



**HAL**  
open science

# Circum-nuclear environment of AGN: molecular tori and outflows

Rémi Poitevineau

► **To cite this version:**

Rémi Poitevineau. Circum-nuclear environment of AGN: molecular tori and outflows. Astrophysics [astro-ph]. Université Paris sciences et lettres, 2023. English. NNT : 2023UPSLO009 . tel-04538485

**HAL Id: tel-04538485**

**<https://theses.hal.science/tel-04538485>**

Submitted on 9 Apr 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



**THÈSE DE DOCTORAT**  
**DE L'UNIVERSITÉ PSL**

Préparée à l'Observatoire de Paris

**Circum-nuclear environment of AGN: molecular tori and outflows**

Soutenue par

**Rémi Poitevineau**

Le 28/09/2023

École doctorale n°127

**Astronomie et  
Astrophysique d'Ile de  
France**

Spécialité

**Astrophysique**

Préparée au

LERMA (Laboratoire d'Études du  
Rayonnement et de la Matière en  
Astrophysique et Atmosphères)

Composition du jury :

Yann Clénet *Président du jury*  
Directeur de recherche,  
LESIA - Observatoire de Paris

Davor Krajnovic *Rapporteur*  
Tenured staff scientists,  
Leibniz-Institut für Astrophysik Potsdam

Roberto Saglia *Rapporteur*  
Professeur,  
Max-Planck Institut für extraterrestrische Physik

Delphine Porquet *Examinatrice*  
Directrice de recherche,  
Laboratoire d'Astrophysique de Marseille

Santiago García-Burillo *Examineur*  
Astronome,  
Observatorio Astronomico Nacional Observatorio  
de Madrid

Matthew Lehnert *Examineur*  
Directeur de recherche,  
Centre de Recherche en Astrophysique de Lyon

David Cornu *Examineur*  
Postdoc,  
LERMA - Observatoire de Paris

Françoise Combes-Bottaro *Directrice de thèse*  
Astronome,  
LERMA - Observatoire de Paris





---

## Résumé

La relation entre la masse du trou noir supermassif (SMBH) situé au centre des galaxies et la masse de leur bulbe ou leur dispersion de vitesse centrale est bien connue. Cela suggère une coévolution entre les SMBH et leurs galaxies hôtes. Les mécanismes détaillés d'alimentation et de rétroaction des Noyaux Actifs de Galaxies (AGN) ne sont pas encore bien compris, et pour les AGN de faible luminosité, les AGN occultés et les galaxies de type tardif, il est difficile de déterminer précisément les masses des trous noirs centraux (BH).

L'objectif de cette recherche de doctorat est d'étudier différents aspects des AGN et leur lien avec les SMBH dans différents environnements galactiques. L'étude présentée s'est concentrée sur deux projets de recherche distincts visant à contribuer à notre compréhension de l'alimentation des AGN, des mécanismes de rétroaction et de la détermination des masses des SMBH.

Le premier projet consistait à examiner la relation entre la masse des SMBH et les propriétés des galaxies radio-bruyantes, en mettant l'accent sur l'évolution en fonction du décalage vers le rouge. Un échantillon de 42 galaxies radio et d'AGN présentant des raies d'émission larges sur une plage de décalage vers le rouge de 0,3 à 4 a été constitué. En croisant les sources du relevé radio FIRST du Very Large Array (VLA) avec des galaxies confirmées spectroscopiquement provenant de relevés à grand champ, tels que SDSS, DES, WISE et GAMA, un ensemble de données complet a ainsi pu être obtenu. Les propriétés analysées comprenaient la masse stellaire, les taux de formation d'étoiles et les caractéristiques des trous noirs tels que la masse des SMBH, le rapport d'Eddington et la puissance du jet. Les résultats ont été comparés à des relations d'échelle existantes dans la littérature, fournissant des informations précieuses sur l'évolution des AGN radio-bruyants et leur lien avec les SMBH.

Le deuxième projet faisait partie du relevé Galaxy Activity, Torus and Outflow Survey (GATOS). Il s'est concentré sur l'étude des mécanismes d'alimentation et de rétroaction des AGN, en particulier dans les AGN de faible luminosité, les AGN occultés et les galaxies de type tardif. La détermination précise des masses des trous noirs centraux dans ces systèmes est difficile du fait de la présence de tores moléculaires sur la ligne de visée.

Cette étude a utilisé l'échantillon principal de GATOS et a utilisé des observations haute résolution de l'Atacama Large Millimeter/submillimeter Array (ALMA) pour étudier les régions circum-nucléaires de galaxies sélectionnées. En analysant l'émission CO(3-2) dans une région d'environ 100 parsecs autour du SMBH, l'étude réalisée visait à utiliser la haute résolution des données pour permettre des estimations de masse des trous noirs pouvant être plus précises que celles obtenues via les relations d'échelle traditionnelles. Une approche novatrice a été développée en utilisant des techniques d'apprentissage supervisé pour estimer les masses des trous noirs en se basant sur des diagrammes position-vitesse et des cartes de moments issues d'observations CO(3-2) d'ALMA. La méthode a été entraînée, validée et testée à l'aide de simulations numériques avec une large gamme de paramètres. Les résultats ont ensuite été appliqués à l'échantillon principal de GATOS, révélant des estimations cohérentes de la masse des trous noirs et fournissant des estimations d'erreur fiables. Ce travail pose les bases d'une approche automatisée pour l'estimation de la masse des trous noirs en utilisant l'apprentissage machine.

---

## Abstract

The relationship between the mass of the supermassive black hole (SMBH) at the center of galaxies and the mass of their bulge or central velocity dispersion is well established, indicating coevolution between SMBHs and their host galaxies. However, the detailed feeding and feedback mechanisms of Active Galactic Nuclei (AGN) remain poorly understood, particularly for low-luminosity AGN, obscured AGN, and late-type galaxies, which makes the precise determination of central black hole masses challenging.

This Ph.D. research aimed to investigate various aspects of AGN and their connection to SMBHs in different galactic environments. Two separate research projects were conducted to contribute to our understanding of AGN feeding, feedback mechanisms, and SMBH mass determination.

The first project focused on examining the relationship between SMBH mass and the properties of radio-loud galaxies, with a specific emphasis on redshift evolution. A sample of 42 radio galaxies and AGN with broad emission lines spanning a redshift range of 0.3 to 4 was compiled. A comprehensive dataset was obtained by cross-matching radio sources from the Very Large Array (VLA) FIRST survey with spectroscopically confirmed galaxies from wide-field surveys, including SDSS, DES, WISE, and GAMA. The analyzed properties included stellar mass, star formation rates, and black hole characteristics such as SMBH mass, Eddington ratio, and jet power. The findings were compared with existing scaling relations from the literature, providing valuable insights into the evolution of radio-loud AGN and their connection to SMBHs.

The second project was part of the Galaxy Activity, Torus, and Outflow Survey (GATOS) and aimed to study the feeding and feedback mechanisms of AGN, particularly in low-luminosity AGN, obscured AGN, and late-type galaxies. The precise determination of central black hole masses in these systems is challenging due to the presence of molecular tori along the line of sight. This study utilized the core sample of GATOS and employed high-resolution observations from the Atacama Large Millimeter/submillimeter Array (ALMA) to investigate the circum-nuclear regions of selected galaxies. By analyzing the CO(3-2) emission within approximately 100 parsecs around the SMBH, the study aimed to obtain more accurate estimates of black hole masses compared to traditional scaling relations. A novel approach was developed using supervised machine learning techniques to estimate black hole masses based on position-velocity diagrams and ALMA CO(3-2) observations. The method was trained, validated, and tested using numerical simulations with a wide range of parameters. The results were then applied to the core sample of GATOS, revealing consistent estimations of black hole masses and providing reliable error estimations. This work lay the foundation for an automated approach to black hole mass estimation using machine learning.

---

## Remerciements

Je tenais tout d’abord à remercier Françoise pour m’avoir donné l’opportunité de réaliser cette thèse. Je suis reconnaissant envers elle pour sa patience, de m’avoir laissé mener cette thèse dans la direction que je voulais et de m’avoir soutenu dans mes décisions.

J’aimerais aussi remercier Gianluca. Sans son investissement, une partie des travaux présentés ici n’aurait jamais abouti. Merci pour sa confiance et son soutien lors de ces dernières années.

Je souhaite ensuite particulièrement remercier David. Sans son expertise en machine learning mon travail aurait été de bien moindre qualité. Merci aussi pour sa bonne humeur, sa patience et sa disponibilité. David a toujours pris le temps pour répondre à mes questions, pour m’aider ou me relire à chaque fois que j’en avais besoin. Travailler de nouveau ensemble serait un immense plaisir.

Merci aussi à Anaëlle pour toutes ses relectures et les conseils qu’elle a pu me donner lors des trois dernières années. J’aimerais également remercier Barbara. Elle a amené avec elle sa bonne humeur et son énergie quand elle est arrivée à mon bureau. Travailler dans le même bureau était un réel plaisir et une bouffée d’air frais.

Je voudrais exprimer ma reconnaissance envers l’équipe administrative du LERMA. Particulièrement à Woihiba, Élise et Murielle qui ont toujours été bienveillantes et efficaces en toutes circonstances. Je remercie aussi mes collègues du LERMA: Romain, Philippe, Benoît, Aristide, Florent, Ian et Juliette. Merci pour la bonne humeur quotidienne et pour toutes nos discussions enrichissantes.

Je voudrais ensuite remercier mes proches, à commencer par mes parents. La pomme ne tombe jamais loin de l’arbre. Ce sont eux qui m’ont donné envie d’être scientifique. Je ne leur dis jamais assez à quel point je tiens à eux, merci de toujours me soutenir.

Un immense merci à mes amis Daniel, Gabriel, Julien V.d.M., Lorenzo, Victor, Sébastien, Julien H. et Adèle. Pour certains, nous sommes amis depuis notre première année à la fac et nous avons passé le plus clair de nos études supérieures ensemble. Sans notre entraide et nos séances de révisions autour de la table du 3ème étage de Condorcet, je n’aurais réalisé mon rêve d’enfance de devenir docteur en astrophysique. Durant ces dix dernières années, j’ai toujours pu compter sur eux. Vous êtes comme une seconde famille pour moi.

Enfin, j’aimerais remercier tout particulièrement Julien H. et Mélanie. Dans leur style respectif, leur soutien m’a été inestimable ces dernières années. Ils m’ont permis de garder le moral plus d’une fois et d’aller de l’avant. Notre amitié est quelque chose de précieux pour moi. Merci pour tout.

---

# CONTENTS

---

<b>List of Figures</b>	<b>xi</b>
<b>List of Tables</b>	<b>xiii</b>
<b>List of Acronyms</b>	<b>xv</b>
<b>1 Introduction to AGN physics</b>	<b>1</b>
1.1 Galaxies . . . . .	2
1.2 Active Galactic Nuclei . . . . .	4
1.2.1 The different types of AGN . . . . .	4
1.2.2 The unification model of AGN . . . . .	5
1.2.3 Global spectrum SED . . . . .	8
1.2.4 SMBH accretion . . . . .	10
1.2.5 Radio galaxies excitation . . . . .	12
1.2.6 Jets formation . . . . .	13
1.3 The Circumnuclear Region . . . . .	14
1.3.1 The dusty torus . . . . .	14
1.3.2 Molecular torus . . . . .	15
1.3.3 Polar biconical outflows . . . . .	16
1.4 AGN Feedback . . . . .	16
1.4.1 The need for feedback . . . . .	16
1.4.2 The two modes of AGN feedback . . . . .	19
1.4.3 Energy or momentum conservation . . . . .	21
1.5 Co-Evolution of Galaxies and SMBH . . . . .	22
1.6 Classical methods for estimating $M_{\text{BH}}$ . . . . .	24
1.6.1 Dynamical estimation . . . . .	24
1.6.2 $M_{\text{BH}}$ - $\sigma$ relation . . . . .	26
1.6.3 Reverberation mapping . . . . .	26
1.6.4 Single epoch method . . . . .	27
1.6.5 Fundamental Plane of BH Activity . . . . .	27

<b>2</b>	<b>SMBH and galaxy co-evolution in RLAGN at <math>z \sim 0.3 - 4</math></b>	<b>29</b>
2.1	Motivations . . . . .	30
2.2	RLAGN sample . . . . .	31
2.2.1	Dark Energy Survey . . . . .	31
2.2.2	Radio, optical, and spectroscopic selection . . . . .	31
2.2.3	IR selection: WISE . . . . .	32
2.2.4	Broad emission lines from SDSS . . . . .	33
2.2.5	RLAGN in GAMA DR3 . . . . .	34
2.2.6	Radio and IR properties . . . . .	35
2.3	BH, jet, accretion, and stellar properties . . . . .	36
2.3.1	BH masses . . . . .	36
2.3.2	Jet power . . . . .	36
2.3.3	Eddington ratio . . . . .	37
2.3.4	SED modeling . . . . .	38
2.3.5	SFR versus $M_\star$ . . . . .	40
2.4	Comparison sample . . . . .	42
2.5	Results . . . . .	44
2.5.1	$M_{\text{BH}} - M_\star$ relation . . . . .	44
2.5.2	Jet power, BH mass, and accretion . . . . .	47
2.5.3	RLAGN and their environments . . . . .	51
2.6	Discussion and conclusions . . . . .	52
<b>3</b>	<b>Introduction to Machine Learning</b>	<b>55</b>
3.1	The different types of ML algorithm . . . . .	56
3.1.1	Definitions . . . . .	56
3.2	Artificial Neuron . . . . .	58
3.2.1	The first Artificial Neurons . . . . .	58
3.2.2	Training a neuron . . . . .	59
3.2.3	The bias node . . . . .	60
3.3	Artificial Neural Network . . . . .	60
3.3.1	The perceptron algorithm . . . . .	60
3.3.2	Multi Layers perceptron . . . . .	61
3.3.3	Back propagation . . . . .	63
3.3.4	Overfitting . . . . .	64
3.4	Network optimization . . . . .	65
3.4.1	Input normalization . . . . .	65
3.4.2	Weight initialization . . . . .	65
3.4.3	Batch size . . . . .	66
3.4.4	Learning rate . . . . .	67
3.4.5	Momentum . . . . .	67
3.4.6	Dropout . . . . .	68
3.4.7	MC dropout . . . . .	68
3.5	Convolutional Neural Network . . . . .	68
3.5.1	Information in 2D images . . . . .	68
3.5.2	Convolution operation . . . . .	72
3.5.3	Convolutional layer . . . . .	73
3.5.4	Multiple convolutional layers . . . . .	73
3.5.5	Activation function . . . . .	74

3.5.6	Pooling layers . . . . .	76
3.5.7	Weights initialization and bias . . . . .	77
3.5.8	Back propagation in a CNN . . . . .	77
3.5.9	Batch normalization . . . . .	80
3.5.10	The use of CNN in astronomy . . . . .	80
<b>4</b>	<b>Estimating SMBH mass with ML for the GATOS core sample</b>	<b>83</b>
4.1	Motivation . . . . .	84
4.2	The GATOS core sample . . . . .	88
4.3	Numerical simulation . . . . .	89
4.3.1	Potentials and mass distributions . . . . .	90
4.3.2	Analytical velocity dispersion . . . . .	91
4.3.3	Analytical velocity . . . . .	94
4.3.4	Positions and velocities initialization . . . . .	96
4.3.5	Matching the ALMA observations . . . . .	96
4.4	ML approach to estimate the $M_{\text{BH}}$ . . . . .	98
4.4.1	Supervised learning approach . . . . .	98
4.4.2	Preparing the inputs . . . . .	99
4.4.3	ML architecture . . . . .	101
4.4.4	Estimating the reliability of models . . . . .	102
4.5	Results on $M_{\text{BH}}$ estimation . . . . .	103
4.5.1	Reliability of ML models . . . . .	103
4.5.2	$M_{\text{BH}}$ estimations results . . . . .	108
4.6	Individual galaxies . . . . .	110
4.6.1	NGC 4388 . . . . .	110
4.6.2	NGC 5506 . . . . .	111
4.6.3	NGC 5643 . . . . .	112
4.6.4	NGC 6300 . . . . .	113
4.6.5	NGC 7314 . . . . .	113
4.6.6	NGC 7465 . . . . .	114
4.6.7	NGC 7582 . . . . .	114
4.7	Discussion . . . . .	115
4.8	Conclusion . . . . .	117
	<b>Bibliography</b>	<b>133</b>
	<b>Appendix</b>	<b>134</b>
<b>A</b>	<b>RLAGN</b>	<b>135</b>
A.1	Tables of the parameters of our RLAGN sample . . . . .	135
A.2	Figures with highlighted populations . . . . .	140
A.3	Radio sources SEDs . . . . .	142
<b>B</b>	<b>GATOS</b>	<b>145</b>
B.1	Moment maps and position-velocity diagrams . . . . .	145
B.2	Relative errors of the $M_{\text{BH}}$ predictions as a function of free parameters . . . . .	146
B.3	Models 2D histograms of their RE . . . . .	150



---

# LIST OF FIGURES

---

1.1	Hubble sequence/tuning fork diagram . . . . .	3
1.2	EHT observations of Sagittarius A* and M87 . . . . .	4
1.3	Schematic view of the unified AGN model . . . . .	6
1.4	Examples of changing-look AGN spectra transition . . . . .	7
1.5	Typical AGN X-ray spectrum . . . . .	9
1.6	HERG and LERG differentiation . . . . .	13
1.7	Schematic view of the dusty and molecular circumnuclear regions . . . . .	15
1.8	Comparison of the total luminosity function of galaxies between the prediction of $\Lambda$ CDM simulations and the observations . . . . .	18
1.9	Illustration of the relationships between the SMBH growth, the star formation and the gas tank available in a galaxy . . . . .	19
1.10	Observations of the two kinds of AGN feedback . . . . .	20
1.11	Schematic view of different phases around a quasar mode AGN momentum-driven outflows and energy-driven ones . . . . .	22
1.12	$M_{\text{BH}}-\sigma$ relation for a sample of 72 galaxies . . . . .	23
2.1	RLAGN sample selection process . . . . .	32
2.2	1.4 GHz Radio luminosity density as a function of redshift for the radio sources in our sample . . . . .	33
2.3	IR colors for the radio sources in our sample . . . . .	34
2.4	Examples of SEDs of the radio sources in our sample with $M_{\text{BH}}$ estimates . . . . .	39
2.5	SFR vs $M_{\star}$ for the galaxies in our sample . . . . .	41
2.6	Scatter plot of $M_{\text{BH}}$ vs $M_{\star}$ for our sample of radio sources and for the galaxies in our comparison sample . . . . .	45
2.7	$M_{\text{BH}}$ to $M_{\star}$ ratio as a function of redshift for our sample of radio sources and for the galaxies in our comparison sample . . . . .	46
2.8	Jet power vs. $M_{\text{BH}}$ for our sample of radio sources and for the galaxies in our comparison sample . . . . .	48
2.9	Jet power vs. the Eddington ratio for our sample of radio sources and for the galaxies in our comparison sample . . . . .	50
3.1	Schema of a artificial neuron . . . . .	59

3.2	Schematic view of a binary neuron with a bias node . . . . .	60
3.3	Schematic view of a perceptron network . . . . .	61
3.4	Example of sigmoid functions and their derivatives for $\beta = 1$ and $\beta = 0.5$ . . . . .	62
3.5	Schema of an MLP with an input layer of 3 neurons, 2 hidden layers composed of 4 neurons each, and an output layer with 2 neurons . . . . .	62
3.6	Example of a square pattern and its corresponding localized target . . . . .	69
3.7	Schematic view of how a 2D image of a squared pattern is flattened to become a dense NN input layer and how an output layer is reshaped into the target 2D image . . . . .	70
3.8	Example of an image containing 2 squared patterns as well as other non-null pixels . . . . .	72
3.9	Example of a convolution operation . . . . .	73
3.10	Example of the ReLU and Leaky ReLU with a $\lambda$ of 0.2 . . . . .	75
3.11	Example of max pooling 2x2 onto a 4x4 matrix . . . . .	76
3.12	Example of a max pooling layer 2x backpropagation on a single layer image . . . . .	78
3.13	Example of the error propagated through a convolution layer . . . . .	79
4.1	SMBH fitting results from <a href="#">Onishi et al. (2017)</a> . . . . .	85
4.2	SMBH fitting results from <a href="#">Combes et al. (2019)</a> . . . . .	86
4.3	Schematic view of the numerical simulation . . . . .	90
4.4	Illustration of the impact of utilizing the 0 <sup>th</sup> order as a multiplicative filter in my NGC 5643 simulations. It includes the following elements: (1) the initial first-moment map from the simulated data cube, (2) the observed 0 <sup>th</sup> order as the multiplicative filter, (3) the updated first-moment map following filter application, and (4) the observed first-moment map for comparison with the simulation results. . . . .	97
4.5	Example of NGC 7582 position-velocity diagram and the first three-moment map obtained from the CO(3-2) ALMA data cubes observations . . . . .	100
4.6	Histograms of the RE of my models when estimating the $M_{\text{BH}}$ of the 5000 simulations from their respective test set. . . . .	105
4.7	NGC 4388's model relative errors on the $M_{\text{BH}}$ estimations made on the test set as a function of the simulation parameter's values . . . . .	106
4.8	Two-dimensional and one-dimensional histograms of the relative errors RE of my models when estimating the $M_{\text{BH}}$ of the test set . . . . .	107
4.9	Probability distributions of the $M_{\text{BH}}$ predicted by my models for the seven galaxies in my sample . . . . .	108
4.10	Comparison of the probability distributions of the $M_{\text{BH}}$ predicted by my models and the ones found in the literature . . . . .	109
4.11	First three moments maps and PVD comparison between NGC 4388 observations and the best-fit model . . . . .	111
4.12	Same as Fig. 4.11 for NGC 5506. . . . .	111
4.13	Same as Fig. 4.11 for NGC 5643. . . . .	112
4.14	Same as Fig. 4.11 for NGC 6300. . . . .	113
4.15	Same as Fig. 4.11 for NGC 7314. . . . .	113
4.16	Same as Fig. 4.11 for NGC 7465. . . . .	114
4.17	Same as Fig. 4.11 for NGC 7582. . . . .	115
A.1	Same figure as Fig. 2.8 but with the outliers in our sample highlighted. . . . .	141
A.2	Same figure as Fig. 2.9 but with the outliers in our sample highlighted. . . . .	141

A.3	SEDs of the radio sources in our sample with $M_{\text{BH}}$ estimates . . . . .	143
B.1	Same as Fig. 4.7 for NGC 5506. . . . .	147
B.2	Same as Fig. 4.7 for NGC 5643. . . . .	147
B.3	Same as Fig. 4.7 for NGC 6300. . . . .	148
B.4	Same as Fig. 4.7 for NGC 7314. . . . .	148
B.5	Same as Fig. 4.7 for NGC 7465. . . . .	149
B.6	Same as Fig. 4.7 for NGC 7582. . . . .	149
B.7	Same as Fig. 4.8 for NGC 5506. . . . .	150
B.8	Same as Fig. 4.8 for NGC 5643. . . . .	151
B.9	Same as Fig. 4.8 for NGC 6300. . . . .	151
B.10	Same as Fig. 4.8 for NGC 7314. . . . .	152
B.11	Same as Fig. 4.8 for NGC 7465. . . . .	152
B.12	Same as Fig. 4.8 for NGC 7582. . . . .	153



---

# LIST OF TABLES

---

2.1	Coefficients for estimating $M_{\text{BH}}$ using broad emission lines and Eq. 1.7 . . . . .	36
4.1	Lower and upper limits of the simulation parameters used to generate the training, validation, and test sample for each galaxy . . . . .	99
4.2	Architecture of the neural network used for the MBH estimations with the PVD as inputs . . . . .	101
4.3	Architecture of the neural network used for the MBH estimations with the moment maps as inputs . . . . .	102
4.4	RMSE, $R^2$ and mean RE of our models on the test sets. The RMSE ( $> 0$ ), $R^2$ (between 0 and 1), and the RE are defined in Section 4.4.4 . . . . .	104
4.5	MBH estimations found in the literature and the one I made in Sect.4.5.2 . . . . .	110
A.1	Properties of the radio sources in our sample . . . . .	136
A.1	Continuation . . . . .	137
A.2	BH masses, accretion, and jet properties of the radio sources in our sample . . . . .	138
A.2	Continuation . . . . .	139
B.1	Table of the parameters used to obtain the fits of the observed moment maps shown in the figures 4.11 to 4.17 . . . . .	146



---

# LIST OF ACRONYMS

---

<b>AGN</b> Active Galactic Nuclei . . . . .	1
<b>SMBH</b> Super Massive Black Hole . . . . .	1
<b>FWHM</b> Full-Width Half Maximum . . . . .	4
<b>RLAGN</b> Radio-Loud AGN . . . . .	5
<b>RQAGN</b> Radio-Quiet AGN . . . . .	5
<b>FRI, FRII</b> Fanaroff-Riley radio galaxies . . . . .	5
<b>ULIRG</b> Ultra-Luminous Infrared Galaxy . . . . .	5
<b>LINER</b> Low-Ionization Nuclear Emission-Line Region galaxy . . . . .	5
<b>NLR</b> Narrow Line Region . . . . .	5
<b>BLR</b> Broad Line Region . . . . .	5
<b>Sy1, Sy2</b> Seyfert galaxy . . . . .	5
<b>NIR</b> Near Infrared . . . . .	7
<b>IR</b> Infrared . . . . .	7
<b>SED</b> Spectral Energy Distribution . . . . .	8

<b>BH</b> Black Hole . . . . .	10
<b>RIAF</b> Radiatively Inefficient Accretion Flow . . . . .	11
<b>ADAF</b> Advection Dominated Accretion Flow . . . . .	11
<b>HERG</b> High-Excitation Radio Galaxy . . . . .	12
<b>LERG</b> Low-Excitation Radio Galaxy . . . . .	12
<b>ALMA</b> Atacama Large Millimeter/submillimeter Array . . . . .	15
<b>GATOS</b> Galaxy Activity, Torus, and Outflow Survey . . . . .	16
<b>BCGs</b> Brightest Cluster Galaxies . . . . .	23
<b>SFR</b> Star Formation Rate . . . . .	22
<b>FPBH</b> Fundamental Plane of BH activity . . . . .	27
<b>UV</b> Ultraviolet . . . . .	7
<b>VLA FIRST</b> Very Large Array Faint Images of the Radio Sky at Twenty-centimeters . . . . .	31
<b>DES</b> Dark Energy Survey . . . . .	31
<b>SDSS</b> Sloan Digital Sky Survey . . . . .	32
<b>WISE</b> Wide-field Infrared Survey Explorer . . . . .	32
<b>S/N</b> Signal-to-Noise ratio . . . . .	32
<b>BOSS</b> Baryon Oscillation Spectroscopic Survey . . . . .	33
<b>GAMA</b> Galaxy And Mass Assembly . . . . .	34
<b>DR</b> Data Release . . . . .	34
<b>LLR</b> Low-Luminosity Radio Sources . . . . .	35
<b>HLR</b> High-Luminosity Radio Sources . . . . .	35

*LIST OF ACRONYMS*

---

<b>QSOs</b> Radio-Loud Quasars . . . . .	35
<b>NED</b> NASA/IPAC Extragalactic Database . . . . .	35
<b>RMS</b> Root Mean Square . . . . .	40
<b>ANNs</b> Artificial Neural Networks . . . . .	55
<b>ML</b> Machine Learning . . . . .	55
<b>CNN</b> Convolutional Neural Network . . . . .	55
<b>NN</b> Neural Network . . . . .	60
<b>MLP</b> Multi-Layer Perceptron . . . . .	63
<b>ReLU</b> Rectified Linear Uni . . . . .	64
<b>MC</b> Monte-Carlo . . . . .	68
<b>SoI</b> Sphere of Influence . . . . .	25
<b>NUGA</b> Nuclei of Galaxies survey . . . . .	88
<b>PVD</b> Position-Velocity Diagram . . . . .	87
<b>RMSE</b> Root Mean Square Error . . . . .	101
<b>RE</b> Relative Errors . . . . .	103
<b>ILR</b> Inner Lindblad Resonance . . . . .	110
<b>WISDOM</b> Wave Interferometric Survey of Dark Object Masses . . . . .	85



# CHAPTER 1

---

## INTRODUCTION TO AGN PHYSICS

---

In this chapter, I will provide the necessary astrophysical background to comprehend the research conducted in this work. I will begin by discussing the fundamental characteristics of galaxies and exploring their various types. Next, I will delve into the concept of Active Galactic Nuclei (AGN), including their classification and key properties. Subsequently, I will focus on the circumnuclear region and its dusty torus. Following that, I will delve into the topic of AGN feedback, understand why it is crucial, and examine its different manifestations. I will then investigate the connection between Super Massive Black Hole (SMBH) and galaxies, exploring how their evolution are intricately linked. Lastly, I will examine several of the most used methods to estimate a SMBH mass.

### Contents

---

<b>1.1 Galaxies</b> . . . . .	<b>2</b>
<b>1.2 Active Galactic Nuclei</b> . . . . .	<b>4</b>
1.2.1 The different types of AGN . . . . .	4
1.2.2 The unification model of AGN . . . . .	5
1.2.3 Global spectrum SED . . . . .	8
1.2.4 SMBH accretion . . . . .	10
1.2.5 Radio galaxies excitation . . . . .	12
1.2.6 Jets formation . . . . .	13
<b>1.3 The Circumnuclear Region</b> . . . . .	<b>14</b>
1.3.1 The dusty torus . . . . .	14
1.3.2 Molecular torus . . . . .	15
1.3.3 Polar biconical outflows . . . . .	16

<b>1.4</b>	<b>AGN Feedback</b>	<b>16</b>
1.4.1	The need for feedback	16
1.4.2	The two modes of AGN feedback	19
1.4.3	Energy or momentum conservation	21
<b>1.5</b>	<b>Co-Evolution of Galaxies and SMBH</b>	<b>22</b>
<b>1.6</b>	<b>Classical methods for estimating <math>M_{\text{BH}}</math></b>	<b>24</b>
1.6.1	Dynamical estimation	24
1.6.2	$M_{\text{BH}}$ - $\sigma$ relation	26
1.6.3	Reverberation mapping	26
1.6.4	Single epoch method	27
1.6.5	Fundamental Plane of BH Activity	27

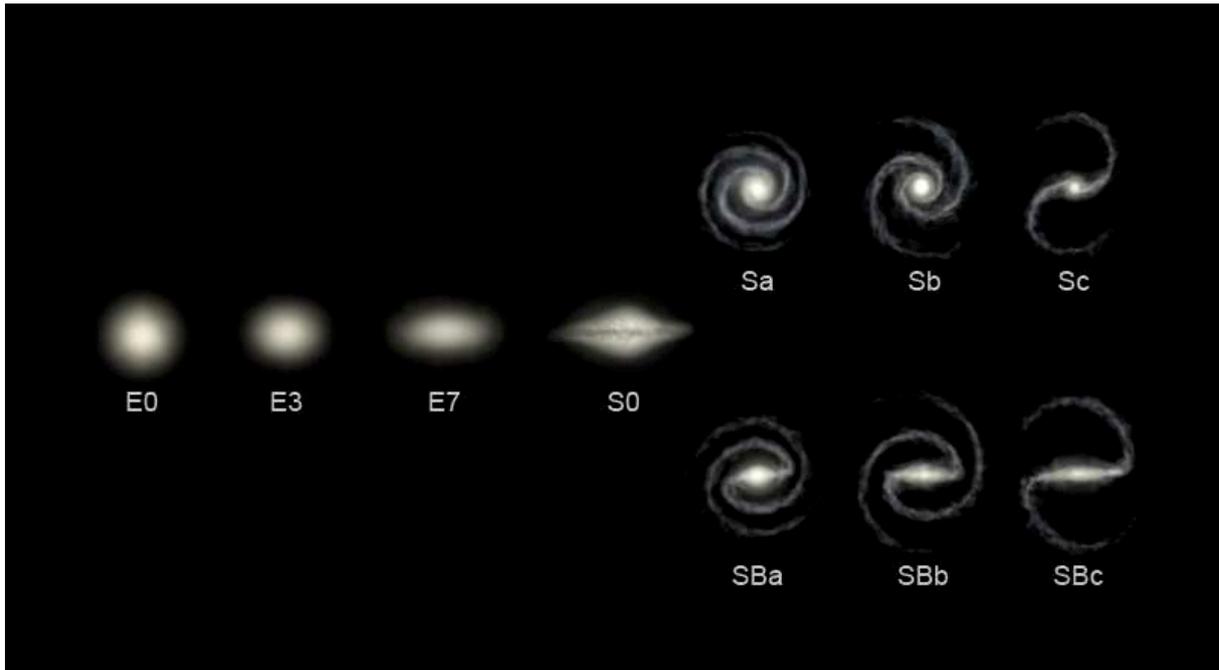
---

## 1.1 Galaxies

Galaxies are systems of stars, gas, dust, and dark matter bound together by gravity. They can be classified based on their morphological features using the Hubble sequence, or Hubble tuning fork diagram (Hubble 1936). It separates galaxies into three categories: elliptical, spiral, and irregular galaxies. These categories are further subdivided into more specific types based on their characteristics.

On the left of the diagram, we find elliptical galaxies, also called early-type galaxies. They are composed of older, red, low-mass stars and contain a sparse interstellar medium, resulting in little or no ongoing star formation within them. Nevertheless, they can experience short but intense episodes of star formation during mergers with other galaxies (Pearson et al. 2019). Elliptical galaxies tend to be found at the center of globular clusters (Dressler 1980) and they are not the most abundant type of galaxies in the Universe (Loveday 1996; Dalya et al. 2021). They are characterized by their ellipsoidal or nearly spherical shape. They are denoted by the letter "E" followed by a number between 0 and 7 representing their degree of elongation. This number is defined as  $10\eta$  where  $\eta = (a - b)/a$ . Here,  $a$  is the length of the major axis, and  $b$  is the one of the minor axis. The greater the number is, the more elongated the galaxy is.

Late-type galaxies, or spiral galaxies are placed on the right of the Hubble sequence. They are composed of a central concentration of stars called bulge, a disk-like structure, and spiral arms that wind outwards from the center. The bulge is a compact, spherical, or ellipsoidal structure located at the center of a galaxy. The bulge can vary in size and shape, ranging from small and nearly spherical to large and elongated. The central bulges of spiral galaxies consist of aging, red stars similar to those found in elliptical galaxies. However, within the rotating disk, there are much younger stars as well as gas and dust, which promote active star formation in the disk. In some cases, dense central components that resemble classical bulges can be formed. They result from the slow transformation of disk gas. These central components, referred to

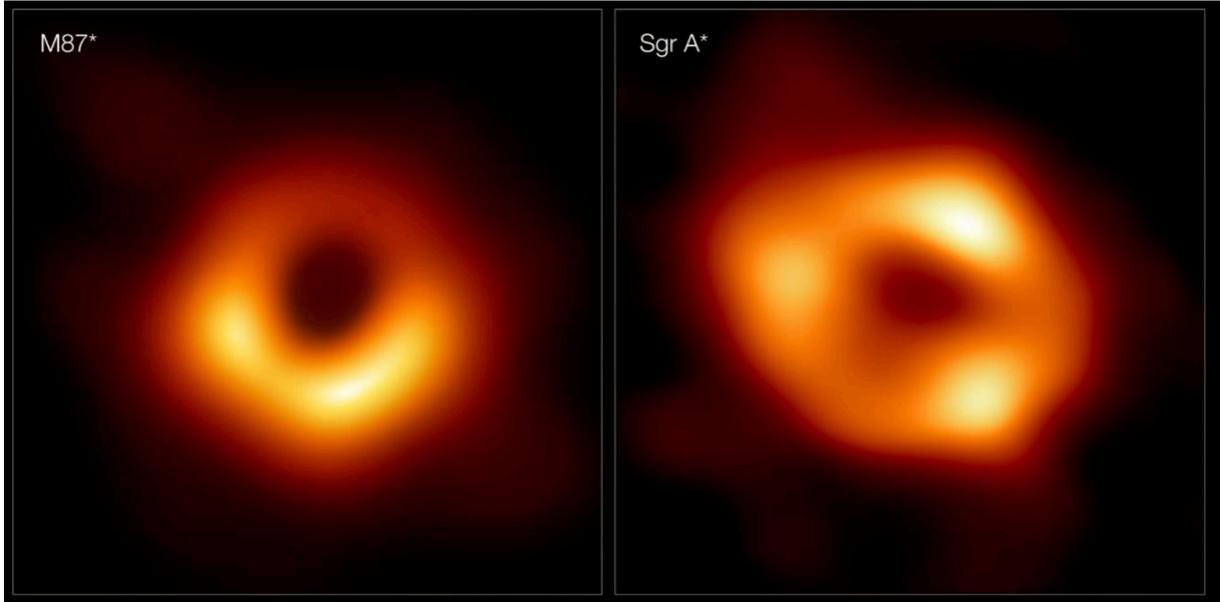


**Figure 1.1:** Hubble sequence/tuning fork diagram. On the left side, we observe the elliptical galaxies, while on the right side, we have the spiral galaxies. Among the spiral galaxies, we find on top the galaxies without bars whereas the barred galaxies are located at the bottom. Image made by Ville Koistinen.

as pseudobulges, exhibit similar characteristics to bulges that originate from galaxy mergers, despite having a distinct formation mechanism. Most galaxies in the Universe are spiral ones (Loveday 1996; Dalya et al. 2021) and most of them have bars (Melvin et al. 2014). Spiral galaxies are denoted by the letter "S". They are characterized by the tightness of their spiral arms and the size of the bulge using the letters "a", "b" or "c". The letter "a" describes spiral galaxies with tighter spiral arms and brighter bulge while "c" describes the ones with looser arms and fainter bulge. They are further divided into two branches based on the presence of a spiral bar (denoted with "B").

At the center of the diagram, between elliptical and spiral galaxies, the lenticular galaxies (S0) can be found. These galaxies have a similar appearance to elliptical while having a bright central bulge but no spiral arms. Some galaxies have irregular structures and do not fit into the Hubble diagram. They are called irregular galaxies.

The Hubble sequence initially offered a fundamental framework for understanding various galaxy morphologies. Still, over time, it has been revised and refined to provide a more comprehensive explanation for the diverse array of galaxies astronomers observe today (de Vaucouleurs 1959; van den Bergh 1976; Kormendy & Bender 1996, 2012; Cappellari et al. 2011; Graham 2019). For example, in 1959, de Vaucouleurs introduced a classification system for barred lenticular galaxies (de Vaucouleurs 1959). This system was primarily based on the contrast and strength of the bars observed in these galaxies. He began by introducing a category for galaxies with weak bars, labeled as the "AB" class. This classification was applied to both spiral and lenticular galaxies and occupied a middle ground between galaxies with distinct and strong bars (designated as "B") and those without bars altogether (designated as "A").



**Figure 1.2:** EHT pictures of the shadow of M87 (left) and Sagittarius A\* (right) ([The Event Horizon Telescope Collaboration et al. 2019, 2022](#)).

## 1.2 Active Galactic Nuclei

### 1.2.1 The different types of AGN

The nucleus of a galaxy can go through bright activity phases and become brighter than the rest of the galaxy ([Padovani et al. 2017](#)). This phenomenon is called AGN. A galaxy with such a nucleus is called an active galaxy. The luminosity of an AGN cannot be explained with only thermal emission. Most of the luminosity comes from the gas being accreted by a central SMBH. The AGN was the strongest proof of the existence of SMBH until more direct observations were made. Today's best proof of the existence of SMBH is the direct observations of the SMBH's shadow of Sagittarius A\* and M87 by the Event Horizon Telescope ([The Event Horizon Telescope Collaboration et al. 2019, 2022](#), , see Fig. 1.2).

Since the first observation by Carl Seyfert in 1943 ([Seyfert 1943](#)), numerous types of AGN have been discovered, each with their unique features. The primary characteristics of AGN are listed below ([Peterson 1997](#); [Beckmann & Shrader 2012](#)):

- The AGN nucleus appears star-like when compared to the host galaxy. They can emit up to  $L \sim 10^{15} L_{\odot}$ , which can exceed the radiation of the entire host galaxy by up to 1000 times.
- Intense and extremely wide emission lines are present in the optical spectrum, with a Full-Width Half Maximum (FWHM) of up to  $\sim 10000 \text{ km.s}^{-1}$ .
- Both their continuum and emission lines can vary. The timescales of variability are generally on the order of months or years, although they can be as short as days for the most

luminous AGN, and even minutes for AGN at smaller wavelengths, such as X-ray or gamma rays.

- In some cases, extended jets made of matter ejected at high velocities can be observed, with lengths ranging from  $10^{-5}$  pc to 100 kpc. See [Mickaelian \(2015\)](#) for a more detailed review of the different type of AGN.

All the different types of AGN can be classified into two main classes: Radio-Quiet AGN (**RQAGN**) and Radio-Loud AGN (**RLAGN**). RLAGN have strong emissions in the radio wavelengths and constitute about 10% of the AGN population. Quasars, Fanaroff-Riley radio galaxies (**FRI**, **FRII**), and blazars (BL-Lacs and FS-RQs- Flat Spectrum Radio Quasars) are classified as RLAGN. On the contrary, RQAGN have a fainter emission in this domain. Seyfert galaxies, Ultra-Luminous Infrared Galaxy (**ULIRG**), and Low-Ionization Nuclear Emission-Line Region galaxy (**LINER**) are classified as RQAGN.

## 1.2.2 The unification model of AGN

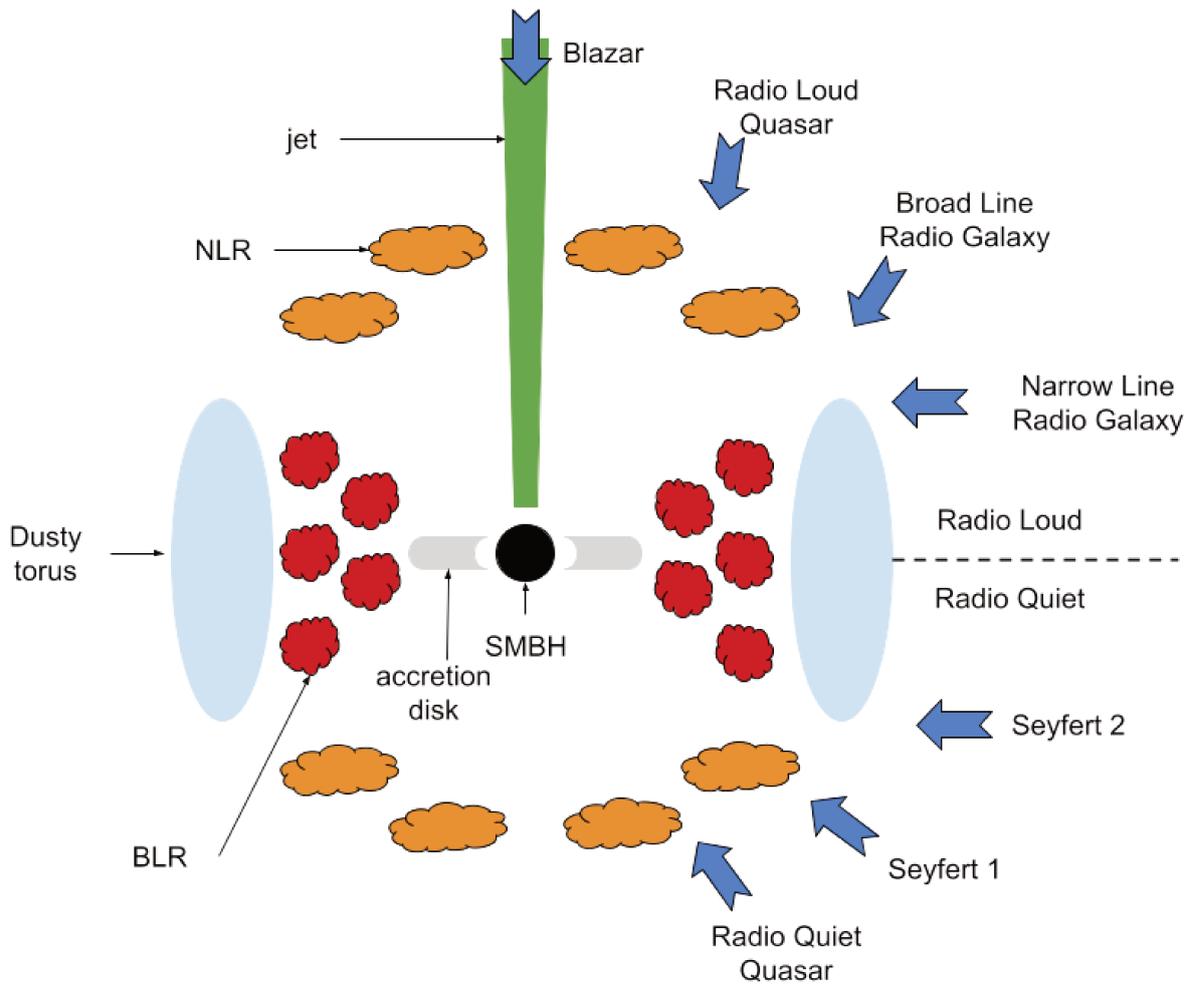
As previously discussed, a wide variety of AGN types exist. However, there must be some shared characteristics among them. The discovery of a dusty torus located in the sub-parsec region, which obscures the central region when viewed edge-on, has contributed to the development of an improved classification system for AGN based on their orientation ([Urry & Padovani 1995](#)).

In addition to the accretion disk, AGN have two distinct gas regions with differing velocities: the Narrow Line Region (**NLR**) and the Broad Line Region (**BLR**). These regions are classified based on the width of their optical emission lines. The BLR is typically less than 1 pc in size and is too small to be resolved. It has gas velocities that can reach up to  $40,000 \text{ km.s}^{-1}$  and a temperature of  $10^3 \text{ K}$ . In contrast, the NLR can be as large as kiloparsecs in Seyfert galaxy ([Sy1](#), [Sy2](#)) and has gas velocities lower than  $1000 \text{ km.s}^{-1}$ .

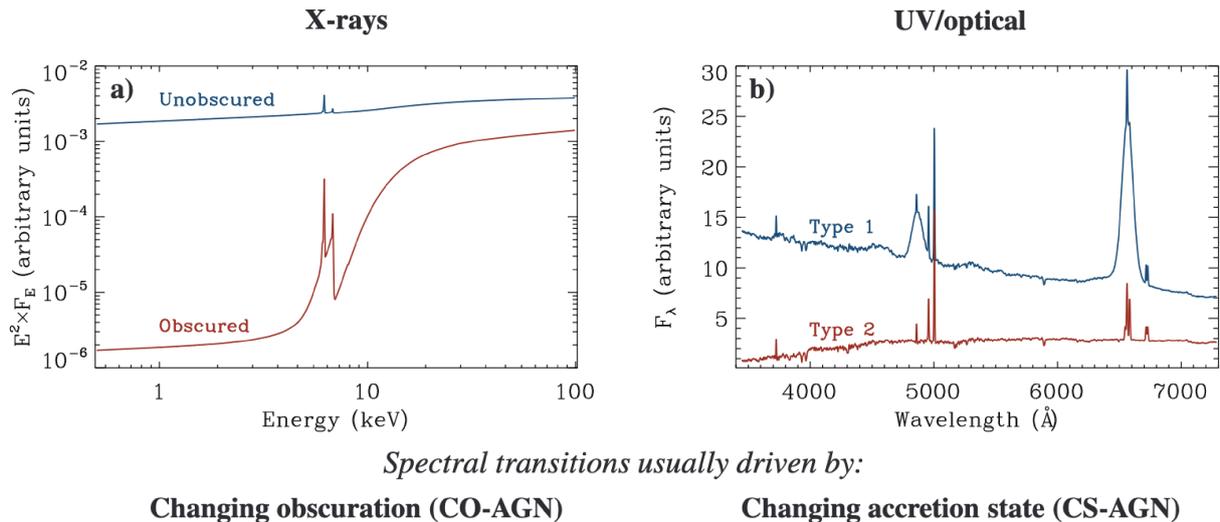
Recombination lines, including hydrogen Lyman and Balmer lines, are considered permitted lines. They are named so because they can be observed in laboratory conditions. In contrast, forbidden lines do not appear in laboratory settings. This is due to their extremely low coefficient of spontaneous emission, which causes the upper energy levels to be de-excited primarily through collisions with other ions ([Zuckerman & Ball 1974](#)). However, in interstellar space, where the density is significantly lower, collisions are rare, allowing for the observation of spontaneous emission. Nevertheless, there is a threshold density beyond which these forbidden lines cannot be observed. In regions with high gas density near the nucleus, these lines are not visible. In the BLR, the gas density exceeds the critical density for forbidden lines, resulting in the observation of only permitted lines. On the other hand, the low density of the NLR permits the presence of forbidden emission lines such as [OIII] ([Vrtilek & Carleton 1985](#)).

In the unification model, AGN can be classified into two main classes:

- Type 1 AGN when the observer is looking face-on and can see both the BLR and the NLR.



**Figure 1.3:** Schematic view of an AGN. The accretion disk and the BLR are surrounded by a dusty torus obscuring the central region when in the line of sight of an observer (Type 2 AGN). When the torus is not in the line of sight, Type 1 AGN, the BLR, and NLR are both visible. When observed face-on, the observer will see a blazar if the galaxy is radio-loud.



**Figure 1.4:** a) Alterations in X-ray emissions demonstrating the shift from an unobscured condition (blue line) to a heavily obscured, Compton-thick state (red line). b) Optical band spectral shift illustrating the transition from a type 1 AGN (blue line) to a type 2 AGN (red line). Figure from Ricci & Trakhtenbrot (2022)

- Type 2 AGN is when an observer is looking edge-on and the torus obscures the radiation from the BLR.

Several observations evidence that support the unification model were found. Antonucci & Miller (1985) detected BLR lines in some Sy2 via the reflection of the obscuring material using spectro-polarimetry. The presence of an obstruction in the BLR of Sy2 was confirmed by the detection of broad lines, such as Pa $\alpha$ , Pa $\beta$ , and Br $\gamma$ , in the Near Infrared (NIR) spectrum. These lines were not observed in the optical range due to higher levels of dust absorption (Veilleux et al. 1997; Lutz et al. 2002). Similar evidence was found in the X-ray range, where soft X-rays are absorbed while hard X-rays are detected, indicating the presence of an obstructing component along the line of sight with a minimum column density of  $N_H = 10^{22} \text{ cm}^{-2}$  (Bassani et al. 1999). Dust grains absorb Ultraviolet (UV) and X-rays and re-emit them in the Infrared (IR). This explains why Sy1 is brighter than Sy2 in UV and X-rays and why Sy2 can be brighter in the IR.

AGN exhibit variability across the entire electromagnetic spectrum (Vanden Berk et al. 2004; Uttley et al. 2005), with these variations occurring over time frames ranging from hours to years (Ulrich et al. 1997). Typically, such variability is characterized by random fluctuations and remains within a range of a few tens of percent.

Over the past two decades, an increasing number of accreting SMBH have displayed more pronounced and coordinated changes in their spectral properties, sometimes leading Sy1 galaxies to transitioned into y2 galaxies or vice versa (McElroy et al. 2016). These changes manifest across a wide range of timescales, thereby questioning the validity of the classical unification model. AGN undergoing these distinct variations are referred to as changing-look AGN and have been observed both in the X-ray and optical/UV spectra.

Changing-look AGNs can be categorized into two types. In the X-ray band, these transformations are primarily attributed to alterations in the line-of-sight column density, typically

arising from gas situated near the SMBH. These are known as changing-obscuration AGN (Mereghetti et al. 2021). In the optical/UV domain, such events are commonly associated with rapid changes in the radiation field driven by accretion. These changes result in the suppression, enhancement, or even the disappearance of the characteristic blue continuum and broad optical/UV lines typical of Type 1 AGNs. These are referred to as "changing-state AGN" (Graham et al. 2020). The fundamental spectral states of AGN is shown in Fig.1.4, illustrating their potential transitions between these states in both the X-ray and optical spectral domains.

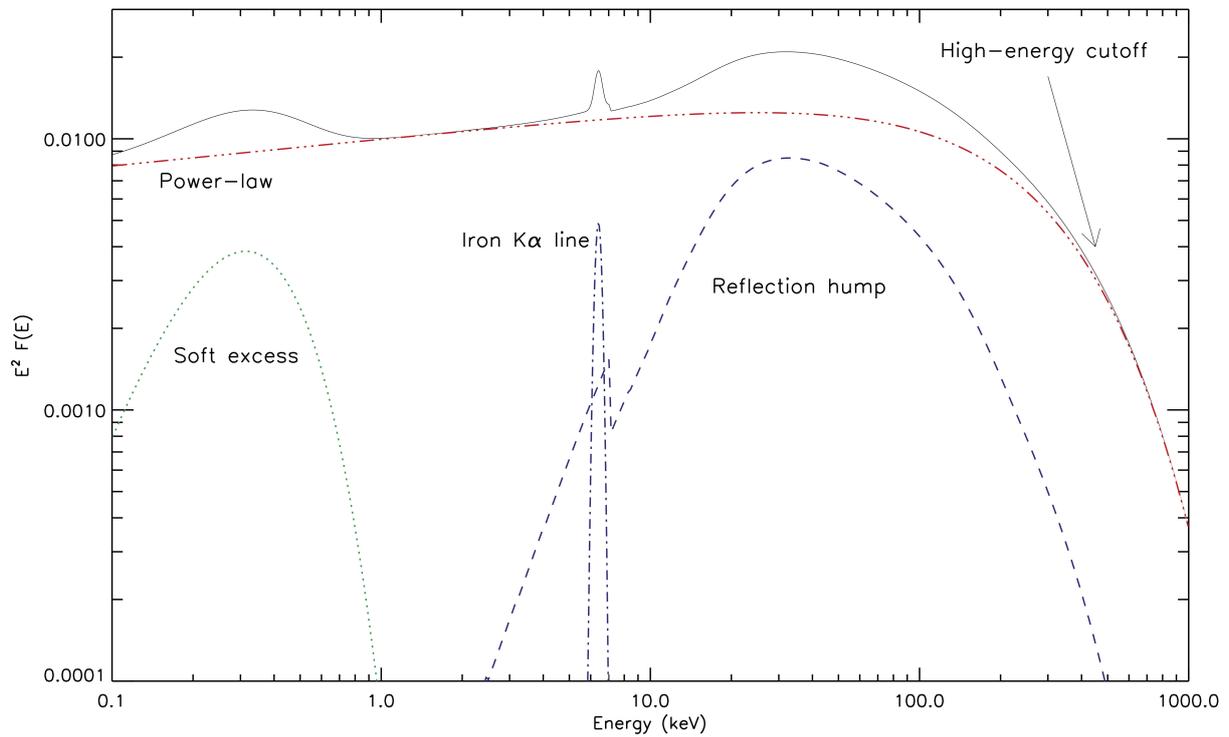
### 1.2.3 Global spectrum SED

The Spectral Energy Distribution (SED) of non-active galaxies is characterized by a continuous thermal distribution coming from the stars of the galaxies. In the case of active galaxies, the AGN emits strongly in many wavelengths which makes their SED very distinct from the non-active ones (Kauffmann et al. 2003; Lacy et al. 2015; Brandt & Alexander 2015; Tadhunter 2016). In the low-frequency, from the radio to the IR domain, a power law can be observed. Then, in the mid to near-IR, the black-body emission coming from the accretion disk begins. According to AGN luminosity, non-stellar radiation dominates more or less over the host stellar and interstellar radiation. This domination is observed when considering a small aperture around the center.

The various physical processes responsible for generating different parts of the spectrum are well established. At low frequencies, particularly in the radio, synchrotron emission is responsible for the observed power law. The spectral slope of the power law provides information about the age of the emission. The spectrum is almost flat close to the AGN, owing to the high-energy electrons present in the vicinity. However, as electrons with higher energy levels lose energy first, the spectrum slopes steepen. In some cases, the synchrotron emission can persist into the optical and X-ray wavelengths.

Near the accreting SMBH, the intense UV-X-ray radiation emitted by the AGN heats the dust up to temperatures exceeding 100 K. Consequently, the dusty molecular torus thermal emission is anticipated to strongly dominate the millimeter part of the spectrum, within the wavelength range of 3-20  $\mu\text{m}$  (Nenkova et al. 2002). Nonetheless, the continuum emission around 10  $\mu\text{m}$  may primarily emanate from the inner, hotter region of the torus, possibly failing to provide a comprehensive representation of the entire torus structure (Schartmann et al. 2008). Additionally, observations of nearby AGN like NGC 1068 have shown that the detected mid-IR continuum frequently results from radiation coming from polar dust or is influenced by scattering, rather than originating directly from the presumed torus itself (Bock et al. 2000; Tristram et al. 2014; Hönig & Kishimoto 2017).

The findings from the IRAS indicated that numerous AGN exhibit robust emissions in the far-IR spectrum, displaying spectral energy distributions akin to those observed in non-AGN star-forming galaxies (Soifer et al. 1987). The resemblance between the emission patterns and the large IRAS beams led to the initial attribution of this emission to star formation occurring within the galaxies hosting AGN. These results convincingly indicated that the IR emissions in the near- and mid-IR range in Seyfert galaxies are mainly produced by dust heated by the central engine, while the far-IR emissions are likely powered by young stars.



**Figure 1.5:** Example of a typical unobscured AGN X-ray spectrum (0.1-1000 keV). The full spectrum is represented by the solid black line while its components are plotted in dashed (the powerlaw up to the high-energy cutoff, the soft excess in green, the iron  $K\alpha$  line and the reflection hump). Figure from (Ricci 2011).

In the UV domain, the spectrum is dominated by the black-body emission coming from the accretion disk, creating a big blue bump (Gierliński & Done 2004; Porquet et al. 2004; Piconcelli et al. 2005) which contains most of the bolometric luminosity. Within the X-ray spectrum of AGNs, characteristic components are commonly observed, which encompass: a soft excess appearing below 2 keV, a power-law continuum extending up to roughly 10 keV, a complex of Fe  $K\alpha$  emission lines typically at around 6.4 keV and a Compton scattering bump that occurs near 20–30 keV (see Fig.1.5).

For the majority of AGN, the soft X-ray excess is not an extension of the black-body emission from the accretion disk (Gierliński & Done 2004; Porquet et al. 2004; Piconcelli et al. 2005), except potentially in low-mass Narrow Line Seyfert 1 galaxies (Done et al. 2012). Various theories have been proposed to explain the origin of this soft excess: possibilities include photo-ionized emission affected by relativistic motion in the accretion disk (Crummy et al. 2006), Comptonization of soft photons by a corona above the disk (Czerny & Elvis 1987), and an effect caused by strong, smeared, partially ionized absorption (Gierliński & Done 2004). While these models offer reasonable explanations for the soft excess, current simulations of line-driven AGN accretion disk winds are unable to replicate the smooth, soft X-ray excess (Schurch et al. 2009), leaving its origin an unresolved matter.

The Fe  $K\alpha$  emission comes from the neutral material in the BLR and the molecular torus. Its analysis provides crucial insights into the physical and dynamic conditions of AGN central engines, ranging from the inner accretion disk to the more distant molecular torus. However, the physical understanding of the broad Fe  $K\alpha$  lines in observations is still not clear. This is mainly due to the frequent presence of a warm absorber (Porquet et al. 2004; Piconcelli et al. 2005; Blustin et al. 2005) which may substantially complicate the analysis of X-ray data by distorting the underlying continuum of the Fe  $K\alpha$  line (Turner & Miller 2009).

Similarly, various sources for the hard X-ray spectra beyond 10 keV in Sy1 have been suggested. These potential sources include relativistic reflection, intricate absorption, Comptonization, or a combination of these factors (Nardini et al. 2011; Noda et al. 2011; Walton et al. 2014; Mehdipour et al. 2015; Bartels et al. 2015). At the highest energies,  $\gamma$ -rays are emitted by non-thermal processes such as inverse Compton scattering in beamed relativistic jets and pair annihilation in the jets.

## 1.2.4 SMBH accretion

The luminosity of an AGN comes from the accretion of matter by an SMBH. As material falls towards the Black Hole (BH), it forms an accretion disk. The SMBH can accrete material from the dense central region of its host galaxy. As material orbits around the SMBH, the friction generated due to the viscosity of the gas causes it to lose angular momentum and spiral towards the SMBH. A substantial portion of the rest mass energy acquired through accretion is emitted in the vicinity of the BH event horizon, resulting in an accretion luminosity  $L$  expressed as

$$L = \eta \dot{M}_{\text{BH}} c^2 \quad (1.1)$$

Here,  $\dot{M}_{\text{BH}}$  denotes the BH accretion rate,  $c$  represents the speed of light in a vacuum, and  $\eta$  stands for the radiative efficiency of the BH. The radiative efficiency typically falls within the

range of 0.05 to 0.42, contingent on the spin of the BH. Comparative analyses of observed AGN luminosities and the deduced BH mass density propose an average value of  $\eta = 0.1 - 0.2$  (Soltan 1982; Fabian & Iwasawa 1999; Yu & Tremaine 2002), aligning with the characteristics of moderately spinning BH. The theoretical maximum luminosity produced by this phenomenon depends on the BH mass ( $M_{\text{BH}}$ ) and is given by:

$$L_{\text{Edd}} = 4\pi Gc \frac{m_p}{\sigma_T} \left( \frac{M_{\text{BH}}}{M_{\odot}} \right) \text{ erg s}^{-1} . \quad (1.2)$$

where  $G$  is the gravitational constant,  $c$  is the speed of light,  $\sigma_T$  is the Thomson scattering cross-section for the electron and  $m_p$  is the mass of a proton. When the SMBH accretes matter quickly enough to reach  $L_{\text{Edd}}$ , the radiation pressure caused by the accretion balances the gravitational force of the SMBH, stopping the accretion. The Eddington ratio compares the nucleus bolometric luminosity to the  $L_{\text{Edd}}$ :

$$\eta = \frac{L_{\text{bol}}}{L_{\text{Edd}}} \quad (1.3)$$

This ratio indicates the accretion rates of an SMBH. The closer to 1 the Eddington ratio is, the more efficiently the SMBH accretes matter. By definition, a nucleus is active when its Eddington ratio is greater than  $10^{-5}$ . When the Eddington ratio is higher than 0.01, the accretion rate is sufficient to build a geometrically thin, optically thick, and radiatively efficient disk. When the Eddington ratio is much lower than 0.01, the accretion becomes inefficient. This regime is called Radiatively Inefficient Accretion Flow (RIAF) or Advection Dominated Accretion Flow (ADAF). The gas heats and inflates, creating a hot corona around the accretion disk, and most of the energy is advected into the BH. This state is often accompanied by a relativistic radio jet, which ejects most of the energy of the AGN through kinetic energy.

This low state is mostly encountered in early-type galaxies with very massive SMBH and with low gas density available in their circumgalactic medium. On the contrary, a high state is frequently found in late-type galaxies rich in gas and with a smaller SMBH. This correlation appears to be true even at high redshift. Thus, the co-evolution of galaxy bulges and SMBH is likely due to secular evolution more than mergers (Kormendy & Ho 2013; Heckman & Best 2014).

Indirect observation of the AGN accretion disk is possible by studying X-rays that are reflected off it. These reflection characteristics, which include phenomena like fluorescent emissions from iron and the scattering of photons through Compton scattering, leave their mark on the original X-ray spectrum. They can provide insights into the geometry of the accretion process (Ross & Fabian 2005; García & Kallman 2010; Wilkins & Fabian 2012; Dauser et al. 2016; Taylor & Reynolds 2018).

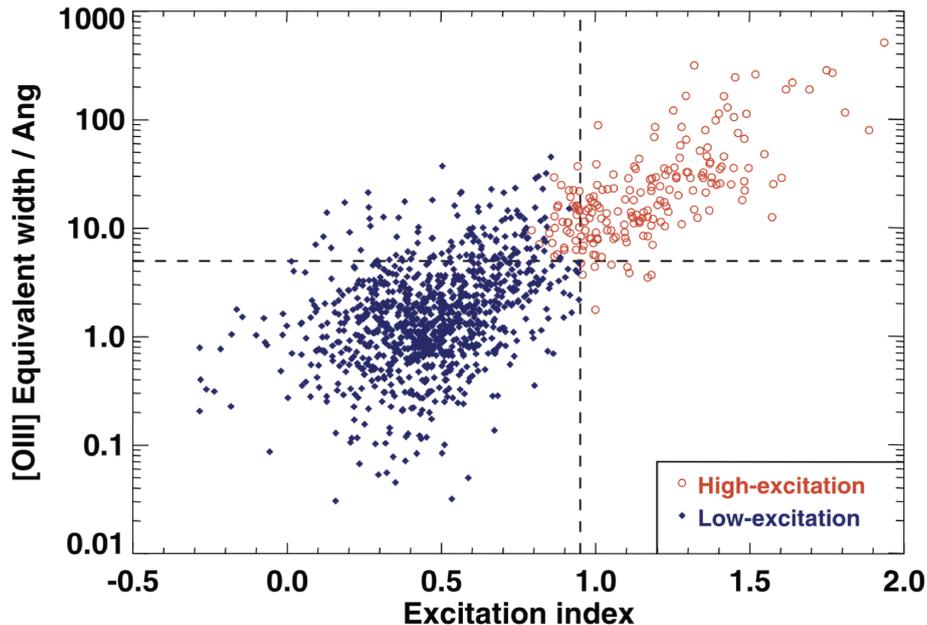
In theory, the accretion of matter with an Eddington ratio exceeding one is considered impossible, assuming spherical accretion and the dominance of Thomson scattering in opacity. However, practical observations reveal that this theoretical limit can be surpassed due to factors like non-spherical geometries and the presence of various instabilities. When an object accumulates matter at a rate surpassing the Eddington limit, this phenomenon is referred to as super Eddington accretion. Rare examples of SMBH accreting above the Eddington limit have been observed (Lanzuisi et al. 2016; Tsai et al. 2018). Still, our understanding of the Super

Eddington regime is still limited and the next generations of instruments will be required to better investigate this regime (Brightman et al. 2019).

## 1.2.5 Radio galaxies excitation

Radio galaxies can be categorized between High-Excitation Radio Galaxy (HERG) and Low-Excitation Radio Galaxy (LERG) (Best & Heckman 2012). In general, HERG exhibit a higher radio luminosity than LERG. The Eddington ratios for HERG range from 0.1 to 0.01 while for LERG it remains below 0.01. Moreover, the HERG are known for their feedback mechanism and triggering of gas outflow. They are usually found in quasar mode, where the radiation pressure ejects wind from the disk. In contrast, the LERG are usually in radio mode, where the gas forms radio jets via shocks.

These two classes are based on the optical lines diagnostic and Baldwin, Phillips, and Terlevich diagram (Baldwin et al. 1981) (Fig.1.6). The presence of the permitted lines  $H_\alpha$  and  $H_\beta$  is also a separating parameter since they exist almost only in HERG. The separation criteria between HERG and LERG works well in most cases but is not absolute since there exist powerful LERG.



**Figure 1.6:** [OIII] equivalent width (in Angstrom) versus the excitation index based on optical lines (Buttiglione et al. 2010). The two approaches are consistent in differentiating between HERG and LERG. Figure from Best & Heckman (2012).

## 1.2.6 Jets formation

RLAGN are typically characterized by jetted outflows that strongly emit at radio wavelengths, mainly via synchrotron emission. These jets are collimated from pc-scales to Mpc-scales. The jets need to be a plasma of electron/proton and electron/positron pairs in order to reproduce the X-ray spectrum of observed radio galaxies with jets.

Currently, observations cannot resolve the region where the jets are formed so the mechanism of their formation is still only theoretical. Several models have been proposed to explain their formation and in many of them, the AGN magnetic fields play an important role. Jets are launched via the rotational kinematic energy of the SMBH. Thus, the jet power is expected to mostly depend on the SMBH spin. In the Blandford & Znajek (1977) model, the BH rotational energy is extracted by a large-scale magnetic field. This makes the vacuum around the BH unstable and relativistic electron/positron pairs are created by the electromagnetic energy. The accretion disk plays a more important role in the Blandford & Payne (1982) model. The cold disk angular momentum is removed magnetically through the hot corona above the disk. The disk field lines are stressed by the rotation and the wind is collimated into a jet. This process is helped by an encircling magneto-hydrodynamic wind leaving the accretion disk at a slower speed. In the Contopoulos et al. (2009) study, the measure of the Faraday rotation has shown that the magnetic field is rotating helicoidally in the jets.

## 1.3 The Circumnuclear Region

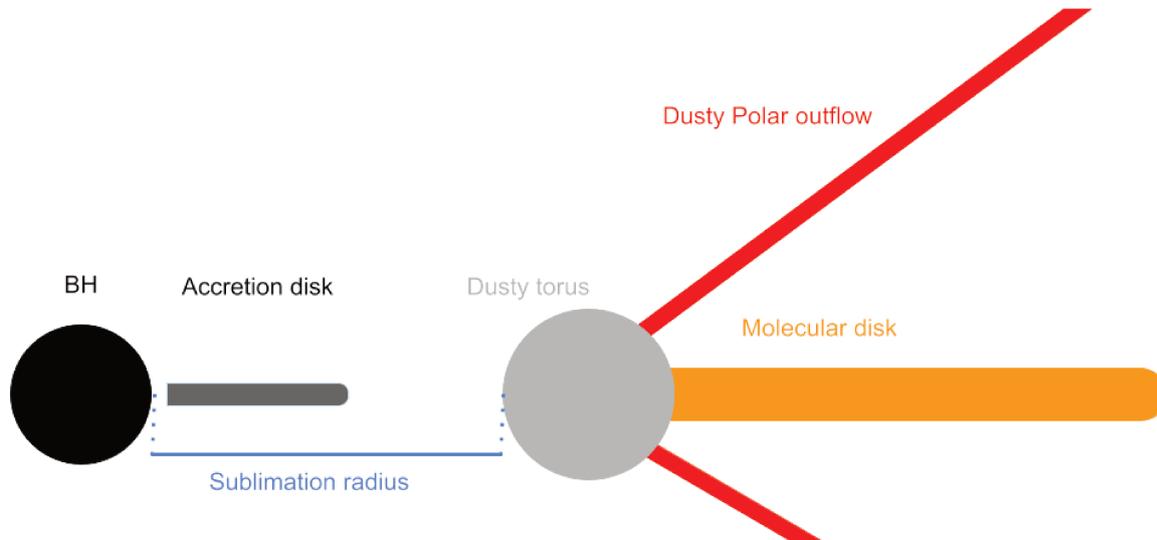
### 1.3.1 The dusty torus

As seen previously in Sect. 1.2.2, AGN exhibit different characteristics. Class 1 AGN have their BLR directly observable while class 2 AGN have their BLR hidden from view. This obscuration arises due to the presence of a thick and approximately axisymmetric structure known as the dusty torus. The dusty torus extends over a spatial scale of 1 – 10 pc (Jaffe et al. 2004; López-Gonzaga et al. 2016; Leftley et al. 2018) and resides just outside the dust-free accretion disk and the BLR. This dust absorbs a important part of the AGN radiation and, reprocesses it to radiate in the IR (Pier & Krolik 1992). It effectively obstructs our line of sight, if it passes through it, preventing direct observation of the BLR and contributing to the obscured nature of class 2 AGN (see Fig. 1.3).

The sublimation radius is the location where dust grains evaporate. This temperature is reached at 1400K for silicate grains and at 1800K for graphite ones. This radius is typically in the order of 1 pc for nearby Seyferts and varies with the square root of the AGN luminosity (Minezaki et al. 2004; Suganuma et al. 2006). At the sublimation radius, the dusty torus must have a doughnut shape (Krolik & Begelman 1988; Urry & Padovani 1995). It must be thick and clumpy, fragmented in clouds. To maintain its thickness, the torus must have a high-velocity dispersion in the vertical direction ( $50 - 100 \text{ km.s}^{-1}$ ) and the clouds inside it must have a low collision rate. These clouds experience heating and surface evaporation on the side facing the nucleus. This implies that they must be dense enough to shield the cold, dense gas at their core against the radiation emitted by the AGN. The collision process, although weak, leads to the necessary dissipation, resulting in the clouds spiraling inward towards the central region. This promotes the accretion of material into the nucleus (Nenkova et al. 2008).

The torus should possess column densities of sufficient magnitude to obscure the optically visible BLR completely. Specifically, a gas column density of at least  $10^{23} \text{ cm}^{-2}$  is necessary. At such densities, a portion of X-ray radiation can still be detected. However, if the gas density surpasses  $10^{24} \text{ cm}^{-2}$ , the X-ray emissions in the 2-10 keV range are obstructed (Kudoh et al. 2023). Moreover, with a density of  $10^{25} \text{ cm}^{-2}$ , even X-ray emissions above 10 keV become obscured. AGNs characterized by these conditions are referred to as mildly and heavily Compton-thick, respectively.

Continuing from the dusty torus, two other structures arise: a molecular torus and polar biconical outflows (see Fig. 1.7).



**Figure 1.7:** Two-dimensional schematic view of the dusty and molecular circumnuclear regions. From left to right we find the BH (in black), the accretion disk (in dark gray), the dusty torus (in gray), the dusty polar outflows (in red), and the molecular torus/disk (in orange). The sublimation radius is indicated in blue. This diagram is not to scale.

### 1.3.2 Molecular torus

The molecular torus, also called molecular disk, extends from  $\sim 10$  pc to  $\sim 100$  pc (Alonso-Herrero et al. 2021). It is much colder than the accretion disk or the polar cones. Since the dust emission is proportional to a high power of the temperature ( $\sim 4$  to  $6$ ), the molecular disk is easier to detect by observing molecular emissions.

The recent advancements in observational capabilities, such as the Atacama Large Millimeter/submillimeter Array (ALMA), have enabled the detection and mapping of molecular disks with high sensitivity and interferometric resolution in the millimeter and submillimeter wavelength range (Imanishi et al. 2018; García-Burillo et al. 2019, 2021). This capability is particularly advantageous when targeting high rotational levels of the CO molecule, typically ranging from CO(3-2) up to CO(6-5). The wide bandwidth of ALMA allows for the simultaneous detection of other dense gas tracers, such as HCO<sup>+</sup> and HCN. In the case of dense and compact nuclei, HCN serves as a valuable tracer for studying the properties of the molecular disks (Aalto et al. 2015).

The size of the molecular disk depends on the molecular tracer used, suggesting that its density varies strongly with the radius. In general, the molecular disks exhibit a relatively thin structure. Yet, they can play a role in the inner regions by actively participating in the molecular wind and contributing to the obscuration needed by the unification model (Audibert et al. 2019).

To date, the observed molecular disks have provided valuable insights into their properties (Combes et al. 2019; García-Burillo et al. 2019, 2021). These tori exhibit random orientations, and their position angles and inclinations are independent of those of the larger galactic disks. This decoupling in both dynamics and morphology indicates that the molecular disk operates as a distinct entity within the galaxy (Combes et al. 2019; García-Burillo et al. 2021). Consequently, it elucidates the phenomenon of radio jets, which extend perpendicular to the accretion

disk and torus, frequently interacting with the extended gaseous disks of the galaxies. Such interactions enhance the connection between the jets and galaxies, facilitating AGN feedback in the radio mode. Moreover, in cases where the kinematics of the molecular disks remain undisturbed by counter-rotating gas or outflows, their dynamics can provide a more reliable means to estimate the  $M_{\text{BH}}$ .

### 1.3.3 Polar biconical outflows

High-resolution interferometric observations conducted using the Very Large Telescope Interferometer and the Keck Observatory on nearby Seyfert galaxies have provided evidence that the torus exhibits an elongated morphology along the perpendicular direction. This elongated shape can be attributed to the heating of dust particles by the central nucleus, which initiates from the sublimation radius. As a result, the heated dust particles join a wind that moves due to the radiation pressure pushing on them. This leads to the formation of a distinctive hollow polar cone structure (Antonucci & Miller 1985), as illustrated in Fig. 1.7.

These outflows are commonly detected through X-ray and UV absorption, as well as optical emission lines (Crenshaw et al. 2003). The central engine has the capability to eject gas to distances of a few hundred parsecs, even photo-ionizing it up to 20-30 kpc (Unger et al. 1987).

Low-luminosity nearby AGN often exhibit a relatively low Eddington ratio, which may limit their ability to generate powerful winds from the accretion disk, as discussed in Sect. 1.4.2. However, as outlined in Sect. 1.2.2, observational evidence still supports the presence of an obscuring structure to account for the high fraction of type 2 AGN, even when the doughnut torus is not directly observed. An alternative explanation proposes that the obscuration arises from dusty outflows, which have been observed in mid-IR images of polar dust (Asmus et al. 2016). IR radiation has a much larger radiation cross-section on dust compared to the Thomson cross-section on dust-free ionized gas. This implies that radiation-driven winds can still be launched beyond the sublimation radius. Radiative hydrodynamical simulations, such as those conducted by Venanzi et al. (2020), have been employed to quantify this phenomenon. Comparisons with nearby Seyferts observed within the Galaxy Activity, Torus, and Outflow Survey (GATOS) project show approximate agreement with predictions regarding the strength of the wind, which depends on both the Eddington ratio and the density column of the dust (Alonso-Herrero et al. 2021).

## 1.4 AGN Feedback

### 1.4.1 The need for feedback

The interconnected growth of SMBHs at the center and their host galaxies indicates that the energy released by the AGN influences the physical properties of the host galaxies in some manner. Further proof of how this was given by Baugh (2006). In his study, he compared the

observed mass function of galaxies to the one of simulated galaxies. The simulation he used has for principle that, in the  $\Lambda$ CDM (cold dark matter) cosmological model, galaxy formation happens in two steps. First, dark matter halos form by collapsing and becoming virialized. Then, just after their formation, the atomic hydrogen gas collapsed, forming stars and galaxies. Simulations based on this model can predict a mass function of matter inside the dark halos but do not take into account any interference from AGN. Then, using a calibrated mass-to-light ratio, he compared the mass function to the observed luminosity function of galaxies, which can be fitted as a Schechter function (Schechter 1976):

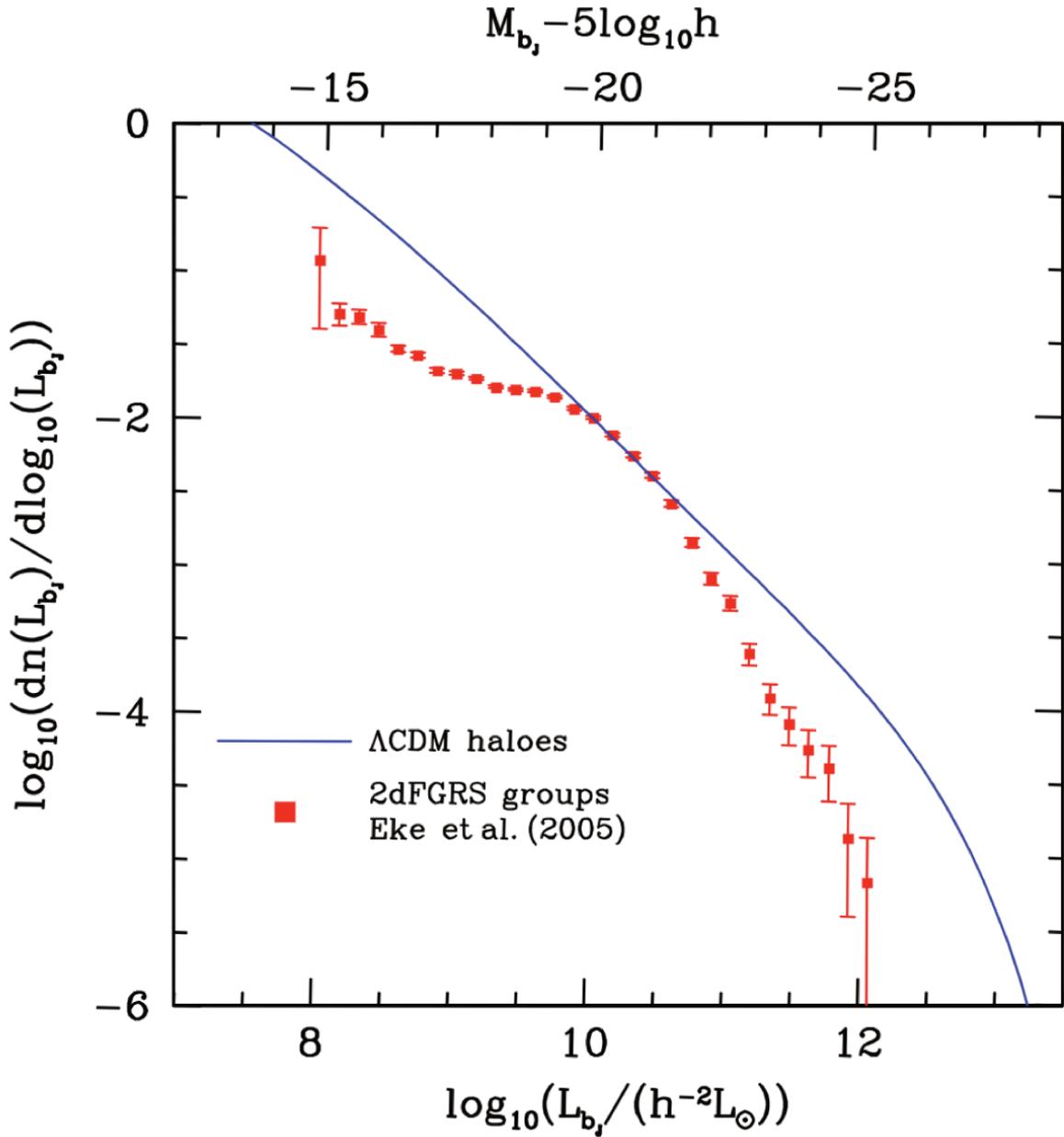
$$\frac{dn}{dL} = \phi_* \left( \frac{L}{L_*} \right) e^{-L/L_*} \quad (1.4)$$

where  $L_*$  is a characteristic galaxy luminosity and  $\phi_*$  is the normalization.

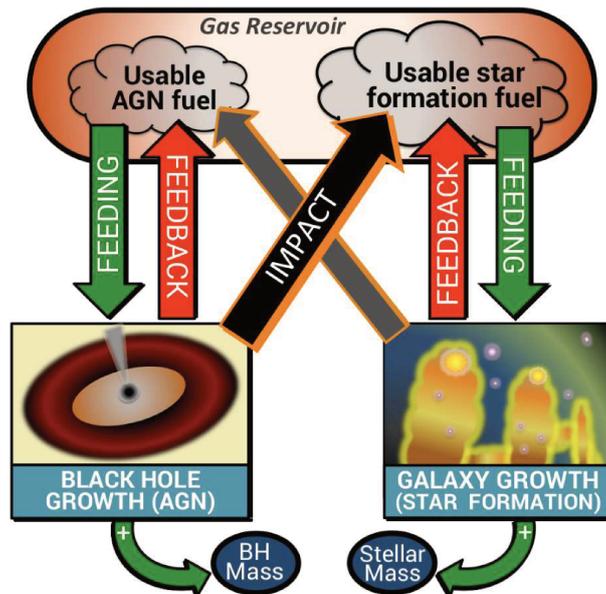
By comparing the results from the simulations and the observations (see Fig. 1.8), it appears that they are not in agreement for the low and high mass ends. In the simulations, baryons inside the dark halos have formed too many stars or galaxies to fit the observations. This result implies that some phenomenon must impact the formation of stars/galaxies.

In low-mass galaxies, the differences observed between simulations and observations can be explained by the influence of supernova explosions and stellar winds. These mechanisms can expel a considerable fraction of the gas from the galaxy's gravitational potential, leading to a reduction in the rate of star formation. This process is commonly known as stellar feedback. In the case of more massive galaxies, stellar feedback alone is insufficient to explain the discrepancies between simulations and observations. Instead, a more potent form of feedback, known as AGN feedback, is necessary. The outflows and winds generated by the AGN can expel a significant amount of gas from the galaxy, thereby reducing the available fuel for both star formation and black hole accretion. Additionally, AGN feedback can heat the gas, preventing its collapse and thereby reducing the overall star formation rate of the galaxy.

These phenomena are called feedback mechanisms because they are self-regulating (Fig. 1.9). Both AGN and star formation rely on cold gas derived from the same gas reservoir within the galaxy's halo. This reservoir gets its supply from gas-rich mergers, recycled material within the galaxy, and gas accretion from the intergalactic medium. The volume and cooling efficiency of this gas determine the available fuel for powering both SMBH growth and star formation. In the case of fueling SMBH growth, the material encounters an additional obstacle in releasing sufficient angular momentum to reach the galaxy's inner region. Both processes emit energy and momentum, affecting the gas availability through ionization, heating, shocking, or expulsion, forming self-regulating feedback mechanisms (Behroozi et al. 2013). These two processes can either positively or negatively affect the available fuel supply for the other process.



**Figure 1.8:** Comparison of the total luminosity function of galaxies between the prediction of  $\Lambda\mathrm{CDM}$  simulations (blue line) and the observations (red squares, Eke et al. (2006)). The simulations are made assuming a mass-to-light ratio of 80%. Figure from Baugh (2006).



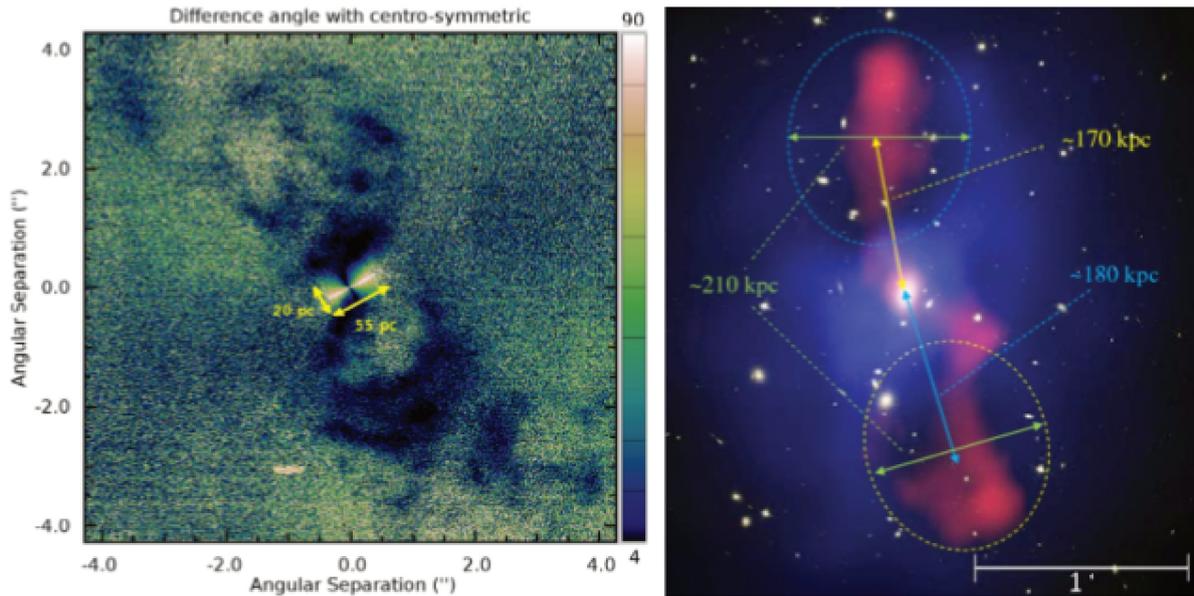
**Figure 1.9:** Illustration of the relationships between the SMBH growth, the star formation and the gas tank available in a galaxy. Figure from [Harrison \(2017\)](#).

## 1.4.2 The two modes of AGN feedback

Two distinct modes of AGN feedback can be distinguished based on the luminosity of the AGN: the radiative mode (also called quasar mode), and the kinetic mode (also known as radio mode) (Fig.1.10).

The quasar mode is characterized by exceptionally high AGN luminosity. In this regime, when a quasar approaches its Eddington limit, the AGN derives its power from a thin accretion disk. The elevated accretion rate associated with such intense luminosity enables the heating and radiation pressure to vaporize the wind emanating from the accretion disk. Eddington's luminosity is attained when radiation pressure balances gravity, causing ionized gas to be hindered from free flow, resulting in a portion of it escaping the galaxy in a high-velocity wind ([Leftley et al. 2019](#); [Venanzi et al. 2020](#)). The coupling between gas and dust in the AGN, facilitated by collisions, allows dust to push the gas, potentially contributing to the establishment of scaling relations ([Fabian 2012](#)). For effective feedback effects, the quasar mode necessitates high densities of relatively cold gas in the vicinity of the black hole, such as large-scale inflows induced by galaxy mergers. In scenarios where the coupling is insufficient, the feedback exerts a negligible impact on the galaxy ([Gabor & Bournaud 2014](#)).

The radio mode occurs in low luminosity AGN when Eddington's ratio is below 0.01. Here, the feedback mechanism is initiated by the pressure and momentum of the relativistic jet, as well as the shocks induced by the interaction between the jet and the surrounding gas. The energy released by the jet inflates a bubble or cocoon, which heats the gas and drives the feedback process. Approximately 10% of all AGN exhibit observable radio jets and are frequently observed in radio galaxies at the center of clusters, where they can stop or moderate the cluster cooling flows. In radio-loud quasars, the ratio between optical and radio brightness,  $R = L_{\text{rad}}/LB$  is a decreasing function of Eddington's ratio. Low luminosity AGN, with Eddington's ratio lower than 0.001, are usually in the ADAF regime and their disk has a small size but a large height.



**Figure 1.10:** Observations of the two kinds of AGN feedback. On the left, we find observations of molecular outflows in NGC 1068 (Gratadour et al. 2015). On the right, we show an image of the central region of the MS0735.6+7421 cluster showing X-ray emission in blue, I-band emission in white, and radio jets in red (McNamara et al. 2009).

Sądowski et al. (2013) showed that wind and jets can provide feedback in this regime.

### 1.4.3 Energy or momentum conservation

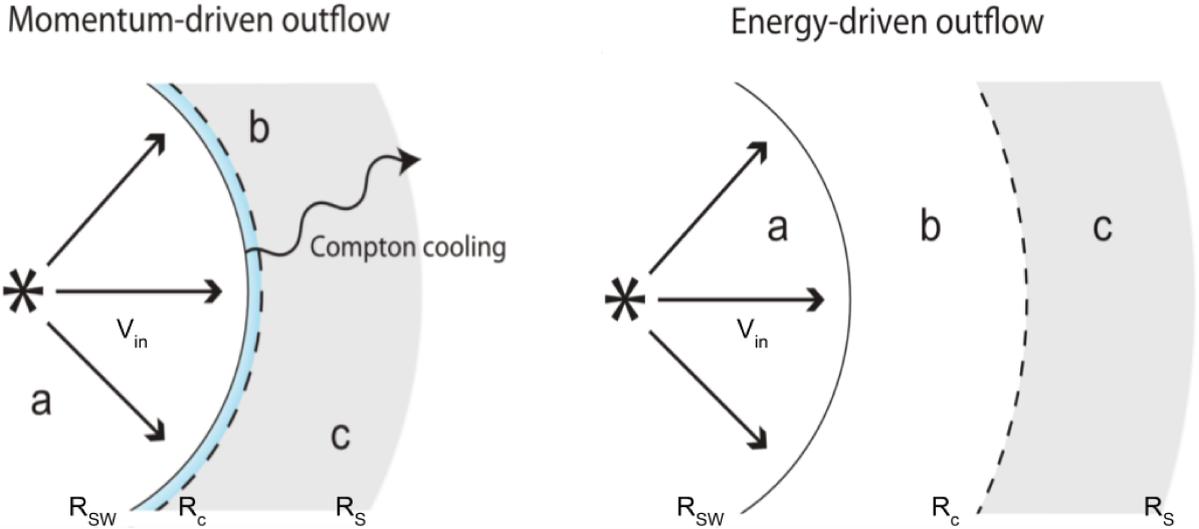
The physical processes able to provide energy and momentum to a gas wind or outflow are numerous (Veilleux et al. 2020). Taking into account all of these phenomena would require sophisticated numerical simulations. However, simple considerations suffice to distinguish the different regimes of the flows. By taking the wind as spherical, it is possible to describe the characteristic shock wave as the one used in supernova models (Weaver 1976; Fukue 2019).

Wind ejected inside the galactic nucleus can reach speeds of up to 10% of the speed of light (Tombesi et al. 2010). The highly supersonic nature of the gas wind compels it to generate a robust shock within the interstellar medium, giving rise to a forward shock wave that penetrates the surrounding ambient interstellar medium, while concurrently experiencing substantial deceleration in an inner reverse shock directed towards the BH. This results in the categorization of three distinct zones in the flow pattern, organized by increasing distance from the AGN (see Fig. 1.11): the unshocked highly supersonic inner wind (zone a); the shocked inner wind material (zone b), having traversed the reverse shock; and a shell composed of interstellar gas, swept up by the forward shock (zone c).

If the forward shock is sufficiently strong, the influence of pressure from the ambient interstellar medium, located just beyond zone c, can be disregarded. In such cases, the dynamics of the shell of swept-up gas (zone c) are entirely dictated by the disparity between the outward pressure force exerted by the shocked inner wind (zone b) and the gravitational inward pull acting on the shell. The latter is contingent on the total mass encompassed by the shell's radius. The thermal energy present in the shocked inner wind may undergo either transfer to the interstellar medium or dissipation through radiative cooling. The subsequent flow pattern's dynamics and the conservation of energy or momentum within the outflow hinge critically on the fate of the thermal energy contained in the hot shocked inner wind. It was first assumed by Silk & Rees (1998) that this cooling mechanism was an adiabatic expansion (energy-conserving). Later, it was argued that the gas is cooled by inverse Compton from the quasar radiation (King 2003; Reeves & Braito 2019) and so it is only a momentum-conserving mechanism.

If the reverse shock efficiently undergoes cooling, the gas in the shocked inner wind within zone b experiences a radiative region until its temperature reverts to its pre-shock value. This process, involving an initially adiabatic shock followed by a cooling region, constitutes an isothermal shock. Due to the effectiveness of cooling, the size of the region of the order of the cooling ( $R_c - R_{sw}$ ) is considerably smaller than the radius at which its temperature returns to its pre-shock value ( $R_c$ ), resulting in a very thin zone b. The post-shock gas subsequently propels an outer shock into the unperturbed interstellar medium, exerting pressure equivalent to the ram-pressure of the pre-shock inner wind. In this scenario, the inner wind can be conceptualized as colliding directly with the interstellar medium, fully transferring its momentum in the interaction. This configuration is termed momentum-driven outflow, as the driving force equals the momentum flux of the inner wind (Costa et al. 2014).

On the contrary, if the reverse shock fails to radiate away its thermal energy, the shell is propelled by the adiabatic expansion of the hot shocked wind bubble. In situations where the entire energy of the inner wind is conserved, the resulting outflow pattern is labeled energy-driven. In this context, a significant portion of the AGN-produced energy is dedicated to transferring



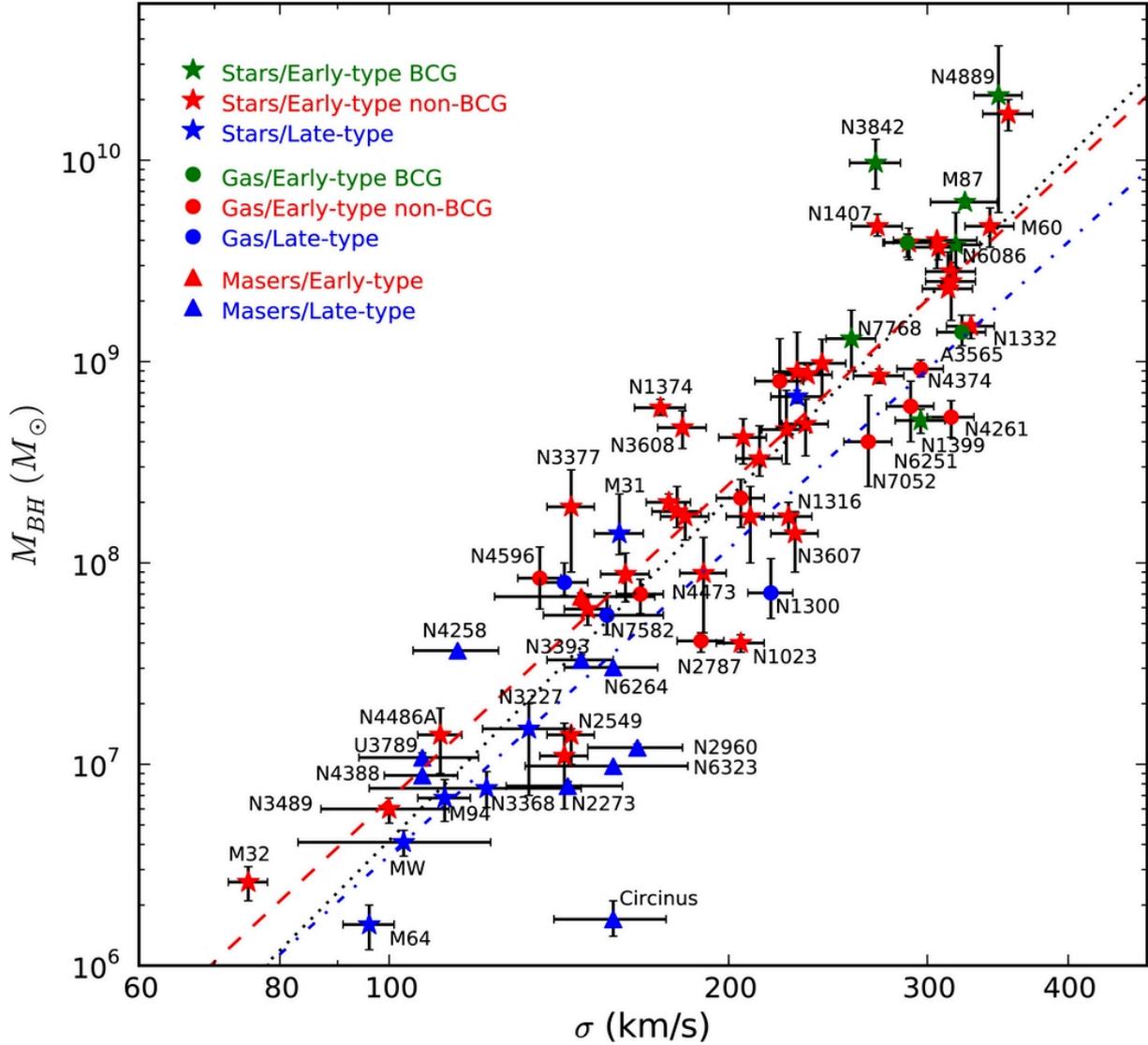
**Figure 1.11:** Schematic view of different phases around a quasar mode AGN momentum-driven outflows (on the left) and energy-driven ones (on the right).  $R_{sw}$  is the reverse shock towards the relativistic wind,  $R_c$  is the discontinuity surface and  $R_s$  is the shock in the interstellar medium. Zone "a" is characterized by the initial speed  $V_{in}$ , then zone "b" of the shock wind, and zone "c" represents the shocked ambient medium. Reproduced from [Costa et al. \(2014\)](#)

momentum to the surrounding medium. Consequently, the shell of the shocked-wind region in the ambient medium is more extended, resulting in a interstellar environment more impacted compared to the momentum-driven regime ([Costa et al. 2014](#)).

## 1.5 Co-Evolution of Galaxies and SMBH

The correlations between the SMBH's mass and their host galaxies suggest that the two evolve in symbiosis. [Magorrian et al. \(1998\)](#) discovered a linear relationship between the mass of the SMBH and the mass of the galactic bulge ( $M_{bulge}$ ), where  $M_{BH} \approx 0.001 M_{bulge}$ . This discovery led to the identification of an even more tightly correlated relationship between  $M_{BH}$  and the central velocity dispersion (see Fig. 1.12, [Gebhardt et al. \(2000\)](#); [Ferrarese & Merritt \(2000\)](#); [Gültekin et al. \(2009\)](#); [McConnell & Ma \(2013\)](#)). An empirical relationship between luminosity and velocity dispersion, with the same power-law exponent, has been observed by [Faber & Jackson \(1976\)](#) and [Falcón-Barroso et al. \(2011\)](#), which can account for the observed  $M_{BH} - \sigma$  relation. The reason why the  $M_{BH}$  relation is more tightly correlated with velocity dispersion than with bulge mass or luminosity is not yet fully understood. The  $M_{BH} - \sigma$  relation suggests that the growth of the BH and the ones of the bulge are connected. While the  $M_{BH}$  correlates with the mass of the bulge, it is not the case with a pseudo-bulge.

[Madau & Dickinson \(2014\)](#) conducted a study on the star formation history throughout cosmic time. Their findings indicate that the peak in Star Formation Rate (SFR) occurred between redshifts of 2 and 1.5. Subsequent to this peak, from  $z = 2$  to the present day, the gradual evolu-



**Figure 1.12:**  $M_{\text{BH}}-\sigma$  relation for a sample of 72 galaxies. Each galaxy type is represented by a distinct color: green for Brightest Cluster Galaxies (BCGs), red for other elliptical and S0 galaxies, and blue for late-type spiral galaxies. The symbols used to indicate the measurement methods are as follows: stars for  $M_{\text{BH}}$  measurements obtained through stellar dynamics, triangles for measurements derived from maser dynamics, and circles for measurements based on gas dynamics. The color-coded power law fits indicate the corresponding relationships for galaxies of the same color. Figure from [McConnell & Ma \(2013\)](#)

tion of main sequence galaxies suggests that the evolution of star formation is mainly regulated by a balance between gas accretion and feedback mechanisms. By comparing the star formation history with BH accretion history from X-ray and IR data, it was discovered that the growth rate of BHs closely mirrors the cosmic star formation history, implying that the two processes are closely linked.

There is, however, still a number of related open issues. For example, it appears that barred galaxies have lower  $M_{\text{BH}}$  than unbarred ones (Graham 2008; Brown et al. 2013). Furthermore, local ellipticals with over-massive SMBH has been observed (Kormendy et al. 1997; van den Bosch et al. 2012; Savorgnan & Graham 2016; Dullo et al. 2021). These over-massive SMBH preferentially occur in galaxy clusters and in brightest cluster galaxies (McConnell & Ma 2013), in particular where it is believed that the environmental effects strip the galaxies from their gas and stop star formation and the growth of bulges. These galaxies are then called massive relics and have particularly old stellar populations (Trujillo et al. 2014; Martín-Navarro et al. 2015; Ferré-Mateu et al. 2015, 2017). Another issue is the very discovery of massive SMBH ( $M_{\text{BH}} \gtrsim 10^9 M_{\odot}$ ) in bright quasars at the epoch of reionization (e.g., Bañados et al. 2018; Farina et al. 2022). These observations are puzzling as they show that extreme SMBH can form within 1 Gyr after the Big Bang. The rapid formation of these high- $z$  SMBH might be explained by invoking some extreme scenarios such as the growth of a  $10^{2-5} M_{\odot}$  seed via super-Eddington accretion (Valiante et al. 2016b; Pezzulli et al. 2017), the direct collapse of an initial gas condensation into a BH of  $\sim 10^5 M_{\odot}$  (Visbal et al. 2014; Regan et al. 2017), or the merger of massive protogalaxies (e.g., Mayer et al. 2010, 2015; Ferrara et al. 2013; Bonoli et al. 2014).

## 1.6 Classical methods for estimating $M_{\text{BH}}$

As they are described in general relativity (Einstein 1916), BH are the simplest objects in the Universe. They are defined by only three quantities: their masses, their spins, and their charges. Thus, determining the  $M_{\text{BH}}$  furnishes us with one-third of the complete information about a BH. This, in conjunction with our prior observation regarding the correlated evolution of SMBH and their host galaxies, emphasizes the increased importance of acquiring accurate measurement of the SMBH mass. However, because BH does not emit or reflect light, it is impossible to directly observe them. Up until the EHT pictures of SMBH shadows (The Event Horizon Telescope Collaboration et al. 2019, 2022), SMBH were detected via their gravitational effects on nearby systems. We present in this section several commonly used methods to estimate the  $M_{\text{BH}}$ .

### 1.6.1 Dynamical estimation

The estimation of SMBH mass through dynamical models involves studying the movements of celestial objects, like stars or gas, around the SMBH. These observed motions, often referred to as tracers, serve as constraints for the dynamical models. These models discern the con-

tribution of the BH to the potential from that of the surrounding galaxy, thus facilitating the inference of the SMBH mass by fitting these models to observations (Kormendy & Richstone 1995; Ferrarese & Merritt 2000; Kormendy 2004; Gebhardt et al. 2000; Kormendy & Ho 2013).

A notable constraint of this methodology is its reliance on resolving the kinematics of objects situated near the SMBH. Consequently, it becomes impractical for galaxies located at considerable distances. Therefore, alternative methodologies are required to estimate SMBH masses in such distant galaxies.

Across various dynamical models, an essential consideration is the necessity for a realistic representation of the gravitational potential of the galaxy. Additionally, the data characterizing the tracers motions must be robust enough to detect the presence of a BH (Gebhardt et al. 2003; Ferrarese & Ford 2005; Kormendy & Ho 2013). Among the multitude of intricate processes connected to SMBHs and their immediate surroundings, the study of stellar phenomena proves more amenable to analysis and modeling. This feasibility arises from stars often being treatable as point masses, predominantly influenced by gravity. The substantial mass disparity between stars and  $M_{\text{BH}}$  simplifies the problem further, allowing stars to be considered test masses on intermediate scales. Here, they operate within the  $M_{\text{BH}}$  potential, where gravitational interactions among stars are negligible, yet they are distant enough to avoid significant tidal effects from the  $M_{\text{BH}}$ .

The molecular gas surrounding the SMBH serves as another dynamical tracer for estimating the  $M_{\text{BH}}$ . Gas-dynamical mass measurements present distinct advantages compared to stellar-dynamical modeling. The simplicity inherent in gas-dynamical modeling stems from the gas Keplerian rotation within a dynamically cold disk. Unlike the intricate orbit-based computations necessary for treating stars, analyzing gas dynamics in a rotating disk proves computationally less demanding. The reduced complexity allows neglecting factors such as orbital anisotropy, triaxiality, or the influence of the dark matter halo. The gas-dynamical approach assumes a thin, rotating disk following circular orbits within the principal plane of the galaxy potential (Marconi et al. 2001; Marconi & Hunt 2003). This methodology aims to calculate a model velocity field aligning with observed velocities, velocity dispersions, and the line emission's surface brightness distribution. The gravitational potential of the galaxy comprises contributions from stars, measured by the projected stellar surface brightness, and assumes a mass-to-light ratio along with the presence of a BH (Melchior & Combes 2011; Combes et al. 2019).

However, gas dynamics introduce several complications (Kormendy & Ho 2013). Properly sampling the BH Sphere of Influence (SoI) necessitates distributing gas across radii. Simultaneously, the gas kinematics must display sufficient orderliness for interpretation. Furthermore, gas differs from stars as it behaves like a collisional fluid, responsive to non-gravitational disturbances such as turbulence, shocks, radiation pressure, and magnetic fields. Another challenge in gas dynamics involves dust absorption potentially rendering the gas distribution opaque. Consequently, assumptions about observing through the gas in projection may not hold under such circumstances. Each galaxy requires meticulous examination to confirm whether the gas has reached an equilibrium configuration, influenced mainly by gravitational effects, considering the opacity caused by dust absorption.

## 1.6.2 $M_{\text{BH}} - \sigma$ relation

The tight correlation between BH and bulge masses drawn from the  $M_{\text{BH}} - \sigma$  relation (e.g. Ferrarese & Merritt 2000; Gebhardt et al. 2000) suggests that there is a mechanical and/or radiative feedback from the AGN on the star formation in galaxies, all along their growth through accretion and mergers. The origin of this coupling is not yet fully understood but there are analytic and numerical models created to constrain the physical mechanism and reproduce the  $M_{\text{BH}} - \sigma$  relation (e.g. Silk & Rees 1998; Granato et al. 2004). Nevertheless, the observational  $M_{\text{BH}} - \sigma$  relation is not well constrained for low-mass galaxies ( $M < 10^9 M_{\odot}$ ). The tight correlation is mostly derived from biased galaxy samples with a high  $M_{\text{BH}}$ , which have a larger SoI easier to resolve with common angular resolutions (e.g. Silk & Rees 1998; Costa et al. 2014). In this work we used the Kormendy & Ho (2013) correction of the  $M_{\text{BH}} - \sigma$  relation:

$$\log M_{\text{BH}}(M_{\odot}) = 8.5 \pm 0.05 + (4.41 \pm 0.29) \log \left( \frac{\sigma}{200 \text{km/s}} \right), \quad (1.5)$$

where  $\sigma$  is the stellar central velocity dispersion. For our sample, it was taken from the Hyperleda compilation (e.g. Makarov et al. 2014). This relation has an intrinsic error of 0.3 dex.

## 1.6.3 Reverberation mapping

BLRs have a radius of light days to light weeks radii which makes them too small to be spatially resolved. However, they can be resolved temporarily with the Reverberation Mapping technique (Blandford & Payne 1982; Bahcall et al. 1972; Lyutyi & Cherepashchuk 1972; Cherepashchuk & Lyutyi 1973). The SMBH accretion disk emits UV radiation that heats the BLR gas clouds, ionizing them and creating emission lines. Thus, when the UV continuum source varies, the emission lines vary too ('reverberate') but with a delay of  $R_{\text{BLR}}/c$ . Therefore, the BLR radius can be estimated by measuring the time delay between the change in emission from the central source and its impact on the BLR. Then, assuming that the width of the emission lines of the BLR is due to the gravitational influence of the central SMBH, we can derive the following relation from the Virial theorem:

$$M_{\text{BH}} = \frac{f (\Delta V^2) R}{G} \quad (1.6)$$

where  $M_{\text{BH}}$  is the mass of the central BH,  $G$  is the gravitational constant,  $\Delta V$  is the virial velocity of the BLR,  $R$  is the BLR radius and  $f$  is a factor depending on the BLR geometry. The fact that the BLR velocity is dominated by the BH was validated by Gaskell (1988).

Most of the uncertainties of this method come from the factor  $f$ . Kollatschny & Zetzl (2011) observed that the BLR structure changes with the emission-line turbulent dispersion and rotational width, implying that  $f$  depends on these two factors. The findings of Gaskell (2008, 2011) indicate another inherent constraint, which suggests that the continua of AGN can exhibit localized flares in off-center regions situated relatively close to the inner boundary of the BLR radius.

Even in the absence of the expensive and time-consuming monitoring observations, it is

possible to use this method, thanks to the AGN-Luminosity vs BLR-size scaling relation, calibrated from reverberation mapping measurements (Wandel et al. 1999). Refining the method with a large reverberation mapping database, and comparing it to the other methods, Peterson et al. (2004) find that the precision of the method is  $\sim 30\%$ , comparable to the gas or stellar kinematical method. This difference can be explained by considering that the reverberation mapping technique is notably influenced by the inclination angle of the BLR region, in contrast to the gravitational method. As demonstrated in the study by Liu et al. (2017), the substantial disparity between gravitational and virial mass estimates can be reconciled by adjusting the parameter  $f$  to a value between 8 and 16 for the virial measurement. This adjustment of  $f$  is contingent upon the inclination of the BLR. The conventional value of  $f \sim 1$  corresponds to an inclination angle of around 30 degrees.

## 1.6.4 Single epoch method

The single epoch method uses the relation between the BLR size and the AGN optical/UV continuum luminosity empirically found from reverberation mapping (Peterson et al. 2004; Kaspi et al. 2007; Bentz et al. 2009), as well as the tight correlation between the continuum luminosity and that of broad emission lines (e.g., Shen et al. 2011). With these considerations, the  $M_{\text{BH}}$  can be expressed as

$$\log\left(\frac{M_{\text{BH}}}{M_{\odot}}\right) = a + b \log\left(\frac{L}{10^{44} \text{ erg s}^{-1}}\right) + c \log\left(\frac{\text{FWHM}}{\text{km s}^{-1}}\right), \quad (1.7)$$

where the coefficients  $a$ ,  $b$ , and  $c$  are empirically calibrated against local AGN with reverberation mapping masses or using different lines.  $L$  and FWHM are the line luminosity and width. At low redshift, below 0.75, the  $\text{H}\beta$  is the most commonly used line. For redshift between 0.75 and 2, the  $\text{Mg II } \lambda 2800$  line is a common choice (McLure & Jarvis 2002). And for  $z \gtrsim 2$ , the best line to use is  $\text{C IV } \lambda 1549$  Vestergaard (2002). Since their first estimation, the relation for these lines had been revised by Vestergaard & Peterson (2006); Wang et al. (2009); Shen et al. (2011); Xiao et al. (2011) and others.

The single epoch method has the advantage of being inexpensive in telescope times since it can estimate the SMBH mass from just a spectrum. However, it is a more indirect method than reverberation mapping and does not contain as many indicators that the measured velocities are virialized ones. This method has an uncertainty of  $\sim 0.3$  dex.

## 1.6.5 Fundamental Plane of BH Activity

The Fundamental Plane of BH activity (FPBH) is a strong correlation between the  $M_{\text{BH}}$ , its 5 GHz radio continuum luminosity, and its 2-10 keV X-ray power-law continuum (Merloni et al. 2003). When accretion disks go through episodes of low-luminosity advection-dominated accretion flow, they are observed to be in a hard state (hard X-ray radiation). In this state, they are often accompanied by relativistic jets (Gallo et al. 2003; Narayan 2005). During this state of inefficient radiation, the accreting material stops flowing in a disk at some distance from

the accreting object. This new configuration results in the creation of a hot, optically-thin gas corona around the BH. Photons, from the now truncated disk, are Compton-upscattered to tens to hundreds of keV when they traverse this corona. Furthermore, the synchrotron mechanism in the relativistic jets produces emission primarily in the radio band (Blandford 1984). The entire broadband combined emissions of the accretion disk, corona, and relativistic jets range from the radio to the X-ray bands. This emission can also be dominated by the jet itself, particularly if accretion rates are higher than  $\sim 0.01\dot{M}_{\text{Edd}}$  (Falcke et al. 2004). A fainter thermal component emanating from the shortened accretion disk may also be identifiable alongside the primary non-thermal corona emission. The jet is linked to the accretion process and it has been shown that the luminosity of the X-ray and radio emissions are correlated and the disk-jet mechanism is independent of the  $M_{\text{BH}}$ . By considering these arguments, Merloni et al. (2003) and Falcke et al. (2004) probed a large sample of galactic BH and SMBH and found a strong correlation between the X-ray luminosity (between 2 et 10 keV), the radio luminosity (5 GHz) and the  $M_{\text{BH}}$ . This correlation is called the fundamental plane of BH activity. According to Gültekin et al. (2019) the  $M_{\text{BH}}$  can be estimated with the relation:

$$\log\left(\frac{M_{\text{BH}}}{10^8 M_{\odot}}\right) = 0.55 \pm 0.22 + (0.19 \pm 0.10)\log\left(\frac{L_R}{10^{38}\text{erg/s}}\right) + (-0.59^{+0.16}_{-0.15})\log\left(\frac{L_x}{10^{40}\text{erg/s}}\right), \quad (1.8)$$

where  $L_x$  is the 2-10 keV X-ray power-law continuum and  $L_5$  the 5 GHz radio luminosity calculated as  $L_R = \nu L_{\nu} = (\nu = 5 \text{ GHz}) \times 4\pi D^2 F_5$ ,  $F_5$  being the 5 GHz radio flux density. This relation has an intrinsic error of  $\sim 0.40$  dex.

This FPBH scaling relation is a useful method to distinguish between X-ray binaries, intermediate mass BH, and SMBH, and to determine the  $M_{\text{BH}}$  in a Type 2 AGN, or in a host galaxy with disturbed morphology, where the  $M_{\text{BH}}\text{-}\sigma$  method is unusable. The only requirement to estimate the  $M_{\text{BH}}$  is to have X-ray and radio data of sufficient quality to resolve the central source and also confirm its spectral characteristics (i.e. that it is indeed in the hard state). The data must have a high angular resolution to avoid contamination from other sources and the host. This is especially important to distinguish low luminosity AGN from X-ray binaries.

# CHAPTER 2

---

## SMBH AND GALAXY CO-EVOLUTION IN RLAGN AT $z \sim 0.3 - 4$

---

This chapter outlines my research focusing on the co-evolution of galaxies and central SMBH within a sample of RLAGN. The results are published in [Poitevineau et al. \(2023\)](#). The research presented here represents the first sample of distant RLAGN with both robust  $M_\star$  and  $M_{\text{BH}}$  estimates and acts as a benchmark for future studies with a larger sample, more statistically representative of the RLAGN population characteristics.

I begin by providing the motivation behind this study and explaining the rationale for investigating this topic. Next, I detail the methodology employed to obtain the RLAGN sample I studied, including the collection of multiwavelength data and the characterization of their properties. I then describe the process of estimating properties such as  $M_{\text{BH}}$ , jet characteristics, accretion rates, and stellar properties. Furthermore, I discuss the inclusion of different samples from existing literature, which serve as a basis for comparison with our findings. Finally, I present our results by examining the scaling relations documented in previous studies, shedding light on the implications of our research.

In this work, the first part of the RLAGN sample selection (from Sect. 2.2 to 2.2.3) as well as the SED fitting and the stellar mass ( $M_\star$ ) estimation was made by Dr. Gianluca Castignani.

### Contents

---

<b>2.1</b>	<b>Motivations</b> . . . . .	<b>30</b>
<b>2.2</b>	<b>RLAGN sample</b> . . . . .	<b>31</b>
2.2.1	Dark Energy Survey . . . . .	31
2.2.2	Radio, optical, and spectroscopic selection . . . . .	31
2.2.3	IR selection: WISE . . . . .	32

2.2.4	Broad emission lines from SDSS . . . . .	33
2.2.5	RLAGN in GAMA DR3 . . . . .	34
2.2.6	Radio and IR properties . . . . .	35
<b>2.3</b>	<b>BH, jet, accretion, and stellar properties . . . . .</b>	<b>36</b>
2.3.1	BH masses . . . . .	36
2.3.2	Jet power . . . . .	36
2.3.3	Eddington ratio . . . . .	37
2.3.4	SED modeling . . . . .	38
2.3.5	SFR versus $M_{\star}$ . . . . .	40
<b>2.4</b>	<b>Comparison sample . . . . .</b>	<b>42</b>
<b>2.5</b>	<b>Results . . . . .</b>	<b>44</b>
2.5.1	$M_{\text{BH}} - M_{\star}$ relation . . . . .	44
2.5.2	Jet power, BH mass, and accretion . . . . .	47
2.5.3	RLAGN and their environments . . . . .	51
<b>2.6</b>	<b>Discussion and conclusions . . . . .</b>	<b>52</b>

## 2.1 Motivations

Altogether, while existing studies show tight coevolution of SMBH, AGN, and their host galaxies with cosmic time, this interplay is still debated and unconstrained. This is at least partially due to the difficulty of building large samples of distant AGN with well-characterized stellar and BH properties and there are still a number of related open issues. One example of these issues are the local ellipticals with overmassive SMBH (Kormendy et al. 1997; van den Bosch et al. 2012; Savorgnan & Graham 2016; Dullo et al. 2021). These overmassive SMBH preferentially occur in galaxy clusters (McConnell & Ma 2013), in particular where environment effects strip the galaxies from their gas and stop star formation and the growth of bulges. Galaxies are then called massive relics and have particularly old stellar populations (Trujillo et al. 2014; Martín-Navarro et al. 2015; Ferré-Mateu et al. 2015, 2017).

Another example is the very discovery of massive SMBH ( $M_{\text{BH}} \gtrsim 10^9 M_{\odot}$ ) in bright quasars at the epoch of reionization (e.g., Bañados et al. 2018; Farina et al. 2022) is a mystery, as it shows that extreme SMBH can form within 1 Gyr after the Big Bang. The rapid formation of these high- $z$  SMBH might be explained by invoking some extreme scenarios such as the growth of a  $10^{2-5} M_{\odot}$  seed via super-Eddington accretion (Valiante et al. 2016b; Pezzulli et al. 2017), the direct collapse of an initial gas condensation into a BH of  $\sim 10^5 M_{\odot}$  (Visbal et al. 2014; Regan et al. 2017), or the merger of massive protogalaxies (e.g., Mayer et al. 2010, 2015; Ferrara et al. 2013; Bonoli et al. 2014).

In order to better understand the growth of SMBH with cosmic time and their coevolution with their host galaxies, we have built a sample of distant RLAGN spanning about 9 Gyr of

cosmic time, between  $z \sim 0.3 - 4$ , with available radio to UV spectrophotometric data. Based on this multiwavelength dataset, I assess the properties of the AGN sample, for example, in terms of  $M_{\text{BH}}$  and  $M_{\star}$ , jet power, and Eddington ratio. As RLAGN are associated with the most massive BHs and host galaxies (e.g., Best et al. 2005; Chiaberge & Marconi 2011; Shen et al. 2011; Shaw et al. 2012), they are excellent sources in which to investigate the galaxy, AGN, and SMBH coevolution in the high-mass regime.

Throughout this work, we adopt a flat  $\Lambda$ CDM cosmology with matter density  $\Omega_{\text{m}} = 0.30$ , dark energy density  $\Omega_{\Lambda} = 0.70$ , and Hubble constant  $h = H_0/100 \text{ km s}^{-1} \text{ Mpc}^{-1} = 0.70$ .

## 2.2 RLAGN sample

We selected a sample of RLAGN by cross-matching the Very Large Array Faint Images of the Radio Sky at Twenty-centimeters (VLA FIRST) source catalog (Becker et al. 1995) with IR to optical spectrophotometric surveys. As further described in the following, the use of IR to UV photometry enables modeling the SED, which ultimately allows us to obtain a good characterization of the galaxy properties, such as the  $M_{\star}$  and the SFR.

### 2.2.1 Dark Energy Survey

We began by examining the Dark Energy Survey (DES) (DES; Collaboration 2005; Dark Energy Survey Collaboration et al. 2016), which is composed of two distinct multiband imaging surveys. The first is a wide-area grizY survey covering  $\sim 5000 \text{ deg}^2$ , while the second is a deep supernova griz survey consisting of six distinct deep fields (Hartley et al. 2021). The DES data release 2 (Abbott et al. 2021)<sup>1</sup> has recently made the coadded source catalog and images publicly available, encompassing six years of DES wide-area survey observations and five years of DES supernova survey observations.

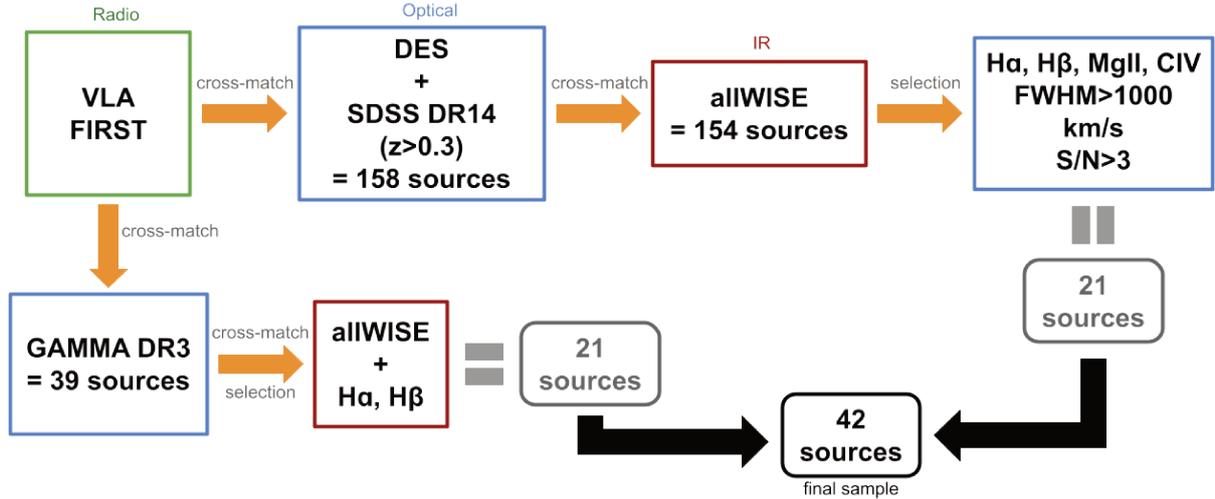
To construct our sample of distant RLAGN, we specifically focused on DES supernova fields located in the equatorial region that overlapped with the VLA FIRST survey conducted at low radio frequencies. The selection procedure closely followed that of Dr. Castignani's previous work (Castignani et al. 2019), which we refer to for additional details. However, his previous study only concentrated on two radio sources, exploring their molecular gas content, cluster environment, as well as stellar and star formation properties. In contrast, the present study considers a more extensive sample, as further elucidated in the subsequent sections.

### 2.2.2 Radio, optical, and spectroscopic selection

In order to construct a sample of extragalactic radio sources, our analysis incorporated data from the VLA FIRST survey (Becker et al. 1995), which observed  $10,000 \text{ deg}^2$  of the North and

---

<sup>1</sup><https://des.ncsa.illinois.edu/releases/dr2>



**Figure 2.1:** Illustration of the RLAGN sample selection process described in Sect.2.2.

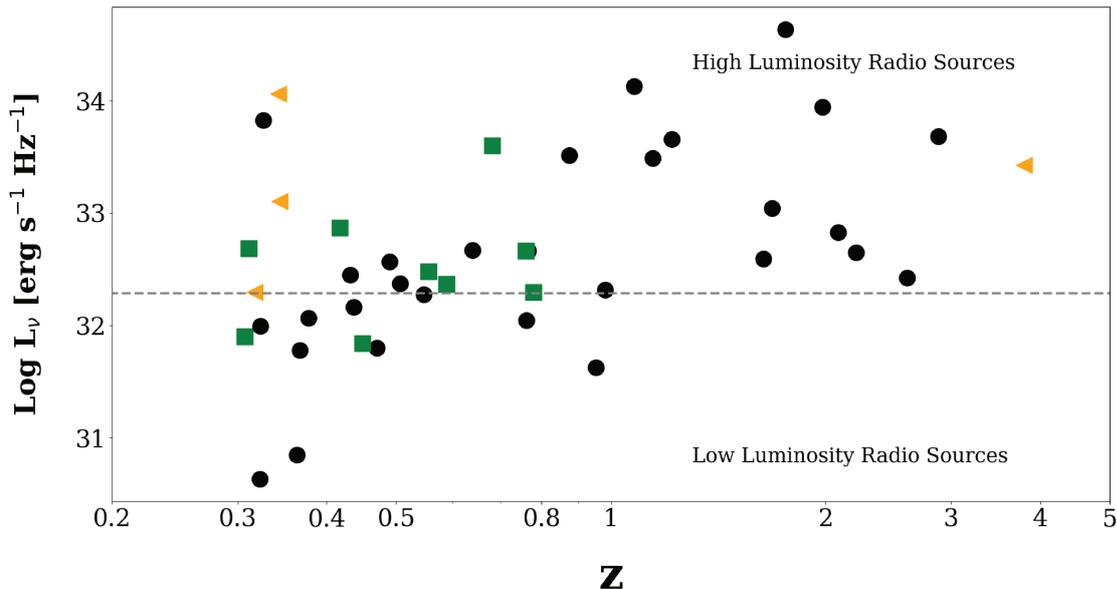
South Galactic Caps at low radio frequencies (1.4 GHz). The survey achieved a point source detection limit of approximately 1 mJy. Our focus was specifically on the Stripe 82 region, covering an equatorial field of  $300 \text{ deg}^2$  that was extensively imaged by the Sloan Digital Sky Survey (SDSS) and overlapped with the VLA FIRST survey. Additionally, we considered DES supernova deep fields numbered 2, 3, and 5 for our search, as outlined in Sect. 2.2.1.

To establish associations between the VLA FIRST radio source catalog and the SDSS DR14 ugriz and DES DR1 grizY source catalogs, a cross-matching process was performed within the designated fields using a search radius of 3 arcsec. This radius was consistent with the positional accuracy of  $\sim 1$  arcsec for FIRST sources. To ensure the selection of secure distant radio sources, our criteria further required the presence of SDSS DR14 spectroscopic redshifts  $z > 0.3$ . Through this search, we successfully identified 158 radio sources with confirmed spectroscopic redshifts, each having unique optical counterparts from both the SDSS and DES datasets.

### 2.2.3 IR selection: WISE

To investigate the presence of IR emission associated with the identified radio sources, we conducted a search using the W4 filter of the Wide-field Infrared Survey Explorer (WISE) (Wright et al. 2010). This was accomplished by cross-matching our radio sources with the allWISE source catalog<sup>2</sup>, employing a search radius of 6.5 arcsec. This search radius was consistent with previous studies on extragalactic radio sources (e.g., Castignani et al. 2013). As a result, we obtained 154 sources with unique counterparts in the WISE catalog, providing W4 magnitudes with a Signal-to-Noise ratio ( $S/N$ )  $> 1$ .

<sup>2</sup><http://wise2.ipac.caltech.edu/docs/release/allwise/>



**Figure 2.2:** Radio luminosity densities for the radio sources in our sample. 1.4 GHz Radio luminosity density as a function of redshift. Sources are distinguished between AGN (black circles), starbursts (orange triangles), and intermediate disks (green squares) according to the color-based classification by Stern et al. (2012) and Jarrett et al. (2017). The horizontal line is at  $L_\nu = 2 \times 10^{32} \text{ erg s}^{-1} \text{ Hz}^{-1}$  and separates low-luminosity radio sources from high-luminosity radio sources.

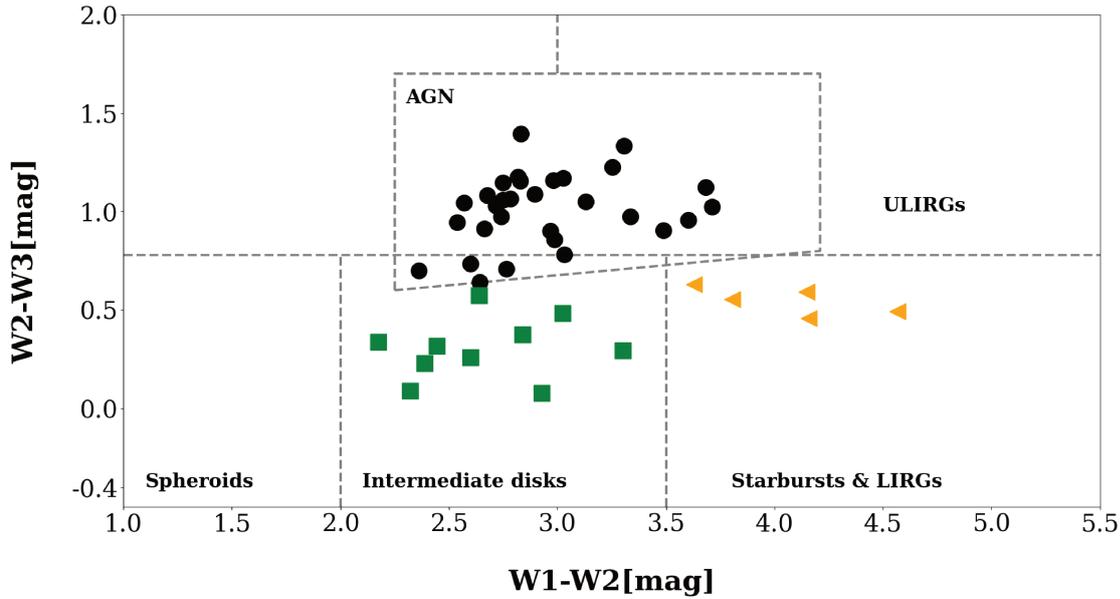
## 2.2.4 Broad emission lines from SDSS

In order to assess the  $M_{\text{BH}}$  of the selected RLAGN, I imposed additional criteria by focusing on sources that exhibit evidence of broad emission lines in their SDSS spectra. Specifically, I selected sources with  $H\alpha$ ,  $H\beta$ ,  $\text{Mg II}$ , or  $\text{C IV}$  emission line fluxes characterized by an  $\text{S/N} > 3$  and a FWHM exceeding  $1000 \text{ km s}^{-1}$ , which is typical for lines originating from the BLR. To identify these sources, I utilized the spZline file<sup>3</sup>, which provides the results of the emission-line fits for the SDSS sources observed by the Baryon Oscillation Spectroscopic Survey (BOSS) (Bolton et al. 2012). The Gaussian line width ( $\sigma$ ) is reported, and I converted it into the FWHM with the formula

$$\text{FWHM} = 2\sigma \sqrt{2\ln(2)} \quad (2.1)$$

This additional spectroscopic selection yielded 21 sources between  $z \sim 0.3 - 3.8$ .

<sup>3</sup>[https://data.sdss.org/datamodel/files/BOSS\\_SPECTRO\\_REDUX/RUN2D/PLATE4/RUN1D/spZline.html](https://data.sdss.org/datamodel/files/BOSS_SPECTRO_REDUX/RUN2D/PLATE4/RUN1D/spZline.html)



**Figure 2.3:** IR colors for the radio sources in our sample. WISE color-color plot. Color coding of the data points is the same as in the top panel. Black dashed lines are overlotted to distinguish the different types of objects (AGN, ULIRGs, spheroids, intermediate disks, starbursts, and LIRGs)

## 2.2.5 RLAGN in GAMA DR3

Similar to our selection criteria for the SDSS spectra described in the previous subsections, we initially identified 632 sources with  $z > 0.3$ , and with available third Data Release (DR) Galaxy And Mass Assembly (GAMA). GAMA DR3 (Baldry et al. 2018) comprises spectra obtained with the AAOmega multi-object spectrograph on the Anglo-Australian Telescope and offers an extensive dataset containing information from more than 200,000 sources. These sources exhibited  $H\alpha$  or  $H\beta$  emission line fluxes with an  $S/N > 3$  as well as line widths  $\text{FWHM} > 1000 \text{ km s}^{-1}$ , as inferred from single Gaussian modeling<sup>4</sup>. We restricted this selection to the equatorial region of  $180 \text{ deg}^2$  covered by both GAMA DR3 and the 1.4 GHz FIRST VLA survey. Among the 632 spectroscopic sources from GAMA, we further limited ourselves to a subsample of 39 galaxies with available 1.4 GHz fluxes from the FIRST VLA survey.

The GAMA DR3 database often includes multiple spectra for individual sources. We carefully examined each of the available spectra and discarded galaxies with uncertain evidence of emission lines, as their associated fits were deemed unreliable. This analysis yielded 21 GAMA DR3 radio sources at  $z \sim 0.3 - 0.8$ . We identified their unique counterparts in the WISE catalog using a search radius of 6.5 arcsec, as described in Sect. 2.2.3.

Combining these galaxies, labeled with the prefix G, with the 21 radio sources possessing SDSS spectra, denoted with the prefix RS, our final sample comprises 42 sources that I refer to hereafter in this study. The key properties of this final sample are listed in Table A.1.

<sup>4</sup><http://www.gama-survey.org/dr3/>

## 2.2.6 Radio and IR properties

I proceed to examine the low-frequency radio luminosities and the IR colors of our sources. To accomplish this, I adopt a similar approach as previous studies (Condon 1989; Chiaberge et al. 2009; Castignani et al. 2014) and assume a power-law for the radio spectrum, given by  $S_\nu \propto \nu^{-\alpha}$ , where  $S_\nu$  represents the radio flux density at the observer frequency  $\nu$ , and the spectral index  $\alpha$  is fixed at  $\alpha = 0.8$ . Subsequently, I convert the 1.4GHz VLA radio fluxes  $S_{1.4\text{GHz}}$  into rest-frame 1.4 GHz luminosity densities using the following formula:

$$L_{1.4\text{ GHz}} = 4\pi S_{1.4\text{ GHz}} D_L(z)^2 (1+z)^{\alpha-1}, \quad (2.2)$$

where  $D_L$  is the luminosity distance. Figure 2.2 displays our sources in the  $L_{1.4\text{ GHz}}$  versus redshift plane. They all have  $L_{1.4\text{ GHz}} \gtrsim 3 \times 10^{30} \text{ erg s}^{-1} \text{ Hz}^{-1}$ , which is typical of RLAGN, while purely starburst galaxies have lower  $L_{1.4\text{ GHz}} < 10^{30} \text{ erg s}^{-1} \text{ Hz}^{-1}$  (Chiaberge et al. 2009).

Moreover, the majority (71%, i.e., 30/42) of our sources exhibit high radio powers exceeding  $L_{1.4\text{GHz}} = 2 \times 10^{32} \text{ erg s}^{-1} \text{ Hz}^{-1}$ . This criterion is employed to distinguish between Low-Luminosity Radio Sources (LLR) and High-Luminosity Radio Sources (HLR), following the approach adopted in previous studies (e.g., Chiaberge et al. 2009; Castignani et al. 2014). It is worth noting that the radio galaxy population exhibits a bimodal distribution in terms of radio power, and the chosen LLR to HLR luminosity threshold corresponds to the fiducial radio power that effectively separates FRI from FRII radio galaxies (Fanaroff & Riley 1974; Zirbel 1996). Furthermore, it is important to consider the influence of the Malmquist bias associated with the VLA FIRST flux limit of  $\sim 1 \text{ mJy}$ , which results in an increased fraction of HLR at higher redshifts, reaching unity for  $z > 1$ .

Fig 2.3 presents the sources in our sample plotted on the WISE color-color diagram, with sources categorized according to the color-based classification by Jarrett et al. (2017), as depicted in the figure. Notably, our sample is observed to populate only three distinct regions within the diagram. The majority (28 out of 42) of our sources are classified as AGN based on their WISE colors. This outcome is anticipated since these sources were specifically selected as distant and high-powered radio sources at  $z > 0.3$ . Based on WISE colors, the remaining sources are reasonably distributed between the intermediate disk (9 out of 42) and starburst (5 out of 42) classes.

Furthermore, as shown in Fig. 2.2, the WISE IR colors of the vast majority of  $z > 1$  sources are consistent with AGN contribution. They also show high 1.4 GHz radio luminosities typical of Radio-Loud Quasars (QSOs). The majority (22 out of 42, i.e., 52%) of our sources are in fact classified as quasars in the NASA/IPAC Extragalactic Database (NED), for instance, with counterparts in the 2dF-SDSS LRG And QSO (2SLAQ; Croom et al. 2009) catalog, or with X-ray counterparts (XMM; Rosen et al. 2016), as outlined in Table A.1.

	$a$	$b$	$c$
H $\alpha$	1.24	0.43	2.1
H $\beta$	1.63	0.49	2.0
Mg II	1.70	0.63	2.0
C IV	1.52	0.46	2.0

**Table 2.1:** Coefficients for estimating  $M_{\text{BH}}$  using broad emission lines and Eq. 1.7. Values for the H $\beta$ , Mg II, and C IV lines are from [Shaw et al. \(2012\)](#). The coefficients for H $\alpha$  come from [Shen et al. \(2011\)](#)

## 2.3 BH, jet, accretion, and stellar properties

### 2.3.1 BH masses

One of the main goals of this work is to investigate the coevolution of central BH with the host galaxies in our RLAGN sample. To do this, I estimated  $M_{\text{BH}}$  using the widely used single epoch method (see Sect. 1.6.4), which is particularly suited for distant type 1 AGN. I used the coefficients obtained for the H $\alpha$ , H $\beta$ , Mg II, and C IV broad emission lines by [Shen et al. \(2011\)](#) and [Shaw et al. \(2012\)](#). These are widely used lines that are redshifted in the optical domain, depending on the redshift of the source. These lines indeed enable estimates of  $M_{\text{BH}}$  over a wide range of redshifts. Similarly to previous studies ([Shaw et al. 2012](#); [Castignani et al. 2013](#)), I used H $\alpha$ , H $\beta$ , and Mg II for sources at  $z < 1$  and the Mg II and C IV lines for sources at higher redshifts. When multiple broad emission lines were available for a given source, I adopted the following order of preference: H $\alpha$ , H $\beta$ , Mg II, and C IV (see, e.g., [Shen & Liu \(2012\)](#)). In Table 2.1 I report the coefficients used in this work, and in Table A.2 I list the  $M_{\text{BH}}$ .

### 2.3.2 Jet power

The sources in our sample are RLAGN which are typically characterized by jetted outflows (see Sect. 1.2.6 for details). By studying jet properties such as its total power, I investigate the complex interplay between the jet, the BH, and the gas accretion onto it, which is commonly referred to as radio-mode AGN feedback (see Sect. 1.4.2). Following the previous work of [Willott et al. \(1999\)](#), I estimate the jet power as:

$$Q_{\text{jet}} = 3 \times 10^{45} \xi^{3/2} \left( \frac{L_{151 \text{ MHz}}}{10^{35} \text{ erg s}^{-1} \text{ Hz}^{-1} \text{ sr}^{-1}} \right)^{6/7} \text{ erg s}^{-1}, \quad (2.3)$$

where  $L_{151 \text{ MHz}}$  is the extended total radio luminosity density at 151 MHz in the rest frame, and  $\xi$  ranges between 10 and 20. I used an intermediate value  $\xi = 15$ . To estimate  $L_{151 \text{ MHz}}$ , I extrapolated the  $L_{1.4 \text{ GHz}}$  luminosity densities assuming an isotropic emission and a power law with  $\alpha = 0.8$ , as further described in Sect. 2.2.6. The resulting jet powers are reported in

Table A.2.

### 2.3.3 Eddington ratio

Our objective is to establish a correlation between the gas accretion onto the BH and the properties of the AGN. To gauge the Eddington ratio, previously defined in eq.1.3, I previously determined the  $M_{\text{BH}}$  (see Section 2.3.1). Thus, the challenge arises when estimating the AGN bolometric luminosity. When estimating the AGN luminosity, the ideal method would have used the total bolometric luminosity ( $L_{\text{bol}}$ ) of the AGN, encompassing X-ray emissions. However, this data is unavailable to us. Additionally, the entire emission is often primarily influenced by the host galaxy, typically elliptical, as observed in our SED fits to estimate parameters like  $M_{\star}$ , SFR, and others (see Sect.2.3.4). Consequently, the bolometric luminosity is not solely attributable to the AGN.

An alternative approach involves utilizing the disk luminosity ( $L_{\text{disk}}$ ) as a proxy, which overwhelmingly dominates the bolometric emission, particularly in bright Type I AGN (Castignani et al. 2013). The continuum emission of the accretion disk is correlated with the BLR. In practice, the BLR reprocesses about 10% of the disk luminosity (Celotti et al. 1997; Müller & Romero 2020), thus  $L_{\text{disk}} \simeq 10L_{\text{BLR}}$ . Furthermore, in the case of jetted AGNs like the one I am investigating, even after removing the galaxy's emission, there still exists non-thermal broad-band emission from the extensive jets. In this scenario, while  $L_{\text{AGN}}$  seeks to estimate the power exclusively due to accretion, employing  $L_{\text{disk}}$  as a proxy appears more appropriate.

Hence, to determine the Eddington ratio, I first need to calculate the BLR luminosity, which will allow us to obtain the disk luminosity. I conducted the estimation of the BLR luminosity by applying the methodologies outlined in Celotti et al. (1997). In their work, they employed the line ratios reported in Francis et al. (1991) to determine the BLR luminosity. These line ratios provided the relative fluxes of emission lines in relation to  $\text{Ly}\alpha$ , with a reference value of 100. The authors make reference to an optical sample, predominantly comprised of radio-quiet sources. However, no significant disparities in the fluxes of broad-line emission lines between radio-quiet and radio-loud objects have been conclusively identified (Corbin 1992; Steidel & Sargent 1991; Boroson & Green 1992; Wills et al. 1993; Zheng et al. 1997). By following these established guidelines, the total normalized BLR luminosity can be calculated as follows:

$$\langle L_{\text{BLR}}^* \rangle = L_{\text{H}\alpha}^* + L_{\text{H}\beta}^* + L_{\text{C IV}}^* + L_{\text{Mg II}}^* + L_{\text{Ly}\alpha}^* + L_{\text{Ly}\beta}^* + L_{\text{H}\gamma}^* + L_{\text{Al III}}^* + L_{\text{Si IV}}^* + L_{\text{C II}}^* + L_{\text{O I}}^* = 390.3 \quad (2.4)$$

Hence, considering the summation of observed luminosities across a specific number of broad lines denoted as  $L_{i,\text{obs}}$ , the total luminosity of the BLR can be estimated as follows:

$$L_{\text{BLR}} = \langle L_{\text{BLR}}^* \rangle \frac{\sum_i L_{i,\text{obs}}}{\sum_i L_{i,\text{est}}^*}, \quad (2.5)$$

in this expression,  $L_{i,\text{obs}}$  represents the luminosity of the  $i$ th observed line within the BLR, while  $L_{i,\text{est}}^*$  denotes the line ratio corresponding to the  $i$ th line as provided in the table presented by Francis et al. (1991).

The resulting Eddington ratios are reported in Table A.2. They are mostly in the range  $\log \eta \sim [-4; -1]$ , with a median = -1.9, as typically found for type 1 RLAGN, but lower than the ratios of type 2 quasars (e.g., Castignani et al. 2013; Kong & Ho 2018). Subsequently, I will see in Sect.2.5.2 that I achieved consistent values for  $L_{disk} \sim 0.01$ , aligning with outcomes from other studies. This consistency reassures us that our  $L_{disk}$  estimation isn't unduly biased towards lower values.

## 2.3.4 SED modeling

The radio sources in our sample have a broad multiwavelength photometric coverage from the UV to the IR, which enables the determination of  $M_{\star}$  and SFR estimates via SED modeling. See Sect. 1.2.3 for more details.

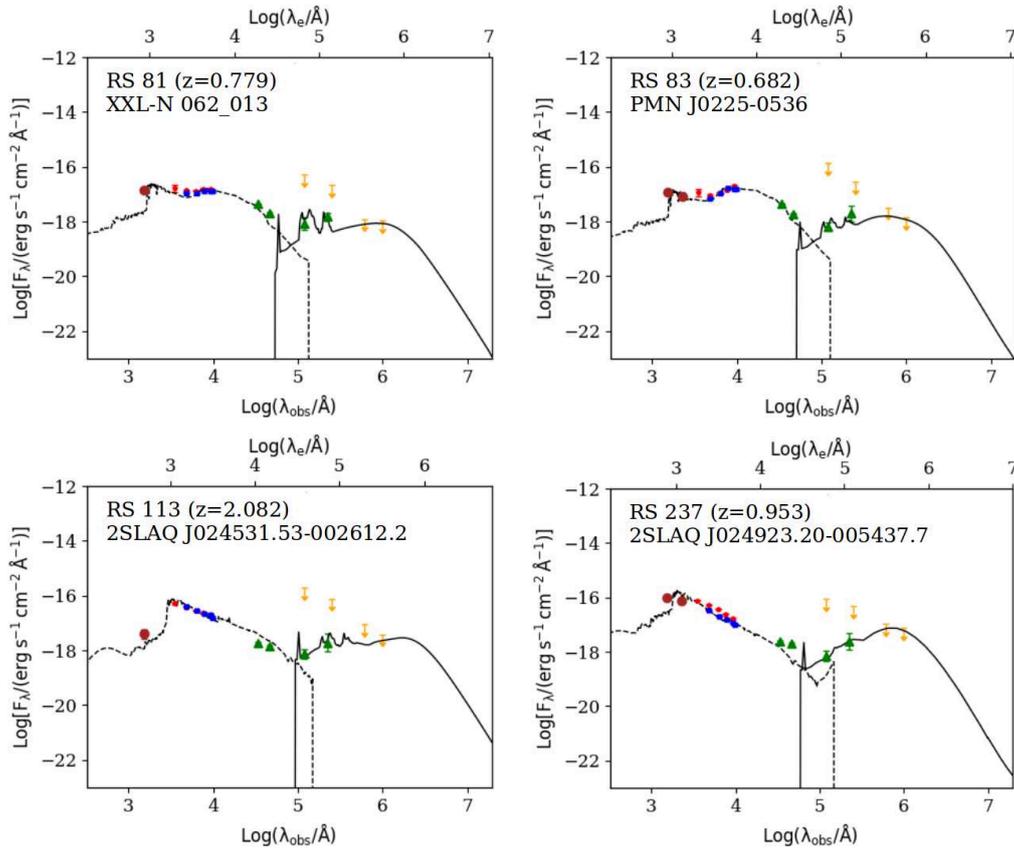
For the GAMA sources in our sample, we considered the SED fits by Driver et al. (2018) performed with MAGPHYS (da Cunha et al. 2008). Photometric data include GALEX (Martin et al. 2005; Morrissey et al. 2007) in the UV, SDSS (York et al. 2000) in the optical, as well as the VISTA Kilo-degree IR Galaxy Survey (VIKING, Edge et al. 2013), WISE (Wright et al. 2010), and *Herschel*-ATLAS (Eales et al. 2010; Valiante et al. 2016a) in the near to far-IR.

For the sources in the DES supernova deep fields, available photometry includes GALEX in the UV, *ugriz* (SDSS) and *grizY* (DES) magnitudes in the optical, WISE data in the near-IR, as well as IRAS upper limits in the far-IR, which we gathered as in Castignani et al. (2019), to which we refer for further details. In this previous work, we followed up two radio sources in molecular gas in dense Mpc-scale environments at  $z = 0.4$  and  $z = 0.6$  within the DES supernova deep fields, while we enlarge the sample here to investigate the coevolution of BHs with radio sources.

Analogously to Castignani et al. (2019), we then performed fits to the SEDs using LePhare (Arnouts et al. 1999; Ilbert et al. 2006). Following the prescriptions provided for the LePhare code, we fit the far-IR data separately to account for possible dust emission, using the Chary & Elbaz (2001) library consisting of 105 templates. The remaining photometric data points at shorter wavelengths were fit using the CE\_NEW\_MOD library, which consists of 66 galaxy templates based on linear interpolation of the four original SEDs of Coleman et al. (1980). We then converted the rest frame (8.0-1000)  $\mu\text{m}$  IR (dust) luminosity into an SFR estimate by using the Kennicutt (1998) relation, calibrated to an initial mass function Chabrier (2003).

$$\text{SFR}(M_{\odot} \text{ yr}^{-1}) = 4.5 \times 10^{-44} L_{FIR} \text{ (ergs s}^{-1}\text{)} \quad (2.6)$$

The SEDs of four of our radio sources are shown in Fig. 2.4. RS 81 and 83 have prominent emission in the optical domain typical of elliptical galaxies, while their IR emission is consistent with dust emission due to star formation. They are indeed classified as intermediate disks based on WISE colors. RS 113 and 237 are WISE AGN and show steep SEDs at near-IR to optical wavelengths, which suggests that the emission is contaminated by nonthermal AGN emission. The other SEDs are shown in the Appendix (see Fig.A.3)



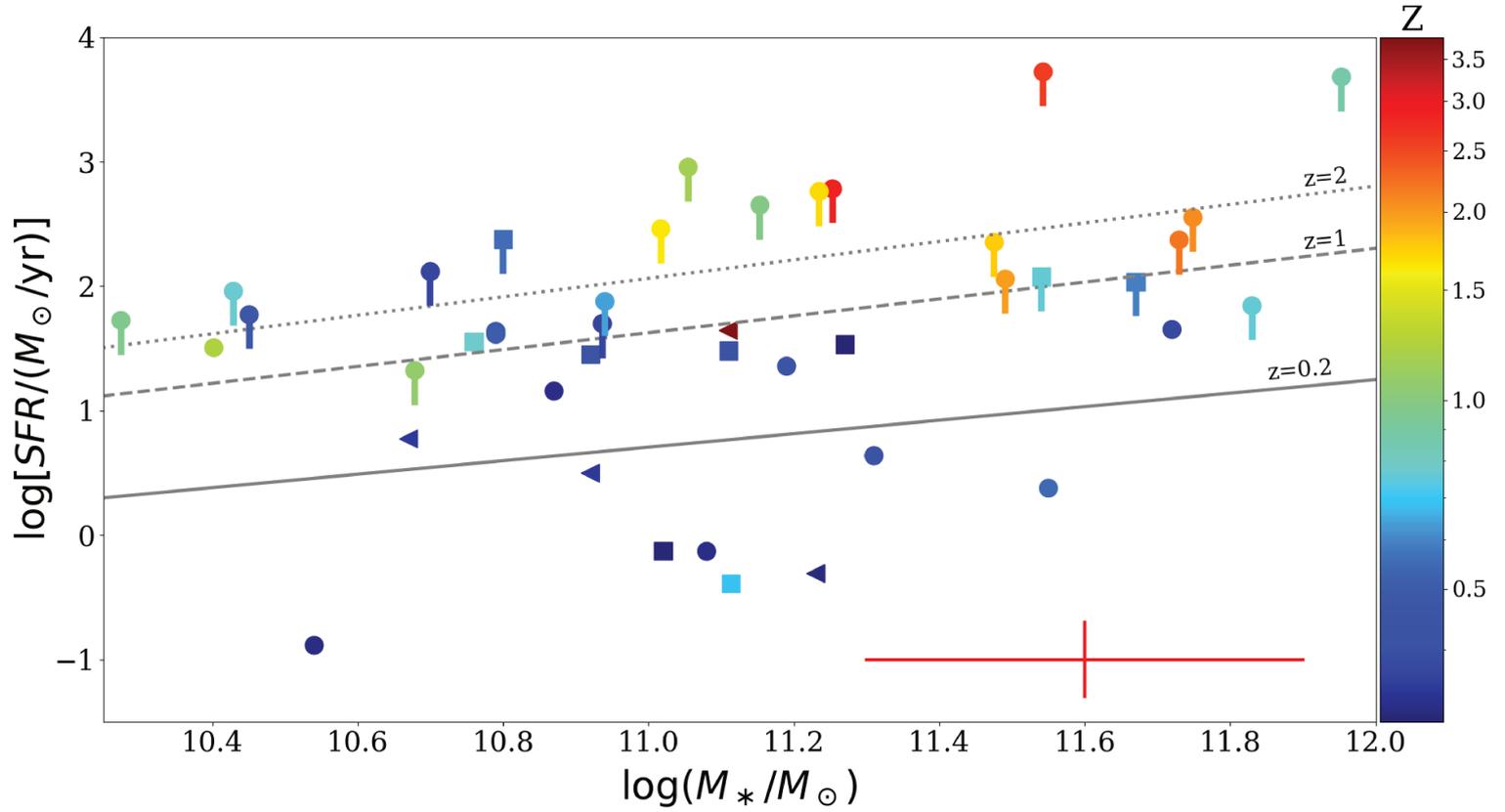
**Figure 2.4:** Examples of SEDs of the radio sources in our sample with  $M_{\text{BH}}$  estimates. The IDs, names, and redshifts of the galaxies are shown at the top left of each panel. Data points are from GALEX (brown dots), SDSS (red pentagons), DES (blue squares), WISE (green triangles), and IRAS (yellow upper limits). Dashed and solid lines are the best-fit models for the stellar and dust components, respectively.

### 2.3.5 SFR versus $M_{\star}$

Figure 2.5 displays the radio sources of our sample in the SFR versus  $M_{\star}$  plane resulting from the SED fits. The sources are color-coded according to the redshift, while the different symbols correspond to the WISE classification.

Overall, the sources are massive, with  $\log(M_{\star}/M_{\odot}) \simeq 10.3 - 12.0$  (median=11.1), which agrees with them being RLAGN, which are indeed typically hosted by massive ellipticals (Best et al. 2005). Our galaxies also mostly lie along the star-forming main sequence, although with a large scatter. The mean specific SFR is  $\text{sSFR}=0.40 \pm 0.44 \text{ Gyr}^{-1}$ , where the Root Mean Square (RMS) dispersion is reported as uncertainty.

Galaxies at higher redshifts tend to have higher SFRs, in agreement with the MS model prescriptions (Speagle et al. 2014). However, as highlighted in Sect. 2.2.6, the fraction of sources with AGN contamination also increases with redshift, which may result in biased-high SFRs. The latter may be the case where the optical-IR SED is steep, and thus the IR emission is likely dominated by the AGN contribution, more than star formation. To overcome this limitation, we conservatively reconsidered the SFR estimates and assigned upper limits when the SFRs largely exceeded  $100 M_{\star}/\text{yr}$  or in the cases of steep-spectrum SEDs (e.g., for RS 113 and 237 mentioned above). Namely, we considered as steep spectra those AGN whose optical-IR SED has a characteristic power-law behavior  $F_{\lambda} \propto \lambda^{-1}$ . I verified a posteriori that these radio sources are indeed mostly located in the upper part of the MS and are classified as WISE AGN.



**Figure 2.5:** Plot of the SFR vs  $M_*$  for the galaxies in our sample. Sources are color-coded according to their redshift, while the different symbols correspond to the different WISE classes (AGN as circles, starbursts as triangles, and intermediate disks as squares), as in Fig. 2.3. Upper limits to the SFRs are indicated with bars. The diagonal lines correspond to the MS model prescriptions by Speagle et al. (2014) at  $z = 0.2, 1,$  and  $2$ . The red cross in the bottom right corner shows the typical uncertainties of  $\sim 0.3$  dex for both SFRs and  $M_*$ .

## 2.4 Comparison sample

To place the AGN in our sample into context, I additionally considered a compilation of sources with available  $M_{\text{BH}}$  and  $M_{\star}$  estimates:

- A selection of 30 galaxies in close proximity was obtained from the research conducted by [Häring & Rix \(2004\)](#). The determination of galaxy masses was achieved by employing Jeans modeling or by adopting existing dynamical models. Correspondingly,  $M_{\text{BH}}$  were acquired from the comprehensive investigation conducted by [Tremaine et al. \(2002\)](#) and the associated references.
- An additional set of 35 nearby galaxies was incorporated into the study, derived from the sample compiled by [McConnell & Ma \(2013\)](#). This particular dataset involved an extensive revision and expansion of galaxy bulge masses, as well as the inclusion of refined dynamical measurements pertaining to  $M_{\text{BH}}$ .
- Furthermore, a selection of 32 types 1 AGN, situated within the redshift range of  $z = 0.3 - 0.9$ , was included in the investigation from the work of [Cisternas et al. \(2011\)](#). These AGN were sourced from the XMM-COSMOS survey and their associated  $M_{\star}$  were determined through meticulous modeling of *HST* images, taking into account the contributions from both the AGN and the host galaxy. The  $M_{\text{BH}}$  assigned to these AGN were derived from  $\text{H}\beta$  emission lines, as reported by [Trump et al. \(2009\)](#).
- In addition, a cohort of 18 broad-line X-ray AGN falling within the redshift range of  $0.5 < z < 1.2$  was incorporated into the study, sourced from the research conducted by [Schramm & Silverman \(2013\)](#). The  $M_{\text{BH}}$  attributed to these AGN were estimated based on the  $\text{Mg II}$  emission lines, while the  $M_{\star}$  estimates were derived utilizing *HST* color-based techniques.
- Furthermore, a cohort of 78 types 1 RQAGN at a redshift range approximately between  $z \simeq 1 - 2$  was included in the study, sourced from the COSMOS survey conducted by [Merloni et al. \(2010\)](#). The  $M_{\star}$  associated with these AGN were determined through SED fitting techniques, while the  $M_{\text{BH}}$  were derived based on the analysis of  $\text{Mg II}$  emission lines observed in VIMOS/VLT spectra.
- A total of 10 types 1 AGN, located within the redshift range of  $1 < z < 2$ , were reported from the COSMOS survey as presented by [Jahnke et al. \(2009\)](#). These AGN were accompanied by  $M_{\star}$  estimates derived from *HST* color-based measurements, while the virial  $M_{\text{BH}}$  were obtained through spectroscopic analyses conducted by the COSMOS Magellan/IMACS and zCOSMOS surveys.
- I incorporated a selection of 53 radio-quiet QSOs, all located at redshifts lower than 3, as provided by [Decarli et al. \(2010\)](#). The determination of the  $M_{\text{BH}}$  was achieved through virial methods utilizing the  $\text{H}\beta$ ,  $\text{Mg II}$ , and  $\text{C IV}$  emission lines. Additionally, the estimation of  $M_{\star}$  involved the assumption of a stellar R-band mass-to-light ratio.

- I acquired data regarding two luminous quasars at a redshift approximately equal to 4 from [Targett et al. \(2012\)](#). The virial  $M_{\text{BH}}$  estimates were derived using C IV emission, while the estimation of  $M_{\star}$  relied on the evolutionary synthesis models proposed by [Bruzual & Charlot \(2003\)](#).
- I presented the findings of nine distant quasars observed at redshifts ranging from approximately 1.4 to 6.4, as documented by [Shields et al. \(2006\)](#). The  $M_{\text{BH}}$  were determined through analysis of broad emission lines, while the dynamical bulge masses were inferred from the widths of CO emission lines.
- Finally, I included seven quasars observed at a redshift of approximately 6, as reported by [Wang et al. \(2010\)](#). The  $M_{\star}$  was determined by calculating the difference between the dynamic mass of the bulge and the mass of the CO molecular gas. For these quasars, I used the  $M_{\text{BH}}$  adopted by the authors, which were estimated based on the AGN continuum luminosity ([Jiang et al. 2006](#); [Wang et al. 2008](#)).

In addition to the sources listed above, the second group of galaxies that I used as a comparison is composed of powerful AGN with available estimates of the  $M_{\text{BH}}$ , jet power, and Eddington ratio:

- I incorporated 44 RLAGN into our study, sourced from [Le et al. \(2018\)](#), exhibiting redshifts below  $z < 0.2$ . These AGN specimens were accompanied by estimations of their respective jet powers as reported by [Le et al. \(2018\)](#), along with  $M_{\text{BH}}$  assessments referenced from [Allen et al. \(2006\)](#) and [Balmaverde et al. \(2008\)](#).
- Furthermore, I incorporated a total of 208  $\gamma$ -ray *Fermi* blazars into our analysis, sourced from [Xiong & Zhang \(2014\)](#), with redshifts ranging from  $0 < z < 3.1$ . These blazars were accompanied by virial  $M_{\text{BH}}$  estimates, primarily obtained through various broad emission lines, while some were derived using scaling relations. Notably, the determination of jet powers  $Q_{\text{jet}}$  was primarily based on the work of [Nemmen et al. \(2012\)](#), utilizing the correlation between extended radio emission and jet power. However, an alternative approach was employed by [Xiong & Zhang \(2014\)](#), whereby  $Q_{\text{jet}}$  was calculated using the scaling relation proposed by [Nemmen et al. \(2012\)](#), which relates the  $\gamma$ -ray luminosity to the kinetic power.
- Finally, I reported a total of 146 radio-loud QSOs sourced from [Liu et al. \(2006\)](#), characterized by redshifts ranging from  $0.1 < z < 2.5$ . These QSOs were further classified as either flat-spectrum (54%) or steep-spectrum radio quasars (46%). The determination of virial  $M_{\text{BH}}$  was accomplished by extracting relevant data from the H $\beta$ , Mg II, or C IV emission lines. Additionally, the authors of [Liu et al. \(2006\)](#) employed low-frequency radio emission, following the methodology outlined in [Punsly \(2005\)](#), to calculate the respective jet powers associated with the QSOs.

These sources outlined above are RLAGN. However, I verified that none of them are included in our sample. While these studies investigated the  $M_{\text{BH}}$  and jet properties of large samples of radio sources, they did not characterize their IR to optical SEDs, as we did here for our smaller sample of radio sources.

## 2.5 Results

In this section, I report different scaling relations for the radio sources in our sample, including  $M_{\text{BH}}$  and  $M_{\star}$ , jet powers, Eddington ratios, and redshifts. I also include the sources from the literature as outlined in Sect. 2.4 as a comparison, as well as the scaling relations derived in previous studies.

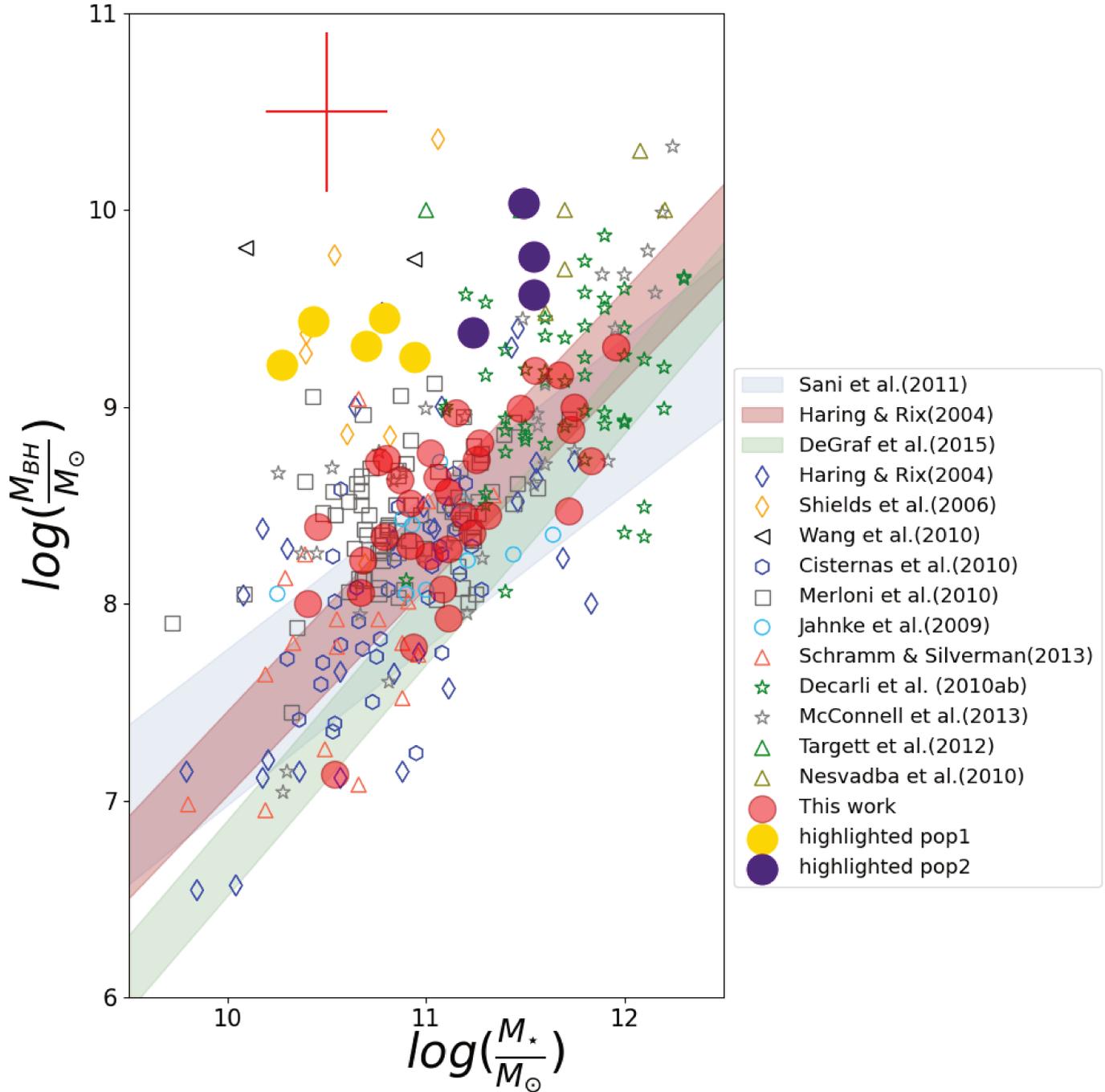
### 2.5.1 $M_{\text{BH}}$ - $M_{\star}$ relation

I start by considering  $M_{\text{BH}}$  and  $M_{\star}$  and their relative evolution with redshift. Fig 2.6 displays the  $M_{\text{BH}}$  versus the  $M_{\star}$ . Interestingly, our radio-loud sources nicely follow the trend previously observed for different samples of both local galaxies and distant AGN, overplotted in the figure, and for those inferred by the scaling relations, which were also reported (Sani et al. 2011; Häring & Rix 2004; DeGraf et al. 2015).

In particular, our sources populate the high  $\log(M_{\text{BH}}/M_{\odot}) \simeq 7.1-10.0$  and high  $\log(M_{\star}/M_{\odot}) \simeq 10.2 - 12.0$  region in Fig. 2.6 densely, which agrees with the fact that RLAGN are almost invariably associated with the most massive galaxies and BHs. Interestingly, a non-negligible fraction of our sources, 9/42 (i.e., 21%), have their  $\log(M_{\text{BH}}/M_{\odot}) > 9$  well above the scaling relations for both local (Häring & Rix 2004; Sani et al. 2011) and distant sources at the median redshift  $z = 0.6$  of our sample (DeGraf et al. 2015). These galaxies are highlighted in gold and purple in function of their stellar mass. This behavior suggests that the growth of  $M_{\text{BH}}$  in RLAGN largely occurs at early  $z > 1$  epochs, while the early  $M_{\star}$  assembly may not be equally effective. The substantial growth of the  $M_{\star}$  may take place even at lower redshifts in order to flatten the observed  $M_{\text{BH}}-M_{\star}$  scaling relation by  $z = 0$ . Previous studies indeed suggested that massive ellipticals may double their  $M_{\star}$  between  $z = 1$  and  $z = 0$  (Ilbert et al. 2010; Lidman et al. 2012).

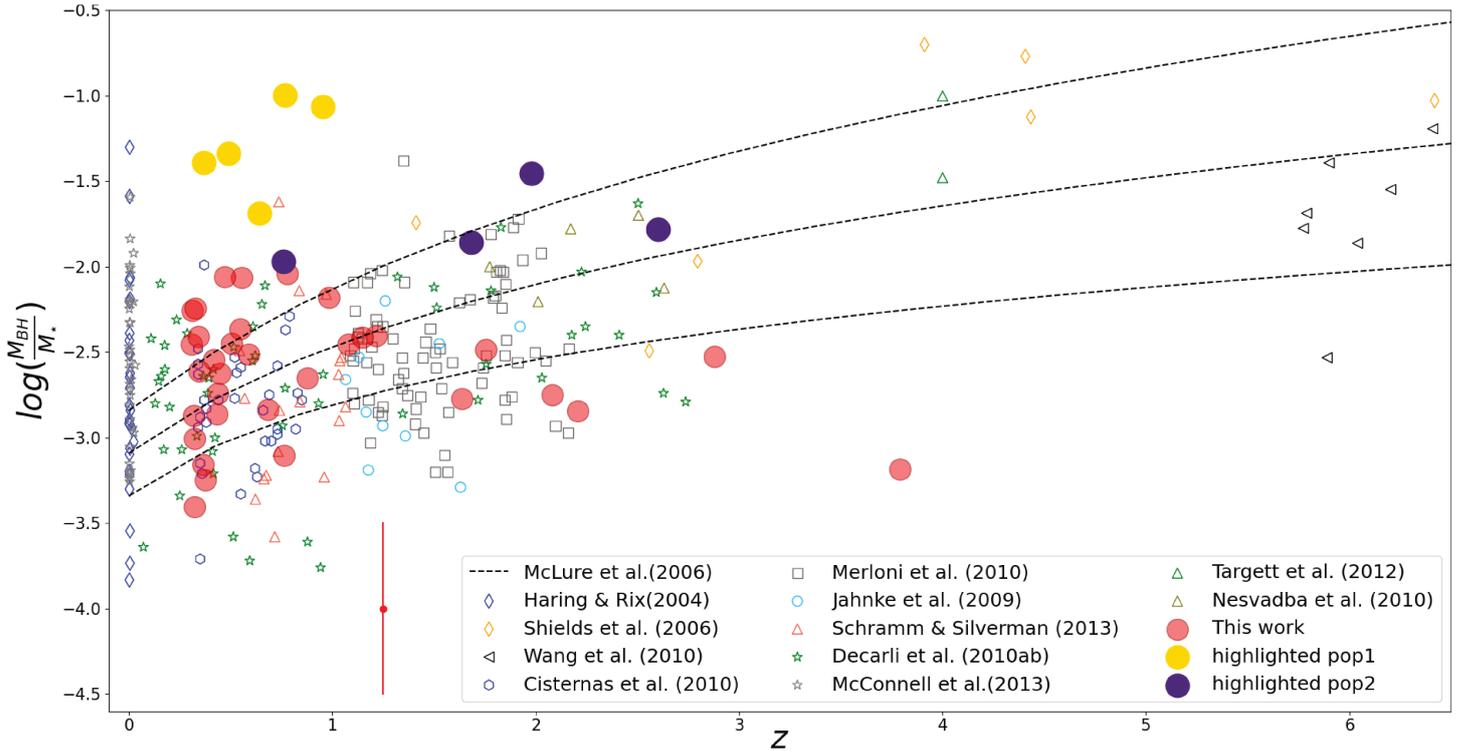
However, this figure presents a challenge in the comparison between the obtained  $M_{\text{BH}}$  through emission-line diagnostics (Fundamental Plane of BHs) and the  $M_{\text{BH}}$  acquired using dynamical models that account for gravitational influence, stars, or gas motions. These values might not be identical or directly comparable. To ensure a consistent comparison of  $M_{\text{BH}}$  using the same estimation method, I examine the evolutionary scenario in Fig. 2.7. This figure illustrates the  $M_{\text{BH}}/M_{\star}$  ratio relative to redshift. Estimating  $M_{\text{BH}}$  using dynamical modeling necessitates resolved observations of the AGN central region, thus feasible primarily at lower redshifts,  $z < 0.3$ . At higher redshifts,  $M_{\text{BH}}$  estimation methods rely on BLR emission lines. By examining the evolution of  $M_{\text{BH}}$  across cosmic time, I can compare estimations that are similar and compatible between them.

I observed in Fig.2.7 that the majority (36 out of 42, i.e., 86%) of our radio sources have  $M_{\text{BH}}/M_{\star}$  ratios that are similar to those of AGN in the comparison sample at similar redshifts, and they agree, at least partially, with model prescriptions by McLure et al. (2006) (overplotted as dashed lines in Fig. 2.7). The ultimate fate of these galaxies by  $z = 0$  remains uncertain, prompting the necessity for conducting high-resolution simulations. It is plausible that these anomalies are associated with scenarios where the stellar mass continues to increase, while the



**Figure 2.6:** Scatter plot of  $M_{\text{BH}}$  vs  $M_*$ . Filled red dots correspond to our sample of radio sources, and the sources in the comparison sample are shown as open symbols. The gold and purple dot correspond to outliers with overmassive  $M_{\text{BH}}$  in our sample. The gold galaxies are characterized by  $\log(M_*/M_\odot) < 11$  and the purple ones with the opposite. Scaling relations are overlaid (Sani et al. 2011; DeGraf et al. 2015; Haring & Rix 2004). The legend at the right displays the adopted color code. The red cross in the top left corner shows the typical uncertainties  $\sim 0.4$  dex and  $\sim 0.3$  dex for the  $M_{\text{BH}}$  and  $M_*$  respectively.

$M_{\text{BH}}$  remains relatively constant unless influenced by further significant mergers. In such cases, the ratio of  $M_{\text{BH}}/M_*$  is likely to decrease. However, arriving at a precise determination regarding these outliers is challenging until I gain a comprehensive understanding of their nature. It is



**Figure 2.7:**  $M_{\text{BH}}$  to  $M_{\star}$  ratio as a function of redshift for our sample of radio sources (filled red dots) and for the galaxies in our comparison sample (open symbols). The gold and purple dots represents the same galaxies that in Fig.2.6. The dashed black lines correspond to the evolutionary model described by McLure et al. (2006), along with the associated  $1\sigma$  uncertainty. The color coding is reported in the legend in the bottom right corner. The error bars to the left of the legend show the typical  $\sim 0.5$  dex uncertainty in  $\log(M_{\text{BH}}/M_{\star})$ .

imperative to acquire high-resolution observations in various wavelengths, such as radio or optical, to commence an interpretation based on physical insights rather than speculative notions outlined in this thesis.

It is worth mentioning that our sample of radio sources is flux-limited. However, I expect the Malmquist bias to have a marginal impact on the  $(M_{\text{BH}}/M_{\star})$  ratio.  $M_{\text{BH}}$  and  $M_{\star}$  both scale with the BLR line and the IR to optical luminosities, respectively, and therefore have a similar dependence on redshift via the luminosity distance. Furthermore, as shown in Fig. 2.7, at fixed redshift, the  $M_{\text{BH}}/M_{\star}$  ratios of both our radio sources and those in the comparison sample span a broad range. Similarly, when plotting  $M_{\star}$  and  $M_{\text{BH}}$  against redshift separately, I did not find any clear trend, as indeed the points are scattered at fixed redshift. These findings suggest that any possible Malmquist bias likely has a subdominant effect.

There are, however, five clear outliers among our radio sources. RS 197 at the highest redshift  $z = 3.79$  has a low  $\log(M_{\text{BH}}/M_{\star}) = -3.2$ , which is well below the expected range of values, according to the McLure et al. (2006) model prescription. Furthermore, a non-negligible fraction (the five sources highlighted in gold, i.e., 12%) of our radio-loud AGN, namely RS 214, RS 237, G 537618, G 721940, and G 746605, at redshifts between  $z \sim 0.37 - 0.95$ , have high  $\log M_{\text{BH}}/M_{\star}$  ratios in the range  $\sim (-1.69; -1.0)$ , which is well above the model predictions displayed in Fig. 2.7, as well as higher than the ratios found in AGN at similar redshifts. Indeed, in the redshift range  $z \sim 0.3 - 3.8$  spanned by our radio sources, there are only 3 out of 186

(i.e., 1.6%) AGN with  $\log M_{\text{BH}}/M_{\star} > -1.69$  in our comparison sample, while the proportion is significantly higher (12%) for our RLAGN.

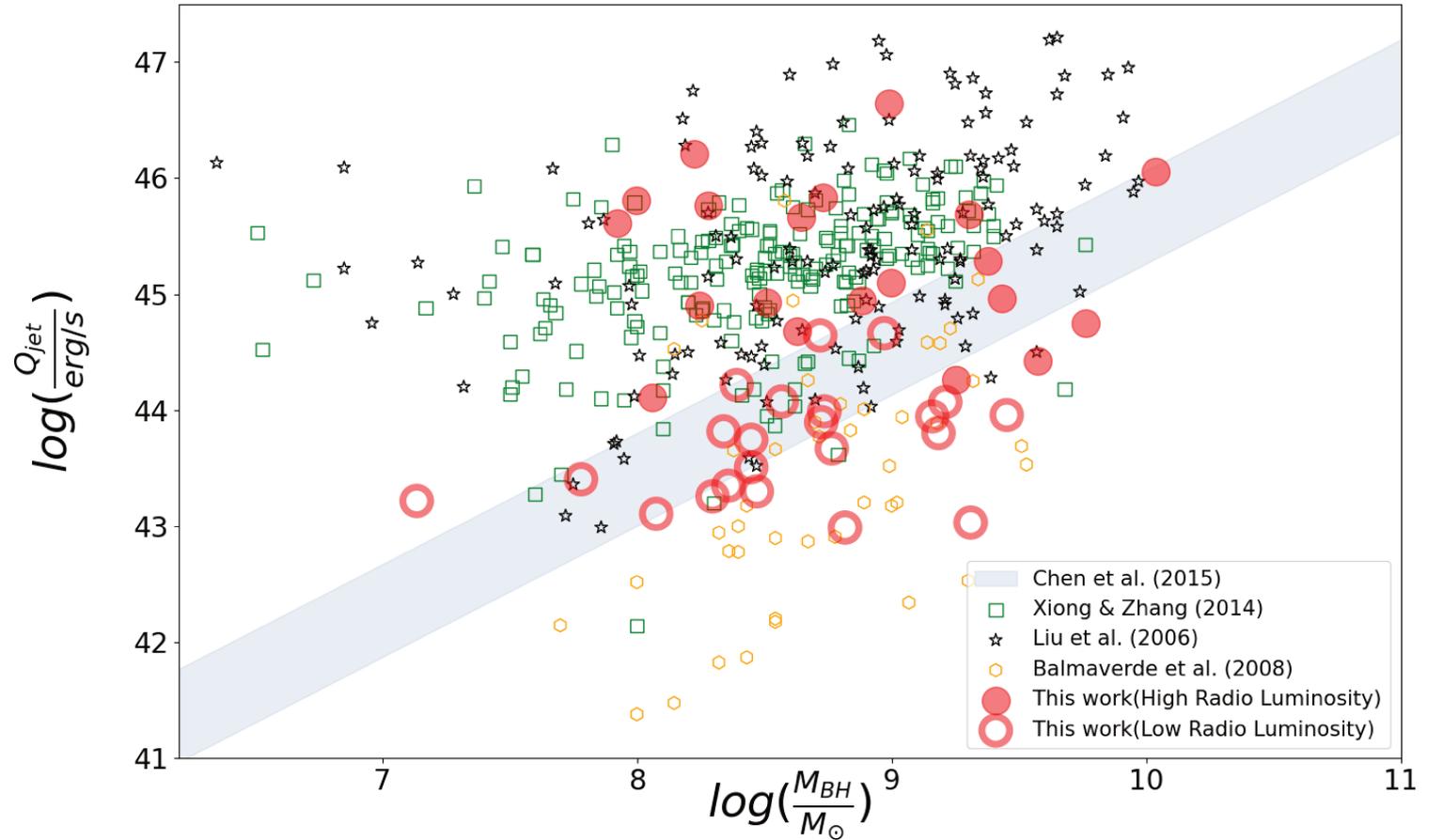
These results suggest that a non-negligible fraction of RLAGN may experience a different  $M_{\star}$  assembly path than RQAGN. I stress that these five radio sources are a subsample of the nine outliers of Fig. 2.6, as discussed above, and have high S/N line fluxes in  $\text{H}\beta$  or  $\text{Mg II}$ , which yielded robust  $M_{\text{BH}}$  estimates. These galaxies were observed as RLAGN, but they may have transitioned from RQAGN at some stage in their evolution, given the existence of AGN duty cycles. However, I do observe a distinction within a subset of our sample. Furthermore, RLAGN typically host the most massive  $M_{\text{BH}}$  and galaxies, indicating that the RLAGN and RQAGN populations exhibit intrinsic differences, regardless of the AGN duty cycles.

The excess of overmassive SMBH in RLAGN may suggest that their stellar and SMBH mass buildup is regulated by their large-scale radio jets. A possible scenario is that SMBH of the subpopulation with high  $M_{\text{BH}}/M_{\star}$  are mature, that is, their mass has been effectively assembled already by redshift  $z = 1$  via accretion (e.g., [Delvecchio et al. 2018](#)). On the other hand, their  $M_{\star}$  growth may not have occurred as effectively as in the overall AGN population, plausibly because of radio-mode AGN feedback ([Fabian 2012](#)). While the accretion of hot gas onto the SMBH sustains the AGN activity and the SMBH growth, the large-scale radio jets may prevent the accretion and cooling of the intergalactic medium gas, which is ultimately responsible for the stellar mass assembly. Altogether, I suggest that radio-mode AGN feedback results in the observed high values for  $M_{\text{BH}}/M_{\star}$  in RLAGN.

To investigate further this scenario, I link accretion and jet properties to the  $M_{\text{BH}}$  in the next sections by considering both the jet power and Eddington ratio of our radio sources. I stress that the usual  $M_{\text{BH}} - M_{\text{bulge}}$  or  $M_{\text{BH}} - \sigma$  relations typically refer to the central spheroid and not to the total  $M_{\star}$  (e.g., [Kormendy & Ho 2013](#)). However, our RLAGN sample is composed of a large majority of early-type galaxies, where the spheroid constitutes most of the  $M_{\star}$  and this approximation is justified. Furthermore, because of the potential AGN contamination to the SED, the  $M_{\star}$  may be biased high. This implies that  $M_{\text{BH}}/M_{\star}$  ratios can be even higher than reported. By considering  $M_{\text{BH}}/M_{\star}$  ratios as lower limits, I would have an even stronger discrepancy, in particular, for the subsample of high  $M_{\text{BH}}/M_{\star}$  radio sources mentioned above, with respect to the model prescriptions and the comparison sample of distant AGN at fixed redshift. All these results seem to corroborate the scenario that SMBH growth is more rapid than stellar mass assembly, and this is particularly true for distant radio sources, in comparison to the overall AGN population.

## 2.5.2 Jet power, BH mass, and accretion

As mentioned in the previous sections, mechanical radio-mode AGN feedback can regulate the cooling of hot gas in the intergalactic medium, and thus the  $M_{\star}$  growth of the host galaxy itself as well as the accretion onto the central SMBH. To better understand the interplay between jet, BH, and accretion properties, in Fig. 2.8 I show the jet power  $Q_{\text{jet}}$  (see Sect. 2.3.2), plotted against the  $M_{\text{BH}}$ . The radio sources of our sample are highlighted, and I also over-plot the comparison sources outlined in Sect. 2.4 ([Liu et al. 2006](#); [Balmaverde et al. 2008](#); [Xiong & Zhang 2014](#); [Chen et al. 2015](#)).



**Figure 2.8:** Jet power vs.  $M_{\text{BH}}$ . The diagonal dashed region represents the model reported in previous studies (Wu & Cao 2008; Chen et al. 2015) that distinguishes between FR I and FR II radio galaxies. Sources from our sample are distinguished between high- and low-luminosity radio sources. Sources from our comparison sample are also shown. I refer to the legend for the color code.

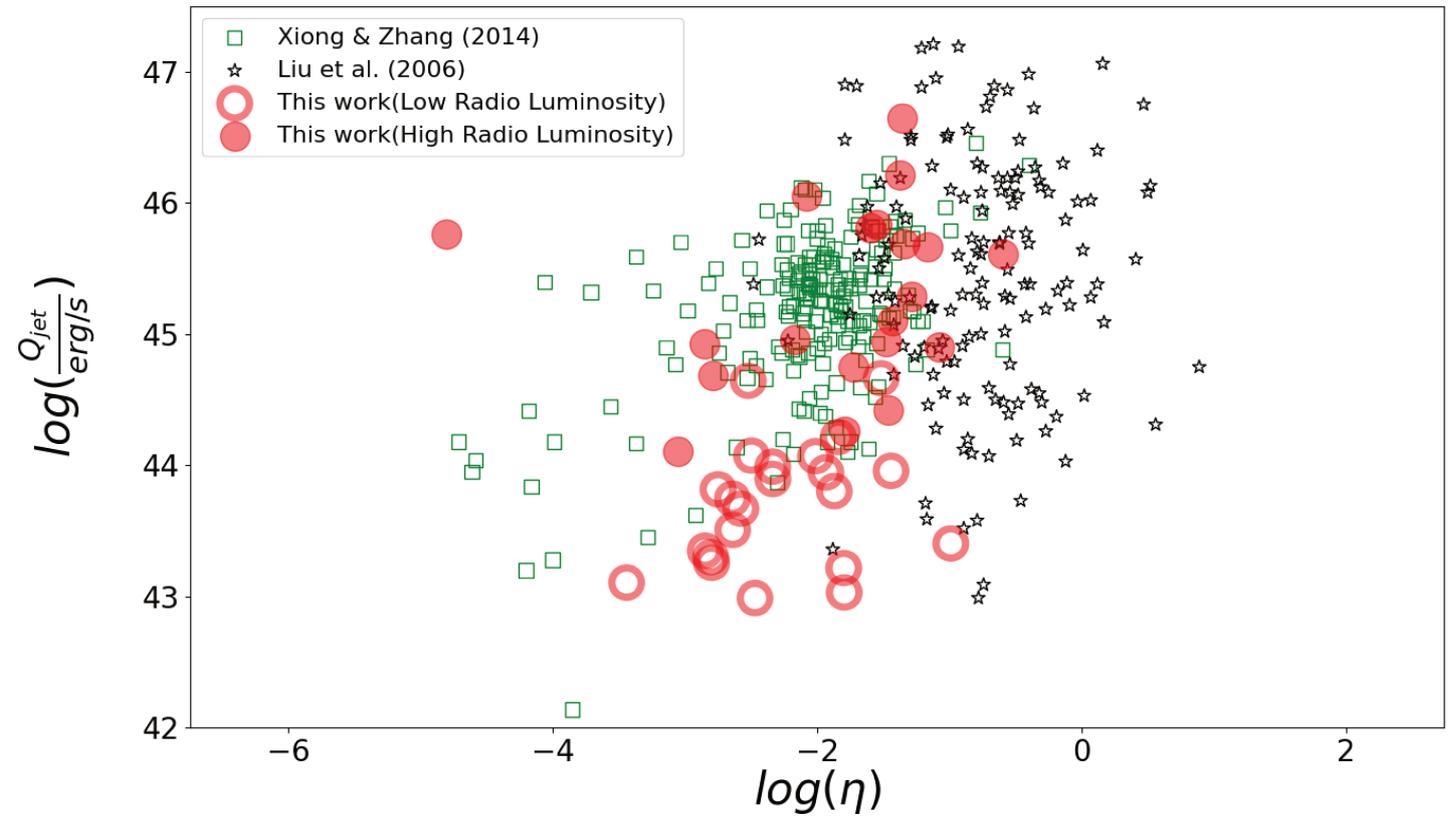
Our RLAGN densely populate the upper right region of the  $Q_{\text{jet}}-M_{\text{BH}}$  plane, which is occupied by sources with high values of both the  $M_{\text{BH}}$  ( $M_{\text{BH}} \gtrsim 10^8 M_{\odot}$ ) and the jet power ( $Q_{\text{jet}} \gtrsim 10^{43}$  erg s $^{-1}$ ). Sources in the comparison sample similarly occupy this region, while they also extend to lower values of  $M_{\text{BH}}$  (Liu et al. 2006; Xiong & Zhang 2014) and jet power (Balmaverde et al. 2008). These results suggest that the distant RLAGN, quite independently of the redshift, is almost invariably associated with massive BHs and powerful radio jets. This agrees with the tight connection existing between BH accretion and jet production in powerful RLAGN (e.g., Ghisellini et al. 2014; Sbarrato et al. 2014; Inoue et al. 2017).

Furthermore, HLR are characterized by a jet power that is typically higher than in LLR. As discussed in Sect. 2.2.6, these two classes indeed have 1.4 GHz rest-frame luminosity densities typical of FR II and FR I radio galaxies, respectively. As  $Q_{\text{jet}}$  increases with the radio luminosity density (Eq. 2.3), high-luminosity radio sources have higher  $Q_{\text{jet}}$  values than low-luminosity sources. Furthermore, the two classes of LLRGs and HLRGs are also delimited in the  $Q_{\text{jet}}-M_{\text{BH}}$  plane by the relation found in previous studies (Wu & Cao 2008; Chen et al. 2015), originally used to distinguish between FR I and FR II radio galaxy populations. The smooth separation of LLRGs and HLRGs in the  $Q_{\text{jet}}$  versus  $M_{\text{BH}}$  plane can be explained by combining the  $M_{\text{BH}}$  versus  $M_{\text{bulge}}$  relation for elliptical galaxies and the relation between  $Q_{\text{jet}}$  and the host galaxy optical luminosity (Ledlow & Owen 1996) that separates the FR I and FR II radio galaxies. The combination of these two relations also implies the observed spread of our sources in Fig. 2.8. I did not find any significant correlation (as measured with the Spearman test) between  $Q_{\text{jet}}$  and  $M_{\text{BH}}$  for our radio sources.

Figure 2.9 displays instead the jet power, plotted against the Eddington ratio  $\eta$  (see Sect. 2.3.3) for our radio sources and the galaxies in the comparison sample (Xiong & Zhang 2014; Liu et al. 2006). Despite the high scatter observed in Fig. 2.9, I can observe that the jet power increases with increasing Eddington ratio, meaning that higher accretion rates favor more powerful jets to be launched by the central engine. For our sample of radio sources, I find that the two quantities are well correlated at a level of  $2.9\text{-}\sigma$  ( $p\text{-value} = 3.30 \times 10^{-3}$ ) by means of the Spearman test. No clear distinction in terms of  $\eta$  is found between the two classes of low- and high-luminosity radio sources, which are distinguished in Fig. 2.9. However, as pointed out in Sect. 2.3.3 our radio sources have an Eddington ratio of  $\log \eta = -1.9$  on average. This value is typical of radiatively efficient accretion disks, such as the Shakura & Sunyaev (1973) optically thick and geometrically thin accretion disk, which is commonly invoked to explain the optical-UV emission in type I AGN (e.g., Ghisellini et al. 2010; Castignani et al. 2013).

I can then estimate the accretion rate  $\dot{M} = L_{\text{disk}}/(\epsilon c^2)$ , where  $\epsilon$  is the mass-to-light conversion efficiency, for which I adopted the standard value  $\epsilon = 0.1$ , which is typical of radiatively efficient disks. For our radio sources, I obtain a median (mean) accretion rate of  $0.16 M_{\odot} \text{ yr}^{-1}$  ( $0.6 M_{\odot} \text{ yr}^{-1}$ ), which corresponds to a substantial SMBH mass growth of  $\Delta M_{\text{BH}} = 1.6 \times 10^6 M_{\odot}$  ( $6.0 \times 10^6 M_{\odot}$ ) over an AGN duty cycle with typical duration of  $\sim 10^7$  yr.

Altogether, the fact that the SMBH of the radio sources in our sample accrete at a sub-Eddington rate, regardless of their redshift, suggests that most of their mass has likely been built up at an epoch prior to their observation. Furthermore, while on one hand, the observed accretion state sustains both the nuclear activity and the SMBH growth at subparsec scales, on the other hand, it also ultimately favors the persistence of large-scale radio jets, which may prevent the host galaxy from accreting gas at kiloparsec scales and thus form stars effectively.



**Figure 2.9:** Plot of the jet power vs. the Eddington ratio ( $\eta$ ) of the associated BH. In addition to our sample, the data found in Xiong & Zhang (2014) and Liu et al. (2006) are plotted.

This radio-mode AGN feedback may be responsible for the presence of overmassive SMBH in our sample of RLAGN. It is worth mentioning that the five  $z \sim 0.37 - 0.95$  radio sources with high  $M_{\text{BH}}/M_{\star}$  ratios that I discussed in Sect. 2.5.1 accrete a sub-Eddington rate of  $\eta \sim 1\%$ , while they have a normal jet power  $Q_{\text{jet}} \sim 10^{44} \text{ erg s}^{-1}$  on average.

### 2.5.3 RLAGN and their environments

A non-negligible fraction ( $17 \pm 5\%$ ) of our radio sources have high  $M_{\text{BH}}/M_{\star}$  ratios (Sect. 2.5.1) and may be early-type galaxies at the center of clusters. For these galaxies, the stellar mass assembly may have been halted, with reduced star formation activity, as typically found in cluster core ellipticals, while their BHs continue to grow via accretion. This interpretation is consistent with earlier studies (e.g., Trujillo et al. 2014) as well as with the substantial fraction (19%) of radio sources in our sample with low SFR  $< 5 M_{\odot}/\text{yr}$ , while many others have SFR upper limits. It is indeed known that cluster-core early-type galaxies tend to be outliers in the  $M_{\text{BH}}$  vs  $M_{\text{bulge}}$  relation (e.g., McConnell & Ma 2013).

Motivated by these studies, I searched in the literature for clusters around the radio sources in our sample, searching by coordinates in the NASA/IPAC Extragalactic Database (NED). NED includes several catalogs of clusters that were identified in wide-field surveys (Goto et al. 2002; Koester et al. 2007; McConnachie et al. 2009; Hao et al. 2010; Durret et al. 2011, 2015; Wen et al. 2012; Radovich et al. 2017; Rykoff et al. 2012; Oguri et al. 2018). Our search yielded three matches. Radio sources RS 49, G 372455, and G 748815 are in the cores (at cluster-centric distances  $\lesssim 0.5 \text{ Mpc}$ ) of the clusters [LIK2015] J034.16359-04.73395 ( $z = 0.89$ ; Lee et al. 2015), WHL J090325.6+011215 ( $z = 0.31$ ; Wen et al. 2012; Wen & Han 2015), and HSCS J142538+002320 ( $z = 0.33$ ; Oguri et al. 2018), respectively. These clusters have redshifts that are consistent with those of the radio sources as well as richness-based masses  $M_{200} \sim (0.9 - 3.0) \times 10^{14} M_{\odot}$ . These are therefore moderately massive clusters at intermediate to high redshifts. The three associated radio sources instead have moderately overmassive BHs  $\log(M_{\text{BH}}/M_{\odot}) \simeq 8.6 - 9.3$ , in particular, in comparison to the  $M_{\star}$  of the systems  $-2.65 \lesssim \log(M_{\text{BH}}/M_{\star}) \lesssim -2.24$  (see Fig. 2.7).

These results support the interpretation described above that the cluster environments tend to prevent the stellar mass assembly of cluster early-type galaxies, resulting in observed overmassive SMBH. Nevertheless, it is worth mentioning that only three radio sources of our sample are found in clusters. However, this is expected as clusters at higher redshifts ( $z \gtrsim 1$ ) or with lower masses  $M_{200} \lesssim 1 \times 10^{14} M_{\odot}$  typical of rich groups are more difficult to find with current surveys and observational facilities. It is, therefore, possible that additional galaxies are hosted in clusters, as distant radio sources are often found in dense megaparsec-scale environments (e.g., Galametz et al. 2012; Castignani et al. 2014; Malavasi et al. 2015; Golden-Marx et al. 2019; Moravec et al. 2020).

These entities could surpass the  $M_{\text{BH}} - M_{\star}$  relation due to various factors such as being remnants of over-massive BH from the high-redshift universe. They have the potential to grow through both minor and major mergers. Minor mergers may increase the galaxy's size with minimal impact on the mass and SMBH mass, while major mergers are likely to augment both the galaxy mass and the SMBH mass. Additionally, RLAGN are often situated at the centers of

intricate galaxy clusters, potentially accelerating the rate of black hole mergers. To investigate this hypothesis further, in-depth observations using the Hubble Space Telescope and analysis of stellar populations could shed light on whether these systems are post-merger structures (e.g., [Morishita et al. 2022](#)).

## 2.6 Discussion and conclusions

I have investigated the evolution of distant RLAGN, as well as their coevolution with their host galaxies and the SMBH at their center. To do this, I built a sample of 42 RLAGN with spectroscopic redshifts between  $z \sim 0.3 - 3.8$  by cross-matching the 1.4 GHz VLA FIRST point-source catalog with available IR to optical spectrophotometric surveys including SDSS and DES in the optical, WISE in the IR, and the GAMA spectroscopic survey. As I am interested in assessing the SMBH masses, the 42 galaxies were further selected by requiring broad emission lines in  $H\alpha$ ,  $H\beta$ ,  $Mg\ II$ , or  $C\ IV$ , with an FWHM  $> 1000\ km\ s^{-1}$ . Based on the available multiwavelength photometry, we modeled the SEDs of the sources in the sample, and then derived estimates of the  $M_\star$  and SFR for all sources, while for GAMA sources, we took them from the literature. We find that the 42 radio sources are broadly consistent with the star-forming main sequence.

For all sources, I then estimated i) the  $M_{BH}$ , based on single-epoch broad-line region spectra, ii) the ratio of  $M_{BH}$  to  $M_\star$ ,  $M_{BH}/M_\star$ , iii) the jet power  $Q_{jet}$ , on the basis of the low-frequency radio continuum emission, and iv) the Eddington ratio  $\eta$ . Although samples of distant AGN with SMBH mass estimates are rapidly growing (e.g., [Shen et al. 2011](#); [Shaw et al. 2012](#); [Dabhade et al. 2020](#); [Rakshit et al. 2020](#); [Gludemans et al. 2021](#); [Li et al. 2022](#)), our study still represents one of the first in which all these quantities are derived simultaneously for a single sample of distant RLAGN.

Our radio sources have  $\log(M_{BH}/M_\odot) \simeq 7.1 - 10.0$  and high  $\log(M_\star/M_\odot) \simeq 10.2 - 12.0$ , which agrees with the fact that RLAGN is almost invariably associated with the most massive galaxies and BH (e.g., [Best et al. 2005](#); [Chiaberge & Marconi 2011](#)). While overall our sources follow the expected trends previously found in the literature, a substantial fraction of our sources, 9 out of 42 (i.e., 21%), have  $\log(M_{BH}/M_\odot) > 9$  well above the values predicted by the scaling relations ([Haring & Rix 2004](#); [Sani et al. 2011](#); [DeGraf et al. 2015](#)). In particular, five sources out of the nine (12% of the full radio source sample) are overmassive outliers, with  $M_{BH}/M_\star > 2\%$ . This fraction is remarkably higher than that of 1.6% found for RQAGN at similar redshifts from the literature. These overmassive SMBH are thus the high- $z$  counterparts of overmassive SMBH found in previous studies of nearby early-type galaxies (e.g., [McConnell & Ma 2013](#); [Trujillo et al. 2014](#)).

Our results imply that the growth of  $M_{BH}$  in at least a non-negligible fraction of RLAGN largely occurs at early epochs, while the early stellar mass assembly may not be so efficient. This population of RLAGN with high  $M_{BH}/M_\star$  ratios have likely experienced a different  $M_\star$  growth than other types of AGN, and I further investigated this scenario in terms of additional complementary probes.

Following early studies on nearby galaxies (e.g., [McConnell & Ma 2013](#); [Trujillo et al.](#)

2014), I found that three of our RLAGN with moderately overmassive SMBH are hosted in clusters from the literature, while clusters and groups around the majority of the remaining RLAGN will likely be detected with forthcoming surveys such as *Euclid* (Euclid Collaboration et al. 2019). These results suggest that the cluster environments tend to prevent the stellar mass assembly of cluster early-type galaxies, possibly via radio-mode AGN feedback.

I found that the nuclear accretion and jet properties of the SMBH of the radio sources in our sample accrete at a sub-Eddington rate ( $\eta \sim 1\%$ ) on average, where higher accretion rates favor more powerful jets to be launched by the central engine. I also find that high jet powers ( $Q_{\text{jet}} \gtrsim 10^{45} \text{ erg s}^{-1}$ ) are invariably associated with high radio luminosity sources ( $L_{1.4 \text{ GHz}} > 2 \times 10^{32} \text{ erg s}^{-1} \text{ Hz}^{-1}$ ). Altogether, the observed accretion state sustains both the nuclear activity and the SMBH growth at sub-pc scales, while it ultimately favors the persistence of large-scale radio jets, which may prevent the host galaxy from accreting gas at kpc scales and thus form stars effectively. Radio-mode AGN feedback may be responsible for the presence of overmassive SMBH in our sample of RLAGN.

This preliminary investigation serves as a reference point for forthcoming research endeavors utilizing cutting-edge surveys. To delve deeper into the analysis of RLAGN hosting excessively SMBH, two potential approaches stand out. Firstly, a focused examination involving targeted observations of both ionized and molecular gas could significantly contribute to exploring the radio-mode AGN feedback hypothesis. This scrutiny aims to unravel the intricate physical processes occurring within RLAGN and how it is coupled with their host galaxies. Secondly, the advent of upcoming surveys spanning the radio-to-optical spectrum, such as those facilitated by the Vera Rubin Observatory and the *Euclid* mission in the IR-optical range, alongside projects like SKA in radio astronomy and its precursors like LOFAR, ASKAP, and MeerKAT, enables future investigations on larger and higher-redshift samples of RLAGN. This advancement in data collection will facilitate comprehensive analyses of a more extensive set of RLAGN, allowing for robust statistical examinations of the AGN population. Additionally, it will better delineate the distinctive physical attributes distinguishing RLAGN from RQAGN.



# CHAPTER 3

---

## INTRODUCTION TO MACHINE LEARNING

---

This chapter introduces the theoretical aspects of Artificial Neural Networks ([ANNs](#)) and Machine Learning ([ML](#)). I begin by discussing the different types of algorithms used in ML. Next, I delve into the behavior of a single artificial neuron, covering its mathematical definition, learning process, and its capabilities and limitations. I then explore how the utilization of multiple neurons within an ANN enables us to create an algorithm capable of approximating any function. After that, I introduce a more advanced ANN architecture, known as Convolutional Neural Network ([CNN](#)), which is specifically designed for efficient image processing. I define the convolution operation and explain how it can be incorporated into an ANN layer. Subsequently, I explore the training process of a CNN architecture. Lastly, I will illustrate the relevance of CNN in astrophysics by showcasing numerous applications within the field. To write this chapter, I used the following works as references: [Hastie et al. \(2001\)](#), [Bishop \(2007\)](#), [Shalev-Shwartz & Ben-David \(2014\)](#), and [Goodfellow et al. \(2016\)](#)

### Contents

---

<b>3.1</b>	<b>The different types of ML algorithm</b> . . . . .	<b>56</b>
3.1.1	Definitions . . . . .	56
<b>3.2</b>	<b>Artificial Neuron</b> . . . . .	<b>58</b>
3.2.1	The first Artificial Neurons . . . . .	58
3.2.2	Training a neuron . . . . .	59
3.2.3	The bias node . . . . .	60
<b>3.3</b>	<b>Artificial Neural Network</b> . . . . .	<b>60</b>
3.3.1	The perceptron algorithm . . . . .	60
3.3.2	Multi Layers perceptron . . . . .	61

3.3.3	Back propagation . . . . .	63
3.3.4	Overfitting . . . . .	64
<b>3.4</b>	<b>Network optimization . . . . .</b>	<b>65</b>
3.4.1	Input normalization . . . . .	65
3.4.2	Weight initialization . . . . .	65
3.4.3	Batch size . . . . .	66
3.4.4	Learning rate . . . . .	67
3.4.5	Momentum . . . . .	67
3.4.6	Dropout . . . . .	68
3.4.7	MC dropout . . . . .	68
<b>3.5</b>	<b>Convolutional Neural Network . . . . .</b>	<b>68</b>
3.5.1	Information in 2D images . . . . .	68
3.5.2	Convolution operation . . . . .	72
3.5.3	Convolutional layer . . . . .	73
3.5.4	Multiple convolutional layers . . . . .	73
3.5.5	Activation function . . . . .	74
3.5.6	Pooling layers . . . . .	76
3.5.7	Weights initialization and bias . . . . .	77
3.5.8	Back propagation in a CNN . . . . .	77
3.5.9	Batch normalization . . . . .	80
3.5.10	The use of CNN in astronomy . . . . .	80

---

## 3.1 The different types of ML algorithm

### 3.1.1 Definitions

Machine learning is a field of artificial intelligence that focuses on developing algorithms that can automatically learn from data and make predictions or decisions that have not been explicitly programmed. The goal of machine learning is to enable computers to learn and improve their performance on a specific task based on experience. There exist 4 different types of ML methods that differ in the way the learning part is handled: supervised learning, unsupervised learning, semi-supervised learning, and reinforcement learning.

**Supervised learning** is a type of machine learning where an algorithm learns to map input data to output labels based on examples of input-output pairs provided in a labeled dataset. The goal is to create a model that can accurately predict the output label for new input data that it has not seen before. In supervised learning, the labeled dataset is typically split into two

parts: a training set and a test set. The algorithm learns from the training set by adjusting its internal parameters to minimize the difference between its predicted output labels and the true labels in the training set. The model is then evaluated on the test set to assess its ability to generalize to new, unseen data. Supervised learning is commonly used in applications such as image recognition, speech recognition, natural language processing, and many others. Some popular supervised learning algorithms include linear regression, logistic regression, decision trees, random forests, support vector machines, and neural networks. One of the advantages of supervised learning is that it can produce highly accurate models that can generalize well to new, unseen data. However, it requires labeled data to train the model, which can be time-consuming and expensive to obtain. It is also important to ensure that the training data is representative of the real-world distribution of data that the model will encounter in practice.

**Unsupervised learning** is a type of machine learning where an algorithm learns patterns and relationships in the data without explicit labels or supervision. In other words, the algorithm is not given specific output values to predict, but instead, it seeks to identify underlying structures and patterns in the input data on its own. In unsupervised learning, the algorithm is typically presented with a dataset, and its task is to discover meaningful patterns within the data. Clustering is a common unsupervised learning technique that aims to group data points based on their proximity or similarity. In clustering, the programmer defines a notion of proximity, usually based on a distance measure in the input space. The goal is to identify meaningful clusters or groups within the data. Various clustering algorithms, such as k-means, hierarchical clustering, and DBSCAN, use different strategies to partition the data into cohesive clusters. By analyzing the distances or similarities between data points, clustering algorithms enable the exploration of data structure, identification of subpopulations or patterns, and facilitate further analysis. The choice of distance measure depends on the specific problem and data characteristics. Clustering plays a crucial role in data exploration, anomaly detection, clustering, and many other applications in machine learning and data analysis. Another common technique is dimensionality reduction, which involves mapping the input space to a lower-dimensional parameter space. The objective is to simplify the data representation while minimizing the loss of relevant details. By reducing the dimensionality, the computational complexity of subsequent analyses can be reduced, and the visual exploration and interpretation of the data become more feasible. Techniques such as principal component analysis, linear discriminant analysis, and t-distributed stochastic neighbor embedding are commonly employed for dimensionality reduction tasks. These methods aim to capture the essential structure and relationships within the data, allowing for more efficient analysis and visualization in various applications such as data compression and visualization. One of the benefits of unsupervised learning is that it can help to uncover hidden structures and relationships in the data that may not be apparent through manual inspection. However, the lack of explicit labels can make it difficult to evaluate the performance of the algorithm and determine the relevance of the groups that it constructs.

**Semi-supervised learning** is a type of machine learning that combines elements of both supervised and unsupervised learning. In semi-supervised learning, the algorithm is given a partially labeled dataset, where only some of the examples have known output labels, and the rest are unlabeled. The goal of semi-supervised learning is to leverage the small amount of labeled data to improve the accuracy and generalization of the model, while also utilizing the much larger amount of unlabeled data to uncover underlying patterns and relationships in the data.

One common approach to semi-supervised learning is to first train an unsupervised learning algorithm on the unlabeled data to extract features or representations that capture the structure of the data. These representations can then be used to improve the performance of the supervised learning algorithm on the labeled data. Semi-supervised learning is commonly used in applications where it may be difficult or expensive to obtain a fully labeled dataset, but where some labeled data is available to guide the learning process. Some examples of applications where semi-supervised learning is used include speech recognition, image classification, and natural language processing. One of the benefits of semi-supervised learning is that it can improve the accuracy of the model without requiring a large amount of labeled data. However, it can also be more complex to implement and may require more computational resources than traditional supervised or unsupervised learning.

**Reinforcement learning** is a type of machine learning where an algorithm learns to make decisions by interacting with an environment and receiving feedback in the form of rewards or penalties for each action it takes. The goal of the algorithm is to maximize the total reward it receives over time. The reinforcement learning algorithm learns through trial and error, exploring different actions in the environment and adjusting its behavior based on the feedback it receives. The algorithm typically uses a technique called value iteration to estimate the value of different actions and choose the best one to take in each situation. Reinforcement learning is commonly used in applications such as robotics, game playing, and autonomous driving, where the algorithm must learn to make decisions in complex and dynamic environments. It is also used in recommendation systems, natural language processing, and other fields where the goal is to optimize a sequence of actions over time. One of the strengths of reinforcement learning is that it can learn to make decisions in situations where there is no pre-existing dataset or labeled training examples. However, it can also require a large number of computational resources and can be difficult to tune and optimize.

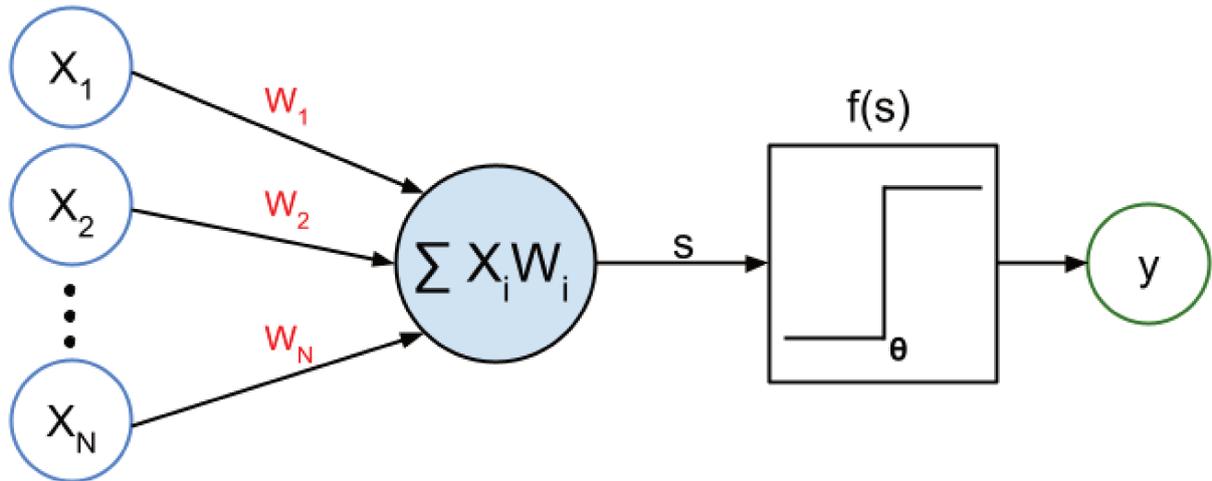
## 3.2 Artificial Neuron

### 3.2.1 The first Artificial Neurons

The beginning of research on ML and ANN can be attributed to [McCulloch & Pitts \(1943\)](#). They attempted to find a mathematical model that can describe the behavior of biological neurons. They modeled a neuron as a node that received an input vector  $X$  of  $N$  dimensions and calculated its scalar product with a vector composed of weights  $W$ .

$$s = \sum_{i=1}^N X_i W_i \quad (3.1)$$

The scalar product result is passed through an activation function to produce the output  $y$ . The simplest example of an activation function is the step function which means that the output is either 0 or 1, depending on whether the scalar product exceeds a certain threshold  $\theta$ .



**Figure 3.1:** Schema of an artificial neuron. The  $X_1$  to  $X_N$  are the inputs and the  $W_i$  their associated weights. The sum of the product of the inputs by their weights ( $s$ ) is computed. Then, we apply a threshold step function ( $f$ ) to it which produce  $y$ . Inspired by the perceptron representation of (Rosenblatt 1958).

$$y = f(s) = \begin{cases} 1 & \text{if } f(s) > \theta \\ 0 & \text{otherwise} \end{cases} \quad (3.2)$$

This neuron can be used to make linear separations in the data set. A schematic view of such binary neuron is illustrated in Fig 3.1.

## 3.2.2 Training a neuron

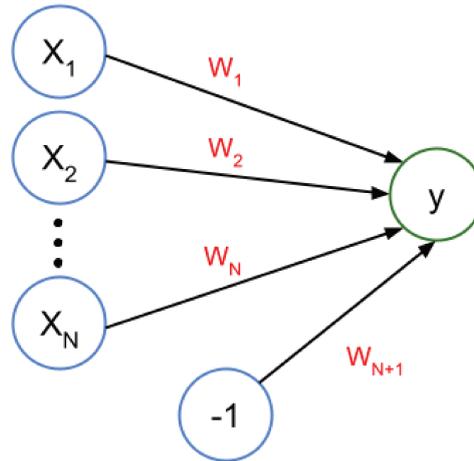
In a supervised ML context, we have a training data set of inputs with known outputs. Let's define a function called "loss function" ( $L$ ) that compares the output of the network to the expected one. A common choice for the loss function is the sum of squares error:

$$L(y, t) = \frac{1}{2}(y - t)^2 \quad (3.3)$$

with  $y$  the neuron output and  $t$  the target. The loss function serves to quantify how good the neural network is to predict the expected outcome. The more the output is accurate, the smaller the loss function is. Thus, what we want is to find the set of weights that minimize the loss function. To train our neural network, for each input in the training data set, the loss function is computed (forward step). Then, we use the gradient of the loss function to slightly modify the weights (update step), such as:

$$W_i \leftarrow W_i - \eta \frac{\partial L}{\partial W_i} \quad (3.4)$$

Where  $\eta$  is the learning rate, a parameter that can be adjusted depending on the problem (see Sect. 3.4.4). The forward and update steps are repeated for each input in the training data set. One pass of all the training data through the algorithm is called an epoch. The weights are initialized by small random values as explained later (see Sect. 3.4.2).



**Figure 3.2:** Schematic view of a binary neuron with a bias node

### 3.2.3 The bias node

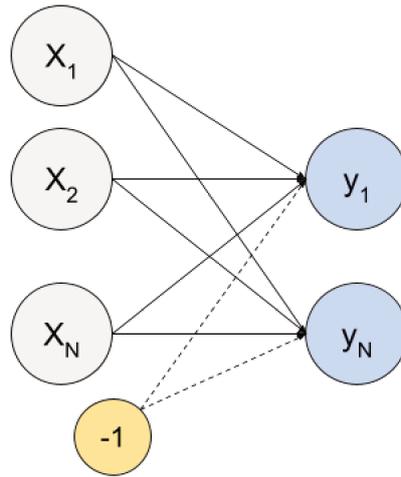
By looking at Eqs. 3.1 and 3.4, we can see that something is missing to properly train a Neural Network (NN). Equation 3.1 describes the set of functions that are linear and have their origin coordinates in  $(0,0)$ . To allow more flexibility in the separation of the classes, an additional neuron is added to the inputs of the neurons. Its value is fixed to a constant value  $(-1)$  by convention) and is connected to the neuron as any other input (see Fig. 3.2). This has for effect of allowing non-zero-origin. Adding such a neuron has the same effect as having an adaptive value for the threshold  $\theta$ . This additional neuron is called a "bias node" because the inputs are sometimes called nodes.

There are two common approaches regarding the bias. The first method initializes the bias at zero and is free to change during the training while the associated weight is fixed at one. Conversely, the second method fixes the bias to a small value but the weight is adaptive. The comparisons between the different methods of weight initialization are mainly empirical but the two methods of bias implementation are supposed to be equivalent.

## 3.3 Artificial Neural Network

### 3.3.1 The perceptron algorithm

As we saw in the previous section, a single neuron is only capable to perform a linear separation. To address more complex problems, we can simply use more neurons at the same time. In 1957, the American Engineer Frank Rosenblatt invented the first ANN, the perceptron (Rosenblatt 1958). It consists of using multiple neurons, each connected to all the inputs but has no connection with the other ones. We call this group of neurons a layer. The Eqs. 3.1, 3.2 and 3.4 become:



**Figure 3.3:** Schematic view of a perceptron network. The gray circles represent the input nodes for an object of the data. The blue circles represent the binary neurons. The weights between the two are represented by the black lines. The bias node is in yellow and its weights are the dashed lines.

$$s_j = \sum_{i=1}^{N+1} X_i W_{ij} \quad (3.5)$$

$$y_j = f(s_j) = \begin{cases} 1 & \text{if } f(s_j) > \theta \\ 0 & \text{otherwise} \end{cases} \quad (3.6)$$

$$W_{ij} \leftarrow W_{ij} - \eta \frac{\partial L}{\partial W_{ij}} \quad (3.7)$$

where  $j$  is the index of the neurons. The sum goes until  $N+1$  to take into account the bias node.

### 3.3.2 Multi Layers perceptron

A perceptron neuron is only capable of parameterizing a linear splitting in the input data space. This way of classification is very limited and too simple to work on more complex data. To upgrade this algorithm, we can make three changes.

The first upgrade we can make is to change the activation function of the neurons. By choosing a step function as an activation function, we limit the outputs of the neuron to either 0 or 1. Instead, we can choose a continuous and differentiable function. The neuron outputs can now be real between 0 and 1 instead of just 0 or 1. A common activation function is the sigmoid (Rumelhart et al. 1986):

$$f(s) = \frac{1}{1 + e^{-\beta s}} \quad (3.8)$$

Where  $\beta$  is a parameter that defines the steepness of the curve. Fig. 3.4 illustrated examples of sigmoid function and their derivative for  $\beta=1$  and 0.5

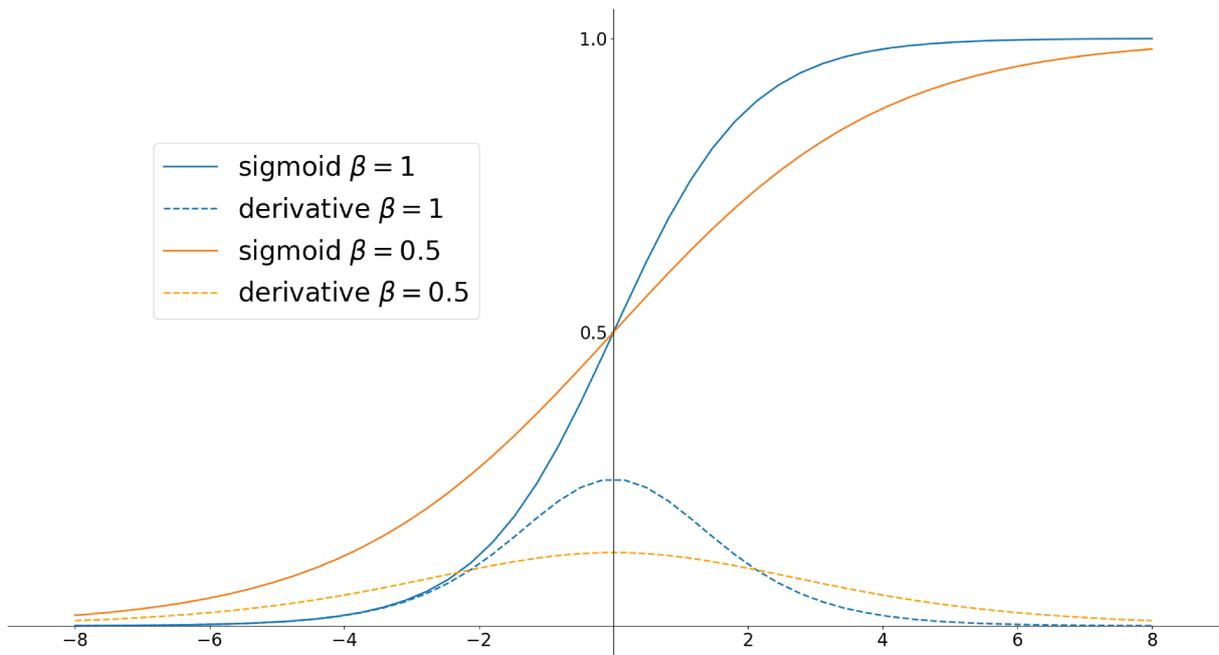


Figure 3.4: Example of sigmoid functions and their derivatives for  $\beta = 1$  and  $\beta = 0.5$ .

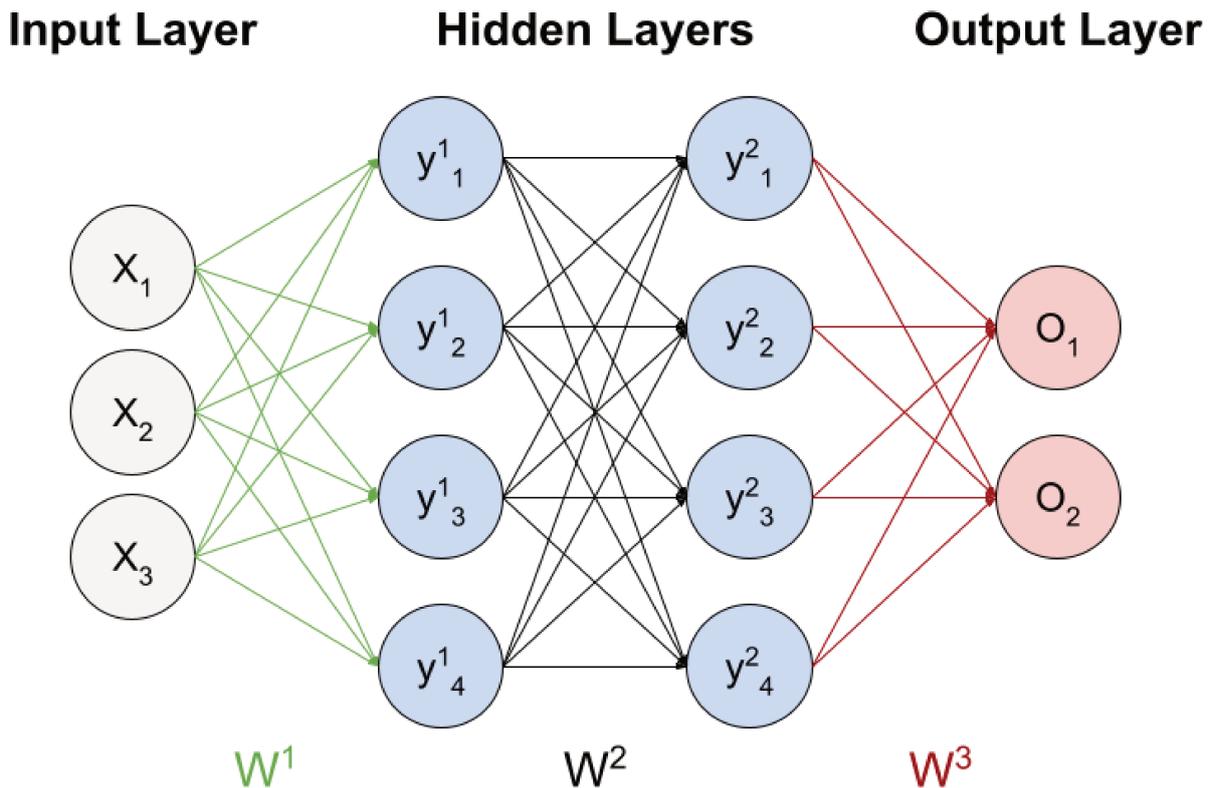


Figure 3.5: Schema of an MLP with an input layer of 3 neurons (gray), 2 hidden layers (blue) composed of 4 neurons each, and an output layer (red) with 2 neurons.  $W^1$  (in green) are the weights between the input layer and the first hidden layer,  $W^2$  (black) are the weights between the two hidden layers, and  $W^3$  (red) are the weights between the second hidden layer and the output layer.  $X_i$  represents the inputs,  $y_i^1$  is the activation function of the first hidden layer,  $y_i^2$  same but for the second hidden layer, and  $o_i$  is the activation of the output layer. Bias nodes are omitted to simplify the schema.

The second change is to add multiple layers to the algorithm. From the second layer to the last, each neuron is connected to all of the neuron outputs of the previous layers. We call the input nodes the input layer, the last layer that computes the desired output is called the output layer, and every layer in between is called a "hidden layer". A network can be considered deep when it has at least one hidden layer. The number of hidden layers and the number of neurons within them can be modified depending on the context. The more complex a problem is, the more and/or bigger layers can be needed. This kind of neural network is called a Multi-Layer Perceptron (MLP). An example of MLP is shown in Fig. 3.5.

At each neuron, a sigmoid is applied to a linear combination of the inputs. It means that, at each layer, a sigmoid function is used on a linear combination of sigmoid functions. The successive application of sigmoid produces a non-linear result that gains in complexity with each layer. In 1989, Cybenko demonstrated that the non-linear combination of a finite number of sigmoidal functions can approximate any continuous function (Cybenko 1989). It means that this type of network is a "Universal Function Approximator". He also demonstrated the Universal Approximation Theorem which says that any continuous function can be approximated by a neural network composed of only one hidden layer with enough neurons.

### 3.3.3 Back propagation

When adding additional layers to a NN, the weights actualization needs to be changed as the loss function is only available for the output of the last layer. The "Backpropagation" algorithm (Rumelhart et al. 1986) computes recursively an error gradient descent from the output layer to the first layer.

The idea is that starting from the output layer, an error is propagated through the network then the weights correction of each layer is calculated one by one. Similarly to Eq. 3.4, the weight  $W_{ij}$  of a layer  $l$  is updated by the formula:

$$W_{ij}^l \leftarrow W_{ij}^l - \eta \frac{\partial L}{\partial W_{ij}^l} \quad (3.9)$$

where  $j$  represents the index on the neurons in the  $l$ -th layer, and  $i$  is the index on the inputs for this neuron. The gradient  $\frac{\partial L}{\partial W_{ij}^l}$  can be develop such as:

$$\frac{\partial L}{\partial W_{ij}^l} = \frac{\partial L}{\partial f_j^l} \frac{\partial f_j^l}{\partial s_j^l} \frac{\partial s_j^l}{\partial W_{ij}^l} \quad (3.10)$$

Where  $s_j^l$  is the weighted sum of the inputs (Eq; 3.5) and  $f_j^l$  is the result of its activation function. If we define the local error of the  $j$ -th neuron of a layer  $l$  by  $\delta_l(j) = \frac{\partial L}{\partial s_j^l}$ , we can write the  $W_{ij}$  update of the same layer by:

$$W_{ij}^l \leftarrow W_{ij}^l - \eta \frac{\partial s_j^l}{\partial W_{ij}^l} \delta_l(j) \quad (3.11)$$

where  $\delta_l(j)$  is:

$$\delta_l(j) = \frac{\partial L}{\partial s_j^l} = \frac{\partial L}{\partial f_j^l} \frac{\partial f_j^l}{\partial s_j^l} = \frac{\partial f_j^l}{\partial s_j^l} \sum_k^{N_{l+1}} W_{kj}^{l+1} \delta_{l+1}(k) \quad (3.12)$$

With  $N_{l+1}$  is the number of neurons in the  $l + 1$  layer. When the activation function of each layer is a sigmoid, this can be simplified by

$$\frac{\partial s_j^l}{\partial W_{ij}^l} = f_j^{l-1} \quad (3.13)$$

The activation function is also simplified:

$$\frac{\partial L}{\partial f_j^N} = f_j^N - Y_j \quad (3.14)$$

Thus, we obtained:

$$\frac{\partial f_j^l}{\partial s_j^l} = \beta f_j^l (1 - f_j^l) \quad (3.15)$$

When a neural network has many layers, the gradients of the loss function have to be propagated back through each layer during the backpropagation using the equations above. As we can see in Eq. 3.15, the derivative of the activation function will always be smaller than one. This causes the gradient to diminish as it is multiplied at each layer during error propagation, making it challenging to update the weights of early layers effectively. This issue is called the vanishing gradient.

The vanishing gradient problem can lead to slow convergence during training, as well as poor performance of the neural network on the test data. In extreme cases, the gradients can become so small that the network stops learning altogether. The vanishing gradient problem is particularly relevant for deep neural networks that use activation functions that have very small derivatives, such as the sigmoid function. One way to alleviate this problem is to use activation functions that have constant derivatives, such as the Rectified Linear Uni (ReLU) function (see Fig. 3.10 and Sect. 3.5.5). Other techniques that can be used to address the vanishing gradient problem include batch normalization (Ioffe & Szegedy 2015a; Santurkar et al. 2018; Yang et al. 2019) and skip connections (Qian et al. 2022).

### 3.3.4 Overfitting

Overfitting occurs when a machine learning model becomes overly complex and fits the training data too closely, resulting in poor generalization to new data. To mitigate this issue, the dataset is typically split into three separate sets: a training set, a validation set, and a test set.

- The training set contains the majority of the data and is used to train the neural network. The validation set is used during training to regularly evaluate the performance of the neural network on data that was not used for training. The test set is used to assess the ability of the neural network to generalize to new, unseen data after training has been completed.

- During training, both the error on the training set and the validation set are regularly calculated. Typically, the training and validation error functions will decrease as training progresses. However, if the neural network starts to overfit the training data, the validation error function will begin to increase. This is when it is necessary to stop the training to prevent further overfitting.
- To assure optimal training, most of the data must be contained in the training set. However, the validation and test set must also be big enough to avoid smaller-number effects and representative of the feature space coverage of the problem.

## 3.4 Network optimization

### 3.4.1 Input normalization

Normalization is a common preprocessing step in ML that involves scaling the input data to a chosen range. The main reason for normalizing the input is to improve the performance of ML models. To optimize the learning process of a NN, the data need to have the same range of values. Otherwise, the model can give more importance to features with larger values, even if they are not necessarily more important for the prediction task. This can result in sub-optimal performance and can also make the model more sensitive to the range and units of measurement of the input data. Furthermore, the input data need to be scaled to the same level as the layer activation. If not, the weights can become too small or too big and will prevent their convergence.

There are several methods for normalizing input data. For example, the standardization method involves subtracting the mean and dividing by the standard deviation of the input features, which results in features with zero mean and unit variance. We can also subtract the mean and divide by the maximum absolute value to get data bounded between -1 and 1 and center around zero. Another method is the min-max scaling. It involves scaling the input features to a fixed range, typically between 0 and 1. It is done by subtracting the minimum value of the input from all values and then dividing the result by the difference between the maximum and minimum values.

### 3.4.2 Weight initialization

In ML, initializing weights refers to the process of setting the initial values of the weights in a NN or other ML model. Initialization is an important step in the training of ML models, as it can have a significant impact on the performance and convergence speed of the model.

On one hand, if the variance of the distribution of the initial weights is smaller than the values of the initial weights, the NN will converge more easily to a solution, but the resulting solution will be more sensitive to the inputs. On the other hand, a higher variance, compared to

the values of the initial weights, allows the network to explore a wider range of solutions, but it can make convergence more difficult. Therefore, to ensure an efficient learning process, the weights are initialized with small random values that have a variance of the same order.

There is no consensus for the best method of weight initialization as it strongly depends on parameters such as the NN depth or the choice of the activation function. A common method (Glorot & Bengio 2010) is to initialize the weights by a random value between  $-1/\sqrt{N}$  and  $1/\sqrt{N}$ , where  $N$  is the inputs vector dimension for the layers.

Combining randomness in the distribution of the initial weight and the stochastic learning method introduces randomness in the learning process. Two identical NN training on the same data will now start at different positions in the phase space, have a different path for optimizing their loss functions, and find a different local minimum.

### 3.4.3 Batch size

The frequency to which we update the weights of a NN is a parameter affecting the training performances. There are 3 common methods that can be used to update the weights during training: stochastic gradient descent, batch gradient descent, and mini-batch gradient descent.

In stochastic gradient descent, the weights are adjusted after each example is shown. The example is drawn with replacement, allowing for multiple weight updates per epoch. This accurate tracking of gradients may lead to noisy convergence. However, this method also allows the use of a larger learning rate which makes the stochastic gradient method usually faster to converge than the other two methods (Wilson & Martinez 2003). Nevertheless, having to update the weights for each example make the method less efficient regarding computational time cost.

The batch gradient method updates the weights once per epoch. It calculates the gradient of the loss function with respect to the weights using all the training samples at once and then updates the weights based on the average gradient over the entire dataset. This avoids the issue of the noisy gradient and requires less calculation time from the hardware. But the batch gradient descent can only take a single step per epoch, which must be in a straight line. It estimates the gradient only at a specific point in the weights phase space, and thus cannot follow curves in the error surface. Furthermore, since the gradient is more stable it is also more susceptible to getting stuck in local minima. When the training sample gets bigger, the batch gradient method must use a smaller learning rate in order for its learning to remain stable.

To address the downsides of both methods, we can use mini-batch gradient descent. This algorithm randomly divides the training set into small batches, called "mini-batches," at the beginning of each epoch. Then, for each mini-batch, the errors are used to update the weights. It allows for more frequent weight updates while reducing noise in the gradient descent through the averaging of multiple errors. This method strikes a balance between the advantages of stochastic and batch gradient descent, making it a commonly used approach in training neural networks.

### 3.4.4 Learning rate

The learning rate determines the magnitude of the weight updates during training. A high learning rate results in large weight updates, leading to faster convergence. However, large weight updates may cause the gradient to oscillate and prevent convergence to the narrow global minimum. Conversely, a low learning rate slows down the learning process, and the network may get trapped in a local minimum. Therefore, choosing an appropriate learning rate is crucial for effective training.

To address this issue, modern optimization algorithms use adaptive learning rates. These algorithms start with a high learning rate that guides the optimization towards the region of the weight space with lower error. The learning rate is then gradually reduced, allowing the optimization to converge towards narrower minima. The use of adaptive learning rates helps balance the trade-off between convergence speed and optimization stability.

### 3.4.5 Momentum

In ML, momentum is a technique used during gradient descent optimization to accelerate the convergence of the learning algorithm (Qian 1999). It is an extension of the standard gradient descent algorithm that takes into account the previous update steps of the weights. The momentum parameter, usually denoted by a symbol like  $\beta$ , is a hyperparameter that controls the contribution of the previous update steps to the current update. A high momentum value means that the previous updates have a greater influence on the current update, while a low momentum value means that only the current update is taken into account. A simple way to use this concept is to add the inertia term to the update equation such as:

$$W_{ij}^t \leftarrow W_{ij}^{t-1} - \Delta W_{ij}^t \quad (3.16)$$

where  $W_{ij}^t$  is the  $j$ -th weight of the  $i$ -th neuron of a layer at an epoch  $t$  and  $\Delta W_{ij}^t$  is the update of this weight. We can write  $\Delta W_{ij}^t$  as:

$$\Delta W_{ij}^t = \eta \frac{\partial E}{\partial W_{ij}^{t-1}} + \alpha \Delta W_{ij}^{t-1} \quad (3.17)$$

where  $E$  is the layer propagated error and  $0 < \alpha < 1$  is a hyperparameter used to scale the momentum.

Momentum can help the optimization algorithm to overcome local minima and saddle points in the loss landscape, and can also help to smooth out noisy gradients. By incorporating information about previous weight updates, momentum can help the optimization algorithm to move more efficiently toward the optimal solution.

There are several variations of the momentum algorithm, including the Nesterov accelerated gradient, which is a modification of momentum that takes into account the expected gradient at the next time step. This can lead to faster convergence and improved performance.

### 3.4.6 Dropout

Regularization refers to all the methods that are used to prevent overfitting and regularize learning from particular neurons. Dropout is a regularization technique introduced by [Srivastava et al. \(2014\)](#). It works by randomly dropping out (setting to zero) a fraction of the neurons activations in a given layer during training. This forces the network to learn redundant representations and prevents any one neuron from becoming too important. In other words, it helps to prevent complex co-adaptations between neurons, which can lead to overfitting.

During training, each neuron in the layer is retained with a probability independently of the other neurons. At test time, all neurons are used, but their outputs are scaled down by the probability of being kept, to compensate for the fact that more neurons are active during training than during testing.

### 3.4.7 MC dropout

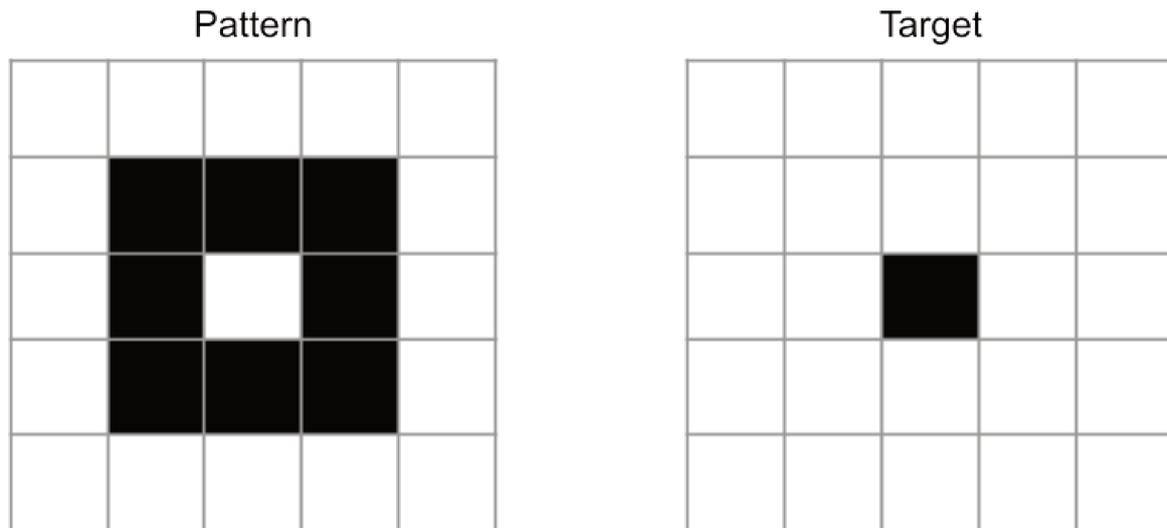
An alternative approach to utilizing dropout is to maintain dropout activation during the prediction phase. It is equivalent to an approximation of the probabilistic deep Gaussian process ([Damianou & Lawrence 2013](#)). For each inference, a different set of weights will be active. This is equivalent to train at the same time multiple networks that partially share their weights. Thus, each inference will give a different result. So, making multiple inferences for the same input will give a probability distribution on the predicted value, similar to what is produced by a Monte Carlo Markov Chain process ([Srivastava et al. 2014](#)). This distribution can be studied to characterize the model uncertainties for that input. This method is called Monte-Carlo (MC) dropout.

Bayesian NNs ([MacKay 1992](#)) also give a probability distribution as an output. They do so by using probability distributions for the weights. It has been demonstrated that MC dropout and Bayesian NNs give similar results. However, MC dropout is much more efficient in terms of computational performance ([Blundell et al. 2015](#); [Gal & Ghahramani 2016](#)).

## 3.5 Convolutional Neural Network

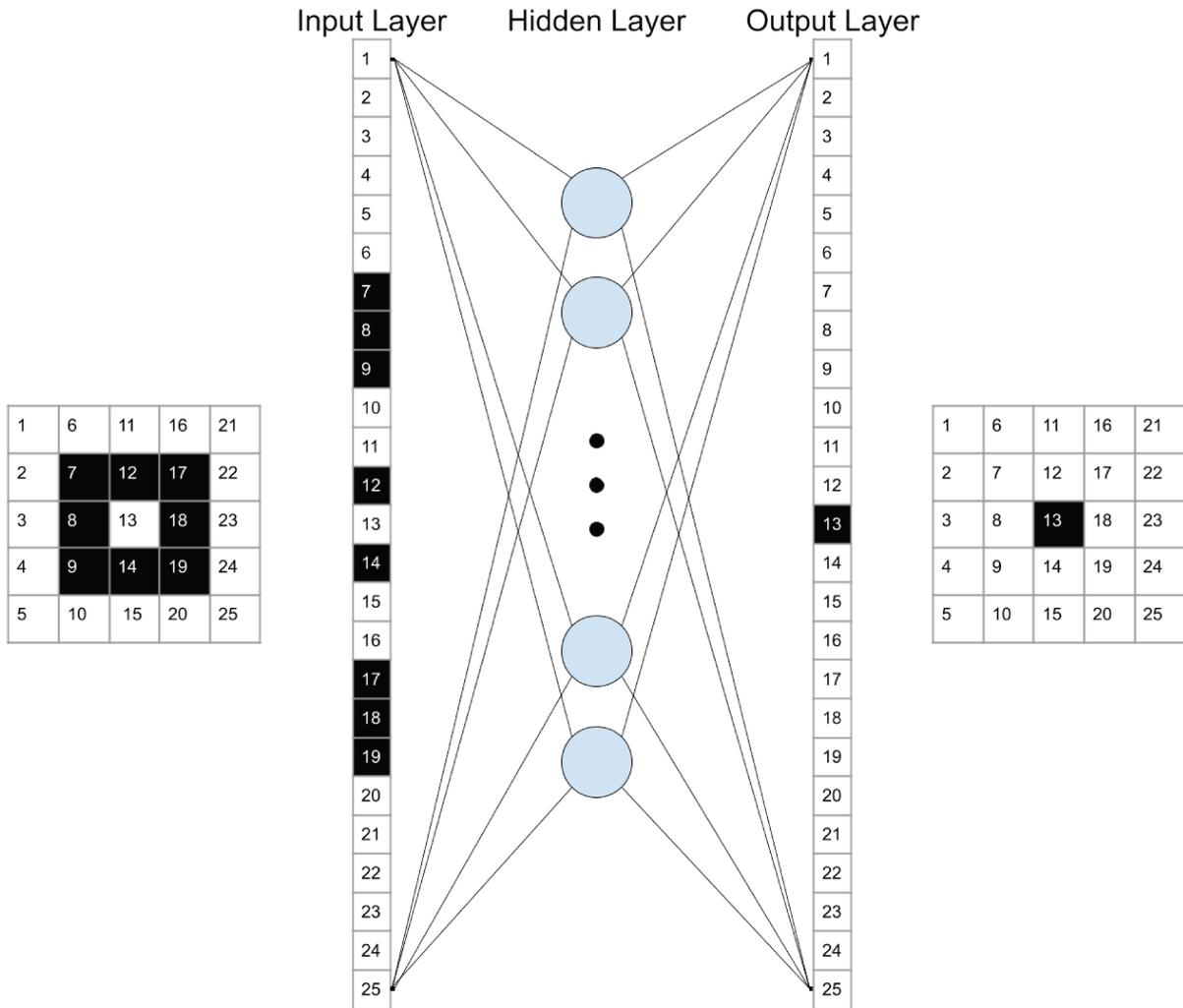
### 3.5.1 Information in 2D images

Images can be described as two-dimensional arrays of pixels, serving as discrete representations of decomposed information. The process of image recognition revolves around identifying cohesive information within these pixel arrays and associating it with higher-level abstractions corresponding to the depicted objects. Spatial coherence plays a crucial role in extracting meaningful information from images.



**Figure 3.6:** Example of a square pattern (left) and its corresponding localized target (right).

To exemplify the challenges associated with spatial coherence, let us consider a scenario where the objective is to develop an algorithm capable of recognizing and locating a specific pattern depicted in Fig. 3.6 if present. In this example, we opted for a 2D binary image as the input. Additionally, we represented the detection task using a binary image of the same dimensions where the pattern's presence is indicated by an activated pixel positioned at the center of the original input image.



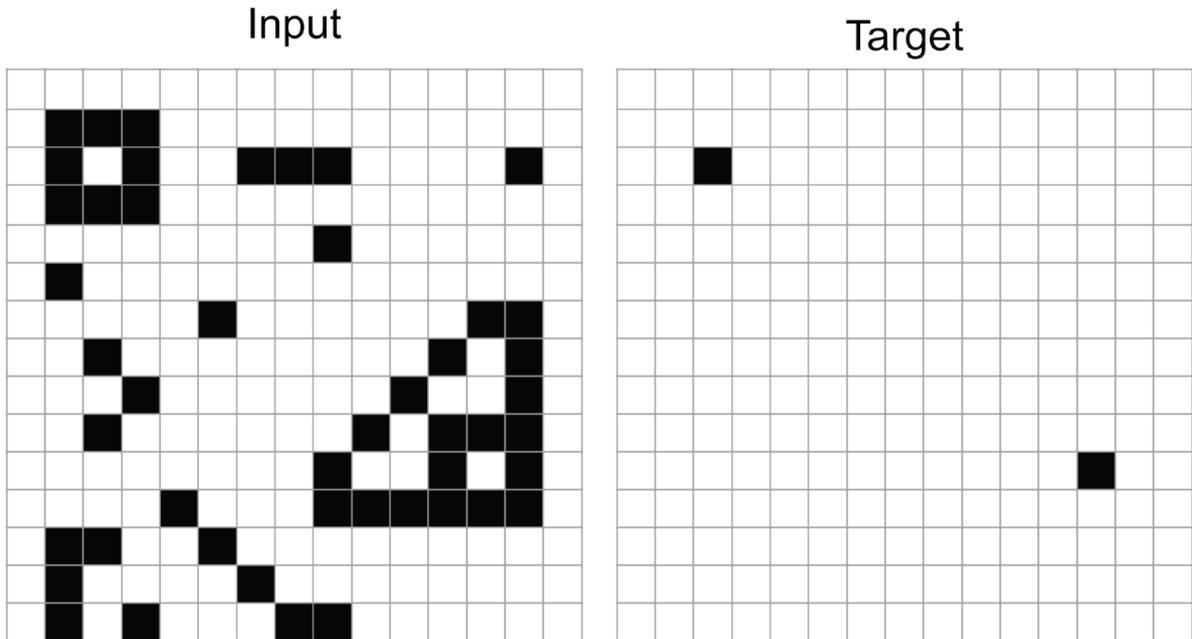
**Figure 3.7:** Schematic view of how a 2D image of a squared pattern is flattened to become a dense NN input layer and how an output layer is reshaped into the target 2D image. From left to right we find, the pattern 2D image, its flattened representation, the dense layer, the output layer, and its 2D image. For better visualization, each pixel is labeled by a number (from 1 to 25).

Having gained an understanding of how ANNs function in the preceding sections, it is natural to explore their applicability in this task. To do so, a straightforward approach would involve employing an ANN with a single hidden layer, which takes the flattened image as input and produces an output vector of the same length (see Fig. 3.7). This configuration allows the network to learn the precise position of the pattern within the image.

However, such a network has a high number of weights which makes its training very long, and a lot of examples are needed to properly constrain the weights. Furthermore, it means that for each pixel, every specific case must be in the training sample, and so the network is very inefficient at generalizing the problem.

Now, let's consider a more complicated case (see Fig. 3.8). We still want to recognize and locate the same pattern on a bigger image where the pattern can appear multiple times. We also added irrelevant information that can be interpreted as noise or other non-studied patterns. Since we want our network to only detect the square pattern, we don't need to show examples of all the possible cases of irrelevant information combinations on all other pixels. Each output neuron has the responsibility of identifying whether it should activate or not based on the presence or absence of the pattern. They have to do this independently, without any assistance from other neurons, by learning to recognize the pattern at all positions in the image. This means that the network needs to learn the same task multiple times, once for each possible position of the pattern in the image. It is essential to have a diverse training dataset where all pixels are activated and deactivated at least once to cover all possible scenarios. This ensures that the neural network becomes proficient at detecting the pattern accurately in various positions within the image. A fully connected network should still be able to perform such a task. However, it is less likely to be more time efficient time-wise than an algorithm that searches for a square centered around each pixel of the image.

However, a more clever solution is possible. Since we look for only one pattern, we could save time and effort by learning this pattern once and searching efficiently for it in the image. Thus, we want to find an operation capable of detecting the pattern in a smaller part of the image and that is equivariant by translation. In practice, such operations are called convolutions and they greatly reduce the number of weights needed to perform pattern recognition by defining a single "filter" that is applied to all possible locations in the image.



**Figure 3.8:** Example of an image containing 2 squared patterns as well as other non-null pixels. On the left, we find the image in itself on the right, the positions of the squares.

### 3.5.2 Convolution operation

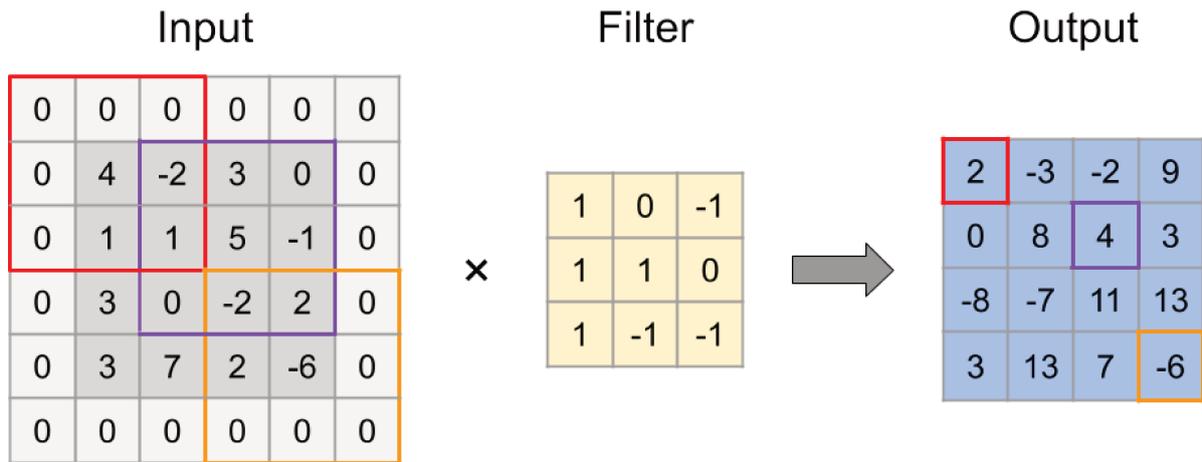
In ML, convolution is a mathematical operation that involves applying a small kernel filter to an input signal or image to produce a feature map as output.

The convolution operation involves sliding the kernel over the input signal and computing the dot product between the kernel and the corresponding input local region. After each application, the kernel is shifted along one axis. How much the kernel moves between each operation is defined by a parameter called the stride. When the kernel has reached the end of the axis, it starts the process again but with a shift along the other axis as defined by the stride. The resulting output of this operation is a new array that contains the aggregated features of the input across all positions. The size of the kernel and the stride are hyperparameters that can be adjusted to control the size and resolution of the output feature map. The convolution operation reduces the size of the output feature map. To control the size of the output image extra rows and columns of zeros around the edge are added to the input data or image. This increases the spatial dimensions of the input and allows the convolutional filter to fit over the edge pixels of the input, preserving the spatial dimensions of the output feature map. This process is called padding. The output dimensions are given by

$$d_{out} = \frac{d_{in} - f_s + 2P}{S} + 1 \quad (3.18)$$

with  $d$  the input/output width/height,  $f_s$  the filter size,  $S$  the stride and  $P$  the padding.

The objective of the Convolution operation is to extract features/patterns, such as edges, from the input image. For example, we showcase a convolution operation of a 3x3 filter with a stride of 1 over a 4x4 input with a padding of 1 (Fig.3.9).



**Figure 3.9:** Example of a convolution operation. Here a 3x3 filter with a stride of 1 is applied on a 4x4 input with a padding of 1. Three sub-region products and their corresponding output pixels are showcased by the red, purple, and orange outlines.

### 3.5.3 Convolutional layer

A convolutional layer is defined by its number of filters, the dimensions of these filters, a stride value, and an activation function.

A convolution operation is calculated on the input for each filter. However, the output generated after applying each filter is subsequently transformed using an activation function. The result of this process is called an activation map. The filter values of a convolutional layer are weights that are learned during the training process.

The output of a convolutional layer is characterized by three dimensions. As demonstrated earlier, each filter convolves with the input data, resulting in a 2D activation map with dimensions of  $w_{out} \times h_{out}$  as described by Eq. 3.18. Considering the presence of  $n_f$  filters, the output of a convolutional layer is obtained by stacking the  $n_f$  activation maps, resulting in an output volume with dimensions of  $w_{out} \times h_{out} \times n_f$ .

Some images have a third dimension ( $d$ ), like RGB images for example. In that case, each convolution of a filter have also a third dimension of the same size. For instance, if your input has a dimension of  $256 \times 256 \times 3$  and you apply to it a  $3 \times 3$  convolution with 16 filters, the filters size will be  $3 \times 3 \times 3$ . Then, if you apply a second convolution of 32,  $3 \times 3$  filters, this convolution filters will have a size of  $3 \times 3 \times 16$ , and so on. This process continues, and each new convolution layer increases the depth of the filters based on the number of filters in the previous layer.

### 3.5.4 Multiple convolutional layers

While a single convolutional layer can effectively identify numerous patterns when equipped with a big enough number of filters, its ability is restricted to patterns of the size corresponding to the filter. Consequently, it frequently struggles to efficiently learn intricate non-linear patterns

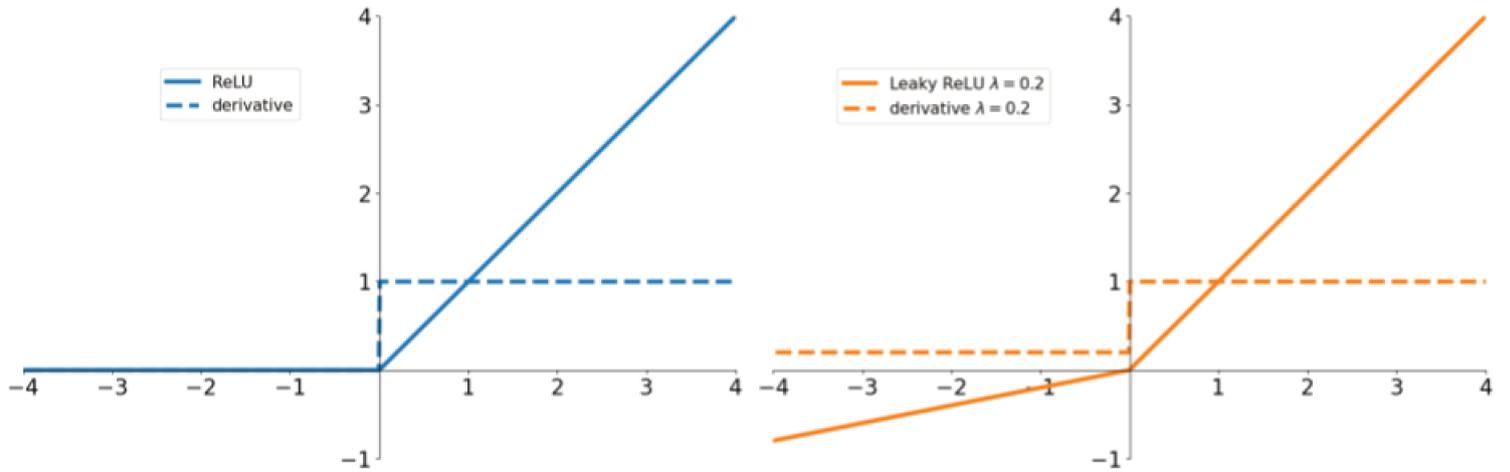
(Lecun et al. 1995). Thus, putting multiple convolutional layers one after the other is common. By using multiple convolutional layers, the network can learn hierarchical representations of the input data, allowing them to capture more complex patterns and relationships. The first convolutional layers detect low-level features such as edges, lines, and corners.

As the input data is passed through more convolutional layers, the network is able to detect larger, more abstract, and high-level features that combine the low-level features detected in previous layers. Additionally, using multiple convolutional layers can help to reduce the spatial dimensionality of the data, which can make the network more efficient and easier to train. By gradually reducing the spatial dimensions of the data, the network can learn more compact and efficient representations of the input data. Moreover, dimensionality reduction is a contributing factor in constructing more complex patterns. It enables small filters in the deeper layers of the network to put together information from a larger region of the input image into a single, smaller filter. This can help to reduce overfitting and improve the generalization performance of the model. Usually, the convolutional filter is linked to some fully connected layers before the output layer. This type of NN architecture is called a CNN.

In a classical CNN architecture, the initial layers typically employ a smaller number of larger filters. As the network progresses deeper, more filters are added, but the size of the filters is often reduced. This design choice serves a specific purpose. By using fewer but larger filters in the early layers, the network can capture low-level features with larger receptive fields. These larger filters are able to detect broader patterns and simple structures in the input data. As the network goes deeper, additional filters are introduced, allowing for the detection of more diverse and fine-grained features. Reducing the size of the filters in deeper layers serves two main advantages. Firstly, it helps to reduce the computational complexity of the network by reducing the number of parameters. Smaller filters require fewer computations, making the network more efficient. Secondly, the first layer can be larger because gradients do not need to propagate through it, allowing the use of larger filters without significant performance losses in training. Additionally, the initial filters function as basic image transformers, making the task easier for subsequent layers. As we move deeper into the network, the filters focus on higher-level combinations, capturing patterns closer to the objects to classify or detect. Gradually increasing the number of filters is essential as we approach the output, where specific patterns are needed for object detection. These patterns become non-sharable, requiring an increased number of filters. For instance, a pattern detecting sharp teeth is useful for classifying animals but not for identifying a car. In contrast, at the beginning, filters searching for edges or enhancing contrast are generally beneficial for all cases. Furthermore, this progressive increase in the number of filters and decrease in filter size allows the classical CNN architecture to learn hierarchical representations of the input data, starting with simple features and gradually building up to more complex and abstract representations. This design choice enables the network to effectively extract and utilize features at different scales and levels of complexity.

### **3.5.5 Activation function**

Adding convolutional layers before the fully connected network increases the depth of the network which means that the network is more susceptible to the vanishing gradient issue (see



**Figure 3.10:** Example of the ReLU (blue) and Leaky ReLU (orange) with a  $\lambda$  of 0.2.

Sect. 3.3.3). In CNN architecture, the most common choice of activation function is the ReLU (Nair & Hinton 2010):

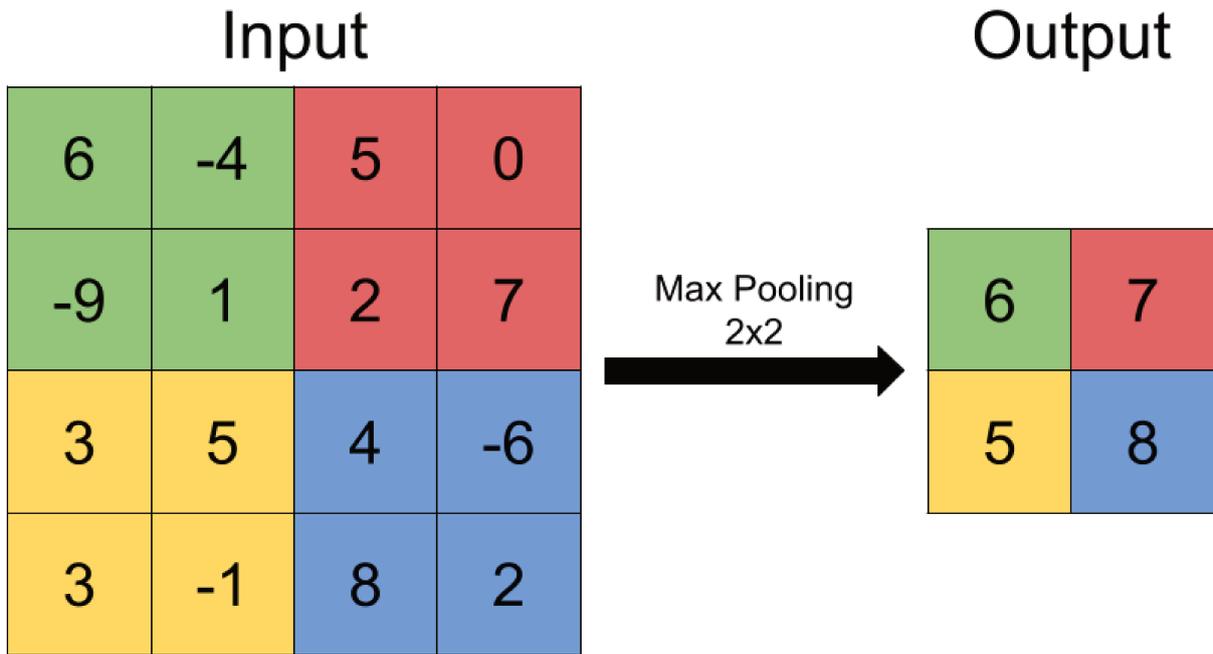
$$y_j = f(s_j) = \begin{cases} s_j & \text{if } s_j \geq 0 \\ 0 & \text{otherwise} \end{cases} \quad (3.19)$$

with the same notation as Eq.3.6. The ReLU function has multiple advantages. It introduces non-linearity in a simple but sufficient way to allow the network to learn more complex and abstract representations of the input data. It also produces a sparse output, as it sets all negative values in the input to zero. This sparsity can help to reduce the computational complexity of the network and prevent overfitting. Furthermore, it helps to mitigate the vanishing gradient problem thanks to its constant gradient (value of 1 when  $y > 0$ ). Another advantage of the ReLU function is that, in its linear regime, the function remains scale-invariant. This property ensures that the activation map exhibits the same difference for inputs between 1 and 10 as it does for inputs between 1001 and 1010. Consequently, ReLU has the ability to detect subtle distinctions across a broad range of input values, making it more versatile and facilitating network convergence. This stands as a notable advantage over activation functions like the sigmoid, where different value ranges may not offer the same level of contrast in their responses.

The ReLU function is not without downsides. Neurons that have a negative output will not be updated as they produce a zero output for all inputs. A variation of the ReLU function called the Leaky ReLU (Maas et al. 2013) is defined as:

$$y_j = f(s_j) = \begin{cases} s_j & \text{if } s_j \geq 0 \\ \lambda s_j & \text{otherwise} \end{cases} \quad (3.20)$$

where  $\lambda$  is a hyperparameter ( $0 < \lambda \ll 1$ ). The Leaky ReLU function introduces a slight slope to the function for negative inputs, allowing a small gradient to be propagated during backpropagation. The slight slope helps the neurons that are never activated to be slowly re-involved in the learning process. Fig.3.10 shows the graphic representation of the ReLU and Leaky ReLU functions. However, the vanishing gradient can affect the propagation to non-activated neurons since the derivative is equal to  $\lambda$ . Thus, the propagation remains primarily



**Figure 3.11:** Example of max pooling 2x2 onto a 4x4 matrix. The different colors indicate the sub-regions for which the pooling takes place.

driven by the activated neurons from the previous layer, which is, in most cases, desirable.

### 3.5.6 Pooling layers

In CNN architecture, convolutional layers are often combined with another type of layer called the pooling layer. They are defined by a kernel size and a pooling operation. It will perform the pooling operation over a region of the image and then slide with a stride equal to the kernel size in that dimension. Thus, it will decompose the input in usually non-overlapping sub-regions of the same size as the kernel. The most common pooling layer is the Max-Pooling one. Here, the pooling operation takes the maximum value of the pixels in the kernel. The pooling layers do not contain weights or activation functions, meaning they do not increase the number of learned parameters. An example of a max-pooling operation is represented in Fig. 3.11. Pooling is applied independently to each channel, thereby preserving the number of channels. In contrast, a convolutional layer combines the previous channels and produces as many output channels as the number of filters it employs. The pooling layer's purpose is to reduce the spatial dimensions of the input feature map while preserving important features. They are designed to select only the most relevant and non-redundant information from their input. Using this approach can significantly speed up the learning and inference processes of a CNN, with an effect on its performance that depends on the problem. Some networks will only have a small or negligible degradation of their performance while others may not work at all.

### 3.5.7 Weights initialization and bias

As seen previously, in a fully connected network, the weights need to be initialized correctly to stabilize and increase the network performance. We also saw that the choice of activation function, bias value, or even network architecture is linked to how the weights are initialized. The same concept also applies to convolutional layers. In fact, it is even more important in the context of deep networks where all of these parameters will have more impact on the prediction performance. When working with a deep network, the sigmoid function is no longer reliable enough and the ReLU (or Leaky ReLU) becomes a better choice than the sigmoid function. Furthermore, an inappropriate deep network weight initialization can prevent it from converging. The primary objective remains to balance the weights, ensuring they are small enough to maintain precision and stability, yet large enough to bring out distinct behaviors for each neuron in the network. The performance comparison between various weight initialization methods is largely based on empirical observations, with each method aiming to achieve a uniform initial weight variance across all layers of the network, independently of their size.

A common weight initialization for deep network is a uniform distribution scaled to the size of the previous layer  $n_{l-1}$  (Glorot & Bengio 2010):

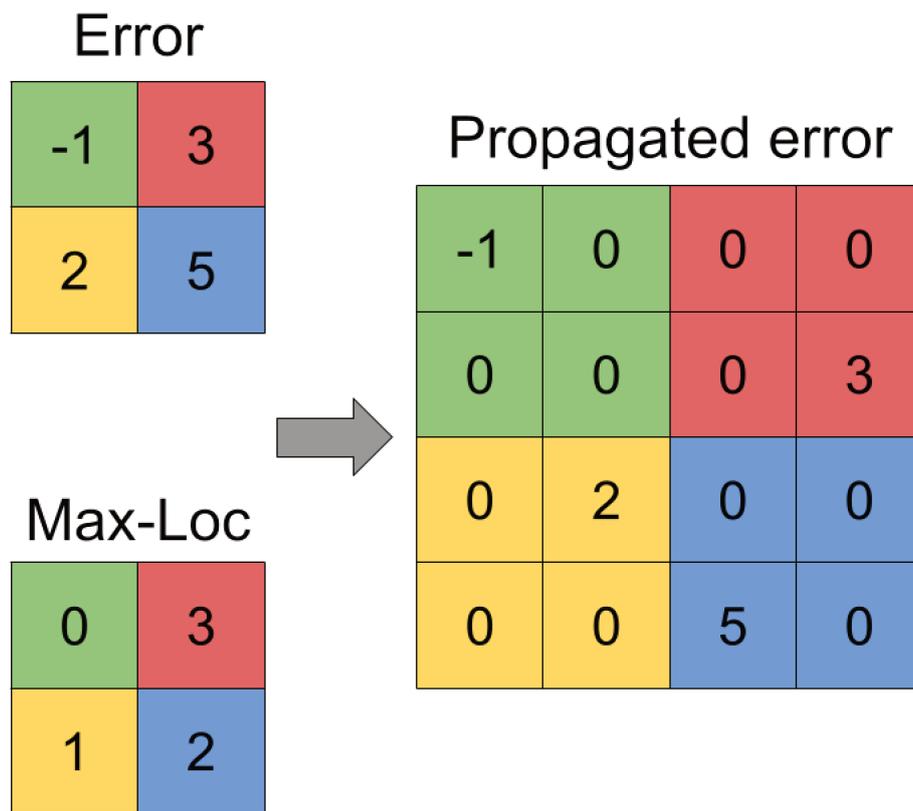
$$\sqrt{\frac{6}{n_{l-1} + n_l}} \quad (3.21)$$

### 3.5.8 Back propagation in a CNN

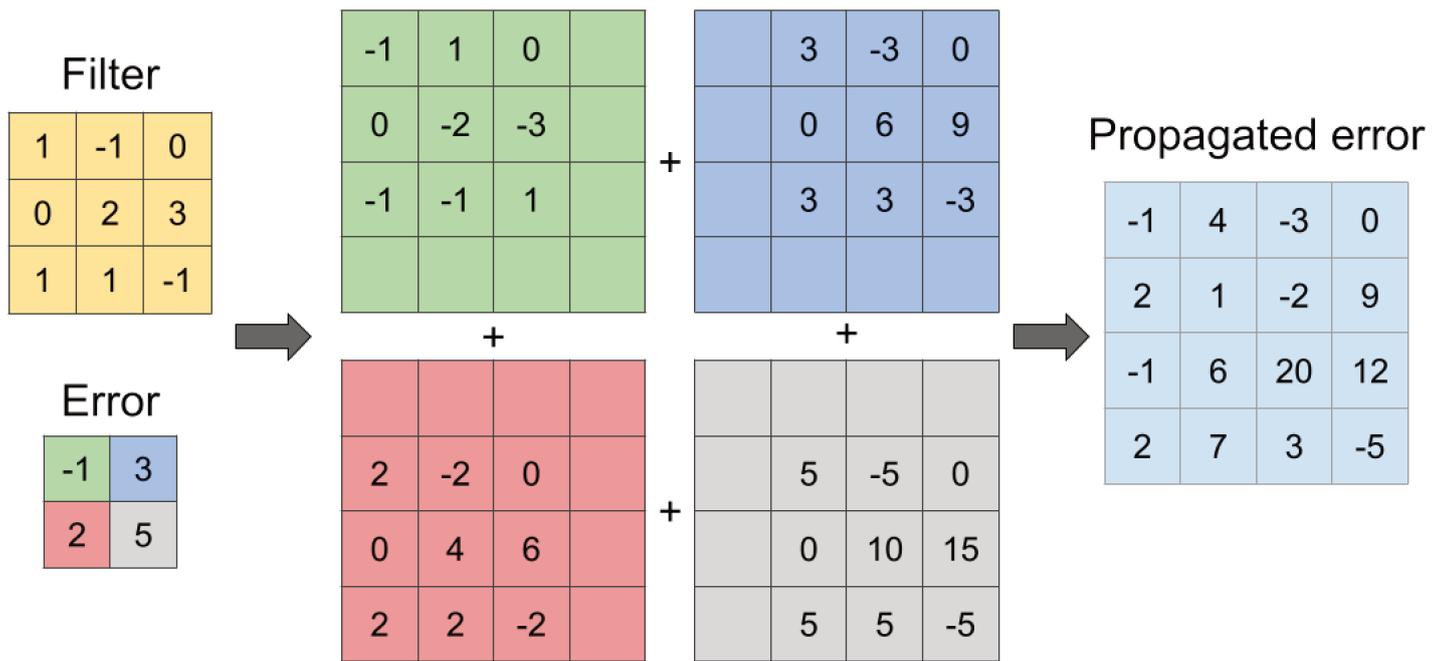
The weights of convolutional layers in a CNN need to be adjusted like the weights of a fully connected NN (LeCun et al. 1998). To understand how it is done, we must look at how convolutional and pooling layers propagate errors.

A pooling layer receives an error with the dimension of its output from the next layer. This error is propagated through the pooling layer in a manner that assigns it to the input neurons that correspond to the maximum values within specific sub-regions. To accomplish this, the position of the neuron with the maximum value needs to be stored in memory. Subsequently, the error signal is directed to this specific position, while all other neuron errors within the sub-region are set to zero. As a result, an error signal is generated that has the same dimensions as the input to the pooling layer. This error signal is then further propagated back to the previous layer of the network (Fig.3.12).

To propagate an error through a convolutional layer, we need to define an operation called transposed convolution (Dumoulin & Visin 2018). The goal of this operation is to upsample the input to a larger spatial size. A transposed convolution consists of sliding the input over the filter and performing element-wise multiplication and summation. An example of a transposed convolution of a  $2 \times 2$  input with a  $3 \times 3$  filter is shown in Fig. 3.13. Most of the time, the projected field of neighboring pixels will overlap. In this scenario, the contribution of each input, weighted by the filter, is summed across overlapping output pixels. During the propagation of an error through a convolutional layer, a transposed convolution is applied. This operation enables the error to be propagated with the same dimensions as the output of the convolutional



**Figure 3.12:** Example of a max pooling layer  $2 \times$  backpropagation on a single layer image. For consistency, the maximum values are from the matrix from Fig. 3.11. The colors indicated the sub-regions. The "Max-Loc" matrix indicates the positions where the maximum values were located before the Max-Pooling operation took place.



**Figure 3.13:** Example of the error propagated through a convolution layer. We represent a transposed convolution of a 2x2 error with a 3x3 filter (in yellow). The details of how each pixel of the error contributes to the results are color-coded. The result is a 4x4 error (in light blue).

layer and subsequently resized to match the dimensions of the input of the convolutional layer.

### 3.5.9 Batch normalization

Batch normalization is a standard technique used in CNN and other types of networks to improve their stability (Ioffe & Szegedy 2015b). By normalizing the inputs to each layer, batch normalization helps to ensure that the network is less sensitive to changes in the distribution of the input data, which can speed up training and improve the accuracy of the model. It also helps to reduce the vanishing gradient problem by rescaling the outputs, which enables the use of deeper network architectures. In a NN, batch normalization layers are layers that normalized their input as follows:

$$\mu = \frac{1}{n} \sum y_i \quad (3.22)$$

$$\sigma^2 = \frac{1}{n} \sum (y_i - \mu)^2 \quad (3.23)$$

$$y_{i,norm} = \frac{y_i - \mu}{\sqrt{\sigma^2 - \epsilon}} \quad (3.24)$$

$$y_{i,new} = \gamma y_{i,norm} + \beta \quad (3.25)$$

Initially, the batch normalization layer calculates the mean ( $\mu$ ) and variance ( $\sigma^2$ ) of the activation values of the previous layer across the batch using Eqs. 3.22) and 3.23 respectively. Subsequently, it applies normalization to its input using Eq. 3.24, ensuring a standard normal deviation across the batch. In this equation,  $\epsilon$  represents a constant utilized for maintaining numerical stability. Finally, the batch normalization layer produces the output ( $y_{i,new}$ ) through Eq. 3.25, where the normalized activation vector is linearly transformed with two trainable parameters,  $\gamma$ , and  $\beta$ . The parameter  $\gamma$  allows for adjustment of the standard deviation, while  $\beta$  allows for adjustment of the bias, shifting the distribution spike horizontally.

### 3.5.10 The use of CNN in astronomy

Over time, astronomical surveys have yielded increasingly voluminous dataset, with future surveys expected to generate even larger volumes of data. The manual analysis of such extensive dataset has become increasingly laborious, prompting astrophysicists to explore innovative data analysis techniques to manage this ever-growing influx of information. ML has emerged as a pivotal solution in this context, enabling significantly faster analysis by automating data analysis processes.

CNN excel at automatically extracting hierarchical and complex features from data. In astrophysics, where dataset often contain intricate spatial or spectral information, CNN can effectively discern patterns, shapes, or structures in images, spectra, or time-series data. Furthermore, CNN are inherently robust to noise and variations in data, making them suitable for handling noisy astronomical observations or images affected by instrumental limitations. This robustness in CNN arises from several architectural aspects. Firstly, CNN utilize local receptive fields, where individual neurons in a layer are connected to small, localized regions within the

input data (see Sect. 3.5.1). This architectural design enables the network to emphasize local patterns and features, thereby diminishing the influence of noise across the entire dataset. Additionally, CNN employ weight sharing, employing a consistent set of weights across different segments of the input data (see Sect. 3.5.2). This shared parameter mechanism empowers the network to recognize patterns irrespective of their specific locations within the input, enhancing the model's resilience to minor variations and noise.

Moreover, CNN architectures frequently integrate pooling layers, such as max pooling, to downsample the data (see Sect. 3.5.6). These layers reduce sensitivity to minor spatial variations or small-scale noise, preserving significant features while discarding less pertinent details. The hierarchical structure of CNN, composed of multiple convolutional layers, aids in learning hierarchical representations of features (see Sect. 3.5.4). Each layer extracts progressively intricate and abstract features from the input, enabling the network to distill essential patterns and information from noisy data.

Finally, the use of non-linear activation functions, like the ReLU (see Sect. 3.5.5), introduces non-linearity into the CNN, enabling it to model complex relationships and diminish the influence of noise on the learned representations. This collective architectural framework equips CNN with robustness, making them highly effective in handling noise and variations inherent in astrophysical dataset.

Thus, CNN have been used in diverse fields of astrophysics. One such field is in the detection of galaxies (Bonaldi et al. 2021; Hartley et al. 2023). Furthermore, CNN have been very successful in the classification of galaxy morphology (Walmsley et al. 2020; Cheng et al. 2020, 2021). These networks have also been used to identify post-merger galaxies (Bickley et al. 2021). Additionally, CNN have been employed in estimating galaxies properties, like their photometric redshifts (Pasquet et al. 2019; Schuldt et al. 2021; Zhou et al. 2022).

Within expansive spectroscopic surveys of stars, CNN holds significant promise in enabling the rapid extraction of physical parameters across millions of spectra. CNN allow for connecting observable such as spectra or stellar magnitudes, and fundamental properties including atmospheric parameters and chemical abundances (Guiglion et al. 2020). Additionally, these networks demonstrate efficiency in classifying stellar spectra (Hon et al. 2017; Liu et al. 2019; Sharma et al. 2020; Zheng et al. 2020).

The anticipated future surveys by the SKA are projected to produce extensive datasets comprising 21 cm maps covering cosmological scales during the epoch of reionization. To grapple with the enormity of this data influx, there's ongoing experimentation on the potential of CNN process 21cm maps to estimate astrophysical and cosmological parameters using 21 cm maps generated from semi-numerical simulations (Gillet et al. 2019; Hassan et al. 2020; Prelogović et al. 2022; Meriot & Semelin 2023).

Detailed understanding of the response of the LIGO and Virgo detectors to true signals in the presence of environmental and instrumental noise. Particular interest is the study of anomalous short-lived non-Gaussian instrumental or terrestrial transients. Their occurrence rate in LIGO and Virgo data can obscure or even mimic true gravitational wave signals. Therefore, successfully identifying and excising these transient anomalies from gravitational wave data is important for the detection and characterization of true signals and for the accurate computation of their significance (George et al. 2018; Fernandes et al. 2023; Bini et al. 2023).



# CHAPTER 4

---

## ESTIMATING SMBH MASS WITH ML FOR THE GATOS CORE SAMPLE

---

In this chapter, I will outline our research focused on the GATOS data. Our objective was to estimate the mass of the SMBH ( $M_{\text{BH}}$ ) within the galaxies comprising the GATOS core sample. To begin, I will present the motivation behind this work, providing an overview of the GATOS collaboration and its core sample. Subsequently, I will delve into the theoretical aspects of the numerical simulation employed in this study, discussing its relevance and implications. Following that, I will provide a detailed account of how ML techniques were utilized to estimate the  $M_{\text{BH}}$  of seven out of ten galaxies within the GATOS core sample. In the subsequent sections, I will describe our findings, presenting the results obtained and comparing them with other  $M_{\text{BH}}$  estimation methods documented in the existing literature.

### Contents

---

<b>4.1</b>	<b>Motivation</b>	<b>84</b>
<b>4.2</b>	<b>The GATOS core sample</b>	<b>88</b>
<b>4.3</b>	<b>Numerical simulation</b>	<b>89</b>
4.3.1	Potentials and mass distributions	90
4.3.2	Analytical velocity dispersion	91
4.3.3	Analytical velocity	94
4.3.4	Positions and velocities initialization	96
4.3.5	Matching the ALMA observations	96
<b>4.4</b>	<b>ML approach to estimate the <math>M_{\text{BH}}</math></b>	<b>98</b>
4.4.1	Supervised learning approach	98

4.4.2	Preparing the inputs . . . . .	99
4.4.3	ML architecture . . . . .	101
4.4.4	Estimating the reliability of models . . . . .	102
<b>4.5</b>	<b>Results on <math>M_{\text{BH}}</math> estimation . . . . .</b>	<b>103</b>
4.5.1	Reliability of ML models . . . . .	103
4.5.2	$M_{\text{BH}}$ estimations results . . . . .	108
<b>4.6</b>	<b>Individual galaxies . . . . .</b>	<b>110</b>
4.6.1	NGC 4388 . . . . .	110
4.6.2	NGC 5506 . . . . .	111
4.6.3	NGC 5643 . . . . .	112
4.6.4	NGC 6300 . . . . .	113
4.6.5	NGC 7314 . . . . .	113
4.6.6	NGC 7465 . . . . .	114
4.6.7	NGC 7582 . . . . .	114
<b>4.7</b>	<b>Discussion . . . . .</b>	<b>115</b>
<b>4.8</b>	<b>Conclusion . . . . .</b>	<b>117</b>

---

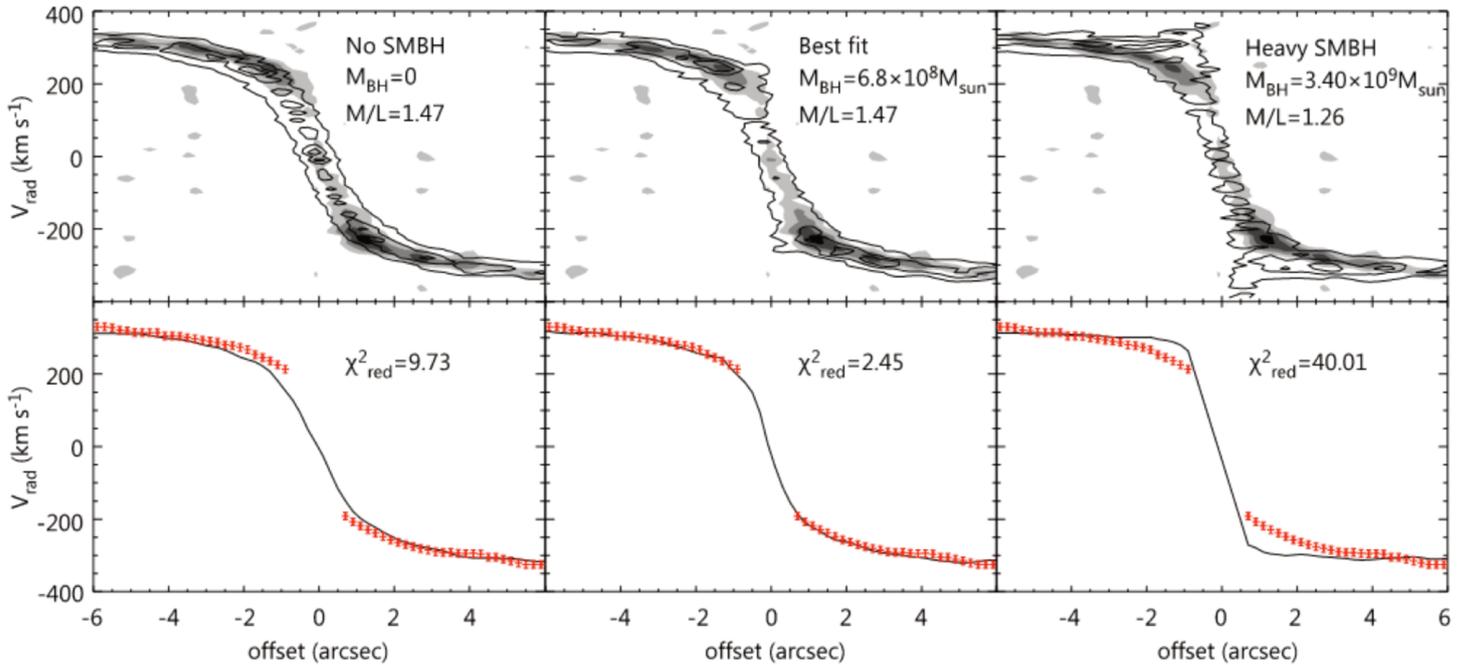
## 4.1 Motivation

For nearby galaxies, dynamics of matter nearby the central SMBH can be used to measure the  $M_{\text{BH}}$  from their kinematics when it can be resolved (see Sect. 1.6.1). Historically, stellar dynamics have been widely used to estimate the  $M_{\text{BH}}$  (Bender et al. 2005; Kormendy & Ho 2013; Saglia et al. 2016; KrAJnović et al. 2018). More recently, the advent of circumnuclear disk observations has opened up the opportunity to employ a similar approach for measuring the mass of SMBH, this time using gas dynamics. The first observation, credited to Harms et al. (1994), was directed at the 100-pc-scale circumnuclear disk within M87, which had been detected through  $\text{H}\alpha$  emission by Ford et al. (1994)). While the single-aperture of the Hubble Space Telescope faint object spectrograph offered constrained spatial sampling, these observations provided compelling evidence that the majority of the observed mass was dark and likely associated with the central SMBH. Gas-dynamical mass measurements offer distinct advantages over stellar-dynamical modeling. Ionized gas emission lines are more easily detectable than stellar absorption lines or even molecular lines, and their simpler line profiles simplify the measurement of velocities and velocity dispersions. In contrast, the stellar-based approach often involves the challenging task of measuring higher-order moments of the line-of-sight velocity distributions, particularly in more massive, lower surface brightness galaxies.

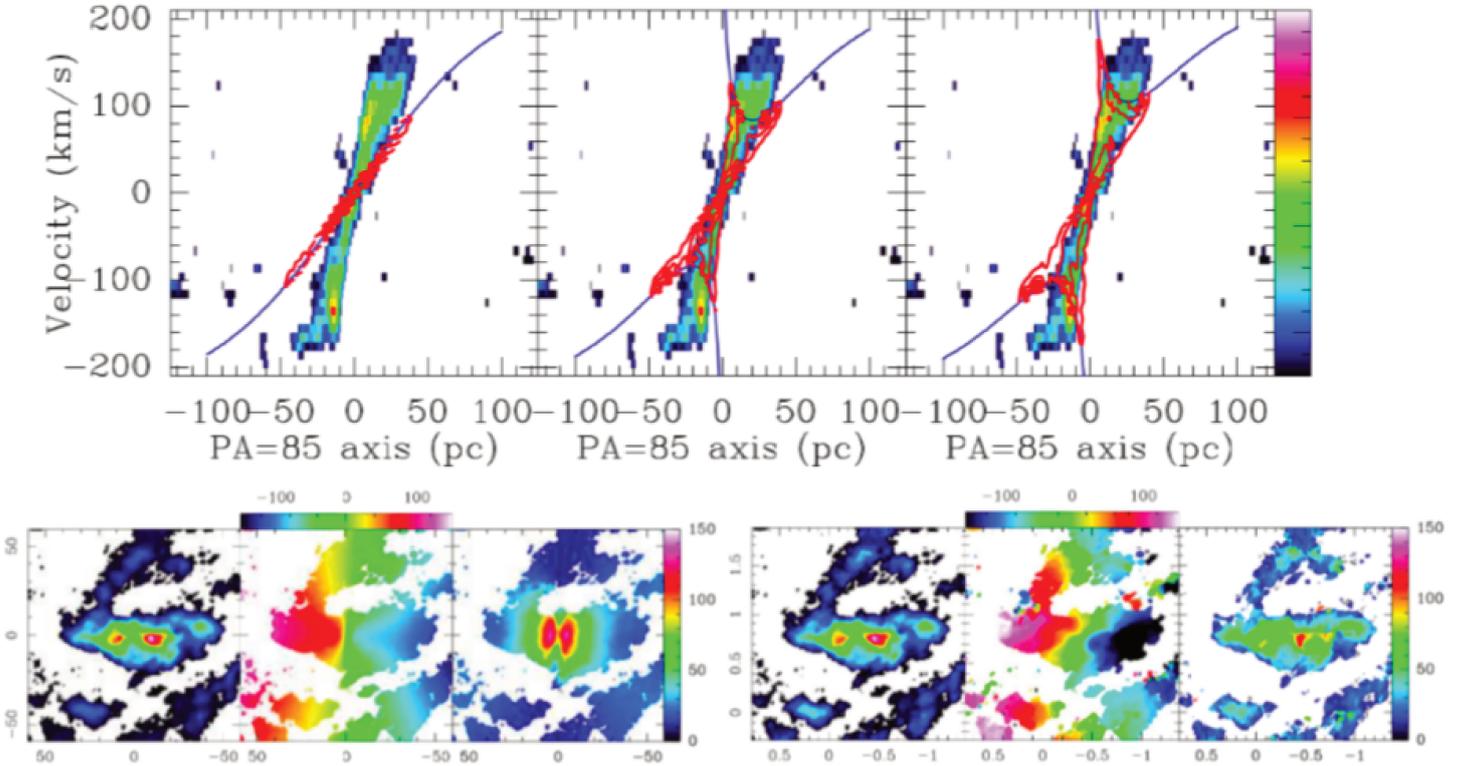
When estimating the SMBH from gas dynamics, the goal is to compute a model velocity field that simultaneously matches the observed velocities and velocity dispersions. Using gas

instead of stars to estimate the SMBH mass also offers a more conceptually straightforward dynamical modeling approach (Macchetto et al. 1997; van der Marel & van den Bosch 1998; Barth et al. 2001; Maciejewski & Binney 2001; Marconi et al. 2001; Marconi & Hunt 2003). In cases where the gas follows Keplerian rotation in a dynamically cold disk, the analysis becomes considerably less complicated and computationally more efficient than the orbit-based techniques required for stellar dynamics. Also, complexities such as orbital anisotropy, triaxiality, or the influence of dark matter halos can be effectively disregarded.

Nowadays, high-resolution observations of parsec-scale molecular gas disks, made possible by ALMA (Combes et al. 2019; García-Burillo et al. 2019; García-Bernete et al. 2021; García-Burillo et al. 2021), present a valuable opportunity for studying the dynamics of these disks and inferring the presence of SMBH located at their centers. Prior research has employed ALMA data cubes to estimate the mass of SMBH. Notably, the work by Combes et al. (2019) and the mm-Wave Interferometric Survey of Dark Object Masses (WISDOM) project, as outlined in Onishi et al. (2017); Davis et al. (2017, 2018), serve as illustrative examples. These investigations hinge on the fitting of dynamical models to observed molecular gas distributions.



**Figure 4.1:** Figures derived from the study conducted by Onishi et al. (2017) depict their results regarding the estimation of  $M_{\text{BH}}$  by fitting dynamical models to ALMA observations of the CO(2-1) molecular gas in NGC 3665. Upper Panels: Depicted are position-velocity diagrams (PVD) observed along the kinematic major axis (displayed in grey-scale), superimposed with contours representing the model PVD. The left panel illustrates a comparison with a model devoid of a BH ( $M_{\text{BH}}=0$ ), the middle panel corresponds to the best-fit model, and the right panel portrays a model featuring an excessively massive BH. Lower Panels: Presented here are the extracted mean velocities along the kinematic major axis, denoted by red points with accompanying error bars. These observed points are juxtaposed with the best-fitting model velocities depicted as black lines. The reduced  $\chi^2$  value for each model is 9.73 (left,  $M_{\text{BH}}=0$ ), 2.45 (middle, best fit), and 40.01 (right, overweight SMBH).



**Figure 4.2:** Figures extracted from the work by [Combes et al. \(2019\)](#) illustrate their findings concerning the estimation of the  $M_{\text{BH}}$  through the fitting of dynamical models to ALMA observations of the CO(3-2) molecular gas in NGC 1672. The uppermost figure presents the PVD for the CO(3-2) line in NGC 1672, utilizing a linear color scale (TM1 only). Overlaid in red are the contours of three different models: one without a BH and with a torus inclination of  $66^\circ$  (left panel), another incorporating a BH derived from the  $M_{\text{BH}} - \sigma$  relation, measuring  $2.5 \times 10^7 M_\odot$  and inclined at  $66^\circ$  (middle panel), and the best-fitting model featuring a BH with a mass of  $5.0 \times 10^7 M_\odot$  and an inclination of  $66^\circ$  (right panel).

The lower panels comprise two sub-figures: The left panel features the first three moments of the model cube for NGC 1672, with offsets in right ascension and declination measured in parsecs. Velocity color scales are expressed in kilometers per second. This cube corresponds to the best PV fit, with a  $M_{\text{BH}}$  of  $5.0 \times 10^7 M_\odot$  and an inclination of  $66^\circ$ . The right panel displays the first three moments for the observations (TM1 only), with right ascension and declination offsets in arcseconds relative to the phase center.

The primary objective of the WISDOM studies was to determine SMBH masses within nearby early-type galaxies exhibiting rapid rotation by applying dynamical models to the observational data. In the case of NGC 3665, as described by [Onishi et al. \(2017\)](#), the approach involved constructing a three-dimensional model of the stellar mass distribution within the galaxy. This model was generated through the de-projection of a two-dimensional representation of the observed surface brightness, assuming a constant M/L. The resultant circular velocity curve, originating from this mass model and the presence of an SMBH, was incorporated into a simulation code that considered the molecular gas distribution and instrumental effects. The SMBH mass was subsequently derived by comparing multiple models to the observed data. The selection of the most appropriate model parameters was determined via Bayesian analysis, wherein the parameter values minimizing the multidimensional  $\chi^2$  space were identified (see Fig.4.1).

In parallel investigations, [Davis et al. \(2017\)](#) examined NGC4697, while [Davis et al. \(2018\)](#) focused on NGC 4429. They leveraged the publicly accessible KINematic Molecular Simulation tool to ascertain optimal model parameters and their associated uncertainties. The Markov chain Monte Carlo technique was employed to fit the data and generate samples from the Bayesian posterior probability distributions of the fitted parameters. All three of these studies reached conclusions aligned with previous SMBH estimations derived from the  $M_{\text{BH}} - \sigma$  relation.

In the study conducted by [Combes et al. \(2019\)](#), the investigation focused on estimating the masses of SMBH in seven nearby Seyfert galaxies. Their methodology involved utilizing a dynamical model, which is elucidated in Section 4.3. The estimation of SMBH masses a least-square fit followed by a visual comparison between the position-velocity diagrams Position-Velocity Diagram (PVD) and the first three moment maps derived from the ALMA data cube and those computed from the simulated data cube (see Fig.4.2).

These approaches leverage the exceptional resolution of observations to obtain SMBH mass estimates that surpass those achieved through scaling relations, which inherently contain associated errors. While fitting dynamical models to individual galaxies is effective, this process is time-consuming and scales linearly with the number of galaxies under investigation. As the volume of observations and data is anticipated to increase in the future, this scalability presents a potential concern. Therefore, it becomes crucial to explore alternative methods for SMBH mass estimation that offer improved time efficiency.

In addressing this challenge, the use of machine learning emerges as a promising solution. Machine learning algorithms have proven their capability to discern intricate patterns within extensive datasets, making them well-suited for handling the increasing volume of astronomical observations. Their adaptability and efficiency in processing vast data sets can significantly reduce the time required for making multiple SMBH mass estimations, circumventing the linear scalability limitations that have posed concerns in traditional dynamical modeling. By utilizing ML capacity to generalize from past data, SMBH mass estimation can be automated and expedited, providing a more time-efficient alternative. Using ML also mitigates the inherent errors associated with scaling relations.

During the past few years, the domain of ML has widely expanded in the scientific community. One of the main domains of the application of ML is image analysis. Efficient and powerful NN architectures, known as CNN, have been developed to effectively handle images. They are used to classify large numbers of galaxies ([Huertas-Company et al. 2020](#)), in view of

future missions like Euclid. They are also used to classify supernovae (Lochner et al. 2016), find strong lensing features (Lanusse et al. 2018), transient phenomena (Mahabal et al. 2019), photometric redshifts (Pasquet et al. 2019) or star clusters (Castro-Ginard et al. 2019). They have not yet been used to estimate the  $M_{\text{BH}}$ .

In this chapter, I present a new ML method to estimate the mass of SMBH in galaxies. The masses obtained with this new method are compared with SMBH masses found in the literature and with other estimations made using the FPBH and the  $M_{\text{BH}} - \sigma$  relation. In Sect. 4.2, the galaxy sample and the selection criteria are described. Sect. 4.3 focuses on the physics behind the numerical simulation used later on. Section 4.4 presents how I combine ALMA data cubes, numerical simulations, and a dedicated ML approach to estimate more precisely  $M_{\text{BH}}$  from the CO(3-2) gas dynamics. A large number of simulated models are used to evaluate the quality of my estimations. In Sect. 4.5, the results found with my ML method are presented. Finally, in Sect. 4.7 and 4.8, I summarize and discuss my findings and possible future improvements to my methodology.

## 4.2 The GATOS core sample

The GATOS international collaboration has for aims to study topics related to the physics taking place in the nuclear region of AGN. These encompass the gas flow cycle, the emission of polar dust, the properties of the torus/obscuring material, and the interplay between star formation activity and AGN phenomena. Here, I study the 10 galaxies from the GATOS core sample studied by García-Burillo et al. (2021). The core sample is composed of nearby galaxies ( $<28\text{Mpc}$ ) and luminous nuclei,  $L_{\text{AGN}}(14 - 150\text{keV}) \geq 10^{42}\text{erg/s}$ . The upper limit on the distance is set to have a sufficient spatial resolution to study molecular tori as small as  $\sim 10\text{pc}$  in radius while the limit on the luminosity is set to avoid overlap with the ongoing ALMA surveys of nearby Seyferts such as the Nuclei of Galaxies survey (NUGA) (e.g. Combes et al. 2019; Audibert et al. 2019).

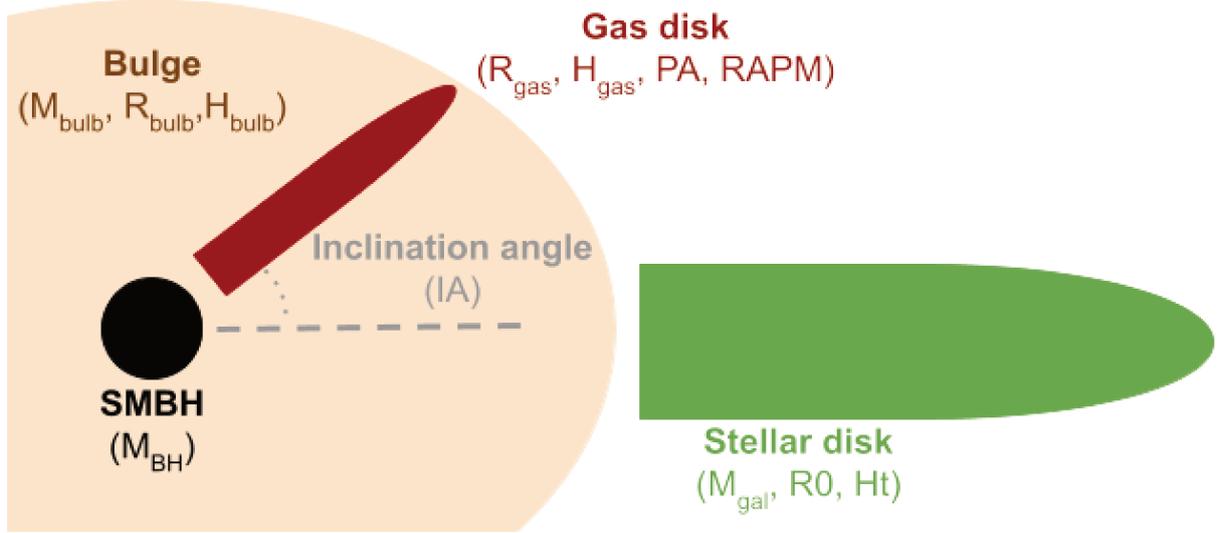
The present work uses the CO(3-2) data cubes from ALMA observations. All the details of my sample and their observations are given in García-Burillo et al. (2021). I do not consider the three additional GATOS galaxies NGC 1068 (e.g. García-Burillo et al. 2019), NGC 1365 (e.g. Combes et al. 2019) and NGC 3227 (e.g. Alonso-Herrero et al. 2019).

Not all 10 data cubes are appropriate for the determination of the central kinematics and  $M_{\text{BH}}$  estimation. Two of the galaxies, NGC 6814 and NGC 7213 do not present enough CO emission in the very center; their central pixels are empty or with some emission with a S/N ratio below 2. For another galaxy, moment maps are not usable for a  $M_{\text{BH}}$  estimation with my ML approach; NGC 4941 has too little gas detected and its distribution is too sparse. There remain therefore 7 galaxies: NGC 4388, NGC 5506, NGC 5643, NGC 6300, NGC 7314, NGC 7465 and NGC 7582.

## 4.3 Numerical simulation

In the following section, I will explore the integration of ML techniques with numerical simulations. As stated in the Sect.4.2, the goal of this chapter is to use ML to estimate the  $M_{\text{BH}}$  of the GATOS galaxies. To do so, I will need numerical simulations presented to create training, validation, and test datasets to train my NN. Further description of the methodology is described in the next section (see Sect.4.4). However, for a coherent reading, I will begin by outlining the theoretical foundation that underlies the numerical simulation in the following section. The goal is to provide an understanding of how the positions, the velocities, and their dispersions are initialized and how I make the simulations match the observations. The equations presented in this section are taken from [Binney & Tremaine \(2008\)](#).

To simulate my data cube I used the model described in [Melchior & Combes \(2011\)](#); [Combes et al. \(2019\)](#) (see Fig.4.3). I employed a gas model, utilizing static simulations of dynamical components represented by particles, with various geometric orientations on the celestial sphere. Our models are axisymmetric homogeneous gas disks (e.g. [Miyamoto & Nagai 1975](#)) with gas particles on nearly circular orbits, with low-velocity dispersion, corresponding to thin cold disks. The sizes of the modeled disks are taken from the observed CO(3-2) disks. To ensure sufficient statistics, I fixed the number of particles at  $10^6$ . Since I simulate only the nuclear disk with a radius of typically  $\sim 100\text{pc}$ , the dark matter contribution to the gravitational potential is negligible. The different parameters of the simulations are: the mass, radius, and height of the galaxy bulge ( $M_{\text{bulb}}, R_{\text{bulb}}, H_{\text{bulb}}$ ); the corresponding properties of the stellar disks ( $M_{\text{gal}}, R_0, H_t$ ); the gas-to-stellar mass ratio within the stellar disk ( $R_{\text{APM}}$ ); the gas disk radius, height, inclination angle, and position angle ( $R_{\text{gas}}, H_{\text{gas}}, IA, PA$ ) and the SMBH mass ( $M_{\text{BH}}$ ).



**Figure 4.3:** Schematic view of the different components and their associated variables of the numerical simulation used in Melchior & Combes (2011); Combes et al. (2019) and described in Sect.4.3.

### 4.3.1 Potentials and mass distributions

To compute the gravitational potential created by a density  $\rho$  of stars and gas, the Poisson equation can be used

$$\nabla^2 \Phi = 4\pi G \rho \quad (4.1)$$

Miyamoto & Nagai (1975) have found a potential-density pair solution to Poisson's equation for a 3D disk:

$$\Phi(R, z) = -\frac{GM}{\sqrt{R^2 + (a + \sqrt{z^2 + b^2})^2}} \quad (4.2)$$

$$\rho(R, z) = \frac{b^2 M}{4\pi} \frac{aR^2 + [a + 3(z^2 + b^2)^{1/2}][a + (z^2 + b^2)^{1/2}]^2}{(R^2 + [a + (z^2 + b^2)^{1/2}]^2)^{5/2} (z^2 + b^2)^{3/2}} \quad (4.3)$$

where  $R$  and  $z$  are the radius and height in the cylindrical coordinate system,  $M$  is the disk mass,  $a$  is the disk scale length and,  $b$  is the disk scale height.

The Miyamoto-Nagai potential is a generalization of the Plummer and Kuzmin potentials. The Plummer spherical solutions (Plummer 1911) can be found from eq.4.2 and 4.3 by posing  $a = 0$

$$\Phi_p(r) = -\frac{GM_p}{\sqrt{r^2 + r_b^2}} \quad (4.4)$$

$$\rho_p(r) = \frac{3b^2 M_p}{4\pi} \frac{1}{(r^2 + r_b^2)^{5/2}} \quad (4.5)$$

where  $r = \sqrt{R^2 + z^2}$  and  $M_p$  is the mass of the Plummer sphere.

With a similar process, one can find the Kuzmin-Plummer disc potential from eq.4.2 and 4.3 by letting  $b = 0$  :

$$\Phi_K(R, z) = -\frac{GM_K}{\sqrt{R^2 + (a + |z|)^2}} \quad (4.6)$$

$$\Sigma_K(R) = \frac{aM_K}{2\pi(R^2 + a^2)^{3/2}} \quad (4.7)$$

where  $M_K$  is the mass of the Kuzmin disk.

The particles in my simulations are influenced by three different gravity potentials: one from the stellar bulge, one from the stellar disc, and one from the SMBH. I will use the Miyamoto-Nagai potential to the bulge and the stellar disk.

## 4.3.2 Analytical velocity dispersion

### 4.3.2.1 Equations of movements

In my simulation, the particles move in three-dimensional space, allowing their motions to extend beyond the equatorial plane. However, the study of axially symmetric disks can be simplified by considering the conservation of z-angular momentum symmetry. This symmetry is reflected in the potential  $\Phi$ , which exhibits symmetry about the plane  $z = 0$ . The motion of these particles is described by the Lagrangian ( $L$ ) and Hamiltonian ( $H$ ) equations, which are formulated in cylindrical coordinate systems such as

$$L = \frac{1}{2} (\dot{R}^2 + (R\dot{\phi})^2 + \dot{z}^2) - \Phi(R, z) \quad (4.8)$$

$$H = \frac{1}{2} \left( p_R^2 + \frac{p_\phi^2}{R^2} + p_z^2 \right) + \Phi(R, z) \quad (4.9)$$

where  $p_R = \dot{R}$ ,  $p_\phi = R^2\dot{\phi}$  and  $p_z = \dot{z}$  are the momenta. Then, the equations of motion as can be written as

$$\dot{p}_R = \frac{p_\phi^2}{R^3} - \frac{\partial\Phi}{\partial R} \quad (4.10)$$

$$\dot{p}_\phi = \frac{d}{dt}(R^2\dot{\phi}) = 0 \quad (4.11)$$

$$\dot{p}_z = -\frac{\partial\Phi}{\partial z} \quad (4.12)$$

here  $p_\phi = L_z$  is a constant and expresses the conservation of angular momentum of the z-axis. From here onward,  $p_\phi$  will be replaced by its numerical value  $L_z$ . Equations 4.10 and 4.12 describe the coupled oscillations of the particles in the  $R$  and  $z$  directions. Then, let us define an effective potential  $\Phi_{\text{eff}}$  such as

$$\Phi_{\text{eff}} = \Phi(R, z) + \frac{L_z^2}{R^3} \quad (4.13)$$

$$\ddot{R} = -\frac{\partial\Phi_{\text{eff}}}{\partial R} \quad (4.14)$$

$$\ddot{z} = -\frac{\partial\Phi_{\text{eff}}}{\partial z} \quad (4.15)$$

Thus, the particles motions are now reduced to a 2D motion in the plan  $(R, z)$  under the effective Hamiltonian

$$H_{\text{eff}} = \frac{1}{2}(p_R^2 + p_z^2) + \Phi_{\text{eff}}(R, z) \quad (4.16)$$

$\Phi_{\text{eff}}$  minimum has physical significance and occur when  $\frac{\partial\Phi_{\text{eff}}}{\partial R} = \frac{L_z^2}{R^3}$  and  $0 = \frac{\partial\Phi_{\text{eff}}}{\partial z}$ . These conditions are satisfied anywhere in the equatorial plane and the first is satisfied at the guiding radius  $R_g$ :

$$\left(\frac{\partial\Phi}{\partial R}\right)_{(R_g, 0)} = \frac{L_z^2}{R_g^3} = R_g \dot{\phi}^2 \quad (4.17)$$

This is the condition for a circular orbit with an angular speed of  $\dot{\phi}$  and thus, the minimum appears at the radius at which the angular momentum is equal to  $L_z$ , and the value of  $\Phi_{\text{eff}}$  at the minimum of this circular energy.

### 4.3.2.2 Nearly circular environment

In disk, many particles have a nearly circular orbit, so it is useful to approximate solutions to Eqs. 4.14 and 4.15 that are valid for such orbits. Thus, a  $x$  coordinate can be introduced as

$$x \equiv R - R_g \quad (4.18)$$

where  $R_g(L_z)$  is the guiding center radius for an orbit of angular momentum  $L_z$  (see Eq. 4.11). Thus,  $(x, z) = (0, 0)$  are the meridional plane coordinate of the minimum in  $\Phi_{\text{eff}}$ . When the  $\Phi_{\text{eff}}$  is expanded in a Taylor series about this point it leads to

$$\Phi_{\text{eff}} = \Phi_{\text{eff}}(R_g, 0) + \frac{1}{2} \left(\frac{\partial^2\Phi_{\text{eff}}}{\partial R^2}\right)_{(R_g, 0)} x^2 + \frac{1}{2} \left(\frac{\partial^2\Phi_{\text{eff}}}{\partial z^2}\right)_{(R_g, 0)} z^2 + O(xz^2) \quad (4.19)$$

To simplify this equation, the epicycle approximation can be used, in which all terms  $\Phi_{\text{eff}}$  of order  $xz^2$  or higher powers of  $x$  and  $z$  are neglected (e.g. [Toomre 1964](#)). Now, two new quantities  $\kappa$  and  $\nu$  can be defined as

$$\kappa^2(R_g) \equiv \left(\frac{\partial^2\Phi_{\text{eff}}}{\partial R^2}\right)_{(R_g, 0)} \quad (4.20)$$

$$\nu^2(R_g) \equiv \left(\frac{\partial^2\Phi_{\text{eff}}}{\partial z^2}\right)_{(R_g, 0)} \quad (4.21)$$

By using these new quantities, Eqs. 4.14 and 4.15 become

$$\ddot{x} = -\kappa^2 x \quad (4.22)$$

$$\ddot{z} = -\nu^2 z \quad (4.23)$$

According to these equations,  $x$  and  $z$  evolve like the displacements of two harmonic oscillators with frequencies  $\kappa$ , called the epicycle frequency, and  $\nu$ , called the vertical frequency. Upon substituting the expression from Eq. 4.15 for  $\Phi_{\text{eff}}$  and further developing, the result is obtained as follows

$$\kappa^2 = \frac{2\Omega}{R} \frac{d}{dR} (R^2\Omega) \quad (4.24)$$

$$\nu^2 = \frac{\partial^2 \Phi}{\partial z^2} \quad (4.25)$$

where  $\Omega$  is the circular frequency and is given by  $L_z^2/R^4$ . The ratio  $\nu^2/\kappa^2$  is a measure of the degree to which the particles are concentrated towards the plane. For example, observations have shown that this ratio is  $\sim 4$  near the Sun.

### 4.3.2.3 Toomre's parameter and velocity dispersion

The Toomre's stability criterion is a relation in a differentially rotating gaseous disk that approximately determined if the system is stable. The Toomre's criterion, also called the Toomre parameter,  $Q$  is written for a stellar disk as

$$Q = \frac{\sigma_r \kappa}{3.36G\Sigma} \quad (4.26)$$

where  $\sigma_r$  is the radial velocity dispersion,  $G$  is Newton's gravitational constant,  $\Sigma$  is the surface density of the disk. In order to ensure the stability of an accretion disk, it is necessary for the value of  $Q$  to be greater than or equal to 1. To achieve this condition,  $\sigma_r$  can be defined in such a way that it sets the Toomre  $Q$  parameter of the disk to 1. Therefore, the value of  $\sigma_r$  can directly determined by using Eq. 4.26

$$\sigma_r = 3.36G \frac{\Sigma_s + \Sigma_g}{\kappa} \quad (4.27)$$

where  $\Sigma_s$  is the surface density of the stars in the disk and  $\Sigma_g$  represents the same quantity but for the gas. The epicyclic theory gives the ratio between tangential and radial velocity dispersion (e.g. Toomre 1964), thus,  $\sigma_\theta$  can be fined from  $\sigma_\phi$

$$\sigma_\theta = \kappa \frac{\sigma_r}{2\Omega} \quad (4.28)$$

### 4.3.2.4 Vertical velocity dispersion

For studying the vertical structure of the thin axisymmetric disk, an assumption can be made that the  $z$ -component of the potential depends solely on  $z$ . The equilibrium state is given by solving

$$\nabla P = -\rho \nabla \Phi \quad (4.29)$$

The pressure  $P = \rho \sigma_z^2$  and the potential is given by Poisson's equation (Eq. 4.1). This allows for the writing of

$$\frac{d}{dz} \left( \frac{1}{\rho} \frac{d\rho}{dz} \right) = -\frac{4\pi G}{\sigma_z^2} \rho \quad (4.30)$$

The solution is of the form of

$$P = \frac{\rho_0}{\text{ch}^2(z/H)} \quad (4.31)$$

with

$$H = \sqrt{\frac{\sigma_z^2}{2\pi G \rho_0}} \quad (4.32)$$

From that point, it can be demonstrated that

$$\sigma_z = \sqrt{2\pi G H \Sigma} \quad (4.33)$$

### 4.3.3 Analytical velocity

In modeling collisionless systems, the verifiability of many predictions relies on the likelihood of locating a particle within a six-dimensional phase-space volume defined by its position  $X(x, y, z)$  and velocity  $V(v_x, v_y, v_z)$ . As a result, a probability distribution function denoted as  $f$  can be established, representing the likelihood, at a given time  $t$  of a random particle having phase space coordinates within the specified range. At any position, the integral

$$\nu(X) = \int d^3V f(X, V) \quad (4.34)$$

gives the probability per unit of volume of finding a particle at  $X$ . Therefore, the probability distribution of stellar velocities at  $X$  can be defined by

$$P_X(V) = \frac{f(X, V)}{\nu(X)} \quad (4.35)$$

Hence, the mean velocity at  $X$  can be defined as

$$\bar{V}(X) = \int V P_X(V) d^3V = \frac{1}{\nu(X)} \int V f(X, V) d^3V \quad (4.36)$$

A property of  $f$  is that it has the same value at a given phase-space point in any canonical coordinate system. Consequently, it is possible to define  $W = (X, V) = (Q, P)$ . As  $f$  evolves through time, the probability must be conserved. The equation for the conservation of probability in phase space is written as

$$\frac{\partial f}{\partial t} + \frac{\partial}{\partial W} \cdot (f\dot{W}) = 0 \quad (4.37)$$

From that point, the collisionless Boltzmann equation can be derived as follows:

$$\frac{\partial f}{\partial t} + \dot{Q} \cdot \frac{\partial f}{\partial Q} + \dot{P} \cdot \frac{\partial f}{\partial P} = 0 \quad (4.38)$$

which can be written in cylindrical systems as

$$\frac{\partial f}{\partial t} + p_R \frac{\partial f}{\partial R} + \frac{p_\phi}{R^2} \frac{\partial f}{\partial \phi} + p_z \frac{\partial f}{\partial z} - \left( \frac{\partial \Phi}{\partial R} - \frac{p_\phi^2}{R^3} \right) \frac{\partial f}{\partial p_R} - \frac{\partial \Phi}{\partial \phi} \frac{\partial f}{\partial p_\phi} - \frac{\partial \Phi}{\partial z} \frac{\partial f}{\partial p_z} = 0 \quad (4.39)$$

Simulating an axisymmetric disk that remains static over time allows for the removal of derivatives concerning  $t$  and  $\phi$

$$p_R \frac{\partial f}{\partial R} + p_z \frac{\partial f}{\partial z} - \left( \frac{\partial \Phi}{\partial R} - \frac{p_\phi^2}{R^3} \right) \frac{\partial f}{\partial p_R} - \frac{\partial \Phi}{\partial z} \frac{\partial f}{\partial p_z} = 0 \quad (4.40)$$

Subsequently, this equation can be multiplied by  $p_R$ ,  $p_\phi$ , or  $p_z$  integrated over the momenta  $p_R = v_R$ ,  $p_\phi = Rv_\phi$ ,  $p_z = v_z$ , and then the momenta can be expressed in terms of velocities to yield:

$$\frac{\partial \overline{v v_R^2}}{\partial R} + \frac{\partial \overline{v v_R v_z}}{\partial z} + v \left( \frac{\overline{v_R^2} - \overline{v_\phi^2}}{R} \frac{\partial \Phi}{\partial R} \right) = 0 \quad (4.41)$$

$$\frac{1}{R} \frac{\partial (R v \overline{v_R v_z})}{\partial R} + \frac{\partial (v \overline{v_z^2})}{\partial z} + v \frac{\partial \Phi}{\partial z} = 0 \quad (4.42)$$

$$\frac{1}{R^2} \frac{\partial (R^2 v \overline{v_r v_\phi})}{\partial R} + \frac{\partial (v \overline{v_z v_\phi})}{\partial z} = 0 \quad (4.43)$$

where  $\nu$  is the vertical frequency

$$\nu^2(R) = \left( \frac{\partial^2 \Phi}{\partial z^2} \right)_{(R, z=0)} \quad (4.44)$$

In this case, if  $f$  is known to be of the form  $f(H, L_z)$ , the mixed moments in the equations above vanish and can be integrated to give

$$\overline{v_R^2}(R, z) = \overline{v_z^2}(R, z) = \frac{1}{\nu(R, z)} \int_z^\infty dz' \nu(R, z') \frac{\partial \Phi}{\partial z'} \quad (4.45)$$

$$\overline{v_\phi^2}(R, z) = \overline{v_R^2} + \frac{R}{\nu} \frac{\partial (v \overline{v_R^2})}{\partial R} + R \frac{\partial \Phi}{\partial R} \quad (4.46)$$

where  $\overline{v_R}$  and  $\overline{v_\phi}$  are the mean radial and tangent velocities. When computing the tangent velocities, the consideration of asymmetric drift becomes necessary. This phenomenon refers to the rotational motion of stars within a galaxy, where the average velocity  $\overline{v_\phi}$  is smaller than

the circular velocity  $v_c$  at the corresponding radius.

The asymmetry in the drift arises from the exponential density profile of the galactic disk. The decline of velocity dispersion  $\sigma_R$  with distance from the center. Consequently, a larger fraction of particles with more eccentric orbits is located closer to the center of the disk, compared to the fraction of particles originating from orbits further out.

The asymmetric drift is defined as

$$v_a \equiv v_c - \bar{v}_\theta \quad (4.47)$$

With  $v_c$  the circular speed,  $v_c^2 = R(\partial\Phi/\partial R)$ , and  $v_a$  the axisymmetric drift.

$$v_a \simeq \frac{\bar{v}_R^2}{2v_c} \left( \frac{\sigma_\phi^2}{\bar{v}_R^2} - 1 - \frac{\partial \ln(\bar{v} \bar{v}_R^2)}{\partial \ln(R)} - \frac{R}{\bar{v}_R^2} \frac{\partial(\bar{v}_R \bar{v}_z)}{\partial z} \right) \quad (4.48)$$

### 4.3.4 Positions and velocities initialization

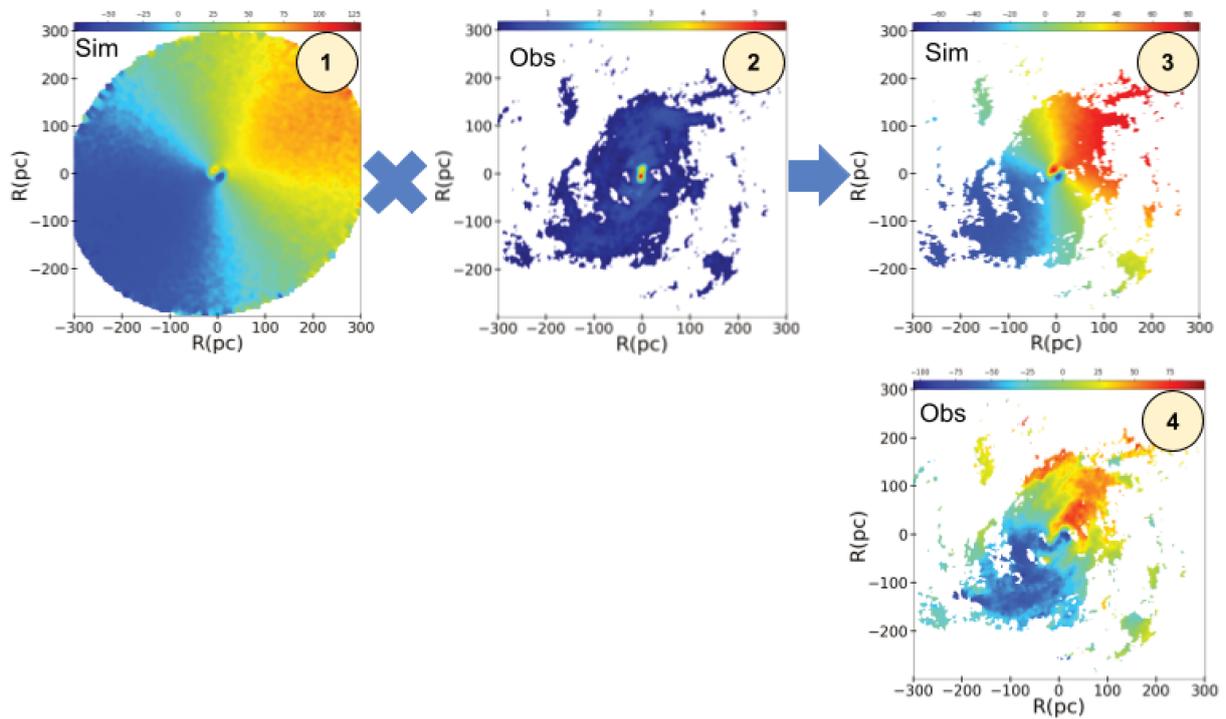
I first randomly initialize the particles' positions in a cylindrical referential center on the SMBH. The angle  $\theta$  is picked from a uniform distribution between 0 and  $2\pi$ . Then, the particles are distributed  $r$  and  $z$  following the Miyamoto & Nagai (1975) density distribution (see Eq.4.3).

Then, I calculate the velocity dispersions (Eqs.4.27, 4.28 and 4.33) and the axisymmetric drift (Eq. 4.48). In my simulation, I want particles to have random velocities within the range given by the velocities dispersions. For each dimension, I generate independent, standard, normally distributed random numbers and then, I multiply it by the velocity dispersion.

### 4.3.5 Matching the ALMA observations

To compare the model with the ALMA observational data, I generated data cubes by projecting the model onto the sky using the best-fit large-scale inclinations and position angles. I then computed the line-of-sight velocity distribution. The pixel size in the data cubes was chosen to match that of the observed data (ranging from 4 to 7 pc, depending on the galaxy), with velocity channels of 10 km/s. The data were smoothed to match the observed beam, resulting in a cube of dimensions (250, 250, 60) for each simulation.

Since the gas distribution exhibits asymmetry and patchiness, which affect the mass-weighted velocity within each observed beam, I applied a normalization process to the model cube. Specifically, I normalized the model cube pixel by pixel in the 2D projection using the zeroth moment map of the CO observations. This normalization serves as a multiplicative filter for my homogeneous gas disks. Consequently, each CO spectrum at every position in the model is normalized to the observed integrated flux at that corresponding position. A demonstration of how the application of a multiplicative filter alters my simulated uniform gas disk to replicate the observed gas distribution is shown in Fig.4.4.



**Figure 4.4:** Illustration of the impact of utilizing the  $0^{th}$  order as a multiplicative filter in my NGC 5643 simulations. It includes the following elements: (1) the initial first-moment map from the simulated data cube, (2) the observed  $0^{th}$  order as the multiplicative filter, (3) the updated first-moment map following filter application, and (4) the observed first-moment map for comparison with the simulation results.

## 4.4 ML approach to estimate the $M_{\text{BH}}$

I have chosen to employ ML techniques to estimate the masses of BH using the ALMA data cubes of the GATOS core sample. These 3D data cubes can be effectively transformed into 2D images, enabling us to apply CNNs for analysis. Although everyday data tends to be more complex, CNNs have proven to be highly effective in such cases, leading us to expect their suitability for astrophysical data as well.

As of now, my sample consists of only 7 galaxies, which implies that applying an ML method to such a small dataset may not yield the most optimal results. However, I anticipate the availability of a larger sample in the future. Additionally, the presence of data in the ALMA archives presents further opportunities for analysis and investigation, thus motivating my efforts toward automating the BH mass estimation process. Although scaling relations offer a certain level of automation, their intrinsic errors restrict the full exploitation of the high-resolution data. The ML method presented in this study serves as a proof of concept that ML can be applied to ALMA data to estimate  $M_{\text{BH}}$  with physical significance and is a first step toward an automated approach that leverages the data resolution to achieve precise  $M_{\text{BH}}$  estimations.

### 4.4.1 Supervised learning approach

The principle of my approach is to run a large number of numerical simulations, fitted to each studied galaxy. With this method, the gravitational potential of the galaxy disk is assumed to be known from HST images, within some uncertainties. The simulations are subsequently conducted by systematically varying the parameters within the defined uncertainty ranges. This includes the mass and scales of the bulge and disk components, as well as the characteristics of the gas and stellar populations. Inclination and position angles are fixed to the values estimated from the observations (García-Burillo et al. 2021), except in some cases, where the nuclear disk is tilted with respect to the large-scale disk. The possible mass of the SMBH is varied in a large range, compatible with the observations. A large number of numerical simulations with different  $M_{\text{BH}}$  is used to train an artificial neural network model for each galaxy. The trained model can then be used to infer the estimated  $M_{\text{BH}}$  from the real observation.

Since there is a limited number of different galaxies in my sample, it is not relevant to train a neural network to infer the  $M_{\text{BH}}$  from the CO(3-2) molecular gas moment maps of any galaxy. In contrast, the galaxy sample is small enough to train a CNN for each galaxy. That means creating a training set, a validation set, and a test set, and then training a different CNN for each different galaxy. For each observed galaxy, I generated 25,000 simulated data cubes. I divided my  $M_{\text{BH}}$ -labeled simulations into three different sets: a training set of 16500 simulations, a validation set of 3500 simulations, and a test set of 5000 simulations.

The different parameters of the simulations are described at the beginning of Sect. 4.3. For each simulation, all the parameters are taken randomly from a uniform distribution between the limits with physical meaning (see Table 4.1), except the  $IA$  and the  $PA$ . I can deduce from the moment maps a range for the radial scale of the gas  $R_{\text{gas}}$  and since the molecular disk is thin, I

Parameters	Min	Max
$M_{\text{bulb}}$	0.1	1.0
$R_{\text{bulb}}$	0.1	0.5
$H_{\text{bulb}}$	0.1	0.5
$M_{\text{gal}}$	1.0	10.0
$R_0$	3.0	10.0
$H_t$	0.001	0.003
$\log M_{\text{BH}}$	4.0	9.0
$RAPM$	0.019	0.021
$R_{\text{gas}}$	$R_{\text{obs}}$	$R_{\text{obs}} + 0.7$
$H_{\text{gas}}$	$R_{\text{obs}}/10$	$(R_{\text{obs}} + 0.7)/10$

**Table 4.1:** Lower and upper limits of the simulation parameters used to generate the training, validation, and test sample for each galaxy. The 12 parameters are described in the second paragraph of the 4.3. The  $RAPM$  parameter has no units,  $M_{\text{bulb}}$  and  $M_{\text{gal}}$  are in units of  $2.25^9 M_{\odot}$ ,  $M_{\text{BH}}$  in  $\log(M/M_{\odot})$  and  $R_{\text{bulb}}$ ,  $H_{\text{bulb}}$ ,  $R_0$ ,  $H_t$ ,  $R_{\text{gas}}$  and  $H_{\text{gas}}$  are in kpc.  $R_{\text{obs}}$  is the molecular disk observed radius in the data (see Fig.4.11 to Fig.4.17 for the values).

choose to keep the ratio  $H_{\text{gas}}/R_{\text{gas}} < 0.1$ . For the stellar disk  $R_0$  and  $M_{\text{gal}}$  are large with respect to the scale probed by my observations/simulations, and their possible variations are small. The  $IA$  and  $PA$  are set constant in all the simulations for a galaxy. They were taken as estimated by [García-Burillo et al. \(2021\)](#) based on the gas kinematics.

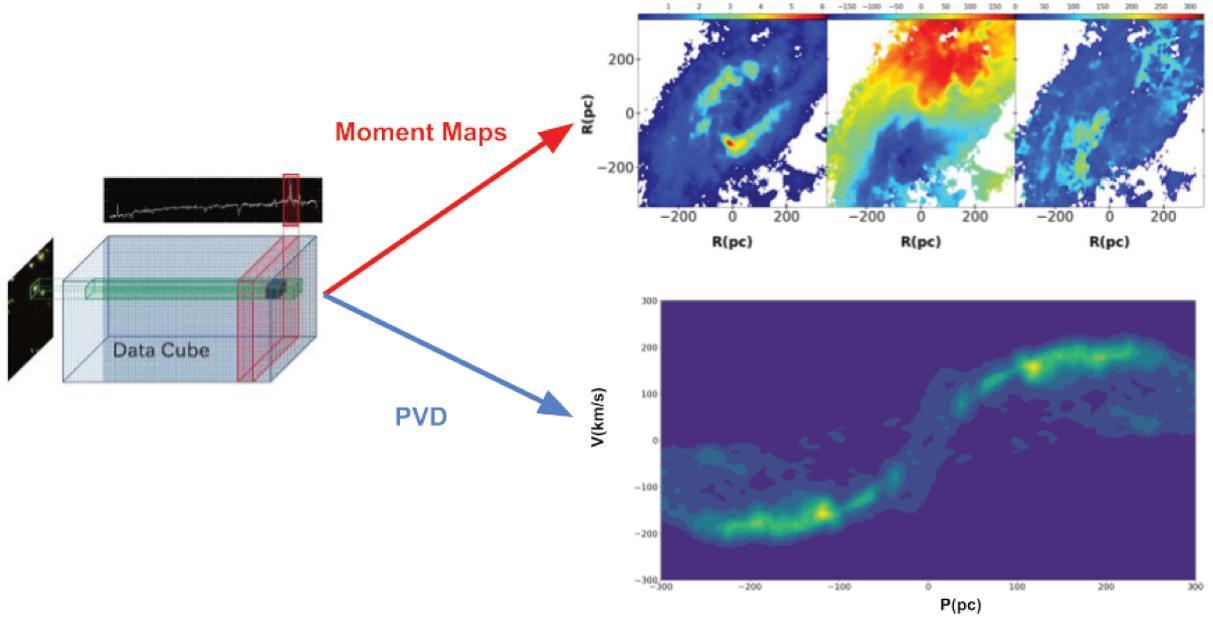
## 4.4.2 Preparing the inputs

The GATOS CO(3-2) molecular gas observations are in the form of spectroscopic data cubes which are three-dimensional arrays that combine spatial and spectral information. Two of the dimensions represent the spatial coordinates and define the position in the sky. The third dimension represents the intensity or flux of light at each spatial position and each wavelength or frequency.

Gathering information directly from spectroscopic data cubes can be a challenging task, especially when these data cubes are utilized as inputs in ML processes ([Hartley et al. 2023](#)). The introduction of a third dimension into the input data significantly amplifies data complexity and can lead to considerable increases in computational requirements. To mitigate the resulting computational load and efficiently extract valuable insights from the data cube, it becomes essential to optimize and fine-tune the ML architectures employed.

However, I can take an easier approach and pre-process the data cubes to put the meaningful information more straightforwardly in the shape of two-dimensional images. This will make it easier for the NN to learn how to estimate the  $M_{\text{BH}}$  from the simulated data cubes. Two commonly used representations to study the CO(3-2) gas dynamics are the PVD and the moment maps (see Fig.4.5).

The PVD allows the gas velocity and concentration as a function of the radius in a chosen direction to be seen in detail. Since the PVD is used to better see the effect of the SMBH



**Figure 4.5:** Example of NGC 7582 position-velocity diagram (PVD) and first three-moment map obtained from the CO(3-2) ALMA data cubes observations. I show on the left how information is contained in a spectroscopic data cube, on the upper right I present the the first three moment maps calculated from the data cube, and on the lower right I show the symmetrized PVD.

on the gas dynamics, I select the PVD along the axis with the higher velocity gradient at the smallest radius possible. Nevertheless, this representation has certain limitations. The PVD solely provides insight into a particular direction of the gas dynamics, thereby leading to the partial depiction of information. Consequently, the selection of a specific direction can influence the bias in the information presented in the PVD. Furthermore, studying the gas in only one direction makes the PVD more sensitive to a lack of gas at a specific radius, such as the center of NGC 4388 (4.11).

In addition to studying the gas in one specific diagram, I consider a representation that encapsulates more information about the entire gas dynamic distribution by calculating the first two moment maps:

The 0th-moment map is the integrated intensity over the spectral axis, proportional to the gas surface density.

$$M_0 = \int I_v dv \quad (4.49)$$

The 1st moment represents the velocity  $V$  weighted by the intensity.

$$M_1 = \frac{\int V I_v dv}{M_0} \quad (4.50)$$

The PVD and the first moment maps are calculated for each simulated cube. Finally, moment by moment, and for the PVD, I normalize the value of each pixel by dividing it by the highest one. The PVD and first-moment maps of the galaxies in my sample are presented in Appendix B.1. Initially, the estimation of the SMBH mass will be conducted using the PVD as

input, followed by employing the first-moment map as input.

### 4.4.3 ML architecture

Layer	Filters	Size	Stride
Conv1	16	7x7	2
Conv2	20	7x7	1
Conv3	20	7x7	2
Conv4	24	7x7	1
Conv5	64	5x5	2
Conv6	80	5x5	1
Conv7	80	5x5	2
Conv8	128	5x5	1
Max Pooling		3x3	2
Dense1	2048		
Dense2	1024		
Dense3	512		

**Table 4.2:** Architecture of the neural network used for the MBH estimations with the PVD as inputs

To make my  $M_{\text{BH}}$  estimations via ML I used a CNN-supervised training method (e.g., [LeCun et al. 2015](#)). Our neural networks were developed using Python’s library Keras ([Chollet et al. 2015](#)). The PVD and the moment maps represent the information differently, so I will use different neural network architectures to get the best out of their distinct characteristics.

The NN taking the PVD as inputs is composed of a succession of eight convolution layers, a max pooling layer, and at the end, three dense layers (i.e. Table 4.2). The NN that takes the first moment maps as inputs are composed of a succession of convolution and max pooling layers followed by three dense layers (i.e. Table 4.3). These NN are the result of many trials and errors.

Dropout was added to the second dense layer in both architectures with a drop rate of 20%. It means that in this layer, 80% of the neurons are randomly selected to remain active at each path through the network. Usually, the dropout technique is used as a regularization method during training to help prevent overfitting ([Srivastava et al. 2014](#)). It can also improve the predictive performance of neural networks in several tasks. Another way to use dropout is to keep it activated at inference time (see Sect. 3.4.7). At each inference, the dropout will randomly select a different combination of neurons, resulting in a different estimation. It is equivalent to an approximation of the probabilistic deep Gaussian process. Thus, making multiple inferences for the same input will not give one value but a probability distribution on the predicted value, similar to what is produced by a Monte Carlo Markov Chain process. This distribution can be studied to characterize the model uncertainties for that input.

I used the Root Mean Square Error ([RMSE](#)) as loss function. It is a common loss function

Layer	Filters	Size	Stride
Conv1	32	11x11	4
Max Pooling 1		3x3	2
Conv2	64	5x5	1
Max Pooling 2		3x3	2
Conv3	64	5x5	1
Max Pooling 3		3x3	2
Conv4	64	5x5	1
Max Pooling 4		3x3	2
Conv5	96	3x3	1
Conv6	96	3x3	1
Conv7	96	3x3	1
Max Pooling 5		3x3	2
Dense1	2048		
Dense2	1024		
Dense3	512		

**Table 4.3:** Architecture of the neural network used for the MBH estimations with the moment maps as inputs

used in regression problems:

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (Y_{pred,i} - Y_{real,i})^2}{n}} \quad (4.51)$$

with  $n$  the number of predictions made,  $Y_{pred,i}$  the predicted values and  $Y_{real,i}$  the actual values. To help stabilize the weights in my models during the training, I normalized the output targets (i.e., the  $M_{BH}$ ) by dividing them by the highest value among the output targets.

I trained my models during 2000 epochs and took the weights of the epoch with the lowest RMSE on the validation set. I used an initial learning rate of 0.0001 and the Adam optimizer (Kingma & Ba 2014). The training was made with a mini-batch size of 128 images.

#### 4.4.4 Estimating the reliability of models

Multiple statistical indicators can be used to evaluate the quality of a regression model. These indicators can be used as metrics to help compare the training results of the different models. A first indication of the quality can be obtained by looking at the value of the loss function on the test set. However, studying the loss function alone is not enough to evaluate a model because it can be highly degenerate. Furthermore, the RMSE function goes from 0 (best fit possible) to  $+\infty$  (worst fit possible). The lack of an upper bound can make the RMSE interpretation difficult but in my case, the error directly measures the average difference between the estimated  $M_{BH}$  and the actual ones.

A more reliable indicator would be the coefficient of determination (Wright 1921)<sup>1</sup>. It is

<sup>1</sup><https://naldc.nal.usda.gov/download/IND43966364/PDF>

a statistical value that represents the proportion of a dependent variable variance that is predictable from the independent variables:

$$R^2 = 1 - \frac{\sum_{i=1}^n (Y_{pred,i} - Y_{real,i})^2}{\sum_{i=1}^n (\overline{Y_{pred}} - Y_{real,i})^2} \quad (4.52)$$

where the notations are the same as the RMSE ones and  $\overline{Y_{pred}}$  is the mean of the predictions. In a regression problem, the coefficient of determination measures how well the model fits the real data. It has the advantage of being confined by construction between 0 and 1, 0 being the worst fit possible and 1 meaning that the model perfectly fits the data. By studying the RMSE and the  $R^2$  I can have a global evaluation of the capacity of a model to make the correct prediction.

To gain more information about the in-depth behavior of a model, I also use the Relative Errors (RE):

$$RE_i = \frac{Y_{pred,i} - Y_{real,i}}{Y_{real,i}} \quad (4.53)$$

Studying the distribution of the relative errors of a model when predicting the test set can reveal any systematic bias during the predictions. Furthermore, since my datasets are made of simulations, I can check if any of the simulation parameters primarily influence the results of the predictions by looking at the relations between the parameter values and the relative errors.

## 4.5 Results on $M_{BH}$ estimation

### 4.5.1 Reliability of ML models

As described above in Sec 2.2, 7 of the 10 galaxies in my sample have sufficiently robust CO(3-2) data in the central regions to enable  $M_{BH}$  estimations with my ML method: NGC 4388, NGC 5506, NGC 5643, NGC 6300, NGC 7314, NGC 7465, and NGC 7582. Before comparing my  $M_{BH}$  estimations with those from the literature, I need to verify their reliability with the tools mentioned in Sect. 4.4.4.

I report in Table 4.4 the measure of the RMSE, the  $R^2$ , and the mean RE of my models on their test set for both the ones estimating the  $M_{BH}$  via the PVD and the ones using the first-moment maps. The analysis reveals that the diverse evaluations of the fitting quality exhibit a high degree of reliability and demonstrate consistency across the various indicators for both architectures. The RMSE values are at worst 0.0367 and the  $R^2$  are all above 0.946, confirming that all the models are capable of making a reliable estimation of the  $M_{BH}$  on simulated data that were not present during the training process.

I show in Fig 4.6, the models RE distributions estimations on the 5000 simulations that composed their respective test sets. As shown in Table 4.4, the mean RE is  $\sim 0$  for the fourteen models, two for each galaxy, one with the PVD as input and one with the moment map as input. The histograms are centered at the value of zero and prominently feature a pronounced peak at this center point. The locations of the centers around 0 show that there is no systematic

ID	RMSE		$R^2$		$\overline{RE}$	
	PVD	Mom	PVD	Mom	PVD	Mom
NGC 4388	0.0341	0.0341	0.986	0.946	0.0020	0.0007
NGC 5506	0.0335	0.0367	0.986	0.947	0.0060	-0.0032
NGC 5643	0.0194	0.0178	0.995	0.996	-0.0019	0.0001
NGC 6300	0.0243	0.0180	0.995	0.996	-0.0002	0.0005
NGC 7314	0.0312	0.0277	0.984	0.990	-0.0029	-0.0041
NGC 7465	0.0246	0.0255	0.993	0.992	0.0010	0.0014
NGC 7582	0.0263	0.0220	0.992	0.994	-0.0002	-0.0031

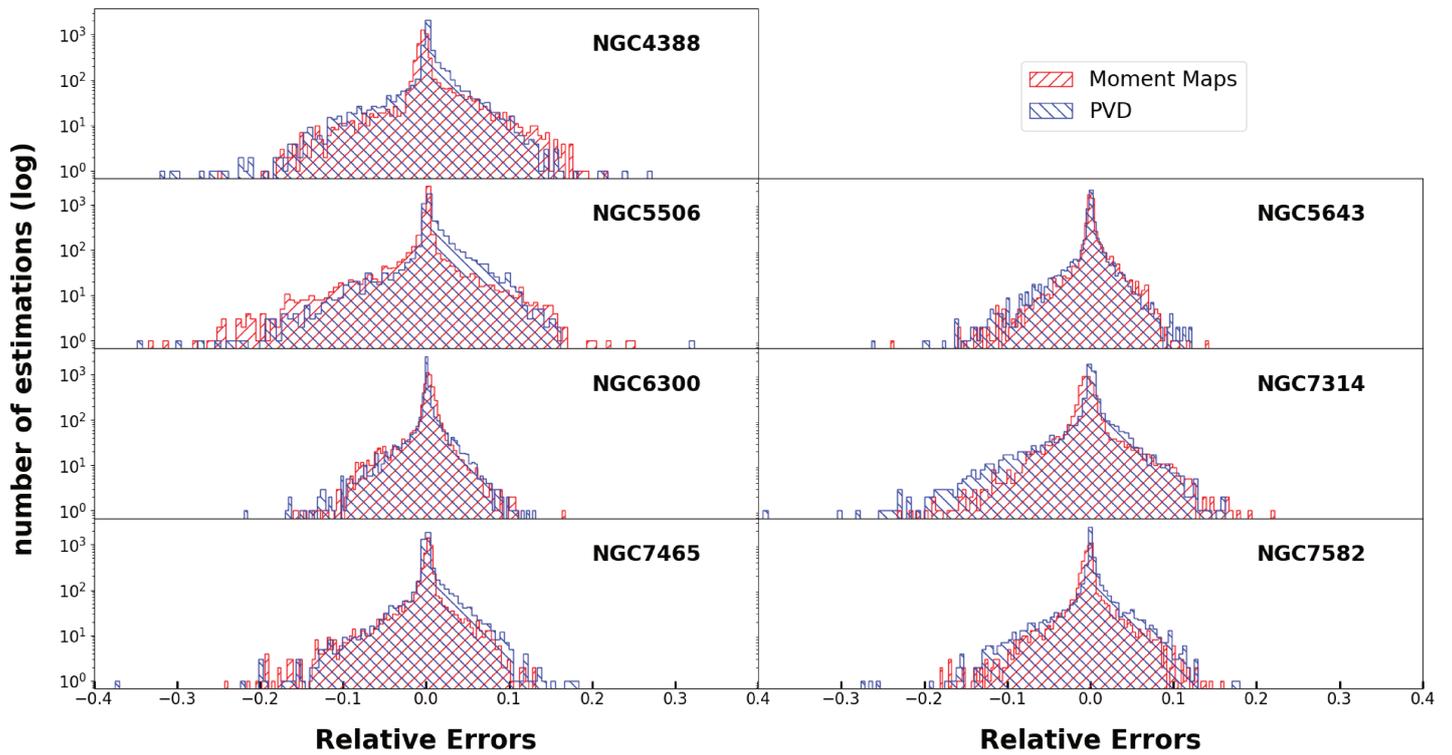
**Table 4.4:** RMSE,  $R^2$  and mean RE of our models on the test sets. The RMSE ( $> 0$ ),  $R^2$  (between 0 and 1), and the RE are defined in Section 4.4.4

bias during the predictions and the relatively symmetrical distributions indicate that there is no constant over/underestimation of the  $M_{\text{BH}}$ .

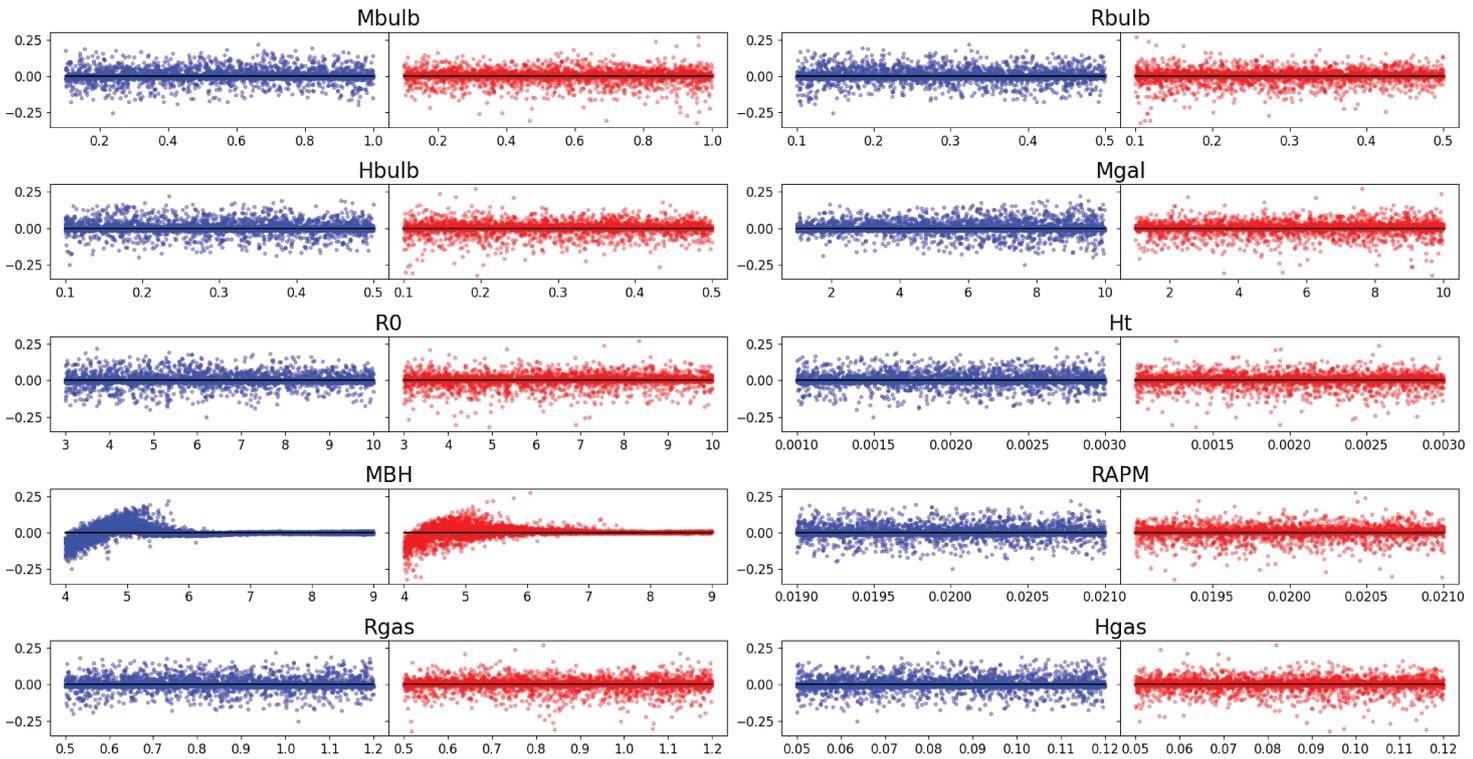
The mock datasets used for training, validation, and testing are produced via a numerical simulation with 10 free parameters. I can use the relative errors to make sure that the confidence of my models  $M_{\text{BH}}$  predictions are not influenced by any other parameters than the  $M_{\text{BH}}$ . For each parameter, I plot the relative errors of the  $M_{\text{BH}}$  predictions as a function of the parameter values. I observe the same behavior for each galaxy so I present here only the case of NGC 6300 (i.e. Fig 4.7), the rest can be found in the Appendix (see Fig.B.1 to B.6). For every parameter, except for the  $M_{\text{BH}}$ , the distribution is uniform and centered around zero. This is strong evidence for considering that the models prediction does not depend on any other parameter other than  $M_{\text{BH}}$ .

To further investigate the behavior of my models, I represent the 2D histograms of the RE as a function of  $M_{\text{BH}}$ . I present here the case of NGC 4388 (i.e. Fig.4.8) the rest can be found in the Appendix (see Fig.B.7 to B.12). All the distributions have the same shape. From  $\log(M_{\text{BH}}/M_{\odot})$  of 4.0 to  $\sim 4.5$  dex, the distribution means increase linearly from  $< 0$  to  $\sim 0$  and their scatter stays constant. At  $\sim 4.5$  dex, the scatters get bigger and I see outliers that are underestimated. Then, the distribution scatters get thinner and the outliers number decreases as the  $M_{\text{BH}}$  predicted increases until  $\sim 6.5$  dex or  $\sim 7$  dex and stays constant afterward. Let us note the exception of the NGC 5643 model using the moment maps. Unlike the other models, its mean is already at 0 for  $M_{\text{BH}}=4$  dex and stays constant with  $M_{\text{BH}}$ . Its scatter is also constant and uniform.

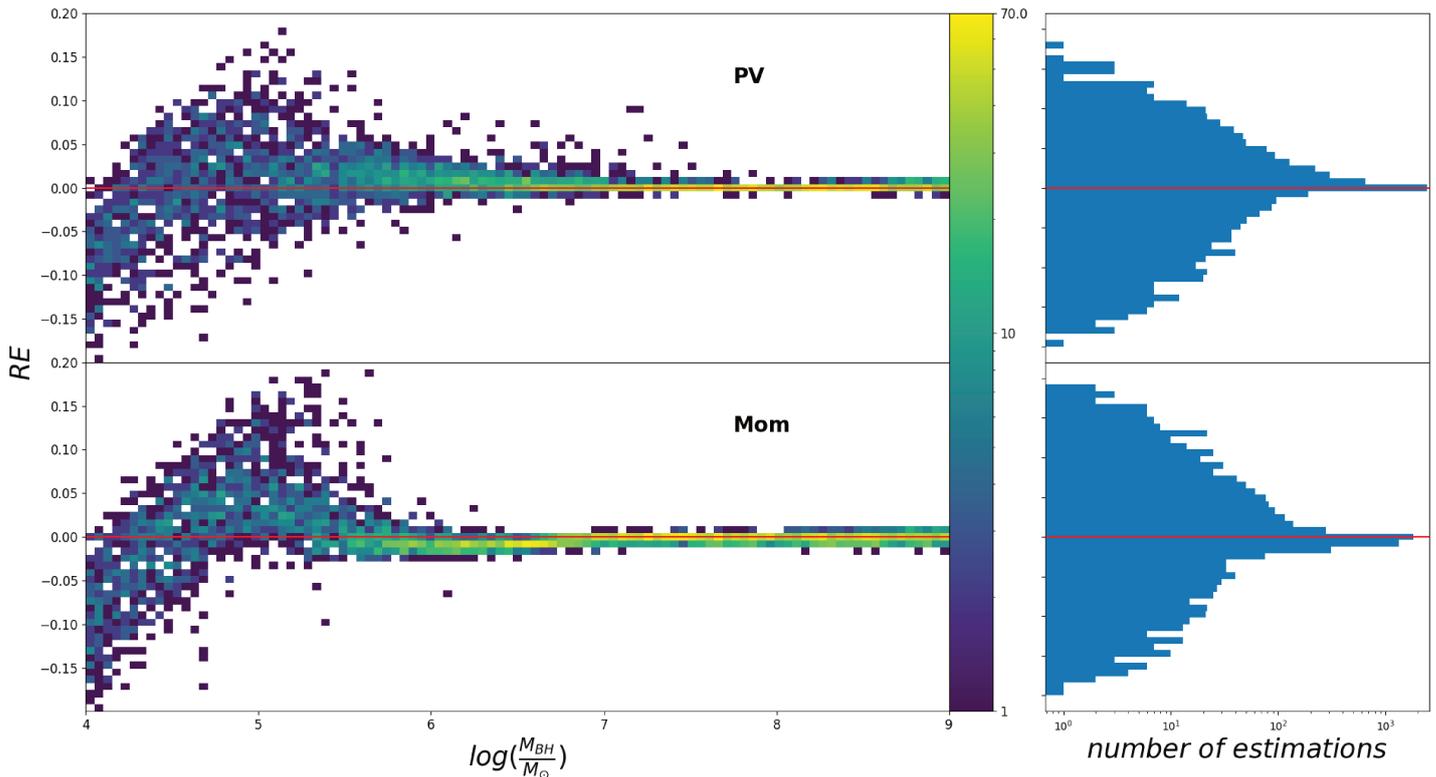
As the  $M_{\text{BH}}$  gets lower, fewer and fewer pixels contains information about it. Thus, it becomes harder for the neural network to extract meaningful information out of the data. This may explain why I observe the high scatter below  $M_{\text{BH}}\sim 6$  dex.



**Figure 4.6:** Histograms of the RE of my models when estimating the  $M_{\text{BH}}$  of the 5000 simulations from their respective test set. I represent in blue the models trained to estimate the  $M_{\text{BH}}$  with the PVD, and in red the models trained with the Moment maps.

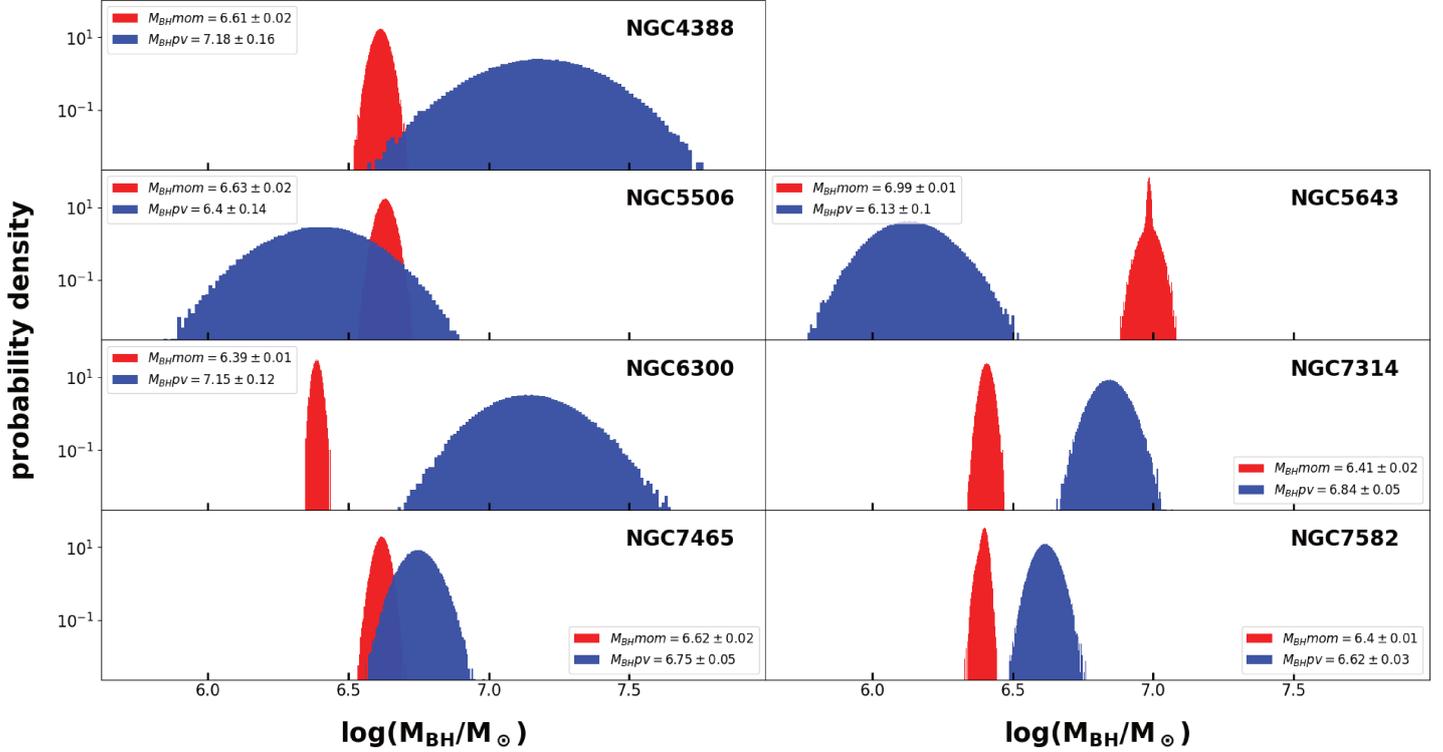


**Figure 4.7:** NGC 4388’s model relative errors on the  $M_{\text{BH}}$  estimations made on the test set as a function of the simulation parameter’s values. I represented in blue the error of the model using the PVD and the one using the first-moment maps in red. The vertical axes represent the relative errors and the horizontal axes the parameter’s values:  $RAPM$  has no units,  $M_{\text{bulb}}$  and  $M_{\text{gal}}$  are in units of  $2.25 \times 10^9 M_{\odot}$ ,  $M_{\text{BH}}$  is in  $\log(M/M_{\odot})$  and  $R_{\text{bulb}}$ ,  $H_{\text{bulb}}$ ,  $R_0$ ,  $H_t$ ,  $R_{\text{gas}}$  and  $H_{\text{gas}}$  are in kpc.



**Figure 4.8:** Two-dimensional and one-dimensional histograms of the relative errors RE of my models when estimating the  $M_{\text{BH}}$  of the test set. On the left, I plot the RE of NGC 4388 models when estimating the  $M_{\text{BH}}$  of the test set as 2D histograms. On the right, I plot the same RE but as a 1D histogram. On top, I plot the PVD model RE, in the middle the first-moment map model RE and at the bottom, I plot the RE difference between the two models. The y-axis represents the RE values. The x-axis of the 2D histograms is in  $\log(\frac{M_{\text{BH}}}{M_{\odot}})$  and the 1D histograms axis is in a number of estimations.

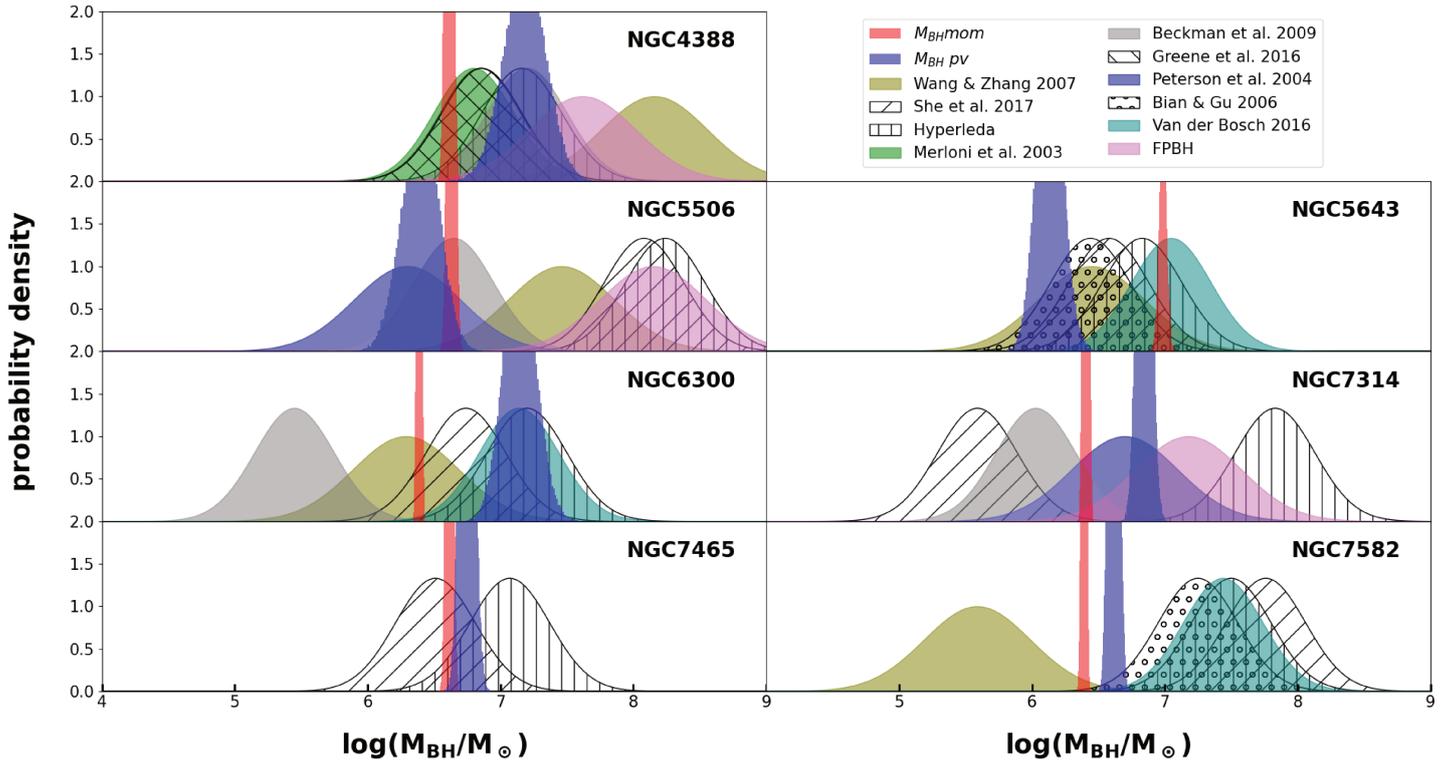
## 4.5.2 $M_{\text{BH}}$ estimations results



**Figure 4.9:** Probability distributions of the  $M_{\text{BH}}$  predicted by my models for the seven galaxies in my sample. I have plotted the estimations made with the PVD as input in blue, and the estimations made with the first moment maps as inputs in red.

The statistical indicators showed that the models succeeded in predicting precisely the  $M_{\text{BH}}$  on unseen simulated data when the  $M_{\text{BH}}$  are above  $\sim 6$  dex. So, if the numerical simulation can accurately emulate the observations, my models should be able to have the same behavior when estimating the  $M_{\text{BH}}$  of their corresponding observations. I show in Fig 4.9 my models estimations on their respective observations. I made  $10^5$  predictions per galaxy. The observed scatter on the  $M_{\text{BH}}$  predictions comes from the MC dropout (as explained in Section 4.4.4). These scatters represent the models confidence in their predictions and do not have a predefined shape. The observed Gaussian shapes come from the fact that my models are confident in their estimations. Moreover, it's important to note that the models are trained on numerical simulations that do not encompass all the underlying physics observed in reality. Thus, the uncertainty shown here cannot be associated with physical ones but only with the models confidence in their estimations.

I can observe that the distributions are all Gaussian, except for the NGC 5643 model using the moment maps. This particular model seems to be the sum of two Gaussians with different dispersions but centered around the same value (6.99 dex). The width of the Gaussians given by the PVD models is broader compared to the moment maps models. Furthermore, the models scatter is constant for the moment map models, whereas it is galaxy-dependent for the PVD models. This can be explained if the PVD shows only a part of the gas dynamic. The



**Figure 4.10:** Comparison of the probability distributions of the  $M_{\text{BH}}$  predicted by my models and the ones found in the literature. I represent in blue the results of the estimation made from the PVD and in red the estimation made from the moment maps. The other Gaussians represent the estimations found in the literature.

effectiveness of the neural network will rely on the extent to which the presented information is valuable for estimating the  $M_{\text{BH}}$ . On the contrary, the moment maps show the entirety of the gas dynamics, so helpful information is always represented. On one hand, the moment maps contain information primarily concentrated in the values of the central pixels, providing a localized representation of the  $M_{\text{BH}}$ -related information. On the other hand, the PVD captures information through the velocity distribution as a function of position. Particularly, in the shape of this distribution in the vicinity of the center. This results in a more sparse distribution of information which makes it more difficult for the network to relate this information to the  $M_{\text{BH}}$ . Another explanation is that I use different architectures for the PVD and the moment maps.

I can observe that, in Fig.4.10, for four out of the seven galaxies, the estimations of the PVD and the moment maps do not overlap. This can be attributed to the same hypotheses as to why the PVD distributions are broader than the moment maps.

I compare my  $M_{\text{BH}}$  estimations with the ones previously made using classical methods in Table 4.5 and Fig 4.10. For the seven galaxies in my sample, the  $M_{\text{BH}}$  predicted by my models is coherent with previous estimations. There exists some overlap with at least one previous estimation, except for NGC 7582. For the latter, my estimations fall in between the one by Wang & Zhang (2007) with reverberation mapping and the three others made using the  $M_{\text{BH}} - \sigma$  relation (e.g. Bian & Gu 2007; van den Bosch 2016; She et al. 2017).

ID	$\log M_{\text{BH}}$ PVD	$\log M_{\text{BH}}$ Mom	M - $\sigma$	Reverberation	FPBH	Ref
NGC4388	$7.18 \pm 0.16$	$6.61 \pm 0.02$	$7.17^{8.16}_{6.85}$	—	7.62	3,4,5,6,7,8
NGC4941	—	—	$6.83^{7.14}_{6.83}$	6.34	—	2,3,6,7
NGC5506	$6.40 \pm 0.14$	$6.63 \pm 0.02$	$7.61^{8.24}_{6.65}$	6.30	8.15	1,3,4,6,7,8,9
NGC5643	$6.13 \pm 0.10$	$6.99 \pm 0.02$	$6.73^{7.05}_{6.44}$	$6.38^{6.45}_{6.30}$	—	2,3,5,6,7
NGC6300	$7.15 \pm 0.12$	$6.39 \pm 0.01$	$6.63^{7.14}_{5.45}$	6.29	—	3,4,5,6,7
NGC6814	—	—	$6.96^{7.32}_{6.42}$	7.28	7.45	1,4,5,7,9
NGC7213	—	—	$8.49^{8.82}_{8.06}$	$7.35^{7.99}_{6.88}$	—	1,3,7
NGC7314	$6.84 \pm 0.05$	$6.41 \pm 0.02$	$6.48^{7.83}_{5.59}$	6.70	7.18	1,4,6,7,9
NGC7465	$6.75 \pm 0.05$	$6.62 \pm 0.02$	$6.79^{7.07}_{6.51}$	—	—	6,7
NGC7582	$6.62 \pm 0.03$	$6.40 \pm 0.01$	$7.56^{7.76}_{7.24}$	5.59	—	2,3,5,6,7

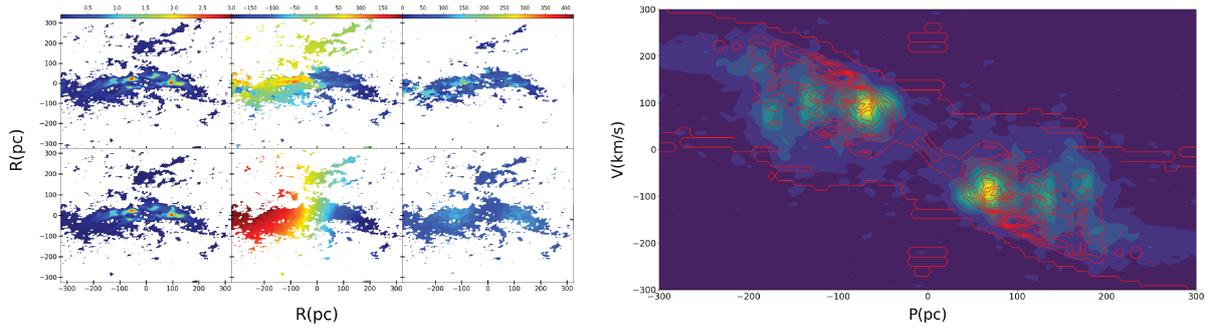
**Table 4.5:** Table summarizing the different MBH estimations found in the literature and the one I made in Section 4.2. The MBH is in  $M_{\odot}$ , and only the log of values are indicated. When multiple MBH estimations using the same method were available, I report here their mean, the highest values in superscript, and the lowest in the subscript. (1) Galaxies ID ; (2) Mean of the MBH probability distribution obtained with ML and the PVD as input (the errors are at  $1\sigma$ ) ; (3) Same as the previous column but with the moment maps as input ; (3) Mean of the MBH estimations with the M -  $\sigma$  (intrinsic errors of 0.3 dex) ; (4) MBH calculated with reverberation mapping (intrinsic errors of 0.4 dex) ; (5) FPBH masses (intrinsic errors of 0.4 dex) ; (6) MBH estimations references (1: [Peterson et al. \(2004\)](#), 2: [Bian & Gu \(2007\)](#), 3: [Wang et al. \(2007\)](#), 4: [Beckmann et al. \(2009\)](#), 5: [van den Bosch \(2016\)](#), 6: [She et al. \(2017\)](#), 7: M- $\sigma$  estimation made with the Hyperleda data, 8: FPBH mass estimated from the XMM-Newton and the FIRST data, 9: FPBH mass estimated from the XMM-Newton and the NVSS data).

## 4.6 Individual galaxies

The three moments of the CO(3-2) cubes from ALMA observations (top panel) are compared with the best fit of my simulations (bottom panel) on the left sides of Figs. 4.11 to 4.17. On the right side, the observed PV diagram is overlaid with the contours of the second model, optimized for this. The models are fitted

### 4.6.1 NGC 4388

The galaxy is almost edge-on ( $IA$  of  $79^\circ$ ), and its nuclear molecular disk reveals a ring of  $\sim 100$  pc radius, which corresponds to the inner Inner Lindblad Resonance (ILR) of the bar. The ring

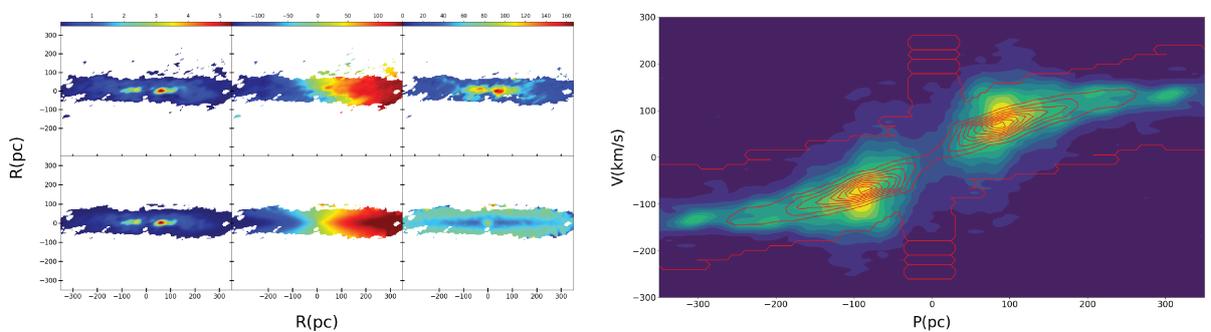


**Figure 4.11:** Left I present the first three moments maps of NGC 4388, obtained from the best-fit model, giving the adopted  $M_{\text{BH}}$ , compared with the NGC 4388 CO(3-2) molecular gas deduced from the [García-Burillo et al. \(2021\)](#) ALMA observations. From left to right, the integrated intensity, the velocity field, and the velocity dispersion; the top is the observation, and the bottom is the best-fit model. On the right, I present the observed PVD, with the best-fit model overlaid in red contours.

is not totally empty and shows clumps that enter the SoI, allowing a  $M_{\text{BH}}$  estimation. Some molecular gas is expelled along the minor axis and might be dragged by the ionized wind and radio jet ([Falcke et al. 1998](#)).

Figure 4.11 shows the model best fit found for NGC 4388. It is clear in this figure that the velocity field of the galaxy is highly asymmetric, and that my symmetric gravitational potential model cannot reproduce the observed first-moment map. While the gas with negative velocity distribution is well reproduced, the positive part is not. On the PVD panel, the model fits better than the symmetrized observation. The observed second moment shows a sudden drop of velocity dispersion at  $P \sim 200$  pc that is not present in the model. Moreover, the model has a large velocity gradient at the center contrary to the observations. This comes from the lack of gas in the SoI, so it is difficult to precisely measure the SMBH mass in this galaxy. It can be noted that the lowest red contour in the PV diagrams is less reliable, related to a low level of emission.

## 4.6.2 NGC 5506



**Figure 4.12:** Same as Fig. 4.11 for NGC 5506.

NGC 5506 is also almost edge-on ( $IA$  of  $80^\circ$ ), and possesses a nuclear molecular ring of radius  $\sim 50$  pc, corresponding to the ILR of its strong bar. There is however dense molecular

gas near the AGN, and  $\text{HCO}^+$  has been detected in the center (García-Burillo et al. 2021). The estimation of  $M_{\text{BH}}$  is therefore possible.

Figure 4.12, reveals in the center of the observed first-moment map that the gas velocity distribution has an S-like shape. This is probably due to the presence of a bar. Except for these details, the model can well reproduce the observed gas velocity distribution. However, it cannot do as well for the velocity dispersion distribution. The model gives a more or less constant velocity dispersion whereas two clumps with high-velocity dispersion are observed near the center and a low-velocity dispersion elsewhere. This effect is also seen in the PVD and could be also due to the bar. The low gas density in the center makes the SMBH mass estimation uncertain. This is reflected in the low-emission red contour of the model, which proposes in general a higher  $M_{\text{BH}}$  than the moment-based model.

### 4.6.3 NGC 5643

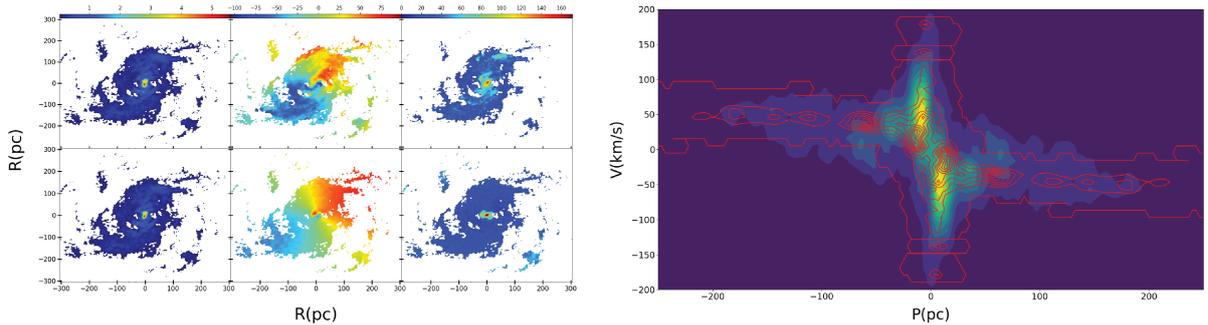
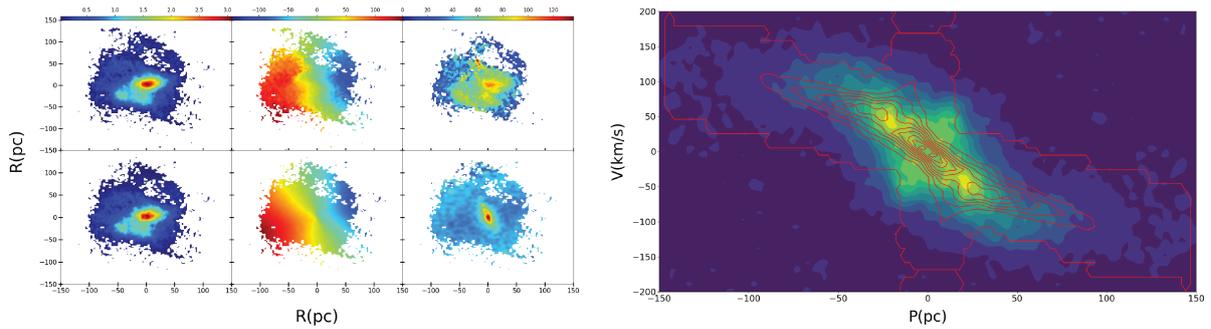


Figure 4.13: Same as Fig. 4.11 for NGC 5643.

This is an almost face-on barred spiral, with an  $IA$  of  $29^\circ$ . The strong stellar bar is almost E-W, with  $PA=85^\circ$ . At  $\sim 100$  pc scale, NGC 5643 CO(3-2) moment zero map shows the presence of a nuclear bar, prolonged by a spiral structure, but the CO bar is almost N-S, with  $PA=5^\circ$ , cf Fig 4.13. The very center reveals a tilted molecular torus, of diameter 26 pc, oriented almost N-S, perpendicular to the ionized gas outflow and radio jet (Alonso-Herrero et al. 2018; García-Bernete et al. 2021). Dense gas traced by  $\text{HCO}^+$  has been detected in the very center (García-Burillo et al. 2021).

The NGC 5643 model reproduces well the shift in the velocity orientation towards the nucleus (Fig.4.13). However, the second-moment map reveals a high-velocity dispersion in the nucleus, which is horizontal in the model while almost vertical in the observation. In the PVD, the model does not distribute the gas clumps at the same positions as in the observed diagram; however, the velocity gradient in the nucleus is reproduced, and the SMBH mass estimation appears robust. Note that the moment-based model leads to a higher  $M_{\text{BH}}$  in this galaxy, which is unusual, maybe due to the strong misalignment.



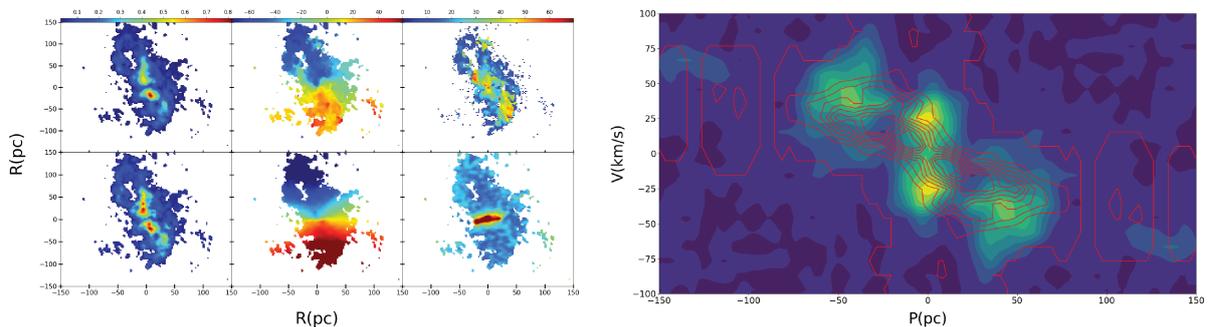
**Figure 4.14:** Same as Fig. 4.11 for NGC 6300.

## 4.6.4 NGC 6300

This is a strongly barred galaxy, moderately inclined ( $57^\circ$ ). At  $\sim 100$  pc scale, the CO(3-2) emission shows a detached nuclear disk, with a central concentration, allowing a determination of  $M_{\text{BH}}$ , cf Fig 4.14. This strong concentration corresponds to a molecular torus of 25 pc radius, also detected in  $\text{HCO}^+$ . The torus oriented at  $PA = 85^\circ$ , is perpendicular to the ionized gas outflow, detected by [Scharwächter et al. \(2016\)](#).

For the NGC 6300 fit (Fig.4.14), the first-moment map is well reproduced by the model, except for the wiggle at the center of the disk, likely due to the bar. The modeled second-moment map underestimates the velocity dispersion at the center but overestimates it near the edges. Furthermore, the central region with high-velocity dispersion is horizontal in the observation while it is almost vertical in the model, a likely consequence of the bar. The PVD panel reveals a reasonable fit, with a sufficient gas sampling of the SoI.

## 4.6.5 NGC 7314



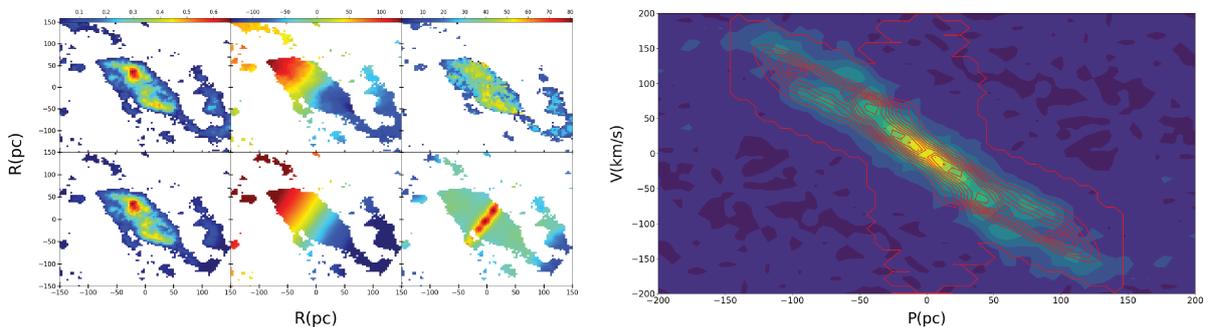
**Figure 4.15:** Same as Fig. 4.11 for NGC 7314.

This barred spiral galaxy looks relatively inclined ( $70^\circ$ ) on the sky, but is kinematically less ( $55^\circ$ ) since its morphological shape is biased by the bar. At  $\sim 100$  pc scale, the molecular disk reveals a central depletion, surrounded by two peaks, suggesting the existence of a ring of radius 20 pc, that could be an ILR resonant ring. Its PVD shows very little gas in the SoI but it is enough to estimate an upper limit of 6.2 dex on the  $\log(M_{\text{BH}}/M_\odot)$ . In this galaxy, the molecular

nuclear disk does not appear strongly misaligned with the large-scale one. It is also oriented along the radio source (Thean et al. 2000), which might be dominated by star formation.

The NGC 7314 fit (Fig.4.15) overestimates the velocity gradient in the outer parts of the disk. The model also overestimates the central velocity dispersion along the minor axis, while the dispersion is more spatially distributed in the observation. The PVD reveals a high central velocity dispersion, leading to a high estimation of the SMBH mass. This is however provided by the lowest emission red contour.

## 4.6.6 NGC 7465



**Figure 4.16:** Same as Fig. 4.11 for NGC 7465.

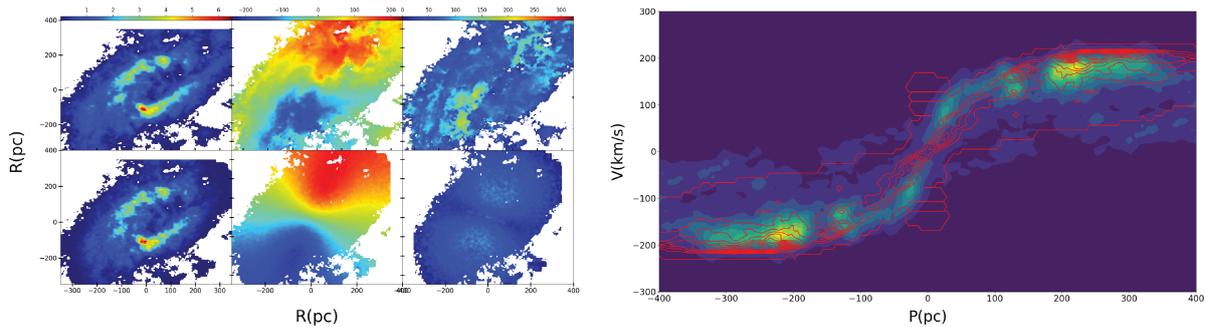
This barred early-type spiral is moderately inclined ( $60^\circ$ ) on the sky and its morphological major axis of  $PA = 160^\circ$  at large-scale is roughly perpendicular to the molecular disk at  $\sim 100$  pc scale, with  $PA = 45^\circ$ . The circum-nuclear disk in the regular rotation is dense, of 70 pc radius, and prolonged by a sparse spiral structure. There is enough dense gas towards the center to estimate  $M_{\text{BH}}$  through the ML method. The inferred  $\log(M_{\text{BH}}/M_\odot)$  is  $5.9 \pm 0.8$  dex.

The model for NGC 7465 (Fig.4.16) reproduces very well the observed first-moment map. The velocity dispersion is however gathered along the minor axis, while it is more spatially distributed in the observed second moment. The low-velocity gradient in the SoI region leads to a low mass estimation for the SMBH.

## 4.6.7 NGC 7582

NGC 7582 is a barred early-type spiral, highly inclined ( $68^\circ$ ), with a  $\sim 100$  pc scale molecular disk aligned with the large-scale galaxy disk, with  $PA = 156^\circ$ . The molecular gas is condensed in a 200 pc radius ring, which corresponds to an ILR ring. However, there is enough CO emission right into the very center, making it possible to estimate  $M_{\text{BH}}$ . There is even a concentration of CO emission in a small disk of radius 30 pc, corresponding to a molecular torus. An outflow of ionized gas is observed in the perpendicular direction (Davies et al. 2016).

The NGC 7582 model fits the observed velocity gradient in the disk relatively well, although its morphology is too regular (Fig.4.17). The observed first-moment map shows that the iso-velocity morphology gets thinner towards the nucleus, with a large central velocity gradient, which might be due to the bar. The velocity dispersion reflects also this morphology and is not



**Figure 4.17:** Same as Fig. 4.11 for NGC 7582.

reproduced in the model. The PVD reveals a low-velocity gradient in the SoI and a low SMBH mass estimation.

## 4.7 Discussion

To provide more robust estimations of  $M_{\text{BH}}$ , there are three ways of improvement: the input data, the numerical simulations, and the ML method itself.

First of all, the model precision is limited by the observation’s spatial resolution. Figure 4.8 demonstrates that NN face challenges in providing accurate estimations when the  $M_{\text{BH}}$  is below approximately 6 dex. This issue arises because, as  $M_{\text{BH}}$  decreases, fewer pixels in the data carry information about it. Consequently, the NN struggles to extract meaningful insights, leading to increased variability in estimations for smaller  $M_{\text{BH}}$  values. Therefore, observations with higher resolutions provide more pixels with relevant  $M_{\text{BH}}$  information, enhancing the NN precision and reducing the sensitivity to small  $M_{\text{BH}}$  variations. The role of the SoI in the context of NN remains somewhat unclear. While the goal is to increase the number of informative pixels about the BH, these pixels don’t need to be situated where the BH dominantly influences gas dynamics. Rather, these pixels need only exhibit a noticeable effect on the NN, even if it’s relatively small.

Another limiting factor comes from the physics included in the numerical simulations used to make the datasets. I use numerical simulations to mimic real observations but, as described in Sect. 3.5.1, I simulated an axisymmetric and homogeneous thin gas disk, with no clumpy structure, nor spiral or bar structures. Furthermore, I have adopted in my models the same  $IA$  and  $PA$  for the gas and stellar disks, while some misalignments of the nuclear molecular disks have been observed with ALMA (Combes et al. 2019). Adding these misalignments and more complex physics, like the presence of a bar, in the simulations would add free parameters to influence directly the  $M_{\text{BH}}$  estimations. This discrepancy in realism can result in the network exhibiting excessive confidence in its predictions, while there is a strong likelihood that the uncertainty in the actual data is being underestimated.

Additionally, as previously elaborated in Section 4.5.2, the probability distributions illustrated in Fig.4.9 lack physical significance. This discrepancy arises from the inherent differences in the underlying physics between the numerical simulations and the actual observational

data. However, by applying a reverse perspective, as the physical fidelity of the simulations approaches that of the observations, the probability distributions of  $M_{\text{BH}}$  become more representative of the actual physical uncertainties. Therefore, enhancing the realism of the numerical simulations serves a dual purpose, it not only improves the accuracy of error estimates but also confers them with a deeper and more grounded physical significance.

The ML method in itself can also be improved. Here, I chose to work with one galaxy at a time which means creating a different training sample and training a different model for each galaxy. This approach has the advantage of simplifying the training set creation and the learning process but it is also very inefficient time-wise and this approach is effective solely due to the limited number of galaxies available in my dataset. It take  $\sim 8$  hours to create a dataset and  $\sim 4$  hours to train one NN. By creating a training sample with the moment maps of multiple galaxies, I could generalize the method to make it work for any galaxy. However, the limited sample of galaxies available allowed us to center the moment maps and PVD around the BH but this will become impossible when working on too many galaxies. Hence, my goal is to design a neural network capable of accurately determining the  $M_{\text{BH}}$ , even when it is not centrally located within the image. To achieve this, one approach would be to develop a neural network that performs BH localization, in the data prior to conducting any mass estimations.

To effectively learn and predict additional physical parameters and/or accurately detect the position of BHs, it is necessary to make modifications to the neural network architecture to address these challenges. Moreover, a larger and more diverse training sample will be required. When aiming to predict multiple parameters, it is essential to ensure that the training sample is sufficiently extensive to comprehensively map the entire feature space. Furthermore, to successfully learn how to localize the BH in the input data, the positional information of the BH must be known for each example. However, the specific size and diversity requirements of the sample are empirical and contingent upon the particularities of the problem at hand. At the very least, several hundred distinct galaxy molecular gas observations are anticipated to be necessary.

Then I can choose to train my NN on the observations or on simulations made from the observations. Training the NN on the observations has the advantage that I do not have problems with the realism of the physics used in the numerical simulations. However, it requires to have  $M_{\text{BH}}$  estimation for all the observed galaxies. In addition, it's important to note that a single galaxy may yield multiple  $M_{\text{BH}}$  estimations derived from various methods, often resulting in discrepancies between these estimates. This situation highlights the necessity for methodological consistency in estimating  $M_{\text{BH}}$  across all galaxies in the sample to avoid potential confusion for the NN. However, this uniformity in estimation methods also implies that the NN is not learning to estimate  $M_{\text{BH}}$  per se but is instead learning to estimate  $M_{\text{BH}}$  based on the predetermined method. Consequently, if the selected method contains inherent flaws or inaccuracies, the NN is likely to incorporate and replicate these shortcomings in its estimations. Working with numerical simulations does have its disadvantages, particularly concerning the intricacies of simulating the physics, as elaborated earlier. However, it also presents the advantage of reducing the volume of observations necessary for the method generalization, as it allows for the creation of multiple simulations per galaxy.

Nevertheless, obtaining the exact position of the BH may not be a prerequisite; instead, it may suffice to ensure that the BH is contained within the image. This approach offers the advantage of simplifying the acquisition of the required sample. However, it is important to

note that training the model using this method may be harder.

## 4.8 Conclusion

The ALMA high sensitivity and spatial resolution allow us to resolve the CO(3-2) molecular disks at the center of nearby galaxies. In the case of the GATOS core sample used in this work, the distances and ALMA angular resolutions allow to resolution of typical physical scales of 10 pc, which probe their SoI. Since the nuclear disks entered the sphere of influence of the BH, I was able to make  $M_{\text{BH}}$  estimations. In this article, I provide a proof of concept for a novel approach to estimate the mass of SMBH using ALMA observations of the circum-nuclear disk. My method employs supervised ML techniques on data obtained from numerical simulations.

Gas dynamics can be studied by looking at the PVD and/or the moments map calculated from a spectral data cube. I developed two artificial neural network architectures. The first one takes the PVD as input, the other one takes the first moment maps, and both of them give an estimation of the  $M_{\text{BH}}$  based on the input. To test the reliability of my models, I used three statistical indicators: the RMSE, the  $R^2$ , and the  $\overline{RE}$ . These statistical indicators confirmed that my fourteen models can make reliable  $M_{\text{BH}}$  estimations on new data. I looked at my models RE distribution when predicting the  $M_{\text{BH}}$  of the test sample. I saw that the distributions were symmetrical and centered around 0, indicating no constant over/underestimation. I also studied the models RE of their  $M_{\text{BH}}$  estimations made on the test set as a function of the simulation parameters values. The resulting distributions are all uniform and center around zero except for the  $M_{\text{BH}}$ . This is a strong indication that the models  $M_{\text{BH}}$  predictions only depend on the  $M_{\text{BH}}$ . I studied more in-depth the behavior of the  $M_{\text{BH}}$  predictions by looking at the predictions RE versus the  $M_{\text{BH}}$ .

I managed to make a  $M_{\text{BH}}$  estimation with both architectures for 7 out of the 10 galaxies in the GATOS core sample: NGC 4388, NGC 5506, NGC 5643, NGC 6300, NGC 7314, NGC 7465, and NGC 7582. I made  $10^5$  predictions per galaxy using MC dropout. The resulting  $M_{\text{BH}}$  estimations are Gaussian-like probability distributions. The scatter of these distributions represents the models confidence in their predictions. I observe that the PVD models give larger scatters from 0.03 dex to 0.16 dex depending on the galaxy. On the other hand, the moment maps models have a constant scatter of  $\sim 0.02$  dex for all galaxies. The narrow error bars observed in my results are likely indicative of the network's overconfidence, which can be attributed to its training on numerical simulations that encompass simplified physics compared to real-world conditions. Nonetheless, all the estimations I made are consistent with the previous studies except for NGC 7582 where my estimate is between the previous ones.

In this study, I showed that even with training sets simulated with simple physics, with a fixed  $IA$  and  $PA$ , and with simple neural network architectures, my approach can produce results coherent with the literature. This work represents the first step toward an automatized method for estimating  $M_{\text{BH}}$ .



---

## BIBLIOGRAPHY

---

- Aalto, S., Martín, S., Costagliola, F., et al. 2015, *A&A*, 584, A42 [15]
- Abbott, T. M. C., Adamów, M., Agüena, M., et al. 2021, *ApJs*, 255, 20 [31]
- Allen, S. W., Dunn, R. J. H., Fabian, A. C., Taylor, G. B., & Reynolds, C. S. 2006, *MNRAS*, 372, 21 [43]
- Alonso-Herrero, A., García-Burillo, S., Hönic, S. F., et al. 2021, *A&A*, 652, A99 [15 and 16]
- Alonso-Herrero, A., García-Burillo, S., Pereira-Santaella, M., et al. 2019, *A&A*, 628, A65 [88]
- Alonso-Herrero, A., Pereira-Santaella, M., García-Burillo, S., et al. 2018, *ApJ*, 859, 144 [112]
- Antonucci, R. R. J. & Miller, J. S. 1985, *ApJ*, 297, 621 [7 and 16]
- Arnouts, S., Cristiani, S., Moscardini, L., et al. 1999, *MNRAS*, 310, 540 [38]
- Asmus, D., Hönic, S. F., & Gandhi, P. 2016, *ApJ*, 822, 109 [16]
- Audibert, A., Combes, F., García-Burillo, S., et al. 2019, *A&A*, 632, A33 [15 and 88]
- Bahcall, J. N., Kozlovsky, B.-Z., & Salpeter, E. E. 1972, *ApJ*, 171, 467 [26]
- Baldry, I. K., Liske, J., Brown, M. J. I., et al. 2018, *MNRAS*, 474, 3875 [34]
- Baldwin, J. A., Phillips, M. M., & Terlevich, R. 1981, *PASP*, 93, 5 [12]
- Balmaverde, B., Baldi, R. D., & Capetti, A. 2008, *A&A*, 486, 119 [43, 47, and 49]
- Bañados, E., Venemans, B. P., Mazzucchelli, C., et al. 2018, *Nature*, 553, 473 [24 and 30]
- Bartels, R., Zandanel, F., & Ando, S. 2015, *A&A*, 582, A20 [10]
- Barth, A. J., Sarzi, M., Rix, H.-W., et al. 2001, *ApJ*, 555, 685 [85]

- Bassani, L., Dadina, M., Maiolino, R., et al. 1999, *ApJs*, 121, 473 [7]
- Baugh, C. M. 2006, *Rep. Prog. Phys.*, 69, 3101 [16 and 18]
- Becker, R. H., White, R. L., & Helfand, D. J. 1995, *ApJ*, 450, 559 [31]
- Beckmann, V. & Shrader, C. R. 2012, *Active Galactic Nuclei* (Wiley-VCH GmbH) [4]
- Beckmann, V., Soldi, S., Ricci, C., et al. 2009, *A&A*, Volume 505, Issue 1, 2009, pp.417-439, 505, 417 [110]
- Behroozi, P. S., Wechsler, R. H., & Conroy, C. 2013, *ApJ*, 770, 57 [17]
- Bender, R., Kormendy, J., Bower, G., et al. 2005, *ApJ*, 631, 280 [84]
- Bentz, M. C., Walsh, J. L., Barth, A. J., et al. 2009, *ApJ*, 705, 199 [27]
- Best, P. N. & Heckman, T. M. 2012, *MNRAS*, 421, 1569 [12 and 13]
- Best, P. N., Kauffmann, G., Heckman, T. M., et al. 2005, *MNRAS*, 362, 25 [31, 40, and 52]
- Bian, W. & Gu, Q. 2007, *ApJ*, 657, 159 [109 and 110]
- Bickley, R. W., Bottrell, C., Hani, M. H., et al. 2021, , 504, 372 [81]
- Bini, S., Vedovato, G., Drago, M., Salemi, F., & Prodi, G. A. 2023, *Classical and Quantum Gravity*, 40, 135008 [81]
- Binney, J. & Tremaine, S. 2008, *Galactic Dynamics: Second Edition* [89]
- Bishop, C. M. 2007, *Pattern Recognition and Machine Learning* (Information Science and Statistics), 1st edn. (Springer) [55]
- Blandford, R. D. 1984, *Annals of the New York Academy of Sciences*, 422, 303 [28]
- Blandford, R. D. & Payne, D. G. 1982, *MNRAS*, 199, 883 [13 and 26]
- Blandford, R. D. & Znajek, R. L. 1977, *MNRAS*, 179, 433 [13]
- Blundell, C., Cornebise, J., Kavukcuoglu, K., & Wierstra, D. 2015, *Weight Uncertainty in Neural Networks*, arXiv:1505.05424 [cs, stat] [68]
- Blustin, A. J., Page, M. J., Fuerst, S. V., Branduardi-Raymont, G., & Ashton, C. E. 2005, *A&A*, 431, 111 [10]
- Bock, J. J., Neugebauer, G., Matthews, K., et al. 2000, *AJ*, 120, 2904 [8]
- Bolton, A. S., Schlegel, D. J., Aubourg, É., et al. 2012, *AJ*, 144, 144 [33]
- Bonaldi, A., An, T., Brüggén, M., et al. 2021, *MNRAS*, 500, 3821 [81]
- Bonoli, S., Mayer, L., & Callegari, S. 2014, *MNRAS*, 437, 1576 [24 and 30]

- Boroson, T. A. & Green, R. F. 1992, *ApJs*, 80, 109 [37]
- Brandt, W. N. & Alexander, D. M. 2015, *A&ARv*, 23, 1 [8]
- Brightman, M., Bachetti, M., Earnshaw, H. P., et al. 2019, *AAS*, 51, 352 [12]
- Brown, J. S., Valluri, M., Shen, J., & Debattista, V. P. 2013, *ApJ*, 778, 151 [24]
- Bruzual, G. & Charlot, S. 2003, *MNRAS*, 344, 1000 [43]
- Buttiglione, S., Capetti, A., Celotti, A., et al. 2010, *A&A*, 509, A6 [13]
- Cappellari, M., Emsellem, E., Krajnović, D., et al. 2011, *MNRAS*, 416, 1680 [3]
- Castignani, G., Chiaberge, M., Celotti, A., Norman, C., & De Zotti, G. 2014, *ApJ*, 792, 114 [35 and 51]
- Castignani, G., Combes, F., Salomé, P., et al. 2019, *A&A*, 623, A48 [31 and 38]
- Castignani, G., Haardt, F., Lapi, A., et al. 2013, *A&A*, 560, A28 [32, 36, 37, 38, and 49]
- Castro-Ginard, A., Jordi, C., Luri, X., Cantat-Gaudin, T., & Balaguer-Núñez, L. 2019, *A&A*, 627, A35 [88]
- Celotti, A., Padovani, P., & Ghisellini, G. 1997, *MNRAS*, 286, 415 [37]
- Chabrier, G. 2003, *PASP*, 115, 763 [38]
- Chary, R. & Elbaz, D. 2001, *ApJ*, 556, 562 [38]
- Chen, Y.-Y., Zhang, X., Xiong, D., & Yu, X. 2015, *AJ*, 150, 8 [47, 48, and 49]
- Cheng, T.-Y., Conselice, C. J., Aragón-Salamanca, A., et al. 2021, *MNRAS*, 507, 4425 [81]
- Cheng, T.-Y., Conselice, C. J., Aragón-Salamanca, A., et al. 2020, *MNRAS*, 493, 4209 [81]
- Cherepashchuk, A. M. & Lyutyi, V. M. 1973, *ApLett*, 13, 165 [26]
- Chiaberge, M. & Marconi, A. 2011, *MNRAS*, 416, 917 [31 and 52]
- Chiaberge, M., Tremblay, G., Capetti, A., et al. 2009, *ApJ*, 696, 1103 [35]
- Chollet, F. et al. 2015, *Keras*, <https://keras.io> [101]
- Cisternas, M., Jahnke, K., Bongiorno, A., et al. 2011, *ApJ*, 741, L11 [42]
- Coleman, G. D., Wu, C. C., & Weedman, D. W. 1980, *ApJs*, 43, 393 [38]
- Collaboration, T. D. E. S. 2005, *arXiv:astro-ph/0510346* [31]
- Combes, F., García-Burillo, S., Audibert, A., et al. 2019, *A&A*, 623, A79 [x, 15, 25, 85, 86, 87, 88, 89, 90, and 115]
- Condon, J. J. 1989, *ApJ*, 338, 13 [35]

- Contopoulos, I., Christodoulou, D. M., Kazanas, D., & Gabuzda, D. C. 2009, *ApJ*, 702, L148 [13]
- Corbin, M. R. 1992, *ApJ*, 391, 577 [37]
- Costa, T., Sijacki, D., & Haehnelt, M. G. 2014, *MNRAS*, 444, 2355 [21, 22, and 26]
- Crenshaw, D. M., Kraemer, S. B., & George, I. M. 2003, *A&A*, 41, 117 [16]
- Croom, S. M., Richards, G. T., Shanks, T., et al. 2009, *MNRAS*, 399, 1755 [35]
- Crummy, J., Fabian, A. C., Gallo, L., & Ross, R. R. 2006, *MNRAS*, 365, 1067 [10]
- Cybenko, G. 1989, *Math. Control Signal Systems*, 2, 303 [63]
- Czerny, B. & Elvis, M. 1987, *ApJ*, 321, 305 [10]
- da Cunha, E., Charlot, S., & Elbaz, D. 2008, *MNRAS*, 388, 1595 [38]
- Dabhade, P., Mahato, M., Bagchi, J., et al. 2020, *A&A*, 642, A153 [52]
- Dalya, G., Diaz, R., Bouchet, F. R., et al. 2021, *VizieR Online Data Catalog*, VII/291 [2 and 3]
- Damianou, A. C. & Lawrence, N. D. 2013, *arXiv e-prints* [68]
- Dark Energy Survey Collaboration, Abbott, T., Abdalla, F. B., et al. 2016, *MNRAS*, 460, 1270 [31]
- Dauser, T., García, J., Walton, D. J., et al. 2016, *A&A*, 590, A76 [11]
- Davies, R. L., Dopita, M. A., Kewley, L., et al. 2016, *ApJ*, 824, 50 [114]
- Davis, T. A., Bureau, M., Onishi, K., et al. 2017, *MNRAS*, 468, 4675 [85 and 87]
- Davis, T. A., Bureau, M., Onishi, K., et al. 2018, *MNRAS*, 473, 3818 [85 and 87]
- de Vaucouleurs, G. 1959, *Handbuch der Physik*, 53, 275 [3]
- Decarli, R., Falomo, R., Treves, A., et al. 2010, *MNRAS*, 402, 2441 [42]
- DeGraf, C., Di Matteo, T., Treu, T., et al. 2015, *MNRAS*, 454, 913 [44, 45, and 52]
- Delvecchio, I., Smolčić, V., Zamorani, G., et al. 2018, *MNRAS*, 481, 4971 [47]
- Done, C., Davis, S. W., Jin, C., Blaes, O., & Ward, M. 2012, *MNRAS*, 420, 1848 [10]
- Dressler, A. 1980, *ApJ*, 236, 351 [2]
- Driver, S. P., Andrews, S. K., da Cunha, E., et al. 2018, *MNRAS*, 475, 2891 [38]
- Dullo, B. T., Gil de Paz, A., & Knapen, J. H. 2021, *ApJ*, 908, 134 [24 and 30]
- Dumoulin, V. & Visin, F. 2018, A guide to convolution arithmetic for deep learning, *arXiv:1603.07285 [cs, stat]* [77]

- Durret, F., Adami, C., Bertin, E., et al. 2015, *A&A*, 578, A79 [51]
- Durret, F., Adami, C., Cappi, A., et al. 2011, *A&A*, 535, A65 [51]
- Eales, S., Dunne, L., Clements, D., et al. 2010, *PASP*, 122, 499 [38]
- Edge, A., Sutherland, W., Kuijken, K., et al. 2013, *The Messenger*, 154, 32 [38]
- Einstein, A. 1916, *Annalen der Physik*, 354, 769 [24]
- Eke, V. R., Baugh, C. M., Cole, S., Frenk, C. S., & Navarro, J. F. 2006, *MNRAS*, 370, 1147 [18]
- Euclid Collaboration, Adam, R., Vannier, M., et al. 2019, *A&A*, 627, A23 [53]
- Faber, S. M. & Jackson, R. E. 1976, *ApJ*, 204, 668 [22]
- Fabian, A. C. 2012, *ARA&A*, 50, 455 [19 and 47]
- Fabian, A. C. & Iwasawa, K. 1999, *MNRAS*, 303, L34 [11]
- Falcke, H., Koerding, E., & Markoff, S. 2004, *A&A*, 414, 895 [28]
- Falcke, H., Wilson, A. S., & Simpson, C. 1998, *ApJ*, 502, 199 [111]
- Falcón-Barroso, J., van de Ven, G., Peletier, R. F., et al. 2011, *MNRAS*, 417, 1787 [22]
- Fanaroff, B. L. & Riley, J. M. 1974, *MNRAS*, 167, 31P [35]
- Farina, E. P., Schindler, J.-T., Walter, F., et al. 2022, arXiv e-prints, arXiv:2207.05113 [24 and 30]
- Fernandes, T., Vieira, S., Onofre, A., et al. 2023, *Classical and Quantum Gravity*, 40, 195018 [81]
- Ferrara, A., Haardt, F., & Salvaterra, R. 2013, *MNRAS*, 434, 2600 [24 and 30]
- Ferrarese, L. & Ford, H. 2005, *SSRv*, 116, 523 [25]
- Ferrarese, L. & Merritt, D. 2000, *ApJ*, 539, L9 [22, 25, and 26]
- Ferré-Mateu, A., Mezcua, M., Trujillo, I., Balcells, M., & van den Bosch, R. C. E. 2015, *ApJ*, 808, 79 [24 and 30]
- Ferré-Mateu, A., Trujillo, I., Martín-Navarro, I., et al. 2017, *MNRAS*, 467, 1929 [24 and 30]
- Ford, H. C., Harms, R. J., Tsvetanov, Z. I., et al. 1994, *ApJL*, 435, L27 [84]
- Francis, P. J., Hewett, P. C., Foltz, C. B., et al. 1991, *ApJ*, 373, 465 [37]
- Fukue, J. 2019, *PASJ*, 71, 99 [21]
- Gabor, J. M. & Bournaud, F. 2014, *MNRAS*, 441, 1615 [19]

- Gal, Y. & Ghahramani, Z. 2016, Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning, arXiv:1506.02142 [cs, stat] [68]
- Galametz, A., Stern, D., De Breuck, C., et al. 2012, ApJ, 749, 169 [51]
- Gallo, E., Fender, R. P., & Pooley, G. G. 2003, MNRAS, 344, 60 [27]
- García, J. & Kallman, T. R. 2010, ApJ, 718, 695 [11]
- García-Bernete, I., Alonso-Herrero, A., García-Burillo, S., et al. 2021, A&A, 645, A21 [85 and 112]
- García-Burillo, S., Alonso-Herrero, A., Ramos Almeida, C., et al. 2021, A&A, 652, A98 [15, 85, 88, 98, 99, 111, and 112]
- García-Burillo, S., Combes, F., Ramos Almeida, C., et al. 2019, A&A, 632, A61 [15, 85, and 88]
- Gaskell, C. M. 1988, ApJ, 325, 114 [26]
- Gaskell, C. M. 2008, in Revista Mexicana de Astronomia y Astrofisica Conference Series, Vol. 32, Revista Mexicana de Astronomia y Astrofisica Conference Series, 1–11 [26]
- Gaskell, C. M. 2011, Baltic Astronomy, 20, 392 [26]
- Gebhardt, K., Bender, R., Bower, G., et al. 2000, ApJ, 539, L13 [22, 25, and 26]
- Gebhardt, K., Richstone, D., Tremaine, S., et al. 2003, ApJ, 583, 92 [25]
- George, D., Shen, H., & Huerta, E. A. 2018, PhRvD, 97, 101501 [81]
- Ghisellini, G., Della Ceca, R., Volonteri, M., et al. 2010, MNRAS, 405, 387 [49]
- Ghisellini, G., Tavecchio, F., Maraschi, L., Celotti, A., & Sbarrato, T. 2014, nat, 515, 376 [49]
- Gierliński, M. & Done, C. 2004, MNRAS, 349, L7 [10]
- Gillet, N., Mesinger, A., Greig, B., Liu, A., & Ucci, G. 2019, MNRAS, 484, 282 [81]
- Glorot, X. & Bengio, Y. 2010, in Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics (JMLR Workshop and Conference Proceedings), 249–256, iSSN: 1938-7228 [66 and 77]
- Gludemans, A. J., Duncan, K. J., Röttgering, H. J. A., et al. 2021, A&A, 656, A137 [52]
- Golden-Marx, E., Blanton, E. L., Paterno-Mahler, R., et al. 2019, ApJ, 887, 50 [51]
- Goodfellow, I., Bengio, Y., & Courville, A. 2016, Deep Learning (MIT Press), book in preparation for MIT Press [55]
- Goto, T., Sekiguchi, M., Nichol, R. C., et al. 2002, AJ, 123, 1807 [51]
- Graham, A. W. 2008, ApJ, 680, 143 [24]

- Graham, A. W. 2019, *MNRAS*, 487, 4995 [3]
- Graham, M. J., Ross, N. P., Stern, D., et al. 2020, *MNRAS*, 491, 4925 [8]
- Granato, G. L., De Zotti, G., Silva, L., Bressan, A., & Danese, L. 2004, *ApJ*, 600, 580 [26]
- Gratadour, D., Rouan, D., Grosset, L., Boccaletti, A., & Clénet, Y. 2015, *A&A*, 581, L8 [20]
- Guiglion, G., Matijević, G., Queiroz, A. B. A., et al. 2020, *A&A*, 644, A168 [81]
- Gültekin, K., King, A. L., Cackett, E. M., et al. 2019, *ApJ*, 871, 80 [28]
- Gültekin, K., Richstone, D. O., Gebhardt, K., et al. 2009, *ApJ*, 695, 1577 [22]
- Hao, J., McKay, T. A., Koester, B. P., et al. 2010, *ApJs*, 191, 254 [51]
- Häring, N. & Rix, H.-W. 2004, *ApJ*, 604, L89 [42, 44, 45, and 52]
- Harms, R. J., Ford, H. C., Tsvetanov, Z. I., et al. 1994, *ApJL*, 435, L35 [84]
- Harrison, C. M. 2017, *Nature Astronomy*, 1, 0165 [19]
- Hartley, P., Bonaldi, A., Braun, R., et al. 2023, *MNRAS*, 523, 1967 [81 and 99]
- Hartley, W. G., Choi, A., Amon, A., et al. 2021, *MNRAS*, 509, 3547 [31]
- Hassan, S., Andrianomena, S., & Doughty, C. 2020, *MNRAS*, 494, 5761 [81]
- Hastie, T., Tibshirani, R., & Friedman, J. 2001, *The Elements of Statistical Learning*, Springer Series in Statistics (New York, NY, USA: Springer New York Inc.) [55]
- Heckman, T. M. & Best, P. N. 2014, *ARA&A*, 52, 589 [11]
- Hon, M., Stello, D., & Yu, J. 2017, *MNRAS*, 469, 4578 [81]
- Hönig, S. F. & Kishimoto, M. 2017, *ApJ*, 838, L20 [8]
- Hubble, E. P. 1936, *Realm of the Nebulae* [2]
- Huertas-Company, M., Guo, Y., Ginzburg, O., et al. 2020, *MNRAS*, 499, 814 [87]
- Ilbert, O., Arnouts, S., McCracken, H. J., et al. 2006, *A&A*, 457, 841 [38]
- Ilbert, O., Salvato, M., Le Floch, E., et al. 2010, *ApJ*, 709, 644 [44]
- Imanishi, M., Nakanishi, K., Izumi, T., & Wada, K. 2018, *ApJL*, 853, L25 [15]
- Inoue, Y., Doi, A., Tanaka, Y. T., Sikora, M., & Madejski, G. M. 2017, *ApJ*, 840, 46 [49]
- Ioffe, S. & Szegedy, C. 2015a, arXiv e-prints, arXiv:1502.03167 [64]
- Ioffe, S. & Szegedy, C. 2015b, arXiv e-prints, arXiv:1502.03167 [80]
- Jaffe, W., Meisenheimer, K., Röttgering, H. J. A., et al. 2004, *Nature*, 429, 47 [14]

- Jahnke, K., Bongiorno, A., Brusa, M., et al. 2009, *ApJ*, 706, L215 [42]
- Jarrett, T. H., Cluver, M. E., Magoulas, C., et al. 2017, *ApJ*, 836, 182 [33 and 35]
- Jiang, L., Fan, X., Hines, D. C., et al. 2006, *AJ*, 132, 2127 [43]
- Kaspi, S., Brandt, W. N., Maoz, D., et al. 2007, *ApJ*, 659, 997 [27]
- Kauffmann, G., Heckman, T. M., Tremonti, C., et al. 2003, *MNRAS*, 346, 1055 [8]
- Kennicutt, Robert C., J. 1998, *ARA&A*, 36, 189 [38]
- King, A. 2003, *ApJ*, 596, L27 [21]
- Kingma, D. P. & Ba, J. 2014, arXiv e-prints, arXiv:1412.6980 [102]
- Koester, B. P., McKay, T. A., Annis, J., et al. 2007, *ApJ*, 660, 239 [51]
- Kollatschny, W. & Zetzl, M. 2011, *Nature*, 470, 366 [26]
- Kong, M. & Ho, L. C. 2018, *ApJ*, 859, 116 [38]
- Kormendy, J. 2004, in *Coevolution of Black Holes and Galaxies*, ed. L. C. Ho, 1 [25]
- Kormendy, J. & Bender, R. 1996, *ApJ*, 464, L119 [3]
- Kormendy, J. & Bender, R. 2012, *ApJS*, 198, 2 [3]
- Kormendy, J., Bender, R., Magorrian, J., et al. 1997, *ApJL*, 482, L139 [24 and 30]
- Kormendy, J. & Ho, L. C. 2013, *ARA&A*, 51, 511 [11, 25, 26, 47, and 84]
- Kormendy, J. & Richstone, D. 1995, *ARA&A*, 33, 581 [25]
- KrAInović, D., Cappellari, M., McDermid, R. M., et al. 2018, *MNRAS*, 477, 3030 [84]
- Krolik, J. H. & Begelman, M. C. 1988, *ApJ*, 329, 702 [14]
- Kudoh, Y., Wada, K., Kawakatu, N., & Nomura, M. 2023, *ApJ*, 950, 72 [14]
- Lacy, M., Ridgway, S. E., Sajina, A., et al. 2015, *ApJ*, 802, 102 [8]
- Lanusse, F., Ma, Q., Li, N., et al. 2018, *MNRAS*, 473, 3895 [88]
- Lanzuisi, G., Perna, M., Comastri, A., et al. 2016, *A&A*, 590, A77 [11]
- Le, T., Newman, W., & Edge, B. 2018, *MNRAS*, 477, 1803 [43]
- LeCun, Y., Bengio, Y., & Hinton, G. 2015, *Nature*, 521, 436 [101]
- LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. 1998, *Proceedings of the Institute of Radio Engineers*, 86, 2278 [77]

- Lecun, Y., Jackel, L., Bottou, L., et al. 1995, Learning algorithms for classification: A comparison on handwritten digit recognition (World Scientific), 261–276 [74]
- Ledlow, M. J. & Owen, F. N. 1996, *AJ*, 112, 9 [49]
- Lee, S.-K., Im, M., Kim, J.-W., et al. 2015, *ApJ*, 810, 90 [51]
- Leftley, J. H., Hönic, S. F., Asmus, D., et al. 2019, *ApJ*, 886, 55 [19]
- Leftley, J. H., Tristram, K. R. W., Hönic, S. F., et al. 2018, *ApJ*, 862, 17 [14]
- Li, J., Silverman, J. D., Izumi, T., et al. 2022, *ApJL*, 931, L11 [52]
- Lidman, C., Suherli, J., Muzzin, A., et al. 2012, *MNRAS*, 427, 550 [44]
- Liu, H. T., Feng, H. C., & Bai, J. M. 2017, *MNRAS*, 466, 3323 [27]
- Liu, W., Zhu, M., Dai, C., et al. 2019, *MNRAS*, 483, 4774 [81]
- Liu, Y., Jiang, D. R., & Gu, M. F. 2006, *ApJ*, 637, 669 [43, 47, 49, and 50]
- Lochner, M., McEwen, J. D., Peiris, H. V., Lahav, O., & Winter, M. K. 2016, *ApJs*, 225, 31 [88]
- López-Gonzaga, N., Burtscher, L., Tristram, K. R. W., Meisenheimer, K., & Schartmann, M. 2016, *A&A*, 591, A47 [14]
- Loveday, J. 1996, *MNRAS*, 278, 1025 [2 and 3]
- Lutz, D., Maiolino, R., Moorwood, A. F. M., et al. 2002, *A&A*, 396, 439 [7]
- Lyutyi, V. M. & Cherepashchuk, A. M. 1972, *Astronomicheskij Tsirkulyar*, 688, 1 [26]
- Maas, A., Hannun, A., & Ng, A. 2013, Proceedings of the 30th International Conference on Machine Learning, 28,3 [75]
- Macchetto, F., Marconi, A., Axon, D. J., et al. 1997, *ApJ*, 489, 579 [85]
- Maciejewski, W. & Binney, J. 2001, *MNRAS*, 323, 831 [85]
- MacKay, D. J. C. 1992, *Neural Computation*, 4, 448 [68]
- Madau, P. & Dickinson, M. 2014, *ARA&A*, 52, 415 [22]
- Magorrian, J., Tremaine, S., Richstone, D., et al. 1998, *AJ*, 115, 2285 [22]
- Mahabal, A., Rebbapragada, U., Walters, R., et al. 2019, *PASP*, 131, 038002 [88]
- Makarov, D., Prugniel, P., Terekhova, N., Courtois, H., & Vauglin, I. 2014, *A&A*, 570, A13 [26]
- Malavasi, N., Bardelli, S., Ciliegi, P., et al. 2015, *A&A*, 576, A101 [51]

- Marconi, A., Capetti, A., Axon, D. J., et al. 2001, *ApJ*, 549, 915 [25 and 85]
- Marconi, A. & Hunt, L. K. 2003, *ApJL*, 589, L21 [25 and 85]
- Martin, D. C., Fanson, J., Schiminovich, D., et al. 2005, *ApJ*, 619, L1 [38]
- Martín-Navarro, I., La Barbera, F., Vazdekis, A., et al. 2015, *MNRAS*, 451, 1081 [24 and 30]
- Mayer, L., Fiacconi, D., Bonoli, S., et al. 2015, *ApJ*, 810, 51 [24 and 30]
- Mayer, L., Kazantzidis, S., Escala, A., & Callegari, S. 2010, *Nature*, 466, 1082 [24 and 30]
- McConnachie, A. W., Patton, D. R., Ellison, S. L., & Simard, L. 2009, *MNRAS*, 395, 255 [51]
- McConnell, N. J. & Ma, C.-P. 2013, *ApJ*, 764, 184 [22, 23, 24, 30, 42, 51, and 52]
- McCulloch, W. S. & Pitts, W. 1943, *Bulletin of Mathematical Biophysics*, 5, 115 [58]
- McElroy, R. E., Husemann, B., Croom, S. M., et al. 2016, *A&A*, 593, L8 [7]
- McLure, R. J. & Jarvis, M. J. 2002, *MNRAS*, 337, 109 [27]
- McLure, R. J., Jarvis, M. J., Targett, T. A., Dunlop, J. S., & Best, P. N. 2006, *MNRAS*, 368, 1395 [44 and 46]
- McNamara, B. R., Kazemzadeh, F., Rafferty, D. A., et al. 2009, *ApJ*, 698, 594 [20]
- Mehdipour, M., Kaastra, J. S., Kriss, G. A., et al. 2015, *A&A*, 575, A22 [10]
- Melchior, A. L. & Combes, F. 2011, *A&A*, 536, A52 [25, 89, and 90]
- Melvin, T., Masters, K., Lintott, C., et al. 2014, *MNRAS*, 438, 2882 [3]
- Mereghetti, S., Balman, S., Caballero-Garcia, M., et al. 2021, *Experimental Astronomy*, 52, 309 [8]
- Meriot, R. & Semelin, B. 2023, arXiv e-prints, arXiv:2310.02684 [81]
- Merloni, A., Bongiorno, A., Bolzonella, M., et al. 2010, *ApJ*, 708, 137 [42]
- Merloni, A., Heinz, S., & di Matteo, T. 2003, *MNRAS*, 345, 1057 [27 and 28]
- Mickaelian, A. M. 2015, *IrJAA*, 2, 1 [5]
- Minezaki, T., Yoshii, Y., Kobayashi, Y., et al. 2004, *ApJL*, 600, L35 [14]
- Miyamoto, M. & Nagai, R. 1975, *Publications of the Astronomical Society of Japan*, 27, 533 [89, 90, and 96]
- Moravec, E., Gonzalez, A. H., Stern, D., et al. 2020, *ApJ*, 888, 74 [51]
- Morishita, T., Chiaberge, M., Hilbert, B., et al. 2022, *ApJ*, 931, 165 [52]
- Morrissey, P., Conrow, T., Barlow, T. A., et al. 2007, *ApJs*, 173, 682 [38]

- Müller, A. L. & Romero, G. E. 2020, *A&A*, 636, A92 [37]
- Nair, V. & Hinton, G. E. 2010, in Proceedings of the 27th International Conference on International Conference on Machine Learning, ICML'10 (Madison, WI, USA: Omnipress), 807–814 [75]
- Narayan, R. 2005, *Astrophysics and Space Science*, Volume 300, Issue 1-3, pp. 177-188, 300, 177 [27]
- Nardini, E., Fabian, A. C., Reis, R. C., & Walton, D. J. 2011, *MNRAS*, 410, 1251 [10]
- Nemmen, R. S., Georganopoulos, M., Guiriec, S., et al. 2012, *Science*, 338, 1445 [43]
- Nenkova, M., Ivezić, Ž., & Elitzur, M. 2002, *ApJL*, 570, L9 [8]
- Nenkova, M., Sirocky, M. M., Ivezić, Ž., & Elitzur, M. 2008, *ApJ*, 685, 147 [14]
- Noda, H., Makishima, K., Yamada, S., et al. 2011, *PASJ*, 63, S925 [10]
- Oguri, M., Lin, Y.-T., Lin, S.-C., et al. 2018, *pasj*, 70, S20 [51]
- Onishi, K., Iguchi, S., Davis, T. A., et al. 2017, *MNRAS*, 468, 4663 [x, 85, and 87]
- Padovani, P., Alexander, D. M., Assef, R. J., et al. 2017, *A&ARv*, 25, 2 [4]
- Pasquet, J., Bertin, E., Treyer, M., Arnouts, S., & Fouchez, D. 2019, *A&A*, 621, A26 [81 and 88]
- Pearson, W. J., Wang, L., Alpaslan, M., et al. 2019, *A&A*, 631, A51 [2]
- Peterson, B. M. 1997, *An Introduction to Active Galactic Nuclei (An Introduction to Active Galactic Nuclei)* [4]
- Peterson, B. M., Ferrarese, L., Gilbert, K. M., et al. 2004, *ApJ*, 613, 682 [27 and 110]
- Pezzulli, E., Volonteri, M., Schneider, R., & Valiante, R. 2017, *MNRAS*, 471, 589 [24 and 30]
- Piconcelli, E., Jimenez-Bailón, E., Guainazzi, M., et al. 2005, *A&A*, 432, 15 [10]
- Pier, E. A. & Krolik, J. H. 1992, *ApJ*, 401, 99 [14]
- Plummer, H. C. 1911, *MNRAS*, 71, 460 [90]
- Poitevineau, R., Castignani, G., & Combes, F. 2023, *A&A*, 672, A164 [29]
- Porquet, D., Reeves, J. N., O'Brien, P., & Brinkmann, W. 2004, *A&A*, 422, 85 [10]
- Prelogović, D., Mesinger, A., Murray, S., Fiameni, G., & Gillet, N. 2022, *MNRAS*, 509, 3852 [81]
- Punsly, B. 2005, *ApJ*, 623, L9 [43]
- Qian, L., Zhou, X., Li, Y., & Hu, Z. 2022, arXiv e-prints, arXiv:2205.11759 [64]

- Qian, N. 1999, *Neural Networks*, 12, 145 [67]
- Radovich, M., Puddu, E., Bellagamba, F., et al. 2017, *A&A*, 598, A107 [51]
- Rakshit, S., Stalin, C. S., & Kotilainen, J. 2020, *ApJs*, 249, 17 [52]
- Reeves, J. N. & Braito, V. 2019, *ApJ*, 884, 80 [21]
- Regan, J. A., Visbal, E., Wise, J. H., et al. 2017, *Nature Astronomy*, 1, 0075 [24 and 30]
- Ricci, C. 2011, PhD thesis, University of Geneva, Switzerland [9]
- Ricci, C. & Trakhtenbrot, B. 2022, arXiv e-prints, arXiv:2211.05132 [7]
- Rosen, S. R., Webb, N. A., Watson, M. G., et al. 2016, *A&A*, 590, A1 [35]
- Rosenblatt, F. 1958, *Psychological Review*, 65, 386, place: US Publisher: American Psychological Association [59 and 60]
- Ross, R. R. & Fabian, A. C. 2005, *MNRAS*, 358, 211 [11]
- Rumelhart, D. E., Hinton, G. E., & Williams, R. J. 1986, *Parallel Distributed Processing* [61 and 63]
- Rykoff, E. S., Koester, B. P., Rozo, E., et al. 2012, *ApJ*, 746, 178 [51]
- Saglia, R. P., Opitsch, M., Erwin, P., et al. 2016, *ApJ*, 818, 47 [84]
- Sani, E., Marconi, A., Hunt, L. K., & Risaliti, G. 2011, *MNRAS*, 413, 1479 [44, 45, and 52]
- Santurkar, S., Tsipras, D., Ilyas, A., & Madry, A. 2018, arXiv e-prints, arXiv:1805.11604 [64]
- Savorgnan, G. A. D. & Graham, A. W. 2016, *MNRAS*, 457, 320 [24 and 30]
- Sbarrato, T., Padovani, P., & Ghisellini, G. 2014, *MNRAS*, 445, 81 [49]
- Schartmann, M., Meisenheimer, K., Camenzind, M., et al. 2008, *A&A*, 482, 67 [8]
- Scharwächter, J., Dopita, M. A., Shastri, P., et al. 2016, in *Astrophysics and Space Science Proceedings*, Vol. 42, *The Universe of Digital Sky Surveys*, ed. N. R. Napolitano, G. Longo, M. Marconi, M. Paolillo, & E. Iodice, 263 [113]
- Schechter, P. 1976, *ApJ*, 203, 297 [17]
- Schramm, M. & Silverman, J. D. 2013, *ApJ*, 767, 13 [42]
- Schuldt, S., Suyu, S. H., Cañameras, R., et al. 2021, *A&A*, 651, A55 [81]
- Schurch, N. J., Done, C., & Proga, D. 2009, *ApJ*, 694, 1 [10]
- Seyfert, C. K. 1943, *ApJ*, 97, 28 [4]
- Shakura, N. I. & Sunyaev, R. A. 1973, *A&A*, 24, 337 [49]

- Shalev-Shwartz, S. & Ben-David, S. 2014, *Understanding Machine Learning - From Theory to Algorithms*. (Cambridge University Press), I–XVI, 1–397 [55]
- Sharma, K., Kembhavi, A., Kembhavi, A., et al. 2020, *MNRAS*, 491, 2280 [81]
- Shaw, M. S., Romani, R. W., Cotter, G., et al. 2012, *ApJ*, 748, 49 [31, 36, and 52]
- She, R., Ho, L. C., & Feng, H. 2017, *ApJ*, 835, 223 [109 and 110]
- Shen, Y. & Liu, X. 2012, *The Astrophysical Journal*, 753, 125, aDS Bibcode: 2012ApJ...753..125S [36]
- Shen, Y., Richards, G. T., Strauss, M. A., et al. 2011, *ApJs*, 194, 45 [27, 31, 36, and 52]
- Shields, G. A., Menezes, K. L., Massart, C. A., & Bout, P. V. 2006, *ApJ*, 641, 683 [43]
- Silk, J. & Rees, M. J. 1998, *A&A*, 331, L1 [21 and 26]
- Sądowski, A., Narayan, R., Penna, R., & Zhu, Y. 2013, *MNRAS*, 436, 3856 [20]
- Soifer, B. T., Sanders, D. B., Madore, B. F., et al. 1987, *ApJ*, 320, 238 [8]
- Soltan, A. 1982, *MNRAS*, 200, 115 [11]
- Speagle, J. S., Steinhardt, C. L., Capak, P. L., & Silverman, J. D. 2014, *ApJs*, 214, 15 [40 and 41]
- Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. 2014, *Journal of Machine Learning Research*, 15, 1929 [68 and 101]
- Steidel, C. C. & Sargent, W. L. W. 1991, *ApJ*, 382, 433 [37]
- Stern, D., Assef, R. J., Benford, D. J., et al. 2012, *ApJ*, 753, 30 [33]
- Suganuma, M., Yoshii, Y., Kobayashi, Y., et al. 2006, *ApJ*, 639, 46 [14]
- Tadhunter, C. 2016, *A&ARv*, 24, 10 [8]
- Targett, T. A., Dunlop, J. S., & McLure, R. J. 2012, *MNRAS*, 420, 3621 [43]
- Taylor, C. & Reynolds, C. S. 2018, *ApJ*, 855, 120 [11]
- The Event Horizon Telescope Collaboration, Akiyama, K., Alberdi, A., Alef, W., et al. 2022, *ApJL*, 930, L12 [4 and 24]
- The Event Horizon Telescope Collaboration, Akiyama, K., Alberdi, A., Alef, W., et al. 2019, *ApJL*, 875, L1 [4 and 24]
- Thean, A., Pedlar, A., Kukula, M. J., Baum, S. A., & O’Dea, C. P. 2000, *MNRAS*, 314, 573 [114]
- Tombesi, F., Cappi, M., Reeves, J. N., et al. 2010, *A&A*, 521, A57 [21]

- Toomre, A. 1964, *ApJ*, 139, 1217 [92 and 93]
- Tremaine, S., Gebhardt, K., Bender, R., et al. 2002, *ApJ*, 574, 740 [42]
- Tristram, K. R. W., Burtscher, L., Jaffe, W., et al. 2014, *A&A*, 563, A82 [8]
- Trujillo, I., Ferré-Mateu, A., Balcells, M., Vazdekis, A., & Sánchez-Blázquez, P. 2014, *ApJL*, 780, L20 [24, 30, 51, and 52]
- Trump, J. R., Impey, C. D., Taniguchi, Y., et al. 2009, *ApJ*, 706, 797 [42]
- Tsai, C.-W., Eisenhardt, P. R. M., Jun, H. D., et al. 2018, *ApJ*, 868, 15 [11]
- Turner, T. J. & Miller, L. 2009, *A&ARv*, 17, 47 [10]
- Ulrich, M.-H., Maraschi, L., & Urry, C. M. 1997, *ARA&A*, 35, 445 [7]
- Unger, S. W., Pedlar, A., Axon, D. J., et al. 1987, *MNRAS*, 228, 671 [16]
- Urry, C. M. & Padovani, P. 1995, *PASP*, 107, 803 [5 and 14]
- Uttley, P., McHardy, I. M., & Vaughan, S. 2005, *MNRAS*, 359, 345 [7]
- Valiante, E., Smith, M. W. L., Eales, S., et al. 2016a, *MNRAS*, 462, 3146 [38]
- Valiante, R., Schneider, R., Volonteri, M., & Omukai, K. 2016b, *MNRAS*, 457, 3356 [24 and 30]
- van den Bergh, S. 1976, *ApJ*, 206, 883 [3]
- van den Bosch, R. C. E. 2016, *ApJ*, 831, 134 [109 and 110]
- van den Bosch, R. C. E., Gebhardt, K., Gültekin, K., et al. 2012, *nat*, 491, 729 [24 and 30]
- van der Marel, R. P. & van den Bosch, F. C. 1998, *AJ*, 116, 2220 [85]
- Vanden Berk, D. E., Wilhite, B. C., Kron, R. G., et al. 2004, *ApJ*, 601, 692 [7]
- Veilleux, S., Goodrich, R. W., & Hill, G. J. 1997, *ApJ*, 477, 631 [7]
- Veilleux, S., Maiolino, R., Bolatto, A. D., & Aalto, S. 2020, *A&ARv*, 28, 2 [21]
- Venanzi, M., Hönig, S., & Williamson, D. 2020, *ApJ*, 900, 174 [16 and 19]
- Vestergaard, M. 2002, *ApJ*, 571, 733 [27]
- Vestergaard, M. & Peterson, B. M. 2006, *ApJ*, 641, 689 [27]
- Visbal, E., Haiman, Z., & Bryan, G. L. 2014, *MNRAS*, 445, 1056 [24 and 30]
- Vrtilek, J. M. & Carleton, N. P. 1985, *ApJ*, 294, 106 [5]
- Walmsley, M., Smith, L., Lintott, C., et al. 2020, *MNRAS*, 491, 1554 [81]

- Walton, D. J., Risaliti, G., Harrison, F. A., et al. 2014, *ApJ*, 788, 76 [10]
- Wandel, A., Peterson, B. M., & Malkan, M. A. 1999, *ApJ*, 526, 579 [27]
- Wang, J.-G., Dong, X.-B., Wang, T.-G., et al. 2009, *ApJ*, 707, 1334 [27]
- Wang, J.-M. & Zhang, E.-P. 2007, *ApJ*, 660, 1072 [109]
- Wang, R., Carilli, C. L., Beelen, A., et al. 2007, *AJ*, 134, 617 [110]
- Wang, R., Carilli, C. L., Neri, R., et al. 2010, *ApJ*, 714, 699 [43]
- Wang, R., Carilli, C. L., Wagg, J., et al. 2008, *ApJ*, 687, 848 [43]
- Weaver, T. A. 1976, *APJS*, 32, 233 [21]
- Wen, Z. L. & Han, J. L. 2015, *ApJ*, 807, 178 [51]
- Wen, Z. L., Han, J. L., & Liu, F. S. 2012, *ApJs*, 199, 34 [51]
- Wilkins, D. R. & Fabian, A. C. 2012, *MNRAS*, 424, 1284 [11]
- Willott, C. J., Rawlings, S., Blundell, K. M., & Lacy, M. 1999, *MNRAS*, 309, 1017 [36]
- Wills, B. J., Brotherton, M. S., Fang, D., Steidel, C. C., & Sargent, W. L. W. 1993, *ApJ*, 415, 563 [37]
- Wilson, D. R. & Martinez, T. R. 2003, *Neural Networks*, 16, 1429, place: Netherlands Publisher: Elsevier Science [66]
- Wright, E. L., Eisenhardt, P. R. M., Mainzer, A. K., et al. 2010, *AJ*, 140, 1868 [32 and 38]
- Wright, S. 1921, *Journal of agricultural research*, 20, 557 [102]
- Wu, Q. & Cao, X. 2008, *ApJ*, 687, 156 [48 and 49]
- Xiao, T., Barth, A. J., Greene, J. E., et al. 2011, *ApJ*, 739, 28 [27]
- Xiong, D. R. & Zhang, X. 2014, *MNRAS*, 441, 3375 [43, 47, 49, and 50]
- Yang, G., Pennington, J., Rao, V., Sohl-Dickstein, J., & Schoenholz, S. S. 2019, arXiv e-prints, arXiv:1902.08129 [64]
- York, D. G., Adelman, J., Anderson, Jr., J. E., et al. 2000, *AJ*, 120, 1579 [38]
- Yu, Q. & Tremaine, S. 2002, *MNRAS*, 335, 965 [11]
- Zheng, W., Kriss, G. A., Telfer, R. C., Grimes, J. P., & Davidsen, A. F. 1997, *ApJ*, 475, 469 [37]
- Zheng, Z.-P., Qiu, B., Luo, A. L., & Li, Y.-B. 2020, *PASP*, 132, 024504 [81]
- Zhou, X., Gong, Y., Meng, X.-M., et al. 2022, *MNARS*, 512, 4593 [81]
- Zirbel, E. L. 1996, *ApJ*, 473, 713 [35]
- Zuckerman, B. & Ball, J. A. 1974, *ApJ*, 190, 35 [5]



# APPENDIX *A*

---

## RLAGN

---

### **A.1 Tables of the parameters of our RLAGN sample**

I report below some tables summarizing the properties of the radio sources in our sample.

ID	RA	Dec.	$z$	$\log \frac{L_{1.4 \text{ GHz}}}{\text{erg s}^{-1} \text{ Hz}^{-1}}$	$\log(L_{\text{dust}}/L_{\odot})$	$\log(M_{\star}/M_{\odot})$	SFR	WISE class	Type	Name
	hh:mm:ss.ss	dd:mm:ss.ss					( $M_{\odot}/\text{yr}$ )			
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
RS 16	02:23:34.94	-06:37:22.99	1.217	$33.65 \pm 0.01$	11.48	10.40	32	AGN	Radio Source	SDSS J022334.90-063722.9
RS 49	02:16:40.74	-04:44:04.97	0.875	$33.51 \pm 0.01$	13.65	11.95	<4801	AGN	QSO	FBQS J0216-0444
RS 52	02:19:38.89	-05:18:05.28	2.207	$32.65 \pm 0.04$	12.34	11.73	<235	AGN	QSO	3XMM J021938.8-051805
RS 60	02:19:54.63	-05:49:22.32	0.322	$30.63 \pm 0.05$	9.08	10.15	0.13	AGN	QSO	SDSS J021954.62-054922.2
RS 62	02:22:47.91	-04:33:31.02	1.635	$32.60 \pm 0.02$	12.43	11.02	<289	AGN	QSO	XXL-N 027_020
RS 79	02:22:55.95	-05:18:15.85	1.756	$34.63 \pm 0.01$	12.32	11.48	<225	AGN	QSO	PKS 0220-055
RS 81	02:26:07.42	-05:32:09.52	0.779	$32.30 \pm 0.01$	11.53	10.76	36	Intermediate disk	X-ray source	XXL-N 062_013
RS 82	02:25:56.39	-05:34:51.44	2.879	$33.68 \pm 0.01$	12.75	11.25	<605	AGN	QSO	XXL-N 062_009
RS 83	02:25:05.12	-05:36:47.94	0.682	$33.60 \pm 0.01$	9.58	11.11	0.41	Intermediate disk	QSO	PMN J0225-0536
RS 104	02:27:12.99	-04:46:36.36	0.981	$32.32 \pm 0.01$	12.62	11.15	<448	AGN	X-ray source	SPIRE 13430
RS 113	02:45:31.51	-00:26:12.32	2.082	$32.83 \pm 0.02$	12.52	11.75	<356	AGN	QSO	2SLAQ J024531.53-002612.2
RS 151	02:27:40.56	-04:02:51.16	2.603	$32.42 \pm 0.07$	13.69	11.54	<5265	AGN	QSO	XXL-N 044_070
RS 159	02:29:15.79	-04:42:15.95	1.074	$34.12 \pm 0.01$	12.29	10.28	<21	AGN	QSO	3XMM J022915.7-044216
RS 177	02:51:56.32	+00:57:06.53	0.471	$31.80 \pm 0.01$	11.74	10.45	<59	AGN	QSO	LBQS 0249+0044
RS 190	02:51:15.50	+00:31:35.45	1.978	$33.93 \pm 0.01$	12.02	11.10	<114	AGN	QSO	WISEA J025115.50+003135.4
RS 195	02:47:06.66	+00:23:18.10	0.363	$30.85 \pm 0.03$	11.67	10.94	<50	AGN	QSO	FBQS J0247+0023
RS 197	02:46:16.61	+00:19:53.11	3.791	$33.46 \pm 0.02$	11.61	11.11	44	Starburst/LIRG	QSO	SDSS J024616.60+001953.6
RS 205	02:53:40.94	+00:11:10.04	1.683	$33.04 \pm 0.01$	12.73	11.23	<577	AGN	QSO	LBQS 0251-0001
RS 206	02:48:54.80	+00:10:53.84	1.145	$33.49 \pm 0.01$	12.92	11.05	<904	AGN	QSO	2SLAQ J024854.80+001053.9
RS 214	02:50:48.66	+00:02:07.46	0.766	$32.67 \pm 0.01$	11.93	10.43	<91	AGN	QSO	FBQS J0250+0002
RS 237	02:49:23.22	-00:54:38.04	0.953	$31.63 \pm 0.06$	11.69	10.27	<53	AGN	QSO	2SLAQ J024923.20-005437.7
G 55673	12:11:55.32	-00:20:19.38	0.436	$33.16 \pm 0.04$	11.55	11.19	23	Starburst/LIRG	G	SDSS J121155.31-002019.4
G 71277	12:16:12.27	00:04:17.91	0.316	$32.29 \pm 0.03$	10.01	11.23	0.49	Intermediate disk	G	SDSS J121612.26+000417.8

ID	RA	Dec.	$z$	$\log \frac{L_{1.4 \text{ GHz}}}{\text{erg s}^{-1} \text{ Hz}^{-1}}$	$\log(L_{\text{dust}}/L_{\odot})$	$\log(M_{\star}/M_{\odot})$	SFR	WISE class	Type	Name
	hh:mm:ss.ss	dd:mm:ss.ss					( $M_{\odot}/\text{yr}$ )			
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
G 165213	12:01:13.77	-02:42:41.34	0.307	$31.90 \pm 0.08$	11.64	11.27	34	AGN	QSO	SDSS J120113.76-024241.3
G 196970	08:53:52.21	-00:45:31.12	0.323	$31.99 \pm 0.06$	10.00	11.08	0.74	Intermediate disk	G	SDSS J085352.21-004531.1
G 208794	08:40:44.47	00:03:05.27	0.449	$31.84 \pm 0.10$	11.59	10.92	28	AGN	G	SDSS J084044.10+000307.2
G 249591	14:08:22.42	02:08:53.70	0.432	$32.45 \pm 0.02$	10.64	11.31	4.3	Intermediate disk	G	SDSS J140822.76+020853.1
G 251343	14:32:57.38	02:03:24.46	0.761	$32.66 \pm 0.01$	12.22	11.54	<1119	Starburst/LIRG	G	SDSS J143257.64+020321.3
G 298359	14:37:31.86	01:18:58.14	0.342	$34.06 \pm 0.01$	10.81	10.92	3.1	Intermediate disk	Radio Source	GAMA J143257.41+020329.1
G 372455	09:03:25.57	01:12:14.86	0.311	$32.68 \pm 0.01$	9.842	11.02	0.75	Intermediate disk	Radio Source	GAMA J090325.48+011214.1
G 537618	12:23:48.39	-00:52:50.43	0.490	$32.57 \pm 0.02$	12.06	10.79	43	Starburst/LIRG	Radio Source	SDSS J122347.89-005249.2
G 714133	14:25:33.03	01:07:37.73	0.556	$32.48 \pm 0.02$	12.35	10.80	<238	AGN	Radio Source	NVSS J142533+010739
G 714228	14:31:20.49	01:14:56.44	0.343	$33.10 \pm 0.01$	10.33	10.67	5.9	AGN	G	SDSS J143120.07+011459.2
G 720847	08:47:02.80	01:30:01.50	0.417	$32.87 \pm 0.01$	10.45	11.11	30	AGN	QSO	WISEA J084702.78+013001.5
G 721940	14:21:30.03	02:13:02.43	0.640	$32.67 \pm 0.01$	11.89	10.94	<75	Intermediate disk	G	SDSS J142130.60+021308.9
G 745066	14:51:22.48	-00:33:41.05	0.377	$32.06 \pm 0.06$	11.81	11.72	45	AGN	QSO	WISEA J145122.47-003341.0
G 746605	12:21:02.95	-00:07:33.74	0.364	$31.78 \pm 0.10$	12.08	10.70	<131	AGN	G	SDSS J122103.51-000749.1
G 748144	11:39:54.20	00:13:47.26	0.589	$32.37 \pm 0.03$	12.09	11.67	<109	AGN	Radio Source	GAMA J113952.95+001348.6
G 748815	14:25:45.91	00:22:42.73	0.326	$33.82 \pm 0.01$	11.31	10.87	14	AGN	QSO	WISEA J142545.90+002242.7
G 804203	09:20:53.32	00:03:53.94	0.506	$32.37 \pm 0.03$	11.93	10.79	41	Starburst/LIRG	G	SDSS J092053.32+000353.9
G 835899	08:42:16.99	01:09:17.87	0.762	$32.05 \pm 0.06$	12.19	11.83	<70	AGN	G	WISEA J084217.25+010834.6
G 887308	14:37:01.00	-01:03:49.03	0.547	$32.28 \pm 0.03$	10.47	11.55	2.4	Intermediate disk	G	SDSS J143702.15-010357.2

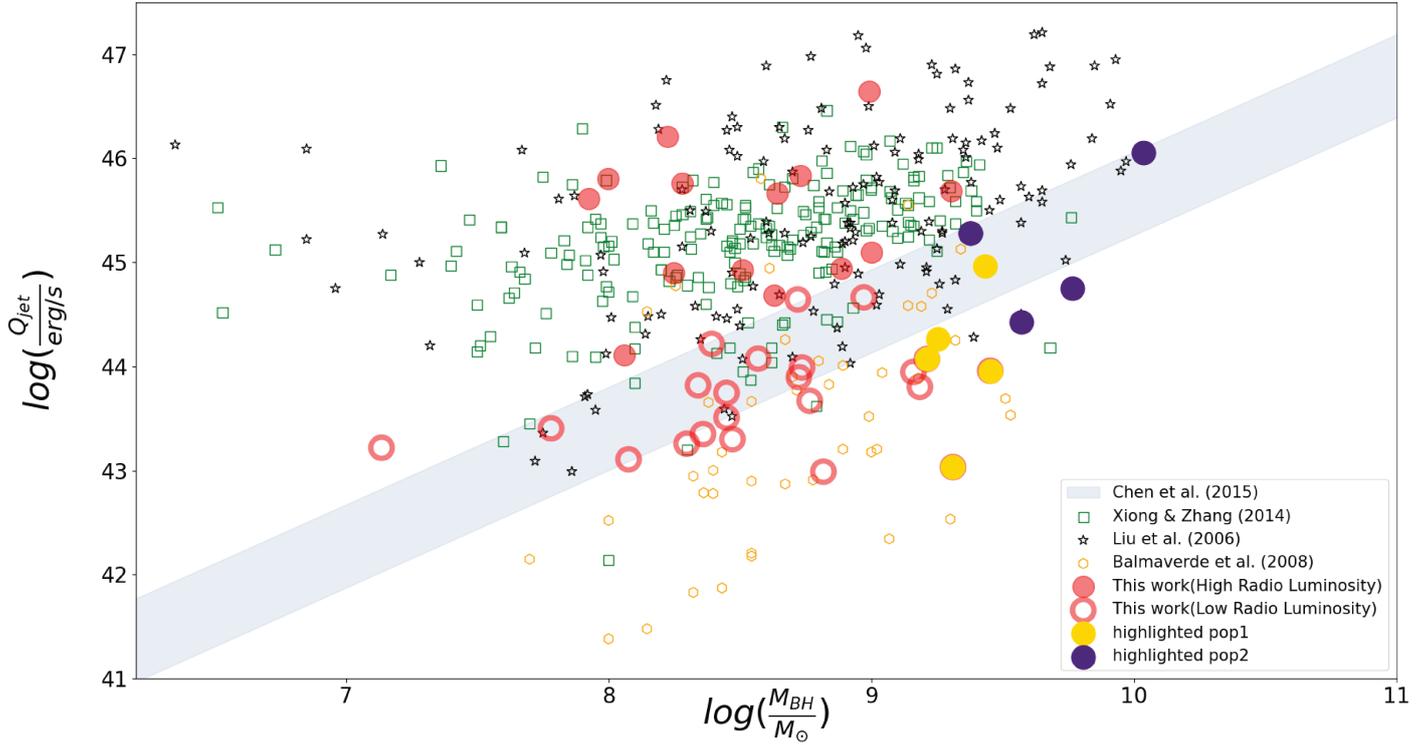
**Table A.1:** Properties of the radio sources in our sample. Column (1): Galaxy ID; (2-3) RA and Dec. coordinates; (4) Spectroscopic redshift; (5) 1.4 GHz rest-frame luminosity density; (6-7) SED-based dust luminosity and stellar mass; (8) SFR; (9) WISE color-based class; (10-11) source type and name as found in the NED.

ID	$z$	Line	FWHM ( $10^3 \text{ km s}^{-1}$ )	$\log(L_{\text{line}}/(\text{erg s}^{-1}))$	$\log(M_{\text{BH}}/M_{\odot})$	$\log(Q_{\text{jet}}/(\text{erg s}^{-1}))$	$\log(L_{\text{BLR}}/(\text{erg s}^{-1}))$	$\log \eta$
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
RS 16	1.22	Mg II	$4.08 \pm 0.42$	$42.53 \pm 0.07$	$8.00 \pm 0.12$	45.81	43.50	-1.60
RS 49	0.88	H $\beta$	$6.77 \pm 0.11$	$44.02 \pm 0.01$	$9.30 \pm 0.04$	45.69	45.06	-1.34
RS 52	2.21	Mg II	$6.00 \pm 0.43$	$43.41 \pm 0.14$	$8.88 \pm 0.13$	44.94	44.51	-1.48
RS 60	0.32	H $\alpha$	$1.74 \pm 0.07$	$41.88 \pm 0.02$	$7.13 \pm 0.04$	43.22	42.44	-1.80
RS 62	1.64	Mg II	$3.20 \pm 0.11$	$43.26 \pm 0.02$	$8.25 \pm 0.08$	44.90	44.27	-1.07
RS 79	1.76	Mg II	$6.48 \pm 0.16$	$43.47 \pm 0.03$	$8.99 \pm 0.08$	46.60	44.73	-1.30
RS 81	0.78	Mg II	$12.47 \pm 2.07$	$42.14 \pm 0.14$	$8.72 \pm 0.18$	44.60	43.30	-2.52
RS 82	2.88	C IV	$5.26 \pm 0.37$	$43.49 \pm 0.12$	$8.73 \pm 0.24$	45.83	44.28	-1.55
RS 83	0.68	H $\beta$	$7.03 \pm 1.67$	$41.87 \pm 0.14$	$8.28 \pm 0.23$	45.70	40.57	-4.80
RS 104	0.98	H $\beta$	$6.31 \pm 0.30$	$43.47 \pm 0.03$	$8.97 \pm 0.06$	44.60	44.56	-1.52
RS 113	2.08	Mg II	$5.89 \pm 0.24$	$43.62 \pm 0.06$	$9.00 \pm 0.09$	45.10	44.67	-1.43
RS 151	2.60	Mg II	$7.68 \pm 0.30$	$44.47 \pm 0.08$	$9.76 \pm 0.09$	44.75	45.13	-1.73
RS 159	1.07	Mg II	$3.95 \pm 0.21$	$42.94 \pm 0.03$	$8.22 \pm 0.10$	46.21	43.95	-1.37
RS 177	0.47	H $\beta$	$5.31 \pm 0.14$	$42.60 \pm 0.01$	$8.39 \pm 0.06$	44.22	43.65	-1.84
RS 190	1.98	C IV	$14.66 \pm 1.29$	$44.40 \pm 0.10$	$10.03 \pm 0.24$	46.05	45.05	-2.08
RS 195	0.36	H $\beta$	$2.90 \pm 0.03$	$42.42 \pm 0.01$	$7.78 \pm 0.06$	43.40	43.89	-0.99
RS 197	3.79	C IV	$1.93 \pm 0.39$	$43.64 \pm 0.09$	$7.92 \pm 0.28$	45.61	44.43	-0.50
RS 205	1.68	Mg II	$7.22 \pm 0.18$	$43.93 \pm 0.02$	$9.38 \pm 0.07$	45.28	45.19	-1.29
RS 206	1.15	Mg II	$4.12 \pm 0.11$	$43.54 \pm 0.01$	$8.64 \pm 0.07$	45.60	44.58	-1.10
RS 214	0.77	H $\beta$	$11.05 \pm 0.48$	$43.42 \pm 0.02$	$9.43 \pm 0.06$	44.96	44.36	-2.10
RS 237	0.95	Mg II	$9.89 \pm 0.99$	$43.24 \pm 0.07$	$9.21 \pm 0.12$	44.07	44.30	-2.01
G 55673	0.44	H $\beta$	$1.20 \pm 0.19$	$41.66 \pm 0.09$	$8.45 \pm 0.16$	43.51	33.91	-2.60
G 71277	0.32	H $\alpha$	$1.05 \pm 0.11$	$41.99 \pm 0.05$	$8.36 \pm 0.09$	43.35	33.61	-2.85
G 165213	0.31	H $\alpha$	$1.19 \pm 0.93$	$42.79 \pm 0.01$	$8.82 \pm 0.71$	42.99	34.45	-2.47
G 196970	0.32	H $\beta$	$1.20 \pm 0.60^a$	$40.89 \pm 0.07$	$8.08 \pm 0.45$	43.11	32.74	-3.44
G 208794	0.45	H $\beta$	$1.20 \pm 0.11$	$41.35 \pm 0.04$	$8.80 \pm 0.12$	43.26	33.6	-2.80

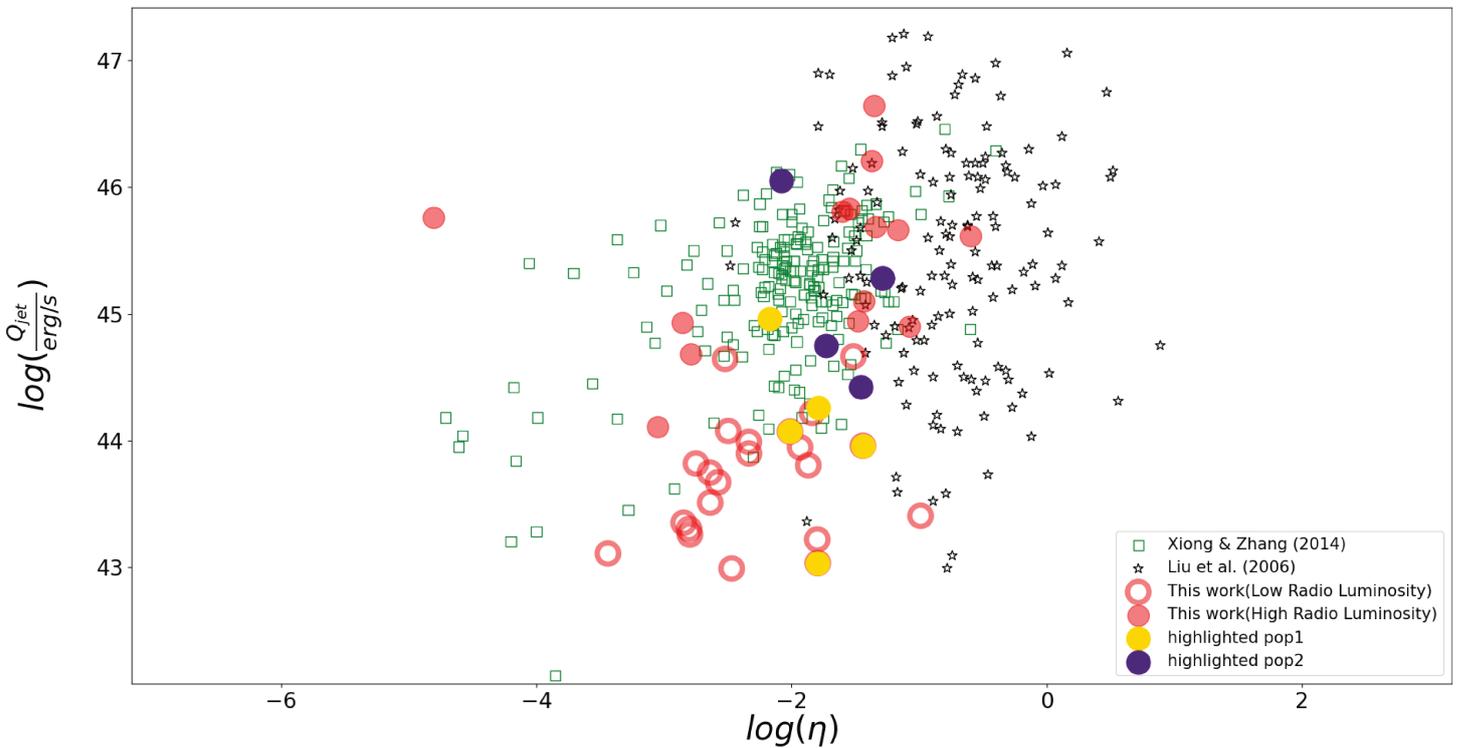
ID	$z$	Line	FWHM ( $10^3 \text{ km s}^{-1}$ )	$\log(L_{\text{line}}/(\text{erg s}^{-1}))$	$\log(M_{\text{BH}}/M_{\odot})$	$\log(Q_{\text{jet}}/(\text{erg s}^{-1}))$	$\log(L_{\text{BLR}}/(\text{erg s}^{-1}))$	$\log \eta$
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
G 249591	0.43	H $\beta$	$1.20 \pm 0.62^a$	$41.66 \pm 0.14$	$8.45 \pm 0.46$	43.75	33.91	-2.64
G 251343	0.76	H $\beta$	$1.19 \pm 0.15^a$	$43.96 \pm 0.02$	$9.57 \pm 0.12$	44.42	36.21	-1.46
G 298359	0.34	H $\alpha$	$1.19 \pm 0.18$	$42.07 \pm 0.06$	$8.51 \pm 0.14$	44.93	33.76	-2.85
G 372455	0.31	H $\alpha$	$1.19 \pm 0.60^a$	$42.66 \pm 0.08$	$8.77 \pm 0.08$	43.67	34.29	-2.58
G 537618	0.49	H $\beta$	$1.10 \pm 0.04$	$43.86 \pm 0.02$	$9.45 \pm 0.05$	43.96	36.11	-1.44
G 714133	0.56	H $\beta$	$1.20 \pm 0.19$	$42.25 \pm 0.01$	$8.74 \pm 0.10$	43.99	34.50	-2.34
G 714228	0.34	H $\beta$	$1.20 \pm 0.11$	$40.86 \pm 0.06$	$8.06 \pm 0.13$	44.11	33.11	-3.05
G 720847	0.42	H $\beta$	$1.19 \pm 0.24$	$41.92 \pm 0.02$	$8.57 \pm 0.19$	44.07	34.17	-2.5
G 721940	0.64	H $\beta$	$1.20 \pm 0.60^a$	$43.31 \pm 0.04$	$9.25 \pm 0.44$	44.26	35.56	-1.79
G 745066	0.38	H $\beta$	$1.19 \pm 0.60^a$	$41.73 \pm 0.02$	$8.47 \pm 0.45$	43.30	33.77	-2.8
G 746605	0.37	H $\beta$	$1.24 \pm 0.09$	$43.37 \pm 0.04$	$9.31 \pm 0.08$	43.03	35.62	-1.79
G 748144	0.59	H $\beta$	$1.22 \pm 0.31$	$43.08 \pm 0.09$	$9.16 \pm 0.23$	43.95	35.33	-1.93
G 748815	0.33	H $\beta$	$1.19 \pm 0.07^a$	$42.05 \pm 0.01$	$8.63 \pm 0.09$	44.68	33.94	-2.79
G 804203	0.51	H $\beta$	$1.20 \pm 0.60^a$	$41.44 \pm 0.11$	$8.34 \pm 0.45$	43.82	33.69	-2.75
G 835899	0.76	H $\beta$	$1.19 \pm 0.16^a$	$42.24 \pm 0.02$	$8.72 \pm 0.13$	43.90	34.49	-2.33
G 887308	0.55	H $\beta$	$1.20 \pm 0.19^a$	$43.17 \pm 0.03$	$9.18 \pm 0.15$	43.80	35.42	-1.87

**Table A.2:** Black hole, accretion, and jet properties of the radio sources in our sample. Column (1): Source ID; (2) spectroscopic redshift; (3-5) broad emission line, FWHM, and line luminosity; (6) single-epoch SMBH mass; (7) jet power; (8) BLR luminosity; and (9) Eddington ratio. Sources marked with the symbol  $^a$  in column (4) have uncertain FWHM errors from GAMA, when the lines were fit with a single Gaussian component. To overcome this limitation, we considered the GAMA fits that include both narrow and broad components of the emission line. We then assumed an FWHM relative error equal to that resulting from the fit to the broad component of the line.

## **A.2 Figures with highlighted populations**

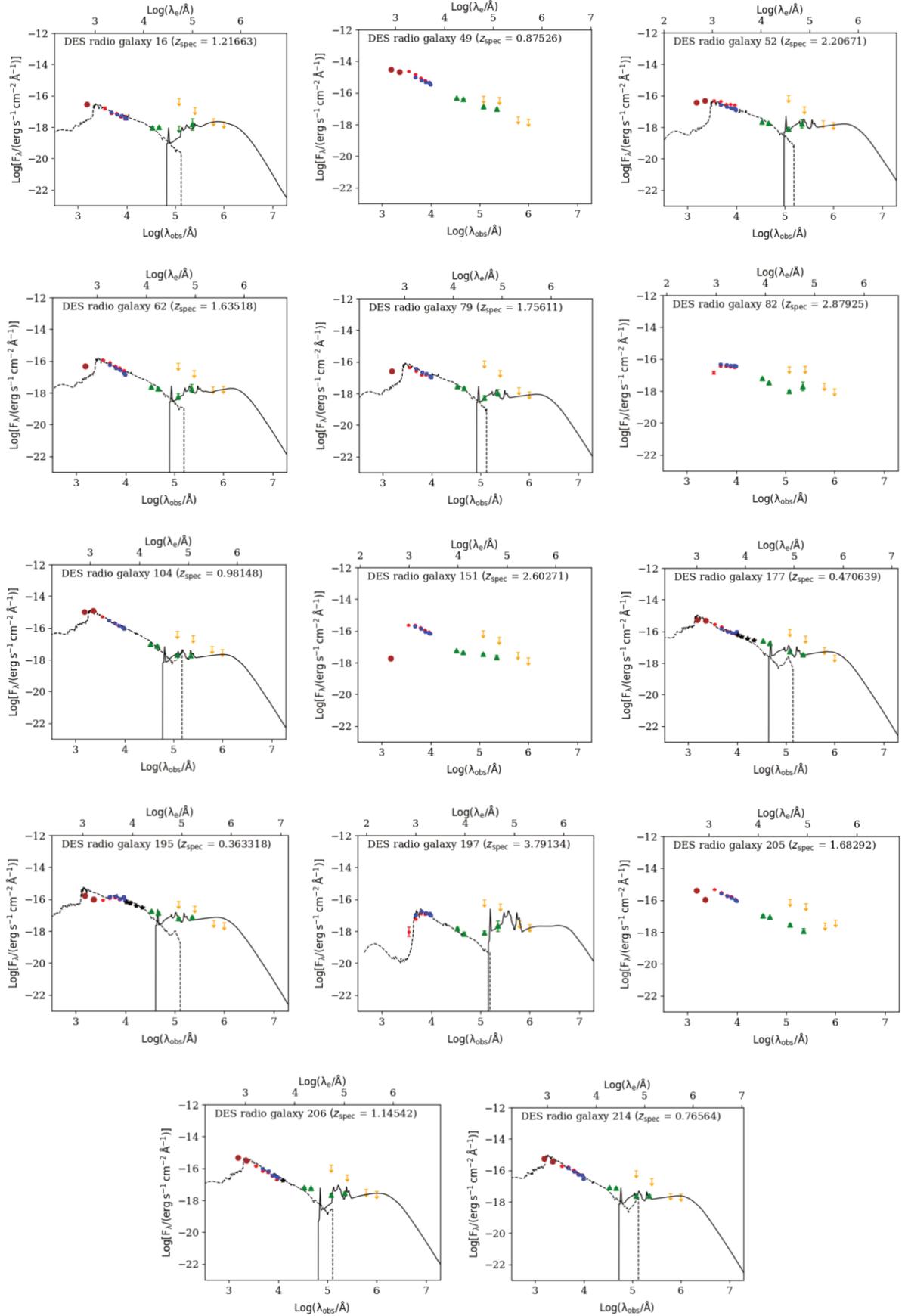


**Figure A.1:** Same figure as Fig. 2.8 but with the outliers in our sample highlighted.



**Figure A.2:** Same figure as Fig. 2.9 but with the outliers in our sample highlighted.

## **A.3 Radio sources SEDs**



**Figure A.3:** SEDs of the radio sources in our sample with  $M_{\text{BH}}$  estimates. The IDs, names, and redshifts of the galaxies are shown at the top left of each panel. Data points are from GALEX (brown dots), SDSS (red pentagons), DES (blue squares), WISE (green triangles), and IRAS (yellow upper limits). Dashed and solid lines are the best-fit models for the stellar and dust components, respectively.



# APPENDIX *B*

---

GATOS

---

## **B.1 Moment maps and position-velocity diagrams**

PARAMETER	NGC4388	NGC5506	NGC5643	NGC6300	NGC7314	NGC7465	NGC7582
$M_{\text{bulb}}$	0.40	0.20	0.10	0.25	0.10	0.60	0.45
$R_{\text{bulb}}$	0.10	0.20	0.10	0.10	0.10	0.15	0.10
$H_{\text{bulb}}$	0.1	0.1	0.1	0.1	0.1	0.1	0.1
$M_{\text{gal}}$	8.0	4.0	8.0	9.0	3.0	1.0	5.0
$R_0$	3.0	3.0	1.0	3.0	3.0	10.0	8.0
$H_t$	0.001	0.001	0.001	0.003	0.001	0.002	0.001
$\log M_{\text{BH}}$	6.61	6.63	6.99	6.39	6.41	6.62	6.40
$RAPM$	0.020	0.021	0.020	0.021	0.020	0.020	0.020
$R_{\text{gas}}$	0.48	0.50	0.55	0.20	0.20	0.20	0.60
$H_{\text{gas}}$	0.12	0.012	0.011	0.03	0.02	0.12	0.08
$PA$	[10,10]	[0,0]	[50,150]	[160,160]	[80,80]	[30,30]	[110,110]
$IA$	79	80	29	57	55	55	59

**Table B.1:** Table of the parameters used to obtain the fits of the observed moment maps shown in the figures 4.11 to 4.17. The 12 parameters are described in the second paragraph of the 4.3:  $RAPM$  has no units,  $M_{\text{bulb}}$  and  $M_{\text{gal}}$  are in units of  $2.25 \times 10^9 M_{\odot}$ ,  $M_{\text{BH}}$  in  $\log(M/M_{\odot})$ ,  $PA$  and  $IA$  are in degrees and  $R_{\text{bulb}}$ ,  $H_{\text{bulb}}$ ,  $R_0$ ,  $H_t$ ,  $R_{\text{gas}}$  and  $H_{\text{gas}}$  are in kpc.

## B.2 Relative errors of the $M_{\text{BH}}$ predictions as a function of free parameters

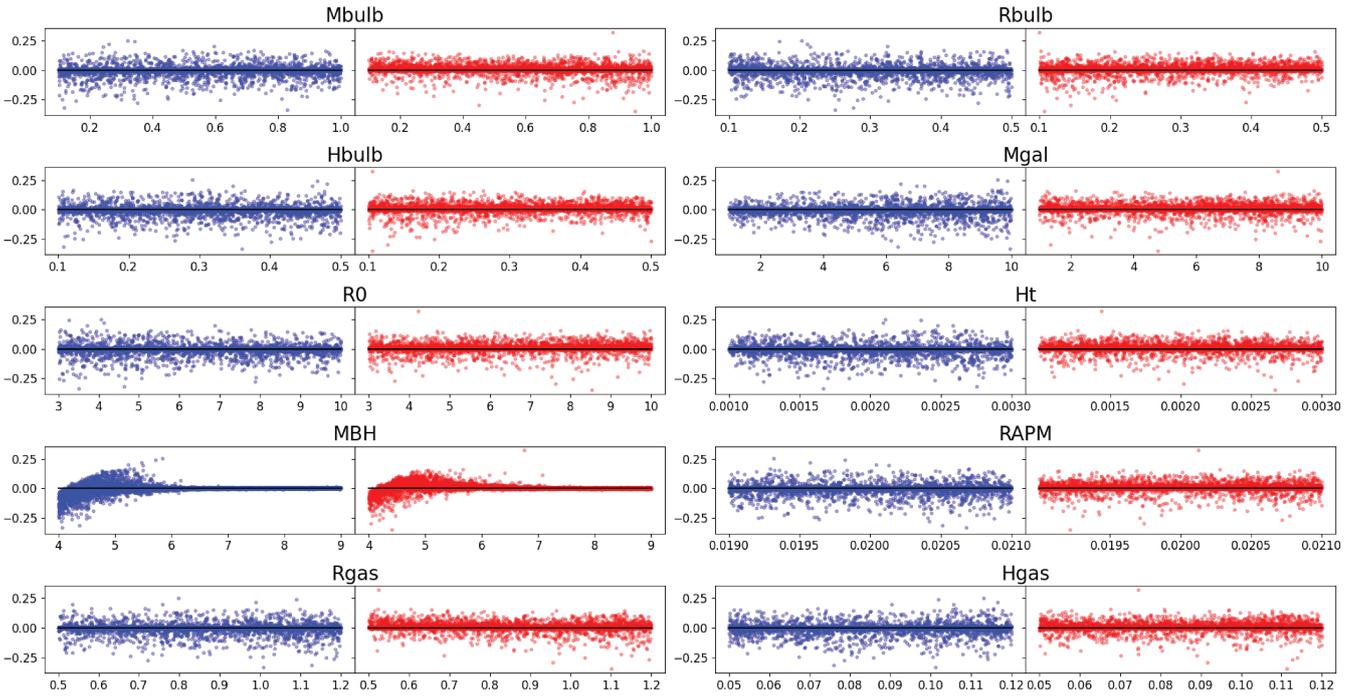


Figure B.1: Same as Fig. 4.7 for NGC 5506.

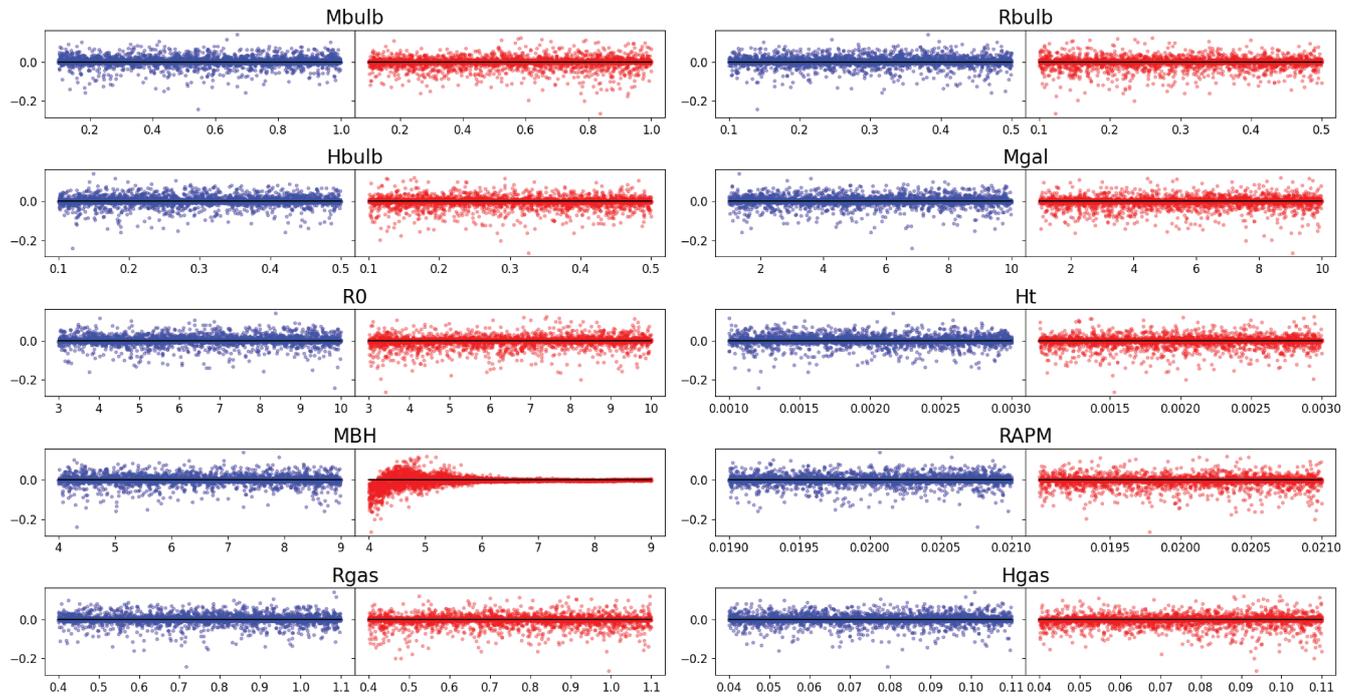
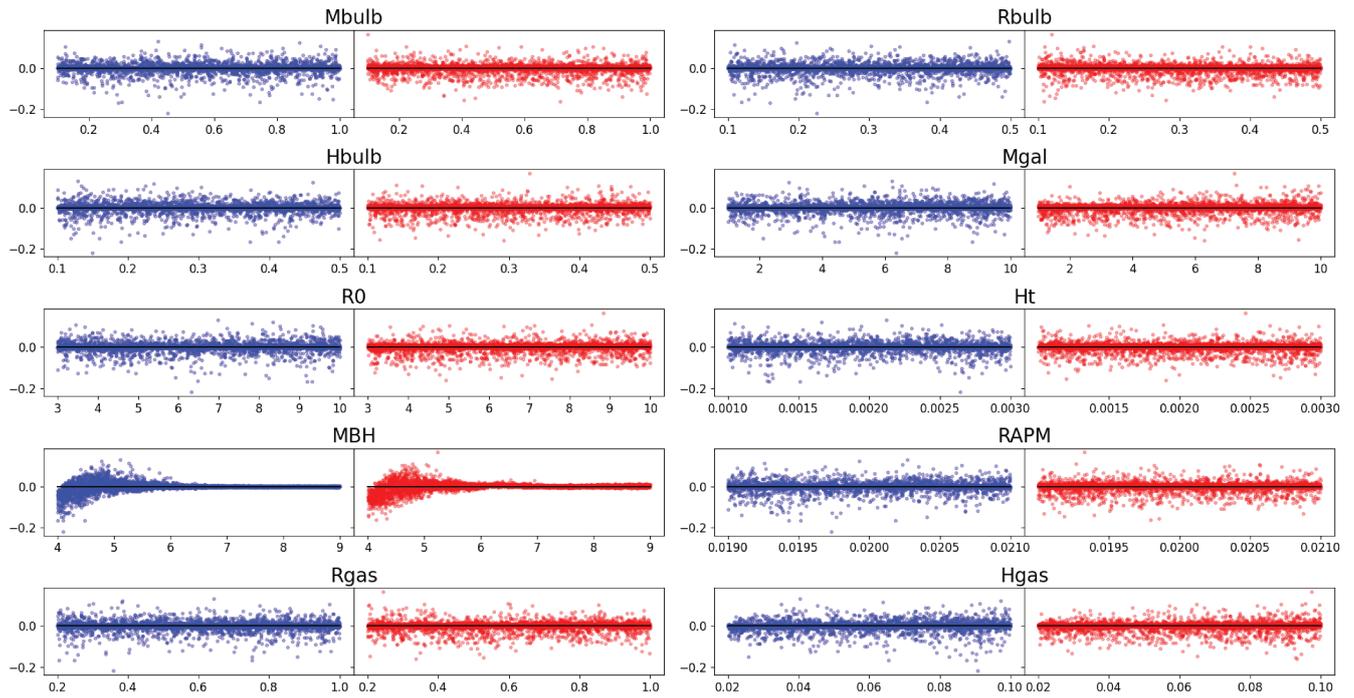
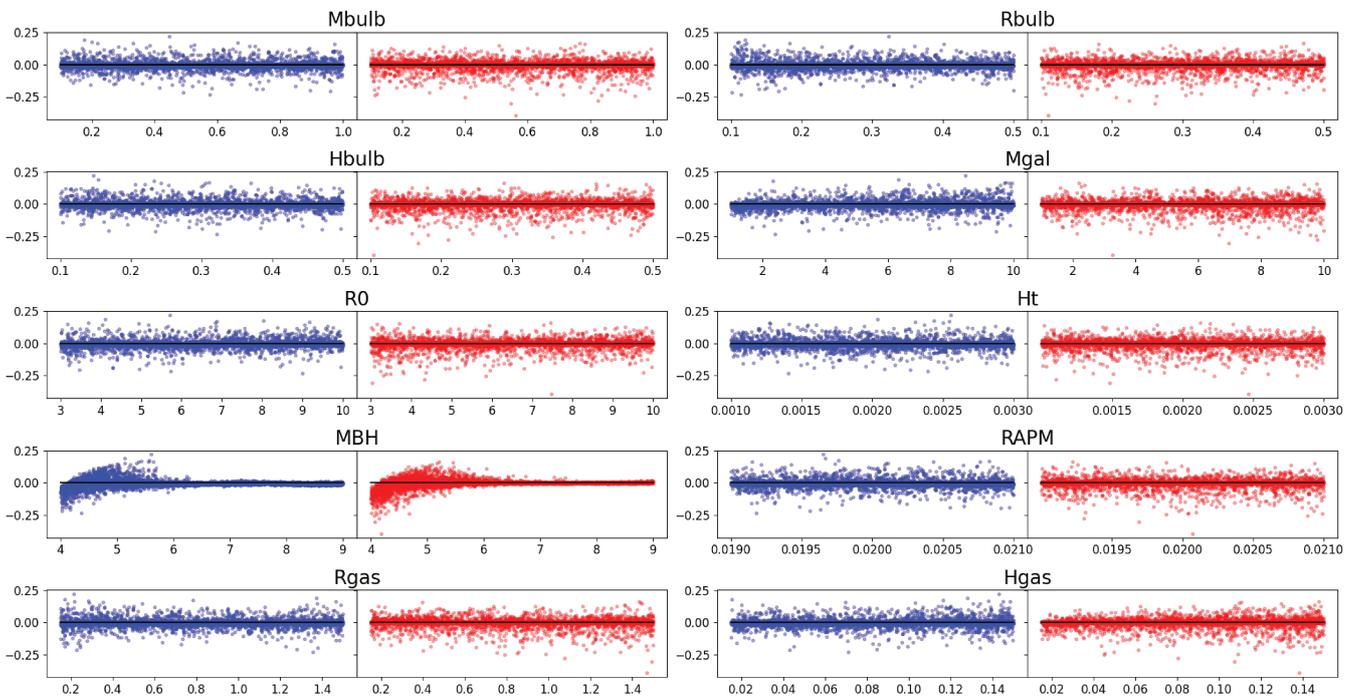


Figure B.2: Same as Fig. 4.7 for NGC 5643.



**Figure B.3:** Same as Fig. 4.7 for NGC 6300.



**Figure B.4:** Same as Fig. 4.7 for NGC 7314.

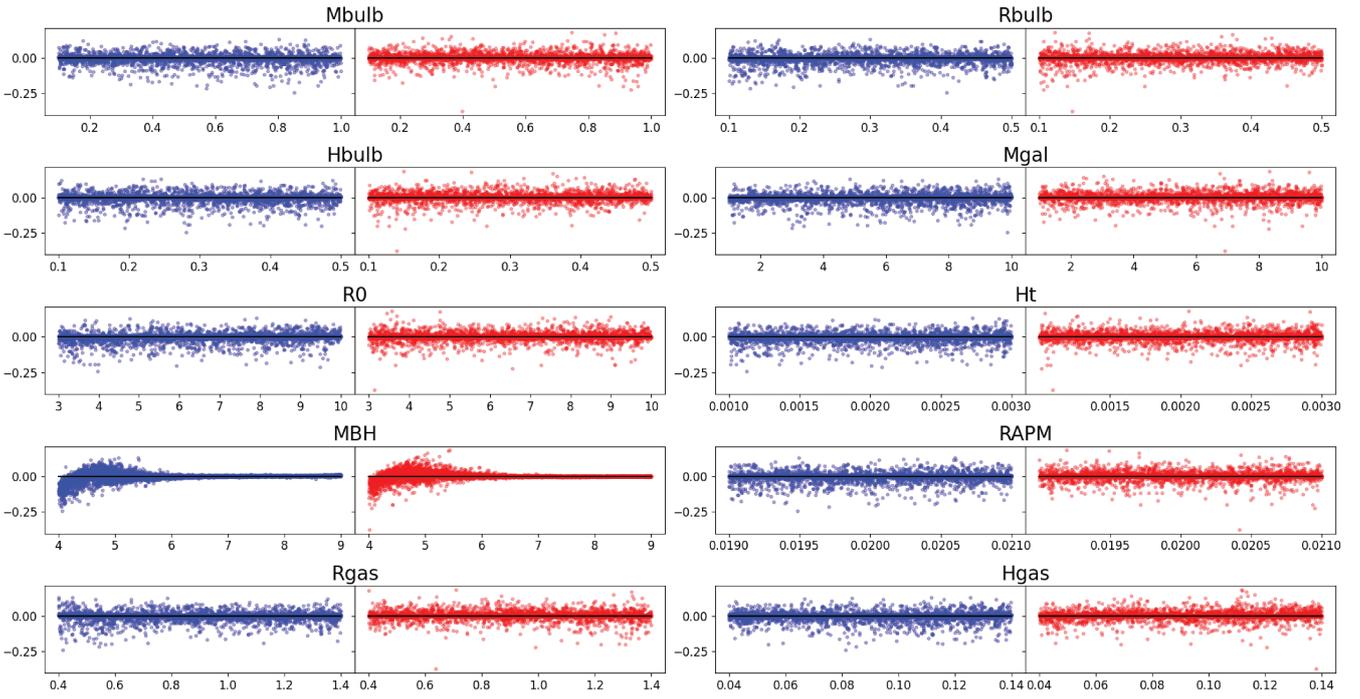


Figure B.5: Same as Fig. 4.7 for NGC 7465.

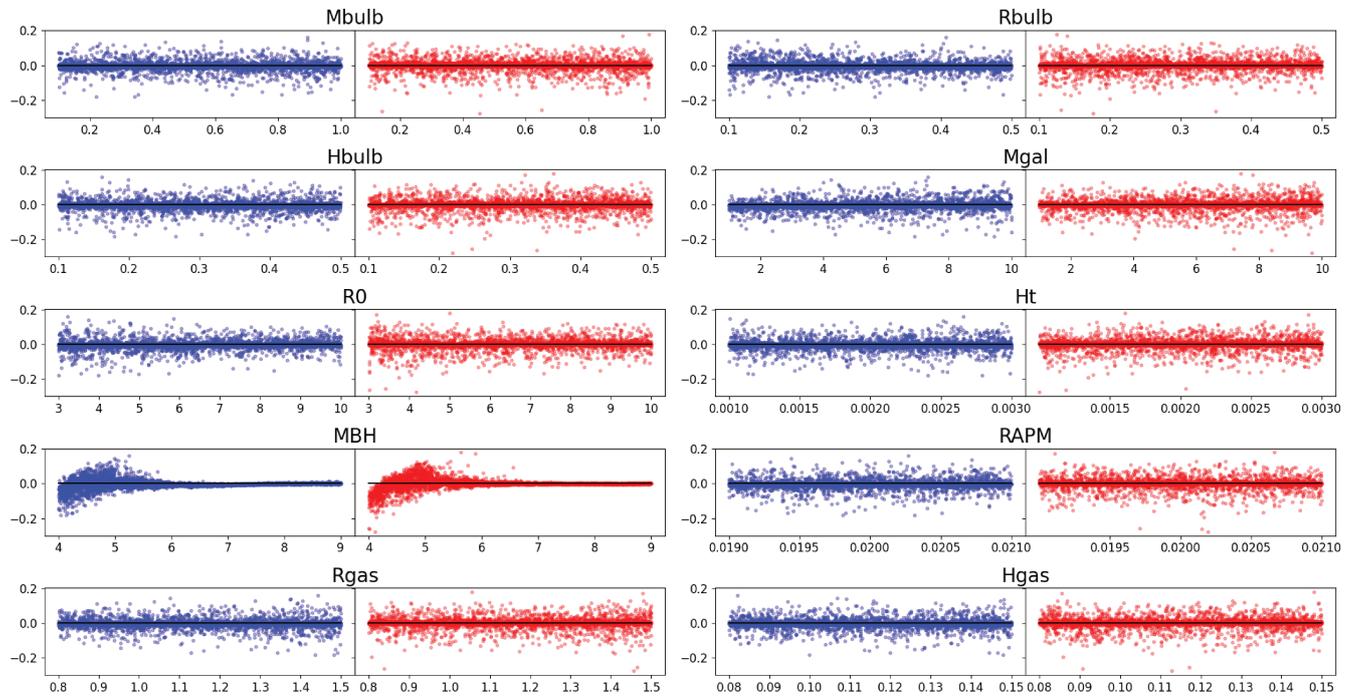


Figure B.6: Same as Fig. 4.7 for NGC 7582.

### B.3 Models 2D histograms of their RE

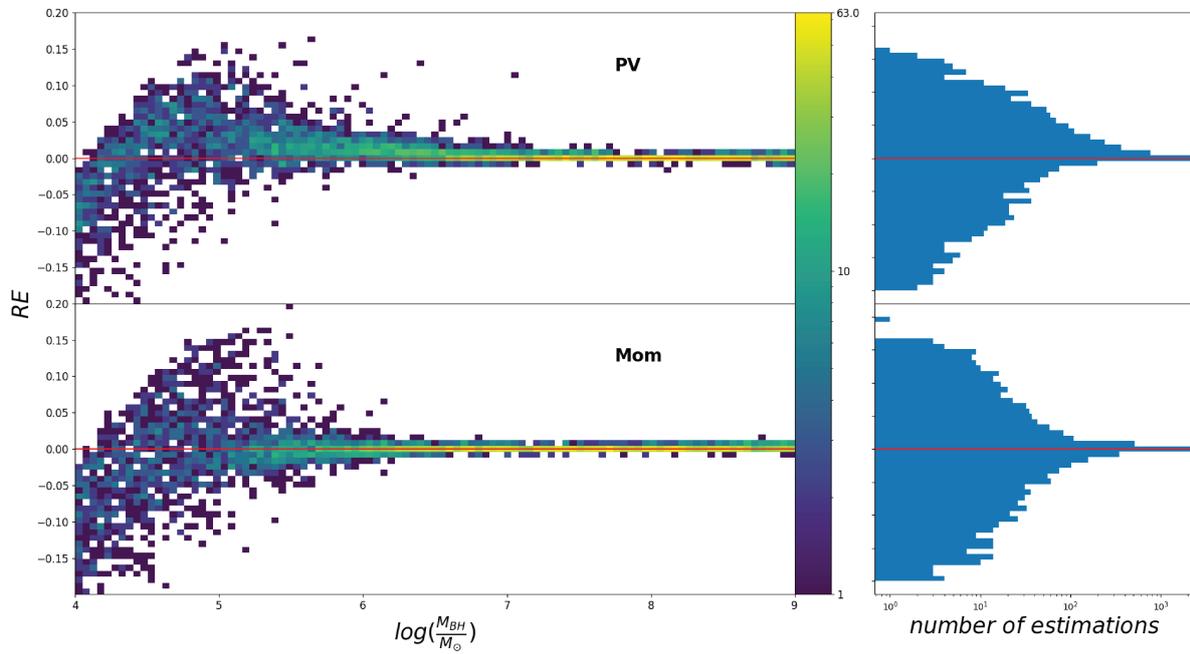
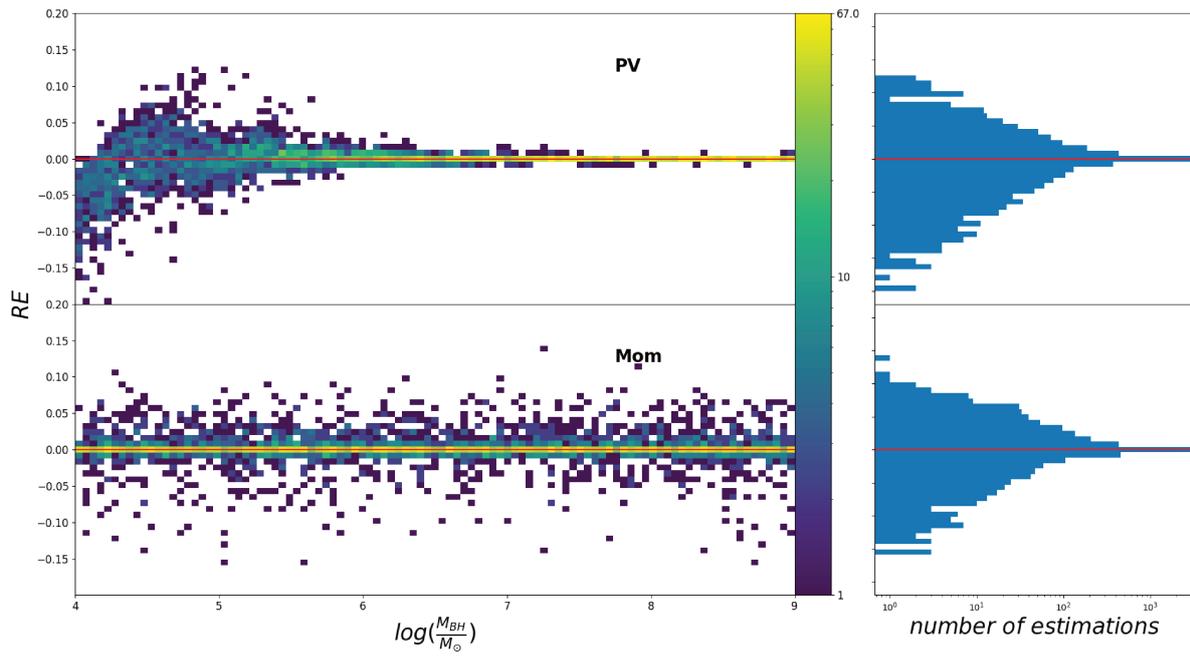
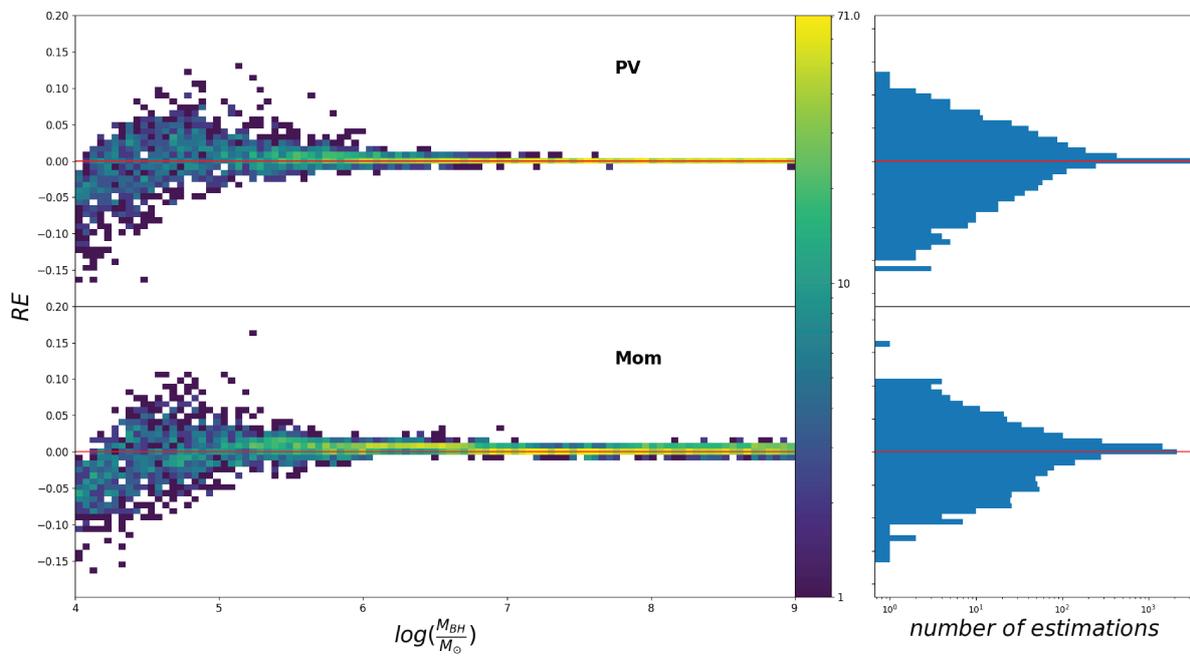


Figure B.7: Same as Fig. 4.8 for NGC 5506.



**Figure B.8:** Same as Fig. 4.8 for NGC 5643.



**Figure B.9:** Same as Fig. 4.8 for NGC 6300.

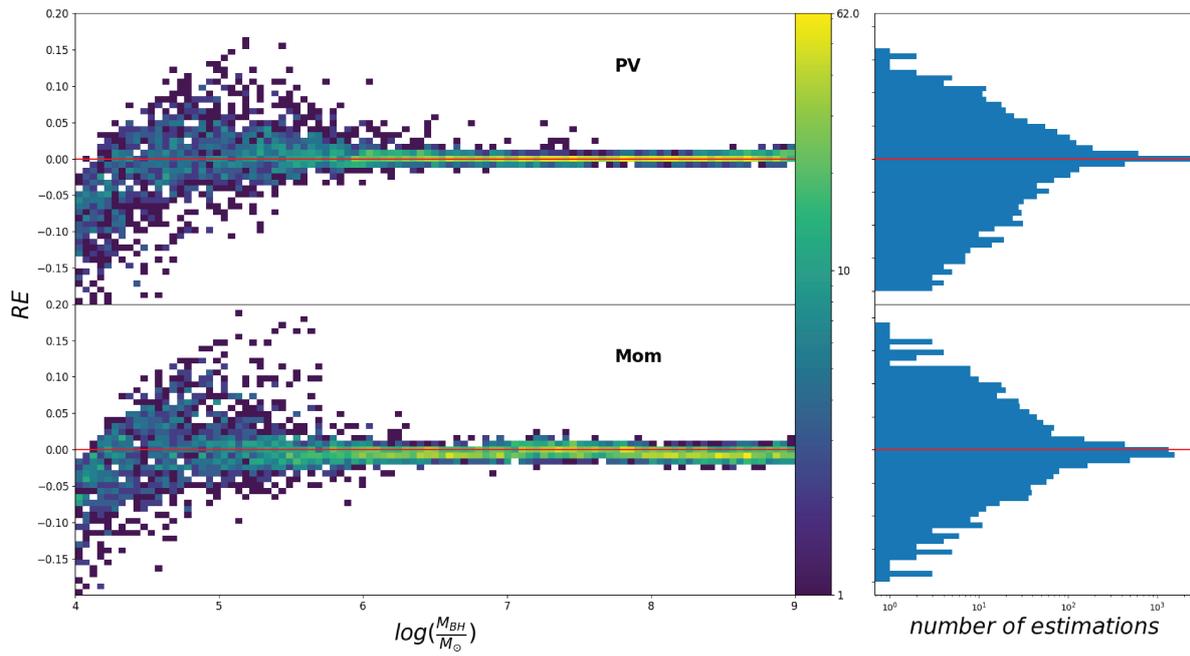


Figure B.10: Same as Fig. 4.8 for NGC 7314.

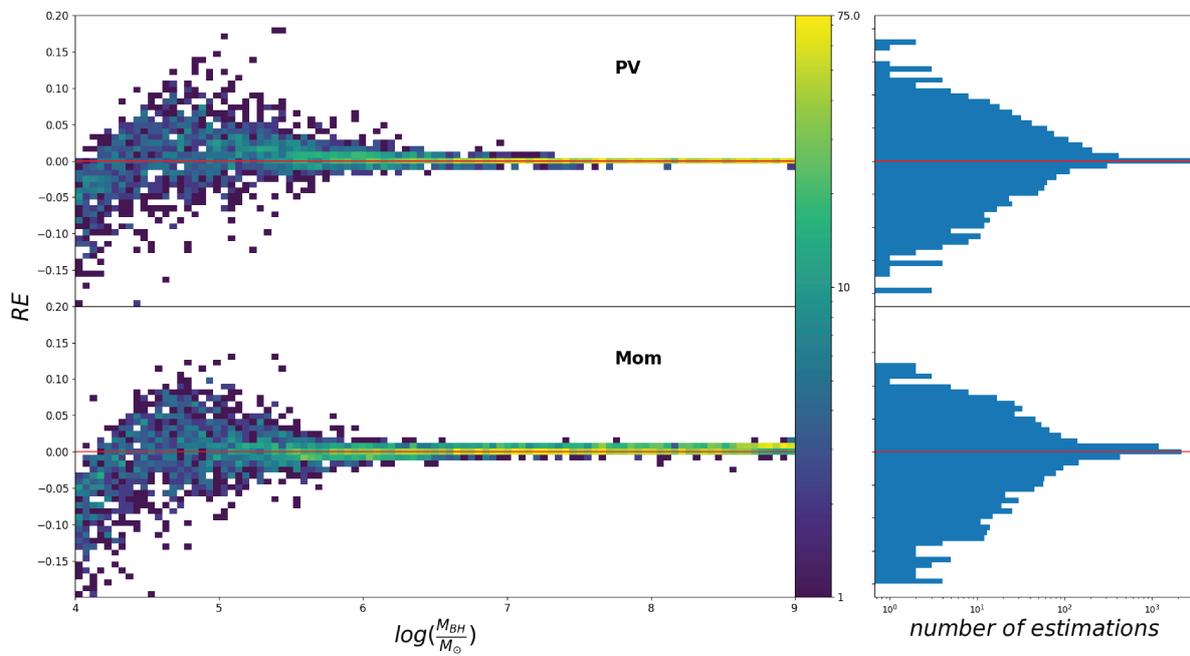
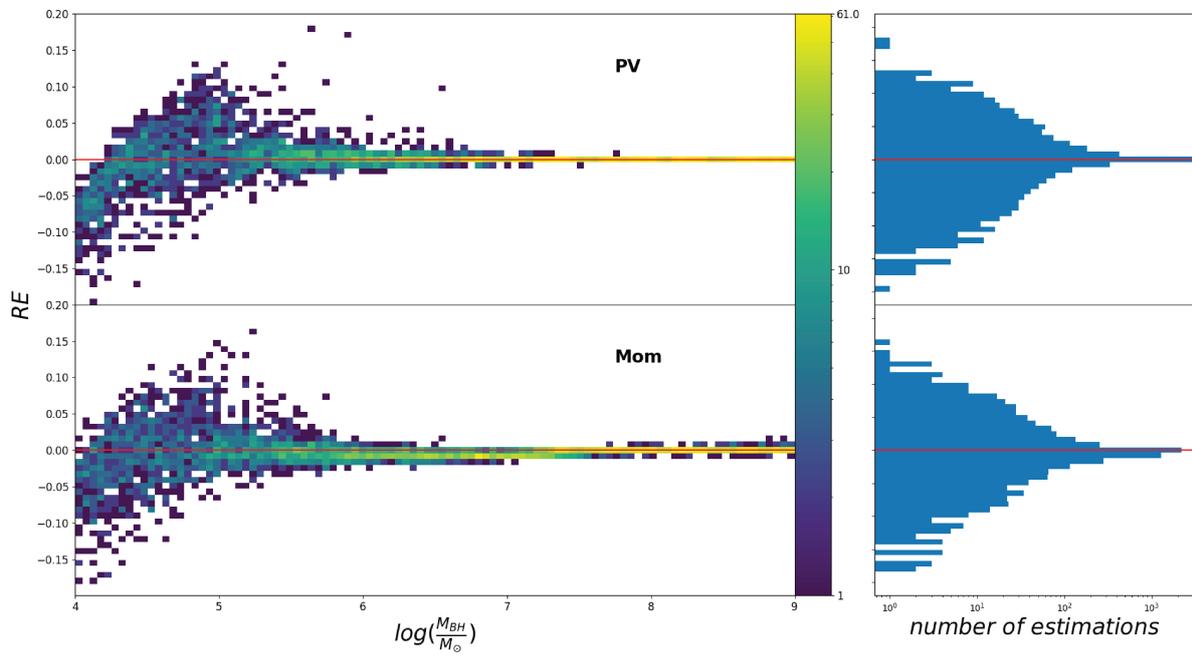


Figure B.11: Same as Fig. 4.8 for NGC 7465.



**Figure B.12:** Same as Fig. 4.8 for NGC 7582.

## RÉSUMÉ

---

La relation entre la masse des trous noirs supermassifs (SMBH) et les propriétés des galaxies est bien établie, indiquant une coévolution entre les SMBH et leurs hôtes. Cependant, les mécanismes détaillés d'alimentation et de rétroaction des Noyaux Actifs de Galaxies (AGN) demeurent incertains, notamment pour les AGN de faible luminosité, les AGN occultés et les galaxies de type tardif, rendant difficile la détermination précise des masses des trous noirs centraux.

Cette recherche doctorale a étudié différents aspects des AGNs et leur connexion avec les SMBHs dans différents environnements galactiques. Deux projets distincts ont été réalisés dans le but de contribuer à notre compréhension de l'alimentation des AGNs, des mécanismes de rétroaction et de la détermination des masses des SMBHs.

Le premier projet examinait la relation entre la masse des SMBHs et les propriétés des galaxies radio-bruyantes, en mettant l'accent sur l'évolution du décalage vers le rouge. Un échantillon de 42 galaxies radio et d'AGN présentant des raies d'émission larges couvrant une plage de décalage vers le rouge de 0,3 à 4 a été constitué. Des données complètes ont été obtenues en croisant les sources radio de l'enquête FIRST du Very Large Array (VLA) avec des galaxies confirmées spectroscopiquement provenant d'enquêtes à large champ, telles que SDSS, DES, WISE et GAMA. L'analyse portait sur la masse stellaire, les taux de formation d'étoiles et les caractéristiques des trous noirs, révélant des informations sur l'évolution et la connexion entre les AGN radio-bruyants et les SMBH.

Le deuxième projet faisait partie de l'enquête Galaxy Activity, Torus, and Outflow Survey (GATOS) et avait pour objectif d'étudier les mécanismes d'alimentation et de rétroaction des AGN dans les AGN de faible luminosité, les AGN occultés et les galaxies de type tardif. La détermination précise des masses des trous noirs centraux dans ces systèmes est un défi. Des observations haute résolution réalisées avec le réseau Atacama Large Millimeter/submillimeter Array (ALMA) ont été utilisées pour étudier les régions circum-nucléaires de galaxies sélectionnées. Une approche novatrice combinant des techniques d'apprentissage automatique supervisé avec des diagrammes position-vitesse et des observations ALMA CO(3-2) a été développée pour estimer les masses des trous noirs. La méthode a été validée et testée à l'aide de simulations numériques, fournissant des estimations fiables et ouvrant la voie à une estimation automatisée de la masse des trous noirs grâce à l'apprentissage automatique.

## MOTS CLÉS

---

AGN, Radio-Loud AGN, Feedback, Tore moléculaire, Trou Noir supermassif, Apprentissage Machine

## ABSTRACT

---

The relation between supermassive black hole (SMBH) mass and galaxy properties is well established, indicating coevolution between SMBHs and their hosts. However, detailed feeding and feedback mechanisms of Active Galactic Nuclei (AGN) remain uncertain, especially for low-luminosity AGN, obscured AGN, and late-type galaxies, making precise determination of central black hole masses challenging.

This Ph.D. research investigated various aspects of AGN and their connection to SMBHs in different galactic environments. Two separate projects were conducted to contribute to our understanding of AGN feeding, feedback mechanisms, and SMBH mass determination.

The first project examined the relation between SMBH mass and properties of radio-loud galaxies, with a focus on redshift evolution. A sample of 42 radio galaxies and AGN with broad emission lines spanning a redshift range of 0.3 to 4 was compiled. Comprehensive data were obtained by cross-matching radio sources from the Very Large Array (VLA) FIRST survey with spectroscopically confirmed galaxies from wide-field surveys, such as SDSS, DES, WISE, and GAMA. The analysis included stellar mass, star formation rates, and black hole characteristics, revealing insights into the evolution and connection between radio-loud AGN and SMBHs.

The second project was part of the Galaxy Activity, Torus, and Outflow Survey (GATOS), focusing on studying the feeding and feedback mechanisms of AGN in low-luminosity AGN, obscured AGN, and late-type galaxies. Precise determination of central black hole masses in these systems is challenging. High-resolution Atacama Large Millimeter/submillimeter Array (ALMA) observations were utilized to study circum-nuclear regions of selected galaxies. A novel approach combining supervised machine learning techniques with position-velocity diagrams and ALMA CO(3-2) observations was developed to estimate black hole masses. The method was validated and tested using numerical simulations, providing reliable estimations and paving the way for automated black hole mass estimation using machine learning.

## KEYWORDS

---

AGN, Radio-Loud AGN, Feedback, Molecular Tori, Supermassive Black Hole, Machine Learning