Facultad de Ciencias

Caracterización Espectral Automatizada de Galaxias Activas altamente Oscurecidas

Automated spectral characterization of highly obscured active galaxies

Trabajo Fin de Master

Master Universitario en Física de Partículas y del Cosmos

Autor: Jorge Luis Orellana Cruz

Directores: Amalia Corral, Francisco J. Carrera

septiembre de 2019

Santander, España

INDEX

Abstract

Los núcleos galácticos activos (AGNs) oscurecidos son un ingrediente clave en los modelos de formación y evolución de galaxias. Los rayos X, gracias a su alto poder de penetración, siguen siendo el método más eficiente para detectarlos y caracterizarlos. En este trabajo proponemos un método automático para derivar la distancia o el redshift a estos AGNs basado únicamente en espectroscopía en rayos X. De este modo aprovechamos que una de las características espectrales más común de estos AGN es la presencia de una línea muy intensa de emisión de fluorescencia del hierro FeKα con energías entre 6.4 y 6.9 keV. El metodo se basa en la detección de picos de emisión en los espectros de rayos X observados, en principio asociados a dicha linea. La mejora consiste en un filtrado para reducir el ruido mediante técnicas de fourier. Esta mejora se ha realizado mediante el uso de simulaciones de AGNs Compton Thick a varios redshift, luminosidades y densidades columna.

Obscured active galactic nuclei (AGNs) are a key ingredient in galaxy formation and evolution models. X-rays, thanks to their high penetration power, remain the most efficient method to detect and characterize them. In this work we propose to develop an automatic method to derive the redshift to these AGNs based on X-ray spectroscopy. Thus, we take advantage of the fact that one of the most common spectral characteristics of these AGNs is the presence of a very intense fluorescence emission line of the iron FeKα at energies between 6.4 and 6.9 keV. The method is based on the detection of emission peaks in the observed X-ray spectra, in principle associated with the FeKα line. The improvement consists of filtering to reduce noise using fourier techniques. This improvement has been performed by using AGN Compton Thick simulations at various redshifts, luminosities and column densities.

keywords: AGN, X-ray data, spectrum, Compton Thick, column density, redshift, signal to noise ratio.

INTRODUCTION

Active Galactic Nuclei (AGN) are powered by the accretion of material onto a super massive black hole (SMBH) and are the brightest persistent objects in the universe. It is been known that the majority of the AGN are found to be obscured, so the study of the complete population requires the identification of both obscured and unobscured AGN. The obscured AGN fraction is found to be a function of the AGN luminosity and redshift. It is more likely to found obscured AGN in a rich environment that contains gas and dust. The lack of a complete detection and study of all AGN types could give a wrong view of the host galaxy and the environments in which AGN are found. We need to study these obscured AGN at distant redshift to understand the accretion history of the universe and evolution of galaxies.

Most of the emission of obscured AGN is not directly detected, this is cause by the presence of material between the accretion disk and the line of sight of the observer. According to R. Hickox and D. Alexander (2018), obscuration is defined as anything that absorbs emission and scatters a large fraction away from the line of sight of the observer. The obscuring medium is composed of gas and dust that opaque the emission coming from the central source. An alternative definition of an obscured AGN is the absence of emission from the (Broad Line Region) BLR in the optical waveband, this means that it is obscuration produced by dust. On the unified model, the accretion disk is surrounded by the torus and it is in this region where molecules and dust grains can form. The form of the torus suggests that the accretion disk emission can be directly detected from some line of sight, but in some cases it is obscured by the torus. There have been some research (for example J. Buchner and F. Bauer (2017)) that explain how obscuration can also come from the host galaxy. It is perfectly known that some AGN emission is suppressed because of the obscuration that the host galaxy produces, but the ability to observe the AGN emission depends on the luminosity and its spectrum, the characteristics that describes the host galaxy and finally the physical characteristics of the material obscuring the AGN.

Some AGN (Type II) are classified as "Compton Thick sources" when the column density reaches values above 10^{24} cm^{-2}, and the direct continuum is complete suppressed below 10 keV. It is only observed in reflected emission. The study of Compton thick (CT) sources is relevant because there is evidence that suggests that a significant fraction of AGN in the universe are obscured by Compton thick gas. Compton thick AGN presents huge amount of absorption in their spectra, but a very pronounced emission line which we can study to determine some main characteristics.

Absorption affects the X-ray emission through two processes, being the first one photoelectric absorption and the second one Compton scattering. Photoelectric absorption is effective at column densities of 10^{21} cm^{-2}, it depends on the energy and does not play a significant role

above 10 keV for $N_H < 10^{24}$ cm^{-2}. Compton scattering becomes significant above $N_H = 10^{24}$ cm^{-2} and does not have a dependence on energies below 150 keV. These absorbed AGN are very difficult to detect, but luckily, there is a strong emission line that we can used to detect them, and moreover, its physical properties. We know this line as the Iron emission line (FeKα) at 6.4 keV.

In this work we will focus on Compton Thick AGN because we need highly absorbed AGN to produce a strong emission line we can detect. We will start applying an existing method developed by (N. Castelló-Mor et al. 2011) that detects the energy were the emission line must be find out. We will evaluate how the method performs over CT AGN. To this end we will use simulated data for the upcoming Athena X-ray mission. We will try to optimize this method for CT AGN and Athena data, so in this work we simulated two sets of data, the first one without noise and absorption, the second one simulating absorption and a corresponding signal to noise ratio. For the analysis it was used a simple absorbed power law with an emission line at the rest frame energy of 6.4 keV. We analyze the data at an energy range of 0.5 to 10 keV and using only z = 1, 2, 3 and 4. At last it was necessary to perform a Fast Fourier Transform to study the signal to noise ratio on some spectra and used a low pass frequency filter to upgrade the detection process.

In chapter I the most relevant information needed in this work about Active Galactic Nuclei is presented. In chapter II we present the Athena mission and its instrumentation followed by the Xspec software implementation and the models needed in the analysis. In chapter III we discuss the Athena simulation spectral data, were we manage to perform it and some studies. Finally in chapter IV we present and discuss the direct method to determine the cosmological redshift, the low pass filter applied to the spectra and finally the results using the new method.

Chapter I

Active Galactic Nuclei

Active Galactic Nuclei (AGN) are powered by the accretion of material onto a super massive black hole (SMBH) and are the brightest persistent objects in the universe. AGN emit huge amounts of energy over the entire electromagnetic spectrum. However, the large radiative emission of most AGN cannot be seen directly because the accretion is hidden behind dust and gas that absorb this emission. This process known as obscuration presents an important challenge for uncovering the AGN population completely and understanding the evolution of super massive black holes. In this chapter a summary of the most relevant information about AGN is presented. Section 1.1 presents the AGN classification and the unification model for all the AGN, section 1.2 presents the most important X-ray emission process. In section 1.3 the main components for the AGN spectrum is presented and section 1.4 explains the physics behind the Iron spectra.

1.1 The AGN Paradigm

An Active Galactic Nucleus is a manifestation of gas accretion observed onto a super-massive black hole. According to Netzer (2013), the definition purely observational of an AGN is adopted by having at least one of the following fulfilled:

1. It contains a compact nuclear region emitting significantly beyond what is expected from stellar process typical of this type of galaxies.

2. It shows the clear signature of a non stellar continuum emitting process in its center.

3. Its spectrum contains strong emission lines with ratios that are typical of excitation by a non stellar radiation field.

4. It shows line and/or continuum variations.

As different kinds of AGN were discovered over the years, the classification has become more complex. We can list: Seyfert 1 galaxies, Seyfert 2 Galaxies, LINERS, QSOs, Quasars, Blazars and Radio Galaxies for example were a schematic representation is shown in figure 1.1. There is a simpler classification based on our line of sight. Type I-AGN are objects with little or no obscuration at all of the central source and also it presents very broad permitted lines. Type

II-AGN are sources that shows a completely obscuration from our line of sight to the center and shows permitted lines with FWHMs that are significantly smaller than the Type I-AGN (C. Megan Urry and Paolo Padovani 1995).

The main purpose is to find the most simple model possible to explain the different appearances we see in AGN since all are believed to share the same structure at their centers. In order to reconcile both theory and observations, unified models for AGN were developed (Antonucci 1993). As it is shown in figure 1.2, the principal components of an AGN are:

- **Super Massive Black Hole:** At the center of all AGNs it is believed that there is a SMBH with a mass in the range $10^5 - 10^9$ M$_\odot$.

- **Accretion Disk:** The accretion of gas around the SMBH produces an optically thick disk of material that emits thermally because of viscosity within the disk. The gas present in the disk has a range of temperatures of T=10^4-10^5 K (R. Hickox and D. Alexander 2018). The accretion of gas around the SMBH is a very efficient process in converting mass into energy.

- **Corona:** Near the innermost parts of the accretion disk, there is a hot corona that consist of very energetic electrons (T=10^9 K), which are responsible for the primary X–ray power law emission.

- **Broad line region (BLR):** In this region, broad permitted lines are produced in the optical/UV band. The BLR consists in many clouds of gas that moves rapidly by the gravitational potential of the SMBH. The size of the BLR is estimated to range between 0.01-0.1 pc and the velocity of the gas is in the range 1000-30000 km s^{-1}.

- **Torus:** It consist on a toroidal structure that extends from about 1 to 100 pc and contains cold gas and dust. The torus absorbs the X-ray spectra of AGNs and also scatters the X-ray photons which produce a Compton hump at an energy of 20 keV (Yuan Liu and Xiaobo Li 2014). The torus emit in the infrared as a broad peak.

- **Narrow line region (NLR):** This region is much further away from the nucleus than the BLR and they are responsible for the narrow emission lines we see in the optical/UV band. It is less turbulent and dense than the BLR with velocities of a few hundreds of Km s^{-1}.

- **Relativistic jet:** These are high energetic jets collimated by the accretion disk. They have speeds close to that of light and they are perpendicular to the accretion disk.
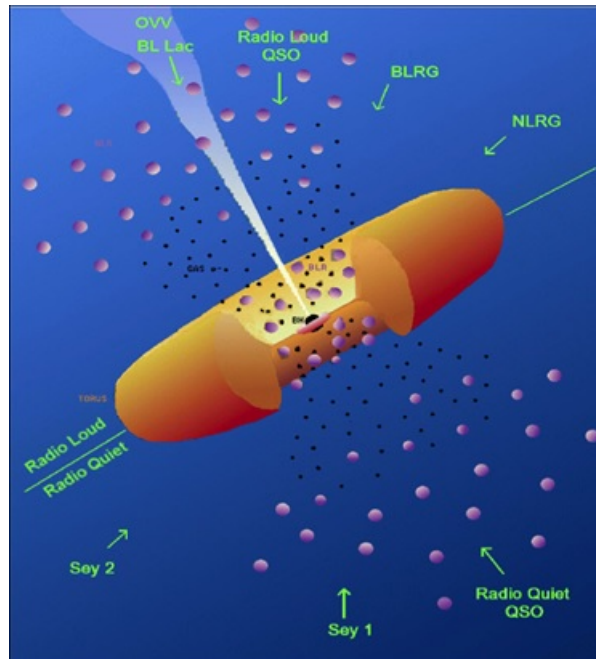
Figure 1.1: *Schematic representation of the AGN unified model. Taken from C. Megan Urry and Paolo Padovani (1995).*
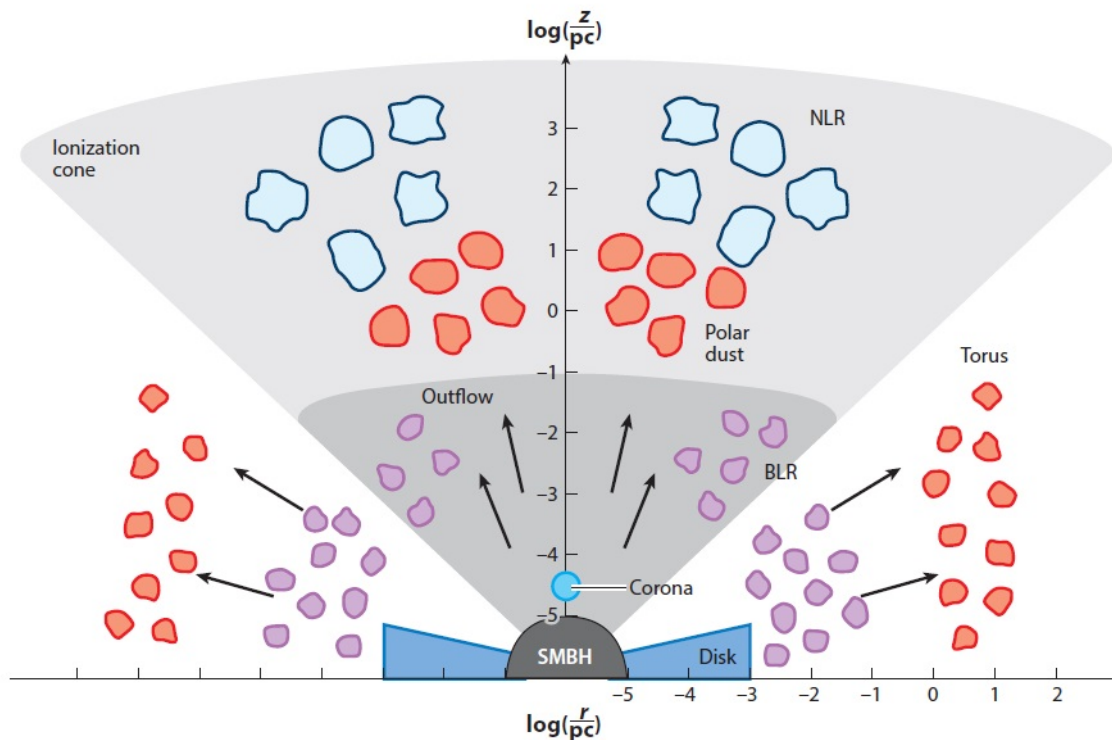

Figure 1.2: *Schematic representation of the AGN physical model, broadly illustrating the scales of the key regions. The accretion disk, corona, BLR, and torus reside within the gravitational influence of the SMBH. Taken from R. Hickox and D. Alexander (2018).*

In the unification model for AGN was established that the accretion disk and broad line region (BLR) are surrounded by a toroidal structure of gas and most of it composed by dust. The dust in the torus obscures the accretion disk for the lines of sight which pass through the torus. The actual geometry of the torus and its physical properties are still unknown, for example the thickness of the torus which determines the covering factor (S. F. Honig and T. Becker 2007).

AGN shows bolometric luminosities from 10^{41} to 10^{48} erg s^{-1}. The emission is present over all the entire electromagnetic spectrum. A large part of this emission appears to arise from a compact region near the nucleus. Because the focus of this work is X-ray spectra of AGN, in next section, a brief description of the primary X–ray emission is presented.

1.2   The AGN Emission

The source of the emission is believed to come from the accretion onto a SMBH. There is strong evidence for the presence of this SMBHs at the center of all galaxies, for example the observed X-ray variability that leads to an efficiency process in converting matter to energy that cannot be achieved by thermonuclear processes such as stars nuclei, the very high central mass derived by reverberation mapping of the broad emission lines, and velocity dispersion of stars that increases towards the nucleus which again implies a high concentration of mass.

As mentioned before, AGN emit a vast amounts of energy over the entire electromagnetic spectrum. Radio emission is strongly present in RL AGN, IR emission is consistent with thermal emission from dust at different temperatures, Optical/UV emission is how we classified obscured AGN, and Gamma ray emission its present in blazars for which most of their bolometric luminosities is on this band. The main discussion of this section is the X-ray emission.

Let discuss first what flux is. Flux, or radiant flux, is the total amount of energy that crosses a unit area per unit time. Flux is measured in erg/cm^2/s or photons/cm^2/s. The flux of an astronomical source depends on the luminosity of the object and its distance from the Earth, it is defined according to the inverse square law:

$$F = \frac{L}{4\pi r^2} \tag{1.1}$$

X-ray emission of an AGN can be approximated to first order by a powerlaw. The density F_ν and frequency ν is:

$$F_\nu \propto \nu^{-\alpha} \tag{1.2}$$

were α is the spectral index typically from 0.8 to 1. But the powerlaw is usually expressed in photon units as:

$$F_E \propto E^{-\Gamma} \tag{1.3}$$

were F_E is the photon flux per unit photon energy E and Γ is the photon index expressed as ($\Gamma = \alpha + 1$).

Over the powerlaw shape, in some data it appears an excess emission at the soft energies (< 2 keV) and in other cases it appears to be suppressed. Through different mechanisms the X-ray are reprocessed by the material surrounding the SMBH. One mechanism is the photo electric absorption by neutral material, which for X-rays the probability of being absorbed by neutral material is

$$P(E) = e^{-\sigma(E)N_H} \tag{1.4}$$

where E is the energy of the photon, σ(E) is the absorption cross-section and N_{H} is the Hydrogen column density, which is defined as the contribution of the various atomic species for a given abundance pattern. It is needed column densities over 10^{20} cm^{-2} in order to detect effects in AGNs with z = 0. As the redshift increases, larger column densities are necessary in order to suppress the X-ray spectral emission at low energies in the observer frame.

The X-rays from AGN appears to be directly associated with the accretion disk and it is believed to be produced in the corona, being the inverse Compton scattering (which is explained by lower energy photons produced and scattered to higher energies by relativistic electrons in the corona) the most important mechanism for producing it. The least energy photons can be produced in the most hottest regions of the disk and its emission is then modified interacting with the material in the accretion disk, such as photoelectric absorption, reflection and scattering.

The Spectral Energy Distribution (SED) of the accretion disk in AGNs is different form those of other astrophysical sources so it is easy to identify(see in figure 1.3). It is worth to mention that since the accretion disk is small and unresolved for even the brightest and closest AGN, we can compare them and identify them. The study of the spectral energy distributions over all the spectral range is an optimal way to characterise the properties of galaxies in general, according to M. A. Prieto et al. (2009). It is important to cover a big spectra range so we can study different physical phenomena such non thermal processes in the X-ray band.
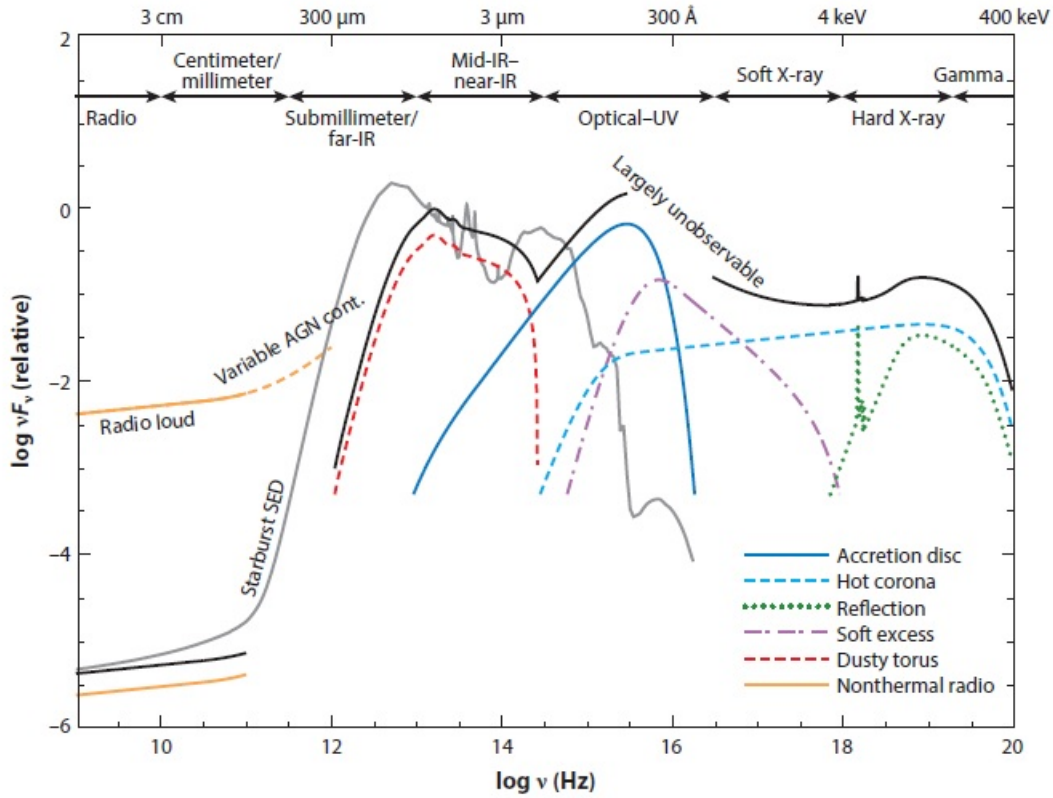
Figure 1.3. Schematic representation of the spectral energy distribution (SED) of an unobscured AGN (black curve), separated into the main physical components (colored curves) and compared to the SED of a star-forming galaxy (gray curve). From R. Hickox and D. Alexander (2018).

1.3 The AGN X-ray Spectra

Figure 1.4 shows and example of the observed components of an AGN spectra. Many AGN show a feature on their spectra that appears to be an excess emission below 2 keV above the power law emission. This feature is called the **soft-excess**. It had been seen that the strength of the soft excess is correlated with the spectral index Γ and with the Eddington ratio λ_{Edd} (see Rozenn Boissay et al. (2016)). They suggested that the soft excess present in 80% of the objects of the sample is likely not due to blurred ionized reflection, but can most probably be explained by warm Comptonization.
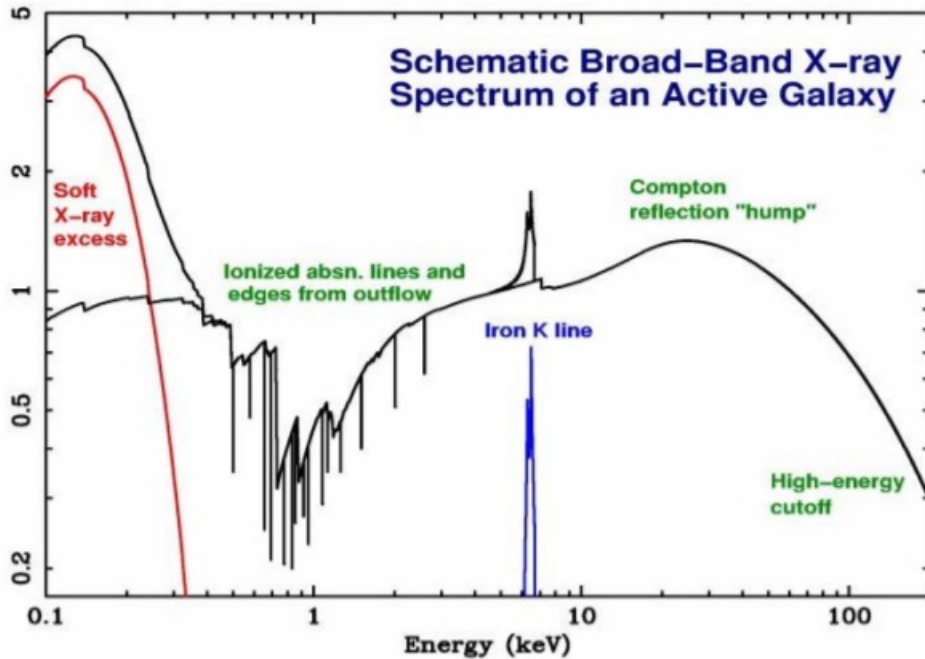
Figure 1.4. A Schematic representation of the broad band X-ray spectrum of an AGN.

Some AGN presents an absorption in their X-ray spectra that is sufficiently strong and that it is difficult to measure the photon index in these kind of objects. Because of the neutral absorption, we see that the photon index is flatter than the one shown in Type-I AGNs. Comastri A. (2004) established that if the X-ray obscuring matter has a column density which is equal to or larger than the inverse of the Thomson cross-section ($N_\mathrm{H} \geq \sigma_T^{-1} = 1.5 \times 10^{24}$ cm^{-2}), then the source is called, by definition, Compton thick. In this AGN the direct continuum is complete suppressed below 10 keV and we observe reflected emission. We study Compton thick AGN because there are evidence that a big population of obscured AGN are Compton thick sources.

1.4 Iron in X-ray spectra

There are some AGN that show two X-ray spectral signatures known in Compton thick obscuration. The first one is a flat X-ray spectrum and the second one is a large equivalent width of the FeKα emission line. This emission line is produced in the obscuring medium, and equivalent width greater than 1 keV is characteristic of Compton thick AGN (N. A. Levenson 2013).

The FeKα line is the most prominent emission line in the rest frame energy range of 6–7 keV, it provides an important view of the phenomena produced in the disk. When the line is produced in the inner parts of the accretion disk, which is close to the SMBH, the effects of the gravitational field and relativistic effects may alter the lines profile. This cause to show a profile to be broad and asymmetric. On the other hand, if the line comes from the outer part of the accretion disk or the dusty torus, it shows a narrow profile. This work is focused on the narrow FeKα emission line and extremely strong in obscured CT AGN.

In more detail, because of the high luminosity and extremely high temperatures, the material is being evaporated from the surface of the disk. The corona radiates X-rays in all directions

and the disk is thick enough that not all the atoms are completely ionized even that it is very hot. Iron is able to retain electrons in the inner parts at temperatures up to million degrees. Now by scattering electrons, the X-rays that hit the disk are absorbed, putting them in an excited state which then it will decay by the emission of X-rays at energy 6.4 keV. This process is known as **fluorescence**. The numerical calculations of these process is shown in Figure 1.5 where it can be seen that iron FeKα dominates over other elements.

The narrow FeKα line is found to be present in all AGN as it is one of the main signatures of X-ray reprocessed radiation from the material near the super massive black hole. There have been many research for unveiling the most significant properties and characteristics of AGN near the SMBH using the FeKα line, see for example Corral (2011) and C. Ricci et al. (2014).
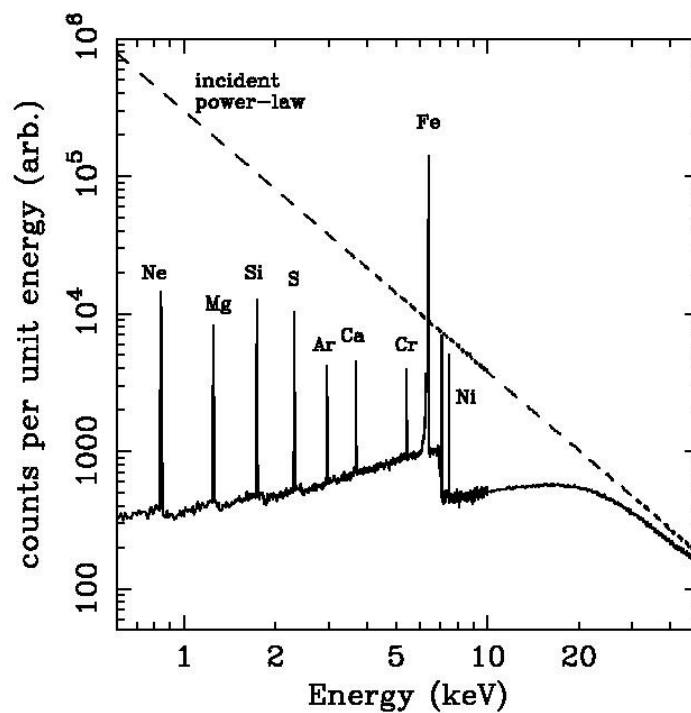

Figure 1.5. Numerical calculations of the effect of irradiating an accretion disk with a hard X-ray spectrum (the incident powerlaw). The resultant reflected X-ray spectrum is dominated by iron, the most abundant and heavy element present, and with the highest fluorescent yield. Taken from Seward & Charles (2010).

Chapter II

X-ray Data Analysis

In this chapter we discuss the main criteria for studying and identifying AGN in X-ray waveband. We also explain the Athena mission and the instrumentation related to it. Finally the main software, methods and models used to analyze X-ray data are presented.

The X-ray band is defined as the energy range 0.2-200 keV. It provides one of the most efficient methods for selecting obscured AGN. The Earth's atmosphere is opaque at X-ray energies so all astronomical X-ray observations have been obtained from space.

Brandt and Alexander (2015) states that X-rays are the best waveband to select obscured AGN because we can separate them from other type of sources, for example, binary stars emit at lower energies than AGN, which can be detected and identify above 10 keV. The signature of absorption indicates that X-ray spectroscopy provides very confident and precise methods in the identification of these obscured AGN, also measuring the amount of absorption at Compton thin column densities. Compton thick absorption signatures are more difficult to detect, but it can be identified in high signal to noise ratio spectra. To observe that a spectra indicates Compton thick absorption it is necessary a strong reflection at E > 10 keV and a very prominent emission line at 6.4 keV with a equivalent with > 1 keV.

There is no single waveband that provides a complete selection of AGN in a confident way. X-rays can penetrate very high amounts of material, and the contamination which are produced by the host galaxy is low in this band. This allow us for a reliable selection of obscured AGN, but the problem is when we search for low luminosity objects. However, Compton thick AGN are weak and can be very difficult to identified at X-ray energies.

High sensitivity is needed to detect and characterize CT AGN up to high z. The Athena mission (presented in next section) will provide unprecedented sensitivity and a bigger field of view. It will allow to study and observe distant AGN obscured and Compton thick, such as the ones that we are studying in this work.
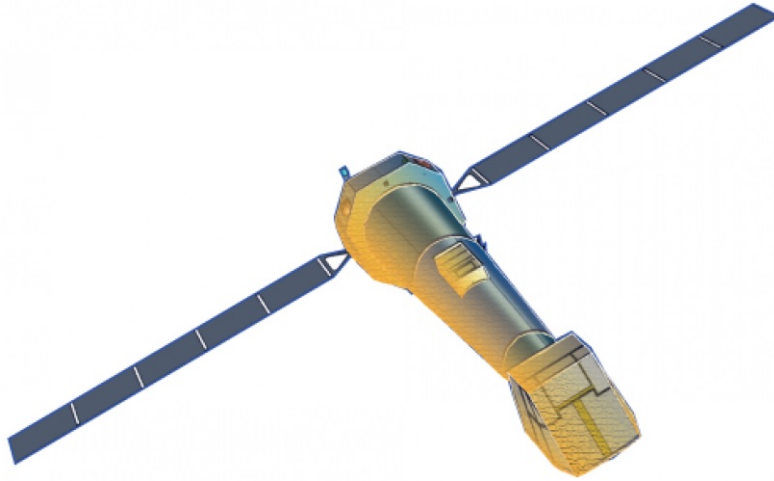
2.1 Athena X-ray Observatory

Figure 2.1: A conceptual design for the Athena spacecraft derived from the ESA CDF study, designed to be accommodated in an Ariane 5 launcher. Credit: ESA and ACO team.

The focus of this work X-ray data is on Athena (Advanced Telescope for High ENergy Astrophysics) shown in figure 2.1. It is a future X-ray telescope of the European Space Agency (ESA), under development for launch around early 2030s. It is the second (L2) large class mission within ESA Cosmic Vision Program. Athena will be at least one order of magnitude more sensitive than the best of the X-ray telescopes up to date such as the *Chandra* and *XMM-Newton*. Athenas main goals are to study the physics of accretion into compact objects and identify the earliest accreting super massive black holes. It will map the hot gas structures and determine their physical properties. Also Athena will trace the growth of these objects in obscured environments and show the influence they provoke in galaxies (X. Barcons et al. 2017).

Athena will consist of a single large aperture incidence X-ray telescope, it will be utilizing a technology developed in Europe, with 12m focal length and 5 arcsec on-axis angular resolution. The focal plane will contain two instruments. The first instrument is X-IFU, the X-ray Integral Field Unit (D. Barret et al. 2016). It will provide spatially resolved high resolution spectroscopy. The instrument is based on Transition Edge Sensors (a type of cryogenic energy sensor) which can deliver the energy resolution necessary for more efficiency compared to current X-ray observatories. The Transition Edge Sensors had proved that posses the required spectral resolution, up to 2.5 eV. In figure 2.2 is shown the effective area of the Athena X-IFU in comparison to other instruments.
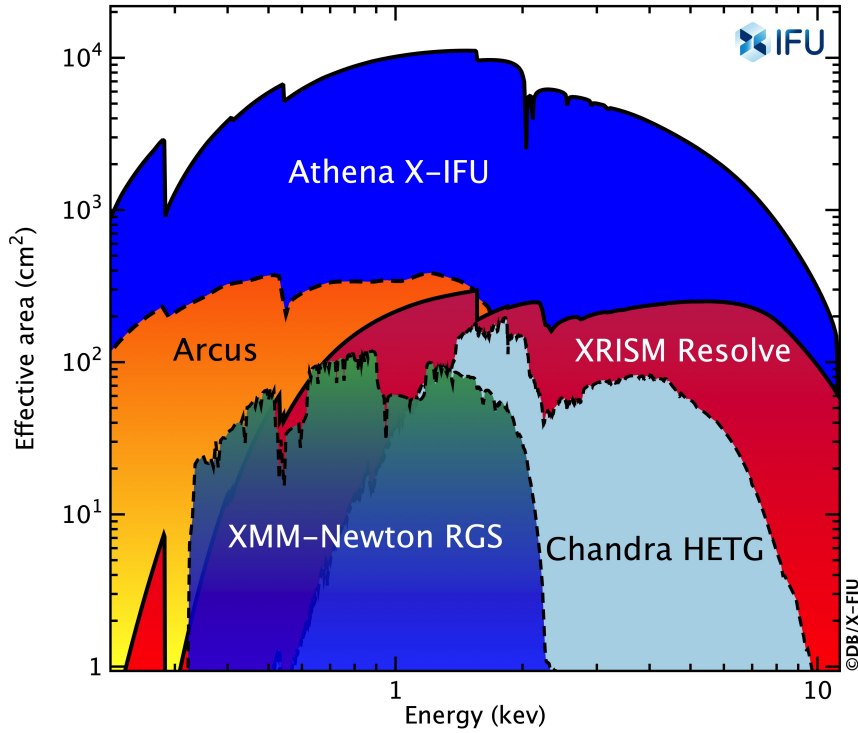
Figure 2.2: Effective area of the Athena X-IFU compared to the one of the XRISM Resolve instrument, both of them are imaging spectrometers. The effective area of the three dispersed spectrometers are also shown, which are XMM-Newton RGS and Chandra HETG, and Arcus. Credits: X-IFU Consortium. Copyrights: DB/X-IFU.

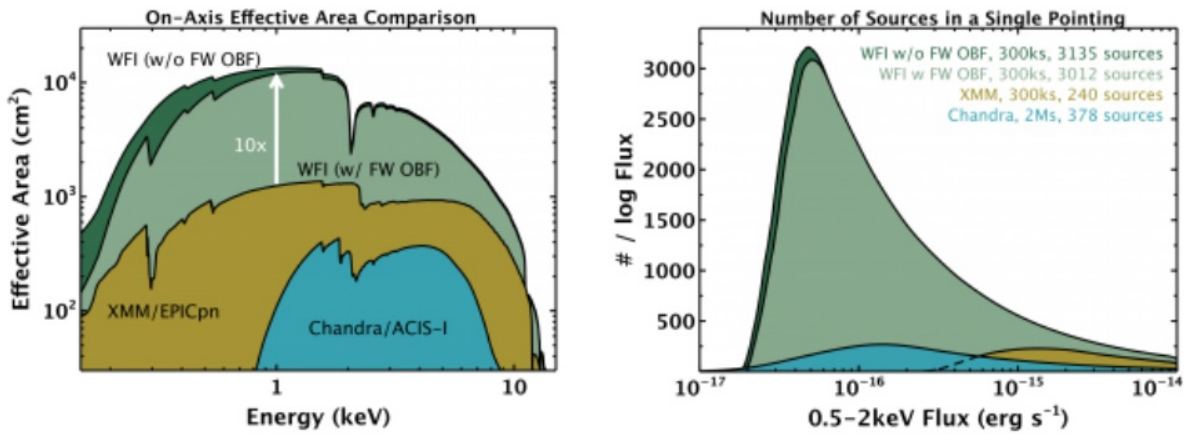*Figure 2.3: On the left panel is shown the Athena/WFI effective area in comparison to XMM-Newton/EPICpn and Chandra/ACIS-I. The Athena/WFI curve assumes an SPO optics with 15 mirror rows, 2.3 mm rib pitch, and Ir+B4C coating. On the right panel is shown the number of sources per log flux that can be detected in a single pointing with Athena/WFI, XMM-Newton/EPICpn, and Chandra/ACIS-I. Credits: A.Rau/WFI Team.*

The second instrument is WFI, the Wide Field Imager which is a Silicon based detector using DEPFET Active Pixel Sensor(APS) technology (A. Rau et al. 2013)). This instrument will provide a 0.2-15 keV imaging energy band over a wide field. The WFI instrument is designed to make optimal use of the collecting area and solid angle by the optical design of the mirror system Athena will posses. The large field of view is achieved through a focal plane that will be composed of many chips. One of them will enable fast computation to work on the measurements of bright targets. Figure 2.3 shows the Athena WFI effective area in comparison to other instruments and the number of sources per log flux.

The data simulated in this work are based on the WFI Athena Instrument. It was used the WFI instrument because of the imaging in the 0.2-15 keV energy band, since we work on the soft band at 0.5-10 energy band. Also because WFI will provide unprecedented simultaneous high time resolution and high count rate capabilities for the observation of bright sources with high efficiency. Thanks to its large field of view, the Athena WFI is expected to detect a large number of CT AGN throughout the mission life time.

2.2 Data Files

In all our work, we use simulated data for the Athena WFI instrument. The X-ray spectral data and calibration files we use in our analysis are described as follows:

- Source spectrum: It is a FITS (Flexible Image Transport System) file containing at least two columns, which are channel energy and number of counts. Quality and grouping columns are introduced before the fitting process and gives the analysis the necessary information to use this grouping during the spectral fit.

- Background spectrum: It is also a FITS file corresponding to the extracted background spectrum. In this case, only the columns corresponding to energy channel and counts are necessary, the information about the grouping is contained in the source spectra.

- Calibration files: These files contain the detector response, which is the probability that a photon of energy E entering the detector will be detected as a count in a determined channel. The detector response is the product of the RMF and ARF matrices. It is specific to the instrument for which the data is being simulated. In our case it corresponds to the lastest Athena WFI calibration file available at the time of this analysis.

It is important to define a Response Matrix. The response of the detector have a signal that depends on the energy of the photon entering the detector. It distributes the photon with a determined energy over channels according to the gain and energy resolution of the detector. The energy and resolution of a detector depends on the design so it is necessary to table the energy resolution while the photon energy works as a parameter. This forms a matrix that shows a data set. The number of counted photon depends on the source being detected, its distance, and the time of the observation. As a consequence this number can vary over many magnitudes. As we explain in the next section, in order to fit a model to the data, it is necessary to bin the source spectrum.

Spectrometers are used to measure the spectrum of a given source, but what the spectrometer obtains is not the actual spectrum, it obtains the photon counts (C) within specific instrument channels, (I). So the output spectrum expressed in counts per channel $C_p(I)$ and the monochromatic input photon flux $F_E(E)$ that is modified by the instrumental response $R(I, E)$ is related by:

$$C_p(I) = \int_0^{+\infty} F_E(E)R(I, E)dE \tag{2.1}$$

The idea is to determine the actual spectrum of a source, $f_E(E)$, by inverting this equation. This is not possible in any analytic way because such inversions tend to be not unique and unstable to small changes in $C_p(I)$. So the alternative is to used a model spectrum $F_E(E)$ described in terms of a few parameters. The model spectrum is selected and predicted counts per channel $C_p(I)$ is computed and compared to the observed counts $C(I)$.

2.3 Spectral Analysis

The X-ray spectral analysis was performed using the Xspec software. Xspec was developed at Cambridge University as a mission independent general analysis program for X-ray spectral data (Arnaud 1996). Xspec takes in consideration theoretical models, manipulates X-ray source data, background data, calibration data and also handles response matrices.

2.3.1 Statistics

There are several statistics implemented in Xspec. In this work we use χ^2 and Cash statistics. The most common fit statistic used for determining the best fit model is χ^2. Since X-ray spectra are in counts per channel, a minimum of 20 sources plus background counts should be selected in order to be able to use χ^2 because it requires Gaussian statistics. χ^2 is defined as:

$$\chi^2 = \sum \frac{(C(I) - C_p(I))^2}{\sigma^2(I)} \tag{2.2}$$

where $\sigma(I)$ is the error for channel estimated by Poisson statistics. A good fit has been achieved when a reduced χ^2 (χ^2 per degree of freedom) value close to 1 is obtained. To obtain the errors in the model parameters, their values are varied until a certain $\Delta\chi^2$ is reach.

It was mention that a minimum number of 20 source counts should be selected in order to use χ^2 statistics. For spectra with less counts it is necessary to use the Cash statistics, it does not provide a goodness of fit criterion, but allows to estimate the model's parameters (the models used on this work are presented in next subsection). Whenever we used Cash statistic, we grouped the source spectra to 1 count/bin. The maximum likelihood-based statistic for Poisson data, given in Cash (1979), is:

$$C = 2\sum_{i=1}^{N} (tm_i) - S_i ln(tm_i) + ln(S_i!) \tag{2.3}$$

were m_i are values of the predicted data rates based on the model (with current parameters) and instrumental response, S_i are the observed counts, t the exposure time.

2.3.2 Spectral models

The spectral fits that are employed in the analysis are a combination of simple models included in Xspec package. There are different kinds of models constructed from individual components. These components may be either additive, multiplicative, mixing, or convolution. Additive components are emission processes such as lines, bremsstrahlung, and more. Multiplicative components multiply the model by a factor that depends on the energy, usually represent absorption processes. At least one additive model must be included in the model and there is no restriction on the number of multiplicative models. What follows is a summary of the models that have been used in our spectral analysis.

- **Power Law**

The *powerlaw* is a simple photon power law additive model and the *zpowerlw* variant computes a redshifted spectrum. The parameters are the photon index (Γ), the normalization at 1 keV (where K is in units of photons $keV^{-1}cm^{-2}s^{-1}$), and is the case of zpowerlaw we used the redshift (z). This models are defined as:

$$\textbf{powerlaw} \longrightarrow M(E) = KE^{-\Gamma} \tag{2.4}$$

$$\textbf{zpowerlaw} \longrightarrow M(E) = K[E(1+z)]^{-\Gamma}/(1+z) \tag{2.5}$$

- **Gaussian emission line**

A simple gaussian line profile. If the width is ≤ 0 then it is treated as a delta function. The zgauss variant computes a redshifted gaussian. This models are defined as:

$$\textbf{gauss} \longrightarrow M(E) = N\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(E-E_L)^2}{2\sigma^2}} \tag{2.6}$$

$$\textbf{zgauss} \longrightarrow M(E) = N\frac{1}{(1+z)\sigma\sqrt{2\pi}}e^{-\frac{(E(1+z)-E_L)^2}{2\sigma^2}} \tag{2.7}$$

where N is the line flux, which is total photons $cm^{-2}s^{-1}$ on the line. σ is the line width and E_L is the central energy in keV.

- **Photoelectric absorption**

A photo-electric absorption using cross-sections set by the xsect command. It is a multiplicative model with N_{H} parameter and z in the red shift variant case, defined as:

$$\textbf{phabs} \longrightarrow M(E) = exp[-N_{\text{H}}\sigma(E)] \tag{2.8}$$

16

$$\textbf{zphabs} \longrightarrow M(E) = exp[-N_{\mathrm{H}}\sigma(E[1+z])] \tag{2.9}$$

where $\sigma(E)$ is the photo-electric cross-section (Not including Thomson scattering).

- **Brightman and Nandra 2011 torus model**

Brightman and Nandra torus model is based on Monte-carlo simulations. It is a model of a source emitting a powerlaw spectrum, at the center of the distribution of a cold neutral medium. In addition it possess a line of sight column density to the observer. It models the material around the X-ray emitting corona and computes the spectrum with photo electric absorption, Compton scattering, and fluorescent line emission (most importantly FeKα). The geometry of the obscuring material modelled is a spherical torus which is essentially a sphere, with a biconical cone cut into it.

It is a table model for Xspec. The torus1006.fits file includes the parameters column density N_{H}, Γ, Torus aperture angle Theta-tor (θ_{oA} in figure 2.4), the angle of sight (or inclination angle θ_i) to the center of the AGN Theta-inc, and the redshift z. A schematic representation is shown in figure 2.4 were we can see the parameters already described.
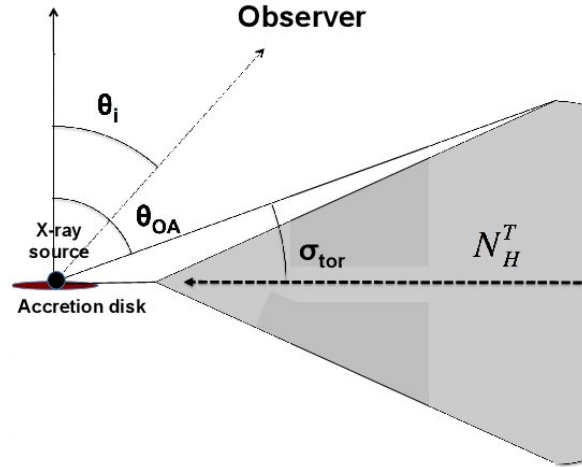

Figure 2.4. A schematic representation of torus model. θ_i is the inclination angle of the observer, θ_{oA} ($\pi/2$ θ_{tor}) is the half-opening angle of the torus. N_{H}^T is the equatorial column density of the torus. By Rozenn Boissay.

These are the files and models that we needed during the Athena simulation process. On next chapter we discuss the process to perform the simulations used in the analysis, also we discuss some simulations based only on a simple powerlaw plus emission line and without noise.

2.4 Castelló Method

Identifying redshift are essential and important to understand the extra galactic sources. Studying the redshift for distant AGN is important for understanding the accretion history of the universe, but it is extremely difficult. So models and methods have to be develop in order to determine AGN redshift.

N. Castelló-Mor et al. (2011) explored the capabilities of the space mission IXO (International X-ray Observatory, a precursor of Athena) for obtaining cosmological redshifts of distant AGN. First they find in which regions of the X-ray luminosity versus redshift plane the FeKα narrow emission line can deliver an accurate redshift as a function of exposure time using a CCD-based Wide Field Imager (IXO/WFI). Down to a 2-10 keV X-ray flux of 10^{-14} cm^{-2} s^{-1}, exposure of 100 ks, 300 ks. They used a method to generate the value of the energy line and then to be passed to Xspec. In order to statistically quantify the best fit model they have considered the statistic $\Delta\chi^2 = \chi^2_{baseline} - \chi^2_{baseline+gauss}$ and F-test probability value obtained from the fit of each simulation. Once the statistics $\Delta\chi^2$ and P_{F-test} have been computed, they define the discriminant $\Delta\chi^2 > 20$. Now to test the method they compared for each source the simulated redshift and the redshift obtained from the fit. They observed that both approaches ($\Delta\chi^2$ and P_{F-test} discrimination) provide good redshift estimation. Being the two methods similarly efficient, the $\Delta\chi^2 > 20$ discrimination is preferred, as it leaves only <1% outliers, while the P_{F-test} discrimination can leave up to 5% outliers.

2.5 Spectral Analysis with the FFT

We presented an analysis of simulated spectra by a powerlaw plus line without noise in chapter IV performing a Fast Fourier Transform in order to consider the frequencies at which the spectra can be cut to improve Castelló method in finding the line. So we explained here how the FFT works in the data analysis.

A fast Fourier transform (FFT) is an algorithm that computes the discrete Fourier transform (DFT) of a sequence. It can also compute the inverse Fast Fourier Transform (IDFT). The Fourier analysis converts a signal and pass it to a different domain. For example if the original domain is time (in our case energy) it converts the signal to frequency and to backwards. A FFT computes rapidly the transformation on the signal by factorizing the DFT matrix into a product of spares factors resulting in a less complicated computing process in contrast of the common Fourier transform.

The DFT (discrete fourier transform) has a forward and inverse form which are defined as follows:

- Forward Discrete Fourier Transform (DFT):

$$X_k = \sum_{n=0}^{N-1} X_n e^{-i2\pi kn/N} \tag{2.10}$$

- Inverse Discrete Fourier Transform (IDFT):

$$X_n = \sum_{k=0}^{N-1} X_k e^{i2\pi kn/N} \tag{2.11}$$

The transformation from $X_n \rightarrow X_k$ is a translation from configuration space (our case energy) to frequency space. It can be very useful in exploring the power spectrum of a signal, and also for transforming certain problems for more efficient computation.

In chapter IV we discuss the use of the FFT in our simulated data.

Chapter III

ATHENA Simulations

In this chapter we present the simulations needed in the analysis. We work in two types of Athena WFI simulations. The first ones are simulations without noise of a simple power-law plus a Gaussian emission line which represent the FeKα emission line. We use these simulations to find out the typical spectral scales over which unresolved 6.4 keV lines at different z are spread by the instrument. The second ones are realistic CT AGN simulations including noise and different redshifts, column densities, exposure times and luminosity.

3.1   Simulations: Power-law plus line no noise

The data we simulated without noise, no background and a simple powerlaw was used to find out the Fast Fourier Transform frequencies at which the FeKα emission line tested at different redshift contribute. Once we can identify the line correctly we can find the redshift of the source. The PSD treatment by the Fast Fourier Transform will be explained in Chapter IV.

The code used to simulated the data was written in PyXspec. PyXspec is an object oriented Python interface to the Xspec spectral fitting program. It provides an alternative scripting language for standard Xspec usage. With PyXspec loaded, a user can run Xspec with Python language scripts. The data was created by using the *fakeit none* command from Xspec, no noise and no background. The procedure to create the fake data is described as follows:

- We simulate a powerlaw in the z version and a gaussian in the z version (zpo+zga) with no noise and background. The powerlaw with parameters $PhoIndex = 1.9$, $norm = 1 \times 10^{-3}$, the gaussian with the Eline according to the desired redshift and sigma $\sigma = 10$ eV.

- Run the fitting process so the modelfile can be generated to used in next step. In our case we use energy range from 0.5-10.0 keV.

- At last, obtain the energies and the unfolded spectrum (which is the incident spectrum before entering the detector in flux units keV cm^{-2} s^{-1} keV^{-1}, eufspec) to make a plot of the data. We also obtained the data to model ratio (hearafter ratio) which we used to characterize the changes on the PSD due to the emission falling at different energies for different redshifst as we explained in chapter IV.
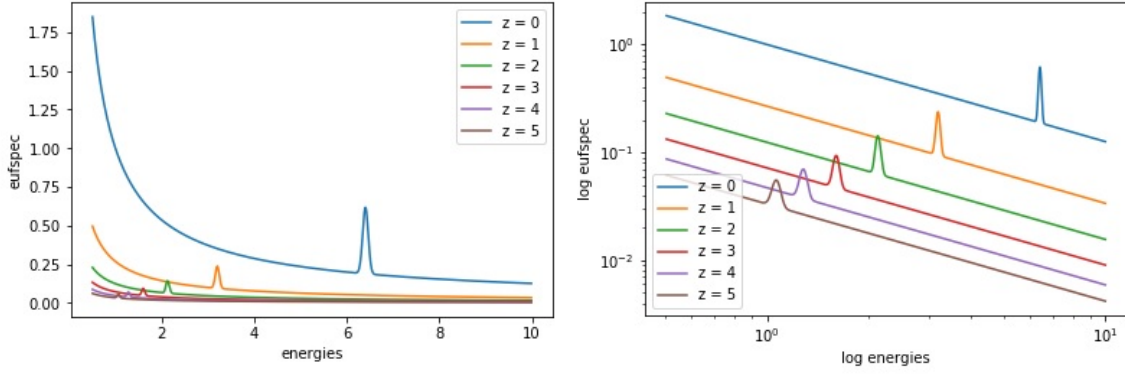
Figure 3.1. Simulations based on a simple powerlaw model and changing the redshift to values
z = 0,1,2,3,4 and 5. On the left panel the data are plotted on x-axis the energy and y-axis
the eufspec data. On the right panel is represented in logarithmic scale. We can notice that
the energy of the emission line correspond to values of 6.4, 3.2, 2.13, 1.6, 1.28 and 1.06 keV
respectively to each z value.

We show in figure 3.1 examples of some of the data simulated at this stage. It consists
on changing the redshift for each spectrum, using values of z = 0,1,2,3,4 and 5, which are in
the range of the values we simulated for Athena (will be explained in section 3.2). Figure 3.1
shows, on the left panel, the x-axis the energies and y-axis the unfolded spectrum. On the right
panel, the spectra are represented in logarithmic scale. Take in consideration that this plot is not
model independent and your unfolded model points will move if the model is changed. These
spectra are used in section 4.2 to estimate the energy scales over which the FeKα emission line
is expected.

3.2   Simulations: Compton Thick AGN

The X-ray data used in this work were simulated with the Athena Simulation code developed
by Francisco Carrera. This code simulates the data with the WFI Instrument over a 12.8 arcsec
diam extraction region, it does not simulate X-IFU spectra. The code simulates (Brightman and
Nandra 2011) torus models from a fixed set for different exposure times (Texp) and rescaled to
the desired values of several quantities.

Athena Simulation code can be modified to use a statistic in the fitting process: cstat (Cash)
or chi (χ^2). We used cstat and for using it, a minimum number of net counts per bin in the
grouped spectra must be set, so for Cash we used 1 total count per bin. The next step was
creating a file with parameters changing within a loop in the code. We work with 4 parameters,
the first parameter was exposure time in seconds with values of 3×10^4, 6×10^4, 7×10^4, 1×10^5,
3×10^5, 4×10^5, 7×10^5 and 1×10^6 s. The second parameter was luminosity with a range from
10^{43} to 10^{45} erg s^{-1}. Our third parameter was to used column density N_{H}, we used values
of 320×10^{22} and 3200×10^{22} cm^{-2} in order to simulate Compton thick AGN. And finally we
compute the redshift z with values of 1, 2, 3 and 4. By using the response matrix, the loop make
the simulation by combining all 4 parameters.

These CT AGN simulations are the set of data that we will analyse to find the FeKα emission line and improve the automated method. By studying the main characteristics we can determine the best subset of data to discuss our results. The procedure will be explained in chapter IV.

3.2.1 Spectral fit of Compton Thick simulations

All the simulations used in this part of the work are AGN Compton thick spectra. This means column densities greater than 10^{24} cm^{-2}. The same code also fits a model based on the torus model developed by (Brightman and Nandra 2011). The torus1006.fits file includes the parameters presented in table 3.1.

The simulated and fitted model used in the CT AGN data simulations is described in equation 3.1.

$$model = phabs * (atable\{torus1006.fits\} + zpo) \tag{3.1}$$

Which includes the torus model of Brightman and Nandra (torus1006.fits), to model the absorbed and reprocessed innermost emission, plus an additional power law (zpo), which accounts for the intrinsic emission scattered into our line of sight, and both photoelectrically absorbed by material in our Galaxy (phabs). This model has been found to be a good representation of the X-ray spectra of CT AGN in many studies.

Model	Parameters	values
Torus1006.fits	NH	320 / 3200
	PhoIndex	1.9
	Theta_tor	30
	Theta_inc	80
	z	1,2,3,4
phabs	NH	
	red shift	1,2,3,4
zpowerlaw	PhoIndex	1.9
	norm	

Table 3.1. Models used to fit the Compton thick AGN spectral data. The table also shows the parameters needed for each model and the values depending on the data.

We perform the fitting process in 340 spectral data. For the brighter CT AGN (L = 10^{45} erg s^{-1}) it was a very good model. In figure 3.2 we present four examples. We selected four spectral data with column density N_H = 3200$\times10^{22}$ cm^{-2}, the Luminosity L = 5$\times10^{45}$ erg s^{-1}, the exposure time T=10^6 s but with different redshift z = 1,2,3,4. We fit the data in the energy range 0.5-10 keV. The FeKα emission line can be seen in all four spectra. Notice that each line corresponds to energies defined as 6.4/(1+z) with the z values of z = 1,2,3,4. This energy values will be important when we discuss the detection of the emission line for different z.
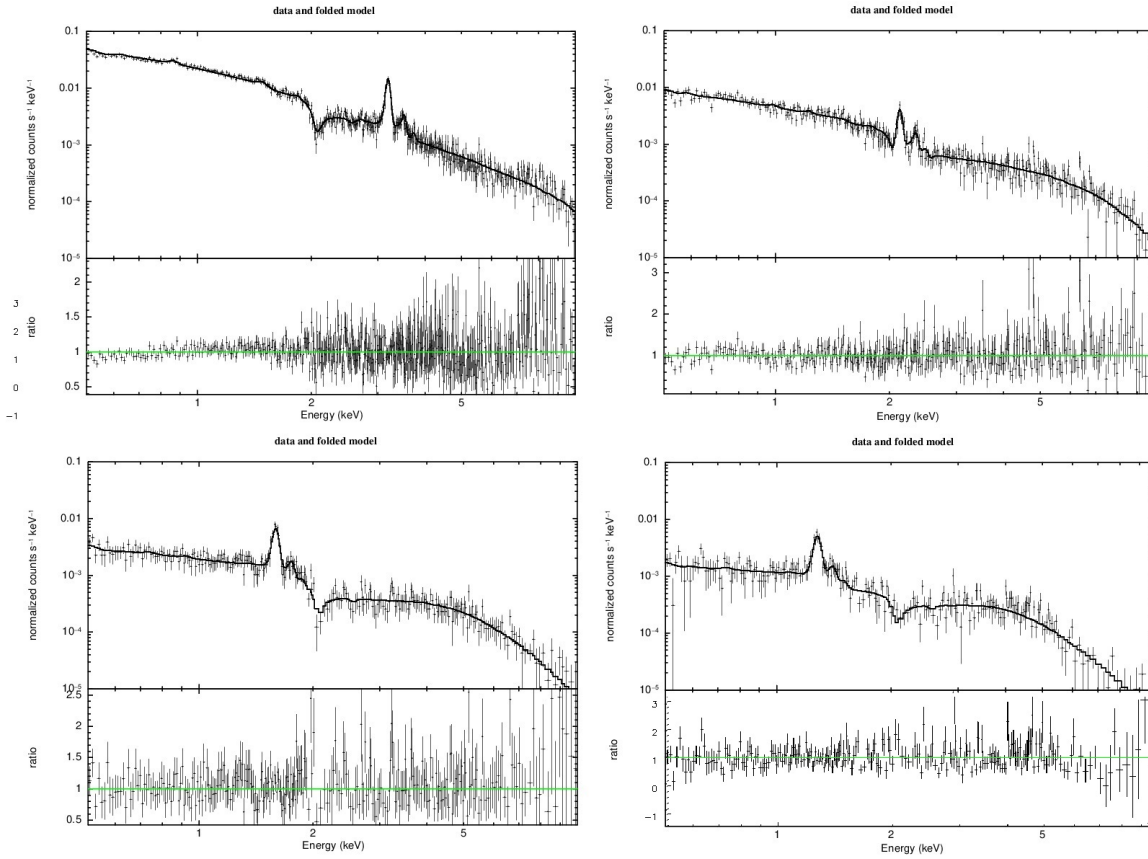
Figure 3.2. Representation of four fitted CT AGN spectral data with equation (3.1). We selected the spectra with N_H = 3200×10^{22} cm^{-2}, L = 5× 10^{45} erg s^{-1} and T = 10^6 s). On the x-axis the energy in keV and on the y-axis the data in normalized counts in s^{-1} keV^{-1} and also the ratio between the observer and the model. All the spectra have the same parameters except for the redshift which is, top left z = 1 , top right z = 2, bottom left z = 3 and bottom right z = 4. Notice the prominent FeKα emission line on the four spectra and how it shifts to lower energies as the redshift increases.

3.3 Discussion

We performed simulations without noise of a simple powerlaw plus a Gaussian emission line which represent the Fekα emission line. We use these simulations to find out the typical spectral scales over which unresolved 6.4 keV lines at different z are spread by the instrument.

During the simulation process we simulated AGN Compton thick spectral data. This means column densities over 10^{24} cm^{-2}. The code implemented to do the simulations required four main parameters. Which are column density, Luminosity, Exposure time and redshift. The code only works for WFI Athena Instrument. To demonstrate AGN Compton thick we fit a model based on a torus model, a simple powerlaw and a photoelectric absorption. The torus model is the one developed by (Brightman and Nandra 2011).

In chapter IV we will used the first simulations in a series of steps in order to study the signal in the fourier domain. It will be necessary to determine a cutoff frequency to improve a method to detect the emission line in the second simulations (the Athena WFI CT AGN simulations) which we will explain in detail. Because we want to improve a method that detects the energy were the emission line must be find out, we will evaluate how the method performs over CT AGN.

Chapter IV

The search for the Iron emission line

The detection of iron emission line is AGNs allows us to obtain a unique signature by the X-ray that come from the central engine and goes to distant material. The analysis of large samples of high z AGN is fundamental to understand the the evolution of SMBH in the center of all galaxies. CT AGN are difficult to detect, but they have a very strong emission line that can be used to identify it.

The main purpose of this work is to improve a method developed by (N. Castelló-Mor et al. 2011) that detects the energy were the emission line must be find out. We will evaluate how the method performs over CT AGN. To this end we will use simulated data for the upcoming Athena X-ray mission. To study in which cases the line is detected or the signal to noise ratio presents higher peaks resulting in the line not being detected, we perform a Fast Fourier Transform analysis to obtain a Power Spectrum Density (PSD) in the powerlaw with no noise simulated data presented in chapter III. The objective to use these simulations is to choose a cut off frequency to introduce a low pass filter and cut the frequencies that cause the detection to be wrong.

4.1   Results by using zseeker, the direct method

The code used to find the cosmological redshift, developed by N. Castelló-Mor et al. (2011), was written in python. The implementation of the code consist on series of steps, all executed with a script called zseeker.py. It defines a initial function called $Steer$ which a number of parameters needs to be taken in consideration. We introduce a *Path data files* which is the path were XSPEC will upload the data. Then we need to determine the *files to fit* were we place the list of spectral data. Finally we establish the *model for signal ratio* were in this case we chose a model $wabs * powerlaw$ in the energy range of 0.5-10 keV.

Now the method to be used. We have selected the function *from peak* which determines the detected line by using a signal to noise ratio threshold. Next we create two data files, first the *zseeker-Eline-detection.csv* to save the data that the code have been able to performe the zseeker process, and the second file is the *zseeker-excluded-files.csv* to save the data that was not suitable to be analyzed. Both files are filled with the information provided by the *files to fit* file. Finally we define the parameters to be used. The defined parameters are: 1) The signal

threshold = 0.8 (this can be in range 0 to 1). It is a top limit for which the method detect peaks, so peaks lower this parameter will not be considered. 2) extensionfile ='.pha' since we used the non grouped spectrum. 3) MethodEl = from-peak (there are two other methods). 4) simflag = defined false if the X-ray data is simulated rather than real. 5) model2fit = wabs*powerlaw which is the model to fit the signal. 6) Emin = 0.5 and Emax = 10.0 are the range selected on this test. and 7) parset = which freeze the parameters given to the fitting model.

Once the setup has been made we run the zseeker script code. It will load the spectral data, ignore some channels and energy range that were established. Then load the model and make the fitting process with the statistic χ^2. The detectEline.py script is a routine based on a looking at peaks in the ratio between the observed data and the model. It finds peaks in the y axis and taking their difference of first order and uses the signal threshold and minimum distance. It is possible to reduce the number of peaks detected in y axis and it must be a numpy array. The parameters to be used are: 1) The y axis which will mark the value of the detected peaks. 2) Thres being a float number in range [0:1]. 3) min-dist which is the minimum distance between each peak. 4) Max is the parameter of the maximum amplitude. 5) Xcut is the range in the x axis were the line must be looked at.

To summarize the code will fit the simulated data with a single power law to obtain a ratio between the data and this model. Detect the line in a given energy range in x and a y axis peak that is greater than the threshold and minimum distance in y. Once the lines are calculated, from measuring the energy E_0 we can obtain the redshift defined as:

$$z = \frac{E_{rf}}{E_0} - 1 \tag{4.1}$$

were E_{rf} is the rest frame emission line energy (established for the Eline at 6.4 keV).

To test the reliability of the method, we compared for each source the simulated redshift z_{sim} and the redshift obtained from the fit z_{fit}. We used a sample of 292 AGN simulations to find a possible emission line. The sample of simulations was taken with all the parameters we simulated. N_H = 320 and 3200$\times10^{22}$ cm^{-2}, L = 10^{43}, 10^{44} and 10^{45} erg s^{-1}, z = 1,2,3,4, and T = 10^4, 10^5, 10^6 s. In 178 AGN was able to detect a signal peak. The rest was not able due to low exposure times and low luminosity. Once we got the z_{fit} we were able to plot it against the simulated redshift. As it it shown in figure 4.1, many of the simulations present a good redshift estimation. From the 178 AGN sample, 108 AGN were found that the simulated redshift correspond to the one of the redshift obtained from the fit. In other words, 61% of the CT AGN simulations were successfully fitted. We perform a displacement (from -0.5 to 0.5 at random) on both simulated and fitted redshift in order to observe and have a complete view of all AGNs.
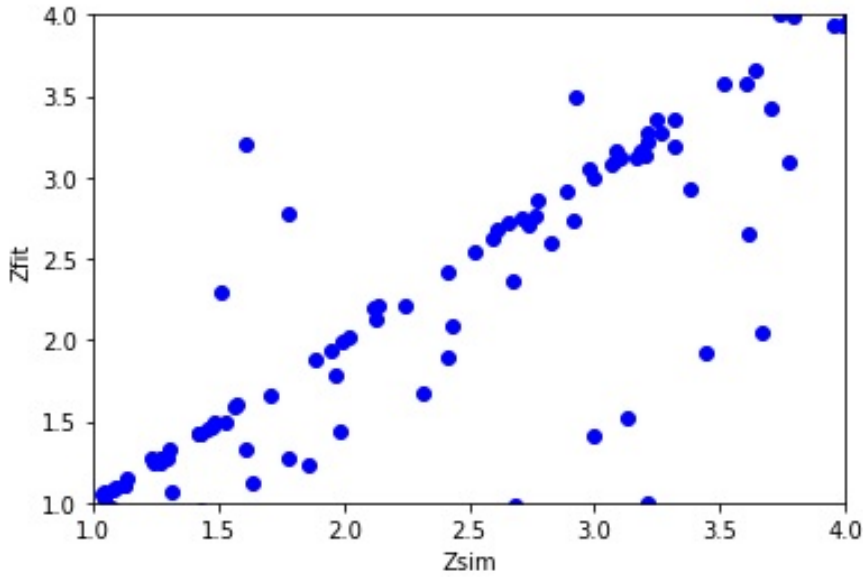
Figure 4.1. Fitted versus simulated redshift. A sample of 178 out of 292 simulations were fitted by the zseeker.py code.

Even though many of the simulations present a good estimation of the redshift, there are some simulations that were not able to be fitted correctly. Here we will discuss 2 cases were the method can find the line or not. In figure 4.2 are shown four examples of how the emission line has been found and the corresponding measure line energy E_0. This examples are simulations of the same $N_H = 3200 \times 10^{22}$ cm^{-2}, but with different luminosities at 10^{43}-10^{45} erg s^{-1}, exposure times at 10^4-10^6 s, and redshift z = 1,2,3,4.

For the simulations in figure 4.2 we used the grouped spectra. We used the selected parameters of $N_H = 3200 \times 10^{22}$ cm^{-2}, L = 10^{45} erg s^{-1} and T = 10^6 s, and changing the redshift. All the lines must be detected according to equation 4.1. So for each z = 1,2,3,4 the energy line values must be E = 3.2, 2.133, 1.66, 1.28, keV respectively. The values of the simulated redshift corresponds to those of the fitted redshift values. The result was that due to the good signal to noise ratio, the emission line peak was detected correctly by Castelló's method. We can see that there are few peaks that can be compared to the emission line. So this signals prove to be good enough to demonstrate what a correct detection looks like.
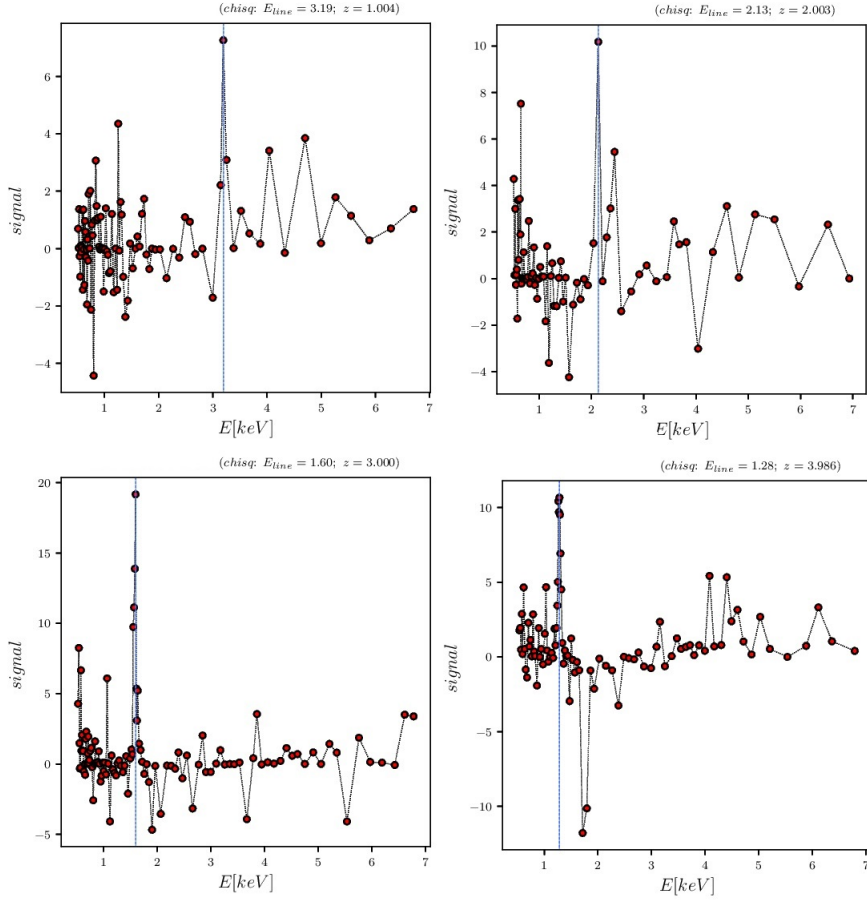
27

Figure 4.2. Signal (ratio) versus Energy (keV) representation of four simulations (All with same $N_H = 3200 \times 10^{22}$ cm^{-2}, $L = 10^{45}$ erg s^{-1} and $T = 10^6$ s) for which zseeker was able to fit the emission line successfully. The blue line shows the peak that the zseeker detected and with that energy calculated the redshift. At $z_{sim} = 1$ (top left) must be $E_0 = 3.2$ keV. $z_{sim} = 2$ (top right) must be $E_0 = 2.13$ keV, $z_{sim} = 3$ (bottom left) must be $E_0 = 1.6$ keV and for $z_{sim} = 4$ (bottom right) must be $E_0 = 1.28$ keV. This are examples of a very good detection of the FeKα emission line due to a very good signal to noise ratio.

A second result can end on the detection of a signal that do not correspond to the FeKα emission line. In figure 4.3 is represented the case in which at low energies (the two left graphics), the signal presents higher peaks than the one of the emission line. This results in an usuccessful detection and, as a consequence, an error in the fitting redshift. Again the selection of the sample consist of four CT AGN spectrum with same $N_H = 3200 \times 10^{22}$ cm^{-2}, but with different luminosities at 10^{43}-10^{45} erg s^{-1}, exposure times at 10^4-$10^6 s$, and redshift z = 1,2,3,4. These simulations are the ones that are shown in figure 4.1 were the z_{fit} is plotted at very high redshift due to the noise peaks at low energies. This case can end as well in higher energies (the two right graphics), were the signal to noise ratio is really bad and presents noisy peaks that are higher than the emission line. In this case, the error in the z_{fit} is due to a noise peaks at higher energies. This is the same as the ones at lower energies, but in figure 4.1 are represented by the ones that were plotted at very low z_{fit}.

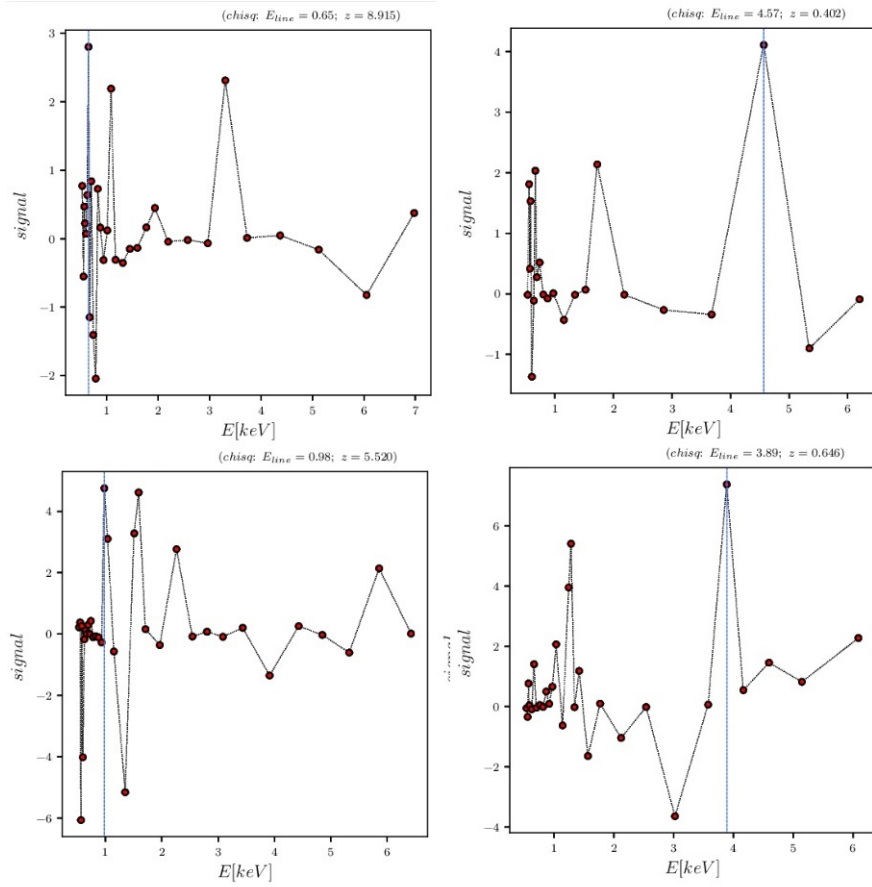*Figure 4.3. Signal (ratio) versus Energy (keV) representation of four simulations for which zseeker was not able to fit the emission line successfully. As it can be seen the z_{fit} due to the bad signal to noise ratio are: top left z_{fit} = 8.915, top right z_{fit} = 10.781, bottom left z_{fit} = 5.520, bottom right z_{fit} = 6.160. The noisy peaks are higher than the emission line producing a none successful detection.*

We discuss the case were the signal to noise ratio peak is lower compared to the one of the FeKα emission line. This case was the more successful one, but the real problem on finding the emission line was due to a peaks greater in the low and higher energy section. Now we need to establish some parameters. We have selected a subsample of spectra with common characteristics to compare in more detail the behaviour of zseeker on the orignal spectra, with its behavior on spectra to which some filter has been applied to improve the signal to noise ratio. To be more accurate, we work with column density $N_{\mathrm{H}} = 320{\times}10^{22}$ cm^{-2} because this AGN presents much less absorption than the ones we used before. We used the same redshift z = 2 so we know that to all spectra the emission line must be detected at the same energies. Because we want data quality we used the largest exposure time simulated T = 1e6 s. The most important parameter is the one we changed, luminosity. We used values of L = 10^{43}, 10^{44}, 10^{45} erg s^{-1}. At higher luminosities the signal to noise ratio must be very good and the detection must be successful, but at lower luminosities the signal to noise ratio will probably present higher peaks and the method detects the wrong signal.
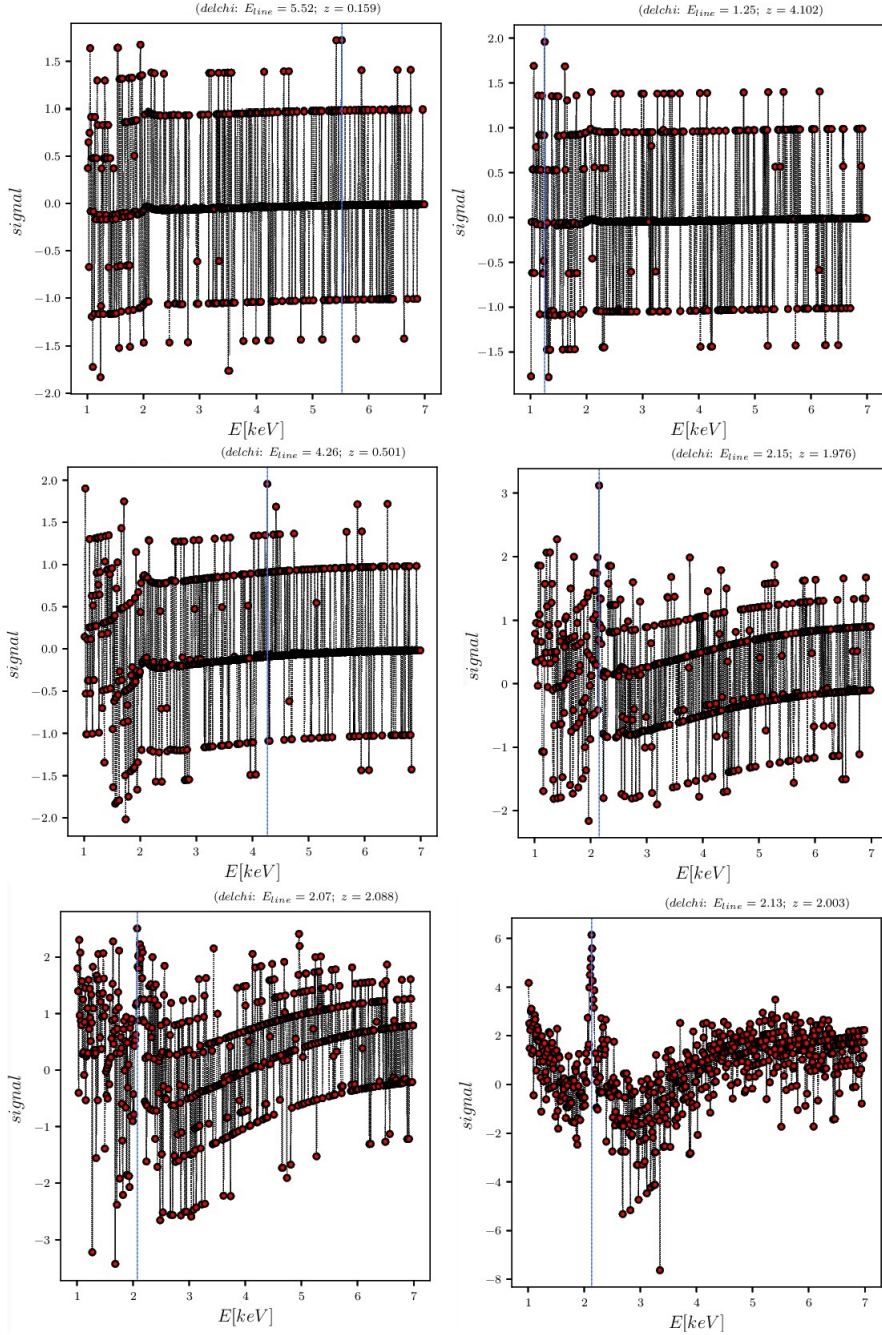
29

Figure 4.4. Signal (ratio) versus Energy (keV) representation of six simulations. The six simulation had been selected to have the parameters $N_H = 320 \times 10^{22}$ cm^{-2}, z = 2, T = 10^6 s, but different luminosities, top two graphics with L = 10^{43} and 5×10^{43} erg s^{-1} respectively, the middle ones with L = 10^{44} and 5×10^{44} erg s^{-1} and the lower ones with L = 1×10^{45} and 5×10^{45} erg s^{-1}. Notice that at lower luminosities ($< 10^{44}$ erg s^{-1}) the line was not detected correctly due to many noise peaks, at L = 10^{44} erg s^{-1} only the right panel presented a successful detection at z = 1.967, finally the higher luminosity samples L $\geq 10^{45}$ erg s^{-1} the detection was the correct one and present less noise peaks.

In figure 4.4 we plotted six CT AGN spectrum with the parameters mentioned above. Out of the six spectrum, 3 were detected correctly at $z_{fit} \sim 2$. As we expected, these spectra are the ones with higher Luminosity (10^{44}, 10^{45} erg s^{-1}), so the noise peaks are lower than the one of the emission line. However, the other 3 spectra, which possess lower Luminosity ($\leq 10^{43}$ erg s^{-1}) were not detected correctly.

Top left and right graphics are the lowest ones in Luminosity values. Both redshifts were wrong detected. The middle left graphics' redshift was also not detected correctly but the middle right was. The last two are the highest in luminosity and both redshift were detected correctly. Again it works better for higher luminosities that for all other parameters fixed, because it correspond to higher Signal to noise ratios.

Taken this information, we need to perform a method to clean the signal so we can obtain better results and improve the number of successful detection. We need to prove that we can improve the results. In next section, we discuss a series of steps to provide a reliable automated method with our upgrade.

4.2   Filtering using FFT, low pass filter

In order to correct the signal to noise ratio so we can have a spectrum more "clean" and reduce the number of spectral data that failed to obtain the peak of the emission line, we need to explore the possibility to use a Fourier transform. Moreover, we can use the Fast Fourier Transform (FFT).

From this point we work on the construction on the low pass filter. By using the powerlaw with no noise simulated data (simulations explained in section 3.1), we compute the ratio from each spectrum (each one with different z). Once we get the ratio we use it to obtain the FFT with the Python-numpy package. Now taking the FFT for each spectrum we obtained the power spectrum density (PSD) as shown in figure 4.5. We can see on the left panel the PSD corresponding to each z and on the right panel in logarithmic scale. This PSD are need to define at which frequencies we can cut off the signal. But first we need to fit a Gaussian model to select the cut off frequency. So in figure 4.6 we show the Gaussian fitting model to each PSD.
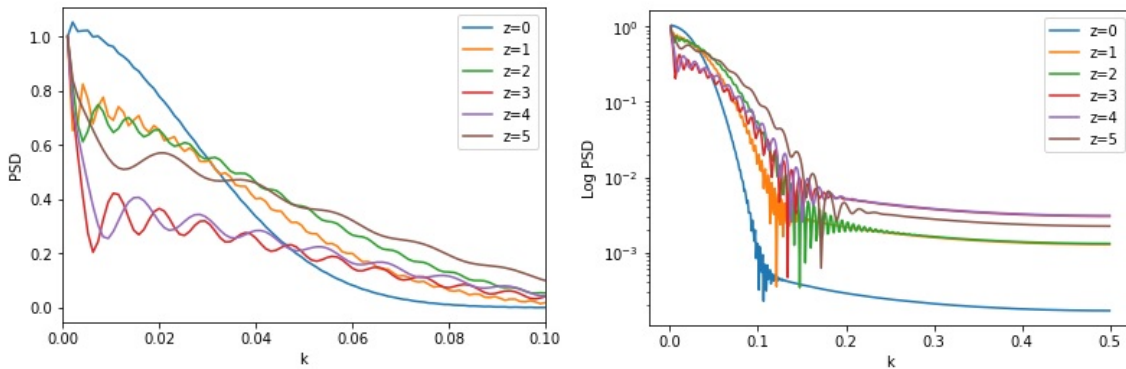

Figure 4.5. Representation of the power spectrum densities (PSD), y-axis, versus the frequency k, x-axis, for each redshift z = 0,1,2,3,4 and 5 on the left panel. On the right panel the PSD

31

are shown in logarithmic scale. We use the from section 3.1 (powerlaw plus line with no noise) to estimate the frequencies scales over which we would expect to see the signal from the FeKα emission line.
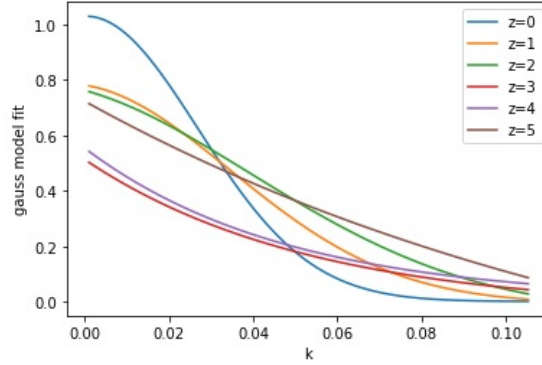

Figure 4.6. Representation of the Gaussian fitted model, y-axis, to each PSD in k frequency k, x-axis. This Gauss model was fitted to obtain the best fit values such as sigma and select at which frequencies we can cut off without compromising the emission line frequency.

Redshift (z)	Sigma(σ)	filter selection (3*σ)
z = 0	0.0265	0.06
z = 1	0.0392	0.09
z = 2	0.0544	0.15
z = 3	0.4336	1.29
z = 4	0.4190	1.23
z = 5	0.0769	0.23

Table 4.1. Best fit sigma values obtained by fitting a Gaussian model to the power spectrum density for each powerlaw with no noise simulated data at different redshift z = 0,1,2,3,4 and 5.

For each Gaussian model we retrieve the sigma values as it is shown in table 4.1. In order to cut the frequencies that we don't need, but not the important ones such as the emission line, we used the highest sigma and multiply it by 3. This will be the cut off frequency on our filter design. In next section we discuss the results on the spectrum selected in figure 4.4.

4.3 Results using the new automated method

The new filtered spectra were transformed in a Xspec format by using the $ftflx2xsp$ command. It creates spectra and response files from an input text file of fluxes and errors. This tool reads a text file (infile) containing one or more spectra and errors and writes out a standard PHA file (phafile) and RSP file (rspfile) which can be used in Xspec. The response file created is a unit

diagonal matrix of the appropriate size. The input text file should have one line for each bin of the spectrum. Each line should have at least four numbers. The first two should be the start and end energy (or wavelength) of the bin and the second two the flux and one sigma error. The units for the energies (or wavelengths) and fluxes in the input text file can be specified using the xunit and yunit parameters.

In figure 4.7 are presented the filtered spectra from the selected sample in figure 4.4. It seems that the both signal and ratio have been amplified but the signal has been amplified more than the noise. It looks like it is a problem with the normalization of the filter design, it is necessary to add a unit step function to the Fourier transform so the signal and noise are not amplified. Even that the signal amplifies more than the noise, no in all the cases the emission line was perfectly detected. Lets analyze each CT AGN spectrum. Recall that all spectra have same column density $N_{\mathrm{H}} = 320 \times 10^{22}$ cm^{-2}, same redshift z = 2, same exposure time T = 10^6 s. The top graphics which correspond to lower luminosity shows no improvement in the line detection. There has been an improvement for spectra at L = 10^{44} erg s^{-1} (middle left). The last two were also detected correctly after the filtering process.
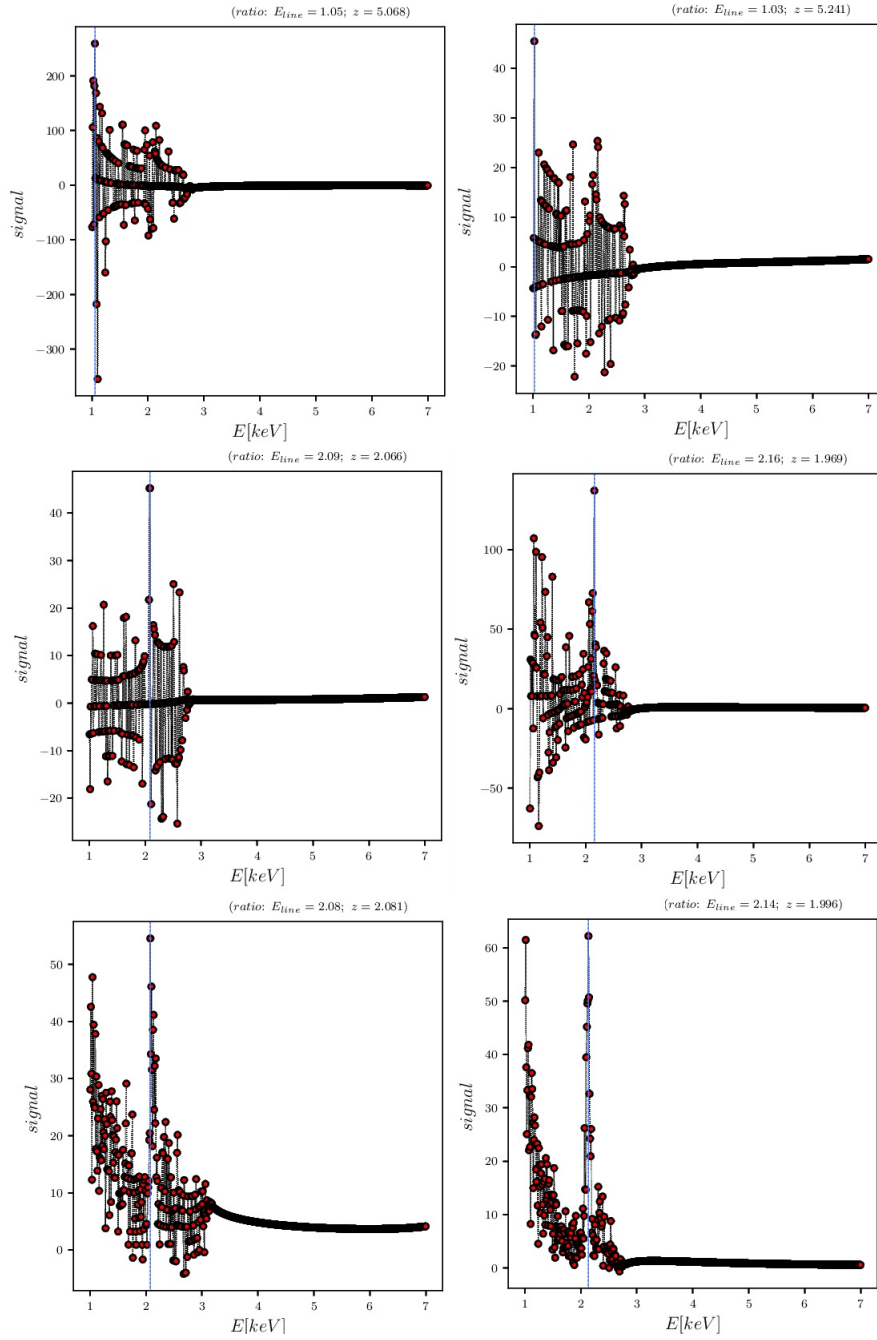
Figure 4.4. Signal (ratio) versus Energy (keV) representation of six filtered spectrum. The six spectrum had been selected to have the parameters $N_H = 320 \times 10^{22}$ cm^{-2}, z = 2, T = 10^6 s, but different luminosities, top two graphics with L = 1×10^{43} and 5×10^{43} erg s^{-1} respectively, the middle ones with L = 1×10^{44} and 5×10^{44} erg s^{-1} and the lower ones with L = 1×10^{45} and 5×10^{45} erg s^{-1}. The filtered spectra present an improvement reducing noisy peaks. At lower luminosities (10^{43} erg s^{-1}) the line was not detected correctly after the filter was applied. At L = 10^{44} erg s^{-1} in the two cases the line was detected correctly since the middle left graphic presents an improvement. Finally the higher luminosity samples L = 10^{45} erg s^{-1} the detection was the correct one.

34

4.4   Discussion and Conclusions

We explained that AGNs are powered by the accretion of material onto a super massive black hole (SMBH) and are the brightest persistent objects in the universe and emit huge amount of energy over the entire electromagnetic spectrum. But the radiative power of most AGN cannot be seen directly, because the accretion is hidden behind dust and gas that absorb this emission and obscure many of the main characteristic properties, this process is known as obscuration. We can study the X-ray emission of an AGN by approximate it to first order power law and a torus model.

Some AGN are classified as "Compton Thick sources" if the column density reaches values above 10^{24} cm^{-2}, and the direct continuum is complete suppressed below 10 keV and it is only observed reflected emission. The study of Compton thick sources is relevant because there is observational evidence that suggests that a large fraction of AGN in the local universe are obscured by Compton thick gas.

The FeKα line, as the most prominent feature in the rest frame energy range of 6–7 keV, provides an important view of the nature of the disk and the central SMBH. The narrow FeKα line is one of the main signatures of the reprocessing of X-ray radiation from the material surrounding super massive black holes, and it has been found to be omnipresent in the X-ray spectra of active galactic nuclei. The detection of FeKα emission line in AGNs allows us to obtain a unique and distinctive spectroscopic feature generated by the X-ray irradiation from the central AGN engine onto cold distant material. The analysis of large samples of high z AGN is fundamental to understand the growth of SMBH in the centers of galaxies. CT AGN are difficult to detect, but they have a very strong emission line that can be used to identify it.

We presented the simulations needed in the analysis working in two types of Athena WFI simulations. The first ones were simulations without noise of a simple power-law plus a Gaussian emission line which represent the FeKα emission line. We use these simulations to find out the typical spectral scales over which unresolved 6.4 keV lines at different z are spread by the instrument. The second ones are realistic CT AGN simulations including noise and different redshifts, column densities, exposure times and luminosity.

We worked on an improvement of a method developed by (N. Castelló-Mor et al. 2011) that detects the energy were the emission line must be find out. We evaluated how the method performs over CT AGN. To study in which cases the line is detect or the signal to noise ratio presents higher peaks resulting in the line not being detected, we performed a Fast Fourier Transform analysis to obtain a Power Spectrum Density (PSD) in the powerlaw with no noise simulated data. The objective of this PSD was to fit a Gaussian model so we can choose a cut off frequency to introduce a low pass filter and cut the frequencies that cause the detection to be wrong.

In the selected sample of CT AGN we observed that the AGN with lower luminosity (10^{43} erg s^{-1}) presented the most noisy peaks compared to the rest of the data. We try to improve the SNR by selecting only lower column densities data because this AGN have less absorption on their spectra. We selected the low NH spectra to see well the effects of the filter, then we kept all other parameters fixed and changed only luminosity which should be a way to increase the

sigal to noise ratio in a controlled way.

The simulations with higher luminosities ($\geq 10^{44}$ erg s^{-1}) were the more successful ones in the line detection process. The noisy peaks were amplified by the low pass filter we implemented to upgrade the method to detect the peaks, buy the signal to noise ratio improves. The FeKα emission line peak was high enough to be detected proving that the low pass filter did not cut off the line frequency.

The low pass filter implemented in this work are good enough to clean one signal and perform a correct line detection. In order to prove that this new improvement to the peak detection method provide a reliable and confident process, it is necessary to study a more variable and bigger sample of CT AGN simulations.

As future works, we need to construct a more complete and effective low pass filter to determine if at lower luminosities we can provide better results and increase the percentage of AGN that their emission line are detected correctly. We can try to implement a more sophisticated methods that can improve the signal we study and show a more clear emission line to be detected so we can calculate the correct redshift.

We also need a better way of detecting and selecting the peaks of emission. This method only choose the highest peak. So one improvement would be to only look peaks at energy above 1 keV in order to remove the low energy peaks. We can constrain at the energy range we know the emission line shows up at different redshifts.

Bibliography

Comastri A. (2004), 'Compton thick agn: The dark side of the x-ray background, supermassive black holes in the distant universe', *ASI, ASSL* **308**.

A. Rau et al. (2013), 'The hot and energetic universe: The wide field imager (wfi) for athena+', *Supporting paper for the science theme "The Hot and Energetic Universe" to be implemented by the Athena+ X-ray observatory,* p. 11.

Antonucci (1993), 'Unified models for active galactic nuclei and quasars', *ARAA* **31**, 473–521.

Arnaud (1996), 'Xspec', *Astronomical Society of the Pacific Conference Series* **101**, 17.

Brandt and Alexander (2015), 'Cosmic x-ray surveys of distant active galaxies: The demographics, physics, and ecology of growing supermassive black holes', *AA* **2**, 120.

Brightman and Nandra (2011), 'An xmm–newton spectral survey of 12 m selected galaxies – i. x-ray data', *Monthly Notices of the Royal Astronomical Society* **413**, 1206–1235.

C. Megan Urry and Paolo Padovani (1995), 'Unified schemes for radio-loud active galactic nuclei', *Astronomical Society of the Pacific* **107**, 803–845.

C. Ricci et al. (2014), 'Iron $k\alpha$ emission in type-i and type-ii active galactic nuclei', *Monthly Notices of the Royal Astronomical Society* **441**, 4.

Corral, A. (2011), "Caracterización espectral en rayos X del acrecimiento y oscurecimiento en Galaxias Activas", PhD thesis, Universidad de Cantabria.

D. Barret et al. (2016), 'The athena x-ray integral field unit (x-ifu)', *Proc. SPIE 10699, Space Telescopes and Instrumentation 2018: Ultraviolet to Gamma Ray* **106**, 41.

J. Buchner and F. Bauer (2017), 'Galaxy gas as obscured - ii. separating the galaxy-scale and nuclear obscurers of active galactic nuclei', *Monthly Notices of the Royal Astronomical Society* **465**, 4348–4362.

M. A. Prieto et al. (2009), 'The spectral energy distribution of the central parsecs of the nearest agn', *Monthly Notices of the Royal Astronomical Society* **402**, 50 pages.

N. A. Levenson (2013), 'Multiwavelegth agn surveys and studies', *International Astronomical Union, IAU Symposium* **304**, 112–118.

N. Castelló-Mor et al. (2011), 'X-ray redshifts with the international x-ray observatory (ixo)', *Advances in space research* **48**, 1304–1310.

Netzer, H. (2013), *"The Physics and Evolution of Active Galactic Nuclei"*, Cambridge University Press.

R. Hickox and D. Alexander (2018), 'Obscured active galactic nuclei', *Annual Review of Astronomy and Astrophysics* **56**, 625–671.

Rozenn Boissay et al. (2016), 'A hard x-ray view of the soft excess in agn', *AA* **2**, 15.

S. F. Honig and T. Becker (2007), 'Active galactic nuclei dust tori at low and high luminosities', *Monthly Notices of the Royal Astronomical Society* **380**, 1172–1176.

Seward, F. D. & Charles, P. A. (2010), *"Exploring the X-ray Universe"*, Cambridge University Press.

X. Barcons et al. (2017), 'Athena: Esa's x-ray observatory for the late 2020s', *Astronomische Nachrichten* **338**, 18.

Yuan Liu and Xiaobo Li (2014), 'A direct comparison of x-ray spectral models for tori in active galactic nuclei', *Monthly Notices of the Royal Astronomical Society: Letters* **448**, 12.