ΕΘΝΙΚΟ & ΚΑΠΟΔΙΣΤΡΙΑΚΟ
ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ
Τμήμα Φυσικής

BSc Thesis

Probing scaling Relations of AGN accretion parameters through the ensemble QSO X-ray variability in XMM-Newton light Curves

Konstantinos Rafail Akalestos

S.N. 1110201900006

Supervisor:

Dr. Antonis Georgakakis, Research Director

Co-supervisor:

Dr. Maria Petropoulou, Assistant Professor

Athens, 2025

For the vast and wonderful world

that gave us life

and keeps us guessing.

Acknowledgments

I would like to express my deepest gratitude to Dr. Antonis Georgakakis for kindly agreeing to supervise this thesis, and for his patience, encouragement, and invaluable support throughout the process. His guidance not only helped me take my first steps into the world of scientific research, but also instilled in me a deep appreciation for the discipline and care required to carry it out.

I would also like to thank the professors and fellow students I met during my undergraduate studies for being some of the brightest minds I've had the pleasure of learning alongside, and for their generosity in sharing their knowledge and perspectives.

Lastly, I want express my deepest and most sincere thanks to my mother, for doing her best through the most impossible moments; to my father, for teaching me the value of living a life meant to be shared; and to my friend Petros Meramveliotakis — one of the most inspiring scientists I have ever known — whose steady support from the very beginning of my studies meant more than words can say, and who had wished to be the first to read this thesis.

Abstract

Active Galactic Nuclei exhibit characteristic stochastic flux fluctuations in their X-ray emission, providing important insights into the accretion physics around supermassive black holes. Despite this, few systematic studies have targeted variability in AGNs with high Eddington ratios and massive black holes, particularly on shorter timescales that closely reflect conditions near the central engine. In this thesis, we study intrinsic X-ray variability through normalized excess variance (NXSV) using a sample of optically-selected SDSS QSOs also found in observations from XMM-Newton's Serendipitous Source Catalog. Our sample spans Eddington ratios $\lambda_{\rm Edd} \approx 10^{-2}$ to the Eddington limit and black hole masses $M_{\rm BH} \approx 10^{8}$–$10^{10} M_{\odot}$. We employ a custom processing pipeline for optimal light curve extraction and utilize a new Bayesian hierarchical methodology fit for the Poisson nature of the light curves to estimate mean NXSV across populations of physically similar QSOs. Our analysis reveals a positive correlation between NXSV and $\lambda_{\rm Edd}$, challenging earlier studies reporting an anti-correlation. This increased short-term X-ray variability at $\lambda_{\rm Edd}$ may result from changes in the corona's geometry or from enhanced turbulence associated with transitions to slim-disk accretion. Conversely, variability shows no clear scaling with $M_{\rm BH}$, although comparisons with theoretical models suggest a subtle correlation could be obscured by the uncertainties inherent to our values.

Contents

List of Figures

List of Tables

Common abbreviations

AGN	**A**ctive-**G**alactic **N**ucleus
BLR	**B**road-**L**ine **R**egion
EEF	**E**ncircled **E**nergy **F**raction
EM	**E**lectro**m**agnetic
FWHM	**F**ull **W**idth at **H**alf **M**aximum
HEALPIX	**H**ierarchical **E**qual **A**rea iso**L**atitude **PIX**elization
ISCO	**I**nnermost **S**table **C**ircular **O**rbit
ISM	**I**nterstellar **S**pace **M**edium
MOC	**M**ulti**O**rder **C**overage
NLR	**N**arrow-**L**ine **R**egion
NXSV	**N**ormalized e**X**ce**S**s **V**ariance
ODF	**O**bservation **D**ata **F**iles
PSD	**P**ower **S**pectral **D**ensity
PSF	**P**oint **S**pread **F**unction
QSO	**Q**uasi-**S**tellar **O**bject
RMS	**R**oot-**M**ean **S**quare
SNR	**S**ignal-to-**N**oise **R**atio
SMBH	**S**upermassive **B**lack **H**ole
UV	**U**ltra**v**iolet
XMM	**X**-ray **M**ulti-**M**irror Mission

Chapter 1

Introduction

1.1 Active Galactic Nuclei

1.1.1 What are AGN

All galaxies are believed to host supermassive black holes (SMBHs) at their centres [e.g. Kormendy and Ho, 2013, review article]. These black holes grow primarily through the accretion of surrounding material, a process that can release tremendous amounts of energy. This energy is observed across a broad range of wavelengths in the Electromagnetic (EM) spectrum and through its interaction with the interstellar medium (ISM) alters its physical conditions thereby directly affecting the host galaxy [e.g. Silk and Rees, 1998, Fabian, 1999, King, 2003]. Simulations of galaxy evolution have shown a consistent need for feedback from acreeting black holes to reconstruct fundamental properties of galaxies, such as their star formation history and stellar mass[e.g. Somerville and Davé, 2015, review article]. Studying the accretion process is therefore essential not only for understanding the formation and growth of these extreme objects but also for having a complete picture of galaxy evolution[Alloin et al., 2006].

1.1.2 Current picture of accretion flow

Given the significance of SMBHs in shaping their host galaxies, there have been intensive efforts to understand the structure and dynamics of accretion flows. However, the scales involved are extremely small. The gravitational radius (R_g), the characteristic radius dominated by the gravitational effects of the black hole, is given by:

$$R_g = \frac{GM}{c^2}, \tag{1.1}$$

where G is the gravitational constant, M is the mass of the black hole, and c is the speed of light. For a black hole with a mass of $10^8 M_\odot$:

$$R_g = \frac{(6.674 \times 10^{-11}\,\mathrm{m^3\,kg^{-1}\,s^{-2}})(10^8 \times 1.989 \times 10^{30}\,\mathrm{kg})}{(3.0 \times 10^8\,\mathrm{m/s})^2} \approx 1.5\,\mathrm{AU} \approx 22 \times 10^7 \mathrm{km}.$$

At a distance of about 100 Mpc, the corresponding angular size θ can be calculated as:

$$\theta = \frac{R_g}{d} = \frac{1.5\,\mathrm{AU}}{100 \times 3.086 \times 10^6\,\mathrm{pc}} \approx 7\,\mathrm{microarcseconds}.$$

Due to the limitations on the angular resolution of our current observatories, direct imaging has been possible for only two black holes so far[1]. For the vast majority of AGN, indirect methods are required to study the accretion flow and its components. One such approach is to study the stochastic flux variations of the radiation emitted by the accretion process. Flux variability is a characterizing feature of any accretion flow and provides a powerful tool to examine their structure beyond the limits of angular resolution since variations in the inner accretion flow are transmitted at the speed of light, thereby allowing us to translate time differ-

[1]Direct imaging of black holes was made possible through the Event Horizon Telescope (EHT), a very long baseline interferometry (VLBI) network operating at millimeter wavelengths. By linking radio observatories across the globe, the EHT effectively creates an Earth-sized telescope with an angular resolution of about 20 microarcsec, enough to image event horizon-scale structures of nearby SMBHs. The EHT has been used to sucessfully image the supermassive black hole at the center of M87 as well as Sagittarius A*, the black hole at the center of the Milky Way.

ences between variability patterns into spatial scales and distances. Recent X-ray polarization studies provide additional insights, but variability remains the dominant method for studying **the inner structure of** AGN. The current picture of the accretion flow around SMBHs**, which is heavily influnced by time domain studies,** includes several distinct components (see figure 1.1):

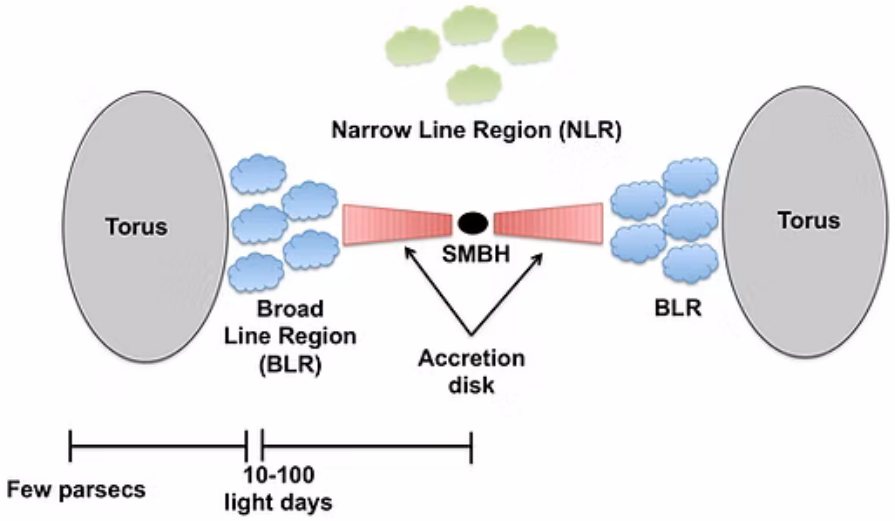

Figure 1.1: Typical structure (not on scale) of an AGN in accordance to the unified model with its distinct physical characteristics highlighted [Ricci, 2011].

1.1.2.1 Central supermassive black hole

The presence of supermassive black holes at the centres of galaxies is inferred from several points of observation. In the case of the Milky Way, Sagittarius A*, the black hole at its center has been studied through precise measurements of stellar movement in its vicinity. These stellar motions are well described by Keplerian orbits, allowing a straightforward application of of Kepler's third law to estimate the black hole mass as:

$$M_{\mathrm{BH}} = \frac{rv^2}{G} \tag{1.2}$$

where r is the star's orbital radius, v is its velocity, and G is the gravitational constant.

For distant AGN, where direct observations of stellar orbits are not feasible, we

rely on broad emission lines observed in their spectra. These lines originate from gas clouds in the *Broad Line Region* (BLR), a region surrounding the centre of the AGN where high-velocity gas emits broad spectral lines. Assuming that the motion of these BLR clouds is dominated by the gravitational influence of an extremely dense object at their centre (i.e. the SMBH) and that this motion follows Keplerian dynamics[2], we can apply the virial theorem to estimate the black hole mass as:

$$M_{\text{BH}} = f\frac{R_{\text{BLR}}v^2}{G}, \tag{1.3}$$

where

- R_{BLR} is the radius of the BLR determined through reverberation mapping,
- f is a geometric factor that accounts for the unknown geometry and kinematics of the BLR,
- v is the velocity of the gas (inferred from the full width at half maximum of the broad emission lines).

Reverberation mapping techniques are often employed to measure R_{BLR} by observing the time delay, τ, between variations in the AGN continuous emission and the corresponding response in the broad emission lines [Alloin et al., 2006]. The radius is then obtained using the relation:

$$R_{\text{BLR}} = c\tau, \tag{1.4}$$

where c is the speed of light. Combining this result with Equation 1.3, we can estimate M_{BH} with reasonable accuracy as:

$$M_{\text{BH}} = \frac{fc\tau\Delta v^2}{G} \tag{1.5}$$

[2]The virial theorem assumes that the BLR clouds are in Keplerian motion around the black hole and that the system is in dynamical equilibrium. While this is a reasonable approximation for some AGN, deviations from these assumptions (e.g., non-Keplerian motion, outflows, or inflows in the BLR) may introduce uncertainties in the estimated mass.

1.1.2.2 Accretion Disk

A piece of evidence for the presence of accretion disks in AGN is a characteristic feature in their spectral energy distribution (SED) shown in (Figure 1.2), known as the *Blue Bump*. This feature, peaking in the ultraviolet (UV) region, is often attributed to thermal emission from a geometrically thin, optically thick accretion disk.

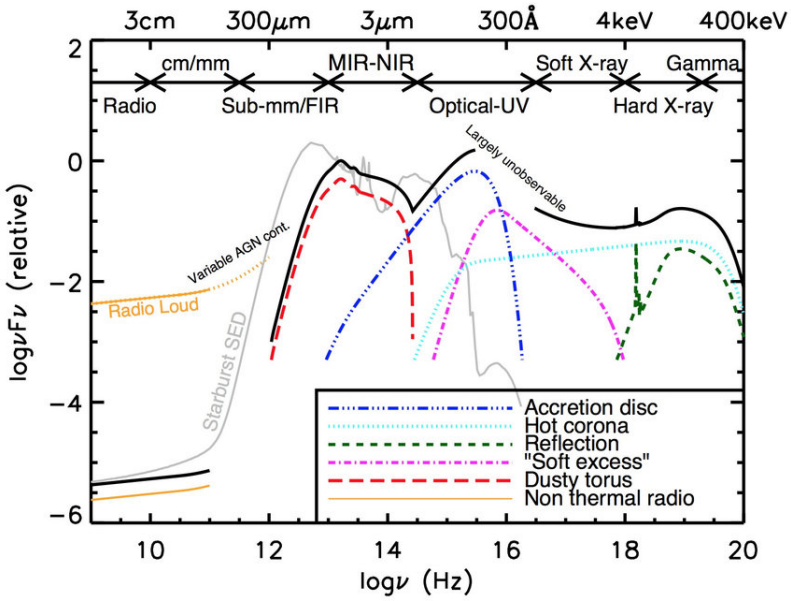

Figure 1.2: A schematic representation of an AGN's Spectral Energy Distribution SED, loosely based on the observed SEDs of radio-quiet quasars. The black solid curve represents the total SED and the various coloured curves (with an arbitrary offset) represent the individual components. Also shown is an example radio–UV SED of a starburst galaxy grey curve; the SED is of M82 taken from Silva et al., 1998). Adapted from [Baldini, 2015].

The presence of this thermal component suggests that most of the gravitational potential energy of the infalling matter is converted into radiation near the black hole, consistent with the predictions of accretion disk models.

In AGN, accretion disks form due to the conservation of angular momentum. As matter falls inward towards the SMBH, viscous forces dissipate energy and enable the gas to lose angular momentum and move further inward. For optically thick, geometrically thin accretion disks, often referred to as Shakura-Sunyaev or

α-disks [Shakura and Sunyaev, 1973], the temperature at a radius r is given by:

$$T(r) = \left(\frac{3GM_{\text{BH}}\dot{M}}{8\pi\sigma r^3} \left[1 - \left(\frac{R_{\text{ISCO}}}{r} \right)^{1/2} \right] \right)^{1/4}, \tag{1.6}$$

where $\dot{M}$ is the mass accretion rate, σ is the Stefan-Boltzmann constant, and R_{ISCO} is the innermost stable circular orbit (ISCO). The temperature increases inward, reaching a maximum near R_{ISCO}.

The total bolometric luminosity L_{bol} of an accreting black hole is related to the mass accretion rate $\dot{M}$ and the radiative efficiency η by:

$$L_{\text{bol}} = \eta \dot{M} c^2, \tag{1.7}$$

where η represents the fraction of the rest mass energy of the accreted matter that is converted into radiation. The value of η depends on the spin of the black hole as it affects the innermost stable circular orbit (ISCO) of the accretion disk, which in turn influences the radiative efficiency of the accretion process. For a non-rotating (Schwarzschild) black hole, the ISCO is approximately $6R_g$ resulting in $\eta \approx 0.06$, while for a maximally rotating (Kerr) black hole, the ISCO can be as close as R_g an in turn η can be as high as 0.42. [Netzer, 2013].

The *Eddington luminosity* L_{Edd}, represents the theoretical upper limit at which radiation pressure outward balances the gravitational pull inward. It is derived by equating the outward radiation force with the inward gravitational force on the infalling matter:

$$L_{\text{Edd}} = \frac{4\pi G M_{\text{BH}} m_p c}{\sigma_T}, \tag{1.8}$$

where M_{BH} is the black hole mass, m_p is the proton mass, and σ_T is the Thomson scattering cross-section for electrons. Substituting the constants, this becomes:

$$L_{\text{Edd}} \approx 1.3 \times 10^{38} \left(\frac{M_{\text{BH}}}{M_\odot} \right) \text{ erg s}^{-1}. \tag{1.9}$$

The Eddington accretion rate $\dot{M}_{\text{Edd}}$ is the mass accretion rate required to sustain the Eddington luminosity, assuming a given radiative efficiency η. From the equation

$L_{\rm bol} = \eta \dot{M} c^2$, substituting $L_{\rm Edd}$ for $L_{\rm bol}$ yields:

$$\dot{M}_{\rm Edd} = \frac{L_{\rm Edd}}{\eta c^2}. \tag{1.10}$$

Substituting $L_{\rm Edd}$ from Eq. 1.12, we get:

$$\dot{M}_{\rm Edd} \approx 2.2 \times 10^{-8} \left(\frac{M_{\rm BH}}{M_\odot}\right) \left(\frac{0.1}{\eta}\right) M_\odot \ {\rm yr}^{-1}. \tag{1.11}$$

In practice, AGN often operate at a fraction of their Eddington luminosity. The ratio of the bolometric luminosity $L_{\rm bol}$, to the Eddington luminosity, known as the Eddington ratio $\lambda_{\rm Edd}$, provides insight into the accretion state of the SMBH:

$$\lambda_{\rm Edd} = \frac{L_{\rm bol}}{L_{\rm Edd}}. \tag{1.12}$$

Thin disks, which are radiatively efficient, operate at accretion rates below or near the Eddington limit. As the accretion rate approaches or exceeds the Eddington mass accretion rate $\dot{M}_{\rm Edd}$, radiation pressure becomes significant, and the thin disk transitions into a *slim disk* [Abramowicz et al., 1988]. In this regime, the disk becomes geometrically thick, and a portion of the gravitational energy is advected inward with the flow rather than being radiated away, leading to reduced radiative efficiency and harder emission spectra. At very low accretion rates, far below $\dot{M}_{\rm Edd}$, the gas density in the disk becomes too low for efficient cooling, resulting in an *advection-dominated accretion flow* (ADAF) or *radiatively inefficient accretion flow* (RIAF) [Narayan et al., 1998]. In such cases, most of the released energy is carried inward by the flow rather than being radiated, leading to a hot, thick disk that emits primarily in the X-ray band. This transition from thin disks to slim disks and eventually to ADAFs helps explain the varied observational properties of AGN.

1.1.2.3 Corona

Observations show that AGN emit a significant fraction of their energy in the X-ray band, with spectra characterized by a power-law shape and a high-energy cutoff at around 100–300 keV (Figure 1.2). This X-ray emission cannot originate from the accretion disk itself, as the disk temperature, even in its innermost regions, reaches only $\sim 10^5$ K, peaking in the UV band. The observed X-rays must therefore be pro-

duced by a separate, hotter component—believed to be a hot, optically thin corona [Deufel et al., 2002, Netzer, 2013].

The primary mechanism behind the X-ray emission in AGN is **inverse Compton scattering**, where low-energy photons gain energy by interacting with high-energy (relativistic) electrons. In AGN, soft UV photons emitted by the accretion disk are thought to be upscattered by hot electrons in the corona, resulting in a characteristic power-law X-ray spectrum. The energy of the scattered photon is approximately given by:

$$E_{\text{scat}} \approx \gamma^2 E_{\text{init}}, \tag{1.13}$$

where γ is the Lorentz factor of the electron and E_{init} is the initial energy of the photon [Padovani et al., 2017]. This process explains the hard X-ray continuum observed in AGN spectra and produces a power-law spectral shape with a constrained photon index Γ, where

$$F_E \propto E^{-\Gamma} \tag{1.14}$$

and Γ typically ranges from 1.5 to 2.5 [Bogensberger et al., 2024a].

The strength of inverse Compton scattering is characterized by the Compton y-parameter, which quantifies the amount of energy gained by the photons during scattering. It is defined as:

$$y = \frac{4kT_e}{m_e c^2}\tau, \tag{1.15}$$

where T_e is the electron temperature, m_e is the electron mass, c is the speed of light, and τ is the optical depth of the corona. For AGN, y is typically close to 1, indicating that significant Comptonization occurs. The temperature of the coronal electrons, T_e, and the optical depth, τ, are the main factors that determine the spectral shape of the X-ray emission.

The observed high-energy cutoff E_{cut} is related to the coronal temperature by:

$$E_{\text{cut}} \approx 2 - 3\, k_B T_e, \tag{1.16}$$

where k_B is the Boltzmann constant. Typical values of T_e inferred from observations range from 10^8 to 10^9 K, corresponding to cutoff energies between 100 and 300 keV. Beyond the cutoff, the flux drops sharply, marking the high-energy limit of the scattering process [Padovani et al., 2017].

The properties of the corona, such as T_e, τ, size, and geometry, can be inferred through various observational methods. X-ray spectroscopy provides direct measurements of the high-energy cutoff $E_{\rm cut}$, which is related to the coronal temperature through equation (1.16), and constrains the optical depth τ through the shape of the power-law spectrum. Reflection features, such as the iron Kα line at 6.4 keV and the Compton hump at 20–30 keV, further help determine the coronal geometry and its distance from the disk [Padovani et al., 2017]. Rapid X-ray variability, on timescales of hours to days, offers insight into the size of the corona, with typical estimates placing it within a few gravitational radii of the black hole. Then, X-ray polarization studies can distinguish between different geometries, such as a compact, patchy corona near the disk or a vertically extended lamppost-like structure, based on the degree and orientation of polarization [Netzer, 2013, Marin et al., 2024]. Together, these methods provide a comprehensive understanding of the corona's physical conditions.

1.1.2.4 Torus

AGN typically exhibit different observational properties based on their orientation relative to the observer. The idea of a dusty obscuring structure originated from observations of type 1 and type 2 AGN, which show the same narrow-line emission but differ in their broad-line and continuum properties. Type 1 AGN exhibit broad emission lines and strong ultraviolet (UV) and optical continuum, while type 2 AGN show only narrow lines and weaker continuum emission [Antonucci, 1993]. These differences can be attributed to the presence of a torus-shaped structure surrounding the accretion disk and broad line region (BLR). When the torus is viewed edge-on, it obscures the central engine and BLR, allowing only narrow lines from the more extended narrow line region (NLR) to be observed, resulting in a type 2 spectrum. Respectively, when viewed face-on, the BLR and accretion disk are directly visible, producing the type 1 spectrum [Netzer, 2013].

Direct imaging of the torus has been achieved in a few nearby AGN using high-resolution interferometric observations at infrared (IR) wavelengths, providing additional evidence for its existence. These observations reveal structures with sizes ranging from a fraction of a parsec to several tens of parsecs, consistent with the expected dimensions of the torus [Tristram et al., 2007, Burtscher et al., 2013].

The torus is thought to be composed of gas and dust with temperatures ranging from a few hundred to about 1,500 K, corresponding to the sublimation temperature of dust grains [Netzer, 2013]. Its inner edge is defined by the dust sublimation radius, which as expected depends on the luminosity of the central source:

$$R_{\text{sub}} \approx 0.4 \left(\frac{L_{\text{bol}}}{10^{45}\ \text{erg s}^{-1}} \right)^{0.5} \text{pc}, \tag{1.17}$$

where L_{bol} is the bolometric luminosity of the AGN[3]. Beyond this radius, the gas cools sufficiently for dust grains to survive.

Observations suggest that the torus extends from a fraction of a parsec to several tens of parsecs [Netzer, 2013]. The dust and gas seem to be distributed in discrete clouds rather than a continuous medium. This clumpiness explains the observed variability in AGN infrared emission and the occasional only partial covering of the central regions.

The torus also reprocesses a significant fraction of the AGN's radiation, absorbing high-energy UV and optical photons from the central engine and re-emitting them in the infrared (IR). The resulting IR emission is a key property of the torus' properties and is often used to indirectly estimate the AGN's bolometric luminosity [Cackett et al., 2021].

1.1.2.5 Broad & narrow line regions

The existence of the broad-line region (BLR) and narrow-line region (NLR) in AGN is inferred from the distinct spectral features observed in their emission-line spectra. AGN typically show both broad and narrow emission lines, with different widths and ionization levels. Broad lines, such as Hα and Hβ, have FWHM values ranging from 10^3 to 10^4 km/s, indicating high-velocity gas close to the black hole. Narrow lines, such as [O III] and [N II], have FWHM values less than 10^3 km/s, suggesting gas at much larger distances, moving at lower velocities. These spectral differences naturally point to two distinct regions of line-emitting gas: a compact, high-density BLR near the black hole, and a more extended, lower-density NLR farther out [Net-

[3]The proportionality constant is an empirical calibration based on interferometric observations of nearby AGN [Kishimoto et al., 2007], while the exponent 0.5 is theoretically expected from radiative equilibrium considerations and has been observationally validated [Suganuma et al., 2006, Kishimoto et al., 2007].

zer, 2013].

Broad-Line Region (BLR): The BLR is located close to the accretion disk, typically within a few light days to a few light weeks from the central black hole [Netzer, 2013]. It consists of high-density clouds ($n_e \sim 10^9 - 10^{11}$ cm^{-3}) moving at high velocities (up to tens of thousands of kilometers per second) under the influence of the black hole's gravitational field[4]. These high velocities result in the broadening of emission lines, with FWHM values ranging from 10^3 to 10^4 km/s.

The BLR emits prominent broad lines, including hydrogen Balmer lines (Hα, Hβ), helium lines, and high-ionization lines such as C IV and Mg II. These emission lines are produced by photoionization, with the ionizing radiation originating from the accretion disk. The size of the BLR can be estimated using the empirical size-luminosity relation derived from reverberation mapping studies [Kaspi et al., 2000]:

$$R_{\text{BLR}} \propto L_{5100}^{0.70\pm0.03}, \tag{1.18}$$

where L_{5100} is the monochromatic luminosity at 5100 Å[5] . For a typical AGN with $L_{5100} = 10^{44}$ erg s^{-1}, R_{BLR} is on the order 10^{-2} pc.

The ionization parameter U, which quantifies the ratio of ionizing photon density to general gas density, is generally higher in the BLR than in the NLR due to the proximity to the ionizing source. U in the BLR typically ranges from 10^{-1} to 1 [Netzer, 2013].

Narrow-Line Region (NLR): The NLR is located much farther from the SMBH, extending from several hundred parsecs to a few kiloparsecs [Netzer, 2013]. The gas in the NLR is less dense ($n_e \sim 10^3 - 10^6$ cm^{-3}) and moves at lower velocities com-

[4]While Keplerian motion is generally assumed for gas in the BLR, there is evidence that non-Keplerian dynamics, such as outflows and radiation-pressure-driven winds, may also play a significant role. In some AGN, high-ionization lines such as C IV exhibit blueshifted profiles, suggesting that at least part of the BLR gas is participating in an outflowing wind rather than being confined to bound Keplerian orbits [Netzer, 2013].

[5]Since direct measurements of the ionizing UV luminosity are challenging due to absorption, the optical luminosity at 5100 Å, L_{5100}, is used as a proxy for the ionizing flux because it scales with the accretion disk's total output. The size of the BLR corresponds to the distance where the ionizing flux is sufficient to maintain photoionization equilibrium, leading to the theoretical scaling $R_{\text{BLR}} \propto L_{5100}^{0.5}$. Empirically, reverberation mapping studies confirm a slightly steeper relation, $R_{\text{BLR}} \propto L_{5100}^{0.6\pm0.1}$, likely due to variations in BLR cloud distribution and geometry [Peterson, 2006, Kaspi et al., 2000].

pared to the BLR, resulting in narrower emission lines with FWHM values typically less than 10^3 km/s.

Prominent narrow lines emitted by the NLR include forbidden lines such as [O III], [N II] and [S II]. These lines are produced primarily by photoionization, with the ionizing radiation originating from the central region. Due to the lower density and larger spatial extent, de-excitation due to collisions is less significant in the NLR, allowing forbidden lines to form. The ionization parameter in the NLR is lower than in the BLR, typically ranging from 10^{-3} to 10^{-1}[Netzer, 2013].

Below is a table (Table 1.1) summarizing and directly comparing the above characteristic values between the BLR and NLR.

Property	**BLR**	**NLR**
Distance from SMBH	10^{-2}–10^{-1} pc	10^{2}–10^{3} pc
Density (n_e)	$10^{9} - 10^{11}$ cm^{-3}	$10^{3} - 10^{6}$ cm^{-3}
Velocity (FWHM)	10^{3}–10^{4} km/s	$< 10^{3}$ km/s
Ionization Parameter (U)	10^{-1} to 1	10^{-3} to 10^{-1}
Prominent Lines	Hα, Hβ, C IV, Mg II	[O III], [N II], [S II]

Table 1.1: Comparison of the characteristic properties of the Broad Line Region and Narrow Line Region in AGN.

1.1.2.6 Relativistic jets

The relativistic jets that are characteristic of some AGN are highly collimated outflows of plasma that have been observed extending over vast distances, from several parsecs to hundreds of kiloparsecs, and emit across the entire electromagnetic spectrum, from radio to gamma rays. Despite decades of observations, the exact mechanisms responsible for jet formation and intense collimation remain poorly understood. [Netzer, 2013].

Observationally, jets are most prominently associated with radio-loud AGN, including blazars and radio galaxies. High-resolution imaging with radio interferometers, such as the Very Long Baseline Array (VLBA), has revealed detailed jet structures suggesting that the plasma flows are not uniform but instead consist of discrete components or regions of varying density and velocity. [Alloin et al., 2006]

1.2 Phenomenology of AGN

The different emission components of the accretion flow have historically led to a large diversity in the observed properties of AGN depending on the wavelength of the observations and the orientation of the accretion flow system relative to the observer. This diversity in the observed properties of AGN has also led to classification schemes devised by astronomers in attempt to better understand the AGN phenomenon. Some of these classifications are arbitrary and are only kept for historical purposes. Others have physical origin and therefore provide information on the structure of accretion flows.

1.2.1 AGN Classification

AGN classification (Figure 1.3) arises from observable differences caused primarily by orientation and, to a lesser extent, intrinsic properties such as luminosity and jet power . The unification paradigm provides a framework to explain these differences by positing that all AGN are manifestations of the same underlying processes with differing observation conditions [Netzer, 2013].

AGN are first divided into two broad categories based on their *radio-wave emissions*:

- **Radio-quiet AGN**, which lack significant jet activity and are primarily observed through their optical, UV, and X-ray emission.
- **Radio-loud AGN**, which feature powerful relativistic jets and produce significant radio emission [Antonucci, 1993].

Within each of these categories, AGN are further classified based on:

1. ***Luminosity***: Based on the bolometric luminosity of the AGN:
   - **Seyferts**: Low to intermediate luminosity AGN ($L_{\text{bol}} \lesssim 10^{45}$ erg s^{-1}), typically found in nearby spiral galaxies.
   - **QSOs (Quasi-Stellar Objects)**: High-luminosity AGN ($L_{\text{bol}} \gtrsim 10^{45}$ erg s^{-1}), able observed at high redshifts and as such able to serve as cosmological probes for early universe conditions. [Netzer, 2013].

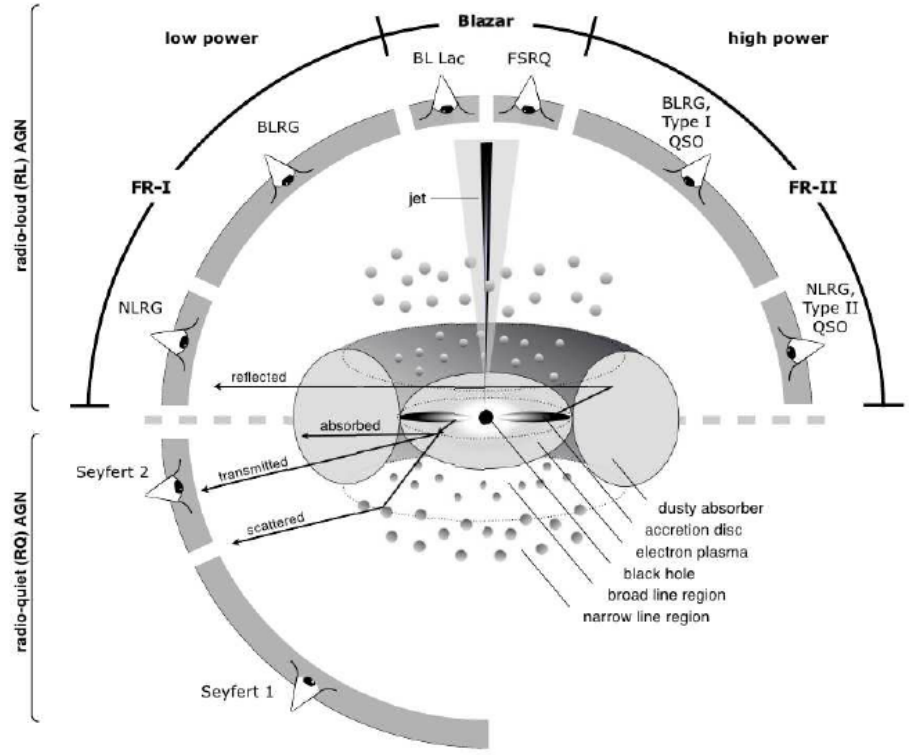

Figure 1.3: Schematic representation of our understanding of the AGN phenomenon in the unified scheme. The type of object we see depends on the viewing angle, whether or not the AGN produces a significant jet emission, and how powerful the central engine is. Graphic by Marie-Luise Menzel [Beckmann and Shrader, 2012].

2. ***Obscuration***: Depending on whether the BLR and accretion disk are directly visible or obscured by the torus:

   - **Type 1 AGN**: BLR and continuum are visible, resulting in broad emission lines and a strong UV/optical continuum.
   - **Type 2 AGN**: BLR and continuum are obscured by the torus, leaving only narrow lines from the NLR visible.

Spectropolarimetric[6] observations of some type 2 AGN reveal broad lines in polarized light, indicating that the BLR is present but obscured, with scattered light reaching the observer. This provides strong evidence supporting the unification model.

[6]Observations that involve measuring and comparing the polarization of light across different wavelengths

Blazars are a subclass of radio-loud AGN where the relativistic jet is pointed almost directly toward the observer. This alignment results in extreme relativistic beaming, causing the observed emission to be dominated by non-thermal radiation from the jet. Blazars exhibit strong variability across all wavelengths and are further divided into:

- **Flat Spectrum Radio Quasars**: High-luminosity blazars with strong broad emission lines and a prominent accretion disk.
- **BL Lacertae objects**: Low-luminosity blazars with weak or no emission lines and spectra dominated by synchrotron emission from the jet.

We will later on (Chapter 1.4) focus on a sample of type-I QSOs in the redshift range $z = 0 - 4$ due to their high Luminosities even at higher redshifts and along with their easier observation conditions as type-I AGN with an unobscured line of sight to the central engine which makes them ideal for long-term observation.

1.3 Variability in AGN

One of the most distinctive features of AGNs is their flux variability, observed as irregular, aperiodic changes in their brightness across a broad range of wavelengths and timescales. This variability, particularly in the X-ray and optical bands, is believed to originate from processes within the accretion flow and corona, regions that lie only a few gravitational radii from the central black hole [Vaughan et al., 2003]. Since these regions cannot be directly imaged, variability studies serve as a key tool for examining the physical structure and dynamics of the innermost parts of AGNs.

While AGNs may exhibit other forms of variability, such as quasi-periodic oscillations [Das and Czerny, 2011], tidal disruption events (TDEs) [Komossa, 2015], or sudden changes in spectral state (e.g., changing-look AGNs) [Veronese et al., 2024], we will focus almost exclusively on stochastic variability, characterized by random, aperiodic flux variations [Georgakakis et al., 2021]. This type of variability is commonly modeled as a ”red noise” process, where fluctuations on longer timescales dominate the variability power spectrum. Such stochastic variability is also be-

lieved to carry information about the physical processes governing accretion [Kelly et al., 2009].

1.3.1 Importance of variability studies

Since the innermost regions of AGNs—where the accretion disk and corona reside—are impossible to resolve directly with current imaging techniques, variability provides an indirect method for examining these regions. By analyzing the stochastic fluctuations in flux, we can gain insight into the structure of the accretion flow, the properties of the corona, and the mechanisms driving the observed variability. The shortest variability timescales observed in AGN light curves set an upper limit on the size of the emitting region, as the signal cannot vary faster than the light-crossing time of the region:

$$t_{\text{cross}} = \frac{R}{c}, \tag{1.19}$$

where R is the characteristic size of the emitting region and c is the speed of light.

For typical AGN with X-ray variability timescales ranging from hundreds of seconds to a few hours, the inferred sizes are on the order of a few gravitational radii from the SMBH. This implies that the X-ray emission originates from a very compact region close to the event horizon [Vaughan et al., 2003, Bogensberger et al., 2024a].

In addition, variability studies are invaluable in constraining models that attempt to explain the physics of accretion. Various models have been proposed to explain the observed flux variations, including instabilities in the accretion disk, changes in coronal conditions, and turbulence in the accretion flow [Kelly et al., 2009]. By comparing observed variability patterns to theoretical predictions, we can test these models and identify the dominant mechanisms responsible for AGN variability.

In the context of accretion disk models for example, variability timescales may provide insights into whether the disk follows the Shakura-Sunyaev [Shakura and Sunyaev, 1973], slim disk [Abramowicz et al., 1988] or ADAF models [Narayan et al., 1998] while short-term X-ray variability better constrains the geometry and physical size of the corona, distinguishing between the lamp-post and patchy corona

models. Additionally, time lags between X-ray and UV/optical variability can be used to infer the radial structure of the disk and test corresponding reprocessing and reverberation models [Cackett et al., 2021].

1.3.2 X-ray Variability in AGN

The significance of the X-ray variability for understanding the inner structure of accretion flows has led to systematic studies to understand the statistical properties of the X-ray flux variations at different timescales and how they depended on the fundamental properties of system, such the mass of the black hole and its Eddington ratio. This is not only important for studying the inner stricture of accretion flows but can also provide constraints to models for the origin of the flux variations.

The most complete characterization of the stochastic flux variability of AGN is by decomposing long and uninterrupted light curves to its Fourier components. The amplitude of the a given Fourier components as a function of the corresponding frequency is refereed to as the power spectral density (PSD).

The PSD of AGN X-ray observations is often used to characterize the distribution of variability power across different timescales. Observations have shown that the PSD typically follows a broken power-law shape (Figure 1.4), with a high-frequency slope of -2 and a low-frequency slope of -1.

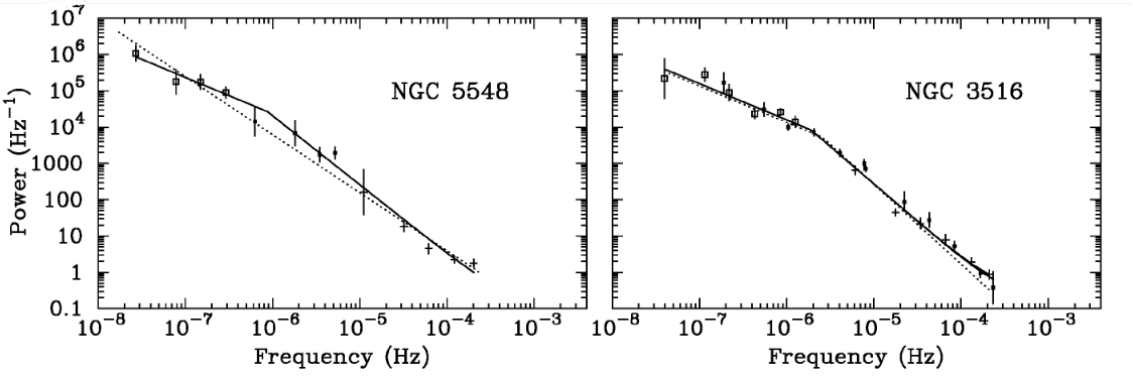

Figure 1.4: PSD plots for the AGN NGC 5548 and NGC 3516, derived from RXTE observations in the 2-10 keV energy range showing both a naive single power-law fit (dotted line) and a more accurate broken power-law fit (solid line). Adapted from [Uttley et al., 2002]

.

The break frequency between the above two regimes is found empirically to be inversely proportional to black hole mass [McHardy et al., 2006]. This is consistent with models suggesting a link between variability and a characteristic timescale within the accretion flow, such as the viscous timescale, which governs how efficiently angular momentum is transported outward and is known to scale linearly with black hole mass: $t_{\text{vis}} \propto M_{BH}$, and as such the break frequency would then scale inversely with it [Netzer, 2013]:

In addition to black hole mass, studies have suggested that both the break frequency as well as the amplitude of the PSD may depend on the Eddington ratio. For example, Ponti et al. [2012] propose that the PSD amplitude is inversely proportional to the Eddington ratio, implying that AGN with higher accretion rates (relative to the Eddington limit) may exhibit smaller amplitude variations. The physical interpretation of this trend is not entirely clear, but it may provide insights into disk or corona instabilities and how variability propagates inward for AGN accreting at different rates.

Despite significant progress in understanding AGN X-ray variability, many questions remain. In particular, the physical conditions of the corona, including its geometry and size, are still under debate. Competing corona models range from compact, centrally located regions to more extended, patchy structures [Padovani et al., 2017]. Furthermore, while there is clear evidence that the variability amplitude inversely correlates with black hole mass, the dependence on the Eddington ratio is less well-constrained. While some models support the dependence of the break frequency on both eddington ratio and black hole mass [McHardy et al., 2006] and even propose a dependence of the PSD normalization on the Eddington ratio [Padovani et al., 2017], other models with constant normalization provide comparable fits to the observed data. Recent studies that highlight the complexity of disentangling intrinsic variability mechanisms from external factors such as obscuration and changes in the line of sight, further complicating the interpretation of observed variability trends [Bogensberger et al., 2024b, Paolillo et al., 2017].

These open questions regarding the correlation between the amplitude of observed X-ray variability and the physical properties of the accretion events such as the Eddington ratio and the black hole mass will be the focus of this thesis.

1.4 Thesis Motivation & Objectives

We are interested in the stochastic intrinsic variability of AGNs in the X-ray band, which is believed to be directly linked to the source's physical characteristics, such as black hole mass and accretion rate. Our objective is to investigate whether AGNs with similar intrinsic properties exhibit comparable variability behavior on timescales of a few hours to a few days. Studying X-ray variability on these timescales is particularly important, as it likely corresponds to physical processes occurring in the innermost regions of the accretion disk and corona.

Understanding this variability is also important for the broader comprehension of AGN's role in galaxy evolution, as active galactic nuclei significantly influence the evolution of their host galaxies through feedback mechanisms that regulate star formation and redistribute energy in the interstellar medium. AGNs can even serve as observational tools for studying cosmic structure, as their high luminosity and characteristic variability make them ideal probes for calibrating cosmological distances.

Distinguishing intrinsic variability mechanisms from external factors, such as obscuration and changes in the line of sight, requires a complex methodology. The low photon count rates in many XMM-Newton observations posed challenges for conventional variability estimation methods, leading us to adopt Bayesian methodologies for a more reliable analysis. These methodologies allow us to derive variability estimates not only for individual AGN but also for ensembles, accounting for trends in populations of similar physical parameters.

A significant part of our work involves the development of a custom light curve extraction pipeline for XMM-Newton observations that would automatically go through steps such as data reduction, proper selection of source and background regions, background subtraction, and light curve generation. Particular emphasis was placed on optimizing the pipeline for sources with low count rates, ensuring that the extracted light curves retain as much variability information as possible. This process is important for gathering a coherent dataset of light curves for our variability analysis.

Chapter 2

Observational Data

2.1 XMM-Newton

2.1.1 Overview

The XMM-Newton space telescope, launched by the European Space Agency (ESA) in December 1999, has played a pivotal role in X-ray astronomy by providing high-sensitivity observations across a broad energy range (0.2–12 keV). Designed for long-duration observations of cosmic X-ray sources its main objective is to observe X-ray emissions from a variety of astrophysical sources, including active galactic nuclei (AGN), galaxy clusters, and stellar coronae.

The observatory follows a highly elliptical orbit with an apogee of 114,000 km and a perigee of 7,000 km with an average period of about 48 hours, allowing for long uninterrupted observations during each orbit.

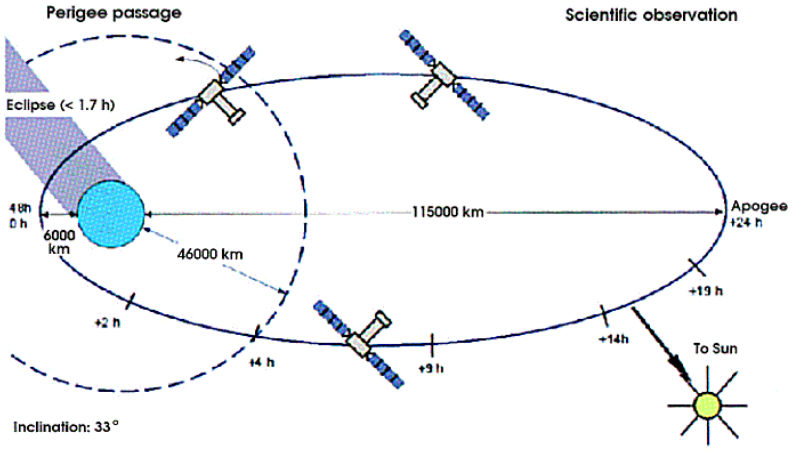

Figure 2.1: Sketch of the highly elliptical XMM-Newton orbit [XMM-Newton Community Support Team].

It consists of three Wolter Type-I[1] X-ray telescopes mounted in parallel, along with an Optical Monitor (OM) for simultaneous optical and UV observations. The X-ray telescopes are designed to focus X-rays via grazing incidence[2]. Each telescope has 58 nested mirrors coated with gold, which ensures high reflectivity for X-rays across the full energy range. The focal length of these telescopes is 7.5 meters, and they achieve an angular resolution ≈ 6 arcseconds on-axis[3]. This is ideal for performing high-sensitivity observations of faint X-ray sources over long exposure times.

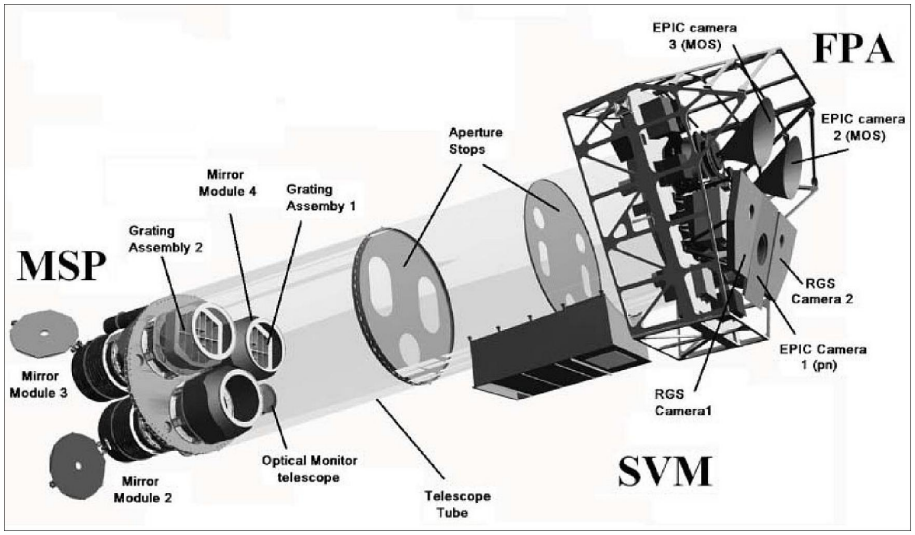

Figure 2.2: Schematic layout of the XMM-Newton observatory, with external shrouds and structure removed for clarity, illustrating the placement of its components: the three X-ray telescopes (mirror modules), the RGS detectors, and the EPIC system with its MOS and PN detectors. The Optical Monitor is also shown slightly obscured, positioned next to the telescope tube [XMM-Newton Community Support Team].

[1]A telescope that uses a combination of parabolic and hyperbolic mirrors to focus X-rays via grazing incidence, minimizing spherical aberrations and maximizing reflection efficiency

[2]A method of reflecting X-rays off a surface at very shallow angles, in order to allow high-energy photons to be reflected without penetrating the surface.

[3]On-axis refers to observations made at the center of the field of view, where the optical alignment provides the best resolution. Off-axis, the resolution decreases due to distortions and aberrations at larger angles.

2.1.2 XMM-Newton detectors

The three telescopes are arranged to support the instruments located at the focal plane. Out of these three telescopes, two feed the Reflection Grating Spectrometers (RGS), while all three provide data to the European Photon Imaging Camera (EPIC) system.

2.1.2.1 European Photon Imaging Camera (EPIC)

The EPIC system comprises two MOS (Metal-Oxide-Semiconductor) cameras and one PN (p-n junction) camera, each positioned behind one of the three X-ray telescopes. The MOS cameras offer high spectral resolution, while the PN camera provides superior time resolution[4] and quantum efficiency[5], making it highly effective for observing bright and rapidly variable sources. The EPIC cameras have a field of view of 30 arcminutes and cover an energy range from 0.15 keV to 15 keV (although quantum efficiency drops sharply after 12 keV(fig. 2.3)). The PN camera can achieve readout times as low as 73 ms, enabling precise measurements of fast-changing phenomena such as flaring AGN and quasars.

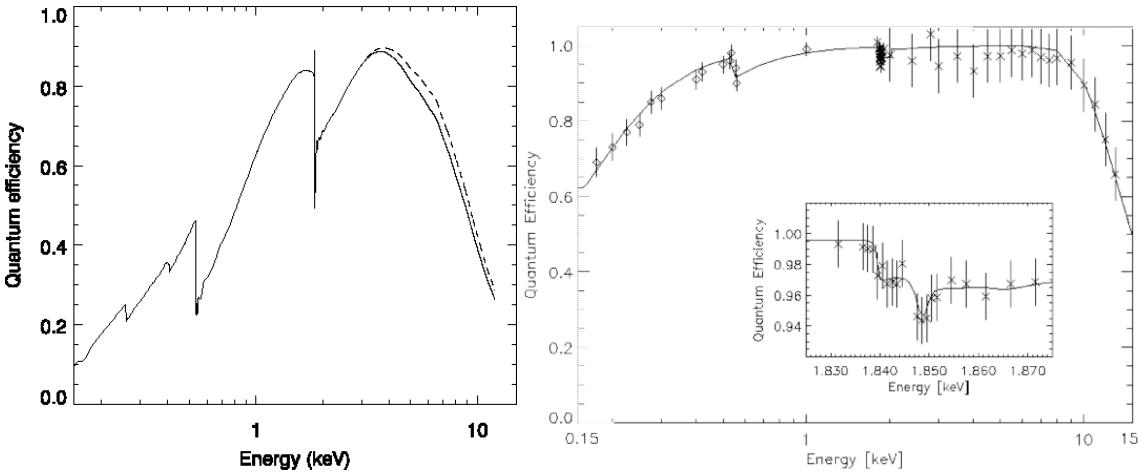

Figure 2.3: **Left:** Quantum efficiency of the EPIC MOS1 (solid line) and MOS2 (dashed line) CCD1 chip as a function of photon energy. **Right:** Quantum efficiency of the EPIC pn CCD chips as a function of photon energy [Strüder et al., 2001].

[4]The ability to measure variations in intensity over short intervals.
[5]The ratio of detected photons to actual incident photons from the source.

2.1.2.2 Reflection Grating Spectrometer (RGS)

The RGS is optimized for high-resolution spectroscopy in the soft X-ray range (0.35–2.5 keV). It works by dispersing incoming X-rays using a grating assembly before they reach the CCD detectors, enabling detailed spectral analysis of hot plasma in galaxy clusters and the inner regions of AGN.

2.1.2.3 Optical Monitor (OM)

The Optical Monitor is a 30 cm optical/UV Ritchey-Chrétien[6] telescope mounted alongside the X-ray telescopes. It operates simultaneously with the EPIC and RGS instruments, providing complementary optical and UV data. The OM covers a wavelength range from 180 nm to 600 nm with a field of view of 17 arcminutes, allowing it to detect optical counterparts of X-ray sources and monitor variability across multiple wavelengths.

[6] A type of reflecting telescope that uses hyperbolic primary and secondary mirrors to minimize optical aberrations, providing a wide, flat field of view.

2.1.3 XMM-Newton Data

As a measure of source intensity, we will be using **counts**, a direct measure of photons detected by the telescope's instruments. Counts form the fundamental unit of measurement in X-ray astronomy, representing the number of photons observed within a given energy range and spatial region, as recorded by the detector. By integrating counts over the exposure time of the observation, we derive **count rates**, which quantify the number of detected photons per second and provide a normalized measure of source intensity, accounting for differences in exposure time.

2.1.3.1 Serendipitous Source catalog

The XMM-Newton Science Survey Centre (XMM-SSC) has undertaken the task of systematically analyzing all imaging-mode observations from the EPIC cameras. This effort has resulted in the **XMM-Newton Serendipitous Source catalog**, an invaluable resource for the scientific community. The catalog contains information on X-ray fluxes, count rates, and detected counts along with their associated errors across different energy bands for $\approx 700,000$ unique sources over $\approx 1,000,000$ detections by XMM-Newton.

The Serendipitous Source catalog also provides various derived products, including source detections, images, and spectra. These datasets enable large-scale statistical studies of X-ray sources observed by XMM. The energy bands defined by the XMM-Newton Serendipitous Source catalog are discretely divided as:

Band	**Energy Range (keV)**
Band 1	0.2–0.5
Band 2	0.5–1.0
Band 3	1.0–2.0
Band 4	2.0–4.5
Band 5	4.5–12.0
Band 6 (Soft band)	0.2–2.0
Band 7 (Hard band)	2.0–12.0
Band 8 (Total band)	0.2–12.0

Table 2.1: Energy bands defined by XMM-Newton catalogs.

Each band can provides unique insight into different physical processes. For variability studies, Band 6 is often preferred due to its high signal-to-noise ratio in soft X-rays, which dominate AGN emission [McHardy et al., 2006, Paolillo et al., 2017].

Fluxes (F) are calculated in units of erg cm^{-2}s^{-1} using the relation:

$$F = CR \cdot \text{ECF},$$

where CR is the count rate (cts s^{-1}), and ECF is the energy conversion factor specific to the camera, filter, and energy band. Errors in flux are propagated from the count rate uncertainties.

2.1.3.2 Point spread function

Not all points in the detector have the same detection capabilities, as the telescope's optical system spreads the light from a point source, creating a characteristic distribution known as the **point spread function (PSF)**. The PSF describes this spatial distribution of light from a point source as detected by a telescope. It characterizes the telescope's imaging performance and accounts for the effects of diffraction, detector imperfections, and optical aberrations. For XMM-Newton, the characteristics of the PSF for each instrument chain can be found in Table 2.2 while images of the on-axis PSF for each of the XMM-Newton instruments can be seen in 2.4

Mirror Module	**2**	**3**	**4**
Instr. Chain[a]	pn	MOS-1+RGS-1	MOS-2+RGS-2
Orbit/Ground	orbit/ground	orbit/ground	orbit/ground
FWHM [″]	< 12.5[7]/6.6	4.3/6.0	4.4/4.5
HEW [″]	16.6/15.1	16.8/13.6	17.0/12.8

Table 2.2: The on-axis in-orbit and on-ground 1.5 keV PSFs of the different X-ray telescopes [XMM-Newton Community Support Team].

The EPIC-PN detector has a broader PSF compared to the MOS cameras, primarily due to its larger pixel size , which results in a lower spatial resolution but higher quantum efficiency. It has a FWHM of approximately 12.5 arcsec and a

[7]The core of mirror module 2 cannot be better resolved in orbit because of the large pn-CCD pixel size.

Half Energy Width[8] (HEW) of about 16.8 arcsec, making it sensitive to faint sources while also being more affected by out-of-time (OOT) events. It should also be noted the effective area of the telescope also decreases with increasing off-axis angle due to vignetting effects. These effects need to be corrected during data reduction to ensure accurate count measurements.

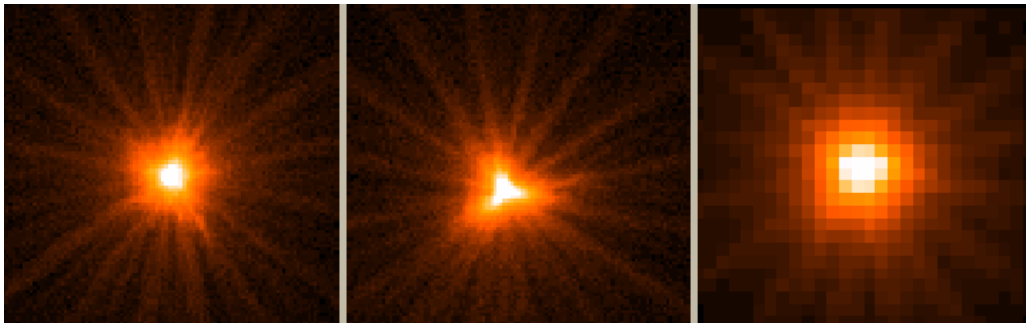

Figure 2.4: On-axis PSF of MOS1, MOS2, and pn detectors (left to right), registered on the same source. The MOS cameras use 1.1 arcsec pixel size and the pn camera used 4.1 arcsec pixel size. Images are 110 arcsec wide with a square root scale to visualize PSF wings [XMM-Newton Community Support Team].

2.1.3.3 Sky footprint

The cross-matching of sources between different catalogs that we are required to perform (see section 2.2.2) for this study along with the inhomogeneous nature of the XMM data releases demands a solid understanding of the complex spatial footprint of each of our catalogs. To address this we will be using sets of **Multi-Order Coverage (MOC)** files. These files are based on the **HEALPix** (Hierarchical Equal Area isoLatitude Pixelization) scheme, a framework for mapping data on the sphere. HEALPix divides the sky into equal-area pixels, enabling efficient data storage and analysis across varying scales. A MOC file encodes the regions of the sky observed by XMM-Newton, helping with identifying overlaps and determining sky coverage (see Figure 2.5.

[8]Half Energy Width (HEW) is the angular diameter of the region that collects 50% of the total reflected photons from a point source. A key measure of a telescope's imaging sharpness and performance.

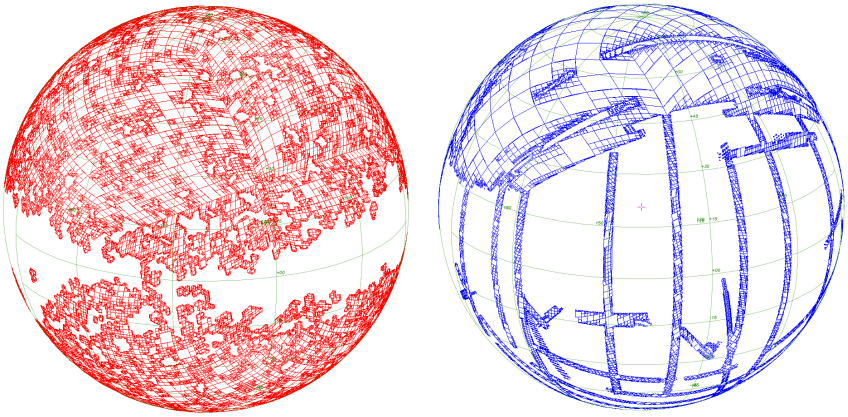

Figure 2.5: Comparison of MOC maps using the HEALPix framework for Different Astronomical Surveys. The left sphere (red) represents the GALEX AIS survey, consisting of approximately 70,000 HEALPix cells, while the right sphere (blue) represents the SDSS survey, with around 225,000 cells [Fernique et al., 2014].

For our analysis, we will be using data from the aforementioned XMM-Newton Serendipitous Source Catalog, designed to record X-ray sources detected incidentally during observations intended for other primary targets, leveraging the wide field of view and high sensitivity of the observatory to capture data on numerous sources beyond the intended study area. More specifically we will be using it's twelfth data release, **4XMM-DR12** [Webb et al., 2020] along with its associated MOC data for an effective area of ~1,283 deg^2. The catalog contains data for 939,270 X-ray detections corresponding to 630,347 unique X-ray sources.

2.2 Catalogs and dataset filtering

2.2.1 SDSS Quasar Catalog: Sixteenth Data Release

The Sloan Digital Sky Survey Quasar Catalog from Data Release 16 (DR16Q)[Lyke et al., 2020], is the most comprehensive catalog of spectroscopically confirmed quasars to date, making it a cornerstone for studies of QSO properties. Released in 2020 as part of the extended Baryon Oscillation Spectroscopic Survey (eBOSS), DR16Q comprises over 750,000 QSOs, with 225,082 of them newly identified, appearing here for the first time. The catalog extends from redshift $z \sim 0$ to $z > 7$, it covers 9,376 square degrees of the sky, with an average surface density of approximately 80

quasars per square degree.

DR16Q includes detailed spectroscopic data for quasars, such as emission-line properties, well calibrated redshift estimates, and multi-wavelength data. It also includes accurate sky positions for each quasar, allowing precise localization and cross-matching with external catalogs. Additionally for the purposes of our analysis, value-added catalogs have been developed on top of DR16Q to derive reliable and consistent estimates for physical parameters such as black hole mass, Eddington ratio (derived through empirical scaling relations of spectral data, see sections (1.1.2.1) and (1.1.2.2)).

Additionally, the SDSS name of each object in DR16Q serves as a unique, almost universal identifier, facilitating cross-matching across datasets and catalogs, including the aforementioned 4XMM-DR12.

2.2.2 Cross-matching and filtering

With the 4XMM-DR12 and DR16Q catalogs in hand, along with their respective MOC files, the next step involves cross-matching these datasets to extract the objects that appear in both catalogs and applying a series of filters based on various factors to refine the combined dataset for analysis.

The initial datasets include **630,347** unique X-ray sources from 4XMM-DR12 and **750,414** QSOs from DR16Q. By applying spatial overlap to the sections of the celestial sphere covered by both catalogs and cross-matching their sky coordinates (with a maximum accepted separation of 5 arcseconds), we identify **25,280** unique objects (QSOs) common to both catalogs.

Spurious Identification Rate :

The spurious identification rate is the probability of chance alignments between unrelated sources due to high source densities in large surveys. The probability of a random SDSS QSO coinciding with an X-ray source within a 5 arcsec radius is estimated using:

$$P_{spur} = 1 - \exp(-N_{QSO} \times \pi \times 5^2), \tag{2.1}$$

which results in a 0.05% probability that any given QSO-X-ray match within 5 arcsec is a chance coincidence. fewer than 1 spurious counterpart is expected per 2000 X-ray sources, supporting that over 99.9% of the identified associations are genuine and alleviating the need to correct for possible contamination of sources.

2.2.2.1 Signal-to-Noise

The first step in refining the matched dataset is filtering based on the signal-to-noise ratio (S/N) of the X-ray sources in order to include only higher-confidence detections and sufficiently bright sources (enabling the study of their light curves) with reliable measurements in our analysis. The S/N is calculated using the source count rates and their associated errors, which are derived from the 4XMM-DR12 catalog.

The signal-to-noise ratio for each source was computed as:

$$\text{S/N} = \frac{\sum_{i=1}^{3} CR_i}{\sqrt{\sum_{i=1}^{3} \delta CR_i{}^2}}, \tag{2.2}$$

where CR is the total count rate summed across energy bands 1, 2, and 3 (see table 2.1), totaling the energy of band 6 (0.2–2 keV), and δCR is the corresponding error. A threshold of S/N > 5 was applied, consistent with standard practices in X-ray astronomy [Vaughan et al., 2003]. This threshold is a trade-off between sufficiently large sample and sufficient number counts to enable analysis of the sources' light curves. Increasing the S/N threshold would improve the LC count statistics but reduce the overall sample size.

These source count rates CR represent the number of photons detected per second from a source in a specific energy band, after correcting for instrumental effects such as detector efficiency, background subtraction, and point spread function (PSF) losses. These rates are derived from the fitted PSF in sub-images of the X-ray data, ensuring accurate measurements that reflect the source's intrinsic X-ray emission. The associated errors are the statistical uncertainties on these rates.

After applying the signal-to-noise threshold of > 5, we are left with a dataset of **14,032** unique quasars observed over 20,414 distinct observations.

2.2.2.2 Duration

The next step in preparing the dataset for variability analysis involves filtering based on observation duration. The initial dataset includes observations with durations ranging from 0 to 140 ks (Figure 2.6). For our analysis, we apply a cutoff below 30 ks, retaining only those observations with durations above this threshold.

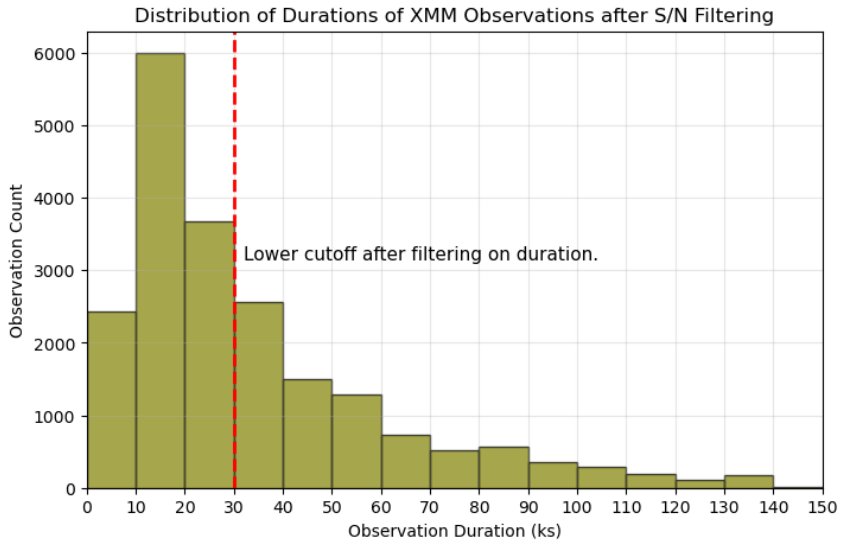

Figure 2.6: Histogram showing the distribution of XMM-Newton observation durations for observations matched between the 4XMM-DR12 and DR16Q catalogs after signal-to-noise filtering. The red dashed vertical line indicates the cutoff point (30 ks) for durations included in the analysis.

This choice of cutoff is primarily motivated by practical considerations rather than a strict scientific reasoning. One consideration is that longer observation durations are more likely to include signs of variability in the QSOs. Variability in AGNs occurs over a range of timescales, but shorter observations may not provide sufficient time for significant fluctuations in brightness to be made apparent. By focusing on longer observations, we ensure that the dataset is better suited for detecting and analyzing such variability. A second consideration is the feasibility of processing a large volume of observational data. The dataset includes a substantial number of observations with durations between 10 and 30 ks, comprising of well over half of the total observations. While these shorter-duration light curves are important, processing such a large subset would be computationally intensive. By applying the 30 ks cutoff, we significantly reduce the number of observations to be

analyzed, while still retaining a dataset with sufficient time coverage to investigate variability across timescales.

After applying a duration cutoff below 30 ks, we are left with a dataset of **8,332** unique quasars.

2.2.2.3 Off-axis angle / PN data / General good data

Off-axis angle: The off-axis angle, measured in arcminutes, represents the distance between the detection position and the on-axis position on the detector. Larger off-axis angles are often with reduced sensitivity and increased uncertainties due to vignetting and other instrumental effects[XMM-Newton Community Support Team]. For our dataset we limited the off-axis angle to a maximum of 12 arcminutes to reduce the impact of vignetting and improve the accuracy of flux and count rate measurements, keeping a sufficient number of sources while ensuring reliable observations.

SUM_FLAG: The 4XMM-DR12 catalog includes a *SUM_FLAG* within it that summarizes the quality of a detection based on several EPIC-specific flags. We selected only detections marked by this flag as ”good data”. This excluded detections affected by proximity to bright sources, low detector coverage, or regions prone to spurious detections.

PN_FLAG: Similarly the catalog includes a flag signifying that an observation includes valid and reliable EPIC-PN data[9] The PN camera provides the highest sensitivity among the EPIC instruments and is useful for detecting weaker X-ray sources, ensuring valid PN detections for our analysis that is often characterized by lower count rates.

After applying filtering for excessive off-axis angles, the sum ”good data” flag and the PN data flag the dataset was reduced to **4234** unique quasars.

[9]The PN flags provide quality checks for detections, including conditions such as low detector coverage, proximity to other sources, presence within extended emission, spurious detections near bright sources, or near hot areas. See Tab 3.1 of XMM-Newton User’s handbook [XMM-Newton Community Support Team].

2.2.3 Filtered dataset

By combining the 4XMM DR12 X-ray catalog with the SDSS DR16Q catalog and applying targeted filtering, we have assembled a robust dataset of ***4234*** unique QSOs corresponding to *5672* distinct observations. These span a broad redshift range (from z=0 to z≈5), with most sources clustering at z<3.

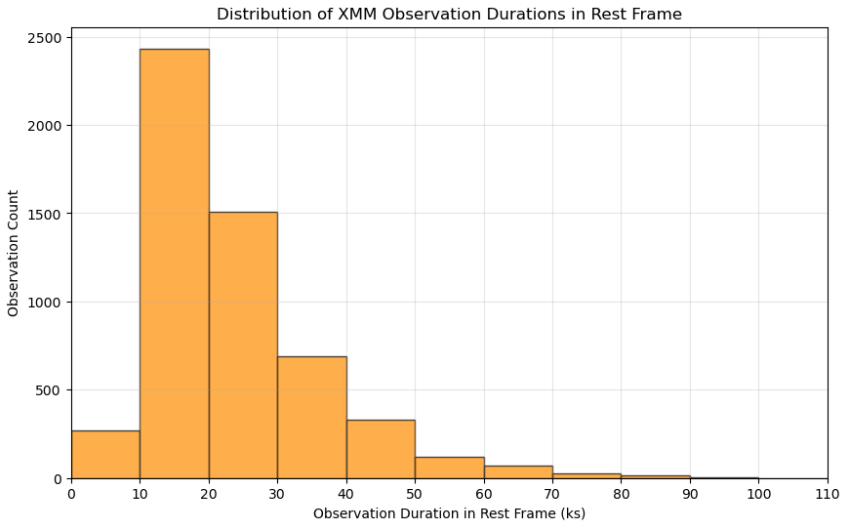

Figure 2.7: Histogram showing the distribution of XMM-Newton observation durations for observations in our final filtered dataset in their rest frame.

The observations have been observed at various exposure times (see Figure 2.8). To account for redshift, we present variability timescales in the sources' rest frame as:

$$t_{\text{rest}} = \frac{t_{\text{obs}}}{1+z} \tag{2.3}$$

This transformation allows us to more effectively express the actual timescales under investigation, highlighting the presence of shorter-timescale observations (see Figure 2.7). These will enable us to probe variability in the inner regions of the central engine, providing a direct examination of processes occurring near the central SMBH.

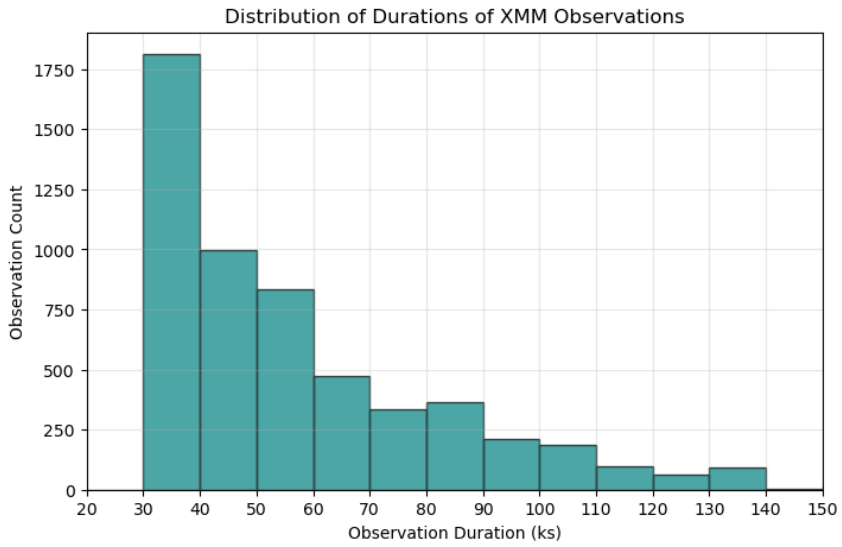

Figure 2.8: Histogram showing the distribution of XMM-Newton observation durations for observations in our final filtered dataset in the instrument's frame.

With with this refined dataset of QSOs in hand we seek to study the possible link between the observed variability of AGN and their physical characteristics—particularly black hole mass and Eddington ratio. To facilitate this, we group the QSOs into bins of similar physical properties (Figure 2.9). This binning allows us to test whether QSOs with similar physical characteristics exhibit consistent variability as implementations of similar physical processes and how variability trends evolve with changes in these parameters. The black hole masses and Eddington ratios used in this analysis are derived from the SDSS DR16Q value added catalog, as showcased in (2.2.1).

As shown in (figure 2.9), the bulk of our sources occupies the range $-2 \leq \log(\lambda_{\text{Edd}}) \leq 0$ and $8 \leq \log(M_{\text{BH}}) \leq 10$. Given the high concentration of data in this region, these bins will serve as the primary focus for our subsequent analysis.

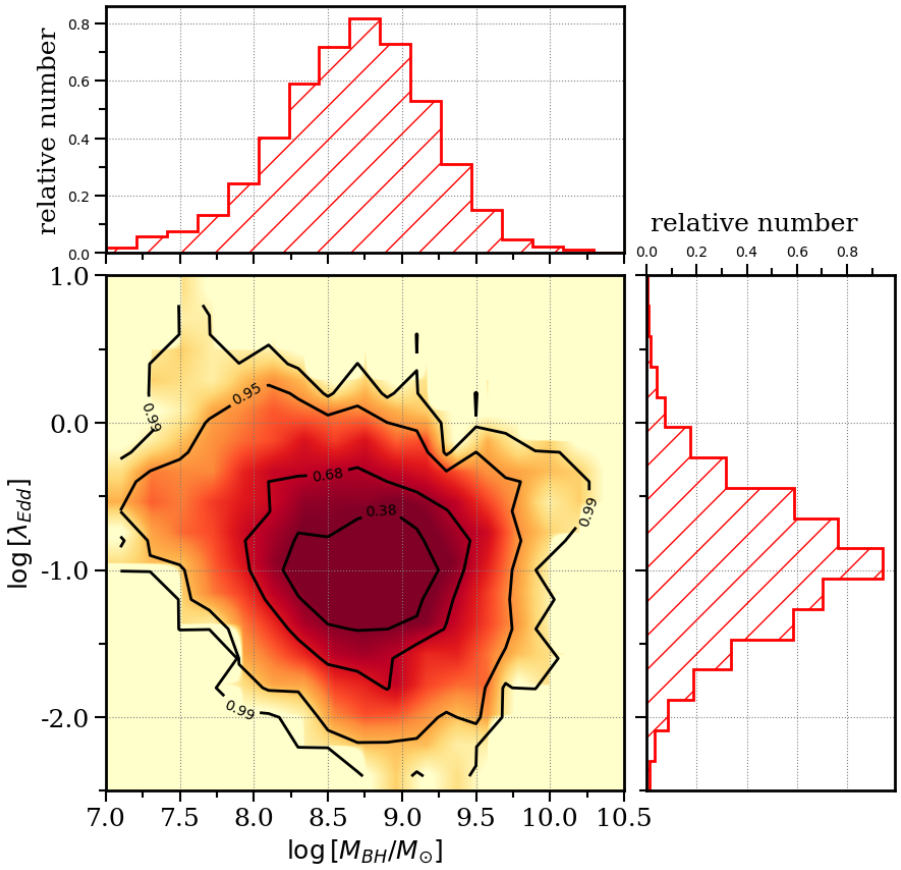

Figure 2.9: Distribution of the filtered dataset of QSOs on the Eddington ratio vs black hole mass plane. The contour levels enclose 34, 68, 95 and 99 per cent of the QSO population. The 1-dimensional projections of this distribution along the black hole mass and Eddington ratio axes are also shown by the histograms on the top and to the right of the main panel respectively.

Chapter 3

Statistical Methodology

3.1 Overview

AGN light curves, while rich in information, are inherently challenging to interpret. Observational constraints, such as limited telescope availability can often result in sparse and uneven sampling. This, combined with the relatively low photon counts typical of many X-ray observations, results in observations that are significantly impacted by the influence of background noise, a fact that complicates the extraction of the intrinsic variability signals we seek. And in addition the effects of redshift also stretch the observed timescales, making it more difficult to compare variability properties across different populations [Paolillo et al., 2017]. Overcoming these challenges requires a combination of statistical techniques and theoretical modeling to isolate the physical variability from the various observational artifacts.

3.2 Normalized Excess Variance as a measure of AGN variability

3.2.1 Excess Variance

The excess variance (σ^2_{XS}) is a statistical measure that can quantify the intrinsic variability of an AGN light curve. It is derived by measuring the spread of flux values around their mean, thereby providing a first-order estimate of the amplitude of variability. In an ideal scenario, where the light curve is noiseless, the excess variance is defined as:

$$\sigma^2_{\text{XS}} = \frac{1}{N}\sum_{i=1}^{N}(F_i - \langle F\rangle)^2, \tag{3.1}$$

where N is the number of observations, F_i is the measured flux at time t_i, and $\langle F\rangle$ is the mean flux over the light curve.

However, in real observations, the measured flux values are affected by instrumental and photon noise. The variability estimate must account for these observational uncertainties to avoid overestimating the intrinsic variability of the source. The commonly used estimator for excess variance is modified to subtract the contribution of measurement errors:

$$\sigma^2_{\text{XS}} = \frac{1}{N}\sum_{i=1}^{N}(F_i - \langle F\rangle)^2 - \sigma^2_{\text{err}}, \tag{3.2}$$

where σ^2_{err} represents the mean squared measurement error across all epochs. It should be noted that this correction assumes Gaussian-distributed errors that are identical for all measurements. In real X-ray observations, however, photon-counting noise often follows a Poisson distribution, meaning the uncertainty in each flux measurement is dependent on the flux itself.

The uncertainty in the excess variance is derived from the propagation of the respective errors and is given by:

$$\Delta\sigma_{\rm XS}^2 = \sqrt{\frac{2}{N}\left(\frac{\sigma_{\rm err}^2}{\langle F\rangle^2} + \frac{\sigma_{\rm XS}^2}{N}\right)}. \tag{3.3}$$

This formula provides an estimate of the statistical error associated with the excess variance, though additional systematic uncertainties may arise from effects such as background fluctuations or telescope calibration errors.

3.2.1.1 Normalized excess variance (NXSV)

The normalized excess variance (NXSV) is a dimensionless form of excess variance, allowing comparison across AGNs with different flux scales:

$$\text{NXSV} = \frac{\sigma_{\rm XS}^2}{\langle F\rangle^2}, \tag{3.4}$$

where the mean flux $\langle F\rangle$ normalizes the variance, ensuring that the variability measurement is independent of the absolute brightness of the source.

NXSV is particularly valuable in AGN variability studies due to its direct relationship with the PSD. Specifically, NXSV can be interpreted as the integral of the PSD over the observational timescale, meaning it captures the total variability power across all accessible frequencies [Vaughan et al., 2003, Paolillo et al., 2017, Bogensberger et al., 2024a]. This property makes NXSV a preferred statistic in cases where direct PSD estimation is challenging, such as when light curves are sparsely sampled or have uneven time coverage.

While NXSV provides a robust measure of the overall amplitude of flux variations, it does not preserve information about specific timescales of variability. Unlike regular PSD-based methods that allow for the identification of characteristic frequencies or breaks, NXSV condenses variability information into a single statistical measure. However, this simplification is often advantageous for large-scale AGN surveys, where individual PSD analysis may not be feasible due to the observational constraints.

The uncertainty in NXSV is propagated from the excess variance uncertainty as:

$$\Delta\text{NXSV} = \frac{\Delta\sigma_{\text{XS}}^2}{\langle F \rangle^2}. \tag{3.5}$$

By normalizing variability amplitude to flux, NXSV allows for meaningful comparisons across AGNs of different luminosities and redshifts.

At low signal-to-noise ratios, NXSV estimates may become negative due to noise over-subtraction. Error bars on NXSV are often calculated using Monte Carlo simulations that account for intrinsic variability and measurement uncertainties [Bogensberger et al., 2024a, Vaughan et al., 2003]. NXSV provides a simple and effective way to quantify the amplitude of variability while accounting for measurement uncertainties that is well-suited for datasets with limited time coverage and irregular sampling, making it particularly useful for X-ray variability studies in AGN surveys.

3.2.1.2 Ensemble NXSV

As mentioned before, individual light curves often suffer from sparse and uneven sampling, leading to unreliable estimates of NXSV. To address this, we introduce the concept of **ensemble NXSV**, which extends variability analysis from individual sources to entire AGN populations. Instead of analyzing each light curve in isolation, we treat all sources as stochastic manifestations of a common variability distribution produces by the same underlying processes. By compiling data across multiple sources, this method attempts to mitigate the effects of sparse sampling, low count rates, and high noise, which can otherwise bias individual NXSV estimates [Paolillo et al., 2017, Bogensberger et al., 2024b].

3.2.2 Bayesian Methodologies in Estimating NXSV

As discussed in Section 3.2.1, traditional methods for estimating NXSV assume Gaussian-distributed, independent measurement errors, which is inconsistent with the Poisson-distributed nature of photon-counting noise in X-ray observations. This mismatch introduces biases, particularly in low-flux regimes where uncertainties in flux measurements are inherently flux-dependent. Additionally, with AGN light curves often suffering from sparse sampling and background fluctuations, the reliability of standard variability estimators is further limited.

Bayesian methodologies address these challenges by explicitly modeling Poisson noise and incorporating hierarchical inference to link single-source variability with ensemble-level distributions. Unlike conventional estimators that provide single-point measurements, Bayesian inference produces posterior distributions of variability parameters , allowing for more reliable uncertainty quantification. This approach is particularly advantageous for ensemble studies of AGN variability, where hierarchical modeling enables a statistically rigorous characterization of population-wide trends despite observational limitations.

3.2.2.1 Single-Source NXSV

To analyze variability at the level of individual AGNs, we model flux as a stochastic process where each measurement fluctuates around a mean value due to intrinsic variations. Specifically, AGN flux variations often exhibit asymmetric, long-tailed distributions , which are well captured by a log-normal distribution [Vaughan et al., 2003]. We therefore assume that the logarithm of the flux follows a normal distribution:

$$\log_{10} F_{X,i} \sim \mathcal{N}(\mu_{\mathrm{LGFlux}}, \sigma^2_{\mathrm{LGFlux}}). \tag{3.6}$$

where μ_{LGFlux} represents the mean logarithmic flux, and σ_{LGFlux} is the standard deviation describing the scatter in log-flux space. This formulation ensures that flux remains strictly positive while aligning with empirical studies of AGN variability.

The normalized excess variance (NXSV) is derived from the variance of the log-normal flux distribution:

$$\log \sigma^2_{\mathrm{NXSV}} = \sigma^2_{\mathrm{LGFlux}} \ln(10). \tag{3.7}$$

This follows from the standard variance transformation of a log-normal distribution:

$$\mathrm{Var}(F_X) = \left(e^{\sigma^2} - 1\right) e^{2\mu+\sigma^2}. \tag{3.8}$$

To estimate these variability parameters, we define the likelihood function for a single-source light curve. Given an observed set of photon counts N_i at times t_i, we assume that flux measurements $F_{X,i}$ follow a Poisson likelihood :

$$P(N_i|F_{X,i}, O_i) = \frac{e^{-\lambda_i}\lambda_i^{N_i}}{N_i!}, \tag{3.9}$$

where the expected photon count λ_i at each epoch is given by:

$$\lambda_i = 10^{\mu_{\text{LGFlux}}} \cdot t_i + B_i. \tag{3.10}$$

This formulation ensures a physically motivated connection between flux variability and observed count rates, incorporating exposure time t_i and background counts B_i. The full likelihood function for an AGN light curve is then:

$$L = \prod_{i=1}^{N} P(N_i|F_{X,i}, O_i) \cdot \mathcal{N}(\log_{10} F_{X,i}|\mu_{\text{LGFlux}}, \sigma^2_{\text{LGFlux}}). \tag{3.11}$$

where the index i indexes epochs in an AGN's light curve.

3.2.2.2 Hierarchical model for estimating Ensemble NXSV

In our ensemble modeling framework, we consider a collection of AGN light curves, each with an associated NXSV value. Instead of analyzing each source in isolation, we seek to infer population-level trends by assuming that the NXSV values of individual AGNs are realizations of the same physical processes and can be taken from a shared statistical distribution.

Observational studies indicate that NXSV values across AGN populations exhibit skewed distributions with long tails, making a log-normal distribution a suitable model for describing ensemble variability [Paolillo et al., 2017]. Thus, we assume that individual AGN NXSV values are drawn from such a log-normal parent distribution:

$$\sigma^2_{\text{NXSV},k} \sim \log\mathcal{N}(\mu_{\sigma^2_{\text{NXSV}}}, \sigma^2_{\sigma^2_{\text{NXSV}}}), \tag{3.12}$$

where $\mu_{\sigma^2_{\rm NXSV}}$ represents the mean log-variance of the AGN population, $\sigma_{\sigma^2_{\rm NXSV}}$ is the scatter in variability amplitudes across sources and k indexes AGNs in the sample.

This formulation assumes that AGNs exhibit a range of variability amplitudes, with some being more variable than others due to differences in the accretion flow and mass of the SMBH (see Chapter 1).

Using this to build on equation 3.11, we extend the model to incorporate an ensemble of AGNs, each following its own stochastic variability process. This leads to the hierarchical likelihood function:

$$L_{\rm ensemble} = \prod_{k=1}^{N_{\rm src}} L_{{\rm single},k} \cdot \log \mathcal{N}(\sigma^2_{{\rm NXSV},k} | \mu_{\sigma^2_{\rm NXSV}}, \sigma^2_{\sigma^2_{\rm NXSV}}), \tag{3.13}$$

where $L_{{\rm single},k}$ refers to the single-source likelihood for AGN k (see equation 3.11), and the final term enforces a log-normal prior on NXSV values. As mentioned before, The index k runs over the number of AGN sources, while i loops through the epochs of each light curve. This hierarchical formulation ensures that individual AGN variability estimates are constrained by ensemble-level priors, reducing statistical noise and improving inference.

The ensemble NXSV estimate is then given by:

$$\sigma^2_{\rm NXSV,ensemble} = 10^{\mu_{\sigma^2_{\rm NXSV}}}, \tag{3.14}$$

which corresponds to the median of the log-normal distribution $\log \mathcal{N}(\mu_{\sigma^2_{\rm NXSV}}, \sigma^2_{\sigma^2_{\rm NXSV}})$.

As we know, AGN X-ray variability follows a red-noise process, in which power decreases at higher frequencies, leading to longer variability timescales [Vaughan et al., 2003]. This results in a distribution of variability amplitudes that often resembles a χ^2 distribution with low degrees of freedom, seemingly producing an extended tail toward high-variance AGNs.

Because NXSV measures the amplitude of variability, its distribution is expected to be skewed rather than Gaussian. A log-normal prior provides a statistically sound approximation for this behavior, ensuring flexibility in capturing the

observed range of AGN variability amplitudes.

The uncertainty in NXSV estimation is quantified using posterior distributions of the hierarchical parameters of the model. Bayesian inference provides credible intervals for each estimate:

$$\Delta\sigma^2_{\text{NXSV,ensemble}} = 10^{\mu_{\sigma^2_{\text{NXSV}}}} \cdot \Delta\mu_{\sigma^2_{\text{NXSV}}}. \tag{3.15}$$

This formulation accounts for both measurement noise and intrinsic scatter in the population variability distribution.

It is also important to note how the single-source model can provide us with additional insights into individual AGN variability by more carefully examining AGN that seem like outliers or edge cases. While the ensemble model enables a broader statistical understanding of AGN variability across populations, the single-source model can even be used as supplementary to it when singular sources seem to effect the ensemble variability in interesting ways that require deeper investigation into their specific variability.

Chapter 4

Data Processing & Light Curve Analysis

4.1 Data Processing & Light curve extraction

4.1.1 Overview

While the XMM Serendipitous Source Catalogue does provide time-binned light curves for each observation, these products are often incompatible with our Bayesian methodology since they are derived from count rates rather than contain information on the discrete photon counts and expected background level at any given time interval. Additionally, we require tighter control and deeper insight into the data-reduction process and as such we opt to reprocess the raw observation data ourselves.

In the pursuit of analyzing variability over large populations of QSOs we sought to develop an automated pipeline to process data from XMM-Newton observations, reducing them and producing reliable light curves for each source and its respective background. This pipeline is built on top of the foundation set by the xmmpype pipeline described in [Georgakakis and Nandra, 2011] which in turn makes extensive use of the XMM-Newton Science Analysis System (SAS) for data analysis, as described below.

4.1.2 Scientific Data Analysis (SAS) System

The Science Analysis Software (SAS) is the official data reduction and analysis package for XMM-Newton observations. It is designed to process the **Observation Data Files** (ODF) which contain raw uncalibrated science events, instrument housekeeping files, radiation monitor files and spacecraft files. SAS consists of two main components: Reduction pipelines, which apply instrument calibrations and produce scientifically useful products such as calibrated event lists, source lists, and time series, and file manipulation tools, which allow users to extract spectra, light curves, and images from the aforementioned products. Definitions of SAS tasks and nomenclature will be provided in this chapter when needed for clarity.

4.1.3 Light curve extraction pipeline

4.1.3.1 The xmmpype pipeline

The *xmmpype* pipeline [Georgakakis and Nandra, 2011] streamlines XMM-Newton data processing by constructing event files[1], detecting sources, applying astrometric corrections, estimating fluxes, and identifying X-ray sources with optical counterparts. We will outline its integration into our work.

Starting off, we select a subset of our refined dataset of observations (as described in Chapter 1.4) depending on the physical characteristics of the QSOs we want to analyze. As explained in the above chapter, this final catalog contains a list of SDSS QSOs along with the corresponding XMM observations (taken from the 4XMM catalog) for all the X-ray sources that lie within 5 arcsec of each QSO.

These XMM observations are assigned a unique **observation identification number** (obsid), with each obsid processed separately for efficiency. The raw ODF corresponding to each selected obsid then undergo reduction via the *EPCHAIN* and *EMCHAIN* tasks within SAS, yielding calibrated event files. These tasks include:

- The creation of raw event lists and good time interval[2] (GTI) data for each CCD.

[1] Event files contain time-ordered X-ray photon detections with arrival time, energy, and detector position.

[2] GTIs are typically defined to exclude times affected by instrumental issues, Earth occultations, and other similar observational constraints.

- Identification of bad pixels,
- Calibration of the raw event lists by flagging trailing events, performing pattern recognition, gain and CTI (charge transfer inefficiency) corrections,
- Assignment of detector and sky coordinates.
- Event pattern filtering since X-ray photons interacting with the detector can deposit charge in one or multiple adjacent pixels. To ensure the highest-quality photon detections, only single- and double-pixel events are kept.

The cleaned event lists for each CCD are merged per EPIC detector to maintain consistency. Images and exposure maps in different filters are created for each EPIC instrument, followed by a combined coadded image of all three. These products, along with high-particle background images (Section 4.1.3.2), serve as the basis for source detection.

HEALPix serves as a spatial indexing scheme to delineate the observation's spatial footprint by defining an observation's field of view while excluding gaps and other non-observed regions. A MOC map is then generated from the defined spatial footprint.

4.1.3.2 Flaring background filtering

X-ray Observations are often affected by flaring particle backgrounds, which can introduce artificial variability. To mitigate this, a filtering process is applied that excludes high-background periods using a methodology similar to that of Nandra et al. [2007].

Sources are masked out using the *EWAVELET* task in SAS, which detects sources in the 0.5–8 keV band with a threshold of five times the local background root mean square (rms). This ensures that AGN variability does not influence the estimation of background fluctuations. The background light curve is then generated by binning events in the 0.2–12 keV energy range into 20-second intervals, providing a time-resolved measurement of background activity. The quiescent background level is estimated by determining the count rate at which the excess variance of the background light curve is minimized. Periods where the background count rate exceeds twice this level are excluded to remove transient flares. It is verified that this

methodology works well for XMM. However, if an entire observation exhibits an elevated background, it is discarded if the quiescent level exceeds 50 and 20 counts per second for PN and MOS detectors, respectively. Additionally, observations with less than 1 ks of remaining good-time intervals after filtering are excluded to ensure sufficient exposure for variability analysis. This filtering ensures that the final event file is free from significant background contamination while preserving as much clean data as possible.

4.1.3.3 Source detection & count extraction

Source detection follows a two-pass process. First, the *emldet* task in SAS is used at a low threshold level to identify source candidate positions. Then, photon counts are extracted at these candidate source positions within an aperture corresponding to the 70% encircled energy fraction (EEF). The expected background level and the mean exposure time within the same aperture are also estimated.

Aperture photometry is then employed to estimate the Poisson probability that a detected source is merely a background fluctuation. Only sources with a background probability $P_{\text{background}} < 4 \times 10^{-6}$ are retained.

Source significance is assessed using a likelihood-based detection algorithm applied to the processed event files. A source is considered statistically significant if its likelihood exceeds a given threshold:

$$L = -\sum_i \ln P(D_i|S), \tag{4.1}$$

where $P(D_i|S)$ represents the probability of the observed data D_i given a source model S. The detected sources are cross-matched with the MOC coverage to confirm that they lie within the observed footprint.

For each detected source, X-ray photon counts are extracted using an elliptical aperture corresponding to a specific EEF, set at 70% of the PSF. This choice balances signal collection and background contamination: a smaller aperture would miss a significant fraction of source photons, while a larger aperture would introduce excessive background noise. Since the PSF of the EPIC detectors is energy-dependent and varies with off-axis angle, the extraction radius is determined using precom-

puted calibration files that provide energy- and position-dependent PSF models. These calibration files ensure that flux extraction remains consistent across different energy bands and detector positions.

To estimate the corresponding background, an elliptical annular region surrounding each source is used. To prevent contamination from nearby sources, all detections within a 4-arcminute radius are masked out before extracting background counts. The annular region has an inner radius of 5 pixels and an outer radius of 15 pixels, ensuring that the background is measured from an area representative of the local conditions. The background counts are then scaled to match the source extraction aperture, and the net source flux is computed by subtracting the scaled background contribution from the total extracted counts.

4.1.3.4 Sensitivity maps

Sensitivity maps are then refined following the methodology of Georgakakis et al. [2008] by incorporating vignetting, flux estimation biases (e.g. Eddington bias[3]) and the fraction of spurious sources expected in any source catalogue. These maps quantify the probability of detecting a source at a given X-ray flux, providing a well-characterized survey sensitivity across the field of view.

To estimate the background distribution, counts in the vicinity of detected sources are masked using elliptical apertures corresponding to 80% of the EEF. The missing pixel values are reconstructed by sampling from the surrounding background, defined by elliptical annuli extending 5 pixels beyond the source region with a width of 15 pixels. The background map is then smoothed using a 20-pixel median filter to reduce small-scale fluctuations while preserving large-scale structure.

Using the smoothed background, the mean expected background counts are computed within the 70% EEF elliptical source detection region. In cases where multiple EPIC detectors or XMM pointings are combined, background counts are summed across all contributing images. This procedure is repeated across the field, yielding a two-dimensional sensitivity map that encodes detection limits as a func-

[3]A selection effect in astronomical surveys where measurement errors cause rarer, more extreme objects to be overrepresented. Since common objects mistaken for rare ones significantly inflate the rare count, while the reverse has a negligible effect, the resulting sample is skewed—especially in cases like stellar brightness, where a small fraction of dim stars misclassified as bright leads to a disproportionate overestimate of bright stars.

tion of position.

To summarize the detection limits, a sensitivity curve is constructed following a Bayesian approach of Georgakakis et al. [2008]. This curve represents the total detector area over which a source of a given flux can be detected. A standard, non-Bayesian approach is also used for comparison, where a single limiting flux is assigned to each detection cell based on a fixed net count threshold. The difference between these approaches is most pronounced at low fluxes, where background noise significantly affects detection probabilities.

Sensitivity at a given flux is determined by exposure time, effective area, and background level and is given by:

$$S = \frac{N_\gamma}{A_{\text{eff}}(E) \times t_{\text{exp}}}, \tag{4.2}$$

where N is the photon count, $A_{eff}(E)$ is the energy-dependent effective area, and t_{exp} is the exposure time.

4.1.3.5 Source regions & masking

Having established the positions, extraction radii, and count rates for each detected source in our observation, the next step is to refine the optimal extraction radius for each source. This is done using an empirical relation which relates the optimal source radius to the source count rate (S. Mateos, private communication):

$$r_{\text{min}} = 18.0 \log_{10}(\text{countrate}) + 65.0 \tag{4.3}$$

This relation ensures that the aperture size is optimized for signal-to-noise ratio, adapting to the brightness of each source. A minimum radius of 10 arcseconds is enforced to prevent excessively small apertures. Using these refined radii, we construct a source mask for the detector, where all detected sources are systematically removed. This mask effectively eliminates contamination from detected sources, providing a source-free background region.

With the source mask established, we cross-match the detected sources in the current obsid with our catalog of QSOs in order to identify which objects of inter-

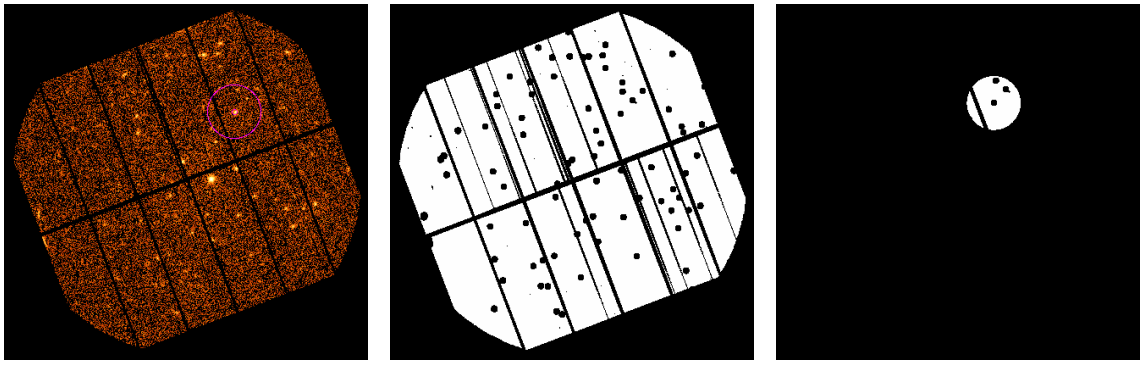

Figure 4.1: Example of an EPIC-pn observation image with counts in log scale. **Left:** The light curve extraction regions in purple, with the inner radius defining the source extraction region and the whole annulus defining the background region. **Middle:** The observation mask with all sources masked out. **Right:** The background light curve extraction region with all sources masked out.

est lie within our observation and extract their corresponding detector coordinates and optimized extraction radii. For each matched quasar, a region file is generated, defining both the source extraction region and its associated background region. The background region is defined as an annulus centered on the source, with an inner radius equal to the source extraction radius and an outer radius approximately five detector pixels larger. It is significant to note, this annular background region also incorporates the source mask, ensuring that any pixels overlapping with other detected sources are excluded. As a result, the background region maintains an annular shape while avoiding contamination from nearby sources.

These two region definitions—the source region and the background annulus—form the basis for the extraction of our light curves (see Figure 4.1).

4.1.3.6 Light curve extraction

Having defined these regions, we analyze each QSO of interest in our observation by extracting and binning time-resolved photon counts from the event files.

For each QSO, the source light curve is obtained by selecting all photon events within the predefined circular source region. Using the SAS task *evselect*, we apply spatial and energy filtering to isolate photons within the Soft X-ray range (0.5-2 keV) using a binning of 1000 seconds for detected photons, leading to light curves of 30 to 140 time bins in accordance with the durations of our sample, ideal for providing

a direct measure of temporal variations in photon arrival rates.

We then extract a background light curve using the annular region surrounding each source with our detected sources removed allowing for local background estimation while avoiding unwanted contamination. The extracted background counts will serve as a reference for further correcting the source light curve, removing any spurious fluctuations caused by instrumental or environmental background variations.

Once the source and background light curves are obtained, we apply background and instrumental corrections using the SAS task *epiclccorr*. This correction process scales the background counts appropriately based on the relative areas of the source and background extraction regions, ensuring that background subtraction is properly normalized. Additionally, vignetting corrections are applied to compensate for the position-dependent sensitivity of the XMM detectors. The correction also accounts for exposure variations, ensuring that gaps in observational coverage do not introduce artificial fluctuations in the light curve. The light curve data contains a column of fractional exposure that represents the fraction of the time bin where valid data was collected. This accounts for observational gaps caused by instrumental effects, high background contamination, or telemetry dropouts. We filter out rows that have 0 or invalid fractional exposure and in further sections when ”correcting for fractional exposure” we refer to correcting time like so:

$$t_{\text{eff}} = t_{\text{bin}} \times \text{FRACEXP} \tag{4.4}$$

Our final output consists of light curves for each QSO in our observation of photon counts binned in 1000-second intervals, generated for each object in both its source and respective background regions, providing the necessary data for further variability analysis. This process from raw data to finalized light curves can be repeated automatically for any amount of obsids with no further input.

4.2 Light curve analysis

Having the light curve data in hand, our goal is now is to apply our methodology (Chapter 3.1) to model NXSV as a measure of the intrinsic variability of our objects.

This analysis can be performed on individual light curves or on an ensemble of sources, grouped in black hole mass ($\log M_{BH}$) and Eddington ratio $\log \lambda_{EDD}$) bins as measures of the QSOs' physical characteristics. We will focus mainly on the latter approach which allows us to study population-level trends in AGN variability to investigate if we can validate any type of correlation between the variability and these characteristics.

4.2.1 Stan

4.2.1.1 Stan as a tool for Bayesian inference

To model the variability of AGN light curves, we employ Stan, a probabilistic code designed for Bayesian statistical modeling and inference using Hamiltonian Monte Carlo (HMC) and its extension, the No-U-Turn Sampler (NUTS). These algorithms are particularly well-suited for high-dimensional parameter spaces and provide efficient exploration of posterior distributions. Stan's ability to perform full Bayesian inference with hierarchical models makes it an ideal choice for estimating the NXSV of our light curves. The hierarchical approach allows us to incorporate population-level priors while still being able to fit individual light curves, thus improving statistical robustness when dealing with low-count X-ray observations.

4.2.1.2 MCMC sampling

Bayesian inference in Stan relies on Markov Chain Monte Carlo (MCMC) sampling, where multiple chains explore the posterior distribution iteratively. The HMC algorithm in Stan employs Hamiltonian dynamics to propose new states in the parameter space by treating parameters as particles moving in a potential energy landscape. This leads to more efficient sampling by reducing the random-walk behavior inherent in simpler MCMC algorithms. The No-U-Turn Sampler (NUTS) further enhances efficiency by automatically tuning the number of steps in the HMC trajectory. This prevents the algorithm from either making excessively short or long moves (which reduce efficiency and increase computational need respectively). The result is an adaptive MCMC method that ensures better convergence while maintaining computational efficiency.

Some pertinent terminology regarding MCMC sampling that will be used later:

- **Warmup steps** (burn-in) are the initial iterations where the sampler adjusts to the posterior distribution, improving efficiency before collecting meaningful samples. These steps are discarded from the final results.
- **Sampling steps** follow the warmup phase and generate the actual posterior samples used for inference. The more sampling steps, the better the estimate of parameter distributions.
- **Chains** are the independent Markov chains running in parallel, each starting from a different point in parameter space. Multiple chains help diagnose convergence—if they all explore similar distributions, we can be more confident in the results.

4.2.2 Variability analysis

4.2.2.1 Data pre-processing

Before being passed to our models, the extracted light curves undergo some minor corrections to ensure accurate statistical modeling. These involve correcting for fractional exposure exposure using equation (4.4) as well as correcting the background light curve by normalizing it using the background-to-source ratio calculated during the creation of the corrected light curves in section 4.1.3.6, ensuring that background counts are scaled correctly to our source counts.

Then we estimate the mean count rate of each source, which serves as an informative prior for the Bayesian model. This estimation is performed using a Poisson likelihood approach. To determine the statistical significance of a detection, we compute the minimum required counts for a source to be considered above background fluctuations. This is estimated using the inverse survival function of the Poisson distribution:

$$N_{\text{min}} = \text{Poisson}^{-1}(p_{\text{false}}, C_{\text{bkg, total}}) \tag{4.5}$$

where p_{false} represents the false-positive threshold, chosen to correspond to a detec-

tion significance of 4e-6 which is approx 4.5σ. We then estimate the most probable count rate by maximizing the Poisson probability mass function:

$$P(C_{\text{total}}|C_{\text{expected}}) = \frac{C_{\text{expected}}^{C_{\text{total}}} e^{-C_{\text{expected}}}}{C_{\text{total}}!} \tag{4.6}$$

where C_{expected} represents the predicted total counts for a given mean rate. This estimated count rate will then be used as a prior in our models.

For both models we will use, we pass the data as individual epochs, meaning that each data point corresponds to a specific time bin rather than an entire light curve. The ensemble model additionally requires an index array to map each epoch to its corresponding AGN.

4.2.2.2 Single light curve analysis

For an individual AGN, we model its X-ray light curve using a log-normal count rate distribution. We pass to our model:

- *Observed photon counts* C_i
- *Effective exposure time* $t_{\text{eff},i}$
- *Background photon counts* B_i
- *A prior on the logarithmic mean count rate* $\log C_{\text{rate}}$, estimated using the methodology described above.

4.2.2.3 Ensemble light curve analysis

Similarly to the individual model, we pass the data as individual epochs which necessitates a means to account for which epochs belong to which AGN. To address this we introduce an index array that maps each epoch to its corresponding AGN. The ensemble model structure consists of:

1. **Epoch-Specific Parameters**:

- ▶ $\log C_{\text{rate},i}$: The log count rate at a given epoch, estimated using the methodology described above.
- ▶ σ_i: The variability amplitude for each epoch.
- ▶ NXSV_i: The normalized excess variance per epoch.

2. **AGN-Specific Parameters**:
   - ▶ $\log C_{\text{mean},s}$: The mean log count rate for AGN s.
   - ▶ σ_s: The overall variability amplitude of an AGN.
   - ▶ NXSV_s: The intrinsic excess variance of an AGN.
3. **Population-Level (Hierarchical) Parameters**:
   - ▶ $\mu_{\sigma^2_{\text{NXSV}}}$: The mean of the log-normal distribution governing NXSV across the population
   - ▶ $\sigma_{\sigma^2_{\text{NXSV}}}$: The scatter in the NXSV distribution

Once the models are defined, inference is performed using 4 parallel chains, each with 4000 warmup steps and 2000 sampling steps, yielding 8000 posterior samples per parameter. We compute **the posterior mean as the final estimated value for each parameter** with uncertainties derived directly from the posterior distribution spread.

Chapter 5

Results & Technical Analysis

5.1 Light curve analysis results

5.1.1 AGN light curves

In this chapter we will present the results of our analysis outlined in chapter 4. Starting off with a representative example of an extracted light curve extracted automatically from raw ODF data using our pipeline (see section 4.1.3). This specific light curve has been binned at 8 kilosecond time intervals (see Figure 5.1) in order to better visualize the amplitude variability while most of our analysis has been done using 1 kilosecond intervals.

5.1.2 Bayesian flux estimates validation

After generating the light curves for our QSOs, these are passed to the ensemble Bayesian model grouped in in logarithmic black hole mass (log M_{BH}) and Eddington ratio (log λ_{Edd}) bins defined to cover the intervals of 8 to 10 and -2 to 0 respectively, using a step of 0.5 dex for both parameters.

Before presenting the results for the median ensemble variability, we first demonstrate the reliability of our Bayesian model at the individual source level. To do this, we compare the inferred posterior count rates with the prior count rate

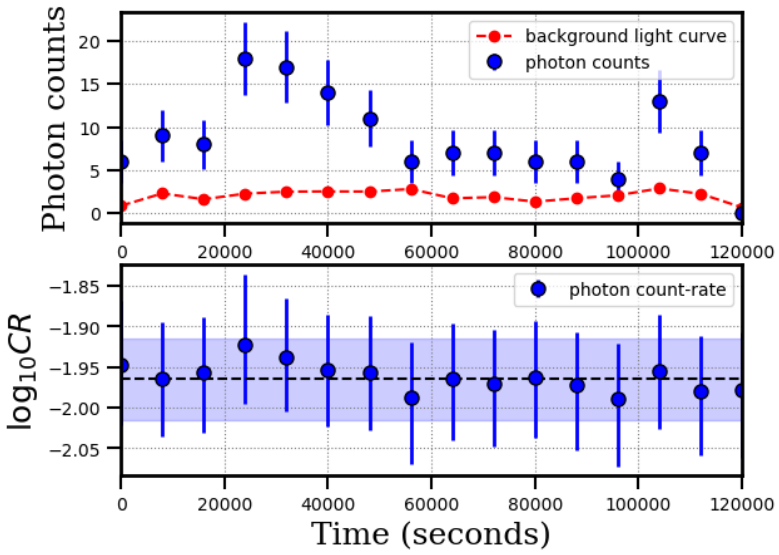

Figure 5.1: Example light curve of the SDSS DR16Q QSO 012635.58-012343.8 made with a time binning of 8 kiloseconds. The top panel plots the observed XMM photon counts (blue filled circles) as a function of the observation epoch in seconds. The errorbars associated with each point are Poisson uncertainties. The red circles connected with a dashed line correspond to the XMM background level. The bottom panel plots the estimated photon count rate of the same source as function of time in seconds. The single epoch rates are inferred from the Bayesian methodology described in section 3.1. The uncertainties correspond to the 68% confidence interval around the median of the posterior distributions of the single epoch photon count rates. The shaded region is the 1σ uncertainty of the mean count rate. The dashed black line shows the mean count rate of the light curve also inferred from the Bayesian approach of Section 3.1.

estimates, as defined in Equation 4.6. Figure 5.2 presents this comparison, with posterior means of inferred count rate on the y-axis versus the prior on the x-axis. Most objects lie near the one-to-one diagonal line, signifying that the posterior did not vastly deviate from the prior, possibly due to it being a good baseline. However, a small subset of sources exhibits deviations of 0.5–1 dex, marking them as potential outliers. The optical spectra of three of these consistent outliers can be seen in Figure 5.3. The spectra exhibit typical broad emission lines characteristic of mostly Type 1 AGN, without obvious signs of extreme absorption or strong outflows. This suggests that their deviation from the prior may not be due to dust obscuration or an unusually absorbed continuum in the optical but could instead arise from factors affecting their X-ray emission such as their increased intrinsic variability.

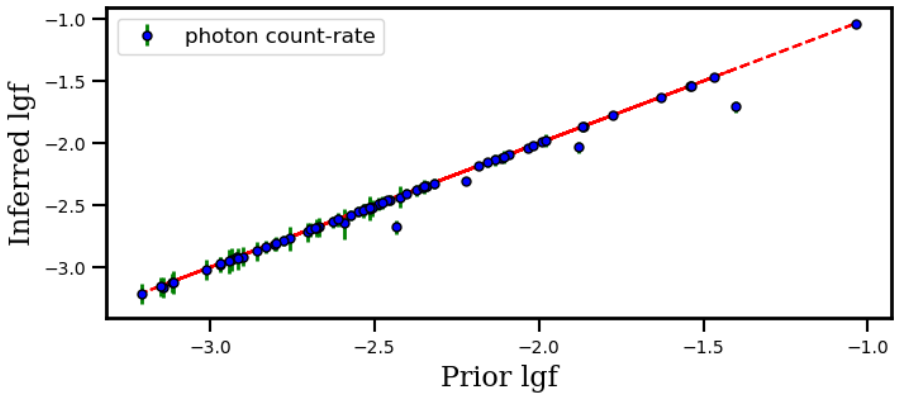

Figure 5.2: Comparison of prior given and model inferred Log Photon Count-Rates for 80 SDSS DR16 QSOs in the range $logM_{BH} = [8.0, 8.5]$ and Eddington ratios in the interval $log\lambda_{Edd} = [-2.0, -1.5]$. Each point corresponds to the mean log photon count-rate of an individual QSO, with the x axis showing the prior value and the y axis displaying the inferred value from the Bayesian ensemble model. Green error bars represent the 68% confidence interval of the inferred posterior distribution. The red dashed line indicates a one-to-one relation between prior and inferred values.

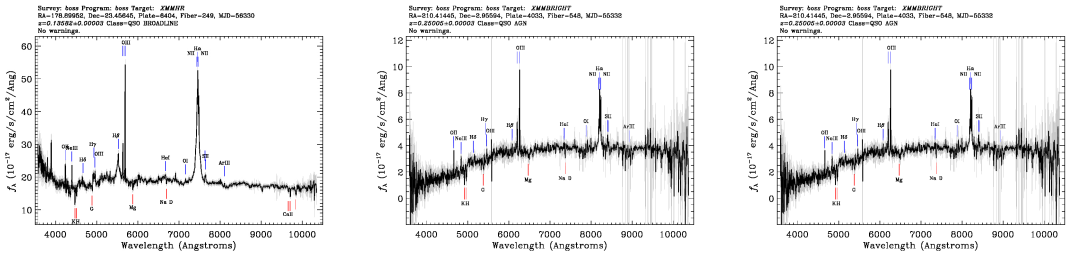

Figure 5.3: Optical spectra of outlier sources in Figure 5.2 found in SDSS DR16: **(Left)** SDSS J115535.88+232723.1, **(Middle)** SDSS J140139.46+025721.3, **(Right)** SDSS J140946.95+260732.7. Each panel plots flux versus observed wavelength, with common AGN emission lines visible.

5.1.3 Log-normal modelling of NXSV

One of the assumptions in our modeling is that the distribution of normalized excess variance ($\sigma_{\rm NXSV}$) follows a log-normal form. This assumption is motivated by the properties of stochastic variability in AGN accretion flows [e.g. Uttley et al., 2005, Kelly et al., 2009]. Additionally, as shown in Vaughan et al. [2003], for a finite number of light curve realizations of a red-noise process, the resulting variances follow a distribution that resembles a χ^2 distribution with a low number of effective degrees of freedom. This distribution features a high-variance tail, which can be approximated by a log-normal function linking the variances of individual sources to the same underlying population.

Figure 5.4 illustrates how our framework incorporates this assumption, showing the posterior draws of $\log_{10}(\sigma_{\rm NXSV})$ alongside the best-fitting log-normal model. The figure provides insight into how the prior structure works, particularly because both the mean and the standard deviation of the log-normal distribution are treated as free parameters in the model. Most of the posterior samples cluster around the log-normal peak, with relatively few sources populating the high-$\sigma_{\rm NXSV}$ tail. While there is some extension toward larger variances, the majority of these points remain within or just above the model's 68% confidence interval, indicating that the assumed log-normal form provides a reasonable approximation. However, as emphasized by Vaughan et al. [2003], the true population-level distribution of excess variance remains uncertain, and our use of a log-normal form serves primarily as a pragmatic choice for linking individual source variances to a shared statistical framework.

Overall, our best-fitting log-normal curve describes the distribution adequately. This result reinforces the idea that short-to-intermediate-timescale AGN flux variations arise from these physically multiplicative mechanisms, such as disc or corona instabilities, rather than purely additive noise. The generally good agreement also suggests that the observed scatter is intrinsic to AGN variability rather than being driven by systematic effects or outliers identified in the previous section.

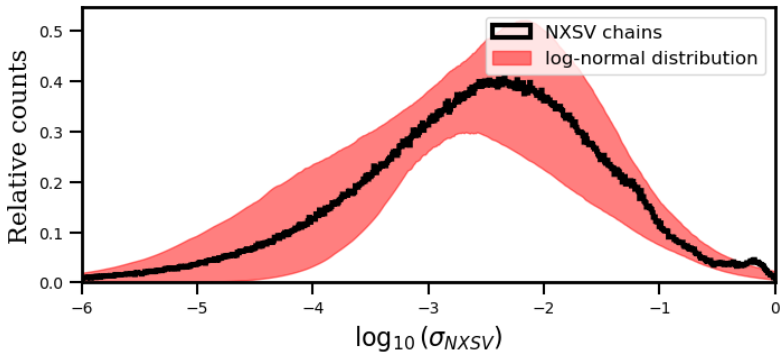

Figure 5.4: Inferred log-normal distribution of the normalised excess variance for the sample of SDSS DRQ16 QSOs with black hole masses in the range $\log M_{BH} = [8.0, 8.5]$ and Eddington ratios in the interval $\log \lambda_{Edd} = [-0.5, 0.0]$. The shaded region shows the 68% confidence interval of the reconstructed log-normal distribution of the population using the posteriors of the model parameters $\mu_{\sigma^2_{\rm NXSV}}, \sigma_{\sigma^2_{\rm NXSV}}$ of Equation (2.11). The histogram is constructed from the inferred $\sigma^2_{{\rm NXSV},k}$ (see Equation (2.11)) of individual light curves.

5.2 NXSV correlation with physical parameters

5.2.1 Eddington ratio dependence

The motivating factor of this thesis has been to assess how the variability amplitude changes with the physical characteristics of the accretion flow of an AGN, namely the **Eddington ratio** and **black hole mass**. Starting off with the former, in contrast to many earlier works suggesting an anti-correlation [e.g. Ponti et al., 2012], our findings show that $\sigma^2_{\rm NXSV}$ actually increases for higher $\log \lambda_{\rm Edd}$ bins, as seen in Figure 5.5. The lower-Eddington objects in our sample appear less variable on these (hours-to-days) timescales, while those approaching $\log \lambda_{\rm Edd} \sim -1$ display significantly larger intrinsic amplitude of flux fluctuations.

These results do, however, come with large errors in some of the lowest-$\lambda_{\rm Edd}$ bins, as evidenced in the right panel of Figure 5.5 stemming from often inconclusive posteriors offered by our model especially those with lower $\lambda_{\rm Edd}$. Hence, part of the apparent discrepancy with prior work could be due to limited statistics and relatively large uncertainties for luminous, fast-accreting AGN. Still, the net effect is suggestive that near-Eddington sources may be prone to additional variability driving mechanisms. Further work on the dataset or on deeper exposures may clarify

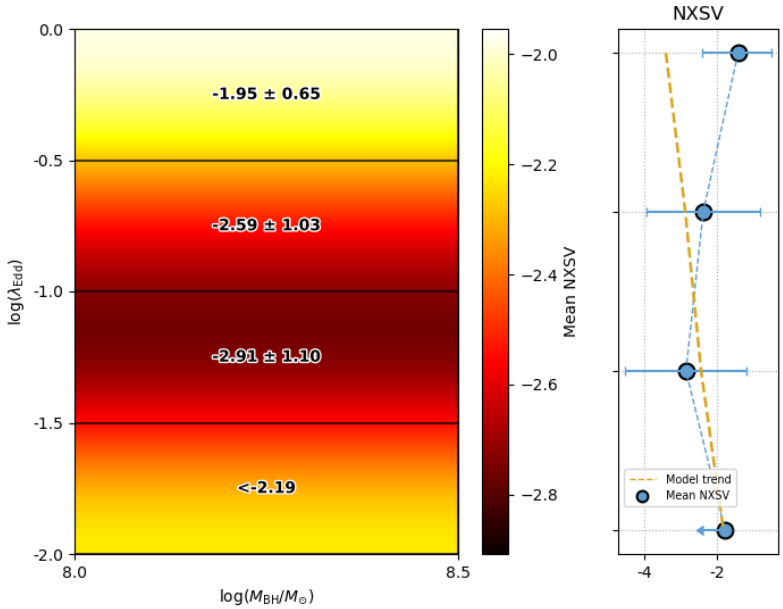

Figure 5.5: Ensemble normalized excess variance of DRQ16 QSOs as a function of Eddington ratio. The vertical and horizontal solid lines in the main panel show how the parameter space is divided in black hole mass and Eddington ratio bins with logarithmic widths of 0.5 dex. The value at the middle of each box is the estimated ensemble normalized excess variance for DRQ16 QSOs within the corresponding black hole mass and Eddington ratio limits. The shading is a smoothed representation of the mean NXSV variations on the 1-dimensional space of stable $\log M_{BH}$ over a varying $log\lambda_{Edd}$ with brighter colours corresponding to higher values. he right panel shows how the mean NXSV varies with Eddington ratio at a fixed black hole mass with its respective error stemming from the 68% confidence interval of the distribution. The value in the $\log \lambda_{Edd} = [-2, -1.5]$ bin is an upper limit, computed such that 99.6% of the posterior distribution falls below it. The yellow trendline showcases a qualitative empirical scaling relation for our data as described in appendix A.

whether this observed increase truly reflects coronal/wind physics or is partially driven by the uncertainties at lower $\lambda_{\rm Edd}$.

5.2.2 Black hole mass dependence

We also explore how $\sigma^2_{\rm NXSV}$ varies as a function of black hole mass. Figure 5.6 similarly presents a two-dimensional map of ensemble-averaged variability amplitudes across bins of $\log M_{\rm BH}$. Although some bins hint that more massive black holes may have somewhat smaller amplitudes, we do not observe a smooth or clearly monotonic trend.

A variety of factors could mask any underlying correlation. Single-epoch mass determinations carry uncertainties of up to 0.5 dex [Netzer, 2013], meaning that different true masses can easily end up in the same nominal bin, masking a weak trend. Additionally, red-noise leakage becomes significant if our observations (spanning days at most) fail to probe the longer characteristic timescales expected for more massive systems [McHardy et al., 2006]. This can artificially lower the measured $\sigma^2_{\rm NXSV}$ for high-$M_{\rm BH}$ objects or otherwise add scatter. Finally, comparing the size of the error bars across these bins reveals that our sensitivity to shallow relationships is limited: the scatter and uncertainties in each bin frequently match or exceed any incremental shift from one mass range to another.

It is also rather noteworthy that the number of QSOs populating each bin in our analysis, roughly on the order of around 100 sources, contributes to, and even possibly dominates, the relatively large error bars we see. Tests of our Bayesian model on samples of increasing size heavily suggest that variance estimates become more reliable as additional sources are incorporated. Consequently, further studies that expand the QSO sample in these particular intervals of $\log M_{\rm BH}$ and $\log \lambda_{\rm Edd}$ will likely reduce the statistical scatter and allow a clearer detection of any shallow trend.

To examine whether a hidden trend could feasibly exist within the error scatter, we apply a variability-black hole linear scaling relation derived in Akylas et al. [2022] by combining $M_{\rm BH}$ estimates obtained from reverberation mapping and stellar velocity dispersion measurements for local Seyfert galaxies. to the $\log M_{\rm BH}$ values of our sources. It should be noted that this relation was derived for the 3–10 keV energy band, while our analysis is conducted in the 0.5–2 keV band. We apply the relation using the 10 ks timescale, as most of our sources have rest-frame durations clustered around this timescale as seen in Figure 2.7. Thus Figure 5.7 contrasts our binned mean NXSV values against the predicted scaling slope. We further investigate this by plotting a linear fit of our NXSV values separately from the expected trend of Akylas et al. [2022] in Figure 5.8. Notably, removing the first bin (which has the largest scatter and the most unreliable posteriors) results in a best-fit linear trend that aligns well with the expected relations.

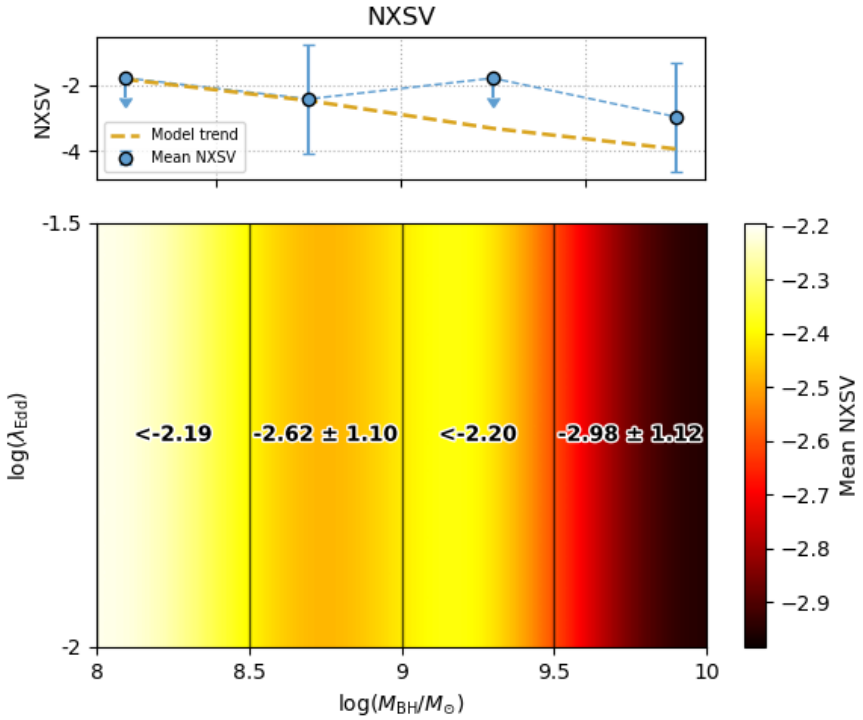

Figure 5.6: Ensemble normalized excess variance of DRQ16 QSOs as a function of Eddington ratio. The vertical and horizontal solid lines in the main panel show how the parameter space is divided in black hole mass and Eddington ratio bins with logarithmic widths of 0.5 dex. The value at the middle of each box is the estimated ensemble normalized excess variance for DRQ16 QSOs within the corresponding black hole mass and Eddington ratio limits. The shading is a smoothed representation of the mean NXSV variations on the 1-dimensional space of stable $\log \lambda_{Edd}$ over a varying $\log M_{BH}$ with brighter colours corresponding to higher values. The top panel shows how the mean NXSV varies with black hole mass at fixed Eddington ratio with its respective error stemming from the 68% confidence interval of the distribution. The values in the $\log M_{BH} = [8.0, 8.5], [[9, 9.5]]$ bins are upper limits, calculated such that 99.6% of the posterior distribution falls below them. The yellow trendline showcases a qualitative empirical scaling relation for our data as described in appendix A.

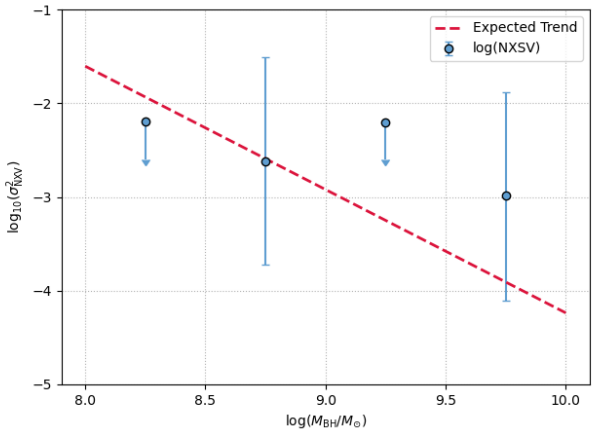

Figure 5.7: Mean NXSV values binned over $\log M_{\rm BH}/M_\odot$ bins compared against the expected scaling relation for a 10 ks timescale in the 3–10 keV energy band, as derived in Akylas et al. [2022]. The data points include error bars reflecting measurement uncertainties while the red dashed line represents the expected trend from the scaling relation. The values in the $\log M_{\rm BH}$ bins of $8.0-8.5$ and $9.0-9.5$ are upper limits, calculated such that 99.6% of the posterior distribution falls below them.

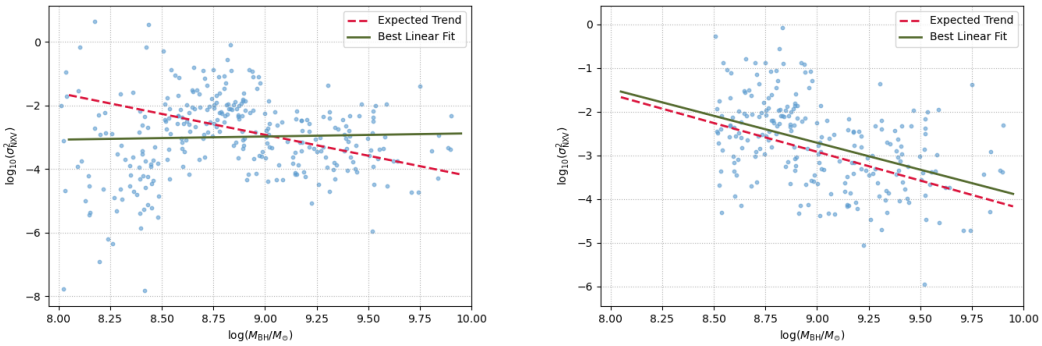

Figure 5.8: **(Left)** Logarithmic mean NXSV values for individual sources in our analysis plotted over $\log M_{\rm BH}/M_\odot$. The red dashed line represents the expected trend from the black hole mass-variability scaling relation for 10 ks timescales in the 3–10 keV energy band, as given in Akylas et al. [2022]. The green solid line indicates the best linear fit to the data. **(Right)** The same dataset but excluding the first bin $\log M_{BH} = [8.0, 8.5]$, where the highest scatter was observed.

Chapter 6

Discussion & Future work

In this thesis, we pursued two closely linked goals: first, to investigate the connection between X-ray variability in highly energetic AGN—quantified via the Normalized Excess Variance (σ^2_{NXV})—and the physical properties properties of the accretion flow such as the black hole mass (M_{BH}) and Eddington ratio (λ_{Edd}); and second, to design and validate an automated pipeline capable of processing raw *XMM-Newton* observations into scientifically usable light curves. This work is based on the recognition that X-ray variability on timescales of hours to days can offer unique insights into the accretion physics unfolding near supermassive black holes, yet extracting that signal reliably poses significant methodological and technical hurdles. By integrating source catalogs (e.g., the XMM-Newton 4XMM Catalog and the SDSS QSO Catalog), implementing systematic filtering strategies (to ensure data quality and adequate temporal coverage), programmatically processing the raw observations into workable light curves and applying a novel Bayesian methodology for variability estimation, we constructed a large and homogeneous dataset of QSO light curves along with estimates on their corresponding variability. Through this comprehensive approach we attempted to not only to probe potential correlations between variability amplitude, black hole mass, and Eddington ratio, but also to establish a standardized workflow for future AGN variability studies.

6.1 Discussion of results

Eddington ratio correlation: Our analysis indicates a tentative *positive* correlation between NXSV and $\lambda_{\rm Edd}$ (Figure 5.5), which contrasts with earlier studies that reported an anti-correlation Ponti et al. [2012]. Several factors may contribute to this discrepancy. At high accretion rates ($\log \lambda_{\rm Edd} \geq -1$), AGN could transition from radiatively efficient thin disks to advection-dominated slim disks Abramowicz et al. [1988], where enhanced turbulence or coronal heating might amplify flux variability. Such processes could obscure or counteract the usual expectation of suppressed fluctuations at higher accretion rates. Moreover, our sample selection differs from prior works that focused on nearby, low-luminosity AGN; by including a larger fraction of high-redshift QSOs with systematically higher $\lambda_{\rm Edd}$, we may be probing a different regime of black hole growth. Differences in rest-frame timescales and energy bands (e.g. soft vs. hard X-rays) also introduce observational biases, as do the evolving size and optical depth of the corona with increasing accretion rate. A patchier or more compact corona at high $\lambda_{\rm Edd}$ can plausibly yield strong short-term flux variability, further supporting a positive trend in NXSV.

Black hole mass correlation: No clear trend emerged between NXSV and $M_{\rm BH}$ (Figure 5.6), it seems the expected anti-correlation between NXSV and black hole mass is subtle enough (Figure 5.7) that it could be obscured within the uncertainties inherent in the observations. It is noteworthy that the trend in our results does loosely align with theoretical predictions of similar sources by Akylas et al. [2022] when excluding the lowest-mass bin that exhibits the strongest scatter of NXSV values (Figure 5.8). The scatter likely stems from several factors. It's possible that the observed timescales (hours to days) may not probe the characteristic variability timescales ($t_{\rm var} \propto M_{\rm BH}$) for high-mass AGN, leading to red-noise leakage and underestimated variability amplitudes. In addition to that, single-epoch $M_{\rm BH}$ estimates are known to carry significant errors (~ 0.5 dex; Netzer [2013]), blurring intrinsic correlations.

Feasibility of pipeline development and use: An important component of this work was the creation of an automated pipeline based on the already existing xmmpype pipeline to process raw XMM-Newton observations into workable light

curves. This pipeline integrates SAS routines for data reduction, source detection, and background filtering, streamlining the extraction of variability metrics. By automating background subtraction and exposure corrections, it minimizes systematic uncertainties, particularly for low-count sources. Its modular design ensures easy integration with xmmpype and usability across different studies. Beyond facilitating the present analysis, the pipeline demonstrates the feasibility of high-throughput AGN variability studies. Its success highlights the potential for future research to prioritize speed and statistical power, overcoming limitations such as small sample sizes that can hinder precise variability estimates by easily and en masse turning observations into light curves compatible with our methodology.

6.2 Future work

With the automated pipeline now demonstrated to scale across extensive *XMM-Newton* datasets, a natural next step is to apply the Bayesian ensemble methodology on significantly larger AGN samples. By pooling together sources within specific bins of black hole mass, Eddington ratio, or even more refined physical parameters, one can expect the statistical power to increase dramatically, thereby suppressing the noise that typically plagues regular single-object variability estimates. As the uncertainties on individual measurements shrink through the inclusion of more objects, this approach may reveal the shallow correlations that may remain hidden in smaller samples.

Another promising direction could lie in combining the pipeline's light-curve products with higher-resolution spectra from both *XMM-Newton* and complementary instruments (e.g. *NuSTAR*, *Chandra*). Such spectroscopic follow-up could characterize the coronal properties (e.g. temperature, optical depth) of the same AGN whose time-domain variability is being scrutinized. This joint analysis of spectral and temporal domains could reveal, for example, whether strong short-term variability correlates with specific coronal configurations or disk states, thereby offering a more holistic view of the accretion process. Finally, future studies might explore narrower-duration windows (e.g. restricting datasets to observations of similar rest-frame length) to avoid mixing variability on disparate timescales—a strategy that could be especially revealing for understanding why certain AGN, particu-

larly at high λ_{Edd}, appear to deviate from traditional variability–mass correlations. By merging these large-scale ensemble analyses with deeper spectral insights and carefully targeted timescale investigations, we can refine our picture of AGN variability and better link it to the fundamental physics governing black hole growth.

To aid in this endeavour, the upcoming large-scale spectroscopic surveys, **4MOST** and **DESI-II** are sure to provide transformative datasets. The 4-metre Multi-Object Spectroscopic Telescope (4MOST) Extragalactic Consortium surveys [de Jong et al., 2019] plan to deliver around 13 million low-resolution spectra for galaxies, AGNs, and QSOs, including detailed measurements of velocity dispersions, black hole masses, star formation rates, metallicities, and star formation histories. In addition to that, the 4MOST AGN Survey will monitor up to 1,000 AGNs and QSOs fortnightly, aiming to use these objects as cosmological standard candles. These datasets will significantly enrich the statistical analyses, correlating AGN variability to galaxy-scale environments, accretion physics, and large-scale cosmic structures. In parallel, DESI-II's Data Release 1 [DESI Collaboration et al., 2024] will feature detailed clustering analyses from tens of millions of galaxies and QSOs, characterized by precise two-point clustering statistics. These data promise significant advancements in cosmological analyses by providing precise measurements of cosmic structure and expansion, thus enhancing our ability to connect AGN variability with broader cosmological and structural contexts.

Appendix A

Empirical Scaling Relations

To interpret the ensemble trends in chapter 5, we compare them against an empirical model of AGN variability described as Model 3 in Georgakakis et al. [2024]. This model is constructed to reflect observed correlations between variability amplitude, black hole mass, and Eddington ratio, particularly the inverse correlation with $\lambda_{\rm Edd}$.

The model assumes that the AGN PSD follows a bending power-law form:

$$\mathrm{PSD}(\nu) = A\nu^{-1}\left(1+\frac{\nu}{\nu_b}\right)^{-1} \tag{A.1}$$

where ν is the temporal frequency, and the break frequency ν_b and normalization A are given by the scaling relations:

$$\nu_b = \frac{580}{M_{\rm BH}/M_\odot}\quad [\mathrm{s}^{-1}], \qquad A = 3\times10^{-2}\cdot\lambda_{\rm Edd}^{-0.8} \tag{A.2}$$

Thus, the break frequency is expected to decrease with increasing black hole mass, shifting the variability power to longer time-scales. The PSD amplitude decreases with Eddington ratio, leading to lower overall variability in more rapidly accreting systems.

In the paper, NXSV values are estimated by generating synthetic light curves from the PSD using the method of Emmanoulopoulos et al. [2013], and integrating

the power over the frequency range corresponding to observed rest-frame timescales (typically from a few days to ~ 10 years). The result is an ensemble prediction for the expected variability amplitude in bins of black hole mass or Eddington ratio.

This model does not attempt to match the exact NXSV values of individual sources. Instead, it captures the expected trends across the parameter space, assuming stochastic, red-noise dominated variability modulated by physical accretion properties.

Bibliography

M. A. Abramowicz, B. Czerny, J. P. Lasota, and E. Szuszkiewicz. Slim Accretion Disks. *The Astrophysical Journal*, 332:646, September 1988. ISSN 0004-637X. doi: 10.1086/166683.

A. Akylas, I. Papadakis, and A. Georgakakis. Black hole mass estimation using X-ray variability measurements in Seyfert galaxies. *Astronomy and Astrophysics*, 666: A127, October 2022. ISSN 0004-6361. doi: 10.1051/0004-6361/202244162.

Danielle Alloin, Rachel Johnson, and Paulina Lira, editors. *Physics of Active Galactic Nuclei at all Scales*, volume 693 of *Lecture Notes in Physics*. Springer Berlin Heidelberg, 2006. ISBN 978-3-540-31207-9. doi: 10.1007/b11550280. URL http://link.springer.com/10.1007/b11550280.

Robert Antonucci. Unified models for active galactic nuclei and quasars. *Annual Review of Astronomy and Astrophysics*, 31:473–521, January 1993. ISSN 0066-4146. doi: 10.1146/annurev.aa.31.090193.002353. URL https://ui.adsabs.harvard.edu/abs/1993ARA&A..31..473A. ADS Bibcode: 1993ARA&A..31..473A.

Umberto Baldini. *JET'S POWER IN RADIO-LOUD ACTIVE GALACTIC NUCLEI: A CASE STUDY ON THE NATURE OF BLAZAR CANDIDATES*. PhD thesis, September 2015.

Volker Beckmann and Chris R. Shrader. *Active Galactic Nuclei*. Wiley, August 2012.

D. Bogensberger, K. Nandra, M. Salvato, T. Liu, J. Wolf, S. Croom, H. Starck, J. Buchner, G. Ponti, J. Ider Chitham, C. Maitra, J. Robrade, A. Merloni, and M. Krumpe. Characterisation of the X-ray point source variability in the eROSITA south ecliptic pole field. *Astronomy and Astrophysics*, 687:A37, July 2024a. ISSN 0004-6361. doi: 10.1051/0004-6361/202449415. URL https://ui.adsabs.harvard.edu/abs/2024A&A...687A..37B. ADS Bibcode: 2024A&A...687A..37B.

David Bogensberger, Kirpal Nandra, and Johannes Buchner. Characterising X-ray variability in light curves with complex sampling patterns: application to the eROSITA south ecliptic pole survey. *A&A*, 687:A21, July 2024b. ISSN 0004-6361, 1432-0746. doi: 10.1051/0004-6361/202346368. URL http://arxiv.org/abs/2401.17278. arXiv:2401.17278 [astro-ph].

L. Burtscher, K. Meisenheimer, K. R. W. Tristram, W. Jaffe, S. F. Hönig, R. I. Davies, M. Kishimoto, J. U. Pott, H. Röttgering, M. Schartmann, G. Weigelt, and S. Wolf. A diversity of dusty AGN tori. Data release for the VLTI/MIDI AGN Large Program and first results for 23 galaxies. *Astronomy and Astrophysics*, 558:A149, October 2013. ISSN 0004-6361. doi: 10.1051/0004-6361/201321890.

Edward M. Cackett, Misty C. Bentz, and Erin Kara. Reverberation mapping of active galactic nuclei: from X-ray corona to dusty torus. *iScience*, 24:102557, June 2021. doi: 10.1016/j.isci.2021.102557. URL https://ui.adsabs.harvard.edu/abs/2021iSci...24j2557C. ADS Bibcode: 2021iSci...24j2557C.

Tapas K. Das and B. Czerny. Modelling the time-resolved quasi-periodic oscillations in active galactic nuclei. *Monthly Notices of the Royal Astronomical Society*, 414 (1):627–633, June 2011. ISSN 0035-8711. doi: 10.1111/j.1365-2966.2011.18427.x.

R. S. de Jong, O. Agertz, A. A. Berbel, J. Aird, D. A. Alexander, A. Amarsi, F. Anders, R. Andrae, B. Ansarinejad, W. Ansorge, P. Antilogus, H. Anwand-Heerwart, A. Arentsen, A. Arnadottir, M. Asplund, M. Auger, N. Azais, D. Baade, G. Baker, S. Baker, E. Balbinot, I. K. Baldry, M. Banerji, S. Barden, P. Barklem, E. Barthélémy-Mazot, C. Battistini, S. Bauer, C. P. M. Bell, O. Bellido-Tirado, S. Bellstedt, V. Belokurov, T. Bensby, M. Bergemann, J. M. Bestenlehner, R. Bielby, M. Bilicki, C. Blake, J. Bland-Hawthorn, C. Boeche, W. Boland, T. Boller, S. Bongard, A. Bongiorno, P. Bonifacio, D. Boudon, D. Brooks, M. J. I. Brown, R. Brown, M. Brüggen, J. Brynnel, J. Brzeski, T. Buchert, P. Buschkamp, E. Caffau, P. Caillier, J. Carrick, L. Casagrande, S. Case, A. Casey, I. Cesarini, G. Cescutti, D. Chapuis, C. Chiappini, M. Childress, N. Christlieb, R. Church, M. R. L. Cioni, M. Cluver, M. Colless, T. Collett, J. Comparat, A. Cooper, W. Couch, F. Courbin, S. Croom, D. Croton, E. Daguisé, G. Dalton, L. J. M. Davies, T. Davis, P. de Laverny, A. Deason, F. Dionies, K. Disseau, P. Doel, D. Döscher, S. P. Driver, T. Dwelly, D. Eckert, A. Edge, B. Edvardsson, D. E. Youssoufi, A. Elhaddad, H. Enke, G. Erfanianfar, T. Farrell, T. Fechner, C. Feiz, S. Feltzing, I. Ferreras, D. Feuerstein, D. Feuillet, A. Finoguenov, D. Ford, S. Fotopoulou, M. Fouesneau, C. Frenk, S. Frey, W. Gaessler, S. Geier,

N. Gentile Fusillo, O. Gerhard, T. Giannantonio, D. Giannone, B. Gibson, P. Gillingham, C. González-Fernández, E. Gonzalez-Solares, S. Gottloeber, A. Gould, E. K. Grebel, A. Gueguen, G. Guiglion, M. Haehnelt, T. Hahn, C. J. Hansen, H. Hartman, K. Hauptner, K. Hawkins, D. Haynes, R. Haynes, U. Heiter, A. Helmi, C. H. Aguayo, P. Hewett, S. Hinton, D. Hobbs, S. Hoenig, D. Hofman, I. Hook, J. Hopgood, A. Hopkins, A. Hourihane, L. Howes, C. Howlett, T. Huet, M. Irwin, O. Iwert, P. Jablonka, T. Jahn, K. Jahnke, A. Jarno, S. Jin, P. Jofre, D. Johl, D. Jones, H. Jönsson, C. Jordan, I. Karovicova, A. Khalatyan, A. Kelz, R. Kennicutt, D. King, F. Kitaura, J. Klar, U. Klauser, J. P. Kneib, A. Koch, S. Koposov, G. Kordopatis, A. Korn, J. Kosmalski, R. Kotak, M. Kovalev, K. Kreckel, Y. Kripak, M. Krumpe, K. Kuijken, A. Kunder, I. Kushniruk, M. I. Lam, G. Lamer, F. Laurent, J. Lawrence, M. Lehmitz, B. Lemasle, J. Lewis, B. Li, C. Lidman, K. Lind, J. Liske, J. L. Lizon, J. Loveday, H. G. Ludwig, R. M. McDermid, K. Maguire, V. Mainieri, S. Mali, and H. Mandel. 4MOST: Project overview and information for the First Call for Proposals. *The Messenger*, 175: 3–11, March 2019. ISSN 0722-6691. doi: 10.18727/0722-6691/5117.

DESI Collaboration, A. G. Adame, J. Aguilar, S. Ahlen, S. Alam, D. M. Alexander, M. Alvarez, O. Alves, A. Anand, U. Andrade, E. Armengaud, S. Avila, A. Aviles, H. Awan, S. Bailey, C. Baltay, A. Bault, J. Behera, S. BenZvi, F. Beutler, D. Bianchi, C. Blake, R. Blum, S. Brieden, A. Brodzeller, D. Brooks, Z. Brown, E. Buckley-Geer, E. Burtin, R. Calderon, R. Canning, A. Carnero Rosell, R. Cereskaite, J. L. Cervantes-Cota, S. Chabanier, E. Chaussidon, J. Chaves-Montero, S. Chen, X. Chen, T. Claybaugh, S. Cole, A. Cuceu, T. M. Davis, K. Dawson, A. de la Macorra, A. de Mattia, N. Deiosso, R. Demina, A. Dey, B. Dey, Z. Ding, P. Doel, J. Edelstein, S. Eftekharzadeh, D. J. Eisenstein, A. Elliott, P. Fagrelius, K. Fanning, S. Ferraro, J. Ereza, N. Findlay, B. Flaugher, A. Font-Ribera, D. Forero-Sánchez, J. E. Forero-Romero, C. S. Frenk, C. Garcia-Quintero, E. Gaztañaga, H. Gil-Marín, S. Gontcho A Gontcho, A. X. Gonzalez-Morales, V. Gonzalez-Perez, C. Gordon, D. Green, D. Gruen, R. Gsponer, G. Gutierrez, J. Guy, B. Hadzhiyska, C. Hahn, M. M. S Hanif, H. K. Herrera-Alcantar, K. Honscheid, J. Hou, C. Howlett, D. Huterer, V. Iršič, M. Ishak, S. Juneau, N. G. Karaçaylı, R. Kehoe, S. Kent, D. Kirkby, F. S. Kitaura, H. Kong, A. Kremin, A. Krolewski, Y. Lai, T. W. Lan, M. Landriau, D. Lang, J. Lasker, J. M. Le Goff, L. Le Guillou, A. Leauthaud, M. E. Levi, T. S. Li, K. Lodha, C. Magneville, M. Manera, D. Margala, P. Martini, M. Maus, P. McDonald, L. Medina-Varela, A. Meisner, J. Mena-Fernández, R. Miquel, J. Moon, S. Moore, J. Moustakas, N. Mudur, E. Mueller, A. Muñoz-Gutiérrez, A. D. My-

ers, S. Nadathur, L. Napolitano, R. Neveux, J. A. Newman, N. M. Nguyen, J. Nie, G. Niz, H. E. Noriega, N. Padmanabhan, E. Paillas, N. Palanque-Delabrouille, J. Pan, S. Penmetsa, W. J. Percival, M. M. Pieri, M. Pinon, C. Poppett, A. Porredon, F. Prada, A. Pérez-Fernández, I. Pérez-Ràfols, D. Rabinowitz, A. Raichoor, C. Ramírez-Pérez, S. Ramirez-Solano, M. Rashkovetskyi, C. Ravoux, M. Rezaie, J. Rich, A. Rocher, C. Rockosi, N. A. Roe, A. Rosado-Marin, A. J. Ross, G. Rossi, R. Ruggeri, V. Ruhlmann-Kleider, L. Samushia, E. Sanchez, C. Saulder, E. F. Schlafly, D. Schlegel, D. Scholte, M. Schubnell, H. Seo, R. Sharples, J. Silber, A. Slosar, A. Smith, D. Sprayberry, T. Tan, G. Tarlé, S. Trusov, R. Vaisakh, D. Valcin, F. Valdes, M. Vargas-Magaña, L. Verde, M. Walther, B. Wang, M. S. Wang, B. A. Weaver, N. Weaverdyck, R. H. Wechsler, D. H. Weinberg, M. White, M. J. Wilson, J. Yu, Y. Yu, S. Yuan, C. Yèche, E. A. Zaborowski, P. Zarrouk, H. Zhang, and C. Zhao. DESI 2024 II: Sample Definitions, Characteristics, and Two-point Clustering Statistics, November 2024.

B. Deufel, C. P. Dullemond, and H. C. Spruit. X-Ray spectra from accretion disks illuminated by protons. *Astronomy & Astrophysics*, 387(3):907–917, June 2002. ISSN 0004-6361, 1432-0746. doi: 10.1051/0004-6361:20020293.

D. Emmanoulopoulos, I. M. McHardy, and I. E. Papadakis. Generating artificial light curves: Revisited and updated. *Monthly Notices of the Royal Astronomical Society*, 433:907–927, August 2013. ISSN 0035-8711. doi: 10.1093/mnras/stt764.

A. C. Fabian. The obscured growth of massive black holes. *Monthly Notices of the Royal Astronomical Society*, 308:L39–L43, October 1999. ISSN 0035-8711. doi: 10.1046/j.1365-8711.1999.03017.x.

Pierre Fernique, Thomas Boch, Tom Donaldson, Daniel Durand, Wil O'Mullane, Martin Reinecke, and Mark Taylor. MOC - HEALPix Multi-Order Coverage map Version 1.0. *IVOA Recommendation 02 June 2014*, page 602, June 2014. doi: 10.5479/ADS/bib/2014ivoa.spec.0602F.

A. Georgakakis, K. Nandra, R. Yan, S. P. Willner, J. M. Lotz, C. M. Pierce, M. C. Cooper, E. S. Laird, D. C. Koo, P. Barmby, J. A. Newman, J. R. Primack, and A. L. Coil. The role of AGN in the colour transformation of galaxies at redshifts z$\approx$ 1. *Monthly Notices of the Royal Astronomical Society*, 385(4):2049–2060, April 2008. ISSN 0035-8711. doi: 10.1111/j.1365-2966.2008.12962.x.

A. Georgakakis, I. Papadakis, and M. Paolillo. Exploring black hole scaling relations via the ensemble variability of active galactic nuclei. *Monthly Notices of the Royal Astronomical Society*, 508:3463–3473, December 2021. ISSN 0035-8711. doi: 10.1093/mnras/stab2818. URL https://ui.adsabs.harvard.edu/abs/2021MNRAS.508.3463G. Publisher: OUP ADS Bibcode: 2021MNRAS.508.3463G.

A. Georgakakis, J. Buchner, A. Ruiz, T. Boller, A. Akylas, M. Paolillo, M. Salvato, A. Merloni, K. Nandra, and T. Dwelly. Ensemble X-ray variability of optically selected QSOs: Dependence on black hole mass and Eddington ratio. *Monthly Notices of the Royal Astronomical Society*, 531:4524–4537, July 2024. ISSN 0035-8711. doi: 10.1093/mnras/stae1383.

Antonis Georgakakis and K. Nandra. A serendipitous XMM survey of the SDSS: The evolution of the colour-magnitude diagram of X-ray AGN from z= 0.8 to 0.1. *Monthly Notices of the Royal Astronomical Society*, 414:992–1010, June 2011. ISSN 0035-8711. doi: 10.1111/j.1365-2966.2011.18387.x.

Shai Kaspi, Paul S. Smith, Hagai Netzer, Dan Maoz, Buell T. Jannuzi, and Uriel Giveon. Reverberation Measurements for 17 Quasars and the Size-Mass-Luminosity Relations in Active Galactic Nuclei. *The Astrophysical Journal*, 533: 631–649, April 2000. ISSN 0004-637X. doi: 10.1086/308704.

Brandon C. Kelly, Jill Bechtold, and Aneta Siemiginowska. ARE THE VARIATIONS IN QUASAR OPTICAL FLUX DRIVEN BY THERMAL FLUCTUATIONS? *The Astrophysical Journal*, 698(1):895, May 2009. ISSN 0004-637X. doi: 10.1088/0004-637X/698/1/895.

Andrew King. Black Holes, Galaxy Formation, and the MBH-σ Relation. *The Astrophysical Journal*, 596(1):L27, September 2003. ISSN 0004-637X. doi: 10.1086/379143.

M. Kishimoto, S. F. Hönig, T. Beckert, and G. Weigelt. The innermost region of AGN tori: Implications from the HST/NICMOS type 1 point sources and near-IR reverberation. *Astronomy and Astrophysics*, 476:713–721, December 2007. ISSN 0004-6361. doi: 10.1051/0004-6361:20077911.

S. Komossa. Tidal disruption of stars by supermassive black holes: Status of observations. *Journal of High Energy Astrophysics*, 7:148–157, September 2015. ISSN 2214-4048. doi: 10.1016/j.jheap.2015.04.006.

John Kormendy and Luis C. Ho. Coevolution (Or Not) of Supermassive Black Holes and Host Galaxies. *Annual Review of Astronomy and Astrophysics*, 51:511–653, August 2013. ISSN 0066-4146. doi: 10.1146/annurev-astro-082708-101811.

Brad W. Lyke, Alexandra N. Higley, J. N. McLane, Danielle P. Schurhammer, Adam D. Myers, Ashley J. Ross, Kyle Dawson, Solène Chabanier, Paul Martini, Nicolás G. Busca, Hélion du Mas des Bourboux, Mara Salvato, Alina Streblyanska, Pauline Zarrouk, Etienne Burtin, Scott F. Anderson, Julian Bautista, Dmitry Bizyaev, W. N. Brandt, Jonathan Brinkmann, Joel R. Brownstein, Johan Comparat, Paul Green, Axel de la Macorra, Andrea Muñoz Gutiérrez, Jiamin Hou, Jeffrey A. Newman, Nathalie Palanque-Delabrouille, Isabelle Pâris, Will J. Percival, Patrick Petitjean, James Rich, Graziano Rossi, Donald P. Schneider, Alexander Smith, M. Vivek, and Benjamin Alan Weaver. The Sloan Digital Sky Survey Quasar Catalog: Sixteenth Data Release. *The Astrophysical Journal Supplement Series*, 250:8, September 2020. ISSN 0067-0049. doi: 10.3847/1538-4365/aba623.

F. Marin, A. Marinucci, M. Laurenti, D. E. Kim, T. Barnouin, A. Di Marco, F. Ursini, S. Bianchi, S. Ravi, H. L. Marshall, G. Matt, C.-T. Chen, V. E. Gianolli, A. Ingram, R. Middei, W. P. Maksym, C. Panagiotou, J. Podgorny, S. Puccetti, A. Ratheesh, F. Tombesi, I. Agudo, L. A. Antonelli, M. Bachetti, L. Baldini, W. Baumgartner, R. Bellazzini, S. Bongiorno, R. Bonino, A. Brez, N. Bucciantini, F. Capitanio, S. Castellano, E. Cavazzuti, S. Ciprini, E. Costa, A. De Rosa, E. Del Monte, L. Di Gesu, N. Di Lalla, I. Donnarumma, V. Doroshenko, M. Dovčiak, S. Ehlert, T. Enoto, Y. Evangelista, S. Fabiani, R. Ferrazzoli, J. Garcia, S. Gunji, J. Heyl, W. Iwakiri, S. Jorstad, P. Kaaret, V. Karas, F. Kislat, T. Kitaguchi, J. Kolodziejczak, H. Krawczynski, F. La Monaca, L. Latronico, I. Liodakis, G. Madejski, S. Maldera, A. Manfreda, A. Marscher, F. Massaro, I. Mitsuishi, T. Mizuno, F. Muleri, M. Negro, S. Ng, S. O'Dell, N. Omodei, C. Oppedisano, A. Papitto, G. Pavlov, M. Perri, M. Pesce-Rollins, P.-O. Petrucci, M. Pilia, A. Possenti, J. Poutanen, B. Ramsey, J. Rankin, O. Roberts, R. Romani, C. Sgrò, P. Slane, P. Soffitta, G. Spandre, D. Swartz, T. Tamagawa, F. Tavecchio, R. Taverna, Y. Tawara, A. Tennant, N. Thomas, A. Trois, S. Tsygankov, R. Turolla, J. Vink, M. Weisskopf, K. Wu, F. Xie, and S. Zane. X-ray polarization measurement of the gold standard of radio-quiet active galactic nuclei: NGC 1068. *Astronomy & Astrophysics*, 689:A238, September 2024. ISSN 0004-6361, 1432-0746. doi: 10.1051/0004-6361/202449760.

I. M. McHardy, E. Koerding, C. Knigge, P. Uttley, and R. P. Fender. Active galactic

nuclei as scaled-up Galactic black holes. *Nature*, 444:730–732, December 2006. ISSN 0028-0836. doi: 10.1038/nature05389.

K. Nandra, A. Georgakakis, C. N. A. Willmer, M. C. Cooper, D. J. Croton, M. Davis, S. M. Faber, D. C. Koo, E. S. Laird, and J. A. Newman. AEGIS: The Color-Magnitude Relation for X-Ray-selected Active Galactic Nuclei. *The Astrophysical Journal*, 660 (1):L11, April 2007. ISSN 0004-637X. doi: 10.1086/517918.

Ramesh Narayan, Rohan Mahadevan, and Eliot Quataert. Advection-Dominated Accretion around Black Holes, March 1998.

Hagai Netzer. *THE PHYSICS AND EVOLUTION OF ACTIVE GALACTIC NUCLEI*. Cambridge University Press, 2013.

P. Padovani, D. M. Alexander, R. J. Assef, B. De Marco, P. Giommi, R. C. Hickox, G. T. Richards, V. Smolčić, E. Hatziminaoglou, V. Mainieri, and M. Salvato. Active galactic nuclei: what's in a name? *Astronomy and Astrophysics Review*, 25: 2, August 2017. ISSN 0935-4956. doi: 10.1007/s00159-017-0102-9. URL https://ui.adsabs.harvard.edu/abs/2017A&ARv..25....2P. Publisher: Springer ADS Bibcode: 2017A&ARv..25....2P.

M. Paolillo, I. Papadakis, W. N. Brandt, B. Luo, Y. Q. Xue, P. Tozzi, O. Shemmer, V. Allevato, F. E. Bauer, A. Comastri, R. Gilli, A. M. Koekemoer, T. Liu, C. Vignali, F. Vito, G. Yang, J. X. Wang, and X. C. Zheng. Tracing the accretion history of supermassive black holes through X-ray variability: results from the ChandraDeep Field-South. *Monthly Notices of the Royal Astronomical Society*, 471: 4398–4411, November 2017. ISSN 0035-8711. doi: 10.1093/mnras/stx1761. URL https://ui.adsabs.harvard.edu/abs/2017MNRAS.471.4398P. Publisher: OUP ADS Bibcode: 2017MNRAS.471.4398P.

B.M. Peterson. The Broad-Line Region in Active Galactic Nuclei. In Danielle Alloin, Rachel Johnson, and Paulina Lira, editors, *Physics of Active Galactic Nuclei at All Scales*, pages 77–100. Springer, Berlin, Heidelberg, 2006. ISBN 978-3-540-34621-0. doi: 10.1007/3-540-34621-X_3.

G. Ponti, I. Papadakis, S. Bianchi, M. Guainazzi, G. Matt, P. Uttley, and N. F. Bonilla. CAIXA: A catalogue of AGN in the XMM-Newton archive. III. Excess variance analysis. *Astronomy and Astrophysics*, 542:A83, June 2012. ISSN 0004-6361. doi: 10.1051/0004-6361/201118326.

Ricci. *AGN in the X-ray band.* PhD thesis, 2011.

N. I. Shakura and R. A. Sunyaev. Black holes in binary systems. Observational appearance. *Astronomy and Astrophysics*, 24:337–355, January 1973. ISSN 0004-6361.

Joseph Silk and Martin J. Rees. Quasars and galaxy formation. *Astronomy and Astrophysics*, 331:L1–L4, March 1998. ISSN 0004-6361. doi: 10.48550/arXiv.astro-ph/9801013.

Laura Silva, Gian Luigi Granato, Alessandro Bressan, and Luigi Danese. Modeling the Effects of Dust on Galactic Spectral Energy Distributions from the Ultraviolet to the Millimeter Band. *The Astrophysical Journal*, 509:103–117, December 1998. ISSN 0004-637X. doi: 10.1086/306476.

Rachel S. Somerville and Romeel Davé. Physical Models of Galaxy Formation in a Cosmological Framework. *Annual Review of Astronomy and Astrophysics*, 53:51–113, August 2015. ISSN 0066-4146. doi: 10.1146/annurev-astro-082812-140951.

L. Strüder, U. Briel, K. Dennerl, R. Hartmann, E. Kendziorra, N. Meidinger, E. Pfeffermann, C. Reppin, B. Aschenbach, W. Bornemann, H. Bräuninger, W. Burkert, M. Elender, M. Freyberg, F. Haberl, G. Hartner, F. Heuschmann, H. Hippmann, E. Kastelic, S. Kemmer, G. Kettenring, W. Kink, N. Krause, S. Müller, A. Oppitz, W. Pietsch, M. Popp, P. Predehl, A. Read, K. H. Stephan, D. Stötter, J. Trümper, P. Holl, J. Kemmer, H. Soltau, R. Stötter, U. Weber, U. Weichert, C. von Zanthier, D. Carathanassis, G. Lutz, R. H. Richter, P. Solc, H. Böttcher, M. Kuster, R. Staubert, A. Abbey, A. Holland, M. Turner, M. Balasini, G. F. Bignami, N. La Palombara, G. Villa, W. Buttler, F. Gianini, R. Lainé, D. Lumb, and P. Dhez. The European Photon Imaging Camera on XMM-Newton: The pn-CCD camera. *Astronomy and Astrophysics*, 365:L18–L26, January 2001. ISSN 0004-6361. doi: 10.1051/0004-6361:20000066.

Masahiro Suganuma, Yuzuru Yoshii, Yukiyasu Kobayashi, Takeo Minezaki, Keigo Enya, Hiroyuki Tomita, Tsutomu Aoki, Shintaro Koshida, and Bruce A. Peterson. Reverberation Measurements of the Inner Radius of the Dust Torus in Nearby Seyfert 1 Galaxies. *The Astrophysical Journal*, 639:46–63, March 2006. ISSN 0004-637X. doi: 10.1086/499326.

K. R. W. Tristram, K. Meisenheimer, W. Jaffe, M. Schartmann, H. W. Rix, Ch. Leinert,

S. Morel, M. Wittkowski, H. Röttgering, G. Perrin, B. Lopez, D. Raban, W. D. Cotton, U. Graser, F. Paresce, and Th. Henning. Resolving the complex structure of the dust torus in the active nucleus of the Circinus galaxy. *Astronomy and Astrophysics*, 474:837–850, November 2007. ISSN 0004-6361. doi: 10.1051/0004-6361:20078369.

P. Uttley, I. M. McHardy, and I. E. Papadakis. Measuring the broad-band power spectra of active galactic nuclei with RXTE. *Monthly Notices of the Royal Astronomical Society*, 332:231–250, May 2002. ISSN 0035-8711. doi: 10.1046/j.1365-8711.2002.05298.x.

P. Uttley, I. M. McHardy, and S. Vaughan. Non-linear X-ray variability in X-ray binaries and active galaxies. *Monthly Notices of the Royal Astronomical Society*, 359:345–362, May 2005. ISSN 0035-8711. doi: 10.1111/j.1365-2966.2005.08886.x.

S. Vaughan, R. Edelson, R. S. Warwick, and P. Uttley. On characterizing the variability properties of X-ray light curves from active galaxies. *Monthly Notices of the Royal Astronomical Society*, 345:1271–1284, November 2003. ISSN 0035-8711. doi: 10.1046/j.1365-2966.2003.07042.x. URL https://ui.adsabs.harvard.edu/abs/2003MNRAS.345.1271V. Publisher: OUP ADS Bibcode: 2003MNRAS.345.1271V.

S. Veronese, C. Vignali, P. Severgnini, G. A. Matzeu, and M. Cignoni. Interpreting the long-term variability of the changing-look AGN Mrk 1018. *Astronomy and Astrophysics*, 683:A131, March 2024. ISSN 0004-6361. doi: 10.1051/0004-6361/202348098.

N. A. Webb, M. Coriat, I. Traulsen, J. Ballet, C. Motch, F. J. Carrera, F. Koliopanos, J. Authier, I. de la Calle, M. T. Ceballos, E. Colomo, D. Chuard, M. Freyberg, T. Garcia, M. Kolehmainen, G. Lamer, D. Lin, P. Maggi, L. Michel, C. G. Page, M. J. Page, J. V. Perea-Calderon, F. X. Pineau, P. Rodriguez, S. R. Rosen, M. Santos Lleo, R. D. Saxton, A. Schwope, L. Tomás, M. G. Watson, and A. Zakardjian. The XMM-Newton serendipitous survey. IX. The fourth XMM-Newton serendipitous source catalogue. *Astronomy and Astrophysics*, 641:A136, September 2020. ISSN 0004-6361. doi: 10.1051/0004-6361/201937353.

XMM-Newton Science Operations Centre Team XMM-Newton Community Support Team. XMM-Newton Users Handbook. https://xmm-tools.cosmos.esa.int/external/xmm_user_support/documentation/uhb/index.html.