Universität Potsdam

Faculty of Science
Institut für Physik und Astronomie

Characterization of the structure of the AGN torus through time-resolved X-ray spectroscopy

Presented by:

**Silvia Carolina
Rueda Vargas**

Supervisors:

**PD Dr. Axel Schwope
Dr. Mirko Krumpe**

2024

A Melissa, mi hermana, a quien dedico cada uno de mis logros. Siempre presente, en mis recuerdos y en mi corazón, tus sueños y tu espíritu viven en mis triunfos.

Abstract

Currently, the properties and morphology of an AGN are explained by the Unified Model, which proposes that the central engine of the AGN is surrounded by a doughnut-shaped structure located a few parsecs away from the supermassive black hole located at the center. This structure is defined as the torus and according to the Unified Model, it is composed of homogeneously distributed gas and dust that provides obscuration to the X-ray source of the AGN. This homogeneous torus model has been challenged by infrared and X-ray studies in the last years, which opt for a clumpy torus where the material is distributed in clouds.

Depending on its hydrogen column density N_{H}, the torus clouds absorb part of the emission from the X-ray corona, mainly the soft photons, while transmitting the rest, corresponding to the hard energy photons. X-ray satellites like *XMM-Newton* detect the transmitted X-ray emission. The above implies that tracking the transmitted X-ray emission from an AGN over time provides insights into the material absorbing it, i.e. the clouds. A decrease in the amount of soft energy photons can be driven by a cloud entering the line of sight, while an increase in this amount can be caused by a cloud leaving the line of sight. The amount of soft and hard energy photons detected from an X-ray source as a function of time, i.e. its spectrum, can be used to search for such cloud events.

Following the above, during this master's thesis, I have reduced and analyzed data of ten AGNs observed with the *XMM-Newton* satellite through the torus, to increase the probability of observing a cloud event. Using these data, I have created two main pipelines: XMM-DH, which automatically carries out the initial processing of the data, time-resolves the observations by splitting them into equally sized time bins, and extracts one single X-ray spectrum per time bin. The second pipeline, XMM-fit, takes the previously extracted spectra per observation and fits them using models drawn from Laha et al. (2020). The fit allows an estimation of the spectral model parameters, such as the photon index and the hydrogen column density for each bin in an observation. By evaluating all results in the time domain, a time evolution curve is created.

The behavior of the N_{H} time evolution curve gives an insight into the material in the torus: a curve with discrete increases and decreases of N_{H} is the signature of an entirely clumpy torus, while a curve that remains constant and presents increases followed by decreases of N_{H} to the baseline value is the signature of a clumpy torus with homogeneous inter-cloud material. After analyzing the results for the ten sources in the sample, no secure cloud events have been identified. This result was expected, as the probability of catching one cloud event during an observation was estimated to be of the order of $10^{-2} - 10^{-4}$. Despite this, two of the sources in the sample present significant intra-observation changes of N_{H} that classify as candidates but not as secure cloud events. Additionally, three of the sources, namely NGC 7172, NGC 4258, and NGC 4507, present inter-observation changes, meaning that the average value of the column density presents a significant change between observations, proving that the material in the torus is not homogeneous.

Abstrakt

Derzeit werden die Eigenschaften und die Morphologie eines AGN durch das Einheitsmodell erklärt. Dieses besagt, dass der zentrale Motor des AGN von einer donutförmigen Struktur umgeben ist, die einige Parsec von dem supermassereichen Schwarzen Loch im Zentrum entfernt ist. Nach dem vereinheitlichten Modell besteht diese Struktur (Torus) aus homogen verteiltem Gas und Staub, der die Röntgenquelle des AGN verdeckt. Dieses homogene Torusmodell wurde in den letzten Jahren durch Infrarot- und Röntgenstudien in Frage gestellt, die auf einen klumpigen Torus hindeuten, in dem das Material in Wolken verteilt ist.

Durch die Quantifizierung der transmittierten Röntgenemission eines AGN, die von der Wasserstoffsäulendichte N_H des Materials in der Sichtlinie abhängt, können Rückschlüsse auf das absorbierende Material, d. h. die Wolken im Torus, gezogen werden. Genauer gesagt kann ein Wolkenereignis durch die zeitliche Veränderung der Menge an Photonen weicher und harter Energie bestimmt werden, die sich im Röntgenspektrum zeigt.

Im Rahmen dieser Masterarbeit habe ich daher Daten von zehn AGNs, die mit dem *XMM-Newton*-Teleskop durch den Torus beobachtet wurden, reduziert und analysiert, um die Wahrscheinlichkeit der Beobachtung eines Wolkenereignisses zu erhöhen. Mit diesen Daten habe ich zwei Hauptpipelines erstellt: XMM-DH, die automatisch die anfängliche Datenverarbeitung durchführt, die Beobachtungen zeitlich auflöst, indem sie in gleich große Zeitabschnitte unterteilt werden, und ein einzelnes Röntgenspektrum pro Zeitabschnitt extrahiert. Die zweite Pipeline, XMM-fit, nimmt die zuvor extrahierten Spektren pro Beobachtung und passt sie mit Modellen aus Laha et al. (2020) an. Der Fit ermöglicht eine Schätzung der Spektralmodellparameter, wie z.B. den Photonindex und die Wasserstoffsäulendichte für jeden Abschnitt einer Beobachtung. Durch die Auswertung aller Ergebnisse im Zeitbereich wird eine Zeitentwicklungskurve erstellt.

Das Verhalten der N_H-Zeitentwicklungskurve gibt Aufschluss über das Material im Torus: Eine Kurve mit diskreten Anstiegen und Abnahmen von N_H ist das Kennzeichen eines vollständig klumpigen Torus, während eine Kurve, die konstant bleibt und Anstiege gefolgt von Abnahmen von N_H auf den Ausgangswert zeigt, das Kennzeichen eines klumpigen Torus mit homogenen Zwischenwolkenmaterial ist. Nach der Analyse der Ergebnisse für die zehn Quellen in der Stichprobe wurden keine sicheren Wolkenereignisse identifiziert. Dieses Ergebnis war zu erwarten, da die Wahrscheinlichkeit, ein Wolkenereignis während einer Beobachtung zu erfassen, von der Größenordnung von $10^{-2} - 10^{-4}$ ist. Trotzdem zeigen zwei der Quellen in der Stichprobe signifikante Änderungen von N_H innerhalb einer Beobachtung, die als Kandidaten, jedoch nicht als sichere Wolkenereignisse klassifiziert werden. Drei der Quellen, nämlich NGC 7172, NGC 4258 und NGC 4507, zeigen Veränderungen zwischen den Beobachtungen, was bedeutet, dass der durchschnittliche Wert der Säulendichte zwischen den Beobachtungen eine signifikante Änderung aufweist und somit beweist, dass das Material im Torus nicht homogen ist.

Contents

References   111

Acknowledgements   116

Statement of originality   117

Chapter 1

Introduction

1.1 Active Galactic Nuclei

Galaxies contain a supermassive black hole (SMBH) in their center with mass ranging between $10^6 - 10^9 M_\odot$. This black hole is usually inactive, meaning that the interactions with its surroundings are limited to gravitational effects, thereby the energy output in galaxies is attributed only to thermal processes associated with stars and interstellar gas. In some galaxies, however, the central SMBH is active and accreting material from a disk surrounding it. This process causes such a highly energetic output in the nucleus that it surpasses that caused by the thermal radiation in the rest of the galaxy. Such a mechanism is called an Active Galactic Nucleus (AGN).

The AGN evolution is closely tied to the host galaxy properties (Trump et al., 2011). On the one hand, the rate at which the material is accreted onto the SMBH depends on the redshift, stellar mass, and star formation rate of the AGN host (Georgakakis et al., 2017). On the other hand, the mass of the SMBH correlates to the host galaxy's velocity dispersion (Ferrarese & Merritt, 2000) and star formation rate (Harrison, 2017), as well as to the AGN's luminosity and the radius of the last stable orbit for orbiting material (Griffin et al., 2019).

1.1.1 Zoo of AGN

Since their discovery and typification, multiple types of AGNs have been observed and classified based on their characteristics, such as radio emission, spectral lines, and the presence of jets. The classification of AGNs into types is known as the Zoo of AGNs and it is mentioned below:

Seyfert 1:

AGNs showing both narrow and broad features in their optical spectrum are classified as Seyfert 1 or Type I. Narrow lines include permitted transitions such as Hα, Hβ, HeI, HeII, MgII and CIV and forbidden such as OIII. Broad lines only include permitted transitions and have typical velocities > 1500 km $\cdot$ s^{-1}.

Seyfert 2:

Galaxies with AGNs showing only narrow permitted and forbidden emission lines in their optical spectra are known as Seyfert 2 or Type II.

Radio galaxies:

When an AGN presents a significant radio emission driven by jets, it is classified as a Radio galaxy. They might present both broad and narrow features, in which case they would be classified as Broad Line Radio Galaxies (BLRG); or only narrow lines, i.e. Narrow Line Radio Galaxies (NLRG).

Quasars:

These types of AGN are extremely luminous ($L > 10^{44}$ erg $\cdot$ s^{-1}) and are mostly located at high redshifts ($z \sim 2$).

Blazars:

This type of object shows considerable variability in the optical continuum during short time scales. Their spectra do not present significant absorption lines, possibly because, according to the unified model, they would be seen through the high-energy jet.

1.1.2 Unification Model of AGN

Despite being collectively referred to as AGNs, the multiple types of sources in the Zoo of AGN were initially considered different objects. This was until 1993 when Robert Antonucci proposed the Unification Model of AGN (Antonucci, 1993).

The Unification Model of AGN provides a simple explanation for the existence of multiple AGN types. According to this model, all AGNs share a common structure and are fundamentally the same object, with their type and observed differences determined by the viewing angle.

The AGN structure proposed by the Unified Model is a central engine composed of an SMBH and an accretion disk, surrounded by fast-moving clouds forming the Broad Line Region (BLR) and a toroidal structure over the equatorial plane called the torus. Additionally, slow-moving clouds form the Narrow Line Region (NLR) at higher distances from the core. A jet might be present along the polar plane when the AGN has a considerable radio emission. According to this model, the diversity of AGNs can be attributed to the torus, which conceals certain features when viewed from specific angles. Figure 1.1 depicts the AGN structure proposed by the Unified Model.

When an AGN is observed edge-on, the torus intersects the line of sight, obscuring the BLR and leaving only the clouds in the NLR visible. Consequently, the optical spectrum of the AGN presents only narrow features, classifying it as a Seyfert 2. Conversely, if the AGN is observed at a lower inclination angle, where the torus is outside the line of sight, both broad and narrow features are present in

the optical spectrum because both the BLR and NLR are visible. In this case, the AGN is classified as a Seyfert 1.

When an AGN has high radio emission and is observed face-on through the jet, its optical spectrum appears featureless. This classifies it as a Blazar: either an Optically Violent Variable AGN (OVV) or a BL Lac object, depending on its flux variability. If the radio galaxy is observed at a higher inclination angle, it is classified as a Broad-Line Radio Galaxy (BLRG). If the line of sight intersects the torus, the AGN is classified as a Narrow-Line Radio Galaxy (NLRG).
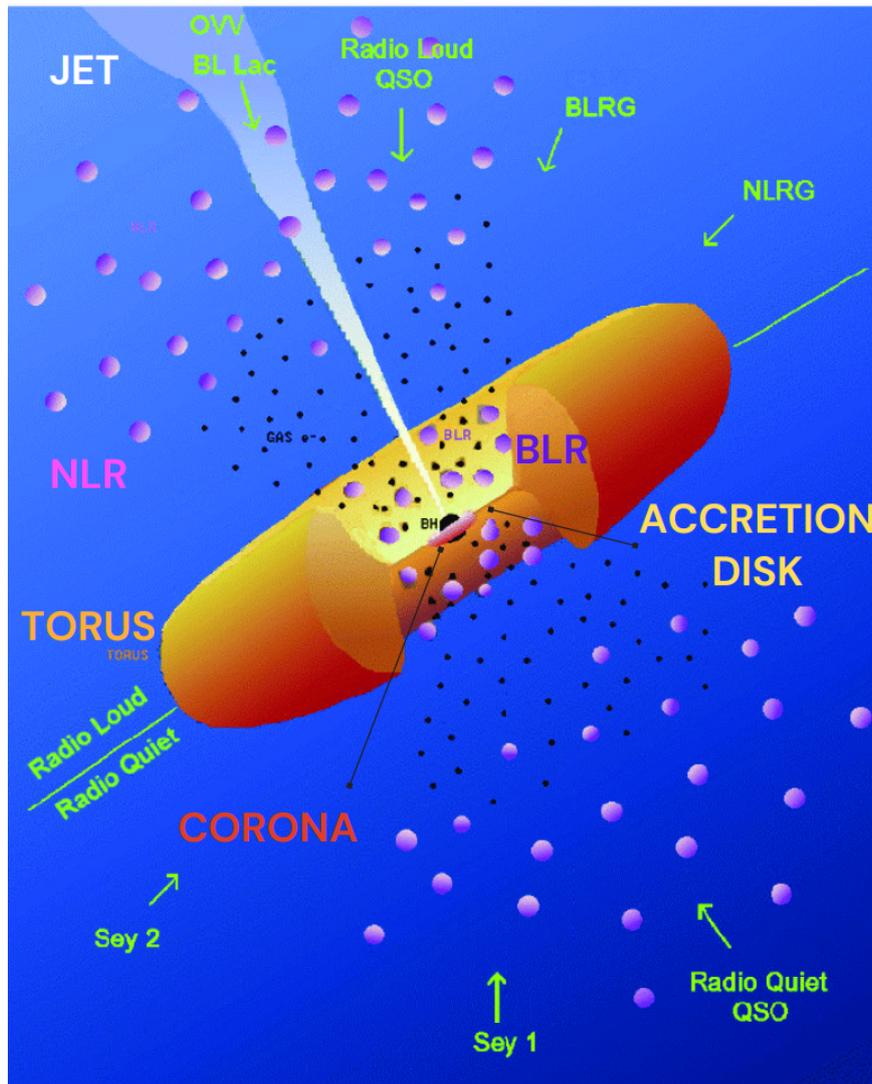

Figure 1.1: Unified model of AGN (Urry & Padovani, 1995). In the center, the supermassive black hole (BH, in black) is surrounded by the X-ray corona and the accretion disk. Two regions of clouds stand out: the Broad Line Region (BLR, in purple) close to the central engine, and the Narrow Line Region (NLR, in pink) further away from it. The dusty homogeneous torus obscuring the central engine is shown in orange, while the jet emerging from it is shown in white. The green names correspond to the type of AGN observed according to the viewing angle.

1.1.3 Structure of a classical AGN

Supermassive black hole

Galaxies with bulges are thought to contain a supermassive black hole (SMBH) at their center. Despite the exact formation mechanism of this type of object is not yet well understood, different theories constrain it (Volonteri, 2010; Kormendy & Ho, 2013).

- **Massive seed black hole** Population III massive stars (~ 260 M$_\odot$ on the main sequence) heat up so quickly that all the energy released by previous burning stages is used up, accelerating the collapse. The nuclear energy released by electron/positron pairs is, however, not enough to counteract the implosion, causing the star to become a black hole. This seed black hole accretes material at a high rate until reaching its supermassive size.

- **Collapse of supermassive objects** Primordial proto-galaxies have a high density and gas content. The lack of metals in these regions drives the efficient collection of material needed for the formation of a supermassive black hole, as opposed to enriched halos where the cooling mechanisms favor fragmentation and star formation.

- **Hierarchical buildup** Intermediate mass black holes (up to 10^5 M$_\odot$) may grow to be supermassive due to continuous mergers and accretion. These primordial black holes were likely to be formed in the early universe within regions where the large density fluctuations cause the gravitational force to overcome the pressure, driving the collapse.

Corona

The corona produces the X-ray emission detected from AGNs above $1 - 2$keV. This structure is composed of highly energetic electrons that up-scatter the optical and UV photons from the accretion disk via inverse Compton scattering, creating a power law spectrum. Although some of the properties of the X-ray corona are still not well understood, studies made with X-ray variability, quasar microlensing, and reverberation mapping have determined this to be compact, of the order of $3 - 10$ R_{g} (where $R_{\mathrm{g}} = 2GM_{\mathrm{BH}}/c^2$ corresponds to the gravitational radius of a black hole) and located around the SMBH (Kamraj et al., 2022).

Accretion disk

The central supermassive black hole of the AGN is surrounded by a disk of gas that extends for a few light days to weeks, depending mainly on the mass of the SMBH. As the disk rotates, the material falls into the SMBH and, according to the virial theorem for bounded systems in equilibrium, $E_{\mathrm{kin}} = -E_{\mathrm{pot}}/2$, half of its potential energy (E_{pot}) is converted to kinetic energy (E_{kin}). The remaining half is lost by mass outflow and radiation (Gaskell, 2007). As matter orbits the SMBH in the accretion disk, the material is heated up due to friction and turbulence and ionized by the gravitational energy and magnetic fields. This causes the emission of thermal

radiation, whose wavelength depends on the temperature of the disk. Regions close to the SMBH have a high temperature, therefore they emit high-energy radiation in the UV range (corresponding to a mean black body temperature of around 10^4 K), while optical radiation is emitted by regions further away at lower temperatures (around 10^3 K). Because of this, the thermal emission from the accretion disk is modeled as a multi-temperature blackbody (Remillard & McClintock, 2006), whose spectral energy distribution peaks in the UV.

The distance from the disk material to the black hole, as well as the accretion rate and mass of the SMBH, determine the optical and geometrical thickness of the disk and the type of rotation of the material.

For AGNs below the Eddington luminosity[1] (low accretion rate), the accretion disk is well fitted by the standard thin disk model (Shakura & Sunyaev, 1973): a geometrically thin and optically thick structure, with fully ionized material at temperatures between $10^5 - 10^7$ K, emitting multi-wavelength blackbody radiation generated mainly in the disk's midplane. In this case, radial pressure gradients are negligible and disks have a Keplerian rotation profile. Within this, angular momentum increases and angular velocity decreases with distance from the black hole, causing differential rotation. As a consequence, a large shear is created, the subsequent effects of which depend on the viscosity of the disk. When there is some viscosity, an outward transfer of angular momentum and generation of heat occurs, otherwise, the disk would rotate forever (Paczynski, 1982).

For AGNs above the Eddington limit (high accretion rate), however, the disk must be characterized by a more complex model called the slim model. In this one, the accretion disk is also optically thick, but the gas is assumed to be hotter, denser, and partially ionized. The conversion of energy to radiation in this case is very inefficient, due to the presence of advection: most of the photons are trapped in the disk, not being able to escape from the surface during the accretion process. The thermal energy is carried to the black hole with the material, rather than radiated away (Cackett et al., 2021; Liu & Qiao, 2022).

When the accretion rate is high, the radiation pressure increases, causing the disk to thicken to radiate away the heat generated. Consequently, the ratio of disk thickness to its radio increases too, and the radiation pressure becomes comparable to the gravitational acceleration in the radial direction. The rotation becomes then non-Keplerian (Paczynski, 1982).

The size of the accretion disk, as well as the location of its inner and outer edges, are highly dependent too on the accretion rate and the mass and spin of the black hole. The latter determines the location of the innermost stable circular orbit, as increased spin allows for more orbits closer to the SMBH. For high accretion rate AGNs, the material in the disk orbits the SMBH inwards up to this innermost stable circular orbit. Beyond that limit, it falls with a ballistic trajectory and is plunged into the central black hole. In contrast, at a low accretion rate, the flow of material is also low, causing very little energy to be radiated away and rather staying within the disk to heat up the material. As a consequence, the inner edge of the accretion disk is moved towards larger radii than the innermost stable circular orbit (Cackett et al., 2021).

---

[1]The Eddington luminosity defines the maximum luminosity an astronomical object can emit, governed by the equilibrium between radiation pressure and gravitational forces.

Broad line region

The Broad Line Region (BLR) is a system of discrete clouds located at short distances from the black hole, typically between 0.01 to 0.1 pc, but further away than the accretion disk. These clouds are predominantly composed of hydrogen and helium, although other elements such as carbon, nitrogen, and oxygen are also present.

When the UV/optical radiation from the accretion disk hits the BLR clouds, photons get photoelectrically absorbed by the atoms in the cloud, and electrons are kicked out from their orbitals. Electrons are then re-absorbed, ionizing the hydrogen atoms which emit radiation as a consequence of the change of energy.

The variation of the gravitational force experienced by each of the clouds causes a velocity dispersion. The effect of the different orbital velocities around the SMBH combined with the Doppler effect due to rotation, reflects on the width of the optical spectral lines observed from this region. When the optical and UV radiation from the accretion disk reaches clouds in the BLR moving in the line of sight towards the observer, the spectral lines are blueshifted (shifted to shorter wavelengths), while they are redshifted when the cloud is moving away. As a result of the combined effect, the line undergoes Doppler broadening, resulting in typical FWHM of 10^4 km s^{-1} (Whittle, 1985).

Narrow line region

The clouds in the Narrow Line Region (NLR) are located at distances further away from the supermassive black hole than the BLR, and their velocities, as a consequence, are lower (< 1000 km $\cdot$ s^{-1}). Due to this, the broadening of the spectral lines produced in this region is not significant, producing lines with typical FWHM of 1000 km $\cdot$ s^{-1} (Whittle, 1985).

Jet

The jet is a collimated stream of plasma, composed of highly energetic particles that travel at relativistic speeds and emit synchrotron radiation. They are present only in around 10% of all AGNs (e.g. Alexander & Hickox (2012)), specifically those with strong radio emissions, as a consequence of the black hole spin and the strong magnetization on the accretion disk. The jets are observed on scales from astronomical units to megaparsecs, emitting in the whole electromagnetic spectrum, from radio to gamma rays (e.g. Blandford et al. (2019)).

Torus

The central engine of an AGN (supermassive black hole and accretion disk) is surrounded by gas and dust, partly located in the BLR and partly in the torus. The torus is a doughnut-shaped structure whose inner boundary is defined by the sublimation temperature of the dust grains outside of the central engine and extends up to a few parsecs.

The cold material in the torus absorbs the incoming radiation from the accretion disk, heats up, and re-emits the radiation at longer wavelengths. As a consequence,

the UV and optical radiation is reprocessed as infrared, and the AGN SED is dominated by the torus from $\sim 1~\mu m$ up to a few tens of microns. The gas in the torus also absorbs the X-ray radiation emitted by the corona, thereby obscuring the central engine.

Although the exact structure of the torus is not well understood yet, multiple studies in IR and X-ray have constrained it to three main possible scenarios, sketched in Fig. 1.2.
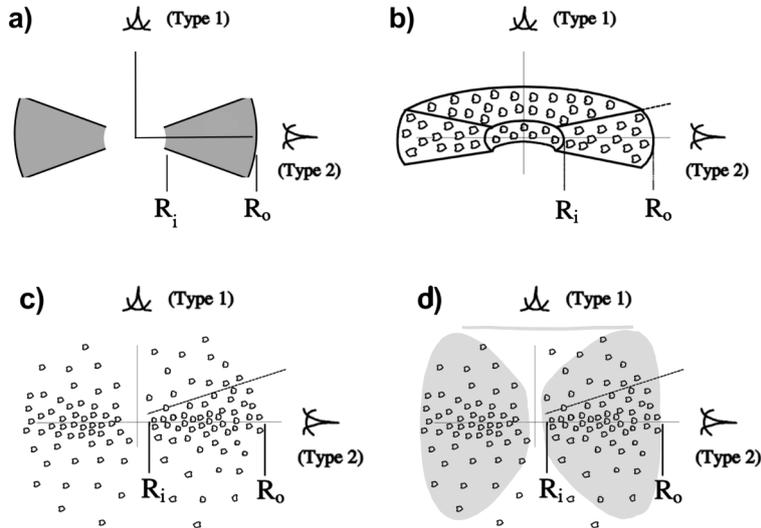

Figure 1.2: Torus models. a) Homogeneous torus, where the material is smoothly distributed over a well-defined doughnut-shaped structure in the equatorial plane of the AGN. b) Clumpy, well-defined edges torus, where the dust and gas form clumps distributed over a well-defined doughnut-shaped structure in the equatorial plane of the AGN. c) Clumpy torus, where the location of the clumps of material is not restricted to the well-defined doughnut-shaped structure. d) Clumpy torus + Inter-cloud medium, where the clumps of material in the torus are embedded in a low-density medium. In the figure, σ represents the angular width of the torus, R_i and R_o its inner and outer edge, respectively, and i the viewing angle of the AGN. Type I and Type II denote the AGN type observed according to its viewing angle. The equatorial plane of the AGN, where the BLR is located, corresponds to the horizontal plane from left to right. The polar plane, where the NLR is located, is perpendicular to the equatorial plane. Modified from Elitzur et al. (2003); Elitzur (2006).

- **Homogeneous torus**: This is the scenario proposed in the Unified Model to explain the observed differences in the optical spectra of different AGNs. According to this model, the fact that some spectra show only narrow lines, while others present a combination of both narrow and broad features, is due to a dense toroidal structure enclosing the central engine, which obscures the region where the broad lines are produced.

  The toroidal structure around the central engine has well-defined inner and outer edges R_i and R_o, respectively, leading to the classification of AGNs into Seyfert 1 (Type I or unobscured) and Seyfert 2 (Type II or obscured). A

representation of this model is shown in Fig. 1.2, panel a. If the AGN happens to be observed edge-on, at a viewing angle i intersecting the angular width σ of the torus, the line of sight crosses the torus and the broad line region is concealed. In that case, only the narrow features are visible in the optical spectrum, and the AGN is classified as a Seyfert 2. On the other hand, if the AGN is observed face-on, at an angle i not intersecting σ, both narrow and broad lines are observed in the optical spectrum. Thus, the AGN is classified as a Seyfert 1. For a comparison of the optical spectrum of Seyfert 1 and 2 AGN, see Fig. 1.3. In summary, in the Homogeneous Torus model, a Seyfert 1 AGN is observed when $0 \leq i < \sigma$ and a Seyfert 2 when $\sigma \leq i < 90°$.
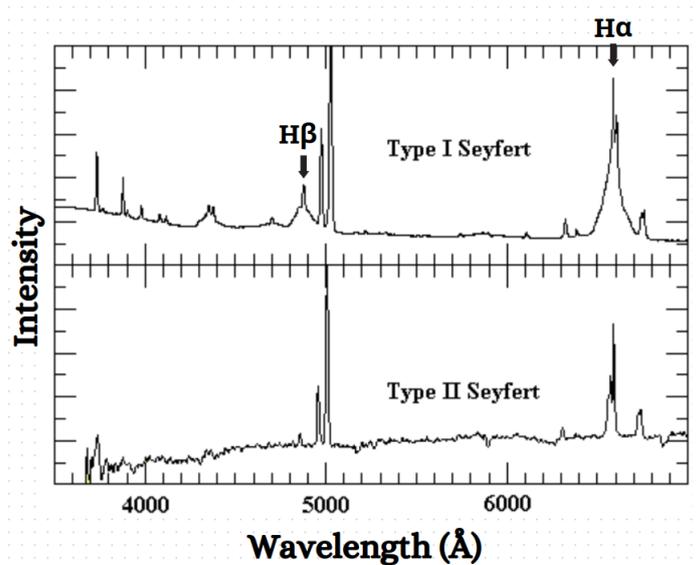
Figure 1.3: Comparison of the optical spectrum of Seyfert 1 (top) and Seyfert 2 (bottom) AGN. Emission lines such as Hα and Hβ are broader in the Seyfert 1 than in the Seyfert 2 spectrum. Taken and modified from Astronomy - Section 3: Distant and Weird Galaxies, https://sites.uni.edu/morgans/astro/course/Notes/section3/new13.html

Although the Unified Model explains multiple characteristics of the AGN structure, new models and infrared (IR) observations challenge its definition of a homogeneous torus, where the material is smoothly distributed. For example, when resolving the heated dust structure of NGC 1068 through mid-infrared interferometry, the temperatures obtained at given distances from the central engine do not correspond to a homogeneous torus. At a distance of around 1.7 pc, a temperature of $T = 320$ K is reported, however, given the bolometric luminosity of $L = 2 \times 10^{45}$ erg s^{-1} in this Seyfert 2 galaxy, a temperature of that order is just expected at a distance of around 26 pc (Jaffe et al., 2004). This inconsistency can be explained by introducing a clumpy structure in the torus of the AGN, as the face of the cloud being illuminated by the AGN would have a much higher temperature than the opposite, dark side. That way, at a distance of around 2 pc in NGC 1068, temperatures will range from 250 K to 950 K. Taking this into account, new models lean

towards a clumpy torus where dust and gas are distributed in discrete clouds. The main reason for this is that due to the high temperatures in the proximity of the SMBH, dust would not be able to exist otherwise (Krolik & Begelman, 1988).

- **Clumpy, well-defined edges torus**: According to this model, the AGN torus is a doughnut-shaped structure with well-defined edges, where the material is completely distributed in clumps (Fig. 1.2, panel b) (García-Bernete et al., 2022). In this model, the observation of a Type I or Type II AGN is, as in the Unified Model, dependent on the viewing angle.

  Although this clumpy, well-defined edges torus presents a solution for the mentioned inconsistency between the estimated dust temperature and its distance from the SMBH, IR observations challenge this model by comparing the emission from Seyfert 1 and Seyfert 2 AGNs. According to a well-defined edges torus model, the obscuration of the central engine is highly anisotropic, however, the infrared emission seems to be isotropic and identical among Seyfert 1 and Seyfert 2 AGNs. This challenge can be explained by introducing a clumpy torus that is not limited to the equatorial plane of the AGN.

- **Clumpy torus**: This model proposes an AGN torus where the material is completely distributed in clumps forming optically thick clouds (Fig. 1.2, panel c). The distribution of these clouds is not limited to a doughnut-shaped region with well-defined edges over the equatorial plane of the AGN, but it is rather described as:

$$N_c(r, \beta) \propto \mathcal{N}_0 e^{(-\beta/\sigma)^2} r^{-q} \tag{1.1}$$

  where $\mathcal{N}_0$ is the average number of clouds encountered along a radial equatorial ray typically between 2 to 15 clouds (Jaffe et al., 2004; Nenkova et al., 2008; Markowitz et al., 2014), r is the radial distance from the center, and β is the viewing angle. The angular distribution of clouds in the torus is parametrized as a Gaussian with width parameter σ, while their radial distribution as an inverse power law with index q (Elitzur, 2006; Almeida et al., 2009).

  The cloud distribution proposed by this clumpy torus model has implications for the AGN obscuration and the AGN type definition. On the one hand, the AGN obscuration depends on the number of clouds along the line of sight, and it is highly anisotropic due to the Gaussian distribution, which increases with the viewing angle β. The SED, on the other hand, does not present significant changes with the viewing angle and becomes highly isotropic for $q = 2$. Therefore, this model solves the anisotropic AGN obscuration with the viewing angle while maintaining the IR emission isotropic (Elitzur et al., 2003).

  On the other hand, as the edges of the torus are not sharp, the probability of finding at least one cloud causing the obscuration of the central engine in the line of sight (eclipse event), is non-zero, regardless of the viewing angle. As a consequence, observing a Type I or Type II AGN following the clumpy model does not depend on the viewing angle, but rather on the probability of

encountering a cloud in the line of sight. This probability depends on the size and location of the clouds. For eclipse events lasting between 0.2 days and 16 years, it has been estimated as $0.006^{+0.160}_{-0.003}$ for sources observed face-on, and $0.110^{+0.461}_{-0.071}$ for sources observed edge-on (Markowitz et al., 2014).

Following the above, there is a finite probability of observing the AGN through an unperturbed line of sight, even over the equatorial plane. In such a case, when the source is observed edge-on, the central engine would not obscured by clouds, and the AGN would be considered a Type I. In contrast, when following the Homogeneous torus model, this same source would always be considered as Type II, as it is observed edge-on (Elitzur et al., 2003).

In addition to the issues that a clumpy structure of the torus naturally solves, it also provides an explanation for one of the types of changing look AGN: the changing-obscuration ones. These types of objects show flux variability over different timescales and the whole electromagnetic spectrum, more specifically showing strong variability in the line of sight column density $\mathcal{N}_H$. In the X-ray, this phenomenon can be attributed to clouds entering and leaving the line of sight, causing changes in the obscuration as the central engine is covered and uncovered (Ricci & Trakhtenbrot, 2022).

- **Clumpy torus + Inter-cloud medium**: According to this model, the AGN torus consists of a multi-phase structure where high-density clumps of material are embedded within a low-density medium (Stalevski et al., 2012; Siebenmorgen et al., 2015). A sketch of this model is shown in Fig. 1.2, panel d. The characteristics of this model are the same as the clumpy model mentioned above, with the main difference being that in this case, the line of sight is constantly intersected by material from the torus, regardless of the viewing angle. As a result, the central engine of the AGN is perpetually obscured to a baseline level due to the inter-cloud medium, and any additional clouds entering the line of sight further add to this baseline level of obscuration.

Torus models proposing a smooth distribution of material in a well-defined toroidal structure over the equatorial plane of AGNs have been dismissed. The scientific community now widely accepts clumpy torus models. However, the question about whether the torus is entirely clumpy or if there is an inter-cloud medium present, remains open.

1.2 X-ray astronomy

X-rays are a type of radiation in the high energy range of the electromagnetic spectrum. Located between the ultraviolet and γ-rays region, X-rays have extremely small wavelengths ranging between 80 nm to 10^{-5} nm, corresponding to energies between 15 eV and 120 MeV (Lenk & Gellert, 1989).

In astrophysical sources, there are five mechanisms by which X-rays can be produced:

- **Bremsstrahlung radiation:** It is produced when the trajectory of an accelerating electric charge, such as an electron, is deflected by the electromagnetic

field of positive ions in a gas, causing the electron to lose energy and emit X-rays. This process is typical in the intracluster medium of galaxy clusters or in hot plasmas such as those in the vicinity of stars.

- **Synchrotron radiation:** This type of radiation is emitted by relativistic electrons when accelerated by the force caused by a magnetic field, perpendicular to their direction of motion. This mechanism is present in objects with strong magnetic fields such as pulsars, Blazars, and supernova remnants.

- **Thermal emission:** This radiation is emitted by a body solely based on its temperature. This corresponds to blackbody radiation, where the body is assumed to absorb all the incident electromagnetic radiation. Examples of this are found in the cores of stars, supernova remnants, or isolated neutron stars.

- **Inverse Compton scattering:** During this process, low-energy photons are scattered off high-energy electrons, gaining energy and reaching X-ray energies, as opposed to the Compton scattering, where the energy is transferred from the photon to the electron. This mechanism is present in AGN and in the interaction between the photons from the cosmic microwave background and the electrons in the hot gas of a galaxy cluster.

- **Characteristic X-ray excitation:** This process takes place when a sufficiently high energetic particle collides with an atom, exciting inner-shell electrons. The removal of those electrons creates vacancies that are filled via the de-excitation of outer-shell electrons, thereby emitting X-rays. This process is important for the X-ray emission of the solar flares.

In AGN, the dominant X-ray production mechanisms are the inverse Compton scattering and the synchrotron radiation. On the one hand, Inverse Compton scattering is present when the low-energy photons emitted by the accretion disk in the optical and UV range interact with the high-energy electrons in the X-ray corona (Section 1.1.3). The electrons transfer part of their energy to the photons, which are then re-emitted at higher energies in the X-ray (Dermer et al., 1992). On the other hand, as mentioned in section 1.1.3, the jets are composed of highly energetic particles. Due to the magnetic field generated by the spinning of the supermassive black hole and the accretion disk, the particles are accelerated at relativistic speeds, emitting synchrotron radiation.

1.2.1   X-ray spectrum in AGN

In its most fundamental structure, the X-ray spectrum of an AGN can be modeled with two components: a primary power law and an exponential cut-off. Fig. 1.4 shows a schematic representation of such a spectrum.
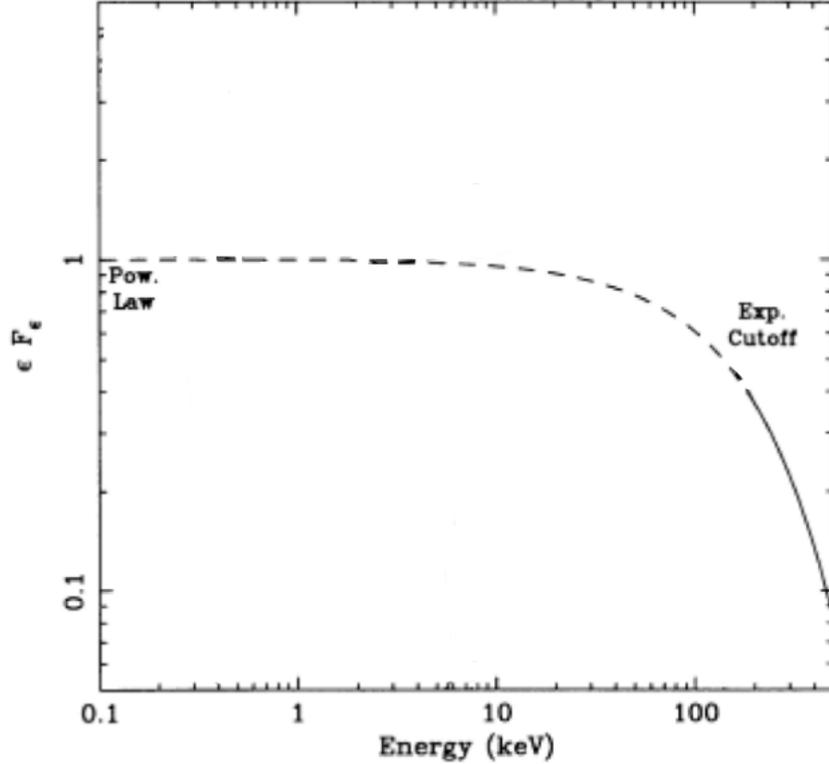
Figure 1.4: Fundamental structure of the X-ray spectrum of an AGN. The y-axis represents the Flux multiplied by the Energy. Modified from Fabian (1999)

- **Power law:** This component is produced by the Comptonization of photons by energetic electrons in a hot plasma located close to the supermassive black hole (the X-ray corona). The power law can be modeled as:

$$I_f \propto f^{-\alpha} \tag{1.2}$$

  where I_f is the intensity or flux density of the radiation, f is the frequency of the radiation, and α is the spectral slope, also given in terms of the photon index as $\Gamma = \alpha + 1$. For AGN, the photon index typically lies in the range of $\Gamma \sim 1.7 - 2.5$ (Ishibashi & Courvoisier, 2010), with slight differences between Seyfert 1 and Seyfert 2 galaxies.

- **High-energy cut-off:** The high-energy cutoff, modeled as an exponential function, is given by the maximum energy the Comptonized photons can reach, due to the energy distribution of the electrons scattering them in the X-ray corona. This parameter is a direct manifestation of the temperature of the Comptonizing plasma in the corona, and its distribution is found to be in the range of $140 - 500$ keV (Baloković et al., 2020).

When considering the entire set of physical processes occurring in an AGN, the spectrum shown in Fig. 1.4 has additional components and can look as shown in Fig 1.5.
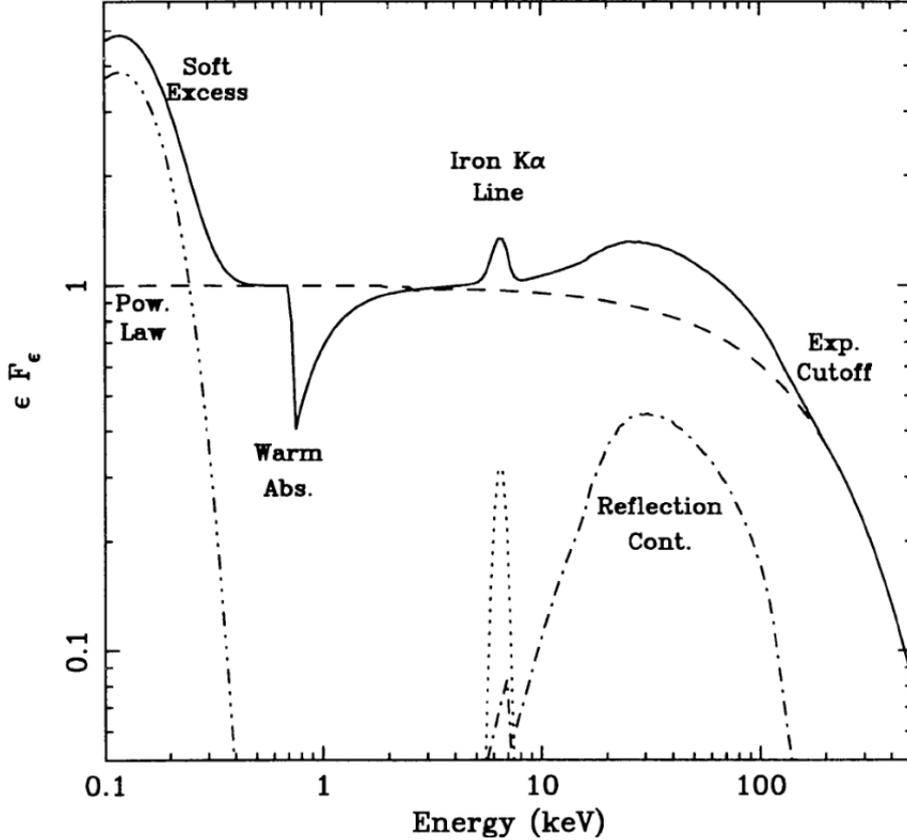
Figure 1.5: X-ray spectrum of an AGN. The y-axis represents the Flux multiplied by the Energy. The solid line represents the resulting X-ray spectrum after taking into account different physical processes evidenced as spectral components such as the Soft Excess (dash-double dotted line), primary power law (dashed line), Reflection continuum (dash-dotted line) and the Iron Line (dotted line) (Fabian, 1999).

- **Soft excess:** This phenomenon occurs in the majority of AGN in the low-energy end of the X-ray spectrum, where the observed emission below 1 keV exceeds the one expected from extrapolating the emission in the higher energy range between $2 - 10$ keV. The origin of this component can be explained by ionized material moving at relativistic speeds close to the SMBH, and according to recent studies (Waddell et al., 2023), it is better modeled by a power law additional to the primary one, than by thermal blackbody emission.

- **Photoelectric absorption:** Additional to the clouds from the broad and narrow line regions, material in the line of sight to the AGN might affect the shape of the soft X-ray spectrum. Through photoelectric absorption, this material absorbs the incoming radiation, showing different effects at different ranges of the spectrum depending on the temperature of the medium. Thereby, absorbers are classified as warm and cold (Halpern, 1984).

  - **Warm absorber:** With temperatures of the order of $10^5 - 10^6$ K, this type of mediums are partially photo-ionised. The effect of the warm absorbers is evidenced in the X-ray spectrum as edges and narrow absorption lines, mainly from C, N, O, Ne, and Fe, such as C v, C vi, O

15

VII, O VIII and Ne X, among others. These features are a direct signature of the structure and kinematics of the gas in the medium (Laha et al., 2011).

– **Cold absorber:** The effect of photo-electric absorption decreases with increasing photon energy. As a consequence, cold absorbers affect mainly the low energy range of the X-ray spectrum, as they absorb soft energy photons first. The primary power-law becomes visible only at higher energies. High column densities of the material lead to higher energy photons being absorbed.

- **Iron Kα line:** This is the most prominent emission line in the X-ray spectrum of Compton thick AGN, observed at an energy of 6.4 keV. This line is produced by the reprocessing of hard X-ray photons by neutral material surrounding the SMBH, most likely the cold and dense one located in the torus (Ricci et al., 2014). The process by which the Iron Kα line is created is the characteristic X-ray excitation of the cosmically abundant iron atoms, by which the K-shell of the atom is photo-ionized by a sufficiently energetic photon, causing the transition of an electron from the L-shell to fill the vacancy.

- **Compton-reflection hump:** This effect is observed in the X-ray spectrum at energies above $20 - 30$ keV, as a result of a combination of photoelectric absorption and Compton scattering in the accretion disk. Photo-electric absorption decreases with increasing photon energy. This implies that low-energy photons (E $\leq$ 20 keV) are more likely to be absorbed near the surface of the material before they are reflected. For high-energy photons (E $\geq$ 30 keV), however, the probability of getting absorbed in the surface is smaller, allowing them to penetrate deeper into the material. As a consequence, these photons undergo multiple Compton scatters before they can escape again, losing part of their energy in each of them. Ultimately, low and high-energy incident photons are reflected and re-emitted at roughly the same energies, creating an excess at intermediate energies (George & Fabian, 1991).

1.2.2 Effect of absorption in X-ray spectrum of AGN

The material in the line of sight between the source of X-rays in the AGN and the observer affects the shape of the X-ray spectrum. This is because the material absorbs a fraction of the incident radiation via photoelectric absorption while transmitting the rest. The fraction of absorbed to transmitted radiation is dependent on the hydrogen column density $N_{\rm H}$ of the material. This variable measures the number of hydrogen atoms H I per cm^2 along the line of sight of the central X-ray source of the AGN. The higher the $N_{\rm H}$, the more soft photons are absorbed by the material, and the fewer are transmitted.

The absorption can be carried out by intrinsic or extrinsic structures to the AGN:

- **Extrinsic absorption**: This is due to material external to the AGN itself, such as that in our galaxy or the CGM. The galactic absorption, due to the

material in the Milky Way, has been estimated as $N_{\text{H}} \sim 10^{20}$ cm^{-2} (Kalberla & Kerp, 2009). Comparably, studies of the CGM of multiple different galaxies (Werk et al., 2014; Voit, 2019) have estimated the material in the CGM to have hydrogen column densities in the range $10^{15} - 10^{18}$ cm^{-2}.

- **Intrinsec absorption**: This is due to structures surrounding and in the AGN itself, such as the host galaxy, with column densities up to 10^{22} cm^{-2} and the torus, with values of N_{H} in the range of $10^{22} - 10^{25}$ cm^{-2}.

Comparing the N_{H} values of the different absorbing structures, it is concluded that the intrinsic absorption is the main contributor of AGN obscuration, primordially attributed to the torus, while the extrinsic absorption is negligible.

The material in the torus absorbs the X-ray radiation emitted by the corona, thereby obscuring the central engine. As mentioned above, the column density of this material determines the likelihood of the X-ray radiation being transmitted: when $N_{\text{H}} < 1.5 \times 10^{24}$ cm^{-2} (Markowitz et al., 2014), part of the X-ray emission, that is the soft X-ray photons, is absorbed, while the rest, corresponding to the hard X-ray ones, is transmitted. This is due to the anti-correlation between the photo-absorption cross-section and the photon energy. In this case, the AGN is referred to as a 'Compton-Thin' source.

As shown in Fig. 1.6, the consequence of the absorption in the X-ray spectrum is a lack of photons in the soft energy band, while the rest of the spectrum components, such as the power-law, remains unchanged at higher energies.

When $N_{\text{H}} > 1.5 \times 10^{24}$cm^{-2} (Markowitz et al., 2014), harder X-ray photons are more likely to be absorbed, thereby the absorption component is shifted to higher energies. In this case, the AGN is referred to as a 'Compton-Thick' source, when other physical effects start to govern the X-ray spectrum.
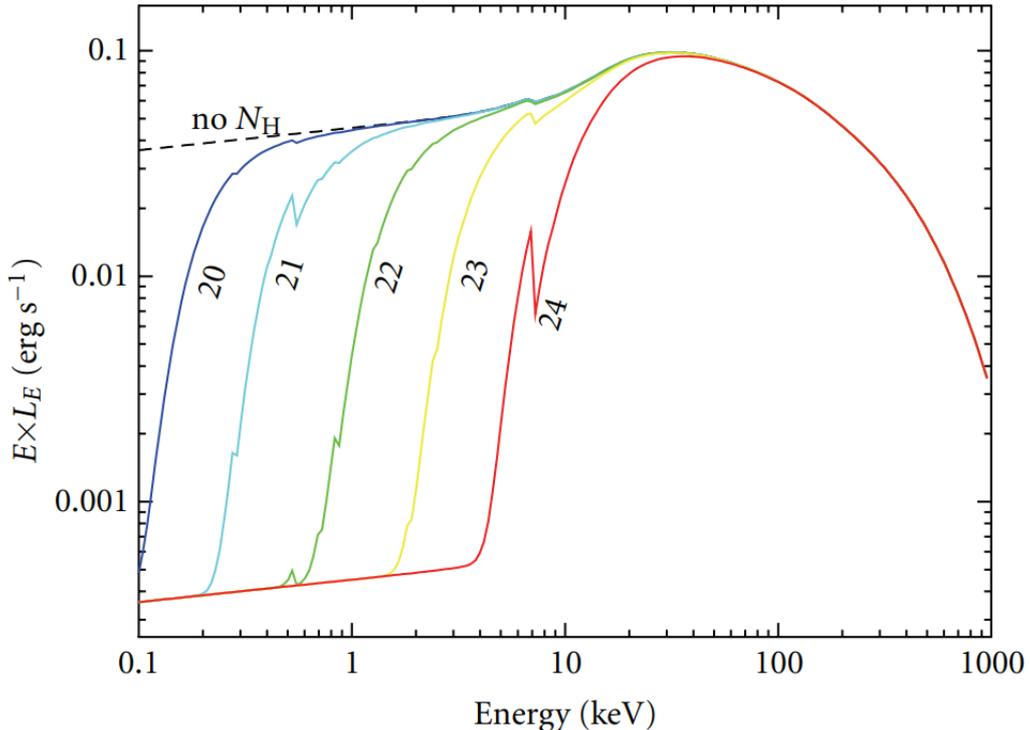
Figure 1.6: Typical X-ray spectrum of an AGN affected by different $N_{\rm H}$ values. The spectrum includes a power law, a Compton reflection hump peaking at ~ 30 keV, a high-energy cutoff at 300 keV, and photoelectric absorption. The numbers in each curve correspond to $\log N_{\rm H}$ in units of cm^{-2}, evidencing the strength of the absorption component. A higher $N_{\rm H}$ is evidenced by a smaller amount of soft photons detected. The y-axis represents the Energy density as the energy multiplied by the luminosity per unit energy (Treister et al., 2012).

Fig. 1.6 implies that for a 'Compton-Thin' AGN, the absorption component of its X-ray spectrum contains information about the absorbing material in the torus.

1.2.3 X-ray telescopes

Due to its high penetration power, X-rays are easily transmitted through a material, and only get reflected at very small incident angles, of a few degrees, below the critical angle which depends on the properties of the medium. Because of this property, X-ray telescopes do not use mirrors or lenses to redirect and focus the incident radiation into the focal point, as optical telescopes do, but rather utilize grazing incidence mirrors that reflect the X-rays at shallow angles. Additionally, X-ray telescopes must be launched into space, as the Earth's atmosphere blocks all X-ray radiation.

The types of mirrors used by X-ray telescopes can have geometries such as the Wolter system. This mechanism is used by telescopes such as *Chandra*, *XMM-Newton*, and *NuStar*, and consists of a paraboloid primary mirror followed by a hyperboloid secondary reflector, forming a mirror shell (Buitrago-Casas et al., 2020).

XMM-Newton

The X-ray Multi-Mirror Mission (*XMM-Newton*) is an X-ray telescope launched by the European Space Agency on December 10th, 1999. It is equipped with three co-aligned telescopes with a field of view of 30 arcmin and spatial resolution of about 6 arcsec (FWHM). Each telescope consists of 58 Wolter I grazing-incidence mirrors, nested in a coaxial and cofocal configuration, with imaging CCD detectors placed in the focal point of each of them. *XMM-Newton* is the most sensitive X-ray telescope to date, as it has the highest effective area among all imaging X-ray instruments.

On board the *XMM-Newton* spacecraft there are three types of instruments, sketched on the structure of the telescope in Fig.1.7:
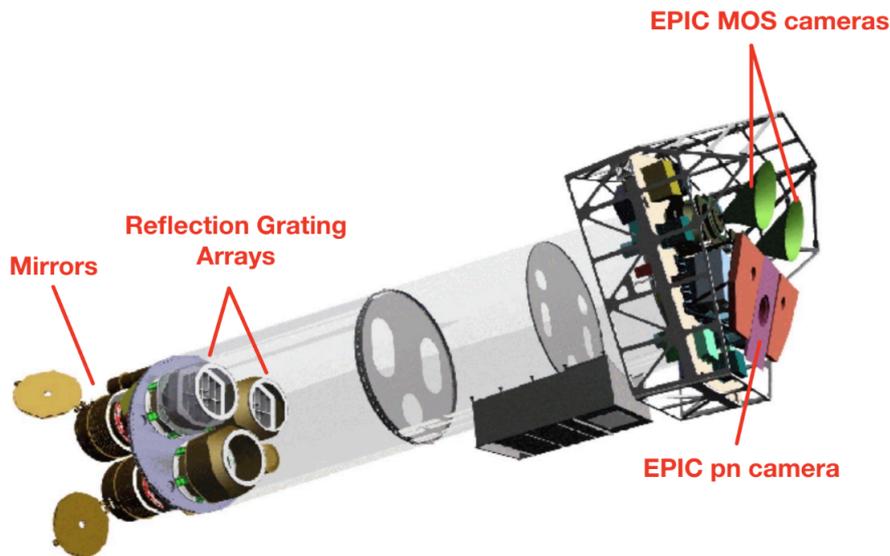
Figure 1.7: Sketch of the structure of the *XMM-Newton* telescope, showing the main structural components (names in red). Taken and modified from ESA (2023).

- **European Photon Imaging Camera (EPIC)**: On board *XMM-Newton*, three X-ray CCD cameras operate in the energy range of $0.15 - 15$ keV: two Metal Oxide Semi-conductor cameras (MOS1 and MOS2), and one pn CCD (Charged Coupled Device). These CCD cameras rapidly detect each photon that reaches the telescope, capturing all its information such as arrival time, energy, and position. Thanks to this, *XMM-Newton* is able to collect the information from hundreds of thousands of photons and store it into a photon event table, which makes part of the instrument's final output.

  The pn camera uses twelve CCDs arranged on a single wafer as shown in Fig. 1.8a. Its complete readout cycle[2] is of 73.3 ms for one individual CCD subunit (Strüder et al., 2001). For comparison, each of the MOS cameras has seven CCDs arranged as shown in Fig. 1.8b. Their basic readout speed is 2.6 s (Turner et al., 2001).

---

[2]The readout time refers to the duration it takes for the detector to collect the photons and transfer its information for processing.
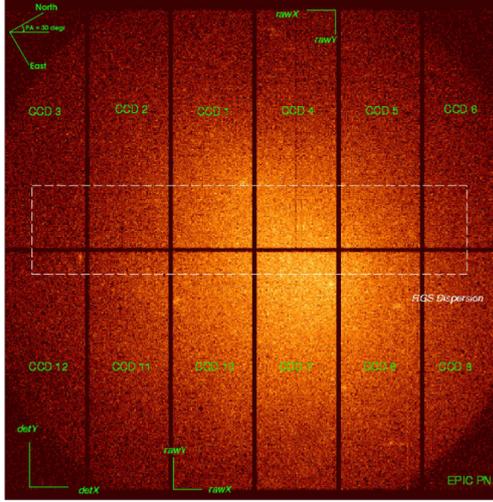
(a) Layout of the pn camera.
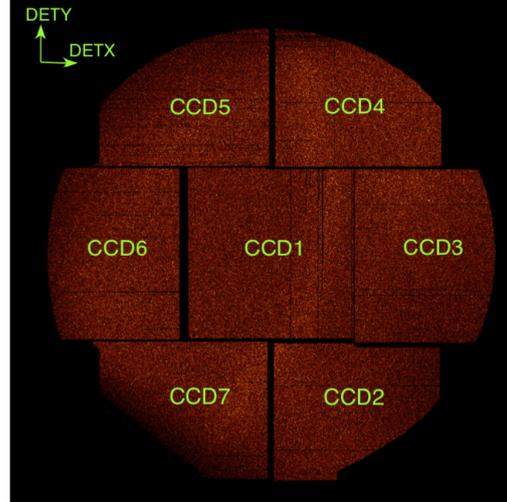
(b) Layout of the MOS cameras.

Figure 1.8: Layouts of the EPIC cameras. The orientation of the `[RAWX,RAWY]` (CCD coordinates) and `[DETX,DETY]` (detector coordinates) axes is shown. The chip array numbering scheme is shown (ESA, 2023).

According to the *XMM-Newton* Users Handbook (ESA, 2023), the EPIC cameras have three operating modes for data acquisition, presented in Fig.1.9a and Fig.1.9b:

– **Full Frame/Extended Full Frame:** All pixels in the twelve CCDs of pn and the seven of the MOS cameras are read out, and the full field of view is covered. Extended Full Frame is only available to pn detector.

– **Partial window:** Either **Large Window mode** or **Small Window mode**. For the MOS cameras, this means that the central CCD can be operated in a different mode than the outer ring of 6 CCDs, which stays in Full Frame mode. For the pn camera, in the Large Window mode, only half of the area in the twelve CCDs is read out, while in Small Window mode only a part of CCD number 4 is used to collect data.

– **Timing mode:** Here, the spatial information is maintained only in one dimension, along the column axis and it is lost along the row direction due to continuous shifting and collapsing of rows to be read out at high speed.

– **Burst mode:** Only available for the pn camera, it is a type of Timing mode offering high time resolution.
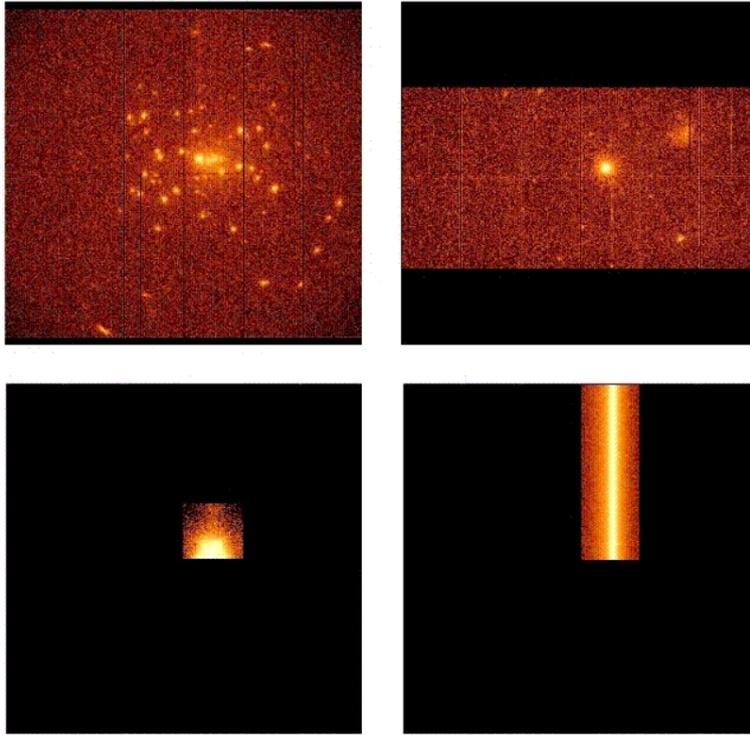
(a) Operating modes for the pn-CCD camera.
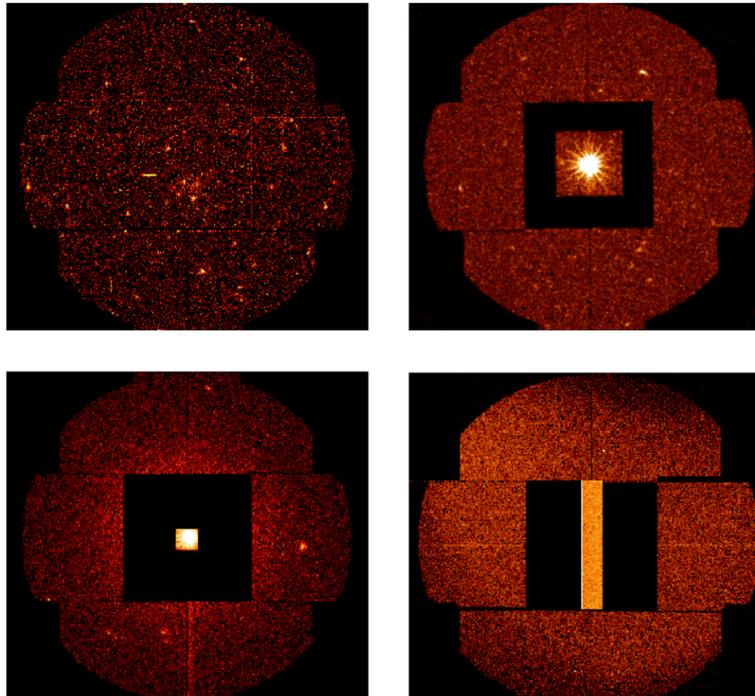
(b) Operating modes for the MOS-CCD cameras.

Figure 1.9: Operating modes of the EPIC cameras. The panels in both figures represent the same mode. Top left: Full Frame/Extended Full Frame. Top right: Large Window mode. Bottom left: Small Window mode. Bottom right: Timing mode (ESA, 2023).

- **Reflection Grating Spectrometer (RGS):** These instruments are used for very high spectral resolution. They are located behind two of the three telescopes on board *XMM-Newton* and operate in the soft energy band, between $0.33 - 2.5$ keV. Each RGS consists of an array of reflection gratings that diffract about half of the incident X-ray radiation into an array of CCD detectors (Den Herder et al., 2001).

- **Optical Monitor (OM):** The optical/UV monitor carries out routine multi-wavelength observations of the targets of the X-ray telescopes, in the range of 170 nm $- 650$ nm of the central 17 arc minute square region of the X-ray field of view (Mason et al., 2001).

1.3 Previous studies of the structure of the AGN torus

As mentioned in section 1.2.2, the X-ray spectrum of an AGN is affected by the obscuring material in the line of sight due to the effect of photoelectric absorption. Multiple studies have been carried out to analyze such material. In particular, research has been focused on probing the clumpy structure of the circumnuclear material around the SMBH of the AGN, considered to be the main contributor to the X-ray absorption.

Some studies place the location of the obscuring clouds in close regions to the SMBH. Risaliti et al. (2009) analyzed an *XMM-Newton* 5-day continuous monitoring of NGC 1365, finding a varying absorption component of $N_\mathrm{H} \sim 9 \times 10^{23}$ cm^{-2}. The rapid time scales of the variations, of the order of hours, attributed the changes to material in the BLR. Lamer & Uttley (2003) have monitored NGC 3227 with the Rossi X-ray Timing Explorer (*RXTE*), finding an absorption event lasting about 100 days due to a gas cloud of $N_\mathrm{H} = 2.6 \times 10^{23}$ cm^{-2} located in the BLR too. In contrast, other studies locate the obscuring clouds further away from the SMBH, more specifically in the torus. Using Centaurus A observations from *RXTE*, Rivers et al. (2011) has tracked an obscuration event lasting ~ 170 days, due to a clump of material of around 8.4×10^{23} cm^{-2} transiting the line of sight. More systematic studies on N_H changes have been performed by Markowitz et al. (2014) and Laha et al. (2020). All of these studies provide evidence that the clumpy model is a more accurate description of the AGN torus, than the homogeneous picture.

1.3.1 Detection of cloud events using time-resolved spectra

Markowitz et al. (2014) has provided the first X-ray-based statistical support for the clumpy-torus model. The study analyzes the X-ray data of a sample of 37 Seyfert 1 and 18 Compton-thin Seyfert 2 AGN obtained from *RXTE*, to identify eclipse events of the X-ray source caused by absorption of the material in the torus. This telescope is sensitive to full-covering, neutral, or lowly ionized clouds with column densities of $N_\mathrm{H} \gtrsim 10^{22}$ cm^{-2} up to 10^{25} cm^{-2}.

To achieve this goal, the X-ray flux light curve has been extracted in multiple energy bands for each source: $2 - 10$ keV, $2 - 4$ keV, $7 - 10$ keV, $4 - 7$ keV, $10 - 18$ keV. Making use of these values, the Hardness Ratio[3] HR1 and HR2 in two different bands have been estimated. Although a candidate obscuration event is evidenced as a sudden increase in the HR light curves, this could also be caused by a change in the photon index Γ of the primary power law, and not necessarily by an increase in the column density N_{H} along the line of sight due to material in the torus or the BLR.

Follow-up time-resolved spectroscopy is performed to confirm the candidate events. To flag a candidate eclipse event as a 'secure' one, the increase in HR must be backed up by an increase of N_{H} and not a flattening of Γ. Heeding this, the Γ_{app} light curve has been extracted. This light curve shows the time evolution of the apparent photon index, which is obtained by fitting the X-ray spectrum to a simple power law with Galactic absorption, neglecting effects such as additional N_{H}, Compton reflection, and Fe K emission. Finally, by binning observations in appropriate time bins and fitting the spectra with appropriate models, the Γ (corresponding to a fit where $N_{\mathrm{H}} = 0$) and the N_{H} (Γ frozen to a given value for each source) light curve have been extracted.

An example of a 'secure' event identified for NGC 3227 is shown in Fig.1.10

---

[3]The Hardness Ratio of an X-ray source is a measure of the relative intensity of X-rays emitted in a high-energy band, compared to those emitted in a low-energy band. Its standard definition is given in Section 3.3, although alternative definitions such as the one of HR1 and HR2 in Markowitz et al. (2014) serve the same purpose.
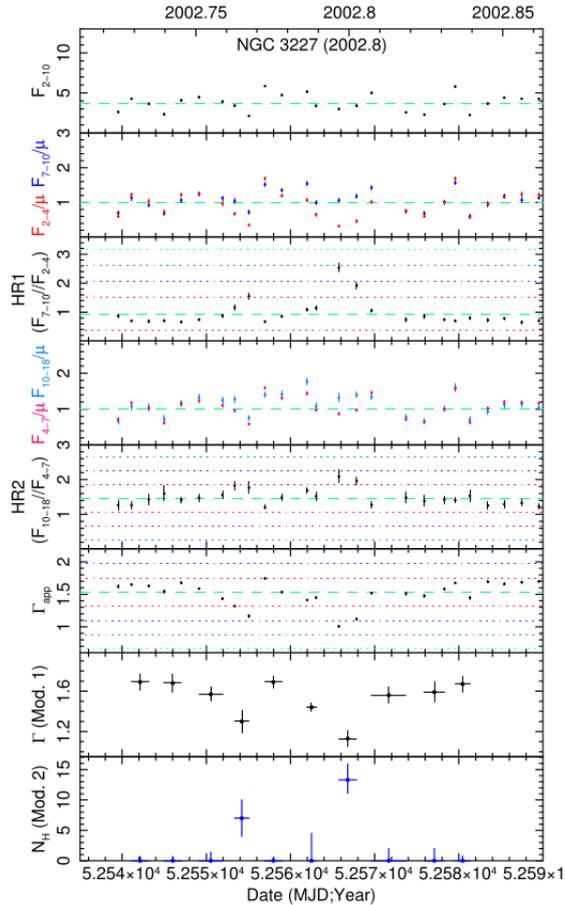
Figure 1.10: NGC 3227. The plot shows a secure occultation event in late 2002, backed by a significant increase in HR1. Top panel: Light curve of the flux F_{2-10} in the $2-10$ keV band, in units of 10^{-11} erg cm^{-2} s^{-1}. Each point corresponds to one observation. Second panel: Mean normalized F_{2-4} (red) and F_{7-10} (blue). Third panel: HR defined as HR1 $= F_{7-10}/F_{2-4}$. Fourth panel: Mean normalized F_{4-7} (red) and F_{10-18} (blue). Fifth panel: HR2 $= F_{10-18}/F_{4-7}$. Sixth panel: Γ_{app} (power law and galactic absorption). Seventh panel: Γ ($N_{\mathrm{H}} = 0$). Bottom panel: N_{H} light curve (Γ frozen at 1.61) in units of 10^{22} cm^{-2}. Taken from Markowitz et al. (2014).

After analysis of all the objects in the sample, Markowitz et al. (2014) has identified 12 X-ray obscuration events: 5 events in 3 Seyfert 1 AGNs, 3 events in 2 Seyfert 1.5, and 4 events in 3 Seyfert 2. Using these results, two main parameters can be constrained: the probability of observing a source undergoing an eclipse event and the distance from each obscuring cloud to the X-ray source.

Using the eclipse event duration and the observation sampling patterns, the probability of observing a source undergoing an eclipse event is estimated as $0.006^{+0.160}_{-0.003}$ for Type I AGNs and $0.110^{+0.461}_{-0.071}$ for Type II.

By using the event's duration, clouds' ionization level, and column density, the distance from each obscuring cloud to the X-ray source is estimated using the method proposed by Lamer & Uttley (2003), as follows.

The ionization parameter[4] ξ is given by:

$$\xi = \frac{L_{\text{ion}}}{n_{\text{H}} r_{\text{cl}}^2} \tag{1.3}$$

where L_{ion} is the ionizing luminosity[5] in the energy range $13.6 \text{ eV} - 13.6 \text{ keV}$, n_{H} is the hydrogen number density of the gas and r_{cl} is the distance from the cloud to the X-ray source.

Assuming each cloud has uniform density and ionization parameter, the cloud diameter is given by:

$$D_{\text{cl}} = N_{\text{H}}/n_{\text{H}} = v_{\text{cl}} t_{\text{D}} \tag{1.4}$$

where v_{cl} and t_{D} are the velocity and crossing time (in days) of the cloud across the line of sight, respectively.

Assuming the cloud is moving in a Keplerian circular orbit of radius r_{cl} around the SMBH of mass M_{BH}, the clouds' velocity can be expressed as $v_{\text{cl}} = \sqrt{GM_{\text{BH}}/r_{\text{cl}}}$ and the hydrogen number density as:

$$n_{\text{H}} = \frac{N_{\text{H}}}{t_{\text{D}}} \sqrt{\frac{r_{\text{cl}}}{GM_{\text{BH}}}} \tag{1.5}$$

where G is the gravitational constant.

Combining equations 1.3 and 1.5 and solving for r_{cl}:

$$r_{\text{cl}} = 4 \times 10^{16} M_7^{1/5} \left(\frac{L_{42} t_{\text{D}}}{N_{\text{H,22}} \, \xi} \right)^{2/5} \text{ cm} \tag{1.6}$$

where $M_7 = M_{\text{BH}}/10^7 \text{ M}_\odot$, $L_{42} = L_{\text{ion}}/10^{42} \text{ erg s}^{-1}$ and $N_{\text{H,22}} = N_{\text{H}}/10^{22} \text{ cm}^{-2}$.

The value of M_{BH} is taken from either reverberation mapping studies or estimations from stellar kinematics or gas dynamics. L_{ion} corresponds to the sum of two components of the luminosity: $L_{0.1-13.6 \text{ keV}}$ in the energy range $0.1 - 13.6 \text{ keV}$ and $L_{0.0136-0.1 \text{ keV}}$ in the range $0.0136 - 0.1 \text{ keV}$, the latter containing significant contributions from the thermal accretion disc continuum emission. For the ionization parameter, it is assumed that $\log \xi = -1$, 0 and $+1$.

The estimated values for r_{cl} lie in the range from tens to hundreds of light days from the central engine of the AGN. For six of the sources in the sample, these distances are commensurate with radii corresponding to the outer region of the BLR or up to ~ 15 times its outer boundary. It is safe to conclude, then, that the location of the X-ray absorbing clouds observed in these sources is more consistent with the torus than with the BLR.

1.3.2 Long-term column density variability

Laha et al. (2020) investigate the X-ray obscuring column density variability of a sample of 20 perpetually obscured Compton-thin Seyfert 2 AGN, observed with

---

[4]The ionization parameter quantifies the degree of ionization of a gas by measuring the ratio of the ionizing photon flux to the gas density Netzer & Woltjer (1990).

[5]The ionizing luminosity quantifies the total amount of energy emitted by a source in a given range, capable of ionizing surrounding material.

XMM-Newton, *Chandra*, and *Suzaku*. This set of instruments allows for a sensitivity of $N_{\mathrm{H}} \sim 10^{20.5-24}$ cm^{-2} of fully and partially covering neutral and lowly ionized gas.

The study aims to assess the change of N_{H} across consecutive observations and different instruments. To do so, the average X-ray spectrum is extracted for every observation per instrument per source. Each spectrum is then modeled with different variations of a best-fit model given, in ISIS[6] notation, by:

$$
\begin{aligned}
&tbabs(1) \times (apec(1) + apec(2) + powerlaw(1) + ztbabs(1) \times \\
&(powerlaw(2) + pexmon(1) + zgauss))
\end{aligned}
\tag{1.7}
$$

or, if an additional partial covering absorption component is required by the data:

$$
\begin{aligned}
&tbabs(1) \times (apec(1) + apec(2) + powerlaw(1) + ztbabs(1) \times \\
&zpcfabs(1) \times (powerlaw(2) + pexmon(1) + zgauss))
\end{aligned}
\tag{1.8}
$$

An explanation of each of the model components is found in Section 3.3.

By fitting each spectrum to an appropriate variation of the best-fit model, the values of the hydrogen absorption column N_{H} and the photon index Γ are obtained for each observation. These values are then concatenated to create the light curve of the parameters and analyze their behavior to identify significant changes over time. The resulting N_{H} and Γ light curve for one of the sources in the sample is observed in Fig. 1.11

---

[6]The *Interactive Spectral Interpretation System* (ISIS) is an astronomical software designed to facilitate the interpretation and analysis of high-resolution X-ray spectra.
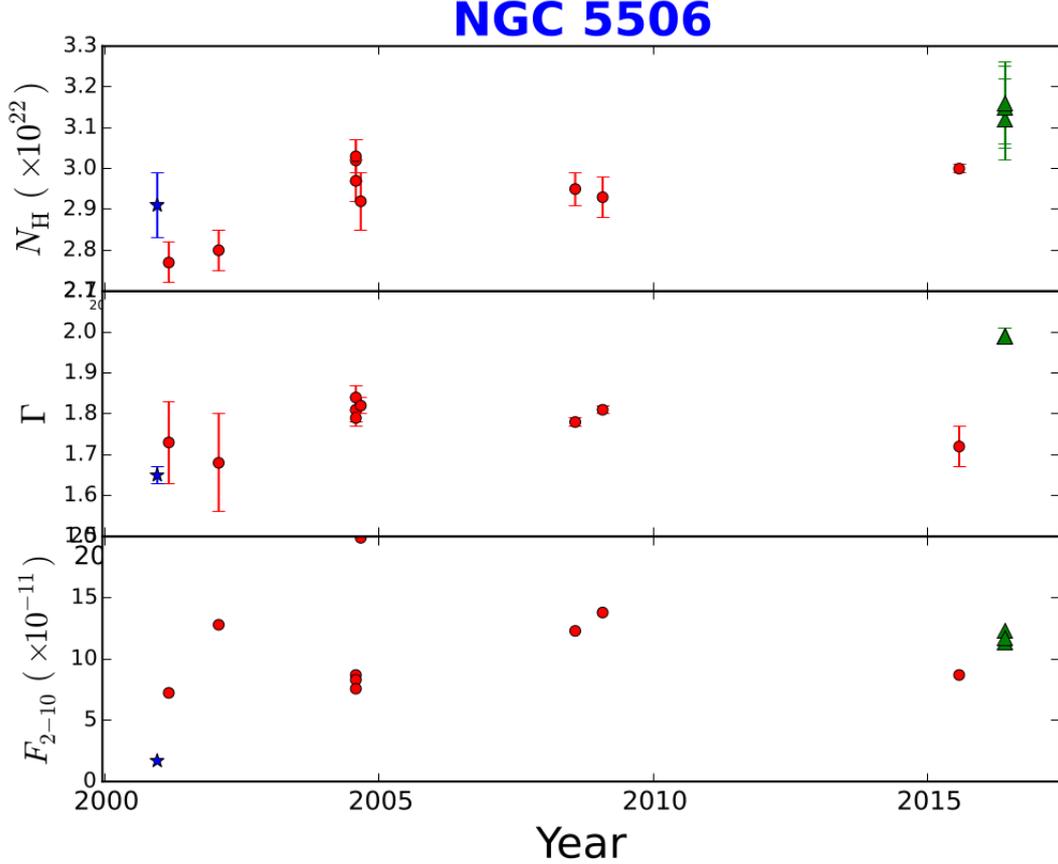
Figure 1.11: Top panel: N_H light curve for the Seyfert 2 AGN NGC 5506. Second panel: Γ light curve. Bottom panel: $2 - 10$ keV unabsorbed flux. The red circles, blue stars, and green triangles represent the data obtained from the *XMM-Newton*, *Chandra*, and *Suzaku* satellites, respectively. Taken from Laha et al. (2020).

When comparing observations for a given source with a single instrument, 3 types of absorbers are identified: full-covering, partial-covering, and constant absorbers. Some sources present more than one type of absorber.

Full-covering absorbers: Eight robust variations of N_H with $|\Delta N_H| \sim 0.1 - 1.9 \times 10^{22}$ cm^{-2} are identified for seven of the sources in the sample, namely Cen A, Fairall 49, MCG-5-23-16, NGC 2992, NGC 5252, NGC 5506, and NGC 7582, on timescales of 2 months to 14.5 years. These events are consistent with full-covering absorbers. Three models are proposed to explain such changes:

- The obscuration is attributed to discrete clumps located in the torus, where an increase or decrease of N_H follows the entry or exit of a single cloud from the line of sight.

- The obscuration is provided by the combination of a time-constant component, such as the dust structures at kiloparsec scales, and a set of compact-scale clumps.

- A non-clumpy, volume-filling, compact-scale inhomogeneous medium, such as the inter-cloud medium in the torus that contains over and under dense regions transiting the line of sight, is considered as the source of the obscuration.

27

Discerning between the three scenarios is, however, impossible, as the sample is restricted in duration and no events are registered from ingress to egress.

Partial-covering absorbers: Eleven of the twenty sources in the sample show signatures of partial-covering absorbers in the range of $1 - 80 \times 10^{22}$ cm^{-2}, with covering fractions between $30\% - 90\%$. The nature of these events can be explained with near-Compton-thick compact scale gas, although further constraints on the size and location of the clouds are limited due to the CCD energy resolution and model complexity.

Constant absorbers: For 13 of the sources in the sample, the absorber column density remains constant over time, down to levels of $|\Delta N_{\mathrm{H}}| \sim 0.1 - 17 \times 10^{22}$ cm^{-2}. Three scenarios are proposed to explain this phenomenon:

- The torus is composed of a large number of low-column density clouds, such that the ingress/egress of a single one does not change the N_{H} or the number of clouds in the line of sight value by a perceptible amount.

- The torus is either a smooth compact structure, or it follows the model proposed by Stalevski et al. (2012) and Siebenmorgen et al. (2015) of a two-phase medium. The latter consists of high-density clouds whose inter-medium is filled by low-density gas. In this study, the high-density clouds would be represented by the absorption events (either full or partial), while the low-density medium to the constant level of N_{H} over time.

- Dusty structures such as lanes or filaments along the line of sight at distances $\gtrsim 0.1$ kpc to several kiloparsecs, belonging to the host galaxy.

One way of discerning between these three scenarios would require a comparison of the observed value of N_{H} in this study to values of N_{H} estimated from optical extinction of known sources of dust at kiloparsec scales and from dust in the torus at parsec scales. These methods, however, have significant limitations that prevent the identification of the nature of the constant absorbers.

Chapter 2

Aims & Scope

Motivation

The Unification Model of AGN proposes that the material in the torus is homogeneously distributed. While this model has long been used to explain the diverse properties of AGN across the electromagnetic spectrum, its definition of the torus has been ruled out by IR and X-ray observations. These observations suggest that the torus material is distributed in clumps. Although clumpy models align more closely with the observations, it remains unclear whether these clumps are embedded in a homogeneous lower-density medium (inter-cloud medium). Thus, determining the type of torus structure that best explains AGN behavior is a significant concern for the scientific community. This understanding is essential for explaining other phenomena such as AGN variability.

Aim

This thesis is focused on answering the question:
What is the structure of the AGN torus: purely clumpy (clouds) or clumps embedded in a homogeneous lower-density medium?

Approach

The main idea behind this thesis is to analyze the X-ray spectra of a sample of Seyfert 2 AGN, which are obscured by material in the torus. By measuring the amount of X-ray absorption caused by this obscuring material and analyzing its time evolution, the distribution of the material in the torus can be constrained.

First, the raw X-ray data of the sample, retrieved from the *XMM-Newton* archive, is reduced and processed by a semi-automated software created for this purpose. Consecutively, each observation is time-resolved by splitting it into multiple time bins and extracting the X-ray spectrum for each of them. The time-averaged spectrum, i.e. the X-ray spectrum for the entire observation time, is also extracted.

By using the best-fit spectral models derived by Laha et al. (2020) for each of the sources in the sample, the time-averaged spectra are fitted to obtain the best-fit parameters for each observation. These parameters are frozen, and two new fits are performed for every time-binned spectrum: the first one, allowing only the

hydrogen column density N_H from the absorption component to vary, and the second one, allowing N_H and the photon index Γ from the primary power law to vary at the same time.

The obtained values for the parameters of interest (N_H and Γ) are concatenated over time. The behavior of the resulting curves is analyzed to identify significant variations indicating N_H transitions. These events include clouds entering or leaving the line of sight, evidenced by an increase or decrease of the column density from its average value, respectively. Evidencing such N_H transitions is proof of inhomogeneous distribution of the material in the torus, proposed by clumpy models.

Overview of this thesis

As mentioned above, in this thesis I make use of the physical effect that the N_H of the absorbing material in the line of sight of an AGN causes on the soft energy range of the X-ray spectrum. This physical effect implies that by tracking the change of the X-ray spectrum of an AGN over time, the change of N_H is also tracked. The behavior of the resulting curve showing the N_H evolution over time ($N_H(t)$ hereafter), gives insights into the structure and distribution of the material in the torus: homogeneous, clouds + inter-cloud medium or clouds only. Obtaining the $N_H(t)$ for the Seyfert 2 AGNs in the sample, is the main result of this thesis.

To achieve the main result of this thesis, I follow a five-step process: data reduction, time binning, spectra extraction, spectra fitting, and $N_H(t)$ curve extraction. An overview of this process follows.

1. **Data reduction:** Obtaining X-ray data from the *XMM-Newton* satellite: multiple observations for multiple Seyfert 2 AGN. The data is processed to remove high-background signal intervals and define the source (+ background) and background (source-free) regions.

2. **Time binning:** Splitting each observation into multiple time bins to track changes in the source within one observation. Each time bin is treated as one observation.

3. **Spectra extraction:** Extraction of the spectrum corresponding to each of the previously obtained time bins in an observation, based on the source (+ background) and background (source-free) regions definition.

4. **Spectra fitting:** Fitting each spectrum to obtain the value of N_H.

5. $N_H(t)$ **curve extraction:** Concatenation of each of the N_H values of a single object to create its $N_H(t)$ curve and analyze it.

There are three main possible scenarios for the $N_H(t)$ curve, sketched in Fig. 2.1, that give insights into the distribution of the material in the torus:

(a) First scenario: Constant $N_{\rm H}$

(b) Second scenario: Stepwise $N_{\rm H}$ increases + decreases

(c) Third scenario: Baseline $N_{\rm H}$ + increases
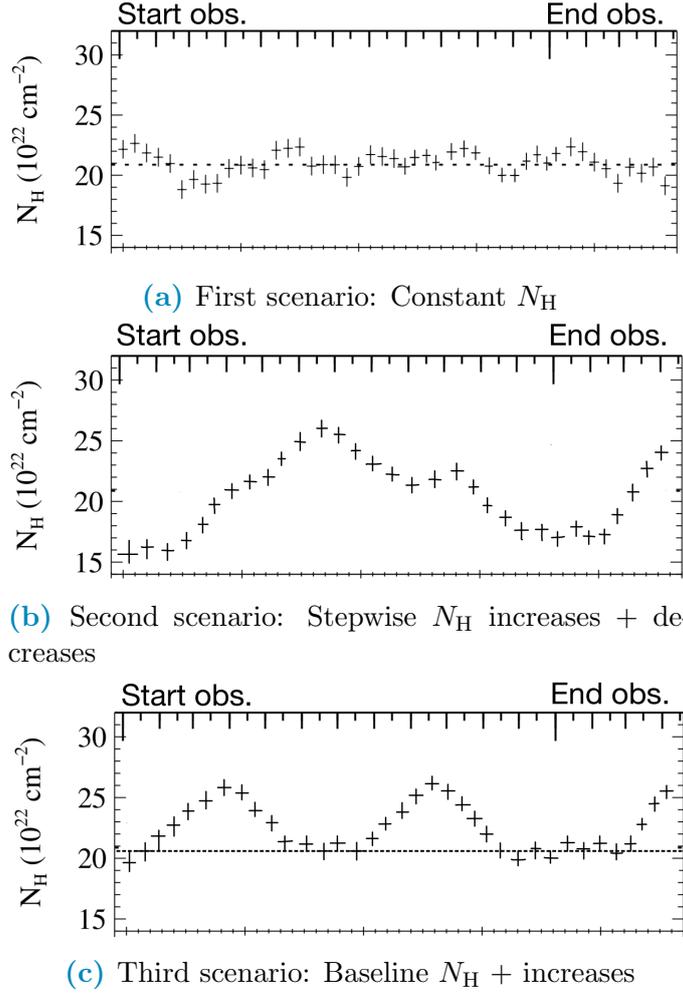
Figure 2.1: Comparison of the three possible behaviors for the $N_{\rm H}(t)$ curve. These plots are artificially created and do not correspond to observed events. They are based on a real cloud event detected by Rivers et al. (2011). The horizontal dotted line represents the baseline $N_{\rm H}$ value, corresponding to the minimum and constant level of X-ray absorption observed in the AGN.

- **Constant $N_{\rm H}$:** In the first scenario, shown in Fig. 2.1a, $N_{\rm H}$ remains constant over time, without considerable changes over or under the baseline value. This type of $N_{\rm H}(t)$ is the signature of either numerous X-ray absorbing clouds with $N_{\rm H} \ll 10^{22}$ cm^{-2} remaining always in the line of sight (Markowitz et al., 2014), or very slow-moving clouds whose ingress or egress to the line of sight is not captured during the observations.

- **Stepwise $N_{\rm H}$ increases + decreases:** The second possibility for $N_{\rm H}(t)$, shown in Fig. 2.1b, includes both discrete increases and decreases of $N_{\rm H}$. This scenario is the evidence of an entirely clumpy torus (Fig.1.2 c.), where multiple clouds with $N_{\rm H} \sim 10^{22}$ cm^{-2} are located along the line of sight at a given moment, and add up to the total column density observed as an increase in the light curve. The decrease in $N_{\rm H}$ corresponds to one or multiple clouds leaving the line of sight and unobscuring the X-ray source. No inter-cloud medium is assumed in this model.

- **Baseline N_H + increases**: The third scenario, sketched in Fig. 2.1c, corresponds to a time evolution where N_H remains constant around the baseline value over time, but sudden increases followed by decreases are evidenced. After such events, N_H returns to its baseline value. If this is the case, the torus is a structure where the material is homogeneously distributed and clouds of higher density are embedded in it, as shown in Fig.1.2 d. The homogeneous component of the torus corresponds to the inter-cloud medium, and it is responsible for the baseline absorption.

Although similar studies of this kind have already been performed (see Section 1.3), this thesis presents two major improvements relative to them:

- By using data from the *XMM-Newton* satellite, the uncertainty of the column density can be tested down to about 10^{21} cm^{-2}, as this is the most sensitive X-ray instrument to date. This allows me to detect obscuring clouds with smaller N_H, compared to the work of Markowitz et al. (2014), where the sensitivity was limited to 10^{22} cm^{-2}. This implies that this thesis provides an improvement of one order of magnitude on the estimation of N_H, compared to previous studies.

- By using time-resolved spectroscopy, I can track the behavior of N_H within a single observation. This opens the possibility of detecting shorter-duration events, compared to those detected by Laha et al. (2020), where only N_H changes between observations were reported, but not within a single observation. Additionally, the use of data from a single X-ray mission in this thesis avoids calibration issues between instruments.

  The spectra obtained in this master thesis have the highest signal-to-noise ratios achieved with X-ray telescopes. Consequently, deriving the spectral fit models is a complex and time-consuming process. Therefore, this thesis is based on the best-fit spectral models derived by Laha et al. (2020) for the sources in the sample.

Chapter 3

Methodology

3.1 Seyfert 2 AGN sample selection

The sample of AGN analyzed in this thesis was chosen according to three main filters:

- Sources previously studied by Laha et al. (2020), which time-averaged X-ray spectrum model is available.

- Sources classified by Laha et al. (2020) and confirmed by the literature as Seyfert 2, as they are more likely to be observed through the torus.

- Sources with at least one *XMM-Newton* observation (as of June 2023) longer than 10 ks to have enough data for time-evolution analysis.

After selecting the sources based on the three previous criteria, the sample consists of ten local Seyfert 2 AGNs located at an average redshift of 0.01 with black hole masses around $\log(M_{\mathrm{BH}}/M_{\odot}) \approx 6.7 - 8.4$. The data collected for the sources in the sample adds up to 1.89 Ms, corresponding to around 21 consecutive days spanning 40 observations.

The data for the ten sources in the sample is listed in Table 3.1.

Source name	R.A. (J2000)	Dec. (J2000)	Obs. ID	Observation date	Duration (s)
IRAS F00521-7054	$00^h53^m56^s.1$	$-70^d38^m04^s$	0301150101	22 Mar 2006	17273
			0301151601	22 Apr 2006	14214
			0795630101	30 Sep 2017	125400
			0795630201	02 Oct 2017	72500
MCG-5-23-16	$09^h46^m48^s.4$	$-33^d36^m13^s$	0112830301	13 May 2001	38436
			0112830401	01 Dec 2001	24909
			0302850201	08 Dec 2005	131710
			0727960101	24 Jun 2013	138399
			0727960201	26 Jun 2013	139408
Mkn 348	$00^h48^m47^s.1$	$+31^d57^m25^s$	0067540201	18 Jul 2002	49498
			0701180101	04 Jan 2013	13115
NGC 2110	$05^h52^m11^s$	$-07^d27^m22^s$	0145670101	05 Mar 2003	59550
NGC 4258	$12^h18^m57^s.5$	$+47d18^m14^s$	0110920101	08 Dec 2000	23311
			0059140101	06 May 2001	12709
			0059140201	17 Jun 2001	13160
			0059140401	17 Dec 2001	15011
			0059140901	22 May 2002	16516
			0203270201	01 Jun 2004	48918
			0400560301	17 Nov 2006	64520
NGC 4507	$12^h35^m36^s.6$	$-39^d54^m33^s$	0006220201	04 Jan 2001	46207
			0653870201	24 Jun 2010	19918
			0653870301	03 Jul 2010	16918
			0653870401	13 Jul 2010	16915
			0653870501	23 Jul 2010	16916
			0653870601	03 Aug 2010	21917
NGC 5252	$13^h38^m15^s.9$	$+04^d32^m33^s$	0152940101	18 Jul 2003	67313
NGC 5506	$14^h13^m14^s.9$	$-03^d12^m27^s$	0013140101	02 Feb 2001	20007
			0013140201	09 Jan 2002	13817
			0201830201	11 Jul 2004	21617
			0201830301	14 Jul 2004	20409
			0201830401	22 Jul 2004	21956
			0201830501	07 Aug 2004	20411
			0554170201	27 Jul 2008	90911
			0554170101	02 Jan 2009	88919
			0761220101	07 Jul 2015	132000
			0880550401	22 Jul 2021	58700
NGC 6300	$17^h16^m59^s.5$	$-62^d49^m14^s$	0059770101	02 Mar 2001	46548
NGC 7172	$22^h02^m01^s.9$	$-31^d52^m11^s$	0147920601	18 Nov 2002	16943
			0202860101	11 Nov 2004	58918
			0414580101	24 Apr 2007	58212
				Total time	**1.89 Ms**

Table 3.1: Sample data overview. Obs. ID represents the unique identifier for each observation within *XMM-Newton*. The Observation Date indicates the start date of the observation, and the Duration specifies the total exposure time from the start to the end date.

3.2 *XMM-Newton* data handling

To carry out the bulk reduction of the data, the time binning, and the extraction of the spectra (steps 1 to 3 mentioned in Chapter 2), a semiautomatic software in Python was created (XMM-DH hereafter, for *XMM-Newton* Data Handling). Each step is carried out by a single executable script: data reduction by REDUCTION.py, time bin creation by BINNING.py, and spectra extraction by SPECTRA.py.

XMM-DH is a wrapper around *XMM-Newton* SAS, which is the Science Analysis System, containing a collection of tasks, scripts, and libraries, specifically designed to reduce and analyze data collected by the *XMM-Newton* satellite (ESA. XMM-Newton, 2023). XMM-DH was written and tested in Python 3.11.0.

The following sections describe the step-by-step procedure to use each part of XMM-DH: REDUCTION.py (Section 3.2.1), BINNING.py (Section 3.2.2), and SPECTRA.py (3.2.3).

Preparation

Before initializing XMM-DH, all the Observation Data Files (ODF) must be available in a single directory. The archive ODF for any observation conducted with *XMM-Newton* can be obtained by visiting the website https://nxsa.esac.esa.int/nxsa-web/#search, searching by the source name and selecting ODF as the download option. This option provides a compressed folder with a .tar.gz extension, which contains all the data for a single observation, including raw science files, instrument housekeeping files, and spacecraft files, among others.

An additional directory, referred to as the Working Directory, must be created to store the products generated by XMM-DH. This directory should also contain the executable scripts from XMM-DH and must be separate from the directory where the original ODF files are stored.

3.2.1 Data reduction

Initialization

To initialize the first XMM-DH pipeline, responsible for the data reduction, the following command is executed in the terminal from the Working Directory: `python3 REDUCTION.py`. The pipeline asks for two initial inputs from the user:

- The Working Directory where all products are stored.

- The ODF directory, where the original ODF files are downloaded. For this master thesis, the ODF directory contains multiple observations for multiple sources (see Section 3.1, Table 3.1).

If and when needed, additional environment variables can be modified by editing the pipeline:

- The variables `SAS_DIR` and `SAS_PATH`, corresponding to the paths where the SAS directory is located.

- `SAS_CCFPATH`, corresponding to the path where the Current Calibration Files (CCF) for XMM-Newton are stored. The directory contains a compilation of all calibration files for *XMM-Newton* data, uniquely identified by an issue number and date of validity.

- `PGPLOT_DIR` and `PGPLOT_PATH`, where plotting tasks are located.

- The path to *XMM-Newton* SAS tasks used in the data reduction process such as `cifbuild`, `odfingest`, `emproc`, `epproc`, `merge`, `evselect`, `tabtigen`, `gtimerge`. More on this below.

After the paths are defined, XMM-DH asks the user for the execution mode. There are two possibilities:

- **Automatic:** XMM-DH runs without asking the user for interaction and pre-determined values are used to produce the final products. Parallel computation is carried out, i.e. execution of multiple processes simultaneously in multiple CPU cores.

- **Personalized:** XMM-DH requires input from the user, who has control over the products by inputting the parameter values in each step. Parallel computation is an option.

From this moment on, XMM-DH operates autonomously, asking for user interaction only if the option was previously selected.

Set-up and files processing

XMM-DH starts by un-compressing the original .tar.gz ODF folders, as well as those .tar files within them (in case the folders have been previously uncompressed, XMM-DH proceeds). Once this is done, the HEASoft[1] and SAS environments are initialized using the `SAS_DIR`, `SAS_PATH` and `SAS_CCFPATH` variables defined previously.

Due to the structure of the `SAS_CCFPATH` directory, it is necessary to find and specify the index of the CCF file corresponding to the observations to be analyzed. As a consequence, XMM-DH continues by executing the task `cifbuild`, which retrieves the observation date of the specified ODF and creates the corresponding Calibration Index File (CIF) for each observation. The CIF file contains information about the location of the calibration files for the given period, and its path is defined by the environment variable `SAS_CCF`.

Having both the ODF and CCF files available, it is possible to run `odfingest`. This task takes excerpts from the instrument housekeeping files and the calibration database to produce an ODF summary file with a .SAS extension. This is an ASCII file containing all the necessary information for SAS to process the observation. The environment variable `SAS_ODF` is then updated to the location of the ODF summary file for each observation.

---

[1]HEASoft is a unified release of FTOOLS and XANADU (IAS. Natural Sciences, 2010). The former is a general package of software to manipulate FITS files, and the latter, a software package comprising high-level, multi-mission tasks for X-ray astronomical data analysis

XMM-DH continues with the execution of `emproc` and `epproc`. These pipeline processing tasks perform the data reprocessing for the MOS and pn camera, respectively, producing photon event tables for all exposures.

From the outputs of these tasks, the photon event tables in Imaging mode are the ones of most interest for this thesis. The photon event tables, named *ImagingEvts* files, are a 12-column table of the calibrated event list file, as shown in Fig. 3.1. Each row corresponds to a photon hitting the detector, and each column contains information about its arrival. The columns are TIME, RAWX, RAWY, DETX and DETY (event position in the focal plane array), X and Y (event position in sky coordinates), PHA (pulse analyzer channel), PI (pulse independent channel), CCDNR (the CCD number), FLAG (a classification of the type of pixel depending on its location and quality: bad pixels, near-the-edge pixels, etc.), and PATTERN (see below for more information).

Select	TIME D s	RAWX I PIXELS	RAWY I PIXELS	DETX I pixel	DETY I pixel	X J pixel	Y J pixel	PHA I CHAN	PI I CHAN	FLAG J	PATTERN B	CCDNR B
All Invert	Modify	Modify	Modify	Modify	Modify	Modify	Modify	Modify	Modify	Modify	Modify	Modify
1	1.540494118055E+08	315	275	315	-563	27253	27283	1046	3440	0	0	1
2	1.540494104291E+08	300	294	-8	-146	27535	26838	1418	4651	0	0	1
3	1.540494107150E+08	574	428	6017	2816	32367	31499	36	138	0	0	1
4	1.540494144555E+08	305	296	92	-102	27611	26917	555	1831	0	0	1

Figure 3.1: Example of the first lines from an original photon even table.

By using the information from different columns, different products are generated. For example, the light curve is obtained by using the TIME column to count the photons arriving at the detector over time. On the other hand, by locating all the photons according to their X and Y position, an image is created, and by using their energy channels, the spectrum is extracted.

In particular, the PATTERN column is important, as it allows the recognition of X-ray events that are split between pixels. Patterns correspond to a number between 0 and 31, each of them corresponding to a different pattern in which the incident photon hit the pixels in the detector, as can be seen in Fig. 3.2:
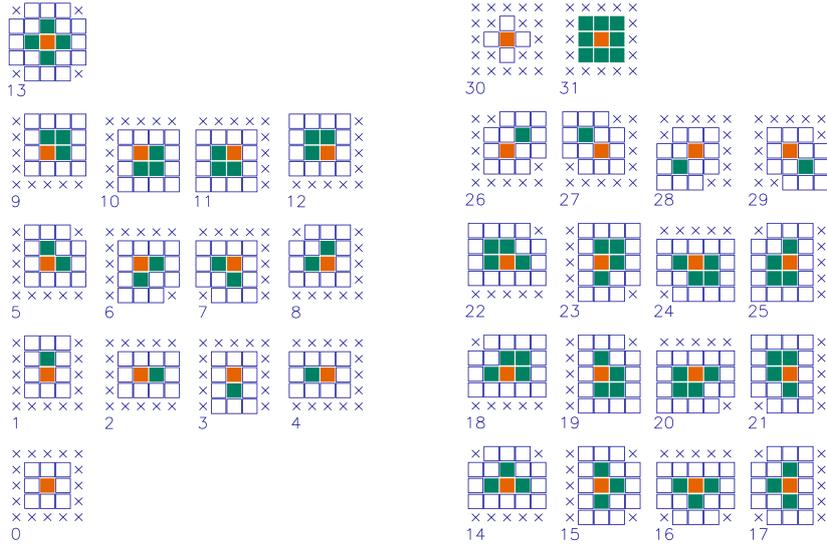
Figure 3.2: Pattern options in *XMM-Newton* EPIC camera Imaging mode. Pattern 0 events are called single, patterns $1-4$ doubles, $5-8$ triples, and so on. Each 5×5 matrix represents a pattern centered on the highest charge pixel colored in red. The pixels above the threshold charge in the pattern are colored in green, while those in white are below the threshold. The crosses indicate indifferent pixels (XMM-Newton Science Analysis System , 2023)

As seen in Fig. 3.2, patterns $0-25$ represent a good compact X-ray pattern, as the highest charge is at the center and isolated. Higher patterns are called diagonal and are not expected from a genuine X-ray source (XMM-Newton Science Analysis System , 2023).

As the *ImagingEvts* files are the ones to be analyzed, XMM-DH moves them to a new sub-folder named `products` within the Working Directory, where all further products are stored. There is usually one single ImagingEvts file per observation and camera (MOS1, MOS2, and pn) combination, whose name contains either the letter S or U. There are, however, cases where a single observation produces two or more different ImagingEvts files, containing either the letter S or U. This case represents one or multiple interruptions of the observation during the exposure time, where the data sets from before and after the interruption are transmitted. As only one ImagingEvts file per observation can be considered further, such observations with multiple ones need to be taken care of.

XMM-DH identifies such cases and informs the user about them by printing a message on the terminal. For the affected observations, the headers of the photon event tables are accessed, and the keywords `RA_PNT`, `DEC_PNT`, `FILTER` and `SUBMODE` are checked. `RA_PNT` and `DEC_PNT` correspond to the pointing coordinates during the observation, given as right ascension and declination. These coordinates indicate the expected location of the source. `FILTER` and `SUBMODE` show the names of instruments and filters used during the observation (see Section 1.2.3). If these four keywords have identical values across all individual photon event tables, it indicates that the pointing remained stable throughout the observation. Consequently, the source maintained a consistent position within the detector, and all segments of the observation were conducted under uniform camera conditions. In such a case, all

the photon event tables from the interrupted observation are merged into one. The resulting table is a sum of the original ones.

In case one or multiple keywords change in the individual photon event tables, meaning a change in the observation conditions with the interruptions, the photon event tables cannot be merged. In that case, XMM-DH identifies the observation with the longest exposure time by reading the keywords `DATE-OBS` and `DATE-END` of each table (which correspond to the start and end time of the exposure, respectively). Only the longest observation is considered further.

First-look images

Using the photon event tables, a 'first-look image' for each observation and camera is extracted. Although these images do not have a specific purpose in the processing, they are used as a first visual check of the data.

The task `evselect` from the SAS software is invoked to perform the image extraction. This task takes as input table the ImagingEvts file and the name of the columns to be used as position columns. The following represents an example call of the `evselect` task in SAS notation:

```
evselect table={file} withimageset=yes
imageset=first_look_image_{obs_id}_{camera}.fits xcolumn=X
ycolumn=Y imagebinning=imageSize ximagesize=600 yimagesize=600
```

After the creation of the first look images, XMM-DH searches into all the products to identify the MOS1, MOS2, and pn images for the same observation ID. The MOS1 and MOS2 images are merged into a single image for MOS, and this one is merged into the pn one to create the EPIC first-look image.

The image merging is done by the task `farith` from the *FTOOLS* environment, which takes as input the images to be merged given an operation such as addition in this case.

```
farith {first_look_image_{obs_id}_MOS}
{first_look_image_{obs_id}_PN} {first_look_image_{obs_id}_EPIC}
ADD clobber=yes
```

One example of a resulting EPIC image is shown in Fig. 3.3:
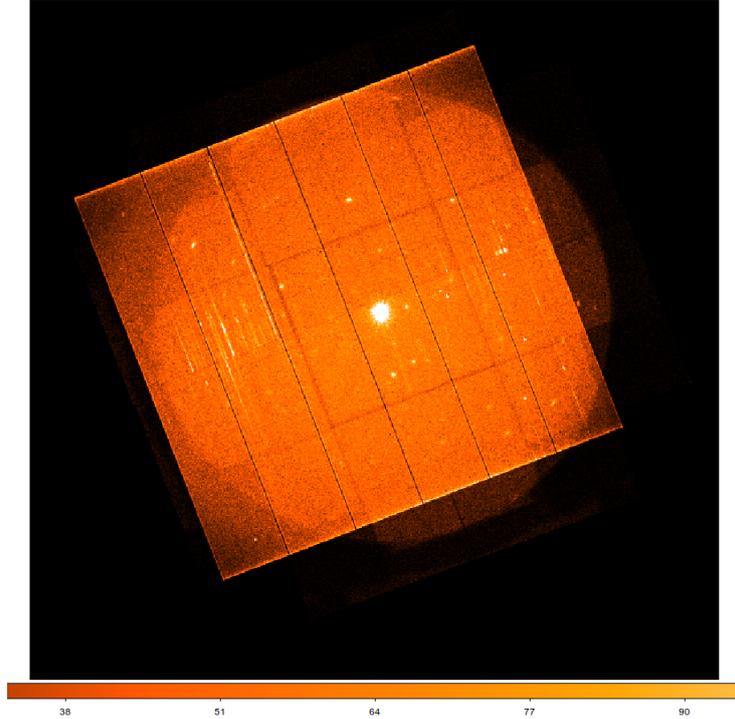
Figure 3.3: First-look image of the EPIC camera of one example observation. The bright source close to the center corresponds to the target of the observation, while the smaller ones around correspond to additional sources in the field of view. This image was created by adding the individual images from the MOS1, MOS2, and pn cameras. The pn chip design is visible.

Data cleaning

The following step in the reduction process is the identification and removal of photon events detected by the instrument in high X-ray background times. Such photons are emitted by two main sources: in the high-energy range, by the highly energetic flares from the sun (e.g. in the Van Allen belts[2]), and in the low-energy range, by the camera electronics. Such photons are removed from the photon event tables by identifying and not taking into account the time intervals when they were detected by the instrument.

To do this, the light curves (photon event table columns RATE vs. TIME) in the high energy range ($10 - 15$ keV) and low energy range ($0.1 - 0.5$ keV) are extracted for each photon event table. For each light curve, a threshold value for the RATE is defined. The time intervals where the count rate is below the given threshold are called Good Time Intervals (GTI). The GTI are considered to be dominated by the source and contain useful data. In contrast, those time intervals whose count rates are above the threshold value are considered to be dominated by external sources and are not taken into account during the analysis.

The GTI definition is carried out by the task `evselect`, defining a filtering expression and using the TIME column. The filtering expression includes the ranges

---

[2]Population of highly energetic protons and electrons trapped in the Earth's magnetic field (Li & Hudson, 2019)

for the PI column (10, 000 − 15, 0000 eV for the high energy range and 100 − 500 eV for the low energy one), and a filter for the PATTERN column to include only single and double events and for the FLAG column to avoid bad pixels.

The `evselect` task has two outputs in this case. One corresponds to the filtered photon event table, meaning the photons that meet the filtering criteria, and the other one, to the light curve plot for such photons.

The execution of the filtering process for the MOS cameras in the low energy range looks as follows:

```
evselect table={file} expression='((#XMMEA_EM && PI in [100:500])
&& (PATTERN<=4) && (FLAG == 0))'
filteredset=background_pet_{obs_id}_{camera}_lowE.fits
filtertype=expression keepfilteroutput=yes updateexposure=yes
filterexposure=yes maketimecolumn=yes timecolumn=TIME
timebinsize=100 makeratecolumn=yes withrateset=yes
rateset=background_lc_{obs_id}_{camera}_lowE.fits
```

And for the pn camera in the high energy range:

```
evselect table={file} expression='((#XMMEA_EP && PI in
[10000:15000]) && (PATTERN<=4) && (FLAG == 0))'
filteredset=background_pet_{obs_id}_{camera}_highE.fits
filtertype=expression keepfilteroutput=yes updateexposure=yes
filterexposure=yes maketimecolumn=yes timecolumn=TIME
timebinsize=100 makeratecolumn=yes withrateset=yes
rateset=background_lc_{obs_id}_{camera}_highE.fits"
```

The next step is to define the counts per second (`RATE`) threshold value for the high and low energy ranges. For each light curve, a conservative and a relaxed threshold value are defined to assess how this definition affects the final scientific result. Considering this, two threshold values are defined per energy range and camera and observation combination.

In the personalized approach to the program, the user is asked to manually introduce all the threshold values. In the automatic approach, the definition of the predetermined values is described below.

For each photon event table leading to a light curve, a histogram for the RATE values is created and the mode[3] of the sample is identified. The conservative RATE threshold value is calculated as $2.0 *$ mode and the relaxed one as $5.0 *$ mode. The threshold value definition is a trade between source counts and clean data: the conservative value includes less data with fewer background counts, and vice-versa for the relaxed one.

An example of a high and low energy light curve is shown in Fig. 3.4

---

[3]The mode is the most frequently occurring value in a set of data.
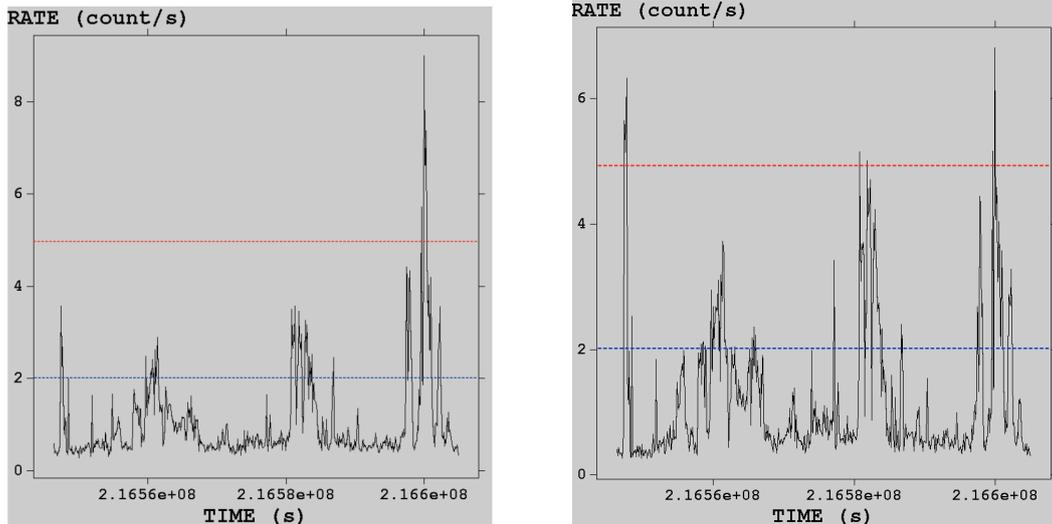
Figure 3.4: Example of a low energy light curve (left) and a high energy light curve (right). The y axis represents the count rate. The dashed blue line represents the conservative `RATE` threshold, while the red one, the relaxed threshold

Following the definition of the threshold RATE values, the execution of the task `tabgtigen` takes place. This task identifies all the time intervals where the value of the column RATE is smaller than the threshold value and stores them in a single GTI table as the one seen in Fig.3.5. The output is the conservative and relaxed GTI tables.

```
tabgtigen table={file} gtiset={obs_id}_{camera}_gti_{energy}.fits
expression='RATE<{thres_value}'
```


Figure 3.5: Example of a GTI table containing a compilation of the good time intervals during which the counts per second are below a given threshold value. The table consists of 2 columns, one for the start and one for the stop time of each interval.

Once the low and high-energy GTI tables are created separately, it is necessary to merge them to have a single GTI table suitable for the entire energy range. To do

this, XMM-DH searches among all the product files and matches every observation and camera and threshold (conservative/relaxed) combination with its corresponding high and low-energy GTI table. The GTI tables are merged using the task `gtimerge` with the argument `AND`, which takes only the intersecting GTI intervals on both tables. An example of this task execution in SAS notation is shown below:

```
gtimerge tables='{gti_lowE} {gti_highE}'
gtitable={{obs_id}_{camera}_gti_combined.fits} mergemode=and
```

The original ImagingEvts tables are then filtered such that the photons emitted outside the Good Time Intervals established in the GTI table are not taken into account. This is done by using the `evselect` task. The filtering expression includes the combined GTI table, the energy range for the EPIC camera in the interval $[200 \text{ eV} - 12,000 \text{ eV}]$, patterns less or equal to 12, and pixels flagged as 0.

The execution of the task for a MOS camera photon event table looks as below:

```
evselect table={table} withfilteredset=yes
filteredset=filtered_{obs_id}_{camera}_0.2_12keV_cons.fits
keepfilteroutput=yes destruct=yes expression='(#XMMEA_EM &&
gti({lowE}:STDGTI,TIME) && gti({highE}:STDGTI,TIME) && (PI in
[200:12000]) && (PATTERN<=12) && (FLAG == 0))'
```

Cleaned images

With the previously obtained filtered photon event tables, it is possible now to create the corresponding images, both for the relaxed and conservative thresholds. The images are created for each EPIC camera. The MOS1 and MOS2 images are merged into a MOS one, which is then merged again to the pn image to form the EPIC camera-cleaned image. A comparison of the first-look image for the EPIC camera and its corresponding cleaned image with a conservative and relaxed threshold RATE value is shown in Fig. 3.6.
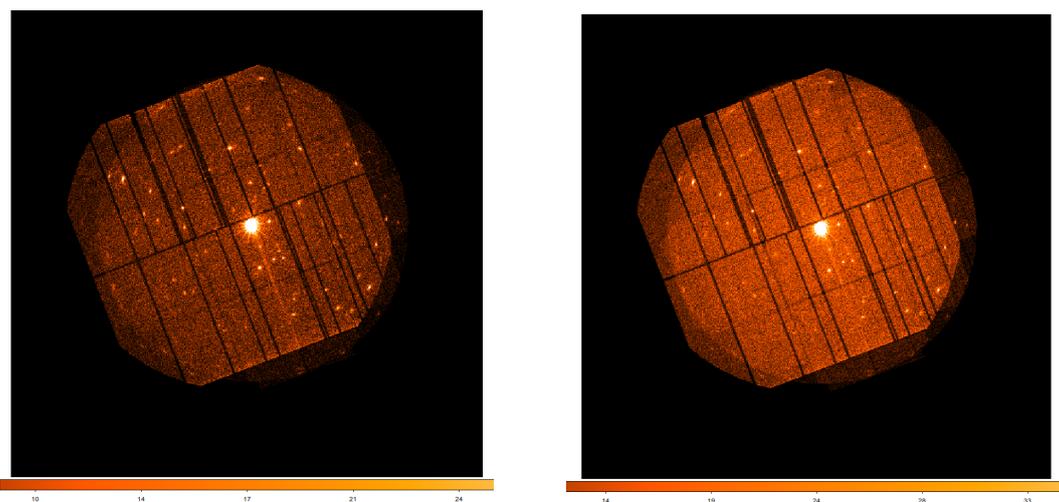
Figure 3.6: Comparison of the final image of the EPIC camera after removing high background intervals with a conservative (left) and a relaxed (right) threshold for the counts per second value. Both images use the same brightness scaling.

As seen in Fig. 3.6 and as expected, the conservative RATE threshold value removes more photons from the first-look image than the conservative one.

Only at this stage, the observations where the primary chip was turned off (and therefore no photons from the source were detected) can be identified. Those observations are not useful for the analysis and need to be discarded to avoid unnecessary computing time. To automatically identify such observations, XMM-DH accesses the header of the cleaned photon event tables and checks the keyword `LIVETIME`. If the central CCD of the camera was turned off during the observation, the keyword value is 0. Observations meeting this condition are not considered further.

In the personalized approach, the user is shown a comparison of the first-look and cleaned images with relaxed and conservative threshold RATE values, as the one shown in Fig. 3.6.

3.2.2   Time bin creation

Once the GTI and clean photon event tables are created, the preparation work for the spectra extraction can be carried out. As this thesis aims to use time-resolved spectra, it is necessary to split the previously created combined GTI tables into shorter-duration time bins. Each of these bins is also defined by a `START` and `STOP` time, allowing XMM-DH to identify the time interval for which a spectrum needs to be extracted. A single spectrum is extracted for each bin.

The GTI binning is carried out only for the pn camera GTI tables (both conservative and relaxed threshold), as this is the most sensitive of the EPIC cameras and therefore is the one that contains the most photons. The pn camera time bins resulting from this process are mirrored to the MOS1 and MOS2 cameras GTI tables. This way, the three EPIC cameras have the same number of bins in each GTI table, with identical `START` and `STOP` times.

As a result of this process, multiple sub-spectra for each EPIC camera are created, allowing the tracking of N_{H} and the flux with higher time resolution over the total exposure time of the observation.

Initialization

To initialize the second XMM-DH pipeline, responsible for time-binning each observation, the following command is executed in the terminal from the Working Directory: `python3 BINNING.py`. The pipeline asks for four initial inputs from the user:

- The Working Directory where all products are stored (same as in the data reduction pipeline).

- The target bin length in seconds

- The maximum allowed pause between GTI in the original table in seconds.

- The tolerance factor for the bin length, given as a percentage. With this value, XMM-DH creates two new variables: the lower `low_lim` and upper `up_lim` limit for the bin duration.

Time bin definition

XMM-DH starts with the calculation of the duration of each GTI, as well as the pause between two consecutive ones. The duration of the `i` GTI is calculated as `STOP[i]-START[i]`. The pause between the `i` and `i+1` GTI is calculated as `START[i+1]-STOP[i]`. For the last interval of the table, the `pause` is `None`. A new four-column table is created with the `START`, `STOP`, `duration` and `pause` for each GTI interval. The red line in Fig. 3.7 shows a visual representation of how the original GTI intervals look like.

Based on the previous table, XMM-DH creates new intervals by progressively merging every GTI with the one immediately following it, until a pause between GTI longer than the maximum allowed pause is encountered. Every GTI merged adds up to the duration of the new long interval.

The `START` time of this new long interval corresponds to the `START` time of the first GTI merged into it. The `STOP` time of the new interval corresponds to the `STOP` time of the last GTI merged before a pause longer than allowed was encountered. The next long interval starts at the `START` time of the next GTI after the long pause. Each of these new long intervals (called a group hereafter) is enumerated with an index, as well as the GTIs within it. The yellow line in Fig. 3.7 shows a graphical representation of how the long intervals are created.

Each group has two durations associated: the first one, `duration_data`, corresponds to the amount of data per group, meaning the cumulative sum of the duration of each one of the original GTI intervals within the group. The second one, `duration_with_hole`, corresponds to the actual duration of the group, including the time intervals with no data. This duration is calculated as the `STOP` minus the `START` time of each group.

Those groups whose `duration_data` is shorter than the initially defined target bin length are not considered further, as they do not contain enough data. The group index is then switched as shown in the green line in Fig. 3.7, from the yellow to the green line.

The next step is splitting each of the new long intervals into bins of the optimal duration defined before. However, as each of the long intervals has a different duration, being strict about this definition may cause a significant loss in the amount of data. Therefore the bin duration is allowed to vary within the range `low_lim<bin_duration<up_lim`. To find the optimal bin duration for each long interval with the least amount of data loss, XMM-DH uses the task `fminbound` on the package `optimize` from the library SciPy in Python. The task finds the optimal bin duration `bin_opt` between the values `x1=low_lim` and `x2=up_lim` that minimizes the function `y = duration_data%x` (in Python notation). This function represents `y` as the remainder of the division of the amount of data per long interval into the number of bins `x`.

Once the value `bin_opt` has been calculated for each group, XMM-DH will start joining each of the original GTI forming it until it has formed a bin with a data duration of `bin_opt`. It is important to note here that an optimal data duration of, for example, 5200 s, does not mean that the subtraction of the `STOP` minus `START` time of the bin is 5200 s, but rather that the data contained in the bin add up to 5200 s. This is because each group has time intervals of no data that do not contribute to the duration of the bin.

To create the optimal-duration bins, XMM-DH starts iterating over each one of the original `i` GTI forming a group. The variable `gti_start` represents the `START` time of each bin, and it is initialized as the `START` time of the first GTI in the group. The following step is to find the `STOP` time of the bin, `gti_stop`, such that the bin data duration is equal to `bin_opt`. To do so, the counter `sum_duration` is initialized to zero, and its value is updated and assessed on each iteration. The value of this variable in the iteration `i` corresponds to its value from the previous iteration `i-1` plus the `duration` of the GTI in the current iteration `i`, i.e. `sum_duration += duration`.

One of three possible scenarios is presented on each iteration `i`:

- If `sum_duration<bin_opt`, then `gti_stop` keeps its `None` value as the bin is still too short. XMM-DH needs to keep adding GTI to get to the optimal duration and find the `gti_stop` for the bin to be created.

- If `sum_duration==bin_opt`, the bin has reached its optimal data duration. In this case, `gti_stop` is set as the `STOP` time of the `i` GTI, and a new bin is created. The variable `gti_start` is updated as the `START` time of the following GTI, and `gti_stop` takes a `None` value again.

- If `sum_duration>bin_opt`, the bin has reached its optimal duration at some point during the bin-merging but has exceeded it. In that case, the value of `sum_duration` in the previous iteration, without adding the duration of the current GTI, is retrieved. With this value on hand, XMM-DH calculates the time duration missing to the bin to reach the value `bin_opt`. That missing time is taken from the current `i` GTI so that `gti_stop` corresponds to its `START` time plus the missing time. This creates a new bin, and `gti_start` is updated as the `gti_stop` of the previous bin, while updating `gti_stop` as `None`.

A representation of the final bins is shown as the blue line in Fig. 3.7.
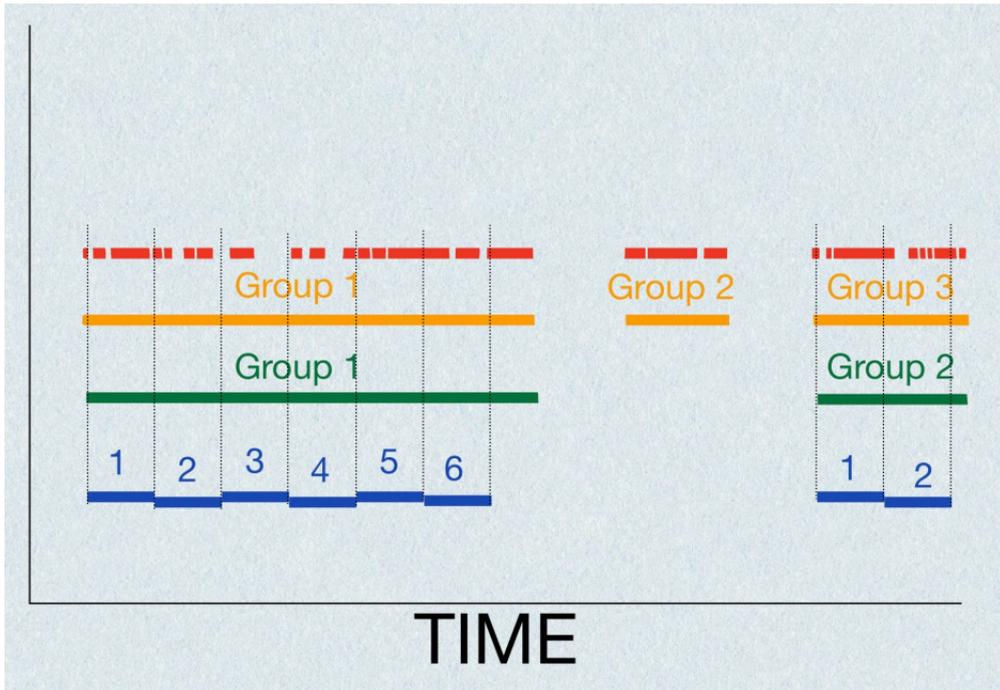
Figure 3.7: Graphical representation of the time binning for each observation. The red line represents the original pn-camera GTI obtained as those intervals where the counts per second are below a given threshold. The spaces between red lines represent the time intervals where the counts per second exceeded the RATE threshold value. The yellow lines represent every group formed by merging consecutive GTI until a pause longer than allowed is encountered. The green line represents the remaining groups after removing groups with a short duration, not enough to form at least one bin. The blue lines represent the final equally-sized time bins resulting from the splitting of every group in green. These blue lines represent only the `START` and `STOP` time of each final bin. The actual data contained in each of them corresponds to the original GTI (red line) in that same interval, marked by the vertical dotted lines. The accumulated duration of the blue bins is shorter than in the red, yellow, and green bins, representing the data lost by the remainder of the bin creation process. An offset is added to consecutive blue lines for visualization purposes. The x axis represents time, while the y axis is only added for visualization purposes.

All the `gti_start` and `gti_stop` times retrieved from the iterations correspond to the `START` and `STOP` times for the new bins, respectively. These time values form a 2-column table saved in 'fits' format. As this table requires the same structure as an original GTI table so SAS can interpret it, the header is modified to include the necessary keywords.

47

3.2.3   Spectra extraction

With all the data products available to this point, it is possible to start the process of spectra extraction.

Initialization

To initialize the third XMM-DH pipeline, responsible for the spectra extraction, the following command is executed in the terminal from the Working Directory: `python3 SPECTRA.py`. The pipeline asks the user to input the path to the Working Directory.

Region definition

The first step before extracting the X-ray spectra is defining the regions to be used as source (+ background) and background (source free). To do this, XMM-DH opens DS9[4] to show the user the clean images obtained in the data reduction (3.2.1) for each of the EPIC cameras.

Using these images, the user must utilize the DS9 region tools to draw and define two types of regions:

- The **source (+background) region** corresponds to the location of the source, which also contains background counts. This is usually defined as a single circle enclosing the source, although alternatives such as annulus are also supported.

- The **background (source free) region** is visually defined as multiple circles around the source and in the same chip, avoiding extra sources in the field of view.

See Section 3.4 for an example of the source (+ background) and background (source free) regions definition for the MOS and pn cameras.

The region definition process creates a file containing the coordinates of all the regions defined above and must be saved in text file format (.txt extension).

Epatplot creation

The epatplot contains information about the EPIC cameras' event pattern statistics. An example of an epatplot is shown in Fig. 3.8. This plot has two panels: the upper one shows the number of photons collected across PI channels, distributed in different event patterns (single, double, triple, quadruple). The bottom panel shows the fraction of photons relative to the total across the PI channels, comparing the data to the expectation models for the single, double, and triple events fractions.

---

[4]DS9 (short for SAOImageDS9) is an astronomical imaging and data visualization application
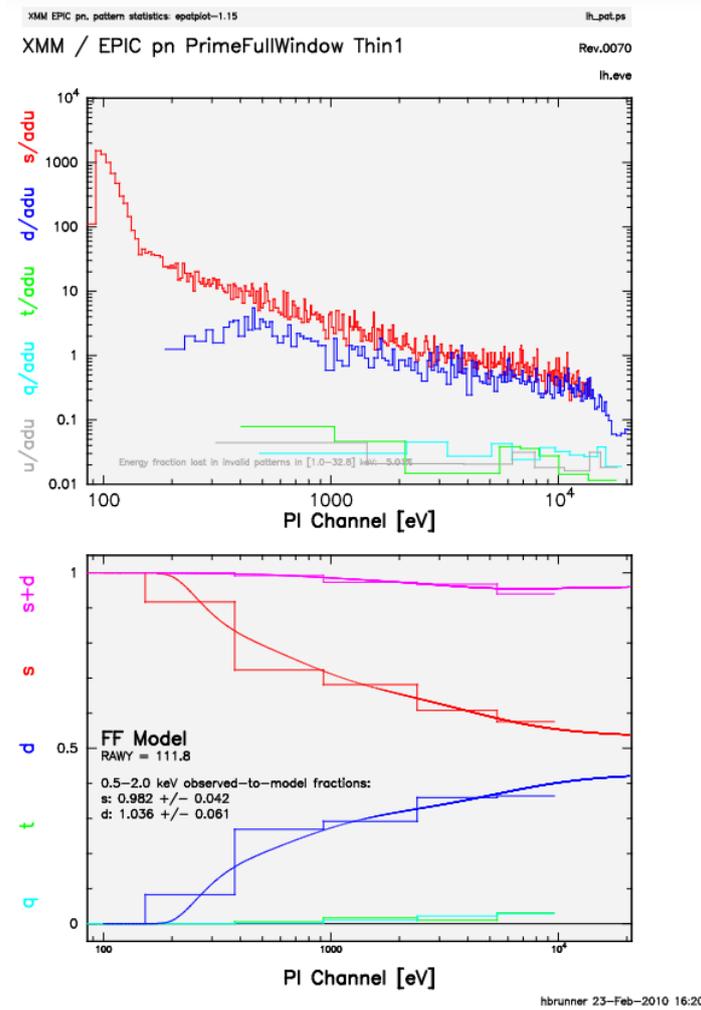
Figure 3.8: Example of an epatplot. The red data corresponds to single events, blue to doubles, green to triples, cyan to quadruples, and gray to higher patterns. The lines show the expected distribution of patterns as a function of energy. The magenta line in the bottom panel represents the sum of single and double events (XMM-Newton, 2023).

The epatplot is used to identify two main effects during the observation: pile up and/or optical and X-ray loading (XMM-Newton, 2023).

- **Pile-up:** This phenomenon affects all the EPIC cameras and occurs when the read-out mode of the camera is inefficient (too slow) for a very bright X-ray source. This phenomenon is observed in two forms:

    – **Pattern pile-up:** Occurs when two or more photons arrive at two or more adjacent pixels during the same readout cycle of the CCD. The electronics of the detector are not able to differentiate the occurrence of multiple events in a single exposure, therefore erroneously combining all of them into a single one of a higher pattern. This is translated into the spectrum as a pattern migration. There is an excess of higher pattern events compared to the expected pattern distribution.

– **Energy pile-up:** Occurs when multiple photons arrive at the same pixel during a read-out cycle. As the electronics cannot distinguish different photons, all of them are erroneously read as a single one of higher energy. This causes an energy distortion in the X-ray spectrum, as photons are moved to higher energies.

Pile-up can be identified in the bottom panel of the epatplot when single events fall below the expected model distribution and double events rise above it. To mitigate this effect in the affected observations, the source region should be adjusted. Specifically, defining the source region as an annulus rather than a circle excludes the central area where higher counts per second are detected. An example of an epatplot showing a piled-up observation is shown in Fig. 3.9.
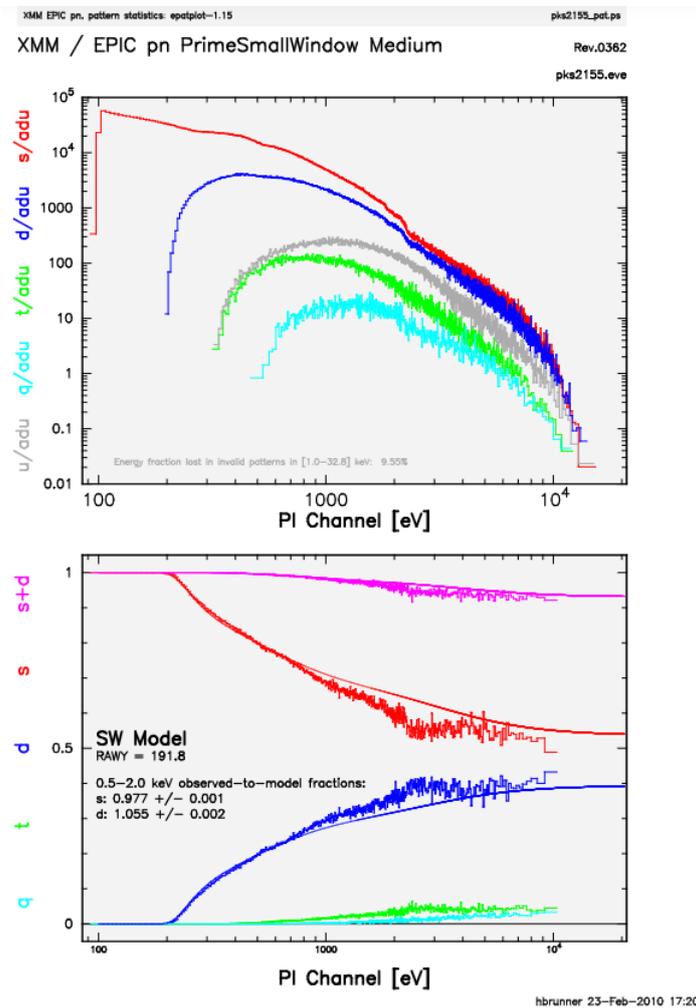
Figure 3.9: Example of an epatplot showing a piled-up observation. The bottom panel shows that the single events detected for the source (red) are below the model distribution (solid red line), while the doubles (blue) are above it (blue solid line). This is expected as a result of a pattern migration XMM-Newton (2023). For more information, see the caption of Fig. 3.8.

- **Optical and X-ray loading:** At the beginning of each exposure, an offset map is calculated onboard to restore the energy scale. This map contains the highest and lowest energy values detected in each pixel, later subtracted from each event hitting that pixel. However, the offset map can be distorted by two main reasons:

  - **Bright sources:** When there is an excessive count rate detected during the offset map calculation, X-ray events are included in the pn offset map. As a result, the counts in the resulting map are so high that part of the source energy is removed when applying it (XMM-Newton Calibration Technical Note, 2013).
  - **Extra source in the field of view:** Any additional source within a 2 arcmin distance from the target contributes optical light to the observation. This causes a uniformity in the offset map, implying an additional value subtracted from the signal.

X-ray loading creates a pattern migration from double to single events and an energy shift from higher to lower energies. An epatplot showing an observation affected by X-ray loading is shown in Fig 3.10.
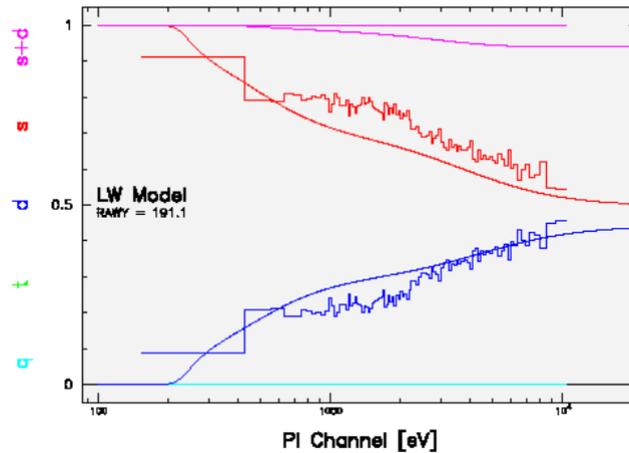
Figure 3.10: Example of the bottom panel of an epatplot showing an observation affected by X-ray loading. The single events detected for the source (red) are above the model distribution (solid red line) implying an excess, while the doubles (blue) are below it (blue solid line). This behavior of the epatplot is expected from a pattern migration (XMM-Newton Calibration Technical Note, 2013). For more information, see the caption of Fig. 3.8.

To create the epatplot, *XMM-Newton SAS* requires the event table containing only those photons emitted by the region of interest in the time intervals of interest. In this case, those correspond to the source region in the GTI, therefore, the cleaned photon event tables obtained from the data reduction are filtered using `evselect`. In this case, the filtering expression includes criteria for the position, time, energy, and pattern of the photons. An example of the execution of this task in SAS notation is shown below:

MOS event tables are filtered to include single, double, triple, and quadruple events.

```
evselect table={table} withfilteredset=true
filteredset=filtered_source_region.fits keepfilteroutput=true
destruct=true expression='(((X,Y) IN {source}) &&
GTI({gti}:STDGTI,TIME) && (PI in [200:12000]) && (PATTERN <=
12))'
```

PN includes only single and double events.

```
evselect table={table} withfilteredset=true
filteredset=filtered_source_region.fits keepfilteroutput=true
destruct=true expression='(((X,Y) IN {source}) &&
GTI({gti}:STDGTI,TIME) && (PI in [200:12000]) && (PATTERN <= 4))'
```

The resulting photon event table from the `evselect` task is used by the task `epatplot` as:

```
epatplot set={filtered_source_region} plotfile=epatplot.ps
```

Once the epatplot for each observation and camera and rate-threshold combination is created, it is analyzed to identify pile-up or X-ray loading, as well as to identify the energy range where the data can be used reliably (more about this in Section 3.3).

Pattern filtering

After correcting the observation for any effects, XMM-DH uses `evselect` to filter the photon event tables to include only the patterns of interest to extract the spectra. For the EPIC MOS cameras, single, double, triple, and quadruple events (`PATTERN <= 12`) are included in the same spectrum, while for EPIC pn only single (`PATTERN == 0`) and double events (`(PATTERN > 0) && (PATTERN <= 4)`) are taken into account into two separate spectra. This separation of patterns for the pn camera is done to account for the higher energy sensitivity of single events.

The pattern definition is done for the MOS camera as:

```
evselect table={file} withfilteredset=true
filteredset=filtered_{obs_id}_{camera}_{en_pet}_pat12.fits
keepfilteroutput=true destruct=true expression='(PATTERN <= 12)
&& (FLAG == 0)'
```

For pn single events as:

```
evselect table={file} withfilteredset=true
filteredset=filtered_{obs_id}_{camera}_0.2_12keV_pat0.fits
keepfilteroutput=true destruct=true expression='(PATTERN == 0) &&
(FLAG == 0)'
```

And for pn double events as:

```
evselect table={file} withfilteredset=true
filteredset=filtered_{obs_id}_{camera}_0.2_12keV_pat14.fits
keepfilteroutput=true destruct=true expression='((PATTERN > 0) &&
(PATTERN <= 4)) && (FLAG == 0)'
```

Average and single spectra extraction

Having filtered the photon event tables, the next step is extracting both the average spectrum for each observation and camera and pattern combination, as well as the single spectrum for each of its time bins. This process is carried out by the task `especget`, which takes as an input the time and space filtering expression for the source and background region and the photon event table from which the spectra are extracted.

The expression to filter the source region for the average spectrum extraction is defined as:

```
{(X,Y) IN {source_region})} && GTI({combined_gti}:STDGTI,TIME)
```

where `source_region` corresponds to the spatial coordinates of the source region, previously defined using DS9 and `combined_gti` corresponds to the GTI table for each observation. The same procedure is followed for the background region expression.

For the single spectrum extraction, the source and background region expression include a filter for the time interval to be used for the extraction. These time intervals correspond to the `start` and `stop` time of every bin obtained in the time binning process (see Section 3.2.2). In this case, the source expression is given as:

```
{source_exp} && (TIME >= {start}) && (TIME <= {stop})
```

The average and single spectra are then then extracted as:

```
especget filestem=av_spec_{obs_id}_{camera}_{en}_{pattern}
srcexp='{source_exp}' backexp='{background_exp}' table={table}
```

Spectra grouping

Once the spectra are extracted, the last step is to group their energy channels. This is done by the task `specgroup`. This task takes as input the source spectrum obtained from the `especget` task, the minimum number of counts per bin, set to 20 in this case, and the bad channels to ignore, defined as $[0:14, 799:800]$ for MOS and $[0:38, 2400:4096]$ for pn. This process is done for each source's average and single spectra.

```
specgroup spectrumset={MOS_spectra} groupedset={grouped_spectra}
mincounts=20 setbad='0:38,2400:4096'
```

```
specgroup spectrumset={pn_spectra} groupedset={grouped_spectra}
mincounts=20 setbad='0:14,799:800'
```

3.3 *XMM-Newton* spectra fitting

To carry out the average and single spectra fitting and time-resolve the value of N_{H} and Γ during each observation to create their time evolution curve (steps 4 to 5 mentioned in Chapter 2), another software (XMM-fit hereafter) was developed. XMM-fit uses PyXspec, the Python module for Xspec, which is an X-ray spectral-fitting program.

The previously extracted spectra require a fitting model. In this thesis, the best-fit spectral models for the sample sources are obtained from Laha et al. (2020). Each source is associated with a single spectral model, which is used by XMM-fit to fit the average spectrum for each observation and determine the best-fit parameter values for that specific observation. These parameter values are subsequently applied to fit the individual spectra extracted for each time bin within an observation.

XMM-fit is developed to analyze one source at a time taking as input the path where the single and averaged grouped spectra are located, the table with the energy ranges to be considered for each observation, and the spectra fitting model for each source. Each of these inputs is defined in this section.

Definition of fitting conditions for average spectrum

The first step carried out by XMM-fit is defining the average spectra set for each observation corresponding to the extracted source spectra for MOS1, MOS2, pn pattern 0, and pn patterns $1 - 4$. This spectral data set is then loaded into PyXspec, one spectrum per data group.

The average spectra data set, in PyXspec notation, is defined as:

```
xspec.AllData(1:1 MOS1_spectrum 2:2 MOS2_spectrum 3:3
pn_pat0_spectrum 4:4 pn_pat1-4_spectrum)
```

The group 1 and 2 correspond to the MOS1 and MOS2 spectrum, and the third and fourth to pn pattern 0 and pn pattern $1 - 4$, respectively.

Additionally, bad energy channels are ignored by setting:

```
xspec.AllData.ignore('bad')
```

These channels correspond to those with quality flags of 1 or 5 (see Section 3.2.1).

The next step is to define the reliable energy ranges to be used for each observation and camera and pattern combination. This is done by analyzing the epatplot (see Section 3.2.3) and identifying the energy range where the expected model distribution well fits the data.

For example, for the MOS1 epatplot shown in Figure 3.11, the energy range chosen is $0.2 - 6.0$ keV, as the bottom panel shows that this is the range where both single and double events (red and blue data, respectively) are well fit by the expected model distribution (red and blue lines for single and double events, respectively).
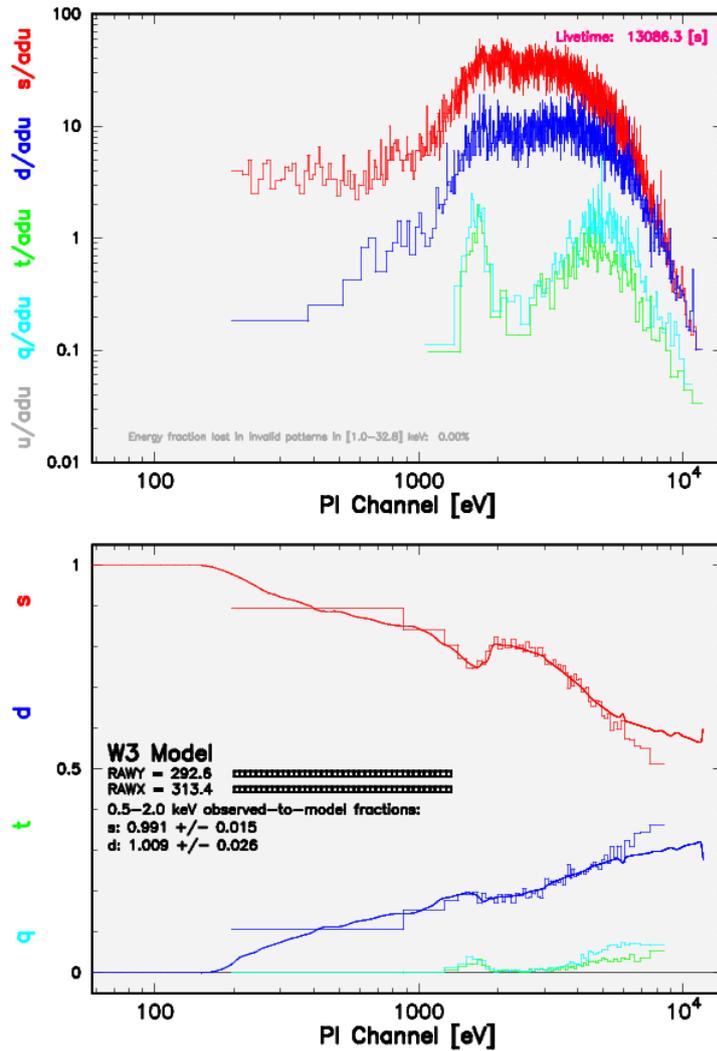
Figure 3.11: Example of an epatplot for MOS1

The same process is followed for the epatplots corresponding to the pn camera.

The energy range definition process is done specially for the pn camera, as it presents a widely known issue in the soft energy range when the Small Window Mode is used. An example epatplot of an observation affected by such an issue is shown in Figure 3.12.
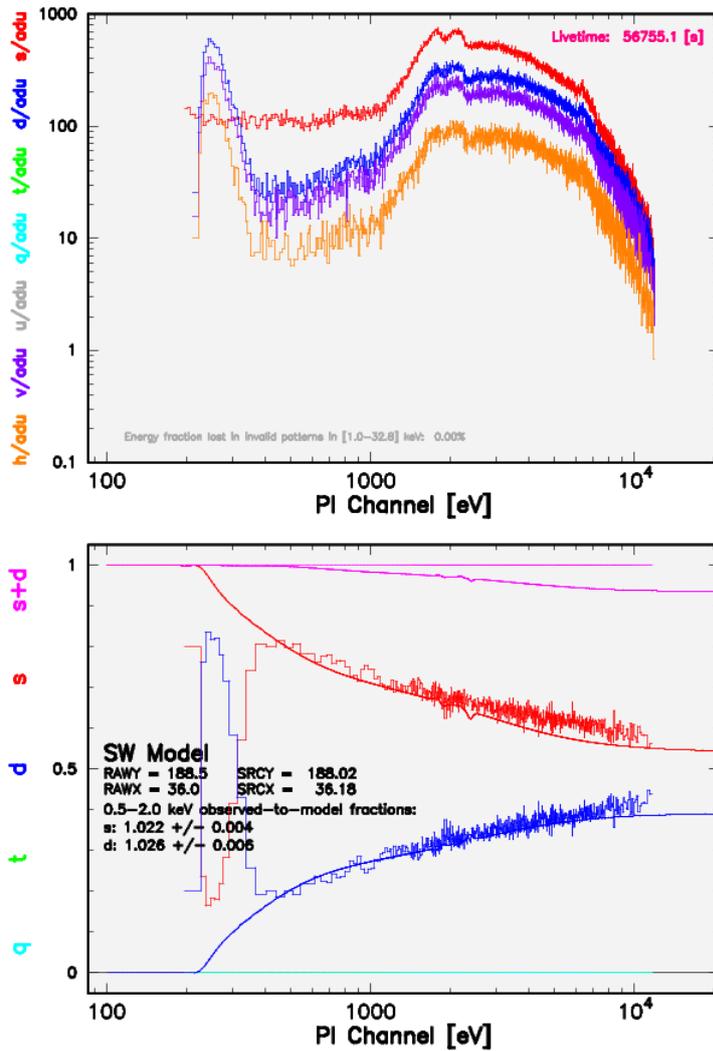
Figure 3.12: Example of a pn camera epatplot for an observation affected by the issue in the soft energy range when using Small Window Mode. For more information, see the caption of Fig. 3.8.

After consultation with *XMM-Newton* pn calibration specialists (private consultation with Frank Haberl), it was established that the issue with the pn camera affects only double but not single events. Because of this, for the epatplot shown in Fig. 3.12, it is safe to assume an energy range of $0.2 - 10.0$ keV for single events, but restrict it to $0.45 - 8.0$ keV for double events. The energy ranges must be defined in a .csv table, which is read by the XMM-fit software. XMM-fit takes the lower and upper boundary of the energy range for each camera and ignores events whose energy is below the lower boundary or above the upper one.

```
xspec.AllData(1).ignore(**-{low_mos1} {up_mos1}-**)
xspec.AllData(2).ignore(**-{low_mos2} {up_mos2}-**)
xspec.AllData(3).ignore(**-{low_pn_0} {up_pn_0}-**)
xspec.AllData(4).ignore(**-{low_pn_14} {up_pn_14}-**)
```

Spectral model definition

In the following step, the spectra-fitting model for each source and the parameter values must be manually added. The former corresponds to a combination of components modeling different physical processes contributing to the X-ray spectrum (e.g. see Table 3.2). The latter corresponds to a 6-element list including the parameter's initial value, the step size for the best-fit value search, and the strict and relaxed lower and upper boundary for the best-fit value.

In this case, as the four data groups from a single average spectra data set correspond to the same source observed under identical conditions, they all use the same spectral model. The parameter values for all data groups are linked, allowing them to vary simultaneously and proportionally during the fit.

Parameters linked to each other and frozen during the fitting are defined at this stage. The spectra model is used to fit the average spectrum of each observation and obtain the parameter values to fix during the fitting of single spectra.

Spectral models for the sample sources

The spectrum model provided by Laha et al. (2020), used to fit the average X-ray spectrum of every observation of a source is shown in Table 3.2.

Source	Spectrum model Laha et al. (2020)
IRAS F00521-7054	`constant*(TBabs*(apec+powerlaw(1)+zTBabs*zpcfabs*(pexmon+powerlaw(2)+zgauss)))`
MCG-5-23-16	`constant*(TBabs*(apec+powerlaw(1)+zTBabs*(pexmon+powerlaw(2)+diskline)))`
Mkn 348	`constant*(TBabs*(apec+apec+powerlaw(1)+zpcfabs*zTBabs*(zgauss+pexmon+powerlaw(2))))`
NGC 2110	`constant*(TBabs*(apec+powerlaw(1)+zTBabs*zpcfabs*(pexmon+powerlaw(2)+zgauss)))`
NGC 4258	`constant*(TBabs*(apec+apec+powerlaw(1)+zTBabs*(zgauss+pexmon+powerlaw(2))))`
NGC 4507	`constant*(TBabs*(apec+apec+powerlaw(1)+zgauss+zTBabs*zpcfabs*(zgauss+pexmon+powerlaw(2))))`
NGC 5252	`constant*(TBabs*(apec+apec+powerlaw(1)+zpcfabs*zTBabs*(zgauss+pexmon+powerlaw(2))))`
NGC 5506	`constant*(TBabs*(zTBabs*(apec+powerlaw(1))+zgauss+zTBabs*(pexmon+powerlaw(2)+zgauss+zgauss)))`
NGC 6300	`constant*(TBabs*(apec+powerlaw(1)+zTBabs*(zgauss+pexrav)))`
NGC 7172	`constant*(TBabs*(powerlaw(1)+apec+zTBabs*zpcfabs*(zgauss+pexmon+powerlaw(2))))`

Table 3.2: Models used to fit average spectra of sources in the sample. These best fit models are taken from Laha et al. (2020).

An overview of the components used in these spectrum models, taken from the *Xspec* manual (Xspec. An X-Ray Spectral Fitting Package, 2024), follows:

- `constant`: Multiplicative component that accounts for the differences between the spectra obtained from MOS1, MOS2, and pn cameras.

- `TBabs`: Multiplicative component that affects the rest of the spectra model, as it accounts for the galactic absorption, i.e., due to the material from the Milky Way in the line of sight. This value is different for each source and was taken from Laha et al. (2020).

- `zTBabs`: Models the redshifted X-ray absorption, for example, due to the AGN's host galaxy or the torus.

- `zpcfabs`: This component models the partial covering absorption at a given redshift.

- `apec`: An additive component that models any thermal emission, for example from collisionally-ionized diffuse gas.

- `powerlaw(1)`: Additive component modeling the secondary soft power-law due to scattered emission.

- `powerlaw(2)`: Additive component modeling the primary hard power-law modeling the inverse Compton-scattered emission from the hot corona.

- `pexrav`: Exponentially cut off power law spectrum reflected from neutral material. It models the Compton reflection hump.

- `pexmon`: This is a combination of the component `pexrav` and the Fe K line.

- `zgauss`: A redshifted gaussian line profile. It describes the higher ionization Fe emission lines.

- `diskline`: Models the line emission from a relativistic accretion disk.

Average spectrum fit performance

Once the spectral model was defined, the appropriate statistics for fitting the data was defined. Given the high number of data points available in each X-ray spectrum of this thesis, Chi-square statistic was used as it provides an appropriate approach for such datasets. In contrast, Cash statistics would be more appropriate for spectra with lower signal-to-noise ratios.

Having defined the initial conditions, XMM-fit fits the average spectrum of each observation to the spectral model of the source.

```
xspec.Fit.statMethod = "chi"
xspec.Fit.perform()
```

Once this is done, the best-fit parameter values obtained for the average spectrum are frozen, i.e. defined to not change during upcoming fits.

The upcoming part in XMM-fit corresponds to the time resolution intended for this thesis to track the change of N_{H} during each observation.

Single spectra fitting

The first step of the time-resolution process is re-defining the spectral set in PyXspec as the MOS1, MOS2, pn pattern 0, and pn pattern $1-4$ spectra for each time bin in an observation. This includes accounting for observations where one of the cameras did not collect data, or where a spectrum in a time interval with no data was extracted. Then, using the same definition as for the average spectra, the data groups' energy ranges are defined and the bad pixels are ignored.

For each new spectral set corresponding to a single time bin, two fits are performed:

1. The first fit (First Fit hereafter) unfreezes (i.e. allows the parameters to vary during the fit) the parameters corresponding to the N_H from the intrinsic absorption component and the normalization of the primary power law. This fit is performed to track the change of N_H within observations, assuming that this is the only parameter that modifies the shape of the spectrum over time.

2. The second fit (Second Fit hereafter) unfreezes the parameters corresponding to the N_H from the intrinsic absorption component and the Γ and normalization of the primary power law. This fit assumes both N_H and Γ as causes of change in the spectrum's shape, considering the degeneracy between those parameters.

For each fit, the uncertainty of the best-fit values for the fitted parameters (N_H and Γ) are calculated at a confidence level of 90% for one parameter, corresponding to 1.646σ or $\Delta\chi^2 = 2.71$. This is the standard for X-ray astronomy. This process is done as:

```
xspec.Fit.error('2.706 n')
```

where `n` corresponds to the index value of a given parameter in the spectra model.

Additionally, for each fit the reduced χ^2 is calculated as $\chi^2/\text{d.o.f}$, where χ^2 is calculated by PyXspec as `xspec.Fit.statistic` and the degrees of freedom as `xspec.Fit.dof`.

Flux and Hardness Ratio calculation

For the First Fit only, the total, soft and hard X-ray flux in the energy bands $0.2 - 10.0$ keV, $0.2 - 2.0$ keV and $2.0 - 10.0$ keV, respectively, are calculated. This and the uncertainty calculation are done as:

```
xspec.AllModels.calcFlux('0.2 10.0 err 20 90')
xspec.AllModels.calcFlux('0.2 2.0 err 20 90')
xspec.AllModels.calcFlux('2.0 10.0 err 20 90')
```

With the flux values in the soft and hard energy bands for each time bin, the Hardness Ratio is calculated as:

$$\mathrm{HR} = \frac{\mathrm{H} - \mathrm{S}}{\mathrm{H} + \mathrm{S}}$$
$$\mathrm{HR} = \frac{F_{2-10} - F_{0.2-2}}{F_{2-10} + F_{0.2-2}} \tag{3.1}$$

The uncertainty of the HR value for each time bin is calculated as:

$$\sigma_{\mathrm{HR}}^2 \simeq \sigma_{\mathrm{H}}^2 \left(\frac{\partial \mathrm{HR}}{\partial \mathrm{H}}\right)^2 + \sigma_{\mathrm{S}}^2 \left(\frac{\partial \mathrm{HR}}{\partial \mathrm{S}}\right)^2$$
$$\sigma_{\mathrm{HR}} \simeq \left(\frac{4}{(\mathrm{H} + \mathrm{S})^4} \left(\sigma_{\mathrm{H}}^2 \mathrm{S}^2 + \sigma_{\mathrm{S}}^2 \mathrm{H}^2\right)\right)^{1/2} \tag{3.2}$$

Once the fitting process is finished, XMM-fit concatenates into time-evolution curves the best-fit values of N_{H} ($N_{\mathrm{H}}(t)$) and Γ for the First and Second Fit, the reduced χ^2, the total, soft and hard X-ray fluxes, and the HR for each time bin of a single source. The resulting time-evolution curves are presented in Chapter 4.

3.4   Software execution for sample sources

This section explains the procedure followed to implement the previously described methodology (Section 3.2 and Section 3.3) to obtain the scientific results (mentioned in Chapter 2 and shown in Chapter 4) for the sample of this thesis (Section 3.1).

The reduction process for the sample sources mentioned in Table 3.1 was done by XMM-DH, using the default values stated in Section 3.2.1: the conservative `RATE` threshold factor of 2.0, and the relaxed one of 5.0. Consecutively, all observations were time-binned (Section 3.2.2) using a target bin length of 5000 s, a maximum allowed pause of half of the target bin length, i.e 2500 s, and a tolerance of 20%. The final time bin length for each observation varies depending on the observation duration and GTIs.

An example of the visualization of the time-binning process for observation 0152940101 of the source NGC 5252 is shown in Figure 3.13
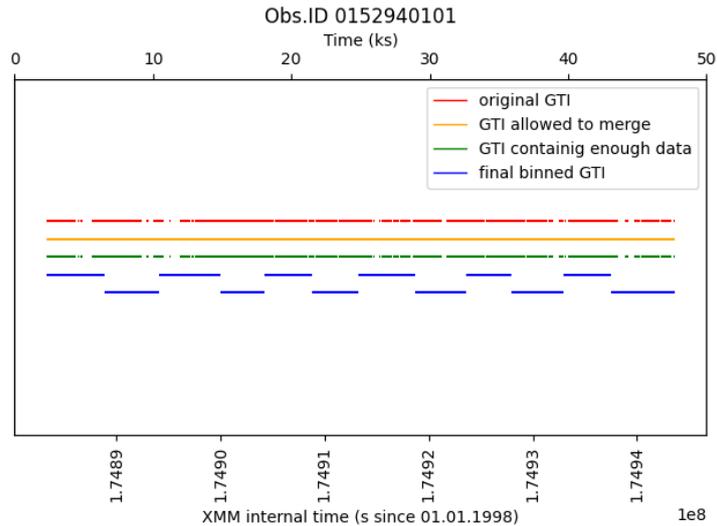
Figure 3.13: Visual representation of the time binning of observation 0152940101. The red line represents the original GTI obtained as those intervals where the counts per second are below a given threshold. The spaces between red lines represent the time intervals where the counts per second exceeded the RATE threshold value. The yellow lines represent every group formed by merging consecutive GTI until a pause longer than allowed is encountered. The green line represents the remaining groups after removing those with a short duration. The blue lines represent the final equally-sized time bins, which result from splitting every group in green. An offset is added to consecutive blue lines for visualization purposes. The x axes represent time evolution during the observation and during *XMM-Newton* time, while the y axis is only added for visualization purposes.

The average and single spectra for each observation and time bin were extracted (Section 3.2.3). An example of the source (+ background) and background (source free) regions definition used for the spectra extraction process of observation 0152940101 of NGC 5252 is shown in Fig. 3.14.
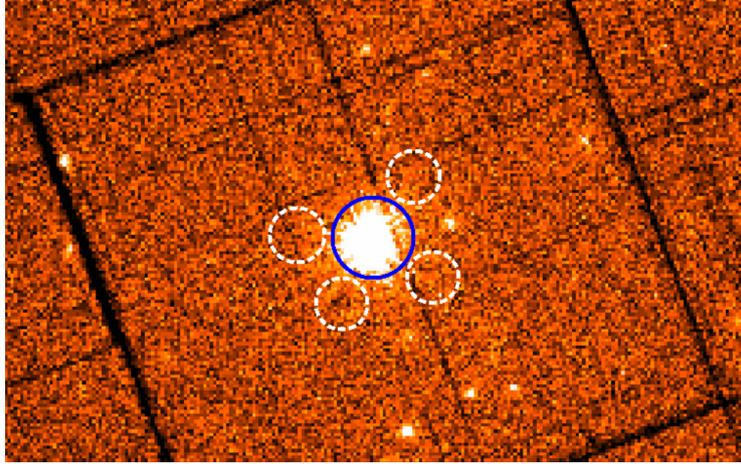
(a) Combined MOS camera image
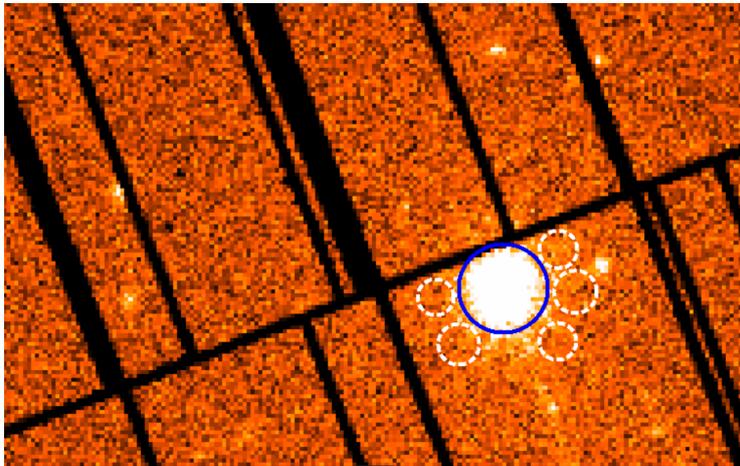
(b) pn camera image

Figure 3.14: Source (blue line) and background (white dashed lines) regions defined for observation 0152940101 of NGC 5252. The images correspond to a zoom-in of a visualization in DS9 of the cleaned images.

The previously extracted average and single spectra were fitted using XMM-fit (Section 3.3). The average spectrum extracted for single events from the EPIC-pn camera in observation 0152940101 of NGC 5252, as well as the Laha et al. (2020) best-fit model, are shown in Figure 3.15.
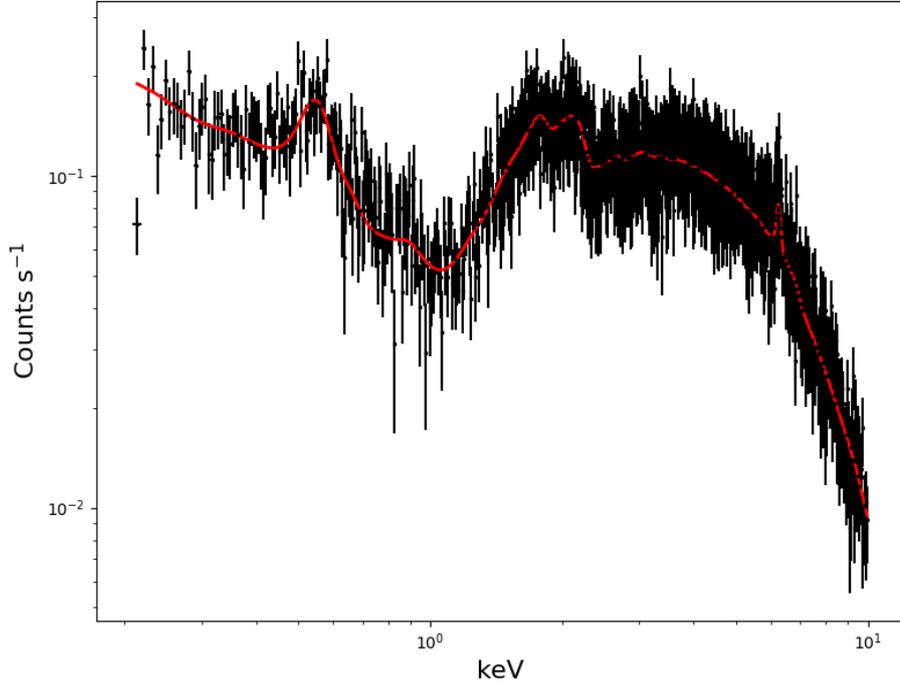
Figure 3.15: Average spectrum for single events from the EPIC-pn camera in observation 0152940101 of NGC 5252, in the energy range $0.2 - 10.0$ keV (black) and best-fit model for the count rate (red).

With the outputs from the spectra fitting procedure, the time evolution curves of the hydrogen column density, the photon index, and the flux were created.

The resulting $N_{\mathrm{H}}(t)$ curve for each source was inspected. For sources where the overall total uncertainty of each data point (i.e the sum of its upper and lower bound of the uncertainty) was greater than 10% of the average N_{H} of the corresponding observation, the observations were re-binned. The re-binning process consisted of increasing the target bin length to increase the signal-to-noise ratio value of the fit in each observation.

The choice of the new time bin length was done under the following procedure:

$$\sigma_{\mathrm{new}} = \alpha \cdot \sigma_{\mathrm{old}} \tag{3.3}$$

where σ_{new} and σ_{old} correspond to the total uncertainty of the N_{H} value with the new and previous time-bin, and α represents the conversion factor from one to the other.

As the uncertainty can be expressed in terms of the signal-to-noise ratio (SNR) as $\sigma = 1/\mathrm{SNR}$, and the SNR in terms of the signal as $\mathrm{SNR} = \sqrt{s}$, equation 3.3 becomes:

$$\frac{1}{\text{SNR}_{\text{new}}} = \alpha \cdot \frac{1}{\text{SNR}_{\text{old}}}$$
$$s_{\text{new}} = \frac{1}{\alpha^2} \cdot s_{\text{old}} \tag{3.4}$$

The above means that, for example, if the uncertainty of the N_{H} value obtained from a time bin of length $s_{\text{old}} = 5$ ks wanted to be reduced in half, i.e $\alpha = 1/2$, the new time bin should have a length of $s_{\text{new}} = 4 * 5$ ks $= 20$ ks, a factor of 4 larger than the initial bin.

As a result of the re-binning process, some observations are not shown in the final $N_{\text{H}}(t)$ curve, as they do not contain enough data to form at least one bin of such an increased length.

After processing all sources' data and analyzing the final $N_{\text{H}}(t)$ curve, some observations were discarded from the sample. An overview of the final sample is shown in Table 3.3.

Source name	Obs. ID
MCG-5-23-16	0112830301
	0112830401
	0302850201
	0727960101
	0727960201
Mkn 348	0067540201
	0701180101*
NGC 2110	0145670101
NGC 4258	0110920101
	0059140101
	0059140201
	0059140401*
	0059140901
	0400560301
NGC 4507	0006220201
	0653870201*
	0653870301*
	0653870401*
	0653870501*
	0653870601*
NGC 5252	0152940101
NGC 5506	0013140101
	0013140201
	0201830201
	0201830301
	0201830401
	0201830501
	0554170201
	0554170101
	0761220101
NGC 7172	0147920601
	0202860101
	0414580101
Total time	**1.25 Ms**

Table 3.3: Final sample data overview after discarding observations. For more information and comparison to the initial sample, see Table 3.1.

* Observations not containing enough data to form at least one bin of the new length, therefore not appearing in the final $N_H(t)$ curve after the re-binning process.

Discarded observations and the reasons are stated below:

- 0203270201 (NGC 4258): The source is not located in the central chip of MOS or pn.

- 0059770101 (NGC 6300): The observation does not contain the necessary HouseKeeping files to reduce the data. Because of this, the source NGC 6300 is not considered further.

- 0880550401 (NGC 5506): The central CCD of the camera was turned off during the observation (`LIVETIME=0`) (see Section 3.2.1 for more information on this).

- 0795630101 & 0795630201 (IRAS F00521-7054): The flux drops significantly compared to other observations of the same source, from $\sim 10^{-12}$ erg cm^{-2} s^{-1} to $\sim 10^{-13}$ erg cm^{-2} s^{-1}. Because of this, the fit model obtained from previous observations does not represent the current one, and the uncertainties of the resulting N_{H} values are considerably high.

- 0301150101 & 0301151601 (IRAS F00521-7054): The average spectrum contains a very low photon count. Because of this, the optimal length of the time bin for time resolution must be increased to at least 10 ks. As a result, these observations contain one single or no time bins. Considering this and the discarding of observations 0795630101 and 0795630201, the source IRAS 00521-7054 is not taken into account within the sample.

- 0701180101 (Mkn 348), 0059140401 (NGC 4258), 0653870201, 0653870301, 0653870401, 0653870501 & 0653870601 (NGC 4507): These observations are considered during all the reduction, time-binning and spectra extraction and fitting, and appear in the final $N_{\mathrm{H}}(t)$ curves for the initial time-binning (target time bin of 5000 s) but do not appear in the final $N_{\mathrm{H}}(t)$ curves obtained after the re-binning of the corresponding source, as they do not contain enough data to form at least one bin of the new length.

The final sample consists of 8 sources, for a total of 33 observations. This corresponds to a total accumulated time of 1.25 Ms after cleaning the data (removing high background intervals) and time-binning the observations.

Chapter 4

Results & Discussion

4.1 Results

After analyzing the resulting $N_H(t)$ and $\Gamma(t)$ curves for each of the sample sources, the N_H transitions within a single observation (intra-observation changes) have been classified as secure and candidate events, and the significant N_H transitions between consecutive observations (inter-observation changes) as global N_H changes.

This chapter shows the results, plots, and analysis for the sample sources undergoing secure events (Section 4.1.1), candidate events (Section 4.1.2), and global N_H transitions (Section 4.1.3). The results for the rest of the sources are shown in Appendix A.

All results shown in this chapter and Appendix A correspond to the relaxed count rate threshold mentioned in Section 3.2.1. The results obtained with the conservative threshold are not shown, as it was observed that the scientific results (time evolution curves for $N_H(t)$, $\Gamma(t)$, flux, and HR) do not depend on the chosen threshold value and the conclusions are the same in both scenarios.

In this chapter, First Fit and Second Fit refer to the two types of fit performed for the single X-ray spectrum of each time bin of an observation mentioned in Section 3.3.

4.1.1 Secure events

To classify an event as secure, six conditions must be met:

1. N_H transition with a significance of at least 3σ relative to the average value outside the transition (before and after).

2. The N_H transition is evidenced in at least three consecutive bins. This discards outliers due to the fit.

3. The value of the photon index Γ remains constant (considering its uncertainty) during the event. This ensures that the intra-observation N_H transition is triggered by a cloud event and not by an intrinsic change in the X-ray corona of the source.

4. The reduced χ^2 during the cloud event remains at the same value as before and after the event, ensuring that the N_H transition is not a statistical effect.

5. The Hard X-ray flux remains constant during the N_{H} transition, while the soft flux varies inversely with N_{H}. This is the behavior expected from a cloud, as the photoelectric absorption due to the material in it would affect soft photons.

6. The Hardness Ratio HR shows the same trend as N_{H}, both increasing or decreasing simultaneously. This is a result of the previous criterion. If a cloud enters the line of sight to the X-ray source absorbing the soft photons, the source becomes harder. And vice-versa for a cloud leaving the line of sight.

After carefully analyzing the results, no secure events could be identified in the entire sample.

4.1.2   Candidate events

To consider an event as a candidate, conditions 3 to 6 for secure events must be met, as well as the following:

1. N_{H} transition with a significance of at least 1σ relative to the average value outside the transition (before and after).

2. The N_{H} transition is evidenced in at least two consecutive bins.

Two candidate events have been identified for two of the sources in the sample: NGC 5252 and NGC 7172. The results for each one of the sources are shown below.

NGC 5252

The $N_{\rm H}(t)$ and $\Gamma(t)$ curves obtained after the First and Second Fit of the observations associated with the source NGC 5252 are shown in Fig. 4.1.
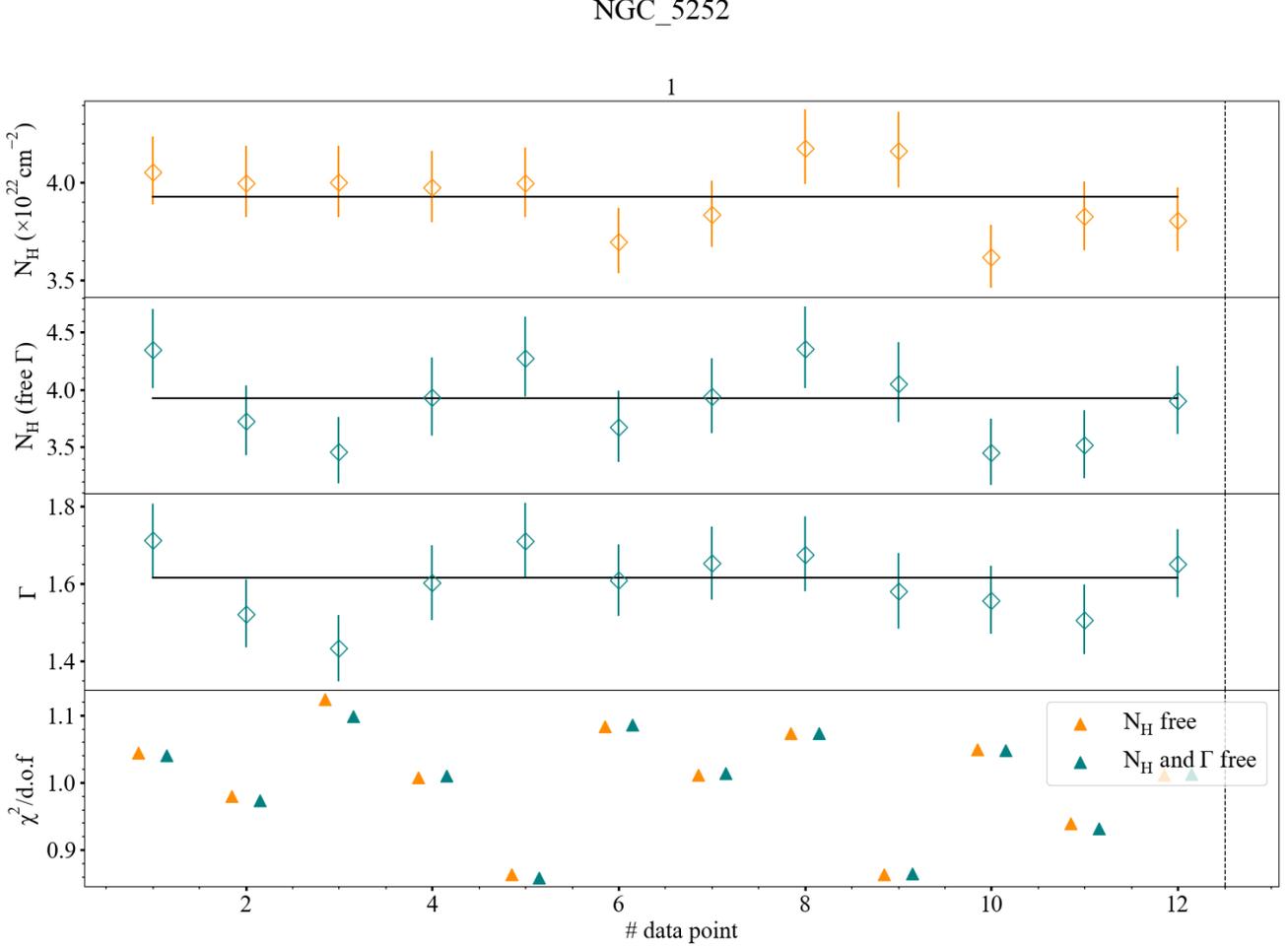
Figure 4.1: $N_{\rm H}(t)$ and $\Gamma(t)$ for source NGC 5252 showing a candidate cloud event from data points 6 to 10. Top panel: $N_{\rm H}(t)$ curve when allowing only $N_{\rm H}$ to vary during the fit (First Fit). Second panel: $N_{\rm H}(t)$ curve when allowing $N_{\rm H}$ and Γ to vary during the fit (Second Fit). Third panel: $\Gamma(t)$ curve when allowing $N_{\rm H}$ and Γ to vary during the fit. Bottom panel: reduced χ^2 for the First and Second Fit. Every point in each panel represents a time bin. The solid black horizontal line represents the best-fit parameter value for the averaged spectrum of that observation (and not a fit to the data) and the dotted black vertical line represents the end of an observation. The numbers in the top x-axis enumerate the observations of the source (useful for sources with multiple observations). The bottom x-axis represents the index of the time bin.

The length of each of the time bins represented as data points in Fig. 4.1 corresponds to 4300 s.

The top panel in Figure 4.1 shows that data points 6 to 10 present an N_{H} transition with respect to the values before data point 6 and after data point 10. To consider this N_{H} transition as a candidate cloud event, the selection criteria for candidate events mentioned previously need to be verified.

As observed, N_{H} remains constant (considering its uncertainty) between data points 1 to 5 (before the transition). It presents a significant variation from data points 6 to 10 (during the transition): N_{H} drops below the average value (data points 6 to 7), rises and remains constant above it (data points 8 to 9), and drops once again (data point 10). N_{H} goes back to its level before the transition in points 11 to 12 (after the transition).

Although with higher deviations from the average value before the N_{H} transition, the trend of $N_{\mathrm{H}}(t)$ with the First Fit is also observed with the Second Fit, shown in the second panel of Fig. 4.1. This indicates that the fit is not the cause of the N_{H} transition.

Using the N_{H} sample collected from the First Fit, the weighted column density average outside the N_{H} transition (data points 1 to 5 and 11 to 12) is estimated as $\overline{N_{\mathrm{H,out}}} = (3.97 \pm 0.05)$ cm^{-2}; and at the peak of the N_{H} transition (data points 8 to 9) as $\overline{N_{\mathrm{H,in}}} = (4.18 \pm 0.12)$ cm^{-2}.

To estimate the weighted column density average, each data point is assigned a proper weight ω given by $\omega_i = 1/\Delta N_{\mathrm{H}i,s}^2$, where $\Delta N_{\mathrm{H}i,s}$ is the symmetrized uncertainty of each N_{H} value at a confidence level of 68%. As the N_{H} values are originally estimated at a confidence level of 90% (see Section 3.3) and the lower and upper bounds of their uncertainty are different, $\Delta N_{\mathrm{H}i,s}$ is calculated by averaging the bounds of the uncertainty and dividing the result by 1.64 (90% belongs to 1.64σ). The weighted average is then calculated as:

$$\overline{N_{\mathrm{H}}} = \frac{\Sigma_i \omega_i N_{\mathrm{H}i}}{\Sigma_i \omega_i} \tag{4.1}$$

And its uncertainty as:

$$\Delta \overline{N_{\mathrm{H}}} = \frac{1}{\Sigma_i \omega_i} \tag{4.2}$$

Considering the above, the peak of the N_{H} transition from data points 8 to 9 has a significance of 1.6σ relative to the average value outside the transition. This meets the selection criteria 1 and 2 for candidate events.

On the other hand, the third and fourth panels of Figure 4.1 show that the value of Γ for data points 6 to 10 remains constant (considering its uncertainty), as well as the reduced χ^2, meeting criteria 3 and 4.

The middle panel of Figure 4.2 shows that the hard flux remains constant over time, compared to the soft flux, which increases from data points 6 to 7, then decreases from data points 8 to 9, and increases once again for data point 10. This corresponds to the inverse behavior of $N_{\mathrm{H}}(t)$ in the top panel of Figure 4.1, as expected from selection criterion 5. Finally, the bottom panel of Figure 4.2 shows that the HR from data points 6 to 10 presents the same behavior as $N_{\mathrm{H}}(t)$ in the top panel of Figure 4.1, meeting selection criterion 6.
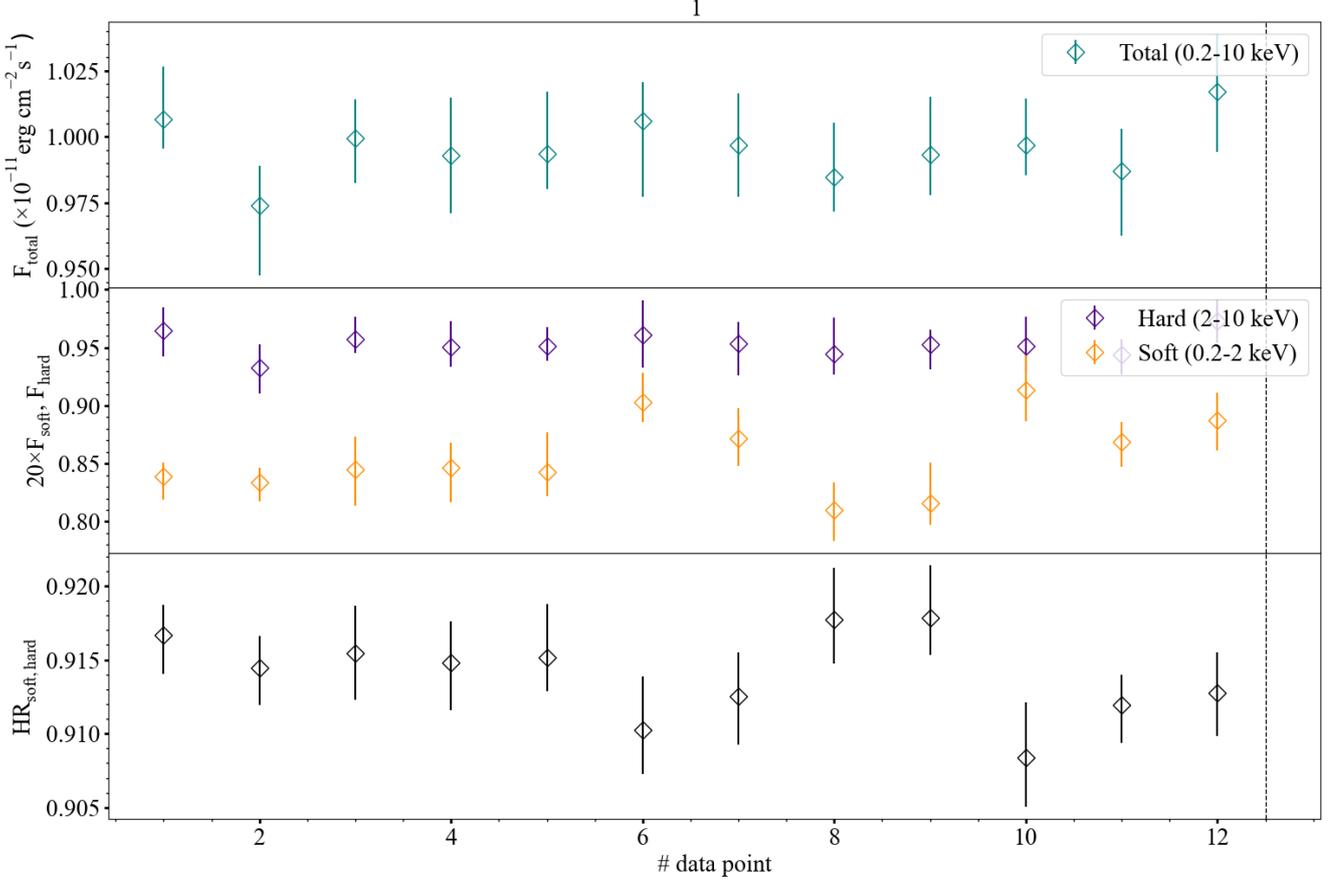
Figure 4.2: Light curves of NGC 5252. Top panel: Total flux light curve ($0.2 - 10$ keV). Middle panel: Hard ($2 - 10$ keV) and soft $0.2 - 2$ keV flux light curve. Bottom panel: HR time-evolution curve. Every point in each panel represents a time bin, corresponding to the ones shown in Figure 4.1. The dotted black vertical line represents the end of an observation. The numbers in the top x-axis enumerate the observations for the source (useful for sources with multiple observations). The bottom x-axis represents the index of the time bin.

As observed in Fig. 4.1 and Fig. 4.2, the mentioned N_H transition in NGC 5252 meets criteria 1 to 6 for candidate cloud events. Thus, the transition might be considered to be driven by a tentative ingress to egress cloud event in 2003. Possibly, the increase in N_H followed by a decrease to its average value is a signature of a cloud entering and immediately leaving the line of sight of the central X-ray source.

It is debated if the candidate cloud event consists of data points 6 to 10, or only data points 8 to 9. The reason for the lack of certainty is the increased level of N_H before the transition, which is not the expected behavior of a cloud event. However, as the soft flux and HR light curve in Fig. 4.2 show a clear transition starting in data point 6 and ending in data point 10, the cloud event is considered to occur in such an interval.

Taking into account the time-binning of the observation, the tentative ingress to egress cloud event has a crossing time of $t_D = (21500 \pm 4300)$ s $= (6 \pm 1)$ h. The peak of the event occurs at data point 8, when $N_{H,peak} = (4.18 \pm 0.12)$ cm^{-2} (68% confidence level), that is when the cloud has entirely entered the line of sight. This corresponds to a maximum change of $\Delta N_H = N_{H,peak} - \overline{N_{H,out}} = (4.18 \pm 0.12) - (3.97 \pm 0.05) = (0.21 \pm 0.13)$ cm^{-2}.

Assuming all clouds in the line of sight of the X-ray source have the same column density adding up to $\overline{N_{H,out}}$, the average number of clouds along any radial equatorial ray can be estimated as $\mathcal{N}_0 = \overline{N_{H,out}}/\Delta N_H = (3.97 \pm 0.05)$ cm^{-2}/(0.21 ± 0.13) cm$^{-2} \approx 19 \pm 12$ clouds.

NGC 7172

Following the same analysis structure as for source NGC 5252, the results for NGC 7172 are presented below.

The resulting $N_H(t)$ and $\Gamma(t)$ curves after fitting the single spectra extracted for each time bin are shown in Figure 4.3.
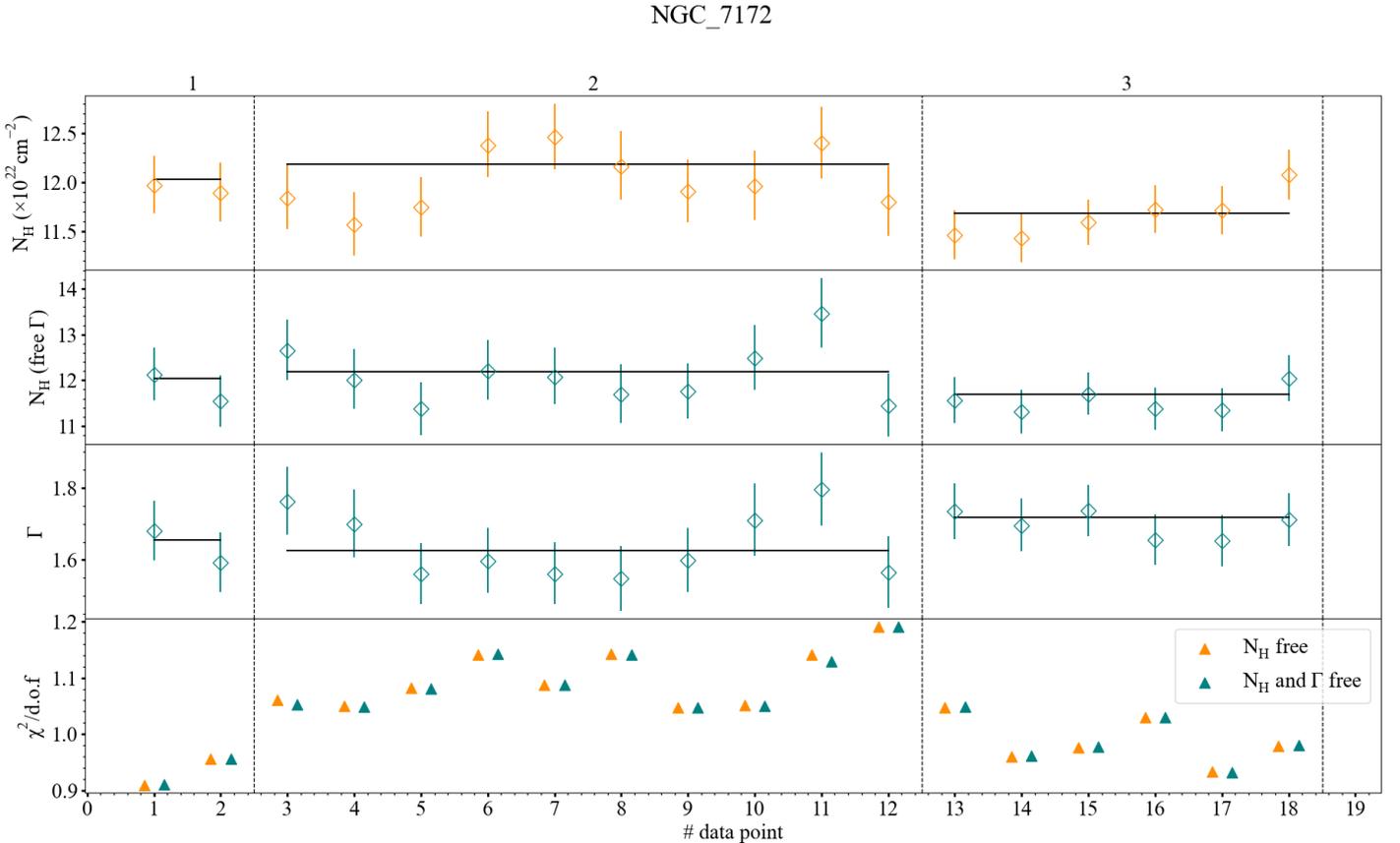
Figure 4.3: $N_H(t)$ and $\Gamma(t)$ for source NGC 7172 showing a candidate cloud event from data points 5 to 9 in observation 2. See Figure 4.1 for a detailed description.

The time bin lengths of each data point for the observations of NGC 7172 shown in Fig. 4.3 are shown in Table 4.1.

Obs. index	Obs. ID	Bin data length (s)
1	0147920601	6000
2	0202860101	4650
3	0414580101	5280

Table 4.1: Time bin length for observations of NGC 7172

The top panel in Figure 4.3 shows an overall increase of N_{H} from data points 5 to 9, corresponding to observation 2. Although not as significant, this increase is also identified in the $N_{\mathrm{H}}(t)$ curve for the Second Fit.

From the N_{H} sample obtained from the First Fit of observation 2, the weighted column density average outside the N_{H} transition (data points 3 to 4 and 10 to 12) is calculated as $\overline{N_{\mathrm{H,out}}} = 12.04 \pm 0.12$, and at the peak of the N_{H} transition (data points 6 to 7) as $\overline{N_{\mathrm{H,in}}} = 12.38 \pm 0.20$. Taking this into account, the peak of the N_{H} transition has a significance of 1.4σ. This meets selection criteria 1 and 2 for candidate events. In this case, the data points before and after the N_{H} transition could be interpreted as an event itself, however, these intervals do not meet conditions 3 to 6 for secure events.

The third panel of Figure 4.3 shows that the value of Γ from data points 5 to 9 remains constant (considering its uncertainty), meeting criteria 3 for secure events. The fourth panel shows that χ^2 does not present any significant variation during the transition, meeting selection criterion 4.

The middle panel of Figure 4.4 shows that the hard flux remains constant over time. Although the changes in the soft flux are imperceptible in the plot, it drops from an average of $\overline{F_{\mathrm{soft}}} = (1.03 \pm 0.02) \times 10^{-13}$ erg cm^{-2} s^{-1} before the event, to $\overline{F_{\mathrm{soft}}} = (1.01 \pm 0.03) \times 10^{-13}$ erg cm^{-2} s^{-1} during it, as expected from criterion 5. Finally, the bottom panel of Figure 4.2 shows that the HR time-evolution curve presents the same behavior as $N_{\mathrm{H}}(t)$ in the top panel of Figure 4.3: it increases from data point 5 to 6, peaks in 7, and decreases again from 8 to 9. This meets criterion 6.
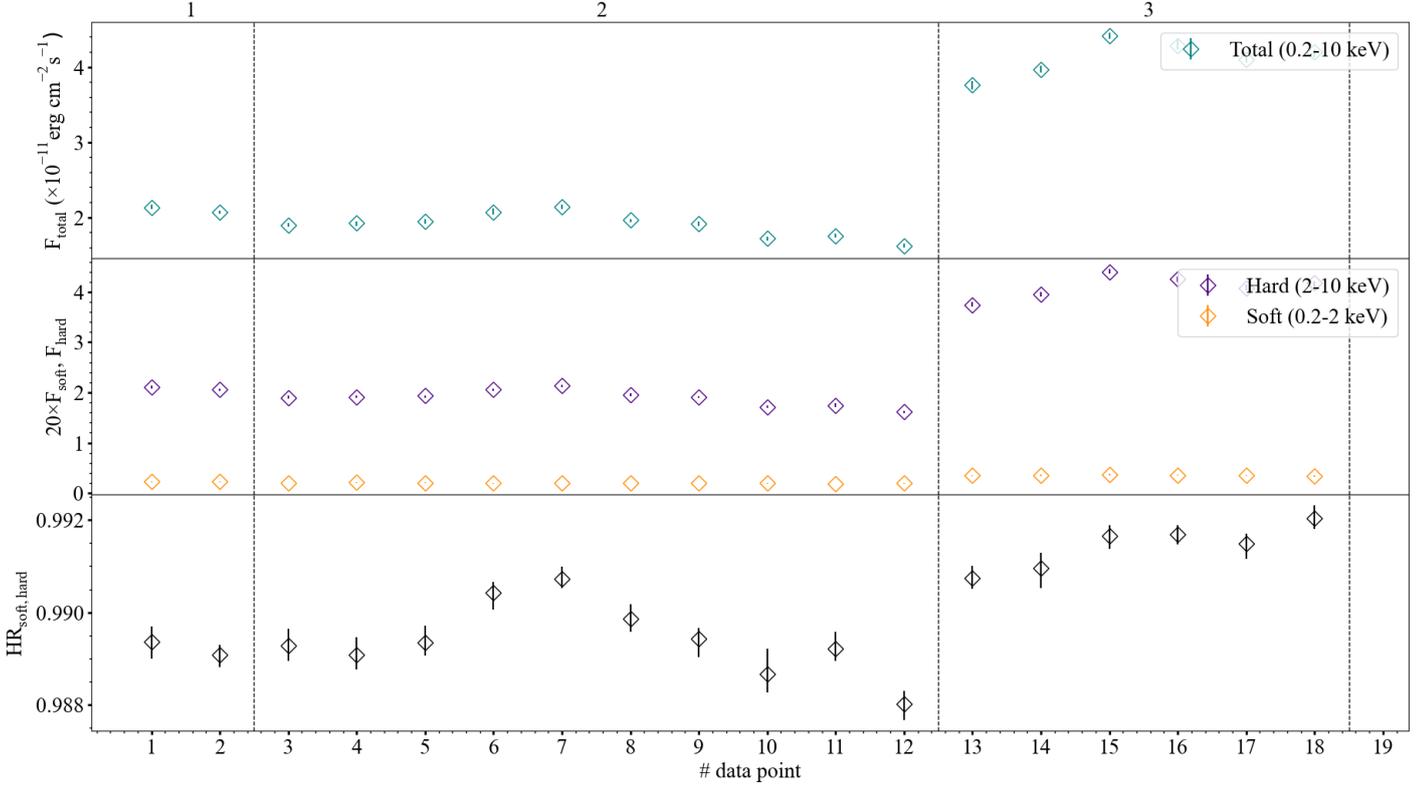
Figure 4.4: Same as in Figure 4.2, for source NGC 7172

As observed in Figures 4.3 and 4.4, the N_H transition in observation 2 of NGC 7172 meets criteria 1 to 6 for candidate cloud events. Thus, as in NGC 5252, the increase in N_H followed by a decrease to its average value in NGC 7172, is a signature of a possible cloud entering and immediately leaving the line of sight of the central X-ray source IN 2004

According to the time-binning of the observation shown in Table 4.1, the tentative ingress to egress cloud event has a crossing time of $t_\mathrm{D} = (23250 \pm 4650)$ s $= (6.4 \pm 1.3)$ h.

The peak of the event occurs at data point 7 when $N_\mathrm{H,peak} = (12.46 \pm 0.20)$ cm^{-2}. This corresponds to a maximum deviation of the hydrogen column density of $\Delta N_\mathrm{H} = (12.46 \pm 0.20) - (12.04 \pm 0.12) = (0.42 \pm 0.23)$ cm^{-2}, leading to an average number of clouds in the line of sight along the equatorial plane of $\mathcal{N}_0 = \overline{N_\mathrm{H,out}}/\Delta N_\mathrm{H} = (12.04 \pm 0.12)$ cm$^{-2}/(0.42 \pm 0.23)$ cm$^{-2} \approx 29 \pm 16$ clouds.

4.1.3   Inferred statistic of global N_H transitions

As mentioned in Section 4.1.1, no secure cloud events were captured during the accumulated exposure time. However, for some of the sources in the sample, the average value of N_H presents significant variations between different observations, typically spaced apart by weeks to months or years. This implies that N_H is indeed

changing on such time scales.

To consider a change in the hydrogen column density as significant, the ΔN_{H} calculated by subtracting the average value of N_{H} for one observation from the value of N_{H} from the immediate consecutive one, must have a significance higher than 3σ.

Although no constraints on the clouds' properties can be drawn from such changes, as no information is available for the time between observations, it is possible to estimate the probability of capturing at least one of such N_{H} transitions during one of the observations of the source.

To do this, two main assumptions are made:

- **Independence of events:** The N_{H} changes are independent of each other, that is, the occurrence of one event does not affect the probability of another event occurring. For example, the fact that a cloud entered the line of sight at a certain time does not influence when the next cloud event will happen.

- **Constant average rate:** The N_{H} changes occur at a constant average rate throughout the observation period. This is based on the assumption that, if the AGN torus is composed of clouds, they are uniformly distributed over the equatorial plane.

The probability of observing an N_{H} transition during a specific observation can be modeled using the Poisson distribution. According to this model, if λ number of events occur in a given period, the probability of observing k events in the same period is given by:

$$P_k = \frac{e^{-\lambda}\lambda^k}{k!} \tag{4.3}$$

To estimate the probability of catching $k = 1$ N_{H} transition during one of the observations, $\lambda_i = R \cdot t_i$ where R is the rate at which N_{H} transitions are observed in the line of sight and t_i is the duration of a single observation. R can be calculated as the number of events n observed in a time T spanning from the first to the last observation of the source.

Following this, the average probability of catching one cloud event in the line of sight of a source observed N times is given by:

$$\overline{P_{\mathrm{event}}} = \frac{1}{N}\left(\sum_{i=1}^{N} e^{-(n/T)\cdot t_i} \cdot (n/T) \cdot t_i\right) \tag{4.4}$$

Three of the sources in the sample present global N_{H} changes between observations: NGC 7172, NGC 4258 and NGC 4507. The results for each source are shown below.

NGC 7172

An overview of the N_H and Γ changes over the years is shown in Figure 4.5.
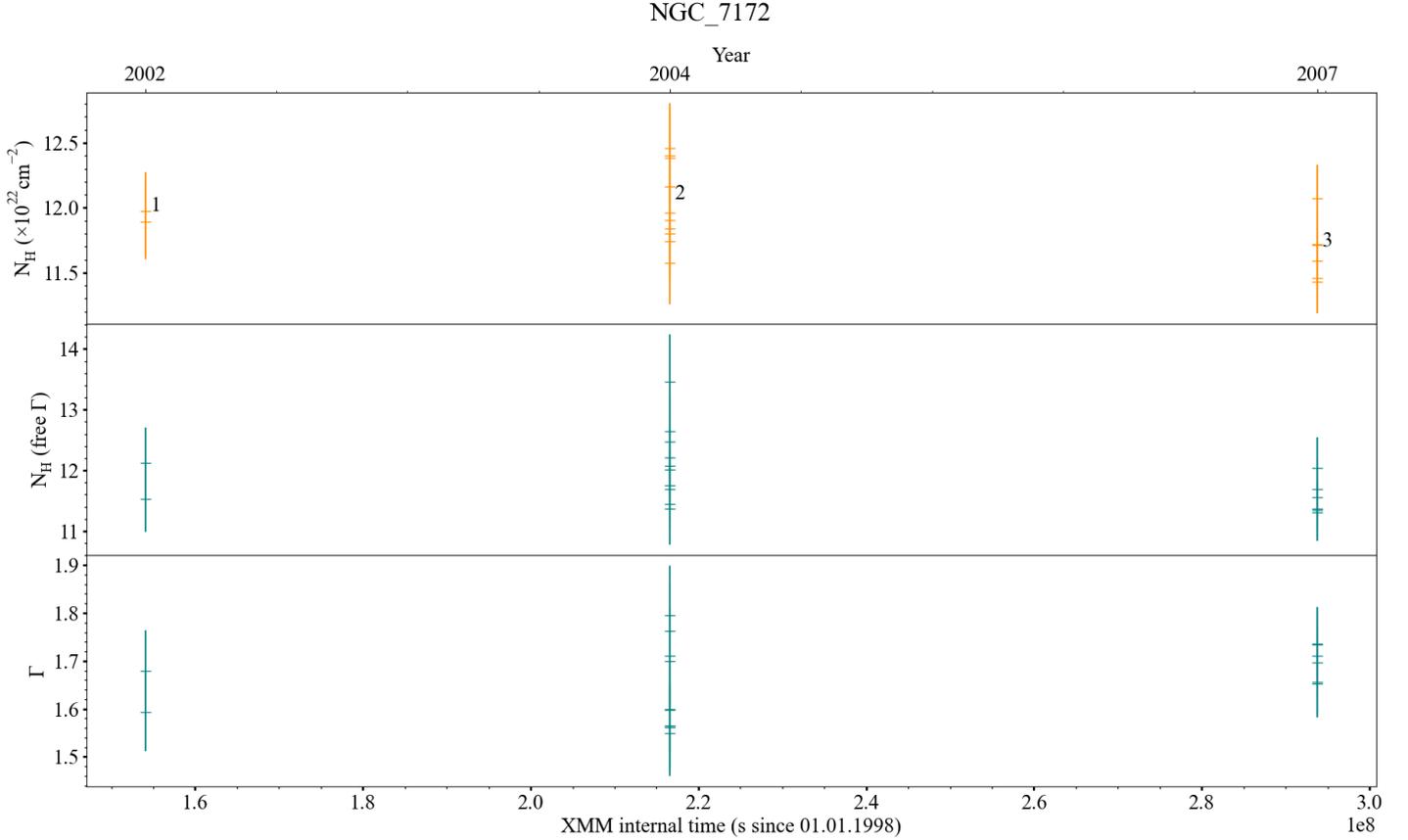
Figure 4.5: Top panel: N_H values for each of the observations time bins when allowing only N_H to vary during the fit. Second panel: N_H value for each time bin, when allowing N_H and Γ to vary during the fit. Third panel: Γ value for each time bin, when allowing N_H and Γ to vary during the fit. The top x-axis shows the year of the observation, and the bottom x-axis represents the internal *XMM-Newton* time at the moment of the observation, measured in seconds since 01.01.1998. The numbers on the top panel represent the index of the observations corresponding to those in the top x-axis of Figure 4.1.

Using the average N_H values estimated from performing the First Fit on the average spectrum of each observation (see top panel of Figure 4.3), the change in the column density from observation 2 to 3, has a significance of 3.1σ, corresponding to $\Delta N_{H,2\to3} = (12.18 \pm 0.13) - (11.69 \pm 0.09) = (0.50 \pm 0.16)$ cm^{-2} (68% confidence level). This means, according to Figure 4.5, that in a time span of around 3 years, from November 2004 to April 2007, at least one N_H transition occurred in the line of sight towards the X-ray source of NGC 7172. As no data is available after 2007, it is not possible to identify if, for example, a new cloud entered the line of sight after the suspected cloud left, or if the column density stayed at its previous level.

Following the above, the number of N_H transitions in the line of sight of NGC 7172 is $n = 1$ during a period of time $T = 1618$ days $= 1.40 \times 10^8$ s, so $R = 7.14 \times 10^{-9}$. Taking into account the duration of each of the $N = 3$ observations of NGC 7172 shown in Table 3.1, the average probability of catching such N_H transition during one of the observations of NGC 7172 was $\overline{P_{\text{event}}} = 0.00031$, according to equation 4.4.

NGC 4258

The resulting $N_H(t)$ and $\Gamma(t)$ curves obtained from the first time-binning (target bin length of 5 ks) of the observations of NGC 4258 are shown below:
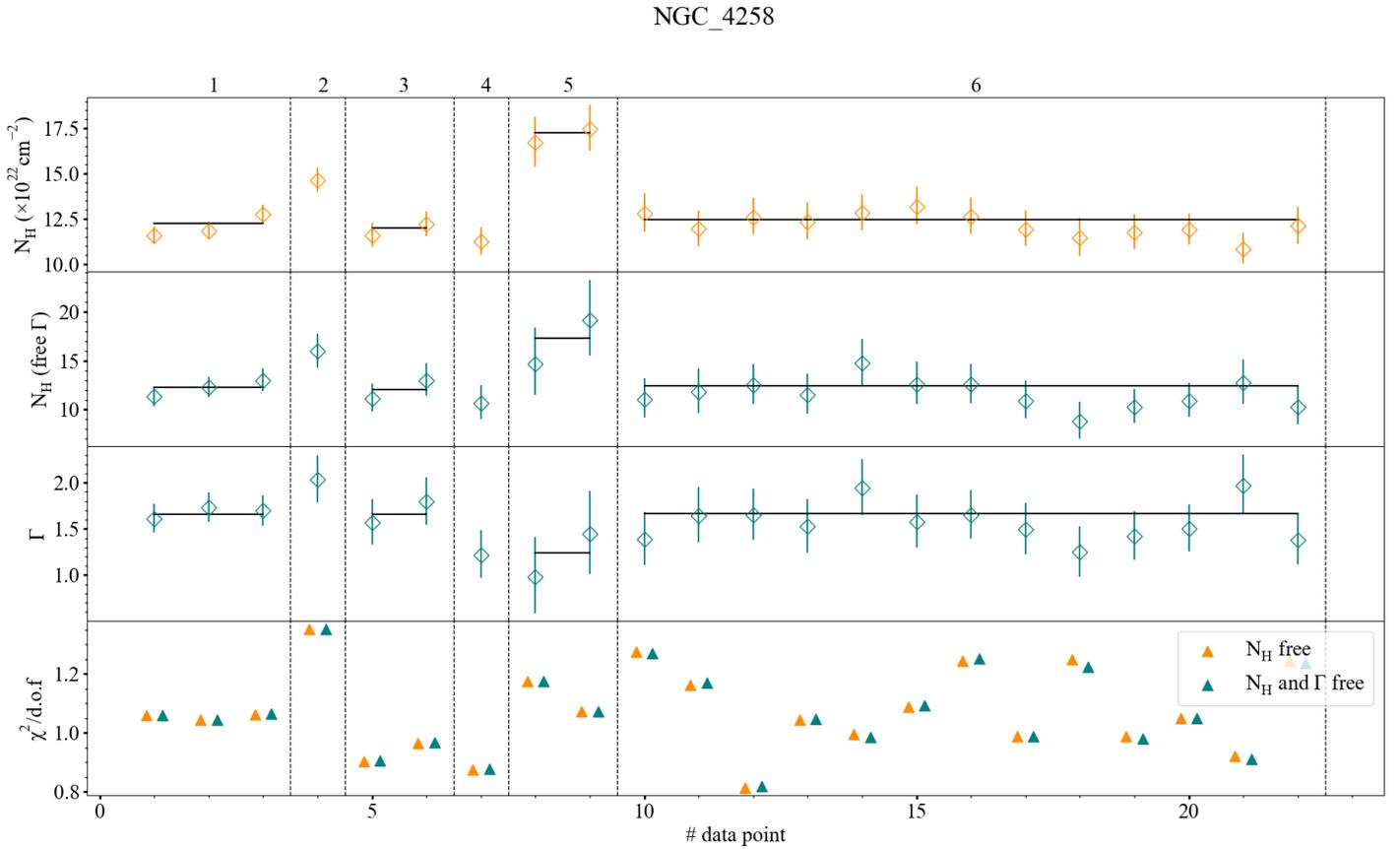
Figure 4.6: Same as in Figure 4.1, for source NGC 4258*
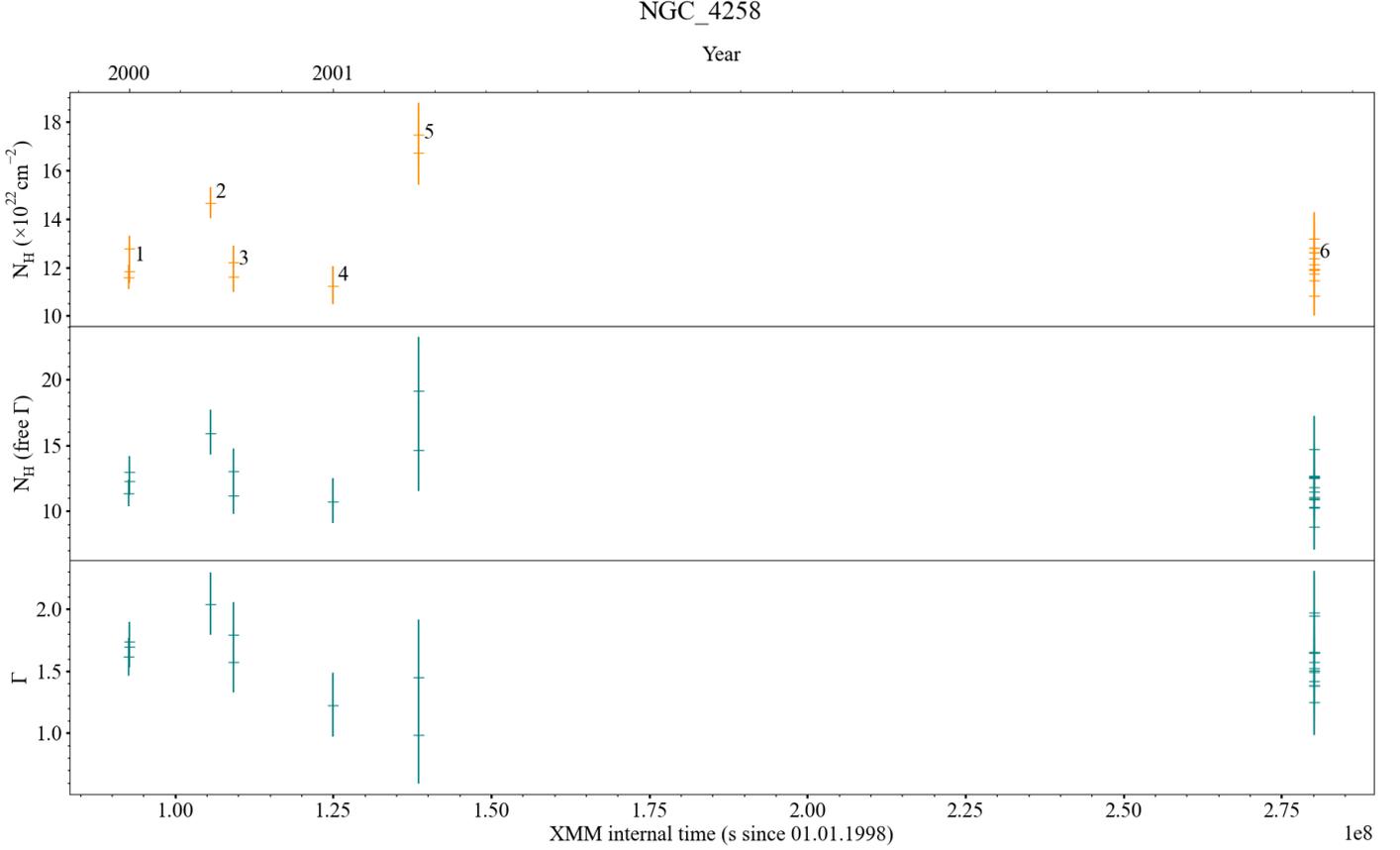
Figure 4.7: Same as in Figure 4.5, for source NGC 4258*

* For this source, the observations binning into bins of 5 ks produced N_{H} uncertainties higher than 10% of the average N_{H} of the observation. Because of this, the observations were re-binned into new bins of appropriate length. These results are shown in Appendix A. In this section, however, the light curves shown correspond to the original binning of the observations (5 ks), as the estimations made are based on the value of N_{H} obtained from the average spectrum, which is independent of the time binning.

The analysis for the global N_{H} transitions captured in the following sources has the same structure as the one for NGC 7172. Based on the average N_{H} estimations for source NGC 4258 and following Figure 4.7, NGC 4258 underwent two significant changes: in a time span of around 1 year, from observation 4 in 2001 to observation 5 in 2002, where $\Delta N_{\mathrm{H},4\rightarrow5} = (17.28 \pm 1.65) - (11.39 \pm 1.17) = (5.88 \pm 2.02)$ cm^{-2}, with a significance of 2.9σ, and in a time span of 4 years, from observation 5 in 2002 to observation 6 in 2006, where $\Delta N_{\mathrm{H},5\rightarrow6} = (12.45 \pm 0.37) - (17.28 \pm 1.61) = (-4.83 \pm 1.65)$ cm^{-2}, with a significance of 2.9σ.

Both changes in N_{H} could be interpreted as the signature of a single cloud event. As observed in the top panel of Fig. 4.6 and Fig. 4.7, before 2002, when observation 5 took place, the value of the average N_{H} remained constant overall with no significant changes. During observation 5, the average N_{H} was estimated to be the highest, compared to the rest of the observations of NGC 4258. In 2006, when the source

was observed again in observation 6, the column density had come back to the same baseline level as before 2002. This is also evidenced as $|N_{H,4\to5}| \approx |N_{H,5\to6}|$. The behavior of the hydrogen column density might correspond to a cloud in the torus entering the line of sight sometime between December 2001 and May 2002, and leaving it sometime between May 2002 and November 2006.

For this source, the number of N_H transitions is $n = 2$ during a period of time $T = 2207$ days $= 1.91 \times 10^8$ s, so $R = 1.05 \times 10^{-8}$. Taking into account the duration of each of the $N = 6$ observations of NGC 4258 shown in Table 3.1, the average probability of catching one cloud event during one of the observations of NGC 4258 was $\overline{P_{\text{event}}} = 0.00025$, according to equation 4.4.

NGC 4507

The resulting $N_H(t)$ and $\Gamma(t)$ curves obtained from the first time-binning of the observations of NGC 4507 (target bin length of 5 ks) are shown in Fig. 4.8.
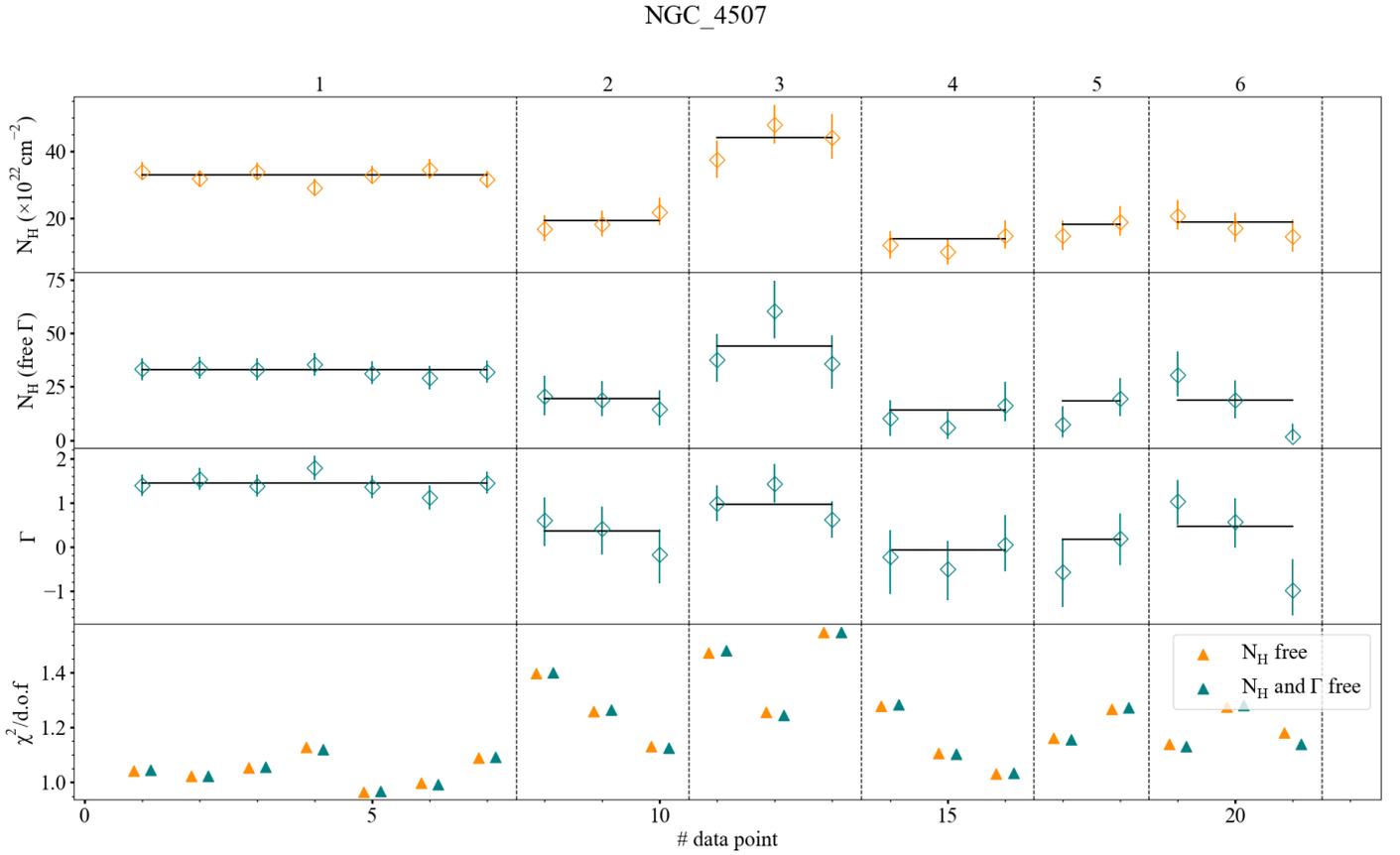
Figure 4.8: Same as in Figure 4.1, for source NGC 4507.
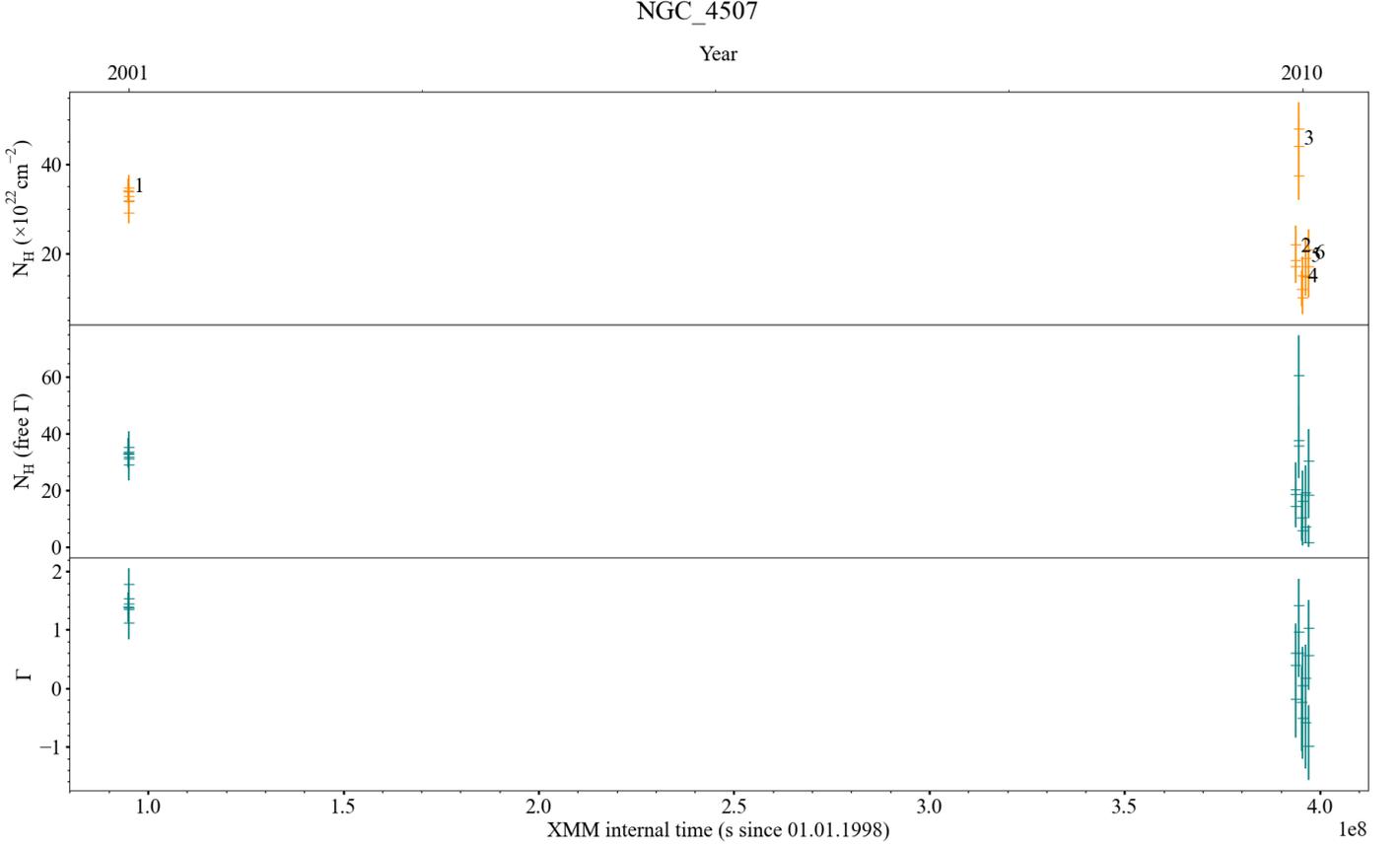
Figure 4.9: Same as in Figure 4.5, for source NGC 4507.

According to the top panel of Fig. 4.8 and following Fig. 4.9, NGC 4507 under-
went two significant changes: in a time span of around 10 days, from observation
2 in June 2010 to observation 3 in July 2010, where $\Delta N_{\text{H},2\to3} = (43.97 \pm 5.21) -$
$(19.31 \pm 3.92) = (24.66 \pm 6.52)$ cm^{-2}, with a significance of 3.8σ, and in a time
span of 10 days again, from observation 3 to observation 4 in July, 2010, where
$\Delta N_{\text{H},3\to4} = (13.99 \pm 3.57) - (43.97 \pm 5.08) = (-29.97 \pm 6.21)$ cm^{-2}, with a signifi-
cance of 4.8σ. As in NGC 4258, both N_{H} changes in NGC 4507 could be interpreted
as the signature of a single cloud event. As observed in the top panel of Fig. 4.8 and
Fig. 4.9, the increase of N_{H} from observation 2 to 3 is of the order of its decrease
from observation 3 to 4, i.e $|N_{\text{H},2\to3}| \approx |N_{\text{H},3\to4}|$, and both changes occur in similar
time frames ($\sim$ 10 days). During observation 3, N_{H} reaches its peak compared to
the rest of the observations of the same source. After observation 4, N_{H} returns to
its baseline level. This behavior of the column density might be driven by a cloud
entering the line of sight of NGC 4507 between June 24th, 2010, and July 3rd, 2010,
and immediately leaving between July 3rd, 2010, and July 13th, 2010.

For this source, the number of N_{H} transitions is $n = 2$ during a period of time
$T = 41$ days $= 3.54 \times 10^6$ s (considering only the observations in 2010), so $R =$
5.65×10^{-7}. Taking into account the duration of each of the $N = 5$ observations of
NGC 4507 (Table 3.1), the average probability of catching one cloud event during
one of the observations of NGC 4507 in 2010 was $\overline{P_{\text{event}}} = 0.0103$, according to
equation 4.4.

80

4.2 Discussion

In section 4.1.1, it was discussed that no significant N_H changes were identified during a single observation. This means that in a cumulative exposure time of 1.25 Ms, no clouds were captured during their ingress and/or egress to and from the line of sight. This scenario corresponds to the behavior of the light curve shown in Fig. 2.1a, which, as discussed in Section 2, might represent an AGN torus composed of either numerous X-ray absorbing clouds with $N_H \ll 10^{22}$ cm^{-2} staying in the line of sight during the entire observing time, or of clouds that are too rare to be observed.

Notwithstanding the above, when loosening the selection criteria, two events are identified for two of the sources in the sample: NGC 5252 and NGC 7172. Both events, although candidates, might be caused by the same process: a cloud entering the line of sight and leaving it after a couple of hours. Although the direct signature of these events is a change in the hydrogen column density while the photon index of the power law remains constant, the candidate events were easier to spot in the hardness ratio and flux light curve. This might be due to the fact that the $N_H(t)$ curve is sensitive to the smallest changes in the column density as it is the only factor accounting, in contrast to the hardness ratio, which accounts for changes in the soft (as a result of changes in N_H) and hard flux.

Candidate events were analyzed following the procedure that would have been followed if a secure event had been identified. On the one hand, the crossing time t_D of the suspected clouds was estimated using the length of the bins forming the event. The value for both candidate events is of the order of a couple of hours, which is at the lower end of the crossing time expected from clouds located in the AGN torus. This, however, does not represent an inconsistency with previous studies (e.g Markowitz et al. (2014), where the cloud events were estimated to have a duration spanning from less than a day to over a year) but it is rather an expected result, as *XMM-Newton* is the most sensitive X-ray satellite, and therefore can constrain events in shorter time domains.

Using this estimate of t_D, the diameter of the cloud could have be calculated following the method proposed by Lamer & Uttley (2003), described in Chapter 1.3. To do so, the first step would have been calculating the distance r_{cl} from the cloud to the supermassive black hole, using equation 1.6. This, however, requires the use of external methods such as reverberation mapping, stellar kinematics, or gas dynamics to estimate the mass of the supermassive black hole M_{BH}, and calculate the total luminosity L_{ion} of the photons capable of ionizing the gas, as well as the ionization parameter ξ. These calculations are, however, outside the scope of this master's thesis and are therefore not carried out.

On the other hand, by estimating the maximum change ΔN_H of the column density during the candidate cloud events, it was possible to give an estimate of the total average number of clouds along any radial equatorial ray, assuming that all clouds have the same N_H. For the first candidate event, in NGC 5252, it was estimated that $\mathcal{N}_0 \approx 19 \pm 12$ clouds. And for NGC 7172, $\mathcal{N}_0 \approx 29 \pm 16$ clouds. Considering their uncertainties, both of the $\mathcal{N}_0$ values are in agreement with previous studies (Nenkova et al., 2008; Markowitz et al., 2014) estimating a total average number of clouds along any radial equatorial ray ranging between 2 and 15 clouds.

When considering global changes of N_H between observations, it is evident that

the torus is not a homogeneous structure where the material is evenly distributed. This is evidenced as N_H undergoes changes of up to more than 100%, as that from observation 2 to 3 of NGC 4507. As no *XMM-Newton* observation covers the time between those two observations, no data is available for the time interval when such changes were occurring, therefore it is not possible to identify their detailed nature. The options include clouds of gas and dust located in the AGN torus, entering and leaving the line of sight of the X-ray source of the AGN as they rotate around the supermassive black hole (Nenkova et al., 2008; Nikutta et al., 2009; Laha et al., 2011; Markowitz et al., 2014), or a multi-phase medium with regions of higher and lower density (Stalevski et al., 2012). If the scenario were a clumpy torus, catching the ingress, egress, or an entire cloud event during one of the observations in the sample would have been highly unlikely. As estimated for the global changes of N_H in Section 4.1.3, this probability was of the order of 10^{-2} to 10^{-4}: around two orders of magnitude smaller than the probability of $0.110^{+0.461}_{-0.071}$ calculated by Markowitz et al. (2014) for Seyfert 2 AGN. The reason for this discrepancy is the difference in the observing strategy between *RXTE* and *XMM-Newton*. *RXTE* conducts short, repeated observations of multiple sources while *XMM-Newton* conducts longer observations of individual objects. As observed from the results of this thesis, the observing strategy of *RXTE* is more suitable to track N_H changes caused by the material in the AGN torus, as they are extremely rare and the AGN needs to be constantly monitored to capture one of them during a single observation. To increase the chance of capturing a cloud event during one observation, it would be necessary to increase the number of Seyfert 2 sources in the sample and observe them for a substantial amount of time, in the order of months.

In the previous study by Markowitz et al. (2014), two cloud events were identified for two of the sources in the sample of this master thesis: NGC 5506 and Mkn 348. Using the data from Table 5 in Markowitz et al. (2014), it is possible to calculate the period of those 2 clouds by assuming they rotate around the supermassive black hole in a Keplerian orbit. For the cloud event of NGC 5506 in 2002.2, the period is ~ 118 years, while for the one of Mkn 348 in 1997-7, it is ~ 5200 years. These are extremely long periods compared to satellite missions times, and combined with the low number of clouds in the torus, make it very improbable to catch repeated cloud events during a single observing run.

Chapter 5

Conclusions & Future Work

This master thesis aimed to constrain the distribution of the gas and dust located in the AGN torus. According to IR and X-ray studies in the last decades, the material in the torus is distributed in high-density clumps that might or might not be embedded in a lower-density homogeneous medium. To constrain these models, I used data from the *XMM-Newton* satellite for eight Seyfert 2 AGN, including a total of 33 observations and 1.25 Ms of data.

To handle the great amount of data from the sample, I developed a software, named XMM-DH, to automatically reduce and remove high background intervals from the data. The software then split every observation into time bins that were subsequently treated as sub-observations. The length of the time bins range between 5 to 20 ks. For the entire observation, as well as for each of the time bins, the X-ray spectrum in the range of $0.2 - 10$ keV was extracted. Using the best-fit spectral models derived from Laha et al. (2020) for the source in the sample, the extracted spectra were fit and the value of the hydrogen column density $N_{\rm H}$ for each of the time bins was estimated. These values were concatenated into time evolution curves (e.g. $N_{\rm H}$), which were assessed to identify any significant changes in $N_{\rm H}$ that might be driven by clouds entering or leaving the line of sight of the X-ray source of the AGNs.

The significant changes identified were classified as secure, candidate, or global events. No secure events were identified for any of the sources of the sample, meaning that none of the changes in the column density within a single observation met the criteria to be attributed with certainty to a cloud. These criteria include a change in $N_{\rm H}$ observed for a significant duration, a photon index Γ that remains constant during the $N_{\rm H}$ change, a reduced χ^2 indicating that the variation of the column density is not a consequence of the fitting performance, a soft flux varying inversely to $N_{\rm H}(t)$ and a hardness ratio proportionally to it. Notwithstanding the above, two candidate cloud events were identified in two sources: NGC 5252 and NGC 7172. Both candidate events can be explained by a tentative cloud entering and immediately leaving the line of sight. These variations meet most of the criteria for secure events, but their duration and peak change are not significant enough to be considered as one. For these candidate events, I followed the procedure that would have been followed if a secure event had been identified. The crossing time t_D of the alleged clouds was estimated as $\sim 6 \pm 1$ hours for the event in NGC 5252 and $\sim 6.45 \pm 1$ hours for the one in NGC 7172. These times are consistent with the literature, which estimates a t_D spanning from hours to a couple of years, based

on observed events. Additionally, the average number $\mathcal{N}_0$ of clouds along a radial equatorial ray in the AGN was estimated. For NGC 5252, this corresponds to 19 ± 12 clouds, while for NGC 7172 to 29 ± 16 clouds. Both estimations, considering their uncertainties, lie within the expected range of previous studies, between 2 to 15 clouds.

The lack of cloud events observed during a single observation does not discard the possibility that the material in the AGN torus is distributed in clouds of higher density than the baseline. This hypothesis still holds, as N_{H} presents significant changes between observations, proving that the events I was after do exist. Although no conclusions can be made about these changes, they put into perspective the large component of luck on which the work here was based: the probability of catching a cloud event occurring during one of the *XMM-Newton* observations was extremely small, of the order of 10^{-2} to 10^{-3}.

This, rather than being a discouragement to try to understand the structure of the AGN torus, is a motivation to continue searching for a model that fits the observations. For a study with fewer time constraints than a master's thesis, this work would be carried out on a much larger scale: the initial sample would be extended to include all Seyfert 2 AGN observed with *XMM-Newton*. Additionally, work would be done to obtain the best-fit spectral models for each of the observations in the sample, to account for sudden changes in the AGN physical processes.

Many of the great achievements in science, like the discovery of the Higgs Boson or the detection of gravitational waves, have been made by scientists who decided to keep trying, even when the odds of success were infinitesimally small. This master's thesis was one more example in that list.

Summary

This thesis has confirmed, once again, that the model describing the AGN torus as a well-defined toroidal structure where the material is homogeneously distributed, does not accurately describe the observations. Instead, clumpy models provide a more accurate approach. However, the nature of the baseline absorber (only clouds or low-density homogeneous medium) remains unconstrained. This is because the probability of observing cloud events in the line of sight is extremely small and catching one during a single observation is a matter of luck.

Appendix A

Results of individual Seyfert 2 AGN

This Appendix shows the resulting plots for sources where no secure, candidate, or global events occurred. The resulting plots obtained after the re-binning of observations with uncertainties higher than 10% of the average $N_{\rm H}$ are also shown here.

For each of the sources, the results include: the average pn spectrum for single events for each observation of the source along with its fitting model obtained from Laha et al. (2020), the figures for the visualization of the binning of each observation, the table containing the lengths of each of the time-bins, the $N_{\rm H}$ and Γ time evolution curves for both fits mentioned in Section 3.3 in a time bin scale, the $N_{\rm H}$ and Γ light curve in real-life time and the flux and hardness ratio light curve. All the plots correspond to the relaxed `RATE` threshold mentioned in Section 3.2.

A.1 MCG 5-23-16

As seen in Fig. A.1, the best-fit spectral model (red line) derived from Laha et al. (2020) for MCG 5-23-16 is well-suited for observations 0112830301 and 0112830401, however, it presents significant discrepancies in the soft energy range of observations 0302850201, 0727960101, and 0727960201. The count rate observed at low energies in these observations seems to have increased after observation 0112830401, creating a hump in the soft energy range that needs an additional model component to describe it. This is not a straightforward process, as the signal-to-noise ratio of these X-ray spectra is very high.

Spectra
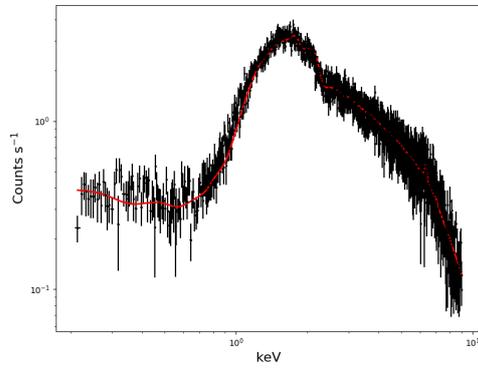
(a) Obs. ID 0112830301
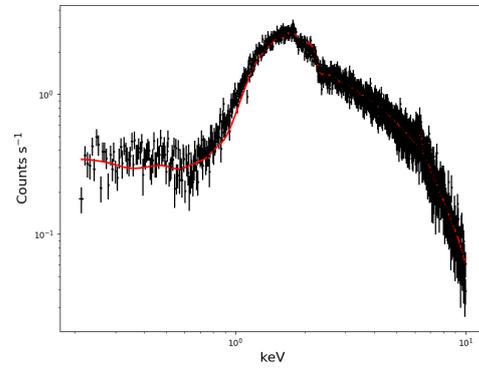
(b) Obs. ID 0112830401
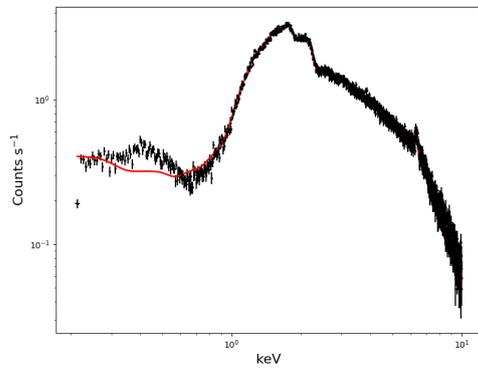
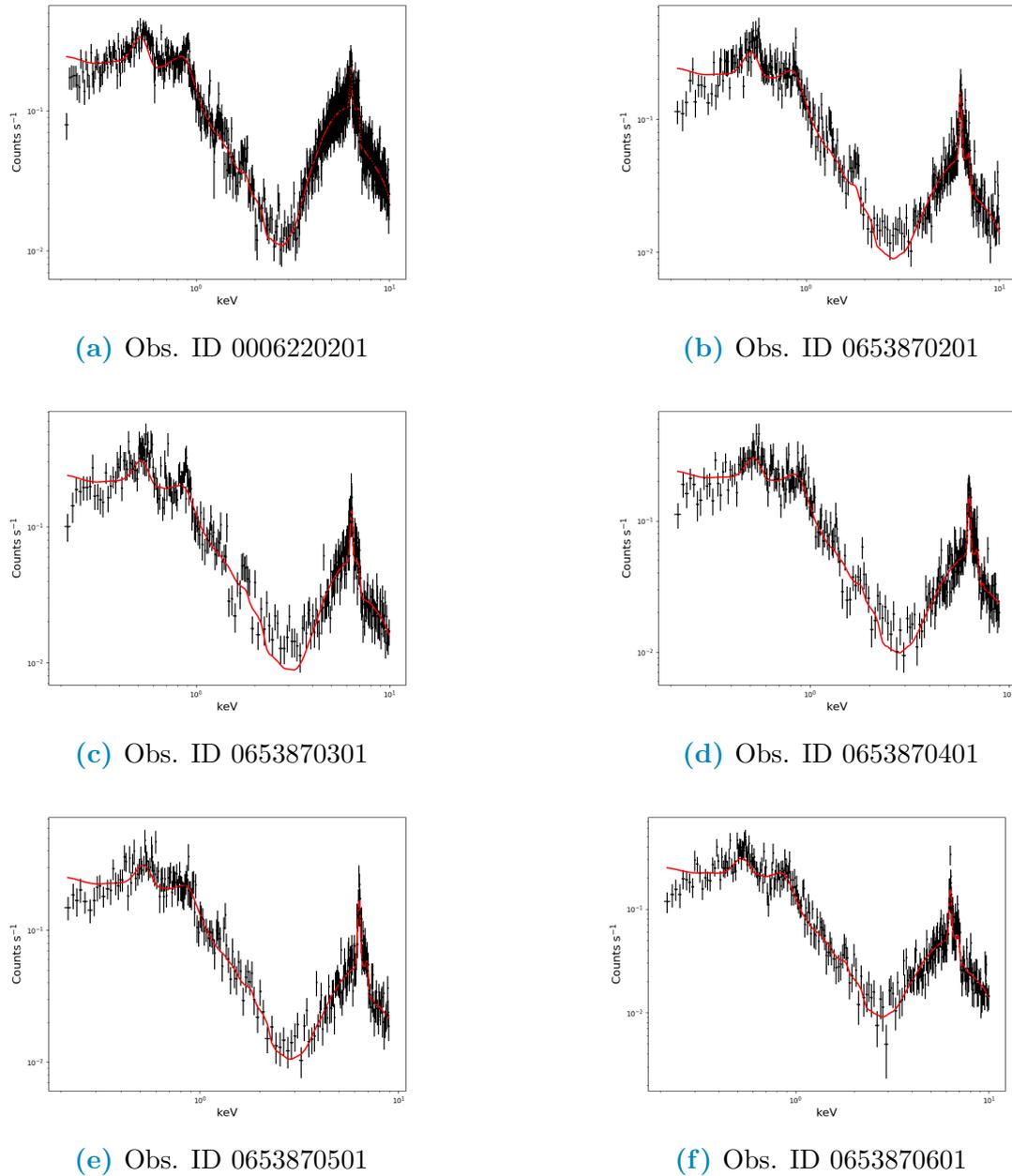(c) Obs. ID 0302850201
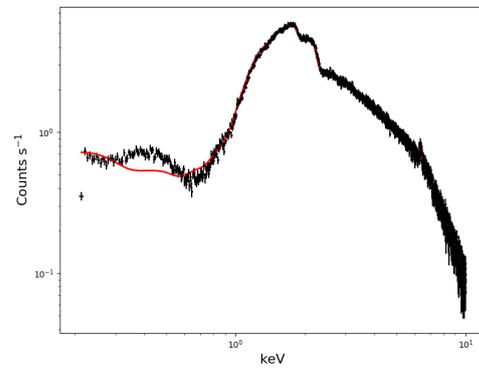
(d) Obs. ID 0727960101
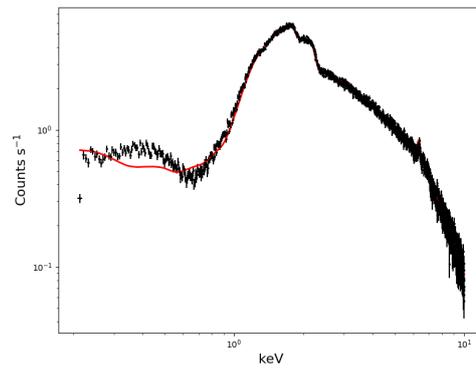
(e) Obs. ID 0727960201

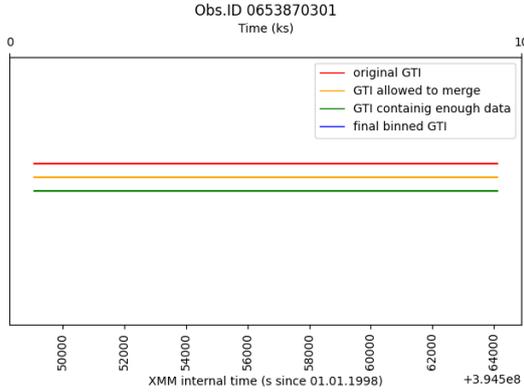**Figure A.1:** Same as in Figure 3.15 for the observations of source MCG 5-23-16.

Time binning
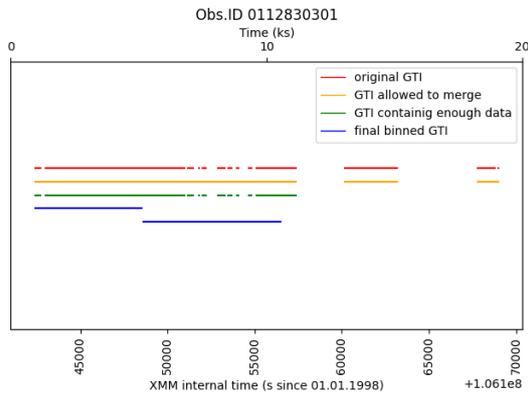
(a) Time binning Obs. ID 0112830301
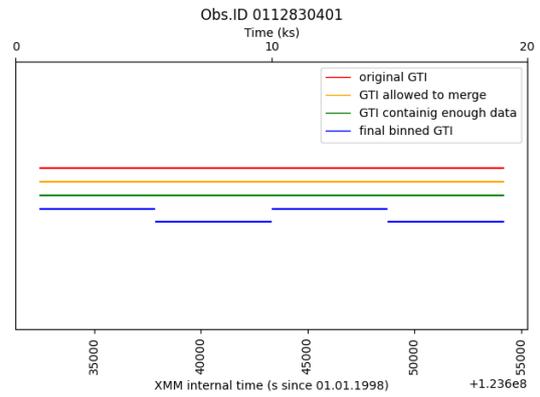
(b) Time binning Obs. ID 0112830401
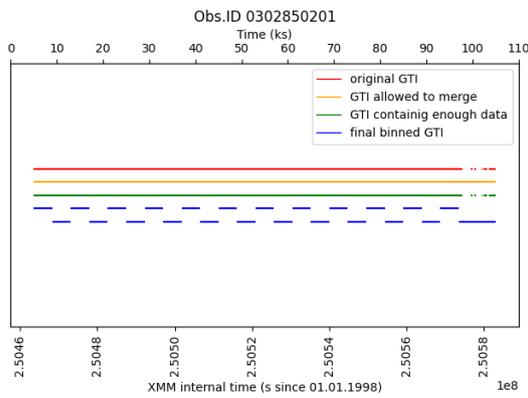
(c) Time binning Obs. ID 0302850201
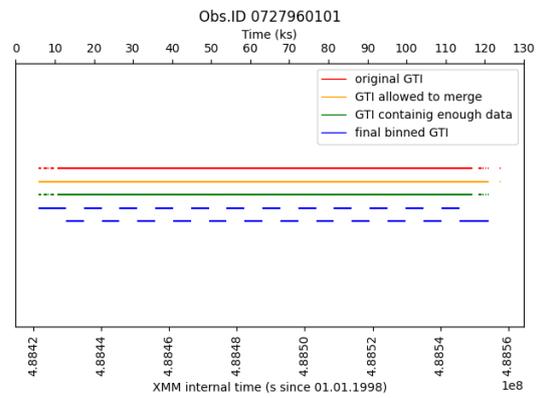
(d) Time binning Obs. ID 0727960101
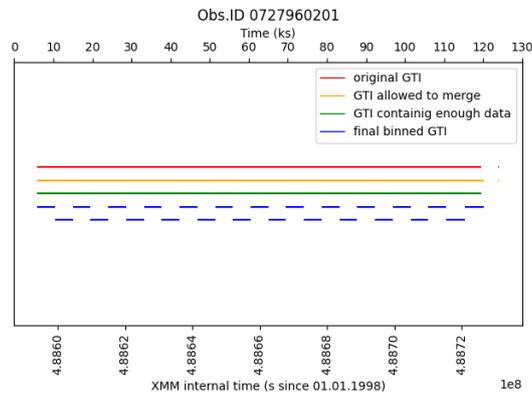
(e) Time binning Obs. ID 0727960201

Figure A.2: Same as in Figure 3.13, for source MCG 5-23-16

Obs. index	Obs. ID	Bin data length (s)
1	0112830301	5999.99
2	0112830401	5449.99
3	0302850201	4779.16
4	0727960101	5270.83
5	0727960201	5283.99

Table A.1: Time bin length for observations of MCG 5-23-16

Time evolution curves
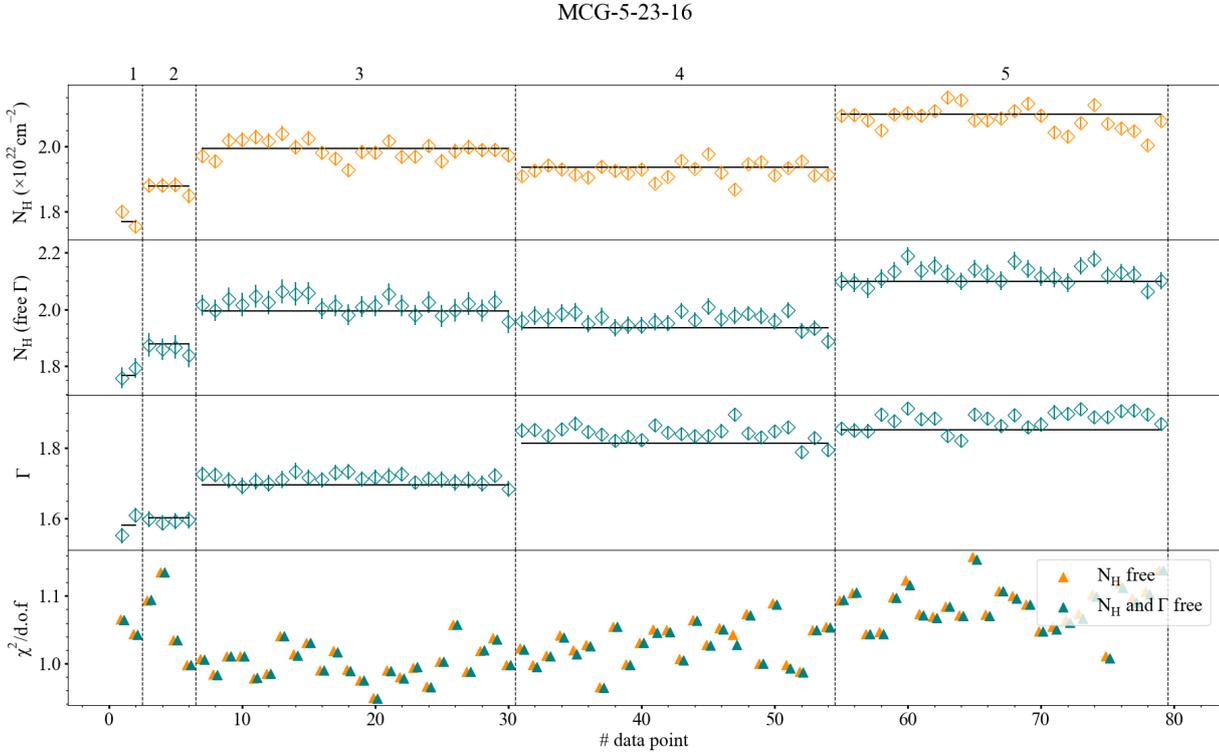
Figure A.3: Same as in Figure 4.1, for source MCG 5-23-16
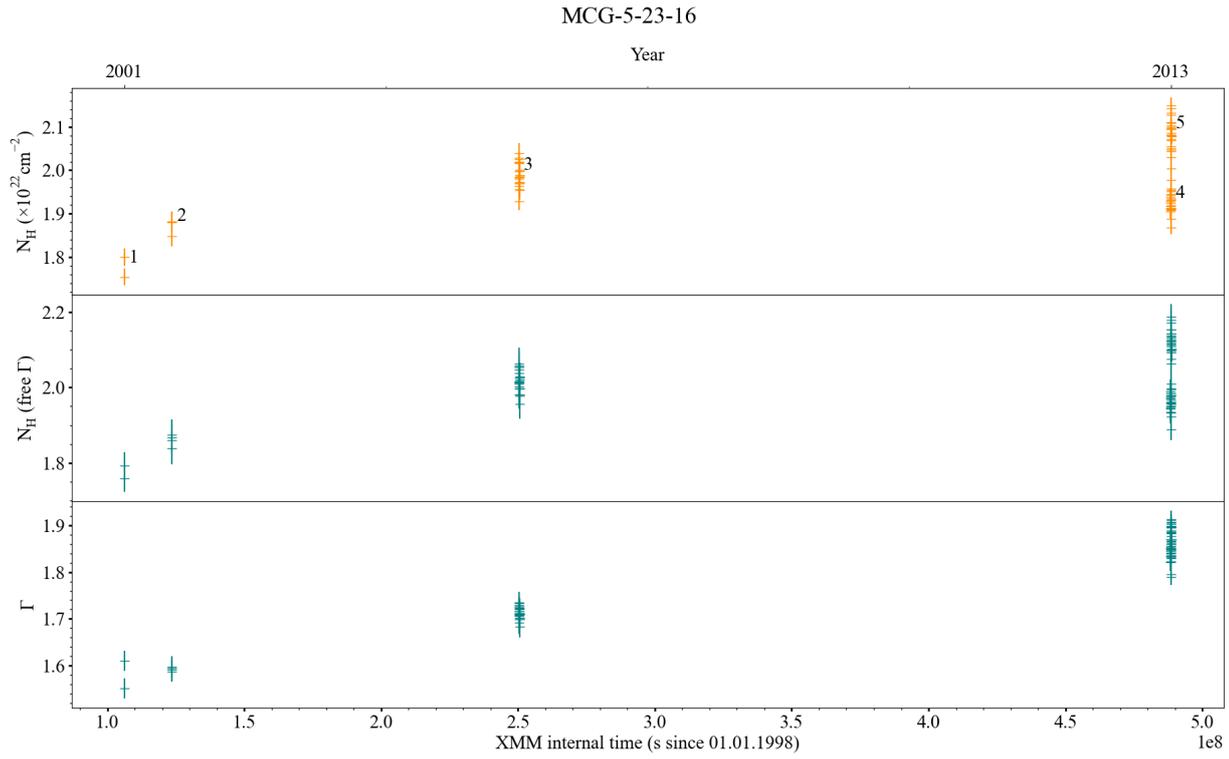
Figure A.4: Same as in Figure 4.5, for source MCG 5-23-16
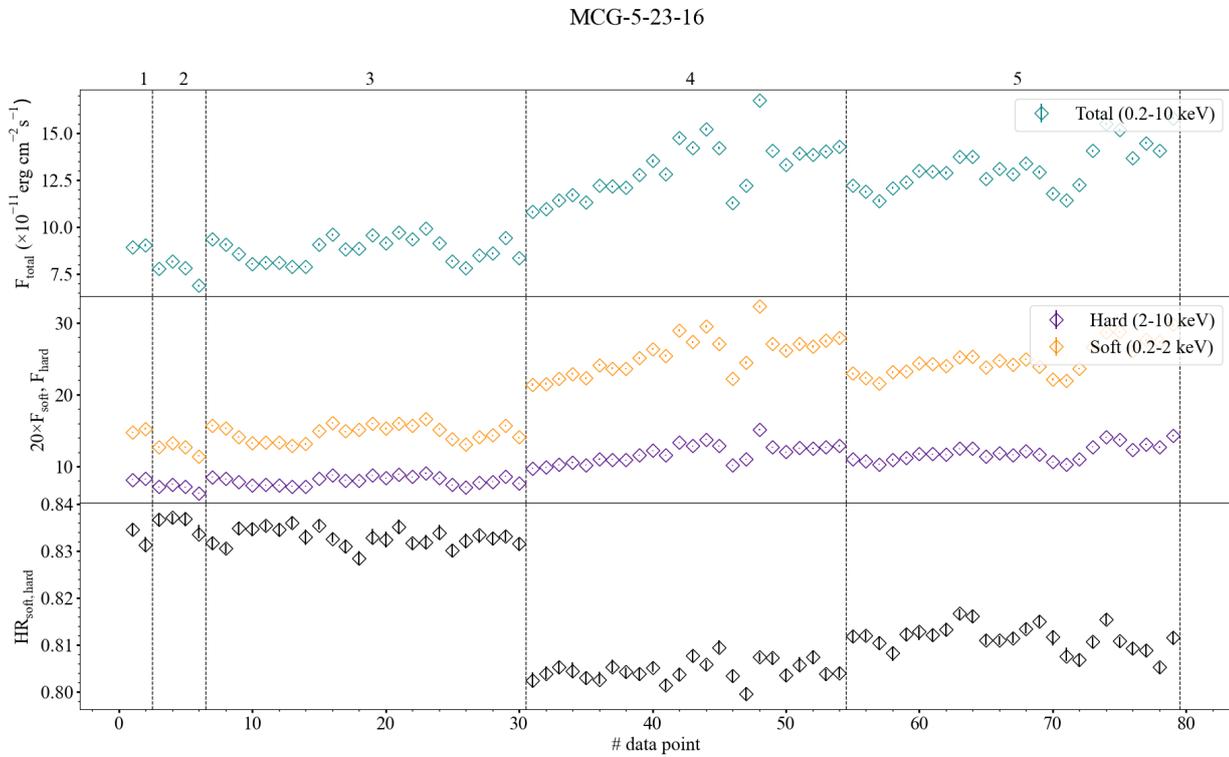
Figure A.5: Same as in Figure 4.2, for source MCG 5-23-16

89

A.2   Mkn 348

Spectra
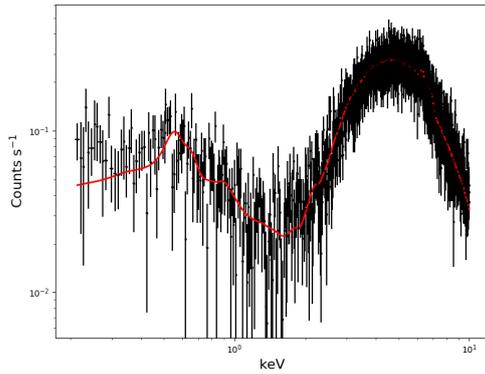
(a) Obs. ID 0067540201
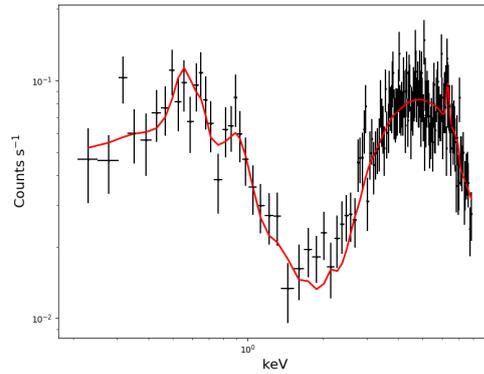
(b) Obs. ID 0701180101

Figure A.6: Same as in Figure 3.15 for the observations of source Mkn 348

Time binning
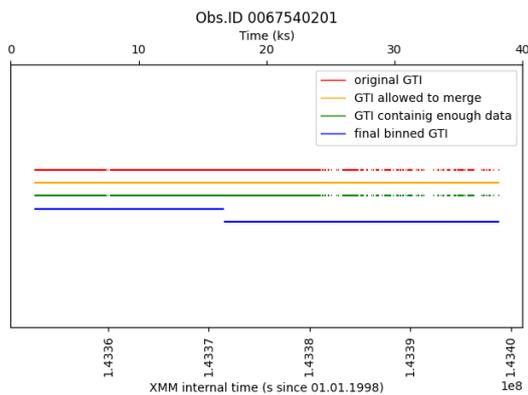
(a) Time binning Obs. ID 0112830301
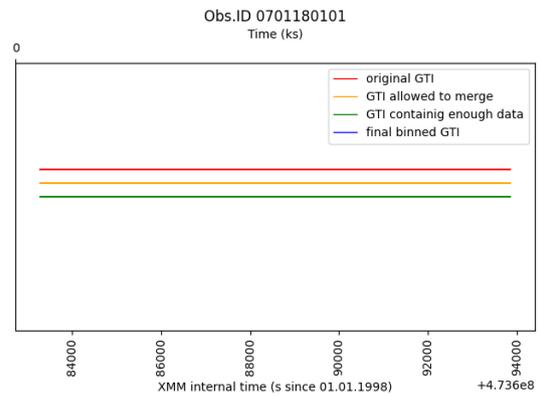
(b) Time binning Obs. ID 0701180101. No final bins are created as the observation is not long enough to create at least one bin of target bin size 20 ks.

Figure A.7: Same as in Figure 3.13, for source Mkn 348

Obs. index	Obs. ID	Bin data length (s)
1	0067540201	18699.99

Table A.2: Time bin length for observations of Mkn 348

Time evolution curves
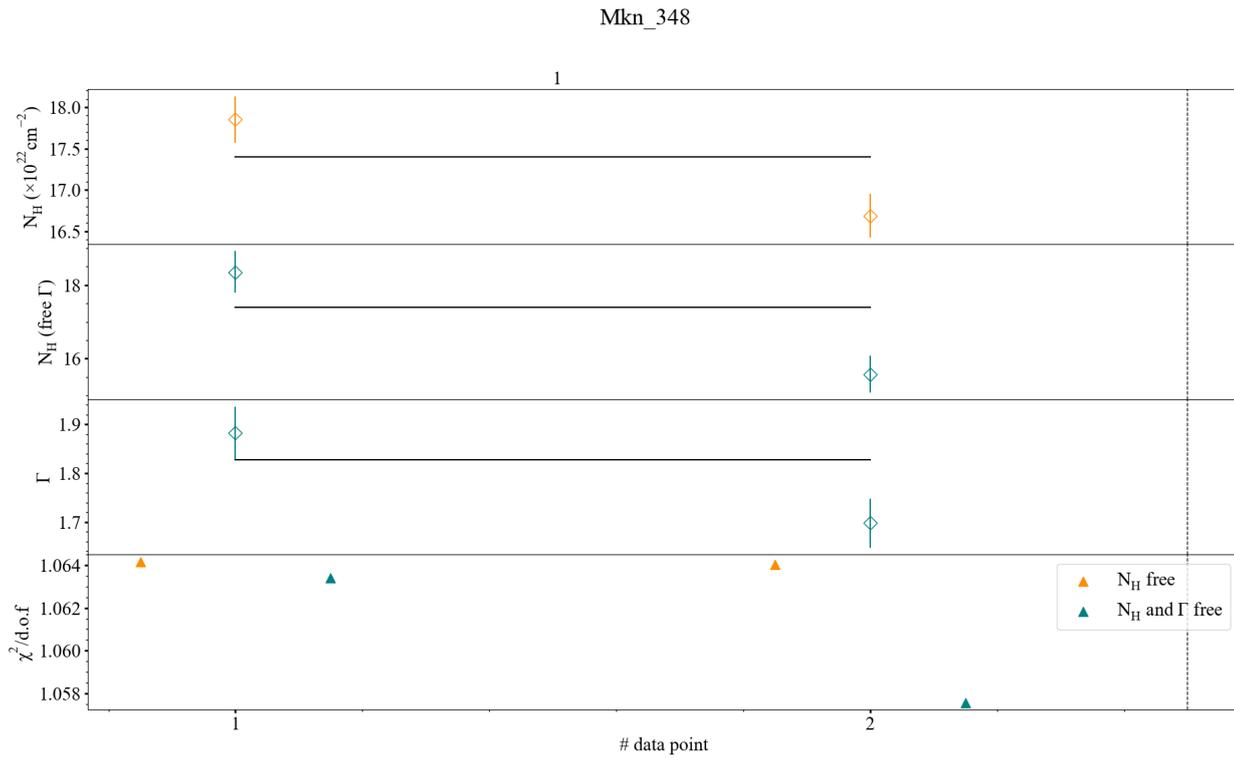
Figure A.8: Same as in Figure 4.1, for source Mkn 348
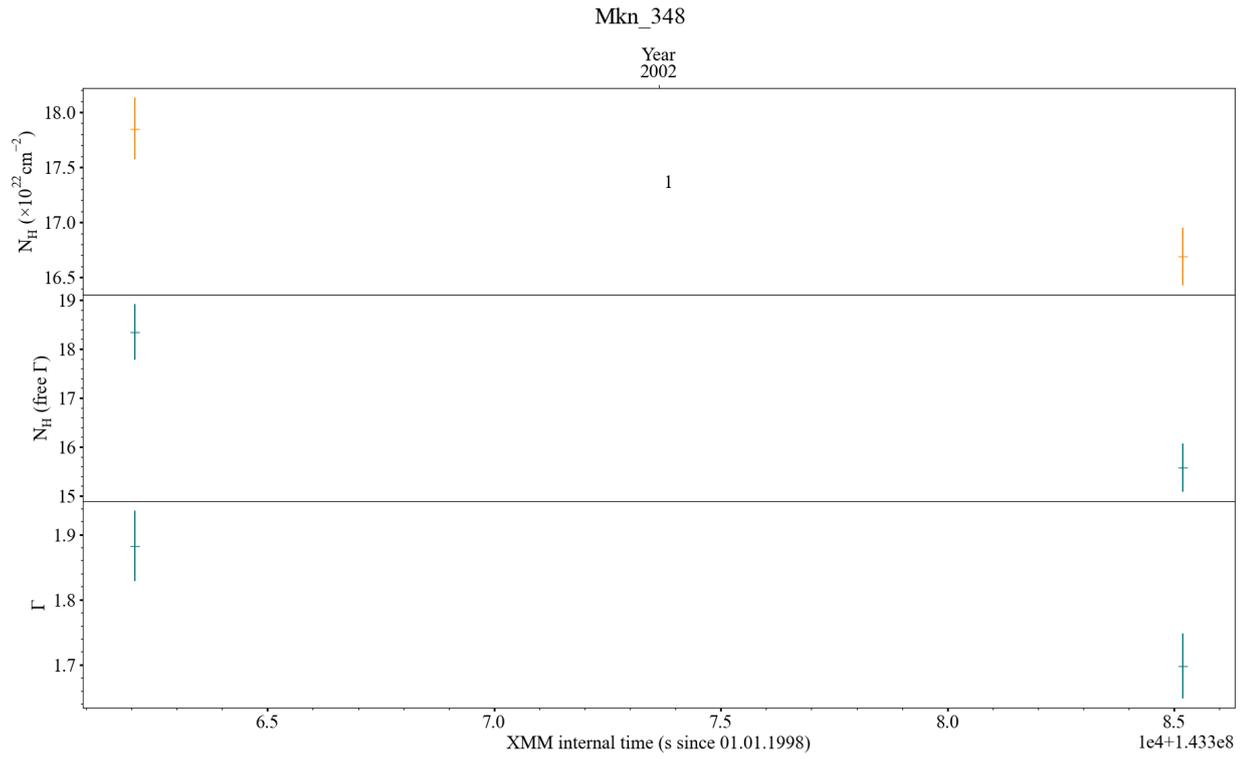
Figure A.9: Same as in Figure 4.5, for source Mkn 348
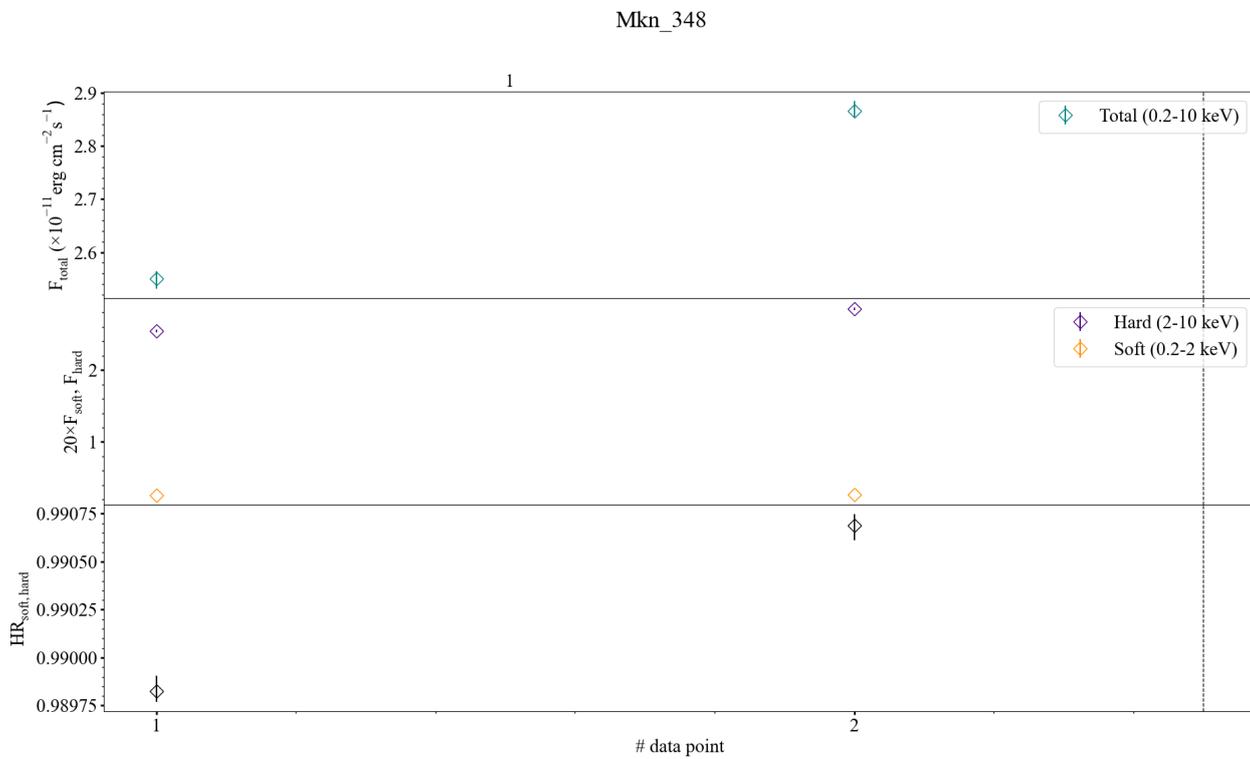
Figure A.10: Same as in Figure 4.2, for source Mkn 348

A.3 NGC 2110

Spectra
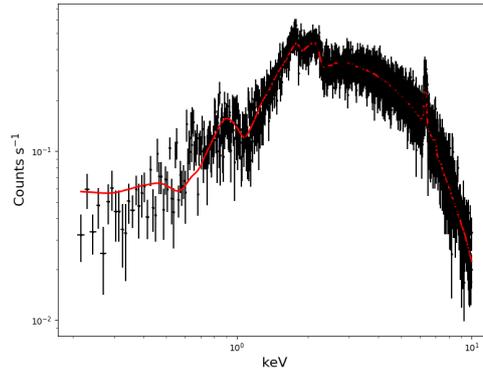
(a) Obs. ID 0145670101

Figure A.11: Same as in Figure 3.15 for the observations of source NGC 2110

Time binning
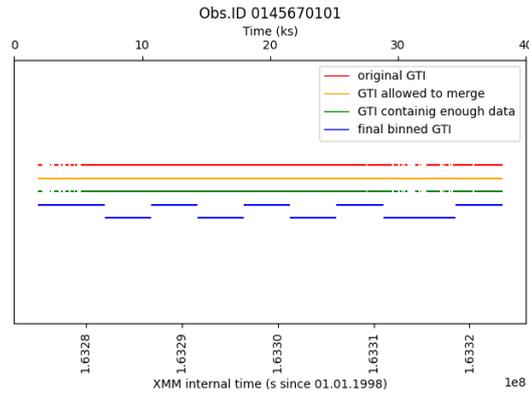
(a) Time binning Obs. ID 0145670101

Figure A.12: Same as in Figure 3.13, for source NGC 2110

Obs. index	Obs. ID	Bin data length (s)
1	0145670101	4822.22

Table A.3: Time bin length for observations of NGC 2110
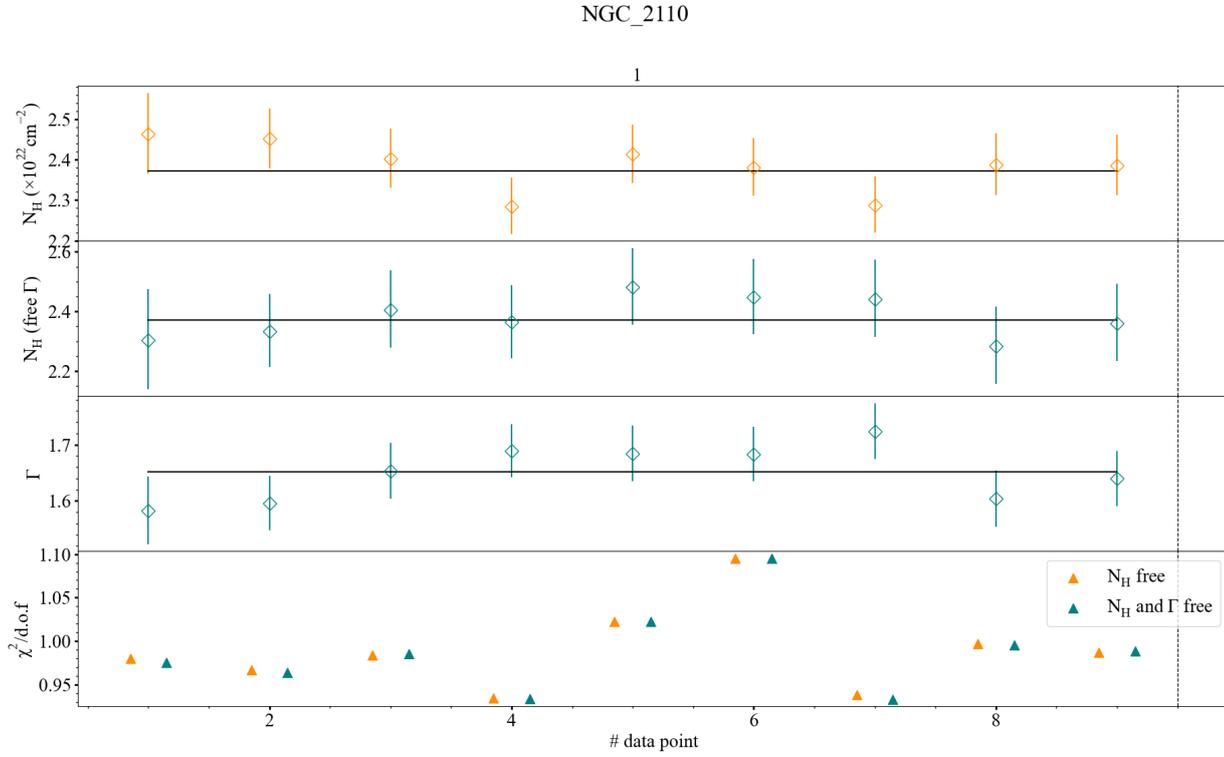
Figure A.13: Same as in Figure 4.1, for source NGC 2110
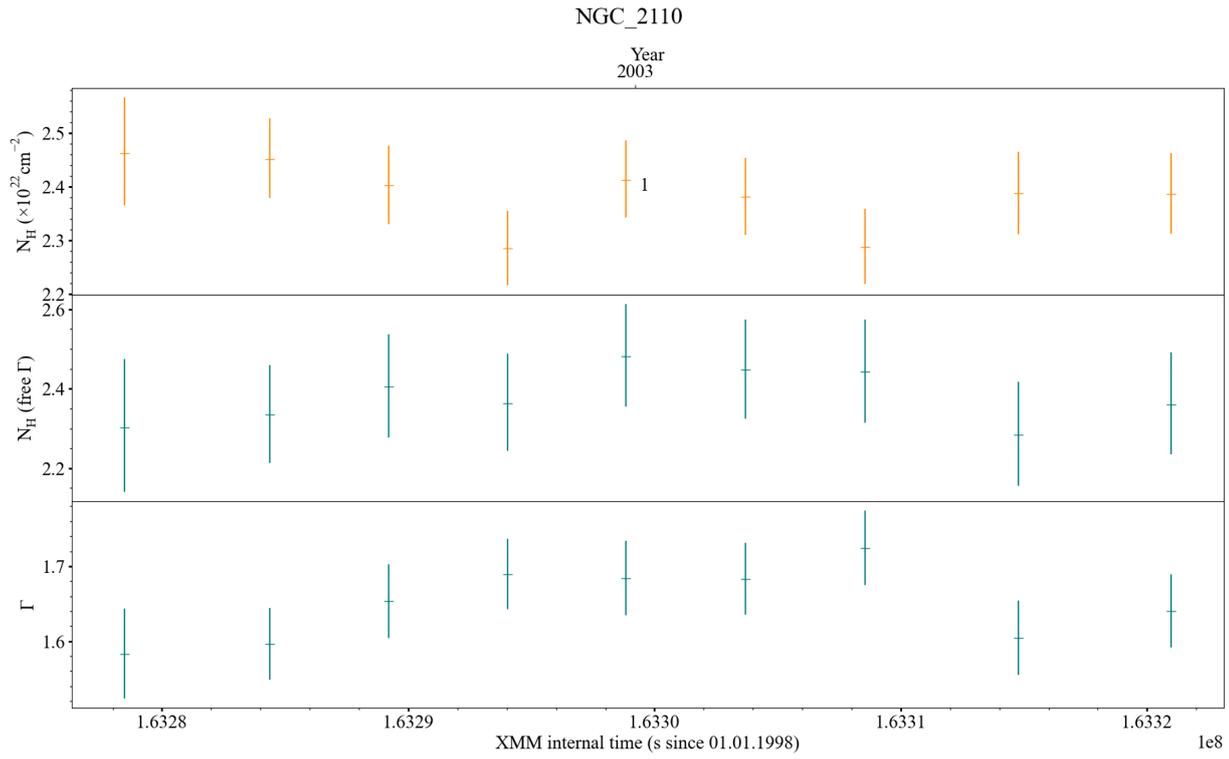
Figure A.14: Same as in Figure 4.5, for source NGC 2110
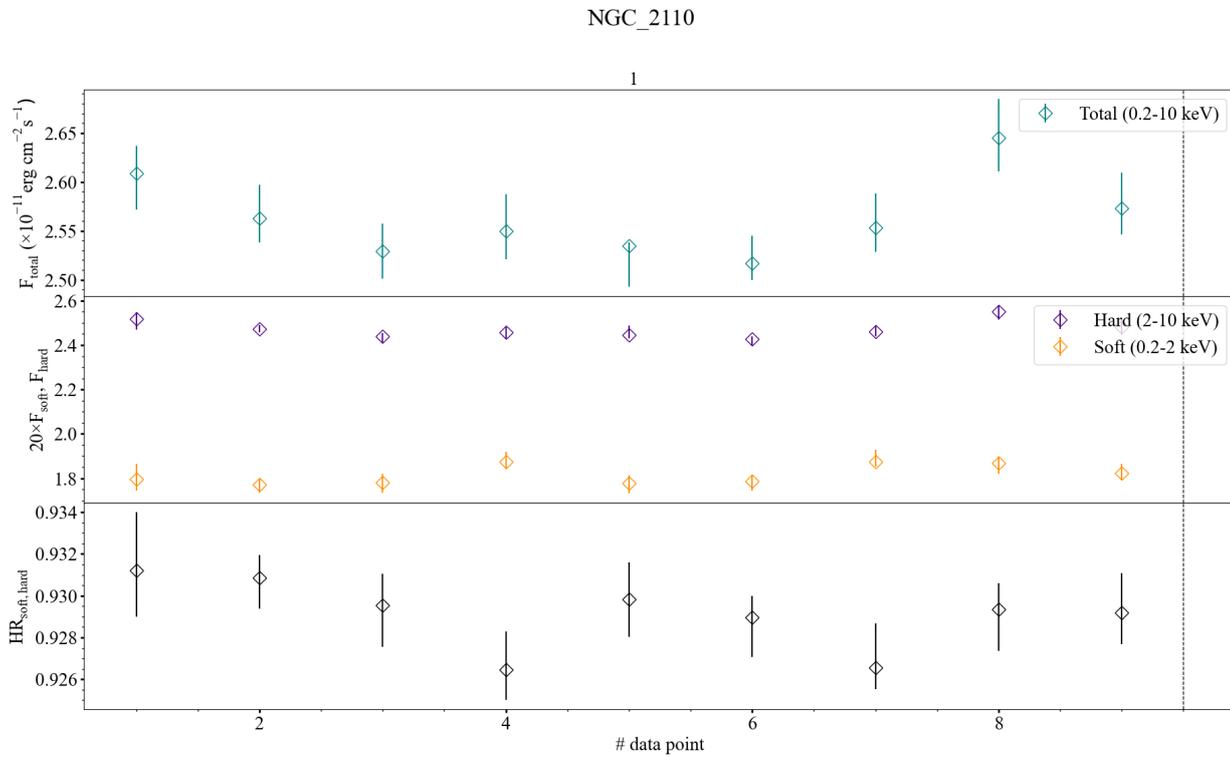
Figure A.15: Same as in Figure 4.2, for source NGC 2110

A.4 NGC 4258

The time evolution curves resulting from the initial binning of this source (5 ks) were shown in Section 4.1.3. The results in this Section correspond to the re-binning to reduce the N_{H} uncertainties to less than 10% of the average value.

Spectra
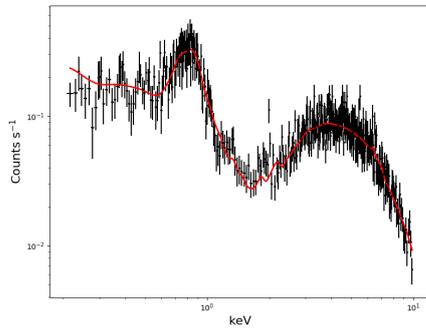
(a) Obs. ID 0110920101
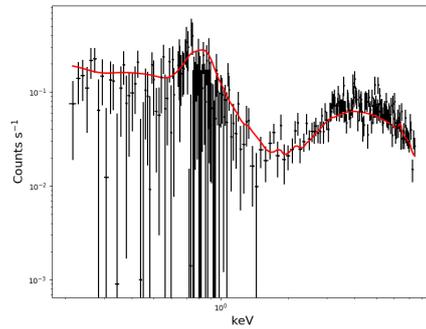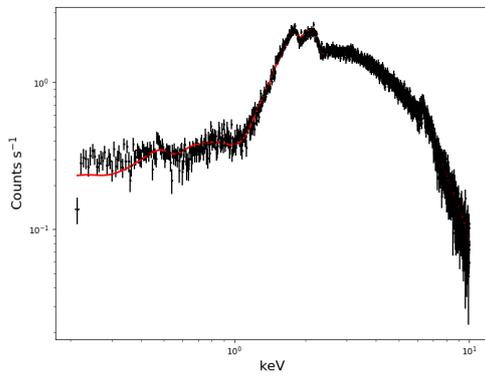
(b) Obs. ID 0059140101
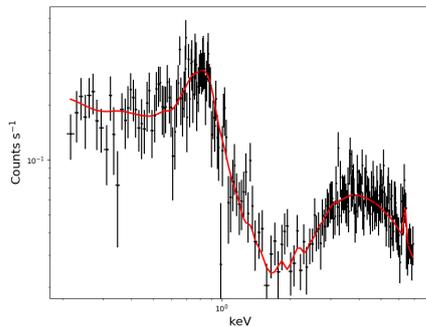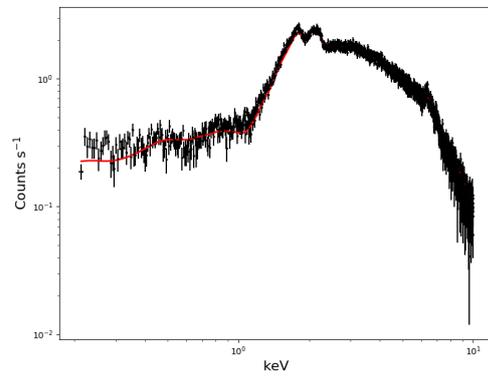
(c) Obs. ID 0059140201
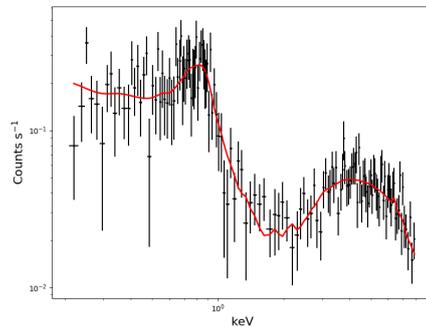
(d) Obs. ID 0059140401
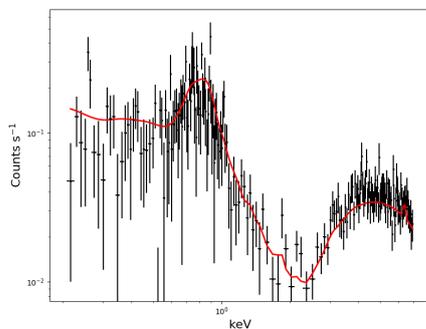
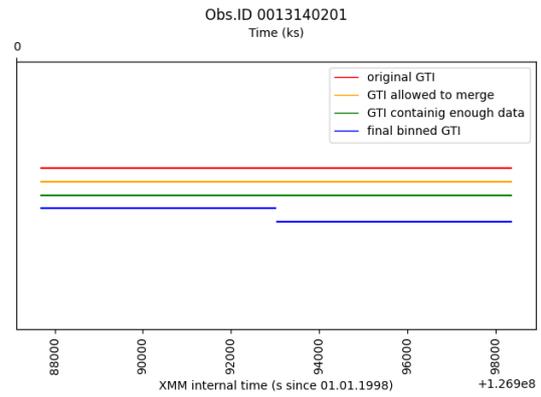(e) Obs. ID 0059140901
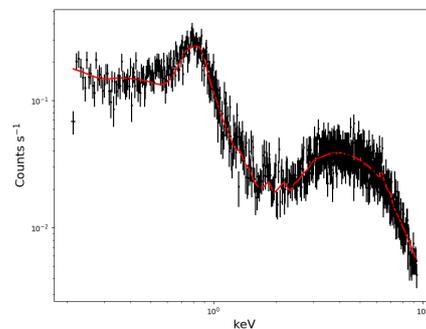
(f) Obs. ID 0400560301

Figure A.16: Same as in Figure 3.15 for the observations of source NGC 4258

Time binning
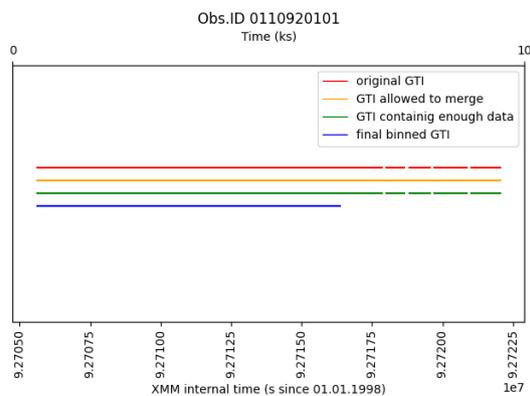
(a) Time binning Obs. ID 0110920101
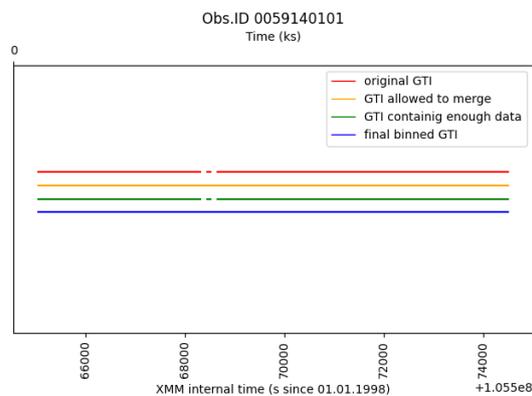
(b) Time binning Obs. ID 0059140101
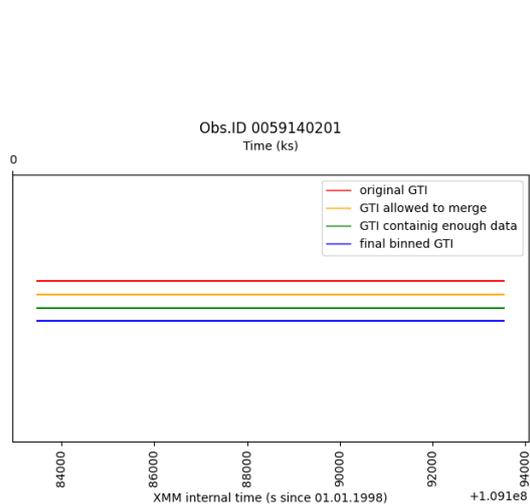
(c) Time binning Obs. ID 0059140201
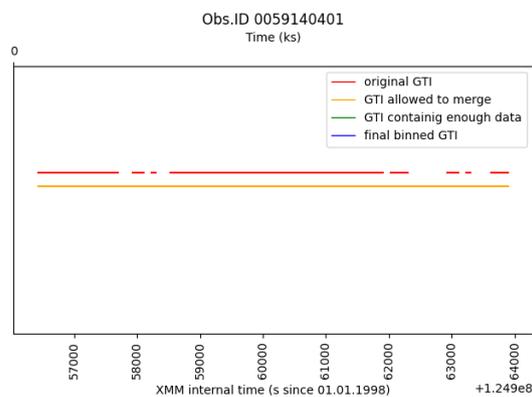
(d) Time binning Obs. ID 0059140401. No final bins are created as the original GTI do not contain enough data to form a single bin.
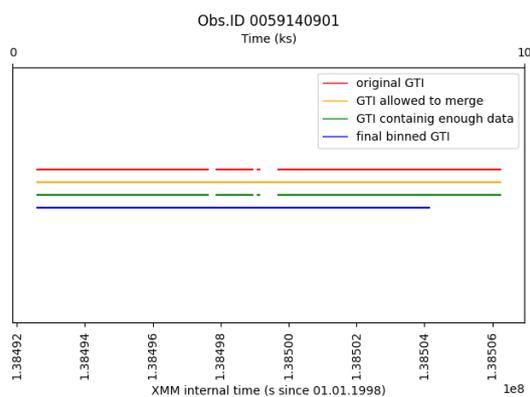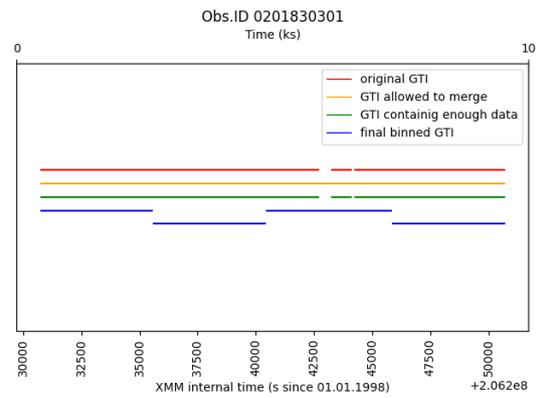

(e) Time binning Obs. ID 0059140901
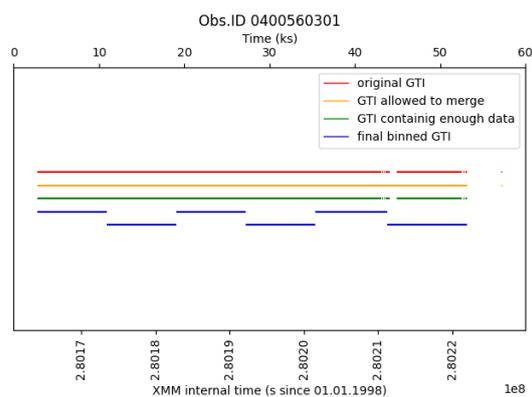
(f) Time binning Obs. ID 0400560301

Figure A.17: Same as in Figure 3.13, for source NGC 4258

Obs. index	Obs. ID	Bin data length (s)
1	0110920101	10799.99
2	0059140101	9299.99
3	0059140201	10099.99
4	0059140901	10799.99
5	0400560301	9383.33

Table A.4: Time bin length for observations of NGC 4258

Time evolution curves
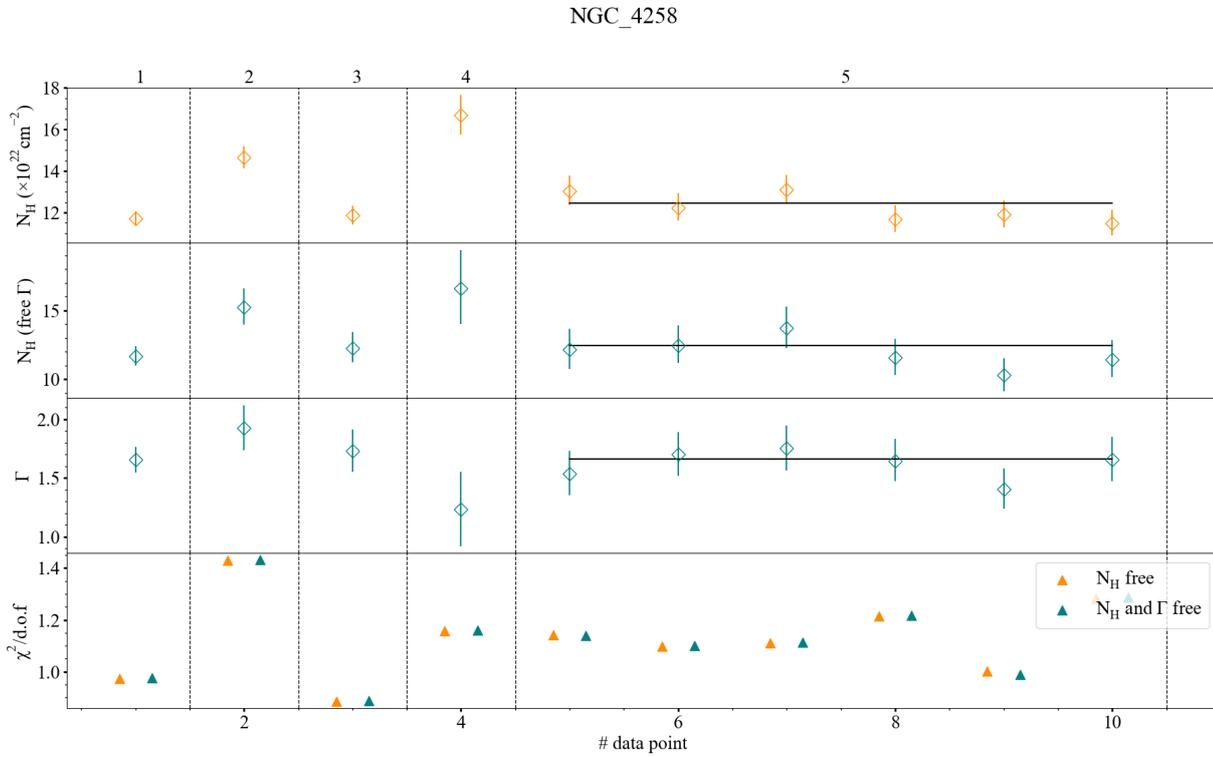
Figure A.18: Same as in Figure 4.1, for source NGC 4258
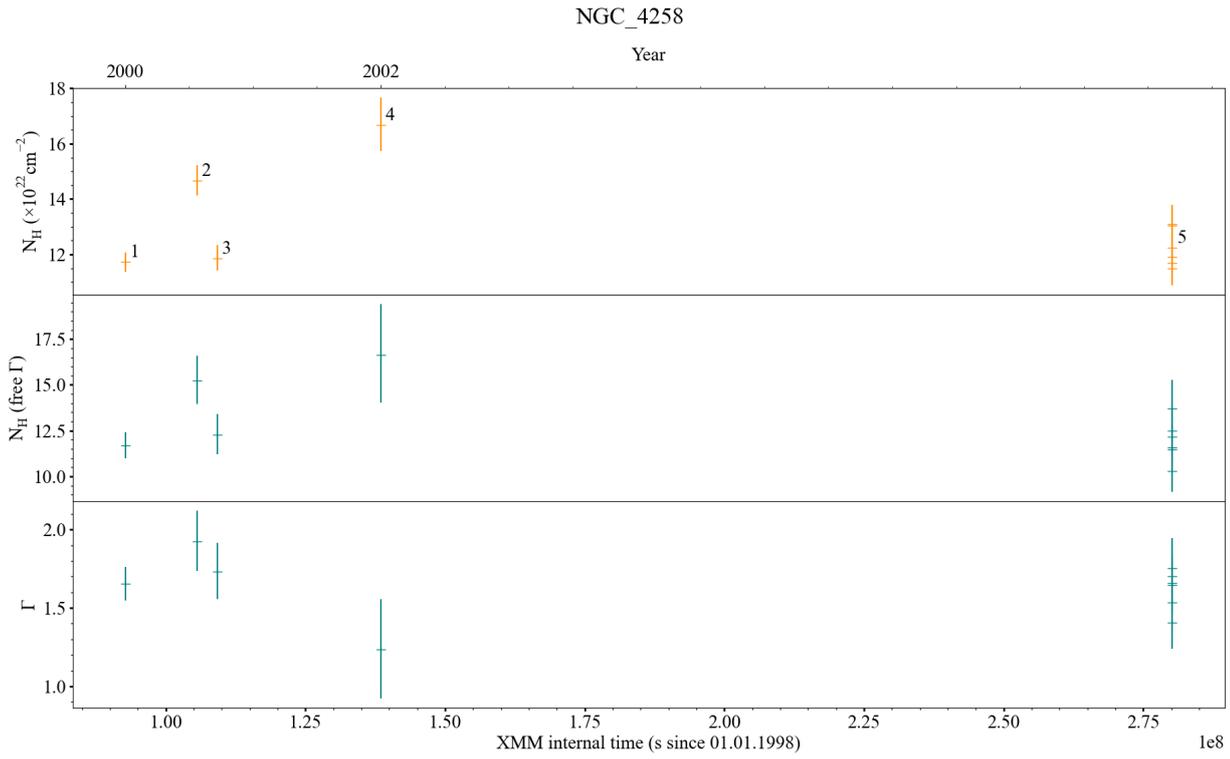
Figure A.19: Same as in Figure 4.5, for source NGC 4258
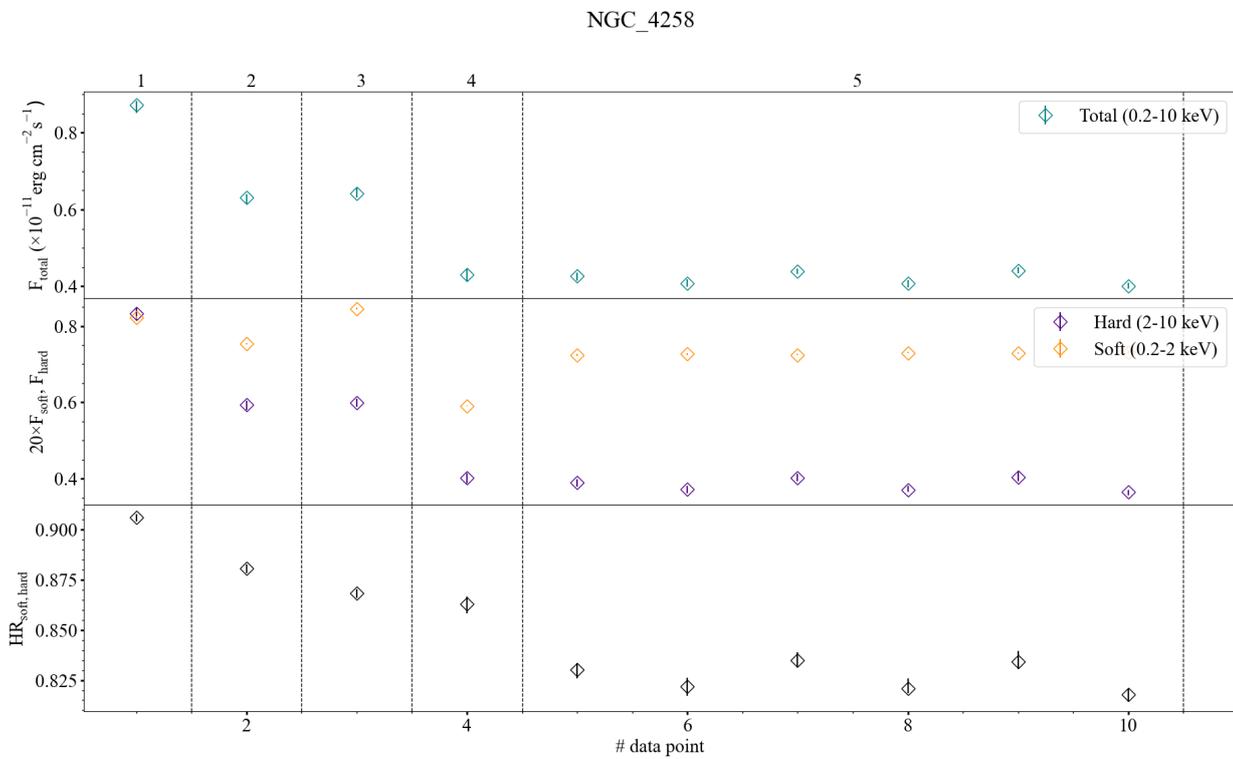
Figure A.20: Same as in Figure 4.2, for source NGC 4258

A.5   NGC 4507

The time evolution curves resulting from the initial binning of this source (5 ks) were shown in Section 4.1.3. The results in this Section correspond to the re-binning to reduce the N_{H} uncertainties to less than 10% of the average value.

Spectra

(a) Obs. ID 0006220201

(b) Obs. ID 0653870201

(c) Obs. ID 0653870301

(d) Obs. ID 0653870401

(e) Obs. ID 0653870501

(f) Obs. ID 0653870601

Figure A.21: Same as in Figure 3.15 for the observations of source NGC 4507

Time binning
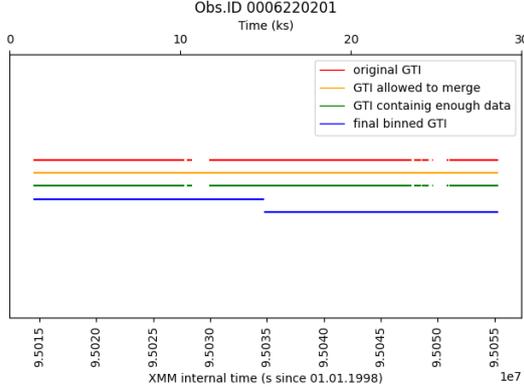
(a) Time binning Obs. ID 0006220201
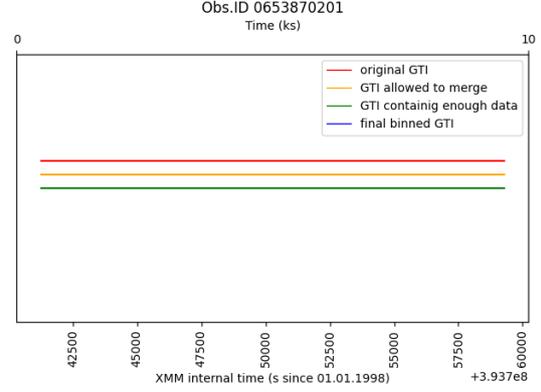
(b) Time binning Obs. ID 0653870201.

(c) Time binning Obs. ID 0653870301.
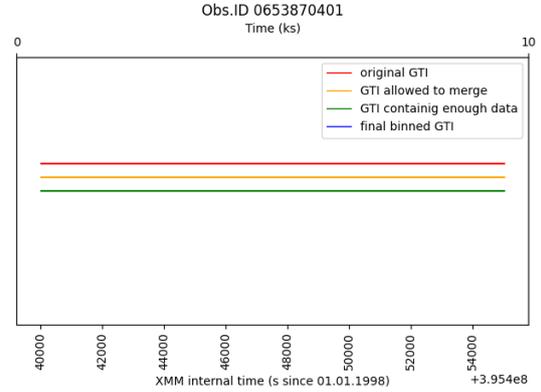
(d) Time binning Obs. ID 0653870401.
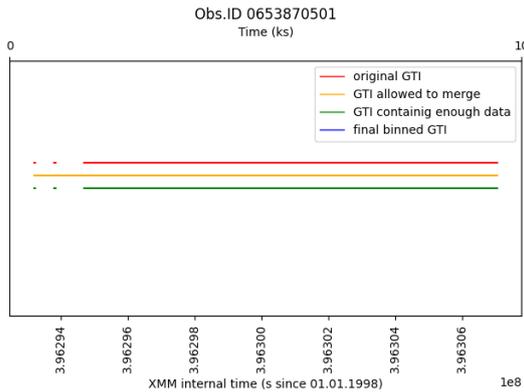
(e) Time binning Obs. ID 0653870501.
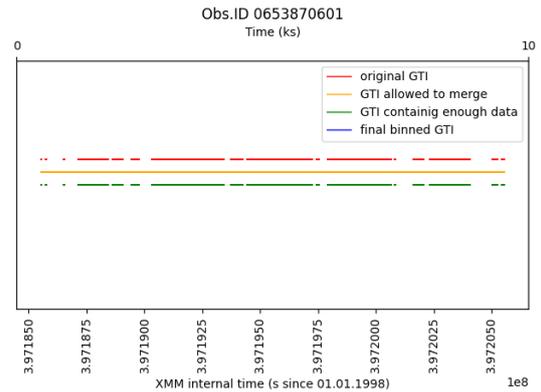
(f) Time binning Obs. ID 0653870601.

Figure A.22: Same as in Figure 3.13, for source NGC 4507. No final bins are created in panels b to f as the observations are not long enough to create a single bin of 20 ks.

Obs. index	Obs. ID	Bin data length (s)
1	0006220201	18649.99

Table A.5: Time bin length for observations of NGC 4507

Time evolution curves
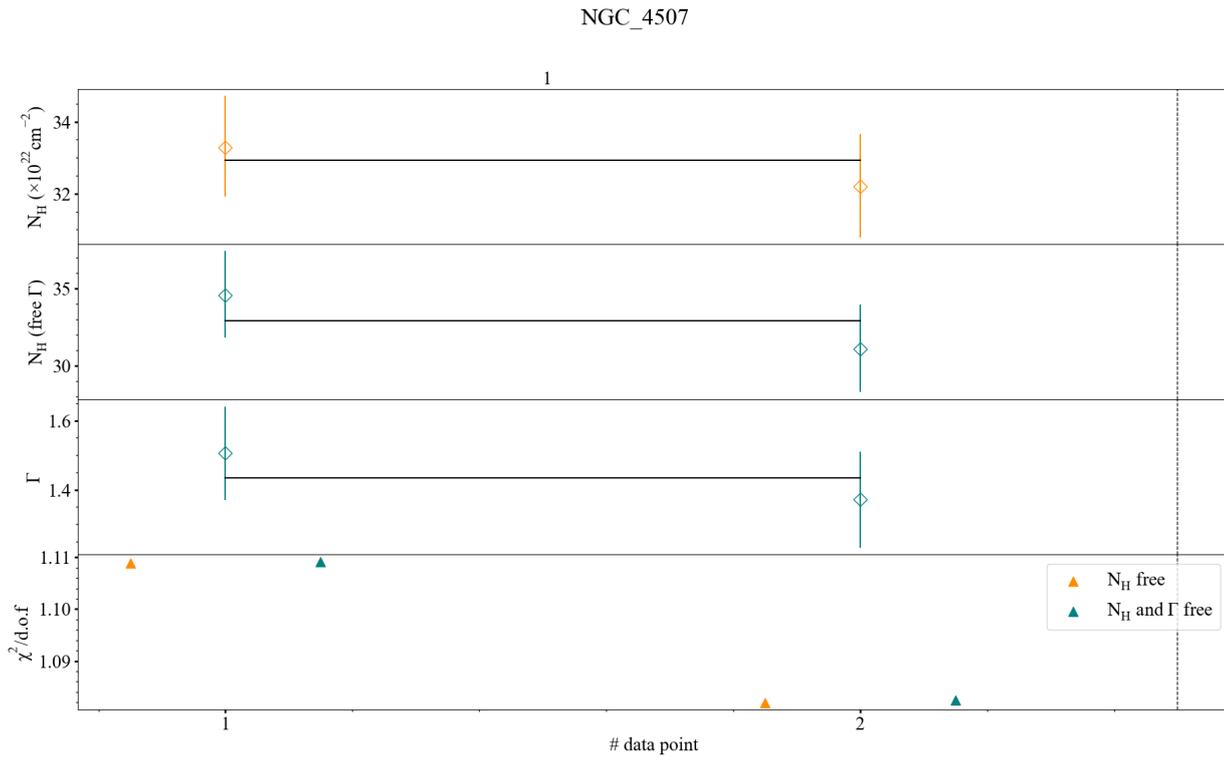
Figure A.23: Same as in Figure 4.1, for source NGC 4507
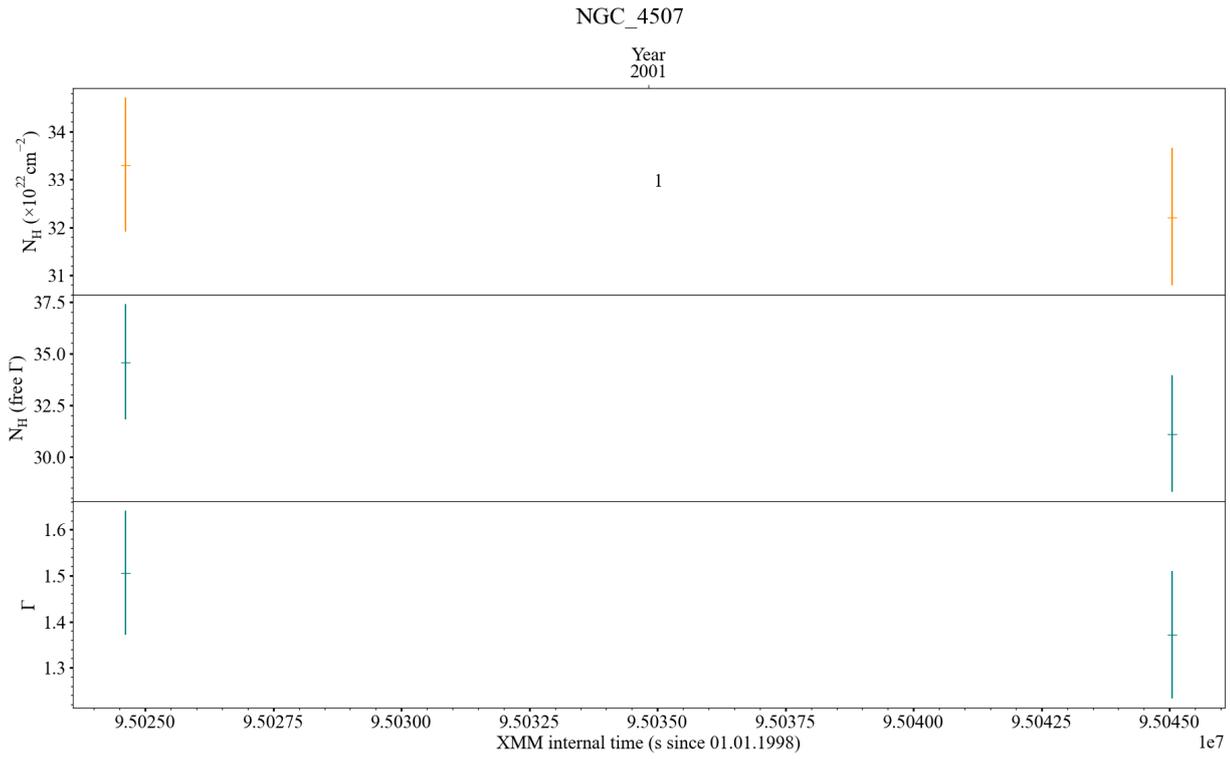
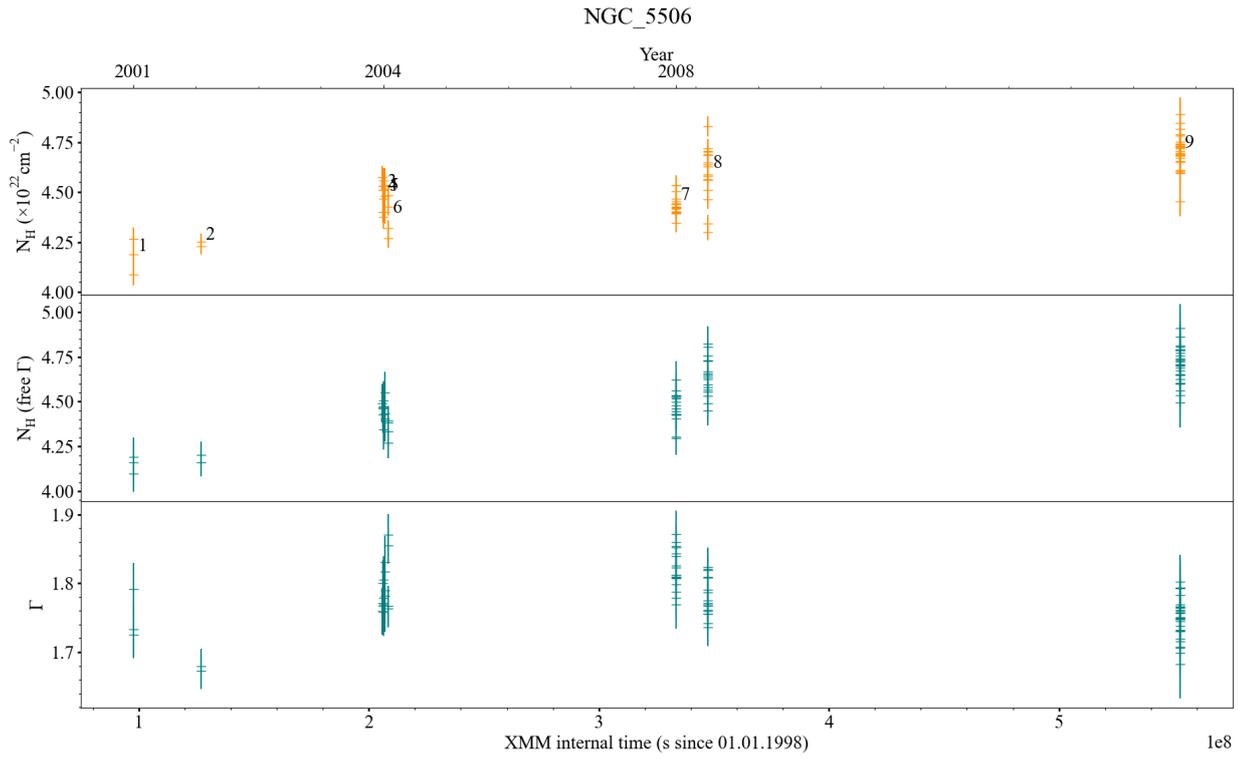**Figure A.24:** Same as in Figure 4.5, for source NGC 4507
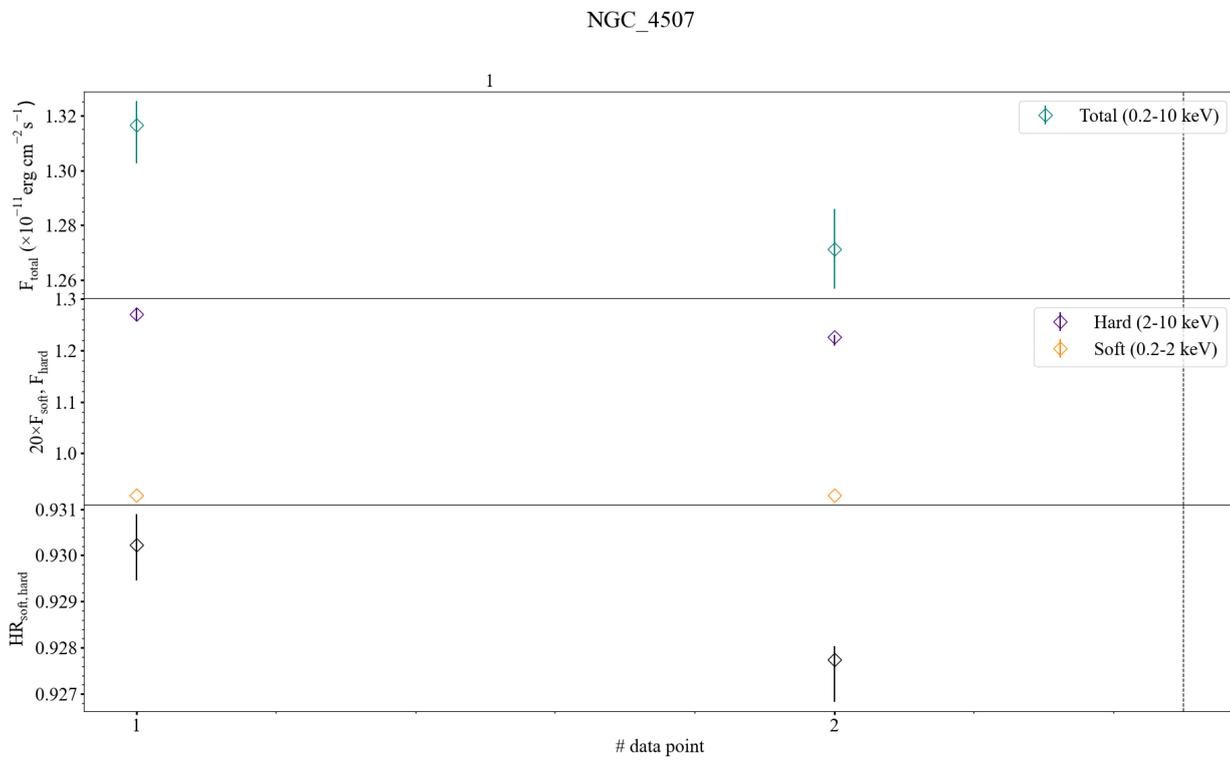
Figure A.25: Same as in Figure 4.2, for source NGC 4507

A.6 NGC 5506

Spectra
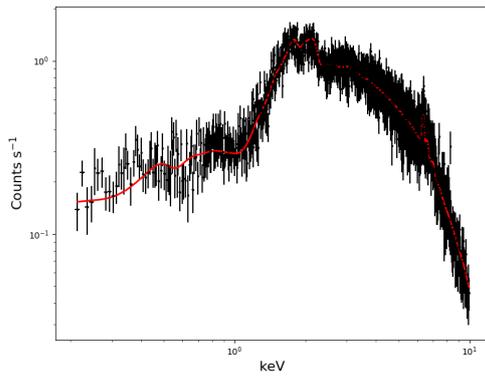
(a) Obs. ID 0013140101
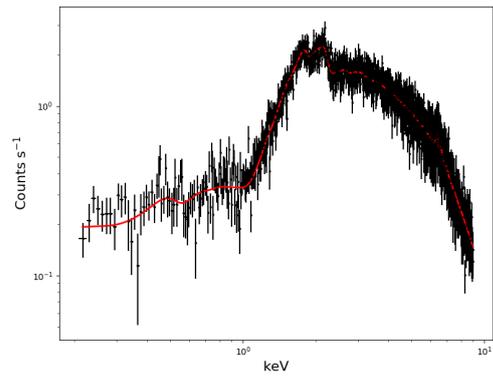
(b) Obs. ID 0013140201
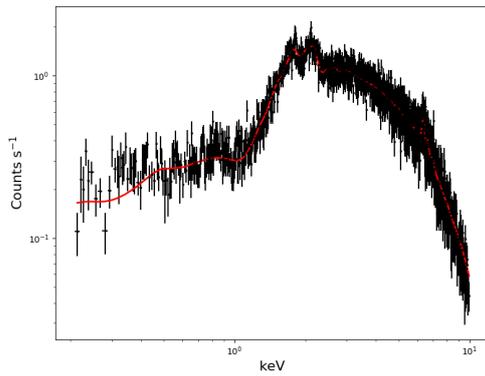
(c) Obs. ID 0201830201
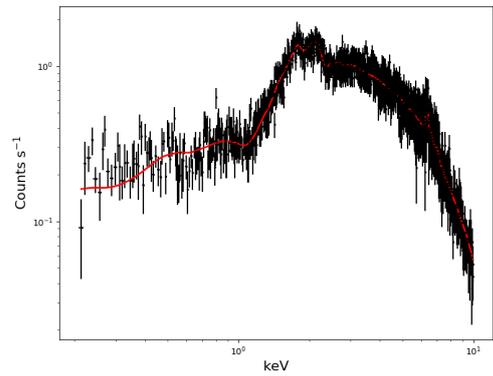
(d) Obs. ID 0201830301
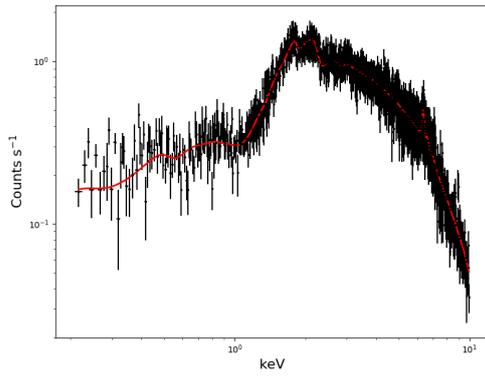
(e) Obs. ID 0201830401
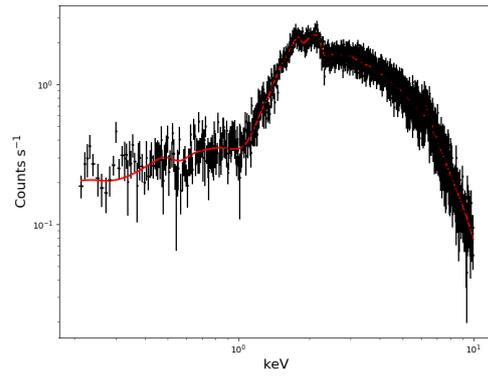
(f) Obs. ID 0201830501

(g) Obs. ID 0554170201

(h) Obs. ID 0554170101
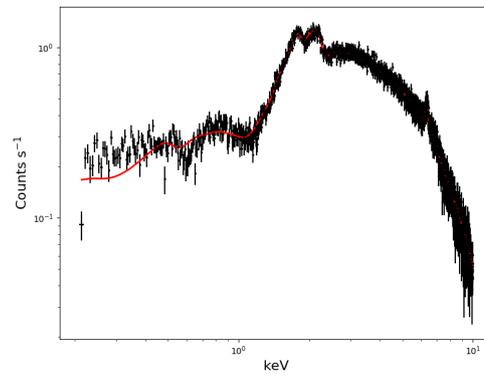
(i) Obs. ID 0761220101

Figure A.26: Same as in Figure 3.15 for the observations of source NGC 5506

Time binning
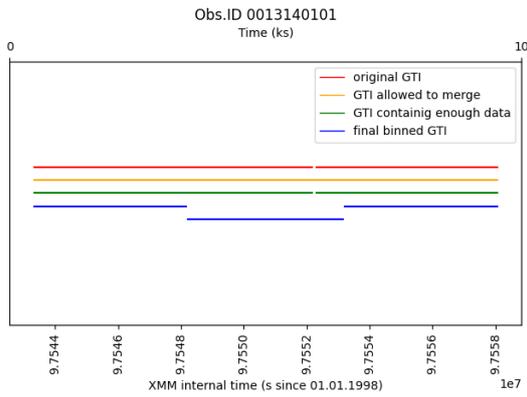
(a) Time binning Obs. ID 0013140101

(b) Time binning Obs. ID 0013140201
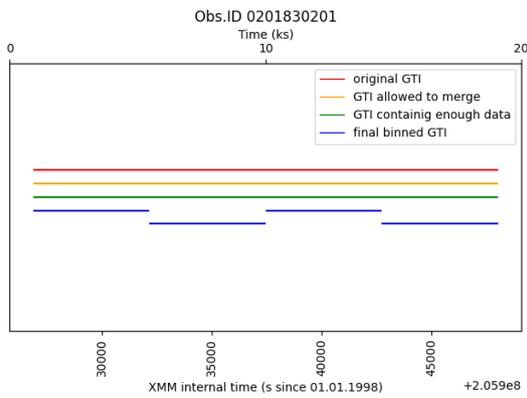
(c) Time binning Obs. ID 0059140201

(d) Time binning Obs. ID 0201830301
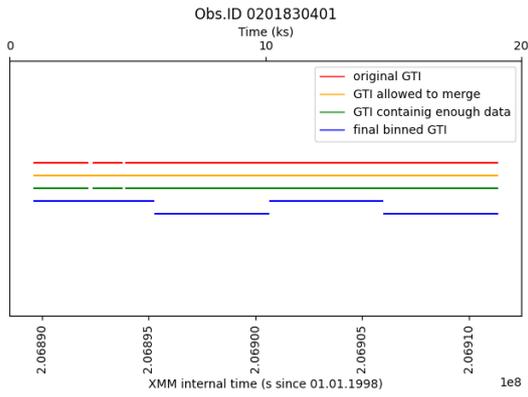
(e) Time binning Obs. ID 0201830401
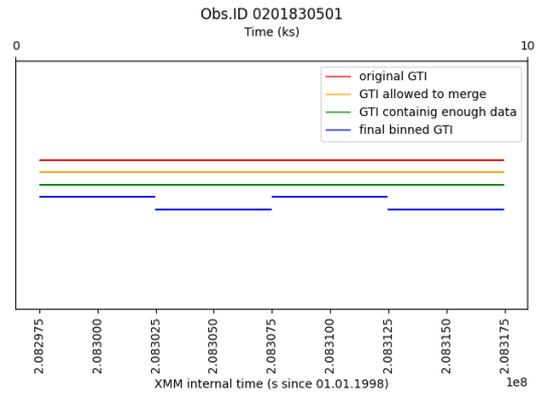
(f) Time binning Obs. ID 0201830501
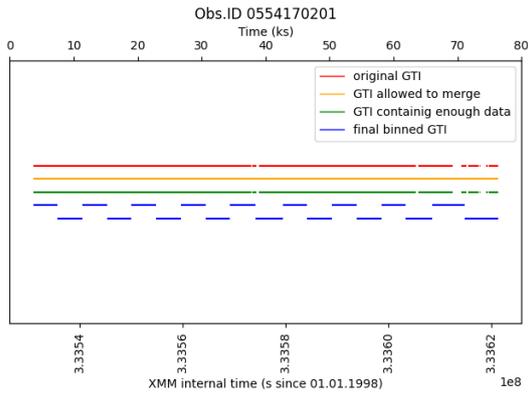
(g) Time binning Obs. ID 0554170201
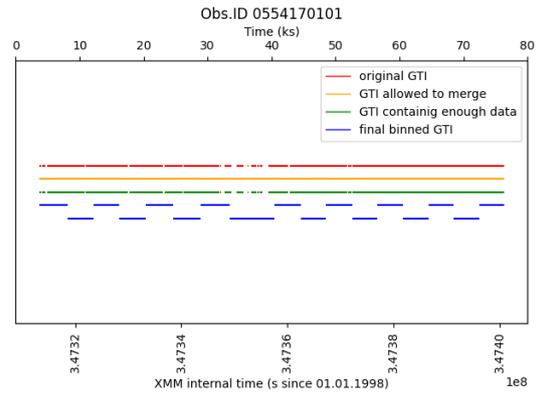
(h) Time binning Obs. ID 0554170101
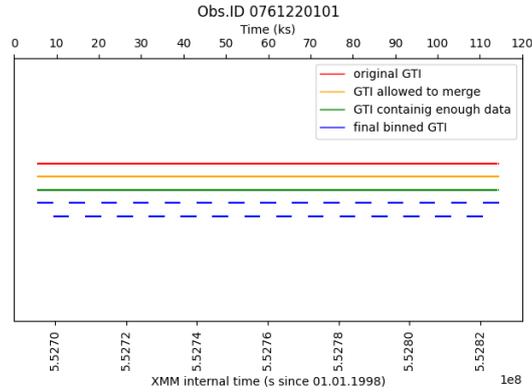
(i) Time binning Obs. ID 0761220101

Figure A.27: Same as in Figure 3.13, for source NGC 5506

Obs. index	Obs. ID	Bin data length (s)
1	0013140101	4899.99
2	0013140201	5349.99
3	0059140201	5299.99
4	0201830301	4849.99
5	0201830401	5374.33
6	0201830501	4999.33
7	0554170201	4794.44
8	0554170101	4770.58
9	0761220101	4496.55

Table A.6: Time bin length for observations of NGC 5506

Time evolution curves
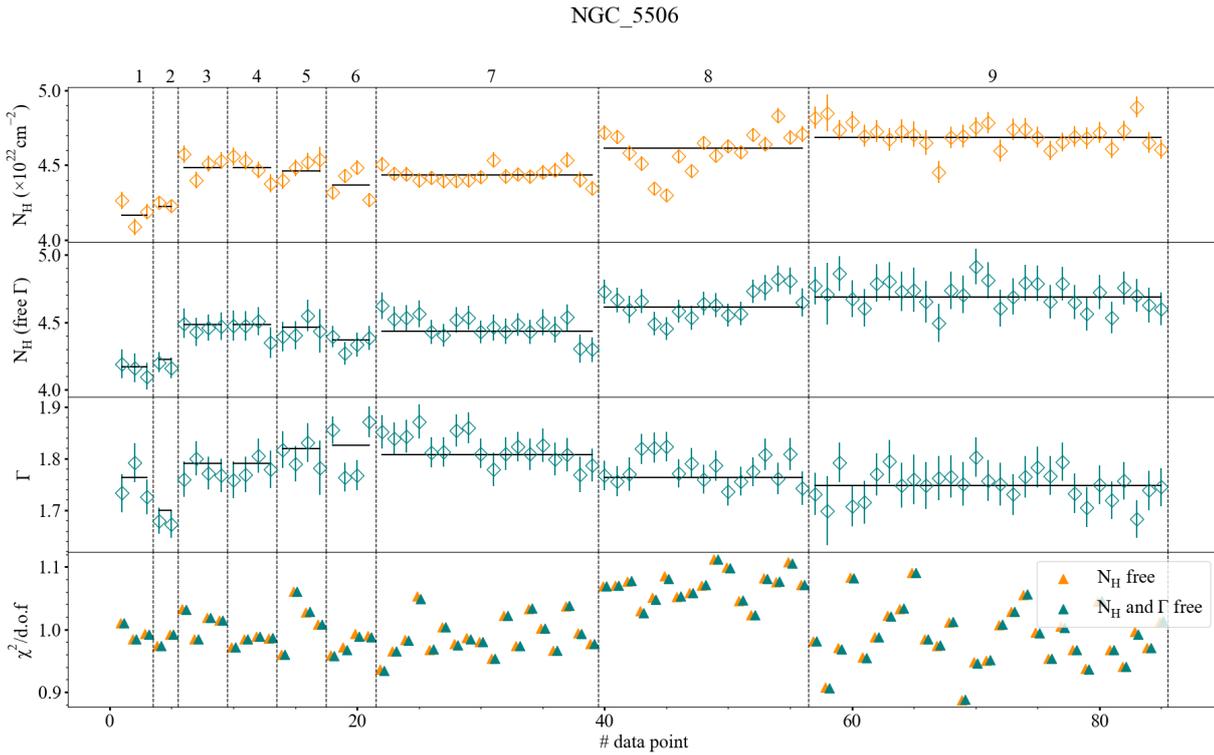
Figure A.28: Same as in Figure 4.1, for source NGC 5506

Figure A.29: Same as in Figure 4.5, for source NGC 5506
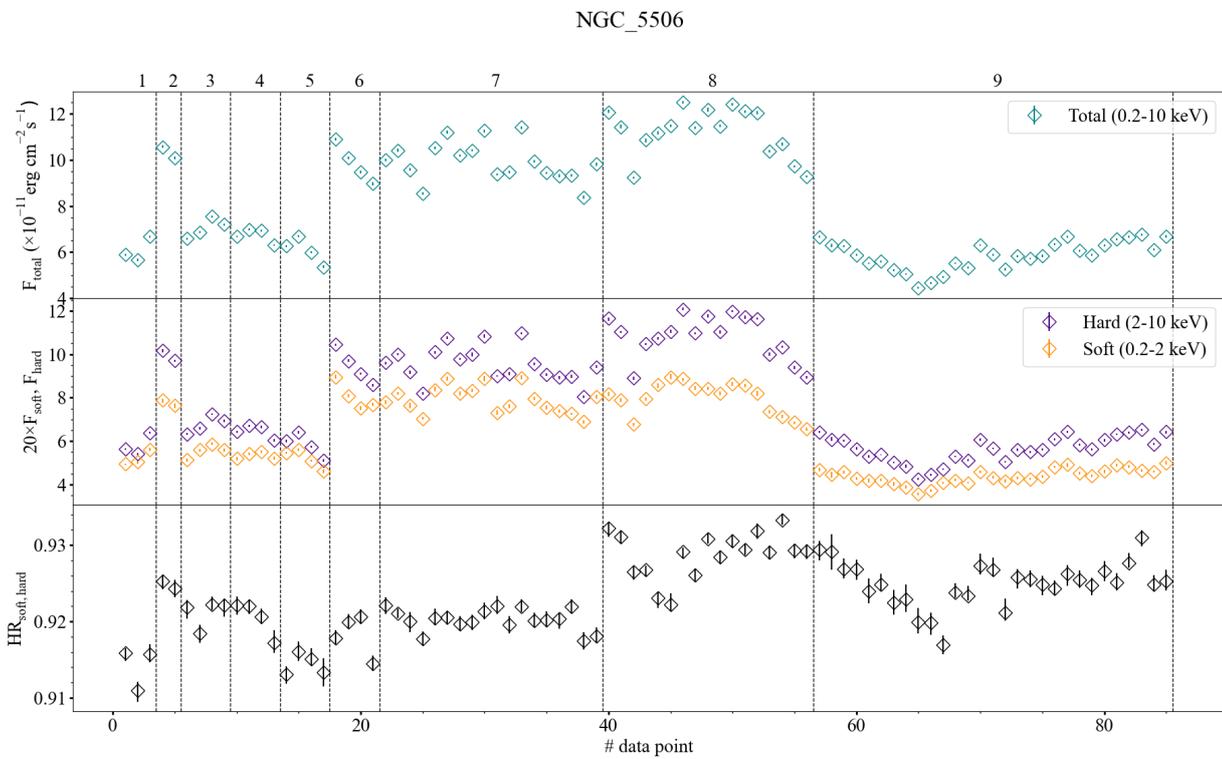
Figure A.30: Same as in Figure 4.2, for source NGC 5506

Bibliography

Alexander D. M., Hickox R. C., 2012, *What drives the growth of black holes?*, New Astronomy Reviews, 56, 93

Almeida C. R., et al., 2009, *The infrared nuclear emission of Seyfert Galaxies on parsec scales: Testing the clumpy torus models*, The Astrophysical Journal, 702, 1127

Antonucci R., 1993, *Unified models for active galactic nuclei and quasars*, Annual review of astronomy and astrophysics, 31, 473

Baloković M., et al., 2020, *NuSTAR Survey of Obscured Swift/BAT-selected Active Galactic Nuclei. II. Median High-energy Cutoff in Seyfert II Hard X-Ray Spectra*, The Astrophysical Journal, 905, 41

Blandford R., Meier D., Readhead A., 2019, *Relativistic jets from active galactic nuclei*, Annual Review of Astronomy and Astrophysics, 57, 467

Buitrago-Casas J., et al., 2020, *Use of a ray-tracing simulation to characterize ghost rays in the FOXSI rocket experiment*, Journal of Instrumentation, 15, P11032

Cackett E. M., Bentz M. C., Kara E., 2021, *Reverberation mapping of active galactic nuclei: From X-ray corona to dusty torus*, Iscience, 24, 102557

Chevallier L., Collin S., Dumont A.-M., Czerny B., Mouchet M., Gonçalves A., Goosmann R., 2006, *The puzzle of the soft X-ray excess in AGN: absorption or reflection?*, arXiv preprint astro-ph/0601464

Den Herder J., et al., 2001, *The reflection grating spectrometer on board XMM-Newton*, Astronomy & Astrophysics, 365, L7

Dermer C., Schlickeiser R., Mastichiadis A., 1992, *High-energy gamma radiation from extragalactic radio sources*, Astronomy and Astrophysics (ISSN 0004-6361), vol. 256, no. 2, p. L27-L30., 256, L27

Done C., Nayakshin S., 2007, *Can the soft excess in AGN originate from disc reflection?*, Monthly Notices of the Royal Astronomical Society: Letters, 377, L59

Done C., Gierlinski M., Sobolewska M., Schurch N., 2007, *The Origin of the soft excess in AGN*, arXiv preprint astro-ph/0703449

ESA 2023, *XMM-Newton Users Handbook*, https://xmm-tools.cosmos.esa.int/external/xmm_user_support/documentation/uhb/

ESA. XMM-Newton 2023, *Introduction to SAS*, https://www.cosmos.esa.int/web/xmm-newton/what-is-sas

Elitzur M., 2006, *Unification issues and the AGN torus*, arXiv preprint astro-ph/0612458

Elitzur M., Nenkova M., Ivezic Z., 2003, *IR emission from AGNs*, arXiv preprint astro-ph/0309040

Fabian A., 1999, in Highlights in X-ray Astronomy. p. 127

Ferrarese L., Merritt D., 2000, *A fundamental relation between supermassive black holes and their host galaxies*, The Astrophysical Journal, 539, L9

García-Bernete I., et al., 2022, *Torus and polar dust dependence on active galactic nucleus properties*, Astronomy & Astrophysics, 667, A140

Gaskell C. M., 2007, *Accretion disks and the nature and origin of AGN continuum variability*, arXiv preprint arXiv:0711.2113

Georgakakis A., Aird J., Schulze A., Dwelly T., Salvato M., Nandra K., Merloni A., Schneider D. P., 2017, *Observational constraints on the specific accretion-rate distribution of X-ray-selected AGNs*, Monthly Notices of the Royal Astronomical Society, 471, 1976

George I., Fabian A., 1991, *X-ray reflection from cold matter in active galactic nuclei and X-ray binaries*, Monthly Notices of the Royal Astronomical Society, 249, 352

Griffin A. J., Lacey C. G., Gonzalez-Perez V., Lagos C. d. P., Baugh C. M., Fanidakis N., 2019, *The evolution of SMBH spin and AGN luminosities for z¡ 6 within a semi-analytic model of galaxy formation*, Monthly Notices of the Royal Astronomical Society, 487, 198

Halpern J., 1984, *Variable X-ray absorption in the QSO MR 2251-178*, Astrophysical Journal, Part 1 (ISSN 0004-637X), vol. 281, June 1, 1984, p. 90-94., 281, 90

Harrison C., 2017, *Impact of supermassive black hole growth on star formation*, Nature Astronomy, 1, 0165

IAS. Natural Sciences 2010, *HEASoft*, https://www.ias.edu/sns/computing/heasoft

Ishibashi W., Courvoisier T.-L., 2010, *X-ray power law spectra in active galactic nuclei*, Astronomy & Astrophysics, 512, A58

Jaffe W., et al., 2004, *The central dusty torus in the active nucleus of NGC 1068*, Nature, 429, 47

Kalberla P. M., Kerp J., 2009, *The Hi distribution of the milky way*, Annual review of astronomy and Astrophysics, 47, 27

Kamraj N., et al., 2022, *X-Ray Coronal Properties of Swift/BAT-selected Seyfert 1 Active Galactic Nuclei*, The Astrophysical Journal, 927, 42

Kormendy J., Ho L. C., 2013, *Coevolution (or not) of supermassive black holes and host galaxies*, Annual Review of Astronomy and Astrophysics, 51, 511

Krolik J. H., Begelman M. C., 1988, *Molecular tori in Seyfert galaxies-Feeding the monster and hiding it*, Astrophysical Journal, Part 1 (ISSN 0004-637X), vol. 329, June 15, 1988, p. 702-711. Research supported by the Ball Corp., Rockwell International Corp., and Exxon Education Foundation., 329, 702

Krumpe M., 2008, Doctoral thesis, Universität Potsdam

Laha S., Dewangan G. C., Kembhavi A. K., 2011, *X-Ray Warm Absorption and Emission in the Polar-scattered Seyfert 1 Galaxy Mrk 704*, The Astrophysical Journal, 734, 75

Laha S., Markowitz A. G., Krumpe M., Nikutta R., Rothschild R., Saha T., 2020, *The variable and non-variable X-ray absorbers in Compton-thin type II active galactic nuclei*, The Astrophysical Journal, 897, 66

Lamer G., Uttley P., 2003, *An absorption event in the X-ray light curve of NGC 3227*, Monthly Notices of the Royal Astronomical Society, 342, L41

Lenk R., Gellert W., 1989, *Physik - Band 2/Ma-Z*. VEB F.A. Brockhaus Verlag, Leipzig, DDR

Li W., Hudson M., 2019, *Earth's Van Allen radiation belts: From discovery to the Van Allen Probes era*, Journal of Geophysical Research: Space Physics, 124, 8319

Liu B., Qiao E., 2022, *Accretion around black holes: The geometry and spectra*, Iscience, 25, 103544

Markowitz A., Krumpe M., Nikutta R., 2014, *First X-ray-based statistical tests for clumpy-torus models: eclipse events from 230 years of monitoring of Seyfert AGN*, Monthly Notices of the Royal Astronomical Society, 439, 1403

Mason K., et al., 2001, *The XMM-Newton optical/UV monitor telescope*, Astronomy & Astrophysics, 365, L36

Middleton M., Done C., Schurch N., 2007, *High Energy spectra of Seyferts and Unification schemes*, arXiv preprint arXiv:0710.5093

Nandra K., Pounds K., 1994, *Ginga observations of the X-ray spectra of Seyfert galaxies*, Monthly Notices of the Royal Astronomical Society, 268, 405

Nenkova M., Sirocky M. M., Nikutta R., Ivezić Ž., Elitzur M., 2008, *AGN dusty tori. II. Observational implications of clumpiness*, The Astrophysical Journal, 685, 160

Netzer H., Woltjer L., 1990, *Active Galactic Nuclei*. Springer-Verlag

Nikutta R., Elitzur M., Lacy M., 2009, *On the 10 μm silicate feature in active galactic nuclei*, The Astrophysical Journal, 707, 1550

Paczynski B., 1982, *Thick accretion disks around black holes/Karl Schwarzschild Lecture 1981*, Mitteilungen der Astronomischen Gesellschaft Hamburg, 57, 27

Padovani P., et al., 2017, *Active galactic nuclei: what's in a name?*, The Astronomy and Astrophysics Review, 25, 1

Peterson B. M., 1997, *An introduction to active galactic nuclei.* Cambridge University Press

Remillard R. A., McClintock J. E., 2006, *X-ray properties of black-hole binaries*, Annu. Rev. Astron. Astrophys., 44, 49

Ricci C., Trakhtenbrot B., 2022, *Changing-look Active Galactic Nuclei*, arXiv preprint arXiv:2211.05132

Ricci C., Ueda Y., Ichikawa K., Paltani S., Boissay R., Gandhi P., Stalevski M., Awaki H., 2014, *The narrow Fe Kα line and the molecular torus in active galactic nuclei: an IR/X-ray view*, Astronomy & Astrophysics, 567, A142

Ricci C., et al., 2017, *BAT AGN spectroscopic survey. V. X-ray properties of the Swift/BAT 70-month AGN catalog*, The Astrophysical Journal Supplement Series, 233, 17

Risaliti G., et al., 2009, *The XMM–Newton long look of NGC 1365: uncovering of the obscured X-ray source*, Monthly Notices of the Royal Astronomical Society: Letters, 393, L1

Rivers E., Markowitz A., Rothschild R., 2011, *An occultation event in Centaurus A and the Clumpy Torus Model*, The Astrophysical Journal Letters, 742, L29

Rivers E., Markowitz A., Rothschild R., 2013, *Full Spectral Survey of Active Galactic Nuclei in the Rossi X-ray Timing Explorer Archive*, The Astrophysical Journal, 772, 114

Shakura N. I., Sunyaev R. A., 1973, *Black holes in binary systems. Observational appearance.*, Astronomy and Astrophysics, 24, 337

Siebenmorgen R., Heymann F., Efstathiou A., 2015, *Self-consistent two-phase AGN torus models-SED library for observers*, Astronomy & Astrophysics, 583, A120

Stalevski M., Fritz J., Baes M., Nakos T., Popović L. Č., 2012, *3D radiative transfer modelling of the dusty tori around active galactic nuclei as a clumpy two-phase medium*, Monthly Notices of the Royal Astronomical Society, 420, 2756

Stalevski M., Ricci C., Ueda Y., Lira P., Fritz J., Baes M., 2016, *The dust covering factor in active galactic nuclei*, Monthly Notices of the Royal Astronomical Society, 458, 2288

Strüder L., et al., 2001, *The European photon imaging camera on XMM-Newton: the pn-CCD camera*, Astronomy & Astrophysics, 365, L18

Treister E., Urry C. M., et al., 2012, *The cosmic history of black hole growth from deep multiwavelength surveys*, Advances in Astronomy, 2012

Trump J. R., et al., 2011, *Accretion rate and the physical nature of unobscured active galaxies*, The Astrophysical Journal, 733, 60

Turner M. J., et al., 2001, *The European photon imaging camera on XMM-Newton: the MOS cameras*, Astronomy & Astrophysics, 365, L27

Urry C. M., Padovani P., 1995, *Unified schemes for radio-loud active galactic nuclei*, Publications of the Astronomical Society of the Pacific, 107, 803

Vasudevan R., Fabian A., Gandhi P., Winter L., Mushotzky R., 2010, *The power output of local obscured and unobscured AGN: crossing the absorption barrier with Swift/BAT and IRAS*, Monthly Notices of the Royal Astronomical Society, 402, 1081

Voit G. M., 2019, *Ambient column densities of highly ionized oxygen in precipitation-limited circumgalactic media*, The Astrophysical Journal, 880, 139

Volonteri M., 2010, *Formation of supermassive black holes*, The Astronomy and Astrophysics Review, 18, 279

Waddell S. G., et al., 2023, *The eROSITA Final Equatorial Depth Survey (eFEDS): Complex absorption and soft excesses in hard X-ray–selected active galactic nuclei*, arXiv preprint arXiv:2306.00961

Werk J. K., et al., 2014, *The COS-Halos survey: physical conditions and baryonic mass in the low-redshift circumgalactic medium*, The Astrophysical Journal, 792, 8

Whittle M., 1985, *The narrow line region of active galaxies–I. Nuclear [O III] profiles*, Monthly Notices of the Royal Astronomical Society, 213, 1

Willingale R., Starling R., Beardmore A. P., Tanvir N. R., O'Brien P. T., 2013, *Calibration of X-ray absorption in our Galaxy*, Monthly Notices of the Royal Astronomical Society, 431, 394

XMM-Newton 2023, *XMM-Newton Users Handbook*, https://xmm-tools.cosmos.esa.int/external/xmm_user_support/documentation/uhb/XMM_UHB.pdf

XMM-Newton Calibration Technical Note 2013, *PN X-Ray Loading*, https://xmmweb.esac.esa.int/docs/documents/CAL-TN-0050-1-1.pdf

XMM-Newton Science Analysis System 2023, *Patterns*, https://xmm-tools.cosmos.esa.int/external/sas/current/doc/emevents/node4.html

Xspec. An X-Ray Spectral Fitting Package 2024, *Users' Guide for version 12.14*, https://heasarc.gsfc.nasa.gov/xanadu/xspec/manual/XspecManual.html

Zhao D., Zhang C., Yuan W., Zhang S., Willingale R., Ling Z., 2017, *Geant4 simulations of a wide-angle x-ray focusing telescope*, Experimental Astronomy, 43, 267

Acknowledgements

Statement of originality

I hereby declare that this thesis is the product of my own work. All the assistance received in preparing this thesis and the sources used have been acknowledged following the guidelines on plagiarism provided by the Senate.

Silvia Carolina Rueda Vargas