Masaryk University
Faculty of Science
University of Turin
Faculty of Science

Ph.D. Thesis

Matej Kosiba

Brno 2024

Machine learning analysis and cataloging of extragalactic sources

Ph.D. Thesis

Matej Kosiba

Supervisors:
Prof. Norbert Werner
&
Prof. Francesco Massaro

Masaryk University, Brno, Czech Republic
&
University of Turin, Turin, Italy

Brno 2024

Bibliografický záznam

Autor:	Mgr. Matej Kosiba Přírodovědecká fakulta, Masarykova Univerzita Ústav teoretické fyziky a astrofyziky & Dipartimento di Fisica, Università di Torino

Název práce:	Analýza strojovým učením a katalogizování extragalaktických objektů
Studijný program:	Doktorský studijní program: Fyzika
Obor:	Astrofyzika
Vedoucí práce:	Prof. Norbert Werner, Ph.D. & Assoc. Prof. Francesco Massaro, Ph.D.
Akademický rok:	2023/2024
Počet stran:	xiii + 185
Klíčová slova:	strojové učení, umělé neurónové sítě, katalógy, kupy galaxií, kosmologie, aktivní jádra galaxií

Bibliographic Entry

Autor:	Mgr. Matej Kosiba Faculty of Science, Masaryk University Department of Theoretical Physics and Astrophysics & Department of Physics, University of Turin

Title of Thesis:	Machine learning analysis and cataloging of extragalactic sources
Degree of study:	Doctoral study program: Physics
Field of Study:	Astrophysics
Supervisors	Prof. Norbert Werner, Ph.D. & Assoc. Prof. Francesco Massaro, Ph.D.
Academic Year:	2023/2024
Number of Pages:	xiii + 185
Keywords:	machine learning, artificial neural networks, catalogs, galaxy clusters, cosmology, active galactic nuclei

Abstrakt

Tato práce se zaměřuje na analýzu strojovým učením a katalogizaci extragalaktických zdrojů. V první části své práce jsem použil umělé neuronové sítě, třídu algoritmů strojového učení pro automatickou detekci kandidátů kup galaxií. Použil jsem obrazová data v rentgenových vlnových délkách z *XMM-Newton* s jejich optickými protějšky z databáze SDSS. Tato práce byla úspěšná; Vytvořil jsem konvoluční neuronovou síť, která dosahovala 90 % průměrné přesnosti v klasifikaci objektů na kupy galaxií a objekty které nejsou kupa galaxií. Druhá část mé práce se zaměřila na vývoj nové metody kosmologické analýzy založené na rozložení galaktických kup, speciálně navržené tak, aby se vyhnula nutnosti používat škálovací vztahy. K tomu jsem použil analytické simulace rentgenově pozorovatelných diagramů, znázorňujících rozložení kup galaxií v čistě pozorovatelných veličinách, count rate, hardness ratio a spektroskopický červený posuv. K odhadu kosmologických parametrů jsem použil Sequential Neural Posterior Estimation, metodu založenou na umnělých neuronových sítích. Tato práce byla úspěšná a ve své práci uvádím, jak lze metodu použít k odhadu kosmologických parametrů Ω_m a σ_8 na základě čistě pozorovatelných patametrů galaktických kup. Ve třetí části mé disertační práce se zaměřuji na katalogizaci tvrdých rentgenových zdrojů, s cílem lépe porozumět zdrojům přispívajícím ke kosmickému rentgenovému pozadí. Provedl jsem multifrekvenční analýzu tvrdých rentgenových zdrojů katalogu 3PBC, abych našel jejich nízkoenergetické protějšky a přiřadil jim klasifikaci. Klasifikace byla založena na jejich multifrekvenčních informacích, jako svítivosti a spektrech. Finálními produkty této práce jsou dva katalogy: revidovaná verze katalogu 3PBC založená na našich multifrekvenčních kritériích a druhé vydání Turin-SyCAT katalogu Seyfertových galaxií. Ve čtvrté části mé práce jsem provedl následnou analýzu populace 218 dosud neidentifikovaných zdrojů, které sme nalezly v předešlé analýze. Jsou to zdroje, které dosud nemněli identifikovaný nízkoenergetický protějšek. Cíl byl stejný: lépe porozumět zdrojové populaci přispívající ke kosmickému rentgenovému pozadí. Abych to udělal, hledal jsem jejich nízkoenergetické protějšky v archivních datech *Swift*-XRT. Výsledným produktem této práce jsou dva katalogy. První je katalog kandidátů na nízkoenergetické protějšky pro 73 zdrojů z 218 doteď neidentifikovaných 3PBC objektů, z nichž 60 má jednoho kandidáta na protějšek a 13 jich má více. Druhým je krátký katalog 10 zdrojů, ke kterým jsem našel i vícefrekvenční informace které byli nutné k jejich klasifikaci.

Abstract

This work focuses on machine learning analysis and cataloging of extragalactic sources. In the first part of my work, I used artificial neural networks, a class of machine learning algorithms for automatically detecting galaxy cluster candidates. I used image data in X-ray wavelengths from *XMM-Newton* with their optical counterparts from the SDSS database. This work was successful; I created a convolutional neural network that achieved 90 % average accuracy, classifying sources into galaxy clusters or non-clusters. The second part of my work focused on developing a new cosmological analysis method based on the distribution of galaxy clusters, specifically designed to bypass the need to use scaling relations. To do this, I used analytical simulations of X-ray observable diagrams, depicting galaxy cluster's distribution in purely observable quantities, X-ray count-rate, hardness ratio, and spectroscopical redshift. I utilized Sequential Neural Posterior Estimation, a class of simulation-based inference techniques based on artificial neural networks for density estimation. This work was successful, and I present in my thesis how the method can be used to estimate the Ω_m and σ_8 cosmological parameters based on purely observable quantities of galaxy clusters. In the third part of my dissertation thesis, I focus on cataloging hard X-ray sources to understand better the sources contributing to the cosmic X-ray background. I have performed a multifrequency analysis of 3PBC hard X-ray sources to find their low-energy counterparts and assign them classification. The classification was based on their multifrequency information, e.g., luminosities and spectra. The final products of this work are two catalogs: the revised version of the 3PBC catalog based on our multifrequency criteria and the second release of the Turin-SyCAT catalog of Seyfert galaxies. In the fourth part of my thesis, I performed a follow-up analysis of a population of 218 yet unidentified sources, the sources for which there was no identified counterpart in my work described in part three. The goal was the same: to better understand the source population contributing to the cosmic X-ray background. To do this, I searched for their low-energy counterparts in archival *Swift*-XRT data. The final product of this work is two catalogs. First is a catalog of candidate counterparts for 73 3PBC sources, of which 60 have a single candidate counterpart, and 13 have multiple. The second is a short catalog of 10 sources with multifrequency information so we could also provide their classifications.

MUNI
SCI

MASARYKOVA UNIVERZITA
PŘÍRODOVĚDECKÁ FAKULTA
KOTLÁŘSKÁ 2, 611 37 BRNO
IČ: 00216224
DIČ: CZ00216224

ZADÁNÍ DISERTAČNÍ PRÁCE

Akademický rok: 2023/2024

Ústav:	Přírodovědecká fakulta
Student:	Mgr. Matej Kosiba
Program:	Fyzika
Specializace:	Astrofyzika

Ředitel *ústavu* PřF MU Vám ve smyslu Studijního a zkušebního řádu MU určuje disertační práci s názvem:

Název práce:	Analýza strojovým učením a katalogizování extragalaktických objektů
Název práce anglicky:	Machine learning analysis and cataloging of extragalactic sources
Jazyk závěrečné práce:	angličtina

Oficiální zadání:
Neural networks are a powerful tool used in many fields of science and technology. They represent a novel approach to data processing, replacing, at least on routine basis, a painstaking human work, which opens new horizons in the reproducibility, quality and quantity of scientific results. Neural networks have been recently applied to tasks such as pattern recognition and enhancing the quality or improving the angular resolution of astrophysical images. The proposed research is focused on the development and application of neural networks in observational extra-galactic astronomy. It includes the detection and classification of objects and the determination of their empirical, statistical and physical characteristics. The student should have a background in extragalactic astrophysics and should be familiar with recent neural network frameworks, as well as standard data processing techniques.

Vedoucí práce:	prof. Mgr. Norbert Werner, Ph.D.
Konzultant:	Mgr. Filip Hroch, Ph.D.
Datum zadání práce:	31. 7. 2019
V Brně dne:	23. 4. 2024

Acknowledgements

First of all, I would like to thank my supervisors, Norbert Werner and Francesco Massaro. We always had a very friendly relationship. They both gave me a lot of freedom to explore scientific topics that I found interesting. They are both unbelievably good, kind people and brilliant scientists. I had a wonderful time working with them personally and professionally.

I have worked with Norbi since my bachelor thesis project, and I am very grateful that he trusted me and gave me the opportunity to work with him on the very advanced topic of machine learning applications, even though I did not have any previous knowledge in the field. I feel like this was one of the most impactful moments in my scientific career. Ever since he has looked after me and helped guide me in the direction of my scientific career.

I am very grateful to Francesco for his decision to take me as his PhD student and set up a co-supervised project between Masaryk University and the University of Turin, which was another hugely impactful moment in my career. I would like to thank Francesco for introducing me to his ongoing CXB and AGN research; he gave me a lot of opportunities in my scientific career and always guided me forward thinking on how to lay the steps so that it would benefit my present and future scientific career the most.

I am also very thankful to Lukáš Burget, whom Norbi introduced me to and who became my Bachelor thesis co-supervisor. Lukáš spent many hours on a weekly basis with me during my bachelor studies for an entire year and taught me how to understand, code, and use machine learning neural network applications from scratch. He also helped me to learn basic Python coding and Linux skills. We also had a very friendly relationship; I have many good memories from our meetings.

I would like to thank Filip Hroch, who was my teacher and later supervisor of my master's thesis. He is a very friendly teacher, making even the hard subjects feel more approachable. He always helped me with any coding, linux, or latex issues. Without him in the faculty at Masaryk University, a great many students would have failed subjects and would not meet deadlines with their thesis; he is spending a lot of his free time helping undergraduate students, which I see as extremely needed and honorous.

I would like to thank Maggie Lieu, who selected me as her student on my ESAC traineeship. Maggie is an excellent scientist, and under her supervision, I progressed tremendously; working with her in the ESAC was one of the best experiences I have had so far in my scientific career. She always had a very friendly approach, and she was always very open to discussing new ideas. I would like to thank my other supervisors back at ESAC, Bruno Altieri, Ivan Valtchanov, and Sarah Kendrew, for their guidance and support throughout the project. I want to thank all the people from the X-CLASS collaboration, which I joined during my ESAC traineeship, for sharing their expertise and for helping to set up the training dataset for my convolutional neural networks, mainly to Nicolas Clerc, Lorenzo Faccioli, Tatyana Sadibekova, and Elias Koulouridis. During my stay in Italy, I also had a wonderful time. I would like to thank Harold A. Peña-

Herazo and Alessandro Paggi for their help with our 3PBC analysis; it would not have been possible to achieve the catalogs, classifications, and results we published without their expertise and help.

In the same year that I started working with Francesco, I was offered an internship by Marguerite Pierre with the aim of developing a new method for cosmological analysis. I am very thankful to Marguerite for trusting me and selecting me to work on such a challenging project. She is an extremely knowledgeable scientist and always loved sharing her deep knowledge about cosmology with me. I would like to thank Marguerite also because she trusted me that after my internship, I would continue the work and finish our project. I am also grateful that she introduced me to her team. Christian Garrel, Marguerite's PhD student, also loved sharing his knowledge. Christian was always there when I needed to ask any code-related or technical questions. I am very thankful to Tom Charnock, who introduced me to simulation-based inference and devoted many hours to helping me understand the IMNN and ABC methods. He also helped us set up the software and hardware and gave me very important insights about how to approach the PhD studies in general, for which I am very grateful. I would also like to thank Nicolas Cerardi, who took over Christian's position in the project and was thus in charge of creating the analytical XOD simulations I used in my work. I am also very thankful to François Lanusse, who taught me about sequential neural posterior estimation techniques and helped me set up the initial working script for the density estimation.

I would like to thank my beloved girlfriend, Simona Axamitová, who always supported me and encouraged me to do whatever where my passion drew me. I am also very thankful that she was able and willing to visit with me wherever my studies required me to travel. I know that this lifestyle is very challenging. She was like my moral and emotional anchor, bringing stability and peace to my heart. I am also thankful to my friends, mainly JP, a fellow PhD student in Norbi's research group, with whom we had a great time together during our studies. I would like to thank Xena, Íris, and Lili Hamas, my dear degus, whom I loved and love from the deepest depths of my heart. I am also very thankful to my family, mainly my brother Tomáš, father Peter, mother Andrea, and grandparents Irena, Vlasta, and Karol. They have always encouraged me throughout my studies and believed I could accomplish any goal I pursued.

Declaration

Hereby I declare that I have prepared my dissertation thesis independently under the guidance of the supervisors with the use of cited works.

Brno, 2024

Matej Kosiba

Contents

1

Introduction

1.1 MACHINE LEARNING

Author's note: In this section, I provide a basic overview of the neural network architectures and an explanation of how their core features work. I designed it to be a complementary reading for Chapter 2 and Chapter 3, where I introduce my scientific publications in which I used artificial neural networks. However, this overview can not cover all features and mechanisms used in neural networks, mainly the techniques used to fine-tune the architecture and other hyperparameters during the training of the networks. Such a detailed text would require an entire stand-alone book, which is out of the scope of my thesis. I would thus like to kindly refer interested readers to some fantastic literature, e.g., (Bishop, 1997, 1994; Bishop, 2006).

Machine learning (ML) is a large domain encompassing many different types of algorithms, e.g., decision trees (Quinlan, 1986), support-vector machines (Cortes and Vapnik, 1995) or artificial neural networks (Mcculloch and Pitts, 1943) to name a few. ML methods can be used for various tasks, such as classification, regression, clustering, object detection, anomaly detection, segmentation, increasing computational speed, or even searching for new physics.

1.1.1 *Overview of recent advancements in the field*

Artificial neural networks (NN) and ML methods, in general, are becoming utilized significantly more in astrophysical research. Fig. 1.1 shows the number of refereed publications since 2000 up to 2023, included, that had `machine learning` (red) or `neural network` (blue) in their keywords. The purple codes the overlap region of the two. We can see that the number of publications has sharply increased since $\sim$2017, mainly due to huge improvements in the state-of-the-art artificial neural network's architectures.

Even though not visible from this figure, the tide started shifting already in 2012 when Alex Krizhevsky, with his convolutional neural network (CNN) architecture (AlexNet) (Krizhevsky et al., 2012), placed first in the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) of the ImageNet benchmark dataset (Deng et al., 2009), creating a state-of-the-art architecture (and a lecture material on the best universities around the world) for many years. Neural networks were held back until that time by a large degree due to software limitations; the core part of the theory was built already, but we did not have enough processing power and GPU memory capacity. Krizhevsky et al., 2012 created a workaround solution by splitting their CNN architecture into two graphics processing units (GPUs) that had multiple places at which they communicated, essentially mimicking the behavior of one larger CNN than what could have been fed to their GTX 580 GPU that had only 3 GB of memory. They also introduced a new concept that has been used as a standard technique

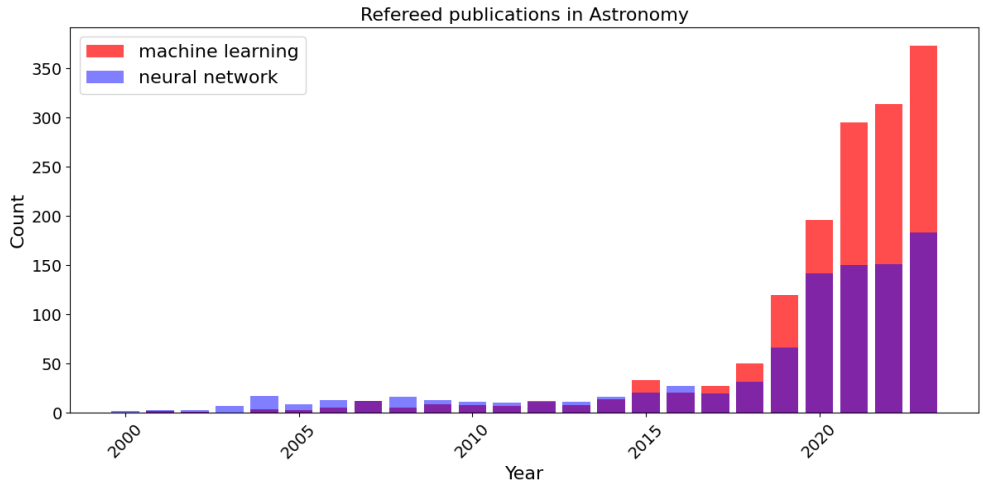

Figure 1.1: Number of refereed publications in astronomy with a 'machine learning' keyword (blue) and 'neural network' keyword (red) since 2000. The Violet color indicates an overlay area. We can see a large increase in publications over the last few years. The data were taken from the NASA Astrophysical Data System (ADS).

ever since, the dropout, a regularisation technique based on randomly switching off a set of neurons in each training iteration. Ever since 2012, the ILSVRC has always been won by CNNs. In 2013, (Zeiler and Fergus, 2013) won the ILSVRC with a modified AlexNet architecture.

In 2014, the ILSVRC was won by GoogleNet (Szegedy et al., 2014). The authors perfected a revolutionary concept, 'network in network' (Lin et al., 2013), in building the CNN architectures, where instead of the usual feed-forward architecture where one layer follows another one in a streamlined fashion (feed-forward neural networks), they created their `inception blocks` where the architecture branches into four paths, each scanning the input to the block under different resolution. These branches' outputs are later concatenated to a single output of the block. They have stacked multiple such inception blocks on top of each other and completely removed the dense layer at the end of the architecture, which eased the majority of the computational time and memory dependency as this region is the bottleneck on the memory load of a neural network. However, in 2014, another state-of-the-art architecture shared the winning placement, the VGG16 (Simonyan and Zisserman, 2014), which beat GoogleNet in the localization challenge. Even though being still a feed-forward neural network, the authors introduced a novel concept of CNN filter sizes, where they used only smaller convolutional filters of 3×3 dimensions (AlexNet also used 3×3 but also 11×11). This was coupled with transforming the information into a depth dimension (z-axis of a 3D array) with max pooling steps followed by an increased number of convolutional filters after each max pooling operation. This concept of smaller filters and progressive increase of depth dimension has been used as a standard architecture building ever since.

The 2015 ILSVRC winner brought us another state-of-the-art masterpiece in the form of ResNet architecture (He et al., 2015), introducing residual skip connections that enabled them to construct very deep neural networks. These

architectures became benchmark models used via transfer learning or training from scratch in a plethora of scientific fields and use cases.

Even though these are advancements in the computer science fields, these innovations hugely impacted astrophysics. This is expected as we are currently in the epoch of big data science, where our astrophysical instruments, both on the Earth and the satellites in space, provide such tremendous quantities of data that astronomy is closely collaborating with data analysis, computer science, and statistics.

Machine learning tools are being applied to all domains of astrophysical research, e.g., X-ray cavity detection in hot atmospheres of galaxy clusters and galaxies (Plšek et al., 2023), simulating galaxy images with deep generative models (Lanusse et al., 2021), deconvolution of images in radio interferometry (Schmidt et al., 2022), denoising of astronomical images (Vojtekova et al., 2021), detecting galaxy clusters in combined X-ray and optical images (Kosiba et al., 2020), detecting solar system objects (Lieu et al., 2019) or estimating galaxy cluster's masses with convolutional neural networks from mock Chandra X-ray images (Ntampaka et al., 2019), to name a few. Another use-case scenario is simulation-based inference machine learning methods, which have been recently used by (Reza et al., 2022) on Quijote latin-hypercube simulations and analytical models to conduct cosmological analysis based on galaxy clusters. Alternatively, (Qiu et al., 2023) used ensemble decision trees, a different ML approach, for their cosmological analysis. They created mock galaxy cluster catalogs based on several hydrodynamical simulation sets, each corresponding to a different cosmology, and trained decision tree models in a classification regime to return an estimate of to which cosmological set a mock galaxy cluster catalog corresponds. de Andres et al., 2024 recently developed an ML method that predicts the overall mass distribution of a galaxy cluster from multi-wavelength observations. They used a U-net architecture (Ronneberger et al., 2015) and trained it on simulations of galaxy clusters, particularly SZ, X-ray, and optical images, having access to the ground truth mass distribution for training and calibration from the simulations. In practice, their U-net can predict the mass of these components of a galaxy cluster when given the images in these three bands.

1.1.2 *Artificial neuron*

We can imagine an artificial neural network as a function consisting of many free parameters. The basic building blocks of a neural network are called neurons (Rosenblatt, 1958). This computational unit simply sums up its weighted inputs, adding a bias to the output of the summation and feeding this number into a non-linear activation function. The non-linear activation function is crucially important as it enables the network to learn complex patterns. Without this step, the entire neural network would perform only a linear regression. The input to the neuron is either the input data or the outputs of neurons in the previous layer. The weights and the bias parameter are the trainable parameters of the neuron that we want to optimize during neural network training. Fig. 1.2 shows a drawing example of an artificial neuron having three inputs.

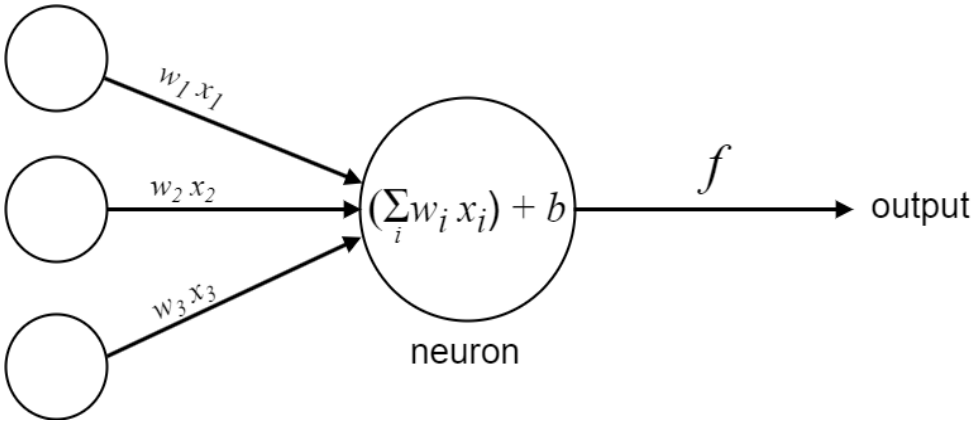

Figure 1.2: Example of an artificial neuron. It has three weighted (w) inputs, x_1, x_2, and x_3, that are summed. The bias term b, a specific trainable parameter for each neuron, is added, and the output of this operation is transformed through a non-linear activation function f, creating the output of the artificial neuron.

1.1.3 *Dense layers*

Artificial neurons are linked together, creating layers. The basic artificial neural network's architecture consists only of dense layers. The term hidden layer refers to all the layers between the input and the output, and the term deep learning refers to neural networks with more hidden layers, typically more than three. The more layers a neural network has, the deeper it is. A dense layer is a stack of artificial neurons connected only with the adjacent layers (or input data), so the layer's neurons do not have any connections inside the layer. Fig. 1.3 shows an example of a neural network's architecture consisting of 4 layers. This neural network has an image as its input that is observed by the 1st layer. The neurons in each layer are visualized as white circles. The network's connections and information flow are depicted with black arrows, similar to Fig. 1.2. The final layer has only two neurons that can be used to, e.g., estimate the probability of some classification category being in the image or for a regression task of some physical quantity from the input image, e.g., the luminosity and observed size of an astrophysical source, this depends on the use case. The word flatten between the input and the first layer indicates a dimensionality change of the image data from the 3D PNG image to a 1D vector. This is a required shape of the input data for a dense layer. Each neuron in the 1st layer has a weighted input connection with each pixel of the flattened input image represented as a 1D vector.

1.1.4 *Convolutional layers*

Convolutional neural networks (CNNs) can still have dense layers in their architecture, but as their name suggests, these neural networks use convolutional layers. Convolutional layers do not have artificial neurons but instead consist of a stack of convolutional filters (LeCun et al., 1999), where the weights are the values of these filters. Each filter also has its bias term. The convolutional layer uses filters to scan across its input in the spatial dimension (x and y axis of a 3D

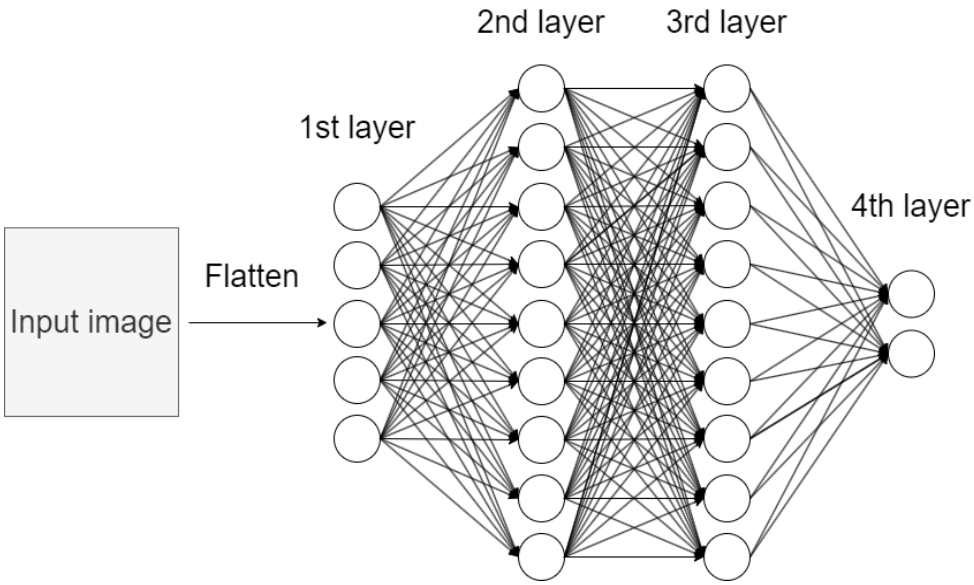

Figure 1.3: Example sketch of a fully connected artificial neural network's layers. The individual white circles represent neurons, and the black arrows, similar to Fig. 1.2, show the weighted inputs to neurons from the previous layer. The dense layer is defined such that each of its neurons is connected with either the input data represented as a flattened 1D vector or neurons of adjacent layers but with none from its own layer.

image). For each scanned place, it performs a convolution on that region across all depth slices (the 3rd dimension of images, which, in the case of astrophysical images, can be used to code different wavelengths). The convolutional filter outputs a single number for each scanned region. The convolutional filter thus creates a 2D convolved representation, a feature map, of its input. The convolutional layer consists of many convolutional filters. Their outputs, the feature maps, are stacked together in the depth dimension, creating a 3D representation of the layer's input. For a case of input data being 3D images, we would use 2D convolutional layers, where each filter automatically sets its depth dimension to fit the third dimension of its input.

Fig. 1.4 shows the convolutional operation in a case of input with dimensions $5\times5\times3$ that is scanned by a convolutional layer with two filters, each of a size 3×3. Note that each convolutional filter gets to have one depth slice for each depth slice of its input. The scanned area is marked with a violet color. I have indicated the result of a convolution operation for each depth slice with a red number next to the first convolutional filter. These numbers are summed together with the filter's bias shown under the filter, and the result is one pixel of the filter's 2D output, the feature map. In this scenario, the convolutional layer has two convolutional filters. Hence, the output of this layer is two 2D feature maps that will be stacked together in the depth dimension, resulting in a $3\times3\times2$ output size. In contrast to the fully connected layer, where the trainable parameters were the weights and biases of the artificial neurons, in the case of convolutional layers, the trainable parameters are the individual values of the layer's convolutional filters together with their bias parameters.

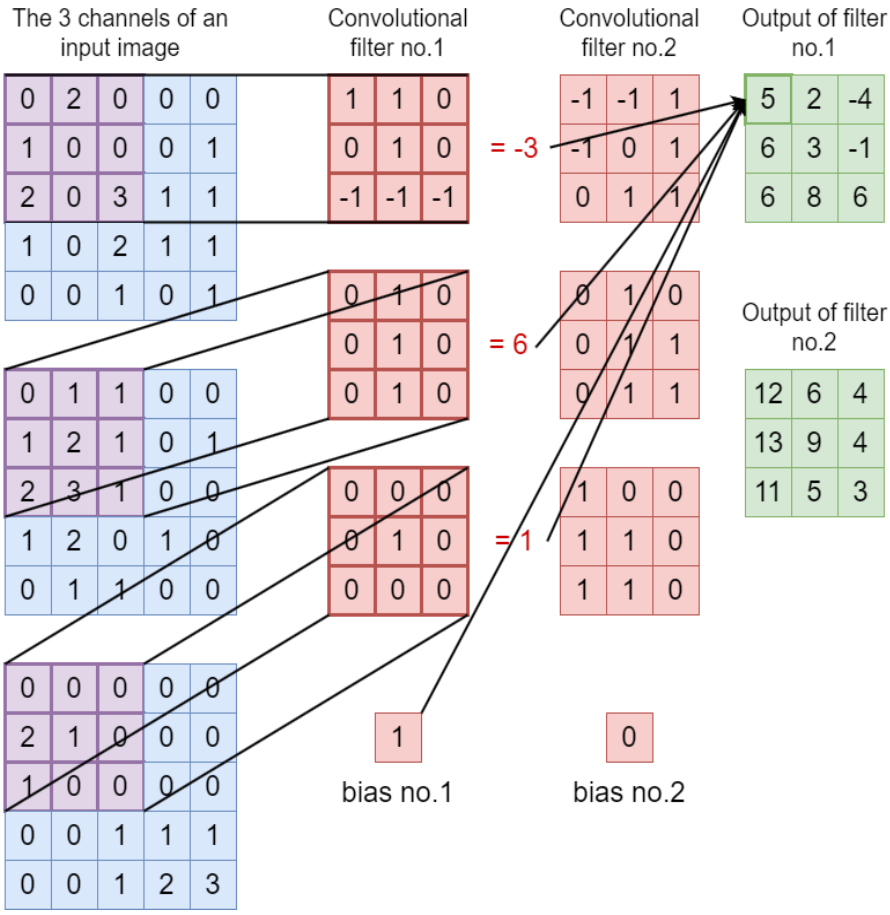

Figure 1.4: Example of a convolutional operation in a convolutional layer consisting of two convolutional filters (red matrixes), each with their bias term below them. This layer is scanning an input of 5×5×3 size (blue matrixes). The situation of the first convolutional filter's scanning of the first position on the input (violet regions) is highlighted together with the convolutional results of each depth slice (red numbers). As the black arrows indicate, these are summed together with the bias parameter to create the first pixel of this filter's output feature map (green).

Giving a complete overview of all details on neural networks would deserve an entire dedicated book, which is out of the scope of this thesis, so I would like to kindly refer interested readers to some fantastic literature, e.g., (Bishop, 1997, 1994; Bishop, 2006). However, I will provide a basic overview of the ML methods I used in my research later in the relevant chapters.

1.1.5 *Max-pooling Layers*

Max-pooling operation is another common feature in the neural network's architectures. It is implemented as an individual layer, but it does not have any trainable parameters. The pooling is essentially used to compress the information in the spatial dimension. Similar to a convolutional layer, the max pooling layer has a filter that scans its input and, in the case of max pooling, outputs only the maximum pixel value in the scanned region. As the filter moves across the image, it creates a compressed representation of the input, whose dimensions

are defined by the max-pooling filter size and the pixel step at which it moves on its input (the stride). Fig. 1.5 shows a 2×2 max pooling filter scanning the input with a stride 2. Its input has a 5×5 size with 1 zero padding, thus creating a 3×3 dimensional compressed representation of its input. The padding is used to control the desired output dimension by artificially enlarging the input. All these parameters, the max-pooling kernel size, the stride, and padding, are called hyperparameters, which must be defined by the user. None of them are, however, trainable parameters tuned during training. Max-pooling works like a bottleneck that shrinks the information in the spatial dimension, keeping only the strongest signal, the most relevant features.

1.1.5.1 *Max-pooling Layers*

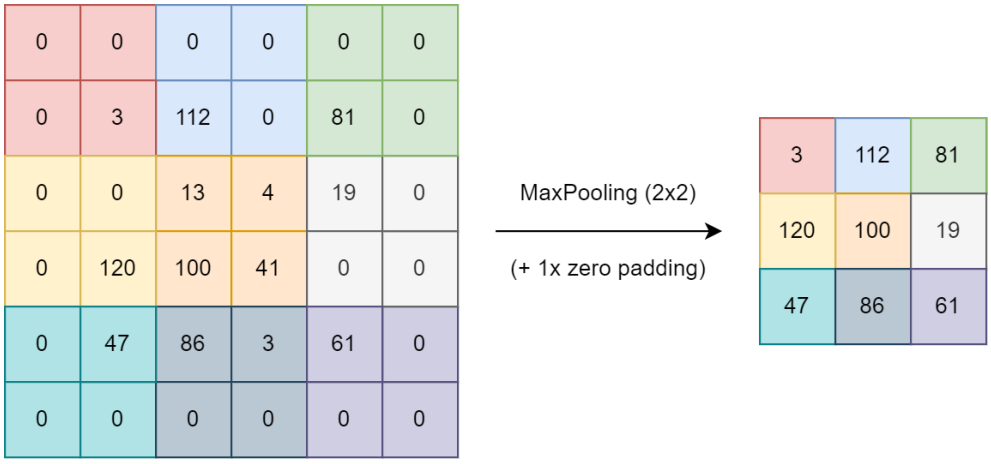

Figure 1.5: Maxpooling operation of a kernel size 2×2 scanning a 2D 5×5 input embedded by padding = 1 with a stride 2. This results in a 3×3 output compressed representation of the input. Color codes the output of individual max-pooling scans of the input data.

1.1.6 *Training*

Before the training starts, all the trainable parameters in the network, e.g., the weights, biases, or convolutional filters, are initialized. Xavier Glorot's initialization (Glorot and Bengio, 2010) or He initialization (He et al., 2015) are some common choices for the initialization, to name a few. The training process of a neural network can be a bit different based on the type of neural network we use and its use case. For example, in an unsupervised clustering task, we would present the neural network data without additional knowledge of the classification categories and essentially train the model to learn the relevant features it has to use to discriminate and categorize data independently without any prior labels. In the case of a supervised classification task, which is the type of problem I considered in my classification of galaxy cluster candidates, we have a dataset with classification categories that are presented to the neural network together with the data. A typical scenario of training a neural network for a supervised classification task on image data is splitting our annotated images into three

datasets: training, validation, and testing. The training data are used to train the network. The validation data are used to monitor its performance during training. The network does not update itself based on its validation prediction. We make it compute predictions on the validation data to see how its performance progresses during training. The network predicts the validation data after a certain number of training iteration steps or after one epoch (once it goes through the entire training dataset). The testing data are never used until the very end of the project when we use them to make the final predictions. This is important because, during the training phase, where we look at how the network performs on the training data and the validation data, we are augmenting its architecture and hyperparameters, essentially tuning it to perform the best specifically on the validation sample, which could in principle introduce undesired biases.

Instead of feeding the neural network with our entire training dataset at once, we commonly let it make predictions only on a small subset of the data, a batch. One of the many parameters of a neural network is a loss function[1] defined by the user based on the task the network is trained to perform. This is the metric used to evaluate the performance of our model. In the case of a supervised multi-class classification, this would be categorical cross-entropy. Let's consider an example of an image classification task of five categories. The prediction output of our network for an input image of the sky would be a vector of five float numbers, all together summing exactly to one, such that they could be interpreted as probability estimates of the source on the image being a particular classification class.

Once our neural network makes predictions for a batch of data and we have a result of the loss computed based on our loss function, the backpropagation algorithm computes gradients of each trainable parameter with respect to the computed loss. Its high efficiency is based on clever use of the chain rule (Leibniz et al., 1920[2]), avoiding the computation of costly derivations. Once we have the gradients, an optimization function (e.g., SGD or ADAM, (Ruder, 2016)) computes the direction and amplitude of the weight change so that the neural network would have a smaller loss if it had to classify these data again. This weight update happens after each batch. The optimization function comes with one of the most important hyperparameters, the learning rate, which is a coefficient directly tuning the amplitude of the weight update's change. Modern neural networks also use clever tricks on how to dynamically change this parameter during training (e.g., Xu et al., 2019) so that the network would better overcome local minima or saddle points in the multidimensional landscape of trainable parameters in its desired way towards the global minima.

1.2 GALAXY CLUSTERS & COSMOLOGY

1.2.1 *Galaxy Clusters*

Galaxy clusters are the largest virialized gravitationally bound objects in the observable Universe. Their main components in the total mass are the galaxies (∼2%), extremely hot gas of low density (∼12%), and a mysterious component,

1 We can also find it called error function or cost function in the literature

2 The chain rule was already mentioned by G. W. F. von Leibniz in 1676, later translated from Latin by J. M. Child in 1920

the dark matter ($\sim 86\,\%$), which particles have not been detected yet and we observe only its indirect gravitational effects (e.g., Sanderson et al., 2013). The smallest galaxy clusters have about 50 galaxies while the largest galaxy clusters consist of more than 1 000 galaxies (e.g., Wen and Han, 2018). There is no sharp boundary on the total mass between galaxy groups and galaxy clusters. Studies in the literature usually report a threshold of $\sim 10^{14}\,M_\odot$. Besides cosmology, galaxy clusters are frequently used as astrophysical laboratories to study highly energetic phenomena on extragalactic scales. The SMBH living in centers of the cluster's member galaxies have such powerful jets that they can interact even with the ICM of the galaxy cluster, depositing energy in the ICM (e.g., Fabian, 2012.

Galaxy clusters grow via mergers that result in shock fronts heating the ICM (Botteon et al., 2018; Ha et al., 2018; Macario et al., 2011). These mergers also provide a tool to study the nature of dark matter because we can investigate how it interacts with itself during collisions. One of the best examples is the bullet cluster (Robertson et al., 2017). This system shows two galaxy clusters that are merging and they already experienced their first encounter. We can thus observe how the galaxies behave in these collosal collisions thanks to optical observations, the ICM thanks to the X-ray observations and the mysterious dark matter thanks to tracking the overall mass distribution via weak lensing that yields us the footprint of the gravitational potential. The galaxies behave like a collisionless fluid. The ICM, however, experienced a collision, which slows it down, thus stripping it from the individual cluster's centers, creating shock fronts radiating at radio wavelengths. Finally, the weak lensing revealed that the overall gravitational potential is centered on the galaxies, however, they constitute only a fraction of the entire baryonic mass budget of the cluster, with the ICM gas being the most massive baryonic component. This strongly indicates the presence of dark matter and its collisionless behavior when encountering such a flythrough in a galaxy cluster merger.

1.2.1.1 *Optical & infrared identification of galaxy clusters*

In the optical and infrared wavelengths, galaxy clusters appear as overdensities of galaxies. George O. Abell liad one of the first cornerstones in the galaxy clusters research, creating the first catalog of galaxy clusters (Abell, 1958). Abell manually worked with the Palomar optical plates with a bare-eye inspection to search for overdensities of galaxies. His first catalog of galaxy clusters contains 2 712 galaxy clusters, focusing on the northern sky. Abell built upon this work, surveying also the southern hemisphere where he identified 1 361 galaxy clusters. After Abell's death[3], his colleagues published the southern and northern catalogs collectively containing 4 073 optically identified galaxy clusters under Abell's name. This work is now known as Abell's catalog of 4 073 rich galaxy clusters (Abell et al., 1989).

Abell defined clusters as overdensities of galaxies that had magnitude larger than $m_3 + 2$, where m_3 was the magnitude of the 3rd brightest cluster's galaxy. The next criterion was the number of galaxies. Abell defined galaxy clusters as galaxy overdensities with at least 50 of the galaxies in the mentioned magnitude range. They have been further subdivided into a few categories based on their

3 George O. Abell passed away on October 7th, 1983.

richness, and the number of cluster's galaxies. Galaxies were included in the galaxy cluster if they were within Abell's counting radius[4]. But to calculate the physical radius, Abell first needed to determine the distance of the galaxy cluster, its redshift. Abell based his distance estimate on the apparent brightness of the 10th brightest cluster galaxy, m_{10}, and finally defined his counting radius as $1.72\,z$ arcminutes. Abell also excluded clusters closer than $z = 0.02$ to cut off nearby galaxy clusters. This was necessary because such clusters' galaxies were distributed across more than a single Palomar plate, which would introduce unwanted biases, making his selection inconsistent when conducted based on the same criteria for such low redshift clusters. He also set an upper limit of 0.2 z, corresponding to $60\,000 \mathrm{km\,s^{-1}}$ based on his estimates due to the magnitude brightness limit of the m_3 galaxy. It was later established that Abell's catalog contains some clusters exceeding this redshift boundary.

Since Abell's first catalog of galaxy clusters, much progress has been made. Technological progress granted us the capabilities to construct telescopes and satellites observing in wavelengths previously unreachable, allowing us to search for galaxy clusters in infrared (e.g., Two Micron All Sky Survey (2MASS) (Skrutskie et al., 2006), Wide-field Infrared Survey Explorer (WISE) (Wright et al., 2010)), X-ray (e.g., Röntgen Satellite (*ROSAT*) (Hasinger et al., 1999), X-ray Multi-Mirror Mission Observatory - Newton satellite (*XMM-Newton*) (Jansen, 1999) and the eROSITA X-ray instrument on board of the Spectrum-Roentgen-Gamma mission (Predehl et al., 2021)) or microwave millimeter wavelengths (e.g., South Pole Telescope (SPT) (Carlstrom et al., 2011)).

Wen et al., 2018 created a catalog of 47 600 galaxy clusters, in which 26 125 were new detections. They also identified new 779 candidate galaxy clusters in the X-ray wavelengths crossmatching their catalog with existing data of the *ROSAT* and *XMM-Newton*.

In the same year, another catalog of optically selected galaxy clusters was published based on The Sloan Digital Sky Survey Data Release 9 (SDSS DS9) (Ahn et al., 2012). This catalog (Banerjee et al., 2018) provides 46 479 galaxy clusters in the redshift range of $0.045 \leq z < 0.641$ of a $\sim$ 11 500 $\mathrm{deg^2}$ survey area.

Just at the time of writing this thesis, Wen and Han, 2024 published a catalog of 1.58 million galaxy clusters using the DESI Legacy Imaging Surveys data. This survey is a joint project of three public projects: the Dark Energy Camera Legacy Survey, the Beijing-Arizona Sky Survey, and the Mayall z-band Legacy Survey. The joined project covers more than 14 000 $\mathrm{deg^2}$ of the northern sky. The DESI Legacy Imaging Survey database provides a catalog of sources and sky images in the g, r, and z bands in the optical wavelengths together with the satellite data of the Wide-field Infrared Survey Explorer (*WISE*) satellite (Wright et al., 2010) in the mid-infrared bands at 3.4, 4.6, 12, and 22 μm.

1.2.1.2 *X-ray identification of galaxy clusters*

In the X-ray wavelengths, galaxy clusters are detected as objects characterized by extended X-ray emission. The source of their extended X-ray emission is nowadays established to be due to bremsstrahlung radiation from their hot, diffuse intracluster gas (Felten et al., 1966). However, nearby galaxies that are

4 Later know nas Abell's radius

massive enough to host X-ray atmospheres are also sources of extended X-ray emission and can appear highly similar to galaxy clusters in the X-ray images. This is why constructing X-ray-selected galaxy cluster catalogs usually involves manual screening of the selected cluster candidates.

There have been many galaxy cluster catalogs constructed based on the Rosat All-Sky Survey data (Truemper, 1992, Voges, 1993) constructed thanks to the *ROSAT* satellite. Piffaretti et al. (2011) created a meta catalog unifying 12 previous catalogs based on the *ROSAT* observations, the Meta-Catalogue of X-ray detected Clusters (MCXC) of 1743 galaxy clusters.

More recently, Xu et al. (2022) decided to implement new dedicated source detection and characterization algorithm developed specifically to search for extended X-ray emission in the X-ray low-count images on the archival *ROSAT* observations. They created a catalog of 944 galaxy clusters and groups, out of which 641 galaxy clusters were already identified based on their X-ray emission, 154 clusters were detected only in optical or infrared wavelengths and they found 149 clusters identified for the first time.

Multiple galaxy cluster catalogs have been constructed based on the *XMM-Newton* observations. The XXL Survey collaboration created a catalog of 365 well-defined galaxy clusters specifically designed for cosmological research (Adami et al., 2018). When using galaxy clusters as cosmological probes, it is essential to have precise knowledge of their mass function and the sample has to have a high degree of purity. Their clusters are identified with a mean number of 6 spectroscopically identified galaxies. This galaxy cluster's catalog is based on the XMM-XXL north and south fields, covering 25 $\deg^2$ each.

In the X-CLASS collaboration, we published a catalog of 1646 X-ray-selected galaxy clusters (Koulouridis et al., 2021). This catalog results from a serendipitous galaxy cluster's search in *XMM-Newton* archival observations up to August 2015. For this catalog, we used XAmin v3.5 (an older version compared to the work of Adami et al., 2018). XAmin cluster detection pipeline is designed for detecting galaxy clusters in the low-count raw *XMM-Newton* X-ray images. Once the XAmin creates a list of galaxy cluster detections, they are split into multiple samples. The interesting one is the 'C1' sample which is designed, based on the XAmin characteristics, such that it should have very little contamination. However, the XAmin v3.5 we used in this work had some problems resolving AGNs, mainly when embedded in an extended X-ray emission of a nearby galaxy or a galaxy cluster, resulting in more false detections. We thus conducted a manual inspection of a few thousand detections. This procedure involved visually comparing the X-ray raw image, the wavelet-smoothed X-ray image, and the optical Digitized Sky Survey POSS-II (DSS2) (Lasker et al., 1996) counterpart image with the X-ray contours superimposed and a zoomed DSS2 image on the central region of the detection. Two experts examined each source. We had three independent experts not participating in the mentioned classification overseeing all classifications. These three were charged to decide on the final classification. The XXL Survey collaboration team also did this manual inspection process. However, they fortunately used a newer version of the XAmin, resulting in much less contamination of their C1 sample.

Very recently, at the time of writing this thesis, the eROSITA team released their first catalog of galaxy clusters (Bulbul et al., 2024), presenting 12 247 op-

tically confirmed galaxy clusters and groups based on 13 116 $\deg^2$ eROSITA western galactic field observations in the $0.2 - 2.3$ keV energy range.

1.2.1.3 *SZ identification of galaxy clusters*

The Sunyaev-Zeldovich effect (Sunyaev and Zeldovich, 1972) describes the interaction of the cosmic microwave background (CMB) photons with the ICM. As the CMB photons interact with the ICM electrons, they are inverse-Compton scattered to higher energies. Therefore, looking at a galaxy cluster at a certain wavelength, we observe a deficit of CMB luminosity and at a higher wavelength, we observe a corresponding increment of the CMB luminosity. This method is particularly useful in detecting high redshift clusters because it is redshift-independent. Its sensitivity depends on the ICM pressure, thus favoring massive unrelaxed clusters and clusters undergoing major mergers with other clusters.

Based on the Sunyaev-Zeldovich effect (Sunyaev and Zeldovich, 1972), (Bleem et al., 2015), using the SPT observing in microwave bands constructed a catalog of galaxy clusters (SPTSZGALCL). The SPTSZGALCL is based on SPT observations covering 2 500 $\deg^2$ and provides 677 galaxy clusters above the signal-to-noise threshold of 4.5.

Planck Collaboration et al., 2014 created The Planck catalog of Sunyaev-Zeldovich sources (PSZ) using the SZ effect in the 15.5 months of the *Planck* satellite all-sky survey data (Planck Collaboration et al., 2020). The PSZ catalog contains 1 227 detections, among which 861 confirmed galaxy clusters, where 178 were confirmed with follow-up observations, and 683 were already known clusters.

Using a similar methodology and 29 months of the Plank observations, Planck Collaboration et al., 2016 updated this catalog and released The second Planck catalog of Sunyaev-Zeldovich sources (PSZ2) containing 1 653 detections, from which 1 203 confirmed galaxy clusters using multi-wavelength search in the available literature, in radio, microwave, infra-red, optical, and X-ray data sets.

At the time of writing this thesis, Bahk and Hwang, 2024 compiled the data of galaxy clusters and galaxies with spectroscopically available redshifts and released an updated galaxy cluster catalog of the second Planck catalog of Sunyaev-Zeldovich sources (PSZ2). They took the centroids of the yet unconfirmed Sunyaev-Zeldovich detections, created a search radius of 4 555 kms^{-1} in the velocity space and of a 15 arcmin radius to search for candidate member galaxies for these SZ detections. If a detection contained more than nine galaxies matching these criteria, they consider them strong candidates. Following this methodology, they updated the PSZ catalog, validating 139 yet unconfirmed SZ detections and updating the redshift information of 399 previously confirmed galaxy clusters.

1.2.2 *Galaxy Clusters as Cosmological probe*

In the hierarchical clustering scenario of structure formation in the Universe, galaxy clusters emerge from the quantum fluctuations in the primordial density field (e.g., Peebles, 1993, Coles and Lucchin, 1995, Bond et al., 1996, Peacock, 1999). Their number density is thus very sensitive to the underlying cosmological parameters of the Universe (e.g., (Kofman et al., 1993; Press and Schechter, 1974)). Galaxy cluster's number density and mass distribution across redshift

is particularly sensitive to Ω_m and σ_8. The Ω_m parameter stands for the mean fraction of the energy density of all forms of matter, baryonic and dark, in the total energy budget of the Universe, while the σ_8 is linked to the amplitude of the fluctuations in the time-space continuum at the birth of the Universe. The number 8 indicates that it is on a scale of $8\,h^{-1}$ Mpc. These parameters are essentially physical characteristics of our Universe. Due to their nature, they determine when (at what redshift) and how the structures in the Universe form. We can thus calculate these cosmological quantities based on large-scale structure observations, the growth rate of galaxy clusters, the mass distribution of galaxy clusters, and their distribution across redshift (e.g., Oukbir and Blanchard, 1992, Eke et al., 1998).

Unfortunately, the mass of a galaxy cluster is not a directly observable characteristic. However, we can use X-ray spectra to calculate the X-ray luminosity and temperature of the hot ICM cluster's gas (Mitchell et al., 1979). The higher the mass of a galaxy cluster, the higher the luminosity and temperature of its ICM. We can use this to our advantage, using scaling relations to derive the total cluster's mass from its temperature or luminosity.

One has to be careful when using scaling relations to derive the cluster's masses, as they can introduce biases because some effects can cause deviations from simple scaling relations. For example, effects such as radiative cooling of the ICM, feedback from supernovae, and AGNs that deposit energy into the ICM originating from cluster's galaxies. Another effect causing discrepancy of mass estimates from scaling relations is the galaxy cluster's mergers with other clusters or smaller groups of galaxies, which cause shocks and heating of the ICM.

1.3 CATALOGING HARD X-RAY SOURCES

In my work presented in Chapter 4 and Chapter 5, I mainly focused on analyzing multi-wavelength archival data with the goal of finding new counterparts to yet unclassified hard X-ray sources of the 3PBC catalog (Cusumano et al., 2010). In this section, I introduce the Cosmic X-ray Background (CXB) and the unified AGN model (Antonucci, 1993, Urry and Padovani, 1995) together with the main types of AGNs. This context is relevant to my research on cataloging hard X-ray sources with their lower-energy counterparts.

1.3.1 *Cosmic X-ray Background*

Scorpius X-1 was the first observed cosmic X-ray source (Giacconi et al., 1962). Together with this observation, we also discovered that the Universe has a component of diffuse X-ray radiation uniform across all directions except the galactic plane, the Cosmic X-ray background (CXB)[5].

The CXB stretches across a wide range of energies, from ∼100 eV up to 1 MeV. Its integrated energy density constitutes ∼10 of the total energy in the cosmic radiation.

In the first decades of CXB research, it was not clear what source was behind its origin. There was a hypothesis explaining the CXB due to a diffuse

5 In older literature, the CXB is called X-ray background (XRB)

intergalactic gaseous source of X-ray radiation emitting the light by thermal bremsstrahlung radiation (e.g., Cowsik and Kobetich, 1972, Marshall et al., 1980, Guilbert and Fabian, 1986, Taylor and Wright, 1989). While this model could fit the observed CXB spectrum well, it struggled to explain how the intergalactic gas would have heated so much to provide the observed CXB radiation at the present time via bremsstrahlung radiation. The second leading explanation was a model describing the CXB radiation as an integrated light of a population of distant X-ray point sources (Giacconi et al., 1979a).

The ROentgen SATellite *ROSAT* (Hasinger et al., 1999) deep images obtained in the early ninety's provided a huge step towards our understanding of the CXB nature, resolving about 80 % of the CXB between 0.5 keV and 2 keV to originate from discrete X-ray point sources of cosmological origin (Hasinger et al., 1998), as suggested by Cavaliere and Fusco-Femiano (1976). A few years later, the *Chandra* X-ray Observatory (Weisskopf et al., 2000) and the *XMM-Newton* satellite (Jansen, 1999) provided even better insights on the CXB. Finally, Chandra's deep-field south observations resolved the majority of the CXB due to individual X-array point sources of various AGN types (Gilli et al., 2007). At hard X-ray energies, *Swift* and *INTEGRAL* resolved 2% of the CXB (Bottacini et al., 2012) and *NuSTAR* resolved 35% of the CXB (Harrison et al., 2016) as being due to hard X-ray point sources. It is now confirmed that the majority of the CXB radiation originates from hard X-ray sources of extragalactic origin. A large fraction of these sources constitute various types of active galactic nuclei (AGNs) (Gilli et al., 2007). Understanding the population of different types of AGNs and their characteristics is thus crucial to comprehend the CXB.

1.3.2 *The unified AGN model*

The unified AGN model is an effort to explain the various types of observed characteristics of AGNs with a single model (Antonucci, 1993, Urry and Padovani, 1995). The model explains different observed characteristics and types of objects based on the geometry, the line-of-sight angle at which we observe the AGN with respect to the AGN's rotation axis, the AGN's luminosity, and the AGN jets' presence or absence. Fig. 1.6 is a sketch of the unified AGN model showing the classification classes of objects that would be observed based on these three physical parameters. The model is first split in top and bottom halves. The top half is radio-loud AGNs (with a jet), and the bottom is radio-quiet (without the jets). There is also a vertical split to low-power and high-power left and right hemispheres, respectively. An observing direction nearly exactly to the axis of an active jet is left for the Blazar category. Each of the mentioned classification classes is further split into subcategories, which are briefly introduced in the following sections. However, before going to the individual classes, we should take a look at the morphology and physical characteristics of the AGNs.

The unified AGN model consists of several components. In the center lies a supermassive black hole (SMBH). If it is active and accretes matter, an accretion disk is formed. Usually, the main source of X-ray radiation in AGN is the interaction of highly energetic electrons in the so-called corona with the photons from the accretion disk. These photons are scattered via the inverse Compton effect to X-ray energies (Haardt and Maraschi, 1993). This X-ray radiation may also get reflected from the accretion disk of the central SMBH. This interaction

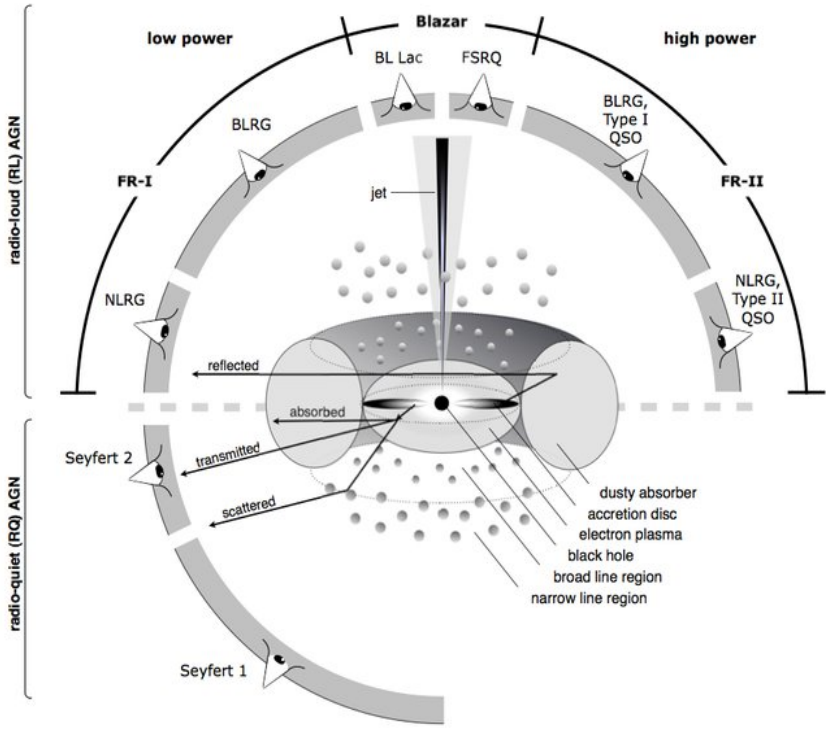

Figure 1.6: Schematic representation of the unified AGN model. The top half of the image shows an active AGN with a jet activity (radio-loud AGN), while the bottom half shows an AGN without jet activity (radio-quite AGN). The figure is also cut into the left (low-power AGN) and right (high-power AGN) halves. The top left represents a radio-loud low-power AGN, and the top right represents a radio-loud high-power AGN, both having a jet activity. The bottom left represents radio-quiet low-power AGN without jet activity. Depending on the angle at which the AGN is observed with respect to the inclination axis of the AGN, the presence of the jet, and the luminosity, we observe different spectral and morphological features. These observational characteristics are used to categorize radio galaxies. Graphic courtesy of Marie-Luise Menzel (MPE), published by Beckmann and Shrader, 2012.

is also responsible for the X-ray fluorescent iron lines (Fabian et al., 2000; Reeves et al., 2001).

At about an order of magnitude larger distance, in the disk plane, lies a torus. At this distance from the SMBH, the temperature allows the existence of dust grains and neutral hydrogen, which form the torus and act as an absorber of optical light and also of X-ray radiation due to the photoelectric effect. The absorbed X-ray and optical light from the central part of the AGN by the torus is subsequently radiated in the infrared spectrum. The torus is relatively thick and extends well above and below the plane of the accretion disk. We can imagine the torus as a doughnut, and in its hollow center lies the SMBH with the accretion disk. Simulations (Krolik and Begelman, 1988) and observations (e.g., Liu et al., 2016) predict the torus to have a clumpy composition.

Going back to the close vicinity of the SMBH, we can find gas clumps scattered spherically around the SMBH that have extreme velocities of thousands $\mathrm{km\,s^{-1}}$ and high-density $n_e \lesssim 10^9\,\mathrm{cm^{-3}}$. These gas clouds are highly ionized and are

responsible for broad emission lines observed in the spectrum. Compared to the distance of this broad emission lines region (BLR), much further away lie gas clumps with an order of magnitude lower speeds and considerably lower density ($n_e \sim 10^3 - 10^6\,cm^{-3}$), responsible for the narrow emission lines observed in the spectrum. This is the narrow emission line region (NLR). Depending on several factors, such as the accretion rate and possibly the spin of the black hole, the SMBH may be launching ultrarelativistic jets. These penetrate the atmosphere of the entire galaxy, reaching up to tens of kiloparsecs away into the intergalactic medium. When they slow down below the speed of sound of the surrounding gaseous medium, they start having fascinating interactions with it and emit synchrotron radiation observed at radio frequencies.

Depending on the observed angle, there are situations in which the region responsible for the broad emission lines will be obscured by the torus. This is not the case for the NLR, in which light is produced at distances much larger where the obscuration does not occur. We will thus not see the broad emission lines in the spectra in a situation where we observe the AGN such that the torus is blocking our view on the central part of the AGN. In a situation where we do not have an obscured view and see directly to the central region of the AGN with a jet, we would observe a source with a much larger luminosity. If the AGN's jet's axis points directly to us, we would observe the most luminous sources also exhibiting relativistic effects due to the relativistic speed of particles in the jet, such as Doppler boosting and relativistic beaming. All these observed characteristics have been used to classify radio galaxies into a generous number of classification classes. It is important to note that many of these classification classes overlap with other ones, which are discussed in the following sections.

1.3.3 *Radio loud and Radio quiet AGNs*

Most of the radio-quiet AGNs are spiral galaxies, while radio-loud AGNs are typically elliptical galaxies, which encountered major mergers in their history. The line and thermal continuum emission of radio-loud and radio-quiet galaxies are very similar. However, the radio-loud galaxies are associated with much larger radio luminosities and distinct radio structures such as lobes and jets, which are created due to the central SMBH interaction with the galaxy's gaseous atmosphere. Based on these observable characteristics, Wilson and Colbert, 1995 propose that the observed differences are not due to the accretion rate and SMBH mass, but instead due to the spin of the SMBH, because the SMBH's spin would be much larger for an elliptical galaxy which's SMBH merged with another one of a similar mass. On the contrary, the supermassive black hole in the AGN of a spiral galaxy would have a much smaller spin as the spin dissipates during the accretion process due to processes such as Penrose process (Penrose and Floyd, 1971) and Blandford–Znajek process (Blandford and Znajek, 1977).

1.3.4 *Fanaroff-Riley classification*

B.L. Fanaroff and J.M. Riley researched 57 extragalactic radio sources (Fanaroff and Riley, 1974) in the radio wavelengths. These were the sources from the 3CR catalog (Spinrad et al., 1985). The studied sources were radio galaxies of low

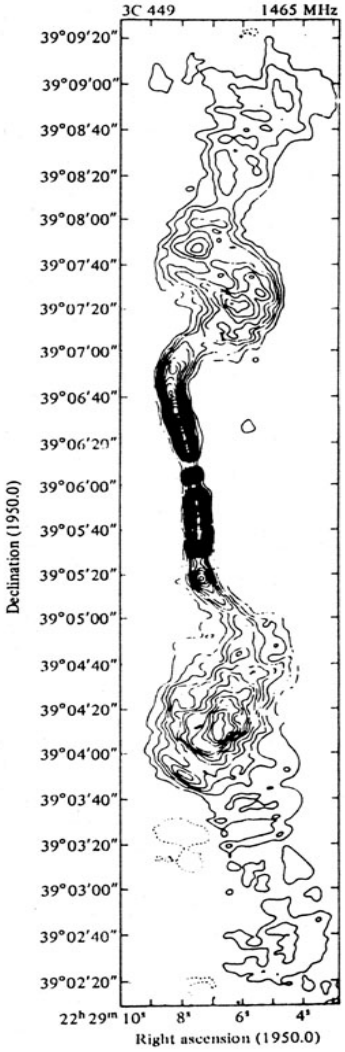

Figure 1.7: VLA map of the FR Class I galaxy 3C 449 at 1465 MHz, with angular resolution 4.8×3.4 arcsec^2 (Perley et al., 1979). We can see large jets that peak at their brightness close to the center of the AGN and decline in brightness as they flow from the galaxy's center.

power (the host galaxy is visible in the optical band) and high power (the optical band shows only a point source; the host galaxy is not visible, quasar). They created a classification system to categorize these sources. First, they calculated the distance of the two spots of the highest surface brightness on opposite sides of the object, D_{RF}. Next, they calculated the total extent of the source up to the lowest brightness contour in the radio map, E_{RF}. Finally, they established a ratio of these two quantities, R_{FR} as

$$R_{FR} = \frac{D_{RF}}{E_{RF}}$$

and defined Class I (FRI) as sources with $R_{FR} < 0.5$ and Class II (FRII) for sources with $R_{FR} > 0.5$. The sources classified in the FRI category turned out to be compact objects with radio emission closer to the galaxy's center. The sources in the FRII category exhibit extended radio structures and highly collimated

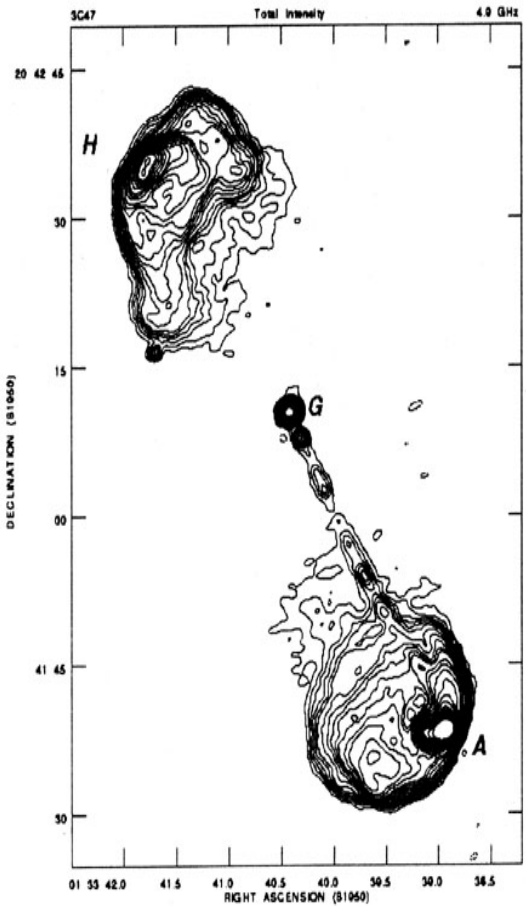

Figure 1.8: 4.9 GHz VLA map of the FR Class II quasar 3C 47 (Bridle et al., 1994) with 1.45×1.13 arcsec^2 resolution. We can observe the bubbles inflated by the get

jets that weaken and interact with the surrounding gaseous envelop that they penetrate later than the FRIs.

Fanaroff and Riley, 1974 found that most of the radio sources in the FRI class had luminosity $L_{178MHz} \leq 2 \times 10^{25} h_{100}^{-2} WHz^{-1} str^{-1}$ while most of the sources in their FRII class had their L_{178MHz} larger than this threshold. They thus discovered that the radio luminosity of these objects correlates with the morphological characteristics of the sources, with the relative position of the regions of high and low surface brightness.

What exactly stands behind these two distinct jet behaviors is still an open question. FRII jets are highly collimated, slowing and interacting with the surrounding medium only at large distances. In contrast, FRI jets start to slow down and interact with the surrounding matter shortly after (in comparison with FRIIs) being launched from the central SMBH. The possible reasons behind these distinct morphological characteristics could be the speed at which the jet is being launched, the density of the surrounding medium, the collimation of the jet, the accretion rate of the SMBH, the mass of the SMBH or the strength of the SMBH magnetic field. It could also be an interplay of these effects or some different mechanisms that have not yet been considered.

Radio galaxies are also classified based on their excitation level into high-excitation radio galaxies (HERG) and low-excitation radio galaxies (LERG). HERGs are nearly always FRII, while LERGs can be either FRI or FRII (Buttiglione et al., 2010).

1.3.5 *FR0s, extension of FR classification*

Recently, a new class of FR galaxies has been established, FR0 (Baldi et al., 2015). The FR0 are sources that do not have any extended radio morphological structures. The FR0s just appear as radio point sources. Compared to the FRIs, the FR0's core-to-total luminosity ratio is about 30× larger (Baldi et al., 2015, Garofalo and Singh, 2019). As FRIs, they are also preferentially found in elliptical galaxies. Baldi et al., 2018 created the first catalog of FR0 galaxies, finding that FR0s are 5× more common than FRIs in their search radius of $z \leq 0.05$.

Several scenarios have been proposed to explain the origin of the FR0. Intuitively, one could assume that these sources are FRIs seen directly in the line of sight of their jet. However, based on the FR0CAT study (Baldi et al., 2018), the authors estimated the abundance of the galaxies according to their FR classifications in the same volume of the local Universe defined by the criteria of the FR0CAT. They found that there were only 21 FRIs and only 1 FRII in this volume, so the 108FR0 completely dominated the population. This ruled out the hypothesis that the FR0 could just be FRIs viewed directly in the axis of their jets. A different scenario suggests that the FR0 have jets of lower Lorentz factors and thus lower jet bulk speeds than that of FRI, possibly due to lower SMBH spin. This could explain why FR0 does not form larger jets because slower jets are more likely to become disrupted by interacting with the surrounding medium before having a chance to launch into the galactic gaseous envelope and beyond.

1.3.6 *Type I and Type II AGNs*

The classification of Type I and Type II AGNs is based on the width of their spectral lines.

The Type I AGNs exhibit broad lines corresponding to velocities of 1 000 – 10 000 $\mathrm{km\,s^{-1}}$. They are further classified into subgroups based on the relative intensity of their broad and narrow spectral lines of the Balmer series.

The Type II AGNs have narrow lines corresponding to gas speed ranging from a few hundred up to 1 000 $\mathrm{km\,s^{-1}}$. The spectral lines are present at near-infrared, optical, or ultraviolet wavelengths.

According to the unified AGN model, this distinction in the spectra characteristics is due to a different angle at which we observe the radio galaxy. The broad line emission originates from gas very close to the central supermassive black hole (SMBH), while the narrow line emission is produced by gas that resides at orders of magnitude larger distances. The observed differences in the broad emission lines are then explained as due to the angle at which we observe the AGN. There are angles at which the central region producing the broad emission lines is completely obscured by the torus, in this scenario, we do not see broad emission lines and we observe a Type II AGN. If we observe at angles closer to the axis of the AGN's jet, we have a direct view of the central part of the AGN,

the accretion disk and the broad emission lines. These objects are classified as Type I AGNs.

This classification is also associated with the classes of Seyfert galaxies and Quasars, which are described in the following sections.

1.3.7 *Quasars*

The first quasars were discovered thanks to the optical follow-up analysis of radio sources. Their optical spectral features initially looked peculiar, as it was not expected to find an optical point source at a high redshift. Schmidt (1963) found that the spectrum consists of well-known features shifted to redder wavelengths implying a very high intrinsic luminosity.

Based on the presence or absence of the broad emission lines in their spectra, quasars are classified into Type I and Type II categories. According to the unified AGN model, this depends on the presence or absence of the obscuring clumpy torus in the line of sight.

Quasars have been classified as radio-quiet[6] or radio-loud based on the ratio of their radio flux density at 5 GHz and optical flux density at 4 400 Å (Kellermann et al., 1989). This classification was later refined by Falcke et al. (1996), who proposed a separate dividing threshold for steep spectrum quasars and flat-spectrum quasars. Radio-quiet quasars are associated mainly with galaxies of lower masses compared to radio-loud quasars (Peacock et al., 1986).

Strittmatter et al. (1980) suggested that two distinct populations could explain the observational characteristics of the radio-loud and radio-quiet quasars. Baloković et al. (2012) investigated this quasar dichotomy. They based their study on 8 307 radio-detected quasars from an SDSS DR7 Quasar Catalog (Schneider et al., 2010) matched with the FIRST survey (Becker et al., 1995). Their best-fit model for this quasar population was a bimodal distribution with a significant overlap that did not describe the data ideally. They thus concluded that the bimodal population is not a suitable description of the quasar population they studied, further stating that the radio properties of quasars are most likely also luminosity and redshift dependent.

1.3.8 *Blazars*

Blazars are characterized by rapid continuum variability at all frequencies from radio to gamma rays, extreme luminosities, and a high linear polarization at optical wavelengths and continuum spectrum of weak or even absent emission and absorption lines, making their redshift classification very challenging (e.g., Urry and Padovani, 1995). When first of these objects were discovered, they were initially classified as stars residing in our Galaxy and only later identified as extragalactic sources of tremendous luminosity. The unified AGN model describes these sources as AGNs which we observe at angles that are very closely aligned to their jet axis. Blazars are subdivided into two main categories: BL Lacertae objects (BL Lac) and Flat Spectrum Radio Quasars (FSRQ). Blazar's radio up to soft X-ray emission is produced by relativistic jet's plasma via synchrotron radiation, while the continuum at higher energies is most likely due to the in-

6 Initially known as radio-weak

verse Compton scattering of photons originating either inside of the jet, from the accretion disk or from the dense gas clouds responsible for the BLR (e.g., Ghisellini and Madau, 1996).

BL Lac Blazars are named after the BL Lacertae, receding in the northern constellation of Lacerta. BL Lacs have a rapid variability, with luminosity changes up to 30 % of their brightness in a time scale of only 24 hours. Over a longer period, their luminosities may change by a few orders of magnitude.

FSRQs are characterized by larger equivalent widths of their optical emission lines ($\gtrsim 5\,\text{Å}$) compared to BL Lacs and higher bolometric luminosities (e.g., Smith et al., 1986). FSRQs thus can exhibit BLR, while BL Lacs have a mostly featureless continuum.

Mondal and Mukhopadhyay (2019) proposed a unified blazar classification model based on two parameters: the strength of the magnetic field and the mass accretion rate. They suggest that BL Lacs are much more magnetically dominated and optically thin than the FSRQs in the jet launching region. To explain the spectral characteristics of blazars, their model describes HSP-BL Lacs as the most optically thin and, thus, with the lowest accretion rate of the blazars classes. The accretion rate is larger for each subsequent class, the ISP and LSP-Blazars until FSRQs which are blazars of the largest accretion rate according to their model. To explain the observed γ-ray luminosities, their model suggests that BL Lacs have a much stronger magnetic field than FSRQs and that the intermediate classes of BL Lacs, the HSP, ISP, and LSP BL Lacs are all objects starting firstly with the strongest magnetic field (HSP-BL Lacs) and gradually as the magnetic field weakens transition through ISP and LSP BL Lacs until reaching FSRQ state.

1.3.9 *Seyfert galaxies*

Seyfert galaxies, despite being low-power radio-quiet AGNs, mainly without the presence of the jet, are still significantly more active than normal galaxies. Seyfert galaxies can be imagined as not-so-bright Quasars, where only the Seyfert's nuclei look more like Quasars, but compared to Quasars, we can still see the Seyfert galaxy in the optical wavelengths and not just a point source as is the case of the Quasars. Seyfert Galaxie's nuclei have bolometric luminosities up to $\sim 10^{40}\ \text{erg}\,\text{s}^{-1}$, which is the threshold I used in my work to distinguish between these two, described in the following chapters.

Nowadays, we know that most Seyfert galaxies are spiral galaxies (Adams, 1977; Heckman, 1978). We classify Seyfert galaxies into Type I and Type II, the same as Quasars, based on the same principle according to the unified AGN model: the presence or absence of the broad emission lines in their spectra.

The unified AGN model's conception that the different spectral characteristics of Seyferts are due to the angle at which they are observed has been recently challenged by Chen and Hwang, 2017, who found that Seyfert Type I are mainly residing in bulge dominant spiral galaxies compared to Seyfert Type II who live in spirals with less prominent bulges. They propose that the difference in Type I and Type II spirals thus has to account also for the different morphology of their host galaxies and should not be solely due to the observed angle of the same type of sources.

1.4 THE QUESTIONS I AIMED TO ADDRESS IN MY RESEARCH

In my first paper (Chapter 2), I address the task of galaxy cluster detection. Even though X-ray-selected galaxy cluster catalogs suffer from much fewer projection effects compared to optically selected clusters, construction of an X-ray-selected catalog is a tedious job where a serendipitous manual inspection of thousands of automatically selected cluster candidates is a necessary step. There are automatic algorithms for cluster detection. However, the XAmin cluster detection pipeline the X-CLASS collaboration had been using at the time I joined this effort had huge contamination with nearby galaxy clusters and X-ray point sources [7]. The contamination with nearby galaxies is not surprising because the XAmin works only with X-ray images, and nearby galaxies with X-ray atmospheres are often indistinguishable from galaxy clusters when considering only X-ray images. Human inspection comparing the optical counterparts can solve this problem, but this process does not scale with modern all-sky surveys (e.g., recent eROSITA data release). My task was thus to create a convolutional neural network cluster detection pipeline that would simultaneously combine information from both, X-ray and optical wavelengths to increase the precision of automatic cluster detection.

In my second paper (Chapter 3), I worked on the development of a new method for cosmological analysis based on purely observable characteristics of galaxy clusters, the X-ray count rate (CR), hardness ratio (HR) coupled with spectroscopic redshift information. The goal was to bypass the need for the use of scaling relations in cosmological analysis with galaxy clusters.

In my third paper (Chapter 4), I analyzed the population of hard X-ray sources provided in the 3PBC catalog. This project aimed to get a complete understanding of the entire hard X-ray source population so we would better understand the nature of the cosmic X-ray background.

In my fourth paper (Chapter 5), I continued working on characterizing the hard X-ray source population. This project aimed to find soft energy counterparts for the yet unidentified hard X-ray sources we catalogized in our previous analysis. We aimed to search for these in the *Swift*-XRT soft X-ray archival data.

7 This issue was later resolved and luckily did not affect the catalog of XXL collaboration that also uses XAmin pipeline.

2

Multiwavelength classification of X-ray selected galaxy cluster candidates using convolutional neural networks

Matej Kosiba,[1,2] *Maggie Lieu,*[2,3] *Bruno Altieri,*[2] *Nicolas Clerc,*[4] *Lorenzo Faccioli,*[5] *Sarah Kendrew,*[6] *Ivan Valtchanov,*[7] *Tatyana Sadibekova,*[5,8] *Marguerite Pierre,*[5] *Filip Hroch,* [1] *Norbert Werner,* [9,1,10] *Lukáš Burget,*[11] *Christian Garrel,* [5] *Elias Koulouridis,*[12,5] *Evelina Gaynullina,*[8] *Mona Molham,*[13] *Miriam E. Ramos-Ceja*[14] *and Alina Khalikova*[8]

[1]Department of Theoretical Physics and Astrophysics, Faculty of Science, Masaryk University, Kotlářská 2, Brno, 611 37, Czech Republic
[2]European Space Astronomy Centre, ESA, Villanueva de la Cañada, E-28691 Madrid, Spain
[3]Centre for Astronomy and Particle Theory, University of Nottingham, UK
[4]IRAP, Université de Toulouse, CNRS, CNES, UPS, (Toulouse), France
[5]AIM, CEA, CNRS, Université Paris-Saclay, Université Paris Diderot, Sorbonne Paris Cite, F-91191 Gif-sur-Yvette, France
[6]European Space Agency, Space Telescope Science Institute, 3700 San Martin Drive, Baltimore MD 21218, USA
[7]Telespazio Vega UK for ESA, European Space Astronomy Centre, Operations Department, 28691 Villanueva de la Cañada, Spain
[8]Ulugh Beg Astronomical Institute of Uzbekistan Academy of Science, 33 Astronomicheskaya str., Tashkent, UZ-100052, Uzbekistan
[9]MTA-Eötvös University Lendület Hot Universe Research Group, Pázmány Péter sétány 1/A, Budapest, 1117, Hungary
[10]School of Science, Hiroshima University, 1-3-1 Kagamiyama, Higashi-Hiroshima 739-8526, Japan
[11]Faculty of Information Technology, Brno University of Technology, Božetěchova 2, Brno, 612 00, Czech Republic
[12]Institute for Astronomy & Astrophysics, Space Applications & Remote Sensing, National Observatory of Athens, GR-15236 Palaia
Penteli, Greece
[13]National Research Institute of Astronomy and Geophysics (NRIAG), 11421 Helwan, Egypt
[14]Max-Planck Institut für extraterrestrische Physik, Postfach 1312, 85741 Garching bei München, Germany

Published in Monthly Notices of the Royal Astronomical Society, Volume 496, Issue 4, August 2020, Pages 4141–4153,
https://doi.org/10.1093/mnras/staa1723

2.1 ABSTRACT

Galaxy clusters appear as extended sources in *XMM-Newton* images, but not all extended sources are clusters. So, their proper classification requires visual inspection with optical images, which is a slow process with biases that are almost impossible to model. We tackle this problem with a novel approach, using convolutional neural networks (CNNs), a state-of-the-art image classification tool, for automatic classification of galaxy cluster candidates. We train the networks on combined *XMM-Newton* X-ray observations with their optical counterparts from the all-sky Digitized Sky Survey. Our data set originates from the X-CLASS survey sample of galaxy cluster candidates, selected by a specially developed pipeline, the `XAmin`, tailored for extended source detection and characterisation. Our data set contains 1 707 galaxy cluster candidates classified by experts. Additionally, we create an official Zooniverse citizen science project, *The Hunt for Galaxy Clusters*, to probe whether citizen volunteers could help in a challenging task of galaxy cluster visual confirmation. The project contained 1 600 galaxy cluster candidates in total of which 404 overlap with the expert's sample. The networks were trained on expert and Zooniverse data separately. The CNN test sample contains 85 spectroscopically confirmed clusters and 85 non-clusters that appear in both data sets. Our custom network achieved the best performance in the binary classification of clusters and non-clusters, acquiring accuracy of 90 %, averaged after 10 runs. The results of using CNNs on combined X-ray and optical data for galaxy cluster candidate classification are encouraging and there is a lot of potential for future usage and improvements.

2.2 INTRODUCTION

Galaxy clusters are massive systems at the peaks of the cosmic web. Their composition, rich in dark matter and hot baryonic gas makes them a potentially powerful tool to constrain cosmological parameters, growth of structure, neutrino mass and sterile neutrinos through cluster number counts, the cluster mass function and the baryon fraction (Allen et al., 2011; Böhringer and Chon, 2016; Mantz et al., 2015).

In recent years, large cluster surveys such as XXL (Pacaud et al., 2016; Pierre et al., 2016), XCS (Mehrtens et al., 2012), X-CLASS (Clerc et al., 2012a; Ridl et al., 2017), *Planck* (Bartlett et al., 2008), redMaPPer (Rykoff et al., 2014), or the SPT-SZ survey (Bleem et al., 2015) have made it possible to statistically improve constraints on cosmology. However one of the challenges in using galaxy clusters for cosmology is understanding and modelling of the cluster selection function (e.g. Pacaud et al., 2006). The selection function has to be modelled in terms of observable parameters (like flux and apparent size), which can then be converted into galaxy cluster mass for a given cosmology and galaxy cluster physics evolution. The selection function of galaxy clusters is not trivial to model and often oversimplified. A selection function should not only take into account the volume and redshift of the survey but also the choice of clusters, which is often more complicated than a cut in flux. In X-ray wavelengths, whilst extended emission is generally a robust indicator of a galaxy cluster, the emission can also be attributed to nearby galaxies, saturated AGN and unresolved double point-sources. For this reason, galaxy cluster candidates are still visually examined

together with optical data, prior to any spectroscopic confirmation (Adami et al., 2018). This process is tedious and out-dated with uncertainties impossible to model. With large X-ray sky surveys such as *e-ROSITA* (Merloni et al., 2012) expecting to discover tens of thousands of new galaxy clusters, combined with large optical surveys including LSST (Ivezic et al., 2008) and *EUCLID* (Racca et al., 2016), the old techniques will become obsolete. We need to prepare for the future with new methods that are able to deal with big data and improved accuracy.

Citizen science projects proved to be a great asset for scientific problems where human classifications are required for large amounts of data (e.g. Lintott et al., 2008; Willett et al., 2013). In the first version of the most well known of all citizen science projects, the Galaxy Zoo (Lintott et al., 2008), citizen volunteers managed to achieve more than 90 % agreement with experts in a task of morphological classification of galaxies. While citizen projects are intended to provide huge manpower in the assessment of large astronomical data sets, the question whether this is an advantage over a limited number of evaluations by experts in the case of the confirmation of galaxy cluster candidates remains to be addressed. This paper scrutinizes this issue by evaluating the citizen volunteers success rate.

Machine learning offers a more constructive approach to the problem. The power of Machine learning has been demonstrated in astronomy for more than two decades, with applications including star-galaxy discrimination (Bertin, 1993; Odewahn et al., 1992), classification of galaxy spectra (Folkes et al., 1996), photometric redshift estimation (Collister and Lahav, 2004) or anomaly detection in X-ray spectra (Ichinohe and Yamada, 2019), to name a few. With the introduction of Convolutional Neural Networks (CNNs, LeCun et al., 1999) and deep learning (E Hinton, 2007), it has been possible to automate human vision tasks such as image recognition (see e.g. Ackermann et al., 2018; Goodfellow et al., 2014; Lieu et al., 2018; Schawinski et al., 2017).

Supervised learning with convolutional neural networks (CNNs) was designed specifically for image classification tasks. If the true labels (classification classes) of the images are known, they can be used to train CNNs. The current way galaxy clusters are classified are liable to false positives and false negatives. Galaxy cluster candidates picked by an automated pipeline are visually analysed by several experts to create an initial catalogue of galaxy clusters, that are later verified with a spectroscopic confirmation. This process will not scale with large data volumes. Citizen science allows us to harness a large number of opinions on each object classification on a short timescale, speeding up the process significantly yet having a reasonable agreement with experts (see e.g. Dieleman et al., 2015; Willett et al., 2013). CNNs can be then trained on classifications made by either experts or citizen volunteers or both, to automate the final classification of galaxy cluster candidates, or even skipping the first step of the pipeline picking the candidate clusters. Applying CNN selection on simulations will enable modelling the selection function.

In this paper, we introduce a citizen science project we created to obtain large numbers of classified objects. We compare the performance of citizen volunteers with experts. We train CNNs on classifications of citizen volunteers and experts and compare their results. CNNs are tested on spectroscopically confirmed galaxy clusters and objects classified as non-clusters by experts.

The structure of the paper is as follows: in Section 2.3 we present our citizen science project and its development together with a description of the observations and the construction of their classifications by the experts, in Section 2.4 we introduce the machine learning methods we use, Section 2.5 presents measurements used to evaluate classification or detection performance, Section 2.6 presents the results of the citizen science campaign as well as the results and discussion of neural networks analysis. Finally, we conclude in Section 2.7.

2.3 THE HUNT FOR GALAXY CLUSTERS

Our citizen science project, *The Hunt for Galaxy Clusters*[1], was launched online as an official Zooniverse project on the 24th of October 2018. There were 1 600 galaxy cluster candidates in the project that have been detected as extended X-ray sources by the `XAmin` wavelet-based pipeline (Pacaud et al., 2006). Each object was classified by at least 30 different volunteers and this was completed by the 29th of April 2019. 1 227 volunteers participated in the project. Classifications of not logged in volunteers, as well as classifications which have been done on each object multiple times by the same volunteer, were not considered.

The project starts with a short tutorial briefly explaining how to navigate in the project's page and how to classify candidate clusters. Each object comes with four images, covering the exact same area of the sky (7×7 arcmin^2): two X-ray and two optical images. Figure 2.1 shows all four images of a galaxy cluster candidate as shown to the volunteers in *The Hunt for Galaxy Clusters*.

Our project uses six questions to help determine the class of a galaxy cluster candidate. Each question has two or three possible answers, and due to the structure of the decision tree (Figure 2.2), only a subset of the questions are answered. Those questions come with help notes, example images, as well as descriptions to each answer. We selected example images very carefully to cover a broad range of objects and/or instrument effects, in order to avoid biases. The Zooniverse volunteer's answers were then used to create a binary classification scheme of *cluster* and *non-cluster*.

2.3.1 *Data*

The data in this work originates from the XMM CLuster Archive Super Survey (X-CLASS) (Clerc et al., 2012b), an X-ray galaxy cluster search in the archival data of the *European Space Agency's* X-ray observatory *XMM-Newton*, combined with corresponding optical counterparts from the Digitized Sky Survey POSS-II (DSS2). We used *XMM-Newton* data obtained between 2000 and 2015, employing selection criteria described in (Clerc et al., 2012b), and excluding the data used by the XXL survey (Pierre et al., 2016).

2.3.2 *X-ray pipeline*

Our sample of galaxy cluster candidates has been constructed using the intermediate `XAmin` 3.5 version (new source models added: double point-source and point + extended source). This version, after the processing of the X-CLASS sur-

1 https://www.zooniverse.org/projects/matej-dot-kosiba/the-hunt-for-galaxy-clusters

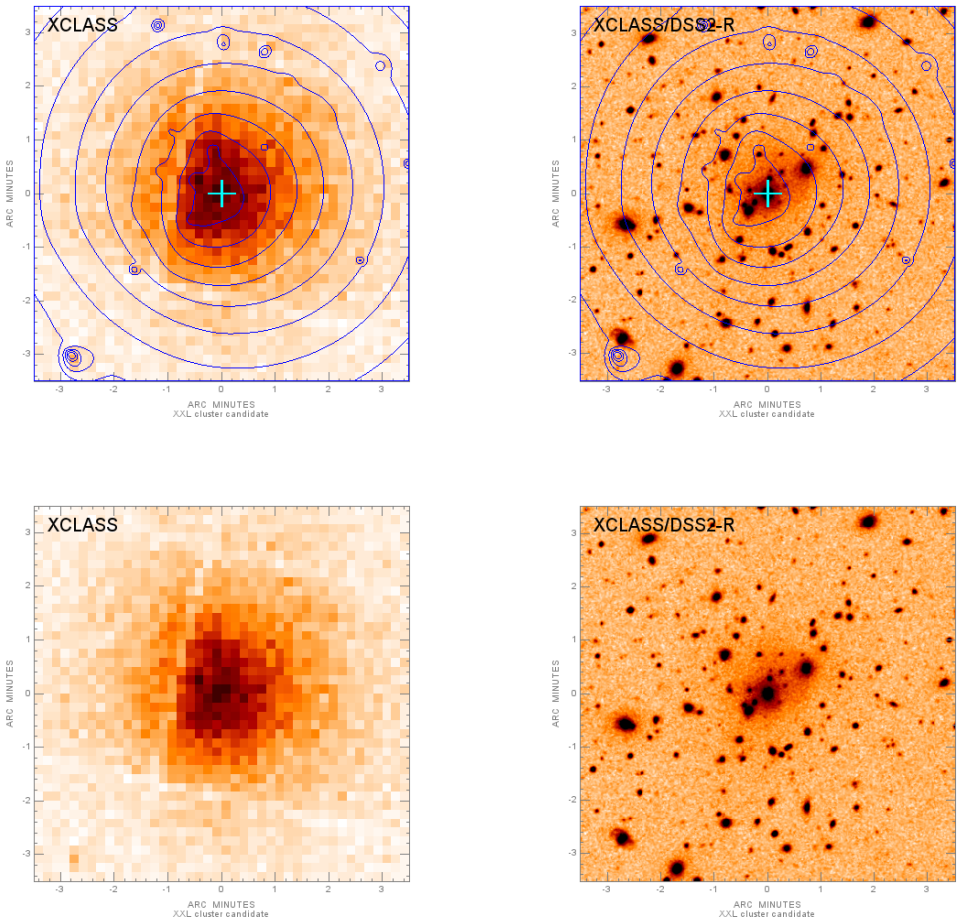

Figure 2.1: *Top left*: raw X-ray image with contours showing the areas of constant X-ray brightness and a cyan cross marking the object selected for classification. *Bottom left*: raw X-ray image without contours and markings. *Right*: corresponding optical images.

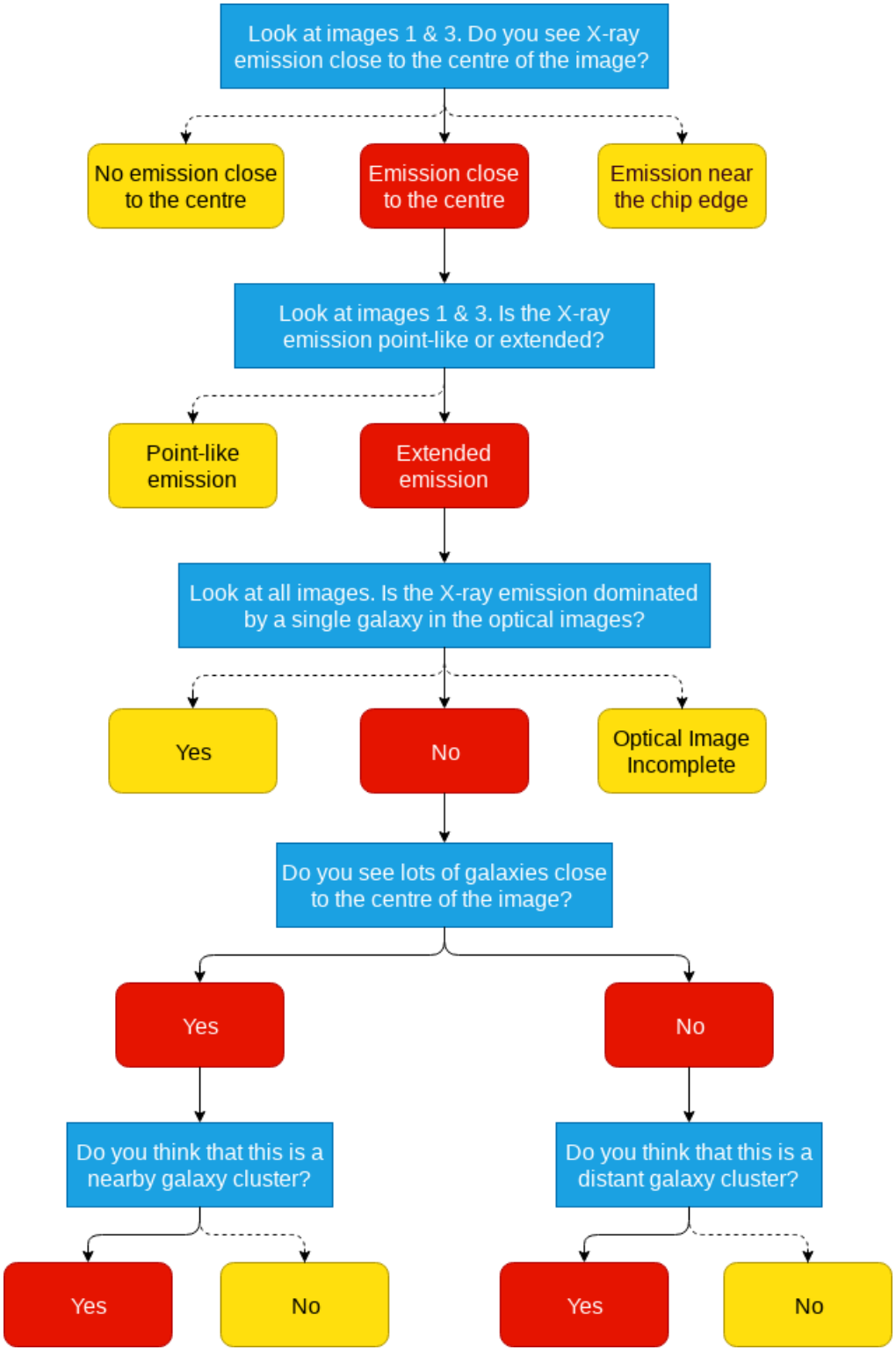

Figure 2.2: The decision tree of *The Hunt for Galaxy Clusters* Zooniverse citizen science project. Blue cells represent questions, red are answers leading to the *cluster* class and yellow are answers leading to the *non-cluster* class.

vey, appeared to suffer from a miss-centering problem randomly affecting a tiny fraction of the point-source population, that led to classify them as extended. In order to remove miss-classified sources, experts then performed an in-depth screening of the putative cluster candidate lists. The screening dealt as well with usual nearby galaxies and saturated AGNs, that both appear extended in the X-ray images

The pipeline is briefly described below. Firstly, a combined MOS1+MOS2+PN image of an *XMM-Newton* (Jansen, 1999) observation is smoothed with a dedicated wavelet smoothing program called mr_filter, described by Starck et al., 1998 and shown in Starck and Pierre, 1998 to effectively recover structures in X-ray images characterised by low numbers of photons.

Secondly, the wavelet smoothed image is analysed by the source extraction software `SExtractor` (Bertin and Arnouts, 1996). It creates a list of candidate sources for further analysis, returning an estimate of their position and their flux.

Note that, since `SExtractor` was developed for optical images which contain many more photons than the X-ray ones, smoothing the X-ray image is a necessity as `SExtractor` would not be able to work with raw data. This smoothing can be performed in several ways; the wavelet smoothing used by `XAmin` is one of the possible ways of smoothing the image and was shown by Valtchanov et al., 2001 to give the best results for X-ray images of diffuse sources like galaxy clusters.

Finally, we characterise the candidate sources found by `SExtractor`. This is done by fitting both a point source model given the *XMM-Newton* PSF computed at the source position and an extended β model (Cavaliere and Fusco-Femiano, 1976) which better describes galaxy clusters. A source is declared to be a point source (AGN or an extended source too faint to be characterised as extended) or an extended source (galaxy cluster) depending on which of these two models best fits the candidates source. The details, including the relevant formulas and the selection criteria for defining an (almost) pure sample of galaxy clusters, are given in Pacaud et al., 2006.

Coordinates of the galaxy cluster candidates picked by `XAmin` are then used to produce normalised images (2.8) with and without X-ray contours to show lines of constant X-ray brightness. These contours are superimposed onto the optical counterpart image, together with a cyan cross mark and are used only for human screening to help visualise the X-ray emission.

2.3.3 *Weighting volunteers classifications*

Since each object is classified by 30 volunteers, we may end up with different classifications for the same galaxy cluster candidate. Each person's classification ability may vary according to the class and the question asked, and there may even be volunteers who purposely create malicious classifications. To mitigate those effects, we weight classifications of each user question-wise. Weighting is done according to the agreement of the majority, so each user has an accuracy

determining a portion of his/her classifications being in agreement with the majority of votes, which is done question-wise,

$$G_i = \frac{C_i}{Q_i}, \qquad i \in 1, ..., 6 \tag{2.1}$$

where G_i is the weight applied for an individual on question i, C_i is the number of answers to question i given by the individual that were in agreement with the majority and Q_i is the total number of answers the individual has made for question i. G_i essentially describes the ability of an individual to classify as the majority of volunteers would. Every classification in the project is then weighted according to the classifying volunteer's accuracy for the specific question. The bottom red leafs of the decision tree (Figure 2.2) are classification ending answers corresponds to the final answers stating that the classified object is a galaxy cluster. Similarly, all yellow leafs corresponds to the final answers stating that the object is not a galaxy cluster. Each galaxy cluster candidate gets 30 votes, each vote is an accuracy of the voting user for the question of his/her classification ending answer (one of bottom red leafs or any yellow leaf). Those 30 weighted scores are summed to galaxy cluster (bottom red leafs) and non-galaxy cluster (yellow leafs) categories. The higher score determines the final Zooniverse weighted classification for the galaxy cluster candidate.

2.3.4 *Classifications of experts*

The galaxy cluster candidates generated by the `XAmin` pipeline are manually classified by the X-CLASS collaboration. Each galaxy cluster candidate is classified by two experts and three moderators make the final classification on conflicting decisions. Figure 2.3 shows how a galaxy cluster candidate is presented to the experts. The images are provided without redshift or sky coordinate information, and the experts make decisions without consulting with each other to avoid any bias. The experts were given the opportunity to classify objects as a low redshift cluster ($0 < z < 0.3$), high redshift cluster ($z > 0.3$), nearby galaxy, point source, star or AGN, double source, artefact, edge, fossil group, high background image, no optical image or dubious source. We create a binary classification scheme where the last four categories in the list are not used, low and high redshift clusters are collectively referred to as clusters and the remaining classes are collectively referred to as non-clusters.

2.4 MACHINE LEARNING APPROACH

Now, we turn our attention to a machine learning approach, which allows us to automatically process astronomical data on much larger scales than what is possible to achieve by human annotations. We use neural networks – a parametric model, that is able to learn to approximate a complex function from training examples of inputs and the corresponding outputs. In our case, each training example consists of combined X-ray and optical image as the input and the corresponding output class label obtained from a human annotator. In our experiments, we consider binary classification, where the class labels are *galaxy cluster* and *non-cluster*, but also multi-class classification with subcategories that

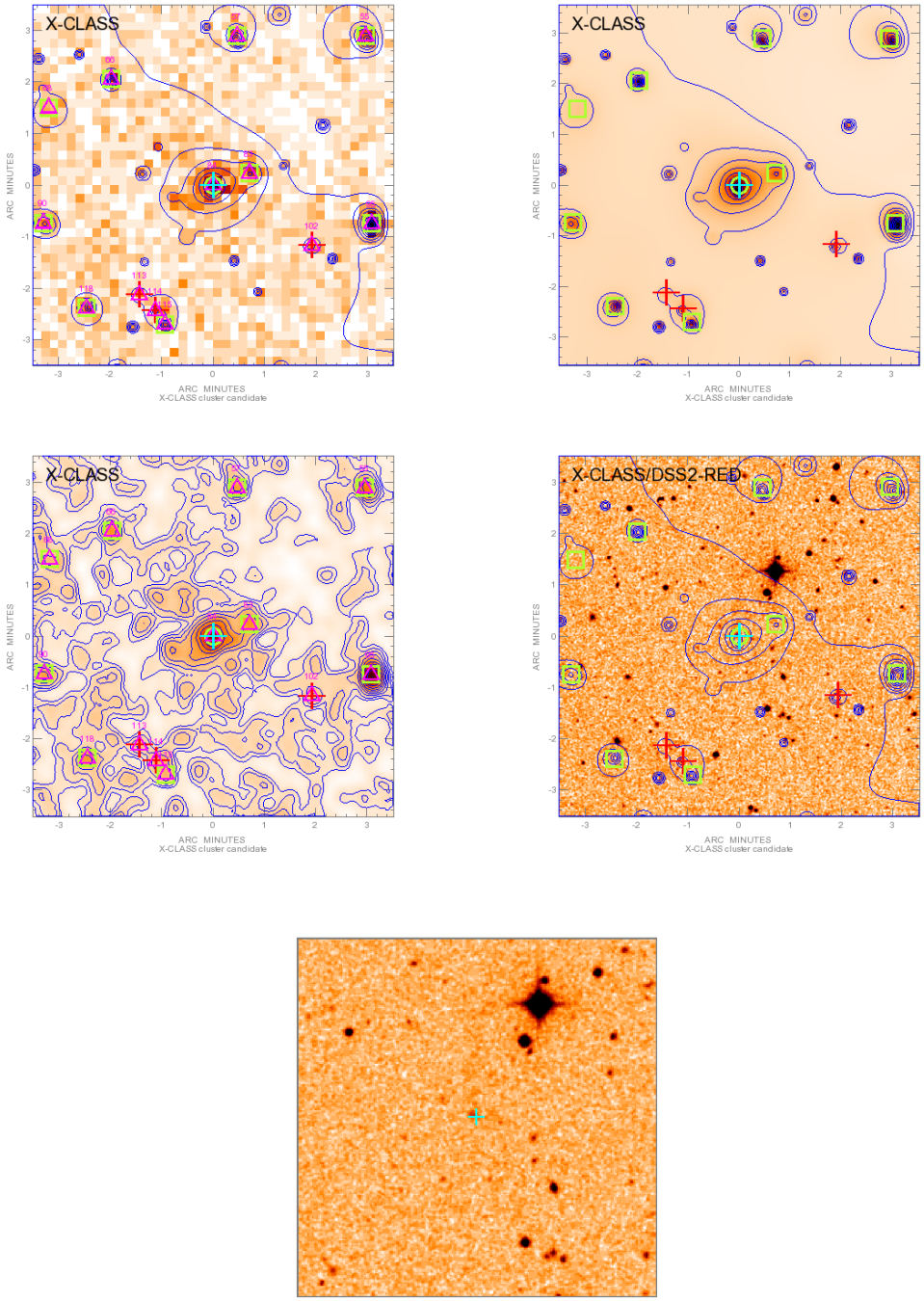

Figure 2.3: Images of a galaxy cluster candidate classified by experts. Top left: an X-ray raw image overplotted with contours showing areas of constant X-ray brightness, and marks produced by the `XAmin` pipeline. Top right and bottom left images are smoothed versions of the X-ray images, wavelet and Gaussian smoothing produced by the `XAmin` pipeline, respectively. The Gaussian smoothed image is overplotted with Gaussian contours, the sigma is chosen to be 3 pixels (with a pixel size of 2.5 arc seconds so the sigma is 7.5 arc seconds). Bottom right: the optical counterpart of the X-ray image with superimposed marks and wavelet X-ray contours. All images cover the exact same area of the sky, 7×7 arcmin^2, except for the bottom panel, where we focus in the central region (4×4 arcmin^2) of the optical image, because with the contours and the symbols it is not easy to see the central cluster brightest galaxy and overdensity of faint galaxies.

will be discussed in Section 2.6.5. From the training examples, our neural networks learn to predict posterior probabilities of all classes given an input image. In our experiments, we evaluate the performance of the neural networks using measures discussed in Section 2.5. For some of the measures, we need to make a hard classification decision for each input image from our evaluation set. In such a case, we simply select the most probable class.

In this work, we use Convolutional Neural Networks (CNN), which is currently the most popular and very effective neural network architecture for image processing (Ciresan et al., 2012; Krizhevsky et al., 2012; Lecun et al., 1989). A deeper knowledge of CNNs is not necessary for interpreting our results and understanding the presented analyses. It is only necessary for understanding some of the technical details. This paper also can not give a complete tutorial to CNNs, therefore, we do not provide a further introduction to CNNs and we kindly refer the interested reader to the relevant textbooks (Bishop, 2006; Goodfellow et al., 2016) or the numerous tutorials available online. We use two CNNs architectures for our experiments: Using the Keras toolkit (Chollet et al., 2015), we build and train our *custom network*, which uses a conventional CNN architecture with interleaving convolutional and pooling layers and final dense layers. The second architecture is MobileNet (Howard et al., 2017). We take these networks as provided by its authors pre-trained on the ImageNet (Deng et al., 2009) data, which is a large data set of millions of real-word images categorised into thousands of classes. We assume that such pre-training can serve as a good initialisation of the CNN parameters, which are further retrained on our training data for galaxy cluster classification.

2.4.1 *Data preprocessing*

For training neural networks, we use images without contours and marks. For each candidate cluster, a pair of X-ray and optical PNG images were merged into a single PNG image. As well as our custom network, we use existing architectures, that were designed to take input images with 3 colour channels. In order to achieve this, we grayscale the X-ray and optical images and stack them together as individual channels, leaving one channel empty (zero-filled) to create a single RGB image. Although training of our custom network can be done with any number of input channels, we use the same 3-channel images as the input to the network unless stated otherwise. By default, we construct the input images as follows: the blue channel contains the grayscaled optical image, the green contains the grayscaled X-ray image and the red is filled with a matrix of zeros (Figure 2.4).

2.4.2 *Data augmentation*

With smaller data sets, the risk of over-fitting increases, resulting in poor generalisation to data outside of the training set. To prevent overfitting, we use data augmentation to reduce the probability that the network will see exactly the same image twice and to essentially increase our training sample size. At each training step, the input image is randomly scaled to a uniform value between $1/1.3$ and 1.3, rotated by a random uniform angle between 0 and $360°$

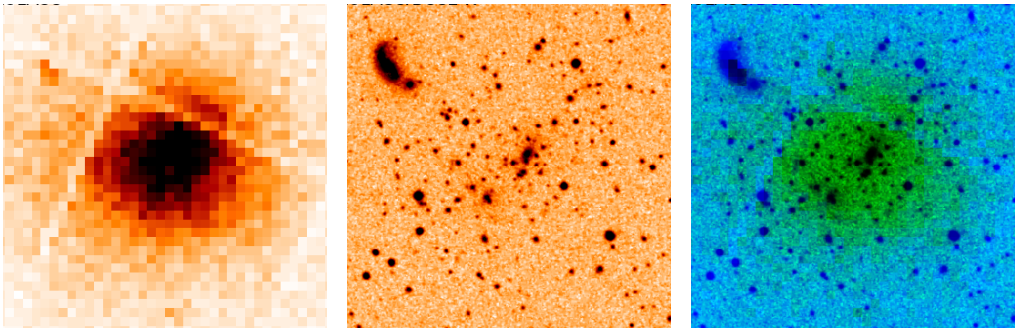

Figure 2.4: Left is the 356×356 pixel X-ray rgb .PNG image, middle is its 356×356 pixel optical .PNG counterpart and right is an rgb .PNG image made by stacking grayscaled optical image as blue channel, grayscaled X-ray image as green channel and the red channel was filled with zeros.

and translated in x and y directions by a random uniform value between −4 and 4 pixels.

2.5 PERFORMANCE MEASUREMENTS

This section describes the measurement methods we chose to evaluate our neural networks compared to a baseline.

Accuracy is the most intuitive performance measurement. It is the ratio of correct predictions to all predictions and is defined as

$$A = \frac{TP + TN}{TP + TN + FP + FN}, \tag{2.2}$$

where TP refers to the number of true positives, in our case the number of clusters correctly classified as clusters, TN is a number of true negatives (number of non-clusters correctly classified as non-clusters), FP is a number of false positives (number of non-cluster incorrectly classified as clusters) and FN states for a number of false negatives (number of clusters incorrectly classified as non-clusters).

Precision is the ratio of the correctly classified positives (i.e. clusters) and all objects classified as positives. This is defined as

$$P = \frac{TP}{TP + FP}. \tag{2.3}$$

Recall is the ratio of the correctly classified positives and all positives examples in the test data. It is defined as

$$R = \frac{TP}{TP + FN}. \tag{2.4}$$

The receiver operating characteristic (ROC) is a performance measurement of detection problems plotted as a true positive rate (recall) against the false positive rate, defined as[2]

$$FPR = \frac{FP}{FP + TN} \tag{2.5}$$

at various thresholds. The area under the curve (AUC) describes the model's capability to distinguish between two classification classes and is independent of the choice of the threshold. When reporting detection performance for a class (from the CNN output) in terms of ROC curve, we compare the posterior probability of the class to a varying detection threshold.

2.6 RESULTS AND DISCUSSION

2.6.1 *The Hunt for Galaxy Clusters results*

The data set of 1 600 galaxy cluster candidates in *The Hunt for Galaxy Clusters* contained 404 objects previously classified by experts.

Table 1 displays a comparison of the unweighted and weighted classifications of the Zooniverse volunteers (Section 2.3.3) based on the agreement with the experts. Figure 2.5 shows ROC curves computed for the whole crossmatch sample of 404 objects classified by both the Zooniverse volunteers and experts and the ROC computed on a subsample of 170 objects, 85 spectroscopically confirmed galaxy clusters and 85 objects classified as non-clusters by experts. This subsample is also used for the testing of the CNNs. The Zooniverse volunteers performed better on the subsample of 170 objects than on the whole crossmatch sample of 404 objects. This could be an indication of a bias towards correctly classifying easier objects since spectroscopically confirmed galaxy clusters tend to be larger.

Figure 2.6 shows the fraction of the Zooniverse volunteer's individual answers in agreement with experts to all Zooniverse answers for classification ending answers, except for *not a nearby cluster* and *not a distant cluster*, which do not have a direct counterpart in the classification of experts. Assuming that the expert classifications are the ground truth, the biggest difficulty for the volunteers seems to be distinguishing extended from point-like X-ray emission. Also, the volunteers inconsistently classified a large fraction of *no emission* classes, suggesting that they struggled to interpret the X-ray images. The huge discrepancy between volunteer's individual classifications and classifications of experts were in the *edge* category, used for galaxy cluster candidates close to the edge of *XMM-Newton's* chips and its field of view. Based upon discussions within the online forum, we assume that this bias could emerge from *XMM-Newton's* grid-like pattern created by small gaps between its individual detectors, which volunteers often mistaken for the edge of the chips. The *nearby galaxy* category was also a difficult question for the volunteers. Again based on the forum discussion we find that volunteers often classified nearby galaxy clusters with a prominent brightest central galaxy as a *nearby galaxy* class, which could lead

2 I made a typo in the original publication in the FPR equation in which I accidentally wrote TN to numerator instead of the correct FP as is written here

Table 1: The results of cluster classification by Zooniverse volunteers on two data sets, 404 objects are those classified by both, scientists and Zooniverse volunteers, the 170 objects data set is a subsample of the 404 objects, where 85 objects are spectroscopically confirmed clusters and 85 are objects classified as non-clusters by experts.

Data set	Zooniverse Classifications	TP	TN	FP	FN	accuracy	precision	recall
404 objects	unweighted	69	150	0	185	0.542	1.000	0.272
404 objects	weighted	102	149	1	152	0.621	0.990	0.401
170 objects	weighted	55	84	1	30	0.818	0.982	0.647

to many nearby galaxy clusters missed. In general, the Zooniverse volunteers preferentially classified objects as $non-clusters$.

Some of the biases could be mitigated in possible future versions of the project if explanations were clearer and more focus was put on example images in the help notes. Possibly the most important biases were often a classification of an X-ray emission as *no emission* and misclassification of an extended X-ray emission as a point-like X-ray emission. This are the main reasons why clusters were missed by the Zooniverse volunteers. We tried to keep in mind the possibility of low scientific knowledge of the volunteers and not to overwhelm the volunteers with huge amounts of information, which could discourage them, but we were still able to provide a detailed explanation of the X-ray emission in the tutorial and the help notes, with nice example images and diagrams to help with the X-ray contours. Small interviews with our beta testers revealed that around 20 % of them did not read the supporting texts. It might be possible that classifications with a lot of disagreement in the interpretation of the X-ray emission preferentially came from volunteers who did not adequately read the supporting material. A questionnaire would be needed to further probe this possibility. These biases could be cut down with simpler and shorter explanations of the X-ray properties, so it would be easier to understand and less information to digest. Another common tendency was the misclassification of nearby clusters that contain prominent BCGs (brightest cluster galaxies), with that of nearby galaxies. This could be reduced with a dedicated pair of images for the two situations in the help notes.

We have to note that even the classifications of experts could be biased towards *low-z* clusters, since we use DSS optical images, which are limited to $z \sim 0.3$.

Another possible bias may come from the fact that spectroscopically confirmed clusters are biased to big clusters, which might affect our interpretation.

To explore if the Zooniverse volunteers were biased finding preferentially most prominent galaxy clusters, we made extent – extension likelihood plane plots (see Appendix:Extent – extent likelihood plots). We found that the galaxy clusters found by the Zooniverse volunteers populate all of the space, not showing bias and their sample of galaxy clusters also can not be recreated by a simple cut in this space.

Even though the Zooniverse volunteers did not show a high accuracy compared to experts, misclassifying many galaxy clusters as other options, the sample of galaxy clusters they selected is pure. This makes us conclude that, via the Zooniverse project, the general public can help scientific research where a very pure sample of galaxy clusters is required, but it did not prove to be helpful in a case where a sample of galaxy clusters should be complete.

2.6.2 *CNN training*

We use two different data sets, one classified by experts and one by the Zooniverse volunteers. We use balanced training batches, containing the same number of classification classes, randomly sampled from the training data. This is to prevent the network from being biased towards the class that occurs most frequently in the training sample.

Regardless of the training data, all the networks were tested on the same data set of 85 spectroscopically confirmed galaxy clusters and 85 objects classified

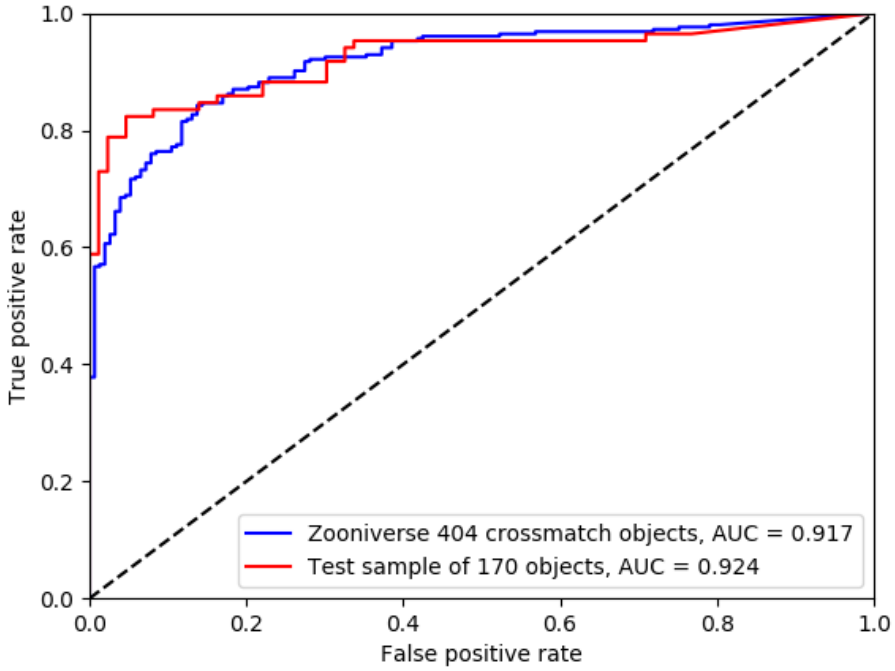

Figure 2.5: The receiver operating characteristic (ROC) curves for the classifications by Zooniverse volunteers, taking the classifications of experts as the ground truth. Closer the curve copies the left vertical and top horizontal axis, better the classifier. The dashed line shows how would the results be if the people guessed totally randomly.

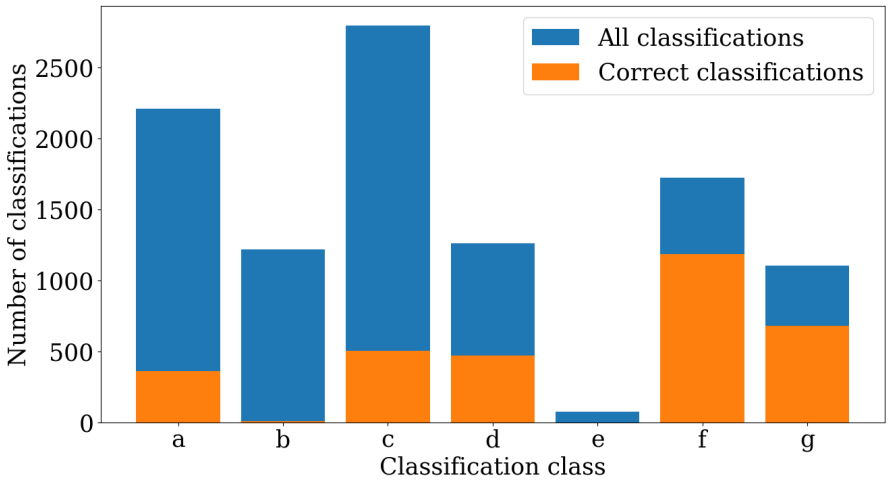

Figure 2.6: A quantification of the Zooniverse classifications for a) no emission, b) edge, c) point, d) nearby galaxy, e) no optical image, f) nearby galaxy cluster, g) distant galaxy cluster, assuming the ground truth is the expert classification.

Table 2: The number of objects in the training, validation and test data sets classified by Zooniverse and experts.

Class	Zooniverse		Experts		
	Train	Validate	Train	Validate	Test
cluster	320	130	845	200	85
non-cluster	880	100	388	104	85
total	1200	230	1233	304	170

as non-clusters by the experts, the 170 test objects. Table 2 and Figure 2.7 describe the numbers of objects used in the training, validation and test data sets, classified by experts and the Zooniverse volunteers for testing on the 170 object test sample. All the networks were trained on grayscaled and combined X-ray and optical images as described in Section 2.4.1 if not stated otherwise.

We experimented with both a custom network (Table 3) and using 3 different state of the art CNN architectures: VGG19 (Simonyan and Zisserman, 2014), InceptionV3 (Szegedy et al., 2015) and MobileNet (Howard et al., 2017). We used those networks with their pre-trained weights, using a large learning rate and unfreezing all the layers. Of the 3 models, MobileNet, pre-trained on the ImageNet (Deng et al., 2009), achieved the best performance and therefore we only discuss this architecture. Similarly, Lieu et al. (2018) found MobileNet to be the superior architecture for classifying solar system objects. The hyperparameters for our custom network and the MobileNet network are given in Table 4. We used Keras (Chollet et al., 2015) with TensorFlow (Abadi et al., 2015) backend. The `lr. red. patience` and `lr. red. factor` are parameters of the `ReduceLROnPlateau` Keras callback. The parameter `lr. red. patience` defines how many epochs without improvement of the validation accuracy (different proxy can be chosen to monitor) have to pass to change the current learning rate by multiplying it with the `lr. red. factor`.

The batches used to train the networks were randomly generated during training, always from the whole training sample. Validation started once a satisfying number of generated batches was presented to the network, this is the training data set size divided by the batch size. This was done to maximise the use of our data while keeping balanced numbers of classes in the yielded training batches, in order to avoid biasing the network.

2.6.3 *CNN results*

We demonstrate that convolutional neural networks are capable of high accuracy, automated galaxy cluster candidate classification. We trained each of our networks 10 times with the exact same hyperparameters, differing only in the seed for generation of random numbers during network's initialisation, the order of random image selection into balanced mini-batches during training and the random sampling of augmentation values applied during training but keeping the same objects in the training, validation and test data sets. The results of indi-

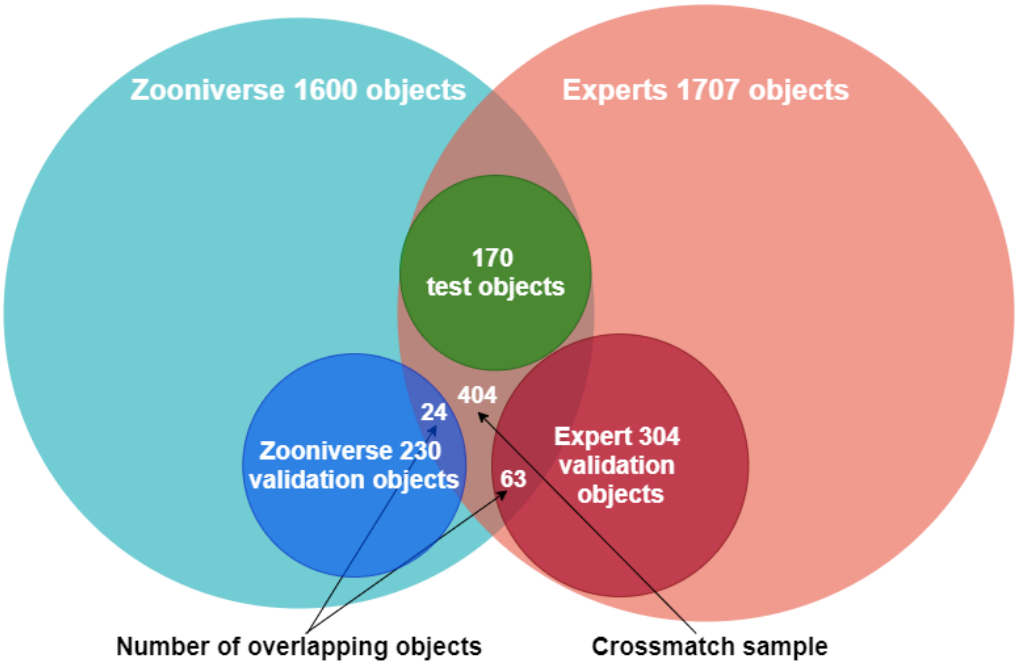

Figure 2.7: A Venn diagram presenting the data sets.

Table 3: The architecture of our custom network which achieved the best performance. Each of the convolutional and dense layers is followed by a ReLU non-linearity with the exception of the final output dense layer which has the softmax for classification.

Layer	Layer type	filter shape / stride	input shape
1	conv	3×3×64/(1, 1)	356×356×3
2	max pool	2×2/(2, 2)	356×356×64
3	conv	3×3×32/(1, 1)	178×178×64
4	max pool	2×2/(2, 2)	178×178×32
5	conv	3×3×32/(1, 1)	89×89×32
6	max pool	2×2/(2, 2)	89×89×32
7	conv	3×3×32/(1, 1)	45×45×32
8	max pool	2×2/(2, 2)	45×45×32
9	conv	3×3×32/(1, 1)	23×23×32
10	max pool	2×2/(2, 2)	23×23×32
11	conv	3×3×32/(1, 1)	12×12×32
12	max pool	2×2/(2, 2)	12×12×32
13	flatten	-	6×6×32
14	dense	256	1152
15	dense	2	256

Table 4: Hyperparameters of our custom network and the MobileNet network. The number of iterations, batches yielded during training, is shown for training on the data set classified by experts.

Hyperparameters	Custom net	MobileNet
Batch size	10	20
Iterations	153 000	3 825
Optimizer	SGD	Adadelta
Nest. Momentum	0.90	-
Rho	-	0.95
Initial lr.	0.0001	1.0
lr. decay	10^{-6}	0.95
Minimal lr.	10^{-4}	0.01
lr. red. patience	14	4
lr. red. factor	0.75	0.85
Dense dropout	0.65	0.65
Output activation	softmax	softmax
Loss function	cat. crossentropy	cat. crossentropy
Input image size	356×356	224×224

vidual runs are averaged and presented together with their standard deviations in Table 5 and Figure 2.8, helping to compare various networks.

To report accuracy (A), precision (P) and recall (R) in Table 5), we need to make hard classification decision for each example image from our test set. Our neural networks are trained to output the probability that the input image is a galaxy cluster. Therefore, we classify input images as galaxy cluster if this probability is higher than 0.5.

Our best-performing custom network (CN-E), trained on the expert classified data set, achieved an average accuracy of $(90 \pm 3)\,\%$. We also explored training on concatenated PNG images, without the grayscaling, so having six channels instead of three, but this did not change the performance significantly.

The MobileNet architecture trained on the data classified by experts achieved an average accuracy of $(88 \pm 2)\,\%$. Perhaps MobileNet has slightly different sensitivity for individual colour channels due to the potential bias in its original training sample. We explored this possibility by training it on two additional channel configurations, X-ray green, optical red, empty blue and X-ray red, optical green, empty blue, but its performance did not change significantly.

Training using the labels obtained in the Zooniverse project resulted in lower performance for both, our custom network (CN-Z) and the MobileNet (MN-Z), achieving average accuracies $(82 \pm 1)\,\%$ and $(79 \pm 2)\,\%$, respectively.

Lastly, we also explored the training of neural networks on single wavelength PNG images. Our custom network using expert labels trained only on the X-ray images without their optical counterparts (CN-E solo X-ray) achieved an average accuracy of $(81 \pm 1)\,\%$. Our custom network using expert labels trained

Table 5: Averaged galaxy cluster candidate classification results of the networks each trained 10 times with the exact same hyperparameters, only with a different seed for generation of random numbers during its initialisation.

network	A ± std	P ± std	R ± std	AUC ± std
CN-E	0.90 ± 0.03	0.89 ± 0.05	0.91 ± 0.03	0.96 ± 0.01
MN-E	0.88 ± 0.02	0.87 ± 0.03	0.91 ± 0.03	0.94 ± 0.01
CN-E solo optical	0.68 ± 0.02	0.64 ± 0.02	0.85 ± 0.04	0.77 ± 0.02
CN-E solo x-ray	0.81 ± 0.01	0.78 ± 0.03	0.86 ± 0.04	0.89 ± 0.01
CN-Z	0.82 ± 0.01	0.96 ± 0.01	0.67 ± 0.02	0.91 ± 0.01
MN-Z	0.79 ± 0.02	0.96 ± 0.03	0.62 ± 0.03	0.86 ± 0.02
CN-E no augm.	0.75 ± 0.02	0.70 ± 0.02	0.87 ± 0.03	0.87 ± 0.01
MN-E no augm.	0.81 ± 0.01	0.75 ± 0.02	0.91 ± 0.01	0.90 ± 0.02

only on the optical images (CN-E solo optical) performed the worse, achieving an accuracy of only $68 \pm 2)\,\%$. This is rather easily understandable knowing that the *XMM-Newton* data are much deeper than the POSS-II images used for the current analysis: while *XMM-Newton* can detect galaxy clusters as extended sources out to $z = 1$ at least, the POSS sensitivity strongly drops beyond $z \sim 0.3$ rendering galaxies are hardly identifiable.

Using augmentation (Section 2.4.2) was critical to achieving good performance, the accuracy of the network CN grayscale would drop from $(90 \pm 3)\,\%$ to $(75 \pm 2)\,\%$ without the augmentation and from $(88 \pm 2)\,\%$ to $(81 \pm 1)\,\%$ for MobileNet.

2.6.4 *Interpreting the results*

We further investigate the results of the best training run of our custom network (CN-E), which can classify even faint clusters and those close to the edge of *XMM-Newton*'s field of view. Figure 2.9 shows some of these randomly selected correctly classified galaxy clusters.

Figure 2.10 shows two objects classified as non-clusters by the experts, but as clusters by our custom network. The top object raised a concern that it was actually a galaxy cluster. We assume that it was classified as a galaxy cluster by our custom network because of the presence of the faint X-ray emission in the centre and that it is a promising candidate for further investigation and spectroscopic redshift confirmation. Figure 2.11 displays images of spectroscopically confirmed galaxy clusters which have been incorrectly classified by our custom network as a *non-cluster* class. The first object from the top is a non-centered galaxy cluster. The second contains a group of nearby galaxies with faint extended X-ray emission, which might have fooled our network. The third is a cluster that falls on a chip gap. The fourth is a galaxy cluster with three prominent nearby galaxies along the line of sight which is probably what fooled our network, and the last

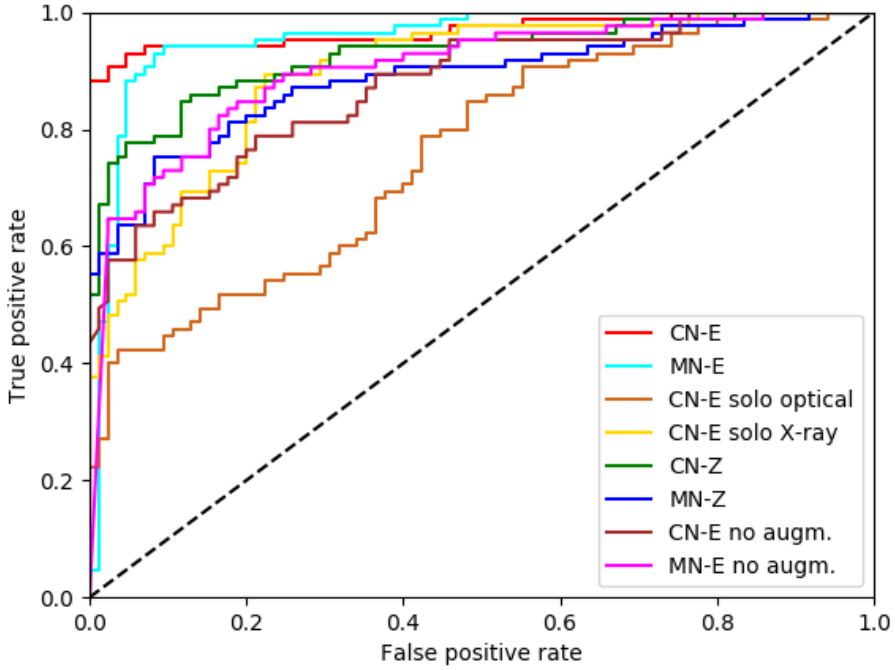

Figure 2.8: ROC curves for the best-performing networks when trained on different data formats. Closer the curve copies the left vertical and top horizontal axis, better the classifier. The dashed line represents how would an untrained, randomly guessing classifier score. Training on optical data only ended up with the poorest results, using only X-ray data achieved much better results, however, the combination of optical and X-ray data resulted in the best performance. CN refers to our custom network, MN to the MobileNet architecture, E to the data set classified by experts, Z to the data set classified by the Zooniverse volunteers.

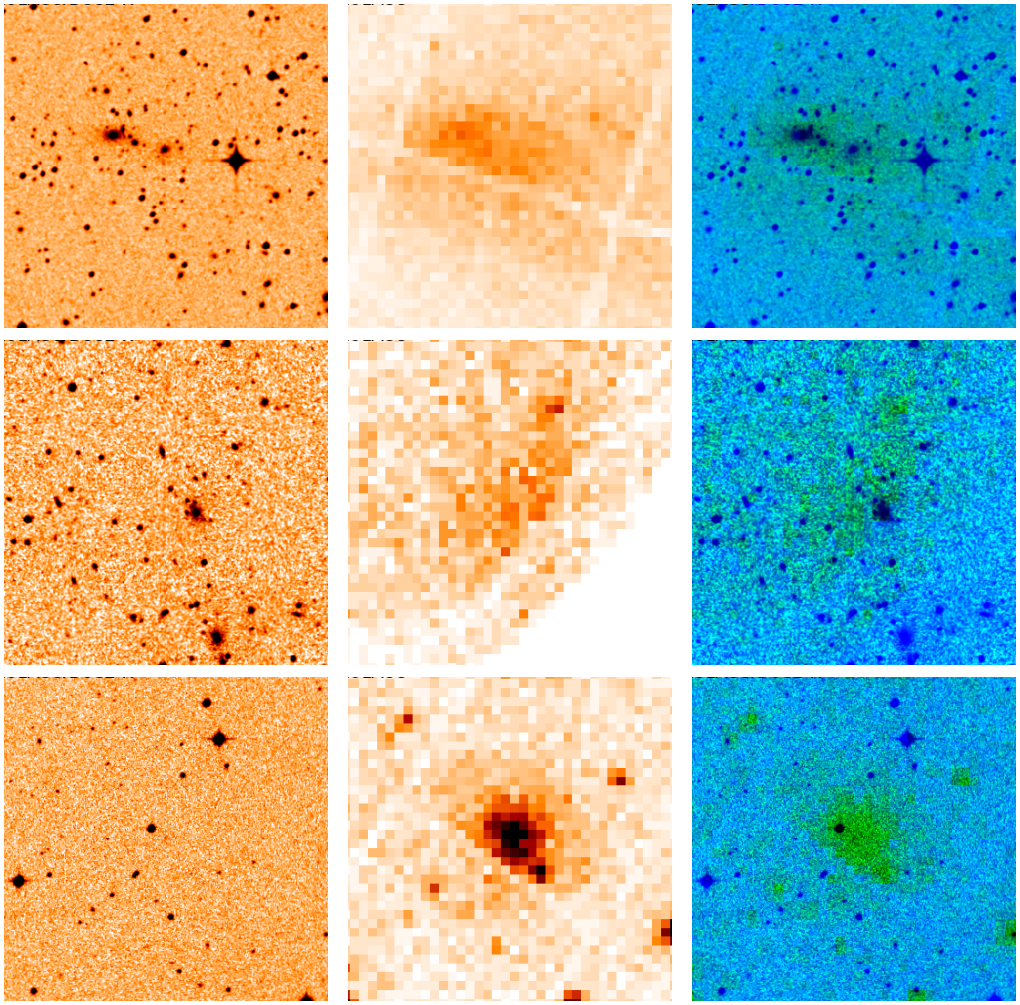

Figure 2.9: Spectroscopically confirmed galaxy clusters correctly classified by our custom network randomly selected from the test sample (TP). *Left:* optical, *middle:* X-ray, *right:* combined.

object appears like a nearby galaxy, which can be hard to classify even for the experts.

Figure 2.12 shows outputs of the selected filters of our custom network for a spectroscopically confirmed nearby galaxy cluster. We can see how the network learned to search for edges and colour patches of X-ray or optical light. Some filters learned to search primarily for X-ray emission and others for optical emission. Most of the filters detected both of the emission components simultaneously. Multiple filters in the same layer usually learned to search for X-ray emission, but their sensitivity is different. There are filters which get activated only by stronger emission, while other filters are more sensitive to X-ray emission. The network uses the filters to probe the presence and extent of the X-ray emission in the input image. Note that the filter output size decreases deeper within the network because of the max-pooling operation applied in the pooling layer after each convolutional layer.

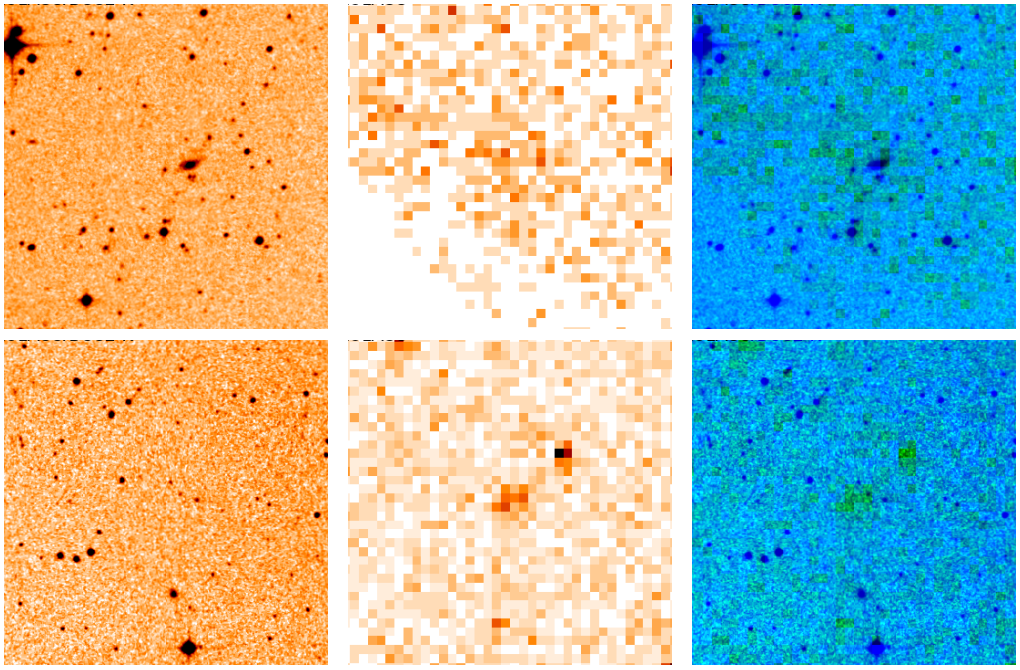

Figure 2.10: Non-galaxy clusters incorrectly classified as galaxy clusters (FP) by our custom network. *Left:* optical, *middle:* X-ray, *right:* combined.

Table 6: Results from the multi-class classification networks.

class	A	P	R	AUC
MN grayscale				
Low-z cluster	0.77	0.62	0.94	0.93
High-z cluster	0.87	0.56	0.22	0.91
Point source	0.87	0.88	0.36	0.89
Nearby galaxy	0.90	0.70	0.73	0.92
Other	0.91	0.65	0.68	0.92
CN grayscale				
Low-z cluster	0.79	0.68	0.81	0.89
High-z cluster	0.84	0.44	0.65	0.89
Point source	0.84	0.75	0.27	0.88
Nearby galaxy	0.89	0.74	0.57	0.85
Other	0.87	0.52	0.64	0.88

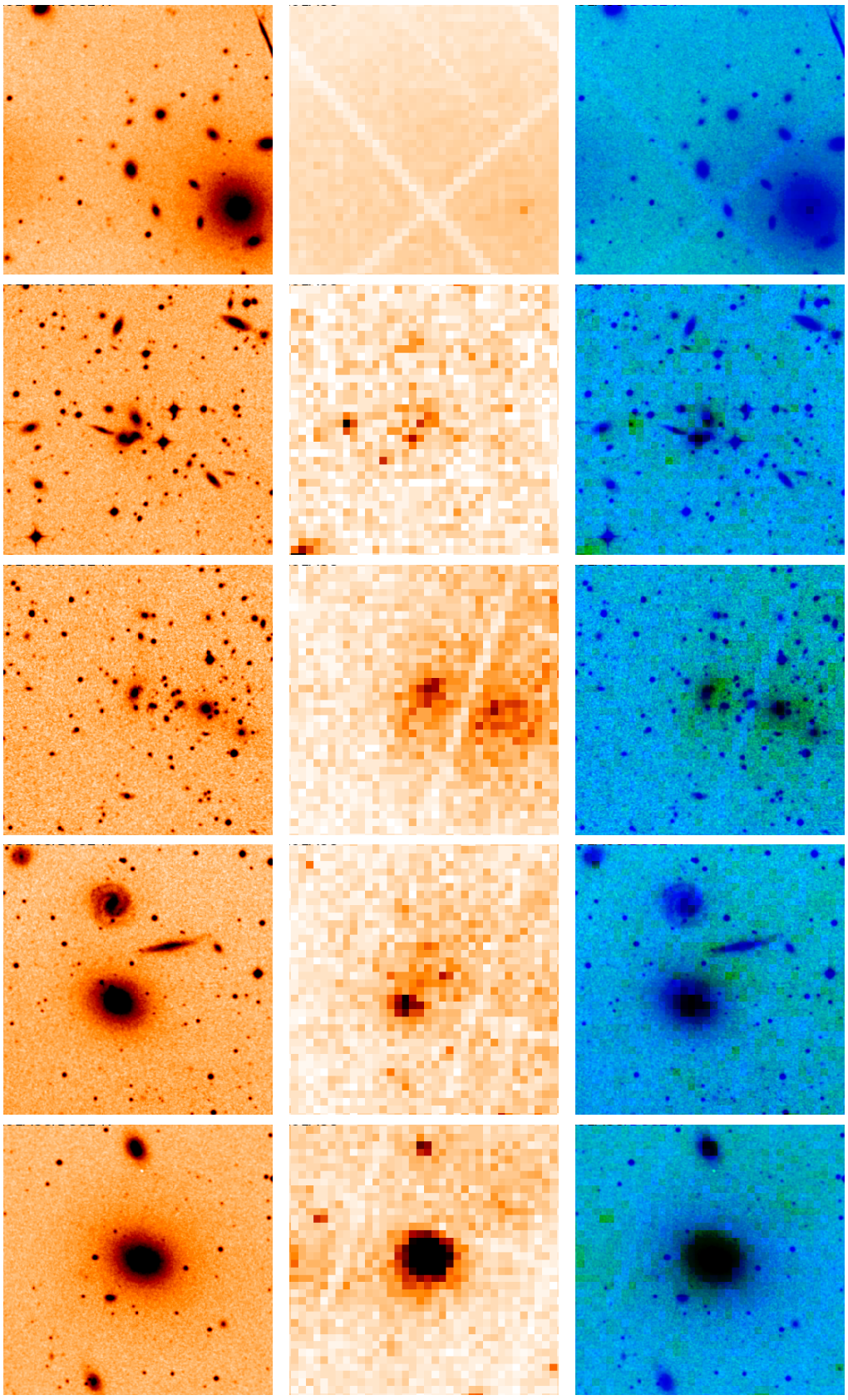

Figure 2.11: Galaxy clusters incorrectly classified as non-galaxy clusters (FN) by our custom network. *Left:* optical, *middle:* X-ray, *right:* combined.

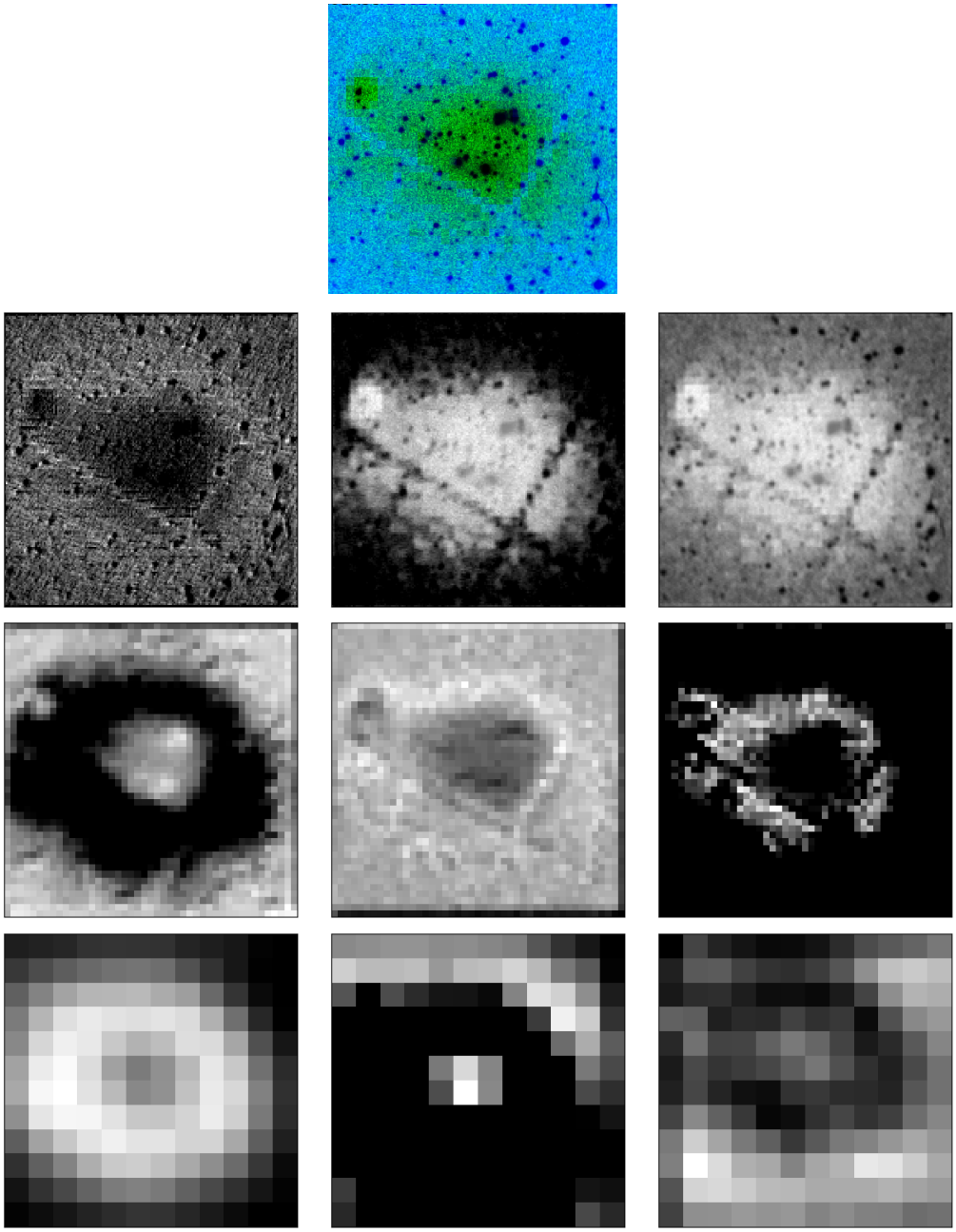

Figure 2.12: Top: Input image to the trained network. Each row from second to last shows outputs (activation maps) of 3 selected filters from 2nd, 4th and 6th convolutional layer of our custom network, respectively.

Table 7: The number of objects in the training, validation and test data sets in a single fold of the 10 fold cross-validation.

Class	Experts		
	Train	Validate	Test
cluster	904	113	113
non-cluster	399	57	114

2.6.5 *Multi-class classification*

We also trained neural networks for multi-class classification using the labels of the experts. We segregated objects into 5 classification classes - *low z cluster*, *high z cluster*, *nearby galaxy*, *point source* (point, star or AGN, double source) and *other* (artefact, edge). The ROC curves and performance measurements were calculated as one versus all problem.

In this regime, the MobileNet architecture and our custom network achieved an AUC and accuracy, averaged over all classes, within 1 sigma. The MobileNet achieved an AUC score of $(91 \pm 2)\,\%$ and accuracy of $(86 \pm 6)\,\%$, and our custom network obtained an AUC of $(88 \pm 2)\,\%$ and $(85 \pm 4)\,\%$ accuracy (Table 6).

In the case of multi-class classification problems, ROC and AUC are plotted for each of the classes separately as one versus all, reducing the problem to the binary case. From the ROC curves (Figure 2.13), we see that the *point source* and *high-z galaxy cluster* were the hardest classes to detect, and in the custom network, the *nearby galaxy* class was the easiest to distinguish. We interpret this as a consequence of nearby galaxies being very distinct from the other classes in the optical. Interestingly, this category did not achieve the best performance for the MobileNet network, however, it still placed among the top-performing classes.

We note that since we have trained the neural networks on a sample of galaxy cluster candidates picked by the `XAmin` pipeline, our sample of point sources is biased towards objects with some spatially extended emission. Thus we can not consider the networks trained for multi-class classification as a reliable point source classifiers since they are not representative of the population and do not reflect the typical appearance of an X-ray point source. If one would like to use our neural networks for point source detection, re-training or fine-tuning of our models on a representative sample of X-ray point sources would be required.

2.6.6 *Cross-validation*

We perform 10-fold cross-validation of CN-E to explore, if the test data set, having all of its galaxy clusters spectroscopically confirmed, shows significant bias compared to the galaxy cluster sample in the training data set. Table 7 contains the number of example images in each data set for a single fold of the cross-validation. The cross-validation accuracy scores between 87 % and 92 % (Table 8, Figure 2.14) and our CN-E achieved accuracy 90 % on average (Table 5,

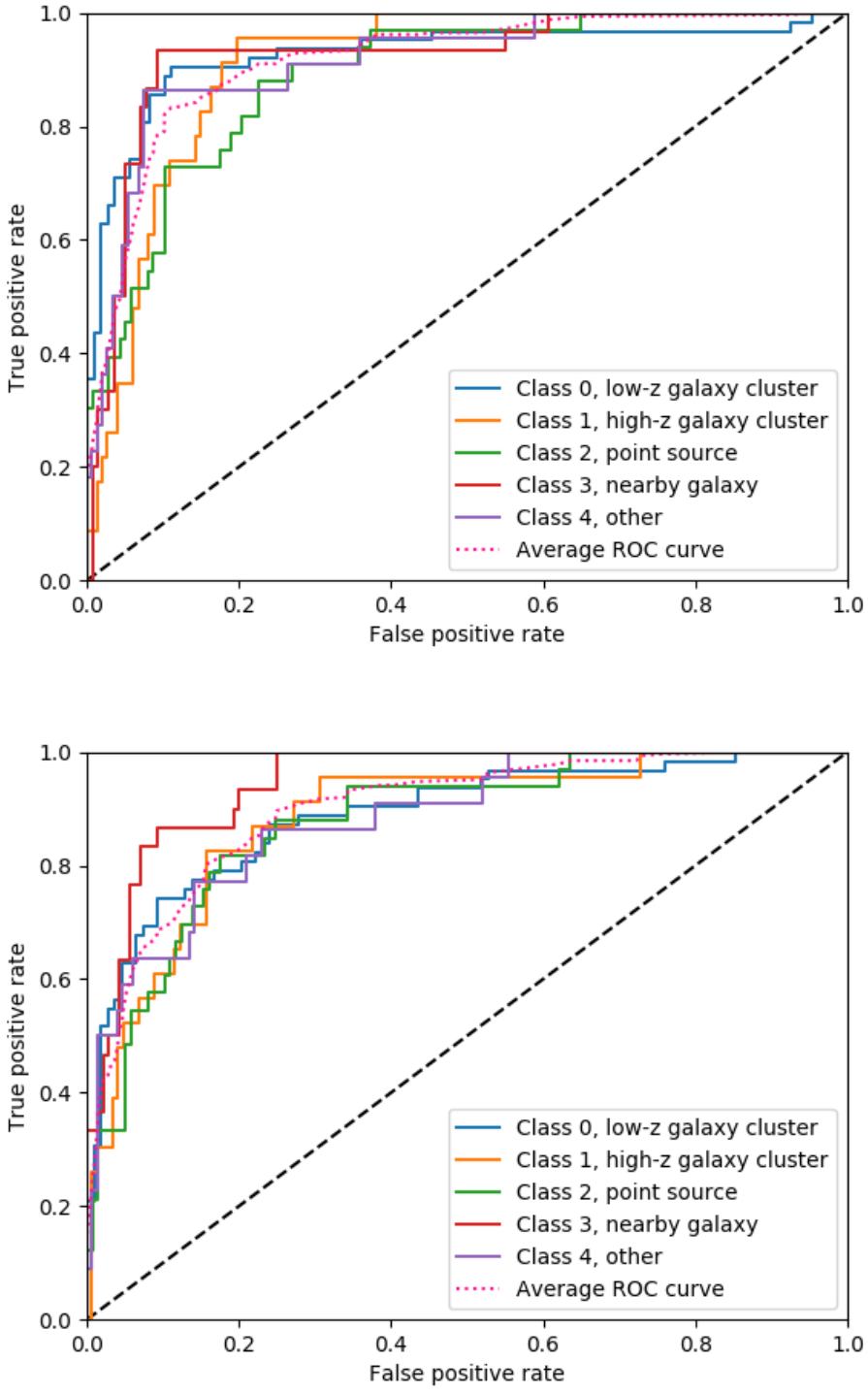

Figure 2.13: ROC curves for multi-class classification performed by the MobileNet architecture (*left*) and our custom network (*right*).

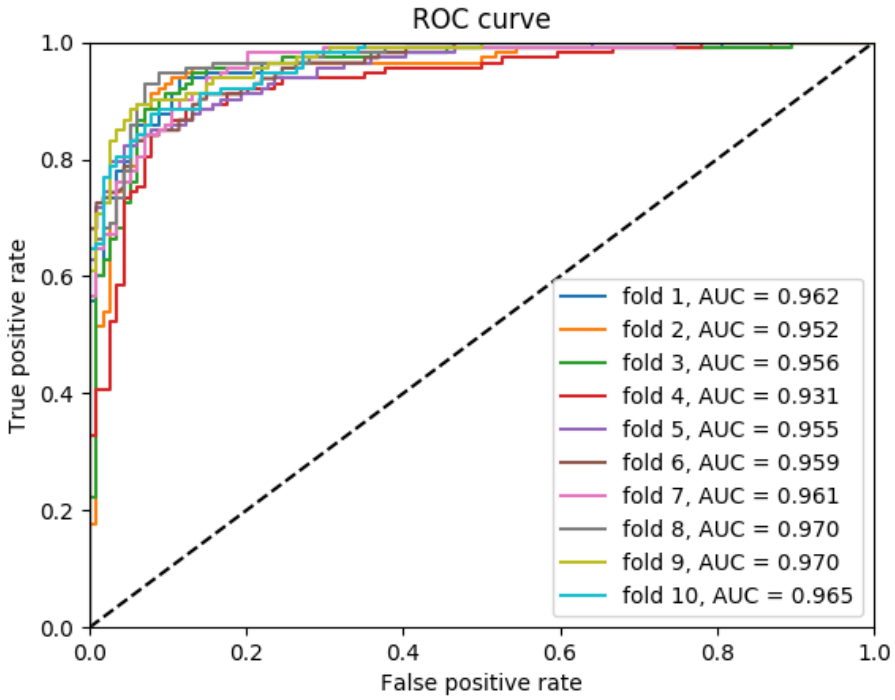

Figure 2.14: ROC curves for 10 fold cross-validation of our custom network trained on expert classifications.

Table 8: Classification results of our custom networks for a 10 fold cross-validation on classifications done by experts.

Fold	A	P	R
1	0.89	0.89	0.88
2	0.92	0.91	0.93
3	0.90	0.90	0.91
4	0.88	0.91	0.83
5	0.87	0.88	0.86
6	0.87	0.87	0.88
7	0.88	0.90	0.86
8	0.92	0.89	0.95
9	0.88	0.84	0.94
10	0.89	0.92	0.87

Figure 2.8). Those results are consistent and the test sample we used does not seem to have any significant bias on the network's performance.

2.7 SUMMARY

In this paper, we have presented convolutional neural networks to classify extended X-ray sources detected by the `XAmin` pipeline. This automated method can be used to replace the traditional manual screening confirmation task of the `XAmin` galaxy cluster candidates, which is often tedious and slow.

Firstly, we built a crowd-sourcing Zooniverse project - *The Hunt for Galaxy Clusters*, to obtain a classification of a large number (1 600) of galaxy cluster candidates in a short time frame (6 months). Our volunteers obtained 62% agreement with experts for identifying clusters and non-clusters in an overlapping sample of 404 objects. We found that the volunteers were often incorrectly classifying objects as point sources or no emission. Out of 254 objects classified as galaxy clusters by experts in the overlapping sample, volunteers agreed on 104 of those (66/146 *low-z* and 38/108 *high-z* galaxy clusters), which is only about 40 %, but they inconsistently classified only 1 non-cluster as a galaxy cluster. In total, the volunteers found 506 clusters from 1 600 candidates. We suspect the reason behind this low performance of the Zooniverse volunteers in *The Hunt for Galaxy Cluster*, if compared to e.g. Galaxy Zoo, to be the complexity of combined X-ray and optical data of galaxy cluster candidates, burdened by multiple projection and instrumental effects (see Section 2.6.1 for discussion of biases the Zooniverse volunteers exhibited). We also tested a hypothesis, that the Zooniverse volunteers would preferentially find prominent galaxy clusters and that their sample could be easily recreated by a cut in the extent – extension likelihood plane (Pacaud et al., 2006), however, the Zooniverse volunteers found galaxy clusters across the entire extent – extension likelihood space (Appendix:Extent – extent likelihood plots), pointing out that their help could be used for a galaxy cluster science.

Next, we trained CNNs on *XMM-Newton* X-ray images combined with their optical counterparts from DSS2, to distinguish galaxy clusters from non-clusters. The cross-validation of our custom network shows consistent results (Table 8, Figure 2.14) with accuracy scoring between 87 % and 92 %. We further developed networks on a fixed training, validation and test samples, the networks trained on Zooniverse classified data having a different training and validation samples than those trained on data classified by experts, but both having the same test sample. Our best network (CN-E) obtained an average accuracy of 90 % (Section 2.6.3). This network used our custom architecture and was trained on labels made by experts. The test sample of 170 objects is composed of 85 spectroscopically confirmed galaxy clusters (62 *low-z* and 23 *high-z*), and 85 galaxy cluster candidates classified as non-clusters by experts. For comparison, a similar network using the MobileNet architecture (MN-E) obtained an average accuracy of 88 % and using the custom architecture with the Zooniverse classifications (CN-Z) gave an average accuracy of 82 % at best.

In this work, we show that CNNs trained using either X-ray only or optical only images had significantly lower performance in reliably identifying galaxy clusters in comparison to using the combined data. While in the X-rays *XMM-Newton* detects galaxy clusters as extended sources to $z = 1$ at least, the optical

POSS-II data sensitivity strongly drops beyond $z \sim 0.3$, making galaxies hardly identifiable. This is evident from the high number of false-positive detections of galaxy clusters (low precision) using the optical only data. The X-ray only network achieved higher accuracy (81 %) than the optical only network (68 %).

Additionally we train our networks for multi-class classification using expert classified labels: *low-z galaxy cluster*, *high-z galaxy cluster*, *point source*, *nearby galaxy* and *other*. In this case, the MobileNet architecture performed slightly, but not significantly, better than our custom network (Table 5).

This project is a pilot study to determine the potential of CNNs for the detection of galaxy clusters. In the future, we intend to apply our methods to large sky surveys such as the new *eROSITA* or LSST and *Euclid*. Their enormous data sets are expected to contain tens of thousands of new galaxy clusters, which will require automated, fast and reliable methods to identify, as human screening of such large data volumes will be impossible. Our methods can also be applied to simulated data. Our custom network can be easily fine-tuned to, e.g., *eROSITA* simulations and deliver an automated search tool for galaxy clusters from X-ray images. Applying our CNN on simulations will also enable modelling of the cluster selection function, important for cosmological studies, which cannot be done with clusters selected by human inspection due to their inconsistent biases.

Acknowledgements

We would like to thank the referee Dr Florence Durret for the valuable comments that helped to improve the paper.

We would like to acknowledge the scientists from the X-CLASS collaboration who manually classified the `XAmin` galaxy cluster candidates, mainly Jean-Baptiste Melin, one of the moderators overseeing all classifications and Edoardo Cucchetti. We are also very thankful to all of our citizen volunteers who participated in *The Hunt for Galaxy Clusters*. We use data generated via the `http://zooniverse.org` platform, development of which is funded by generous support, including a Global Impact Award from Google, and by a grant from the Alfred P. Sloan Foundation. The Digitised Sky Surveys were produced at the Space Telescope Science Institute under U.S. Government grant NAG W-2166. The images of these surveys are based on photographic data obtained using the Oschin Schmidt Telescope on Palomar Mountain and the UK Schmidt Telescope. The plates were processed into the present compressed digital form with the permission of these institutions. The Second Palomar Observatory Sky Survey (POSS-II) was made by the California Institute of Technology with funds from the National Science Foundation, the National Geographic Society, the Sloan Foundation, the Samuel Oschin Foundation, and the Eastman Kodak Corporation.

We implement our machine learning codes using Keras (Chollet et al., 2015) with a TensorFlow (Abadi et al., 2015) backend, and data augmentation using scikit-learn (Pedregosa et al., 2011). We also used Numpy (Oliphant, 2006), Matplotlib (Hunter, 2007) and Astropy (Astropy Collaboration et al., 2013, Astropy

Table 9: Threshold values used by the `XAmin` pipeline, std and median are the standard deviation and the median of the image data.

	X-ray	Optical
min cut	0	median - std
max cut	median $\times$ 14	median $+ 5 \times$ std

Table 10: PNG image channel values as constructed by the `XAmin` pipeline. *pix* refers to the pixel value after cutting.

Channel	Pixel value	Normalised pixel value
R	pix >= 176	255
	pix < 176	pix $\times$ 255 / 176
G	pix >= 120	(pix - 120) $\times$ 255 / (255 - 120)
	pix. < 120	0
B	pix >= 190	(pix - 190) $\times$ 255 / (255 - 190)
	pix < 190	0

Collaboration et al., 2018) Python3 (Van Rossum and Drake, 2009a) packages. Our codes are open source[3].

Matej Kosiba is supported by the *European Space Agency* traineeship and the ERASMUS program and travel funding from the ESAC science faculty, Maggie Lieu was supported by *European Space Agency* research fellowship at the European Space Astronomy Centre and a research fellowship at the University of Nottingham.

2.8 APPENDIX: IMAGE PREPROCESSING

The output of the `XAmin` pipeline is an image with the following normalisation: if a pixel value is lower than min cut, it is attributed a value of 255; if a pixel is greater than max cut it is attributed a value of 0; and 255 $\times$ (1 - (data - min cut) / (max cut - min cut)) otherwise Table 9. To produce the .png images used in the neural networks, `XAmin` applies the normalisation separately to each of the channels according to Table 10.

2.9 APPENDIX: EXTENT – EXTENSION LIKELIHOOD PLANE PLOTS

The extent – extension likelihood plane plots (Figure 2.15, Figure 2.16) of our C1 sample of galaxy cluster candidates as described in (Pacaud et al., 2006), were used to analyse the Zooniverse sample of galaxy clusters and investigate our initial hypothesis, that the Zooniverse volunteers will preferentially find most

3 https://github.com/matej-kosiba/CNN-multiwavelength-classification-of-X-ray-selected-galaxy-cluster-candidates

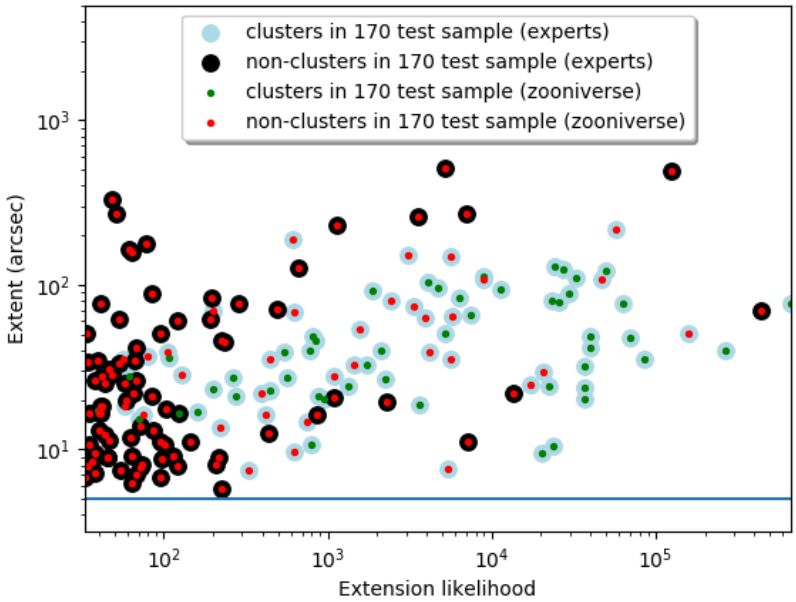

Figure 2.15: Extent – extension likelihood plane for objects of the 170 test sample classified by experts and the Zooniverse volunteers.

prominent galaxy clusters. We find that the sample of the Zooniverse galaxy clusters span the entire extent – extension likelihood plane and can not be recreated by a simple cut in this space. Please note however that the `XAmin` v3.5 we used to make the C1 sample had an issue fitting the point source peak, resulting in many non-clusters in the C1 region on the plots and that it is not the same pipeline as the XXL collaboration used before.

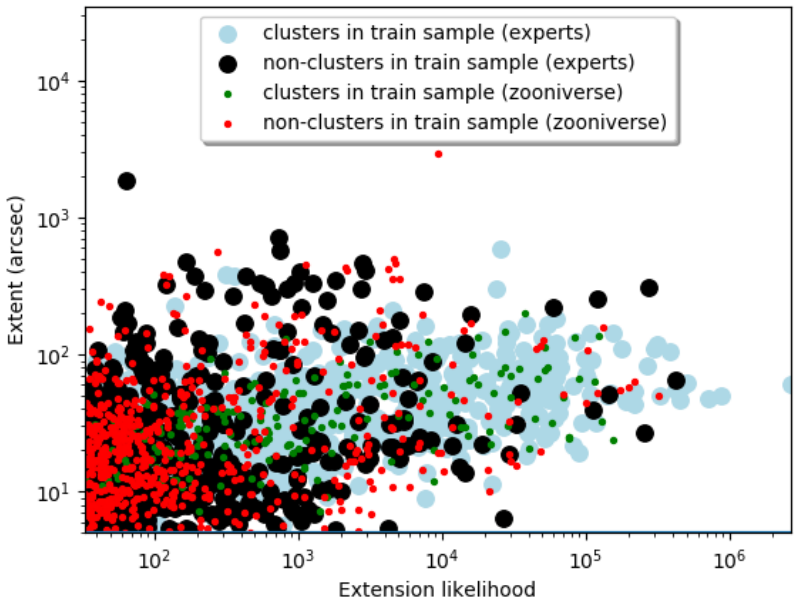

Figure 2.16: Extent – extension likelihood plane for objects of the experts train sample and the Zooniverse train sample.

3

The cosmological analysis of X-ray cluster surveys: VI. Inference based on analytically simulated observable diagrams

M. Kosiba,[1,2] *N. Cerardi,*[3] *M. Pierre,*[4] *F. Lanusse,*[4] *C. Garrel,*[5] *N. Werner,*[1] *T. Charnock,*[3] *and M. Shalak*[3]

[1]Department of Theoretical Physics and Astrophysics, Faculty of Science, Masaryk University, Kotlářská 2, Brno, 611 37, Czech
[2]Dipartimento di Fisica, Università degli Studi di Torino, via Pietro Giuria 1, I-10125 Torino, Italy
[3]Université Paris Cité, Université Paris-Saclay, CEA, CNRS, AIM, F-91191, Gif-sur-Yvette, France
[4]Université Paris-Saclay, Université Paris Cité, CEA, CNRS, AIM, 91191, Gif-sur-Yvette, France
[5]Max Planck Institute for Extraterrestrial Physics, Giessenbachstrasse 1, D-85748 Garching, Germany

Submitted to Astronomy & Astrophysics, April, 2024

3.1 INTRODUCTION

Galaxy clusters are the largest virialized systems in the Universe, sitting on the peaks of the primordial density fluctuation field (e.g. Bahcall, 1988). The mass distribution and number density of galaxy clusters across space and redshift are governed by the cosmological parameters of the Universe, making them potentially powerful cosmological probes e.g. Eke et al., 1996. Clusters in X-rays are less prone to projection effects than galaxy over-densities in the optical (e.g. Bhatiani et al., 2022) and provide key physical information on the intra-cluster medium (ICM), which is a good tracer of the underlying dark matter distribution (e.g. Fischer et al., 2023; Holland et al., 2015).

However, cluster cosmology is usually considered a delicate enterprise, given that the cluster mass, a key parameter that enters the cosmological analysis, is not a directly observable parameter. Rather than using mass proxies, like the X-ray temperature, luminosity, or gas mass, our approach is to forward model a set of observed properties. More specifically, we consider the distribution of the count rates (CR) and hardness ratios (HR) observed by the XMM-Newton satellite in several redshift bins (X-ray Observable Diagrams, XOD). The main idea is that the CR and HR values are analogous quantities to the X-ray fluxes and the temperatures of the galaxy clusters. Together with the redshift information, these quantities thus carry both physical and cosmological information. This method, ASpiX, has proven an efficient way of performing cluster cosmological

analyses (Clerc et al., 2012b). It enables easy modelling of the cluster selection function and has the merit of cleaning several potential systematic uncertainties that are otherwise cosmology-dependent. Moreover, it enables the inclusion of all detected clusters in the analysis because HR measurements require much fewer photons than measuring a temperature.

To date, the method has been tested on simulations (Pierre et al., 2017; Valotti et al., 2018) and used to extract the cosmological constraints from the XMM-XXL cluster sample (Garrel et al., 2022, Garrel et al. (2024), submitted). All these applications used a Markov Chain Monte Carlo (MCMC) algorithm to infer the cosmological parameters and related uncertainties. Moreover, cosmology-dependent priors derived from the same cluster sample were imposed on the cluster scaling relations, linking the mass with the cluster luminosity, temperature, size, and associated scatter values.

The present paper is the sixth of the series investigating the potential of the ASpiX concept. Our goal here is to use a deep learning approach to explore to what extent we can infer cosmology without any prior knowledge of the cluster evolutionary physics, i.e., on the coefficients of the scaling relations. We start by creating XODs for a wide range of Ω_m and σ_8 values, assuming random values for the scaling relation coefficients, which will be handled as nuisance parameters: this allows us to produce an unbiased simulated XOD sample. We then train a sequential neural posterior estimation (SNPE) (Papamakarios and Murray, 2016) network, a novel machine learning (ML)-based technique of the simulation-based inference family (SBI), to infer the cosmological parameters Ω_m and σ_8 from the XODs. We consider several survey sizes independently to estimate the effect of the shot noise on cluster counts, thus on the representation of the XODs and subsequently on the inferred cosmological parameters.

This paper is structured as follows: Section 3.2 describes the formalism used to create the XOD sample. Section 3.3 describes the SNPE method. We present our main results in Section 3.4. In Section 3.5, we discuss the impact of various factors on our cosmological predictions: the size of the training set for the neural networks, the effect of imposing a prior on the cluster redshift distribution, and of decreasing the number of free parameters by assuming self-similar evolution. We discuss here also what steps would need to be taken when using this method on a real observed XOD of our Universe. We conclude in Section 3.6.

In this paper, we shall consider uniform XMM-Newton extragalactic surveys of various sizes, all performed with an exposure time of 10 ks at any sky position. We take WMAP9 (Hinshaw et al., 2013) as our fiducial cosmology.

3.2 CREATING THE SET OF X-RAY OBSERVABLE DIAGRAMS

Given an XMM-Newton extragalactic survey area, size, and exposure time, the XOD is a summary statistic of the entire detected cluster population. Namely, a 3D representation of the HR, CR, and z observed parameters. CR is the total (MOS1+MOS2+pn) count rate in the [0.5-2] keV band, HR is the ratio of the number of counts in the [1.0-2.0]/[0.5-1.0] bands; while measurement errors can be attributed to CR and HR, we consider that spectroscopic redshifts are available with negligible uncertainty.

3.2.1 *Cosmological model and galaxy cluster mass function*

To simulate the X-ray cluster population, we assume a flat ΛCDM cosmology ($w = -1$) and the Tinker et al., 2008 halo mass function (HMF). This provides us with the number of galaxy clusters (dn) formed per unit mass (dM) per redshift bin (dz) per unit solid angle (dΩ). The ranges of considered Ω_m and σ_8 values are given in Table 11.

3.2.2 *Scaling relation formalism*

The next step is to attribute to each halo an X-ray luminosity and a temperature; this is done through scaling relations. We adopt the formalism described by Pacaud et al., 2018:

$$\frac{M_{500}}{M_0 \times 10^{13} M_\odot h^{-1}} = \left(\frac{T_{300kpc}}{1keV} \right)^{\alpha_{MT}} E(z)^{\gamma_{MT}} \tag{3.1}$$

$$\frac{L_{500}}{L_0 \times 10^{41} ergs^{-1}} = \left(\frac{T_{300kpc}}{1keV} \right)^{\alpha_{LT}} E(z)^{\gamma_{LT}} \tag{3.2}$$

where L_{500} is the rest-frame luminosity in the [0.5-2] keV band, within R_{500}; $T_{300\mathrm{kpc}}$ stands for a generic measure of the temperature within a radius of 300 kpc. The ranges adopted for the six free parameters (normalization, slope, evolution) are given in Tab. 11. Following Pacaud et al., 2018, we encapsulate all the scatter in the L-T relation, i.e. $\sigma_{LT} = 0.67$, and $\sigma_{MT} = 0$.

Table 11: Sampling range for parameters used to simulate the CR-HR-z diagrams. The first two are the cosmological parameters of interest for the present study; the latter six define the scaling relations. For all sets of simulated XOD, the eight parameters are drawn at random, assuming a uniform probability within the given ranges. The adopted ranges are ±50% of the fiducial values; except for γ_{LT} for which the current uncertainties are much larger.

coefficients	central value	min value	max value
Ω_m	0.279	0.1395	0.4185
σ_8	0.821	0.4195	1.2315
M_0	2.6	1.3	3.9
α_{MT}	1.67	0.835	2.505
γ_{MT}	-1.0	-1.5	-0.5
L_0	8.24	4.12	12.36
α_{LT}	3.17	1.585	4.755
γ_{LT}	0.47	0	1.2

3.2.3 *X-ray observable properties*

After each (M, z) halo is assigned a luminosity and a temperature, we derive the corresponding observed total XMM-Newton count rates in the three [0.5-2],

[1-2], and [0.5-1] keV bands of interest; this is done using the APEC plasma model folded with the XMM-Newton responses for the three EPIC detectors as described e.g. in Garrel et al., 2022.

3.2.4 *XMM-Newton survey design and selection function*

In the present study, we consider several (hypothetic) survey sizes ranging from 1 000 deg^2 to infinity (see Sec. 3.2.6 and Sec. 3.5.3). This allows us to study the impact of the shot noise in the XODs: the larger the survey, the larger the number of clusters in the dCR/dHR/dz bins, hence the smaller the statistical fluctuations in each bin.

To simulate the survey selection function, we assume a simple CR cut. This allows us to significantly simplify the modeling of the selection process, which in reality depends on both the CR and the cluster apparent size (Pacaud et al., 2006). This hypothesis has no incidence for the purpose of the present study. We set the cut to CR = 0.02 c/s, which corresponds to $\sim$ 200 counts for a 10 ks XMM-Newton exposure; this is a safe limit inspired by the XXL survey statistics (Adami et al., 2018) and yields for the fiducial cosmology and the Pacaud et al., 2018 scaling relations around 4 clusters / deg^2.

3.2.5 *Construction of the XOD sample and the dn/dz selection*

The CR-HR-z XODs are defined as 64×64×10 arrays for the CR, HR, and z dimensions, respectively (Fig. 3.1) and are later compressed to a 16×16×10 resolution used for the SNPE ML model (Fig. 3.2). To create an XOD, firstly, a set of two cosmological and six scaling relation parameters are drawn from the allowed parameter space (Tab. 11) based on random-uniform prior distributions. We then compute the number of halos in the survey volume for a 1 000 deg^2 area, derive the CR for each halo, and apply the CR selection, to include or not the halo in our simulated detected sample. Because the analytical derivation of the dn/dM/dz distribution is computer-time demanding, we apply an intermediate test to exclude a priori unphysical combinations of the eight free parameters - this has a high probability of occurring given that the parameters are drawn independently at random. For this, we use the fiducial XOD based on the observed cluster's distribution in our Universe. We calculate the number of clusters that pass the CR cut in the second redshift bin (i.e. the most populated one). If this number is incompatible within 3σ with the real-life value (i.e. that of our fiducial diagram), the combination is discarded and we consider another set of eight parameters. If this number is compatible with the observed cluster counts, we compute the number of clusters that pass the CR cut in the fourth redshift bin; we perform the same comparison as for the second bin and, similarly, decide whether to keep the set of eight parameters. We perform this test every two redshift bins to validate a parameter combination as plausible quickly. We stop when we have identified 70 000 combinations of the eight parameters that verify the dn/dz constraint for every two bins over the full redshift range. We find that this only occurs for 1.7% of the random combinations. The entire process takes about 12 hours to simulate, running in parallel on $\sim$110 cores. The sample size of 70 000 was arbitrarily chosen and later proved to be a sufficiently large number

to ensure good performance (Sec. 3.5.1). Fig. 3.3 (left) shows the distribution of the Ω_m and σ_8 values of the 70 000 $\mathrm{d}n/\mathrm{d}z$ accepted XODs.

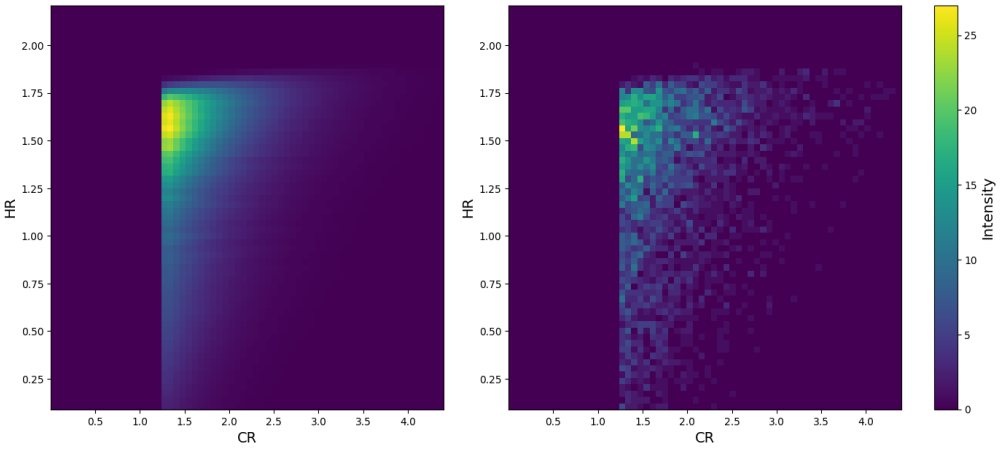

Figure 3.1: Left: Fiducial XOD for a 1 000 deg^2 survey area (simulated for central values of Table 11), integrated over the redshift dimension. The XOD has a 64×64×10 resolution. Right: a particular Poisson realization of the left-hand-side image.

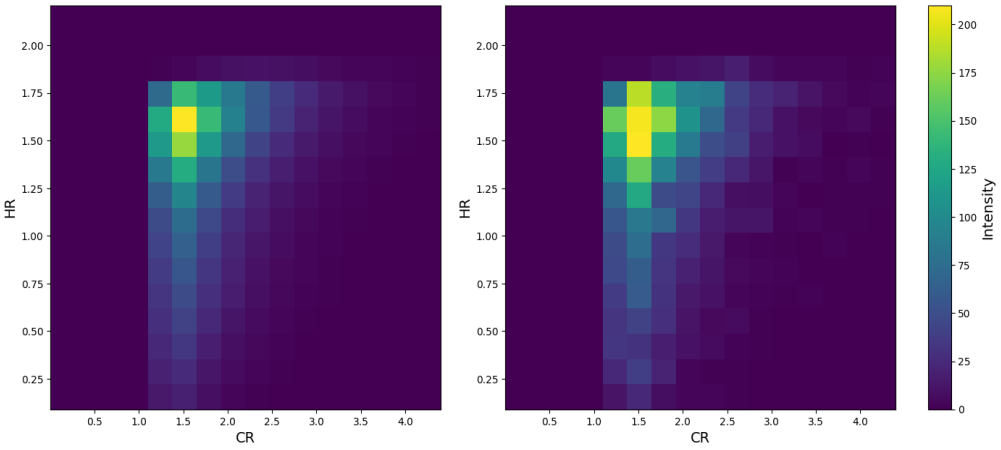

Figure 3.2: Left: Fiducial XOD for a 1 000 deg^2 survey area (simulated for central values of Table 11), integrated over the redshift dimension. The XOD has been downsampled to a 16×16×10 resolution, the shape used for the ML model. Right: a particular Poisson realization of the left-hand-side image.

3.2.6 *Modelling the impact of survey area and of stochastic noise*

In average, the number of detected clusters scales proportionally to the survey area, but the true number of clusters in each XOD pixel is affected by shot noise. (In all this study, we neglect the effect of sample variance on the cluster number counts). We describe below, how noise is implemented in the diagrams.

The XODs are first simulated for a 1 000 deg^2 survey area with a 64×64×10 resolution, for a given set of cosmological + cluster physics parameters. The $\mathrm{d}n/\mathrm{d}z$ 3σ test is applied on this realisation. If the XOD passes the test, the pixel

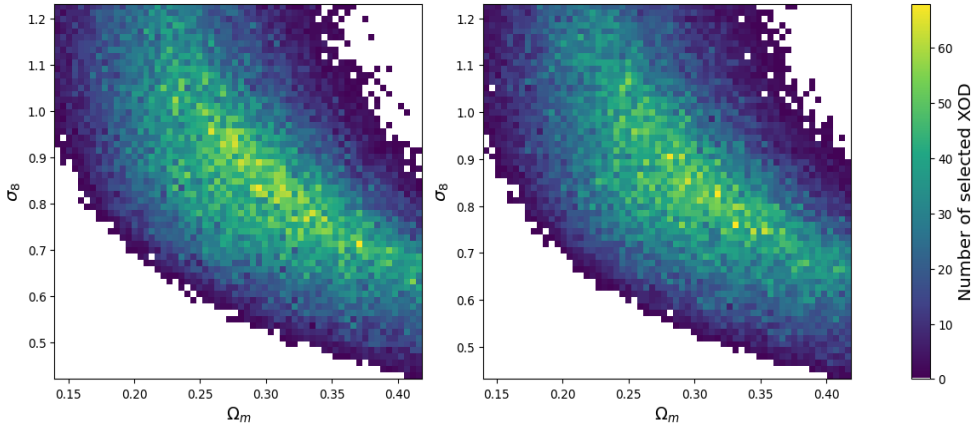

Figure 3.3: Comparison of the Ω_m and σ_8 distribution for XODs from the D1 and D3 datasets Left: D1 with 8 free parameters. Right: D3 with 6 free parameters, having fixed cluster evolution to be self-similar. Both datasets verify the $\mathrm{d}n/\mathrm{d}z$ test. Each dataset contains 70 000 XODs. The white zones show the regions excluded by each selection. Fixing cluster evolution does not significantly restrict the a priori range of possible cosmological values.

values are multiplied 10, to obtain the XOD for 10 000 deg^2. After that, the XODs are downsampled to the 16×16×10 resolution used for the ML model and we apply a Poisson noise model on every XOD pixel with the Numpy (Harris et al., 2020) Python (Van Rossum and Drake, 2009b) library. Fig. 3.2 shows an XOD before noising (left) and after the noise (right) at a 16×16×10 resolution.

In addition, for the purpose of investigating the limits and possible biases of our modelling, we consider two other - unrealistic - survey sizes : 100 000 deg^2 and 'infinite'. The infinite realisation is not affected by noise, while the noising of 100 000 deg^2 follows that of 10 000. We note that since the $\mathrm{d}n/\mathrm{d}z$ test is applied on the 1 000 realisation only, it is more permissive for the parameter sets compatible with the 10 000 and 100 000 surveys.

3.2.7 *Datasets*

In addition to the main dataset described previously, we also created three other datasets of XODs (Tab. 12) to examine how the $\mathrm{d}n/\mathrm{d}z$ selection and how decreasing the number of free parameters by fixing cluster evolution parameters at fiducial values impacts the final accuracy of the cosmological predictions.

The main dataset for this work is the 70 000 $\mathrm{d}n/\mathrm{d}z$ accepted XODs created by the steps described previously (D1 dataset). The distribution of its parameters is shown in Fig. 3.4. The distribution of the parameter values in the dataset shows some forbidden regions in the parameter space, because of the $\mathrm{d}n/\mathrm{d}z$ selection.

The second XOD dataset inhibits the $\mathrm{d}n/\mathrm{d}z$ subselection (D2 dataset). We do not show the corresponding parameter distribution as it is just the input random uniform one.

The third XOD dataset uses the $\mathrm{d}n/\mathrm{d}z$ subselection and, in addition, assumes that cluster evolution is self-similar (i.e. γ_{LT} and γ_{MT} are fixed at their central fiducial values). this is the D3 dataset (Tab. 11). Fig. 3.5 shows the distribution of this dataset's $\mathrm{d}n/\mathrm{d}z$ accepted XOD parameters.

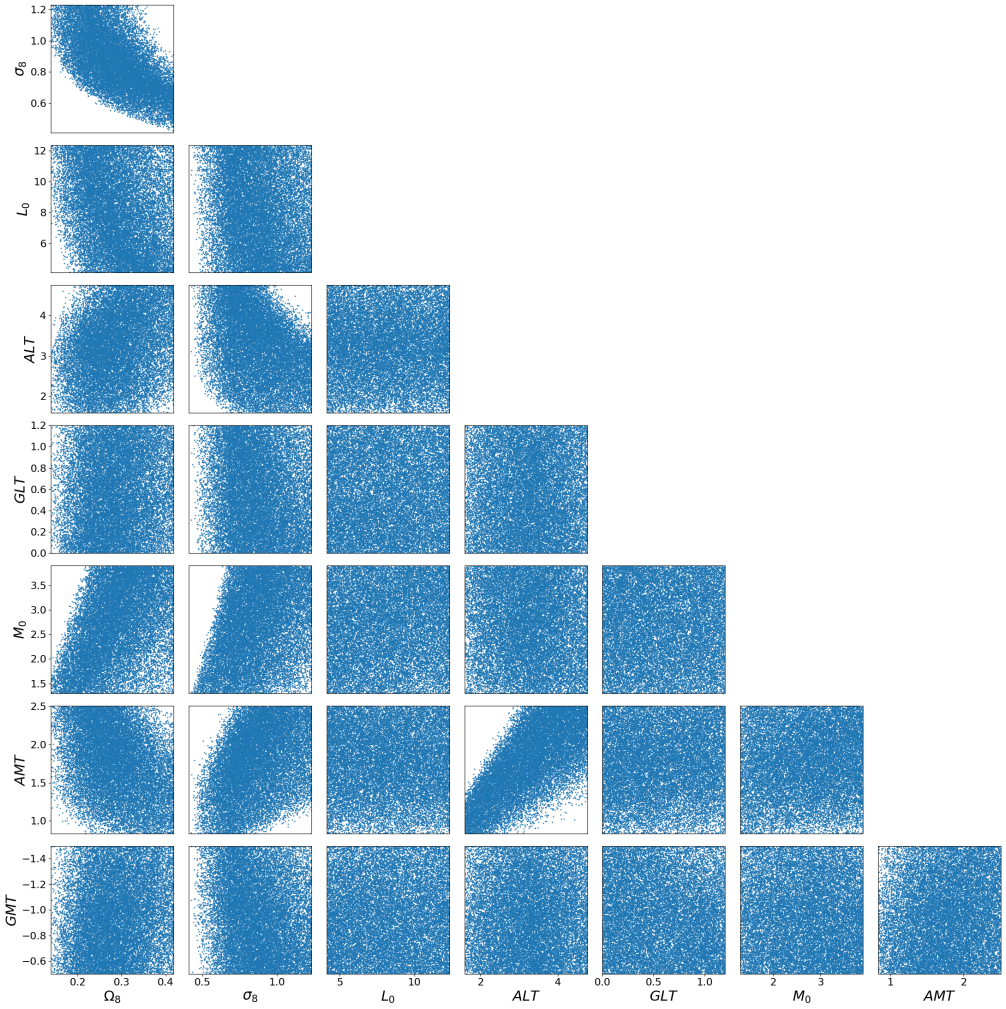

Figure 3.4: Prior distribution of the two cosmological and six scaling relations parameters after $\mathrm{d}n/\mathrm{d}z$ subselection for D1 dataset. The blue dots represent parameters of XODs that were accepted by the $\mathrm{d}n/\mathrm{d}z$ selection on this dataset. We can see that the $\mathrm{d}n/\mathrm{d}z$ selection flags forbidden regions compared to the initial random uniform selection from this parameter space.

The last data set (D4) inhibits the $\mathrm{d}n/\mathrm{d}z$ subselection and assumes cluster self-similar evolution.

Thanks to these four data sets, we can investigate the effect of restricting the possible ranges of XOD, by applying well-justified constraints on the cluster scaling relation coefficients, otherwis totally random : (i) the XOD must verify the the observed detected-cluster redshifty distribution and (ii) clusters evolve self-similarly.

The datasets are listed in Tab. 12. They all contain 70 000 XODs.

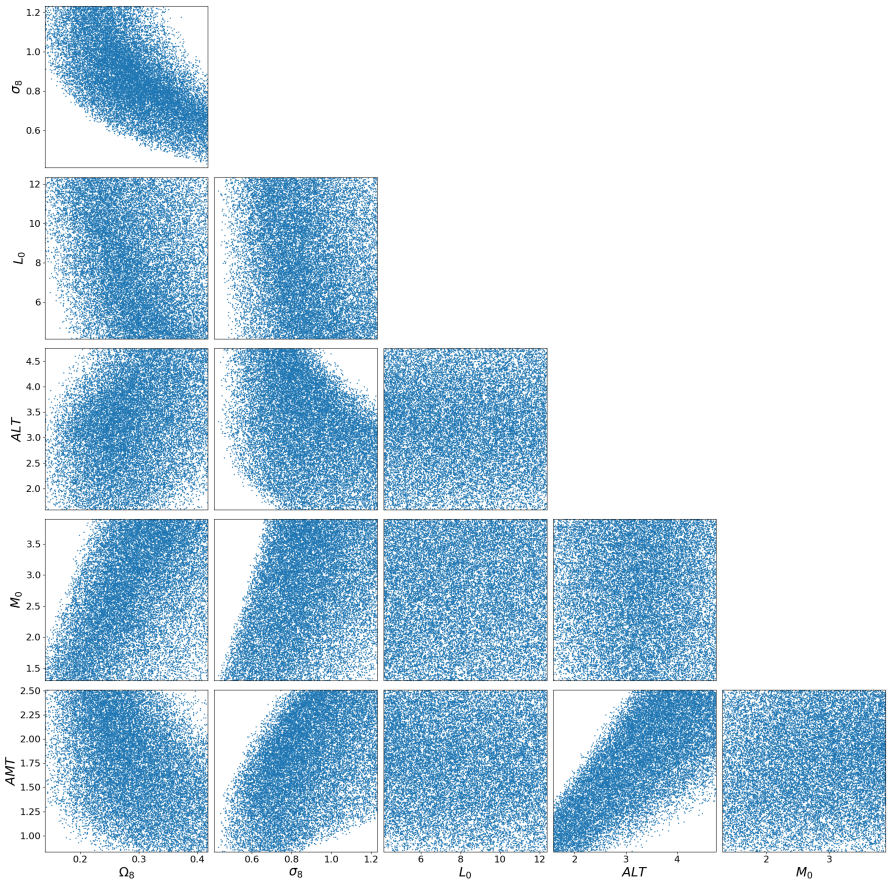

Figure 3.5: Prior distribution having two cosmological and four scaling relations parameters after $\mathrm{d}n/\mathrm{d}z$ subselection for the ResNet4 training dataset with fixed γ_{MT} and γ_{LT} scaling relation parameters (D3 dataset). The data points represent XODs accepted in this dataset by the $\mathrm{d}n/\mathrm{d}z$ selection. There are no new forbidden regions compared to the distribution of the D1 dataset (Fig. 3.4).

3.3 SEQUENTIAL NEURAL POSTERIOR ESTIMATION

Simulation-based inference (SBI) is a broad domain encompassing various models. We use the Sequential Neural Posterior Estimation model, SNPE (Papamakarios and Murray, 2016), from the Python `sbi` package (Tejero-Cantero et al., 2020). This model uses neural density estimation techniques (Magdon-Ismail and Atiya, 1998) to conduct likelihood-free inference. In this section, we give a non-exhausting basic overview of the SNPE method, and we kindly refer interested readers for more in-depth details to (Papamakarios and Murray, 2016).

The SNPE and SBI methods, in general, usually work with a compressed version of the data. This is done to preserve a significant amount of relevant information about the data while reducing the dimensionality of the data as

Table 12: The four datasets used for training the networks. Column 2 indicates whether the coefficients of the cluster scaling relations were selected such as to match the observed redshift distribution. Column 3 shows whether cluster evolution was fixed to self-similar. Each dataset contains 70 000 XODs.

Acronym	$\mathrm{d}n/\mathrm{d}z$	γ_{MT} & γ_{LT}
D1	Yes	Free
D2	No	Free
D3	Yes	Fixed
D4	No	Fixed

much as possible. Compressing the data also keeps the data size for the inference manageable.

We compress our XODs by training a small ResNet model with two resolution levels and two ResNet blocks per level trained under mean square error regression loss as a regressor to estimate the Ω_m and σ_8 from the XODs. These estimates are our new parameter space. The XODs enter the SNPE method, which is represented as their compressed version provided by our trained ResNet model.

We aim to estimate the posterior probability distribution of the Ω_m and σ_8 for a given XOD. We want to get the posterior probability distribution $p(\theta \,|\, x = x_0)$, where x_0 corresponds to our target XOD diagram. The posterior probability distribution is proportional as

$$p(\theta \,|\, x = x_0) \propto p(x = x_0 \,|\, \theta) p(\theta) \tag{3.3}$$

where $p(\theta)$ is our prior and $p(x = x_0 \,|\, \theta)$ is the likelihood of our simulator model. In this sense, the simulator model is the pipeline simulating XODs plus the regressor neural network compressing them into the new parameter space. The data, x, in our work, represents the compressed version of the XOD.

With the simulation-based approach of the SNPE, we avoid the explicit computation of the likelihood and its necessary assumptions. The SNPE implements a Mixture Density Network (MDN) (Bishop, 1994) trained to output a posterior probability distribution as a mixture of 10 Gaussian components. The MDN consists of feed-forward neural networks trained to compute the components' mixing coefficients, means, and covariance matrices. In this work, we use the SNPE-a (Papamakarios and Murray, 2016), after the training procedure, the MDN models the posterior distribution.

Our XOD simulations are not sampled from a simple prior. We estimate the empirical proposal prior distribution of the simulations and account for it in the SNPE procedure to be able to set our desired prior.

3.4 RESULTS

In this section, we present the main results of this work as an estimation of cosmological parameters, Ω_m and σ_8 for a target XOD for a survey size of 1 000 deg^2. The dataset used for this experiment is the D1 (see Sec. 3.2.7). Fig. 3.6 shows the SNPE's cosmological prediction based on this data for a target test XOD. We compare the SNPE's prediction with a well-established method, a

Fisher analysis, focusing on the same setting as the D1 dataset, with $\mathrm{d}n/\mathrm{d}z$, and including evolution parameters. We compute the Fisher information matrix on the fiducial model parameters. Here, we note an XOD as a set of bins $\lambda_i(\theta)$, with i running over z–CR–HR bins, and its likelihood $\mathcal{L}$ can be expressed, assuming that the bins are independent and follow a Poisson distribution, as:

$$F_{\alpha\beta} = \sum_i \frac{1}{\lambda_i} \frac{\partial \lambda_i}{\partial \theta_\alpha} \frac{\partial \lambda_i}{\partial \theta_\beta}. \tag{3.4}$$

The derivatives are numerically computed following the methodology presented in Cerardi et al., 2023, ensuring the output's stability. As the Fisher matrix is computed around a particular point in the parameter space, it does not take account of any proposal distribution. Thus, to represent the effect of the $\mathrm{d}n/\mathrm{d}z$ selection, we compute the covariance matrix of the $\mathrm{d}n/\mathrm{d}z$ training sample and invert it to obtain the Fisher Proposal matrix. Naturally, this induces a loss of information as it only keeps the multi-variate Gaussian signal in the proposal distribution, but it is sufficient for the test we intend to perform here. The Fisher matrices can be added together and then inverted to recover the constraints on Ω_m and σ_8. We show the corresponding ellipse in Fig.3.6 and an estimated posterior from the SNPE for an XOD with its Ω_m and σ_8 near the fiducial model. The contours agree well, although the Fisher contours are slightly broader. This could be caused by (i) the prior information drawn from the proposal information being degraded and (ii) the Fisher computation only reflecting the properties of the local derivatives. While the Fisher prediction is anchored to the fiducial model and then well-centered, it is normal that the noisy realization of the target diagram offsets the SBI contours. The general agreement between the two constraining methods allows us to increase our confidence in the main results of this paper.

3.5 DISCUSSION

In this section, we first discuss the impact of the size of the training sample on the neural network performances. Next, we show what levels of accuracy can be expected as a function of the survey area. We discuss the effect of the $\mathrm{d}n/\mathrm{d}z$ selection and of the self-similar evolution assumption. Lastly, we discuss which steps would be taken when using this method on a real observed XOD instead of a simulated one.

3.5.1 *Training data size*

In deep learning studies, the larger the training set, the better in principle. However, simulating XOD is expensive, we thus investigate the trade-off between the size of the training samples and the returned cosmological accuracy. Our neural network training consists of two steps (ResNet and SNPE) and we thus have to validate the training-data size for each of them. First, the ResNet regressor compresses the XODs in a new parameter space and second, the SNPE MDN estimates the posterior probability distribution of Ω_m and σ_8 for a given ResNet-compressed XOD. For doing this, we fix the training size of one network on a

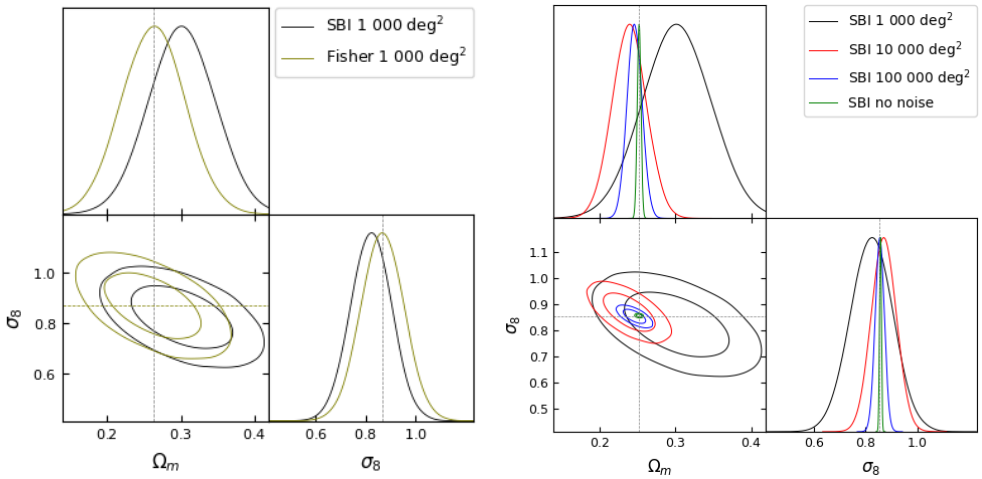

Figure 3.6: Left: Comparison of the SNPE's cosmological prediction with Fisher prediction on the same target XOD corresponding to 1 000 deg^2 survey area. Right: Comparison of the SNPE's cosmological results for four different survey areas. The SNPE was trained on our D1 dataset, with dn/dz selection and free γ parameters. The y-axis in the 1D posterior probability distribution plots indicates the probability density at each parameter value shown on the x-axis. Each 1D probability density function is normalized in a way that the total area under its curve equals 1.

larger volume while varying the size of the other. The results are presented in the next two figures.

Fig. 3.7 shows how the accuracy on the final prediction varies with the ResNet's training dataset size while keeping the SNPE training size set as 10 000 XODs. The quality of the final prediction does not significantly improve beyond 10 000 training samples.

Fig. 3.8 shows how the error on the final prediction varies with the SNPE's training dataset size while keeping the ResNet training size set to 40 960 XODs.

The quality of the final prediction is more sensitive to the ResNet's training size, while the SNPE can already perform well even with a few hundred training examples. This is expected because Papamakarios and Murray (2016) designed the SNPE to perform well even with a few hundred training samples. Based on these results, we concluded that our initial guess of creating 70 000 XODs per dataset is enough and that we do not need to enlarge the sample. We set as default values 20 480 for the ReNet4's training size and 10 000 XODs for the SNPE's training size. If not stated otherwise, the results presented in this paper are always produced with these training sizes.

3.5.2 *Testing the SNPE's calibration*

To test whether the SNPE density estimator is well-trained and calibrated, we re-train it eight times with the same settings of all hyperparameters, and let it make predictions for the same target XOD. This is done for all four survey areas. In these tests, its training data size is set at 10 000, and the XODs are always compressed with the same ResNet of the corresponding survey size that is not re-trained. The only difference between these tests is the seed used to draw random

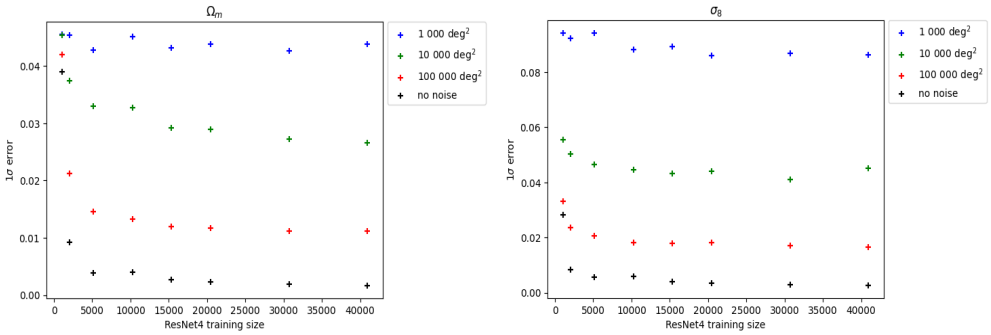

Figure 3.7: Amplitude of the 1 σ errors as a function of the ResNet's training size, for a training size of the SNPE set fixed at 10 000 XODs. We can see that the precision does not significantly improve after 20 000 training XODs.

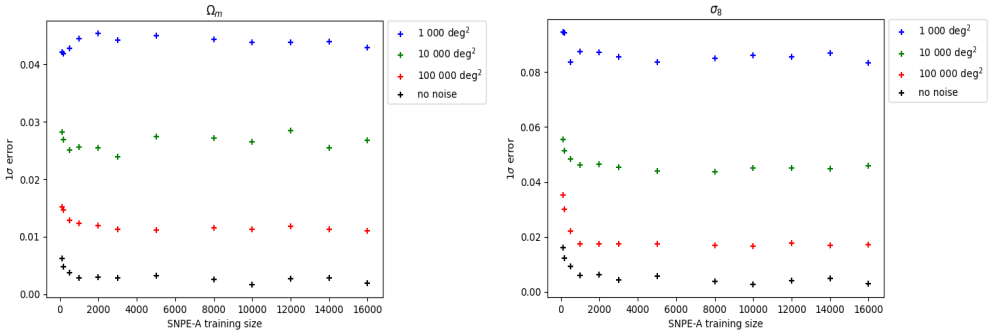

Figure 3.8: Amplitude of the 1 σ errors as a function of the SNPE's training size, for a fixed training size of the ResNet set at 40 960 XODs. We can see that the SNPE already performs well with a few hundred XODs as a training size. This was to be expected as it was intentionally developed to work in this way. Even though working already with smaller training XODs, we conservatively decided to use 10 000 XODs as our final training sample, given that we already simulated 70 000 XODs.

numbers from distributions when initializing the SNPE's network's layers. If the ML model is adequately trained, it will not have any significant deviations from its predictions when re-trained in this fashion. Fig. 3.9 shows four calibration tests for four different models, each trained for a different noise level represented as a survey size. We can see that the results are very consistent.

3.5.3 *Survey area*

As described in Sec 3.2.6, we are working with four different survey sizes, i.e., noise levels in the XOD pixels: namely 1 000 deg^2, 10 000 deg^2, 100 000 deg^2, and an infinite survey area (no noising at all). Even though the last two settings are 'thought experiments', they are potentially useful in understanding possible biases in our methodology. We perform these tests to see what levels of accuracy we can expect for a different survey area and to probe the effect of noise on the accuracy of our cosmological predictions.

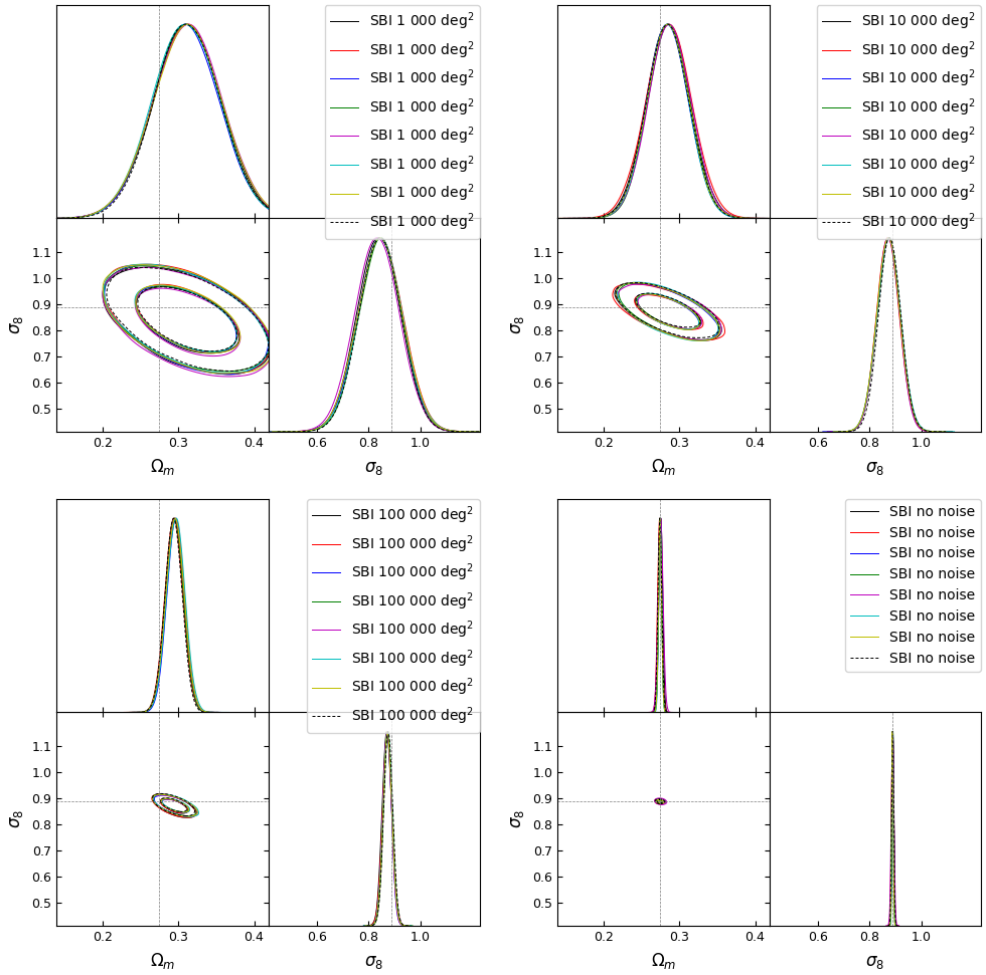

Figure 3.9: Calibration tests of the density estimator on four different survey areas. Each contour represents an output of a newly re-trained SNPE with the same training parameters. The only difference is in the seeds used for generating random numbers, e.g., initialization of weights in the neural network. The tests are done for the same testing XOD. We can see that the SNPE is well-trained in that regard it produces consistent results and does not show large deviations only due to small differences in the initialization of its neural networks.

Fig. 3.6 (right) shows how, within our framework, the accuracy of the cosmological parameters depends on the survey size. Tab. 14 summarises the cosmological constraints. Considering the first three columns, it appears that the cosmological accuracy roughly scales as the square root of the area, that is, as the square root of the number of clusters, which is sound. We observe that no noise at all in the XOD (infinite area) still results in non-zero errors. We may interpret this as a possible limitation of the numerical accuracy when constructing the XODs. Alternatively, it could also be due to keeping the same $16\times16\times10$ resolution for all noise levels. Decreasing the resolution in the case of an infinite survey area should lead to smaller errors in this regime. We tested this hypothesis by re-training the ML models for a $32\times32\times10$ XOD resolution. Fig. 3.10 shows the final posterior probability cosmological predictions for ML models trained on $32\times32\times10$ XOD resolution (blue) and the standard $16\times16\times10$ XOD resolution

(yellow) for two target XODs (left and right figures). Fig. 3.11 shows the same but for a setting with no Poisson noise (infinite deg^2). Tab. 13 shows the relative $1\,\sigma$ accuracy of these predictions. We can see that increasing the resolution does not significantly change the accuracy of our cosmological predictions. For the T1 XOD, the relative $1\,\sigma$ has a bit better accuracy for the 16×16×10 resolution in the case of 100 000deg^2 survey area and only for Ω_m. However, we observe an opposite trend for the T2 target XOD, which had a bit better accuracy on the Ω_m but for the case of 32×32×10 resolution of 100 000 deg^2 survey area. For the T2 target XOD, there was also a small improvement in the accuracy of our cosmological prediction for the 16×1610 resolution in the case of no noise at all (infinite deg^2 survey area). Based on these results, we conclude that increasing the resolution to 32×32×10 does not play a significant role in the accuracy of our cosmological predictions, when shot noise becomes negligible. The plausible interpretation is that, given that we work at the CR level, cluster spectral features (in particular emission lines) are somewhat erased by the rather low XMM-EPIC spectral resolution, containing itself sharp discontinuities[1]. This may cause the observed remaining weak degeneracy.

Table 13: This table shows the relative $1\,\sigma$ errors of our resolution tests for two target XODs T1 and T2 for two different XOD resolutions, 32×32×10 and 16×16×10 (the standard resolution) in a case of two different levels of Poissonian noise, corresponding to 100 000 deg^2 and an infinite deg^2 survey areas.

	10^5 deg^2				infinite deg^2			
	32×32×10		16×16×10		32×32×10		16×16×10	
	Ω_m	σ_8	Ω_m	σ_8	Ω_m	σ_8	Ω_m	σ_8
T1	3.7	1.9	3.2	2.0	0.7	0.3	0.7	0.4
T2	4.2	2.2	4.8	2.1	1.1	0.6	0.9	0.4

Table 14: Relative $1\,\sigma$ errors (in %) from SNPE for a single target diagram. Separate SNPE and regressor are trained for each dataset (D1-D4) and each survey area. The results are computed for the same XOD test target, always of the appropriate survey size, simulated with the γ evolution parameters fixed at the fiducial values and with the $\mathrm{d}n/\mathrm{d}z$ selection. This choice allowed us to use it fairly for the ML model trained on each of our four datasets.

	10^3 deg^2		10^4 deg^2		10^5 deg^2		infinite deg^2	
	Ω_m	σ_8	Ω_m	σ_8	Ω_m	σ_8	Ω_m	σ_8
D1	15.2	10.0	9.6	5.6	4.1	2.1	0.8	0.5
D2	17.6	13.4	11.2	7.2	7.9	5.0	4.1	2.1
D3	12.2	5.5	5.3	2.1	2.0	0.8	0.4	0.2
D4	15.0	10.5	7.6	4.7	4.0	2.3	2.4	0.9

1 https://xmm-tools.cosmos.esa.int/external/xmm_user_support/documentation/uhb/effareaonaxis.html

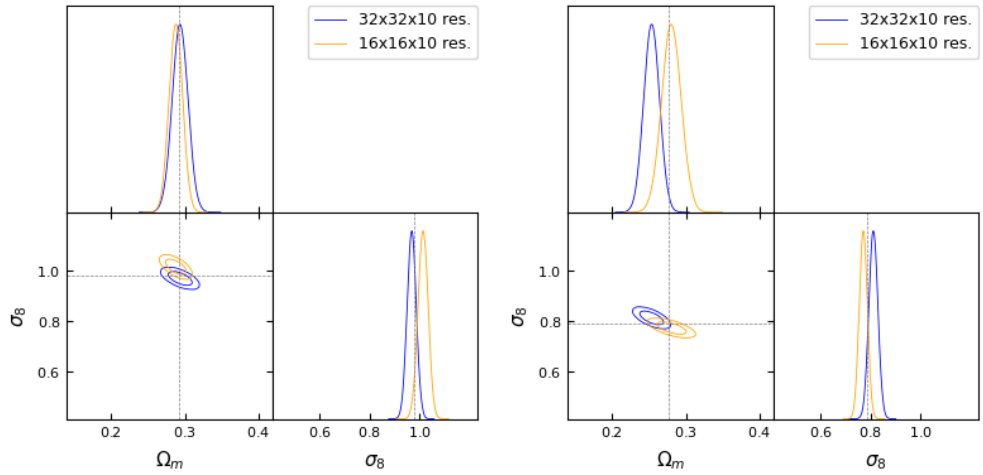

Figure 3.10: Tests of the accuracy of our cosmological predictions based on the resolution of the XODs in the case of 100 000 deg^2 survey area for two target testing XODs T1 (left) and T2 (right). The XOD resolution is color-coded.

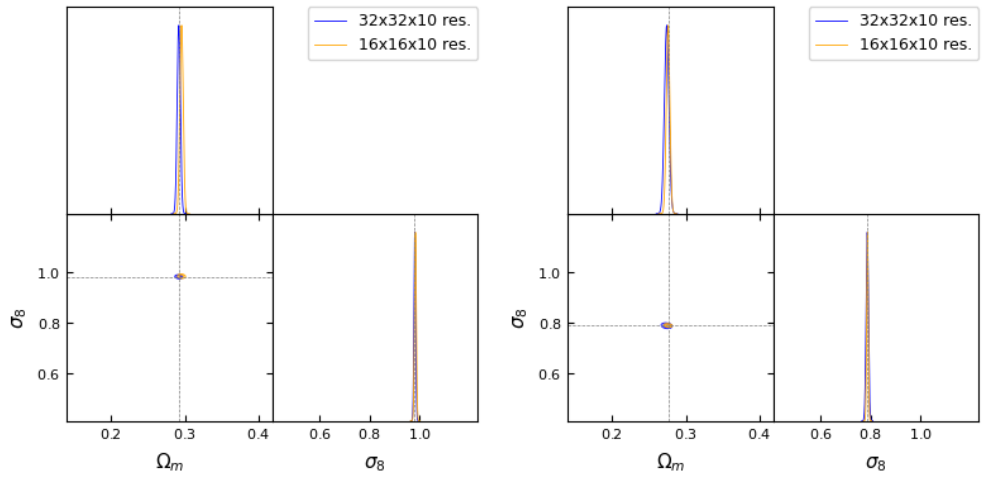

Figure 3.11: Tests of the accuracy of our cosmological predictions based on the resolution of the XODs in the case of no Poisson noise (infinite deg^2 survey area) for two target testing XODs T1 (left) and T2 (right). The XOD resolution is color-coded.

3.5.4 *dn/dz subselection*

In this section, we explore how the dn/dz selection improves the accuracy of our cosmological predictions. This selection effectively excludes all XODs whose redshift distribution is incompatible with the observations (i.e., here, with the dn/dz of our fiducial model) at the 3σ level. However, the test is rather permissive, so XODs that pass the dn/dz test can show very different combinations of the eight free parameters due to the high level of degeneracy for such a large number of free parameters. Compared to a normal uniform prior distribution, the dn/dz restricts the range of possible parameter combinations (Fig. 3.4), introducing forbidden areas in the parameter space.

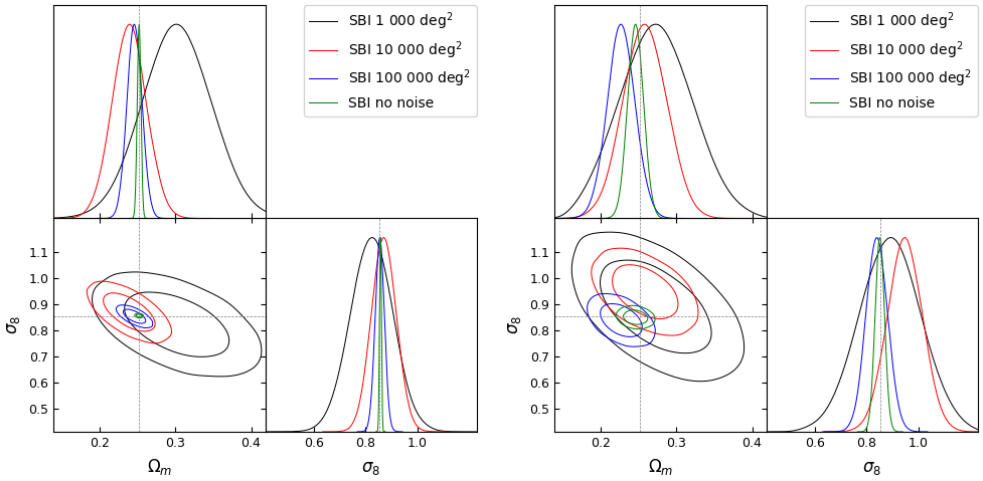

Figure 3.12: Posterior probability distribution, trained on the D1 dataset ($\mathrm{d}n/\mathrm{d}z$ selection, left) and on the D2 dataset (no $\mathrm{d}n/\mathrm{d}z$ selection, right) (Tab. 12). These datasets have fixed γ cluster evolution parameters. Four MDNs are trained per dataset each on a different survey area. These results are computed for the same target testing XOD, always of the appropriate survey size, that was produced letting all 8 parameters free and verifying the $\mathrm{d}n/\mathrm{d}z$ selection.

To understand how the $\mathrm{d}n/\mathrm{d}z$ selection improves the final predictions, we compare the ResNet and the SNPE trained on the D1 and D2 sets. (see Sec. 3.2.7). A separate density estimator must be trained on the corresponding dataset for each survey area. The predictions are made for the same target XOD modulated by the noise level pertaining to the survey area. The results are displayed in Fig. 3.12. The relative errors are reported in Tab. 14. We can see that plugging the $\mathrm{d}n/\mathrm{d}z$ selection decreases the size of the error bars on cosmological parameters. This difference is most prominent for smaller noise levels corresponding to larger survey areas. Hence, filtering out unrealistic distributions at the $\mathrm{d}n/\mathrm{d}z$ level is a quick and efficient method to improve the constraining power of this technique.

3.5.5 *Constraining scaling relations - fixing gammas*

In this section, we investigate how the performance of our cosmological predictions improves when reducing the number of free parameters to six, assuming cluster self-similar evolution (i.e. γ_{MT} and γ_{LT} are fixed to their fiducial values). By doing this, the potential degeneracy between cluster physics and cosmology is lowered. The test uses data sets D3 and D4. Fig. 3.13 and Fig. 3.14 show the outcome, compared with the 8-parameter realizations, including or excluding the $\mathrm{d}n/\mathrm{d}z$ selection, respectively. Tab. 14 shows the relative $1\,\sigma$ errors. Clearly, as expected, fixing cluster evolution improves the accuracy of our cosmological predictions. We also note a change in the direction of the degeneracy.

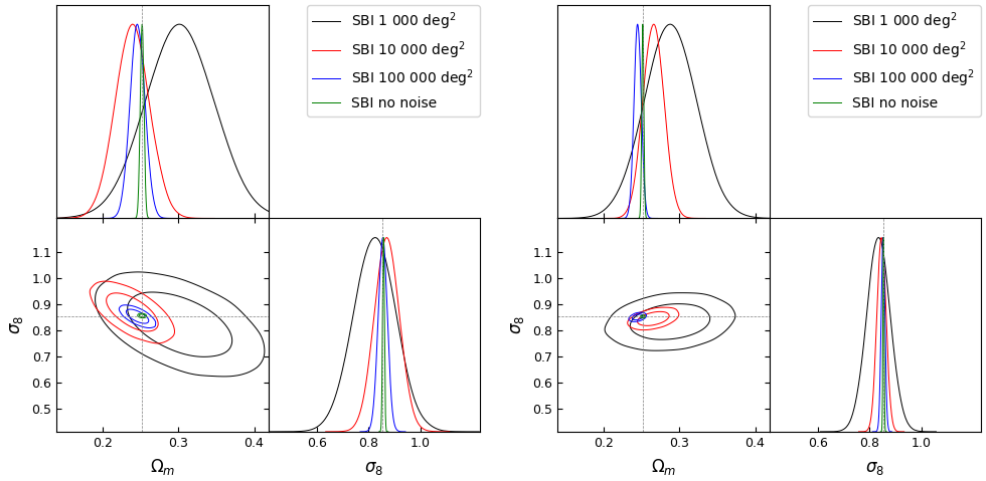

Figure 3.13: Posterior probability distribution estimated trained on D1 dataset (left) and training on D3 dataset (right) (Tab. 12). These datasets both have the $\mathrm{d}n/\mathrm{d}z$ selection but differ in the choice of fixing the cluster evolution γ parameters on their fiducial values. Four MDNs are trained per dataset each on a different survey area. These results are computed for the same target testing XOD, always of the appropriate survey size, that was produced having all 8 parameters free and with the $\mathrm{d}n/\mathrm{d}z$ selection. We can see that fixing the γ parameters improves the accuracy of the SNPE's cosmological predictions.

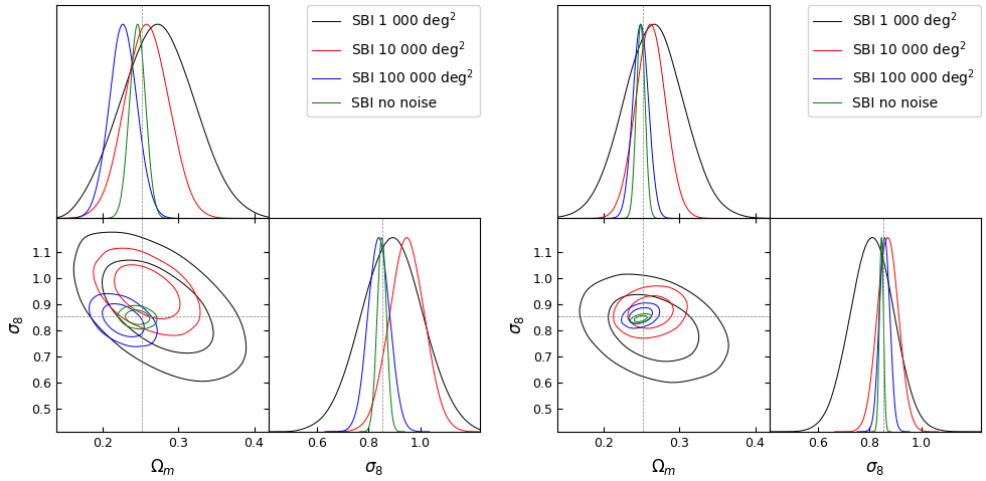

Figure 3.14: Posterior probability distribution estimated trained on D2 dataset (left) and training on D4 dataset (right) (Tab. 12). These datasets both come with no $\mathrm{d}n/\mathrm{d}z$ selection but differ in the choice of fixing the γ parameters. Four MDNs are trained per dataset each on a different survey area. These results are computed for the same target testing XOD, always of the appropriate survey size, that was produced having all 8 parameters free and with the $\mathrm{d}n/\mathrm{d}z$ selection. We can again observe that using the $\mathrm{d}n/\mathrm{d}z$ selection and fixing the γ parameters improves the accuracy of the SNPE's cosmological predictions.

3.5.6 *Applying the method to real observations*

Now, the question is: what steps would be necessary to apply our method to a real observed XOD? So far in the paper, we showcased the method performance

on testing XODs which have been inside the parameter range used to create the training simulations (Tab. 11). When used on an XOD composed from an observed galaxy cluster sample, with the underlying true Ω_m and σ_8 being well within the range of the training simulations parameters, we would see results similar to those demonstrated in this work. Given the current constraints on Ω_m and σ_8 from various probes, we can confidently expect a real XOD to lie well within the boundaries of our study. However, in principle, we have to address the case where the true cosmological parameters would be located outside of the simulation training range. Since neural regression models are known to extrapolate badly, the posterior estimation would probably fail in this situation. This could result in a final estimated posterior being placed towards the edges of the training sample distribution in the direction of the true underlying cosmological parameters. If we obtained results like this on a real observed XOD, we would need to create a second training sample, now centred on the central values of the estimated posterior probability distribution, and to retrain the neural networks involved, the ResNet regressor as well as the MDN density estimator on this new dataset. These newly trained networks would give us a new estimate of the cosmological parameter posterior probability distribution. We would need to repeat this process until the final estimated posterior would be well-centered within the training sample range. We show such a situation in Fig.3.15 where the true Ω_m and σ_8 cosmological parameters used to simulate this testing XOD are close to the borders of our simulation box.

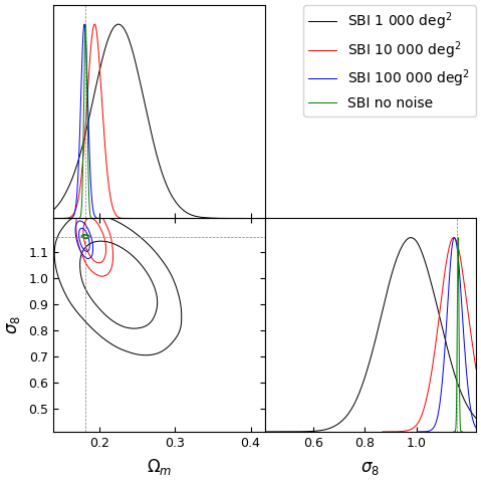

Figure 3.15: This figure shows the SNPE's cosmological results for four different survey areas computed for a target XOD with its Ω_m and σ_8 values close to the borders of our simulation box. The regressor and density estimator neural networks were trained on our D1 dataset, with $\mathrm{d}n/\mathrm{d}z$ selection and free γ parameters.

Other effects will have to be considered when applying this method to a real sample. Firstly, using a CR cut to mimic the selection function is very simplistic and does not correctly render the detection probability within an X-ray survey, which is two-dimensional (flux vs apparent size) (Clerc et al., 2024; Pacaud et al., 2006). However, the high flexibility of the ResNet regression models should make the training just as efficient on more complex selection functions. Secondly,

measurement errors will have to be included. For X-ray surveys, the number of photons for individual clusters can be as low as a few tens of counts, and hence, the observed CR and HR are noisy estimates of the true quantities. Implementing an error model is not a showstopper for the inference method: basically, this can be modeled by two-dimensional filtering of the XOD (Clerc et al., 2012b).

3.6 CONCLUSIONS

This paper is a proof of concept investigating the cosmological constraining power of an X-ray cluster survey using only three observed parameters: the X-ray flux in a given band (XMM-Newton count rate, CR), an X-ray colour (XMM-Newton hardness ratio, HR), and the redshift. The cosmological inference is based on a neural network trained on all possible realizations of the CR-HR-z distribution of the detected cluster population (XOD diagrams). We have no knowledge of cluster masses or X-ray temperatures. This means we discard any a priori knowledge of the cluster scaling relations. Practically, to set up the training sample for the network, we compute several tens of thousands of combinations of the cosmological parameters (Ω_m and σ_8), with ad hoc scaling relations for which the coefficients are drawn at random. This formalism allows us to compute CR(M, z) and HR(M, z) values of a cluster sample for any cosmology. We note that CR and HR, being purely instrumental measures, we need this intermediate analytical step based on scaling relations to compute the desired quantities. Since the cosmological parameters and the scaling relation coefficients are independently and randomly drawn, we can say that our modeling is scaling-relation free, that it covers a very large parameter space and, consequently, encompasses all possible effects of cluster physics (e.g., cool-cores or feedback).

The performances of our approach are sensible: cosmological constraints scale roughly as the square root of the surveyed area, imposing that the simulated XODs verify the observed $\mathrm{d}n/\mathrm{d}z$ cluster distribution, and fixing the cluster evolutionary parameters leads to better constraints.

The absolute constraining power of our method can be compared to the recent eROSITA cosmological results (Ghirardini et al., 2024). Their cosmological sample has 5 259 massive clusters over 12 791 deg^2; our 1 000 deg^2 realization yields about 4 000 clusters. Our predicted uncertainties on Ω_m and σ_8 are ~ 2.2 and ~ 4.3 times larger, respectively. This is not surprising because (i) the eROSITA clusters are more massive compared to our 1 000 deg^2 because our sample has a flux cut based on the XXL analysis, thus having a similar cluster mass distribution, also having $\sim$ 4 000 clusters on a 1 000 deg^2 compared to the similar 5 259 eROSITA clusters over 12 791 deg^2 and (ii) several assumptions are made on the eROSITA cluster scaling relations (from optical data sets) to calibrate the individual cluster masses (Ghirardini et al., 2024).

Applying our method to real data would require implementing a two-dimensional selection function in the CR vs cluster-apparent-size plane. Practically, this means that we need to introduce a supplementary scaling relation (M_{500}-R_c) in the formalism that produces the XODs; this is not a problem as long as the coefficients of this new relation would be drawn at random, the same way we do with the scaling relations in this study. Modeling the measurement errors would also be needed.

The present study treats the scaling relation coefficients as nuisance parameters; this is a fair approach, but because these parameters amount to six, and we exclude any a priori knowledge on them, there remains a significant degeneracy with the cosmological parameters. This explains why our predicted errors on the cosmological parameters appear rather large. A much more efficient approach is to create the XOD training sample from numerical simulations spanning a large range of cosmologies and for various feedback assumptions (e.g., the CAMELS simulations, Villaescusa-Navarro et al., 2021). In this case, the number of parameters ruling the cluster's physical properties is drastically reduced (two AGN feedback parameters) and, moreover, these parameters should not be considered as nuisance parameters. In the end, the SBI, relying on such a training sample, faces much less degeneracy and delivers the cosmological parameters plus the feedback parameters, which are the truly relevant physics parameters (paper VII, Cerardi et al. in prep).

Acknowledgements

We are very grateful for Tom Charnock's huge help and contribution to this project in its early stages. M. K. and N. W. are supported by the GACR grant 21-13491X. M. K.'s stays in France and Italy were supported by the French Alternative Energies and Atomic Energy Commission internship, the ERASMUS program, and the Italian Government Scholarship issued by the Italian MAECI.

4

A multifrequency characterization of the extragalactic hard X-ray sky

M. Kosiba,[1,2] *F. Massaro,*[2,3,4] *H. A. Peña-Herazo,*[5] *N. Masetti,*[6,7] *A. Paggi,*[2,3,4] *V. Chavushyan,*[8] *E. Bottacini,*[9,10] *and N. Werner*[1]

[1]Department of Theoretical Physics and Astrophysics, Faculty of Science, Masaryk University, Kotlářská 2, Brno, 611 37, Czech Republic
[2]Dipartimento di Fisica, Università degli Studi di Torino, via Pietro Giuria 1, I-10125 Torino, Italy.
[3]Istituto Nazionale di Fisica Nucleare, Sezione di Torino, I-10125 Torino, Italy
[4]INAF–Osservatorio Astrofisico di Torino, via Osservatorio 20, I-10025 Pino Torinese, Italy
[5]East Asian Observatory, 660 North A'ohōkū Place, Hilo, Hawaii 96720, USA.
[6]INAF - Osservatorio di Astrofisica e Scienza dello Spazio, via Piero Gobetti 101, 40129 Bologna, Italy
[7]Departamento de Ciencias Físicas, Universidad Andrés Bello, Fernández Concha 700, Las Condes, Santiago, Chile
[8]Instituto Nacional de Astrofísica, Óptica y Electrónica, Apartado Postal 51-216, 72000 Puebla, México.
[9]Dipartimento di Fisica e Astronomia G. Galilei, Univeristà di Padova, Padova, Italy.
[10]Eureka Scientific, 2452 Delmer Street Suite 100, Oakland, CA 94602-3017, USA.

Published in Astronomy & Astrophysics, Volume 670, id.A171, 18 pp. 2023. doi:10.1051/0004-6361/202243848.

4.1 ABSTRACT

Nowadays we know that the origin of the Cosmic X-ray Background (CXB) is due to the integrated emission of nearby active galactic nuclei. Thus to obtain a precise estimate of the contribution of different source classes to the CXB it is crucial to have a full characterization of the hard X-ray sky.

We present a multifrequency analysis of all sources listed in the 3^{rd} release of the Palermo *Swift*-BAT hard X-ray catalog (3PBC) with the goal of (i) identifying and classifying the largest number of sources adopting multifrequency criteria, with particular emphasis on extragalactic populations and (ii) extracting sources belonging to the class of Seyfert galaxies to present here the release of the 2^{nd} version of the Turin-SyCAT.

We outline a classification scheme based on radio, infrared and optical criteria that allows us to distinguish between unidentified and unclassified hard X-ray

sources, as well as to classify those sources belonging to the Galactic and the extragalactic populations.

Our revised version of the 3PBC lists 1176 classified, 820 extragalactic and 356 Galactic ones, 199 unclassified and 218 unidentified sources. According to our analysis, the hard X-ray sky is mainly populated by Seyfert galaxies and blazars. For the blazar population, we report trends between the hard X-ray and the gamma-ray emissions since a large fraction of them have also a counterpart detected by the *Fermi* satellite. These trends are all in agreement with the expectations of inverse Compton models widely adopted to explain the blazar broadband emission. For the Seyfert galaxies, we present the 2^{nd} version of the Turin-SyCAT including a total of 633 Seyfert galaxies, with 282 new sources corresponding to an increase by ∼80 % with respect to the previous release. Comparing the hard X-ray and the infrared emissions of Seyfert galaxies we confirm previous statistical results in the literature.

Finally, we highlight that the SWIFT archive has already extensive soft X-ray data available to search for potential counterparts of unidentified hard X-ray sources. All these datasets will be reduced and analyzed in a forthcoming analysis to determine the precise position of low energy counterparts in the 0.5 – 10 keV energy range for 3PBC sources that can be targets of future optical spectroscopic campaigns, necessary to obtain their precise classification.

4.2 INTRODUCTION

The Cosmic X-ray background (CXB) was discovered when the earliest X-ray astronomical rocket experiments were carried out (see e.g. Giacconi et al., 1962). It appeared as a diffuse component of X-ray radiation distributed in all directions. In the last decades, after its discovery, several different scenarios were proposed to interpret its origin, such as considering spanning new types of faint discrete X-ray sources whose integrated emission could have been responsible for the CXB (e.g. (Gilli et al., 1999), Gilli et al., 2001) up to diffuse radiative processes occurring in the intergalactic space, as for example exotic emission from dark matter particle decay (see e.g. Abazajian et al., 2001). However, the solution to this puzzle arose thanks to deep images obtained first with the ROentgen SATellite *ROSAT* (Hasinger et al., 1999), collected in the early ninety's, and more recently with *Chandra* (Weisskopf et al., 2000), all revealing that about 80 % of CXB is resolved (Hasinger et al., 1998), between 0.5 keV and 2 keV, as suggested by Cavaliere and Fusco-Femiano, 1976. At hard X-ray energies the fraction of resolved CXB by *Swift* and *INTEGRAL* is of 2% (Bottacini et al., 2012) and by *NuSTAR* is of 35% (Harrison et al., 2016). Thus, the origin of the CXB is nowadays established to be mainly due to the high energy emission of the extragalactic discrete sources, whose large fraction belongs to different classes of active galactic nuclei (AGNs) (Gilli et al., 2007).

The first survey in the hard X-ray band was carried by the *UHURU* satellite (a.k.a SAS-1 Giacconi et al., 1971). Since the discovery of the CXB, many surveys were performed in the soft and hard X-ray bands, including Forman et al. (1978) who produced a catalogue of 339 X-ray sources observed by the *UHURU* satellite in the 2–20 keV energy band. Levine et al. (1984), using the X-ray and Gamma-ray detector HEAO-A4 on board the *HEAO 1* satellite (Rothschild et al., 1979) presented an all-sky survey in 13–180 keV range detecting 77 new sources. The

hard X-ray component of the CXB radiation, observable between 3 keV and up to 300 keV, shows a distinct peak at $\sim$30 keV (Gruber et al., 1999) being extremely uniform across the sky with the only exception of an overdensity along the Galactic plane (Krivonos et al., 2007a; Revnivtsev et al., 2006; Valinia and Marshall, 1998) and it is again strictly connected with AGN population emitting in the hard X-rays (Frontera et al., 2007).

Measurements carried out in the hard X-rays thanks to recent satellites, as the INTErnational Gamma-Ray Astrophysics Laboratory *INTEGRAL* (Winkler et al., 2003), with its Imager on Board the *INTEGRAL* Satellite *IBIS* (Ubertini et al., 2003) and Neil Gehrels *Swift* Observatory (Gehrels et al., 2004) with its Burst Alert Telescope (BAT Barthelmy, 2004) on board, significantly improved our understanding of the origin of the CXB and refined its measurement. This has been also possible thanks to improvements achieved in the preparation of hard X-ray source catalogues (see e.g. Beckmann et al., 2006; Bird et al., 2010; Bottacini et al., 2012; Cusumano et al., 2010; Krivonos et al., 2007a; Krivonos et al., 2022; Markwardt et al., 2005; Oh et al., 2018; Sazonov et al., 2007; Tueller et al., 2008) and the association of hard X-ray sources with their low energy counterparts (Bär et al., 2019; Koss et al., 2019; Malizia et al., 2010; Smith et al., 2020, e.g.) and their optical spectroscopy follow-ups (e.g. Marchesini et al., 2019a; Masetti et al., 2008, 2006b, 2012, 2013; Parisi et al., 2014; Rojas et al., 2017).

There are three major catalogues built on observations collected in the last decade with two major space missions still active: (i) the Palermo *Swift*-BAT hard X-ray catalogue (Cusumano et al., 2010) based on 54 months of the Swift-BAT operation, currently updated to its 3^{rd} release and with a 4^{th} release ongoing[1], (ii) the *Swift*-BAT all-sky hard X-ray survey, that published the 105 month *Swift*-BAT catalogue (see e.g. Oh et al., 2018), and (iii) the *INTEGRAL* IBIS catalogue in the energy range 17-100 keV (Bird et al., 2016), performed using the *INTEGRAL* Soft γ-ray Imager (ISGRI) (Lebrun et al., 2003), the low energy CdTe γ-ray detector on (*IBIS*) telescope (Winkler et al., 2003).

Here we focus on the investigation of the 3^{rd} release of the Palermo *Swift*-BAT hard X-ray catalogue (hereinafter 3PBC), with particular emphasis on extragalactic sources, since the release of the next version is currently ongoing. The 3PBC is based on the data reduction and detection algorithms of the first Palermo *Swift*-BAT Catalog hard X-ray catalogue (Cusumano et al., 2010; Segreto et al., 2010). The 3PBC is available only online[2] thus we point for the reference the publication of its 2^{nd} release, the 2PBC (Cusumano et al., 2010). On the other hand, the 2PBC provides data in three energy bands, namely: 15 – 30 keV, 15 – 70 keV, 15 – 150 keV for a total of 1256 sources above 4.8 σ level of significance, where 1079 hard X-ray sources have an assigned soft X-ray counterpart, while the remaining 177 are still unassociated. The total source number increased in 3PBC to 1593 when considering a signal to noise ratio above 3.8, which is the catalogue release we analysed here. Please note, that only three 3PBC sources are detected at signal to noise ratio lower than 5 and that all Turin-SyCAT sources with a 3PBC counterpart are detected in the 3PBC at signal to noise ratio above 6. The 3PBC catalogue covers 90 % of the sky down

1 https://www.ssdc.asi.it/bat54/
2 http://bat.ifc.inaf.it/bat_catalog_web/66m_bat_catalog.html

to a flux limit of 1.1×10^{-11} erg cm^{-2} s^{-1}, decreasing to ∼ň,50 % when decreasing the flux limit to 0.9×10^{-11} erg cm^{-2} s^{-1}.

First, we verified source classification for all associated counterparts, adopting a multifrequency approach. This analysis was corroborated by checking if additional studies, available in the literature and carried out after the last 3PBC release, allowed us to obtain a more complete overview of source populations emitting in the hard X-rays. Then, our final goal was to explore in detail those extragalactic sources being identified as Seyfert galaxies (Antonucci and Miller, 1985) to (i) release the 2^{nd} version of the Turin-SyCAT (Peña-Herazo et al., 2022) and thus (ii) refine our statistical analysis on the correlation found between the infrared (IR) and the hard X-ray fluxes for this extragalactic population. Additionally we also aim at investigating possible connections between the hard X-ray and the gamma-ray emission in those blazars detected by Fermi-LAT. It is worth noting that given our final goal, the classification task performed on the Galactic sources is mainly devoted to exclude them from the final sample of new Seyfert galaxies.

The present work will be also relevant for the association of hard X-ray sources with their low energy counterpart, which will be included in the next releases of hard X-ray catalogues. In addition, we highlight that the proposed investigation will also provide a more complete overview of those sources, that lack an assigned low energy counterpart as still unidentified.

Finally, we remark that the reason underlying the choice of working with the 3PBC rather than subsequent versions of hard X-ray catalogues is mainly motivated by the opportunity of having more multifrequency information available in the literature. However a comparison with other recent catalogues as: the 105 month *Swift*-BAT catalog[3] (Oh et al., 2018) and the *INTEGRAL* hard X-ray catalogue (Bird et al., 2016) are also included in the present analysis.

The structure of the paper is outlined as follows: in Section 4.3 , we described various catalogues and surveys used to search for multifrequency information related to high and low energy counterparts of hard X-ray sources; in Section 4.4 , we present our multifrequency classification scheme adopted to label source counterparts. Then Section 4.5 focuses on the main results on the characterization of the extragalactic hard X-ray sources while Section 4.6 is entirely devoted to the second release of the Turin-SyCAT catalogue and the statistical analysis for the IR – hard X-ray connection. Finally, our summary, conclusions and future perspectives are given in Section 4.7.

We used cgs units unless stated otherwise. We adopted ΛCDM cosmology with $\Omega_M = 0.286$, and Hubble constant $H_0 = 69.6\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$ (Bennett et al., 2014) to compute cosmological corrections, the same used for the 1^{sth} release of the Turin-SyCAT (Peña-Herazo et al., 2022). WISE magnitudes are in the Vega system and are not corrected for the Galactic extinction, as it would not change the magnitude. As shown in our previous analyses (D'Abrusco et al., 2014; D'Abrusco et al., 2019; Massaro and D'Abrusco, 2016), such correction affects mainly the magnitude at 3.4 μm for sources lying at low Galactic latitudes (i.e., $|b| < 20^\circ$), and it ranges between 2 % and 5 % of their magnitude values, thus not significantly affect our results. We indicate the WISE magnitudes at 3.4, 4.6, 12, and 22 μm as W1, W2, W3, W4, respectively. For all WISE

3 https://heasarc.gsfc.nasa.gov/W3Browse/swift/swbat105m.html

magnitudes of sources flagged as extended in the AllWISE catalogue (i.e., extended flag "ext_flg" greater than 0) we used values measured in the elliptical aperture. Sloan Digital Sky Survey (SDSS) (Abdurro'uf et al., 2021; Blanton et al., 2017) and Panoramic Survey Telescope & Rapid Response System (Pan-STARRS) (Chambers et al., 2016) magnitudes are in the AB system. Given the large number of acronyms used here, mostly due to different classifications and telescopes used, we summarized them in Table 15.

Table 15: Table of Acronyms used in the text.

Acronym	Meaning
ATNF	australian telescope national facility
CXB	cosmic X-ray background
AGN	active galactic nuclei
BLL	BL-Lac object
BZG	Galaxy Dominated Blazars
BZU	blazar of uncertain type
CV	cataclysmic variable
FSRQ	flat spectrum radio quasar
HERG	high excitation radio galaxy
LERG	low excitation radio galaxy
LINER	Low-ionization nuclear emission-line region galaxy
NOV	novae
PN	planetary nebulae
PSR	pulsar
QSO	quasi stellar objects
RDG	radio galaxy
SNR	supernovae remnant
WD	white dwarf
XBONG	X-ray bright optically normal galaxy

4.3 HUNTING COUNTERPARTS OF HARD X-RAY SOURCES: CATALOGUES AND SURVEYS

This section provides a basic overview of all major catalogues used to carry out cross-matching analysis across the whole electromagnetic spectrum. Here we considered several (i) low energy and multifrequency catalogues, listing sources detected in radio, infrared and optical surveys and or based on literature analyses and (ii) high energy catalogues, based on hard X-rays and γ-ray surveys.

It is worth noting that the 3PBC catalogue is based on a moderate shallow survey thus we expect relatively bright sources in the hard X rays to be also bright at lower energies, at least for the extragalactic population of 3PBC sources, being

mainly constituted by AGNs. This limits the number of catalogues used to perform the cross-matching analysis and we used the same adopted in the original 3PBC analysis. Our analysis has been also augmented by using NED[4] and SIMBAD[5] databases, where we queried all sources having a low energy counterpart listed in the 3PBC before providing a final classification to verify the presence of updated literature information that is not reported in the catalogues adopted for the cross-matching analysis. All catalogues used in the current analysis are listed in Table 21.

4.3.1 *Low energy catalogues for cross-matching analysis*

At low frequencies, from radio to X-ray energies below 10 keV, we mainly considered:

1. The Revised Third Cambridge catalog[6] (3CR, (Spinrad et al., 1985)). This catalogue provides radio and optical data for 298 extragalactic sources, being the most powerful at low radio frequencies. It includes their positions, magnitudes, classification and redshifts with only 25 sources being still unidentified (Maselli et al., 2016; Massaro et al., 2013b; Missaglia et al., 2021). A vast suite of multifrequency observations at radio, infrared, optical and X-ray energies are also available for more than 90 % of the whole 3CR (see e.g. Maselli et al., 2016; Massaro et al., 2015c; Stuardi et al., 2018). The 3CR catalogue was create with a flux density limit $S_{178} \geq 2 \times 10^{-26}\,\mathrm{W\,m\,(Hz)^{-1}}$ at 178 MHz, spanning across the northern hemisphere with declination above -5 degrees. The 3CR catalogue has been also augmented by a vast suite of multifrequency observations carried out in the last decades that provides all information necessary to have a completed overview of the source classification (Balmaverde et al., 2021, 2019; Hilbert et al., 2016; Kotyla et al., 2016; Madrid et al., 2006; Massaro et al., 2015c; Missaglia et al., 2021; Privon et al., 2008; Stuardi et al., 2018).

2. The Fourth Cambridge Survey catalog (4C) [7] is based on the radio survey which used the large Cambridge interferometric telescope at the Mullard Radio Astronomy Observatory at frequency 178 $\mathrm{Mc\,s^{-1}}$, detecting sources which have flux density $S_{178} \geq 2 \times 10^{-26}\,\mathrm{W\,m\,(Hz)^{-1}}$. It is published in two papers, the first one listing 1219 sources at declination between $+20°$ and $+40°$ (Pilkington and Scott, 1965), while the second one includes 3624 sources in two declination ranges, $-07°$ to $+20°$ and $+40°$ to $+80°$ (Gower et al., 1967).

3. The Australia Telescope National Facility (ATNF) [8] Pulsar catalog (Manchester et al., 2005) is a complete catalog listing more than 1500 pulsars (PSR). Accretion powered X-ray PSRs are not included in this catalogue, because they have different periods, unstable on short timescales. The catalogue is based on the PSR database of 558 PSRs (Taylor et al., 1993)

4 https://ned.ipac.caltech.edu/
5 http://simbad.cds.unistra.fr/simbad/
6 https://ned.ipac.caltech.edu/uri/NED::InRefcode/1985PASP...97..932S
7 http://astro.vaporia.com/start/fourc.html
8 https://heasarc.gsfc.nasa.gov/W3Browse/all/atnfpulsar.html

which was further supplemented by more recent PSR databases (Edwards et al., 2001; Manchester et al., 2001) to establish the ATNF PSR catalogue.

4. The Catalog of Galactic Supernovae Remnants (SNRs) [9] (Green, 2017), which is an updated version of the original catalogue of galactic SNRs (Green, 1984), currently listing 295 SNRs built on the available results published in literature updated to 2016.

5. The 4^{th} edition of the catalogue of High mass X-ray binaries in the Galaxy [10] (Liu et al., 2006) provides 114 sources, updated with 35 new sources detected, most of them being X-ray binaries having a Be type star or a supergiant star as an optical companion.

6. The 7^{th} edition [11] of the catalogue of cataclysmic variables (CVs), low mass X-ray binaries and related objects (original paper, (Ritter and Kolb, 2003)) lists 1166 cataclysmic variables, 105 low-mass X-ray binaries, and 500 related objects. For a total of 1771 sources. The sources are provided with coordinates, apparent magnitudes, orbital parameters, stellar parameters and other characteristics. The entire catalogue is split into three tables provided online.

7. The 4^{th} edition of the catalogue of Low mass X-ray binaries in the Galaxy and Magellanic Clouds[12] (Liu et al., 2007) contains 187 sources, updated by 44 newly discovered sources. The companion star of a Low mass X-ray binary is typically a K or M type dwarf star. Small percentages of the companion stars are G type, red giants, or white dwarfs, and even smaller percentages of companions are A and F type stars. Sources are provided with their optical counterparts, spectra, X-ray luminosities, system parameters, stellar parameters of the components and other parameters.

8. The Catalog and Atlas of Cataclysmic Variables[13] (CVcat, (Downes et al., 2005)) presented its final release in January 2006 listing 1600 sources. The catalogue provides all types of cataclysmic variables like novae, dwarf-novae, nova-like variables, sources classified only as CVs, interacting binary WDs and possible supernovae. This catalogue contains also all objects that have been classified as CVs at some point in the past and are no longer considered being CVs. Those stars are labelled as NON-CV and are provided also with relevant references.

9. To cross-match the sources with galaxy clusters we used only the Abell catalogue of rich galaxy clusters[14] (Abell et al., 1989). This catalogue was conducted by a manual all-sky search for overdensities of galaxies on photographic plates. The catalogue contains 4073 rich galaxy clusters, with at least 30 galaxies in magnitude range between m_3 and $m_3 + 2$, where m_3 is the magnitude of the third brightest cluster galaxy.

9 https://heasarc.gsfc.nasa.gov/W3Browse/all/snrgreen.html
10 https://heasarc.gsfc.nasa.gov/w3browse/all/hmxbcat.html
11 https://heasarc.gsfc.nasa.gov/W3Browse/all/ritterlmxb.html
12 https://heasarc.gsfc.nasa.gov/W3Browse/all/lmxbcat.html
13 https://heasarc.gsfc.nasa.gov/W3Browse/all/cvcat.html
14 https://heasarc.gsfc.nasa.gov/W3Browse/all/abell.html

4.3.2 *High energy surveys for cross-matching analysis*

We also compared our analysis and classification of 3PBC sources with those of two hard X-ray catalogues (energies larger than 10 keV) and one of the latest releases of the Fermi catalogue of γ-ray sources. The former comparison allows us also to obtain more information about the source classification in particular for the Galactic population, while the latter one allows us to look for any trend between the hard X-ray and the γ-ray emission for the class of blazars.

1. The 105 month *Swift*-BAT catalog[15] (Oh et al., 2018) is created from data of a uniform hard X-ray all-sky survey in 14-195 keV band. It was developed using the same detector as the 3PBC catalogue, but implementing different source algorithms to build X-ray images, data reduction and source detection. Over 90 % of the sky is covered down to a flux limit of 8.40×10^{-12} erg cm^{-2} s^{-1} and over 50 % of the sky is covered down to a flux limit of 7.24×10^{-12} erg cm^{-2} s^{-1}. The catalogue provides 1632 hard X-ray sources detected above the 4.8 σ level, presenting 422 new detections compared to the previous version of 70 month *Swift*-BAT catalogue (Baumgartner et al., 2013). The catalogue contains 1132 extragalactic sources, out of which 379 are Seyfert I and 448 Seyfert II type galaxies, 361 are Galactic sources and 139 are unidentified sources. Objects in the 105 month *Swift*-BAT catalogue are identified together with their optical counterparts by searching the NED and SIMBAD databases and archival X-ray data (e.g., *Swift-XRT*, *Chandra*, *ASCA*, *ROSAT*, *XMM-Newton*, and *NuSTAR*).

2. The *INTEGRAL* IBIS survey hard X-ray catalogue[16], (Bird et al., 2016)[17] consists of 939 sources detected above a 4.5 σ significance threshold in energy band 17 – 100, using the (IBIS) hard X-ray telescope (Winkler et al., 2003). The catalogue showed 120 previously undiscovered soft γ-ray emitters.

3. The second release of the fourth *Fermi*-LAT catalog of γ-ray sources[18] (4FGL-DR2, (Ballet et al., 2020), using the Large Area Telescope (LAT) on the *Fermi* Gamma-ray space telescope mission (Atwood et al., 2009), reports 723 new sources, increasing up to 5064 γ-ray sources. The catalogue processed the first 10 years of the data in the energy range between 50 MeV to 1 TeV. The largest class of Galactic sources in the 4FGL-DR2 is constituted by PSRs listing 292 sources, while in the extragalactic sample is dominated by blazars with 2226 identified and/or associated BL Lac objects and Flat spectrum radio quasars, and 1517 additional blazar candidates of uncertain type.

4.3.3 *Multifrequency catalogues for low energy associations*

1. The current 5^{th} edition of Roma-BZCAT catalogue of blazars based on multi-frequency surveys and extensive review of literature[19] (Massaro et

15 https://heasarc.gsfc.nasa.gov/W3Browse/swift/swbat105m.html
16 https://heasarc.gsfc.nasa.gov/W3Browse/all/ibiscat.html
17 https://heasarc.gsfc.nasa.gov/W3Browse/integral/intibisass.html
18 https://heasarc.gsfc.nasa.gov/W3Browse/fermi/fermilpsc.html
19 http://www.ssdc.asi.it/bzcat

al., 2015a) lists coordinates and multifrequency data for 3561 sources which are either confirmed blazars or sources exhibiting blazar-like behaviour. All sources included in the Roma-BZCAT are detected at radio frequencies. According to the Unified AGN model (Antonucci, 1993; Urry and Padovani, 1995) blazars are AGNs whose jet happens to be closely aligned with our line of sight, exhibiting strong variations, apparent superluminal motion and emission extending across all electromagnetic spectrum.

2. The Turin-SyCAT (Peña-Herazo et al., 2022) multifrequency catalogue of Seyfert galaxies was built using optical, infrared and radio selection criteria. Seyfert galaxies are AGNs, which are distinguished as a type 1 and type 2 based on the observer's angle (Antonucci and Miller, 1985). All objects included in its 1^{st} release have an optical spectroscopic classification, allowing us to establish precisely their redshifts and class. The catalogue presents 351 Seyfert galaxies, out of which 233 are type 1 and 118 are type 2. The analysis presented here, the 2^{nd} release of the Turin-SyCAT, increased their number substantially by 80% to 633 Seyfert galaxies. Details can be find in Section 5.

4.4 CLASSIFICATION

For classifying the sources considered in the presented analysis, we adopted the following step-by-step analysis, as shown in Figure 4.1 and according to the criteria outlined below.

4.4.1 *Scheme*

We start inspecting the 3PBC (Cusumano et al., 2010) catalogue. If a 3PBC source has an assigned counterpart we just adopted the multifrequency criteria reported below in this section to classify it. Then, for sources belonging to the extragalactic population, we also verified its redshift estimate. In particular, for all extragalactic sources, being the main focus of the current analysis, the presence of the optical spectrum or a description of it published in the literature is mandatory to consider it as *classified*.

For all sources lacking an assigned low energy counterpart in the original 3PBC, thus being unassociated, we perform the cross-matching analysis with all catalogues reported in section 4.3 and we also checked updated information in NED and SIMBAD databases, if any. If no reliable X-ray counterpart, is found within the BAT positional uncertainty region, we flagged the 3PBC source as *unidentified*. On the other hand, if a potential counterpart is found, as in the previous step, we adopted the multifrequency criteria to classify it and eventually provide a redshift estimate, and, when successful, the associated source is indicated as *classified*.

Moreover, all associated sources that do not have optical spectrum available for their low energy counterpart and/or lack relevant information to determine their classification were labelled as *unclassified*.

All classified sources were then split into two main samples distinguishing between Galactic and extragalactic populations. Populations of the main classification categories are visualised in Figure 4.2 and compared in Figure 4.5.

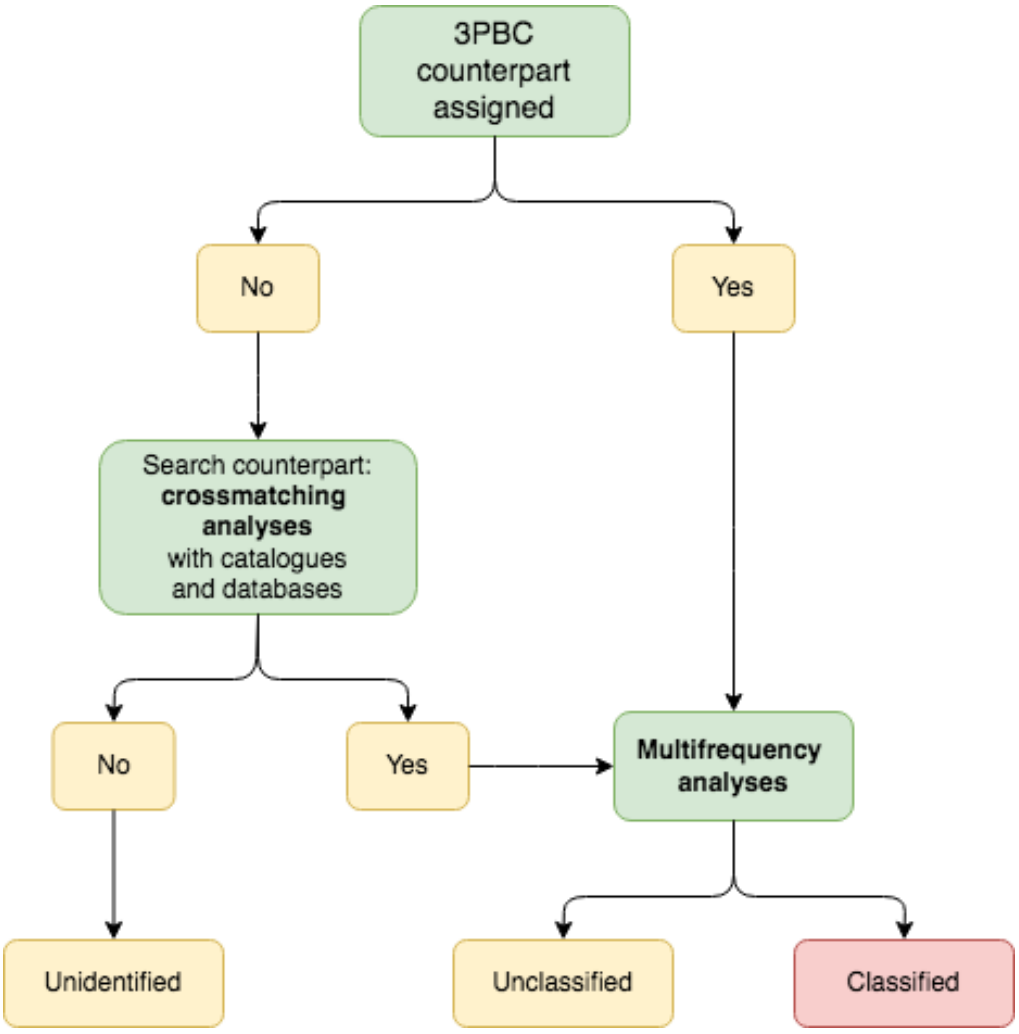

Figure 4.1: The decision tree adopted in our analysis to distinguish between classified, unclassified and unidentified hard X-ray sources (see Section 3.1 for a complete description)

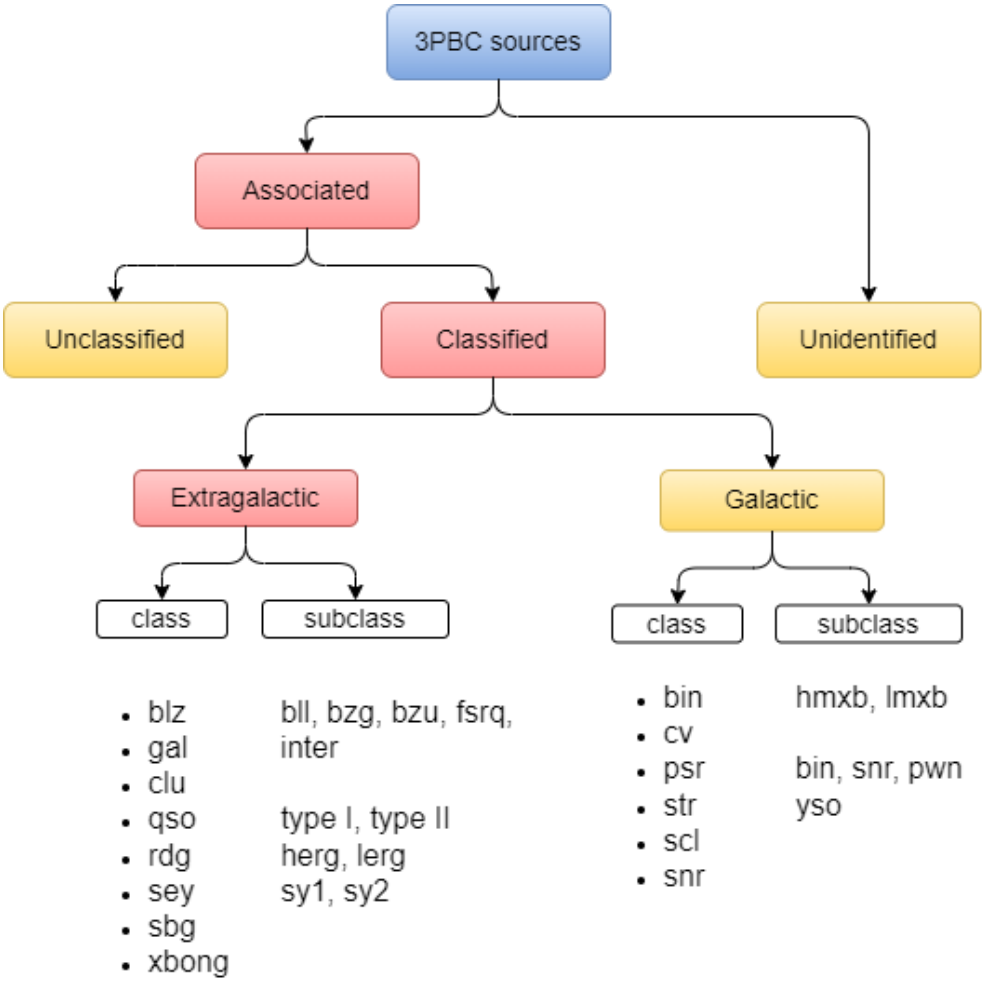

Figure 4.2: Flowchart of the classification classes. Red cells represent categories leading to extragalactic sources, yellow are categories ending in non-extragalactic sources. Details about the numbers of sources in each class and subclass are described in Tab. 16 and Tab. 17. Note that classes and subclasses with only a few members are not included for clarity reasons.

We identified 9 classes and a few sub-classes for both the extragalactic and the Galactic sources discussed in detail in the following subsections, see Table 16 and Table 17.

4.4.2 *Criteria, classes & distributions*

4.4.2.1 *Extragalactic sources*

The largest fraction of sources identified in the extragalactic hard X-ray sky belongs mainly to the two classes of Seyfert galaxies and blazars (Oh et al., 2018, Paliya et al., 2019, Ajello et al., 2009), of which the latter account for 10% - 20% of the entire survey population (Diana et al., 2022). Thus, given the possibility to use both the Roma-BZCAT and the Turin-SyCAT (Peña-Herazo et al., 2022), built on the basis of multifrequency criteria, for all 3PBC classified sources that belongs to these two catalogues we adopted the same classification

reported therein. Moreover, we also used their classification schemes to identify new blazars and Seyfert galaxies, to help future releases.

Blazars (class symbol: *blz*) are the largest known population of γ-ray sources (Massaro et al., 2012a), dominated by non-thermal radiation over the whole electromagnetic spectrum (Massaro et al., 2009; Urry and Padovani, 1995). Their observational features also include high and variable polarization, superluminal motions, very high observed luminosities coupled with a flat radio spectrum (Healey et al., 2007; Hovatta et al., 2012) peculiar infrared colours (D'Abrusco et al., 2012a; Massaro et al., 2011) and a rapid variability from the radio to X-ray bands with weak or absent emission lines (Stickel et al., 1991). Blandford and Rees, 1978 suggested that radiation of blazars could be interpreted as arising from a relativistic jet closely aligned with the line of sight. Blazars were thus classified into 4 categories: BL Lac objects (subclass symbol: *bll*), with featureless optical spectra or presenting only relatively weak and narrow emission lines mainly due to their host galaxies. BL Lacs with their optical-UV spectral energy distribution dominated by the emission of their host galaxy have a subclass symbol: *bzg*. Flat spectrum radio quasars (subclass symbol: *fsrq*) show typical broad emission lines over a blue continuum. Sources exhibiting blazar-like broad-band features, but lacking and optical spectroscopic classification are classified as blazars of uncertain type (subclass symbol: *bzu*). According to the nomenclature of the Roma-BZCAT BL Lacs and FSRQs are labelled as BZBs and BZQs, respectively, while, to avoid confusion here they are marked with the classification symbols *bll* and *fsrq*. This choice was adopted because, given the recent optical spectroscopic campaigns devoted to the search for γ-ray blazars (Landoni et al., 2015; Massaro et al., 2014; Peña-Herazo et al., 2017, 2019) a few more blazars, not yet listed in the Roma-BZCAT, were found as low energy counterparts of 3PBC sources and thus, to avoid confusion, we did not use the Roma-BZCAT nomenclature. No further BZGs and/or BZUs were discovered in our analysis and thus no different classification symbols with respect to those of the Roma-BZCAT were used in these cases.

Names for blazar-like counterparts of 3PBC sources were collected from the Roma-BZCAT if the source is listed therein otherwise in the final table the name reported in one of the major radio surveys as NVSS (Condon et al., 1998) and/or SUMSS (Bock et al., 1999; Mauch et al., 2003), taken from the NED database.

Seyfert galaxies (class symbol: *sey*) were originally defined mainly by their morphology (Seyfert, 1943) as galaxies with high surface brightness nuclei. Nowadays, they are identified spectroscopically as (mostly spiral) galaxies with strong, highly ionized emission lines. Seyfert galaxies come in two flavours distinguished by the presence (or absence) of broad lines emission in their optical spectra (Khachikian and Weedman, 1974; Khachikian and Weedman, 1971). Type 1 Seyfert galaxies (subclass symbol: *sy1*) have both narrow and broad emission lines superimposed to their optical continuum. The former lines originate from a low density ionised gas with density ranging between $\sim 10^3$ and 10^6 cm^{-3} and line widths corresponding to velocities of several hundred kilometres per second (e.g. (Vaona et al., 2012)), while broad lines are located only in permitted transitions, correspondent to electron densities of $\sim 10^9$ cm^{-3} and velocities of 10^4 km s^{-1} (e.g. Kollatschny and Zetzl, 2013). Type 2 Seyfert galaxies (subclass symbol: *sy2*) show only narrow lines in their optical spectra (e.g. Capetti et al., 1999; Miyaji et al., 1992; Weedman, 1977).

To classify Seyfert galaxies we adopted all the same criteria reported in the Turin-SyCAT (Peña-Herazo et al., 2022) in terms of (i) presence of the optical spectrum in the literature, (ii) radio, infrared and optical luminosities, (iii) radio morphology. This was chosen because we include the new Seyfert galaxies discovered here in the 2^{nd} release of the Turin-SyCAT as described in the following sections.

Names for Seyfert-like counterparts of 3PBC sources were collected from the 1^{st} edition of the Turin-SyCAT if the source is listed therein otherwise, are reported in the main table with a NED name taken mainly out of one of the following catalogues: 1RXS (Voges et al., 1999), 2MASSS (Skrutskie et al., 2006), 2MASX (Jarrett et al., 2000) and BAT105 (Oh et al., 2018). All Seyfert galaxies were then renamed according to the Turin-SyCAT nomenclature.

All extragalactic sources that did not fall into the blazar and Seyfert classes mainly belong to other two major classes: quasars and radio galaxies.

Quasars (QSOs), (class symbol: *qso*) are AGNs with bolometric luminosities above $\sim 10^{40}$ erg s^{-1}. They have broad spectral energy distribution and are emitting from radio up to hard X-ray energies, having variable flux densities almost at all frequencies, mid-IR emission due to the dusty torus and broad emission lines superimposed to an optical blue continuum Schmidt, 1969. For this extragalactic source class, we also distinguished type 1 and type 2 QSOs on the basis of the presence of broad emission lines in their optical spectra according to the same criteria adopted for the Seyfert galaxies (Khachikian and Weedman, 1974). Then to distinguish a Seyfert galaxy from a QSO we also considered the same thresholds used to create the Turin-SyCAT (Peña-Herazo et al., 2022), indicating QSOs as sources with both (i) radio luminosity above 10^{40} erg s^{-1} and (ii) mid-IR luminosity estimate at 3.4μ m above $10^{11} L_{\odot}$. Names for the QSOs counterparts of 3PBC sources were collected from is reported in the final table Tab. 18 and Tab. 19 with a NED name taken mainly from the following catalogues: 1RXS (Voges et al., 1999), 2MASSS (Skrutskie et al., 2006), 2MASX (Jarrett et al., 2000) and 1SXPS (Evans et al., 2014a).

Radio galaxies (RDGs), (class symbol: *rdg*) are radio-loud AGNs whose radio emission is at least 100 times that of normal elliptical galaxies and extends beyond tens of kpc scale(Moffet, 1966; Urry and Padovani, 1995), thus being neatly distinct from the Seyfert galaxies. On the other hand, to distinguish between QSO and RDG we adopted a radio morphological criterion where the latter ones clearly present diffuse radio emission at a large scale when radio maps are available to check it. We used the same criteria and classification scheme recently adopted by (Capetti et al., 2017a,b). If the source was not listed with those names, we took the NED name mainly from 3C (Spinrad et al., 1985), 4C (Gower et al., 1967; Pilkington and Scott, 1965) or 7C (Hales et al., 2007) catalogues.

We firstly classified RDGs on the basis of their radio morphologies at 1.4 GHz distinguishing between classical FR I and FR II sources (Fanaroff and Riley, 1974). On the other hand, we also considered the two subclasses of radio galaxies defined on the basis of their optical emission lines, distinguishing between high excitation radio galaxies (HERGs), (subclass symbol: *herg*) and low excitation radio galaxies (LERGs), (subclass symbol: *lerg*) (Hine and Longair, 1979). HERGs are almost always FRIIs, while LERGs can be either FRIs or FRIIs (Buttiglione et al., 2010).

We also considered as extragalactic sources of hard X-rays the: galaxy clusters (class symbol: *clu*), the largest gravity bounded structures in the Universe, composed primarily of dark matter, highly ionised and extremely hot intra-cluster gas of low density and galaxies (Giodini et al., 2009). Their X-ray emission is mainly due to bremsstrahlung radiation of relatively hot particles in their intra-cluster medium in the soft X-rays (i.e., between 0.5 and 10 keV Nevalainen et al., 2003), although a tail of this emission is also detectable at higher energies (Ajello et al., 2010). Since it is well known that some galaxy clusters were also detected by the BAT instrument on board of SWIFT (Ajello et al., 2010) we reported 3PBC sources associated with them mainly when the cross-match with the Abell catalogue indicated the possible presence of a galaxy cluster within the hard X-ray positional uncertainty region.

Finally, we highlight that a handful of extragalactic sources, not belonging to the five major classes listed above, fall into the following categories, being classified as starburst galaxies (class symbol: *sbg*), (Searle et al., 1973; Weedman et al., 1981), galaxies forming stars at unusually fast rates (10^3 times faster than in an average galaxy), X-ray bright optically normal galaxies (class symbol: *xbong*), which are normal galaxies, not hosting an AGN, but having substantial X-ray luminosity (Comastri et al., 2002; Elvis et al., 1981; Yuan and Narayan, 2004), low-ionization nuclear emission-line region galaxies (class symbol: *liner*) (Singh et al., 2013) and normal galaxies (class symbol: *gal*), the latter not hosting an AGN but in a few cases interacting with nearby companions. Names of the 3PBC counterparts for those sources were collected mainly from 2MASX (Jarrett et al., 2000) and 2MASS (Skrutskie et al., 2006) catalogues.

We list a preview of the first 10 sources included in the main table, our revised version of the 3PBC catalogue in Tab. 18 and Tab. 19 where we provide 3PBC catalogue name, coordinates, counterpart name, counterpart coordinates, spectroscopic redshifts, the classification in our class and subclass system and the WISE counterpart name. We show examples of spectra of a few objects in Figure4.3 and Figure4.3.

4.4.2.2 *Galactic sources*

In our Milky Way, most of the sources emitting in the hard X-rays are X-ray binaries (Grimm et al., 2002), while the second dominant class of hard X-ray sources are the cataclysmic variables (Revnivtsev et al., 2008).

X-ray binaries (BINs) (class symbol: *bin*) are systems of double stars containing compact stellar remnants, such as neutron stars, pulsars or black holes and a normal star which can range a variety of masses (e.g. Charles and Coe, 2003; Knigge et al., 2011). The compact stellar remnant accretes material from its companion, creating continual or transient X-ray emission. X-ray binaries are classified based on their companion star distinguishing between low mass X-ray binaries (subclass symbol: *lmxb*) having the companion star of mass $\lesssim 1$ $M_\odot$ and high mass X-ray binaries (subclass symbol: *hmxb*) usually accompanied by a star of mass $\gtrsim 10$ $M_\odot$, where the accretion happens directly from a stellar wind of the companion star. Names for the BINs counterparts of 3PBC sources were collected mainly from the following catalogues: IGR (Bird et al., 2004), 1H, SWIFT (Ajello et al., 2010) and (RX+XTE+SAX) (Bade et al., 1992; Frontera et al., 2009; Voges et al., 1999).

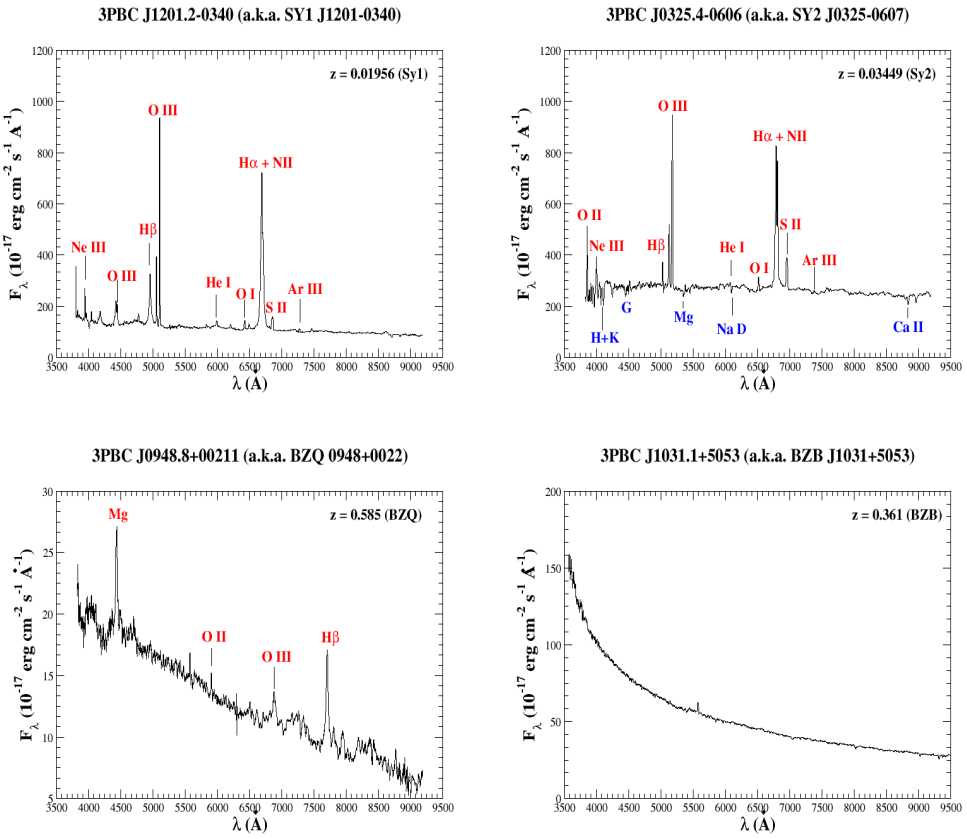

Figure 4.3: Example images of the spectra of objects in our revised analysis of the 3PBC. Top left: Type I Seyfert galaxy 3PBCJ1201.2-0340. Top right: Type II Seyfert galaxy 3PBCJ0325.4-0606. Middle left: Flat spectrum radio quasar (fsrq) 3PBCJ0948.8+0021. Middle right: BL Lac Blazar 3PBCJ1031.1+5053 (BZB in Roma-BZCAT (Massaro et al., 2009) nomenclature). The redshift of the BZB is not collected from the spectrum, since it is featurless, but was taken from literature as reported in the Roma-BZCAT.

Cataclysmic variables (CVs), (class symbol: *cv*) are binary systems composed of a main-sequence companion star and a compact stellar remnant which is a white dwarf (WD) (Revnivtsev et al., 2008). The accretion happens almost always via filling the Roche-lobe of the companion star and subsequent formation of an accretion disk around the WD (Warner, 1995). Their X-ray emission can originate from a variety of processes depending on the type of the CV. CVs which do not have strong magnetic fields accrete matter closer to the surface of the WD and produce sporadic eruptions. They are categorised in 3 subclasses, based on their accretion rate and explosions. For 4 sources belonging to the CV class we also indicated if they are symbiotic stars or novae, however, given their relative low number with respect to all CVs identified we did not labelled these as subclasses and we only report the source class. Names for the CVs counterparts of 3PBC sources were collected mainly from the following catalogues: CV (Downes et al., 2005), IGR (Bird et al., 2004), 1RXS (Voges et al., 1999) and 2MASS (Skrutskie et al., 2006).

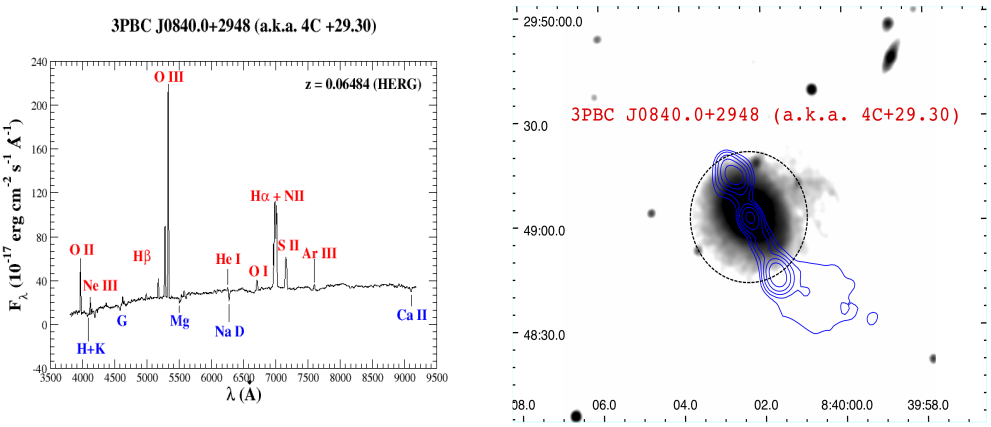

Figure 4.4: Bottom left: High excitation radio galaxy (HERG) 3PBCJ0840.0+2948. Bottom right: The images of the HERG galaxy. Radio contours are drawn at level of 0.0005 mJy/beam increasing by a factor of 2. Here background image is PANSTARRS R band while radio contours are at 3 GHz from the VLASS survey extending beyond the galaxy. The main spectral emission and/or absorption lines are marked at each figure.

The hard X-ray sky is also populated by isolated X-ray pulsars (PSR), (class symbol: *psr*), not being hosted in X-ray binaries. Since they can be indeed hosted in pulsar wind nebulae (subclass symbol: *pwne*) or supernova remnants we highlight the presence of this extended emission around the PSR in the subclass column. On the other hand, if the hard X-ray emission is indeed due to a supernova remnant not hosting a neutron star then we adopted a different class (subclass symbol: *snr*), in these cases, their hard X-ray emission is due to the thermal radiation of plasma heated in shocks, coupled with non-thermal synchrotron radiation (see e.g., Vink, 2012). Names for the PSRs counterparts of 3PBC sources were collected mainly from the ATNF PSR catalogue or other radio surveys.

As occurred for the extragalactic hard X-ray population a handful of sources were also identified belonging to normal stars (class symbol: *str*), (coming with a subclass: *yso* for young stellar objects) and star clusters (class symbol: *scl*). X-ray emission from main sequence stars of masses > 10 $M_{\odot}$ can be due to discrete ionized metal lines in their spectrum. Young stellar objects, protostars and T Tauri stars also exhibit X-ray radiation, predominantly emerging from magnetic coronae accreting material where shocks occur (Güdel and Nazé, 2009). On the other hand, star clusters can appear as an amalgamation of point-like sources and extended X-ray emission. Their point-like component can be produced by hot stars and/or SNR, lasting a few thousand years while their extended component is produced by star cluster wind, formed by the interaction of stellar winds of massive O or B type stars, Wolf-Rayet stars and supernovae explosions (e.g. Cantó et al., 2000, Law and Yusef-Zadeh, 2004, Oskinova, 2005). In addition to them we also reported the classification for one microquasar (class symbol: *mqso*), namely: 3PBC J0804.7-2748. Microquasars are similar to quasars, but on a much smaller case. Their radiation comes from a stellar mass black hole or a neutron star accreting matter from a normal star (Mirabel, 2010). In addition, we report one planetary nebula (class symbol: *pn*): 3PBC J1701.5-4306. Planetary

Table 16: Extragalactic classes and subclasses.

Class symbol	Class number	Subclass symbol	Subclass number
blz	129	bll	30
		bzg	7
		bzu	24
		fsrq	68
gal	10	interacting	3
		-	7
clu	27		
liner	1		
qso	26	type 1	18
		type 2	1
		?	7
rdg	25	herg	21
		lerg	3
		?	1
sey	593	sy1	325
		sy2	268
sbg	5		
xbong	4		

nebulae are the ejected red giant's atmosphere ionized by the leftover star's core, forming at the end of life of stars with initial masses in range ~ 1 to 8 solar masses. Lastly we also labelled the Galactic centre Sgr A* with the symbol: *galcent*.

4.4.2.3 *Sky distributions*

Starting from the total number of 1593 sources listed in the 3PBC catalogue (Cusumano et al., 2010) we found that according to our analysis there are 218 unidentified hard X-ray sources ($\sim$13.7 %) and 199 unclassified sources ($\sim$12.5 %), see Figure 4.5. The classified sources are distinguished into two main groups: Galactic objects including 356 sources ($\sim$22.2 %) and extragalactic objects having 820 sources ($\sim$51.5 %).

We show the sky distribution of 3PBC sources via the Hammer-Aitoff projection for both unclassified and unidentified cases in Figure 4.6 and for classified sources, distinguishing between Galactic and extragalactic ones in Figure 4.7. Given the source distributions for both unidentified and unclassified sources that appear to be quite uniform over the whole sky, we could expect that a large fraction of them could have extragalactic origin. This could imply that the lack of classified counterparts is mainly due to missing follow up spectroscopic observations, thus strengthening the need to complete optical campaigns carried out to

Table 17: Galactic classes and subclasses. Please note that classes: galcent, mqso and pn were ommited due to small member count (1 each).

Class symbol	Class number	Subclass symbol	Subclass number
bin	231	hmxb	117
		lmxb	108
		?	6
cv	87		
psr	21	bin	1
		-	5
		snr	12
		pwn	3
str	12	-	6
		yso	1
		?	1
scl	2		
snr	4		

date (**references**). Then fractions of other classes for extragalactic sources are shown in Figure 4.9 and for galactic classes in Figure 4.8.

4.5 CHARACTERIZING THE EXTRAGALACTIC HARD X-RAY SKY

Our revised analysis of the 3PBC lists 735 extragalactic sources, classified into 9 classes: 129 blazars (*blz*), 10 galaxies (*gal*), 27 galaxy clusters (*clu*), 1 low-ionization nuclear emission-line region galaxy (*liner*), 26 quasars (*qso*), 25 radio galaxies (*rdg*), 593 seyfert galaxies (*sey*), 5 star-burst galaxies (*sbg*) and 1 X-ray bright optically normal galaxy (*xbong*). Table 16 reports those numbers together with the number of sources in their associated subclasses.

The most abundant class of extragalactic sources are Seyfert galaxies (Figure 4.9), while the second largest population of extragalactic sources emitting in the hard X-rays is constituted by blazars. The hard X-ray luminosity, K-corrected, is shown in Figure 4.10 as a function of the redshift with particular emphasis to the two classes of Seyfert galaxies and blazars. We used the measured spectral index reported in the 3PBC for K-correlation computation.

Once we assigned the coordinates of each counterpart we also crossmatched the 3PBC catalogue with the AllWISE survey[20] (Cutri et al., 2021a) and we found that adopting an association radius of 3.3, as typically used in other analyses (D'Abrusco et al., 2019; Massaro et al., 2012b) we found 1279 mid-IR coutnerparts in the 1593 3PBC sources. It is worth noting that associating sources within this angular separation corresponds to a chance probability of having a spurious match lower than $\sim$2 % (Massaro et al., 2015b, 2013a).

20 https://wise2.ipac.caltech.edu/docs/release/allwise/

Table 18: Preview of the first four columns of the main catalog table of our revised 3PBC catalog, showing the first 10 sources. Column description: (1) Name in 3PBC catalogue, (2) right ascension as reported in 3PBC, (3) declination as reported in 3PBC, (4) counterpart name. The last six columns are described in Tab. 19.

3PBC	ra3PBC	dec3PBC	counterpart
3PBC J0000.9-0708	0.228	-7.134	2MASS J00004877-0709115
3PBC J0001.7-7659	0.429	-76.986	2MASX J00014596-7657144
3PBC J0002.5+0322	0.636	3.367	SY1 J0002+0322
3PBC J0002.5+0322	0.853	27.638	2MASX J00032742+2739173
3PBC J0002.5+0322	1.009	70.312	SY2 J0004+7020
3PBC J0006.3+2012	1.584	20.205	SY1 J0006+2013
3PBC J0010.4+1058	2.624	10.976	5BZQ J0010+1058
3PBC J0016.7-2611	4.194	-26.2	
3PBC J0017.4+0519	4.37	5.326	HS 0014+0504
3PBC J0017.8+8135	4.454	81.591	5BZQ J0017+8135

Table 19: Preview of the last six columns of the main catalog table of our revised 3PBC catalog, showing the first 10 sources. Column description: (5) right ascension of the counterpart, (6) declination of the counterpart, (7) redshift, (8) associated class, (9) associated subclass, (10) name in WISE catalog.

ractp	decctp	zctp	class	subclass	WISE
0.203216	-7.153221	0.03748	sey	sy2	J000048.77-070911.6
0.441917	-76.953972	0.05839	sey	sy1	J000146.08-765714.2
0.610046	3.351961	0.025518	sey	sy1	J000226.42+032106.8
0.864283	27.654828	0.03969	sey	sy2	J000327.41+273917.0
1.008228	70.32175	0.096	sey	sy2	J000401.97+701918.3
1.581389	20.202968	0.025785	sey	sy1	J000619.53+201210.6
2.629166	10.974888	0.089100	blz	fsrq	J001031.00+105829.5
0.	0.	0.	uhx		
4.344167	5.352778	0.11	sey	sy1	J001722.71+052111.4
4.28525	81.58561	3.387000	blz	fsrq	J001708.50+813508.1

We also used the counterpart coordinates to carry out a crossmatch between all blazars listed in the 3PBC and those associated in the 4FGL catalogue. There are 92 out of 129 blazars with a *Fermi* counterpart and with known redshift, with 25 of them belonging to the BL Lac class and 52 to that of FSRQs. For all these γ-ray emitting blazars we also found two neat trends/correlations between their hard X-ray and γ-ray emissions as highlighted in Figure 4.11. The first trend is between their hard X-ray and γ-ray fluxes, where a mild correlation is also reported: 0.52 is the measured value for the correlation coefficient for the whole blazar sample. The p-chance for all correlations are below 10^{-5} level of significance due to the high numbers of sources used to compute the correlation coefficients. Then a second trend was indeed found between the photon indices of blazars measured in the 3PBC and in the 4FGL catalogues.

Both trends highlighted for the blazar population emitting in the hard X-rays are expected given the nature of their emission (e.g. Acharyya et al., 2021). For BL Lac objects the steep hard X-ray spectra could be due to emission arising from the tail of their synchrotron (Maraschi et al., 1992) component, and the flat γ-ray spectra are related to the peak of their inverse Compton bump at γ-ray energies (Dermer, 1995; Maraschi et al., 1999; Marscher and Gear, 1985). On the other hand for the FSRQs both the hard X-ray and the γ-ray emission is due to their inverce Compton component peaking in the γ-ray band (Acharyya et al., 2021). Then we also note that even if the broadband spectral energy distributions (SEDs) of BL Lacs is mainly interpreted as due Synchrotron Self Compton emission while that of FSQRs to external Compton radiation (Abdo et al., 2010), relativistic particles responsible for both SED bumps are the same and thus we could expect that fluxes in the hard X-rays and in the γ-rays are, on average, connected (e.g., Wolter et al., 2008).

4.6 SECOND RELEASE OF THE TURIN-SYCAT

We found 282 new Seyfert galaxies resulting from our analysis of the extragalactic hard X-ray sky presented in the previous sections. Including new sources in the Turin-SyCAT 1^{st} release, we list 633 Seyfert galaxies: 351 type 1 and 282 type 2 for the 2^{nd} release of the Turin-SyCAT presented here. We added 118 type 1 and 164 type 2 Seyfert galaxies and we also present here an updated analysis of the infrared - hard X-ray connection including all new sources.

Sources added in the 2^{nd} release of the Turin-SyCAT were selected according to the same procedure as in Peña-Herazo et al. (2022). These strict selection criteria allow us to have a negligible fraction of contaminants since we selected only extragalactic sources with a Seyfert-like optical spectrum and having:

1. a published optical spectrum;

2. a luminosity in Radio lower than $<10^{40}$ erg s^{-1} if a counterpart is listed in the two major radio surveys i.e., NVSS and SUMSS Condon et al., 1998; Mauch et al., 2003, respectively;

3. a counterpart in the AllWISE Source catalog with a mid-IR luminosity at $3.4\mu\,m$ less than 3×10^{11} L. This was mainly adopted to avoid selection of QSOs.

In Figure 4.12 we present the redshift distribution of Turin-SyCAT 2^{nd} release. The source number for both classes drastically drops after z > 0.2 as occurs for those listed in the first release, and the source with the highest redshift is SY2 J0304-3026 at 0.436. We compare the redshift distribution of all Seyfert galaxies (Figure 4.13, only Type 1 Seyfert galaxies (Figure 4.14) and only Type 2 Seyfert galaxies (Figure 4.15) between the 1^{st} release of the Turin-Sycat and it's presented 2^{nd} release.

With respect to the previous Turin-SyCAT 1^{st} release, we modified the name of SY2 J2328+0330 in SY2 J2329+0331 having a WISE counterpart J232903.90+033159.9 and since a new Seyfert type 2 galaxy was associated with its mid-IR counterpart J232846.65+033041.1 thus being named as SY2 J2328+0330.

We list all sources included in the Turin-SyCAT 2^{nd} release in Table 20 where we provide SyCAT name, coordinates, spectroscopic redshifts, WISE counterpart and 3PBC counterpart names, a flag to indicate if the source is also associated in BAT105 catalogue, and a flag to point out those added in this 2^{nd} release.

On the basis of the new Seyfert galaxies discovered here, we revisited the connection between their hard X-ray and the mid-IR emission (Assef et al., 2013). This connection is related to the reprocessed radiation from the dust of all energy absorbed from the optical and UV wavelengths in the central engine of Seyfert galaxies (e.g. Elvis et al., 2009). The high-energy emission measures an intrinsic radiated luminosity above ∼10 keV, while WISE 12 μm and 22 μm is related to the reprocessed radiation from the dust of all energy absorbed from the optical and UV wavelengths.

Mid-IR fluxes show a significant correlation with the hard X-ray fluxes, similar to those highlighted using Seyfert galaxies listed in the Turin-SyCAT 1^{st} release, as shown in Figure 4.16. Comparing integrated fluxes as F_{12} and F_{HX} we found a linear correlation coefficient of 0.54 (correspondent to a slope of 1.09 ± 0.10 given the measured dispersion) for Seyfert 1 and 0.45 (slope of 1.20 ± 0.16) for Seyfert 2 galaxies, respectively. This is in agreement with results presented on the statistical analysis of Seyfert galaxies listed in the Turin-SyCAT 2^{nd} release where we measured a correlation coefficient of 0.57, with a slope of 1.02 ± 0.10 and a coefficient of 0.52 (slope of 0.93 ± 0.16), for Seyfert 1 and 2 galaxies, respectively. On the other hand, also comparing mid-IR at lower frequencies with the hard X-ray flux (i.e., F_{22} vs F_{HX}) we found a correlation coefficient of 0.55 (with a slope of 1.11 ± 0.10) for type 1 Seyfert galaxies and 0.46 (slope of 1.08 ± 0.17) for type 2 Seyfert galaxies.

Considering both classes together, since they show similar mid-IR to hard X-ray ratios, we found a correlation coefficient of 0.51 and a slope of 1.1 ± 0.08 for both hard X-ray flux F_{HX} correlation with F_{12} and F_{22}, all in agreement with previous results based on the Turin-SyCAT 1^{st} release.

We also cross-matched sources listed in Turin-SyCAT 2^{nd} release with the Point Source catalogue of the Infrared Astronomical Satellite (*IRAS*)[21], using the positional uncertainties reported therein. We obtained 67 new matches for a total of 216 Seyfert galaxies with an IRAS counterpart, being 89 type 1 and 127 type 2 counterparts at both 60 μm and 100 μm, respectively. Then, as occurred in our previous analysis (Peña-Herazo et al., 2022), we also tested possible trends between the infrared fluxes, at 60 μm and at 100 μm, and the hard X-ray one.

21 https://heasarc.gsfc.nasa.gov/W3Browse/iras/iraspsc.html

We found no clear correlation as evident in Figure 4.17 and again these results are in agreement with our previous findings based on Turin-SyCAT 1^{st} release. Moreover, we did not expect any correlation while inspecting trends between infrared and hard X-ray fluxes since the cold dust, mainly responsible for the emission at 60 μm and 100 μm is not significantly affected by the behaviour of the central AGN but it is mainly linked to the star formation occurring in Seyfert galaxies (Rodriguez Espinosa et al., 1987).

The strict multi-frequency selection criteria that we used to select Turin-SyCAT sources allowed us to minimise the possible contamination of other source classes, thus strengthening our results. Thus we remind that we visually inspected all Turin-SyCAT galaxies' optical spectra, allowing us to measure their redshifts and establish their proper optical classification.

4.7 SUMMARY, CONCLUSIONS AND FUTURE PERSPECTIVES

The CXB is nowadays established to constitute mainly of an integrated emission of discrete sources, primarily arising from AGNs (Gilli et al., 2007). Having a precise knowledge of the population and properties of various types of AGNs is thus crucial to improve our knowledge of the CXB.

In this work, we focus on the analysis of the 3PBC catalogue (Cusumano et al., 2010), which provides data in three energy bands, namely: 15 – 30 keV, 15 – 70 keV, 15 – 150 keV, in particular focusing on extragalactic source population, aiming also at discovering new Seyfert galaxies that can be included in the presented Turin-SyCAT 2^{nd} release. The 3PBC provides 1593 sources above signal to noise ration 3.8, approximately 57 % sources appear to have a clear extragalactic origin while 19 % belong to our Milky Way and the remaining 24 % are yet unknown. Results of our multifrequency investigation are also based on those recently found for the 105 month *Swift*-BAT catalogue (Oh et al., 2018) and the *INTEGRAL* IBIS hard X-ray survey in energy range 17-100 keV (Bird et al., 2016). The 3PBC catalogue presented 861 extragalactic sources, 269 galactic, 230 uncertain and 233 unclassified according to our nomenclature. In our revision of the 3PBC catalogue, we reduced the uncertain sources to 55 (reduction by $\sim 76\%$) and unclassified to 196 (reduction by $\sim 16\%$).

Thanks to our analysis we (i) developed a multifrequency classification scheme for hard X-ray sources, that can be later adopted also to investigate different high energy surveys, (ii) investigate the main properties of sources populating the extragalactic hard X-ray sky and finally, extract other Seyfert galaxies now included in the 2^{nd} release of the Turin-SyCAT catalogue presented here.

We started with the 1593 sources of the 3PBC catalogue, comparing them with various other catalogues mentioned in the paper and adopting the following classification scheme criteria. Firstly, we checked if the 3BC source has an assigned counterpart, if not, we performed the multifrequency crossmatching analyses across the available literature to search for counterparts. Sources without counterparts were assigned with *unidentified* category. Those which were found with counterparts, together with sources already having a counterpart in the 3PBC catalogue, were further inspected with multifrequency analyses. Sources lacking

Table 20: All Turin-SyCAT sources with names of their counterparts (First ten rows, full table available in the on-line material)

IDv2 (1)	IDv1 (2)	SyCAT (3)	R.A. (4)	Dec. (5)	z (6)	WISE (7)	3PBC (8)	F_{HX} (9)	in 3PBC (10)	in BAT105 (11)	SyCAT v2 (12)
1		SY2 J0000-0709	0.203216	-7.153221	0.03748	J000048.77-070911.6	3PBC J0000.9-0708	1.25E-11 ± 1.4E-12	true	true	true
2		SY1 J0001-7657	0.441917	-76.953972	0.05839	J000146.08-765714.2	3PBC J0001.7-7659	1.09-11 ± 1.5E-12	true	true	true
3	1	SY1 J0002+0322	0.6102917	3.352	0.025518	J000226.41+032107.0	3PBC J0002.5+0322	1.39E-11 ± 1.9E-12	true	true	false
4		SY2 J0003+2739	0.864283	27.654828	0.03969	J000327.41+273917.0	3PBC J0003.4+2738	1.82E-11 ± 2.6E-12	true	true	true
5	2	SY2 J0004+7020	1.00817	70.32175	0.096	J000401.97+701918.2	3PBC J0004.0+7018	1.1E-11 ± 1.30E-12	true	true	false
6	3	SY1 J0006+2013	1.58133	20.202917	0.025785	J000619.53+201210.6	3PBC J0006.3+2012	1.76E-11 ± 1.40E-12	true	true	false
7		SY1 J0017+0521	4.344167	5.352778	0.11	J001722.71+052111.4	3PBC J0017.4+0519	8.69E-12 ± 1.5E-12	true		true
8		SY2 J0021-1910	5.281417	-19.168222	0.09558	J002107.53-191005.4	3PBC J0021.1-1909	1.76E-11 ± 1.60E-12	true	true	true
9	4	SY2 J0025+6821	6.38542	68.3622	0.012	J002532.37+682144.9	3PBC J0025.5+6822	1.79E-11 ± 1.40E-12	true	true	false
10	5	SY1 J0025-1859	6.4265	-19.002917	0.24622	J002542.34-190010.1	3PBC J0025.6-1859	1.05E-11 ± 1.60E-12	true	false	false

Column description: (1) Unique catalog identified (ID) from SyCAT 2^{nd} version; (2) Unique catalog identified (ID) from SyCAT 1^{st} version; (3) Right Ascencion J2000; (4) Declination J2000; (5) Redshift; (6) name in WISE; (7) name in 3PBC; (8) Flux; (9) name in BAT105; (10) name in IBIS4CAT.

sufficient information to assign their class were put to *unclassified* category, the rest to *classified* category. We further distinguish the classified sources to Galactic and extragalactic and we purely focus on the extragalactic sources in this work.

Results obtained from our analysis can be outlined as follows.

1. The final revised 3PBC catalogue we present in this study lists 1176 classified, 820 extragalactic and 356 Galactic ones, 218 unidentified and 199 unclassified sources, respectively. Compared to the original 3PBC, which has 233 unidentified and 300 unclassified sources, we decreased this fraction from ∼33 % (533 sources) to ∼26 % (417 sources).

2. The hard X-ray sky is mainly populated by nearby AGNs, where the two largest known populations of associated AGNs are: Seyfert galaxies (∼ 79%) and blazars (∼ 17%).

3. We report the trends between the hard X-ray and the gamma-ray emissions of those blazars that are also listed in the 4FGL as expected by the models widely adopted to explain their broadband SED.

4. In the presented 2^{nd} release of the Turin-SyCAT we list 633 Seyfert galaxies, 282 new ones added here thus correspondent to increase their number by ∼80 % with respect to its 1^{st} release.

5. We updated the statistical analysis carried out comparing the hard X-ray and the IR emissions of Seyfert galaxies. All results obtained are in agreement with those previously found even if now the analysis appears more robust since it was performed with a sample of Seyfert galaxies increased by ∼80 % with respect to the 1^{st} release of the Turin-SyCAT.

Finally, we already checked the presence of SWIFT observations carried out using the X-ray telescope on board for the sample of unidentified hard X-ray sources and we found that more than 95 % of them have at least a few ksec exposure time available. Thus the next step of the present analysis will be to search for the potential soft X-ray counterpart of these 3PBC unidentified sources to obtain their precise position necessary to carry out optical spectroscopic campaigns aimed at identifying the whole sky seen between 15 and 150 keV.

Acknowledgements

M. K. and N. W. are supported by the GACR grant 21-13491X. E. B. acknowledges NASA grant 80NSSC21K0653. F. M. wishes to thank Dr. G. Cusumano for introducing him to the Palermo BAT Catalog project. We would like to thank A. Capetti for his work done on the 1^{st} version of the Turin-SyCAT, which was relevant for this work. This investigation is supported by the National Aeronautics and Space Administration (NASA) grants GO0-21110X, GO1-22087X and GO1-22112A. This research has made use of the NASA/IPAC Infrared Science Archive, which is funded by the National Aeronautics and Space Administration and operated by the California Institute of Technology. Funding for the Sloan Digital Sky Survey IV has been provided by the Alfred P. Sloan Foundation, the U.S. Department of Energy Office of Science, and the Participating

Table 21: Table of catalogues used in the cross-matching analysis.

Acronym	Catalogue Name	Reference
4FGL-DR2	The second release of the fourth *Fermi*-LAT catalog of γ-ray sources	1
3PBC	The 3^{rd} *Swift*-Bat Hard X-ray catalog	2
BAT105	The 105-month Palermo *Swift*BAT catalog	3
INTEGRAL	The IBIS soft gamma-ray sky after 1000 *INTEGRAL* orbits	4
Homa-BZCAT	5^{th} edition of Roma-BZCAT catalog of blazars	5
3CR	The Revised Third Cambridge catalog	6
4C	The Fourth Cambridge Survey	7, 8
SyCAT	The Turin-SyCAT catalog	9
CVcat	The Catalog and Atlas of Cataclysmic Variables	10
SNRcat	The Catalog of Galactic Supernovae Remnants	11
hmxb	The 4^{th} edition of the catalog of High mass X-ray binaries in the Galaxy	12
lmxb	The 4^{th} edition of the catalog of Low mass X-ray binaries in the Galaxy and Magellanic Clouds	13
Rlmxb	The 7^{th} edition of the catalog of cataclysmic binaries, low mass X-ray binaries and related objects	14
ANTF	The Australian Telescope National Facility Pulsar Catalog	15
Abellcat	Abell catalog of rich galaxy clusters	16

(1) Ballet et al. (2020); (2) Cusumano et al. (2010); (3) Oh et al. (2018); (4) Bird et al. (2016); (5) Massaro et al. (2015a); (6) Spinrad et al. (1985); (7) Pilkington and Scott (1965); (8) Gower et al. (1967); (9) Peña-Herazo et al. (2022); (10) Downes et al. (2005); (11) Green (2017); (12) Liu et al. (2006); (13) Liu et al. (2007); (14) Ritter and Kolb (2003); (15) Manchester et al. (2005); (16) Abell et al. (1989).

Institutions. SDSS-IV acknowledges support and resources from the Center for High Performance Computing at the University of Utah. The SDSS website is www.sdss.org. SDSS-IV is managed by the Astrophysical Research Consortium for the Participating Institutions of the SDSS Collaboration including the Brazilian Participation Group, the Carnegie Institution for Science, Carnegie Mellon University, Center for Astrophysics | Harvard & Smithsonian, the Chilean Participation Group, the French Participation Group, Instituto de Astrofísica de Canarias, The Johns Hopkins University, Kavli Institute for the Physics and Mathematics of the Universe (IPMU) / University of Tokyo, the Korean Participation Group, Lawrence Berkeley National Laboratory, Leibniz Institut für Astrophysik Potsdam (AIP), Max-Planck-Institut für Astronomie (MPIA Heidelberg),Max-Planck-Institut für Astrophysik (MPA Garching), Max-Planck-Institut für Extraterrestrische Physik (MPE), National Astronomical Observatories of China, New Mexico State University, New York University, University of Notre Dame, Observatário Nacional / MCTI, The Ohio State University, Pennsylvania State University, Shanghai Astronomical Observatory, United Kingdom Participation Group, Universidad Nacional Autónoma de México, University of Arizona, University of Colorado Boulder, University of Oxford, University of Portsmouth, University of Utah, University of Virginia, University of Washington, University of Wisconsin, Vanderbilt University, and Yale University. The Pan-STARRS1 Surveys (PS1) and the PS1 public science archive have been made possible

through contributions by the Institute for Astronomy, the University of Hawaii, the Pan-STARRS Project Office, the Max-Planck Society and its participating institutes, the Max Planck Institute for Astronomy, Heidelberg and the Max Planck Institute for Extraterrestrial Physics, Garching, The Johns Hopkins University, Durham University, the University of Edinburgh, the Queen's University Belfast, the Harvard-Smithsonian Center for Astrophysics, the Las Cumbres Observatory Global Telescope Network Incorporated, the National Central University of Taiwan, the Space Telescope Science Institute, the National Aeronautics and Space Administration under Grant No. NNX08AR22G issued through the Planetary Science Division of the NASA Science Mission Directorate, the National Science Foundation Grant No. AST–1238877, the University of Maryland, Eotvos Lorand University (ELTE), the Los Alamos National Laboratory, and the Gordon and Betty Moore Foundation. This publication makes use of data prod- ucts from the Wide-field Infrared Survey Explorer, which is a joint project of the University of California, Los Angeles, and the Jet Propulsion Laboratory/California Institute of Technology, funded by the National Aeronautics and Space Administration. TOPCAT and STILTS astronomical software (Taylor, 2005) were used for the preparation and manipulation of the tabular data and the images.

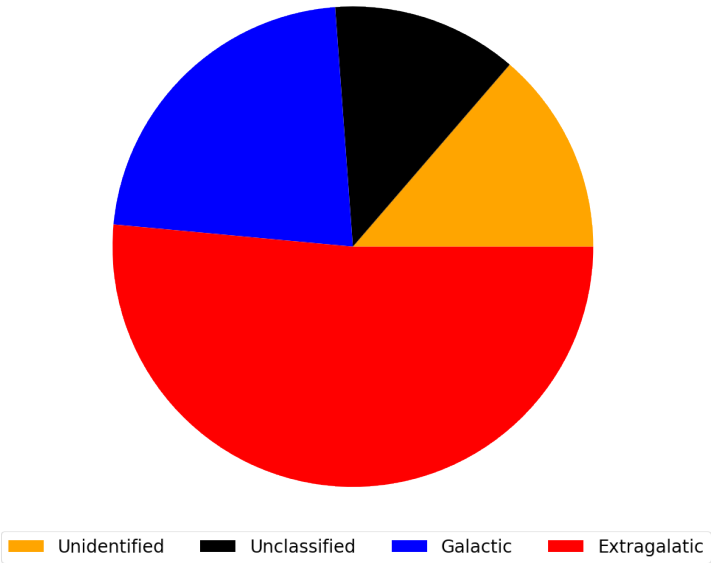

Figure 4.5: Pie-chart showing the fractions of the main classification categories derived thanks to our revised analysis of the 3PBC. The representation of the categories is following: Unidentified (218), Unclassified (199), Galactic (356), Extragalactic (820).

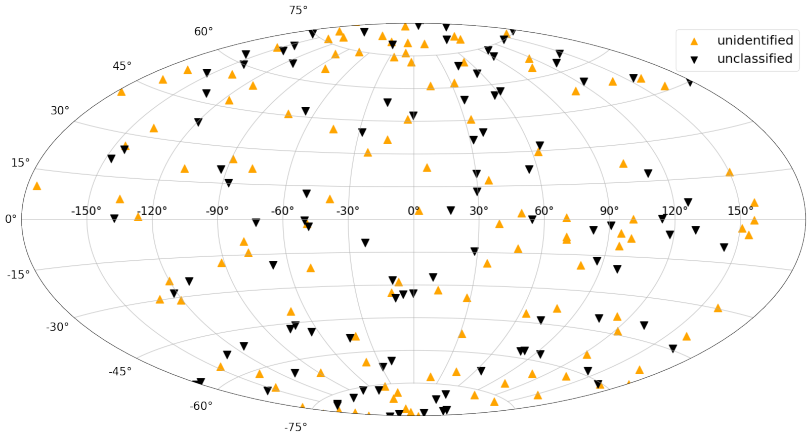

Figure 4.6: Hammer-Aitoff projection derived thanks to our revised analysis of the 3PBC, showing unidentified and unclassified sources.

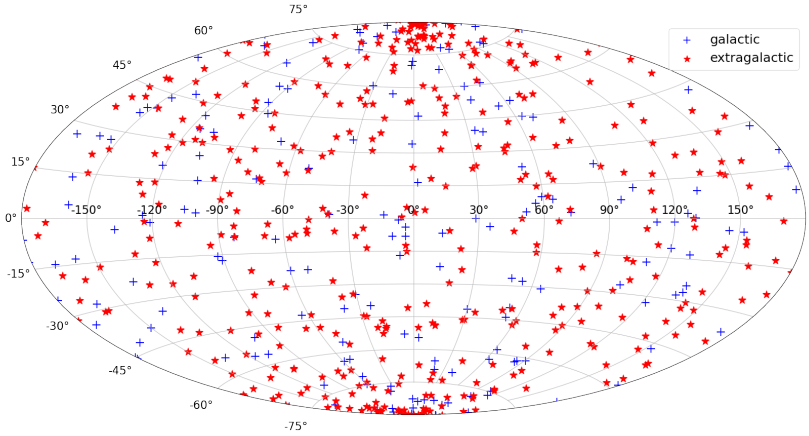

Figure 4.7: Hammer-Aitoff projection based on our revised analysis of the 3PBC, showing galactic and extragalactic sources.

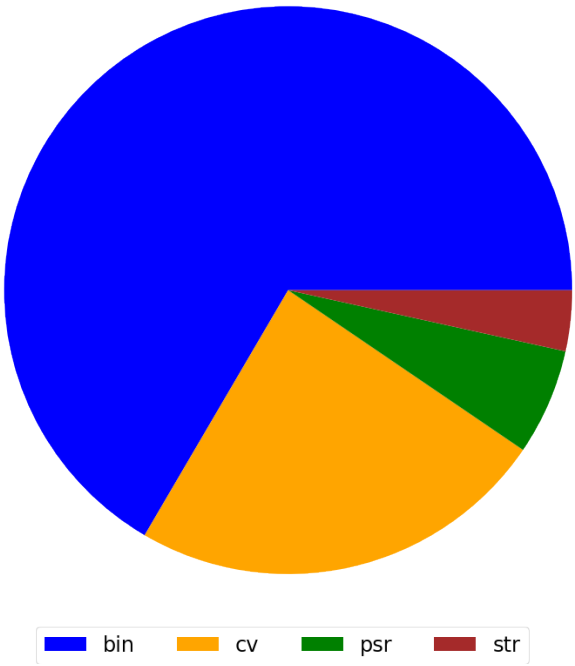

Figure 4.8: Pie-chart showing the fractions of galactic classes derived thanks to our revised analysis of the 3PBC. The classes are represented as follows: bin (231), cv (87), psr (21), str (8). Note, that classes galcent, mqso and pn are omitted for a small contribution (1 member each).

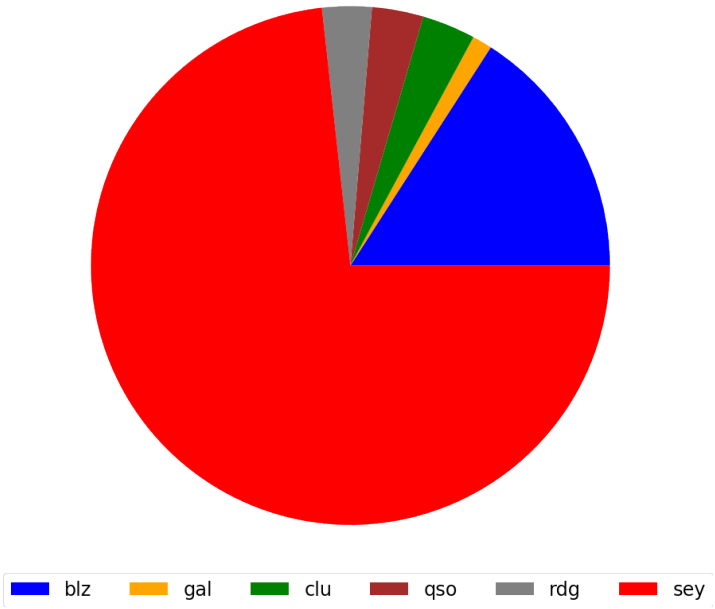

Figure 4.9: Pie-chart showing the fractions of extragalactic classes derived thanks to our revised analysis of the 3PBC. Individual classes have following representation: blz (129), gal (10), clu (27), qso (26), rdg (25), sey (593). Note that the liner subclass is omitted for a small contribution (1 member).

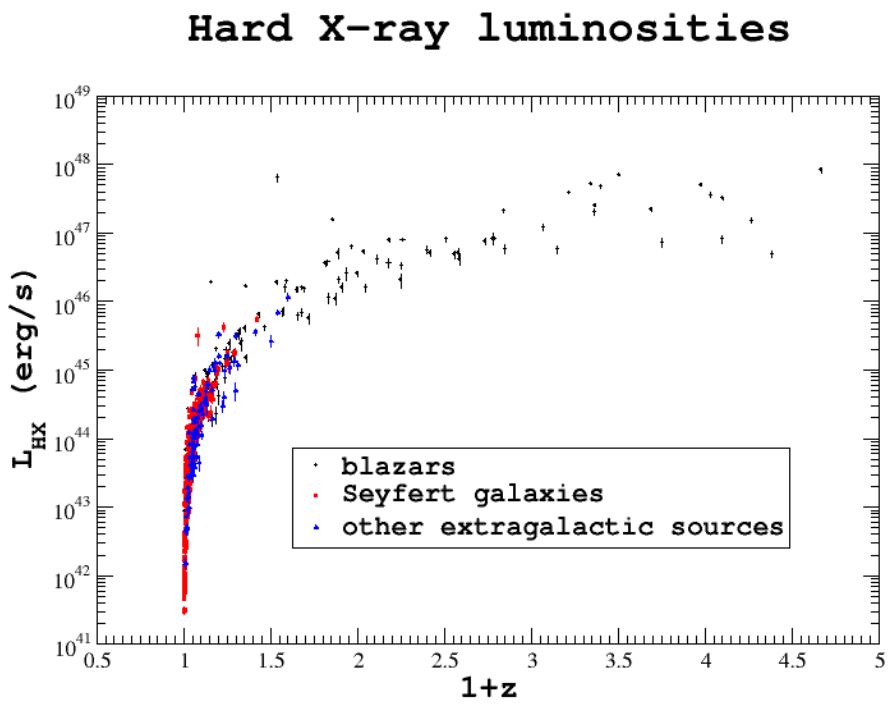

Figure 4.10: K-corrected hard X-ray luminosity computed using the 10-150 keV flux reported in the 3PBC. Given the high numbers of sources used to compute the correlation coefficients, the p-chance for all correlations are below 10^{-5} level of significance. Seyfert galaxies are reported as red squares, blazars as blue circles while all other extragalactic sources are marked as green triangles.

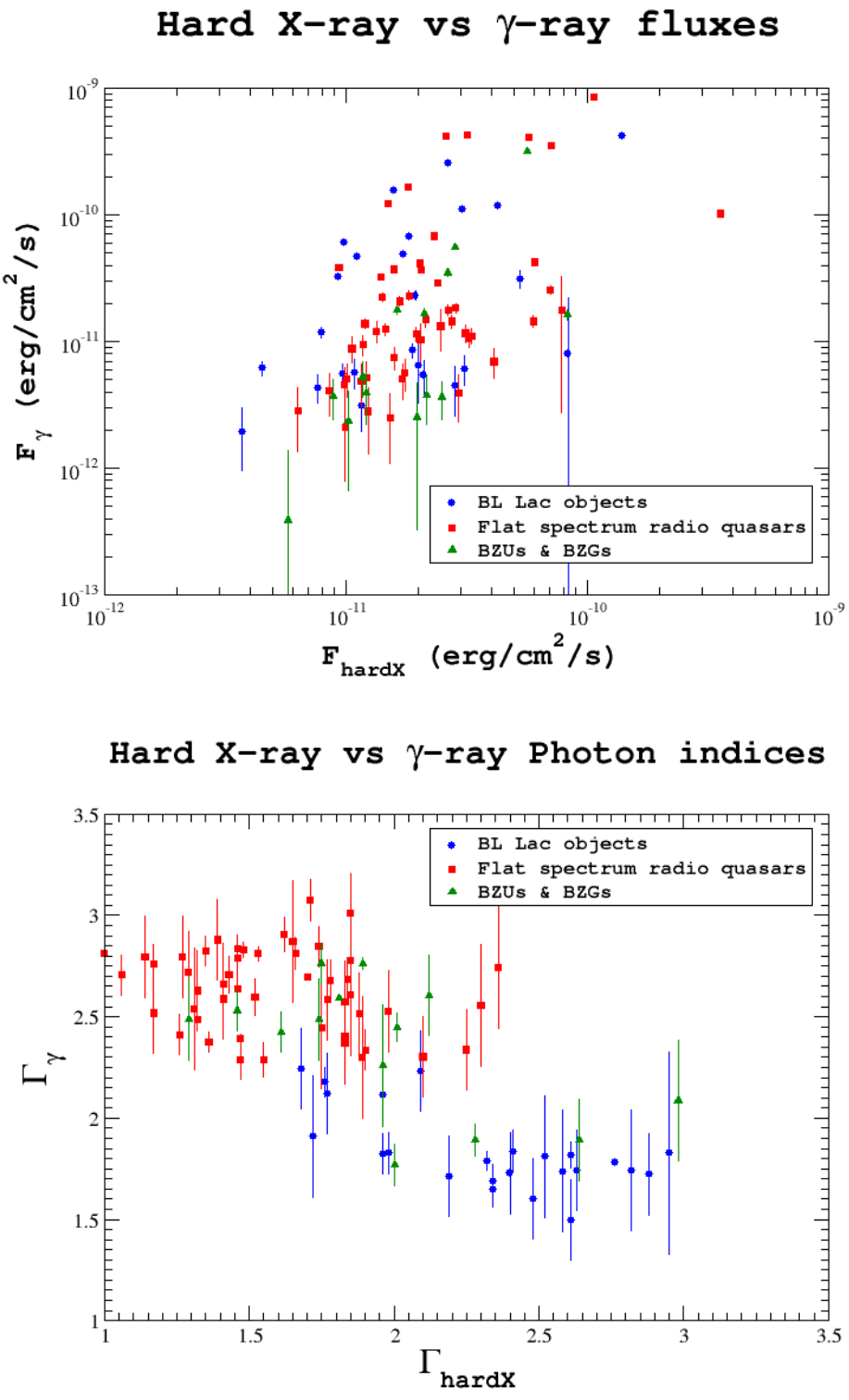

Figure 4.11: Top: Hard X-ray vs γ-ray and vs photon indices (bottom) correlation of all blazars in our sample. We can see, that BL Lacs are steeper in hard X-rays and flatter in γ-rays (bottom), however, their Hard X-ray vs γ-ray flux distribution (top) appears similar.

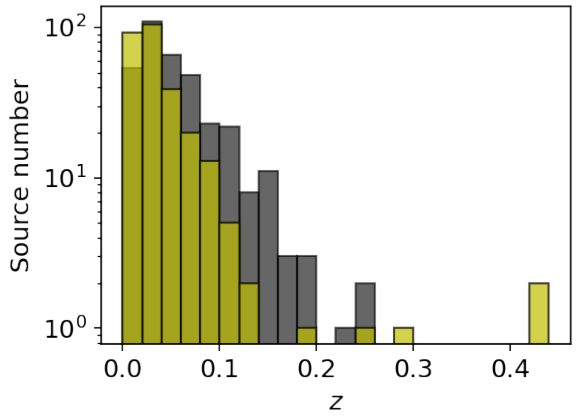

Figure 4.12: Redshift distribution of Turin-SyCAT Seyfert galaxies. Type 1 in black, type 2 in yellow.

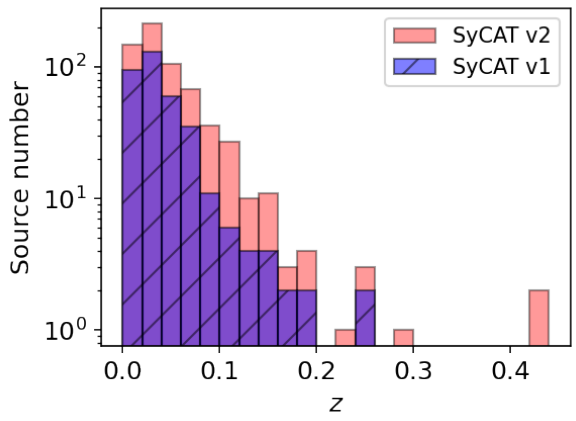

Figure 4.13: Redshift distribution of all Seyfert galaxies from 1^{st} release of the Turin-SyCAT compared to those listed in the 2^{nd} release.

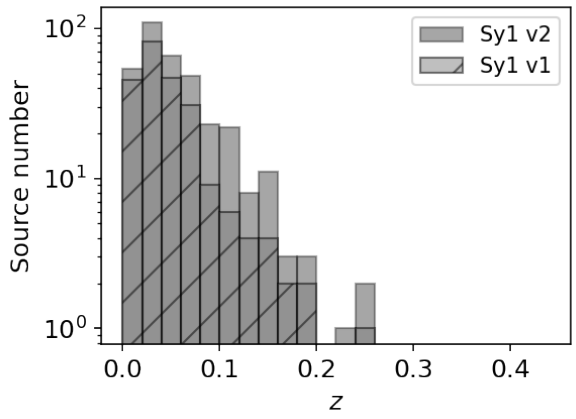

Figure 4.14: Redshift distribution of Type 1 Seyfert galaxies from 1^{st} release of the Turin-SyCAT compared to the presented 2^{nd} release.

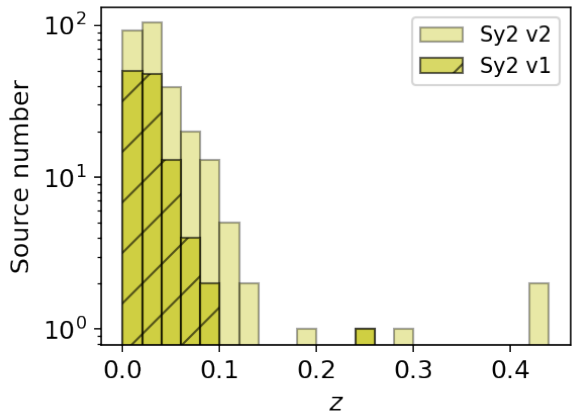

Figure 4.15: Redshift distribution of Type 2 Seyfert galaxies from 1^{st} release of the Turin-SyCAT compared to the presented 2^{nd} release.

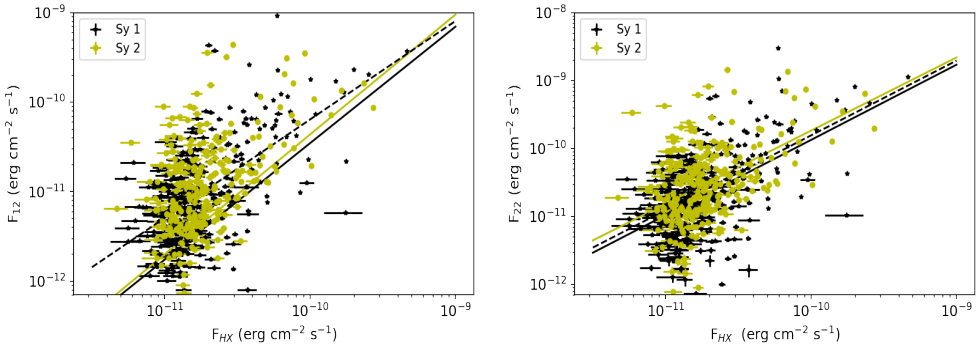

Figure 4.16: Mid-IR fluxes at 12 μm (Left panel) and 22 μm (Right panel) as a function of hard X-rays flux. The different lines are the correlations found between the W3 integrated flux and that in the hard X-ray band from the 3PBC, for both Seyfert 1 and 2 galaxies, marked in black and yellow, respectively. The dashed black line corresponds to the regression line computed for the whole sample while the straight black and yellow lines mark that for type 1 and type 2 Seyfert galaxies, respectively. The correlation coefficients are reported in Section 4.6.

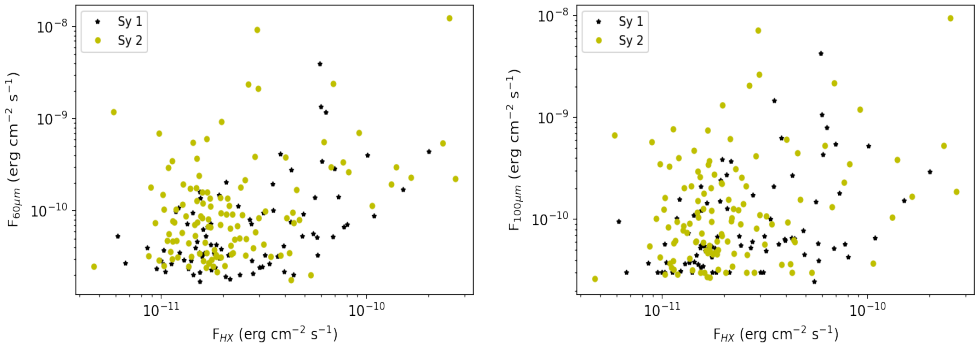

Figure 4.17: Fluxes at 60 μm (left panel) and 100 μm (right panel) as a function of hard X-rays flux. Seyfert 1 and 2 galaxies marked in black and yellow, respectively. No neat trend is evident between the two emissions.

5

Swift-XRT follow-up analysis of unidentified hard X-ray sources

M. Kosiba,[1,2] *F. Massaro,*[2,3,4] *A. Paggi,*[2,3,4] *H. A. Peña-Herazo,*[5] *N. Masetti,*[6,7] *V. Chavushyan,*[8] *E. Bottacini,*[9,10] *and N. Werner*[1]

[1]Department of Theoretical Physics and Astrophysics, Faculty of Science, Masaryk University, Kotlářská 2, Brno, 611 37, Czech Republic
[2]Dipartimento di Fisica, Università degli Studi di Torino, via Pietro Giuria 1, I-10125 Torino, Italy.
[3]Istituto Nazionale di Fisica Nucleare, Sezione di Torino, I-10125 Torino, Italy
[4]INAF–Osservatorio Astrofisico di Torino, via Osservatorio 20, I-10025 Pino Torinese, Italy
[5]East Asian Observatory, 660 North A'ohōkū Place, Hilo, Hawaii 96720, USA.
[6]INAF - Osservatorio di Astrofisica e Scienza dello Spazio, via Piero Gobetti 101, 40129 Bologna, Italy
[7]Departamento de Ciencias Físicas, Universidad Andrés Bello, Fernández Concha 700, Las Condes, Santiago, Chile
[8]Instituto Nacional de Astrofísica, Óptica y Electrónica, Luis Enrique Erro 1, Tonantzintla, Puebla 72840, México.
[9]Dipartimento di Fisica e Astronomia G. Galilei, Univeristà di Padova, Padova, Italy.
[10]Eureka Scientific, 2452 Delmer Street Suite 100, Oakland, CA 94602-3017, USA.

Submitted to Astronomy & Astrophysics, April 2024

5.1 ABSTRACT

It is currently established that the sources contributing to the cosmic X-ray background (CXB) emission are mainly nearby active galactic nuclei (AGN), in particular those that are obscured. Thus, it is important to fully identify the hard X-ray sky source population to accurately characterize the individual contribution of different AGNs to the overall CXB emission.

We present a follow-up analysis of all the 218 sources marked as unidentified in our previous revision of the third release of the Palermo *Swift*-BAT hard X-ray catalog (3PBC) based on our multifrequency classification scheme. These 218 sources were classified as unidentified in our previous analyses because they lack an assigned low-energy counterpart.

We searched for soft X-ray counterparts of these 218 3PBC sources in archival *Swift*-XRT observations obtained between 2005 January 1st and 2018 August 1st. In particular, we found 1213 archival *Swift*-XRT observations for 192 of the 218 unidentified sources.

We found 93 possible *Swift*-XRT counterparts inside of the *Swift*-BAT positional uncertainty regions. These correspond to 73 3PBC sources, where 60 have only a single *Swift*-XRT detection, and 13 sources have multiple detections. We present a catalog of all the detected possible counterparts of the yet unidentified hard X-ray sources to the community as a catalog for future spectroscopic follow-up targets, together with a short catalog of our classification of the ten sources for which we found available spectra.

5.2 INTRODUCTION

The hard X-ray sky, at energies greater than ∼10 keV, was systematically observed by several telescopes in the past decades. *Uhuru* was the first X-ray satellite (Giacconi et al., 1971), launched in 1970, which delivered an all-sky hard X-ray survey (Forman et al., 1978) in 2 – 20 keV band listing 339 sources. Subsequently, Levine et al. (1984), thanks to the X-ray and Gamma-ray detector HEAO-A4 on board the *HEAO 1* satellite (Rothschild et al., 1979) conducted an all-sky survey in 13–180 keV range providing 77 newly detected sources. The INTErnational Gamma-Ray Astrophysics Laboratory *INTEGRAL* (Winkler et al., 2003) with its Imager on Board the *INTEGRAL* Satellite *IBIS* (Ubertini et al., 2003) was launched in 2002, observing in energy range from 15 keV up to 10 MeV. Finally, the Neil Gehrels *Swift* Observatory (Gehrels et al., 2004), launched in 2004, carried an all-sky hard X-ray survey at 14 – 195 keV using the Burst Alert Telescope (BAT) (Barthelmy, 2004). Significant improvements in the soft X-ray background were possible mainly thanks to NASA's HEAO-2 *Einstein* observatory (Giacconi et al., 1979b), the German-US-UK X-ray observatory *ROSAT* (Röntgen Satellite) (Hasinger et al., 1999), the *Chandra* X-ray observatory (Weisskopf et al., 2000) and the *XMM-Newton* observatory (Jansen et al., 2001).

The *INTEGRAL* and *Swift* have allowed for the creation of many catalogs focusing on the hard X-ray sky (see, e.g., Beckmann et al., 2006; Bird et al., 2016; Bottacini et al., 2012; Churazov et al., 2007; Cusumano et al., 2010; Krivonos et al., 2007b; Krivonos et al., 2021, 2022, 2017; Markwardt et al., 2005; Mereminskiy et al., 2016; Oh et al., 2018; Sazonov et al., 2007; Tueller et al., 2008) and catalogs providing the association of hard X-ray sources with their low-energy counterparts (e.g., Bär et al., 2019; Koss et al., 2019; Malizia et al., 2010; Smith et al., 2020). They were also necessary for the optical spectroscopy follow-up observations (e.g., Marchesini et al., 2019a; Masetti et al., 2006a, 2008, 2006c, 2012, 2013; Parisi et al., 2014; Rojas et al., 2017). These missions are still operational nowadays and deliver new scientific results, e.g., the *INTEGRAL*-IBIS 17-yr hard X-ray all-sky survey (Krivonos et al., 2022), the AGN catalog and optical spectroscopy for the second data release of the *Swift*-BAT AGN Spectroscopic Survey (BASS DR2) (Koss et al., 2022) and the upcoming catalog based on the *Swift*-BAT 157-month survey (Lien et al., 2023).

Multiple catalogs based on the Burst Alert Telescope (BAT) on board the *Swift* observatory exist. Koss et al. (2022) constructed the AGN catalog and optical spectroscopy for the second data release of the *Swift*-BAT AGN Spectroscopic Survey (BASS DR2). They provide 1449 optical spectra corresponding to 858 hard-X-ray-selected AGNs in the *Swift*-BAT 70-month observations. Their AGN sample is spectroscopically complete, with 857/858 AGNs reported with

redshifts. Oh et al. (2018) created the 105-month *Swift*-BAT catalog of hard X-ray sources. This catalog covers over 90 % of the sky with a sensitivity of 8.40×10^{-12} erg s^{-1} cm^{-2} and 7.24×10^{-12} erg s^{-1} cm^{-2} over 50 % of the sky in the 14 – 195 keV band providing 1632 hard X-ray detections above 4.8 σ significance threshold. Cusumano et al. (2010) created the Palermo *Swift*-BAT Hard X-ray catalog. This work links to the 2PBC a catalog release after 54 months of sky survey. The 3^{rd} release of the Palermo *Swift*-BAT hard X-ray catalog (3PBC) is currently ongoing and is available only online[1]. This is the catalog version we analyze in this study. The 3PBC lists 1256 sources detected above 4.8 σ level of significance in the 15 – 150 keV energy range. Their number increases to 1593 total sources when considering a threshold on the signal-to-noise ratio above 3.8. The catalog covers nearly 90 % of the sky to a flux limit of 1.1×10^{-11} erg cm^{-2} s^{-1}, decreasing to ∼50 % when decreasing this flux threshold to 0.9×10^{-11} erg cm^{-2} s^{-1}.

We recently conducted a refined analysis (Kosiba et al., 2023, hereinafter paper I) of all sources listed in the 3PBC catalog. Our refined analysis is based on the multifrequency classification scheme we developed to analyze hard X-ray sources, mainly focusing on extragalactic source populations (Peña-Herazo et al., 2022). Findings of our refined analysis of the 3PBC were also based on results reported in the 105-month *Swift*-BAT catalog (Oh et al., 2018) and the INTEGRAL hard X-ray survey above 100 keV with its 11-year release (Krivonos et al., 2015), which we used to cross-match the 3PBC sources for counterparts to obtain luminosities and spectra if available.

We found that approximately 57 % of the sources listed in the 3PBC have an extragalactic origin, while 19 % belong to our Milky Way. The remaining 24 % are yet unknown. In particular, our final revised version of the 3PBC catalog lists 1176 classified sources, 218 unidentified, and 199 unclassified. Of the 1176 classified sources, 820 have an extragalactic origin, and 356 have a Galactic origin. Compared to the original 3PBC, which has 233 unidentified and 300 unclassified sources, we decreased these fractions from ∼33% (533 sources) to ∼26% (417 sources).

Our study also allowed us to discover new Seyfert galaxies included in the Turin-SyCAT 2^{nd} release (paper I). In the 2^{nd} release of the Turin-SyCAT, there are 633 Seyfert galaxies, 282 new ones added thanks to our refined analysis and corresponding to an increase of ∼80% with respect to the Turin-SyCAT 1^{st} release (Peña-Herazo et al., 2022).

Moreover, trends between the hard X-ray and the gamma-ray emissions of those blazars listed in the 3PBC with a counterpart in the second release of the fourth *Fermi*-LAT catalog (4FGL-DR2) were also found (paper I), as expected from emission models widely adopted to explain their broadband SED (e.g., Marscher and Gear, 1985; Marscher and Travis, 1996; Massaro et al., 2006).

In this work, we examine the population of the 218 unidentified hard X-ray sources listed in our revised version of the 3PBC, i.e., those lacking an assigned counterpart at lower energies than the BAT energy range. We analyzed all soft X-ray observations (between 0.5 and 10 keV) available in the archive of the X-ray telescope (XRT) (Burrows et al., 2005) on board the *Swift* Observatory and found available data for 192 of the 218 3PBC sources, which is the sample we further

1 http://bat.ifc.inaf.it/bat_catalog_web/66m_bat_catalog.html

analyze in this study. This analysis aims to search for potential counterparts in the soft X-ray data of the Swift-XRT for the 192 yet unidentified 3PBC sources. The final goal of the present analysis is to provide a catalog of all unidentified hard X-ray sources having at least one candidate counterpart if detected in the soft X-ray band, that could be targeted with follow-up spectroscopic observations to obtain its final classification, as successfully carried out in the last decades (e.g., Koss et al., 2017; Landi et al., 2017; Marchesini et al., 2019a; Masetti et al., 2006a, 2008, 2006c, 2012, 2013; Parisi et al., 2014; Rojas et al., 2017; Tomsick et al., 2020).

The paper is organized as follows. Section 5.3 describes the *Swift*-XRT data reduction and data analysis procedure. Then Section 5.4 is devoted to our results while details on the multifrequency comparison are illustrated in Section 5.5. Finally, Section 5.6 is dedicated to our summary, conclusions, and future perspectives. X-ray images for all analyzed BAT sources are reported in the Appendix.

As previously adopted in the paper I, we used cgs units unless stated otherwise. We also adopted ΛCDM cosmology with $\Omega_M = 0.286$, and Hubble constant $H_0 = 69.6\,\mathrm{km}\,s^{-1}\,Mpc^{-1}$ (Bennett et al., 2014) to compute cosmological corrections through the whole manuscript.

5.3 *SWIFT*-XRT OBSERVATIONS

5.3.1 *Sample selection*

The 3PBC catalog lists 1 593 hard X-ray sources, all detected with a signal-to-noise ratio (S/N) above 3.8 in the 15–150 keV energy range. Our analysis, presented in paper I, identified 218 3PBC sources lacking an assigned counterpart at lower energies.

In this work, we searched the *Swift*-XRT archive, and we found that 192 out of these 218 hard X-ray sources have at least one X-ray observation with exposure time larger than 250 sec in the 0.5–10 keV energy range. We found a total of 1213 such observations that have been reduced and analyzed here according to the standard procedures described below. In Fig. 5.1, we report the distribution of the exposure time for all selected observations. The 1213 *Swift*-XRT observations have a mean of 1462 s and variance $\sim 6 \times 10^6$ s with a total exposure time of 1.77×10^6 s.

All observations we reduced and analyzed in this study were performed between April 2005 and December 2022.

5.3.2 *Data Reduction*

Data reduction procedures applied here for all *Swift*-XRT observations are the same previously adopted for similar analyses (see e.g. Marchesini et al., 2020, 2019b; Massaro et al., 2008a,b, 2023a; Paggi et al., 2013) and procedures of the *Swift*-XRT X-Ray point source catalogs (D'Elia et al., 2013; Evans et al., 2014b, 2020). Thus, we only describe the basic information and refer to the above references for a more detailed description.

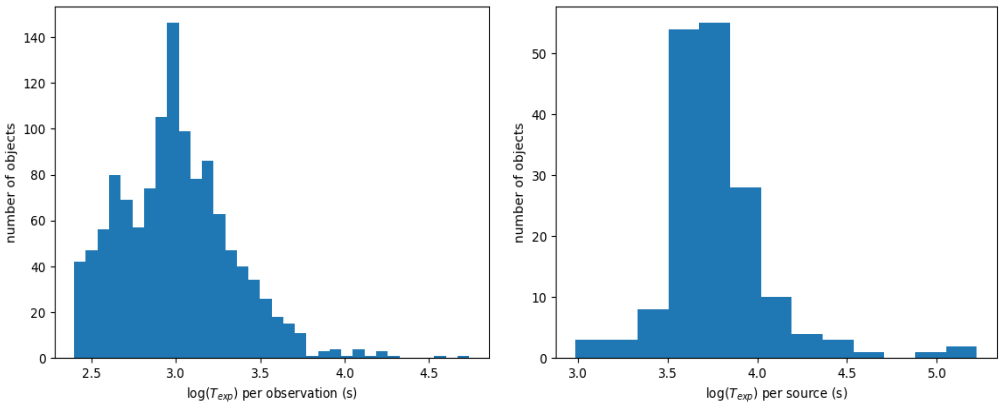

Figure 5.1: Distribution of T_{exp} of the 1213 *Swift*-XRT observations analyzed in this work, with mean 1462 s (left panel) and the total T_{exp} per source of the 192 sources for which we found the *Swift*-XRT observations with mean 8864 s (right panel).

We downloaded raw *Swift*-XRT data from the archive[2]. Then we run the XRTPIPELINE task, developed as part of the *Swift* X-Ray Telescope Data Analysis Software (XRTDAS) and distributed within the HEAsoft package (version 6.30.1) (Nasa High Energy Astrophysics Science Archive Research Center (Heasarc), 2014). This allowed us to obtain clean event files for all *Swift*-XRT observations. The entire analysis and all X-ray images shown in the present manuscript are restricted to the 0.5-10 keV energy range unless stated otherwise.

We subsequently calibrated these cleaned event files with the usual filtering criteria and using calibration files provided in the High Energy Astrophysics Science Archive Research Center (HEASARC) calibration database (CALDB) version (v.20220907)[3]. Using the XSELECT task we excluded all time intervals with count rates higher than 40 photons/sec as well as those with CCD temperature exceeding -50°C, in regions located at the edges of the XRT detector (D'Elia et al., 2013). Then, the XSELECT task was also used to merge all cleaned and filtered event files for those sources with multiple observations. Finally, it is worth mentioning that the entire analysis was carried out using the XIMAGE software[4] to merge the corresponding exposure maps (Giommi et al., 1992a).

5.3.3 *Data Analysis*

To detect sources, we used the sliding cell DETECT (DET) algorithm available in the XIMAGE software package (Giommi et al., 1992b) on all merged event files as well as on the single event files for sources with only one observation. We set a threshold on the S/N equal to 3 for claiming detection of an X-ray source in the 0.5-10 keV energy range, as recently performed in the analysis of Massaro et al., 2023b.

Then, to identify and characterize the 3PBC sources, we labeled them using three different X-ray detection flags (XDF) on the basis of the following criteria.

2 https://heasarc.gsfc.nasa.gov/docs/archive.html
3 https://heasarc.gsfc.nasa.gov/docs/heasarc/caldb/caldb_supported_missions.html
4 https://heasarc.gsfc.nasa.gov/xanadu/ximage/ximage.html

- *x* flag: is used for 3PBC sources that have a single soft X-ray source within their BAT positional uncertainty region (see, e.g., 3PBC J1039.4-4903 shown in the left panel of Figure 5.2)
- *m* flag: indicates 3PBC sources with more than one soft X-ray source (multiple detections) within their BAT positional uncertainty region (see, e.g., 3PBC J0819.2-2509 shown in the central panel of Figure 5.2)
- *u* flag: is adopted for 3PBC sources with no soft X-ray counterparts detected in their merged event files within their BAT positional uncertainty region (see, e.g., 3PBC J1834.7-0345 shown in the right panel of Figure 5.2).

We measured several parameters for all detected possible X-ray counterparts of the 3PBC hard X-ray sources in merged event files. In particular, we obtained coordinates of the distributions of X-ray photons from each source using the XRTCENTROID task. We also measured n_{90}, the number of photons within a circular region centered on the X-ray coordinates with radius 120,207 arcsec (51 pixels), which is enclosing 90% of the *Swift*-XRT point spread function (PSF).

5.4 A SOFT X-RAY PERSPECTIVE OF THE HARD X-RAY SKY

5.4.1 *Outline of the main goal*

The main goal of this analysis is to search for possible counterparts to the sample of yet unidentified hard X-ray sources identified in our previous work (Kosiba et al., 2023). We search for the counterparts in soft X-ray wavelengths of the *Swift*-XRT data. The final product of this analysis is to create and provide a catalog of soft X-ray counterparts we found in the *Swift*-XRT data for the yet unidentified hard X-ray sources. We release this catalog along with this publication. In this paper, we also describe the sources we analyze in this study.

5.4.2 *Overview of results*

We analyzed all available *Swift*-XRT data, selected according to the criteria previously described, for a total of 1213 observations corresponding to 192 3PBC sources with a total exposure time 1.77×10^6 s. Considering only the *Swift*-XRT detections inside the BAT positional uncertainty region above the S/N = 3 we adopted, we found 93 soft X-ray sources. These 93 soft X-ray sources correspond to 73 unique 3PBC sources. From those, 13 3PBC sources are associated with multiple soft X-ray detected sources (*m* flag), and the remaining 60 3PBC sources are associated with a single soft X-ray detected source (*x* flag). Those are the final results presented in this analysis.

We note that all 3PBC sources with at least one soft X-ray counterpart detected within their positional uncertainty region have an integrated exposure time above 975 seconds, about four times longer than the minimum selected value. The distribution of X-ray count rates and that of their positional uncertainty in the 0.5–10 keV energy range computed using the XRTCENTROID task for the 93 *Swift*-XRT detected sources are shown in Fig. 5.3 (left) and Fig. 5.3 (right), respectively.

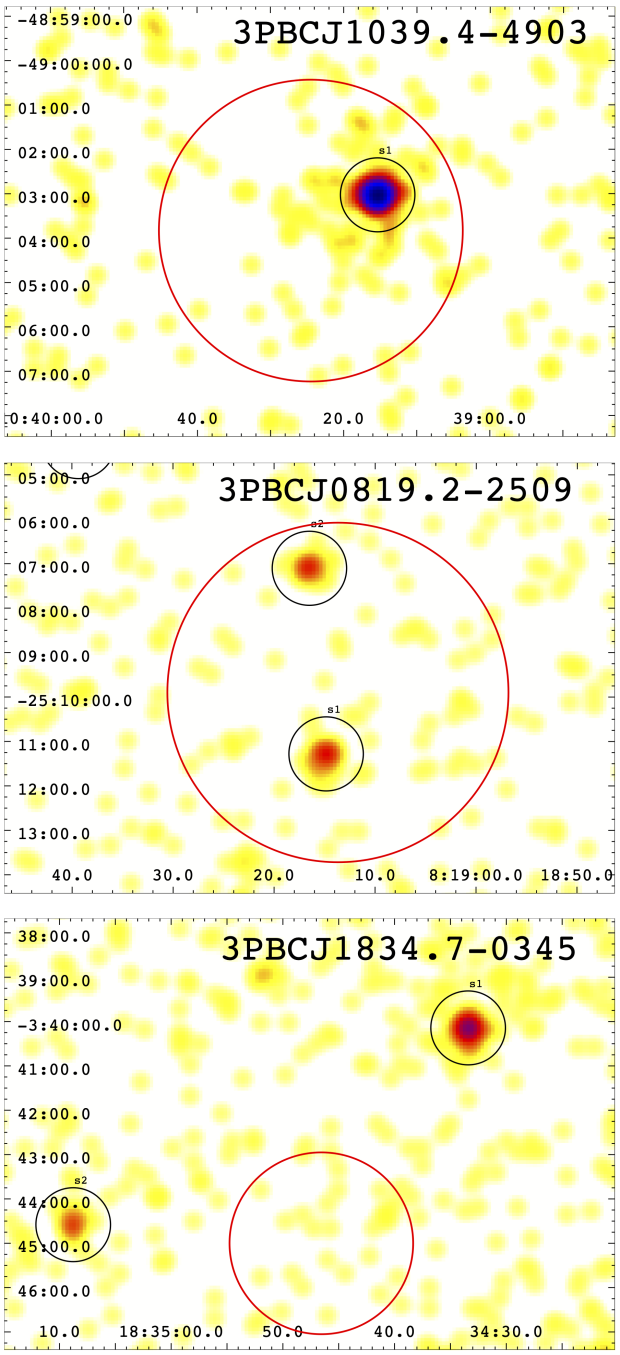

Figure 5.2: Each figure in this image depicts an example of our XDF flags that label the 3PBC sources. These figures are the XRT merged event files with the red circle indicating BAT positional uncertainty region and the black circles highlighting the position of a soft X-ray source detected in the *Swift*-XRT archive. The top panel is an example of a 3PBC source with a single *Swift*-XRT source found within the BAT positional uncertainty region (red circle), flag *x*. The center panel is the case of multiple detected *Swift*-XRT counterparts inside of the BAT positional uncertainty, flag *m*. The bottom panel is an example of no *Swift*-XRT counterparts detected within the BAT positional uncertainty region, flag *u*.

For calculating the BAT positional uncertainty region (red circles in Fig. 5.2), we took the values reported in the 3PBC catalog (Cusumano et al., 2010).

In this section, we are focusing on the m XDF flagged sources (Fig. 5.4 and Fig. 5.5). We discuss these 13 3PBC sources separately to detail their potential soft X-ray counterparts.

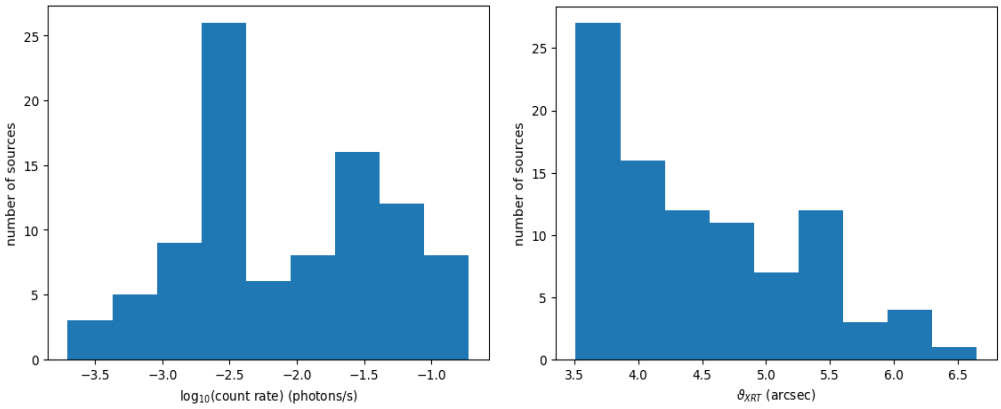

Figure 5.3: Distribution of the count rate in logarithmic scale, with mean 0.028 photon-s/s (left panel) and the XRT positional uncertainty (ϑ_{XRT}) with mean value 4.5 arcsec (right panel). Both are in the 0.5-10 keV energy range for all 93 *Swift*-XRT detected counterparts.

In Appendix 5.7.1, we report the first five columns of our final table for the first ten sources. Those are the 3PBC name, the XRT counterpart name, the XDF flag, the right ascension and declination of the centroid, and the uncertainty on the centroid's position. The remaining seven columns of our final table are also reported in the Appendix 5.7.1, listing the angular separation between XRT and BAT, count rate, uncertainty on the count rate, S/N, number of observations, integrated exposure time (T_{exp}) and counterpart's name in WISE.

Finally, Appendix 5.7.2 includes all X-ray images obtained from cleaned and merged event files with a red dashed circle showing the BAT positional uncertainty and, when present, black circles indicating all soft X-ray counterparts detected in our *Swift*-XRT analysis. The green cross marks soft X-ray detections with a WISE counterpart. The WISE counterparts have been searched within a circular region corresponding to the positional uncertainty of the *Swift*-XRT.

5.4.3 *3PBC sources with multiple candidate counterparts*

This section shows images of 3PBC sources with a brief discussion for which we find multiple XRT PC counterparts consistent within the positional uncertainties (m flag).

Three more sources deserve a more detailed description: 3PBC J1430.3+2303 (m flag), 3PBC J1620.1-5001 (x flag), and 3PBC J1730.0-3436 (x flag). All these sources appear to have extended X-ray sources close to the BAT positional uncertainty region. Thus, we adopted the following approach to consider potential X-ray counterparts and avoid spurious detected objects.

For 3PBC J1620.1-5001 and 3PBC J1730.0-3436 (Appendix 5.7.2), we only indicate in the main table the X-ray detected source having the highest signal-

to-noise ratio among those automatically detected. In particular, both sources lie close to the Galactic plane and close to star-forming regions and could be unknown supernova remnants deserving further investigation. However, in the case of 3PBC J1730.0-3436, artifacts in the merged event file prevented us from conducting a detailed analysis. On the other hand, we noticed that for 3PBC J1620.1-5001, the BAT positional uncertainty region is not centered on the extended X-ray source, thus suggesting that it is not the soft X-ray counterpart of the hard X-ray object. A forthcoming paper will present a multifrequency analysis of 3PBC J1620.1-5001 (Kosiba et al., 2023).

Finally, we considered the case of 3PBC1430.3+2303, for which, given the diffuse X-ray emission clearly detected, we only selected as potential counterparts those targets detected as described in the previous section but also having a mid-infrared counterpart, marked with the green cross in Fig. 5.4 and Fig. 5.5. It is worth noting that among them, SWXRTJ143016.094+230343.862 seems to be associated with the galaxy cluster MSPM 05080, thus indicating that the possible origin of this extended X-ray emission is that arising from its intracluster medium.

5.4.3.1 *3PBC J0022.2+2539*

This source has three XRT PC counterparts Fig. 5.4. While sources s1 and s3 are faint with S/Ns of 4.7 and 3.1, respectively, source s2 is much brighter with a S/N of 37 and a count rate of 0.158 ± 0.004 s^{-1}. In addition, source s2 is also detected by WISE (J002203.09+254003.2) and in SDSS (J002203.09+254003.1) with magnitude $r = 17.0$.

5.4.3.2 *3PBC J0218.5-5005*

This source has three XRT PC counterparts Fig. 5.4. The brightest of these is s1 with an S/N of 5.9, while the faintest is s2 with an S/N of 3.8, also detected by WISE (J021822.70-500557.5).

5.4.3.3 *3PBC J0536.1-3205*

This source has two XRT PC counterparts Fig. 5.4. The brightest being s1 with an S/N of 4.1, also detected by WISE (J053618.88-320533.0).

5.4.3.4 *3PBC J0709.5-3538*

This source has two XRT PC counterparts Fig. 5.4, with s1 being by far the brightest, with a S/N of 28.6 and a countrate of 0.176 ± 0.006 s^{-1}, also detected by WISE (J070932.05-353746.5) with a spectrum reported in Rojas et al., 2017.

5.4.3.5 *3PBC J0800.7-4308*

This source also has two XRT PC counterparts Fig. 5.4. The source s1 is the faintest, with a S/N of 5.8 and a countrate 0f 0.0036 ± 0.0006 s^{-1}, while s2 has a S/N of 27.6 and a countrate 0f 0.065 ± 0.002 s^{-1}. Both sources have a WISE counterpart, J080045.83-430939.3, and J080039.96-431107.2, respectively.

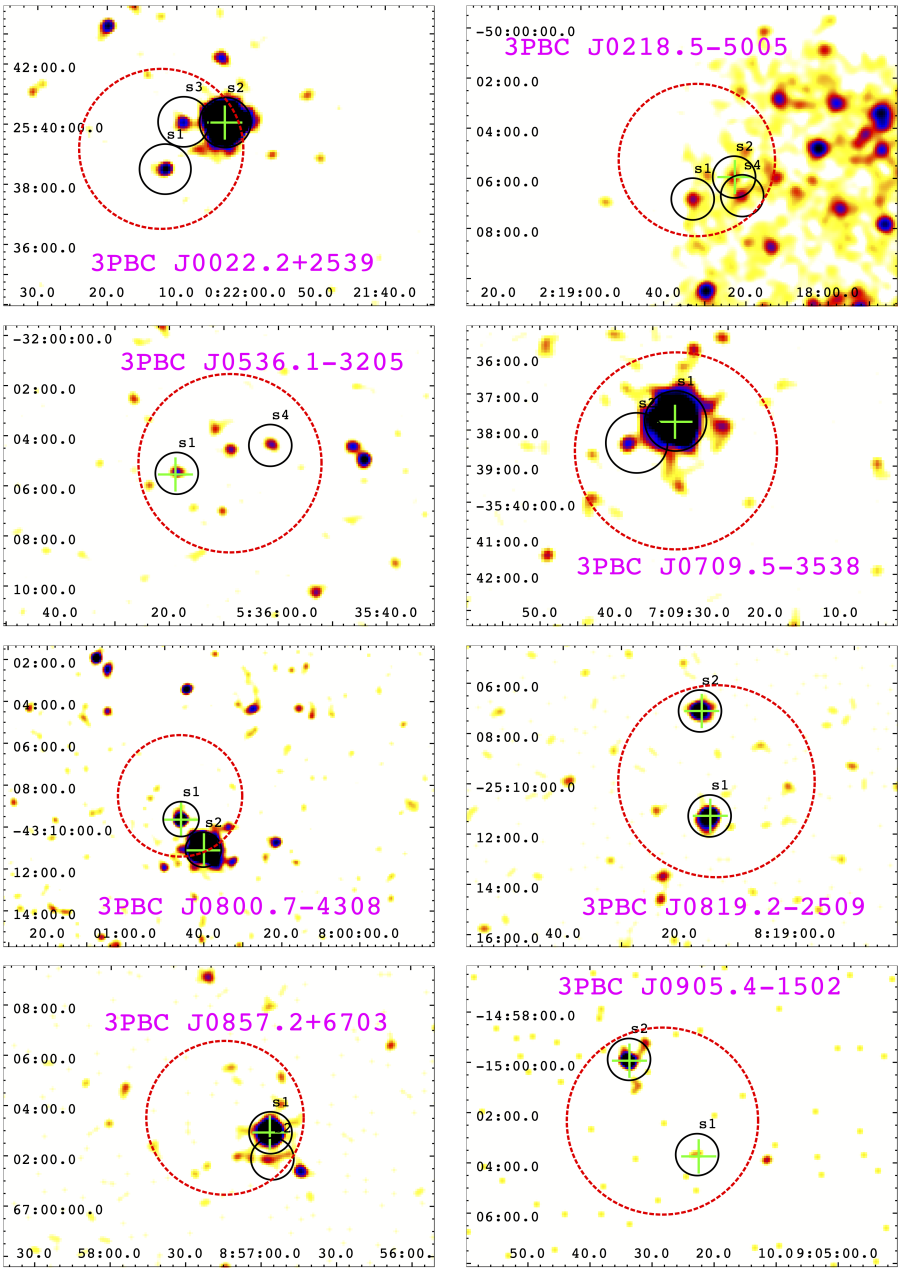

Figure 5.4: Images of the first 8 of the 13 3PBC sources with more than one soft *Swift*-XRT source (XDF flag *m*) detected inside of the BAT positional uncertainty region (red dashed circle). The soft X-ray detections are indicated with a black circle. The black circle indicates the position of the soft X-ray source, not its positional uncertainty. If the soft X-ray detection is also marked with a green cross, it indicates that it has a WISE counterpart.

5.4.3.6 *3PBC J0819.2-2509*

This source has two XRT PC counterparts Fig. 5.4, with the brightest being s1 with an S/N of 14.0 and the faintest being s2 with an S/N of 7.4. Both sources have a WISE counterpart, J081914.73-251116.6, and J081916.20-250706.4, respectively, with s1 having a redshift of 0.00557 with a spectrum reported in (Strauss et al., 1992).

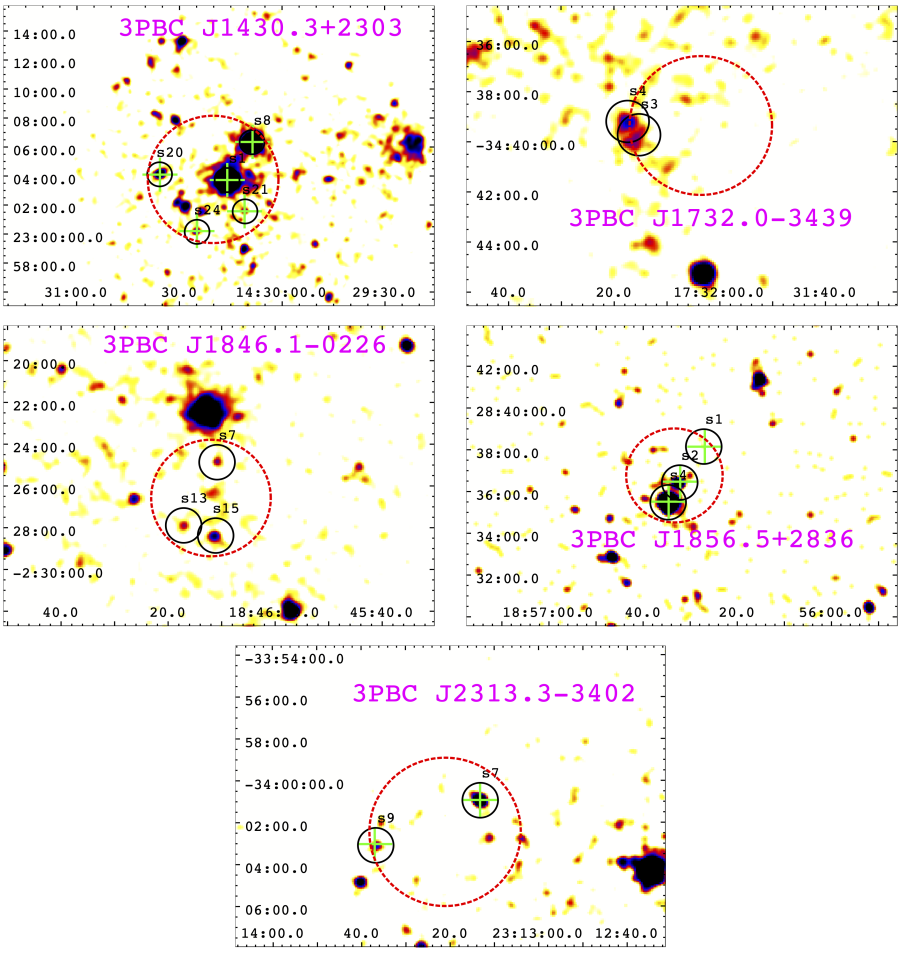

Figure 5.5: Images of the last 5 of the 13 3PBC sources with more than one soft *Swift*-XRT source (XDF flag *m*) detected inside of the BAT positional uncertainty region (red dashed circle). The soft X-ray detections are indicated with a black circle. The black circle indicates the position of the soft X-ray source, not its positional uncertainty. If the soft X-ray detection is also marked with a green cross, it indicates that it has a WISE counterpart.

5.4.3.7 *3PBC J0857.2+6703*

For this source Fig. 5.4, we find two XRT PC counterparts: the brightest being s1 with an S/N of 17.6 and the faintest being s2 with an S/N of 3.1. The source s1 has a WISE counterpart, namely J085656.49+670257.3.

5.4.3.8 *3PBC J0905.4-1502*

This source Fig. 5.4 has two XRT PC counterparts s1 and s2, with similar S/Ns, 3.8 and 3.9, respectively. Both are detected by WISE (J090522.48-150344.6 and J090533.64-145956.2, respectively), with s2 reported in NED with a redshift of 0.088.

5.4.3.9 *3PBC J1430.3+2303*

For the 3PBCJ1430.3+2303 (Fig. 5.5), given the presence of diffuse X-ray emission, the automatic algorithm described in the previous section detected many spurious sources. We thus selected as potential soft X-ray counterparts only the sources with a mid-infrared counterpart, marked with the green cross. It is worth noting that among them, SWXRTJ143016.094+230343.862 seems to be associated with the galaxy cluster MSPM 05080, thus indicating that the possible origin of this extended X-ray emission is that arising from its intracluster medium.

5.4.3.10 *3PBCJ1732.0-3439*

This source Fig. 5.5 has two XRT PC counterparts, s3 and s4, with S/Ns of 3.3 and 4.7, respectively, and count rates of 0.0044 ± 0.0013 and 0.0087 ± 0.0018 $\mathrm{ct\,s^{-1}}$, respectively.

5.4.3.11 *3PBCJ1846.1-0226*

This source Fig. 5.5 has three XRT PC counterparts, s7, s13 and s15, with similar S/Ns of 4.2, 4.5 and 4.9, respectively. The source s13 is also detected in SDSS (J184617.13-022753.4) with a magnitude $r = 17.6$.

5.4.3.12 *3PBCJ1856.5+2836*

This source Fig. 5.5 has three XRT PC counterparts, s1, s2 and s4, with S/Ns of 3.6, 3.5 and 9.9, respectively. The sources are all detected by WISE (J185626.89+283809.3, J185632.13+283628.8, and J185634.58+283531.3), although the faintest s2 only has an upper limit of 8.3 mag in W_4 band.

5.4.3.13 *3PBCJ2313.3-3402*

This source Fig. 5.5 has two XRT PC counterparts, s7, and s9, with the brightest being the former, with an S/N of 8.4. Both sources are detected by WISE (J231313.21-340056.2 and J231337.01-340302.1), although the fainter s9 only has an upper limit in W_3 and W_4 bands.

5.4.4 *Comparison with 2SXPS*

Finally, we compared our results with those that can be obtained by simply crossmatching all 218 hard X-ray sources listed in the 3PBC with the latest release of the Second *Swift*-XRT Point Source Catalog[5] (2SXPS Evans et al., 2020). The 2SXPS catalog has a sky coverage of $3,790$ $\mathrm{deg^2}$, listing 206,335 point sources detected by XRT in the $0.3-10$ keV energy range. Here, we briefly summarize the procedure used to build the 2SXPS.

The 2SXPS was built based on all XRT observations taken between 2005, January 1st, and 2018, August 1st, with an exposure of at least 100 s in PC (photon-counting) mode. Source detection was performed with the sliding-cell technique with a S/N threshold set to 1.5, in comparison to our choice of $\mathrm{S/N} = 3$,

5 https://heasarc.gsfc.nasa.gov/W3Browse/swift/swift2sxps.html

yielding the final catalog of 206,335 XRT PC sources listed in the 2SXPS. The catalog contains a “clean" subsample, listing 146,768 sources without analysis flags (see (Evans et al., 2020) for more details). In the following, we will consider this “clean" subsample.

Due to the different dataset and analysis procedures used in the present work and in Evans et al. (2020), we expect differences in the XRT PC source detections. In fact, cross-matching the 2SXPS “clean" sample with the 3PBC sources considered here (see Sect. 5.3.1), taking into account both BAT and 2SXPS positional uncertainties, we find 126 2SXPS counterparts to 90 3PBC sources, 68 3PBC sources with a single 2SXPS match and 22 3PBC sources with multiple 2SXPS matches. In the left panel of Fig. 5.6, we compare the exposures of the XRT PC observations used in the present analysis (blue distribution) and in the 2SXPS observations for which we find the 126 2SXPS counterparts to the 3PBC sources considered in this work. We see that 2SXPS datasets and the dataset used in our analysis span a similar exposure range.

In addition, we find XRT PC counterparts to 7 3PBC sources without 2SXPS counterparts, while in the 2SXPS catalog, there are counterparts to 22 3PBC sources for which we did not find XRT PC counterparts. However, we note that the S/Ns for these 22 sources are less than 2.5, below the S/N threshold of 3 that we adopted for the present analysis. In addition, we find 78 sources positionally compatible between our sample of 93 XRT PC sources and the 126 2SXPS counterparts to the 3PBC sources.

In the right panel of Fig. 5.6, we compare the count rate evaluated in the present analysis with the count rate reported in the 2SXPS catalog for these 78 common sources. We stress that to account for the different energy bands adopted in the present analysis ($0.5-10$ keV) and in the 2SXPS catalog ($0.3-10$ keV), we rescaled the count rate of the 2SXPS catalog by a factor 0.86 evaluated via PIMMS[6] tool assuming a power-law spectrum with a 1.8 slope. We see that the two estimates are in good agreement at low count rate values ($< 10^{-3}$ cps). In contrast, above 10^{-2} cps, the 2SXPS count rates appear systematically larger than those evaluated in the present analysis. The two-count rate estimates, however, are compatible at $2\,\sigma$ level.

5.5 MULTIFREQUENCY COMPARISON

To search for additional information regarding all detected soft X-ray candidate counterparts we crossmatched their position, derived with the XRTCENTROID task - taking into account their positional uncertainties - with three main catalogs/surveys: (i) the NASA Extragalactic Database (NED)[7]; (ii) the SIMBAD Astronomical Database[8] as well as (iii) the ALLWISE catalog (Cutri et al., 2021b) based on the all-sky survey performed with WISE telescope (Wright et al., 2010) and (iv) the spectroscopic catalog of the Sloan Digital Sky Survey (SDSS) data release 16 (DR16) (Blanton et al., 2017). We used the positional uncertainty region of the *Swift*-XRT for all soft X-ray detected sources to search and claim an association with its wise counterpart.

6 https://cxc.harvard.edu/toolkit/pimms.jsp
7 http://ned.ipac.caltech.edu
8 http://simbad.u-strasbg.fr/simbad/

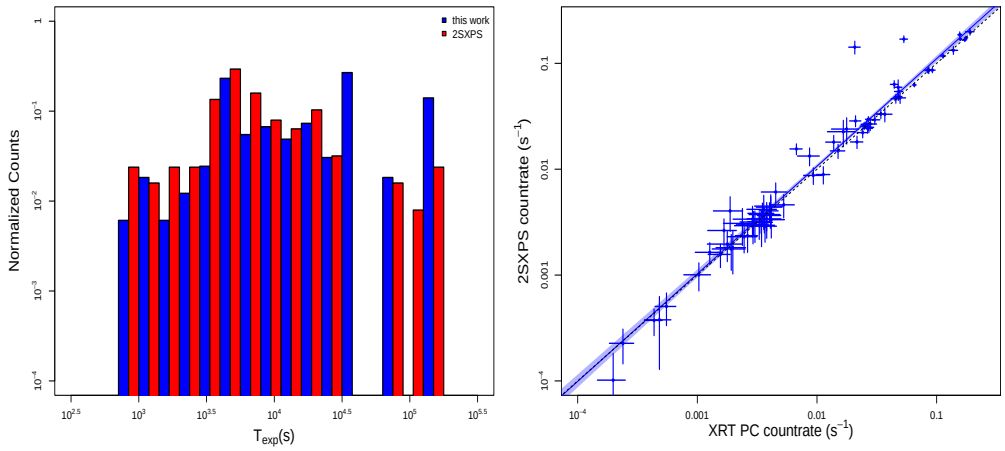

Figure 5.6: Left panel: the comparison between the exposure of the dataset used in the present work (blue distribution) and in the 2SXPS catalog (red distribution). Right panel: XRT PC count rates as reported in the 2SXPS catalog versus those evaluated in the present work for the 78 sources in common between the two analyses (see Sect. 5.4.4). The 2SXPS count rate has been rescaled to match the energy band used in the present analysis. The blue line indicates the linear regression to the logarithmic data, while the light blue shaded area represents the $1-\sigma$ uncertainty around the best-fit relation. The black dashed line indicates the $y = x$ relation. See Sect.5.4.4 for more details.

Our crossmatching analysis revealed that 84 of 93 soft X-ray potential counterparts have an identification reported in NED. In addition, 74 of the 93 have photometry available in WISE. We found that 10 out of the 93 have spectra in the SDSS DR16. We classified them according to their spectral characteristics and we report the results in Tab. 24.

5.5.1 *A mid-infrared perspective*

From the 93 detected soft X-ray possible counterparts corresponding to the 73 unique 3PBC sources, 74 sources have been detected in at least one WISE band. There are 74 sources detected in both the W1 and W2 bands, 66 in W3, and 52 in W4. Fig. 5.7 shows the distribution of the angular separation between the *Swift*-XRT centroid and the WISE centroid of the 74 counterparts that have the WISE detection. We took all sources that are detected in the W1, W2, and W3 bands, so 66 sources, and plotted them on a color-color diagram (Fig. 5.8). The gray background sources in Fig. 5.8 are 3 000 WISE sources in the mid-IR sky selected randomly in a region of 0.5 deg radius around Galactic coordinates (50.411113,-45.668864) and (50.411113,45.668864). The black dots correspond to the 66 *Swift*-XRT detected sources with reported luminosities in the first three WISE bands. The mid-IR colors of this sample of 66 sources are not in good agreement with the mid-IR colors of stars but are more consistent with AGNs, mainly Seyfert galaxies and QSOs.

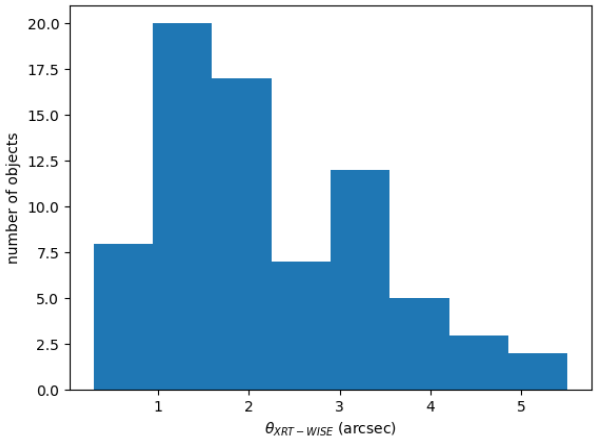

Figure 5.7: Distribution of the angular separation between the *Swift*-XRT centroid and the WISE centroid ($\theta_{XRT-WISE}$) of the 74 soft X-ray detections with the WISE counterpart.

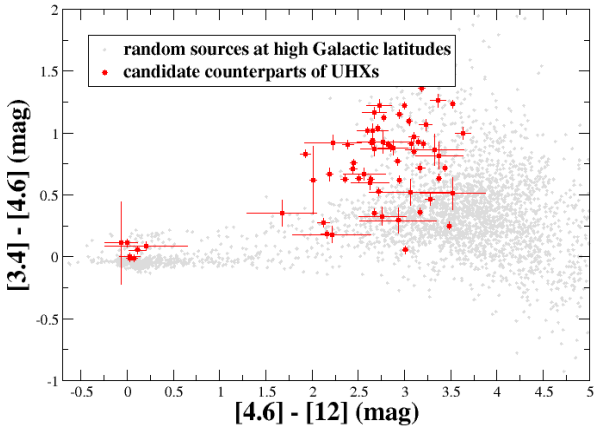

Figure 5.8: This figure shows the [3.4]-[4.6]-[12] µm color-color diagram of WISE thermal sources and blazars. The gray dots represent a sample of 3 000 randomly selected mid-IR sources in a region of 0.5 deg radius around Galactic coordinates (50.411113,-45.668864) and (50.411113,45.668864). The 66 *Swift*-XRT detected sources with available luminosities in the first three WISE bands are marked as black dots. The mid-IR colors of this sample of 66 sources do not agree with the mid-IR colors of stars. Instead, they are more consistent with AGNs, mainly Seyfert galaxies and QSOs. The sources with [4.6]-[12] mag > 2 are AGNs and QSOs, while the concentration of sources around 0 [4.6]-[12] mag are mostly normal elliptical galaxies and stars.

5.5.2 *Archival optical spectra*

According to previous analyses carried out during past follow-up spectroscopic campaigns (see, e.g., Kosiba et al., 2023; Massaro and D'Abrusco, 2016; Peña-Herazo et al., 2020, 2022), we adopted a conservative criterion to provide spectroscopic identification of selected X-ray counterparts. We adopted the same

classification scheme described in paper I, and we only considered reliable redshift measurements, those for which we could verify the presence of a published image of the optical spectrum or a description of the published spectrum with emission and/or absorption lines clearly reported in a table format or the publication manuscript. We found spectra in the SDSS DR16 archive for 10 soft X-ray sources. We show the spectra in Fig. 5.9 and the classification of those sources in Appendix 5.7.1. The most prominent spectral lines among their spectra were the Hα + [N II], [O III], and [O II]. We classified four sources as AGN Type 2, one as AGN Type 1, four as QSOs and one as a star-forming galaxy, which agrees with the WISE colors of all sources and previous results.

Additionally, we compute the Baldwin, Phillips & Terlevich (BPT) diagrams (Baldwin et al., 1981) for sources with narrow lines with the objective of classifying them as either Seyfert 2 or star-forming galaxies. We present the BPT diagram in Figure 5.10. Our findings indicate that all narrow-line sources, except for SDSS J143010.96+230134.7, fall outside the region corresponding to star-forming galaxies on the plot. This region is delineated by the theoretical line of (Kewley et al., 2001; Kewley et al., 2013). Consequently, we classify these sources as Seyfert 2 galaxies. For 3PBCJ1504.1-6019 and 3PBCJ0800.7-4308, the analysis we carried out is in agreement with that of Landi et al., 2017, since we found the same soft X-ray sources lying within the positional uncertainty of these BAT unidentified objects. Inspecting NED and SIMBAD databases, we found the spectra for the soft X-ray counterparts of the UHXs 3PBCJ1329.7-1052 and 3PBCJ1854.4-3436. These are MCG -02-34-058 and ESO 396- G 007, lying at z=0.021648 and z=0.019483 (Jones et al., 2009), respectively. The *Swift*-XRT counterparts we assigned for the sources 3PBCJ2136.1+2002, 3PBCJ2155.3+6204, and 3PBCJ2238.8+4050 are the same as previously assigned in the literature at which optical spectroscopic follow-up observations revealed these to be three active galaxies lying at z=0.081 (Sy1), z=0.058 (Sy1) and z=0.055 (LINER), respectively, as reported in (Rojas et al., 2017). The yet unidentified source 3PBCJ0024.1-6823 having the *Swift*-XRT counterpart SWXRTJ002406.457-682052.549 could also be associated with the radio source PKS 0021-686, a gamma-ray blazar candidate selected based on its mid-IR colors (D'Abrusco et al., 2012b, 2014; D'Abrusco et al., 2019).

5.6 SUMMARY AND CONCLUSIONS

The main goal of the present analysis is to prepare a catalog of candidate soft X-ray *Swift*-XRT counterparts detected in the 0.5–10 keV energy range to list potential targets for the optical spectroscopic campaign, aiming at the classification of the yet unknown sources in the hard X-ray sky so we could obtain a more complete overview of it. We found archival *Swift*-XRT observations for 192 of the 218 3PBC sources marked as unidentified in our previous analyses. Those were the hard X-ray sources lacking an assigned low-energy counterpart. In this work, we searched for possible counterparts at soft X-ray energies for those 192 3PBC sources. If found, we carried out the literature search and multiwavelength analyses as done in the paper I.

We found that only in 172 out of 192 sources, there is at least one soft X-ray detected source above our S/N threshold of 3 present in the cleaned and merged event file, and only in 73 of the 3PBC sources we find at least one soft

X-ray candidate counterpart detected within the BAT positional uncertainty region. In particular, for 13 3PBC sources, there are multiple detected soft X-ray objects, while all remaining 60 3PBC sources have only a single *Swift*-XRT detected object. Thus, including multiple matches, the total number of *Swift*-XRT detected possible counterparts inside the BAT positional uncertainty, listed in our final catalog, is 93, sampling 73 3PBC hard X-ray sources.

Our X-ray results are in agreement with those achieved simply crossmatching the catalog of 3PBC unidentified sources with the 2SXPS, with only marginal differences, as reported in Section 5.4 mainly due to (i) longer exposure times and new observations collected after its release that were considered in our analysis and (ii) a small difference in the detection threshold chosen between the two analyses.

We found available spectra in the literature for 10 detected counterparts. For those, we carried the same multifrequency analyses as in paper I. We found four sources to be quasars, four sources to be Type 2 AGN, and one source to be Type 1 AGN, and one star-forming galaxy. The present analyses thus decreased the 218 3PBC unidentified hard X-ray sources sample to 143, which remain unidentified, lacking any low-frequency counterpart. This corresponds to a decrease by a factor of $\sim 34\,\%$. From the 73 3PBC sources for which we found at least one assigned candidate counterpart, 10 were classified according to our multifrequency criteria, becoming identified. The remaining 65 sources are left as unclassified, indicating that they lack spectroscopic information for their classification and are thus excellent candidates for future spectroscopic follow-up observations.

Along with this publication, we provide a catalog table of all 93 soft X-ray detections and a short table with our classification of the sources with spectra. The soft X-ray sources we found in this analysis can be targets of future spectroscopic campaigns aimed at classifying them to obtain redshift values and confirm that most of those having mid-IR detection are AGNs, as expected by the mid-IR plot.

Acknowledgements

M. K. and N. W. are supported by the GACR grant 21-13491X. E. B. acknowledges NASA grant 80NSSC21K0653. M. K. was supported by the Italian Government Scholarship issued by the Italian MAECI. This work was partially supported by CONACyT (National Council of Science and Technology) research grants 280789 (Mexico) Funding for the Sloan Digital Sky Survey V has been provided by the Alfred P. Sloan Foundation, the Heising-Simons Foundation, the National Science Foundation, and the Participating Institutions. SDSS acknowledges support and resources from the Center for High-Performance Computing at the University of Utah. The SDSS web site is www.sdss.org. SDSS is managed by the Astrophysical Research Consortium for the Participating Institutions of the SDSS Collaboration including the Brazilian Participation Group, the Carnegie Institution for Science, Carnegie Mellon University, Center for Astrophysics | Harvard & Smithsonian (CfA), the Chilean Participation Group, the French Participation Group, Instituto de Astrofísica de Canarias, The Johns Hopkins University, Kavli Institute for the Physics and Mathematics of the Universe (IPMU) / University of Tokyo, the Korean Participation Group, Lawrence Berke-

ley National Laboratory, Leibniz Institut für Astrophysik Potsdam (AIP), Max-Planck-Institut für Astronomie (MPIA Heidelberg), Max-Planck-Institut für Astrophysik (MPA Garching), Max-Planck-Institut für Extraterrestrische Physik (MPE), National Astronomical Observatories of China, New Mexico State University, New York University, University of Notre Dame, Observatório Nacional / MCTI, The Ohio State University, Pennsylvania State University, Shanghai Astronomical Observatory, United Kingdom Participation Group, Universidad Nacional Autónoma de México, University of Arizona, University of Colorado Boulder, University of Oxford, University of Portsmouth, University of Utah, University of Virginia, University of Washington, University of Wisconsin, Vanderbilt University, and Yale University.

5.7 APPENDIX

5.7.1 *Appendix: Catalog of soft Swift-XRT candidate counterparts of the UHXs*

Here we report the first 10 lines of the two main tables, showing parameters derived from our analysis, for soft X-ray sources detected by *Swift*-XRT that could be candidate counterparts of the sample of unidentified 3PBC hard X-ray sources, namely: Tab. 22 and Tab. 23.

Complementary to the Fig. 5.9 and Sec. 5.5.2 we provide here a table of our spectral classification of the 10 sources for which we found available spectra in the SDSS archive (Tab. 24).

5.7.2 *Appendix: Images of all x flagged 3PBC sources*

This section shows images of all the 3PBC sources with exactly one soft *Swift*-XRT detection inside the BAT positional uncertainty region (red dashed circle). The green cross, if present, indicated that the source has a WISE counterpart. The black circle only highlights the position of the *Swift*-XRT detection, not its positional uncertainty region. One exception is 3PBCJ1730.0-3436, which has a cyan circle instead of black because the black would not be visible on top of the source's emission.

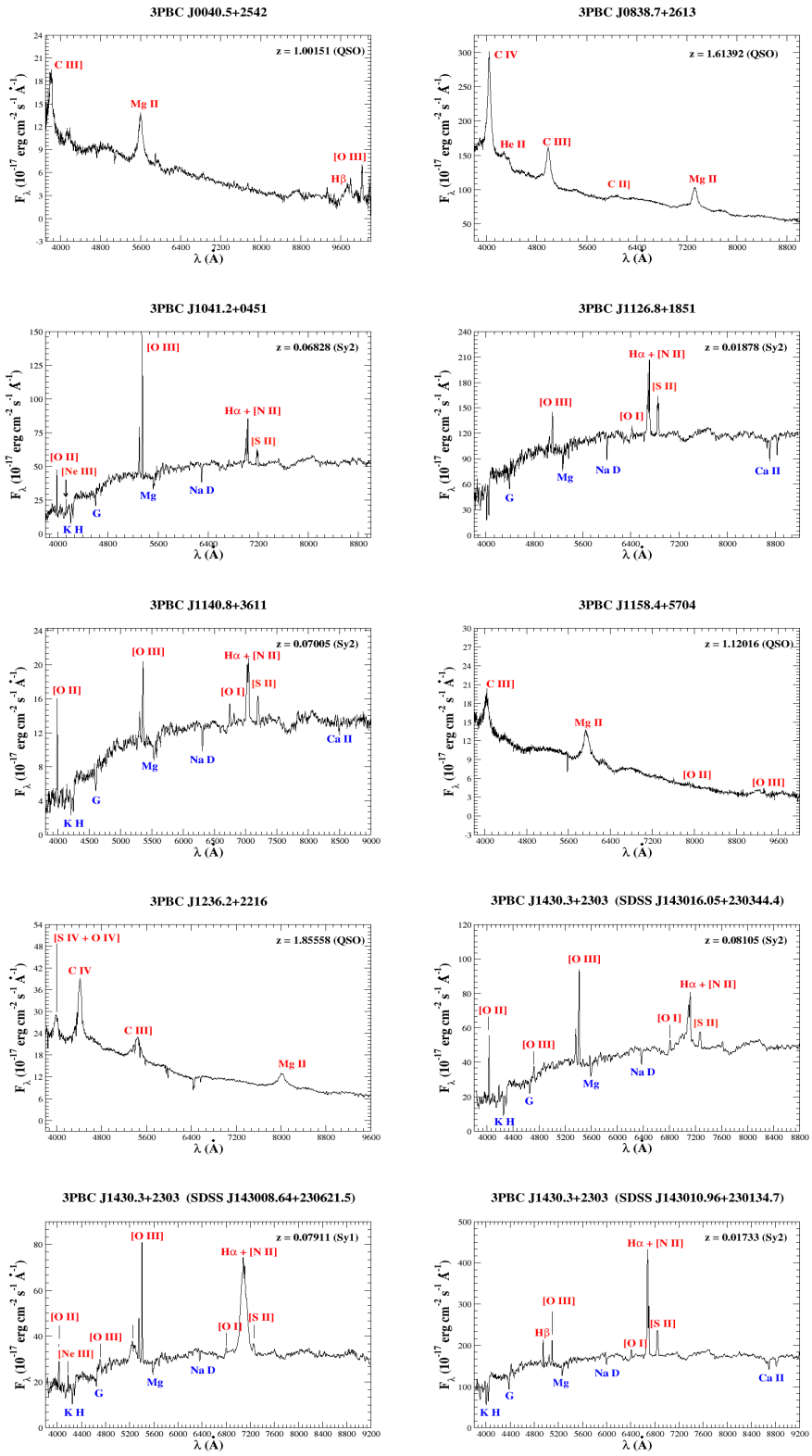

Figure 5.9: This figure shows the ten spectra we collected from the SDSS archive for the soft X-ray *Swift*-XRT sources detected in the BAT positional uncertainty region of the 3PBC sources. The upper left spectrum (J0040.5+2542) corresponds to a quasar. The upper middle spectrum (J0838.7+2613) is a quasar spectrum as well. We classified the upper right spectrum (J1041.2+0451) as a Type 2 AGN. The second row shows spectra of Type 2 AGN (left and middle) and quasar (right). The third row shows the spectra of a quasar (left), Type 2 AGN (middle), and Type 1 AGN (right). The last spectrum image on the bottom corresponds to a star-forming galaxy. The main spectral emission and/or absorption features are marked in each figure. We are reporting the Sloan spectra with the same redshift precision since none of the fitting they performed had warnings.

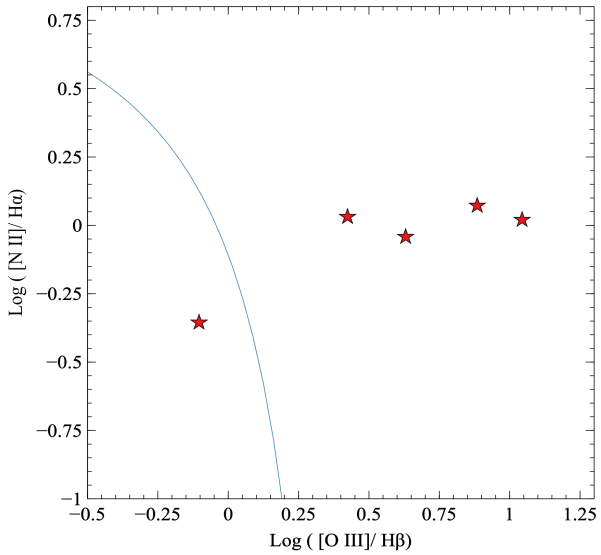

Figure 5.10: The BPT Diagram for distinguishing Type 2 and Star-Forming Galaxies. Error bars are all less than 0.009 and not visible in the plot. Note that all sources, except for SDSS J143010.96+230134.7, are above the theoretical line of Kewley et al., 2013 (right part of the graph separated by the blue line), which serves to discriminate between starburst regions and harder ionization sources.

Table 22: First ten lines of our final table of all 93 detected soft X-ray *Swift*-XRT sources within the BAT positional uncertainty in the total of 73 3PBC sources marked as unidentified in our previous analyses (paper I). Here are the first seven columns listing the 3PBC name with '3PBC' omitted to keep the table shorter, the name of an XRT candidate counterpart, which always starts with 'SWXRT' but was omitted here to keep the table smaller, the X-ray detection flag, the right ascension of the centroid, the declination of the centroid, the σ_{XRT} stands for XRT centroid positional uncertainty and $\vartheta_{XRT-BAT}$ is the angular separation between the XRT centroid and BAT centroid. The first row is the header, and the second row shows the units, if applicable.

3PBC name	XRT detection	XDF	RA centroid	DEC centroid	σ_{XRT}	$\vartheta_{XRT-BAT}$
			hh:mm:ss.ss	dd:mm:ss.ss	arcsec	arcsec
J0016.7-2611	J001637.005-261425.118	x	00:16:37.14	-26:14:26.22	4.8	193.5
J0022.2+2539	J002211.671+253830.384	m	00:22:11.73	25:38:31.64	5.0	39.8
J0022.2+2539	J002203.037+254003.324	m	00:22:03.09	25:40:02.51	3.6	134.1
J0022.2+2539	J002208.994+254004.792	m	00:22:09.44	25:39:57.24	4.3	59.9
J0024.1-6823	J002406.457-682052.549	x	00:24:06.47	-68:20:52.42	3.8	141.1
J0040.5+2542	J004024.511+254302.390	x	00:40:24.50	25:43:03.91	5.6	133.0
J0122.0-6105	J012204.938-610706.365	x	01:22:05.08	-61:07:07.16	4.0	103.4
J0132.5-7426	J013251.760-742547.319	x	01:32:52.00	-74:25:47.20	3.7	102.6
J0154.1-5034	J015422.390-503235.190	x	01:54:22.34	-50:32:41.76	5.4	166.4
J0158.9+2644	J015848.911+264247.120	x	01:58:48.68	26:42:47.38	5.8	115.1

Table 23: First ten lines of our final table of all 93 detected soft X-ray *Swift*-XRT sources within the BAT positional uncertainty in the total of 73 3PBC sources marked as unidentified in our previous analyses (paper I). Here are the last eight columns list the following parameters: the count rate, the delta of the count rate, Signal to Noise (S/N) in full band ([0.5, 10] keV), hard band ([2, 10] keV) and soft band ([0.5, 2] keV), the number of observations, N_{obs}, the approximated total integrated exposure time, T_{exp} and the name of the source if found in the WISE catalog. The first row is the header, and the second row shows the units, if applicable.

count rate	$\delta_{count\,rate}$	S/N full	S/N hard	S/N soft	N_{obs}	T_{exp}	WISE Name
photons/s	*photons/s*	[0.5, 10] keV	[2, 10] keV	[0.5, 2] keV		s	
3.33e-03	7.10e-04	4.7			4	8718	J001637.05-261426.7
2.63e-03	5.60e-04	4.7		3.8	6	12211	
1.58e-01	4.30e-03	37.1	22.5	31.4	6	12211	J002203.09+254003.2
1.35e-03	4.30e-04	3.1			6	12211	
4.77e-02	3.50e-03	13.8			5	5326	J002406.72-682054.5
2.89e-03	9.00e-04	3.2			5	4596	J004024.38+254303.4
1.90e-01	1.60e-02	11.9			1	975	J012205.12-610705.0
2.69e-02	1.50e-03	18.3			5	15292	J013251.50-742545.4
3.68e-03	9.10e-04	4.0			4	6035	J015422.40-503240.0
1.78e-03	5.70e-04	3.2			5	8247	J015848.81+264250.6

Table 24: Multiwavelength classification of the ten counterparts with available spectra in the SDSS. There are three sources with the same 3PBC name, each corresponding to a different *Swift*-XRT soft X-ray detection.

3PBC name	SDSS name	z	Class
3PBC J0040.5+2542	J004024.39+254303.4	1.0015	QSO
3PBC J0838.7+2613	J083850.15+261105.4	1.6139	QSO
3PBC J1041.2+0451	J104115.61+045313.8	0.0683	Sy 2
3PBC J1126.8+1851	J112655.08+184957.4	0.0188	Sy 2
3PBC J1140.8+3611	J114054.58+360957.0	0.0701	Sy 2
3PBC J1158.4+5704	J115843.32+570536.6	1.1202	QSO
3PBC J1236.2+2216	J123622.60+221834.8	1.8559	QSO
3PBC J1430.3+2303	J143016.05+230344.4	0.0811	Sy 2
3PBC J1430.3+2303	J143008.64+230621.5	0.0791	Sy 1
3PBC J1430.3+2303	J143010.96+230134.7	0.0173	Sy 2

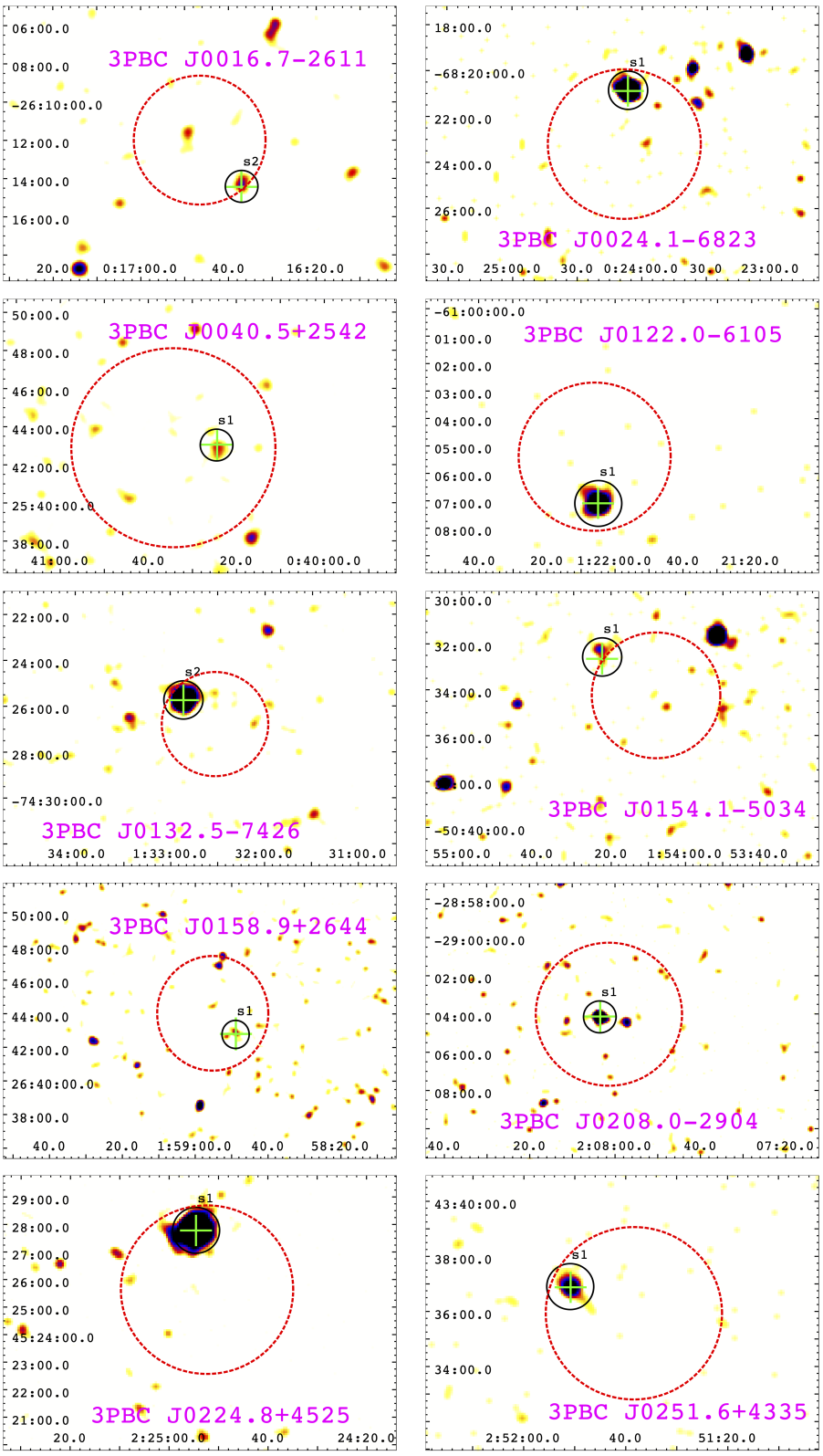

Figure 5.11: Images of 3PBC sources with exactly one soft *Swift*-XRT source (XDF flag *x*) detected inside of the BAT positional uncertainty region (red dashed circle). The soft X-ray detections are indicated with a black circle. The black circle indicates the position. It does not show the positional uncertainty of the source. If the soft X-ray detection is also marked with a green cross, it indicates that it has a WISE counterpart.

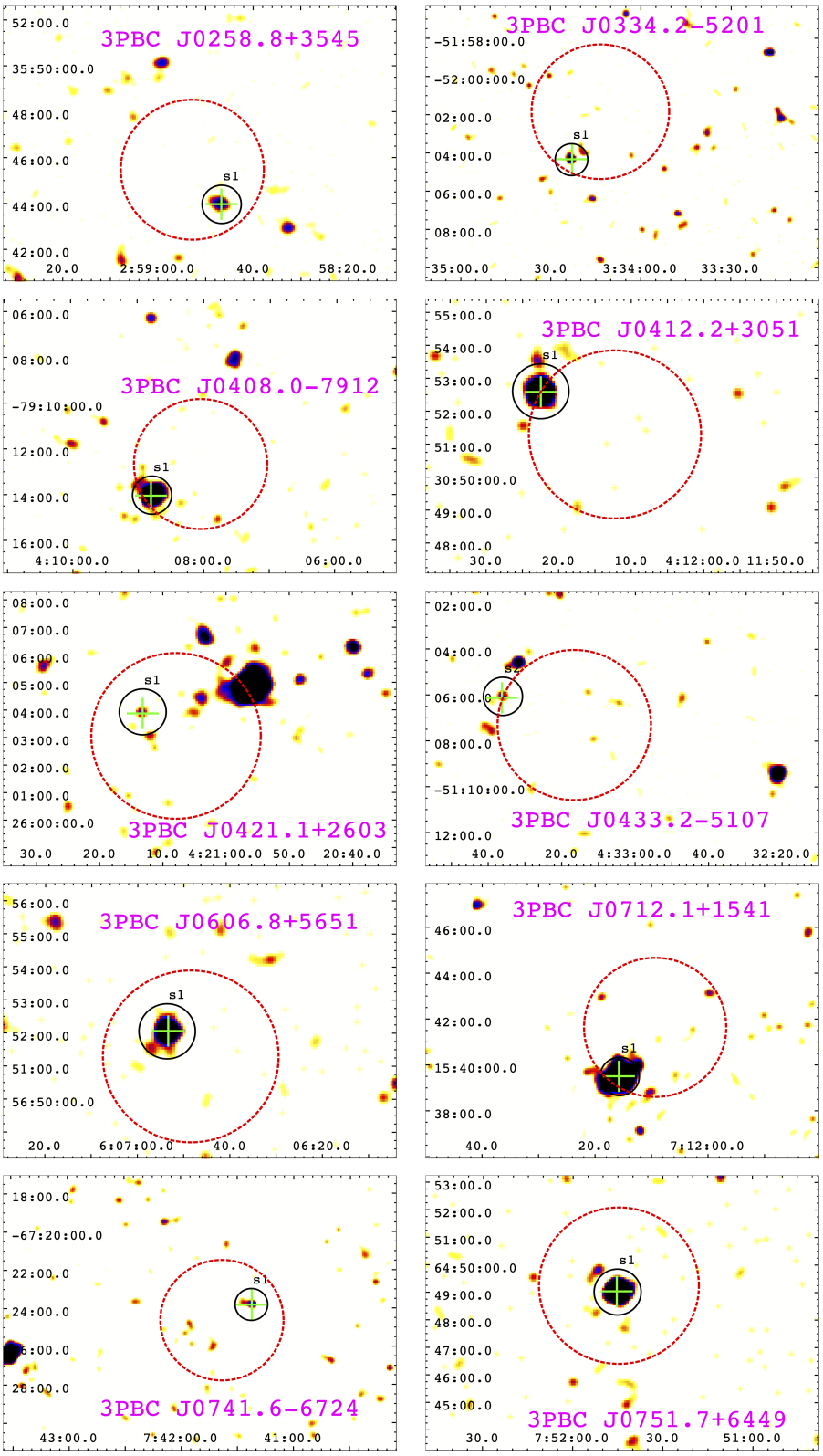

Figure 5.12: Images of 3PBC sources with exactly one soft *Swift*-XRT source (XDF flag *x*) detected inside of the BAT positional uncertainty region (red dashed circle). The soft X-ray detections are indicated with a black circle. The black circle indicates the position. It does not show the positional uncertainty of the source. If the soft X-ray detection is also marked with a green cross, it indicates that it has a WISE counterpart.

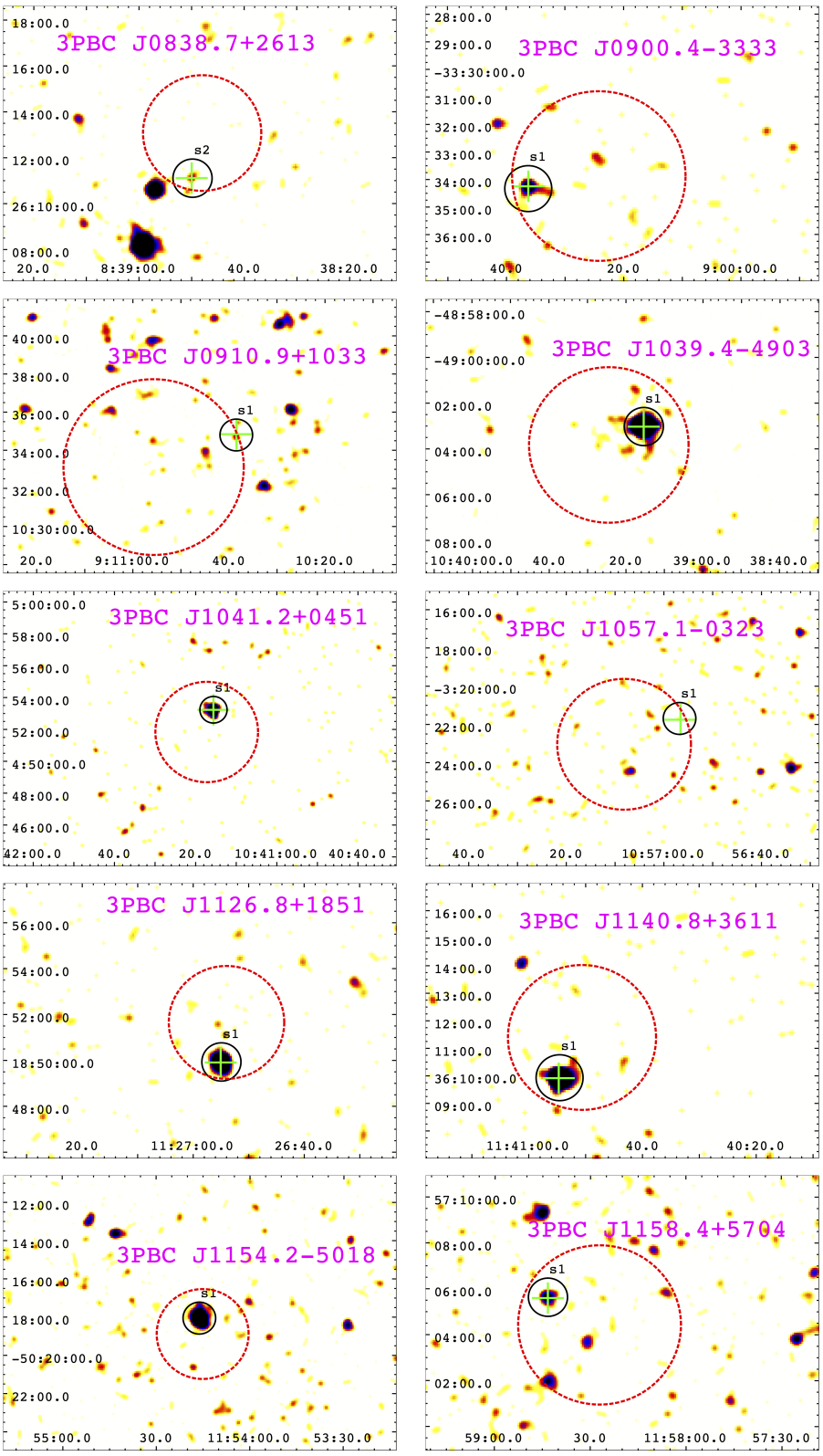

Figure 5.13: Images of 3PBC sources with exactly one soft *Swift*-XRT source (XDF flag x) detected inside of the BAT positional uncertainty region (red dashed circle). The soft X-ray detections are indicated with a black circle. The black circle indicates the position. It does not show the positional uncertainty of the source. If the soft X-ray detection is also marked with a green cross, it indicates that it has a WISE counterpart.

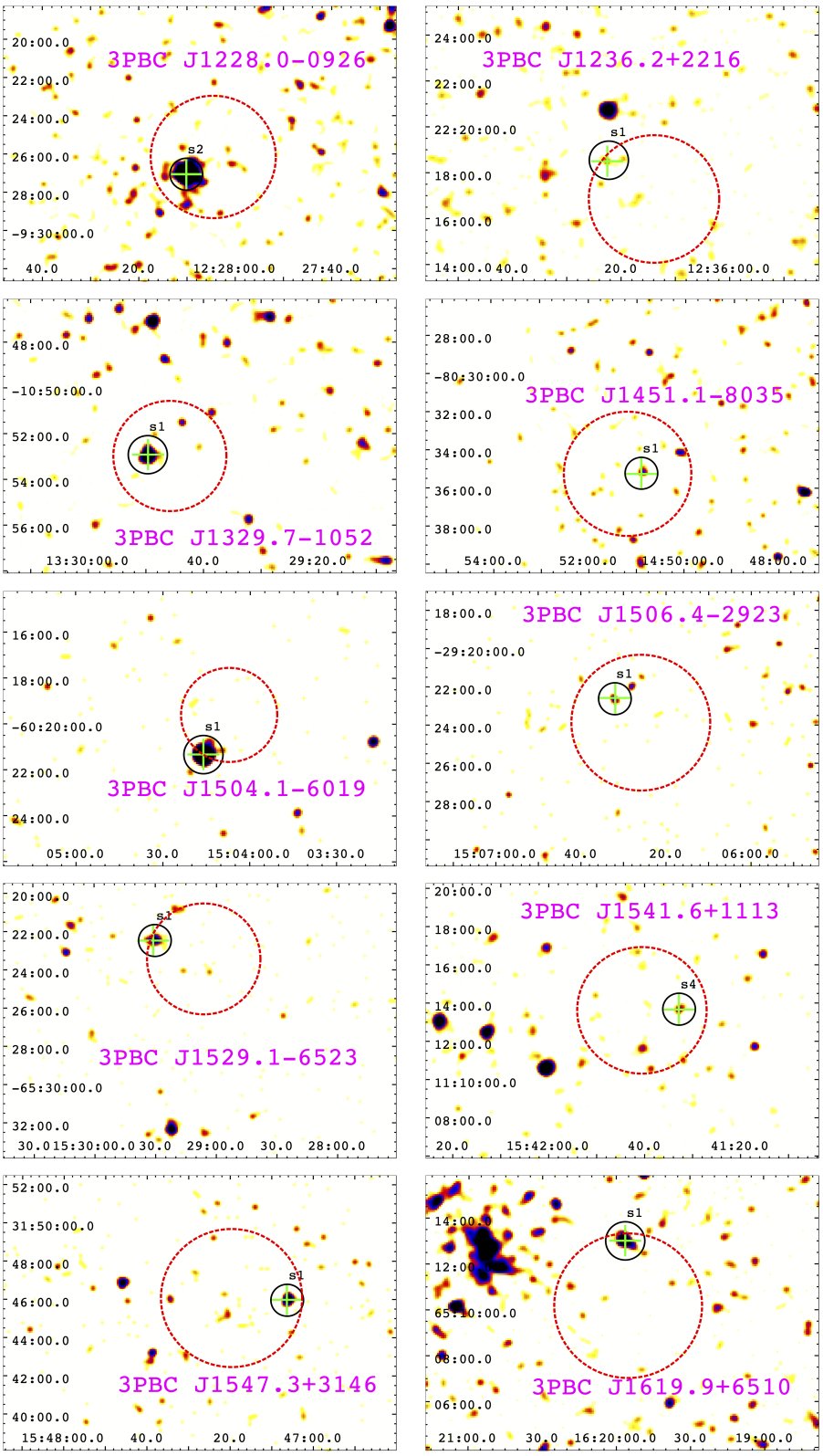

Figure 5.14: Images of 3PBC sources with exactly one soft *Swift*-XRT source (XDF flag *x*) detected inside of the BAT positional uncertainty region (red dashed circle). The soft X-ray detections are indicated with a black circle. The black circle indicates the position. It does not show the positional uncertainty of the source. If the soft X-ray detection is also marked with a green cross, it indicates that it has a WISE counterpart.

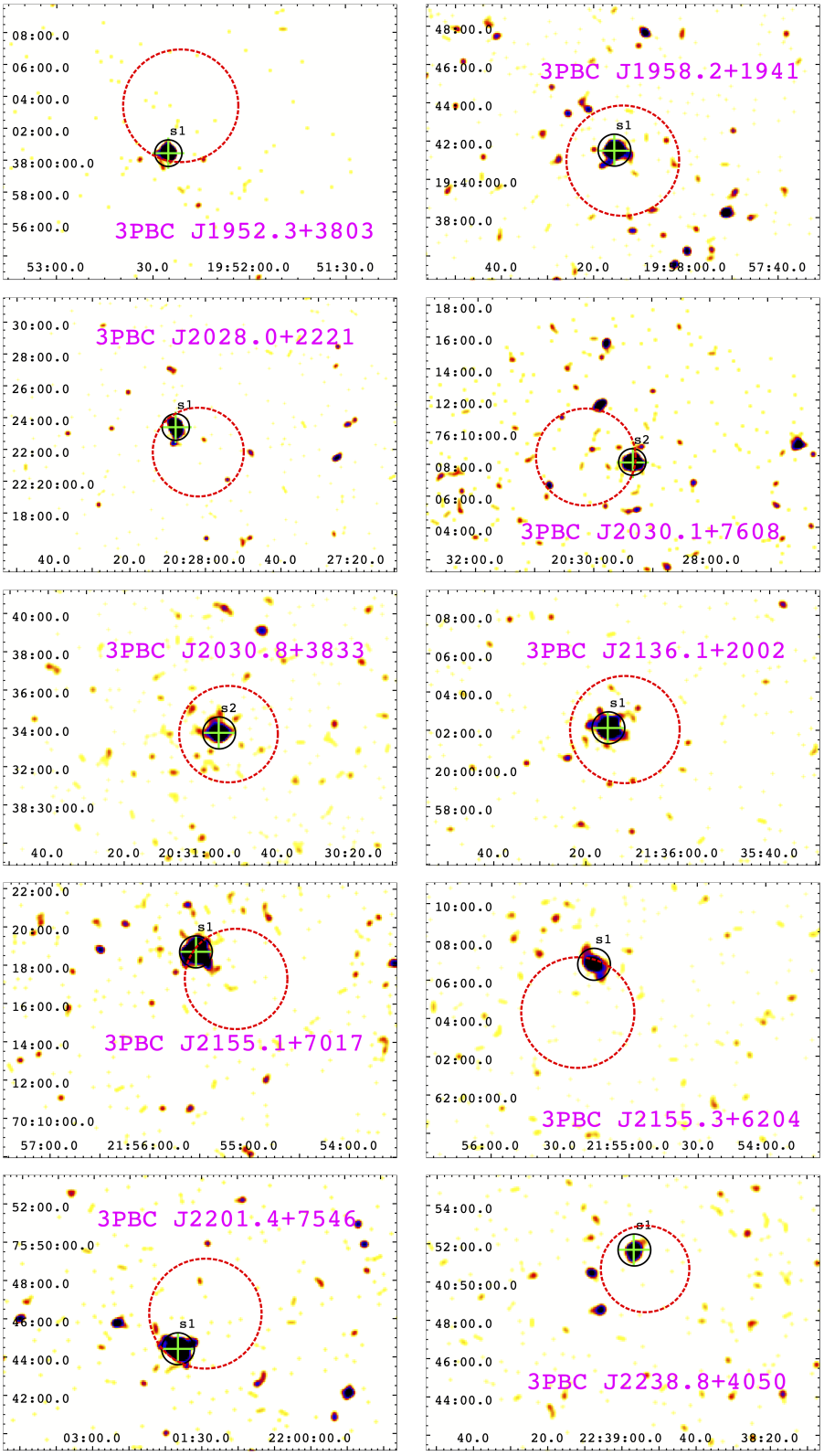

Figure 5.15: Images of 3PBC sources with exactly one soft *Swift*-XRT source (XDF flag *x*) detected inside of the BAT positional uncertainty region (red dashed circle). The soft X-ray detections are indicated with a black circle. The black circle indicates the position. It does not show the positional uncertainty of the source. If the soft X-ray detection is also marked with a green cross, it indicates that it has a WISE counterpart.

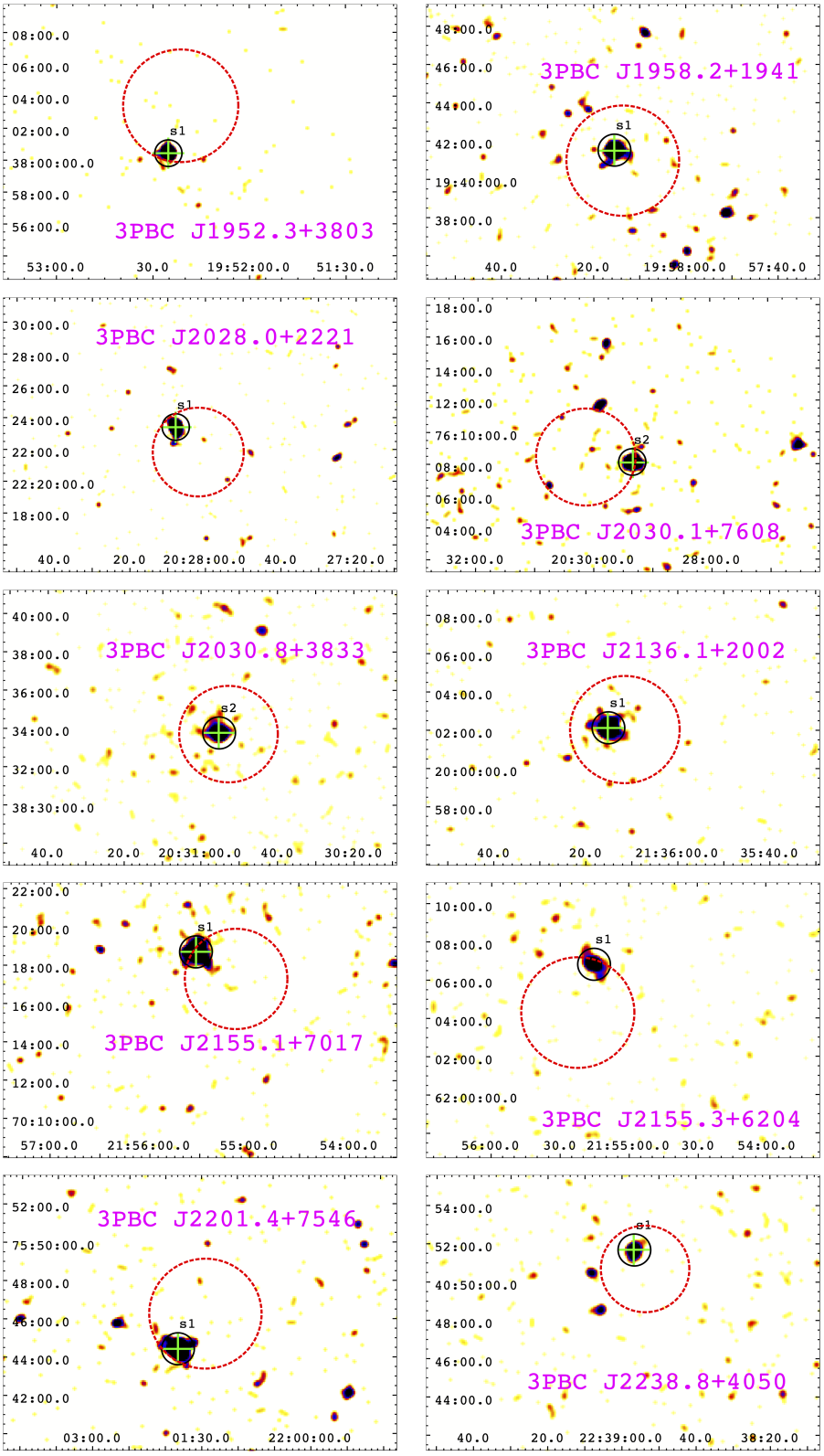

Figure 5.16: Images of 3PBC sources with exactly one soft *Swift*-XRT source (XDF flag *x*) detected inside of the BAT positional uncertainty region (red dashed circle). The soft X-ray detections are indicated with a black circle. The black circle indicates the position. It does not show the positional uncertainty of the source. If the soft X-ray detection is also marked with a green cross, it indicates that it has a WISE counterpart.

6

Summary & Future Perspective

6.1 Multiwavelength classification of X-ray-selected galaxy cluster candidates using convolutional neural networks

The aim of my Multiwavelength classification of X-ray-selected galaxy cluster candidates using convolutional neural networks project (Kosiba et al., 2020) was to create a galaxy cluster detection method that could replace the serendipitous manual classification of galaxy cluster candidates. I have developed a custom convolutional neural network architecture to classify galaxy cluster candidates in combined X-ray and optical wavelengths. This combination was especially complementary because cluster selection based on any of the two wavelengths alone has its own intrinsic biases.

- I trained my custom convolutional neural network on images of 1 100 cluster candidates and tested on a sample of 85 spectroscopically confirmed galaxy clusters and 85 sources classified as non-cluster detections by X-CLASS collaboration. It achieved 90 % averaged accuracy over ten runs.
- I trained this network also in a multi-class classification regime, classifying candidate clusters to categories *nearby cluster*, *distant cluster*, *point source*, *nearby galaxy* and *other*. My custom network achieved $\sim$ 85 % overall accuracy in this mode.
- I also used state-of-the-art architectures available during this research (e.g., MobileNet, ResNet, VGG19) in a transfer learning regime. The MobileNet CNN performed slightly better in this regime than our custom network, achieving $\sim$ 86 % overall accuracy. This result was within 1 σ with our custom network.
- I also created a Zooniverse citizen science project, *The Hunt for Galaxy Clusters* to acquire a large sample of classified galaxy cluster candidates that could potentially enhance our CNNs.
- Regarding the Zooniverse project, I communicated with the general public on social media platforms, working on public outreach promotion of the project and the forum discussion with the volunteers (more than 1 400 individuals and more than 3 000 comments). Once finished, I coded and developed a statistical analysis to analyze the metadata provided by the Zooniverse platform and created a metric to evaluate the citizen classifications. I compared them with the expert classifications of the X-CLASS collaboration on an overlap sample of 404 sources. The Zooniverse volunteers achieved only 62 % agreement with the X-CLASS experts. However, they provided a very clean sample of galaxy clusters, incorrectly classifying only 1 non-cluster as a galaxy cluster. Their bias was mainly in preferential classification towards *point source* or *no emission* Zooniverse categories.

I have created a convolutional neural network achieving 90 % accuracy for the classification of galaxy cluster candidates. This alone, even though providing good results, would not be sufficient for automatically creating a galaxy cluster catalog aimed at cosmological studies because these require even higher purity of the cluster sample. Retraining my network on a larger sample, possibly using simulations to cover some peculiar types of sources and XMM artifacts in greater detail, could also be a beneficial step forward in achieving better performance. Using spectroscopic information on the galaxies in the optical images would also help the training. I would also like to point out that my neural network was tested on a sample of 85 spectroscopically confirmed clusters but also on 85 objects that were classified as non-clusters by us in the X-CLASS collaboration, and we can not be 100 % sure that these X-CLASS classifications are actually correct due to the nature of the problem. This is especially a place where the simulations would be useful.

6.2 THE COSMOLOGICAL ANALYSIS OF X-RAY CLUSTER SURVEYS: VI. INFERENCE BASED ON ANALYTICALLY SIMULATED OBSERVABLE DIAGRAMS

The goal of my study; The cosmological analysis of X-ray cluster surveys: VI. Inference based on analytically simulated observable diagrams (Kosiba et al., 2024, submitted) was to develop a novel method for cosmological analysis. A method that could predict the Ω_m and σ_8 cosmological parameter based purely on galaxy cluster's observable characteristics, the X-ray count rate, hardness ratio, and spectroscopic redshift information independently of scaling relations.

- I have developed a scaling-relations-free method of cosmological analysis based on observable characteristics of galaxy clusters (CR, HR, z) employing simulation-based inference tasked to predict the posterior probability distribution of Ω_m and σ_8.
- For this task, I used analytically simulated X-ray observable diagrams (XOD) of galaxy cluster's distribution in a CR, HR, z space provided by colleagues. Next, I trained a ResNet-based model as a regressor to compress the XOD diagrams. I used the SNPE-A simulation-based inference method to train a mixture density neural network that models the posterior probability distribution of the Ω_m and σ_8 for a target diagram.
- Finally, this method is independent of scaling relations of galaxy clusters because these were drawn from a random uniform distribution during simulations and are used only to get the desired observable cluster quantities of the analytically simulated halos. The scaling relations never enter the ML-based model at any step. The method thus makes predictions independently of any apriori knowledge of scaling relations.
- I used this method to estimate the posterior probability distribution of the Ω_m and σ_8 for a target simulated XOD in a case of a $1\,000\,\mathrm{deg}^2$ and $10\,000\,\mathrm{deg}^2$. The XODs have a flux cut as a selection function designed to resemble the cluster number count of the XXL survey statistics.
- Following relative 1 σ errors were achieved on one of the target testing simulations: $1\,000\,\mathrm{deg}^2$: 15.2% for Ω_m and 10.0% for σ_8; $10\,000\,\mathrm{deg}^2$: 9.6% for Ω_m and 5.6% for σ_8.

The aim to create a novel approach for cosmological analysis that would be independent of the use of scaling relations and mass measurements was achieved. So far, I tested this method on simulated XODs for a scenario of $1\,000\,\mathrm{deg}^2$ and $10\,000\,\mathrm{deg}^2$ survey area. In the follow-up analysis, we plan to use this method to measure the Ω_m and σ_8 from a real XOD. To do this, we will need to implement some additional features, e.g., our selection function based on a simple flux-cut designed to mimic the cluster number count of the XXL survey is satisfactory for the present analysis, but it will need to be refined for a more realistic selection function that will need to be two dimensional (flux vs apparent size). Next, we will have to consider measurement errors in the simulation process of our XODs.

6.3 A MULTIFREQUENCY CHARACTERIZATION OF THE EXTRAGALACTIC HARD X-RAY SKY, PRESENTING THE SECOND RELEASE OF THE TURIN-SYCAT

The goal of my project, multifrequency characterization of the extragalactic hard X-ray sky, presenting the second release of the Turin-SyCAT (Kosiba et al., 2023), was to get a full characterization of the hard X-ray background. To do this, I aimed to complete soft energetic counterparts for all hard X-ray 3PBC sources and provide their classification based on our multifrequency criteria (e.g., spectral features, luminosities).

- The main results of this work are two catalogs, the revised version of the 3PBC and the second release of the Turin-SyCAT catalog of Seyfert galaxies.
- The revised version of the 3PBC catalog is a stand-alone catalog that reports 1 176 classified sources, from which 820 were of extragalactic origin and 356 were Galactic sources. Considering the remaining 417 sources, 218 were classified as unidentified (lacking low-energy counterpart) and 199 as unclassified (not meeting the classification criteria of our multi-wavelength analysis, e.g., due to missing spectra in the literature). This corresponds to a decrease of the fraction of 3PBC classified uncertain sources by 76 % and of the 3PBC unclassified category by 16 %.
- Concerning the first Turin-SyCAT, I found 282 new Seyfert galaxies. I published them together with the former sources in the second release of the Turin-SyCAT that are presented in this publication, now containing 633 Seyfert galaxies, an increase of 80 % with respect to the first version.

I managed to find some new low-energy counterparts and mainly to provide detailed classification based on spectral characteristics (the width of spectral lines) and luminosities for sources that lacked this information in the original 3PBC catalog. However, in the revised 3PBC catalog, 218 unidentified and 199 unclassified sources remained. These were the objects of study for my last project.

6.4 *SWIFT*-XRT FOLLOW-UP ANALYSIS OF UNIDENTIFIED HARD X-RAY SOURCES

The last project of my PhD research is the *Swift*-XRT follow-up analysis of unidentified hard X-ray sources (Kosiba et. al., submitted 2023). This is a follow-up analysis of the previous project with the same main aim: to provide a complete overview of the cosmic X-ray background. This project aimed to search for low-energy counterparts in the *Swift*-XRT archival soft X-ray data of the 218 yet unidentified sources we provided in the revised 3PBC catalog of the previous project.

- A total of 1213 archival *Swift*-XRT observations have been found for 192 of the 218 unidentified sources obtained between 2005, January 1st, and 2018, August 1st.
- From these 93 possible *Swift*-XRT counterparts were detected inside of the *Swift*-BAT positional uncertainty regions. These correspond to 73 3PBC sources, where 60 have only a single *Swift*-XRT detection, and 13 sources have multiple detections.
- The main result of this work is a catalog of these low-energy counterparts, which I present to the scientific community as potential candidates for future spectroscopic follow-up targets.
- I also provide a short catalog of 10 of these counterparts for which I found available spectra in the literature, and I present them together with corresponding classifications.

In this follow-up study, we managed to find candidates for low-energy counterparts for 73 of the 218 yet unidentified sources. There are also the 199 yet unclassified sources, which are identified with a counterpart, but we are missing some information (e.g., spectra or luminosities) to identify these based on our multifrequency criteria. My contribution to this field was to decrease the number of unknown hard X-ray sources, but there are still some left that need to be analyzed in future work.

Bibliography

[1]M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng, *TensorFlow: large-scale machine learning on heterogeneous systems*, Software available from tensorflow.org, 2015.

[2]K. Abazajian, G. M. Fuller, and W. H. Tucker, "Direct Detection of Warm Dark Matter in the X-Ray," **562**, 593–604 (2001).

[3]A. A. Abdo, M. Ackermann, M. Ajello, W. B. Atwood, M. Axelsson, L. Baldini, J. Ballet, G. Barbiellini, D. Bastieri, B. M. Baughman, K. Bechtol, R. Bellazzini, B. Berenji, E. D. Bloom, G. Bogaert, E. Bonamente, A. W. Borgland, J. Bregeon, A. Brez, M. Brigida, P. Bruel, T. H. Burnett, G. A. Caliandro, R. A. Cameron, P. A. Caraveo, J. M. Casandjian, E. Cavazzuti, C. Cecchi, Ö. Çelik, A. Chekhtman, C. C. Cheung, J. Chiang, S. Ciprini, R. Claus, J. Cohen-Tanugi, J. Conrad, S. Cutini, C. D. Dermer, A. de Angelis, F. de Palma, S. W. Digel, E. d. C. e. Silva, P. S. Drell, R. Dubois, D. Dumora, C. Farnier, C. Favuzzi, S. J. Fegan, E. C. Ferrara, W. B. Focke, M. Frailis, L. Fuhrmann, Y. Fukazawa, S. Funk, P. Fusco, F. Gargano, D. Gasparrini, N. Gehrels, S. Germani, B. Giebels, N. Giglietto, F. Giordano, M. Giroletti, T. Glanzman, G. Godfrey, I. A. Grenier, M. H. Grondin, J. E. Grove, L. Guillemot, S. Guiriec, Y. Hanabata, A. K. Harding, M. Hayashida, E. Hays, R. E. Hughes, G. Jóhannesson, A. S. Johnson, R. P. Johnson, W. N. Johnson, M. Kadler, T. Kamae, H. Katagiri, J. Kataoka, M. Kerr, J. Knödlseder, M. L. Kocian, F. Kuehn, M. Kuss, J. Lande, L. Latronico, M. Lemoine-Goumard, F. Longo, F. Loparco, B. Lott, M. N. Lovellette, P. Lubrano, G. M. Madejski, A. Makeev, M. Marelli, E. Massaro, W. Max-Moerbeck, M. N. Mazziotta, W. McConville, J. E. McEnery, C. Meurer, P. F. Michelson, W. Mitthumsiri, T. Mizuno, A. A. Moiseev, C. Monte, M. E. Monzani, A. Morselli, I. V. Moskalenko, S. Murgia, P. L. Nolan, J. P. Norris, E. Nuss, T. Ohsugi, N. Omodei, E. Orlando, J. F. Ormes, M. Ozaki, D. Paneque, J. H. Panetta, D. Parent, V. Pavlidou, T. J. Pearson, V. Pelassa, M. Pepe, M. Pesce-Rollins, F. Piron, T. A. Porter, S. Rainò, R. Rando, M. Razzano, S. Razzaque, A. Readhead, A. Reimer, O. Reimer, T. Reposeur, J. L. Richards, S. Ritz, L. S. Rochester, A. Y. Rodriguez, R. W. Romani, M. Roth, F. Ryde, H. F. W. Sadrozinski, D. Sanchez, A. Sander, P. M. Saz Parkinson, J. D. Scargle, C. Sgrò, M. S. Shaw, E. J. Siskind, D. A. Smith, P. D. Smith, G. Spandre, P. Spinelli, M. Stevenson, M. S. Strickman, D. J. Suson, H. Tajima, H. Takahashi, T. Tanaka, J. B. Thayer, J. G. Thayer, D. J. Thompson, L. Tibaldo, O. Tibolla, D. F. Torres, G. Tosti, A. Tramacere, P. Ubertini, Y. Uchiyama, T. L. Usher, V. Vasileiou, N. Vilchez, V. Vitale, A. P. Waite, P. Wang, B. L. Winer, K. S. Wood, H. Yasuda, T. Ylinen, J. A. Zensus, M. Ziegler, Fermi LAT Collaboration, E. Angelakis, T. Hovatta, E. Hoversten, Y. Ikejiri, K. S. Kawabata, Y. Y. Ko-

valev, Y. A. Kovalev, T. P. Krichbaum, M. L. Lister, A. Lähteenmäki, N. Marchili, P. Ogle, C. Pagani, A. B. Pushkarev, K. Sakimoto, M. Sasada, M. Tornikoski, M. Uemura, M. Yamanaka, T. Yamashita, Fermi LAT Collaboration, and Multifrequency Campaign Collaboration, "PKS 1502+106: A New and Distant Gamma-ray Blazar in Outburst Discovered by the Fermi Large Area Telescope," **710**, 810–827 (2010).

[4]Abdurro'uf, K. Accetta, C. Aerts, V. Silva Aguirre, R. Ahumada, N. Ajgaonkar, N. Filiz Ak, S. Alam, C. Allende Prieto, A. Almeida, F. Anders, S. F. Anderson, B. H. Andrews, B. Anguiano, E. Aquino-Ortiz, A. Aragon-Salamanca, M. Argudo-Fernandez, M. Ata, M. Aubert, V. Avila-Reese, C. Badenes, R. H. Barba, K. Barger, J. K. Barrera-Ballesteros, R. L. Beaton, T. C. Beers, F. Belfiore, C. F. Bender, M. Bernardi, M. A. Bershady, F. Beutler, C. Moni Bidin, J. C. Bird, D. Bizyaev, G. A. Blanc, M. R. Blanton, N. F. Boardman, A. S. Bolton, M. Boquien, J. Borissova, J. Bovy, W. N. Brandt, J. Brown, J. R. Brownstein, M. Brusa, J. Buchner, K. Bundy, J. N. Burchett, M. Bureau, A. Burgasser, T. K. Cabang, S. Campbell, M. Cappellari, J. K. Carlberg, F. Carneiro Wanderley, R. Carrera, J. Cash, Y.-P. Chen, W.-H. Chen, B. Cherinka, C. Chiappini, P. D. Choi, S. D. Chojnowski, H. Chung, N. Clerc, R. E. Cohen, J. M. Comerford, J. Comparat, L. da Costa, K. Covey, J. D. Crane, I. Cruz-Gonzalez, C. Culhane, K. Cunha, Y. S. Dai, G. Damke, J. Darling, J. Davidson James W., R. Davies, K. Dawson, N. De Lee, A. M. Diamond-Stanic, M. Cano-Diaz, H. Dominguez Sanchez, J. Donor, C. Duckworth, T. Dwelly, D. J. Eisenstein, Y. P. Elsworth, E. Emsellem, M. Eracleous, S. Escoffier, X. Fan, E. Farr, S. Feng, J. G. Fernandez-Trincado, D. Feuillet, A. Filipp, S. P. Fillingham, P. M. Frinchaboy, S. Fromenteau, L. Galbany, R. A. Garcia, D. A. Garcia-Hernandez, J. Ge, D. Geisler, J. Gelfand, T. Geron, B. J. Gibson, J. Goddy, D. Godoy-Rivera, K. Grabowski, P. J. Green, M. Greener, C. J. Grier, E. Griffith, H. Guo, J. Guy, M. Hadjara, P. Harding, S. Hasselquist, C. R. Hayes, F. Hearty, J. Hernandez, L. Hill, D. W. Hogg, J. A. Holtzman, D. Horta, B.-C. Hsieh, C.-H. Hsu, Y.-H. Hsu, D. Huber, M. Huertas-Company, B. Hutchinson, H. S. Hwang, H. J. Ibarra-Medel, J. Ider Chitham, G. S. Ilha, J. Imig, W. Jaekle, T. Jayasinghe, X. Ji, J. A. Johnson, A. Jones, H. Jonsson, I. Katkov, D. Khalatyan Arman, K. Kinemuchi, S. Kisku, J. H. Knapen, J.-P. Kneib, J. A. Kollmeier, M. Kong, M. Kounkel, K. Kreckel, D. Krishnarao, I. Lacerna, R. R. Lane, R. Langgin, R. Lavender, D. R. Law, D. Lazarz, H. W. Leung, H.-H. Leung, H. M. Lewis, C. Li, R. Li, J. Lian, F.-H. Liang, L. Lin, Y.-T. Lin, S. Lin, C. Lintott, D. Long, P. Longa-Pena, C. Lopez-Coba, S. Lu, B. F. Lundgren, Y. Luo, J. T. Mackereth, A. de la Macorra, S. Mahadevan, S. R. Majewski, A. Manchado, T. Mandeville, C. Maraston, B. Margalef-Bentabol, T. Masseron, K. L. Masters, S. Mathur, R. M. McDermid, M. Mckay, A. Merloni, M. Merrifield, S. Meszaros, A. Miglio, F. Di Mille, D. Minniti, R. Minsley, A. Monachesi, J. Moon, B. Mosser, J. Mulchaey, D. Muna, R. R. Munoz, A. D. Myers, N. Myers, S. Nadathur, P. Nair, K. Nandra, J. Neumann, J. A. Newman, D. L. Nidever, F. Nikakhtar, C. Nitschelm, J. E. O'Connell, L. Garma-Oehmichen, G. Luan Souza de Oliveira, R. Olney, D. Oravetz, M. Ortigoza-Urdaneta, Y. Osorio, J. Otter, Z. J. Pace, N. Padilla, K. Pan, H.-A. Pan, T. Parikh, J. Parker, S. Peirani, K. Pena Ramirez, S. Penny, W. J. Percival, I. Perez-Fournon, M. Pinsonneault, F. Poidevin, V. J. Poovelil, A. M. Price-Whelan, A. B. d. A.

Queiroz, M. J. Raddick, A. Ray, S. Barboza Rembold, N. Riddle, R. A. Riffel, R. Riffel, H.-W. Rix, A. C. Robin, A. Rodriguez-Puebla, A. Roman-Lopes, C. Roman-Zuniga, B. Rose, A. J. Ross, G. Rossi, K. H. R. Rubin, M. Salvato, S. F. Sanchez, J. R. Sanchez-Gallego, R. Sanderson, F. A. Santana Rojas, E. Sarceno, R. Sarmiento, C. Sayres, E. Sazonova, A. L. Schaefer, R. Schiavon, D. J. Schlegel, D. P. Schneider, M. Schultheis, A. Schwope, A. Serenelli, J. Serna, Z. Shao, G. Shapiro, A. Sharma, Y. Shen, M. Shetrone, Y. Shu, J. D. Simon, M. F. Skrutskie, R. Smethurst, V. Smith, J. Sobeck, T. Spoo, D. Sprague, D. V. Stark, K. G. Stassun, M. Steinmetz, D. Stello, A. Stone-Martinez, T. Storchi-Bergmann, G. S. Stringfellow, A. Stutz, Y.-C. Su, M. Taghizadeh-Popp, M. S. Talbot, J. Tayar, E. Telles, J. Teske, A. Thakar, C. Theissen, D. Thomas, A. Tkachenko, R. Tojeiro, H. Hernandez Toledo, N. W. Troup, J. R. Trump, J. Trussler, J. Turner, S. Tuttle, E. Unda-Sanzana, J. A. Vazquez-Mata, M. Valentini, O. Valenzuela, J. Vargas-Gonzalez, M. Vargas-Magana, P. V. Alfaro, S. Villanova, F. Vincenzo, D. Wake, J. T. Warfield, J. D. Washington, B. A. Weaver, A.-M. Weijmans, D. H. Weinberg, A. Weiss, K. B. Westfall, V. Wild, M. C. Wilde, J. C. Wilson, R. F. Wilson, M. Wilson, J. Wolf, W. M. Wood-Vasey, R. Yan, O. Zamora, G. Zasowski, K. Zhang, C. Zhao, Z. Zheng, Z. Zheng, and K. Zhu, "The Seventeenth Data Release of the Sloan Digital Sky Surveys: Complete Release of MaNGA, MaStar and APOGEE-2 Data," arXiv e-prints, arXiv:2112.02026, arXiv:2112.02026 (2021).

[5]G. O. Abell, "The Distribution of Rich Clusters of Galaxies.," **3**, 211 (1958).

[6]G. O. Abell, J. Corwin Harold G., and R. P. Olowin, "A Catalog of Rich Clusters of Galaxies," **70**, 1 (1989).

[7]A. Acharyya, P. M. Chadwick, and A. M. Brown, "Locating the gamma-ray emission region in the brightest Fermi-LAT flat-spectrum radio quasars," **500**, 5297–5321 (2021).

[8]S. Ackermann, K. Schawinski, C. Zhang, A. K. Weigel, and M. D. Turp, "Using transfer learning to detect galaxy mergers," **479**, 415–425 (2018).

[9]C. Adami, P. Giles, E. Koulouridis, F. Pacaud, C. A. Caretta, M. Pierre, D. Eckert, M. E. Ramos-Ceja, F. Gastaldello, S. Fotopoulou, V. Guglielmo, C. Lidman, T. Sadibekova, A. Iovino, B. Maughan, L. Chiappetti, S. Alis, B. Altieri, I. Baldry, D. Bottini, M. Birkinshaw, M. Bremer, M. Brown, O. Cucciati, S. Driver, E. Elmer, S. Ettori, A. E. Evrard, L. Faccioli, B. Granett, M. Grootes, L. Guzzo, A. Hopkins, C. Horellou, J. P. Lefèvre, J. Liske, K. Malek, F. Marulli, S. Maurogordato, M. Owers, S. Paltani, B. Poggianti, M. Polletta, M. Plionis, A. Pollo, E. Pompei, T. Ponman, D. Rapetti, M. Ricci, A. Robotham, R. Tuffs, L. Tasca, I. Valtchanov, D. Vergani, G. Wagner, J. Willis, and XXL Consortium, "The XXL Survey. XX. The 365 cluster catalogue," **620**, A5, A5 (2018).

[10]T. F. Adams, "A Survey of the Seyfert Galaxies Based on Large-Scale Image-Tube Plates," **33**, 19 (1977).

[11]C. P. Ahn, R. Alexandroff, C. Allende Prieto, S. F. Anderson, T. Anderton, B. H. Andrews, É. Aubourg, S. Bailey, E. Balbinot, R. Barnes, J. Bautista, T. C. Beers, A. Beifiori, A. A. Berlind, V. Bhardwaj, D. Bizyaev, C. H. Blake, M. R. Blanton, M. Blomqvist, J. J. Bochanski, A. S. Bolton, A. Borde, J.

Bovy, W. N. Brandt, J. Brinkmann, P. J. Brown, J. R. Brownstein, K. Bundy, N. G. Busca, W. Carithers, A. R. Carnero, M. A. Carr, D. I. Casetti-Dinescu, Y. Chen, C. Chiappini, J. Comparat, N. Connolly, J. R. Crepp, S. Cristiani, R. A. C. Croft, A. J. Cuesta, L. N. da Costa, J. R. A. Davenport, K. S. Dawson, R. de Putter, N. De Lee, T. Delubac, S. Dhital, A. Ealet, G. L. Ebelke, E. M. Edmondson, D. J. Eisenstein, S. Escoffier, M. Esposito, M. L. Evans, X. Fan, B. Femenía Castellá, E. Fernández Alvar, L. D. Ferreira, N. Filiz Ak, H. Finley, S. W. Fleming, A. Font-Ribera, P. M. Frinchaboy, D. A. García-Hernández, A. E. García Pérez, J. Ge, R. Génova-Santos, B. A. Gillespie, L. Girardi, J. I. González Hernández, E. K. Grebel, J. E. Gunn, H. Guo, D. Haggard, J.-C. Hamilton, D. W. Harris, S. L. Hawley, F. R. Hearty, S. Ho, D. W. Hogg, J. A. Holtzman, K. Honscheid, J. Huehnerhoff, I. I. Ivans, Ž. Ivezić, H. R. Jacobson, L. Jiang, J. Johansson, J. A. Johnson, G. Kauffmann, D. Kirkby, J. A. Kirkpatrick, M. A. Klaene, G. R. Knapp, J.-P. Kneib, J.-M. Le Goff, A. Leauthaud, K.-G. Lee, Y. S. Lee, D. C. Long, C. P. Loomis, S. Lucatello, B. Lundgren, R. H. Lupton, B. Ma, Z. Ma, N. MacDonald, C. E. Mack, S. Mahadevan, M. A. G. Maia, S. R. Majewski, M. Makler, E. Malanushenko, V. Malanushenko, A. Manchado, R. Mandelbaum, M. Manera, C. Maraston, D. Margala, S. L. Martell, C. K. McBride, I. D. McGreer, R. G. McMahon, B. Ménard, S. Meszaros, J. Miralda-Escudé, A. D. Montero-Dorta, F. Montesano, H. L. Morrison, D. Muna, J. A. Munn, H. Murayama, A. D. Myers, A. F. Neto, D. C. Nguyen, R. C. Nichol, D. L. Nidever, P. Noterdaeme, S. E. Nuza, R. L. C. Ogando, M. D. Olmstead, D. J. Oravetz, R. Owen, N. Padmanabhan, N. Palanque-Delabrouille, K. Pan, J. K. Parejko, P. Parihar, I. Pâris, P. Pattarakijwanich, J. Pepper, W. J. Percival, I. Pérez-Fournon, I. Pérez-Ràfols, P. Petitjean, J. Pforr, M. M. Pieri, M. H. Pinsonneault, G. F. Porto de Mello, F. Prada, A. M. Price-Whelan, M. J. Raddick, R. Rebolo, J. Rich, G. T. Richards, A. C. Robin, H. J. Rocha-Pinto, C. M. Rockosi, N. A. Roe, A. J. Ross, N. P. Ross, G. Rossi, J. A. Rubiño-Martin, L. Samushia, J. Sanchez Almeida, A. G. Sánchez, B. Santiago, C. Sayres, D. J. Schlegel, K. J. Schlesinger, S. J. Schmidt, D. P. Schneider, M. Schultheis, A. D. Schwope, C. G. Scóccola, U. Seljak, E. Sheldon, Y. Shen, Y. Shu, J. Simmerer, A. E. Simmons, R. A. Skibba, M. F. Skrutskie, A. Slosar, F. Sobreira, J. S. Sobeck, K. G. Stassun, O. Steele, M. Steinmetz, M. A. Strauss, A. Streblyanska, N. Suzuki, M. E. C. Swanson, T. Tal, A. R. Thakar, D. Thomas, B. A. Thompson, J. L. Tinker, R. Tojeiro, C. A. Tremonti, M. Vargas Magaña, L. Verde, M. Viel, S. K. Vikas, N. P. Vogt, D. A. Wake, J. Wang, B. A. Weaver, D. H. Weinberg, B. J. Weiner, A. A. West, M. White, J. C. Wilson, J. P. Wisniewski, W. M. Wood-Vasey, B. Yanny, C. Yèche, D. G. York, O. Zamora, G. Zasowski, I. Zehavi, G.-B. Zhao, Z. Zheng, G. Zhu, and J. C. Zinn, "The Ninth Data Release of the Sloan Digital Sky Survey: First Spectroscopic Data from the SDSS-III Baryon Oscillation Spectroscopic Survey," **203**, 21, 21 (2012).

[12]M. Ajello, L. Costamante, R. M. Sambruna, N. Gehrels, J. Chiang, A. Rau, A. Escala, J. Greiner, J. Tueller, J. V. Wall, and R. F. Mushotzky, "The Evolution of Swift/BAT Blazars and the Origin of the MeV Background," **699**, 603–625 (2009).

[13]M. Ajello, P. Rebusco, N. Cappelluti, O. Reimer, H. Böhringer, V. La Parola, and G. Cusumano, "Galaxy Clusters in the Swift/BAT Era. II. 10 More Clusters Detected Above 15 keV," **725**, 1688–1706 (2010).

[14]S. W. Allen, A. E. Evrard, and A. B. Mantz, "Cosmological Parameters from Observations of Galaxy Clusters," **49**, 409–470 (2011).

[15]R. R. J. Antonucci and J. S. Miller, "Spectropolarimetry and the nature of NGC 1068.," **297**, 621–632 (1985).

[16]R. Antonucci, "Unified models for active galactic nuclei and quasars.," **31**, 473–521 (1993).

[17]R. J. Assef, D. Stern, C. S. Kochanek, A. W. Blain, M. Brodwin, M. J. I. Brown, E. Donoso, P. R. M. Eisenhardt, B. T. Jannuzi, T. H. Jarrett, S. A. Stanford, C. W. Tsai, J. Wu, and L. Yan, "Mid-infrared Selection of Active Galactic Nuclei with the Wide-field Infrared Survey Explorer. II. Properties of WISE-selected Active Galactic Nuclei in the NDWFS Boötes Field," **772**, 26, 26 (2013).

[18]Astropy Collaboration, A. M. Price-Whelan, B. M. Sipőcz, H. M. Günther, P. L. Lim, S. M. Crawford, S. Conseil, D. L. Shupe, M. W. Craig, N. Dencheva, A. Ginsburg, J. T. Vand erPlas, L. D. Bradley, D. Pérez-Suárez, M. de Val-Borro, T. L. Aldcroft, K. L. Cruz, T. P. Robitaille, E. J. Tollerud, C. Ardelean, T. Babej, Y. P. Bach, M. Bachetti, A. V. Bakanov, S. P. Bamford, G. Barentsen, P. Barmby, A. Baumbach, K. L. Berry, F. Biscani, M. Boquien, K. A. Bostroem, L. G. Bouma, G. B. Brammer, E. M. Bray, H. Breytenbach, H. Buddelmeijer, D. J. Burke, G. Calderone, J. L. Cano Rodríguez, M. Cara, J. V. M. Cardoso, S. Cheedella, Y. Copin, L. Corrales, D. Crichton, D. D'Avella, C. Deil, É. Depagne, J. P. Dietrich, A. Donath, M. Droettboom, N. Earl, T. Erben, S. Fabbro, L. A. Ferreira, T. Finethy, R. T. Fox, L. H. Garrison, S. L. J. Gibbons, D. A. Goldstein, R. Gommers, J. P. Greco, P. Greenfield, A. M. Groener, F. Grollier, A. Hagen, P. Hirst, D. Homeier, A. J. Horton, G. Hosseinzadeh, L. Hu, J. S. Hunkeler, Ž. Ivezić, A. Jain, T. Jenness, G. Kanarek, S. Kendrew, N. S. Kern, W. E. Kerzendorf, A. Khvalko, J. King, D. Kirkby, A. M. Kulkarni, A. Kumar, A. Lee, D. Lenz, S. P. Littlefair, Z. Ma, D. M. Macleod, M. Mastropietro, C. McCully, S. Montagnac, B. M. Morris, M. Mueller, S. J. Mumford, D. Muna, N. A. Murphy, S. Nelson, G. H. Nguyen, J. P. Ninan, M. Nöthe, S. Ogaz, S. Oh, J. K. Parejko, N. Parley, S. Pascual, R. Patil, A. A. Patil, A. L. Plunkett, J. X. Prochaska, T. Rastogi, V. Reddy Janga, J. Sabater, P. Sakurikar, M. Seifert, L. E. Sherbert, H. Sherwood-Taylor, A. Y. Shih, J. Sick, M. T. Silbiger, S. Singanamalla, L. P. Singer, P. H. Sladen, K. A. Sooley, S. Sornarajah, O. Streicher, P. Teuben, S. W. Thomas, G. R. Tremblay, J. E. H. Turner, V. Terrón, M. H. van Kerkwijk, A. de la Vega, L. L. Watkins, B. A. Weaver, J. B. Whitmore, J. Woillez, V. Zabalza, and Astropy Contributors, "The Astropy Project: Building an Open-science Project and Status of the v2.0 Core Package," **156**, 123, 123 (2018).

[19]Astropy Collaboration, T. P. Robitaille, E. J. Tollerud, P. Greenfield, M. Droettboom, E. Bray, T. Aldcroft, M. Davis, A. Ginsburg, A. M. Price-Whelan, W. E. Kerzendorf, A. Conley, N. Crighton, K. Barbary, D. Muna, H. Ferguson, F. Grollier, M. M. Parikh, P. H. Nair, H. M. Unther, C. Deil,

J. Woillez, S. Conseil, R. Kramer, J. E. H. Turner, L. Singer, R. Fox, B. A. Weaver, V. Zabalza, Z. I. Edwards, K. Azalee Bostroem, D. J. Burke, A. R. Casey, S. M. Crawford, N. Dencheva, J. Ely, T. Jenness, K. Labrie, P. L. Lim, F. Pierfederici, A. Pontzen, A. Ptak, B. Refsdal, M. Servillat, and O. Streicher, "Astropy: A community Python package for astronomy," **558**, A33, A33 (2013).

[20]W. B. Atwood, A. A. Abdo, M. Ackermann, W. Althouse, B. Anderson, M. Axelsson, L. Baldini, J. Ballet, D. L. Band, G. Barbiellini, J. Bartelt, D. Bastieri, B. M. Baughman, K. Bechtol, D. Bédérède, F. Bellardi, R. Bellazzini, B. Berenji, G. F. Bignami, D. Bisello, E. Bissaldi, R. D. Blandford, E. D. Bloom, J. R. Bogart, E. Bonamente, J. Bonnell, A. W. Borgland, A. Bouvier, J. Bregeon, A. Brez, M. Brigida, P. Bruel, T. H. Burnett, G. Busetto, G. A. Caliandro, R. A. Cameron, P. A. Caraveo, S. Carius, P. Carlson, J. M. Casandjian, E. Cavazzuti, M. Ceccanti, C. Cecchi, E. Charles, A. Chekhtman, C. C. Cheung, J. Chiang, R. Chipaux, A. N. Cillis, S. Ciprini, R. Claus, J. Cohen-Tanugi, S. Condamoor, J. Conrad, R. Corbet, L. Corucci, L. Costamante, S. Cutini, D. S. Davis, D. Decotigny, M. DeKlotz, C. D. Dermer, A. de Angelis, S. W. Digel, E. do Couto e Silva, P. S. Drell, R. Dubois, D. Dumora, Y. Edmonds, D. Fabiani, C. Farnier, C. Favuzzi, D. L. Flath, P. Fleury, W. B. Focke, S. Funk, P. Fusco, F. Gargano, D. Gasparrini, N. Gehrels, F. X. Gentit, S. Germani, B. Giebels, N. Giglietto, P. Giommi, F. Giordano, T. Glanzman, G. Godfrey, I. A. Grenier, M. H. Grondin, J. E. Grove, L. Guillemot, S. Guiriec, G. Haller, A. K. Harding, P. A. Hart, E. Hays, S. E. Healey, M. Hirayama, L. Hjalmarsdotter, R. Horn, R. E. Hughes, G. Jóhannesson, G. Johansson, A. S. Johnson, R. P. Johnson, T. J. Johnson, W. N. Johnson, T. Kamae, H. Katagiri, J. Kataoka, A. Kavelaars, N. Kawai, H. Kelly, M. Kerr, W. Klamra, J. Knödlseder, M. L. Kocian, N. Komin, F. Kuehn, M. Kuss, D. Landriu, L. Latronico, B. Lee, S. H. Lee, M. Lemoine-Goumard, A. M. Lionetto, F. Longo, F. Loparco, B. Lott, M. N. Lovellette, P. Lubrano, G. M. Madejski, A. Makeev, B. Marangelli, M. M. Massai, M. N. Mazziotta, J. E. McEnery, N. Menon, C. Meurer, P. F. Michelson, M. Minuti, N. Mirizzi, W. Mitthumsiri, T. Mizuno, A. A. Moiseev, C. Monte, M. E. Monzani, E. Moretti, A. Morselli, I. V. Moskalenko, S. Murgia, T. Nakamori, S. Nishino, P. L. Nolan, J. P. Norris, E. Nuss, M. Ohno, T. Ohsugi, N. Omodei, E. Orlando, J. F. Ormes, A. Paccagnella, D. Paneque, J. H. Panetta, D. Parent, M. Pearce, M. Pepe, A. Perazzo, M. Pesce-Rollins, P. Picozza, L. Pieri, M. Pinchera, F. Piron, T. A. Porter, L. Poupard, S. Rainò, R. Rando, E. Rapposelli, M. Razzano, A. Reimer, O. Reimer, T. Reposeur, L. C. Reyes, S. Ritz, L. S. Rochester, A. Y. Rodriguez, R. W. Romani, M. Roth, J. J. Russell, F. Ryde, S. Sabatini, H. F. W. Sadrozinski, D. Sanchez, A. Sander, L. Sapozhnikov, P. M. S. Parkinson, J. D. Scargle, T. L. Schalk, G. Scolieri, C. Sgrò, G. H. Share, M. Shaw, T. Shimokawabe, C. Shrader, A. Sierpowska-Bartosik, E. J. Siskind, D. A. Smith, P. D. Smith, G. Spandre, P. Spinelli, J. L. Starck, T. E. Stephens, M. S. Strickman, A. W. Strong, D. J. Suson, H. Tajima, H. Takahashi, T. Takahashi, T. Tanaka, A. Tenze, S. Tether, J. B. Thayer, J. G. Thayer, D. J. Thompson, L. Tibaldo, O. Tibolla, D. F. Torres, G. Tosti, A. Tramacere, M. Turri, T. L. Usher, N. Vilchez, V. Vitale, P. Wang, K. Watters, B. L. Winer, K. S. Wood, T. Ylinen, and M. Ziegler, "The Large Area Telescope on the Fermi Gamma-Ray Space Telescope Mission," **697**, 1071–1102 (2009).

[21]N. Bade, D. Engels, H. Fink, H. J. Hagen, D. Reimers, W. Voges, and L. Wisotzki, "ROSAT active galactic nuclei identified on objective prism plates.," **254**, L21–L24 (1992).

[22]N. A. Bahcall, "Large-scale structure in the universe indicated by galaxy clusters.," **26**, 631–686 (1988).

[23]H. Bahk and H. S. Hwang, "UPCluster-SZ: The Updated Catalog of Galaxy Clusters from the List of Planck Sunyaev-Zeldovich Sources," arXiv e-prints, arXiv:2403.03818, arXiv:2403.03818 (2024).

[24]R. D. Baldi, A. Capetti, and F. Massaro, "FR0CAT: a FIRST catalog of FR 0 radio galaxies," **609**, A1, A1 (2018).

[25]R. D. Baldi, A. Capetti, and G. Giovannini, "Pilot study of the radio-emitting AGN population: the emerging new class of FR 0 radio-galaxies," **576**, A38, A38 (2015).

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

[27]J. Ballet, T. H. Burnett, S. W. Digel, and B. Lott, "Fermi Large Area Telescope Fourth Source Catalog Data Release 2," arXiv e-prints, arXiv:2005.11208, arXiv:2005.11208 (2020).

[28]B. Balmaverde, A. Capetti, A. Marconi, G. Venturi, M. Chiaberge, R. D. Baldi, S. Baum, R. Gilli, P. Grandi, E. T. Meyer, G. Miley, C. O'Dea, W. Sparks, E. Torresi, and G. Tremblay, "The MURALES survey. III. Completing the MUSE observations of 37 3C low-z radio galaxies," **645**, A12, A12 (2021).

[29]B. Balmaverde, A. Capetti, A. Marconi, G. Venturi, M. Chiaberge, R. D. Baldi, S. Baum, R. Gilli, P. Grandi, E. Meyer, G. Miley, C. O'Dea, W. Sparks, E. Torresi, and G. Tremblay, "The MURALES survey. II. Presentation of MUSE observations of 20 3C low-z radio galaxies and first results," **632**, A124, A124 (2019).

[30]M. Baloković, V. Smolčić, Ž. Ivezić, G. Zamorani, E. Schinnerer, and B. C. Kelly, "Disclosing the Radio Loudness Distribution Dichotomy in Quasars: An Unbiased Monte Carlo Approach Applied to the SDSS-FIRST Quasar Sample," **759**, 30, 30 (2012).

[31]P. Banerjee, T. Szabo, E. Pierpaoli, G. Franco, M. Ortiz, A. Oramas, and B. Tornello, "An optical catalog of galaxy clusters obtained from an adaptive matched filter finder applied to SDSS DR9 data," **58**, 61–71 (2018).

[32]R. E. Bär, B. Trakhtenbrot, K. Oh, M. J. Koss, O. I. Wong, C. Ricci, K. Schawinski, A. K. Weigel, L. F. Sartori, K. Ichikawa, N. J. Secrest, D. Stern, F. Pacucci, R. Mushotzky, M. C. Powell, F. Ricci, E. Sani, K. L. Smith, F. A. Harrison, I. Lamperti, and C. M. Urry, "BAT AGN Spectroscopic Survey - XIII. The nature of the most luminous obscured AGN in the low-redshift universe," **489**, 3073–3092 (2019).

[33]S. D. Barthelmy, "Burst Alert Telescope (BAT) on the Swift MIDEX mission," in X-ray and gamma-ray instrumentation for astronomy xiii, Vol. 5165, edited by K. A. Flanagan and O. H. W. Siegmund, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series (Feb. 2004), pp. 175–189.

[34]J. G. Bartlett, A. Chamballu, J.-B. Melin, M. Arnaud, and Members of the Planck Working Group 5, "The Planck cluster survey," Astronomische Nachrichten **329**, 147 (2008).

[35]W. H. Baumgartner, J. Tueller, C. B. Markwardt, G. K. Skinner, S. Barthelmy, R. F. Mushotzky, P. A. Evans, and N. Gehrels, "The 70 Month Swift-BAT All-sky Hard X-Ray Survey," **207**, 19, 19 (2013).

[36]R. H. Becker, R. L. White, and D. J. Helfand, "The FIRST Survey: Faint Images of the Radio Sky at Twenty Centimeters," **450**, 559 (1995).

[37]V. Beckmann and C. Shrader, "The AGN phenomenon: open issues," in Proceedings of "an integral view of the high-energy sky (the first 10 years)" - 9th integral workshop and celebration of the 10th anniversary of the launch (integral 2012). 15-19 october 2012. bibliotheque nationale de france (Jan. 2012), p. 69.

[38]V. Beckmann, S. Soldi, C. R. Shrader, N. Gehrels, and N. Produit, "The Hard X-Ray 20-40 keV AGN Luminosity Function," **652**, 126–135 (2006).

[39]C. L. Bennett, D. Larson, J. L. Weiland, and G. Hinshaw, "The 1% Concordance Hubble Constant," **794**, 135, 135 (2014).

[40]E. Bertin, *Science with astronomical near-infrared sky surveys: proceedings of the les houches school, centre de physique des houches, les houches, france, 20–24 september, 1993* (Springer Netherlands, 1993), pp. 49–51.

[41]E. Bertin and S. Arnouts, "SExtractor: Software for source extraction.," **117**, 393–404 (1996).

[42]S. Bhatiani, X. Dai, R. D. Griffin, J. M. Nugent, C. S. Kochanek, and J. N. Bregman, "Optical Confirmation of X-Ray-selected Galaxy Clusters from the Swift AGN and Cluster Survey with MDM and Pan-STARRS Data. III," **259**, 9, 9 (2022).

[43]A. J. Bird, E. J. Barlow, L. Bassani, A. Bazzano, A. Bodaghee, F. Capitanio, M. Cocchi, M. Del Santo, A. J. Dean, A. B. Hill, F. Lebrun, G. Malaguti, A. Malizia, R. Much, S. E. Shaw, J. B. Stephen, R. Terrier, P. Ubertini, and R. Walter, "The First IBIS/ISGRI Soft Gamma-Ray Galactic Plane Survey Catalog," **607**, L33–L37 (2004).

[44]A. J. Bird, A. Bazzano, L. Bassani, F. Capitanio, M. Fiocchi, A. B. Hill, A. Malizia, V. A. McBride, S. Scaringi, V. Sguera, J. B. Stephen, P. Ubertini, A. J. Dean, F. Lebrun, R. Terrier, M. Renaud, F. Mattana, D. Götz, J. Rodriguez, G. Belanger, R. Walter, and C. Winkler, "The Fourth IBIS/ISGRI Soft Gamma-ray Survey Catalog," **186**, 1–9 (2010).

[45]A. J. Bird, A. Bazzano, A. Malizia, M. Fiocchi, V. Sguera, L. Bassani, A. B. Hill, P. Ubertini, and C. Winkler, "The IBIS Soft Gamma-Ray Sky after 1000 Integral Orbits," **223**, 15, 15 (2016).

[46]C. Bishop, "Bayesian neural networks," English, Journal of the Brazilian Computer Society **4**, 61–68 (1997).

[47]C. Bishop, *Mixture density networks*, English, WorkingPaper (Aston University, 1994).

[48]C. M. Bishop, *Pattern recognition and machine learning (information science and statistics)* (Springer-Verlag, Berlin, Heidelberg, 2006).

[49]R. D. Blandford and M. J. Rees, "Some comments on radiation mechanisms in Lacertids.," in Bl lac objects, edited by A. M. Wolfe (Jan. 1978), pp. 328–341.

[50]R. D. Blandford and R. L. Znajek, "Electromagnetic extraction of energy from Kerr black holes," Monthly Notices of the Royal Astronomical Society **179**, 433–456 (1977).

[51]M. R. Blanton, M. A. Bershady, B. Abolfathi, F. D. Albareti, C. Allende Prieto, A. Almeida, J. Alonso-García, F. Anders, S. F. Anderson, B. Andrews, E. Aquino-Ortíz, A. Aragón-Salamanca, M. Argudo-Fernández, E. Armengaud, E. Aubourg, V. Avila-Reese, C. Badenes, S. Bailey, K. A. Barger, J. Barrera-Ballesteros, C. Bartosz, D. Bates, F. Baumgarten, J. Bautista, R. Beaton, T. C. Beers, F. Belfiore, C. F. Bender, A. A. Berlind, M. Bernardi, F. Beutler, J. C. Bird, D. Bizyaev, G. A. Blanc, M. Blomqvist, A. S. Bolton, M. Boquien, J. Borissova, R. van den Bosch, J. Bovy, W. N. Brandt, J. Brinkmann, J. R. Brownstein, K. Bundy, A. J. Burgasser, E. Burtin, N. G. Busca, M. Cappellari, M. L. Delgado Carigi, J. K. Carlberg, A. Carnero Rosell, R. Carrera, N. J. Chanover, B. Cherinka, E. Cheung, Y. Gómez Maqueo Chew, C. Chiappini, P. D. Choi, D. Chojnowski, C.-H. Chuang, H. Chung, R. F. Cirolini, N. Clerc, R. E. Cohen, J. Comparat, L. da Costa, M.-C. Cousinou, K. Covey, J. D. Crane, R. A. C. Croft, I. Cruz-Gonzalez, D. Garrido Cuadra, K. Cunha, G. J. Damke, J. Darling, R. Davies, K. Dawson, A. de la Macorra, F. Dell'Agli, N. De Lee, T. Delubac, F. Di Mille, A. Diamond-Stanic, M. Cano-Díaz, J. Donor, J. J. Downes, N. Drory, H. du Mas des Bourboux, C. J. Duckworth, T. Dwelly, J. Dyer, G. Ebelke, A. D. Eigenbrot, D. J. Eisenstein, E. Emsellem, M. Eracleous, S. Escoffier, M. L. Evans, X. Fan, E. Fernández-Alvar, J. G. Fernandez-Trincado, D. K. Feuillet, A. Finoguenov, S. W. Fleming, A. Font-Ribera, A. Fredrickson, G. Freischlad, P. M. Frinchaboy, C. E. Fuentes, L. Galbany, R. Garcia-Dias, D. A. García-Hernández, P. Gaulme, D. Geisler, J. D. Gelfand, H. Gil-Marín, B. A. Gillespie, D. Goddard, V. Gonzalez-Perez, K. Grabowski, P. J. Green, C. J. Grier, J. E. Gunn, H. Guo, J. Guy, A. Hagen, C. Hahn, M. Hall, P. Harding, S. Hasselquist, S. L. Hawley, F. Hearty, J. I. Gonzalez Hernández, S. Ho, D. W. Hogg, K. Holley-Bockelmann, J. A. Holtzman, P. H. Holzer, J. Huehnerhoff, T. A. Hutchinson, H. S. Hwang, H. J. Ibarra-Medel, G. da Silva Ilha, I. I. Ivans, K. Ivory, K. Jackson, T. W. Jensen, J. A. Johnson, A. Jones, H. Jönsson, E. Jullo, V. Kamble, K. Kinemuchi, D. Kirkby, F.-S. Kitaura, M. Klaene, G. R. Knapp, J.-P. Kneib, J. A. Kollmeier, I. Lacerna, R. R. Lane, D. Lang, D. R. Law, D. Lazarz, Y. Lee, J.-M. Le Goff, F.-H. Liang, C. Li, H. Li, J. Lian, M. Lima, L. Lin, Y.-T. Lin, S. Bertran de Lis, C. Liu, M. A. C. de Icaza Lizaola, D. Long, S. Lucatello, B. Lundgren, N. K. MacDonald, A. Deconto Machado, C. L. MacLeod, S. Mahadevan, M. A. Geimba Maia, R. Maiolino, S. R. Majewski, E. Malanushenko, V. Malanushenko, A. Manchado, S. Mao, C. Maraston, R. Marques-Chaves, T. Masseron, K. L. Masters, C. K. McBride, R. M. McDermid, B. McGrath, I. D. McGreer, N. Medina Peña, M. Melendez, A. Merloni, M. R. Merrifield, S. Meszaros, A. Meza, I. Minchev, D. Minniti, T. Miyaji, S. More, J. Mulchaey, F. Müller-Sánchez, D. Muna, R. R. Munoz, A. D. Myers, P. Nair, K. Nandra, J. Correa do Nascimento, A. Negrete, M. Ness, J. A. Newman, R. C. Nichol, D. L. Nidever, C. Nitschelm, P. Ntelis, J. E. O'Connell, R. J. Oelkers, A. Oravetz, D. Oravetz, Z. Pace, N. Padilla, N. Palanque-Delabrouille, P. Alonso Palicio,

K. Pan, J. K. Parejko, T. Parikh, I. Pâris, C. Park, A. Y. Patten, S. Peirani, M. Pellejero-Ibanez, S. Penny, W. J. Percival, I. Perez-Fournon, P. Petitjean, M. M. Pieri, M. Pinsonneault, A. Pisani, R. Poleski, F. Prada, A. Prakash, A. B. d. A. Queiroz, M. J. Raddick, A. Raichoor, S. Barboza Rembold, H. Richstein, R. A. Riffel, R. Riffel, H.-W. Rix, A. C. Robin, C. M. Rockosi, S. Rodríguez-Torres, A. Roman-Lopes, C. Román-Zúñiga, M. Rosado, A. J. Ross, G. Rossi, J. Ruan, R. Ruggeri, E. S. Rykoff, S. Salazar-Albornoz, M. Salvato, A. G. Sánchez, D. S. Aguado, J. R. Sánchez-Gallego, F. A. Santana, B. X. Santiago, C. Sayres, R. P. Schiavon, J. da Silva Schimoia, E. F. Schlafly, D. J. Schlegel, D. P. Schneider, M. Schultheis, W. J. Schuster, A. Schwope, H.-J. Seo, Z. Shao, S. Shen, M. Shetrone, M. Shull, J. D. Simon, D. Skinner, M. F. Skrutskie, A. Slosar, V. V. Smith, J. S. Sobeck, F. Sobreira, G. Somers, D. Souto, D. V. Stark, K. Stassun, F. Stauffer, M. Steinmetz, T. Storchi-Bergmann, A. Streblyanska, G. S. Stringfellow, G. Suárez, J. Sun, N. Suzuki, L. Szigeti, M. Taghizadeh-Popp, B. Tang, C. Tao, J. Tayar, M. Tembe, J. Teske, A. R. Thakar, D. Thomas, B. A. Thompson, J. L. Tinker, P. Tissera, R. Tojeiro, H. Hernandez Toledo, S. de la Torre, C. Tremonti, N. W. Troup, O. Valenzuela, I. Martinez Valpuesta, J. Vargas-González, M. Vargas-Magaña, J. A. Vazquez, S. Villanova, M. Vivek, N. Vogt, D. Wake, R. Walterbos, Y. Wang, B. A. Weaver, A.-M. Weijmans, D. H. Weinberg, K. B. Westfall, D. G. Whelan, V. Wild, J. Wilson, W. M. Wood-Vasey, D. Wylezalek, T. Xiao, R. Yan, M. Yang, J. E. Ybarra, C. Yèche, N. Zakamska, O. Zamora, P. Zarrouk, G. Zasowski, K. Zhang, G.-B. Zhao, Z. Zheng, Z. Zheng, X. Zhou, Z.-M. Zhou, G. B. Zhu, M. Zoccali, and H. Zou, "Sloan Digital Sky Survey IV: Mapping the Milky Way, Nearby Galaxies, and the Distant Universe," **154**, 28, 28 (2017).

[52]L. E. Bleem, B. Stalder, T. de Haan, K. A. Aird, S. W. Allen, D. E. Applegate, M. L. N. Ashby, M. Bautz, M. Bayliss, B. A. Benson, S. Bocquet, M. Brodwin, J. E. Carlstrom, C. L. Chang, I. Chiu, H. M. Cho, A. Clocchiatti, T. M. Crawford, A. T. Crites, S. Desai, J. P. Dietrich, M. A. Dobbs, R. J. Foley, W. R. Forman, E. M. George, M. D. Gladders, A. H. Gonzalez, N. W. Halverson, C. Hennig, H. Hoekstra, G. P. Holder, W. L. Holzapfel, J. D. Hrubes, C. Jones, R. Keisler, L. Knox, A. T. Lee, E. M. Leitch, J. Liu, M. Lueker, D. Luong-Van, A. Mantz, D. P. Marrone, M. McDonald, J. J. McMahon, S. S. Meyer, L. Mocanu, J. J. Mohr, S. S. Murray, S. Padin, C. Pryke, C. L. Reichardt, A. Rest, J. Ruel, J. E. Ruhl, B. R. Saliwanchik, A. Saro, J. T. Sayre, K. K. Schaffer, T. Schrabback, E. Shirokoff, J. Song, H. G. Spieler, S. A. Stanford, Z. Staniszewski, A. A. Stark, K. T. Story, C. W. Stubbs, K. Vanderlinde, J. D. Vieira, A. Vikhlinin, R. Williamson, O. Zahn, and A. Zenteno, "Galaxy Clusters Discovered via the Sunyaev-Zel'dovich Effect in the 2500-Square-Degree SPT-SZ Survey," **216**, 27, 27 (2015).

[53]D. C. J. Bock, M. I. Large, and E. M. Sadler, "SUMSS: A Wide-Field Radio Imaging Survey of the Southern Sky. I. Science Goals, Survey Design, and Instrumentation," **117**, 1578–1593 (1999).

[54]H. Böhringer and G. Chon, "Constraints on neutrino masses from the study of the nearby large-scale structure and galaxy cluster counts," Modern Physics Letters A **31**, 1640008, 1640008 (2016).

[55]J. R. Bond, L. Kofman, and D. Pogosyan, "How filaments of galaxies are woven into the cosmic web," **380**, 603–606 (1996).

[56]E. Bottacini, M. Ajello, and J. Greiner, "The Deep Look at the Hard X-Ray Sky: The Swift-INTEGRAL X-Ray (SIX) Survey," **201**, 34, 34 (2012).

[57]A. Botteon, F. Gastaldello, and G. Brunetti, "Shocks and cold fronts in merging and massive galaxy clusters: new detections with Chandra," **476**, 5591–5620 (2018).

[58]A. H. Bridle, D. H. Hough, C. J. Lonsdale, J. O. Burns, and R. A. Laing, "Deep VLA Imaging of Twelve Extended 3CR Quasars," **108**, 766 (1994).

[59]E. Bulbul, A. Liu, M. Kluge, X. Zhang, J. S. Sanders, Y. E. Bahar, V. Ghirardini, E. Artis, R. Seppi, C. Garrel, M. E. Ramos-Ceja, J. Comparat, F. Balzer, K. Böckmann, M. Brüggen, N. Clerc, K. Dennerl, K. Dolag, M. Freyberg, S. Grandis, D. Gruen, F. Kleinebreil, S. Krippendorf, G. Lamer, A. Merloni, K. Migkas, K. Nandra, F. Pacaud, P. Predehl, T. H. Reiprich, T. Schrabback, A. Veronica, J. Weller, and S. Zelmer, "The SRG/eROSITA All-Sky Survey: The first catalog of galaxy clusters and groups in the Western Galactic Hemisphere," arXiv e-prints, arXiv:2402.08452, arXiv:2402.08452 (2024).

[60]D. N. Burrows, J. E. Hill, J. A. Nousek, J. A. Kennea, A. Wells, J. P. Osborne, A. F. Abbey, A. Beardmore, K. Mukerjee, A. D. T. Short, G. Chincarini, S. Campana, O. Citterio, A. Moretti, C. Pagani, G. Tagliaferri, P. Giommi, M. Capalbi, F. Tamburelli, L. Angelini, G. Cusumano, H. W. Bräuninger, W. Burkert, and G. D. Hartner, "The Swift X-Ray Telescope," **120**, 165–195 (2005).

[61]S. Buttiglione, A. Capetti, A. Celotti, D. J. Axon, M. Chiaberge, F. D. Macchetto, and W. B. Sparks, "An optical spectroscopic survey of the 3CR sample of radio galaxies with z $<$ 0.3 . II. Spectroscopic classes and accretion modes in radio-loud AGN," **509**, A6, A6 (2010).

[62]J. Cantó, A. C. Raga, and L. F. Rodríguez, "The Hot, Diffuse Gas in a Dense Cluster of Massive Stars," **536**, 896–901 (2000).

[63]A. Capetti, D. J. Axon, F. D. Macchetto, A. Marconi, and C. Winge, "The Origin of the Narrow-Line Region of Markarian 3: An Overpressured Jet Cocoon," **516**, 187–194 (1999).

[64]A. Capetti, F. Massaro, and R. D. Baldi, "FRICAT: A FIRST catalog of FR I radio galaxies," **598**, A49, A49 (2017).

[65]A. Capetti, F. Massaro, and R. D. Baldi, "FRIICAT: A FIRST catalog of FR II radio galaxies," **601**, A81, A81 (2017).

[66]J. E. Carlstrom, P. A. R. Ade, K. A. Aird, B. A. Benson, L. E. Bleem, S. Busetti, C. L. Chang, E. Chauvin, H. M. Cho, T. M. Crawford, A. T. Crites, M. A. Dobbs, N. W. Halverson, S. Heimsath, W. L. Holzapfel, J. D. Hrubes, M. Joy, R. Keisler, T. M. Lanting, A. T. Lee, E. M. Leitch, J. Leong, W. Lu, M. Lueker, D. Luong-Van, J. J. McMahon, J. Mehl, S. S. Meyer, J. J. Mohr, T. E. Montroy, S. Padin, T. Plagge, C. Pryke, J. E. Ruhl, K. K. Schaffer, D. Schwan, E. Shirokoff, H. G. Spieler, Z. Staniszewski, A. A. Stark, C. Tucker, K. Vanderlinde, J. D. Vieira, and R. Williamson, "The 10 Meter South Pole Telescope," **123**, 568 (2011).

[67]A. Cavaliere and R. Fusco-Femiano, "X-rays from hot plasma in clusters of galaxies," **49**, 137–144 (1976).

[68]N. Cerardi, M. Pierre, P. Valageas, C. Garrel, and F. Pacaud, "The cosmological analysis of X-ray cluster surveys V. The potential of cluster counts in the $1 < z < 2$ range," arXiv e-prints, arXiv:2312.04253, arXiv:2312.04253 (2023).

[69]K. C. Chambers, E. A. Magnier, N. Metcalfe, H. A. Flewelling, M. E. Huber, C. Z. Waters, L. Denneau, P. W. Draper, D. Farrow, D. P. Finkbeiner, C. Holmberg, J. Koppenhoefer, P. A. Price, A. Rest, R. P. Saglia, E. F. Schlafly, S. J. Smartt, W. Sweeney, R. J. Wainscoat, W. S. Burgett, S. Chastel, T. Grav, J. N. Heasley, K. W. Hodapp, R. Jedicke, N. Kaiser, R. P. Kudritzki, G. A. Luppino, R. H. Lupton, D. G. Monet, J. S. Morgan, P. M. Onaka, B. Shiao, C. W. Stubbs, J. L. Tonry, R. White, E. Bañados, E. F. Bell, R. Bender, E. J. Bernard, M. Boegner, F. Boffi, M. T. Botticella, A. Calamida, S. Casertano, W. P. Chen, X. Chen, S. Cole, N. Deacon, C. Frenk, A. Fitzsimmons, S. Gezari, V. Gibbs, C. Goessl, T. Goggia, R. Gourgue, B. Goldman, P. Grant, E. K. Grebel, N. C. Hambly, G. Hasinger, A. F. Heavens, T. M. Heckman, R. Henderson, T. Henning, M. Holman, U. Hopp, W. H. Ip, S. Isani, M. Jackson, C. D. Keyes, A. M. Koekemoer, R. Kotak, D. Le, D. Liska, K. S. Long, J. R. Lucey, M. Liu, N. F. Martin, G. Masci, B. McLean, E. Mindel, P. Misra, E. Morganson, D. N. A. Murphy, A. Obaika, G. Narayan, M. A. Nieto-Santisteban, P. Norberg, J. A. Peacock, E. A. Pier, M. Postman, N. Primak, C. Rae, A. Rai, A. Riess, A. Riffeser, H. W. Rix, S. Röser, R. Russel, L. Rutz, E. Schilbach, A. S. B. Schultz, D. Scolnic, L. Strolger, A. Szalay, S. Seitz, E. Small, K. W. Smith, D. R. Soderblom, P. Taylor, R. Thomson, A. N. Taylor, A. R. Thakar, J. Thiel, D. Thilker, D. Unger, Y. Urata, J. Valenti, J. Wagner, T. Walder, F. Walter, S. P. Watters, S. Werner, W. M. Wood-Vasey, and R. Wyse, "The Pan-STARRS1 Surveys," arXiv e-prints, arXiv:1612.05560, arXiv:1612.05560 (2016).

[70]P. A. Charles and M. J. Coe, "Optical, ultraviolet and infrared observations of X-ray binaries," arXiv e-prints, astro-ph/0308020, astro–ph/0308020 (2003).

[71]Y.-C. Chen and C.-Y. Hwang, "Morphology of Seyfert galaxies," **362**, 230, 230 (2017).

[72]F. Chollet et al., *Keras*, `https://keras.io`, 2015.

[73]E. Churazov, R. Sunyaev, M. Revnivtsev, S. Sazonov, S. Molkov, S. Grebenev, C. Winkler, A. Parmar, A. Bazzano, M. Falanga, A. Gros, F. Lebrun, L. Natalucci, P. Ubertini, J. P. Roques, L. Bouchet, E. Jourdain, J. Knödlseder, R. Diehl, C. Budtz-Jorgensen, S. Brandt, N. Lund, N. J. Westergaard, A. Neronov, M. Türler, M. Chernyakova, R. Walter, N. Produit, N. Mowlavi, J. M. Mas-Hesse, A. Domingo, N. Gehrels, E. Kuulkers, P. Kretschmar, and M. Schmidt, "INTEGRAL observations of the cosmic X-ray background in the 5-100 keV range via occultation by the Earth," **467**, 529–540 (2007).

[74]D. C. Ciresan, U. Meier, J. Masci, and J. Schmidhuber, "Multi-column deep neural network for traffic sign classification.," Neural Networks **32**, 333–338 (2012).

[75]N. Clerc, J. Comparat, R. Seppi, E. Artis, Y. E. Bahar, F. Balzer, E. Bulbul, T. Dauser, C. Garrel, V. Ghirardini, S. Grandis, C. Kirsch, M. Kluge, A. Liu, F. Pacaud, M. E. Ramos-Ceja, T. H. Reiprich, J. S. Sanders, J. Wilms, and X. Zhang, "The SRG/eROSITA All-Sky Survey: X-ray selection function models for the eRASS1 galaxy cluster cosmology," arXiv e-prints, arXiv:2402.08457, arXiv:2402.08457 (2024).

[76]N. Clerc, T. Sadibekova, M. Pierre, F. Pacaud, J.-P. Le Fèvre, C. Adami, B. Altieri, and I. Valtchanov, "The cosmological analysis of X-ray cluster surveys - II. Application of the CR-HR method to the XMM archive," **423**, 3561–3583 (2012).

[77]N. Clerc, T. Sadibekova, M. Pierre, F. Pacaud, J.-P. Le Fevre, C. Adami, B. Altieri, and I. Valtchanov, "VizieR Online Data Catalog: X-ray clusters from XMM (Clerc+, 2012)," VizieR Online Data Catalog **742** (2012).

[78]P. Coles and F. Lucchin, *Cosmology. The origin and evolution of cosmic structure* (1995).

[79]A. A. Collister and O. Lahav, "ANNz: Estimating Photometric Redshifts Using Artificial Neural Networks," **116**, 345–351 (2004).

[80]A. Comastri, M. Brusa, P. Ciliegi, M. Mignoli, F. Fiore, R. Maiolino, P. Severgnini, A. Baldi, S. Molendi, C. Vignali, F. La Franca, G. Matt, and G. C. Perola, "On the nature of X-ray Bright Optically Normal galaxies," arXiv e-prints, astro-ph/0203019, astro–ph/0203019 (2002).

[81]J. J. Condon, W. D. Cotton, E. W. Greisen, Q. F. Yin, R. A. Perley, G. B. Taylor, and J. J. Broderick, "The NRAO VLA Sky Survey," **115**, 1693–1716 (1998).

[82]C. Cortes and V. Vapnik, "Support-vector networks," Machine learning **20**, 273–297 (1995).

[83]R. Cowsik and E. J. Kobetich, "Comment on Inverse Compton Models for the Isotropic X-Ray Back-Ground and Possible Thermal Emission from a Hot Intergalactic Gas," **177**, 585 (1972).

[84]G. Cusumano, V. La Parola, A. Segreto, C. Ferrigno, A. Maselli, B. Sbarufatti, P. Romano, G. Chincarini, P. Giommi, N. Masetti, A. Moretti, P. Parisi, and G. Tagliaferri, "The Palermo Swift-BAT hard X-ray catalogue. III. Results after 54 months of sky survey," **524**, A64, A64 (2010).

[85]R. M. Cutri, E. L. Wright, T. Conrow, J. W. Fowler, P. R. M. Eisenhardt, C. Grillmair, J. D. Kirkpatrick, F. Masci, H. L. McCallon, S. L. Wheelock, S. Fajardo-Acosta, L. Yan, D. Benford, M. Harbut, T. Jarrett, S. Lake, D. Leisawitz, M. E. Ressler, S. A. Stanford, C. W. Tsai, F. Liu, G. Helou, A. Mainzer, D. Gettngs, A. Gonzalez, D. Hoffman, K. A. Marsh, D. Padgett, M. F. Skrutskie, R. Beck, M. Papin, and M. Wittman, "VizieR Online Data Catalog: AllWISE Data Release (Cutri+ 2013)," VizieR Online Data Catalog, II/328, II/328 (2021).

[86]R. M. Cutri, E. L. Wright, T. Conrow, J. W. Fowler, P. R. M. Eisenhardt, C. Grillmair, J. D. Kirkpatrick, F. Masci, H. L. McCallon, S. L. Wheelock, S. Fajardo-Acosta, L. Yan, D. Benford, M. Harbut, T. Jarrett, S. Lake, D. Leisawitz, M. E. Ressler, S. A. Stanford, C. W. Tsai, F. Liu, G. Helou, A. Mainzer, D. Gettngs, A. Gonzalez, D. Hoffman, K. A. Marsh, D. Padgett, M. F. Skrutskie, R. Beck, M. Papin, and M. Wittman, "VizieR Online Data Catalog: AllWISE Data Release (Cutri+ 2013)," VizieR Online Data Catalog, II/328, II/328 (2021).

[87]R. D'Abrusco, F. Massaro, M. Ajello, J. E. Grindlay, H. A. Smith, and G. Tosti, "Infrared Colors of the Gamma-Ray-detected Blazars," **748**, 68, 68 (2012).

[88]R. D'Abrusco, F. Massaro, M. Ajello, J. E. Grindlay, H. A. Smith, and G. Tosti, "Infrared Colors of the Gamma-Ray-detected Blazars," **748**, 68, 68 (2012).

[89]R. D'Abrusco, F. Massaro, A. Paggi, H. A. Smith, N. Masetti, M. Landoni, and G. Tosti, "The WISE Blazar-like Radio-loud Sources: An All-sky Catalog of Candidate γ-ray Blazars," **215**, 14, 14 (2014).

[90]R. D'Abrusco, N. Álvarez Crespo, F. Massaro, R. Campana, V. Chavushyan, M. Landoni, F. La Franca, N. Masetti, D. Milisavljevic, A. Paggi, F. Ricci, and H. A. Smith, "Two New Catalogs of Blazar Candidates in the WISE Infrared Sky," **242**, 4, 4 (2019).

[91]V. D'Elia, M. Perri, S. Puccetti, M. Capalbi, P. Giommi, D. N. Burrows, S. Campana, G. Tagliaferri, G. Cusumano, P. Evans, N. Gehrels, J. Kennea, A. Moretti, J. A. Nousek, J. P. Osborne, P. Romano, and G. Stratta, "The seven year Swift-XRT point source catalog (1SWXRT)," **551**, A142, A142 (2013).

[92]J. Deng, W. Dong, R. Socher, L. Li, K. Li, and L. Fei-Fei, "Imagenet: a large-scale hierarchical image database," in 2009 ieee conference on computer vision and pattern recognition (2009), pp. 248–255.

[93]C. D. Dermer, "On the Beaming Statistics of Gamma-Ray Sources," **446**, L63 (1995).

[94]A. Diana, A. Caccianiga, L. Ighina, S. Belladitta, A. Moretti, and R. Della Ceca, "The evolution of the heaviest supermassive black holes in jetted AGNs," **511**, 5436–5447 (2022).

[95]S. Dieleman, K. W. Willett, and J. Dambre, "Rotation-invariant convolutional neural networks for galaxy morphology prediction," **450**, 1441–1459 (2015).

[96]R. A. Downes, R. F. Webbink, M. M. Shara, H. Ritter, U. Kolb, and H. W. Duerbeck, "A Catalog and Atlas of Cataclysmic Variables: The Final Edition," Journal of Astronomical Data **11**, 2 (2005).

[97]G. E Hinton, "Learning multiple layers of representation," **11**, 428–34 (2007).

[98]R. T. Edwards, M. Bailes, W. van Straten, and M. C. Britton, "The Swinburne intermediate-latitude pulsar survey," **326**, 358–374 (2001).

[99]V. R. Eke, S. Cole, and C. S. Frenk, "Cluster evolution as a diagnostic for Omega," **282**, 263–280 (1996).

[100]V. R. Eke, S. Cole, C. S. Frenk, and J. Patrick Henry, "Measuring Omega_0 using cluster evolution," **298**, 1145–1158 (1998).

[101]M. Elvis, E. J. Schreier, J. Tonry, M. Davis, and J. P. Huchra, "Two optically dull galaxies with strong nuclear X-ray sources.," **246**, 20–27 (1981).

[102]M. Elvis, F. Civano, C. Vignali, S. Puccetti, F. Fiore, N. Cappelluti, T. L. Aldcroft, A. Fruscione, G. Zamorani, A. Comastri, M. Brusa, R. Gilli, T. Miyaji, F. Damiani, A. M. Koekemoer, A. Finoguenov, H. Brunner, C. M. Urry, J. Silverman, V. Mainieri, G. Hasinger, R. Griffiths, C. M. Carollo, H. Hao, L. Guzzo, A. Blain, D. Calzetti, C. Carilli, P. Capak, S. Ettori, G. Fabbiano, C. Impey, S. Lilly, B. Mobasher, M. Rich, M. Salvato, D. B. Sanders, E. Schinnerer, N. Scoville, P. Shopbell, J. E. Taylor, Y. Taniguchi, and M. Volonteri, "The Chandra COSMOS Survey. I. Overview and Point Source Catalog," **184**, 158–171 (2009).

[103]P. A. Evans, J. P. Osborne, A. P. Beardmore, K. L. Page, R. Willingale, C. J. Mountford, C. Pagani, D. N. Burrows, J. A. Kennea, M. Perri, G. Tagliaferri, and N. Gehrels, "1SXPS: A Deep Swift X-Ray Telescope Point Source Catalog with Light Curves and Spectra," **210**, 8, 8 (2014).

[104]P. A. Evans, J. P. Osborne, A. P. Beardmore, K. L. Page, R. Willingale, C. J. Mountford, C. Pagani, D. N. Burrows, J. A. Kennea, M. Perri, G. Tagliaferri, and N. Gehrels, "1SXPS: A Deep Swift X-Ray Telescope Point Source Catalog with Light Curves and Spectra," **210**, 8, 8 (2014).

[105]P. A. Evans, K. L. Page, J. P. Osborne, A. P. Beardmore, R. Willingale, D. N. Burrows, J. A. Kennea, M. Perri, M. Capalbi, G. Tagliaferri, and S. B. Cenko, "2SXPS: An Improved and Expanded Swift X-Ray Telescope Point-source Catalog," **247**, 54, 54 (2020).

[106]A. C. Fabian, "Observational Evidence of Active Galactic Nuclei Feedback," **50**, 455–489 (2012).

[107]A. C. Fabian, K. Iwasawa, C. S. Reynolds, and A. J. Young, "Broad Iron Lines in Active Galactic Nuclei," **112**, 1145–1161 (2000).

[108]H. Falcke, W. Sherwood, and A. R. Patnaik, "The Nature of Radio-intermediate Quasars: What Is Radio-loud and What Is Radio-quiet?" **471**, 106 (1996).

[109]B. L. Fanaroff and J. M. Riley, "The morphology of extragalactic radio sources of high and low luminosity," **167**, 31P–36P (1974).

[110]J. E. Felten, R. J. Gould, W. A. Stein, and N. J. Woolf, "X-Rays from the Coma Cluster of Galaxies," **146**, 955–958 (1966).

[111]M. S. Fischer, N.-H. Durke, K. Hollingshausen, C. Hammer, M. Brüggen, and K. Dolag, "The role of baryons in self-interacting dark matter mergers," **523**, 5915–5933 (2023).

[112]S. R. Folkes, O. Lahav, and S. J. Maddox, "An artificial neural network approach to the classification of galaxy spectra.," **283**, 651–665 (1996).

[113]W. Forman, C. Jones, L. Cominsky, P. Julien, S. Murray, G. Peters, H. Tananbaum, and R. Giacconi, "The fourth Uhuru catalog of X-ray sources.," **38**, 357–412 (1978).

[114]F. Frontera, C. Guidorzi, E. Montanari, F. Rossi, E. Costa, M. Feroci, F. Calura, M. Rapisarda, L. Amati, D. Carturan, M. R. Cinti, D. D. Fiume, L. Nicastro, and M. Orlandini, "The Gamma-Ray Burst Catalog Obtained with the Gamma-Ray Burst Monitor Aboard BeppoSAX," **180**, 192–223 (2009).

[115]F. Frontera, M. Orlandini, R. Landi, A. Comastri, F. Fiore, G. Setti, L. Amati, E. Costa, N. Masetti, and E. Palazzi, "The Cosmic X-Ray Background and the Population of the Most Heavily Obscured AGNs," **666**, 86–95 (2007).

[116]D. Garofalo and C. B. Singh, "FR0 Radio Galaxies and Their Place in the Radio Morphology Classification," **871**, 259, 259 (2019).

[117]C. Garrel, M. Pierre, P. Valageas, D. Eckert, F. Marulli, A. Veropalumbo, F. Pacaud, N. Clerc, M. Sereno, K. Umetsu, L. Moscardini, S. Bhargava, C. Adami, L. Chiappetti, F. Gastaldello, E. Koulouridis, J.-P. Le Fevre, and M. Plionis, "The XXL survey. XLVI. Forward cosmological analysis of the C1 cluster sample," **663**, A3, A3 (2022).

[118]N. Gehrels, G. Chincarini, P. Giommi, K. O. Mason, J. A. Nousek, A. A. Wells, N. E. White, S. D. Barthelmy, D. N. Burrows, L. R. Cominsky, K. C. Hurley, F. E. Marshall, P. Mészáros, P. W. A. Roming, L. Angelini, L. M. Barbier, T. Belloni, S. Campana, P. A. Caraveo, M. M. Chester, O. Citterio, T. L. Cline, M. S. Cropper, J. R. Cummings, A. J. Dean, E. D. Feigelson, E. E. Fenimore, D. A. Frail, A. S. Fruchter, G. P. Garmire, K. Gendreau, G. Ghisellini, J. Greiner, J. E. Hill, S. D. Hunsberger, H. A. Krimm, S. R. Kulkarni, P. Kumar, F. Lebrun, N. M. Lloyd-Ronning, C. B. Markwardt, B. J. Mattson, R. F. Mushotzky, J. P. Norris, J. Osborne, B. Paczynski, D. M. Palmer, H. S. Park, A. M. Parsons, J. Paul, M. J. Rees, C. S. Reynolds, J. E. Rhoads, T. P. Sasseen, B. E. Schaefer, A. T. Short, A. P. Smale, I. A. Smith, L. Stella, G. Tagliaferri, T. Takahashi, M. Tashiro, L. K. Townsley, J. Tueller, M. J. L. Turner, M. Vietri, W. Voges, M. J. Ward, R. Willingale, F. M. Zerbi, and W. W. Zhang, "The Swift Gamma-Ray Burst Mission," **611**, 1005–1020 (2004).

[119]V. Ghirardini, E. Bulbul, E. Artis, N. Clerc, C. Garrel, S. Grandis, M. Kluge, A. Liu, Y. E. Bahar, F. Balzer, I. Chiu, J. Comparat, D. Gruen, F. Kleinebreil, S. Krippendorf, A. Merloni, K. Nandra, N. Okabe, F. Pacaud, P. Predehl, M. E. Ramos-Ceja, T. H. Reiprich, J. S. Sanders, T. Schrabback, R. Seppi, S. Zelmer, X. Zhang, W. Bornemann, H. Brunner, V. Burwitz, D. Coutinho, K. Dennerl, M. Freyberg, S. Friedrich, R. Gaida, A. Gueguen, F. Haberl, W. Kink, G. Lamer, X. Li, T. Liu, C. Maitra, N. Meidinger, S. Mueller, H. Miyatake, S. Miyazaki, J. Robrade, A. Schwope, and I. Stewart, "The SRG/eROSITA All-Sky Survey: Cosmology Constraints from Cluster Abundances in the Western Galactic Hemisphere," arXiv e-prints, arXiv:2402.08458, arXiv:2402.08458 (2024).

[120]G. Ghisellini and P. Madau, "On the origin of the gamma-ray emission in blazars," **280**, 67–76 (1996).

[121]R. Giacconi, J. Bechtold, G. Branduardi, W. Forman, J. P. Henry, C. Jones, E. Kellogg, H. van der Laan, W. Liller, H. Marshall, S. S. Murray, J. Pye, E. Schreier, W. L. W. Sargent, F. Seward, and H. Tananbaum, "A high-sensitivity X-ray survey using the Einstein Observatory and the discrete source contribution to the extragalactic X-ray background.," **234**, L1–L7 (1979).

[122]R. Giacconi, G. Branduardi, U. Briel, A. Epstein, D. Fabricant, E. Feigelson, W. Forman, P. Gorenstein, J. Grindlay, H. Gursky, F. R. Harnden, J. P. Henry, C. Jones, E. Kellogg, D. Koch, S. Murray, E. Schreier, F. Seward, H. Tananbaum, K. Topka, L. Van Speybroeck, S. S. Holt, R. H. Becker, E. A.

Boldt, P. J. Serlemitsos, G. Clark, C. Canizares, T. Markert, R. Novick, D. Helfand, and K. Long, "The Einstein (HEAO 2) X-ray Observatory.," **230**, 540–550 (1979).

[123]R. Giacconi, E. Kellogg, P. Gorenstein, H. Gursky, and H. Tananbaum, "An X-Ray Scan of the Galactic Plane from UHURU," **165**, L27 (1971).

[124]R. Giacconi, H. Gursky, F. R. Paolini, and B. B. Rossi, "Evidence for x Rays From Sources Outside the Solar System," **9**, 439–443 (1962).

[125]R. Gilli, A. Comastri, and G. Hasinger, "The synthesis of the cosmic X-ray background in the Chandra and XMM-Newton era," **463**, 79–96 (2007).

[126]R. Gilli, G. Risaliti, and M. Salvati, "Beyond the standard model for the cosmic X-ray background," **347**, 424–433 (1999).

[127]R. Gilli, M. Salvati, and G. Hasinger, "Testing current synthesis models of the X-ray background," **366**, 407–417 (2001).

[128]S. Giodini, D. Pierini, A. Finoguenov, G. W. Pratt, H. Boehringer, A. Leauthaud, L. Guzzo, H. Aussel, M. Bolzonella, P. Capak, M. Elvis, G. Hasinger, O. Ilbert, J. S. Kartaltepe, A. M. Koekemoer, S. J. Lilly, R. Massey, H. J. McCracken, J. Rhodes, M. Salvato, D. B. Sanders, N. Z. Scoville, S. Sasaki, V. Smolcic, Y. Taniguchi, D. Thompson, and COSMOS Collaboration, "Stellar and Total Baryon Mass Fractions in Groups and Clusters Since Redshift 1," **703**, 982–993 (2009).

[129]P. Giommi, L. Angelini, P. Jacobs, and G. Tagliaferri, "XIMAGE a Multi-Mission X-ray Image Analysis Package," in Astronomical data analysis software and systems i, Vol. 25, edited by D. M. Worrall, C. Biemesderfer, and J. Barnes, Astronomical Society of the Pacific Conference Series (Jan. 1992), p. 100.

[130]P. Giommi, L. Angelini, P. Jacobs, and G. Tagliaferri, "XIMAGE a Multi-Mission X-ray Image Analysis Package," in Astronomical data analysis software and systems i, Vol. 25, edited by D. M. Worrall, C. Biemesderfer, and J. Barnes, Astronomical Society of the Pacific Conference Series (Jan. 1992), p. 100.

[131]X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in Proceedings of the thirteenth international conference on artificial intelligence and statistics, Vol. 9, edited by Y. W. Teh and M. Titterington, Proceedings of Machine Learning Research (2010), pp. 249–256.

[132]I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative Adversarial Networks," ArXiv e-prints (2014).

[133]I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*, `http://www.deeplearningbook.org` (MIT Press, 2016).

[134]J. F. R. Gower, P. F. Scott, and D. Wills, "A survey of radio sources in the declination ranges —07° to 20° and 40° to 80°," **71**, 49 (1967).

[135]D. A. Green, "Statistical studies of supernova remnants.," **209**, 449–478 (1984).

[136]D. A. Green, "VizieR Online Data Catalog: A Catalogue of Galactic Supernova Remnants (Green 2017)," VizieR Online Data Catalog, VII/278, VII/278 (2017).

[137]H. J. Grimm, M. Gilfanov, and R. Sunyaev, "The Milky Way in X-rays for an outside observer. Log(N)-Log(S) and luminosity function of X-ray binaries from RXTE/ASM data," **391**, 923–944 (2002).

[138]D. E. Gruber, J. L. Matteson, L. E. Peterson, and G. V. Jung, "The Spectrum of Diffuse Cosmic Hard X-Rays Measured with HEAO 1," **520**, 124–129 (1999).

[139]M. Güdel and Y. Nazé, "X-ray spectroscopy of stars," **17**, 309–408 (2009).

[140]P. W. Guilbert and A. C. Fabian, "A hot intergalactic medium?" **220**, 439–451 (1986).

[141]J.-H. Ha, D. Ryu, and H. Kang, "Properties of merger shocks in merging galaxy clusters," The Astrophysical Journal **857**, 26 (2018).

[142]F. Haardt and L. Maraschi, "X-Ray Spectra from Two-Phase Accretion Disks," **413**, 507 (1993).

[143]S. E. G. Hales, J. M. Riley, E. M. Waldram, P. J. Warner, and J. E. Baldwin, "A final non-redundant catalogue for the 7C 151-MHz survey," **382**, 1639–1642 (2007).

[144]C. R. Harris, K. J. Millman, S. J. van der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N. J. Smith, R. Kern, M. Picus, S. Hoyer, M. H. van Kerkwijk, M. Brett, A. Haldane, J. F. del Río, M. Wiebe, P. Peterson, P. Gérard-Marchant, K. Sheppard, T. Reddy, W. Weckesser, H. Abbasi, C. Gohlke, and T. E. Oliphant, "Array programming with NumPy," Nature **585**, 357–362 (2020).

[145]F. A. Harrison, J. Aird, F. Civano, G. Lansbury, J. R. Mullaney, D. R. Ballantyne, D. M. Alexander, D. Stern, M. Ajello, D. Barret, F. E. Bauer, M. Baloković, W. N. Brandt, M. Brightman, S. E. Boggs, F. E. Christensen, A. Comastri, W. W. Craig, A. Del Moro, K. Forster, P. Gandhi, P. Giommi, B. W. Grefenstette, C. J. Hailey, R. C. Hickox, A. Hornstrup, T. Kitaguchi, J. Koglin, B. Luo, K. K. Madsen, P. H. Mao, H. Miyasaka, K. Mori, M. Perri, M. Pivovaroff, S. Puccetti, V. Rana, E. Treister, D. Walton, N. J. Westergaard, D. Wik, L. Zappacosta, W. W. Zhang, and A. Zoglauer, "The NuSTAR Extragalactic Surveys: The Number Counts of Active Galactic Nuclei and the Resolved Fraction of the Cosmic X-Ray Background," **831**, 185, 185 (2016).

[146]G. Hasinger, R. Burg, R. Giacconi, M. Schmidt, J. Trumper, and G. Zamorani, "The ROSAT Deep Survey. I. X-ray sources in the Lockman Field," **329**, 482–494 (1998).

[147]G. Hasinger, I. Lehmann, R. Giacconi, M. Schmidt, J. Trümper, and G. Zamorani, "The ROSAT Deep Surveys," in Highlights in x-ray astronomy, Vol. 272, edited by B. Aschenbach and M. J. Freyberg (Jan. 1999), p. 199.

[148]K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," ArXiv e-prints (2015).

[149]S. E. Healey, R. W. Romani, G. B. Taylor, E. M. Sadler, R. Ricci, T. Murphy, J. S. Ulvestad, and J. N. Winn, "CRATES: An All-Sky Survey of Flat-Spectrum Radio Sources," **171**, 61–71 (2007).

[150]T. M. Heckman, "Peculiar nuclei and their relation to galaxy type.," **90**, 241–243 (1978).

[151]B. Hilbert, M. Chiaberge, J. P. Kotyla, G. R. Tremblay, C. Stanghellini, W. B. Sparks, S. Baum, A. Capetti, F. D. Macchetto, G. K. Miley, C. P. O'Dea, E. S. Perlman, and A. Quillen, "Powerful Activity in the Bright Ages. I. A Visible/IR Survey of High Redshift 3C Radio Galaxies and Quasars," **225**, 12, 12 (2016).

[152]R. G. Hine and M. S. Longair, "Optical spectra of 3CR radio galaxies.," **188**, 111–130 (1979).

[153]G. Hinshaw, D. Larson, E. Komatsu, D. N. Spergel, C. L. Bennett, J. Dunkley, M. R. Nolta, M. Halpern, R. S. Hill, N. Odegard, L. Page, K. M. Smith, J. L. Weiland, B. Gold, N. Jarosik, A. Kogut, M. Limon, S. S. Meyer, G. S. Tucker, E. Wollack, and E. L. Wright, "Nine-year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Cosmological Parameter Results," **208**, 19, 19 (2013).

[154]J. G. Holland, H. Böhringer, G. Chon, and D. Pierini, "Optical and X-ray profiles in the REXCESS sample of galaxy clusters*," **448**, 2644–2664 (2015).

[155]T. Hovatta, M. L. Lister, M. F. Aller, H. D. Aller, D. C. Homan, Y. Y. Kovalev, A. B. Pushkarev, and T. Savolainen, "MOJAVE: Monitoring of Jets in Active Galactic Nuclei with VLBA Experiments. VIII. Faraday Rotation in Parsec-scale AGN Jets," **144**, 105, 105 (2012).

[156]A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications," ArXiv e-prints (2017).

[157]J. D. Hunter, "Matplotlib: a 2d graphics environment," Computing in science & engineering **9**, 90–95 (2007).

[158]Y. Ichinohe and S. Yamada, "Neural network-based anomaly detection for high-resolution X-ray spectroscopy," **487**, 2874–2880 (2019).

[159]Z. Ivezic, J. A. Tyson, B. Abel, E. Acosta, R. Allsman, and A. et al., "LSST: from Science Drivers to Reference Design and Anticipated Data Products," ArXiv e-prints (2008).

[160]F. A. Jansen, "XMM: advancing science with the high-throughput X-ray spectroscopy mission.," ESA Bulletin **100**, 9–12 (1999).

[161]F. Jansen, D. Lumb, B. Altieri, J. Clavel, M. Ehle, C. Erd, C. Gabriel, M. Guainazzi, P. Gondoin, R. Much, R. Munoz, M. Santos, N. Schartel, D. Texier, and G. Vacanti, "XMM-Newton observatory. I. The spacecraft and operations," **365**, L1–L6 (2001).

[162]T. H. Jarrett, T. Chester, R. Cutri, S. Schneider, M. Skrutskie, and J. P. Huchra, "2MASS Extended Source Catalog: Overview and Algorithms," **119**, 2498–2531 (2000).

[163]D. H. Jones, M. A. Read, W. Saunders, M. Colless, T. Jarrett, Q. A. Parker, A. P. Fairall, T. Mauch, E. M. Sadler, F. G. Watson, D. Burton, L. A. Campbell, P. Cass, S. M. Croom, J. Dawe, K. Fiegert, L. Frankcombe, M. Hartley, J. Huchra, D. James, E. Kirby, O. Lahav, J. Lucey, G. A. Mamon, L. Moore, B. A. Peterson, S. Prior, D. Proust, K. Russell, V. Safouris, K.-I. Wakamatsu, E. Westra, and M. Williams, "The 6dF Galaxy Survey: final redshift release (DR3) and southern large-scale structures," **399**, 683–698 (2009).

[164]K. I. Kellermann, R. Sramek, M. Schmidt, D. B. Shaffer, and R. Green, "VLA Observations of Objects in the Palomar Bright Quasar Survey," **98**, 1195 (1989).

[165]L. J. Kewley, M. A. Dopita, R. S. Sutherland, C. A. Heisler, and J. Trevena, "Theoretical Modeling of Starburst Galaxies," **556**, 121–140 (2001).

[166]L. J. Kewley, C. Maier, K. Yabe, K. Ohta, M. Akiyama, M. A. Dopita, and T. Yuan, "The Cosmic BPT Diagram: Confronting Theory with Observations," **774**, L10, L10 (2013).

[167]E. Y. Khachikian and D. W. Weedman, "An atlas of Seyfert galaxies.," **192**, 581–589 (1974).

[168]E. E. Khachikian and D. W. Weedman, "A spectroscopic study of luminous galactic nuclei.," Astrofizika **7**, 389–406 (1971).

[169]C. Knigge, M. J. Coe, and P. Podsiadlowski, "Two populations of X-ray pulsars produced by two types of supernova," **479**, 372–375 (2011).

[170]L. A. Kofman, N. Y. Gnedin, and N. A. Bahcall, "Cosmological Constant, COBE Cosmic Microwave Background Anisotropy, and Large-Scale Clustering," **413**, 1 (1993).

[171]W. Kollatschny and M. Zetzl, "The shape of broad-line profiles in active galactic nuclei," **549**, A100, A100 (2013).

[172]M. Kosiba, H. A. Peña-Herazo, F. Massaro, N. Masetti, A. Paggi, V. Chavushyan, E. Bottacini, and N. Werner, "A multifrequency characterization of the extragalactic hard X-ray sky. Presenting the second release of the Turin-SyCAT," **670**, A171, A171 (2023).

[173]M. Kosiba, M. Lieu, B. Altieri, N. Clerc, L. Faccioli, S. Kendrew, I. Valtchanov, T. Sadibekova, M. Pierre, F. Hroch, N. Werner, L. Burget, C. Garrel, E. Koulouridis, E. Gaynullina, M. Molham, M. E. Ramos-Ceja, and A. Khalikova, "Multiwavelength classification of X-ray selected galaxy cluster candidates using convolutional neural networks," **496**, 4141–4153 (2020).

[174]M. J. Koss, C. Ricci, B. Trakhtenbrot, K. Oh, J. S. den Brok, J. E. Mejía-Restrepo, D. Stern, G. C. Privon, E. Treister, M. C. Powell, R. Mushotzky, F. E. Bauer, T. T. Ananna, M. Baloković, R. E. Bär, G. Becker, P. Bessiere, L. Burtscher, T. Caglar, E. Congiu, P. Evans, F. Harrison, M. Heida, K. Ichikawa, N. Kamraj, I. Lamperti, F. Pacucci, F. Ricci, R. Riffel, A. F. Rojas, K. Schawinski, M. J. Temple, C. M. Urry, S. Veilleux, and J. Williams, "BASS. XXII. The BASS DR2 AGN Catalog and Data," **261**, 2, 2 (2022).

[175]M. Koss, B. Trakhtenbrot, C. Ricci, I. Lamperti, K. Oh, S. Berney, K. Schawinski, M. Baloković, L. Baronchelli, D. M. Crenshaw, T. Fischer, N. Gehrels, F. Harrison, Y. Hashimoto, D. Hogg, K. Ichikawa, N. Masetti, R. Mushotzky, L. Sartori, D. Stern, E. Treister, Y. Ueda, S. Veilleux, and L. Winter, "BAT AGN Spectroscopic Survey. I. Spectral Measurements, Derived Quantities, and AGN Demographics," **850**, 74, 74 (2017).

[176]M. Koss, B. Trakhtenbrot, C. Ricci, M. Powell, C. M. Urry, D. Stern, and K. Oh, "The BAT AGN Spectropic Survey (BASS) DR2, A Thousand New Spectra," in American astronomical society meeting abstracts #233, Vol. 233, American Astronomical Society Meeting Abstracts (Jan. 2019), p. 431.04.

[177]J. P. Kotyla, M. Chiaberge, S. Baum, A. Capetti, B. Hilbert, F. D. Macchetto, G. K. Miley, C. P. O'Dea, E. S. Perlman, W. B. Sparks, and G. R. Tremblay, "The Environment of z > 1 3CR Radio Galaxies and QSOs: From Proto-clusters to Clusters of Galaxies?" **826**, 46, 46 (2016).

[178]E. Koulouridis, N. Clerc, T. Sadibekova, M. Chira, E. Drigga, L. Faccioli, J. P. Le Fèvre, C. Garrel, E. Gaynullina, A. Gkini, M. Kosiba, F. Pacaud, M. Pierre, J. Ridl, K. Tazhenova, C. Adami, B. Altieri, J. C. Baguley, R. Cabanac, E. Cucchetti, A. Khalikova, M. Lieu, J. B. Melin, M. Molham, M. E. Ramos-Ceja, G. Soucail, A. Takey, and I. Valtchanov, "The X-CLASS survey: A catalogue of 1646 X-ray-selected galaxy clusters up to z ∼ 1.5," **652**, A12, A12 (2021).

[179]R. Krivonos, M. Revnivtsev, A. Lutovinov, S. Sazonov, E. Churazov, and R. Sunyaev, "INTEGRAL/IBIS all-sky survey in hard X-rays," **475**, 775–784 (2007).

[180]R. Krivonos, M. Revnivtsev, A. Lutovinov, S. Sazonov, E. Churazov, and R. Sunyaev, "INTEGRAL/IBIS all-sky survey in hard X-rays," **475**, 775–784 (2007).

[181]R. Krivonos, S. Tsygankov, A. Lutovinov, M. Revnivtsev, E. Churazov, and R. Sunyaev, "INTEGRAL 11-year hard X-ray survey above 100 keV," **448**, 3766–3774 (2015).

[182]R. A. Krivonos, A. J. Bird, E. M. Churazov, J. A. Tomsick, A. Bazzano, V. Beckmann, G. Bélanger, A. Bodaghee, S. Chaty, E. Kuulkers, A. Lutovinov, A. Malizia, N. Masetti, I. A. Mereminskiy, R. Sunyaev, S. S. Tsygankov, P. Ubertini, and C. Winkler, "15 years of galactic surveys and hard X-ray background measurements," **92**, 101612, 101612 (2021).

[183]R. A. Krivonos, S. Y. Sazonov, E. A. Kuznetsova, A. A. Lutovinov, I. A. Mereminskiy, and S. S. Tsygankov, "INTEGRAL/IBIS 17-yr hard X-ray all-sky survey," **510**, 4796–4807 (2022).

[184]R. A. Krivonos, S. S. Tsygankov, I. A. Mereminskiy, A. A. Lutovinov, S. Y. Sazonov, and R. A. Sunyaev, "New hard X-ray sources discovered in the ongoing INTEGRAL Galactic plane survey after 14 yr of observations," **470**, 512–516 (2017).

[185]A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems 25*, edited by F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger (Curran Associates, Inc., 2012), pp. 1097–1105.

[186]J. H. Krolik and M. C. Begelman, "Molecular Tori in Seyfert Galaxies: Feeding the Monster and Hiding It," **329**, 702 (1988).

[187]R. Landi, L. Bassani, A. Bazzano, A. J. Bird, M. Fiocchi, A. Malizia, F. Panessa, V. Sguera, and P. Ubertini, "Investigating the X-ray counterparts to unidentified sources in the 1000-orbit INTEGRAL/IBIS catalogue," **470**, 1107–1120 (2017).

[188]M. Landoni, F. Massaro, A. Paggi, R. D'Abrusco, D. Milisavljevic, N. Masetti, H. A. Smith, G. Tosti, L. Chomiuk, J. Strader, and C. C. Cheung, "Optical Spectroscopic Observations of γ-Ray Blazar Candidates. III. The 2013/2014 Campaign in the Southern Hemisphere," **149**, 163, 163 (2015).

[189]F. Lanusse, R. Mandelbaum, S. Ravanbakhsh, C.-L. Li, P. Freeman, and B. Póczos, "Deep generative models for galaxy image simulations," **504**, 5543–5555 (2021).

[190]B. M. Lasker, J. Doggett, B. McLean, C. Sturch, S. Djorgovski, R. R. de Carvalho, and I. N. Reid, "The Palomar–ST ScI Digitized Sky Survey (POSS–II): Preliminary Data Availability," in Astronomical data analysis software and systems v, Vol. 101, edited by G. H. Jacoby and J. Barnes, Astronomical Society of the Pacific Conference Series (Jan. 1996), p. 88.

[191]C. Law and F. Yusef-Zadeh, "X-Ray Observations of Stellar Clusters Near the Galactic Center," **611**, 858–870 (2004).

[192]Y. LeCun, P. Haffner, L. Bottou, and Y. Bengio, "Object recognition with gradient-based learning," in Shape, contour and grouping in computer vision (1999), pp. 319–.

[193]F. Lebrun, J. P. Leray, P. Lavocat, J. Crétolle, M. Arquès, C. Blondel, C. Bonnin, A. Bouère, C. Cara, T. Chaleil, F. Daly, F. Desages, H. Dzitko, B. Horeau, P. Laurent, O. Limousin, F. Mathy, V. Mauguen, F. Meignier, F. Molinié, E. Poindron, M. Rouger, A. Sauvageon, and T. Tourrette, "ISGRI: The INTEGRAL Soft Gamma-Ray Imager," **411**, L141–L148 (2003).

[194]Y. Lecun, B. Boser, J. Denker, D. Henderson, R. Howard, W. Hubbard, and L. Jackel, "Backpropagation applied to handwritten zip code recognition," English (US), Neural Computation **1**, 541–551 (1989).

[195]G. von Leibniz, J. Child, and C. Gerhardt, *The early mathematical manuscripts of leibniz: translated from the latin texts published by carl immanuel gerhardt with critical and historical notes* (Open court publishing Company, 1920).

[196]A. M. Levine, F. L. Lang, W. H. G. Lewin, F. A. Primini, C. A. Dobson, J. P. Doty, J. A. Hoffman, S. K. Howe, A. Scheepmaker, W. A. Wheaton, J. L. Matteson, W. A. Baity, D. E. Gruber, F. K. Knight, P. L. Nolan, R. M. Pelling, R. E. Rothschild, and L. E. Peterson, "The HEAO1 A-4 catalog of high-energy X-ray sources.," **54**, 581–617 (1984).

[197]A. Lien, H. Krimm, C. Markwardt, N. Collins, S. Barthelmy, K. Oh, M. Koss, T. Parsotan, and S. Cenko, "The BAT 157 month survey catalog and beyond," in American astronomical society meeting abstracts, Vol. 55, American Astronomical Society Meeting Abstracts (Jan. 2023), p. 254.07.

[198]M. Lieu, L. Conversi, B. Altieri, and B. Carry, "Detecting solar system objects with convolutional neural networks," ArXiv e-prints (2018).

[199]M. Lieu, L. Conversi, B. Altieri, and B. Carry, "Detecting Solar system objects with convolutional neural networks," **485**, 5831–5842 (2019).

[200]M. Lin, Q. Chen, and S. Yan, "Network In Network," arXiv e-prints, arXiv:1312.4400, arXiv:1312.4400 (2013).

[201]C. J. Lintott, K. Schawinski, A. Slosar, K. Land, and B. et al., "Galaxy Zoo: morphologies derived from visual inspection of galaxies from the Sloan Digital Sky Survey," **389**, 1179–1189 (2008).

[202]J. Liu, Y. Liu, X. Li, W. Xu, L. Gou, and C. Cheng, "Clumpy tori around type II active galactic nuclei as revealed by X-ray fluorescent lines," **459**, L100–L104 (2016).

[203]Q. Z. Liu, J. van Paradijs, and E. P. J. van den Heuvel, "A catalogue of low-mass X-ray binaries in the Galaxy, LMC, and SMC (Fourth edition)," **469**, 807–810 (2007).

[204]Q. Z. Liu, J. van Paradijs, and E. P. J. van den Heuvel, "Catalogue of high-mass X-ray binaries in the Galaxy (4th edition)," **455**, 1165–1168 (2006).

[205]G. Macario, M. Markevitch, S. Giacintucci, G. Brunetti, T. Venturi, and S. S. Murray, "A shock front in the merging galaxy cluster a754: x-ray and radio observations," The Astrophysical Journal **728**, 82 (2011).

[206]J. P. Madrid, M. Chiaberge, D. Floyd, W. B. Sparks, D. Macchetto, G. K. Miley, D. Axon, A. Capetti, C. P. O'Dea, S. Baum, E. Perlman, and A. Quillen, "Hubble Space Telescope Near-Infrared Snapshot Survey of 3CR Radio Source Counterparts at Low Redshift," **164**, 307–333 (2006).

[207]M. Magdon-Ismail and A. Atiya, "Neural networks for density estimation," in Advances in neural information processing systems, Vol. 11, edited by M. Kearns, S. Solla, and D. Cohn (1998).

[208]A. Malizia, L. Bassani, V. Sguera, J. B. Stephen, A. Bazzano, M. Fiocchi, and A. J. Bird, "XMM-Newton observations of unidentified INTEGRAL/IBIS sources," **408**, 975–982 (2010).

[209]R. N. Manchester, G. B. Hobbs, A. Teoh, and M. Hobbs, "The Australia Telescope National Facility Pulsar Catalogue," **129**, 1993–2006 (2005).

[210]R. N. Manchester, A. G. Lyne, F. Camilo, J. F. Bell, V. M. Kaspi, N. D'Amico, N. P. F. McKay, F. Crawford, I. H. Stairs, A. Possenti, M. Kramer, and D. C. Sheppard, "The Parkes multi-beam pulsar survey - I. Observing and data analysis systems, discovery and timing of 100 pulsars," **328**, 17–35 (2001).

[211]A. B. Mantz, A. von der Linden, S. W. Allen, D. E. Applegate, P. L. Kelly, R. G. Morris, D. A. Rapetti, R. W. Schmidt, S. Adhikari, M. T. Allen, P. R. Burchat, D. L. Burke, M. Cataneo, D. Donovan, H. Ebeling, S. Shandera, and A. Wright, "Weighing the giants - IV. Cosmology and neutrino mass," **446**, 2205–2225 (2015).

[212]L. Maraschi, G. Fossati, F. Tavecchio, L. Chiappetti, A. Celotti, G. Ghisellini, P. Grandi, E. Pian, G. Tagliaferri, A. Treves, A. C. Breslin, J. H. Buckley, D. A. Carter-Lewis, M. Catanese, M. F. Cawley, D. J. Fegan, S. Fegan, J. Finley, J. Gaidos, T. Hall, A. M. Hillas, F. Krennrich, R. W. Lessard, C. Masterson, P. Moriarty, J. Quinn, J. Rose, F. Samuelson, T. C. Weekes, C. M. Urry, and T. Takahashi, "Simultaneous X-Ray and TEV Observations of a Rapid Flare from Markarian 421," **526**, L81–L84 (1999).

[213]L. Maraschi, G. Ghisellini, and A. Celotti, "A Jet Model for the Gamma-Ray–emitting Blazar 3C 279," **397**, L5 (1992).

[214]E. J. Marchesini, N. Masetti, E. Palazzi, V. Chavushyan, E. Jiménez-Bailón, V. M. Patiño-Álvarez, V. Reynaldi, A. F. Rojas, I. Saviane, I. Andruchow, L. Bassani, A. Bazzano, A. J. Bird, A. Malizia, D. Minniti, L. Monaco, J. B. Stephen, and P. Ubertini, "Optical spectroscopic classification of 35 hard X-ray sources from the Swift-BAT 70-month catalogue," **364**, 153, 153 (2019).

[215]E. J. Marchesini, A. Paggi, F. Massaro, N. Masetti, R. D'Abrusco, and I. Andruchow, "The γ-ray sky seen at X-ray energies. II. The Swift hunt of Fermi BL Lac objects among unidentified γ-ray sources," **638**, A128, A128 (2020).

[216]E. J. Marchesini, A. Paggi, F. Massaro, N. Masetti, R. D'Abrusco, I. Andruchow, and R. de Menezes, "The γ-ray sky seen at X-ray energies. I. Searching for the connection between X-rays and γ-rays in Fermi BL Lac objects," **631**, A150, A150 (2019).

[217]C. B. Markwardt, J. Tueller, G. K. Skinner, N. Gehrels, S. D. Barthelmy, and R. F. Mushotzky, "The Swift/BAT High-Latitude Survey: First Results," **633**, L77–L80 (2005).

[218]A. P. Marscher and W. K. Gear, "Models for high-frequency radio outbursts in extragalactic sources, with application to the early 1983 millimeter-to-infrared flare of 3C 273.," **298**, 114–127 (1985).

[219]A. P. Marscher and J. P. Travis, "Synchrotron self-Compton interpretation of multiwaveband observations of gamma-ray bright blazars.," **120**, 537–540 (1996).

[220]F. E. Marshall, E. A. Boldt, S. S. Holt, R. B. Miller, R. F. Mushotzky, L. A. Rose, R. E. Rothschild, and P. J. Serlemitsos, "The diffuse X-ray background spectrum from 3 to 50 keV.," **235**, 4–10 (1980).

[221]A. Maselli, F. Massaro, G. Cusumano, V. La Parola, D. E. Harris, A. Paggi, E. Liuzzo, G. R. Tremblay, S. A. Baum, and C. P. O'Dea, "Swift observations of unidentified radio sources in the revised Third Cambridge Catalogue," **460**, 3829–3837 (2016).

[222]N. Masetti, E. Mason, L. Bassani, A. J. Bird, E. Maiorano, A. Malizia, E. Palazzi, J. B. Stephen, A. Bazzano, A. J. Dean, P. Ubertini, and R. Walter, "Unveiling the nature of INTEGRAL objects through optical spectroscopy. II. The nature of four unidentified sources," **448**, 547–556 (2006).

[223]N. Masetti, E. Mason, L. Morelli, S. A. Cellone, V. A. McBride, E. Palazzi, L. Bassani, A. Bazzano, A. J. Bird, P. A. Charles, A. J. Dean, G. Galaz, N. Gehrels, R. Landi, A. Malizia, D. Minniti, F. Panessa, G. E. Romero, J. B. Stephen, P. Ubertini, and R. Walter, "Unveiling the nature of INTEGRAL objects through optical spectroscopy. VI. A multi-observatory identification campaign," **482**, 113–132 (2008).

[224]N. Masetti, L. Morelli, E. Palazzi, G. Galaz, L. Bassani, A. Bazzano, A. J. Bird, A. J. Dean, G. L. Israel, R. Landi, A. Malizia, D. Minniti, F. Schiavone, J. B. Stephen, P. Ubertini, and R. Walter, "Unveiling the nature of INTEGRAL objects through optical spectroscopy. V. Identification and properties of 21 southern hard X-ray sources," **459**, 21–30 (2006).

[225]N. Masetti, L. Morelli, E. Palazzi, G. Galaz, L. Bassani, A. Bazzano, A. J. Bird, A. J. Dean, G. L. Israel, R. Landi, A. Malizia, D. Minniti, F. Schiavone, J. B. Stephen, P. Ubertini, and R. Walter, "Unveiling the nature of INTEGRAL objects through optical spectroscopy. V. Identification and properties of 21 southern hard X-ray sources," **459**, 21–30 (2006).

[226]N. Masetti, P. Parisi, E. Jiménez-Bailón, E. Palazzi, V. Chavushyan, L. Bassani, A. Bazzano, A. J. Bird, A. J. Dean, G. Galaz, R. Landi, A. Malizia, D. Minniti, L. Morelli, F. Schiavone, J. B. Stephen, and P. Ubertini, "Unveiling the nature of INTEGRAL objects through optical spectroscopy. IX. Twenty two more identifications, and a glance into the far hard X-ray Universe," **538**, A123, A123 (2012).

[227]N. Masetti, P. Parisi, E. Palazzi, E. Jiménez-Bailón, V. Chavushyan, V. McBride, A. F. Rojas, L. Steward, L. Bassani, A. Bazzano, A. J. Bird, P. A. Charles, G. Galaz, R. Landi, A. Malizia, E. Mason, D. Minniti, L. Morelli, F. Schiavone, J. B. Stephen, and P. Ubertini, "Unveiling the nature of INTEGRAL objects through optical spectroscopy. X. A new multi-year, multi-observatory campaign," **556**, A120, A120 (2013).

[228]E. Massaro, P. Giommi, C. Leto, P. Marchegiani, A. Maselli, M. Perri, S. Piranomonte, and S. Sclavi, "Roma-BZCAT: a multifrequency catalogue of blazars," **495**, 691–696 (2009).

[229]E. Massaro, A. Maselli, C. Leto, P. Marchegiani, M. Perri, P. Giommi, and S. Piranomonte, "The 5th edition of the Roma-BZCAT. A short presentation," **357**, 75, 75 (2015).

[230]E. Massaro, A. Tramacere, M. Perri, P. Giommi, and G. Tosti, "Log-parabolic spectra and particle acceleration in blazars. III. SSC emission in the TeV band from Mkn501," **448**, 861–871 (2006).

[231]F. Massaro and R. D'Abrusco, "The Infrared-Gamma-Ray Connection: A WISE View of the Extragalactic Gamma-Ray Sky," **827**, 67, 67 (2016).

[232]F. Massaro, R. D'Abrusco, M. Ajello, J. E. Grindlay, and H. A. Smith, "Identification of the Infrared Non-thermal Emission in Blazars," **740**, L48, L48 (2011).

[233]F. Massaro, R. D'Abrusco, M. Landoni, A. Paggi, N. Masetti, M. Giroletti, H. Otí-Floranes, V. Chavushyan, E. Jiménez-Bailón, V. Patiño-Álvarez, S. W. Digel, H. A. Smith, and G. Tosti, "Refining the Associations of the Fermi Large Area Telescope Source Catalogs," **217**, 2, 2 (2015).

[234]F. Massaro, R. D'Abrusco, A. Paggi, N. Masetti, M. Giroletti, G. Tosti, H. A. Smith, and S. Funk, "Unveiling the Nature of the Unidentified Gamma-Ray Sources. V. Analysis of the Radio Candidates with the Kernel Density Estimation," **209**, 10, 10 (2013).

[235]F. Massaro, R. D'Abrusco, G. Tosti, M. Ajello, D. Gasparrini, J. E. Grindlay, and H. A. Smith, "The WISE Gamma-Ray Strip Parameterization: The Nature of the Gamma-Ray Active Galactic Nuclei of Uncertain Type," **750**, 138, 138 (2012).

[236]F. Massaro, R. D'Abrusco, G. Tosti, M. Ajello, A. Paggi, and D. Gasparrini, "Unidentified γ-Ray Sources: Hunting γ-Ray Blazars," **752**, 61, 61 (2012).

[237]F. Massaro, P. Giommi, G. Tosti, A. Cassetti, R. Nesci, M. Perri, D. Burrows, and N. Gerehls, "Swift observations of IBL and LBL objects," **489**, 1047–1054 (2008).

[238]F. Massaro, D. E. Harris, E. Liuzzo, M. Orienti, R. Paladino, A. Paggi, G. R. Tremblay, B. J. Wilkes, J. Kuraszkiewicz, S. A. Baum, and C. P. O'Dea, "The Chandra Survey of Extragalactic Sources in the 3CR Catalog: X-ray Emission from Nuclei, Jets, and Hotspots in the Chandra Archival Observations," **220**, 5, 5 (2015).

[239]F. Massaro, D. E. Harris, G. R. Tremblay, E. Liuzzo, A. Bonafede, and A. Paggi, "A Chandra Snapshot Survey for 3C Radio Galaxies with Redshifts between 0.3 and 0.5," **206**, 7, 7 (2013).

[240]F. Massaro, N. Masetti, R. D'Abrusco, A. Paggi, and S. Funk, "Optical Spectroscopic Observations of Blazars and γ-Ray Blazar Candidates in the Sloan Digital Sky Survey Data Release Nine," **148**, 66, 66 (2014).

[241]F. Massaro, A. Tramacere, A. Cavaliere, M. Perri, and P. Giommi, "X-ray spectral evolution of TeV BL Lacertae objects: eleven years of observations with BeppoSAX, XMM-Newton and Swift satellites," **478**, 395–401 (2008).

[242]F. Massaro, S. V. White, A. García-Pérez, A. Jimenez-Gallardo, A. Capetti, C. C. Cheung, W. R. Forman, C. Mazzucchelli, A. Paggi, N. P. H. Nesvadba, J. P. Madrid, I. Andruchow, S. Cellone, H. A. Peña-Herazo, R. Grossová, B. Balmaverde, E. Sani, V. Chavushyan, R. P. Kraft, V. Reynaldi, and C. Leto, "Powerful Radio Sources in the Southern Sky. I. Optical Identifications," **265**, 32, 32 (2023).

[243]F. Massaro, S. V. White, A. Paggi, A. Jimenez-Gallardo, J. P. Madrid, C. Mazzucchelli, W. R. Forman, A. Capetti, C. Leto, A. García-Pérez, C. C. Cheung, V. Chavushyan, N. P. H. Nesvadba, I. Andruchow, H. A. Peña-Herazo, E. Sani, R. Grossová, V. Reynaldi, R. P. Kraft, B. Balmaverde, and S. Cellone, "Powerful Radio Sources in the Southern Sky. II. A Swift X-Ray Perspective," **268**, 32, 32 (2023).

[244]T. Mauch, T. Murphy, H. J. Buttery, J. Curran, R. W. Hunstead, B. Piestrzynski, J. G. Robertson, and E. M. Sadler, "SUMSS: a wide-field radio imaging survey of the southern sky - II. The source catalogue," **342**, 1117–1130 (2003).

[245]W. Mcculloch and W. Pitts, "A logical calculus of ideas immanent in nervous activity," Bulletin of Mathematical Biophysics **5**, 127–147 (1943).

[246]N. Mehrtens, A. K. Romer, M. Hilton, E. J. Lloyd-Davies, C. J. Miller, S. A. Stanford, M. Hosmer, B. Hoyle, C. A. Collins, A. R. Liddle, P. T. P. Viana, R. C. Nichol, J. P. Stott, E. N. Dubois, S. T. Kay, M. Sahlén, O. Young, C. J. Short, L. Christodoulou, W. A. Watson, M. Davidson, C. D. Harrison, L. Baruah, M. Smith, C. Burke, J. A. Mayers, P.-J. Deadman, P. J. Rooney, E. M. Edmondson, M. West, H. C. Campbell, A. C. Edge, R. G. Mann, K. Sabirli, D. Wake, C. Benoist, L. da Costa, M. A. G. Maia, and R. Ogando, "The XMM Cluster Survey: optical analysis methodology and the first data release," **423**, 1024–1052 (2012).

[247]I. A. Mereminskiy, R. A. Krivonos, A. A. Lutovinov, S. Y. Sazonov, M. G. Revnivtsev, and R. A. Sunyaev, "INTEGRAL/IBIS deep extragalactic survey: M81, LMC and 3C 273/Coma fields," **459**, 140–150 (2016).

[248]A. Merloni, P. Predehl, W. Becker, H. Böhringer, T. Boller, H. Brunner, M. Brusa, K. Dennerl, M. Freyberg, P. Friedrich, A. Georgakakis, F. Haberl, G. Hasinger, N. Meidinger, J. Mohr, K. Nandra, A. Rau, T. H. Reiprich, J. Robrade, M. Salvato, A. Santangelo, M. Sasaki, A. Schwope, J. Wilms, and t. German eROSITA Consortium, "eROSITA Science Book: Mapping the Structure of the Energetic Universe," ArXiv e-prints (2012).

[249]I. F. Mirabel, "Microquasars: Summary and Outlook," in *Lecture notes in physics, berlin springer verlag*, Vol. 794, edited by T. Belloni (2010), p. 1.

[250]V. Missaglia, F. Massaro, E. Liuzzo, A. Paggi, R. P. Kraft, W. R. Forman, A. Jimenez-Gallardo, J. P. Madrid, F. Ricci, C. Stuardi, B. J. Wilkes, S. A. Baum, C. P. O'Dea, J. Kuraszkiewicz, G. R. Tremblay, A. Maselli, A. Capetti, E. Sani, B. Balmaverde, and D. E. Harris, "Hidden Treasures in the Unknown 3CR Extragalactic Radio Sky: A Multiwavelength Approach," **255**, 18, 18 (2021).

[251]R. J. Mitchell, R. J. Dickens, S. J. B. Burnell, and J. L. Culhane, "The X-ray spectra of clusters of galaxies and their relationship to other cluster properties.," **189**, 329–361 (1979).

[252]T. Miyaji, A. S. Wilson, and I. Perez-Fournon, "The Radio Source and Bipolar Nebulosity in the Seyfert Galaxy NGC 3516," **385**, 137 (1992).

[253]A. T. Moffet, "The Structure of Radio Galaxies," **4**, 145 (1966).

[254]T. Mondal and B. Mukhopadhyay, "FSRQ/BL Lac dichotomy as the magnetized advective accretion process around black holes: a unified classification of blazars," **486**, 3465–3472 (2019).

[255]Nasa High Energy Astrophysics Science Archive Research Center (Heasarc), *HEAsoft: Unified Release of FTOOLS and XANADU*, Astrophysics Source Code Library, record ascl:1408.004, Aug. 2014.

[256]J. Nevalainen, R. Lieu, M. Bonamente, and D. Lumb, "Soft X-Ray Excess Emission in Clusters of Galaxies Observed with XMM-Newton," **584**, 716–728 (2003).

[257]M. Ntampaka, J. ZuHone, D. Eisenstein, D. Nagai, A. Vikhlinin, L. Hernquist, F. Marinacci, D. Nelson, R. Pakmor, A. Pillepich, P. Torrey, and M. Vogelsberger, "A Deep Learning Approach to Galaxy Cluster X-Ray Masses," **876**, 82, 82 (2019).

[258]S. C. Odewahn, E. B. Stockwell, R. L. Pennington, R. M. Humphreys, and W. A. Zumach, "Automated star/galaxy discrimination with neural networks," **103**, 318–331 (1992).

[259]K. Oh, M. Koss, C. B. Markwardt, K. Schawinski, W. H. Baumgartner, S. D. Barthelmy, S. B. Cenko, N. Gehrels, R. Mushotzky, A. Petulante, C. Ricci, A. Lien, and B. Trakhtenbrot, "The 105-Month Swift-BAT All-sky Hard X-Ray Survey," **235**, 4, 4 (2018).

[260]T. E. Oliphant, *A guide to numpy*, Vol. 1 (Trelgol Publishing USA, 2006).

[261]L. M. Oskinova, "Evolution of X-ray emission from young massive star clusters," **361**, 679–694 (2005).

[262]J. Oukbir and A. Blanchard, "X-ray clusters in open universes.," **262**, L21–L24 (1992).

[263]F. Pacaud, N. Clerc, P. A. Giles, C. Adami, T. Sadibekova, M. Pierre, B. J. Maughan, M. Lieu, J. P. Le Fèvre, S. Alis, B. Altieri, F. Ardila, I. Baldry, C. Benoist, M. Birkinshaw, L. Chiappetti, J. Démoclès, D. Eckert, A. E. Evrard, L. Faccioli, F. Gastaldello, L. Guennou, C. Horellou, A. Iovino, E. Koulouridis, V. Le Brun, C. Lidman, J. Liske, S. Maurogordato, F. Menanteau, M. Owers, B. Poggianti, D. Pomarède, E. Pompei, T. J. Ponman, D. Rapetti, T. H. Reiprich, G. P. Smith, R. Tuffs, P. Valageas, I. Valtchanov, J. P. Willis, and F. Ziparo, "The XXL Survey. II. The bright cluster sample: catalogue and luminosity function," **592**, A2, A2 (2016).

[264]F. Pacaud, M. Pierre, J. B. Melin, C. Adami, A. E. Evrard, S. Galli, F. Gastaldello, B. J. Maughan, M. Sereno, S. Alis, B. Altieri, M. Birkinshaw, L. Chiappetti, L. Faccioli, P. A. Giles, C. Horellou, A. Iovino, E. Koulouridis, J. P. Le Fèvre, C. Lidman, M. Lieu, S. Maurogordato, L. Moscardini, M. Plionis, B. M. Poggianti, E. Pompei, T. Sadibekova, I. Valtchanov, and J. P. Willis, "The XXL Survey. XXV. Cosmological analysis of the C1 cluster number counts," **620**, A10, A10 (2018).

[265]F. Pacaud, M. Pierre, A. Refregier, A. Gueguen, J.-L. Starck, I. Valtchanov, A. M. Read, B. Altieri, L. Chiappetti, P. Gandhi, O. Garcet, E. Gosset, T. J. Ponman, and J. Surdej, "The XMM Large-Scale Structure survey: the X-ray pipeline and survey selection function," **372**, 578–590 (2006).

[266]A. Paggi, F. Massaro, R. D'Abrusco, H. A. Smith, N. Masetti, M. Giroletti, G. Tosti, and S. Funk, "Unveiling the Nature of the Unidentified Gamma-Ray Sources. IV. The Swift Catalog of Potential X-Ray Counterparts," **209**, 9, 9 (2013).

[267]V. S. Paliya, M. Koss, B. Trakhtenbrot, C. Ricci, K. Oh, M. Ajello, D. Stern, M. C. Powell, C. M. Urry, F. Harrison, I. Lamperti, R. Mushotzky, L. Marcotulli, J. Mejía-Restrepo, and D. Hartmann, "BAT AGN Spectroscopic Survey. XVI. General Physical Characteristics of BAT Blazars," **881**, 154, 154 (2019).

[268]G. Papamakarios and I. Murray, "Fast ϵ-free Inference of Simulation Models with Bayesian Conditional Density Estimation," arXiv e-prints, arXiv:1605.06376, arXiv:1605.06376 (2016).

[269]P. Parisi, N. Masetti, A. F. Rojas, E. Jiménez-Bailón, V. Chavushyan, E. Palazzi, L. Bassani, A. Bazzano, A. J. Bird, G. Galaz, D. Minniti, L. Morelli, and P. Ubertini, "Accurate classification of 75 counterparts of objects detected in the 54-month Palermo Swift/BAT hard X-ray catalogue," **561**, A67, A67 (2014).

[270]H. A. Peña-Herazo, R. A. Amaya-Almazán, F. Massaro, R. de Menezes, E. J. Marchesini, V. Chavushyan, A. Paggi, M. Landoni, N. Masetti, F. Ricci, R. D'Abrusco, C. C. Cheung, F. La Franca, H. A. Smith, D. Milisavljevic, E. Jiménez-Bailón, V. M. Patiño-Álvarez, and G. Tosti, "Optical spectroscopic observations of low-energy counterparts of Fermi-LAT γ-ray sources," **643**, A103, A103 (2020).

[271]H. A. Peña-Herazo, E. J. Marchesini, N. Álvarez Crespo, F. Ricci, F. Massaro, V. Chavushyan, M. Landoni, J. Strader, L. Chomiuk, C. C. Cheung, N. Masetti, E. Jiménez-Bailón, R. D'Abrusco, A. Paggi, D. Milisavljevic, F. La Franca, H. A. Smith, and G. Tosti, "Optical spectroscopic observations

of gamma-ray blazar candidates. VII. Follow-up campaign in the southern hemisphere," **362**, 228, 228 (2017).

[272]H. A. Peña-Herazo, F. Massaro, V. Chavushyan, E. J. Marchesini, A. Paggi, M. Landoni, N. Masetti, F. Ricci, R. D'Abrusco, D. Milisavljevic, E. Jiménez-Bailón, F. La Franca, H. A. Smith, and G. Tosti, "Optical spectroscopic observations of gamma-ray blazar candidates. IX. Optical archival spectra and further observations from SOAR and OAGH," **364**, 85, 85 (2019).

[273]H. A. Peña-Herazo, F. Massaro, V. Chavushyan, N. Masetti, A. Paggi, and A. Capetti, "Turin-SyCAT: A multifrequency catalog of Seyfert galaxies," **659**, A32, A32 (2022).

[274]J. A. Peacock, L. Miller, and M. S. Longair, "The statistics of radio emission from quasars.," **218**, 265–278 (1986).

[275]J. A. Peacock, *Cosmological Physics* (1999).

[276]F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine Learning in Python," Journal of Machine Learning Research **12**, 2825–2830 (2011).

[277]P. J. E. Peebles, *Principles of Physical Cosmology* (1993).

[278]R. Penrose and R. M. Floyd, "Extraction of Rotational Energy from a Black Hole," Nature Physical Science **229**, 177–179 (1971).

[279]R. A. Perley, A. G. Willis, and J. S. Scott, "The structure of the radio jets in 3C 449.," **281**, 437–442 (1979).

[280]M. Pierre, F. Pacaud, C. Adami, S. Alis, B. Altieri, N. Baran, C. Benoist, M. Birkinshaw, A. Bongiorno, M. N. Bremer, M. Brusa, A. Butler, P. Ciliegi, L. Chiappetti, N. Clerc, P. S. Corasaniti, J. Coupon, C. De Breuck, J. Democles, S. Desai, J. Delhaize, J. Devriendt, Y. Dubois, D. Eckert, A. Elyiv, S. Ettori, A. Evrard, L. Faccioli, A. Farahi, C. Ferrari, F. Finet, S. Fotopoulou, N. Fourmanoit, P. Gandhi, F. Gastaldello, R. Gastaud, I. Georgantopoulos, P. Giles, L. Guennou, V. Guglielmo, C. Horellou, K. Husband, M. Huynh, A. Iovino, M. Kilbinger, E. Koulouridis, S. Lavoie, A. M. C. Le Brun, J. P. Le Fevre, C. Lidman, M. Lieu, C. A. Lin, A. Mantz, B. J. Maughan, S. Maurogordato, I. G. McCarthy, S. McGee, J. B. Melin, O. Melnyk, F. Menanteau, M. Novak, S. Paltani, M. Plionis, B. M. Poggianti, D. Pomarede, E. Pompei, T. J. Ponman, M. E. Ramos-Ceja, P. Ranalli, D. Rapetti, S. Raychaudury, T. H. Reiprich, H. Rottgering, E. Rozo, E. Rykoff, T. Sadibekova, J. Santos, J. L. Sauvageot, C. Schimd, M. Sereno, G. P. Smith, V. Smolčić, S. Snowden, D. Spergel, S. Stanford, J. Surdej, P. Valageas, A. Valotti, I. Valtchanov, C. Vignali, J. Willis, and F. Ziparo, "The XXL Survey. I. Scientific motivations - XMM-Newton observing plan - Follow-up observations and simulation programme," **592**, A1, A1 (2016).

[281]M. Pierre, A. Valotti, L. Faccioli, N. Clerc, R. Gastaud, E. Koulouridis, and F. Pacaud, "The cosmological analysis of X-ray cluster surveys. III. 4D X-ray observable diagrams," **607**, A123, A123 (2017).

[282]R. Piffaretti, M. Arnaud, G. W. Pratt, E. Pointecouteau, and J. B. Melin, "The MCXC: a meta-catalogue of x-ray detected clusters of galaxies," **534**, A109, A109 (2011).

[283]J. D. H. Pilkington and J. F. Scott, "A survey of radio sources between declinations 20° and 40°," **69**, 183 (1965).

[284]Planck Collaboration, P. A. R. Ade, N. Aghanim, C. Armitage-Caplan, M. Arnaud, M. Ashdown, F. Atrio-Barandela, J. Aumont, H. Aussel, C. Baccigalupi, A. J. Banday, R. B. Barreiro, R. Barrena, M. Bartelmann, J. G. Bartlett, E. Battaner, K. Benabed, A. Benoît, A. Benoit-Lévy, J. P. Bernard, M. Bersanelli, P. Bielewicz, I. Bikmaev, J. Bobin, J. J. Bock, H. Böhringer, A. Bonaldi, J. R. Bond, J. Borrill, F. R. Bouchet, M. Bridges, M. Bucher, R. Burenin, C. Burigana, R. C. Butler, J. F. Cardoso, P. Carvalho, A. Catalano, A. Challinor, A. Chamballu, R. R. Chary, X. Chen, H. C. Chiang, L. Y. Chiang, G. Chon, P. R. Christensen, E. Churazov, S. Church, D. L. Clements, S. Colombi, L. P. L. Colombo, B. Comis, F. Couchot, A. Coulais, B. P. Crill, A. Curto, F. Cuttaia, A. Da Silva, H. Dahle, L. Danese, R. D. Davies, R. J. Davis, P. de Bernardis, A. de Rosa, G. de Zotti, J. Delabrouille, J. M. Delouis, J. Démoclès, F. X. Désert, C. Dickinson, J. M. Diego, K. Dolag, H. Dole, S. Donzelli, O. Doré, M. Douspis, X. Dupac, G. Efstathiou, P. R. M. Eisenhardt, T. A. Enßlin, H. K. Eriksen, F. Feroz, F. Finelli, I. Flores-Cacho, O. Forni, M. Frailis, E. Franceschi, S. Fromenteau, S. Galeotta, K. Ganga, R. T. Génova-Santos, M. Giard, G. Giardino, M. Gilfanov, Y. Giraud-Héraud, J. González-Nuevo, K. M. Górski, K. J. B. Grainge, S. Gratton, A. Gregorio, E. Groeneboom N., A. Gruppuso, F. K. Hansen, D. Hanson, D. Harrison, A. Hempel, S. Henrot-Versillé, C. Hernández-Monteagudo, D. Herranz, S. R. Hildebrandt, E. Hivon, M. Hobson, W. A. Holmes, A. Hornstrup, W. Hovest, K. M. Huffenberger, G. Hurier, N. Hurley-Walker, A. H. Jaffe, T. R. Jaffe, W. C. Jones, M. Juvela, E. Keihänen, R. Keskitalo, I. Khamitov, T. S. Kisner, R. Kneissl, J. Knoche, L. Knox, M. Kunz, H. Kurki-Suonio, G. Lagache, A. Lähteenmäki, J. M. Lamarre, A. Lasenby, R. J. Laureijs, C. R. Lawrence, J. P. Leahy, R. Leonardi, J. León-Tavares, J. Lesgourgues, C. Li, A. Liddle, M. Liguori, P. B. Lilje, M. Linden-Vørnle, M. López-Caniego, P. M. Lubin, J. F. Macías-Pérez, C. J. MacTavish, B. Maffei, D. Maino, N. Mandolesi, M. Maris, D. J. Marshall, P. G. Martin, E. Martínez-González, S. Masi, M. Massardi, S. Matarrese, F. Matthai, P. Mazzotta, S. Mei, P. R. Meinhold, A. Melchiorri, J. B. Melin, L. Mendes, A. Mennella, M. Migliaccio, K. Mikkelsen, S. Mitra, M. A. Miville-Deschênes, A. Moneti, L. Montier, G. Morgante, D. Mortlock, D. Munshi, J. A. Murphy, P. Naselsky, F. Nati, P. Natoli, N. P. H. Nesvadba, C. B. Netterfield, H. U. Nørgaard-Nielsen, F. Noviello, D. Novikov, I. Novikov, I. J. O'Dwyer, M. Olamaie, S. Osborne, C. A. Oxborrow, F. Paci, L. Pagano, F. Pajot, D. Paoletti, F. Pasian, G. Patanchon, T. J. Pearson, O. Perdereau, L. Perotto, Y. C. Perrott, F. Perrotta, F. Piacentini, M. Piat, E. Pierpaoli, D. Pietrobon, S. Plaszczynski, E. Pointecouteau, G. Polenta, N. Ponthieu, L. Popa, T. Poutanen, G. W. Pratt, G. Prézeau, S. Prunet, J. L. Puget, J. P. Rachen, W. T. Reach, R. Rebolo, M. Reinecke, M. Remazeilles, C. Renault, S. Ricciardi, T. Riller, I. Ristorcelli, G. Rocha, C. Rosset, G. Roudier, M. Rowan-Robinson, J. A. Rubiño-Martín, C. Rumsey, B. Rusholme, M. Sandri, D. Santos, R. D. E. Saunders, G. Savini, M. P. Schammel, D. Scott, M. D.

Seiffert, E. P. S. Shellard, T. W. Shimwell, L. D. Spencer, S. A. Stanford, J. L. Starck, V. Stolyarov, R. Stompor, R. Sudiwala, R. Sunyaev, F. Sureau, D. Sutton, A. S. Suur-Uski, J. F. Sygnet, J. A. Tauber, D. Tavagnacco, L. Terenzi, L. Toffolatti, M. Tomasi, M. Tristram, M. Tucci, J. Tuovinen, M. Türler, G. Umana, L. Valenziano, J. Valiviita, B. Van Tent, L. Vibert, P. Vielva, F. Villa, N. Vittorio, L. A. Wade, B. D. Wandelt, M. White, S. D. M. White, D. Yvon, A. Zacchei, and A. Zonca, "Planck 2013 results. XXIX. The Planck catalogue of Sunyaev-Zeldovich sources," **571**, A29, A29 (2014).

[285]Planck Collaboration, P. A. R. Ade, N. Aghanim, M. Arnaud, M. Ashdown, J. Aumont, C. Baccigalupi, A. J. Banday, R. B. Barreiro, R. Barrena, J. G. Bartlett, N. Bartolo, E. Battaner, R. Battye, K. Benabed, A. Benoît, A. Benoit-Lévy, J. P. Bernard, M. Bersanelli, P. Bielewicz, I. Bikmaev, H. Böhringer, A. Bonaldi, L. Bonavera, J. R. Bond, J. Borrill, F. R. Bouchet, M. Bucher, R. Burenin, C. Burigana, R. C. Butler, E. Calabrese, J. F. Cardoso, P. Carvalho, A. Catalano, A. Challinor, A. Chamballu, R. R. Chary, H. C. Chiang, G. Chon, P. R. Christensen, D. L. Clements, S. Colombi, L. P. L. Colombo, C. Combet, B. Comis, F. Couchot, A. Coulais, B. P. Crill, A. Curto, F. Cuttaia, H. Dahle, L. Danese, R. D. Davies, R. J. Davis, P. de Bernardis, A. de Rosa, G. de Zotti, J. Delabrouille, F. X. Désert, C. Dickinson, J. M. Diego, K. Dolag, H. Dole, S. Donzelli, O. Doré, M. Douspis, A. Ducout, X. Dupac, G. Efstathiou, P. R. M. Eisenhardt, F. Elsner, T. A. Enßlin, H. K. Eriksen, E. Falgarone, J. Fergusson, F. Feroz, A. Ferragamo, F. Finelli, O. Forni, M. Frailis, A. A. Fraisse, E. Franceschi, A. Frejsel, S. Galeotta, S. Galli, K. Ganga, R. T. Génova-Santos, M. Giard, Y. Giraud-Héraud, E. Gjerløw, J. González-Nuevo, K. M. Górski, K. J. B. Grainge, S. Gratton, A. Gregorio, A. Gruppuso, J. E. Gudmundsson, F. K. Hansen, D. Hanson, D. L. Harrison, A. Hempel, S. Henrot-Versillé, C. Hernández-Monteagudo, D. Herranz, S. R. Hildebrandt, E. Hivon, M. Hobson, W. A. Holmes, A. Hornstrup, W. Hovest, K. M. Huffenberger, G. Hurier, A. H. Jaffe, T. R. Jaffe, T. Jin, W. C. Jones, M. Juvela, E. Keihänen, R. Keskitalo, I. Khamitov, T. S. Kisner, R. Kneissl, J. Knoche, M. Kunz, H. Kurki-Suonio, G. Lagache, J. M. Lamarre, A. Lasenby, M. Lattanzi, C. R. Lawrence, R. Leonardi, J. Lesgourgues, F. Levrier, M. Liguori, P. B. Lilje, M. Linden-Vørnle, M. López-Caniego, P. M. Lubin, J. F. Macías-Pérez, G. Maggio, D. Maino, D. S. Y. Mak, N. Mandolesi, A. Mangilli, P. G. Martin, E. Martínez-González, S. Masi, S. Matarrese, P. Mazzotta, P. McGehee, S. Mei, A. Melchiorri, J. B. Melin, L. Mendes, A. Mennella, M. Migliaccio, S. Mitra, M. A. Miville-Deschênes, A. Moneti, L. Montier, G. Morgante, D. Mortlock, A. Moss, D. Munshi, J. A. Murphy, P. Naselsky, A. Nastasi, F. Nati, P. Natoli, C. B. Netterfield, H. U. Nørgaard-Nielsen, F. Noviello, D. Novikov, I. Novikov, M. Olamaie, C. A. Oxborrow, F. Paci, L. Pagano, F. Pajot, D. Paoletti, F. Pasian, G. Patanchon, T. J. Pearson, O. Perdereau, L. Perotto, Y. C. Perrott, F. Perrotta, V. Pettorino, F. Piacentini, M. Piat, E. Pierpaoli, D. Pietrobon, S. Plaszczynski, E. Pointecouteau, G. Polenta, G. W. Pratt, G. Prézeau, S. Prunet, J. L. Puget, J. P. Rachen, W. T. Reach, R. Rebolo, M. Reinecke, M. Remazeilles, C. Renault, A. Renzi, I. Ristorcelli, G. Rocha, C. Rosset, M. Rossetti, G. Roudier, E. Rozo, J. A. Rubiño-Martín, C. Rumsey, B. Rusholme, E. S. Rykoff, M. Sandri, D. Santos, R. D. E. Saunders, M. Savelainen, G. Savini, M. P. Schammel, D. Scott, M. D. Seiffert, E. P. S. Shellard, T. W. Shimwell, L. D. Spencer, S. A. Stan-

ford, D. Stern, V. Stolyarov, R. Stompor, A. Streblyanska, R. Sudiwala, R. Sunyaev, D. Sutton, A. S. Suur-Uski, J. F. Sygnet, J. A. Tauber, L. Terenzi, L. Toffolatti, M. Tomasi, D. Tramonte, M. Tristram, M. Tucci, J. Tuovinen, G. Umana, L. Valenziano, J. Valiviita, B. Van Tent, P. Vielva, F. Villa, L. A. Wade, B. D. Wandelt, I. K. Wehus, S. D. M. White, E. L. Wright, D. Yvon, A. Zacchei, and A. Zonca, "Planck 2015 results. XXVII. The second Planck catalogue of Sunyaev-Zeldovich sources," **594**, A27, A27 (2016).

[286]Planck Collaboration, N. Aghanim, Y. Akrami, F. Arroja, M. Ashdown, J. Aumont, C. Baccigalupi, M. Ballardini, A. J. Banday, R. B. Barreiro, N. Bartolo, S. Basak, R. Battye, K. Benabed, J. P. Bernard, M. Bersanelli, P. Bielewicz, J. J. Bock, J. R. Bond, J. Borrill, F. R. Bouchet, F. Boulanger, M. Bucher, C. Burigana, R. C. Butler, E. Calabrese, J. F. Cardoso, J. Carron, B. Casaponsa, A. Challinor, H. C. Chiang, L. P. L. Colombo, C. Combet, D. Contreras, B. P. Crill, F. Cuttaia, P. de Bernardis, G. de Zotti, J. Delabrouille, J. M. Delouis, F. X. Désert, E. Di Valentino, C. Dickinson, J. M. Diego, S. Donzelli, O. Doré, M. Douspis, A. Ducout, X. Dupac, G. Efstathiou, F. Elsner, T. A. Enßlin, H. K. Eriksen, E. Falgarone, Y. Fantaye, J. Fergusson, R. Fernandez-Cobos, F. Finelli, F. Forastieri, M. Frailis, E. Franceschi, A. Frolov, S. Galeotta, S. Galli, K. Ganga, R. T. Génova-Santos, M. Gerbino, T. Ghosh, J. González-Nuevo, K. M. Górski, S. Gratton, A. Gruppuso, J. E. Gudmundsson, J. Hamann, W. Handley, F. K. Hansen, G. Helou, D. Herranz, S. R. Hildebrandt, E. Hivon, Z. Huang, A. H. Jaffe, W. C. Jones, A. Karakci, E. Keihänen, R. Keskitalo, K. Kiiveri, J. Kim, T. S. Kisner, L. Knox, N. Krachmalnicoff, M. Kunz, H. Kurki-Suonio, G. Lagache, J. M. Lamarre, M. Langer, A. Lasenby, M. Lattanzi, C. R. Lawrence, M. Le Jeune, J. P. Leahy, J. Lesgourgues, F. Levrier, A. Lewis, M. Liguori, P. B. Lilje, M. Lilley, V. Lindholm, M. López-Caniego, P. M. Lubin, Y. Z. Ma, J. F. Macías-Pérez, G. Maggio, D. Maino, N. Mandolesi, A. Mangilli, A. Marcos-Caballero, M. Maris, P. G. Martin, M. Martinelli, E. Martínez-González, S. Matarrese, N. Mauri, J. D. McEwen, P. D. Meerburg, P. R. Meinhold, A. Melchiorri, A. Mennella, M. Migliaccio, M. Millea, S. Mitra, M. A. Miville-Deschênes, D. Molinari, A. Moneti, L. Montier, G. Morgante, A. Moss, S. Mottet, M. Münchmeyer, P. Natoli, H. U. Nørgaard-Nielsen, C. A. Oxborrow, L. Pagano, D. Paoletti, B. Partridge, G. Patanchon, T. J. Pearson, M. Peel, H. V. Peiris, F. Perrotta, V. Pettorino, F. Piacentini, L. Polastri, G. Polenta, J. L. Puget, J. P. Rachen, M. Reinecke, M. Remazeilles, C. Renault, A. Renzi, G. Rocha, C. Rosset, G. Roudier, J. A. Rubiño-Martín, B. Ruiz-Granados, L. Salvati, M. Sandri, M. Savelainen, D. Scott, E. P. S. Shellard, M. Shiraishi, C. Sirignano, G. Sirri, L. D. Spencer, R. Sunyaev, A. S. Suur-Uski, J. A. Tauber, D. Tavagnacco, M. Tenti, L. Terenzi, L. Toffolatti, M. Tomasi, T. Trombetti, J. Valiviita, B. Van Tent, L. Vibert, P. Vielva, F. Villa, N. Vittorio, B. D. Wandelt, I. K. Wehus, M. White, S. D. M. White, A. Zacchei, and A. Zonca, "Planck 2018 results. I. Overview and the cosmological legacy of Planck," **641**, A1, A1 (2020).

[287]T. Plšek, N. Werner, M. Topinka, and A. Simionescu, "CAvity DEtection Tool (CADET): Pipeline for detection of X-ray cavities in hot galactic and cluster atmospheres," 10.1093/mnras/stad3371 (2023).

[288]P. Predehl, R. Andritschke, V. Arefiev, V. Babyshkin, O. Batanov, W. Becker, H. Böhringer, A. Bogomolov, T. Boller, K. Borm, W. Bornemann, H. Bräuninger,

M. Brüggen, H. Brunner, M. Brusa, E. Bulbul, M. Buntov, V. Burwitz, W. Burkert, N. Clerc, E. Churazov, D. Coutinho, T. Dauser, K. Dennerl, V. Doroshenko, J. Eder, V. Emberger, T. Eraerds, A. Finoguenov, M. Freyberg, P. Friedrich, S. Friedrich, M. Fürmetz, A. Georgakakis, M. Gilfanov, S. Granato, C. Grossberger, A. Gueguen, P. Gureev, F. Haberl, O. Hälker, G. Hartner, G. Hasinger, H. Huber, L. Ji, A. v. Kienlin, W. Kink, F. Korotkov, I. Kreykenbohm, G. Lamer, I. Lomakin, I. Lapshov, T. Liu, C. Maitra, N. Meidinger, B. Menz, A. Merloni, T. Mernik, B. Mican, J. Mohr, S. Müller, K. Nandra, V. Nazarov, F. Pacaud, M. Pavlinsky, E. Perinati, E. Pfeffermann, D. Pietschner, M. E. Ramos-Ceja, A. Rau, J. Reiffers, T. H. Reiprich, J. Robrade, M. Salvato, J. Sanders, A. Santangelo, M. Sasaki, H. Scheuerle, C. Schmid, J. Schmitt, A. Schwope, A. Shirshakov, M. Steinmetz, I. Stewart, L. Strüder, R. Sunyaev, C. Tenzer, L. Tiedemann, J. Trümper, V. Voron, P. Weber, J. Wilms, and V. Yaroshenko, "The eROSITA X-ray telescope on SRG," **647**, A1, A1 (2021).

[289]W. H. Press and P. Schechter, "Formation of Galaxies and Clusters of Galaxies by Self-Similar Gravitational Condensation," **187**, 425–438 (1974).

[290]G. C. Privon, C. P. O'Dea, S. A. Baum, D. J. Axon, P. Kharb, C. L. Buchanan, W. Sparks, and M. Chiaberge, "WFPC2 LRF Imaging of Emission-Line Nebulae in 3CR Radio Galaxies," **175**, 423–461 (2008).

[291]L. Qiu, N. R. Napolitano, S. Borgani, F. Zhong, X. Li, M. Radovich, W. Lin, K. Dolag, C. Tortora, Y. Wang, R.-S. Remus, and G. Longo, "Cosmology with Galaxy Cluster Properties using Machine Learning," arXiv e-prints, arXiv:2304.09142, arXiv:2304.09142 (2023).

[292]J. R. Quinlan, "Induction of decision trees," Machine learning **1**, 81–106 (1986).

[293]G. D. Racca, R. Laureijs, L. Stagnaro, J.-C. Salvignol, J. Lorenzo Alvarez, G. Saavedra Criado, L. Gaspar Venancio, A. Short, P. Strada, T. Bönke, C. Colombo, A. Calvi, E. Maiorano, O. Piersanti, S. Prezelus, P. Rosato, J. Pinel, H. Rozemeijer, V. Lesna, P. Musi, M. Sias, A. Anselmi, V. Cazaubiel, L. Vaillon, Y. Mellier, J. Amiaux, M. Berthé, M. Sauvage, R. Azzollini, M. Cropper, S. Pottinger, K. Jahnke, A. Ealet, T. Maciaszek, F. Pasian, A. Zacchei, R. Scaramella, J. Hoar, R. Kohley, R. Vavrek, A. Rudolph, and M. Schmidt, "The Euclid mission design," in Space telescopes and instrumentation 2016: optical, infrared, and millimeter wave, Vol. 9904, (July 2016), 99040O.

[294]J. N. Reeves, M. J. L. Turner, K. A. Pounds, P. T. O'Brien, T. Boller, P. Ferrando, E. Kendziorra, and S. Vercellone, "XMM-Newton observation of an unusual iron line in the quasar Markarian 205," **365**, L134–L139 (2001).

[295]M. Revnivtsev, S. Molkov, and S. Sazonov, "Map of the Galaxy in the 6.7-keV emission line," **373**, L11–L15 (2006).

[296]M. Revnivtsev, S. Sazonov, R. Krivonos, H. Ritter, and R. Sunyaev, "Properties of the Galactic population of cataclysmic variables in hard X-rays," **489**, 1121–1127 (2008).

[297]M. Reza, Y. Zhang, B. Nord, J. Poh, A. Ciprijanovic, and L. Strigari, "Estimating Cosmological Constraints from Galaxy Cluster Abundance using Simulation-Based Inference," in Machine learning for astrophysics (July 2022), p. 20.

[298]J. Ridl, N. Clerc, T. Sadibekova, L. Faccioli, F. Pacaud, J. Greiner, T. Krühler, A. Rau, M. Salvato, M.-L. Menzel, H. Steinle, P. Wiseman, K. Nandra, and J. Sanders, "Cosmology with XMM galaxy clusters: the X-CLASS/GROND catalogue and photometric redshifts," **468**, 662–684 (2017).

[299]H. Ritter and U. Kolb, "Catalogue of cataclysmic binaries, low-mass X-ray binaries and related objects (Seventh edition)," **404**, 301–303 (2003).

[300]A. Robertson, R. Massey, and V. Eke, "What does the Bullet Cluster tell us about self-interacting dark matter?" **465**, 569–587 (2017).

[301]J. M. Rodriguez Espinosa, R. J. Rudy, and B. Jones, "Star Formation in Seyfert Galaxies," **312**, 555 (1987).

[302]A. F. Rojas, N. Masetti, D. Minniti, E. Jiménez-Bailón, V. Chavushyan, G. Hau, V. A. McBride, L. Bassani, A. Bazzano, A. J. Bird, G. Galaz, I. Gavignaud, R. Landi, A. Malizia, L. Morelli, E. Palazzi, V. Patiño-Álvarez, J. B. Stephen, and P. Ubertini, "The nature of 50 Palermo Swift-BAT hard X-ray objects through optical spectroscopy," **602**, A124, A124 (2017).

[303]O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," arXiv e-prints, arXiv:1505.04597, arXiv:1505.04597 (2015).

[304]F. Rosenblatt, "The perceptron: a probabilistic model for information storage and organization in the brain.," Psychological review **65 6**, 386–408 (1958).

[305]R. Rothschild, E. Boldt, S. Holt, P. Serlemitsos, G. Garmire, P. Agrawal, G. Riegler, S. Bowyer, and M. Lampton, "The cosmic X-ray experiment aboard HEAO-1.," Space Science Instrumentation **4**, 269–301 (1979).

[306]S. Ruder, "An overview of gradient descent optimization algorithms," arXiv e-prints, arXiv:1609.04747, arXiv:1609.04747 (2016).

[307]E. S. Rykoff, E. Rozo, M. T. Busha, C. E. Cunha, A. Finoguenov, A. Evrard, J. Hao, B. P. Koester, A. Leauthaud, B. Nord, M. Pierre, R. Reddick, T. Sadibekova, E. S. Sheldon, and R. H. Wechsler, "redMaPPer. I. Algorithm and SDSS DR8 Catalog," **785**, 104, 104 (2014).

[308]A. J. R. Sanderson, E. O'Sullivan, T. J. Ponman, A. H. Gonzalez, S. Sivanandam, A. I. Zabludoff, and D. Zaritsky, "The baryon budget on the galaxy group/cluster boundary," **429**, 3288–3304 (2013).

[309]S. Sazonov, M. Revnivtsev, R. Krivonos, E. Churazov, and R. Sunyaev, "Hard X-ray luminosity function and absorption distribution of nearby AGN: INTEGRAL all-sky survey," **462**, 57–66 (2007).

[310]K. Schawinski, C. Zhang, H. Zhang, L. Fowler, and G. K. Santhanam, "Generative adversarial networks recover features in astrophysical images of galaxies beyond the deconvolution limit," **467**, L110–L114 (2017).

[311]K. Schmidt, F. Geyer, S. Fröse, P. S. Blomenkamp, M. Brüggen, F. de Gasperin, D. Elsässer, and W. Rhode, "Deep learning-based imaging in radio interferometry," **664**, A134, A134 (2022).

[312]M. Schmidt, "3C 273 : A Star-Like Object with Large Red-Shift," **197**, 1040 (1963).

[313]M. Schmidt, "Quasistellar Objects," **7**, 527 (1969).

[314]D. P. Schneider, G. T. Richards, P. B. Hall, M. A. Strauss, S. F. Anderson, T. A. Boroson, N. P. Ross, Y. Shen, W. N. Brandt, X. Fan, N. Inada, S. Jester, G. R. Knapp, C. M. Krawczyk, A. R. Thakar, D. E. Vanden Berk, W. Voges, B. Yanny, D. G. York, N. A. Bahcall, D. Bizyaev, M. R. Blanton, H. Brewington, J. Brinkmann, D. Eisenstein, J. A. Frieman, M. Fukugita, J. Gray, J. E. Gunn, P. Hibon, Ž. Ivezić, S. M. Kent, R. G. Kron, M. G. Lee, R. H. Lupton, E. Malanushenko, V. Malanushenko, D. Oravetz, K. Pan, J. R. Pier, I. Price Ted N., D. H. Saxe, D. J. Schlegel, A. Simmons, S. A. Snedden, M. U. SubbaRao, A. S. Szalay, and D. H. Weinberg, "The Sloan Digital Sky Survey Quasar Catalog. V. Seventh Data Release," **139**, 2360, 2360 (2010).

[315]L. Searle, W. L. W. Sargent, and W. G. Bagnuolo, "The History of Star Formation and the Colors of Late-Type Galaxies," **179**, 427–438 (1973).

[316]A. Segreto, G. Cusumano, C. Ferrigno, V. La Parola, V. Mangano, T. Mineo, and P. Romano, "The Palermo Swift-BAT hard X-ray catalogue. I. Methodology," **510**, A47, A47 (2010).

[317]C. K. Seyfert, "Nuclear Emission in Spiral Nebulae.," **97**, 28 (1943).

[318]K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," arXiv e-prints, arXiv:1409.1556, arXiv:1409.1556 (2014).

[319]R. Singh, G. van de Ven, K. Jahnke, M. Lyubenova, J. Falcón-Barroso, J. Alves, R. Cid Fernandes, L. Galbany, R. García-Benito, B. Husemann, R. C. Kennicutt, R. A. Marino, I. Márquez, J. Masegosa, D. Mast, A. Pasquali, S. F. Sánchez, J. Walcher, V. Wild, L. Wisotzki, and B. Ziegler, "The nature of LINER galaxies:. Ubiquitous hot old stars and rare accreting black holes," **558**, A43, A43 (2013).

[320]M. F. Skrutskie, R. M. Cutri, R. Stiening, M. D. Weinberg, S. Schneider, J. M. Carpenter, C. Beichman, R. Capps, T. Chester, J. Elias, J. Huchra, J. Liebert, C. Lonsdale, D. G. Monet, S. Price, P. Seitzer, T. Jarrett, J. D. Kirkpatrick, J. E. Gizis, E. Howard, T. Evans, J. Fowler, L. Fullmer, R. Hurt, R. Light, E. L. Kopan, K. A. Marsh, H. L. McCallon, R. Tam, S. Van Dyk, and S. Wheelock, "The Two Micron All Sky Survey (2MASS)," **131**, 1163–1183 (2006).

[321]K. L. Smith, R. F. Mushotzky, M. Koss, B. Trakhtenbrot, C. Ricci, O. I. Wong, F. E. Bauer, F. Ricci, S. Vogel, D. Stern, M. C. Powell, C. M. Urry, F. Harrison, J. Mejia-Restrepo, K. Oh, J. Baek, and A. Chung, "BAT AGN spectroscopic survey - XV: the high frequency radio cores of ultra-hard X-ray selected AGN," **492**, 4216–4234 (2020).

[322]P. S. Smith, T. J. Balonek, P. A. Heckert, and R. Elston, "The Optical and Near-Infrared Polarization Properties of the OVV Quasar 3C 345," **305**, 484 (1986).

[323]H. Spinrad, S. Djorgovski, J. Marr, and L. Aguilar, "A third update of the status of the 3 CR sources : further new redshifts and new identifications of distant galaxies.," **97**, 932–961 (1985).

[324]J.-L. Starck, F. D. Murtagh, and A. Bijaoui, *Image Processing and Data Analysis* (July 1998), p. 297.

[325]J.-L. Starck and M. Pierre, "Structure detection in low intensity X-ray images," **128**, 397–407 (1998).

[326]M. Stickel, P. Padovani, C. M. Urry, J. W. Fried, and H. Kuehr, "The Complete Sample of 1 Jansky BL Lacertae Objects. I. Summary Properties," **374**, 431 (1991).

[327]M. A. Strauss, J. P. Huchra, M. Davis, A. Yahil, K. B. Fisher, and J. Tonry, "A Redshift Survey of IRAS Galaxies. VII. The Infrared and Redshift Data for the 1.936 Jansky Sample," **83**, 29 (1992).

[328]P. A. Strittmatter, P. Hill, I. I. K. Pauliny-Toth, H. Steppe, and A. Witzel, "Radio observations of optically selected quasars," **88**, L12–L15 (1980).

[329]C. Stuardi, V. Missaglia, F. Massaro, F. Ricci, E. Liuzzo, A. Paggi, R. P. Kraft, G. R. Tremblay, S. A. Baum, C. P. O'Dea, B. J. Wilkes, J. Kuraszkiewicz, W. R. Forman, and D. E. Harris, "The 3CR Chandra Snapshot Survey: Extragalactic Radio Sources with Redshifts between 1 and 1.5," **235**, 32, 32 (2018).

[330]R. A. Sunyaev and Y. B. Zeldovich, "The Observations of Relic Radiation as a Test of the Nature of X-Ray Radiation from the Clusters of Galaxies," Comments on Astrophysics and Space Physics **4**, 173 (1972).

[331]C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the Inception Architecture for Computer Vision," ArXiv e-prints (2015).

[332]C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going Deeper with Convolutions," arXiv e-prints, arXiv:1409.4842, arXiv:1409.4842 (2014).

[333]G. B. Taylor and E. L. Wright, "Effects of a Hot Intergalactic Medium," **339**, 619 (1989).

[334]J. H. Taylor, R. N. Manchester, and A. G. Lyne, "Catalog of 558 Pulsars," **88**, 529 (1993).

[335]M. B. Taylor, "TOPCAT & STIL: Starlink Table/VOTable Processing Software," in Astronomical data analysis software and systems xiv, Vol. 347, edited by P. Shopbell, M. Britton, and R. Ebert, Astronomical Society of the Pacific Conference Series (Dec. 2005), p. 29.

[336]A. Tejero-Cantero, J. Boelts, M. Deistler, J.-M. Lueckmann, C. Durkan, P. J. Gonçalves, D. S. Greenberg, and J. H. Macke, "Sbi: a toolkit for simulation-based inference," Journal of Open Source Software **5**, 2505 (2020).

[337]J. Tinker, A. V. Kravtsov, A. Klypin, K. Abazajian, M. Warren, G. Yepes, S. Gottlöber, and D. E. Holz, "Toward a Halo Mass Function for Precision Cosmology: The Limits of Universality," **688**, 709–728 (2008).

[338]J. A. Tomsick, A. Bodaghee, S. Chaty, M. Clavel, F. M. Fornasini, J. Hare, R. Krivonos, F. Rahoui, and J. Rodriguez, "Chandra observations of high-energy x-ray sources discovered by integral," The Astrophysical Journal **889**, 53 (2020).

[339]J. Truemper, "ROSAT: A New Look at the X-ray Sky," **33**, 165 (1992).

[340]J. Tueller, R. F. Mushotzky, S. Barthelmy, J. K. Cannizzo, N. Gehrels, C. B. Markwardt, G. K. Skinner, and L. M. Winter, "Swift BAT Survey of AGNs," **681**, 113–127 (2008).

[341]P. Ubertini, F. Lebrun, G. Di Cocco, A. Bazzano, A. J. Bird, K. Broenstad, A. Goldwurm, G. La Rosa, C. Labanti, P. Laurent, I. F. Mirabel, E. M. Quadrini, B. Ramsey, V. Reglero, L. Sabau, B. Sacco, R. Staubert, L. Vigroux, M. C. Weisskopf, and A. A. Zdziarski, "IBIS: The Imager on-board INTEGRAL," **411**, L131–L139 (2003).

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

[343]A. Valinia and F. E. Marshall, "RXTE Measurement of the Diffuse X-Ray Emission from the Galactic Ridge: Implications for the Energetics of the Interstellar Medium," **505**, 134–147 (1998).

[344]A. Valotti, M. Pierre, A. Farahi, A. Evrard, L. Faccioli, J. L. Sauvageot, N. Clerc, and F. Pacaud, "The cosmological analysis of X-ray cluster surveys. IV. Testing ASpiX with template-based cosmological simulations," **614**, A72, A72 (2018).

[345]I. Valtchanov, M. Pierre, and R. Gastaud, "Comparison of source detection procedures for XMM-Newton images," **370**, 689–706 (2001).

[346]G. Van Rossum and F. L. Drake, *Python 3 reference manual* (CreateSpace, Scotts Valley, CA, 2009).

[347]G. Van Rossum and F. L. Drake, *Python 3 reference manual* (CreateSpace, Scotts Valley, CA, 2009).

[348]L. Vaona, S. Ciroi, F. Di Mille, V. Cracco, G. La Mura, and P. Rafanelli, "Spectral properties of the narrow-line region in Seyfert galaxies selected from the SDSS-DR7," **427**, 1266–1283 (2012).

[349]F. Villaescusa-Navarro, D. Anglés-Alcázar, S. Genel, D. N. Spergel, R. S. Somerville, R. Dave, A. Pillepich, L. Hernquist, D. Nelson, P. Torrey, D. Narayanan, Y. Li, O. Philcox, V. La Torre, A. Maria Delgado, S. Ho, S. Hassan, B. Burkhart, D. Wadekar, N. Battaglia, G. Contardo, and G. L. Bryan, "The CAMELS Project: Cosmology and Astrophysics with Machine-learning Simulations," **915**, 71, 71 (2021).

[350]J. Vink, "Supernova remnants: the X-ray perspective," **20**, 49, 49 (2012).

[351]W. Voges, "The rosat all-sky survey," Advances in Space Research **13**, 391–397 (1993).

[352]W. Voges, B. Aschenbach, T. Boller, H. Bräuninger, U. Briel, W. Burkert, K. Dennerl, J. Englhauser, R. Gruber, F. Haberl, G. Hartner, G. Hasinger, M. Kürster, E. Pfeffermann, W. Pietsch, P. Predehl, C. Rosso, J. H. M. M. Schmitt, J. Trümper, and H. U. Zimmermann, "The ROSAT all-sky survey bright source catalogue," **349**, 389–405 (1999).

[353]A. Vojtekova, M. Lieu, I. Valtchanov, B. Altieri, L. Old, Q. Chen, and F. Hroch, "Learning to denoise astronomical images with U-nets," **503**, 3204–3215 (2021).

[354]B. Warner, *Cataclysmic variable stars*, Vol. 28 (1995).

[355]D. W. Weedman, "Seyfert galaxies.," **15**, 69–95 (1977).

[356]D. W. Weedman, F. R. Feldman, V. A. Balzano, L. W. Ramsey, R. A. Sramek, and C. C. Wuu, "NGC 7714 - The prototype star-burst galactic nucleus," **248**, 105–112 (1981).

[357]M. C. Weisskopf, H. D. Tananbaum, L. P. Van Speybroeck, and S. L. O'Dell, "Chandra X-ray Observatory (CXO): overview," in X-ray optics, instruments, and missions iii, Vol. 4012, edited by J. E. Truemper and B. Aschenbach, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series (July 2000), pp. 2–16.

[358]Z. L. Wen and J. L. Han, "A catalog of 1.58 million clusters of galaxies identified from the DESI Legacy Imaging Surveys," arXiv e-prints, arXiv:2404.02002, arXiv:2404.02002 (2024).

[359]Z. L. Wen and J. L. Han, "A sample of 1959 massive galaxy clusters at high redshifts," **481**, 4158–4168 (2018).

[360]Z. L. Wen, J. L. Han, and F. Yang, "A catalogue of clusters of galaxies identified from all sky surveys of 2MASS, WISE, and SuperCOSMOS," **475**, 343–352 (2018).

[361]K. W. Willett, C. J. Lintott, S. P. Bamford, K. L. Masters, and S. et al., "Galaxy Zoo 2: detailed morphological classifications for 304 122 galaxies from the Sloan Digital Sky Survey," **435**, 2835–2860 (2013).

[362]A. S. Wilson and E. J. M. Colbert, "The Difference between Radio-loud and Radio-quiet Active Galaxies," **438**, 62 (1995).

[363]C. Winkler, T. J. L. Courvoisier, G. Di Cocco, N. Gehrels, A. Giménez, S. Grebenev, W. Hermsen, J. M. Mas-Hesse, F. Lebrun, N. Lund, G. G. C. Palumbo, J. Paul, J. P. Roques, H. Schnopper, V. Schönfelder, R. Sunyaev, B. Teegarden, P. Ubertini, G. Vedrenne, and A. J. Dean, "The INTEGRAL mission," **411**, L1–L6 (2003).

[364]A. Wolter, V. Beckmann, G. Ghisellini, F. Tavecchio, and L. Maraschi, "The Hard Synchrotron X–ray Spectrum of the TeV BL Lac 1ES 1426+428," in Extragalactic jets: theory and observation from radio to gamma ray, Vol. 386, edited by T. A. Rector and D. S. De Young, Astronomical Society of the Pacific Conference Series (June 2008), p. 302.

[365]E. L. Wright, P. R. M. Eisenhardt, A. K. Mainzer, M. E. Ressler, R. M. Cutri, T. Jarrett, J. D. Kirkpatrick, D. Padgett, R. S. McMillan, M. Skrutskie, S. A. Stanford, M. Cohen, R. G. Walker, J. C. Mather, D. Leisawitz, I. Gautier Thomas N., I. McLean, D. Benford, C. J. Lonsdale, A. Blain, B. Mendez, W. R. Irace, V. Duval, F. Liu, D. Royer, I. Heinrichsen, J. Howard, M. Shannon, M. Kendall, A. L. Walsh, M. Larsen, J. G. Cardon, S. Schick, M. Schwalm, M. Abid, B. Fabinsky, L. Naes, and C.-W. Tsai, "The Wide-field Infrared Survey Explorer (WISE): Mission Description and Initial On-orbit Performance," **140**, 1868–1881 (2010).

[366]W. Xu, M. E. Ramos-Ceja, F. Pacaud, T. H. Reiprich, and T. Erben, "Catalog of X-ray-selected extended galaxy clusters from the ROSAT All-Sky Survey (RXGCC)," **658**, A59, A59 (2022).

[367]Z. Xu, A. M. Dai, J. Kemp, and L. Metz, "Learning an Adaptive Learning Rate Schedule," arXiv e-prints, arXiv:1909.09712, arXiv:1909.09712 (2019).

[368]F. Yuan and R. Narayan, "On the Nature of X-Ray-Bright, Optically Normal Galaxies," **612**, 724–728 (2004).

[369]M. D Zeiler and R. Fergus, "Visualizing and Understanding Convolutional Networks," ArXiv e-prints (2013).

[370]D. de Andres, W. Cui, G. Yepes, M. De Petris, A. Ferragamo, F. De Luca, G. Aversano, and D. Rennehan, "The three hundred project: mapping the matter distribution in galaxy clusters via deep learning from multiview simulated observations," **528**, 1517–1530 (2024).