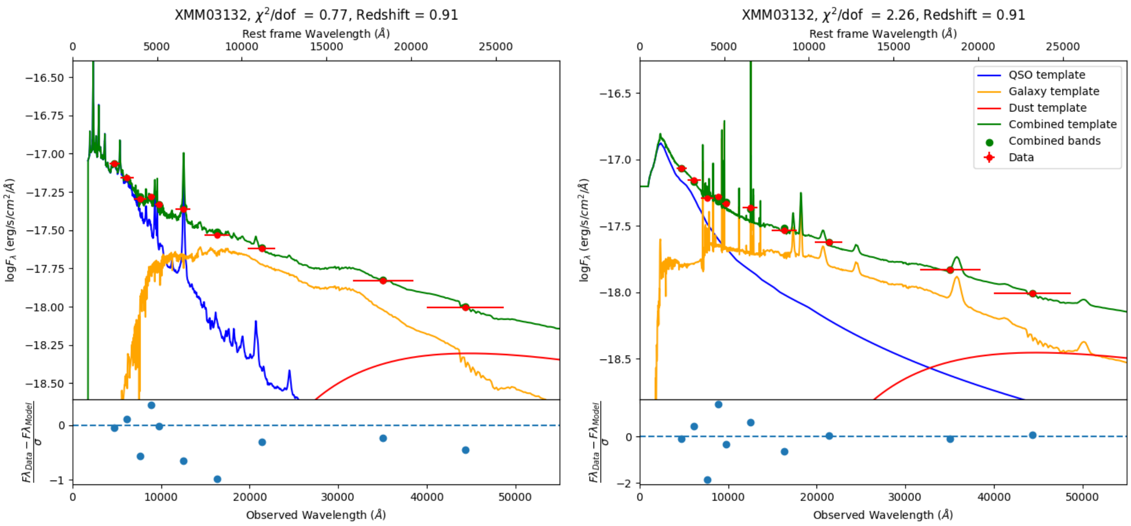

Fig. 4.14 The maximum likelihood SED fits produced using the T21 (left) and R06 (right) AGN templates. For the majority of the sample of discrepant stellar mass fits, the inferred AGN properties for the T21 and R06 fits are consistent, with the galaxy component changing instead to compensate for the lack of significant emission lines within the R06 template.

AGN contributions with similar AGN luminosities to accurately model the HSC g-,r- and i-bands.

Whilst the galaxy is able to provide additional flux via emission lines, the lines present in the galaxy SED have a lower equivalent width when compared to the AGN Hα line. In a smaller number of cases (11% of fits), L_{3000} is increased to provide this additional flux. The AGN luminosity change, along with the fits discussed in Section 4.3.1, may explain the 11% of objects in which L_{3000} is discrepant between the two fits, with the R06 fit typically inferring higher luminosities. Similarly, for a small number of cases, the less steep continuum slope of the R06 template can have a significant effect on the hot dust luminosity inference. The continuum difference may therefore explain the 7% of $L_{3\mu m}$ discrepant objects discussed in Section 4.2.2. The total combination of these factors leads to the differences in stellar mass and AGN property inferences seen in Figs. 4.4 and 4.5.

4.4 Comparison of median spectra

To understand how the difference in AGN template can affect our sample of objects as a whole, we also look at the median AGN and galaxy contributions from all of the R06 and T21 fits, shown in Fig 4.15. In each case, median spectra were produced by first converting the maximum likelihood AGN (excluding the hot dust component) and host galaxy contributions

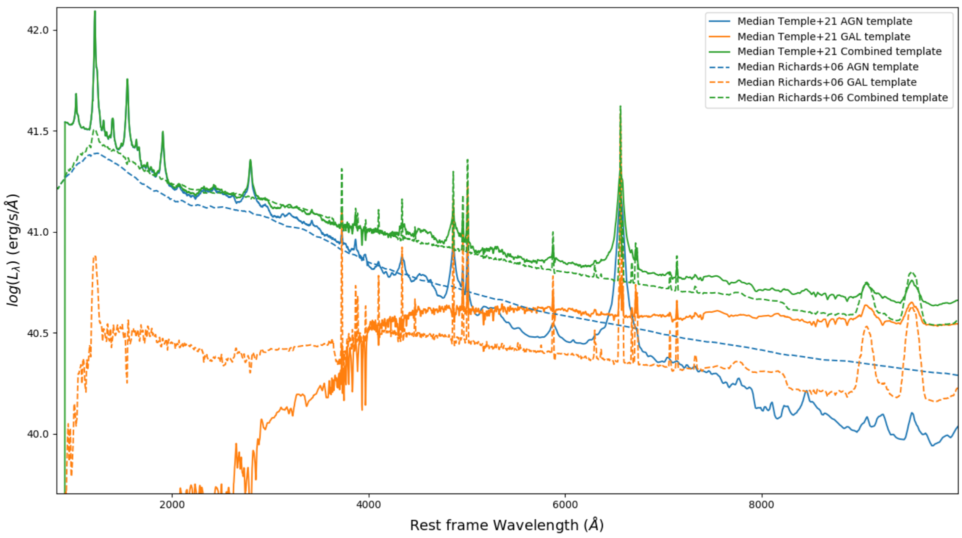

Fig. 4.15 The median rest frame spectra of the galaxy (orange) and AGN (blue) and combined (green) contributions to the R06 (dashed lines) and T21 (solid lines) SED fits. In the case of both AGN templates, the host dust component is not included in the plot. The median R06 galaxy SED is younger than the T21 template, with higher rates of star formation resulting in prominent emission lines compensating for the lack of emission lines in the R06 AGN SED.

from each object into the rest frame. For each point on the wavelength axis, the median luminosity was then calculated from all objects, thus producing a single median galaxy spectrum.

As was seen within the individual SED fits shown in Fig. 4.11, the median galaxy of the R06 fits is younger, with a greater number of emission lines. This relates to the findings of Section 4.3.1, where the smoothing of emission lines within the R06 SED makes use of a young galaxy to provide these emission lines instead. Simultaneously, the differences in both the galaxy and AGN median spectra tie into the difference in AGN continuum between the two templates. The median AGN templates in Fig. 4.15 show the steeper T21 template provides less luminosity at longer wavelengths. The steeper slope combines well with the average galaxy from the T21 fits, which, being comparatively older and quiescent compared to its R06 fit counterpart, provides greater flux within this region, thus balancing out the luminosity contributions.

One feature of the T21 template was to provide an accurate determination of the contamination from the host galaxy to the AGN continuum. Such contamination has previously been found to account for $> 5\%$ of the flux of the total SED at $1\mu m$ (Temple et al., 2021).

The median AGN and host galaxy contributions may also show the effects of the additional removal of host galaxy contamination from the T21 template. It can be seen that the contribution from the T21 template is lower at $1\mu m$ compared to the R06 fit. If the R06 AGN template does still contain some contamination from the host galaxy, this might explain why the median galaxy contribution is lower at $1\mu m$.

From our analysis of the R06 template, and comparison to the T21 SED, our work shows the opposite conclusion to that of Merloni et al. (2010) on the nature of host galaxy contamination within the R06 template. Merloni et al. (2010) produced a new mean SED from point-like type-1 COSMOS AGNs for comparison with the R06 template they used in SED fitting. They found that their new mean SED had a less pronounced dip within the $1\mu m$ region compared to the R06 template. From this result, Merloni et al. (2010) therefore suggested that the use of the R06 template in SED fitting 'maximises' their host galaxy stellar mass estimates, as the additional dip around $1\mu m$ in the R06 template will be compensated for by the additional host galaxy template. This maximisation of host galaxy stellar mass is suggested as one possible reason that Merloni et al. (2010) find an positive offset from the relation shown by Häring & Rix (2004) between SMBH and host galaxy masses. Conversely, our findings suggest that the R06 template doesn't compensate for the host galaxy enough, and in fact a better fit is possible using the T21 template, which itself has a lower flux contribution at $1\mu m$ than the R06 template. We would therefore argue that COSMOS type-1 AGN template produced by Merloni et al. (2010) contains further host galaxy contamination than the R06 template, and that both of these templates still lead to an underestimation of host stellar mass due to this contamination.

4.5 Conclusions

- From a sample of 559 objects, SED fitting was performed using two AGN SEDs produced by Temple et al. (2021) and Richards et al. (2006). The results of these fits showed a significant change in stellar mass values for around 39% of the sample, with the R06 template typically providing lower stellar mass estimates than the T21 template. For a subset of objects, there was a significant decrease in the contribution from the galaxy template within the SED fit performed using the R06 AGN template. Based on the Akaike information criterion, the decrease in galaxy contribution was significant enough that a AGN-only fit was preferred for 11% of the R06 sample, compared to only 0.7% for the T21 fits. For these objects, reliable host galaxy stellar mass inferences could no longer be made.

- Comparison of our stellar mass inferences was made against previous work by Bongiorno et al. (2012), which had also made use of the R06 template within their SED fits. The smaller photometric range used within our SED fits led to a greater spread to lower stellar mass estimates for the fits produced using the R06, but did not affect the T21 fits in the same way. This may be related to the additional removal of host galaxy contributions within the T21 template, as well its higher resolution when compared to the R06 template.

- Considering the sample as a whole showed that the T21 template on average provided better quality fits. Only 3% of the sample of objects with reasonable fits ($\chi^2_{red} < 3$) were found to provide R06 fits with smaller χ^2_{red} than the respective T21 fit.

- Looking at the sample of objects fit well by the T21 template, but poorly fit by the R06 template, we found two main features were responsible for the observed stellar mass discrepancies:

  1) To compensate for the lack of clear AGN emission lines within the R06 template due to its low resolution, a younger, star forming galaxy replaced the quiescent galaxy within the T21 fit. For around 7% of the sample, this had the additional effect of switching the contributions of the AGN and host galaxy within the R06 fit, leading to discrepant E(B-V) inferences.

  2) The difference in the continuum of the R06 and T21 templates led to a change in the quality of the fit. For a subset of these objects, the less steep continuum slope of the R06 template led to an AGN dominant fit, such that the stellar mass inferences were no longer meaningful as a comparison to the T21 fit. The continuum slope of the R06 AGN was found to be worse at reproducing the observed photometry when compared to the steeper T21 slope.

- For the majority of objects with a significant difference in stellar mass inferences (80%), the difference between the R06 and T21 template fits was found to be a combination of the two factors above. Despite the fact that only the AGN SEDs were switched within the SED fitting code, it was the properties of the host galaxy SED, specifically the stellar mass, that were typically found to have the most significant changes. These changes were shown via a comparison of the median AGN and host galaxy spectra, produced from the maximum likelihood fits from both the R06 and T21 template runs.

Overall, the outcome of the work outlined within this chapter highlights the need for the careful consideration of the templates used when inferring AGN and host galaxy properties.

We have focused on two AGN templates that both rely on the use of empirical data, which, when attempting to produce a ‘pure’ AGN template, can be influenced by host galaxy contamination. Whilst the R06 template did attempt to correct for the inclusion of this contamination (as outlined in Section 4.1.1) the more recent work from T21 has shown not only that an updated template can be significantly different in the extent of its host galaxy removal, but also that these differences can have a large effect on our host galaxy and AGN inferences. Additionally, other aspects such as the resolution of the templates used can also greatly affect inferences, and the smoothing of emission lines between bands can lead to poor fits to the observed data. Whilst a AGN template such as R06 may be well established, it is important to understand the method used in its creation in order to fully appreciate the biases that its usage in SED fitting may impart. Such an issue can be applied further to the use of larger AGN fitting code suites that may make use of a range of possible AGN and host galaxy templates. Whilst a specific combination of AGN and host galaxy templates may provide a best match to a set of data, it is important to ensure that the way in which the chosen templates were produced is understood, as properties such as low resolution or host galaxy contamination can have a significant effect on the inferred results, in particular host galaxy stellar mass estimates.

In general, two main courses of action can be recommended based on the results of this chapter. The first is that when performing SED fitting, it is best to use the most up-to-date high-resolution AGN templates that are able to accurately recreate observations. Understanding how these templates were produced can help to avoid, or at least explain, possible biases in property inferences that might not have been considered in previous SED fitting work. Finally, it is important to keep in account the differences in the templates used by other work used for comparison, as these can be the reason for major discrepancies in results.

Chapter 5

Summary and Future work

5.1 Thesis summary

The main aims of this thesis have been to explore AGN and host galaxy properties through the use of spectral energy distribution modelling. To this end, work has focused on the development of an SED fitting code, making use of a newly produced AGN SED (Temple et al. 2021; hereby referred to as T21). This is the first time the T21 AGN SED has been used within an SED fitting code. Detailed analysis of the T21 template and SED fitting code was therefore initially required to ensure that any host galaxy and AGN properties could be deemed reliable:

- Initial tests were made to observe the prevalence of bimodality or lack of convergence within our SED fits. Our findings revealed cases of bimodality due to the complexity of the parameter space (e.g. Fig. 3.4). Within Chapter 2, a parallel-tempering MCMC optimisation was therefore introduced. This method used multiple temperatures to aid walkers in escaping local maxima and place them close to the global maxima. The bulk of the posterior can then be fully explored using an MCMC run. Further diagnostics were also performed in order to find a useful statistic that would allow all bimodal and unconverged fits to be removed from large samples of objects. Our findings suggested that the complex nature of high parameter space MCMC fitting made this difficult. Thus, the decision was made to include the distributions of all of the samples within analysis throughout this thesis.

- In Chapter 3, SED fitting was performed on 711 luminous X-ray AGN at $0.7 < z < 4.5$ using 10-bands of optical and infra-red photometric data for objects within the XMM-SERVS catalogue (Chen et al., 2018). Our fits provided 510 reliable (reduced $\chi^2 < 3$)

inferences on AGN and host galaxy properties. The AGN optical (3000Å) luminosity inferred from SED fitting was found to correlate with the measured X-ray (2-10 keV) luminosity, in good agreement with previous work (Marconi et al., 2004). Using X-ray hardness as a proxy for AGN obscuration, we also studied the differences in the host galaxy properties of obscured and unobscured AGN. Both populations have consistent stellar masses ($\log_{10}(M_*/M_\odot)$ = 10.88 $\pm 0.09 M_\odot$ and $\log_{10}(M_*/M_\odot)$ = 10.8 $\pm 0.1 M_\odot$ for unobscured and obscured AGN respectively). However, comparison of AGN optical luminosity to hot dust luminosity at 3μm showed an offset when compared to previous work (Jun & Im, 2013). It is theorised that this offset may be related to the greater removal of galaxy contamination within the T21 template, when compared to the AGN template used by Jun & Im (2013). We also found a positive correlation of AGN X-ray 2-10keV luminosity with the host galaxy stellar mass, in agreement with previous work by Suh et al. (2019).

- One of the more unique features of the T21 AGN template is the ability to vary the strength and asymmetry (which are correlated) of strong emission lines via the addition of a single free parameter. The limits of the emission line variations for this parameter are set by the extremes observed within SDSS spectra (Temple et al., 2021). Emission line type was therefore included as a free parameter within our SED fitting code in order to test if such information could be inferred from photometry alone. For 18.8% of our sample, non-standard emission lines were found to be preferred over the average emission lines seen in AGN at z = 2, according to the AIC of the fit. In order to confirm these results, comparisons were made between our best fit SEDs, and observed SDSS spectra where available. Comparison with SDSS showed good agreement with both emission line shape and strength for 91% of the sub-sample of non-standard emission line fits. For further analysis, objects within this sub-sample were then used to test α_{ox} vs emission line type, which is linked to the Baldwin relation (Baldwin, 1977). It was found that the presence of weaker, more blueshifted emission lines inferred from the SED fits were associated with more negative values of α_{ox}. While correlation between the hardness of the ionising SED and emission line properties has been known for some time, in this case, it was possible to derive this correlation purely from broadband photometry.

- As this thesis has focused on the ability to infer AGN and host galaxy properties from SED fitting, it is important to consider of the extent to which the SED components, such as the chosen AGN template, can change our inferences. Chapter 4 features a case study in which an alternative AGN SED is considered within our fitting code. The chosen AGN SED, produced by Richards et al. (2006) (hereby referred to as R06) has been used previously on samples similar to the XMM-SERVS catalogue (Bongiorno et al., 2012; Marconi et al., 2004). The R06 AGN template differs from the T21 template in both continuum slope and resolution. The low resolution of the R06 template means that, unlike the T21 template, individual emission lines cannot be seen within the spectrum. The differences between these templates were found to have a significant effect on the stellar mass inferences of AGN host galaxies, with the R06 template preferring significantly lower host galaxy mass fits than the T21 template for 39% of our sample. Example SEDs showed that, in the case of the R06 fits, a young, star forming galaxy was chosen to compensate for the lack of emission lines within the AGN template. Within the T21 SED fitting run, the same photometric bands were fit using the AGN emission lines, which provided lower χ^2 values, and thus were deemed a better representation of the observed photometry (as shown in Fig. 4.11).

- Based on our SED fits, we also present an average galaxy SED for both our T21, and R06 fits. Comparisons of the two average galaxy templates showed that the lack of AGN emission lines within the R06 template leads to younger average host galaxy. We believe the observed discrepancies might also be linked to additional host galaxy contamination within the R06 template, a feature that has been also been suggested by Assef et al. (2010).

5.2 Future work

5.2.1 Extension into the UV/MIR

Limitations on the reliability of CFHT *u*-band photometry (see Section 3.3.1) within our matched XMM-SERVS/VISTA VIDEO catalogues did not make extension into the near UV possible. Additionally, one of the aims of this thesis was to see to what extent it was possible to constrain host galaxy and AGN properties within a limited wavelength range. The reasoning for this was that the nature of current and upcoming OIR surveys are surpassing the sky-coverage and photometric precision of surveys within other wavelength ranges. Thus, such surveys would rely heavily on the OIR range for host galaxy and AGN property inferences for objects too faint within other wavelength surveys. However, as previous work has reported (Ciesla et al., 2015), and we have discussed in Section 2.3.3, SED fitting using only optical and NIR photometry poorly constrains host galaxy properties such as age, and consequently star formation rates. Additional photometry, such as the CFHT u-band, may aid in separating bright, unobscured AGN from star forming host galaxies.

Star formation rate estimates could provide useful information into the relation between AGN and their host galaxies, especially on the role of AGN feedback in quenching star formation. Future work could therefore build on the work shown within this thesis by extending the wavelength range further, adding u-band or MIR photometry into the fitting code, providing it can be reliably matched to the current XMM-SERVS based catalogue.

5.2.2 Photometric redshift estimates

Along with the inclusion of extra photometry, additional information could also be gained from our SED fitting code through the determination of photometric redshift estimates. The addition of a photometric redshift free parameter would remove our reliance on spectroscopic redshift estimates that have been required for the objects included within this thesis. For the XMM-SERVS catalogue alone, reliable photometric redshift estimates may allow fits for an additional 4500 objects (Chen et al., 2018). However, as has been discussed throughout the development of our SED fitting code (see Sections 2.3.1, 2.3.3), the addition of a free parameter such as redshift within SED fitting can greatly complicate the quality of results, adding additional complexities that could introduce further bimodality. To help to avoid this, the total complexity of the parameter space could be further simplified via an aggregation of host galaxy properties.

The AGN template developed by Temple et al. (2021) was created with the aim of providing a spectrum that accurately reproduced the observed colours of AGN from SDSS, whilst retaining the minimum number of required parameters (Temple et al., 2021). Such aims were achieved through the combining of numerous spectral features into single parameter value. For example, emission line features including strength, width, and asymmetry were all combined into a single parameter that represented the strength of all emission lines within the spectrum. The extremes for the emission line values were based on observations of objects within SDSS spectra.

In order to reduce the number of free parameters within our fit, it could therefore be possible to apply a similar method to our galaxy parameter values. In this case, multiple galaxy properties, specifically the e-folding time, τ, v-band optical depth, τ_v, and host galaxy age may be combined into a single free parameter. In general, our galaxy-property parameter space is much larger than the space real galaxies lie in. For example, our parameter space includes very old, very highly star-forming galaxies which are unlikely to exist in significant numbers in the real Universe. By forming a 3-dimensional plot of τ, τ_v, and age of observed galaxies, we can form a 2-dimensional surface that incorporates all three properties, whilst simplifying the exploration of the parameter space. Such a method would remove two free parameters from our fits, and thus reduce the complexity of the parameter space, even with the addition of a photometric redshift parameter.

For the AGN SED, the optical luminosity, E(B-V), and hot dust luminosity would also remain as free parameters within the fits. These, along with the redshift, host galaxy stellar mass and τ, τ_v and age aggregate, would provide fits with six total free parameters. As the sub-selection of objects used within this thesis all have spectroscopic redshift values available, our sample would therefore serve as a useful initial test for confirming the reliability of photometric redshift estimates.

5.2.3 Target selection for future spectroscopic surveys

Using our SED fitting code, and the further development of including redshift as a free parameter, work within this thesis could also be applied to larger object samples as a method of AGN target selection. The wide Chandra Deep Field-South (W-CDF-S) (Ni et al., 2019) would be well suited to this purpose, given it contains approximately 2 Million objects with data from the HSC (Aihara et al., 2018), VISTA (Jarvis et al., 2013) and SERVS (Mauduit et al., 2012) surveys. If successful, our SED fitting code would be useful in providing a method of separating out possible AGN from inactive galaxies, without the need for X-ray observations. One of the main aims within the study of AGN is to build a complete picture of their co-evolution with host galaxies across the Universe. Increasing the AGN sample size available is key to achieving this goal.

SED fitting using data from these photometric surveys could also be used to identify possible AGN targets for the most up-to-date spectroscopic instruments. The multi-object spectrographs 4MOST (4-meter Multi-Object Spectroscopic Telescope) (Merloni et al., 2019) and MOONS (The Multi-Object Optical and Near-infrared Spectrograph) Maiolino et al. (2020a) will be invaluable for further study of host galaxy and SMBHs relations. Both spectrographs will provide a significantly larger sample of dusty SMBHs for analysis, allowing for the study of the typical properties of these objects. The MOONS infrared spectrograph, for example, will provide high quality spectra across the 0.6-1.8 μm wavelength range. One of the key scientific objectives for MOONS includes the collection of data for $\sim$1 million galaxies with redshifts z > 1, and its $\sim$500 square arc-minute field includes some of the same regions previously viewed within the VISTA photometric catalogue. The total

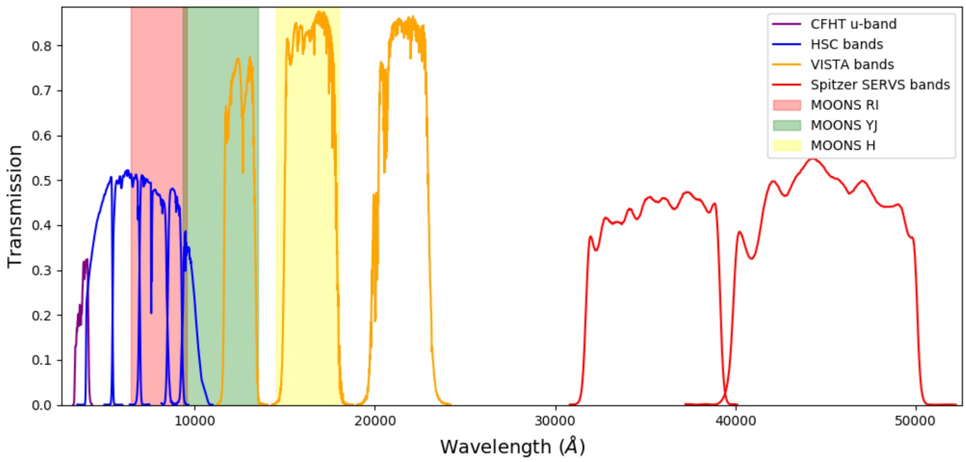

Fig. 5.1 The three regions in which spectra will be collected the using MOONS (Maiolino et al., 2020a), in comparison with the transmissions curves, including atmospheric absorption and detector quantum efficiencies, of the bands used throughout this thesis (HSC, VISTA VIDEO and Spitzer SERVS), along with the CFHT *u*-band

spectral range for MOONS is shown in figure 5.1, along with the transmission curves the XMM-SERVS based catalogue photometry, and additional CFHT u-band. As there is direct crossover of this survey wavelength range, and the photometry used within this thesis, further comparison can also be made between the best fit spectra shown within Chapter 3, and the actual spectra of these objects from MOONS.

5.2.4 Machine learning

In order provide a useful method of AGN target selection for future spectroscopic surveys, it is a requirement to computationally achieve a significant number of fits using our SED code. For large scale upcoming surveys such as LSST (Ivezić et al., 2019), data will be available for around 20 billion galaxies. The SED fitting code developed throughout this thesis could be applied to these massive samples to provide both AGN identification, along with inactive galaxy inferences based on our AIC designation. However, one current limiting factor is the time taken per object to produce a useful fit. For our current SED fitting code, it can take two minutes per object for walkers to converge. Such a timescale would not therefore be suited for the large number of objects required for a useful target selection sample for LSST, 4MOST or MOONS. To this end, the use of machine learning could greatly increase the sample size and speed of SED fits. A machine learning algorithm could be trained on half of the SED fits produced within this thesis, being provided with the photometry and inferences produced for each object, in order to produce an complex function based on these inputs. The function could then be tested on the second half of our SED fits, and compared to our actual inferences for reliability.

If successful, the use of machine learning could greatly increase the number of objects it would be possible to fit (Williams et al., 2018), increasing our sample size of AGN and host galaxy inferences. Such an increase in sample size would allow for the development of a more complete picture of AGN in the Universe, and a significant expansion of the sample size used throughout this thesis.

References

Abbott T. M. C., et al., 2021, ApJS, 255, 20

Aihara H., et al., 2018, PASJ, 70, S8

Aihara H., et al., 2019, PASJ, 71, 114

Angione R. J., 1973, AJ, 78, 353

Antonucci R., 1993, ARA&A, 31, 473

Assef R. J., et al., 2010, ApJ, 713, 970

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

Baldwin J. A., Phillips M. M., Terlevich R., 1981, PASP, 93, 5

Barišić I., et al., 2017, ApJ, 847, 72

Barnes J. E., Hernquist L., 1992, ARA&A, 30, 705

Bennett J. S., Sijacki D., 2022, MNRAS, 514, 313

Best P. N., Kauffmann G., Heckman T. M., Brinchmann J., Charlot S., Ivezić Ž., White S. D. M., 2005, MNRAS, 362, 25

Bicknell G. V., Sutherland R. S., van Breugel W. J. M., Dopita M. A., Dey A., Miley G. K., 2000, ApJ, 540, 678

Bongiorno A., et al., 2012, MNRAS, 427, 3103

Boquien M., Burgarella D., Roehlly Y., Buat V., Ciesla L., Corre D., Inoue A. K., Salas H., 2019, A&A, 622, A103

Bower R. G., Benson A. J., Malbon R., Helly J. C., Frenk C. S., Baugh C. M., Cole S., Lacey C. G., 2006, MNRAS, 370, 645

Bowler R. A. A., Jarvis M. J., Dunlop J. S., McLure R. J., McLeod D. J., Adams N. J., Milvang-Jensen B., McCracken H. J., 2020, MNRAS, 493, 2059

Brown M. J. I., Duncan K. J., Landt H., Kirk M., Ricci C., Kamraj N., Salvato M., Ananna T., 2019, MNRAS, 489, 3351

Bruzual G., Charlot S., 2003, MNRAS, 344, 1000

Calistro Rivera G., Lusso E., Hennawi J. F., Hogg D. W., 2016, ApJ, 833, 98

Calzetti D., Kinney A. L., Storchi-Bergmann T., 1994, ApJ, 429, 582

Capelo P. R., Volonteri M., Dotti M., Bellovary J. M., Mayer L., Governato F., 2015, MNRAS, 447, 2123

Chabrier G., 2003, PASP, 115, 763

Chen C.-T. J., et al., 2018, MNRAS, 478, 2132

Cicone et al., 2014, A&A, 562, A21

Ciesla L., et al., 2015, A&A, 576, A10

Coleman B., et al., 2022, MNRAS, 515, 82

Conroy C., Gunn J. E., 2010, ApJ, 712, 833

Conroy C., Gunn J. E., White M., 2009, ApJ, 699, 486

Conroy C., Dutton A. A., Graves G. J., Mendel J. T., van Dokkum P. G., 2013, ApJ, 776, L26

Croton D. J., et al., 2006, MNRAS, 365, 11

Curtis-Lake E., Chevallard J., Charlot S., Sandles L., 2021, MNRAS, 503, 4855

Da Cunha E., Charlot S., Dunne L., Smith D., Rowlands K., 2011, Proceedings of the International Astronomical Union, 7, 292–296

Davies J. J., Pontzen A., Crain R. A., 2022, MNRAS, 515, 1430

De Barros S., Nayyeri H., Reddy N., Mobasher B., 2013, in Cambresy L., Martins F., Nuss E., Palacios A., eds, SF2A-2013: Proceedings of the Annual meeting of the French Society of Astronomy and Astrophysics. pp 531–535 (arXiv:1310.3405)

Dey A., van Breugel W., Vacca W. D., Antonucci R., 1997, ApJ, 490, 698

Dubois Y., Peirani S., Pichon C., Devriendt J., Gavazzi R., Welker C., Volonteri M., 2016, MNRAS, 463, 3948

Dullemond C. P., van Bemmel I. M., 2005, A&A, 436, 47

Ellison S. L., Mendel J. T., Scudder J. M., Patton D. R., Palmer M. J. D., 2013, MNRAS, 430, 3128

Elvis M., et al., 1994, ApJS, 95, 1

Fabian A. C., 2012, ARA&A, 50, 455

Falomo R., Pian E., Treves A., 2014, A&ARv, 22, 73

Fioc M., Rocca-Volmerange B., 1997, A&A, 326, 950

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

Fritz J., Franceschini A., Hatziminaoglou E., 2006a, MNRAS, 366, 767

Fritz J., Franceschini A., Hatziminaoglou E., 2006b, MNRAS, 366, 767

Galametz A., et al., 2013, ApJS, 206, 10

Gao F., et al., 2020, A&A, 637, A94

Gardner J. P., et al., 2006, Space Sci. Rev., 123, 485

Graham M. J., Djorgovski S. G., Drake A. J., Mahabal A. A., Chang M., Stern D., Donalek C., Glikman E., 2014, MNRAS, 439, 703

Granato G. L., De Zotti G., Silva L., Bressan A., Danese L., 2004, ApJ, 600, 580

Gültekin K., et al., 2009, ApJ, 698, 198

Guo Y., et al., 2013, ApJS, 207, 24

Habouzit M., et al., 2022, MNRAS, 509, 3015

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

Hasinger G., 2008, A&A, 490, 905

Hickox R. C., Alexander D. M., 2018, ARA&A, 56, 625

Hönig S. F., 2019, ApJ, 884, 171

Hubeny I., Blaes O., Krolik J. H., Agol E., 2001, ApJ, 559, 680

Ishibashi W., Fabian A. C., 2012, MNRAS, 427, 2998

Ivezić Ž., et al., 2019, ApJ, 873, 111

Jahnke K., Macciò A. V., 2011, ApJ, 734, 92

Jarvis M. J., et al., 2013, MNRAS, 428, 1281

Johnson B. D., Leja J., Conroy C., Speagle J. S., 2021, ApJS, 254, 22

Jun H. D., Im M., 2013, ApJ, 779, 104

Juneau S., et al., 2014, ApJ, 788, 88

Kormendy J., Ho L. C., 2013, ARA&A, 51, 511

Krawczyk C. M., Richards G. T., Mehta S. S., Vogeley M. S., Gallagher S. C., Leighly K. M., Ross N. P., Schneider D. P., 2013, ApJS, 206, 4

Kriek M., Conroy C., 2013, ApJ, 775, L16

Krolik J. H., Begelman M. C., 1988, ApJ, 329, 702

Laloux B., et al., 2022, Monthly Notices of the Royal Astronomical Society, 518, 2546

Lanzuisi G., et al., 2017, A&A, 602, A123

Laor A., Netzer H., 1989, MNRAS, 238, 897

López-Navas E., Arévalo P., Bernal S., Graham M. J., Hernández-García L., Lira P., Sánchez-Sáez P., 2023, MNRAS, 518, 1531

Lower S., Narayanan D., Leja J., Johnson B. D., Conroy C., Davé R., 2020, ApJ, 904, 33

Luo B., et al., 2016, The Astrophysical Journal Supplement Series, 228, 2

Lusso E., et al., 2010, A&A, 512, A34

Lusso E., et al., 2011, A&A, 534, A110

Lyke B. W., et al., 2020, ApJS, 250, 8

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

Madau P., Pozzetti L., Dickinson M., 1998, ApJ, 498, 106

Magliocchetti M., et al., 2020, MNRAS, 493, 3838

Magnelli B., et al., 2012, A&A, 539, A155

Magorrian J., et al., 1998, AJ, 115, 2285

Maiolino R., et al., 2020a, The Messenger, 180, 24

Maiolino R., et al., 2020b, The Messenger, 180, 24

Marconi A., Hunt L. K., 2003, ApJ, 589, L21

Marconi A., Risaliti G., Gilli R., Hunt L. K., Maiolino R., Salvati M., 2004, MNRAS, 351, 169

Marshall A., Auger-Williams M. W., Banerji M., Maiolino R., Bowler R., 2022, MNRAS, 515, 5617

Matthews J. H., Knigge C., Higginbottom N., Long K. S., Sim S. A., Mangham S. W., Parkinson E. J., Hewitt H. A., 2020, MNRAS, 492, 5540

Mauduit J. C., et al., 2012, PASP, 124, 1135

McNamara B. R., Nulsen P. E. J., 2007, ARA&A, 45, 117

Merloni A., et al., 2010, ApJ, 708, 137

Merloni A., et al., 2019, The Messenger, 175, 42

Merritt D., Ferrarese L., 2001, MNRAS, 320, L30

Moreno J., et al., 2019, MNRAS, 485, 1320

Morganti R., 2017, Frontiers in Astronomy and Space Sciences, 4, 42

Mountrichas G., Buat V., Georgantopoulos I., Yang G., Masoura V. A., Boquien M., Burgarella D., 2021, arXiv e-prints, p. arXiv:2106.11579

Müller-Sánchez F., Prieto M. A., Hicks E. K. S., Vives-Arias H., Davies R. I., Malkan M., Tacconi L. J., Genzel R., 2011, ApJ, 739, 69

Mushotzky R. F., Done C., Pounds K. A., 1993, ARA&A, 31, 717

Ni Q., Timlin J., Brandt W. N., Yang G., 2019, Research Notes of the American Astronomical Society, 3, 5

Padovani P., et al., 2017, A&ARv, 25, 2

Peng C. Y., Impey C. D., Rix H.-W., Kochanek C. S., Keeton C. R., Falco E. E., Lehár J., McLeod B. A., 2006, ApJ, 649, 616

Percival W., et al., 2019, in Canadian Long Range Plan for Astronomy and Astrophysics White Papers. p. 20, doi:10.5281/zenodo.3758532

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

Pier E. A., Krolik J. H., 1992, ApJ, 401, 99

Pilbratt G. L., et al., 2010, A&A, 518, L1

Polletta M., et al., 2007, ApJ, 663, 81

Pouliasis E., et al., 2019, MNRAS, 487, 4285

Predehl P., et al., 2021, A&A, 647, A1

Reber G., 1944, ApJ, 100, 279

Richards G. T., et al., 2001, AJ, 121, 2308

Richards G. T., et al., 2006, ApJS, 166, 470

Richards G. T., et al., 2011, AJ, 141, 167

Richards G. T., McCaffrey T. V., Kimball A., Rankine A. L., Matthews J. H., Hewett P. C., Rivera A. B., 2021, The Astronomical Journal, 162, 270

Rodrigo C., Solano E., Bayo Arán A., 2017, in Astronomical Society of India Conference Series. pp 63–68

Rosario D. J., 2019, FortesFit: Flexible spectral energy distribution modelling with a Bayesian backbone, Astrophysics Source Code Library, record ascl:1904.011 (ascl:1904.011)

Salvatier J., Wiecki T. V., Fonnesbeck C., 2016, PeerJ Computer Science, 2, e55

Salvato M., Ilbert O., Hoyle B., 2019, Nature Astronomy, 3, 212

Sanders D. B., Soifer B. T., Elias J. H., Madore B. F., Matthews K., Neugebauer G., Scoville N. Z., 1988, ApJ, 325, 74

Scannapieco C., et al., 2012, MNRAS, 423, 1726

Schaerer D., de Barros S., 2009, A&A, 502, 423

Schmidt M., 1963, Nature, 197, 1040

Schneider D. P., et al., 2003, AJ, 126, 2579

Shin J., Woo J.-H., Chung A., Baek J., Cho K., Kang D., Bae H.-J., 2019, ApJ, 881, 147

Sijacki D., Vogelsberger M., Genel S., Springel V., Torrey P., Snyder G. F., Nelson D., Hernquist L., 2015, MNRAS, 452, 575

Smit R., et al., 2014, ApJ, 784, 58

Stalevski M., Ricci C., Ueda Y., Lira P., Fritz J., Baes M., 2016, MNRAS, 458, 2288

Stern D., et al., 2005, ApJ, 631, 163

Suh H., et al., 2019, ApJ, 872, 168

Temple M. J., Hewett P. C., Banerji M., 2021, MNRAS, 508, 737

Thorne J. E., et al., 2022, MNRAS, 509, 4940

Timlin J. D., Brandt W. N., Ni Q., Luo B., Pu X., Schneider D. P., Vivek M., Yi W., 2020, MNRAS, 492, 719

Timlin J., Zhu S., Brandt W. N., Laor A., 2021, Research Notes of the American Astronomical Society, 5, 101

Trevese D., Boutsia K., Vagnetti F., Cappellaro E., Puccetti S., 2008, A&A, 488, 73

Tristram K. R. W., et al., 2007, A&A, 474, 837

Truemper J., 1982, Advances in Space Research, 2, 241

Ulrich M.-H., Maraschi L., Urry C. M., 1997, ARA&A, 35, 445

Urry C. M., Padovani P., 1995, PASP, 107, 803

Vanden Berk D. E., et al., 2006, AJ, 131, 84

Veillet C., 2007, in American Astronomical Society Meeting Abstracts #210. p. 170

Veilleux S., Osterbrock D. E., 1987, ApJS, 63, 295

Veilleux S., Cecil G., Bland-Hawthorn J., 2005, ARA&A, 43, 769

Walcher J., Groves B., Budavári T., Dale D., 2011, Ap&SS, 331, 1

Wang F., et al., 2021, ApJ, 907, L1

Werner M. W., et al., 2004, ApJS, 154, 1

Williams W. L., et al., 2018, MNRAS, 475, 3429

Wright E. L., et al., 2010, AJ, 140, 1868

Wylezalek D., Zakamska N. L., Greene J. E., Riffel R. A., Drory N., Andrews B. H., Merloni A., Thomas D., 2018, MNRAS, 474, 1499

Yuan S., Strauss M. A., Zakamska N. L., 2016, MNRAS, 462, 1603

Zou F., Yang G., Brandt W. N., Xue Y., 2019, ApJ, 878, 11

Appendix A

Determining mean stellar masses for different populations from MCMC samples

Our SED modelling has been undertaken independently for each object in this study, but it is also useful to determine properties of *populations* of objects. For concreteness, we will consider the case of inferring the mean stellar masses of the populations of AGNs with harder and softer X-ray emission, as discussed in Section 3.3.3. In principle, we could simultaneously investigate all objects with a *hierarchical* model, assuming that each object's stellar mass is drawn from a Gaussian distribution for example, and sample for the population *hyperparameters* (e.g., the mean and variance of the Gaussian distribution from which each object in a population might be drawn).

However, we have already obtained sampled realisations of population stellar masses from our independent MCMC analyses, by considering the set of samples from a given step of each object's Markov chain. We can therefore calculate the *sample mean* $M_{\rm pop,i}$ (and *standard error on the sample mean*, $\sigma_{\rm pop,i}$) for each ith MCMC step of all obscured or unobscured AGN, for example. This gives us a distribution of samples for the population mean stellar masses (i.e., the set of all $M_{\rm pop,i}$), with uncertainties, that also implicitly propagates the uncertainties on the individual stellar masses from the SED fitting. In practice, we find that the standard error on the sample mean, $\sigma_{\rm pop}$, is nearly the same for all MCMC steps, and we therefore quote the inference on the population stellar mass as the mean of the samples $M_{\rm pop,i}$ with an uncertainty given by the square root of the sum of the variance of these samples and $\sigma^2_{\rm pop}$.