Pontificia Universidad Católica de Chile
Facultad de Física
Instituto de Astrofísica

CENSUS OF ACCRETING SMBHS AT z > 4 ACROSS THE SOUTHERN HEMISPHERE

BY

Laura Natalia Martínez Ramírez

Tesis presentada al Instituto de Astrofísica de la Pontificia Universidad Católica de Chile para optar al grado académico de Doctor en Astrofísica.

Supervisors : Dr. Felipe Barrientos (PUC)
Dr. Hans-Walter Rix (Uni-Heidelberg)
Dr. Franz Bauer (U. Tarapacá)
Dr. Eduardo Bañados (MPIA)
Correctors : Dr. Jorge González (PUC)
Dr. Chiara Mazzucchelli (UDP)
Dr. Matthias Bartelmann (Uni-Heidelberg)

October 27, 2025
Santiago, Chile

Dissertation

submitted to the

Combined Faculty of Mathematics, Engineering and Natural Sciences

of Heidelberg University, Germany

for the degree of

Doctor of Natural Sciences

Put forward by

Laura Natalia, Martínez Ramírez

born in: Bucaramanga, Colombia

Oral examination: 12th of December, 2025

Census of accreting SMBHs at z > 4 across the southern hemisphere

Referees: Franz Bauer

Hans-Walter Rix

"De esta lucha entre la idea creadora de un gran talento y el deseo permanente del adversario de destruirla suele surgir la obra medular, las partidas de ajedrez que se perpetúan a través de los siglos, como modelos de un arte perfecto."

Roberto G. Grau, 1996

Tratado general de Ajedrez, Tomo III: Conformación de peones.

A mis papás, por ser un ejemplo de superación
e inculcarme el valor del conocimiento.

Agradecimientos

El camino hacia el doctorado se asemeja mucho a una inmersion en lo más profundo del campo, allí donde la señal de televisión es solo estática y tiene un poquito del fondo de radiación de microondas. Como lo dicen las malas lenguas: es un camino pedregoso, fangoso, y con subidas y bajadas; aunque también tiene sus trayectos pavimentados. Hoy quiero reconocer el rol vital de todas las personas que fueron mi bocadillo, ese azúcar que me dió la energía para seguir adelante cuando las fuerzas y el desconsuelo me abatían, y que ahora, merecen celebrar este resultado, un conglomerado de esfuerzos visibles e invisibles.

Gracias a mis supervisores Franz, Eduardo y Gabriela, por tantas enseñanzas, por ofrecerme un espacio y la oportunidad de crecer como científica, como divulgadora y también como persona. Gracias por confiar en mis capacidades cuando yo no lograba ver mi potencial, por apoyarme en mis planes (siempre óptimistas), por su empatía en los momentos difíciles, y por ser esos modelos que me motivaron a tratar de construir una mejor versión de mí. A mis grupos de investigación en Santiago y en Heidelberg, a Silvia, Julien, Raphael y al profesor Ezequiel por las discusiones científicas, ideas, cuestionamientos y preguntas que permitieron mejorarar mi ciencia.

Siempre estaré eternamente agradecida con el MAS por permitirme buscar la felicidad en un aula de clase, en un telescopio apuntando al sol en un parque, o en un poco de algodón y pintura, cuando las frustaciones de la ciencia me apagaban la llama. Gracias Maka, Pepa, Tracy, Cata, Pauli y Sofi por ser ese constante recordatorio de que me gusta lo que hago, por mostrarme el lado más hermoso y humano de Chile, por apañarme y por tanto amor. A mis compañeros de ruta, que se convirtieron en mis amigos y cómplices. A Maca, Jenn, Karina, Marcy, Claudio, Juan, Julio, Álvaro y Devika; gracias por ser mi familia y mi casa en Chile, por el apoyo, por ser parte de mis tantas primeras veces y por tantas aventuras juntos. A mis amigos de Heidelberg: Fran, Lukrecia, Anna, Jeff, Javier, Roel, Lynn, Jan; y al team del MPIA: Dimitris, Jonas, Anirudh, Johanna, David, Daniele,

Stefan, Simon y Rogelio, gracias por acogerme, por recargarme con tantas risas y buena energía, por librarme tantas veces de mi cerebro ensimismado y por la preocupación, en particular en estos últimos meses que fueron una prueba de fuego para mí. A mis amigos de la UIS: Steven, Rolando, Juan y Vero, gracias por estar incondicionalmente conmigo, indelebles y atemporales. Gracias a ustedes decidí emprender este gran viaje.

A mi ancla, Alessandro que no me soltó durante mis peores momentos y me apoyó con valentía aún cuando él mismo moría de miedo. Gracias por convencerme de creer en mí, por celebrar mis triunfos, por tanta paciencia, escucha, amor y empatía; por tu óptimismo contagioso y por empujarme a ser más constante. A mis papás, por todos los sacrificios que me permitieron llegar hasta aquí, por enseñarme a soñar en grande, a ser resiliente, a dar lo mejor de mí y confiar en que el trabajo duro siempre traerá sus frutos tarde o temprano. Gracias a mis hermanos Cata, Paola y John; y a mis cuñados Rapha y Vivi por su apoyo incondicional, por emocionarse con mis chocoaventuras, animarme en los momentos de adversidad y hacerme sentir acompañada aún con miles de kilómetros de distancia de por medio. A mi motorcito Jeremy, gracias por llegar mi vida a llenarla de tanta amor. La fortuna de verte crecer y poder enseñarte tantas cosas me mueven a querer ser cada vez mejor y a ser un referente para tí.

Nada de esto hubiera sido posible sin ellos, y a ellos les debo lo que soy al final de este camino. Mi mejor forma de agradecerles es seguir caminando, persiguiendo un próximo destino. El camino no ha dejado de ser pedregoso ni tedioso, pero ahora yo estoy equipada.

Financiamiento: esta tesis fue financiada parcialmente por la beca de doctorado ANID número 21220337, el Instituto Milenio de Astrofísica MAS, la beca de Estadía de Cotutela Doctoral en el Extranjero, conducente a doble grado de la Vicerrectoría de Investigación de la Pontifica Universidad Católica de Chile y el Max Planck Institute For Astronomy a través de la subvención “ASTR00171338”. El trabajo de los supervisores fue financiado parcialmente por FONDECYT Regular 1200495, ANID-Chile BASAL CATA FB210003, y Millennium Science Initiative Program – ICN12_009.

Disponibilidad de datos: todos los datos observacionales utilizados en este estudio están disponibles públicamente.

Contents

1 Introduction **29**

1.1 Active galactic nuclei 29

1.2 Quasars in the reionization era 32

1.3 Traditional catalog-based selection of high-z quasars 35

1.4 Machine learning for quasar discoveries 38

1.5 Characterization of quasars and active galaxies 41

1.5.1 Photometric SED fitting 42

1.5.2 Broad emission line measurements 46

1.6 This thesis 50

2 Quasar discoveries with self-supervised learning applied to images **54**

2.1 Introduction 54

2.2 Methodology 56

2.2.1 Pre-selection of *i*-dropout candidates 56

2.2.2 Constrastive Learning for the selection of quasars candidates . . . 61

2.2.3 2-dimensional embedded space 63

2.2.4 Assessment of the quasar nature of promising candidates 74

2.3 Results 81

2.3.1 New $z > 6$ quasars 81

2.3.2 Physical properties of the newly discovered quasars 87

2.3.3 Comparison with the literature quasars 92

2.4 Conclusions 97

3 Investigating MIR variability for high-z quasar selection 100

3.1 Introduction 100

3.1.1 Metrics for variability modeling 103

3.2 Methodology 106

3.2.1 Compilation of light curves 106

3.2.2 Computation of variability features with ALeRCE light-curve classifier 108

3.3 Results 111

3.3.1 General trends of quasars vs calibrator star populations 111

3.3.2 Individual-quasar variability estimation 118

3.3.3 Power spectral density analysis 123

3.4 Summary 128

4 Radio-to-X-ray modelling of AGN SEDs 131

4.1 Introduction 131

4.2 Methodology 133

4.2.1 Flexibility for adding new filters and models 133

4.2.2 Flexibility for including priors 134

4.2.3 Physical models from the radio to the X-rays 134

4.2.4 Data: Radio-to-X-ray SEDs of local AGN 148

4.2.5 Fitting setup and model comparison strategy 152

4.3 Results 154

4.3.1 Comparing torus models 157

4.3.2 Comparing accretion-disk models 163

4.3.3 Comparing model combinations 167

4.3.4 Effects of photometric uncertainties 168

4.3.5 Current advances and limitations on the modeling of the torus . . 169

4.3.6 Current advances and limitations on accretion disk models 172
4.3.7 The AGN contribution to the SED from the radio to X-rays 178
4.4 Conclusions 181

5 Summary and future work **185**

Appendices **188**

A Supervised learning with Modulos IA **190**

B Latent space representation of HSC quasar candidates **197**

C Combining optical and NIR images into CL **201**

D Structure function parameter space for unWISE light curves **206**

Bibliography **208**

List of Figures

1.1 Schematic representation of the unified AGN model from Thorne et al. (2022), illustrating the classification framework based on viewing angle, radio emission strength, and accretion rate (which traces the electromagnetic power). The upper left and right quadrants correspond to low- and high-excitation radio galaxies (LERGs and HERGs), typically associated with the kinetic (radio) and radiative (quasar) modes, respectively. The lower left and right quadrants represent radio-quiet Seyfert galaxies and quasars. Common AGN subclasses are labeled accordingly. Colored bars indicate the expected presence and relative strength (transparency) of broad or narrow emission lines, and of MIR, radio, X-ray, or γ-ray emission. . . 30

1.2 Two-dimensional intensity map compiling the published spectra of $z \gtrsim 5.3$ quasars, sorted by increasing redshift from bottom to top. Emission lines such as Ly α, O I+ S III, S IV and C IV; along with the Ly α forest and the GP troughs are labeled. The progressive transition from transmission spikes to GP troughs with increasing redshift traces the evolution of the cosmic reionization process. Figure taken from Fan et al. (2023). 34

1.3 Sky distribution in celestial coordinates of spectroscopically confirmed quasars at $z > 5$ up to date (including the new ones from this thesis shown as unfilled circles), color-coded by redshift. Higher concentrations are observed in the Northern Hemisphere and around $dec \sim \pm 5°$, the latter dominated by HSC discoveries of faint quasars. 37

1.4 Schematic representation of the dominant components shaping the emission of active galaxies according to the current AGN unification model. From the most inner region: the hot corona and accretion disk emission dominating the optical, UV and X-rays (blue curves), the hot-dust torus in the NIR and MIR (purple curve) and the radio jet (light-blue curve). The host galaxy components extended to kpc scales and contributing to the SED are: the stellar population (yellow curve) and the cold dust from star-forming regions (green curve). The upper panel shows an example of an SED modeling with the best-fitted models, along with the photometric data. Artistic illustration of the AGN taken from: `https://www.iaa.csic.es/~jlgomez/Jose_L._Gomez/Animations_files/agn_wolli_2_1.jpg`. 44

1.5 Transversal view of AGN emission components relevant for the reverberation mapping technique: the accretion disk, the corona represented by a light blue circle, and the BLR region by pink clouds. *Left panel:* different light travel paths of the UV continuum (blue arrow) and the re-processed BLR lines (green and red arrows) emissions. *Right panel:* effects of the light travel times on the stochastic variability observed in the continuum and emission-line fluxes. The time delay provides an estimate of the BLR size. Credits: Francisco R. Villatoro, available in `https://francis.naukas.com/files/2019/06/D20190613-nature-astronomy-s41550-019-0815-y-principle-1.png`. 47

2.1 z-band $10'' \times 10''$ sized stamps illustrating examples of the artifacts identified in the photometry through visual inspection of the tensor. Light blue circles indicate the $1''$ and $3''$ radius apertures used in the compactness criteria calculation. The $C_{1''/3''}$ values are presented above the corresponding postage stamp. 60

2.2 Architecture of the self-supervised contrastive learning framework used in this work. The top and left sections illustrate the image preprocessing and the tensor assembly steps. The structure of network is shown sequentially from left to right, starting with data augmentation, then the encoder block, and lastly the projection head. Red and green arrows indicate the contrastive loss computation, which measures similarity between random pairs of augmented images. 62

2.3 Embedded Euclidean space generated by UMAP for LS DR10 i-dropout sources after training the CL algorithm. All sources are represented as grey countours given by the number density; known M, L and T dwarfs are: brown, black and red crosses, respectively; and Gaia DR3 stars are yellow crosses. Spectroscopically confirmed quasars from the literature and the new quasars from this work are represented by filled circles and stars, respectively, all of them color-coded by their redshift. Quasar candidates are selected from the lower part on the left island where the source density is low and the contamination ratio with ultracool dwarfs is $\sim$ 1:1. The two regions with the highest concentration of quasars are presented as zoom-in panels and highlighted with black squares. 65

2.4 Color-coded embedded space maps showing catalog photometric features: z-band magnitude (*top left*), $i - z$ color (*top right*), $z - J$ color (*bottom left*) and z-W1 color (*bottom right*). *Top left*: magnitude gradients across both regions underscore the crucial role of the brightness in shaping the low-dimensional representation. *Top right*: Multiple regions of strong $i - z$ color indicate that while this feature may not fully dominate the embedding, it still plays a key role—evidenced by patterns and some clustering that overlap with the positions of spectroscopically confirmed quasars. *Bottom left*: The $z - J$ color map reveals populations consistent with blue quasars ($z - J \leq 0.5$), quasars with intermediate colors ($z - J \sim 1.0$–1.5) and a population with red NIR slopes ($z - J > 2$), supporting the potential of the model to uncover diverse quasar types. *Bottom right*: Red z-W1 colors seems to reflect the W1 blending due to companions and potential faint T dwarf populations. 67

2.5 Continuation Fig 2.4. Color-coded embedded space maps showing the g-band depth (*left*), and z-band PSF size (*right*). *Left*: Subtle observational systematics, such as variations in background noise texture, may influence the latent space structure as evidence by the g-band depth pattern. *Right*: Larger PSFs point to a mix of slightly extended or noisy z-band images. . 68

2.6 Mean pixel-by-pixel z, r and g band fluxes (RGB mapped) within 15×15 binned embedded space. The number of sources within each bin contributing to the mean are shown in white in the left upper side of each image. The red dashed lines highlight a region with the highest number of known quasars, with a zoom-in binned embedded space on the right side. The number of known quasars in each bin at the zoom-in plot is shown in orange, below the number of sources in white. 69

2.7 Embedded space distribution of spectroscopically confirmed stars, color-coded by the signal-to-noise ratio (SNR) of Gaia DR3 proper motion measurements. Higher SNR values correspond to stronger detections of motion, highlighting a gradient in proper motion across the embedded space. 70

2.8 2D KDE density maps for MLT ultracool dwarfs (*upper*), $z > 5.7$ quasar population (*central*); and quasar-to-brown-dwarf density ratio map (*lower*). White countours highlight the source density in the embedded space. . . . 73

2.9 Histogram of the $\chi^2_{\rm BD}/\chi^2_{\rm quasars}$ ratio of the population of spectroscopically confirmed ultracool dwarf (red curve) and quasars (blue curve) compilations from the literature. The vertical dashed black line hightlights the ratio threshold allowing to recover 86% of the know quasars with 28% of contamination from UCD. The internal panel shows the zoom-in distribution at $\chi^2_{\rm BD}/\chi^2_{\rm quasars} < 6$. 78

2.10 Example of an SED fitting result and the photometric redshift probability distribution function (*top panels*). LS DR10, VHS DR7 and WISE postage stamps of the photometry used (*bottom panel*) for the source LS J000-79. *Top left*: The blue curve represents the best-fit model given by the QSO1 template (see Salvato et al. 2022 for details) at redshift 6.096, while the red curve shows the best-fit brown dwarf template from Meisner et al. (2021). The observed photometry represented by black squares with error bars correspond to the DECam, VHS and WISE catalogs and their uncertainties, while the yellow circles are the expected flux densities assuming the best quasar model. *Top right*: photometric redshift probability distribution function. Despite the distribution extending over $z = [0, 12]$, we limited the plot to the range $z = [4, 9]$ for better visualization. 79

2.11 Optical discovery spectra of the 16 new quasars reported in this work, including the quasar LS J1451-0445 from the DESI data release (Abdul-Karim et al. 2025). The black curves show the observed spectra, the red curves the best-fit quasar templates used for redshift estimation, and blue dashed lines indicate the adopted best-fit position of Ly-α. Each panel lists the source name, spectroscopic redshift, and the best-fit template in the upper left corner. The typical redshift uncertainty is 0.03 (see Section 2.3.1). Quasars are sorted from top to bottom by decreasing redshift. 85

2.12 Continuation Fig. 2.11. 86

2.13 Normalized EW(Ly α+ N v) distribution for the 16 quasars at $z = 5.93 - 6.45$ identified through our CL selection. The solid green and dashed red curves show the best-fit log-normal distributions for PS1 quasars (Bañados et al. 2016) and SDSS quasars (Diamond-Stanic et al. 2009), with mean values of log(EW/Å) = 1.542 and 1.803, and standard deviations of 0.391 and 0.205 Å, respectively. The vertical dashed gray line marks the limit for WLQs (EW $<$ 15.4 Å; Diamond-Stanic et al. 2009). While most of our sample overlaps with the PS1 distribution at log(EW/Å) $<$ 2, a substantial excess appears in the high-EW tail, potentially revealing a distinct population of strong-line quasars. 91

2.14 Color–color diagrams showing the positions of the newly discovered quasars (stars), and known high-redshift quasars at the same redshift range of our discoveries $5.94 \leq z \leq 6.45$ (filled circles) both color-coded by redshift; and L, and T dwarfs (black and red crosses, respectively) used as labels in the embedded space. Top: $i - z$ vs. $z -$ W1. Bottom: $z - J$ vs. W1 − W2. While $i - z$ vs. $z -$ W1 shows a total overlap of the new quasars with the T dwarf population, the $z - J$ color shows a clear distinction. 93

3.1 Light curves of brown dwarf SIMP0136 (spectra type T2.5) observed in the J-band with the 1.6-m telescope at the Observatoire du Mont-Mégantic over four epochs (adapted from Artigau et al. (2009)); and of WISE 0855-0714, a Y dwarf, monitored with Spitzer/IRAC in bands 1 and 2 at 3.6 μm and μm, respectively (taken from Esplin et al. (2016)). The cyan and black curves show the best-fit models: a two-harmonic sinusoidal model with a period of 2.388 h for SIMP 0136 (*upper panel*), and double sine model with periods of 9.7 and 10.8 h for IRAC bands 2 and 1, respectively (*lower panel*). 103

3.2 Structure function parameters A$-\gamma$ plane for $\sim$ 60-epoch light curves of $\sim$ 15000 sources from SDSS stripe 82 survey. Spectroscopically confirmed quasars and RR Lyrae are shown as blue and red points, respectively, while color-selected F and G stars as gray points. The solid lines highlight the limits to select quasars with a quasar completeness of 93%. The histograms of A and γ values for the three populations in colors are presented along the horizontal and vertical axis, respectively. Figure taken from Schmidt et al. (2010a). 104

3.3 Mean IR magnitudes versus redshift for the 285 quasars with at least 15-epoch unWISE light curves. Band 1 and 2 mean measurements are represented by black and blue filled circles, respectively. As the sample is dominated by bright sources ($<$ 20) in both bands, low-amplitude variability is expected. 107

3.4 W1 light curve of the quasar J233008.71+095743.70 at z = 5.3 (red dashed line with black squares) and 10 calibrating stars (gray dashed lines with diamonds) from the field within a 0.1° radius. While the light curves of the stars exhibit fluctuations in the order of 0.1 magnitude between consecutive epochs, the one of the quasar shows a prominent peak of 1 magnitude likely to trigger some of the variability parameters. 108

3.5 Distributions of variability features with discriminative potential from band 1 (*top*) and band 2 (*bottom*) light curves. Gray filled histograms show the stellar distributions, and blue lines the quasar distributions. From left to right: $\log_{10}$(`ExcessVar`), $\log_{10}$(`MeanVariance`), `Amplitude`, and `Std`. In W1, quasars exhibit a systematic shift toward larger values and extended high-value tails, indicative of strongly variable sources. In contrast, the W2 distributions largely overlap with those of stars, reflecting the reduced sensitivity of this band to variability. 114

3.6 Variability features as a function of redshift from top to bottom: `iqr`, `Amplitude`, `MaxSlope`, `MedianAbsDev`, `PercentAmplitude`, `Std`, and `ExcessVar`. Red points represent quasars deviating by more than 1σ from the mean distribution, while gray points correspond to the remaining quasar population. Most methods identify more than 20 quasars with significant variability. Amplitude- and standard deviation–based features show the largest deviations at low redshift and a decreasing trend with redshift, possibly indicating a redshift dependence. 115

3.7 Variability features as function of the mean light curve magnitude for the overall distribution of calibrator stars: `iqr`, `MaxSlope`, `Std`, and `ExcessVar`. The first three panels show a clear positive correlation with magnitude, indicating that increasing photometric uncertainties can replicate variability signatures in otherwise non-variable stars. In contrast, the excess variance displays both higher positive and negative amplitudes for the intrinsic variability of faint sources. 117

3.8 `iqr_W1` values for quasars (red circles) that exceed their local variability baseline, compared with calibrator stars (gray crosses with 1σ error bars). The interquartile range quantifies variability as the difference between the 3rd and 1st quartile magnitudes. The dashed line marks the $\mu + 3\sigma$ threshold of the global calibrator star distribution. 119

3.9 W1 light curve of the false positive variable quasar J030642.51+185315.82 at $z = 5.36$ (red dashed line with black squares) and 10 calibrating stars (gray dashed lines with diamonds) from the field within a 0.1° radius. A prominent peak is produced by one epoch with two different measurements of the magnitude, while the remaining light curve seems nearly constant. This phenomenon is seen not only in the quasar light curve but also in one of its calibrator stars. 120

3.10 Power spectral density of the quasar `J135649.41-264230.23` at $z = 5.72$ (blue points), with the median-binned PSD over 20 bins (black curve) and the corresponding window function (gray curve). The decline at $\log_{10}(f) < -2.2$ traces intrinsic quasar variability, while multiple peaks at higher frequencies coincide with those in the window function and are therefore attributed to sampling artifacts. Then our analysis is restricted to the low-frequency regime, with the upper frequency limit set, source-by-source, at the point where the PSD begins to follow the window function. 124

3.11 Observed-frame W1 light curves of three representative variable quasars illustrating the diversity of PSD behaviors. *Top*: `J135649.41-264230.23` ($z = 5.72$) shows red-noise–like variability with a high amplitude (> 0.5 mag) over long timescales. *Middle*: `J040819.23-563228.82` ($z = 6.10$) displays a repeating rise–decline pattern of nearly constant amplitude, producing a strong PSD peak at $\log_{10}(f) = -2.55$. *Bottom*: `J151444.91+211419.84` ($z = 5.91$) exhibits a flat, white-noise–like PSD at the first epochs, and later two cycles that seem superimposed on a decreasing magnitude trend. 127

4.1 Examples of the model templates included in AGNFITTER-RX for the physical components responsible for the emission of the host galaxy and its star-forming regions. At radio frequencies, the host-galaxy SED is dominated by the synchrotron emission associated with the diffusion of cosmic rays from SNR and PWN acceleration sites through the galaxy ISM, and the IR emission is dominated by the cold dust emission (green lines), both of which are associated with regions of high star-formation. The NIR–optical–UV emission is dominated by the stellar emission (yellow lines). The upper panel shows a subsample of semi-empirical templates by Chary & Elbaz (2001) and Dale & Helou (2002) for the emission from star-forming regions and the stellar population synthesis models by Bruzual & Charlot (2003) for SFH of 0.05×10^9 yr, and different ages and metallicities. The lower panel shows theoretical models for cold dust by Schreiber et al. (2018) for increasing dust temperatures and $f_{\rm PAH}$ smoothly joined to the radio emission estimated with the FIR-radio correlation (eq. 4.1). . . . 136

4.2 Examples of the model templates included in AGNfitter-rx for three distinct physical components of the nuclear emission (AGN) in active galaxies. At radio frequencies, the nuclear SED is dominated by the synchrotron emission (light blue curves) associated with jets and core emission at $\sim 10^{-1} - 10^{6}$ parsec scales. The MIR emission is dominated by the hot dusty "torus" emission (purple curves) arising at ~ 10 pc scales. The optical-UV-X-ray emission is dominated by the accretion disk and X-ray corona emission (dark blue curves) at $\sim 10^{-2}$ pc scales. The first panel shows a SPL in radio, a subsample of torus templates by Silva et al. (2004) with different column densities, and the accretion disk model by Richards et al. (2006b) with an increasing reddening parameter. The second panel presents SPL emission with different slopes, torus models by Nenkova et al. (2008) for different inclination and opening angles, and optical depth of 10, and accretion disk models by Slone & Netzer (2012) for diverse black holes masses and accretion rates. 139

4.3 Continuation Fig 4.2. The first panel shows a DPL with $curv$ and $\log \nu_t$ as free parameters; torus templates by Stalevski et al. (2016) with optical depth of 3 and diverse inclination and opening angles; and accretion disk models by Kubota & Done (2018) defined by black holes masses and accretion rates. The second panel shows a DPL with $\log \nu_t = 11.5$ Hz and different values of α_1 and α_2; torus model by Hönig & Kishimoto (2017) depending on inclination angles, index of cloud distribution and cloud fractions; and the accretion disk template by Temple et al. (2021b). 140

4.4 Examples of the best SED fittings of the different AGN types: The blazar Mrk 421 (top panel), the Seyfert1 PG0052+251 (middle panel), and the Seyfert 1.5 Mrk 290 (bottom panel). The yellow and green solid curves show the emission from the stellar population and cold dust of the galaxy, respectively. Purple, dark blue, and light blue solid curves show the torus, accretion disk, and radio AGN emission models, respectively. Ten models constructed from combinations of parameters randomly selected from the posterior PDFs (hereafter referred to as realizations) are plotted as shaded areas and the residuals of each fit realization are presented in the graphs below each source. A flexible energy balance prior was assumed so that the cold dust IR emission is at least comparable to the dust-absorbed stellar emission. 155

4.5 Continued from Figure 4.4. Examples of the best SED fittings for the blazar WCom, the Seyfert 1 Mrk876, the Seyfert 1.5 NGC7469 and the Seyfert 2 Ark 564. 156

4.6 Performance analysis of different torus models in SED fitting: The models presented are the homogeneous S04 (upper panel), the clumpy and windy CAT3D (central upper panel), the clumpy NK08 (central lower panel), and the two-phase SKIRTOR model (lower panel). The left panel shows the histograms of the logarithm of the likelihood values corresponding to each model and the dashed line in each plot indicates the median value of the distribution. Residuals from the best fits for each galaxy of the sample are shown in the right panel. The plot shows the existence of templates that manage to capture some SED features (high density of points around zero) while systematically failing in modeling specific regions of the observed SED (high dispersion). The dash-dotted line shows the 10 μm feature and the dashed lines show the limits of 0σ, 3σ, and -3σ. 158

4.7 $\alpha_{\rm MIR}$-$\alpha_{\rm NIR}$ spectral slope space covered by the different torus models included in AGNFITTER-RX with different complexity levels. The gray circles show estimates of the spectral slope covered by our sample of AGN, computed with *Spitzer* and WISE observed photometry. 159

4.8 Performance analysis of different accretion-disk models in SED fitting: The models shown are the semi-empirical R06 (upper panel), the semi-empirical with emission lines THB21 (central upper panel), the theoretical α-disk SN12 (central lower panel), and the theoretical three-component KD18 model (lower panel). The left panel shows the histograms of the logarithm of the likelihood values from each model and the dashed line in each plot indicates the median value of the distribution. Residuals from the best fits for each active galaxy of the sample are shown in the right panel. The effect of overlapped emission lines, such as the doublet of [N II] $\lambda\lambda$ 6549,6585 Å and H α λ 6563 Å, is large, reducing residuals to about zero. The dash-dotted line indicates the 0.65 μm broad emission lines feature and the dashed lines the limits of 0σ, 3σ, and -3σ. 164

4.9 $\alpha_{\rm ox}$-$\alpha_{\rm B\text{-}V}$ spectral slope space covered by the different accretion disk models included in AGNFITTER-RX. The gray circles show estimates of the spectral slope covered by our sample of AGN, computed with *Swift*, GALEX and *XMM-Newton* observed photometry. 165

4.10 Normalized histogram of the total distribution of inclination angle estimates by the NK08 (left panel), SKIRTOR (central panel) and CAT3D (right panel) model for the 36 galaxies of the sample classified as Seyfert 1 (light blue) and Seyfert 2 (magenta). The dotted contours represent the distribution of the angles for the one-parameter model, while the solid histograms represent the three-parameter models. The solid vertical lines highlight the median values of the distributions. 169

4.11 Representation of the upper right quadrant of the cross-section of the toroidal distribution of hot nuclear dust in NK08 (upper panels) and SKIRTOR (lower panels) models. The shaded color areas with decreasing transparency represent the 16th, 50th and 84th percentiles of the torus opening angle in Seyfert 1 (blue) and Seyfert 2 (magenta) AGNs. While the black dotted lines represent the median values of the viewing angle and the dotted gray shaded areas are the 16th and 84th percentiles. 171

4.12 AGN fraction contributing to the total IR luminosity ($8 - 1000\,\mu$m) for the 36 galaxies of the sample and each torus model tested. The gray area highlights the SEDs with missing data in the FIR. The different models used do not have a large effect on the overall AGN fraction in this range. . 173

4.13 Comparison of the log-scale black hole mass estimates made via AGNfitter-rx using KD18 and SN12 accretion disk models (y-axis) and broad Hα emission line (x-axis) for 17 AGNs in our sample with available information in BASS DR2 catalog. The dotted line corresponds to the 1-to-1 relation. 174

4.14 Bolometric luminosities are computed by adding integrated luminosities of the AGN components vs bolometric luminosities based on the inferred black hole mass and accretion rate by the accretion disk models. The gray dashed line indicates the one-to-one ratio and the color bar is the accretion disk reddening parameter. 177

4.15 AGN fraction contributing to the total optical luminosity ($400-500$nm) for the 36 galaxies of the sample and each accretion disk model tested. The flexibility in modeling X-rays and the effects of UV degeneracy causes outliers more often with the KD18 accretion disk model. 178

4.16 AGN fraction distribution as a function of frequency for Sy1 (upper panel) and Sy2 (lower panel) populations. The white dots within the violin plots highlight the median of the distribution while the black line is the range between the 25th and 75th percentiles. The blue and red curves present the average radio-to-X-ray SED, including BC03_metal, S17, CAT3D, THB21, and combinations of SPL and DPL radio models, for type 1 and type 2 AGNs, respectively. 179

A.1 $r - i$ vs. $i - z$ plane for high-redshift quasars and the main contaminants in color selections. Quasars at $4 < z < 7.2$ are shown as green filled circles, low-redshift quasars ($z < 1$) as blue filled circles, ultracool dwarfs and main-sequence stars as red crosses, mildly dust-reddened galaxies as yellow empty circles, and dusty galaxies up to $z \sim 3$ as red open diamonds. Spectroscopically confirmed quasars at $z > 5.3$ from Fan et al. (2023) are marked with black filled diamonds. Black lines indicate the adopted color cuts designed to select quasars in three regimes: $z < 5.7$, $5.7 < z < 6$, and $z > 6$. These cuts were applied to candidate selection from DELVE DR2. 192

A.2 Confusion matrix for the best-performing classification of DELVE DR2 sources, obtained with a random forest model (validation entropy score ~ 0.991 with identity transformation). The training set includes six classes: quasars in three redshift bins, ultracool dwarfs, low-mass stars, and luminous red galaxies. The method achieves high accuracy overall, with most predictions lying along the diagonal. The algorithm occasionally face troubles to classify quasars at $z > 5.7$ being assigned to the neighboring $5.7 < z < 6$ category. 194

A.3 Permutation feature importance for the best supervised ML solution corresponding to a random forest classifier. The critical features driving the classification include the $i - z$ and $z - W1$ colors, consistent with expectations from quasar selection in the literature. Spurious importance is assigned to the g-band RMS, which is non-physical given the non-detection of all sources in g, while other useful contributions arise from magnitude variance across multiple exposures tracked by the RMS, and proper motion estimates, which help to flag artifacts and stellar contaminants. 195

A.4 z-band magnitude vs. $i - z$ color diagram for the final sample of 482,754 quasar candidates selected with *Modulos*. The overdensity at $i - z \lesssim 0.5$ suggests a possible high contamination from ultracool dwarfs. 196

B.1 Embedded Euclidean space generated by UMAP for HSC DR3 i and z-dropout sources after training the CL algorithm. All sources are represented as grey countours given by the number density; known M, L and T dwarfs are: brown, black and red crosses, respectively; and Gaia DR3 stars are yellow crosses. Spectroscopically confirmed quasars from the literature are represented by filled circles color-coded by their redshift. Known quasars appear to be on a parameter space distinct from ultracool dwarfs, and following an "evolutionary track" at the extreme left of the embedded space. New quasar candidates are selected from these regions. 199

B.2 Coadded HSC optical cutouts (*left*) and single-epoch Y-band images (*right*) of three 9moving objects located in the $z \gtrsim 6.6$ quasar subregion of the latent space. These sources, initially resembling potential lensed quasar candidates, are instead consistent with asteroids. 200

C.1 Comparison of K (*top*) and $W1$ (*bottom*) reprojection stamps computed with different methods. From left to right: original downloaded cutouts, cropped 10″ stamps, adaptive reprojection with a Hann kernel, adaptive reprojection with a Gaussian kernel, interpolation-based reprojection, and exact reprojection. For VHS (K band), the differences among methods are minimal. In contrast, the lower resolution of WISE relative to LS DR10 makes clearly visible the impact of the different approaches on the shape of the point-like source. 203

C.2 Embedded Euclidean space generated by UMAP for LS DR10 i-band dropout sources after training the CL algorithm with optical plus VHS J and K imaging (*top*) and with the addition of the $W1$ band (*bottom*). Source number densities are shown as grey countours; known M, L and T dwarfs are marked as brown, black and red crosses, respectively; and Gaia DR3 stars as yellow crosses. Spectroscopically confirmed quasars from the literature are represented by filled circles, color-coded by their redshift. Unlike representations derived from optical-only images, quasars here are more dispersed and overlap significantly with stars and T dwarfs. 204

D.1 Structure function parameter space (`SF_ML_amplitude` vs. $\log_{10}$(`SF_ML_gamma`)) for quasars (blue) and stars (gray). Horizontal bands at `SF_ML_amplitude` $\sim$ –0.5 and 3, and diagonal loci where amplitude and slope correlate linearly, are artifacts of the sparse sampling and limited number of epochs in unWISE light curves. These features indicate fitting degeneracies rather than intrinsic variability, and result in significant overlap between quasars and stars. As a consequence, the SF model provides no discriminative power for our dataset. 207

Resumen

Los cuásares son las fuentes no transitorias más luminosas del universo y proporcionan una visión única del crecimiento de los agujeros negros supermasivos (SMBH), la evolución de las galaxias y las propiedades del medio intergaláctico (IGM). Su descubrimiento en los primeros miles de millones de años despúes del Big Bang (con corrimientos al rojo $z > 6$) es retador, ya que son poco frecuentes y difíciles de distinguir fotométricamente de contaminantes mucho más numerosos. La población actual, compuesta principalmente por cuásares brillantes y azules, se concentra en el hemisferio norte.

Esta tesis explora enfoques en múltiples longitudes de onda y el dominio temporal para la selección y caracterización de cuásares en $z > 4$, utilizando datos desde radio hasta rayos X. Probamos cortes de color, aprendizaje automático (supervisado y auto-supervisado), ajuste de la distribución espectral de energía (SED) y análisis de variabilidad. Un método de aprendizaje por contraste aplicado a las imágenes del DESI Legacy Survey DR10 logró una tasa de éxito del 45%, dando lugar a 16 nuevos cuásares confirmados espectroscópicamente en $z \gtrsim 6$. El análisis de variabilidad de 285 cuásares en $z > 5.3$ con curvas de luz de unWISE identificó 19 con variabilidad significativa (3σ), convirtiéndose en una de las primeras detecciones de variabilidad óptica, en el sistema de referencia en reposo, más allá de $z \sim 5$.

También desarrollamos AGNFITTER-RX, una herramienta bayesiana de ajuste de SED extendida desde radio hasta rayos X. Su aplicación a una muestra de AGN en $z < 0.7$ demuestra una caracterización física fiable y consistente con análisis espectroscópicos. En conjunto, estas herramientas mejorarán el descubrimiento y la caracterización de cuásares

de alto desplazamiento al rojo en la próxima era de Rubin/LSST, Euclid, Roman, JWST, 4MOST y DESI.

Palabras clave: cuásares: alto desplazamiento al rojo – galaxias: activas – cuásares: variabilidad – técnicas: fotométricas – métodos: estadísticos

Abstract

Quasars are the most luminous, non-transient sources in the Universe and provide unique insight into supermassive black hole (SMBH) growth, galaxy evolution, and intergalactic medium (IGM) properties. Their discovery within the first Gyr ($z > 6$) is challenging because they are rare and difficult to distinguish photometrically from much more numerous contaminants. The current population, mainly bright and blue quasars, is concentrated in the northern hemisphere.

This thesis explores multiwavelength and time-domain approaches for quasar selection and characterization at $z > 4$, using data from radio to X-rays. We tested color cuts, machine learning (supervised and self-supervised), spectral energy distribution (SED) fitting, and variability analysis. A contrastive learning method applied to DESI Legacy Survey DR10 imaging achieved a 45% success rate, leading to 16 new spectroscopically confirmed quasars at $z \gtrsim 6$. Variability analysis of 285 $z > 5.3$ quasars with unWISE light curves identified 19 with significant (3σ) variability, marking one of the first detections of rest-frame optical variability beyond $z \sim 5$.

We also developed AGNfitter-rx, an extended Bayesian SED-fitting tool from radio to X-rays. Its application to a $z < 0.7$ AGN sample demonstrates reliable physical characterization consistent with spectroscopic analyses. Together, these tools will enhance high-redshift quasar discovery and characterization in the upcoming era of Rubin/LSST, Euclid, Roman, JWST, 4MOST, and DESI.

Key words: quasars: high-redshift – galaxies: active – quasars: variability – techniques: photometric – methods: statistical

Zusammenfassung

Quasare sind die leuchtstärksten, nicht-transienten, astronomischen Quellen im Universum und bieten einzigartige Einblicke in das Wachstum supermassereicher Schwarzer Löcher (SMBH), die Galaxienentwicklung und die Eigenschaften des intergalaktischen Mediums (IGM). Die Suche nach Quasaren in den ersten Milliarden Jahren nach dem Urknall –bei Rotverschiebung $z > 6$– ist eine Herausforderung, da sie äußerst selten sind und photometrisch nur schwer von den viel zahlreicheren Kontaminanten zu unterscheiden sind. Die derzeit bekannte Population,überwiegend helle und blaue Quasare, konzentriert sich auf die nördliche Hemisphäre.

Diese Arbeit exploriert multiwellenlängen- und zeitbereichsbasierte Ansätze zur Suche und Charakterisierung von Quasaren bei $z > 4$, unter Verwendung von Daten vom Radiobereich bis zu Röntgenstrahlen. Getestet wurden photometrische Farbselektion, maschinelles Lernen (überwacht und selbstüberwacht), Spektralenergieverteilungs-Anpassung (SED) und Variabilitätsanalyse. Eine kontrastive Lernmethode, angewendet auf DESI-Legacy-Survey-DR10-Bilddaten, erreichte eine Erfolgsrate von 45 % und führte zur spektroskopischen Bestätigung von 16 neuen Quasaren bei $z \gtrsim 6$. Die Variabilitätsanalyse von 285 Quasaren mit $z > 5.3$ auf Basis von unWISE-Lichtkurven identifizierte 19 Objekte mit signifikanter (3σ) Variabilität und stellt damit eine der ersten Entdeckungen von optischer Variabilität im Ruhesystem jenseits von $z \sim 5$ dar.

Darüber hinaus wurde AGNfitter-rx entwickelt, ein erweitertes bayessches SED-Anpassungswerkzeug vom Radio bis zum Röntgenbereich. Die Anwendung auf eine Testpopulation Aktiver Galaxienkerne (AGN) mit $z < 0.7$ zeigt eine zuverlässige physikalische

Charakterisierung, die mit spektroskopischen Analysen übereinstimmt. Zusammen werden diese Werkzeuge die Entdeckung und Charakterisierung von Quasaren mit hoher Rotverschiebung in der kommenden Ära von Rubin/LSST, *Euclid*, *Roman*, *JWST*, 4MOST, und DESI deutlich verbessern.

Schlüsselwörter: Quasare: hohe Rotverschiebung – Galaxien: aktiv – Quasare: Variabilität – Techniken: photometrisch – Methoden: statistisch

Chapter 1

Introduction

1.1 Active galactic nuclei

Active galactic nuclei (AGN) are bright, compact structures at the centers of massive galaxies (< 10 pc; Netzer 2015), powered by the accretion of matter onto supermassive black holes (SMBHs, M• > 10^5 M⊙; Netzer 2015). They produce luminous emission (with bolometric luminosities up to $L_{\rm bol} \approx 10^{48}$ erg s^{-1}; Padovani et al. 2017) spanning the entire electromagnetic spectrum from radio to X-rays (Blandford et al. 1990; Peterson 1993; Urry & Padovani 1995; Netzer 2015; Padovani et al. 2017), and even into the γ-ray regime in the case of blazars (e.g., Ackermann et al. 2011). AGNs were first identified through their broad optical emission lines (Seyfert 1943), revealing gas clouds rotating at high velocities (1000–5000 km s^{-1}; Carroll & Ostlie 2017). Since then, selection techniques have expanded to every spectral band, each dominated by different physical processes. The growing number of detections, combined with significant discrepancies among AGN properties, has led to the emergence of more than 15 main classifications and up to 50 subcategories (e.g., Padovani et al. 2017).

The AGN unification scheme seeks to explain the wide variety of observed AGN properties through a universal set of physical components, governed by a small number of parameters such as viewing angle, accretion rate, and black hole mass (Netzer 2015). Its

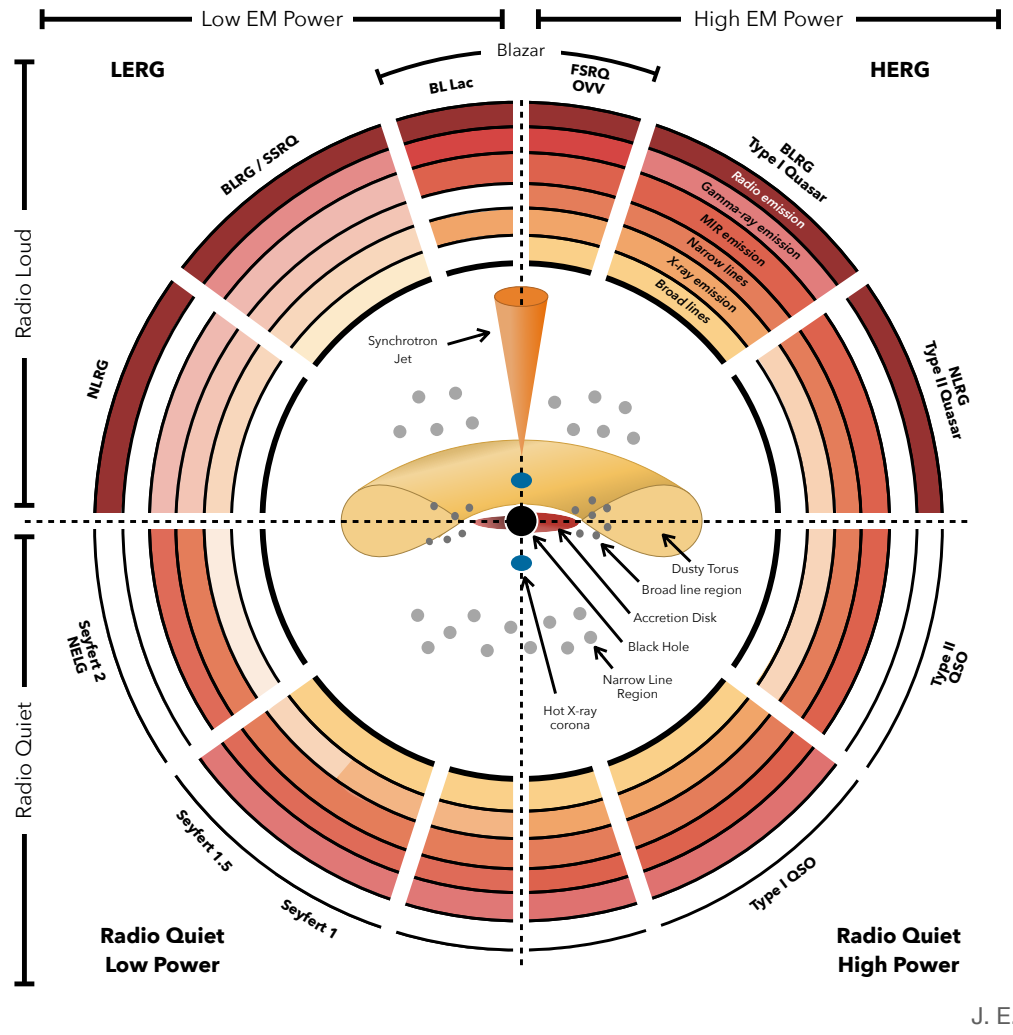

Figure 1.1: Schematic representation of the unified AGN model from Thorne et al. (2022), illustrating the classification framework based on viewing angle, radio emission strength, and accretion rate (which traces the electromagnetic power). The upper left and right quadrants correspond to low- and high-excitation radio galaxies (LERGs and HERGs), typically associated with the kinetic (radio) and radiative (quasar) modes, respectively. The lower left and right quadrants represent radio-quiet Seyfert galaxies and quasars. Common AGN subclasses are labeled accordingly. Colored bars indicate the expected presence and relative strength (transparency) of broad or narrow emission lines, and of MIR, radio, X-ray, or γ-ray emission.

key ingredients include: an optically thick accretion disk with a hot, optically thin corona above and below it; dense, high-velocity gas clouds within ∼1 pc of the SMBH forming the broad-line region (BLR); a toroidal distribution of hot dust around the disk; a more extended, low-density ionized gas on kpc scales that produces narrow emission lines; and sometimes radio jets. Fig. 1.1 illustrates these components, along with the primary AGN classifications based primarily on viewing geometry, radio power, and dominant energy output. The latter divides AGNs in two main groups: the radiative (or quasar) mode, in which most of the energy is emitted as multiwavelength electromagnetic (EM) radiation (high EM power in Fig. 1.1); and the kinetic (or jet) mode, where the output is dominated by mechanical feedback from relativistic jets (Fabian 2012) (low EM power). Radiative-mode AGNs are very efficient accretors with Eddington ratios $\lambda_{\rm Edd} \gtrsim 0.01$, whereas jet-mode AGNs typically exhibit much lower accretion rates Netzer (2015). However, this distinction becomes less clear at high redshift, where powerful jets are often found in sources with high accretion rates, suggesting that both modes can coexist in the early Universe (e.g.; Ghisellini et al. 2014; Sbarrato et al. 2021; Ighina et al. 2025).

AGN can be broadly classified based on their emission properties, accretion power, and radio morphology. Seyfert galaxies are lower-luminosity AGNs, with types 1, 1.2, 1.5, and 2 reflecting the decreasing visibility of their broad-line regions. LINERs (Low-Ionization Nuclear Emission-line Regions) show spectra dominated by low-ionization lines. Blazars are jet-dominated AGNs viewed close to the jet axis, with strong variability and non-thermal emission; BL Lacertae objects are a subclass with very weak or absent emission lines. In the radio, AGNs are also categorized by their jet morphology as Fanaroff–Riley (FR) type I (bright core, diffuse low-luminosity jets) or FR type II (bright hotspots in the lobes, high-luminosity collimated jets) (Fanaroff & Riley 1974). Together, these classifications reflect the diversity of accretion rates, orientation effects, and feedback modes in the AGN population, providing a framework to situate quasars as the most radiatively efficient and luminous AGNs.

Quasars populate the high EM power (radiative) regime in the AGN unification frame-

work. They are the most luminous non-transient sources in the Universe, outshining their host galaxies and dominating emission across the electromagnetic spectrum. Because their radiation is concentrated on the compact nuclear region, quasars appear point-like or quasi-stellar in optical images. Optical surveys are inherently biased toward bright, unobscured systems (with Eddington luminosity ratio of $L/L_{\rm Edd} > 0.01$; Padovani et al. 2017), which has made quasars one of the most extensively studied AGN classes, particularly in the context of evolutionary studies.

1.2 Quasars in the reionization era

Their extreme luminosities and the central role of SMBH accretion in galaxy evolution make quasars powerful probes of the early Universe and the interplay between black holes and their hosts. The energy and momentum released by quasars strongly interact with the surrounding gas, influencing the evolution of their host galaxies and the environment (Fabian 2012). This phenomenon, known as AGN feedback, is crucial in cosmological simulations to reproduce realistic galaxy population properties (e.g., Kauffmann & Haehnelt 2000; Granato et al. 2004; Di Matteo et al. 2005; Springel et al. 2005; Kaviraj et al. 2017; Bustamante & Springel 2019). Empirical correlations between SMBH mass and galaxy properties, such as stellar bulge mass, velocity dispersion and luminosity support a co-evolutionary scenario between quasars and their hosts (e.g., Tremaine et al. 2002; Marconi & Hunt 2003; Häring & Rix 2004; McConnell & Ma 2013; Kormendy & Ho 2013).

However, observations of some high-redshift quasars suggest that this relation may evolve with redshift, although the nature and extent of this evolution remain uncertain. In particular, a subset of SMBHs at early cosmic times seems to be more massive than expected from the local M•–M$*$ correlation (e.g., Wang et al. 2016b; Decarli et al. 2018; Harikane et al. 2023), potentially suggesting a faster growth of black holes relative to their host galaxies, or the existence of more massive black hole seeds. Expanding the quasar population at $z > 5$ and pushing the magnitude frontier to fainter sources, along with a

comprehensive physical characterization is essential to test these coevolution scenarios at early cosmic times.

Cosmological simulations indicate that at $z \sim 6$, bright and massive black holes are located in the most massive dark matter haloes inside overdense regions, likely evolving into galaxy clusters (Costa et al. 2014; Overzier et al. 2009). Observational results are not so conclusive. On one hand, some HST, ALMA, MUSE, and JWST observations have revealed overdensities in the field of high-z quasars by selecting galaxies via photometric selection (e.g., Utsumi et al. 2010; Morselli et al. 2014; Mignoli et al. 2020; Overzier 2022) or bright [CII] (e.g., Decarli et al. 2017; Neeleman et al. 2019; Venemans et al. 2020), Lyman-alpha (e.g., Meyer et al. 2022) or [OIII] emission (e.g., Wang et al. 2023). These companion galaxies are likely located in gas-rich environments and, as found by Wang et al. (2023), can be part of filamentary structures. On the other hand, other observations of quasar environments have found no overdensities and even underdensities (e.g., Bañados et al. 2013; Simpson et al. 2014; Mazzucchelli et al. 2017; Eilers et al. 2024; Kim et al. 2009). Diverse and larger quasar samples, as well as systematic characterization of their environments are needed to assess these trends statistically and determine whether bright quasars are indeed reliable signposts of large scale structure formation (Fan et al. 2023).

Quasars at $z \gtrsim 5$ serve as powerful backlights for studying the intergalactic medium (IGM), offering key insights into its ionization state and chemical enrichment during the epoch of cosmic reionization. At these redshifts, the fraction of neutral hydrogen in the IGM is large enough to produce strong absorption of photons with energies above the ionization threshold (Lyman limit at 13.6 eV or 912 Å). Besides the continuum absorption, hydrogen atoms also absorb photons matching the energies of the electronic transitions in neutral hydrogen (Lyman series), with the Lyman-alpha (Ly α) transition at 1216 Å being the most probable due to its resonant nature. This creates the Lyman-alpha forest, a series of absorption lines in quasar spectra between $912 \text{ Å} < \lambda < 1216 \text{ Å}$.

The evolution of the Ly α forest provides a powerful timeline for tracing cosmic reionization. The transition from transmission spikes of the forest to dark gaps, known as

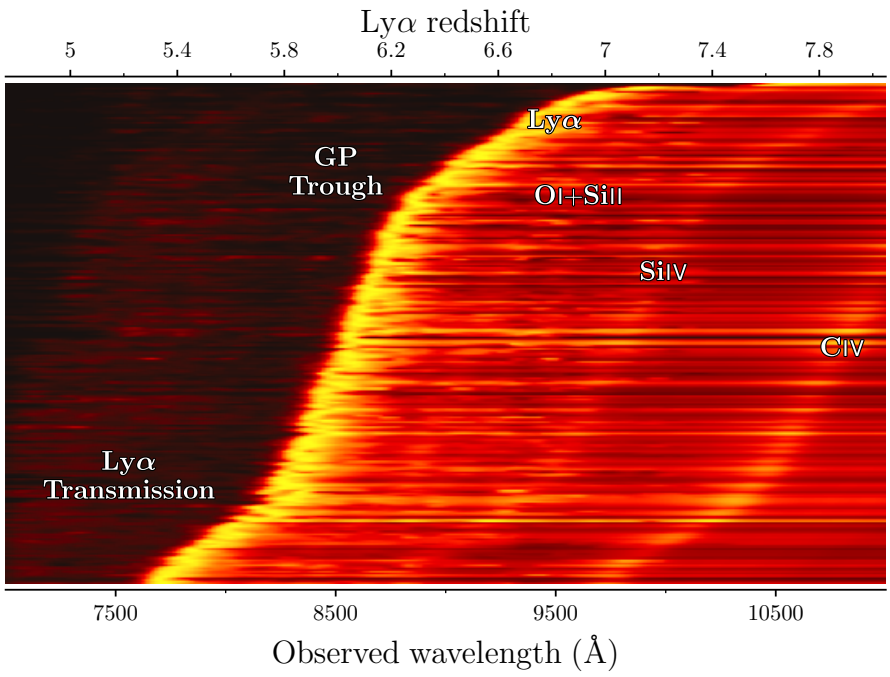

Figure 1.2: Two-dimensional intensity map compiling the published spectra of $z \gtrsim 5.3$ quasars, sorted by increasing redshift from bottom to top. Emission lines such as Ly α, O I+ S III, S IV and C IV; along with the Ly α forest and the GP troughs are labeled. The progressive transition from transmission spikes to GP troughs with increasing redshift traces the evolution of the cosmic reionization process. Figure taken from Fan et al. (2023).

Gunn-Peterson (GP) absorption troughs, indicates that the Universe is almost totally ionized by $z \sim 5.3-6$ (e.g., Bosman et al. 2018; Fan et al. 2023). Fig. 1.2 presents an intensity map compiling the published spectra of quasars at $z \gtrsim 5.3$ from Fan et al. (2023) review, highlighting key spectral features: the Ly α emission line, the Ly α forest transmission spikes, and the GP troughs, as well as prominent metal lines such as O I+ S III, S IV and C IV. The redshift evolution of these features is evident, with the Ly α forest dominating at lower redshifts and the GP troughs becoming increasingly prevalent toward higher redshift. In addition to probing cosmic reionization history, quasar spectra allow to track the chemical evolution and metal enrichment of the IGM through absorption lines from heavy elements (e.g., Ryan-Weber et al. 2006; Becker et al. 2009; D'Odorico et al. 2013; Matejek

& Simcoe 2012; Chen et al. 2017; Cooper et al. 2019).

1.3 Traditional catalog-based selection of high-z quasars

Quasar spectra exhibit a strong flux decrement blueward of rest-frame $\sim$ 1216,Å, produced by absorption from neutral hydrogen in the IGM through the Lyman series and Lyman continuum. At $z \gtrsim 3$, this Ly α forest becomes detectable in optical broad-band photometry. By $z \gtrsim 5$, the increasing neutral fraction of the IGM causes these absorption lines to blend together, producing a distinctive Ly α break. Thus, quasars at high-redshift appear undetectable or faint in photometric filters covering wavelengths below Ly α, creating red colors in photometric indices involving such filters: this is the basis of the photometric 'dropout' selection technique (e.g., Warren et al. 1987; Steidel et al. 1999). However, foreground ultracool M, L, T dwarfs also exhibit red spectra riddled with strong molecular absorption bands that can mimic the dropout signature and a point-like morphology resembling that of quasars. Moreover, while the number density of quasars declines exponentially at $z > 5$ (Jiang et al. 2016), dropping to fewer than one with $M_{1450A} < -26$ per cGpc^3 at $z \sim 6$ (see Fig. 14 in Schindler et al. 2023); M, L and T dwarfs associated with the thin and thick disk remain 2–4 orders of magnitude more numerous (e.g., Barnett et al. 2019), making them the primary astrophysical contaminants in high-redshift quasar searches.

The era of wide-field optical surveys in early 2000s started with the Sloan Digital Sky Survey (SDSS; York et al. 2000), triggering quasar discoveries at $z > 5$ (e.g., Fan et al. 1999, 2001) via the aforementioned color drop-out selection. Deeper surveys such as the Canada-France-Hawaii Telescope Legacy Survey (CFHTLS), Pan-STARRS1 (PS1, Chambers et al. 2016) and Dark Energy Survey (DES, Abbott et al. 2018) significantly expanded the sample (e.g., Bañados et al. 2016, 2023; Jiang et al. 2016; Wang et al. 2019; Reed et al. 2017; Mazzucchelli et al. 2017; Wolf et al. 2024; Belladitta et al. 2019, 2020; Belladitta et al. 2025).

The vast majority of these quasars have been selected by relatively conservative color criteria to minimize contamination. This approach yielded estimated space densities of luminous quasars ($M_{1450\text{Å}} < -26$) of 0.39 cGpc^{-3} at $z \sim 6.7$, based on 17 quasars in the range $6.45 < z < 7.05$ (Wang et al. 2019); and 1.35 cGpc^{-3} at $z \sim 6$, derived from 52 quasars at $5.7 < z < 6.4$ (Jiang et al. 2016). These samples have enabled us to model the bright end of the quasar luminosity function (QLF), which describes the number density of quasars as a function of redshift and luminosity, and provides key constraints on the growth history of SMBHs. As a consequence of the stringent selections, most known quasars at $z > 5$ exhibit similar properties: typically unobscured, with massive black holes ($\sim 4 \times 10^7 \mathrm{M}_\odot$ to $\sim 10^{10} \mathrm{M}_\odot$; Fan et al. 2023) and moderate-to-high accretion rates (median $\mathrm{L_{Bol}/L_{Edd}}$ = 0.79 and 0.72; Fan et al. 2023; Mazzucchelli et al. 2023), leaving potentially important populations underrepresented, such as red or obscured quasars, and objects with unusual emission line properties (e.g., Maiolino et al. 2004; Willott et al. 2007; Matsuoka et al. 2018, 2025b).

Efforts over more limited portions of the sky, such as the Subaru High-z Exploration of Low-Luminosity Quasars (SHELLQs) project ($\sim$ 1100 deg^2; Matsuoka et al. 2016, 2018, 2019, 2025a) have extended the census toward lower luminosities ($M_{1450A} < -22$). The SHELLQs survey revealed that the QLF exhibits a pronounced break and flattens at the faint-end ($M_{1450\text{Å}} \gtrsim -24.9$). When combined with the brighter SDSS quasars, the SHELLQs sample yields a density of 0.90 cGpc^{-3} for luminous quasars ($M_{1450\text{Å}} < -26$) at $z \sim 6.1$, based on 110 quasars in the range $5.7 < z < 6.5$ (Matsuoka et al. 2018). A similar analysis combining SHELLQs with the bright PS1 quasars gives a consistent estimate of 1.15 cGpc^{-3} at $z \sim 5.9$, from 173 quasars at $5.7 < z < 6.2$ (Schindler et al. 2023).

Meanwhile, the James Webb Space Telescope (JWST) is revealing interesting objects such as an AGN hosted by a compact, low-metallicity galaxy at $z \sim 5.55$ (Übler et al. 2023), and a $z \sim 7.15$ offset AGN (Übler et al. 2024). JWST has also uncovered populations of faint and obscured AGNs characterized by V-shaped spectral energy distributions (SEDs), red continua, and broad Balmer emission lines (e.g., Harikane et al. 2023; Matthee et al.

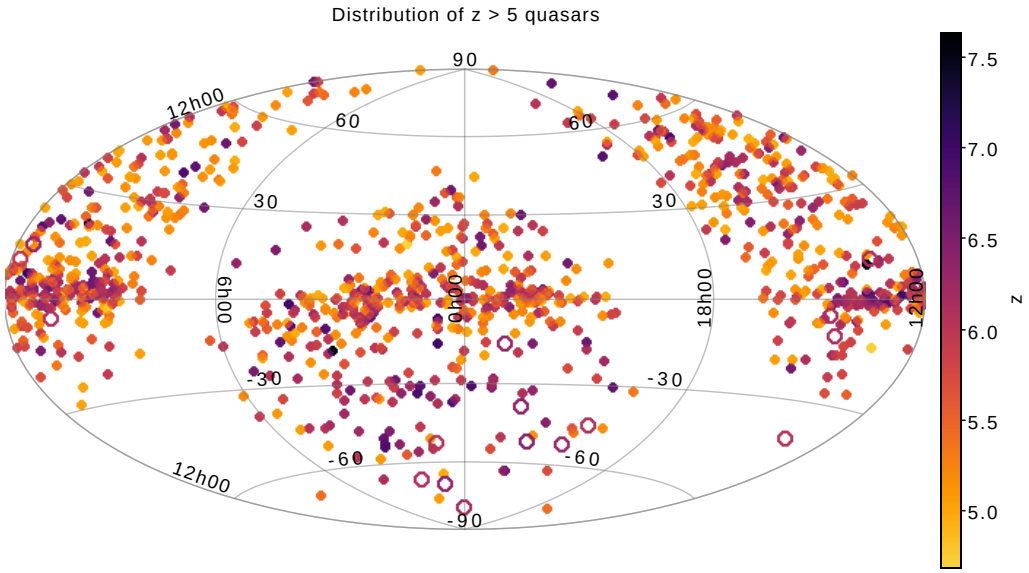

Figure 1.3: Sky distribution in celestial coordinates of spectroscopically confirmed quasars at $z > 5$ up to date (including the new ones from this thesis shown as unfilled circles), color-coded by redshift. Higher concentrations are observed in the Northern Hemisphere and around $dec \sim \pm 5°$, the latter dominated by HSC discoveries of faint quasars.

2024; Greene et al. 2024; Hviding et al. 2025). These objects, nicknamed as 'little red dots' (LRDs), are significantly more numerous than quasars: faint LRDs with $-17 < M_{1450\text{Å}} < -22$ at $4.5 < z < 6.5$ have spatial densities up to ~100 times higher than quasars of comparable luminosity (see Fig. 4 in Kokorev et al. 2024). In contrast, brighter LRDs ($M_{5100\text{Å}} < -23.5$) at $z \sim 4.8$ have densities of ~ 10 cGpc^{-3} (Ma et al. 2025), nearly an order of magnitude lower than that of quasars with similar optical luminosities at $z \sim 5$ (Niida et al. 2020).

The bulk of $z \gtrsim 6$ quasar selections have been done with SDSS, PS1, HSC, and DES photometry, resulting in a distribution of quasars highly concentrated in the Northern Hemisphere (dec > -30), with mostly DES sources populating the south, as illustrated by Fig. 1.3. This imbalance is reinforced by the historically superior multi-object spectroscopic coverage in the north, primarily driven by the dual-beam SDSS spectrographs (e.g., BOSS; York et al. 2000; Dawson et al. 2012) and, more recently, by the Dark Energy Spectroscopic

Instrument (DESI; Abareshi et al. 2022). The high multiplexing capabilities of these instruments (640, 1000, and 5000 fibers for SDSS, BOSS, and DESI, respectively) enable the simultaneous observation and confirmation of thousands of quasars per exposure. In contrast, southern facilities such as the AAOmega spectrograph (392 fibers; Saunders et al. 2004; Sharp et al. 2006) or instruments employing multi-slit masks, including the Gemini Multi-Object Spectrograph (GMOS; Hook et al. 2004), the Cerro Tololo Ohio State Multi-Object Spectrograph (COSMOS; Martini et al. 2014), and the Inamori Magellan Areal Camera and Spectrograph (IMACS; Dressler et al. 2011), offer significantly lower multiplexing efficiency. However, the advent of new-generation instruments such as the 4-metre Multi-Object Spectroscopic Telescope (4MOST; De Jong et al. 2012) at the VISTA telescope and the Multi-Object Optical and Near-infrared Spectrograph (MOONS; Cirasuolo et al. 2014) at the Very Large Telescope, providing up to 2436 and 1000 simultaneous spectra, respectively, will help to homogenize spectroscopically confirmed quasar distribution across hemispheres in the coming years.

Relaxing the traditional selection conditions, such as the requirement for a flat continuum redward of the Ly α break, opens the possibility of identifying new types of luminous quasars, including those with heavy dust obscuration or prominent narrow Ly α emission. Given the low spatial density of quasars expected from luminosity function predictions, such searches must be conducted over wide areas using large-scale photometric surveys. However, the high contamination fraction from stars and artifacts at faint magnitudes requires the development of more sophisticated selection techniques.

1.4 Machine learning for quasar discoveries

Machine learning (ML), particularly traditional supervised methods like Random Forests (RF) and XGBoost, has increasingly been applied to quasar selection, demonstrating strong performance up to $z \sim 5$ (e.g., Nakoneczny et al. 2019; Yèche et al. 2020; Jin et al. 2019; Schindler et al. 2019). Even unsupervised clustering approaches (e.g.,

Logan & Fotopoulou 2020) and comparative studies of unsupervised, semi-supervised, and fully-supervised methods (e.g., Clarke et al. 2020) have been explored for common star/galaxy/quasar classification in large surveys.

Over the past five years, ML techniques have been extended to the high-redshift regime. For instance, Wenzl et al. (2021) applied a 11-class RF classifier to Pan-STARRS1 and WISE data to identify $z > 4.7$ quasars. Their model achieved completeness and efficiency values of up to $66 \pm 7\%$ and $78^{+10}_{-8}\%$ for quasars at $z < 5.6$, and $83^{+6}_{-9}\%$ and $94^{+5}_{-8}\%$ at $z > 5.6$. The resulting catalog contained 515 candidates, including 225 previously known quasars and 193 predicted new ones: 148 expected at $z < 5.6$ and 45 at $z > 5.6$. Spectroscopic follow-up of 37 candidates confirmed 18 new quasars at $z < 5.6$ and 2 at $z > 5.6$.

A complementary implementation by Yue et al. (2023) used a probabilistic RF to search for lensed quasars and quasar pairs at $z \sim 5$, based on optical, near-infrared (NIR), and mid-infrared (MIR) photometry, and combined with morphological constraints and visual inspection downselections. Their campaign identified 88 priority 1 and 136 priority 2 candidates; spectroscopic follow-up of 76 of these led to the discovery of one lensed quasar and one close quasar pair at $z \sim 5$, four quasar systems at $2 < z < 4$, three more reported also by Lemon et al. (2023), and the quasar pair presented in Yue et al. (2021).

An important caveat is that all the above efforts rely heavily on catalog-level photometric measurements and flux ratios, which are sensitive to contamination from artifacts and systematics in the source-extraction process. Moreover, supervised methods are often constrained by the availability of balanced and representative training samples, critical to accomplish at $z > 6$, where spectroscopically confirmed quasars are scarce. Unless complemented or superseded by simulated data (e.g., Yue et al. 2021, 2023), rare populations such as red, or lensed high-redshift quasars are typically underrepresented. As a result, classification reliability and completeness are often poorly assessed and can be strongly overestimated, especially if confounding artifacts are not properly accounted for.

A recent attempt to apply multi-band imaging to the discovery of strongly lensed quasars identified 36 candidates using a 4-class convolutional neural network (CNN)

classifier complemented with spectral energy distribution (SED) modeling preselection and visual inspection downsizing (Andika et al. 2023). As in Yue et al. (2021, 2023), the galaxy-quasar lens training set is built on mock data. The selection claims an impressive low false-positive rate of 2%, however, systems remain at the candidate stage pending spectroscopic confirmation (Andika et al. 2023). Byrne et al. (2024) present the first application in this field of contrastive learning (CL), a fully self-supervised representation-learning method, to optical imaging (DES DR2, Abbott et al. 2021), leading to the discovery of three quasars at $z = 5.94-6.07$. Unlike conventional supervised methods, which require labeled training sets and can suffer from bias, CL learns by contrasting data samples against each other without any prior labels, making it well-suited to uncover rare or atypical quasar populations.

One of the most widely adopted and promising CL architectures in astronomical applications is the Simple framework for Contrastive Learning of visual Representations (SimCLR; Chen et al. 2020). This framework is based on a single CNN encoder without a memory bank, which constrains the training to the batch size rather than the entire dataset. In astronomy, SimCLR has been successfully applied to diverse tasks: for example, Sarmiento et al. (2021) used it to analyze MaNGA-byproduct two-dimensional maps, enabling interpretation of stellar population and kinematic properties of galaxies unhampered by nonphysical effects. Similarly, Eisert et al. (2024) employed CL to compare galaxy representations from the IllustrisTNG (TNG50 and TNG100) cosmological simulations and from real HSC observations, quantifying their resemblance and inferring physical properties of HSC galaxies from simulated ones. In another application, Vega-Ferrero et al. (2024) trained a CL model on mock images from the TNG50 simulation to investigate the morphology of $3 < z < 6$ galaxies observed with JWST/NIRCam as part of the CEERS survey.

Despite these promising applications, machine learning for high-redshift quasar selection still has significant room for growth. The availability of larger, deeper, and expanded multi-wavelength datasets from surveys such as LSST, Euclid, Roman, and JWST will

enable more robust discoveries. Future approaches could combine data types like imaging, spectroscopy, and time-domain data, and explore novel methods such as transformer-based architectures and active learning strategies (e.g., Thuruthipilly et al. 2024; Möller et al. 2025). These innovations have the potential to uncover rare or atypical quasars, improving both completeness and reliability at the highest redshifts.

1.5 Characterization of quasars and active galaxies

A natural step after discovering AGNs is characterization of their physical properties to place them in context, both from physical and demographic standpoints. The precision and extent of this characterization depend strongly on the available data, whether photometric or spectroscopic, and on the wavelength range covered. For large high-redshift samples, panchromatic photometry and rest-frame UV spectroscopy remain the most accessible and commonly used approaches, given their broad availability across wide-area surveys. Nonetheless, complementary techniques such as integral field spectroscopy (IFS) enable spatially resolved studies of AGNs (e.g., Übler et al. 2023, 2024; Travascio et al. 2024; Cazzoli et al. 2022; Venturi et al. 2023), while radio interferometry allows probing of kpc-scale jets in radio AGNs (e.g., Gómez et al. 2016; Spingola et al. 2020; Morganti et al. 2013; Badole et al. 2020; Momjian et al. 2018). At the highest redshifts, IFS, radio, and X-ray observations are also becoming increasingly feasible, though typically limited to small, targeted samples.

In this thesis, we focus on two complementary approaches: panchromatic photometry and rest-frame UV spectroscopy. Section 1.5.1 introduces the SED-fitting methodology, which provides a comprehensive characterization of AGN emission across all components, outlining its scope, limitations, and the current state of theoretical and empirical models within the unified framework. Section 1.5.2 focuses on estimating black hole masses from broad emission lines, a more specialized approach particularly valuable for studying SMBH demographics and obtaining accurate redshifts, which are essential for evolutionary

analyses. The challenges and limitations of applying this technique to high-redshift samples are also discussed.

1.5.1 Photometric SED fitting

The emission from quasars and AGNs spans the entire electromagnetic spectrum, from radio to X-rays, and can extend up to γ-rays in blazars (e.g., Blandford et al. 1990; Peterson 1993; Urry & Padovani 1995; Netzer 2015; Padovani et al. 2017; Ackermann et al. 2011). This emission is multiscale, spanning sub-pc to kpc sizes, and is often difficult to spatially distinguish from the host-galaxy components. Thus, a composite spectral energy distribution (SED) accounting for the galaxy and the AGN combined emission is observed. This makes it challenging to assess the physical properties of both components, limiting our understanding of more complex processes such as black hole growth and the co-evolution of galaxies and AGNs.

Distinct physical processes taking place in AGNs leave their unique imprints and dominate in different energy regimes within the SED, as represented by Fig. 1.4. The optically thick gaseous accretion disk around the SMBH produces an optical to UV spectrum, usually modeled by a multi-temperature blackbody. Additionally, the spectrum includes a soft X-ray excess and a hard X-ray tail produced by Compton up-scattering of disk photons in the hot optically thin corona, as well as the reprocessed/reflected coronal flux on the inner disk (e.g., Kubota & Done 2018). Obscuring dust clouds surrounding the disk in the shape of a torus are responsible for absorbing optical/UV photons along certain lines-of-sight, which are reprocessed into the IR regime with a predominant bump in the mid-infrared (MIR). In Compton-thick clouds ($\log N_H \geq 24$), the high-column-density dust also down-scatters X-rays due to Compton recoil, and produces a strong iron K_α line and a reflection spectrum. While these have often been associated with the inner wall of the torus (Matt et al. 2003; Comastri 2004), recent results based on X-ray variability and velocities seem to indicate that they are produced from outer BLR clouds (Andonie et al. 2022). Finally, the large-scale and powerful jets release radio emissions originating in the interactions of

relativistic electrons or positrons with the magnetic fields present in the jets, as well as internal shocks associated with bright radio knots.

The most widely used method to unravel the physical information from composite SEDs is the fitting with a combination of emission models (e.g., Berta et al. 2013; Ciesla et al. 2015; Calistro Rivera et al. 2016; Leja et al. 2018; Yang et al. 2020; Thorne et al. 2022). As deeper photometric surveys of increasing spectral and spatial coverage and resolution emerge, along with more advanced theoretical models, many of the parameterized emission models need to be updated and validated. One example is the smooth dusty torus models, which are not able to reproduce the observed silicate emission line in some Seyfert 2 galaxies and need higher-than-dust-sublimation temperatures to be consistent with observations (Tanimoto et al. 2019). Also, extinction produced by dust along the polar direction on scales of up to hundreds of parsecs (Asmus et al. 2016; Hönig 2019; Stalevski et al. 2019; Calistro Rivera et al. 2021) is not accounted for in the traditional torus models. Similarly, simple theoretical accretion disk models do not consider the observed high-velocity winds (Nardini et al. 2015; Tombesi et al. 2015), predict a steeper-than-observed UV power-law index (Davis et al. 2007), are still inconclusive in regards to the mechanisms driving soft X-rays (Kubota & Done 2018), and do not fully explore the very low Eddington ratio regimes.

The depth of information extracted from data in the SED-fitting approach is limited by the dimensionality of the parameter space, which is a product of both the number of photometric bands and the model complexity. A complex parameter space introduces degeneracies and correlations between parameters (Calistro Rivera et al. 2016), which can compromise the physical interpretation if not properly addressed. X-ray emission, strongly linked to the accretion disk, can help to break some of these degeneracies. For instance, luminous UV emission can be equally well reproduced by (1) an unobscured accretion disk combined with a subdominant stellar component, or (2) an obscured accretion disk with a bright host galaxy dominated by a young stellar population (Calistro Rivera et al. 2016). In such cases, X-ray data can help distinguish between the two scenarios.

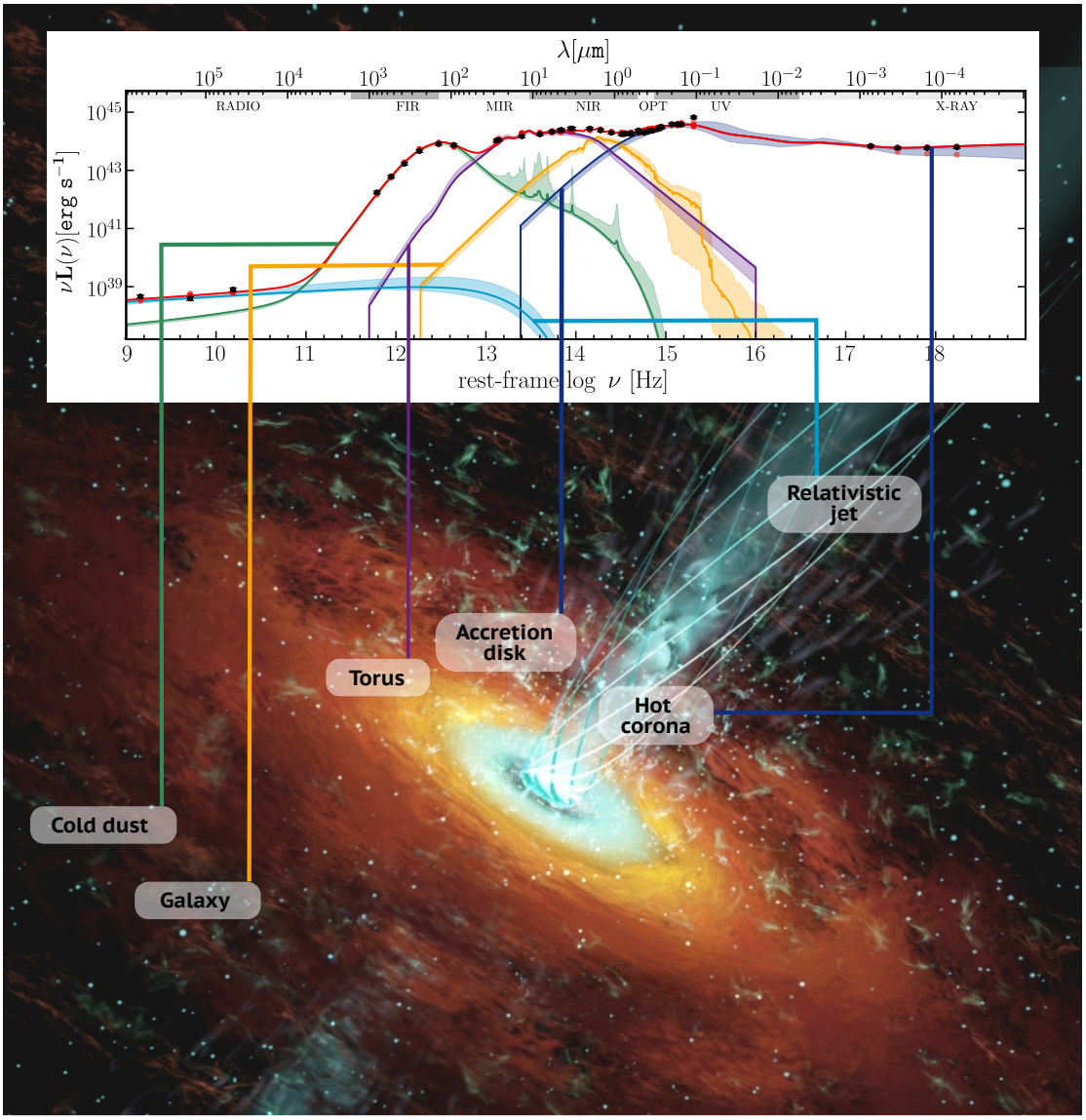

Figure 1.4: Schematic representation of the dominant components shaping the emission of active galaxies according to the current AGN unification model. From the most inner region: the hot corona and accretion disk emission dominating the optical, UV and X-rays (blue curves), the hot-dust torus in the NIR and MIR (purple curve) and the radio jet (light-blue curve). The host galaxy components extended to kpc scales and contributing to the SED are: the stellar population (yellow curve) and the cold dust from star-forming regions (green curve). The upper panel shows an example of an SED modeling with the best-fitted models, along with the photometric data. Artistic illustration of the AGN taken from: `https://www.iaa.csic.es/~jlgomez/Jose_L._Gomez/Animations_files/agn_wolli_2_1.jpg`.

High-energy X-ray emission also exhibits significant penetrating power and is thus largely unaffected by obscuration. In high-luminosity systems, the galaxy's X-ray contribution from high- and low-mass X-ray binaries (HMXBs, LMXBs) or star formation is negligible compared to that of the AGN. Conversely, in low-luminosity systems, correlations between X-ray emission and the star-formation rate (SFR) can help constrain the host-galaxy contribution (Mineo et al. 2014), allowing more accurate estimates of the AGN's intrinsic X-ray power.

It is well-established that radio emission is a powerful tracer of both star formation and AGN activity, as it is unaffected by dust. In particular, the radio emission originating in cosmic-ray electrons from star formation in the galaxy is correlated to the galactic far-infrared (FIR) luminosity (Helou et al. 1985; De Jong et al. 1985), offering an avenue to disentangle the origin of the overall radio emission. While this remains challenging for radio-quiet AGN, requiring high-resolution imaging and spectra index analysis (e.g.; Kawamuro et al. 2022), it is more straightforward for radio-loud sources. The radio spectra associated with AGN emission can provide insights into feedback mechanisms and the geometry and structure of outflows (Panessa et al. 2019; Silpa et al. 2022; Calistro Rivera et al. 2023).

Although X-ray and radio regimes are not commonly included in galaxy SED-fitting codes, their diagnostic power as tracers of AGN activity makes them highly valuable for improving the robustness of multiwavelength modeling of AGN SEDs. However, incorporating these regimes is challenging because radio emission can arise from a combination of AGN jets, lobes, star formation, and synchrotron self-absorption, each with distinct spectral shapes, while X-ray flux depends on coronal emission, absorption, and scattering, which vary with orientation, obscuration, and intrinsic AGN properties. Sparse coverage, heterogeneous depth, instrumental systematics, and differences in point spread function size across bands further complicate the integration of these data into standard SED fits. Accounting for these effects requires detailed physical templates, special care on the data handling and deep understanding of non-thermal processes, which are not yet fully imple-

mented in most SED-fitting tools.

1.5.2 Broad emission line measurements

The gravitational potential of a SMBH causes the gas clouds in the BLR to move at high velocities, leaving distinctive signatures in the rest-frame UV/optical spectra of quasars. Broad emission lines, with full width at half maximum (FWHM) > 1000 km s^{-1}, appear due to the Doppler effect, offering a powerful tool for estimating black hole masses (Peterson et al. 2004; Vestergaard & Osmer 2009). Assuming the BLR gas is virialized, the gravitational potential energy is related to the kinetic energy, allowing the SMBH mass to be estimated from the velocity of the gas (V), measured from the emission-line width, and the size of the BLR.

Since AGN are intrinsically non-stochastic variable sources, time-domain observations can be used to determine the BLR size. As discussed in Section 1.1, the optical and UV continuum traces the accretion disk emission from the immediate vicinity of the black hole, while the broad emission lines originate from a region extending up to a few parsecs. Then, the variations in the continuum emission manifest as changes in the emission-line fluxes after a time delay (τ). This delay reflects the light-travel time between the accretion disk and the BLR, providing a measure of the BLR size. This technique is known as reverberation mapping (Blandford & McKee 1982; Peterson 1993), and an schematic representation of it is presented in Fig. 1.5. Within this formalism, the SMBH mass can be expressed as:

$$\mathrm{M}_{BH} = \frac{c\tau V^2}{G}, \tag{1.1}$$

where G is the gravitational constant and V is computed from the FWHM of the emission line.

The extensive application of this technique has not only enabled SMBH mass measurements for many AGNs, but also revealed a correlation between the BLR size and the AGN luminosity by the form $R_{BLR} \propto L_{\lambda}^{\alpha}$, where L_{λ} is a monochromatic continuum luminosity

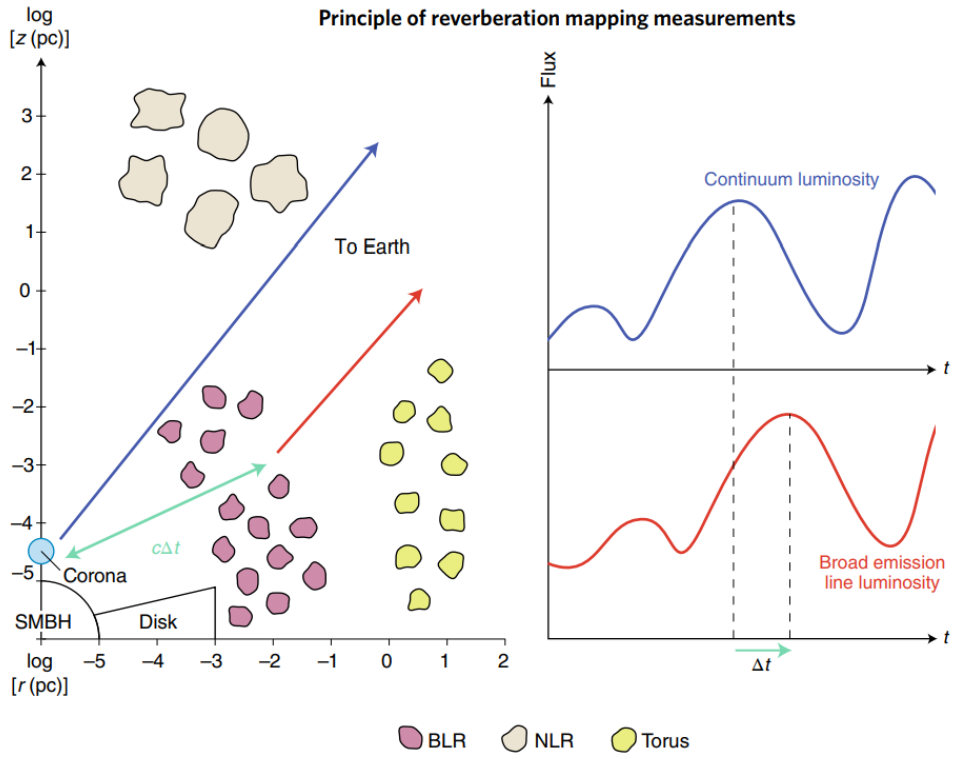

Figure 1.5: Transversal view of AGN emission components relevant for the reverberation mapping technique: the accretion disk, the corona represented by a light blue circle, and the BLR region by pink clouds. *Left panel:* different light travel paths of the UV continuum (blue arrow) and the re-processed BLR lines (green and red arrows) emissions. *Right panel:* effects of the light travel times on the stochastic variability observed in the continuum and emission-line fluxes. The time delay provides an estimate of the BLR size. Credits: Francisco R. Villatoro, available in `https://francis.naukas.com/files/2019/06/D20190613-nature-astronomy-s41550-019-0815-y-principle-reverberation-mapping.png`.

at a given wavelength (Koratkar & Gaskell 1991; Kaspi et al. 2000, 2005; Bentz et al. 2006, 2009a). This empirical relation made it possible to estimate black hole masses dispensing with extensive monitoring campaigns, which are challenging and unfeasible for large samples, by using single-epoch spectra instead (Shen et al. 2008; Fine et al. 2008; McGill et al. 2008). In this case, the SMBH mass can be written as:

$$\mathrm{M}_{BH} \propto R_{BLR} V^2 \tag{1.2}$$

$$\mathrm{M}_{BH} = \frac{f L^{\alpha} V^2}{G}, \tag{1.3}$$

where f is the virial factor that accounts for stochastic and systematic uncertainties (Park et al. 2012).

While photoionization models predict $\alpha \approx 0.5$ (e.g.; Davidson 1972); empirical calibrations, well established only for a few broad emission lines such as H β λ4861 Å (e.g., Bentz et al. 2013; Kaspi et al. 2000; Vestergaard & Peterson 2006), Mg II λ2800 Å (e.g., McLure & Jarvis 2002; McLure & Dunlop 2004; McGill et al. 2008), C IV λ1549 Å (e.g., Vestergaard 2002; Vestergaard & Peterson 2006), and H α λ6565 Å (e.g., Dalla Bontà et al. 2025), find α values in the range $0.5 - 0.7$. These relations are typically defined using continuum luminosities at 3000 Å, 5100 Å, and the H α and H β line luminosities (McGill et al. 2008). The most reliable R_{BLR}-L_{λ} relation is based on H β and $L_{5100\text{Å}}$, since it has been calibrated on larger samples AGN and covers a wide range of luminosities (10^{43}erg s^{-1} $< L_{5100\text{Å}} < 10^{46}$erg s^{-1}; Mejía-Restrepo et al. 2018). Then, calibrations with other lines seek to reconcile the results based on H β.

However, neither H β nor H α are longer observable in the optical regime at redshifts $z \gtrsim 0.8$ and $z \gtrsim 0.3$, respectively, requiring NIR spectroscopy to estimate SMBH masses (McLure & Dunlop 2004). NIR spectroscopy from ground is significantly affected by telluric absorption features from molecules in Earth's atmosphere. At $z > 6$, this is even more critical as all rest-frame optical lines are redshifted into the mid-infrared (MIR), beyond the reach of ground-based instruments. Consequently, mass estimates at these redshifts rely on alternative rest-frame UV emission lines, primarily C IV and Mg II (De Rosa et al. 2014; Farina et al. 2022), if lying within a NIR region free of atmospheric attenuations.

It is worth noting that C IV is a high-ionization line that poses several challenges for the study of high-redshift quasars. The first issue arises in luminous quasars, where the C IV emission line is often strongly blueshifted (> 2000 km s^{-1}; Gaskell 1982; Richards et al.

2002; Schindler et al. 2020), indicating the presence of non-virial motions, and exhibits asymmetric profiles (Richards et al. 2011a). This blueshift is typically smaller at low redshift, suggesting an evolutionary trend, and is commonly attributed to strong winds launched from the accretion disk that contribute significantly to the high-ionization line emission (e.g., Richards et al. 2011a).

A second complication arises from the Baldwin effect, in which the equivalent width (EW) of high-ionization broad emission lines anticorrelates with the continuum luminosity (Baldwin 1977). This effect is particularly relevant for high-redshift quasars, where Richards et al. (2011a) also found that both the blueshift and the asymmetry of the C IV line correlate with its EW. Current interpretations attribute these properties to radiation-driven, non-virialized outflows or to C IV disk winds coexisting with a high-opacity, hot-dust torus (Meyer et al. 2019).

Given the above, the C IV line is considered less reliable for black hole mass estimates (Park et al. 2017; Mejía-Restrepo et al. 2018). Since mass estimates based on low-ionization lines are closely consistent with those derived from H β (typically within ≈ 0.2 dex; Greene & Ho 2005; Trakhtenbrot & Netzer 2012), the low-ionization Mg II line is widely used as an alternative. McLure & Jarvis (2002) demonstrated that the UV continuum luminosity at $L_{3000,\text{Å}}$ correlates strongly with the BLR radius, and combined this relation with the Mg II line to estimate UV-based black hole masses in high-redshift quasars (see also Barth et al. 2003; Willott et al. 2003). Later Vestergaard & Osmer (2009) recalibrated the Mg II-based relation using high-quality spectra from the SDSS DR3 quasar sample (Schneider et al. 2005), covering monochromatic luminosities at 1350 Å, 2100 Å, 3000 Å, and 5100 Å. We adopt the Mg II-based BH mass relation from Vestergaard & Osmer (2009), defined as:

$$M_{\text{BH}} = 10^{6.86} \left[\frac{\text{FWHM(Mg II)}}{1000\ \text{kms}^{-1}} \right]^2 \left[\frac{L_{3000\text{Å}}}{10^{44}\text{ergs}^{-1}} \right]^{0.5} \tag{1.4}$$

with a 0.55 dex scatter.

The unprecedented sensitivity and wavelength coverage of JWST have enabled rest-frame NIR emission-line diagnostics for BH mass estimates in the early Universe. Using

the Mid-infrared Resolution Spectrograph (MRS; 4.9–27.9 μm), Bosman et al. (2024) detected the broad Paβ 1.2822 μm, Paα 1.8756 μm, and Hα lines in the $z = 7.0848$ quasar J1120+0641, enabling one of the most robust rest-frame NIR measurements of a SMBH at this epoch. The Paschen lines (Paα and Paβ) are particularly valuable as black hole mass tracers because they are much less affected by dust extinction than Balmer lines (Kim et al. 2010); however, direct calibration via reverberation mapping is challenging due to the long timescales required for those campaigns.

The Near Infrared Spectrograph (NIRSpec; 0.6–5.3 μm) has also opened access to rest-frame optical emission-line regimes across a wide redshift range. JWST/NIRSpec observations have allowed BH mass estimates for faint AGNs at $z = 4.015 - 6.936$ using the Hα line width and luminosity together with the Greene & Ho (2005) scaling relation (Harikane et al. 2023). Similarly, Larson et al. (2023) derived BH mass of the AGN CEERS_1019 at $z = 8.679$, and Übler et al. (2024) did so for the AGN system ZS7 at $z = 7.15$, both based on Hβ line using the same Greene & Ho (2005) relation. Given its wavelength coverage, NIRSpec can probe the Hβ line for quasars and AGNs up to redshifts ~ 10 (Fan et al. 2023).

1.6 This thesis

We explore modern and flexible approaches for the selection and characterization of reionization-era quasars. This work is motivated by current and forthcoming wide-field photometric and spectroscopic surveys such as the DESI Legacy Survey (in particular DR10; Dey et al. 2019), the DECam Local Volume Exploration Survey (DR12; Drlica-Wagner et al. 2022), the Vera C. Rubin Observatory's Legacy Survey of Space and Time (LSST; Ivezić et al. 2019), the Euclid Wide Survey (EWS; Scaramella et al. 2022), and the High Latitude Survey from the Roman Space telescope (Schlieder et al. 2024), as well as high multiplexing spectroscopic facilities like 4MOST, MOONS, and DESI.

These large-area and high-cadence datasets, covering both deep photometric and time-

domain regimes, represent outstanding opportunities for breakthrough discoveries but also pose major challenges for data-mining and candidate selection. The dedicated tools designed and tested in this work will serve as scalable pipelines capable of efficiently processing the data volumes expected from upcoming facilities such as LSST, and providing robust target lists for spectroscopic follow-up.

In this thesis, we investigate a range of techniques to identify quasars at $z > 6$ in a more flexible and less biased way than traditional methods, while minimizing contamination from other astrophysical sources. Addressing this challenge requires a deeper understanding of the contaminating populations and the reasons they are misclassified as quasars. To this end, we employ and test multiple approaches, including color selections, supervised and self-supervised machine learning methods, SED fitting, visual inspection of photometric stamps, and variability analysis at different stages of the project.

After establishing an effective "recipe" for quasar discovery, currently guided by the known population from the literature but extendable toward less biased searches, we focus on improving tools for physical characterization. Robust physical characterization is essential to turn discoveries into meaningful scientific results, enabling studies of the co-evolution of SMBHs and their host galaxies, the physical conditions of the Universe within the first billion years after the Big Bang, the assembly of SMBH mass across cosmic time, and the role of potentially underrepresented quasar populations in early cosmic evolution. Consequently, the main goals of this thesis are:

- To define an initial preselection of quasar candidates at $z > 4$ from DESI Legacy Survey DR10 using optical photometry and catalog-based constraints.
- To evaluate the performance of supervised and self-supervised machine learning algorithms using catalog and imaging data, together with variability analysis of unWISE light curves, to refine quasar selection.
- To identify robust quasar candidates at high-redshift via SED fitting and visual inspection to minimize contamination from ultracool dwarfs, asteroids, and image

artifacts.

- To establish calibration baselines for the variability analysis by processing light curves of non-variable stars surrounding known quasars from the literature, using the ALeRCE light-curve classifier module.

- To perform optical spectroscopic follow-up observations to confirm the quasar nature of candidates and to assess contaminant populations.

- To expand a Bayesian SED-fitting tool to include radio and X-ray regimes for the photometric characterization of local quasars and AGNs, in preparation for its application to high-redshift sources.

The thesis is organized as follows. In Chapter 2, we present the selection of $z \gtrsim 6$ quasar candidates based on self-supervised contrastive learning applied to DESI Legacy Survey DR10 (LS DR10 henceforth) imaging data. We demonstrate the effectiveness of this approach with the discovery and spectroscopic confirmation of 16 new quasars at $z = 5.94 - 6.45$, some of them exhibiting unusual properties such as narrow, and strong Ly α emission, and a mild observed-frame red near-infrared continua. We also discuss the image properties crucial for the CL selection, and the prioritization stage based on SED fitting.

In Chapter 3, we analyze the variability of 285 quasars from Fan et al. (2023) with available unWISE light curves, along with 10 non-variable calibration stars around each quasar, using different metrics from the ALeRCE light-curve classifier. These results, complemented by a power-spectrum analysis, are compared to previous studies based on the Zwicky Transient Facility (ZTF; Bellm et al. 2018) and SDSS time-domain data from low-redshift quasars.

In Chapter 4, we focus on AGN characterization by presenting AGNfitter-rx, a Bayesian tool for modeling the radio-to-X-ray SEDs of active galaxies. This framework enables detailed characterization of the accretion disk, hot dusty torus, relativistic jets and

core radio emission, hot corona, stellar populations, cold dust, and star-forming regions. We introduce new theoretical and semi-empirical models for these emission components and apply AGNfitter-rx to a sample of 36 AGNs at $z \lesssim 0.7$. The analysis includes the inferred physical properties and model comparisons. Finally, in Chapter 4.4, we summarize the main results and discuss potential directions for future work.

Throughout this thesis we adopt ΛCM cosmology with $H_0 = 67.4 \mathrm{kms}^{-1}\mathrm{Mpc}^{-1}$, $\Omega_m = 0.315$ and $\Omega_\Lambda = 0.685$ (Collaboration et al. 2020); and report magnitudes in the AB system.

Chapter 2

Quasar discoveries with self-supervised learning applied to images

This chapter is based on Martínez-Ramírez, L. N., Wolf, J., Belladitta, S., Bañados, E., Bauer, F. E., Hviding, R. E., et al., Submitted (2025)

2.1 Introduction

As discussed in Section 1.4, applications of supervised ML techniques to quasar selection has been extended to the high redshift regime ($z \gtrsim 5$) over the past decade. An outstanding example is the Random Forest classifier developed by Wenzl et al. (2021), which led to the discovery of 20 new quasars with 54% success rate. Nonetheless, careful inspection reveals that their sample is dominated by prototypical bright, blue quasars (see Fig. 11 in Wenzl et al. 2021), with only two sources at $z > 5.6$, underscoring potential inherent limitations of supervised approaches. The reliance on photometric data, prone to artifacts, combined with the small size of the high-redshift population, currently about 328 quasars at $z \gtrsim 6$ (computed from Fan et al. 2023, and private communication), poses serious hurdles for uncovering underrepresented quasar populations and for effectively

exploiting the discovery potential of present and upcoming wide-field surveys.

The more recent and novel attempt to incorporate imaging data into high-z quasar selection is the approach by Byrne et al. (2024) via self-supervised CL framework. Their preselection with relaxed $i - z > 1$ color, lack of morphological restriction, and allowing g and r bands detections, was intended to enhance the discovery of lensed quasars while avoiding visual inspection. As a byproduct, this strategy enabled the recovery of non-lensed quasars previously missed by traditional color-based selections. However, the requirement of detections in J, K, $W1$ and $W2$ bands definitively biases the method towards bright sources, particularly given the relatively shallow $W2$ depth of 18.4 AB mag (Wright et al. 2010). Even more critically, these detection constraints are combined with a $W1 - W2$ color requirement, limiting the sample to just 7438 objects. This number is too small for an unsupervised approach to adequately optimize the contrastive loss, resulting in limited representation quality. While these issues stem primarily from the preselection strategy, they indicated that the CL performance, already demonstrated to be good, can be boosted just by providing a better training sample.

Empirical studies show that CL can perform as well as, or even better than, supervised training, producing representations that transfer more effectively and remain robust across different datasets (Chen et al. 2020; Karthik et al. 2021). Also, the work by Sarmiento et al. (2021) demonstrated that CL is more robust against nonphysical correlations related to observational effects in astronomical data, than unsupervised dimensionality reduction algorithms such as principal component analysis (PCA; Pearson 1901). Building on this promise, we extend CL-based quasar selection to test whether it enables the recovery of diverse high-redshift quasars with lower contamination and fewer selection biases than supervised alternatives.

In this part of the thesis, we perform $z \gtrsim 5.5$ quasar selection based on CL, a self-supervised ML algorithm mapping high-dimensional data sets into a low-dimensional feature space representation. This method learns the most informative correlations in the data by measuring the similarity between data points without any labels and taking

advantage of data augmentation methods (Huertas-Company et al. 2023). The data for this application of the CL consist of multi-band optical images from the DESI Legacy Survey DR10 (hereafter LS DR10; Dey et al. 2019). This input format with multiple channels containing two-dimensional spatial information is similar to those used in previous successful astronomical applications of CL (e.g., Sarmiento et al. 2021; Byrne et al. 2024).

In section 2.2.1, we describe the preparation of the data set for the CL training. We start with a catalog-based preselection of *i*-dropout candidates, followed by the assembly of the imaging tensor. The implementation of the CL framework is detailed in section 2.2.2, while the analysis of the resulting latent space in 2.2.3. The candidate prioritization strategy and follow-up observations are described in section 2.2.4. In section 2.3.1, We demonstrate the effectiveness of our method with the spectroscopic confirmation of 16 new quasars, along with the reidentification of 3 previously published quasars, out of 40 observed candidates, corresponding to a 45% success rate. Several of these quasars exhibit interesting properties discussed in 2.3.2, such as narrow Ly α emission lines, red near-infrared (NIR) continua and high Ly α EWs. Finally in section 2.3.3, we compare our selection with others, mainly based on color constraints, reported in the literature.

2.2 Methodology

2.2.1 Pre-selection of *i*-dropout candidates

To ensure effective training of our CL model, we first apply a preselection to the input dataset. Although larger batch sizes typically improve CL performance (Chen et al. 2020), the underlying neural network can easily get sidetracked by the most dominant features, such as sizes or extended morphology, if the training set is not properly delimited. To mitigate this and encourage the learning of the most fundamentally relevant features, we restrict the input to red, nearly point-like sources. This preselection filters out contaminants such as extended galaxies, asteroids, satellites, and diffraction spikes in the photometry

while minimizing the impact of irrelevant features.

The procedure to select $z > 5.5$ quasar candidates is divided in three major steps: (1) The selection of a preliminary set of candidates from photometric optical catalogs by applying color constraints; (2) the training of the CL algorithm with g, r, i and z band optical images to produce an embedded low-dimensional representation of i-dropout sources, and the subsequent identification of candidates and contaminants guided by labeled data; (3) SED fitting procedure using quasar, galaxy and ultracool dwarf templates to prioritize for follow up observations the sources with reliable photometry and SEDs more consistent with a quasar model.

Photometric catalog

We selected our candidates from LS DR10, which offers deep optical imaging over $> 20\,000$ deg^2 of sky, including substantial coverage of the underexplored Southern Hemisphere, where high-redshift quasar discoveries remain sparse (e.g., Reed et al. 2017; Belladitta et al. 2019; Belladitta et al. 2025; Wolf et al. 2024; Onken et al. 2022) and next generation facilities are being built (e.g., The Extreme Large Telescope (ELT) and The square kilometer array (SKA) radio telescope, respectively; Padovani & Cirasuolo 2023; Hall et al. 2008). The DR10 release combines imaging of: the Southern sky from the DECam Legacy Survey (DECaLS, Dey et al. 2019) using the Dark Energy Camera (DECam, Flaugher et al. 2015) on the Blanco 4m telescope and other programs using DECam such as the DECam eROSITAS survey (DeROSITAS, Zenteno et al. 2025), the DECam Local Volume Exploration Survey (DELVE, Drlica-Wagner et al. 2021), the Blanco Imaging of the Southern Sky Survey (BLISS; PI: Soares-Santos Mau et al. 2019) and the Dark Energy Survey (DES, Abbott et al. 2018); the Northern sky from the Mayall z-band Legacy Survey (MzLS, Silva et al. 2016) and the Beijing-Arizona Sky Survey (BASS, Zou et al. 2017); and full-sky mid-infrared (MIR) photometry from the Wide-field Infrared Survey Explorer (WISE, Wright et al. 2010): the unblurred coadds of WISE (unWISE, Mainzer et al. 2014) and Near-Earth Object Wide-field Infrared Survey Explorer Reactivation Mission

(NEOWISE, Mainzer et al. 2014). LS DR10 photometry reaches median 5σ point-source depths of $g_{\rm DECam} = 24.7$, $r_{\rm DECam} = 23.9$, $i_{\rm DECam} = 23.6$, $z_{\rm DECam} = 23.0$ AB magnitudes in the southern hemisphere. Similar but slightly shallower depths are achieved in the northern hemisphere.

Since machine learning algorithms face problems when dealing with the lack of data, we request valid observations in all optical bands ($n_{\rm obs} > 0$), limiting the selection of targets to an area of 15,342 deg^2 and declinations $< 30°$ (due to the limited i-band coverage). We focus on candidates for which the SED fitting refinement step is possible and informative, thus requiring detection in at least one NIR band from the VISTA Hemisphere Survey DR7 (VHS, McMahon 2012), UKIRT Infrared Deep Sky Surveys DR11 (UKIDSS, Lang 2014) or UKIRT Hemisphere Survey DR2 (UHS, Dye et al. 2018) surveys and a signal-to-noise ratio (S/N) > 3 in WISE W1. While WISE bands are already included in LS DR10 catalog, NIR bands were added by crossmatching our catalog with the previously mentioned surveys within a 1.5″ radius. Due to the previous criteria, the catalog is limited to 5σ depths of VHS DR7 J = 21.7-22.3 and Ks = 21.8-22.1 mag, UKIDSS DR11 depth of H= 20.2 mag, UHS DR2 depth of J=20.5 and Ks=20.0 mag, and unWISE depth of $W1$ = 20.72 mag. Besides, the low spatial resolution (2.75″/pixel) of WISE W1 renders this band prone to source blending and therefore, potential flux overestimation.

The constraints accounting for the i-dropout sources are as follow:

$$(\mathrm{S/N})_{\rm z} > 7 \tag{2.1}$$

$$(\ (i_{\rm PSF} - z_{\rm PSF} > 1.5 \ \text{ and } \ (\mathrm{S/N})_{\rm i} > 3) \ \text{ or} \tag{2.2}$$

$$(i_{\rm limPSF} - z_{\rm PSF} > 1.5 \ \text{ and } \ (\mathrm{S/N})_{\rm i} < 3)\ ) \tag{2.3}$$

$$(\mathrm{S/N})_{\rm g,r} < 3 \tag{2.4}$$

$$i_{\rm flux\_ivar} \neq 0 \tag{2.5}$$

where $i_{\rm PSF}$ and $z_{\rm PSF}$ are i and z band magnitudes; and $i_{\rm flux\_ivar}$ the inverse variance of the flux in the i band. All of them are products from *The Tractor* algorithm (Lang et al. 2016)

used in LS DR10 for the source extraction, based on the convolution of the images to a PSF model. The magnitude limit for the i-band (i_{limPSF}) is computed as:

$$i_{\mathrm{limPSF}} = 22.5 - 2.5 \log \left(\frac{3}{\sqrt{i_{\mathrm{flux\_ivar}}}} \right). \tag{2.6}$$

We set the color threshold at 1.5, which is 0.5 mag lower than in previous selections (e.g., Reed et al. 2017; Bañados et al. 2016). This choice enables us to recover 127 out of the 166 known quasars at $z > 5.3$ from the literature (Fan et al. 2023) that lie within the LS DR10 footprint, have complete optical coverage, and satisfy conditions (1)–(6). Among the 39 missed sources, only 15 lie at $z > 6$, the redshift range on which this work is focused.

The recovery rate decreases when imposing infrared constraints: from 127 to 106 when requiring a significant detection in WISE W1, and further to 69 when requiring both NIR and WISE W1 detections. In other words, the NIR+W1 criterion removes $\sim 46\%$ of known quasars. However, without any IR constraints, the candidate list would swell to 3,254,837 sources, which is beyond the computational scope of this work. Moreover, such a large sample would have far less reliable prioritization for follow-up, as most sources would have too few photometric points for meaningful SED fitting. Even requiring only the W1 requirement (losing only $\sim 17\%$ of known quasars) still results in 1,850,912 candidates, a size that remains challenging to process and likely to have a high contamination rate.

We also adopted maskbits that select clean sources (including masked point-like and/or unsaturated sources).[1] We explicitly do not select objects by source type nor morphology features. We note that adopting the type "PSF" to focus on the most point-like objects would only reduce the catalog by 17%, but exclude 11 known $z \sim 6$ quasars from the literature. The resulting sample consists of 165 253 objects.

[1] i.e., maskbits < 2 or = 32, 64 ,96, 128, 160, 256, 512, 768, 2048, 2176.

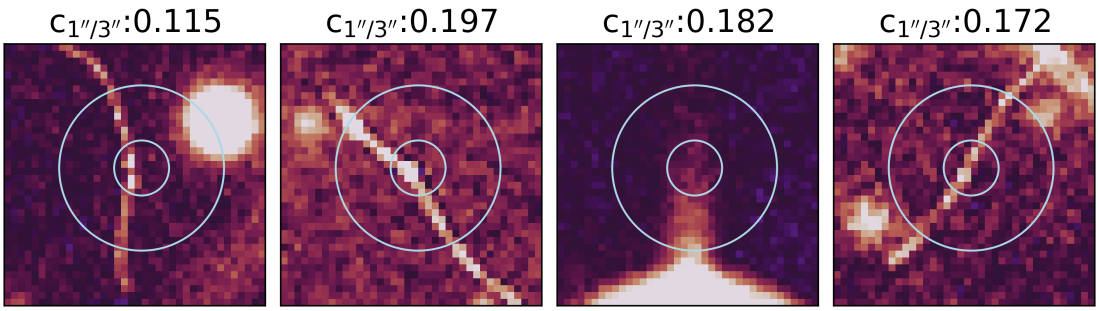

Figure 2.1: z-band $10'' \times 10''$ sized stamps illustrating examples of the artifacts identified in the photometry through visual inspection of the tensor. Light blue circles indicate the $1''$ and $3''$ radius apertures used in the compactness criteria calculation. The $C_{1''/3''}$ values are presented above the corresponding postage stamp.

Multi-band imaging tensor

We compiled $10'' \times 10''$ sized cutouts from the g, r, i and z band images for each of the 165 253 targets from the LS DR10 archive. Given the native DECam pixel scale of $0\farcs262$ per pixel, this results in 38×38 pixel^2 images and a tensor dimension of $165253\times4\times38\times38$.

During our inspection of tensor slices, we discovered clear artifacts, mainly cosmic rays and diffraction spikes (see Fig. 2.1 for examples), that the maskbits conditions at the catalog level did not filter out. To identify and remove these extended sources, we calculate the ratio of flux within a $1''$ radius aperture to the flux within a $3''$ radius aperture ($C_{1''/3''}$). If this ratio exceeds 0.3, the source is considered compact. Conversely, we reject sources where 30% or less of the total flux within the $3''$ aperture is confined to the inner $1''$ aperture, as this indicates an extended structure. This flexible 30% threshold was calibrated on a sample of known quasars at $z > 5.3$ (Fan et al. 2023), all with ratios above 45%, except for the quasar pair J2037–4537 (Yue et al. 2021) with a 32% ratio, prompting us to lower our limit to avoid excluding similar systems. This step reduced the training sample used as input to the CL pipeline to 130472 sources by removing these extended unphysical sources in line with the literature.

2.2.2 Constrastive Learning for the selection of quasars candidates

With the imaging tensor input of the CL defined, a preprocessing was carried out consisting of:

1) masking nan-value pixels, if present, with zeros, given that the pixel flux distributions across all bands are dominated by background noise and peak at zero.

2) normalizing the tensor using a single global factor derived from the 99th percentile of the brightest band across all sources, which in our case corresponds to the z-band. To compute this factor, we first build per-band distributions of pixel flux values (within a $1''$ radius from the center) across the full sample. We then determine the 99th percentile of each band's distribution and identify the brightest among them. This value is adopted as the normalization factor and applied uniformly to all sources in the tensor. By using one global factor rather than source-by-source scaling, we preserve both the relative brightness distribution of the targets and the intrinsic SED shape across bands.

3) random shuffling of the sources indexes in the tensor to prevent any potential spatial or data quality bias, as the original tensor was generated in increasing right ascension order, and the training sample is divided into batches.

We train a modified version of the self-supervised model simCLR (a simple framework for contrastive learning of visual representations, Chen et al. 2020) built on *Keras 2.15.0*. We implement the following data augmentations: horizontal and vertical flips, and random rotation within $\pm 90°$, to produce transformed versions of each image. Two augmented views of the same input source are referred to as positive pairs, while negative pairs are augmented views of different input sources. These pairs drive the training, encouraging the network to cluster similar sources in the latent space while pushing away dissimilar ones. The encoder producing this low-dimensional representation of the images consists of four convolutional neural network (CNN) layers that progressively reduce the spatial dimensions of the four channels to a 128 length feature vector. This was followed by a Flatten layer and a Dense layer, which applies a transformation to the vector combining the local spatial information into a more compact and complex embedded representation.

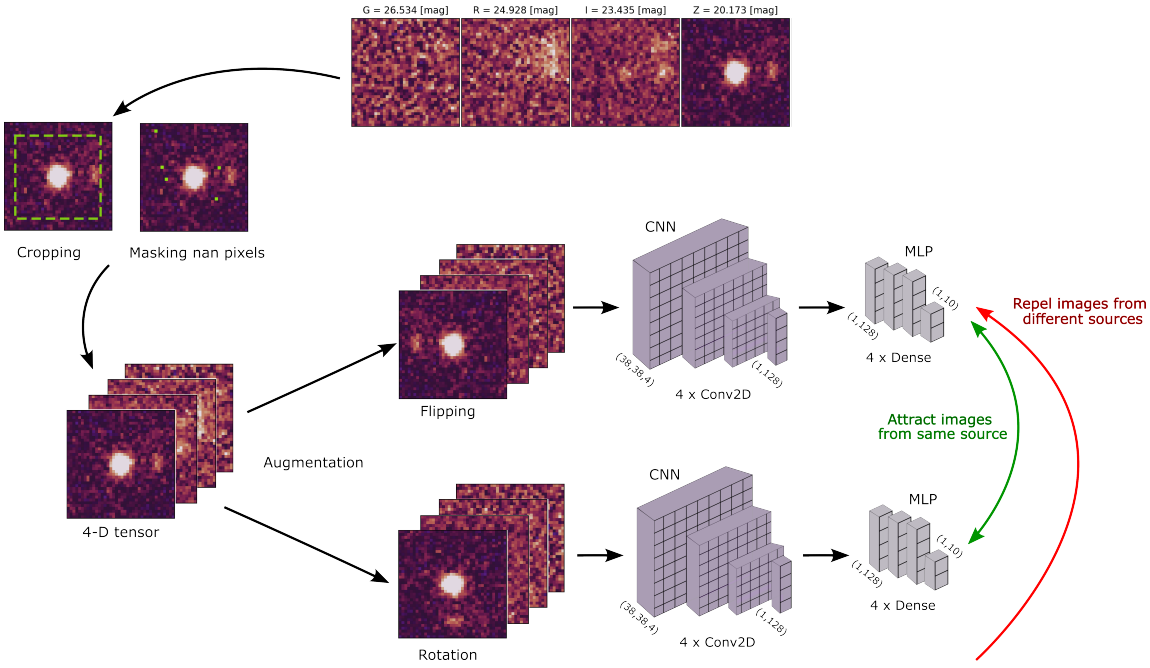

Figure 2.2: Architecture of the self-supervised contrastive learning framework used in this work. The top and left sections illustrate the image preprocessing and the tensor assembly steps. The structure of network is shown sequentially from left to right, starting with data augmentation, then the encoder block, and lastly the projection head. Red and green arrows indicate the contrastive loss computation, which measures similarity between random pairs of augmented images.

The stacked layers allow the network to capture increasingly complex visual features: from basic edges and shapes to higher-level structures like smudges and PSF-like features. All convolutional and dense layers use a ReLU activation function to introduce non-linearity and emphasize informative patterns.

The projection head maps the 128-dimensional vectors resulting from the encoder into a 10-dimensional latent space, where the contrastive loss is computed by comparing the similarity between random pairs of sources. This small multilayer perceptron (MLP), composed of four dense layers, is designed to prevent overfitting of trivial features and enhance the separation between positive and negative pairs in the embedded space. Fig. 2.2 shows the architecture of our CL implementation. The combination of hyperparameters achieving the best results was: temperature of 0.1 for the normalized temperature-scaled

cross entropy contrastive loss; width of 128 for the latent space; maximum batch size given by the limits of our GPU resources, 10000; and 3000 epochs. Other architecture choices, such as the number of CNN and Dense layers in the encoder and MLP, were selected to balance representation power and overfitting, following similar astronomy applications (e.g., 3 CNN + 3 Dense layers in Sarmiento et al. 2021; 4 CNN + 3 Dense layers in Byrne et al. 2024).

All the training experiments were carried out on one of the two GPU nodes of the *Astronodes* cluster at the *Max Planck Institute für Astronomie*. Each node is equipped with NVIDIA Tesla V100S GPUs (32 GB HBM2 memory), and runs CUDA 12.4 (driver version 550.90.07). The runtime was usually between $\sim$ 4.2 hours for the optimized hyperparameters values, and $\sim$ 8 hours for tests with more epochs, higher temperatures, batch sizes or widths of the latent space.

2.2.3 2-dimensional embedded space

We applied the unsupervised dimensionality reduction algorithm known as Uniform Manifold Approximation and Projection (UMAP, McInnes et al. 2018) to reduce the representation dimensions from 128 to 2 in the embedded space generated by simCLR.

As part of the downstream analysis, we cross-matched spectroscopically classified data within a 1.5″ radius to facilitate a comprehensive interpretation of the latent space, which helped us identify potential clusters of high-z quasar candidates and contaminants. This included spectroscopically confirmed quasars at $z > 5.3$ from Fan et al. (2023), M dwarfs from West et al. (2011), a compilation of LT ultracool dwarfs from the literature (Albert et al. 2011; Burningham et al. 2008, 2010, 2013; Chiu et al. 2008; Day-Jones et al. 2013; Deacon et al. 2011; Kendall et al. 2007; Kirkpatrick et al. 2010, 2011; Knapp et al. 2004a; Liu et al. 2002; Lodieu et al. 2007, 2009, 2012; Mace et al. 2013; Matsuoka et al. 2011; Pinfield et al. 2008; Thompson et al. 2013; Radigan et al. 2013; Schmidt et al. 2010b; Scholz 2010a,b; Best et al. 2015; Cardoso et al. 2015; Marocco et al. 2015; Tinney et al. 2018) and stars from Gaia DR3 (Collaboration & Vallenari 2023) with a high probability

of being a single star (PSS > 0.99) and statistically significant proper motion (pm/e_pm > 3). UMAP clustering was fine-tuned in order to optimize the separation between labeled quasars and contaminants adopting the following parameters: n_neighbors = 30, min_dist = 0.02, metric = euclidean.

The resulting two-dimensional embedded space representation of the *i*-dropout candidates, alongside known populations of stars, quasars, and MLT dwarfs is shown in Fig. 2.3. The multiple tests we performed to find the optimal UMAP parameters revelead that, regardless of their values, the embedded space naturally separates into two prominent regions: the right "island," dominated by stars and ML dwarfs, and the left island, primarily populated by T dwarfs and high-redshift quasars. Notably, we identify a concentration of quasars at redshifts z = 6–6.4 near the bottom of the left island, while more sparsely distributed $z > 6.8$ quasars appear toward the top.

To investigate the dominant properties influencing the arrangement of sources in the embedded space, we color-code the data points by catalog magnitudes, color indices, depth and PSF size. Figures 2.4 and 2.5 highlight several relevant features: the z-band magnitude; $i - z$, $z - J$ and $z - W1$ colors; the g-band depth and z-band PSF size. These maps reveal trends that aid the interpretation of the algorithm. For instance, source brightness is a major driver of the representation, as evidenced by the gradient in z magnitude: it increases from bottom to top within the left island and from right to left within the right island. The $i - z$ color map highlights regions with strong spectral breaks, where most confirmed quasars and stars are located in Fig. 2.3. Notably, the area around (x,y) = (2, -3) containing quasars shows more moderate $i - z$ colors ($\sim$ 1.5), suggesting that our more flexible selection criteria successfully include sources that might be excluded by stricter cuts.

Although the NIR bands were not part of the CL input, we include $z - J$ color maps to investigate potential differences among quasar populations. The region near (x,y) = (5, -2.5) corresponds to typical blue quasars with flat NIR slopes ($z - J \leq 0.5$), whereas the region around (x,y) = (2, -3) contains a mix of intermediate colors ($z - J \sim$ 1.0–1.5). Meanwhile, the region near (x,y) = (-2, -1) and the vertical structure toward the top of the

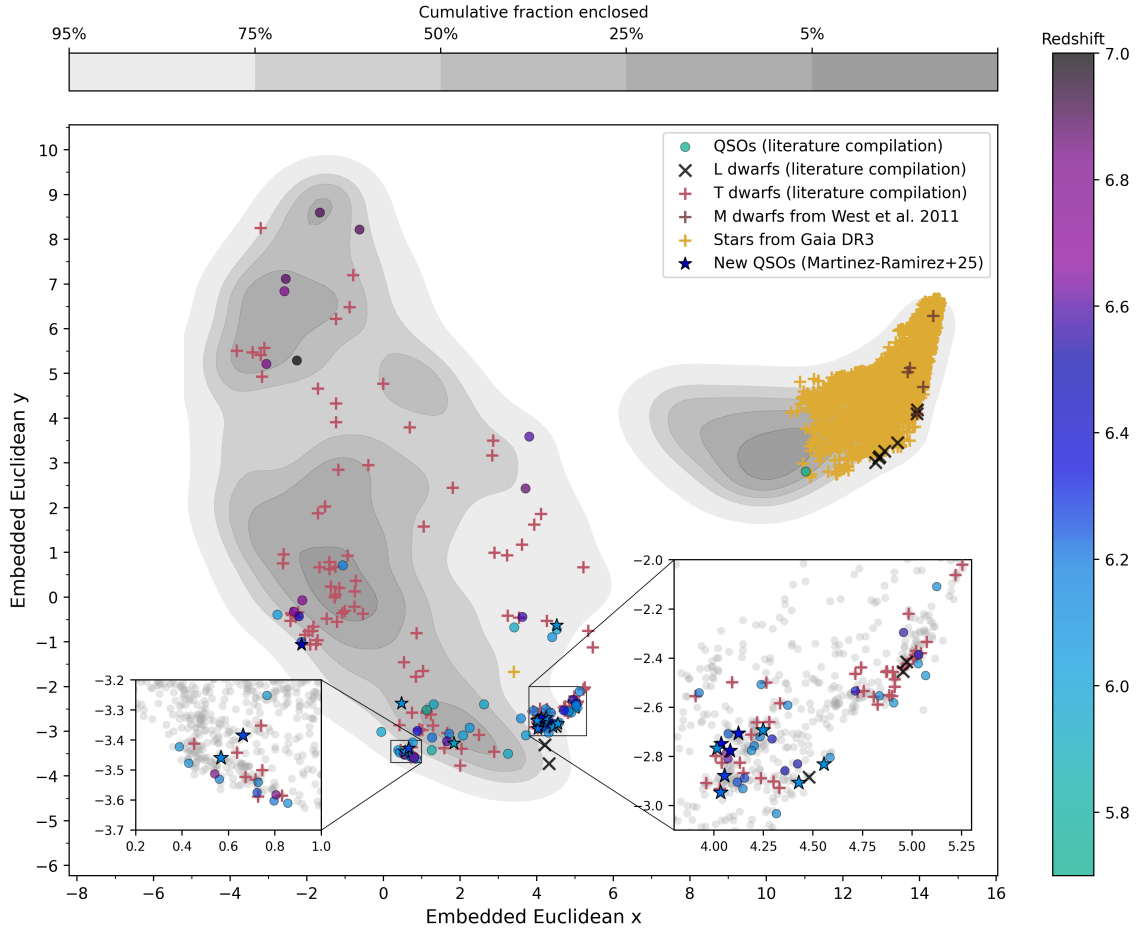

Figure 2.3: Embedded Euclidean space generated by UMAP for LS DR10 *i*-dropout sources after training the CL algorithm. All sources are represented as grey countours given by the number density; known M, L and T dwarfs are: brown, black and red crosses, respectively; and Gaia DR3 stars are yellow crosses. Spectroscopically confirmed quasars from the literature and the new quasars from this work are represented by filled circles and stars, respectively, all of them color-coded by their redshift. Quasar candidates are selected from the lower part on the left island where the source density is low and the contamination ratio with ultracool dwarfs is $\sim$ 1:1. The two regions with the highest concentration of quasars are presented as zoom-in panels and highlighted with black squares.

map appear to host a distinct population of quasars with red NIR slopes. We also examine the z–W1 color, commonly employed for quasar selection, where colors > -0.2 select quasars at $z \gtrsim 5.7$ (e.g., Yang et al. 2023), with redder values corresponding to higher redshifts. Consistently, we observe z–W1 ~ 0 near (x,y) = (5, –2), a region containing several spectroscopically confirmed quasars and extreme i–z breaks—suggesting a high-confidence selection region. Other areas hosting known quasars exhibit z–W1 values between $\sim 0 - 0.6$, still consistent with $z > 6$ quasars. Anomalously red z–W1 colors (~ 3) appear on the left side of the small island, likely caused by W1 flux contamination from nearby sources, as the lower spatial resolution of WISE compared to LS DR10 can lead to blending. Meanwhile, the region with z–W1 ~ 2.5 on the right side of the large island coincides with fainter z-band fluxes, weaker i–z breaks and two known T dwarfs, suggesting that it may be hosting a larger population of T dwarfs.

The g-band depth map offers insight into the impact of data quality on the latent space representation. Sources with deeper g-band images, characterized by lower background noise, tend to cluster along the left edges of both latent-space islands. Upon visual inspection of the optical stamps of random sources from these regions, we did not find any trend on the background level; in fact, sources with both high and low background levels are often located adjacent to each other. However we notice that the noise structure appears more peaked and irregular in the upper regions of the embedded space, becoming flatter and more homogeneous toward the bottom. This suggests that subtle observational systematics, particularly related to the noise texture, may influence the latent representation. Nevertheless, the effect is not dominant; there is no smooth gradient or sharply defined clustering of deep g-band sources, indicating that this feature likely plays only a minor role in shaping the embedded space. This effect of the imaging quality is evidenced also with the z-band PSF size map, as the smaller sizes indicating better resolution and less blurring overlap the deeper g-band areas and conversely.

The fact that the z–J color seems a strong separator of the embedded space into the two distinct "islands" is puzzling, as the J-band imaging was not included in the contrastive

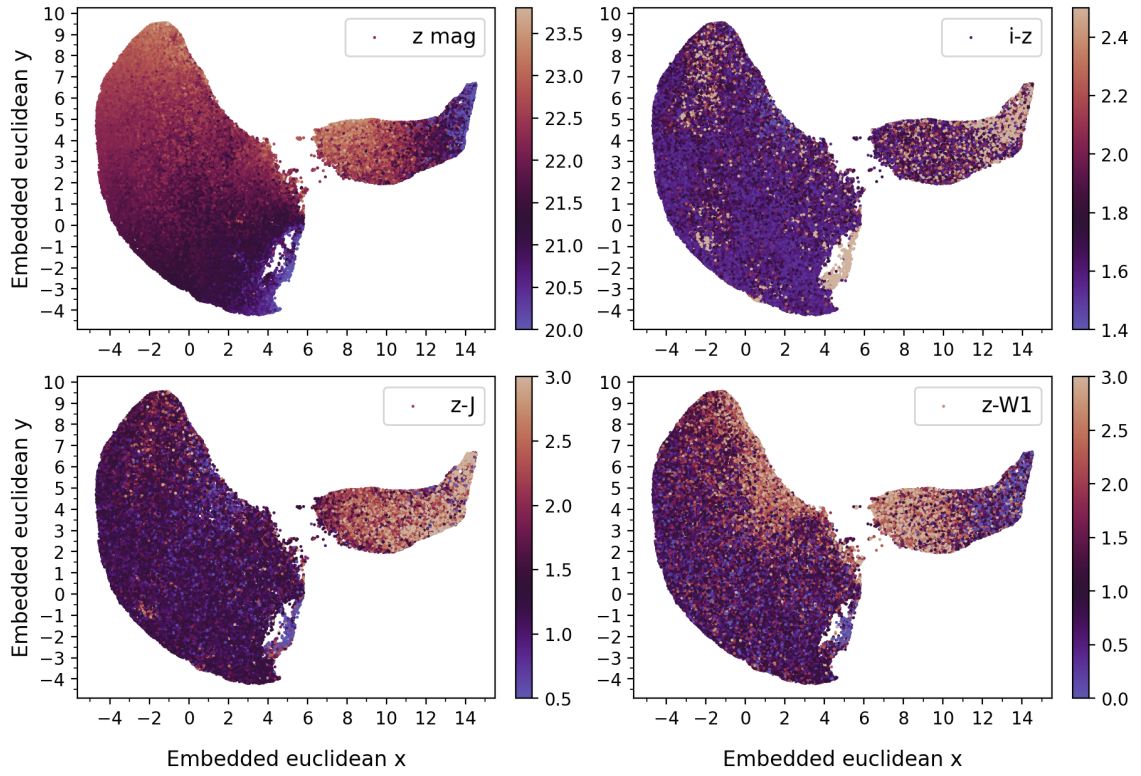

Figure 2.4: Color-coded embedded space maps showing catalog photometric features: z-band magnitude (*top left*), $i-z$ color (*top right*), $z-J$ color (*bottom left*) and z-W1 color (*bottom right*). *Top left*: magnitude gradients across both regions underscore the crucial role of the brightness in shaping the low-dimensional representation. *Top right*: Multiple regions of strong $i-z$ color indicate that while this feature may not fully dominate the embedding, it still plays a key role—evidenced by patterns and some clustering that overlap with the positions of spectroscopically confirmed quasars. *Bottom left*: The $z-J$ color map reveals populations consistent with blue quasars ($z-J \leq 0.5$), quasars with intermediate colors ($z-J \sim 1.0$–1.5) and a population with red NIR slopes ($z-J > 2$), supporting the potential of the model to uncover diverse quasar types. *Bottom right*: Red z-W1 colors seems to reflect the W1 blending due to companions and potential faint T dwarf populations.

learning input. Therefore, some other feature present in the optical images must be driving this separation. To investigate further, we proceeded to inspect the image cutouts.

We divide the embedded space into 15×15 bins, computed the mean pixel-by-pixel flux for each band for sources within each bin, and plot the resulting mean z, r and g band

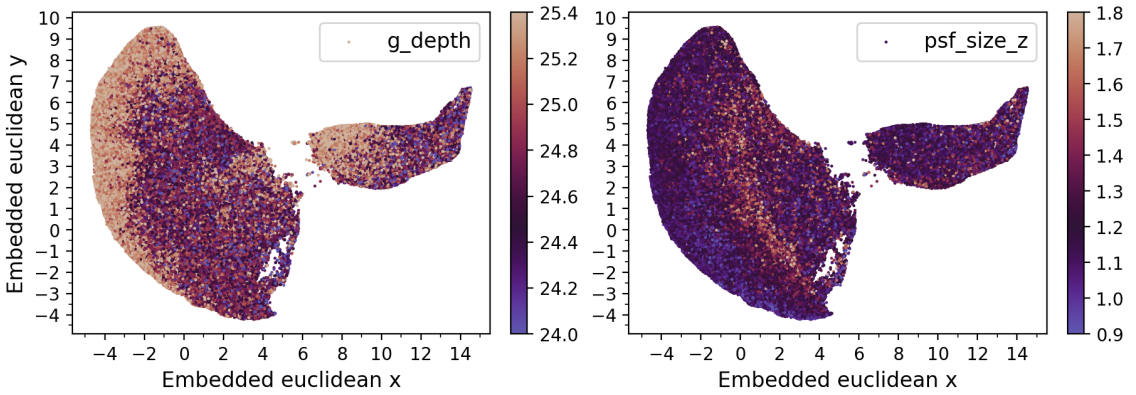

Figure 2.5: Continuation Fig 2.4. Color-coded embedded space maps showing the g-band depth (*left*), and z-band PSF size (*right*). *Left*: Subtle observational systematics, such as variations in background noise texture, may influence the latent space structure as evidence by the g-band depth pattern. *Right*: Larger PSFs point to a mix of slightly extended or noisy z-band images.

channels (RGB mapped) at the location of each bin in the latent space (see Fig. 2.6). Since the LS DR10 images are weighted coadds of multiple epochs, temporal effects such as source motion or transient events may introduce spurious morphological features. These could influence the contrastive learning algorithm, potentially driving some of the trends observed in the embedded space. To assess this possibility, we inspect single-epoch cutouts of random sources in the subsequent analysis to validate our interpretations based on the coadded data.

The requirements of our catalog-based preselection are aimed at targeting sources with non-detections in the g and r bands but detectable emission in z, without imposing any morphological constraints. When mapping the z, r and g bands into RGB colors, we notice that the stacked-average images exhibit extended emission (red shadows) and contribution from r-band (yellow shadows). The red structures populate the island on the right side of the embedded space and appear as faint rings or arcs around the central bright source, as well as elongated bright sources overlapping the stellar population distribution in Fig. 2.6.

Gaia DR3 reveals that the elongated sources are primarily stars with high proper motions. Fig. 2.7 shows the stars in the embedded space color-coded them by their

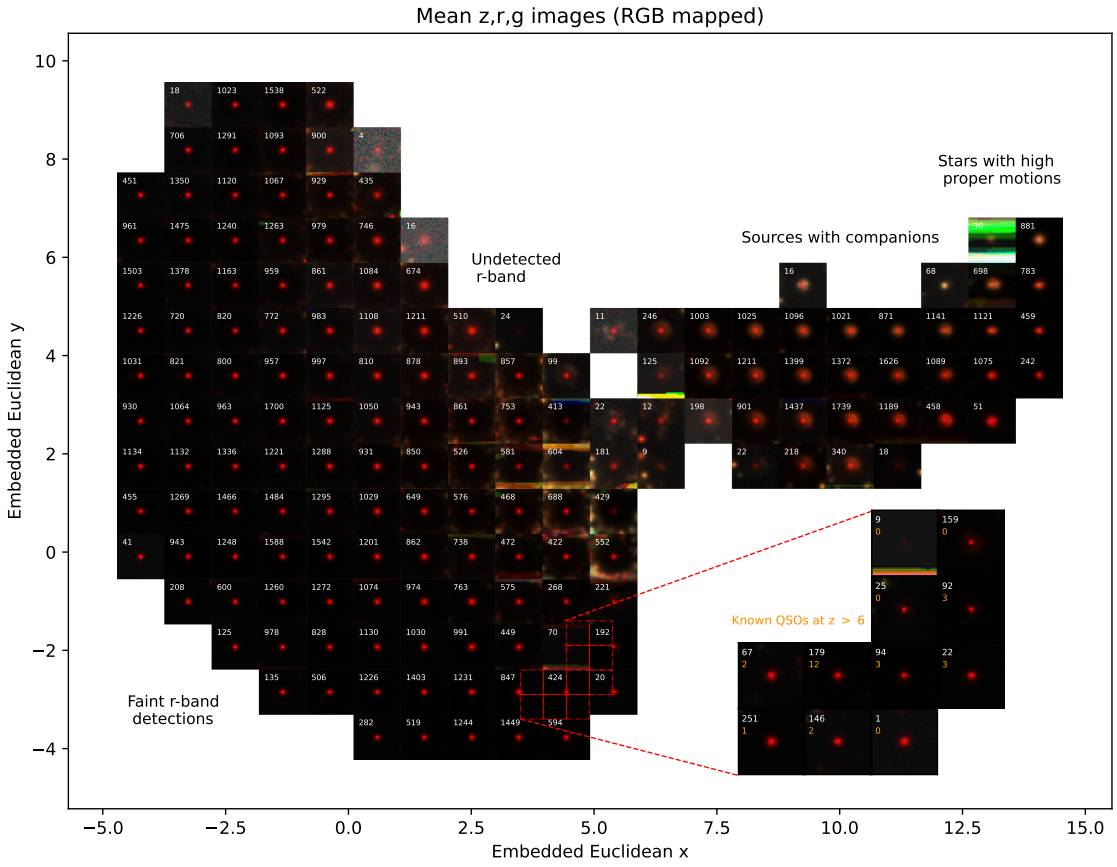

Figure 2.6: Mean pixel-by-pixel z, r and g band fluxes (RGB mapped) within 15×15 binned embedded space. The number of sources within each bin contributing to the mean are shown in white in the left upper side of each image. The red dashed lines highlight a region with the highest number of known quasars, with a zoom-in binned embedded space on the right side. The number of known quasars in each bin at the zoom-in plot is shown in orange, below the number of sources in white.

proper motion signal-to-noise ratio (SNR), defined as the proper motion divided by its uncertainty. We find consistent trends in both RA and Dec components, with increasing proper motion SNR toward the upper right of the embedded space. Inspection of individual epochs shows that these stars are in fact point-like, but appear duplicated in the Legacy Survey DR10 catalog due to positional shifts across observing epochs. In these cases, the catalog erroneously lists two nearby sources: one faint object, centered in our stamps and

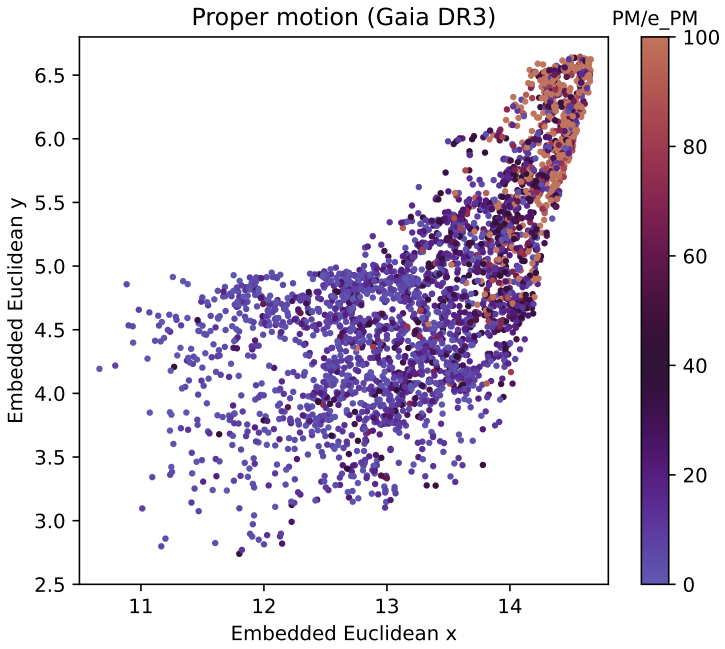

Figure 2.7: Embedded space distribution of spectroscopically confirmed stars, color-coded by the signal-to-noise ratio (SNR) of Gaia DR3 proper motion measurements. Higher SNR values correspond to stronger detections of motion, highlighting a gradient in proper motion across the embedded space.

detected only in z, mimicking the color signature of a high-redshift quasar; and a second, brighter source detected in all bands but offset in position. Because LS DR10 performs forced photometry anchored to the z-band detections, this positional mismatch leads to underestimation or loss of flux in the bluer bands, artificially producing a spectral break in otherwise stellar SEDs. This explains why some of these high proper motion stars were selected by our i-dropout preselection.

The apparent elongation seen in Fig. 2.6 is not intrinsic to individual sources but a result of our visualization method. Each pixel in the figure represents a stack-averaged cutout of all sources within a given bin in the embedded space. In regions dominated by high proper motion stars, these small positional shifts accumulate in the averaging process, producing blurred or elongated structures. While some individual exposures show mild

extension, likely due to motion, variability, or poor seeing, the dominant effect in the averaged images stems from the collective misalignment. Other moving objects, such as satellites and asteroids, can also produce real elongation in coadds, but their motion occurs on much shorter timescales (minutes).

Moving to the left side of the right island, we identify binary systems where the central target, detected in the z band or both i and z bands, is the source selected in our analysis, while its companion is visible in all bands. In this region, the source density increases significantly, and the overlapping light from multiple companions creates a ring-like effect on average. Cross-matching with our labeled sources reveals the existence of one of the quasars from the pair at $z = 5.66$ (Yue et al. 2021) with a spatial separation of $1\farcs24$ (7.3 kpc at $z = 5.66$), along with stars exhibiting low proper motions, perhaps in binary systems or by chance alignment with a companion. We do not rule out the presence of additional quasar pairs or even gravitationally lensed sources in this region. However, given the high density of sources, over 2000 within a small 1×1 latent-space unit region centered on the quasar pair, and the significant contamination from stars (about 50%, with $\sim$ 10% spectroscopically confirmed and the remaining having $>$ 90% probability of being stars), it is challenging to reliably identify quasars here. Consequently, we exclude candidates from this region for the time being. This demonstrates how environmental effects influence the latent space representation, even within a small $5''$ radius.

Since the large island on the left appears relatively homogeneous in the RGB map, we examine intensity maps of each band individually. We find that the left side of this island is dominated by faint central flux in the r-band (comparable to the noise), and a clear detection in the i-band. In contrast, the right side exhibits undetected r-band, and fainter or nonexistent central emission in the i-band. Given that the known quasars located there are at $z \sim 6 - 6.4$, it is possible that a non-negligible Lyman-α forest can contribute to the i-band below the Ly-α break. Based on the labels and the high source density in the center, we also anticipate significantly higher contamination from unlabeled T dwarfs in this region, yet to be confirmed.

We identify multiple isolated regions with high concentrations of known quasars and moderate contamination by ultracool dwarfs. The explored latent space is large and rather poorly covered by labeled data, highlighting strong potential biases among currently confirmed objects. A comprehensive strategy would require spectroscopic confirmation of > 30000 objects across the parameter space populated by quasars, given that the region with `euclidean_y` < 0 alone already exceeds this number of candidates. As this is not currently feasible, we instead propose a more immediate, quantitative criteria that can be employed to find potentially similar objects to the existing labels by computing a 2D Gaussian Kernel Density Estimation (KDE) of quasars and ultracool dwarfs based on the spectroscopically confirmed labels and normalized by the total number of labels. We iteratively adjusted the thresholds in quasar density ($\rho_{\rm QSO}$) and density ratio ($\rho_{\rm QSO}/\rho_{\rm BD}$) to define a quasar-candidate region that maximizes the recovery of known quasars while limiting the number of candidates. The final selection, $\rho_{\rm QSO} > 0.0002$ per latent-space-unit^2 and $\rho_{\rm QSO}/\rho_{\rm BD} > 0.65$, includes 54 of the 69 known quasars and results in 24,212 candidates. The quasar and brown dwarf KDE density maps, along with the density ratio map are presented in Fig. 2.8.

While our initial candidate selection focused on the lower region of the embedded space, densely populated by known quasars, a more relaxed threshold (e.g., $\rho_{\rm QSO} > 0.0001$) would extend the search toward the upper region, where $z > 6.6$ quasars are more scattered and the contamination from T dwarfs is lower. This part of the space, however, is itself densely populated (source density > 0.01) by faint i-dropout sources. Confirming them would require highly competitive observations with large-aperture ground-based telescopes and NIR spectrographs, making such targets more challenging and risky to pursue. Furthermore, the previously discussed approaches have largely been guided by existing spectroscopic samples and thus inevitably inherit their selection biases. A fully unbiased search will require probing less-explored regions of the latent space where spectroscopic labels are scarce or absent. In this context, the analysis of Figs. 2.4 and 2.5 becomes essential: by revealing trends in catalog-based features such as red NIR slopes or z–W1 colors, it offers

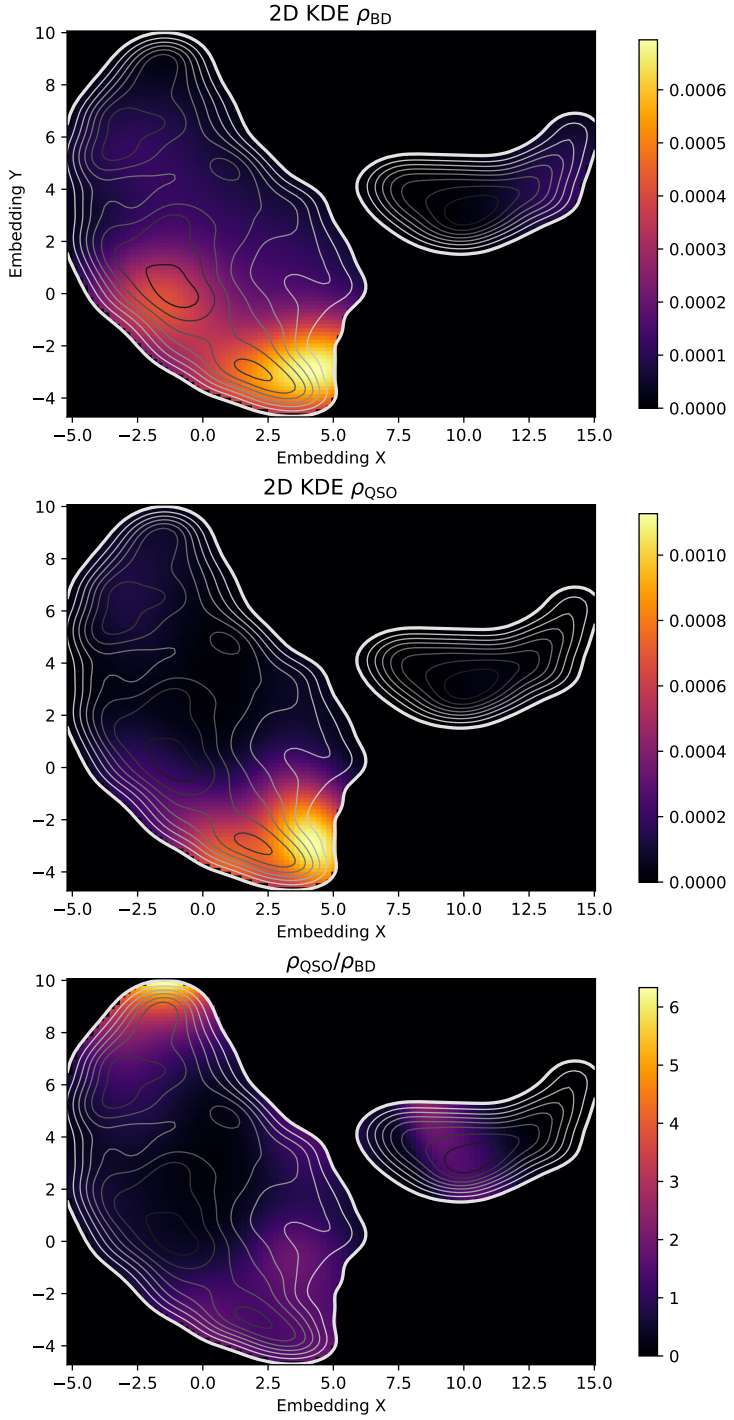

Figure 2.8: 2D KDE density maps for MLT ultracool dwarfs (*upper*), $z > 5.7$ quasar population (*central*); and quasar-to-brown-dwarf density ratio map (*lower*). White countours highlight the source density in the embedded space.

a pathway to identify promising regions even in the absence of spectroscopic labels. This paves the way for a more complete and less label-dependent quasar census in future work.

As part of developing the contrastive learning framework, we performed an initial test using a smaller sample of *i*- and *z*-dropout candidates drawn from a 14 deg^2 region of the Hyper Suprime-Cam Subaru Strategic Program DR3, making use of its five-band optical imaging. This experiment demonstrated the adaptability of our pipeline to data from different surveys and allowed us to evaluate the impact of a more flexible catalog selection and the inclusion of an additional photometric band on the latent space representation. Further details on the sample selection and embedded space are provided in Appendix B.

Encouraged by the positive outcome of adding an extra information channel with HSC, we conducted an additional test by incorporating *J*- and *K*-band images from VHS DR7 and *W*1 from allWISE into the southern sources of the LS DR10 tensor used in the main analysis. Because DECam and VHS have different spatial resolutions, image reprojection was required to ensure a homogeneous tensor shape across all bands. The steps for building this extended NIR tensor and the corresponding latent space results are presented in Appendix C. In this case, quasars overlapped significantly with M and L dwarfs, as well as stars, suggesting that additional fine-tuning will be necessary to fully exploit the added near-infrared channels.

2.2.4 Assessment of the quasar nature of promising candidates

Prioritization of quasar candidates through SED fitting

To demonstrate the effectiveness of our CL approach and optimize the usage of valuable telescope time, we try to target candidates most likely to yield successful spectroscopic confirmations. This led us to implement an additional prioritization step to select the most promising sources from the high-density quasar regions in the embedded space. We performed SED fitting on the 24212 candidates with *eazy-py* (Brammer et al. 2008)

version 0.6.9[2] using quasar templates along with templates for common contaminants such as ultracool dwarfs, stars, and galaxies.

The set of ultracool dwarfs templates include: 1) cloud-free, solar-composition, and substellar atmosphere Sonora templates; 2) a compilation of low-resolution spectra of M3-M9, L0-L9, T0-T8 dwarfs from the IRTF SpeX spectrograph; 3) low temperature, low metallicity, and cloud-free atmosphere templates from the Lowz Library; 4) AMES-Dusty and 5) AMES-Cond atmosphere models with and without dust opacity, respectively; and 5) NextGen model atmosphere grid, which covers high-temperature low-mass stars and brown dwarfs. We also consider possible contamination by main-sequence stars by incorporating templates from the TRDS Pickles Atlas. In Table 2.1 we provide technical details of the templates used for the SED fitting downselection step including: number of templates, type and wavelength coverage in μm for all the templates; and surface gravity ($\log g$), effective temperature ($T_{\rm eff}$) and metallicity ([M/H]) only for the theoretical ones. Also we report on assumptions and the cleaning process we performed to reach the final number of templates.

To account for red galaxies (early type spectra and high-z) as potential contaminants, we include: five templates from the Flexible Stellar Population Synthesis (FSPS) library, modeling the emission of a range of galaxy types such as star-forming, quiescent and dusty, with diverse star formation histories (SFHs; Conroy et al. 2009; Conroy & Gunn 2010); three galaxy templates with redshift-dependent SFHs and low dust attenuation from the CORR_SFHZ_13 library, based on JWST data of high-redshift galaxies (Larson et al. 2023; Carnall et al. 2023); and, to consider extreme emission line galaxies, the rest-frame version of the best-fit template to the NIRSpec spectrum of a $z = 8.5$ galaxy (ID= 4590 in Carnall et al. 2023), which features high-equivalent-width [OIII] and Hβ emission lines producing a distinctive U-shaped SED in the F277W, F356W, and F444W bands.

The set of quasar and AGN emission models to fit are composed by a compilation of 35 individual and combined templates of the host galaxy and AGN emission with different

[2]See `https://github.com/gbrammer/eazy-py`

Table 2.1: Templates included in *eazy-py* to account for stellar contaminants.

Name (1)	**Reference** (2)	**Type** (3)	**Number** (4)	log **g** [cm s^2] (5)	$\mathbf{T}_{\mathrm{eff}}$ [K] (6)	**Coverage** [μm] (7)
Sonora 2018	Marley et al. (2018)	Theoretical	420	[3.25, 5.5]	[200, 2400]	[0.4, 40]
Lowz Library	Meisner et al. (2021)	Theoretical	374[a]	3.5, 5	[500, 1600]	[0.1, 10]
AMES-Dusty	Allard et al. (2001) Baraffe et al. (2003) Grevesse et al. (1993) Partridge & Schwenke (1997)	Theoretical	16	4.5	[500, 2000]	[0.1, 10]
AMES-Cond	Allard et al. (2001) Chabrier et al. (2000) Grevesse et al. (1993) Partridge & Schwenke (1997)	Theoretical	11	4.5	[800,1900]	[0.1, 10]
NextGen	Allard et al. (1997) Baraffe et al. (1997, 1998) Grevesse et al. (1993) Hauschildt et al. (1999) Schryber et al. (1995)	Theoretical	8	4.5	[2200, 9800]	[0.01, 921]
SpeX Prism Library[b]	Burgasser et al. (2006) Burgasser (2014)	Empirical	149			[0.6, 2.5]
L and T dwarf data archive [c]	Knapp et al. (2004a) Chiu et al. (2006) Golimowski et al. (2004)	Empirical	148			[0.1, 2.5]
TRDS Pickles Atlas[d]	Pickles (1998b)	Empirical	131		[2950, 39810]	[0.1, 2.5]

[a] This subsample is defined by a carbon-to-oxygen ratio C/O = 0.55, vertical eddy diffusion coefficient $\log_{10} K_{zz} = 2.0$ and metallicity [M/H]= [-2.5, 1].

[b] Extended to 0.1 and 100 μm by assuming Wien and Rayleigh-Jeans emission tails, respectively.

[c] Find compilation at `http://svo2.cab.inta-csic.es/theory/newov2/index.php?`.

[d] TRDS Pickles Atlas account for all normal spectral types and luminosity classes at solar abundance, as well as metal-weak and metal-rich F-K dwarf and G-K giant stars.

Notes: Col (1): Set name; Col (2): references; Col (3): type of spectra: from a theoretical grid or empirical observed; Col (4): number of templates; Col (5): logarithm of the surface gravity; Col (6): Effective temperature; Col (7): lower and upper limits of the wavelength coverage.

contributions (Salvato et al. 2022, see their Appendix B for more details and references). To be more inclusive, we also consider less traditional objects like the "little red dot" J0647-1045, whose NIRSpec spectrum is reproduced by templates of an obscured AGN at $z = 4.50$ hosted by a star-forming galaxy (Killi et al. 2024).

For all the targets, we include the following filters in the *eazy-py* translate file: DECam g, r, i and z bands; WISE bands $W1$ and $W2$; and UKIDSS or VHS Y, J, H, and Ks bands, the latter subject to data availability. The setup parameters used are: the maximum photometric redshift `Z_MAX` = 12, and the redshift grid interval `Z_STEP` = 0.01. We run the main fitting function in multiprocessing mode to fit galaxy and quasar templates, one template at a time, with a simultaneous photometric redshift estimation. To fit the stellar templates, redshift estimation is not required, and we simply run the function `fit_phoenix_stars` with our customized templates mentioned above. The output contains χ^2 and coefficients representing the contribution to the observed SED for all the templates in the set, the photometric redshift probability distribution (including the maximum-likelihood value $z_{\rm ml}$), rest-frame U, B, V and J band-magnitudes, luminosities, and stellar masses, star formation rates and other physical parameters associated to parametric FSPS and CORR_SFHZ galaxy templates.

To calibrate our SED-fitting criterion, we tested the recovery of the 56 and 205 known quasars and ultracool dwarfs, respectively, available in our catalog. While candidate selection is restricted to a smaller latent-space region, we use the full labeled set for calibration to ensure sufficient statistics. Recovery is quantified based on the ratio of the χ^2 values obtained from quasar or AGN (quasar+host galaxy) templates, $\chi^2_{\rm quasar}$, compared to those from stellar contaminant templates (ultracool dwarfs or stars), $\chi^2_{\rm BD}$. Applying a threshold of $\chi^2_{\rm BD}/\chi^2_{\rm quasars} > 1$ yielded a precision of 0.72, a recall[3] of 0.86, and a contamination rate of 28%, successfully recovering 48 quasars (see Fig. 2.9). In the opposite case, we selected ultracool dwarfs with a precision of 0.96, a recall of 0.91 and a

[3]Precision = TP/(TP+FP) and Recall = TP/(TP+FN), where TP are the true positives, FP the false positives and FN the false negatives. They are metrics to evaluate the performance in a classification task.

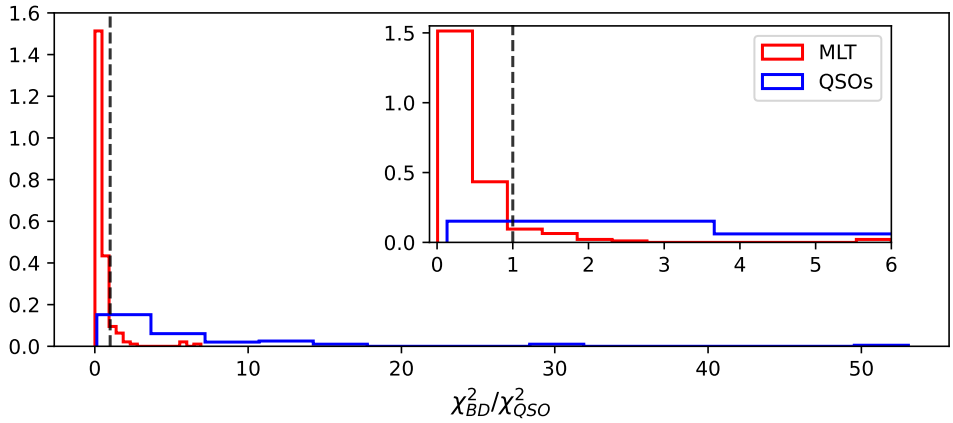

Figure 2.9: Histogram of the $\chi^2_{\rm BD}/\chi^2_{\rm quasars}$ ratio of the population of spectroscopically confirmed ultracool dwarf (red curve) and quasars (blue curve) compilations from the literature. The vertical dashed black line hightlights the ratio threshold allowing to recover 86% of the know quasars with 28% of contamination from UCD. The internal panel shows the zoom-in distribution at $\chi^2_{\rm BD}/\chi^2_{\rm quasars} < 6$.

contamination rate of 4%, recovering 185 objects.

To optimize available resources for follow-up observations, we focus our search on candidates with $\chi^2_{\rm BD}/\chi^2_{\rm quasars} > 1$, maximum-likelihood photometric redshifts $z_{\rm ml} > 5.5$ and single, narrow peaks in their redshift distribution given by z_{84}-$z_{16} < 1$, where z_{84} and z_{16} are the 84th and 16th percentiles. These constraints reduce the initial set of 24212 candidates from the high-density quasar loci in the embedded space to 1438 targets.

To establish priorities, we evaluated the observability of our targets at the time and location of the spectroscopic follow-up observing runs available to us in past semesters. We also conducted visual inspection of the g, r, i, z, J, Ks, and $W1$ stamps and SED fitting results of the 1438 targets, sorted by decreasing $\chi^2_{\rm quasars}$. Fig. 2.10 shows an example of the SED fitting, the photometric redshift probability distribution function and the stamps of a high-priority candidate that ended up being confirmed as a quasar at $z = 6.21$ (see Section 2.3.1). In this stage, we performed visual inspection of the stamps and removed 299 sources with spurious detections caused by cosmic rays, diffraction spikes, or luminous

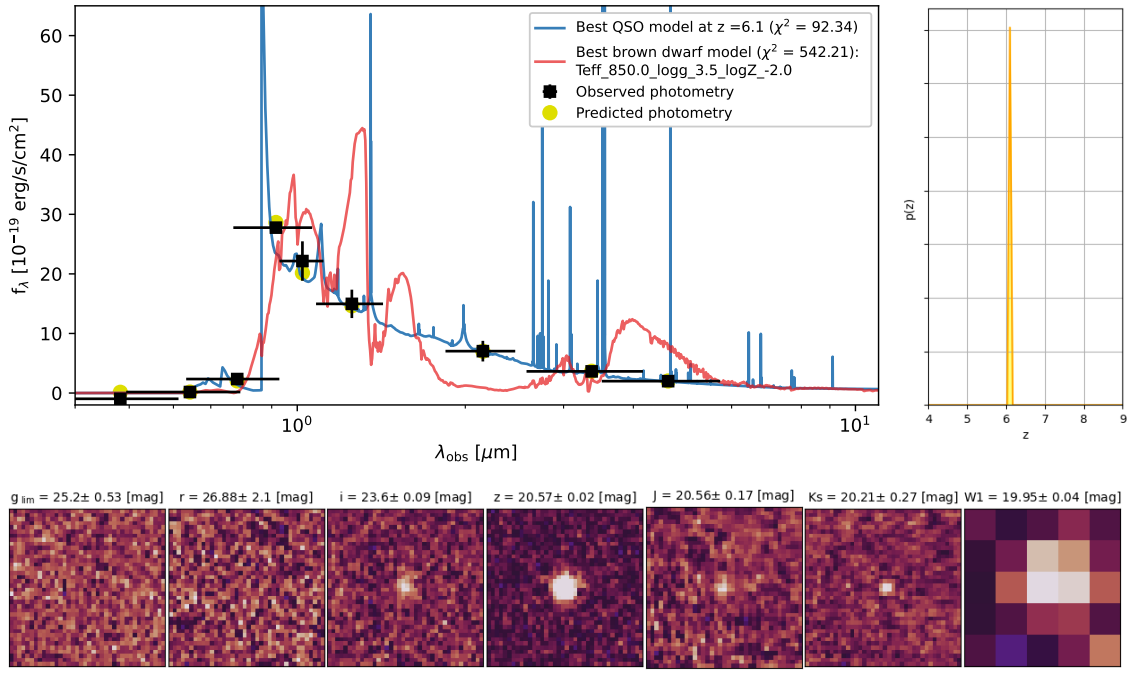

Figure 2.10: Example of an SED fitting result and the photometric redshift probability distribution function (*top panels*). LS DR10, VHS DR7 and WISE postage stamps of the photometry used (*bottom panel*) for the source LS J000-79. *Top left*: The blue curve represents the best-fit model given by the QSO1 template (see Salvato et al. 2022 for details) at redshift 6.096, while the red curve shows the best-fit brown dwarf template from Meisner et al. (2021). The observed photometry represented by black squares with error bars correspond to the DECam, VHS and WISE catalogs and their uncertainties, while the yellow circles are the expected flux densities assuming the best quasar model. *Top right*: photometric redshift probability distribution function. Despite the distribution extending over $z = [0, 12]$, we limited the plot to the range $z = [4, 9]$ for better visualization.

streaks (e.g., satellite trails). These artifacts are characterized by sharp and irregular shapes that usually show up in one band (the z-band in our case). Table 2.2 summarizes the main steps for the selection and prioritization of candidates.

Table 2.2: Downselection of candidates

Condition	Size of the sample
All LS DR10 sources	2,827,055,986
Color cut, maskbits conditions, and VHS/UKIDSS/UHS and W1 detection	165,253
Compactness criteria	130472
Density threshold in the embedded space	24212
SED fiting χ^2 ratio, $z_{\rm ml}$ and photometric redshift distribution	1438
No artifacts in optical stamps	1139

Follow-up observations

We carried out spectroscopic follow-up observations between October 2024 and July 2025 with different telescopes and instruments: the Double Spectrograph (DBSP; Oke & Gunn 1982 and the Next Generation Palomar Spectrograph (NGPS; Jiang et al. 2018) mounted at the 200-inch Hale telescope at the Palomar Observatory, the ESO Faint Object Spectrograph and Camera (EFOSC2; Buzzoni et al. 1984) at the New Technology Telescope at La Silla Observatory, the Multi-object Double Spectrograph (MODS; Pogge et al. 2010) at the Large Binocular Telescope (LBT); and the Low Dispersion Survey Spectrograph (LDSS3) at the Magellan II Clay Telescope. All observations employed long-slit spectroscopy with slit widths of $1''$, $1\farcs2$, or $1\farcs5$.

Standard calibrations such as bias substraction, flat fielding, arc lamps for wavelength calibrations, and fluxing with spectrophotometric standard stars were included in the data reduction process with the Python package *PypeIt* (Prochaska et al. 2020) and the Image Reduction and Analysis Facility (*IRAF*; Tody 1986, 1993). The resulting spectra were produced by combining individual science exposures after the background substraction and calibration process, and before the spectra extraction, as an unweighted co-add. The

final 1-D spectra are normalized to the LSDR10 z-band magnitude.

2.3 Results

2.3.1 New $z > 6$ quasars

The wavelength coverage of the optical spectra allow us to identify the presence of the Lyman-α break, confirming the nature of the quasar candidates and assessing their redshift. We performed follow-up observations for 40 candidates, 22 of them turned out to be contaminants. In Table 2.3 we compile the sample of quasar candidates that did not exhibit a Ly-α break and present a likely brown dwarf spectra. A detailed analysis of the spectrum of these sources is beyond the scope of this paper.

During our spectroscopic follow-up campaign, we discovered 15 new quasars at $z \geq 6$, and 3 quasars that were already reported in the literature[4]. Additionally, one of our candidates was confirmed by DESI as a quasar at $z = 6.03$ in its first public data release (Abdul-Karim et al. 2025). Since this source is not reported in the literature and is part of our selection, we included it and its spectrum in the following sections, with a total of 16 new quasars. The details of the follow-up campaigns where the new quasars were discovered are presented in Table 2.4.

The photometric catalog properties, along with the main results from the *eazy-py* run for the new quasars are presented in Table 2.5. In Figs. 2.11 and 2.12 we show the compilation of optical discovery spectra (observed frame), exhibing a wide range of Ly-α line profiles and continuum slopes. For redshift estimation, we extended the approach presented in Bañados et al. (2023). We fit the spectra with a set of diverse quasar templates, but now including the SMC reddening law (Prevot et al. 1984) to account for obscuration. The free parameters of the χ^2 minimization algorithm are the redshift, the normalization and the color excess (E_{B-V}). The template set is composed by the following:

[4]DESI J011553.41+031829.3 discovered by Yang et al. 2023, and J1111+0640 and J1257+1033 by Yang et al. 2024

Table 2.3: Spectroscopically confirmed to not be high-z quasars

Name	RA	DEC	Date	Telescope	Type	$z_{\rm ml}$
(1)	(2)	(3)	(4)	(5)	(6)	(7)
LS J012806.35-070949.28	22.02645	-7.16369	2024 Sep. 8	Hale/DBSP	EXP	6.89
LS J013153.69-091341.41	22.97369	-9.22817	2024 Oct. 4	Hale/DBSP	PSF	6.60
LS J023459.52-184732.42	38.74801	-18.79234	2024 Dec. 2	LBT/MODS	PSF	7.03
LS J040409.74-010111.57	61.04060	-1.01988	2024 Dec. 2	Hale/DBSP	PSF	7.07
LS J040940.43-044153.29	62.41846	-4.69797	2024 Oct. 4	Hale/DBSP	PSF	6.64
LS J082810.33-024514.58	127.04305	-2.75405	2024 Dec. 2	LBT/MODS	REX	6.29
LS J091505.37-162230.52	138.77237	-16.37507	2025 May. 1	Clay/LDSS3	PSF	7.08
LS J093306.52-004703.13	143.27715	-0.78420	2024 Dec. 5	LBT/MODS	PSF	5.96
LS J101348.72-244710.12	153.45299	-24.78617	2025 May. 1	Clay/LDSS3	PSF	7.05
LS J113244.38+153820.65	173.18492	15.63907	2025 April. 24	LBT/MODS	PSF	6.34
LS J115515.06+024328.27	178.81276	2.72452	2025 March. 5	Keck/LRIS	PSF	6.98
LS J130209.38+064549.53	195.53908	6.76376	2025 April. 24	LBT/MODS	PSF	7.02
LS J132838.18-340410.74	202.15908	-34.06965	2025 May. 1	Clay/LDSS3	REX	6.73
LS J133245.75-391414.68	203.19063	-39.23741	2025 May. 1	Clay/LDSS3	PSF	7.13
LS J140247.81+102132.87	210.69923	10.35913	2025 April. 20	LBT/MODS	PSF	6.56
LS J150722.36-232544.21	226.84315	-23.42895	2025 May. 1	Clay/LDSS3	PSF	6.01
LS J152443.10-262001.87	231.17957	-26.33393	2025 May. 1	Clay/LDSS3	PSF	6.85
LS J160255.50-082640.24	240.73126	-8.44451	2025 May. 1	Clay/LDSS3	PSF	5.94
LS J190808.07-691041.42	287.03364	-69.17804	2025 May. 1	Clay/LDSS3	PSF	5.94
LS J193812.38-570418.00	294.55158	-57.07166	2025 May. 1	Clay/LDSS3	PSF	6.97
LS J200926.15-600141.31	302.35896	-60.02814	2025 May. 1	Clay/LDSS3	PSF	6.00
LS J230435.88-800915.08	346.14950	-80.15419	2025 May. 1	Clay/LDSS3	PSF	5.93

Notes: Col (1): Brown dwarf name with convention: "LS" for DESI Legacy Survey DR10; Col (2) and Col (3): coordinates (J2000) in degrees, Col (4): Date of the observations; Col (5): Telescope and instrument used for the optical spectroscopic follow-up; Col (6): Morphological model from LS DR10 catalog: PSF=stellar, REX=round exponential galaxy, or EXP=exponential; Col (7): maximum-likelihood photometric redshift estimated by *eazy-py*.

Table 2.4: Spectroscopic follow-up observations of the 16 new quasars

Name	RA	Dec	Date	Telescope/ Instrument	Exp. time	z
	deg	deg			seconds	
(1)	(2)	(3)	(4)	(5)	(6)	(7)
LS J145109.79-044542.12	222.79078	-4.76170	2022 May. 31[a]	Mayall/DESI	3363	6.16
LS J230129.72-153020.4	345.37384	-15.50566	2024 Oct. 4	Hale/DBSP	1200	6.32
LS J020801.31-664713.7	32.00547	-66.78714	2024 Oct. 25	NTT/EFOSC2	5400	6.09
LS J222343.78-381526.8	335.93243	-38.25753	2024 Oct. 25	NTT/EFOSC2	4800	6.36
LS J000622.24-793548.1	1.55765	79.59670	2024 Oct. 28	NTT/EFOSC2	3600	6.21
LS J010449.12-685756.8	16.20466	-68.96577	2024 Oct. 28	NTT/EFOSC2	3000	6.38
LS J103511.29-051537.9	158.79705	-5.26053	2025 Feb. 16	LBT/MODS2[b]	3000	6.09
LS J113000.56+142043.97	172.50233	14.34555	2025 Feb. 21	Hale/NGPS	2700	6.28
LS J133204.89+110208.94	203.02036	11.03582	2025 Apr. 20	LBT/MODS	3600	6.11
LS J143510.65-105325.11	218.79436	-10.89031	2025 Apr. 23	LBT/MODS	2700	6.13
LS J114156.14+100636.90	175.48392	10.11025	2025 Apr. 23	LBT/MODS	2700	5.94
LS J133014.01-402508.92	202.55836	-40.41915	2025 May. 1	Clay/LDSS3	1800	6.07
LS J201119.04-443609.39	302.82931	-44.60261	2025 May. 1	Clay/LDSS3	1800	6.08
LS J203704.37-515240.27	309.26819	-51.87785	2025 May. 1	Clay/LDSS3	2800	6.25
LS J215501.13-511151.11	328.75473	-51.19753	2025 May. 1	Clay/LDSS3	1200	6.45
LS J013938.24-520945.73	14.90935	-52.16270	2025 Jul. 28	NTT/EFOSC2	3600	6.05

[a] This source was confirmed as a quasar with DESI DR1 data release.

[b] Note that although MODS is composed by two identical two-channel spectrographs, at the time of this observing run only MODS2 was available.

Notes: Col (1): Quasar name with convention: "LS" for DESI Legacy Survey DR10, and then the coordinates in HMS and DMS format. Col (2) and Col (3): coordinates in degrees, Col (4): Date of the observations; Col (5): Telescope and instrument used for the optical spectroscopic follow-up; Col (6) Total exposure time; Col (7) spectroscopic redshift estimated with the spectrum fitting with an uncertainty of 0.03 (see Section 2.3.1).

Table 2.5: Photometric properties of the new 16 quasars

Name	*i*	*z*	*Y*	*J*	*H*	*Ks*	*W1*	*W2*	$z_{\rm ml}$	$\chi^2_{\rm BD}/\chi^2_{\rm quasars}$
LS J1451-0445	23.69 ± 0.17	20.83 ± 0.01	20.94 ± 0.23	20.63 ± 0.17	20.21 ± 0.19		20.02 ± 0.06	20.02 ± 0.14	6.1	4.37
LS J2301-1530	24.75 ± 0.80	20.65 ± 0.03	20.16 ± 0.08	20.07 ± 0.10		19.74 ± 0.18	19.78 ± 0.05	19.91 ± 0.13	6.37	2.89
LS J0208-6647	23.44 ± 0.14	20.17 ± 0.01		20.72 ± 0.10		19.85 ± 0.14	19.18 ± 0.02	18.82 ± 0.03	6.11	5.53
LS J2223-3815	23.56 ± 0.20	20.57 ± 0.03		19.84 ± 0.04		19.77 ± 0.12	19.92 ± 0.05	19.92 ± 0.11	6.30	2.25
LS J0006-7935	23.60 ± 0.09	20.57 ± 0.02	20.58 ± 0.16	20.56 ± 0.17		20.21 ± 0.27	19.95 ± 0.04	19.92 ± 0.08	6.10	5.87
LS J0104-6857	23.75 ± 0.21	20.50 ± 0.02	20.83 ± 0.19	20.89 ± 0.22		19.81 ± 0.19	19.90 ± 0.04	19.33 ± 0.05	6.11	3.71
LS J1035-0515	23.34 ± 0.22	20.14 ± 0.01	20.51 ± 0.15	19.68 ± 0.07	19.14 ± 0.07	18.55 ± 0.05	18.32 ± 0.01	18.19 ± 0.03	6.43	2.16
LS J1130+1420	24.66 ± 0.39	21.05 ± 0.03	21.13 ± 0.21		20.61 ± 0.20	20.48 ± 0.24	20.31 ± 0.08	20.20 ± 0.16	6.20	21.09
LS J1332+1102	24.17 ± 0.36	20.80 ± 0.04	20.50 ± 0.13	20.29 ± 0.11	20.03 ± 0.13	19.85 ± 0.12	19.95 ± 0.05	20.38 ± 0.18	6.22	2.92
LS J1435-1053	24.64 ± 0.53	21.27 ± 0.04		20.49 ± 0.21			20.51 ± 0.09	21.22 ± 0.42	6.21	1.24
LS J1141+1006	24.08 ± 0.31	21.14 ± 0.02	21.19 ± 0.22		20.95 ± 0.27	20.58 ± 0.22	20.93 ± 0.14	21.50 ± 0.56	6.07	2.70
LS J1330-4025	22.80 ± 0.08	20.13 ± 0.01		20.58 ± 0.25		19.94 ± 0.25	19.72 ± 0.04	19.46 ± 0.07	6.03	11.61
LS J2011-4436	23.90 ± 0.26	20.95 ± 0.03		20.44 ± 0.10		19.59 ± 0.10	19.32 ± 0.03	19.26 ± 0.06	6.31	2.90
LS J2037-5152	25.55 ± 0.70	21.08 ± 0.04		21.00 ± 0.17		20.32 ± 0.21	20.40 ± 0.08	19.75 ± 0.09	6.18	1.30
LS J2155-5111	25.45 ± 0.34	21.35 ± 0.03		21.05 ± 0.22			20.65 ± 0.09	20.41 ± 0.16	6.24	2.51
LS J0139-5209	23.24 ± 0.08	20.65 ± 0.01		20.71 ± 0.12		20.27 ± 0.23	20.35 ± 0.07	20.64 ± 0.19	6.05	1.83

Notes: Col (1): Abbreviated quasar name; Col (2)-Col (3): optical magnitudes from LS DR10, Col (4)- Col (7): NIR magnitudes from UKIDSS DR11/VHS DR7/UHS DR2; Col (8) -Col (9): MIR magnitudes from WISE; Col (10): maximum-likelihood photometric redshift estimated by *eazy-py*; Col (11): stellar contaminant-to-quasar χ^2 ratio used for the prioritization of the candidates.

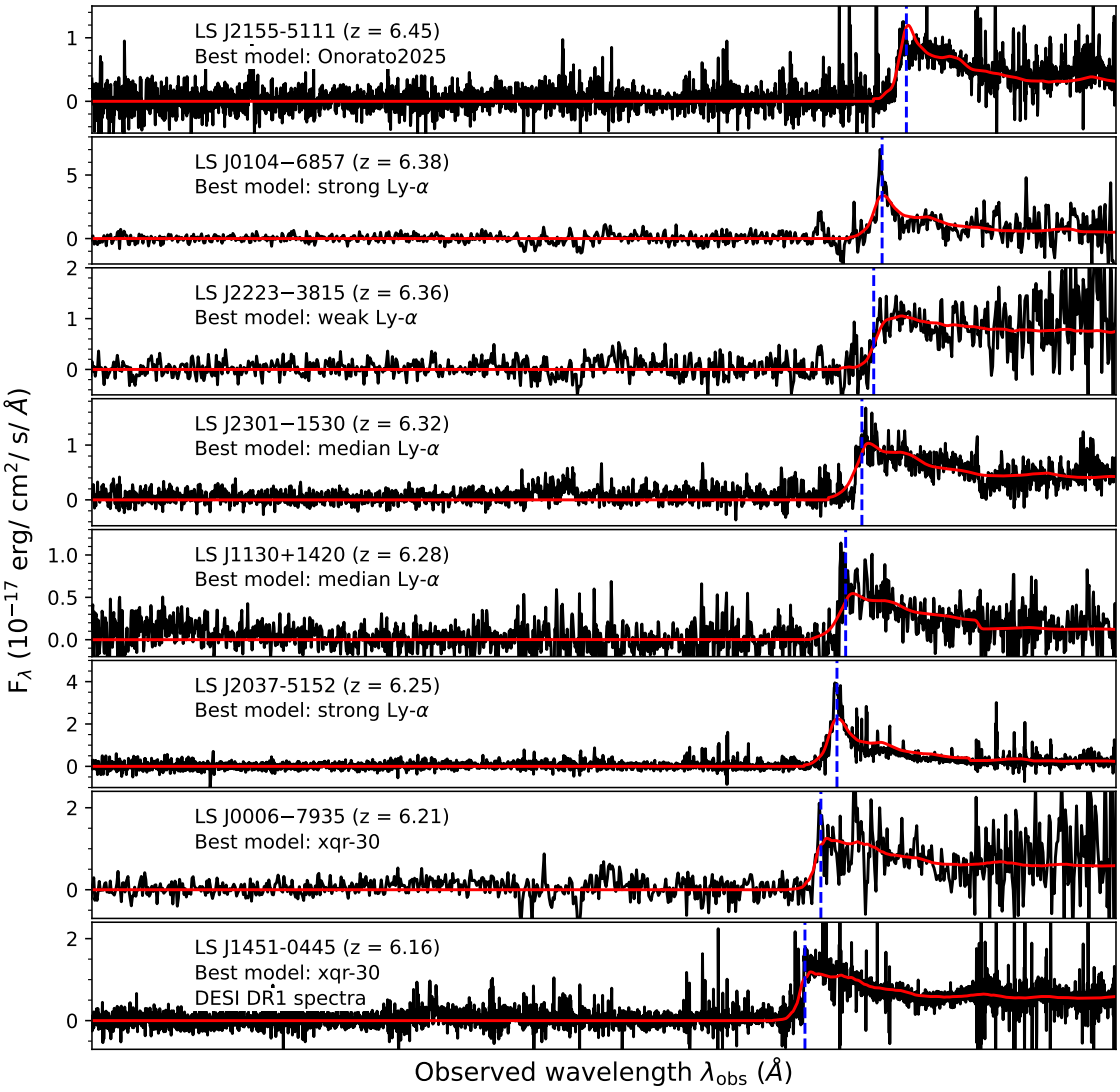

Figure 2.11: Optical discovery spectra of the 16 new quasars reported in this work, including the quasar LS J1451-0445 from the DESI data release (Abdul-Karim et al. 2025). The black curves show the observed spectra, the red curves the best-fit quasar templates used for redshift estimation, and blue dashed lines indicate the adopted best-fit position of Ly-α. Each panel lists the source name, spectroscopic redshift, and the best-fit template in the upper left corner. The typical redshift uncertainty is 0.03 (see Section 2.3.1). Quasars are sorted from top to bottom by decreasing redshift.

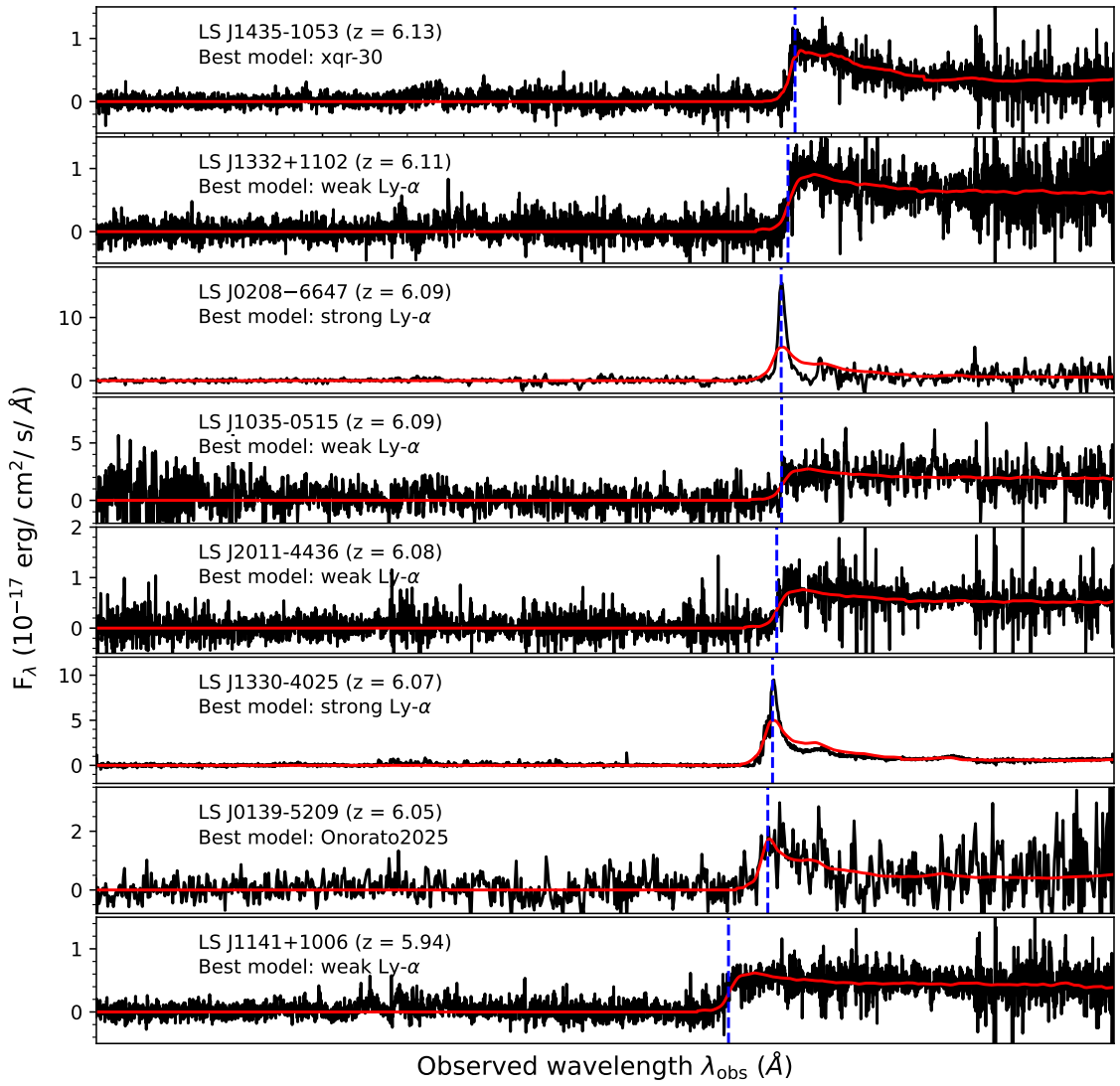

Figure 2.12: Continuation Fig. 2.11.

- *median* Ly-α, which is the median spectrum of 117 $z \sim 6$ quasars from the Panoramic Survey Telescope & Rapid Response System 1 (PS1; Kaiser et al. 2002, 2010) presented in Bañados et al. (2016).

- *strong* Ly-α, which is the median PS1 spectrum of $z \sim 6$ quasars with the 10% largest rest-frame Ly α + NV equivalent width from Bañados et al. (2016).

- *weak* Ly-α, which is the median PS1 spectrum of the $z \sim 6$ quasars with the 10%

smallest rest-frame Ly α + NV equivalent width from Bañados et al. (2016).

- onorato2025, which is the weighted mean of 33 $z > 6.5$ spectra presented in Onorato et al. (2025).

- xqr-30, which is the median of 42 $z \sim 6$ high quality quasars spectra observed with X-Shooter and reported in D'Odorico et al. (2023).

We selected the combination of parameters and template producing the lowest χ^2 for each source in the wavelength range between $8000 - 8800$ Å, depending on the Ly-α location, and ~ 9600 Å, avoiding the highly noisy regions bluewards of Ly-α and the red extreme of the spectra. The resulting redshift, χ^2 and name of the best template are presented in Table 2.6. We do not include E_{B-V} as all the values are < 0.04, indicating that none of the sources seems to be significantly reddened. We assume a median uncertainty on the spectral redshift estimate with the fitting method of 0.03 as presented by Bañados et al. (2023).

The sample of quasars spans a redshift range $z = [5.94, 6.45]$ and displays diverse types of Ly-α emission, from weak to strong lines. Four quasars (LS J010-68, LS J203-51, LS J020-66 and LS J133-40) exhibit prominent and relatively narrow, high EW Ly-α lines. The reduced χ^2 values indicate that some models underfit the data, notably for LS J0208-6647, where the fit fails redward of the Ly-α line, and for LS J1035-0515, which has the noisiest spectrum in the sample. Conversely, nine sources exhibit very low reduced χ^2 values (below 0.2), suggesting uncertainties or degeneracies in the fitted parameters, likely due to low S/N and poor spectral quality.

2.3.2 Physical properties of the newly discovered quasars

The availability of optical, NIR and MIR broad-band photometry allows us to carry out a first diagnostic on the quasar parameter space that we are exploring. We start with widely used physical quantities, such as the bolometric luminosity and the magnitudes

Table 2.6: Redshift and physical properties of the new quasars

Name	z	χ^2/dof	**Template**	**log** λL_λ (1350 Å) erg s^{-1}	**log** $L_{\rm bol}$ erg s^{-1}	$\mathbf{m}_{1450}$	$\mathbf{M}_{1450}$	**FWHM (Ly-α)** km/s	$\mathbf{EW}_{\mathbf{Ly\alpha+NV}}$ Å
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
LS J2155-5111	6.45	0.11	Onorato25	45.31	45.89	21.27	−25.53		82
LS J0104-6857	6.38	0.43	*strong* Ly-α	45.65	46.22	20.43	−26.36	1613	90
LS J2223-3815	6.36	0.12	*weak* Ly-α	45.62	46.20	20.49	−26.29		15
LS J2301-1530	6.32	0.04	*median* Ly-α	45.59	46.17	20.58	−26.20		160
LS J1130+1420	6.28	0.04	*median* Ly-α	45.42	46.00	20.98	−25.78		87
LS J2037-5152	6.25	0.14	*strong* Ly-α	45.41	45.99	21.01	−25.74	2656	212
LS J0006-7935	6.21	0.28	xqr-30	45.61	46.19	20.51	−26.24		32
LS J1451-0445	6.16	0.22	xqr-30	45.60	46.19	20.51	−26.22		153
LS J1435-1053	6.13	0.03	xqr-30	45.32	45.91	21.21	−25.52		247
LS J1332+1102	6.11	0.08	*weak* Ly-α	45.51	46.09	20.74	−25.98		22
LS J0208-6647	6.09	2.82	*strong* Ly-α	45.76	46.34	20.12	−26.60	1146	170
LS J1035-0515	6.09	3.51	*weak* Ly-α	45.77	46.35	20.08	−26.63		8
LS J2011-4436	6.08	0.09	*weak* Ly-α	45.45	46.03	20.90	−25.82		45
LS J1330-4025	6.07	0.38	*strong* Ly-α	45.78	46.36	20.08	−26.63	1713	149
LS J0139-5209	6.05	0.49	Onorato25	45.41	45.99	21.00	−25.71		167
LS J1141+1006	5.94	0.04	*weak* Ly-α	45.36	45.94	21.10	−25.58		25

Notes: Col (1): Abbreviated quasar name; Col (2): spectroscopic redshift estimated with the spectrum fitting with an uncertainty of 0.03 (see Section 2.3.1), Col (3): reduced χ^2 of the best-fitted template; Col (4): Name of the best-fit template; Col (5): monochromatic luminosity at rest-frame 1350 Å; Col (6): bolometric luminosity, Col (7-8): apparent and absolute magnitudes at rest-frame 1450 Å.

at rest-frame wavelength 1450 Å, which are useful for bolometric corrections and quasar characterization. However, we do not derive these quantities directly from the best-fit spectral templates due to several limitations. First, the spectra are generally noisy, and our fits exclude data at observed wavelengths > 9600 Å (corresponding to rest-frame wavelengths > 1385 Å), which affects the reliability of continuum slope estimates. Consequently, flux predictions in these regions are not robust. Furthermore, our current templates do not reproduce succesfully the sharp Ly-α emission features along with the underlying continuum, often leading to underestimated fluxes in the rest-frame UV regime for the peaky Ly-α systems. Besides, we wanted to use a method that would be consistent for all the quasars, regardless of the quality of the single spectrum, and consistently with the literature for fair comparisons.

We follow the approach presented in Bañados et al. (2016), Mazzucchelli et al. (2017) and Belladitta et al. (2025), modeling the quasar continuum as a power-law with slope $\alpha_\nu = -0.44$ (Berk et al. 2001). Since all our sources have reliable z-band photometry, we use this magnitude at the pivot wavelength 9168.85 Å (for DECam) to extrapolate the apparent magnitude at 1450 Å (m_{1450}) and the monochromatic luminosity at 1350 Å ($\lambda L_\lambda(1350\,\text{Å})$). While m_{1450} traces the accretion disk output, $\lambda L_\lambda(1350\,\text{Å})$ provides a basis for estimating the bolometric luminosity. To compute the absolute magnitude M_{1450}, we adopt the redshifts derived from our spectral fitting. All these quantities are listed in Table 2.6.

To estimate the total radiative output due to SMBH accretion, we adopt a bolometric correction based on the monochromatic luminosity at 1350 Å following Shen et al. (2008):

$$L_{\text{bol}} = (3.81 \pm 0.38) \times \lambda L_\lambda(1350\,\text{Å}). \tag{2.7}$$

For the sources exhibiting prominent Ly α emission (LS J0104–6857, LS J2037–5152, LS J0208–6647, and LS J1330–4025), we estimate the FWHM of the Ly α line. First, we shifted the spectra to the rest-frame. Then, we model the continuum emission following the approach of Diamond-Stanic et al. (2009), fitting a power law of the form $f_\lambda = C\lambda^\beta$,

where $\beta = \alpha - 2$, to spectral regions free of strong emission lines (at rest-frame wavelengths 1285–1295, 1315–1325, 1340–1375, 1425–1470, 1680–1710, 1975–2050, and 2150–2250 Å). However, because our spectra are noisy at rest-frame > 1385 Å, we cannot reliably constrain the continuum slope. Therefore, we fix the power-law index to $\beta = -1.56$, consistent with the quasar slope we adopted to estimate m_{1450} and $\lambda L_\lambda(1350\,\text{Å})$. After subtracting the continuum component, we locate the peak of the Ly α emission and fit a single Gaussian profile to the red side of the line. The blue side is strongly affected by IGM absorption and resonant scattering, which distorts the line profile and makes it unsuitable for fitting. Fitting only the red wing provides a more reliable estimate of the line width.

The resulting FWHM values are in the range $1147 - 2657$ km s^{-1}. According to the classification of Matsuoka et al. (2018), which defines narrow Ly α quasars as those with FWHM < 500 km s^{-1}, these sources do not meet the criterion. However, the prominent Ly α emitters represent 25% of our sample and display significantly narrower line widths compared to the rest. As virial black hole mass estimates scale with the square of the broad emission line width, narrower lines may indicate lower black hole masses. Near-infrared spectroscopy will be essential to confirm this interpretation by providing measurements of other broad emission lines and estimates of the black hole mass.

In order to systematically investigate the emission-line properties of our quasar sample, we compute the EW of the blended emission from Ly α λ1216 Å, N v λ1240 Å and Si ii λ1263 Å, following the approach by Diamond-Stanic et al. (2009) and Bañados et al. (2016) applied to SDSS DR5 and PS1 high-redshift quasars. We first determine the continuum using the same power-law model described earlier and then compute the EW by integrating the flux excess over the continuum in the rest-frame wavelength range 1160–1290 Å.

We find that only two of our quasars meet the definition of weak emission-line quasars (WLQs), with EW < 15.4 Å as set by Diamond-Stanic et al. (2009). This represents 12.5% of our sample, a fraction comparable to the 13.7% reported by Bañados et al. (2016) for 124 PS1 quasars at $5.6 \lesssim z \lesssim 6.7$. Although the size of our quasar sample prevents a robust statistical analysis, we compare our EW distribution of Ly α+ N v+ Si ii to the

log-normal fits derived for SDSS DR5 quasars at $3 < z < 5$ and PS1 quasars. Fig. 2.13 presents the normalized EW distribution of our 16 quasars, along with the SDSS and PS1 log-normal models from Diamond-Stanic et al. (2009) and Bañados et al. (2016), respectively. While 56.25% of our sample with log(EW/Å) < 2 closely follows the PS1 distribution, a significant fraction (43.75%) exhibits enhanced line strengths, occupying the high-EW tails of both reference distributions. This fraction is notably higher than the 17% and 12% predicted by the SDSS and PS1 fits, respectively.

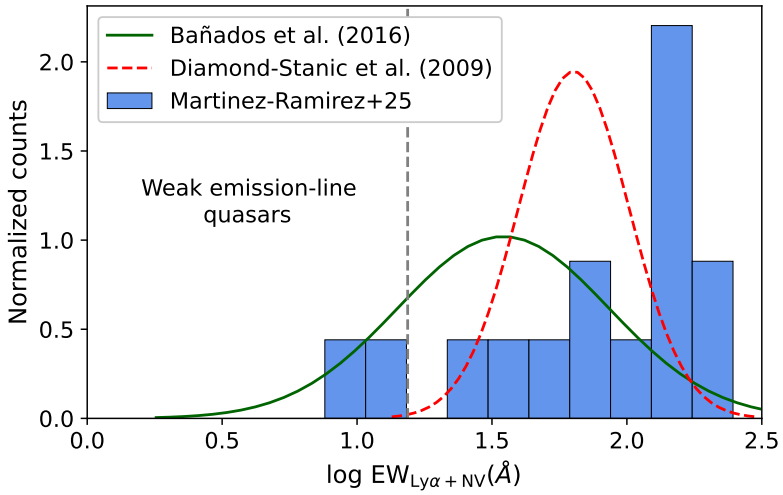

Figure 2.13: Normalized EW(Ly α+ N v) distribution for the 16 quasars at $z = 5.93 - 6.45$ identified through our CL selection. The solid green and dashed red curves show the best-fit log-normal distributions for PS1 quasars (Bañados et al. 2016) and SDSS quasars (Diamond-Stanic et al. 2009), with mean values of log(EW/Å) = 1.542 and 1.803, and standard deviations of 0.391 and 0.205 Å, respectively. The vertical dashed gray line marks the limit for WLQs (EW < 15.4 Å; Diamond-Stanic et al. 2009). While most of our sample overlaps with the PS1 distribution at log(EW/Å) < 2, a substantial excess appears in the high-EW tail, potentially revealing a distinct population of strong-line quasars.

Diamond-Stanic et al. (2009) reported an increasing fraction of WLQs with redshift, which implies that our sample would be expected to show overall lower EW values than SDSS and be broadly consistent with PS1. We performed an Anderson–Darling test

comparing our sample to the PS1 EW distribution, obtaining a statistic of 56.92 and a p-value of 0.001. This indicates a substantial difference between the two distributions and a very low probability of obtaining such a statistic if our sample were drawn from the PS1 distribution, leading us to reject the null hypothesis. The EW measurements therefore suggest that our selection method is recovering a population of quasars with systematically stronger Ly α+N v+ Si II emission than typically found in previous surveys. The EW values, along with the other physical quantities derived in this section are summarized in Table 2.6. A sample with at least several dozen quasars ($\gtrsim$ 50) within this selection framework will be crucial to determine whether our selection is probing a different less-explored region of quasar parameter space, or if the current results arise from small-number statistics.

2.3.3 Comparison with the literature quasars

To place our newly discovered sources in context, we examine color–color diagrams to assess the parameter space occupied by them (see Fig. 2.14). Table 2.5 shows that all the sources from our sample have valid i, z, $W1$ and $W2$ band magnitudes, while J is available for 14 of the sources. We therefore compute colors using these bands to enable a more homogeneous comparison.

13 of the 16 quasars exhibit strong spectral breaks with $i - z > 3$ and $z - \mathrm{W1} > 0.5$, overlapping entirely with the region populated by T dwarfs. While a significant fraction of known $z \sim$ 6–6.7 quasars occupy the same region of color space, previous selections typically rely on additional data or strict constraints to separate them from contaminants. In contrast, our method successfully identifies quasars in regions heavily contaminated by ultracool dwarfs by leveraging the i and z band imaging as input. This demonstrates the effectiveness of our contrastive learning approach in isolating quasars without requiring extensive multi-wavelength information.

The $z - J$ colors indicate that $\sim$ 38% of the newly discovered quasars exhibit mildly red NIR slopes ($z - J > 0.4$), though still consistent with the colors of known quasars at similar redshifts. We noticed a strong concentration of sources with $0.4 < z - J < 0.8$.

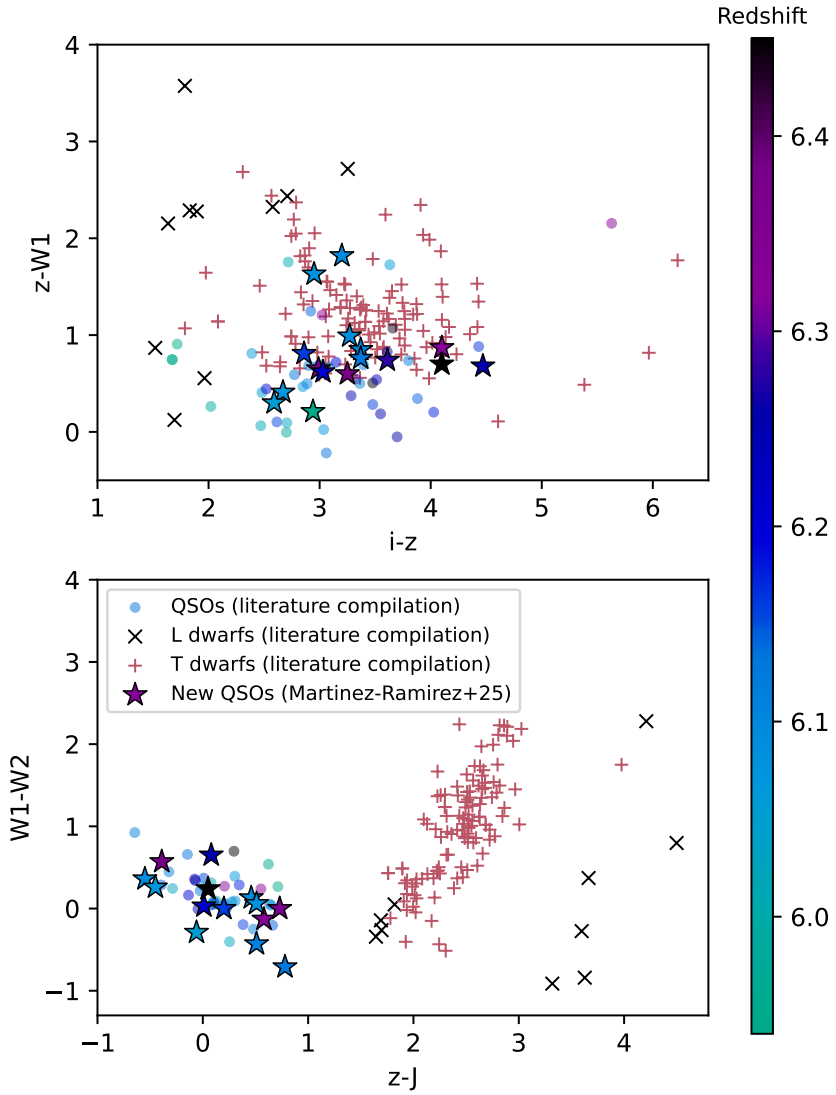

Figure 2.14: Color–color diagrams showing the positions of the newly discovered quasars (stars), and known high-redshift quasars at the same redshift range of our discoveries $5.94 \leq z \leq 6.45$ (filled circles) both color-coded by redshift; and L, and T dwarfs (black and red crosses, respectively) used as labels in the embedded space. Top: $i - z$ vs. $z - \mathrm{W1}$. Bottom: $z - J$ vs. $\mathrm{W1} - \mathrm{W2}$. While $i - z$ vs. $z - \mathrm{W1}$ shows a total overlap of the new quasars with the T dwarf population, the $z - J$ color shows a clear distinction.

For quasars within this color range, we compute an average redshift of $z = 6.27$ for the known sample from the literature, compared to $z = 6.18$ for our discoveries. This offset

may suggest that our selection is recovering a population of quasars with slightly redder NIR colors. While this is not a direct outcome of the CL approach itself, it likely reflects the more flexible constraints adopted during our catalog-based preselection.

Quasars with redder NIR slopes are more likely to overlap with the locus of UCD in traditional color–color diagrams, and are therefore often excluded or deprioritized by traditional selections that require flat or blue continua redward of Ly α to minimize contamination (e.g., Bañados et al. 2016; Wang et al. 2017; Findlay et al. 2012). Physically, redder slopes can arise from several effects: modest dust extinction in the quasar host galaxy, an intrinsically softer continuum from the accretion disk, or strong line contributions in the J band (e.g., from C IVλ1549 or C III]λ1909 depending on redshift). Previous studies have suggested that red quasars may represent either dust-obscured young transitional phases of SMBHs growth (e.g., Glikman et al. 2024) or quasars viewed at larger inclinations (e.g., Martín et al. 2020). Thus, the concentration of our new discoveries in this color regime may be pointing toward a less-explored population systematically missed by past color-cut selections.

To understand why our newly discovered quasars were overlooked by previous selections, we investigated the survey coverage and selection criteria applied to these sources. None of the 16 newly confirmed quasars have X-ray or radio detections[5], excluding them from selections relying on these bands. Also, of the 16 only three overlap with the DES footprint, three with PS1, and none with SDSS DR12, making incomplete spatial overlap a primary factor. All of our sources are classified as PSF-like, indicating that traditional selection methods based on point-source morphology would not discard them. Beyond coverage, color and S/N requirements along with prioritization strategies also explain the missing sources. In the following, we provide a breakdown of the different selections from the literature:

[5]The X-ray and radio dataset checked are: eROSITA DR1, LOFAR DR2, EMU, RACS (low,mid,high), FIRST, NVSS, VLASS, and Meerkat-MALS.

PS1-based selections: The DESI selection by Yang et al. (2023) relied on PS1 i-band imaging, which immediately excludes 13 of our quasars due to lack of PS1 coverage. Of the remaining three, LS J1332+1102 failed their $W1 - W2 > -0.14$ color cut, and LS J2301-1530 and LS J1035–0515 were excluded by the flat continuum criterion ($z_{\rm P1} - y_{\rm P1} < 1$). Similarly, the PS1-based selection of Bañados et al. (2016) included only three of our quasars. LS J1332+1102 failed their $-0.2 < W1 - W2 < 0.85$ color prioritization criterion. Although LS J2301-1530 and LS J1035–0515 meet the selection thresholds, they may have been deprioritized due to low S/N, or other quality cuts. The selection of $z \sim 7$ quasar candidates by Wang et al. (2019), combining DESI Legacy imaging survey (DELS), PS1, UKIRT/UHS, and WISE data, also includes only three of our quasars. All of them (LS J2301-1530, LS J1332+1102, and LS J1035–0515) were excluded by the z-dropout requirement ($z_{\rm P1} - y_{\rm P1} > 1.5$).

Comprehensive multi-method selection: The recent work by Belladitta et al. (2025) explore a broader set of quasar selections strategies, including: 1) PS1 i-dropouts with NRAO VLA Sky Survey + ALLWISE detections (Belladitta et al. 2023), 2) Radio-selected quasars candidates from the Million Quasar catalog (Flesch 2023), 3) PS1-based i-dropout selection from Bañados et al. (2023), 4) $z \gtrsim 6.6$ selection using DELS + PS1 and 5) $z > 6.5$ selection from PS1 and ALLWISE by Belladitta et al. (2025), 6) DES/VHS/CatWISE i-dropouts (Wolf et al. 2024), and 7) DES-based photometric candidates (Yang & Shen 2022).

None of our quasars are detected in radio, excluding them from selections (1) and (2). For selection (3), LS J2301-1530, LS J1332+1102, and LS J1035–0515 failed the $z_{\rm P1} - y_{\rm P1} < 0.5$ cut. Selection (4) excluded LS J2301-1530 and LS J1332+1102 based on $z_{\rm P1} - z_{\rm DE} > 0.8$, and LS J1035–0515 based on $z_{\rm DE} - y_{\rm P1} > 0.5$. Selection (5) impose the same criteria from Wang et al. (2019), then it excludes the same three sources (LS J2301-1530, LS J1332+1102, and LS J1035–0515) due to the same z-dropout condition. Selection (6) missed LS J2037–5152 and LS J2155–5111 due to the absence of i-band

detections, which are necessary for their color criteria. Despite the fact that LS J0139-5209 meets the photometric constraints, it may have been rejected at the SED fitting stage or deprioritized due to quality flags. Finally, none of our sources were matched to the candidates catalog from selection (7).

DECaLS and DES selections: Wang et al. (2017) combined DECaLS with SDSS i-band photometry; however, their sample does not overlap with any of our recently confirmed quasars. Reed et al. (2017), using DES data, included LS J2037–5152, LS J2155–5111 and LS J0139-5209 in their footprint: the first two were excluded for being fainter than their $z_{\rm PSF} \leq 21$ magnitude threshold; the last one met their color selection, but likely was never followed up with spectroscopy to confirm it.

Contaminant morphologies and classification notes

Our catalog-based preselection does not impose strict morphological constraints on i-dropouts, allowing for the inclusion of both point-like and extended sources. All confirmed quasars in our sample are classified as point-like ("PSF" type) in the LS DR10 catalog. Among the contaminants, 19 are also point-like and are thus consistent with being ultracool dwarfs. One contaminant is modeled with an exponential profile, and two are best fit by round exponential galaxy models; suggesting they could be compact, dusty red galaxies at low redshift or other types of extended contaminants. Spectroscopic analysis is required to confirm their nature.

Interestingly, about half of the contaminants have photometric redshifts $z_{\rm phot} \gtrsim 6.7$, a range where contamination from T dwarfs is particularly common. This further supports the interpretation that a significant fraction of our contaminants are ultracool dwarfs misclassified due to their colors and morphology.

2.4 Conclusions

In this work, we report the discovery of 16 new quasars at z = 5.94–6.45 using an self-supervised contrastive learning (CL) approach applied to DESI Legacy Survey DR10 optical images. Our method combines a flexible catalog-based preselection of i-dropouts, free from strict morphological or color constraints, with representation learning, enabling us to recover sources that would otherwise be overlooked, including quasars with slightly red NIR slopes, softer Lyα breaks, and even potential quasar pairs. Such successful outcomes stem from the interplay of our modern self-supervised image-based method with traditional methods such as dropout color cuts, SED fitting, and aperture photometry constraints.

Starting from a sample of 165,253 i-dropout candidates, we trained a CL model using only four DESI LS DR10 bands and produced a low-dimensional latent space representation where quasars naturally cluster. Within this embedded space, we identified overdensities of spectroscopically confirmed quasars at $z \sim 6.0-6.3$, with contamination rates of $\sim 50\%$ from known M, L, and T dwarfs from the literature. We also uncovered trends linked to brightness and morphology, the primary drivers of the latent representation: point-like and elongated sources occupy distinct regions, and a smooth gradient in brightness is observed. Importantly, quasars appear across a broad range of $i - z \sim 1.5 - 2.5$ and $z - J$ colors, including regions heavily contaminated by ultracool dwarfs, underscoring the strength of CL in disentangling complex parameter spaces.

To prioritize candidates for follow-up, we coupled our CL approach with SED fitting, achieving a $\sim 45\%$ success rate in our spectroscopic campaigns. Among the confirmed quasars, we find a diversity in Lyα profiles, including four with relatively narrow lines (FWHM $\sim$ 1146–2656 km s^{-1}), and six with red NIR slopes ($z - J > 0.4$) with slightly lower average redshifts compared to known quasars with similar colors. Our EW analysis of the Lyα+N v+ Si II complex reveals that $\sim 44\%$ of our sample exhibits enhanced line strenghts. To robustly determine whether our selection systematically recovers quasars

with high EWs, a larger spectroscopic sample, on the order of 50–100 quasars, will be required. Such numbers will soon be within reach with DESI and 4MOST optical spectra, providing the statistical power to confirm whether the high-EW tail we observe reflects a genuine population rather than small-number statistics. With the currently available optical bands and a more flexible, data-driven search, we have shown that it is possible to probe parameter space heavily contaminated by T dwarfs and to recover quasars with systematically stronger emission features than those typically reported in the literature, which were overlooked by previous selections.

As discussed in Section 2.2.1, relaxing the NIR detection requirement in the input catalog for the CL pipeline could potentially double the number of quasar discoveries. Many of these currently missed candidates are likely to be revealed by ongoing surveys such as *Euclid*, thanks to its deeper and more uniform NIR coverage. Notably, even within the more traditional regions of parameter space populated by known quasars, our selection moves beyond the standard color-cut paradigm by recovering sources with atypical properties. Alternative strategies, such as extending the search into less-explored regions of the latent space, where spectroscopic labels are scarce but trends in catalog properties offer guidance (see Section 2.2.3), holds the potential to uncover even more unusual populations, albeit at higher observational risk.

The set of artifacts and brown dwarf contaminants identified in our selection provides valuable training data to improve the algorithm's performance, enhance the label-based interpretation of the embedded space, and refine the SED-fitting prioritization. Future iterations could make the algorithm more robust against artifacts by including them in the augmentation process or by incorporating them as a separate class in a semi-supervised framework. Similarly, by expanding the stellar training set with high-quality optical and NIR spectra of these newly discovered ultracool dwarfs, and especially focusing on atmospheric compositions that mimic quasar SEDs, we can reduce contamination rates and explore alternative photometric or morphological indicators to help discriminate these sources more reliably.

Looking ahead, this approach can be naturally extended to ongoing and upcoming wide-field surveys such as Rubin/LSST, *Euclid*, and Roman, where the combination of deep, multi-band photometry and large-area coverage will dramatically increase the discovery potential for high-redshift quasars. With *Euclid* already demonstrating the capability of the NISP spectroscopy to pinpoint rare and luminous quasars at $z > 5$ (e.g., Bañados et al. 2025), our method is well-positioned to scale to these future datasets, leveraging both imaging and low-resolution spectral data to push the quasar frontier deeper into the reionization era and place stronger constraints on the formation and early growth of supermassive black holes.

Chapter 3

Investigating MIR variability for high-z quasar selection

3.1 Introduction

AGNs are highly variable astrophysical objects, exhibiting non-periodic stochastic fluctuations in flux on timescales from hours to years (Martin Gaskell & Klimek 2003). The AGN variability is a multiwavelength phenomena, observed in radio, UV/optical, and X-rays bands, both in the continuum emission as well as in emission lines (Ulrich et al. 1997). The physical processes driving these fluctuations are not fully determined, but they are commonly associated with changes in the disk temperature (Ross et al. 2018), accretion rate or inhomogeneities/instabilities in the accretion disk or the material around the BH (Sánchez-Sáez et al. 2018).

Extensive efforts have been undertaken to link physical properties such as the BH mass, accretion rate, or luminosity with the different parameters modeling the variability. Systematic studies have found evidence of an anti-correlation between the amplitude of the variability and both the rest-frame emission wavelength and Eddington ratio (Bauer et al. 2009; Kelly et al. 2013; Simm et al. 2016; Rakshit & Stalin 2017; Li et al. 2018; Sánchez-Sáez et al. 2018), a negative correlation with bolometric luminosity (Bauer et al.

2009; Kelly et al. 2009; MacLeod et al. 2010; Li et al. 2018), and a positive correlation with redshift (Li et al. 2018). However associations with BH mass seem more uncertain, as some analyses suggest a negative correlation with the excess variance in X-ray variability (Papadakis 2004) or the amplitude of the optical variability (Kelly et al. 2009), while other works found positive correlations with the amplitude of optical variability (Bauer et al. 2009; Wilhite et al. 2008; MacLeod et al. 2010)

Variability studies have been key in building the current AGN paradigm, providing constraints on the location and properties of the different emission components, as determined by the time delays and correlations between spectral changes (Ulrich et al. 1997). For example, the finite light travel time introduces delays between variations observed from the UV-optical continuum emission associated with the inner accretion disk and induced variations from ionized emission lines in the so-called broad line region. This phenomenon is the basis of the reverberation mapping technique to assess the geometry and kinematics of the broad line region clouds, and by assuming them as a virial system, to estimate BH masses (e.g., Blandford & McKee 1982; Peterson 1993; Kaspi et al. 2007; Bentz et al. 2009b).

Due to its stochastic nature, variability presents itself as a promising selection technique to find AGNs. In particular optical variability has been widely used for quasar (i.e., bright AGN) selections from multi-epoch surveys such as the MACHO survey (e.g. at $0.2 < z < 2.8$; Geha et al. 2003), Pan-STARRS 1 (e.g. at $2.5 < z < 3$; Schmidt et al. 2010a), SDSS (e.g. up to $z \sim 3.5$; Butler & Bloom 2011; Palanque-Delabrouille et al. 2011; Peters et al. 2015; Guo et al. 2020), Catalina Real-tine Transient Survey (CRTS, e.g. up to $z \sim 4$; Graham et al. 2014), DES (e.g. up to $z \sim 4$; Tie et al. 2017), Optical Gravitational Lensing Experiment (OGLE, e.g. at $z \sim 5$; Kozłowski et al. 2019) and the Zwicky Transient Facility (ZTF, e.g. up to $z \sim 4$; Sánchez-Sáez et al. 2023; Bernal et al. 2025; Nakoneczny et al. 2025). Variability has even been used to uncover under-sampled populations, such as lensed quasars and quasar pairs (Dux et al. 2024). Unlike traditional optical color selections, which are prone to pollution from stars and

brown dwarfs, particularly at $2.5 < z < 3$ (Schmidt et al. 2010a; Richards et al. 2006a), $3 < z < 3.5$ (Fan et al. 1999; Worseck & Prochaska 2011); and $z \gtrsim 5.2$ (Wang et al. 2016b; Yang et al. 2017), respectively, variability offers a more robust, independent pathway to identify quasars.

Brown dwarfs become significant contaminants in photometric quasar selection at $z \gtrsim 5.2$, where late-type M dwarfs occupy similar regions in the $r - i$ versus $i - z$ color diagram (e.g.; Wang et al. 2016b). At even higher redshifts ($z > 6$), contamination is dominated by L and T dwarfs, whose colors overlap with those of high-z quasars in both optical ($i - z$ vs. $z - y$; Matsuoka et al. 2016; Bañados et al. 2016) and NIR to MIR colors ($y - J$ vs. $J - W1$; Wang et al. 2019). However, variability can help to distinguish these objects. Brown dwarfs exhibit flux variability induced by their rotation and inhomogeneities in surface features, such as temperature and cloud patterns (Artigau 2018). The presence of large-scale storms, bands or in general weather-like patterns on the surface produces wavelength-dependent modulations (Tinney & Tolley 1999). Also, due to the influence of the surface gravity on the dust settlement affecting the thickness and altitude of dust clouds, there is a positive correlation between the variability amplitude and the surface gravity (e.g., Metchev et al. 2015).

Assuming that old brown dwarfs (> 120 Myr) have typical radii of $0.85 - 1.2$ jupyter-radius (Chabrier & Baraffe 2000) and rotational velocities > 20 km/s (Mohanty & Basri 2003), the expected rotation periods range between 2 to 12 hours. Fig. 3.1 shows examples of brown dwarf light curves in the NIR and MIR regimes, with peak-to-peak amplitudes of ~ 50 mmag and $\sim 550 - 700$ mmag, respectively. Multiple studies on L dwarfs have found variability amplitudes of the order of tens of mmag with timescales from minutes to weeks in optical bands (e.g.; Tinney & Tolley 1999; Gelino et al. 2002; Bailer-Jones & Mundt 2001), while L/T transition brown dwarfs often show amplitudes of several hundred mmag and periods of a few hours (Enoch et al. 2003). This periodic, short-timescale variability (e.g., Harding et al. 2013) makes brown dwarfs easily distinguishable from quasars, whose variability is stochastic and occurs over longer timescales. While the $\sim$ 6-month cadence of

unWISE time-series data is insufficient to detect the periodic behavior of ultracool dwarfs, the higher temporal sampling of LSST, with a cadence ranging from minutes to years, will, in principle, allow the detection of these objects.

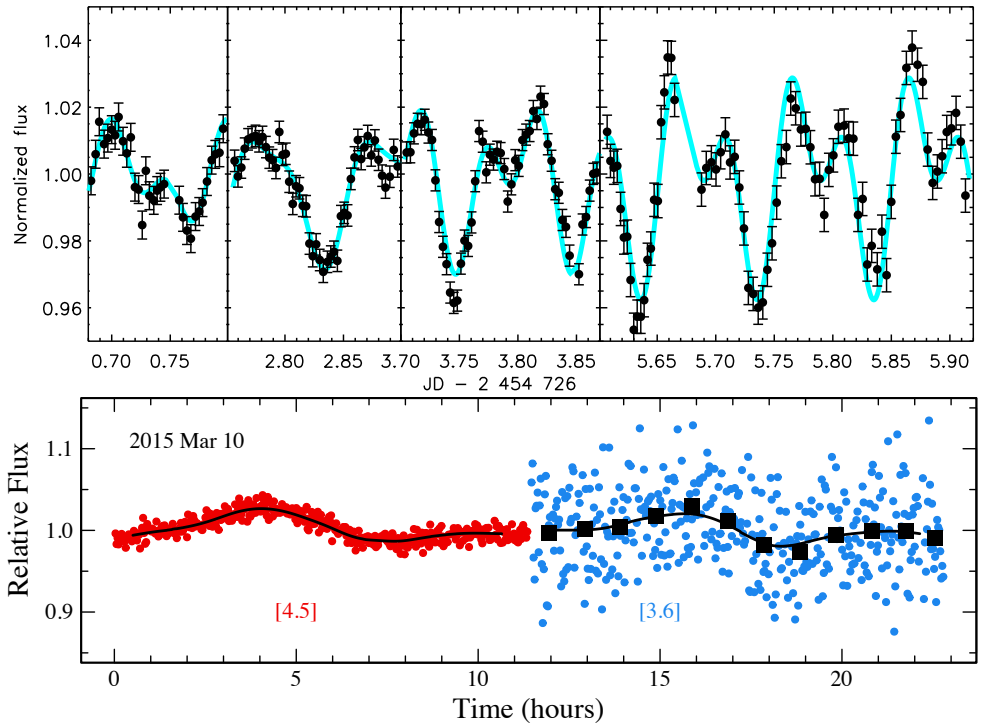

Figure 3.1: Light curves of brown dwarf SIMP0136 (spectra type T2.5) observed in the J-band with the 1.6-m telescope at the Observatoire du Mont-Mégantic over four epochs (adapted from Artigau et al. (2009)); and of WISE 0855-0714, a Y dwarf, monitored with Spitzer/IRAC in bands 1 and 2 at 3.6 μm and μm, respectively (taken from Esplin et al. (2016)). The cyan and black curves show the best-fit models: a two-harmonic sinusoidal model with a period of 2.388 h for SIMP 0136 (*upper panel*), and double sine model with periods of 9.7 and 10.8 h for IRAC bands 2 and 1, respectively (*lower panel*).

3.1.1 Metrics for variability modeling

The most widely adopted approaches in the literature for modeling variability in quasars are the structure function (SF) and the damped random walk (DRW) model. On one hand,

the SF is an empirical approach that models the variability amplitude between epochs separated by a given timescale (Rumbaugh et al. 2018). It has been shown in Schmidt et al. (2010a) not only that SF can be well-modeled by a power law with amplitude A and power-law index γ as free parameters; but also that quasars can be clearly distinguished from contaminants such as RR Lyrae, F and G stars, by their location in the A$-\gamma$ plane (see Fig. 3.2). As most contaminants do not exhibit long-term variability growth, this results in low γ values. While this γ distribution ranges from 0.15 to 0.5 with a peak at $\sim$ 0.3, mid-IR light curves exhibit steeper SF with $\gamma \sim 0.45$, resulting in lower variability amplitudes for small time difference between two observations (Kozłowski et al. 2016)

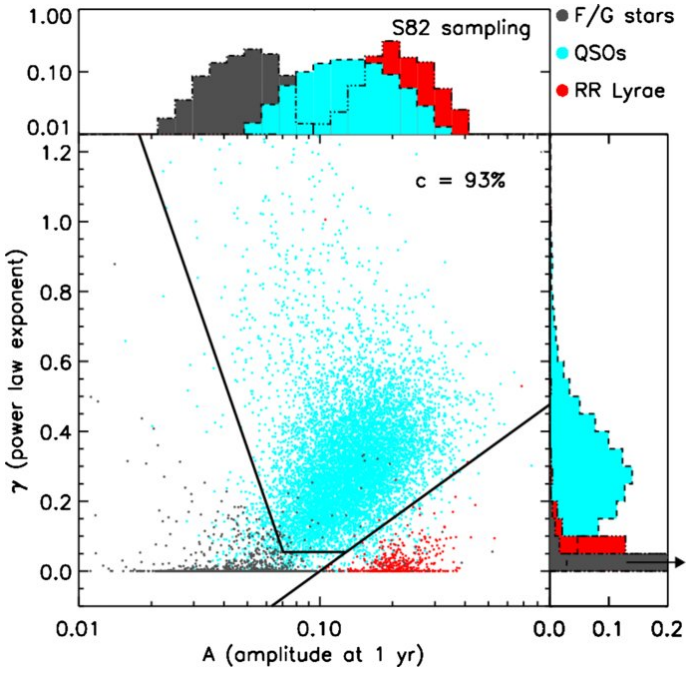

Figure 3.2: Structure function parameters A$-\gamma$ plane for $\sim$ 60-epoch light curves of $\sim$ 15000 sources from SDSS stripe 82 survey. Spectroscopically confirmed quasars and RR Lyrae are shown as blue and red points, respectively, while color-selected F and G stars as gray points. The solid lines highlight the limits to select quasars with a quasar completeness of 93%. The histograms of A and γ values for the three populations in colors are presented along the horizontal and vertical axis, respectively. Figure taken from Schmidt et al. (2010a).

On the other hand, the DRW models light curves as a continuous time stochastic process

with a damping term responsible for boosting deviations toward the mean value (MacLeod et al. 2010). This formalism is defined by three free parameters: the mean value of the quasar magnitude, the characteristic timescale and the long-term variability amplitude (Kelly et al. 2009). In a physical context, the characteristic timescale is interpreted as the time required to dampen local perturbations, such as disk orbital or thermal timescales (Kelly et al. 2009). This approach is robust enough to efficiently select quasars from other variable sources, in particular for light curves with a cadence of days (Zu et al. 2013).

Over the last years, pipelines have been developed to enable rapid classification of light curves, in preparation for the high data stream that will come from the Vera C. Rubin Observatory Legacy Survey of Space and Time (LSST; Ivezić et al. 2019). One of these LSST brokers is the Automatic Learning for the Rapid Classification of Events (ALeRCE; Sánchez-Sáez et al. 2021). ALeRCE has been tested and already demonstrated promising results in processing the Zwicky Transient Facility (ZTF) alert stream. Its light curve classifier module is built on a balanced random forest algorithm with a two-level approach: the first one classifies each source as periodic, stochastic, or transient; while the second one further classifies the target among 15 astrophysical variable classes. The algorithm ingests 152 features resulting from modeling the light curves with different approach as the SF, DRW, the Mexican hat power spectrum (MHPS), the Periodogram pseudo entropy, among other widely used statistical metrics (Sánchez-Sáez et al. 2021).

Optical emission at rest-frame $4000 \lesssim \lambda \lesssim 6000$ Å, from quasars at $5.3 < z < 7.64$ (based on the most recent and public high-redshift quasar compilation by Fan et al. 2023) is redshifted into the MIR, falling within one or both of the WISE bands: W1 (3.4 μm) and W2 (4.6 μm). The upper limit of 6000 Å, is such that, for the redshift range of the mentioned quasar sample, we ensure that the rest-frame optical emission is fully sampled by the W1–W2 wavelength coverage, the two deepest WISE bands, and the only ones included in the unWISE time-domain light curves used in this work.

The unTimely Catalog (Meisner et al. 2023), a full-sky time-domain catalog built from the unWISE coadds spanning from 2010 through 2020 (Meisner et al. 2018), provides

a golden opportunity to investigate rest-frame-optical variability in high-redshift quasars. Due to the depth limitations of the unWISE coadds, our sample is restricted to very luminous sources, which, owing to the known anticorrelation between luminosity and variability amplitude, are expected to exhibit relatively low variability. Nevertheless, this dataset offers a valuable laboratory for assessing the feasibility of extending variability-based selection methods to discover new high-redshift quasars. In this part of the thesis, we employ the features associated with the LSST broker ALeRCE (Sánchez-Sáez et al. 2021) to model the variability of unWISE light curves for quasars at $z > 5.3$ from Fan et al. (2023), additionally using nearby non-variable stars as determined by Gaia-DR3 (Collaboration & Vallenari 2023) to check and potentially improve the photometric calibration of the quasar light curves.

3.2 Methodology

3.2.1 Compilation of light curves

We constructed the IR light curves using photometry from the unTimely Catalog via the unTimely Catalog Explorer tool[1]. This tool supports box-search queries and produces as output: multi-epoch unWISE coadd cutouts, light-curve files and plots, and unWISE time-series image blinks. The key advantage of unTimely is that it provides cleaned, robust, forced-photometry light curves optimized for faint sources, with improved artifact rejection, which greatly facilitates variability analysis. In contrast, official WISE single-exposure (L1b) photometry extends the time baseline by ~ 3 years, but suffer from higher noise and systematic effects.

We queried the coordinates of the 531 known quasars at $z > 5.3$ from the Fan et al. (2023) catalog within a $5''$ radius, resulting in 476 sources with available unWISE W1 and W2 light curves. To ensure meaningful variability measurements and facilitate the

[1] `https://github.com/fkiwy/unTimely_Catalog_explorer`

exploration of different metrics, we restricted the sample to light curves with at least 15 epochs, resulting in a final working set of 285 quasars. The redshift–magnitude distribution of the final 285-quasar sample is shown in Fig. 3.3.

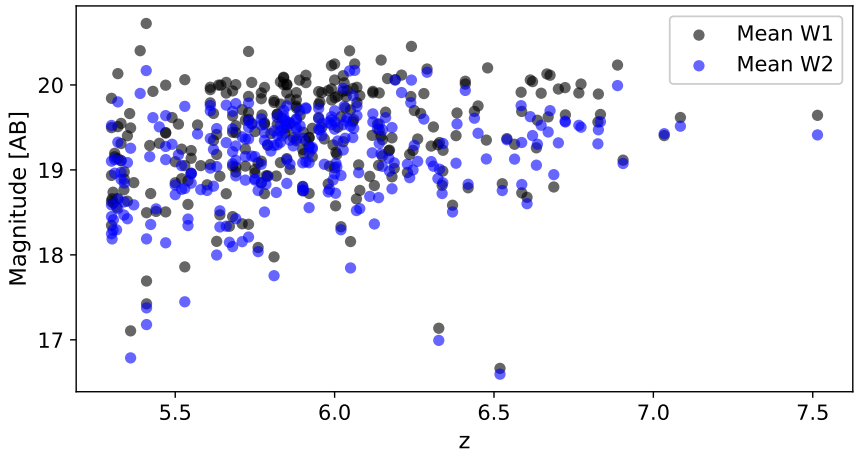

Figure 3.3: Mean IR magnitudes versus redshift for the 285 quasars with at least 15-epoch unWISE light curves. Band 1 and 2 mean measurements are represented by black and blue filled circles, respectively. As the sample is dominated by bright sources (< 20) in both bands, low-amplitude variability is expected.

To establish a baseline and characterize the typical variability level of noise fluctuations, we used Gaia stars with available unWISE light curves as non-variable calibrators. For each quasar, we searched for Gaia DR3 stars within a 0.1° radius satisfying the following criteria: (1) `astrometric_excess_noise` < 10; (2) `classprob_dsc_combmod_star` > 0.8; (3) `pm` < 50; and (4) `phot_variable_flag` $\neq$ `VARIABLE`. These conditions ensure (1) reliable astrometric positions consistent with parallax and proper motion; (2) at least an 80% probability of the source being a star according to the Discrete Source Classifier; (3) minimal coordinate offsets between Gaia and unWISE due to low proper motion; and (4) the absence of known variability. We retrieved unWISE light curves for all stars meeting these criteria and further selected those with < 1 mag difference (in both bands) compared to the corresponding quasar. We required a minimum of 10 and a maximum of 25 calibrator stars per quasar. After applying these constraints, the sample was reduced to

264 quasars with suitable calibrators. Fig 3.4 shows an example of W1 light curve of the quasar J233008.71+095743.70 at $z = 5.3$ and its 10 calibrating stars.

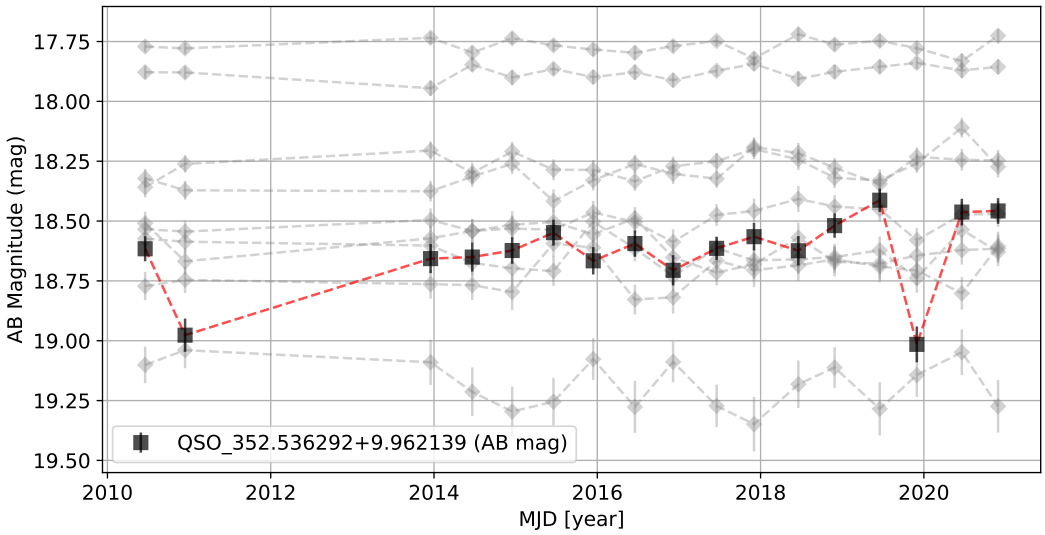

Figure 3.4: W1 light curve of the quasar J233008.71+095743.70 at $z = 5.3$ (red dashed line with black squares) and 10 calibrating stars (gray dashed lines with diamonds) from the field within a 0.1° radius. While the light curves of the stars exhibit fluctuations in the order of 0.1 magnitude between consecutive epochs, the one of the quasar shows a prominent peak of 1 magnitude likely to trigger some of the variability parameters.

3.2.2 Computation of variability features with ALeRCE light-curve classifier

To characterize variability in the unWISE light curves, we extracted parameters from several features implemented in the ALeRCE light-curve (LC) classifier: `MHPS`, `GPDRW`, `IAR`, and `TurboFats`. These features provide complementary descriptions of the variability signal, ranging from stochastic-process modeling to statistical time-series metrics. A brief overview of each method is given below:

- **Mexican Hat Power Spectrum (MHPS):** Arévalo et al. (2012) developed this method to compute low-resolution power spectra from irregularly sampled data. It

uses a Mexican hat filter to isolate fluctuations on specific scales, and then estimates the variance of the convolved data.

- **Damped Random Walk (DRW):** This stochastic model describes light curves as a random process with a damping term, often applied to quasar variability. For the LC-classifier a Gaussian process regression is used to infer the characteristic variability timescale and short-term amplitude (Graham et al. 2017).

- **Irregular Autoregressive (IAR):** Proposed by Eyheramendy et al. (2018), this is a discrete-time representation of the continuous autoregressive model of order 1 tailored to irregular time-sampled data. Its key parameter is the autocorrelation function, which quantifies correlations across different segments of the light curve.

- **TurboFats:** This module extends the Feature Analysis for Time Series package (FATS; Nun et al. 2015), compiling a wide range of variability metrics. These include simple statistical descriptors (e.g., median, standard deviation, skewness, kurtosis), as well as more specific time-series features such as autocorrelation length, Con index (Kim et al. 2011), continuous time auto regressive model (CAR; Pichara et al. 2012) parameters, among others.

Tables 3.1 and 3.2 summarize the full set of parameters extracted with these methods for both W1 and W2 light curves. It should be noted that the current implementation in the ALeRCE LC classifier for MHPS and IAR models includes hard-coded parameters that were found to be suboptimal for the WISE cadence and noise properties. In the following sections, we evaluate which of these features best separate quasars from non-variable calibrator stars, and thus serve as robust indicators of significant variability.

Table 3.1: Variability features from light-curve classifier used in this work

Features	Description	Reference
MHPS_high_W1	Variance for a timescale of 10 days.	Arévalo et al. (2012)
MHPS_low_W1	Variance for a timescale of 100 days.	
MHPS_ratio_W1	Ratio of MHPS_low_W1/MHPS_high_W1	
GP_DRW_tau	Relaxation time	Graham et al. (2017)
GP_DRW_sigma	Amplitude of the variability at short timescales	
IAR_phi	Level of autocorrelation	Eyheramendy et al. (2018)
Amplitude	Half of the difference between the median of the maximum and minimum 5% magnitudes	Richards et al. (2011b)
ExcessVar	Intrinsic variability amplitude	Allevato et al. (2013)
Pvar	Probability that the source is intrinsically variable	McLaughlin et al. (1996)
SF_ML_amplitude	rms magnitude difference of the SF over 1 yr	Richards et al. (2011b)
SF_ML_gamma	Log. gradient of the mean change in magnitude	
AndersonDarling	Test of whether data is drawn from a specific distribution	Nun et al. (2015)
Autocor_length	Lag value where the autocorrelation becomes negligible	Kim et al. (2011)
Beyond1Std	Percentage of points that lie beyond 1σ from the mean	Richards et al. (2011b)
Con	Number of three consecutive points beyond 2σ	Kim et al. (2011)
Eta_e	Ratio of the mean of the squares of successive mag differences to the variance	Kim et al. (2014)
Gskew	Median-based measure of the skew	Nun et al. (2015)
LinearTrend	Slope of a linear fit to the light curve	Richards et al. (2011b)
MaxSlope	Maximum slope between consecutive observations	Richards et al. (2011b)
Meanvariance	Ratio of the standard deviation to the mean	Nun et al. (2015)
MedianAbsDev	Median discrepancy of the data from the median	Richards et al. (2011b)

For a more detailed description of the parameters see Table 2 in Sánchez-Sáez et al. (2021).

Table 3.2: Continued from Table 3.1

Features	Description	Reference
MedianBRP	Fraction of points within amplitude/10 of the median	Richards et al. (2011b)
PairSlopeTrend	Fraction of increasing differences minus fraction of decreasing differences over the last 30 measures	Richards et al. (2011b)
PercentAmplitude	Largest percentage diff. between max/min and median	Richards et al. (2011b)
iqr	Difference between the 3rd and the 1st quartile	Kim et al. (2014)
Rcs	Range of a cumulative sum	Kim et al. (2011)
Skew	Skewness measure	Richards et al. (2011b)
SmallKurtosis	Small sample kurtosis	Richards et al. (2011b)
Std	Standard deviation	Nun et al. (2015)
StetsonK	Robust kurtosis measure	Kim et al. (2011)

3.3 Results

3.3.1 General trends of quasars vs calibrator star populations

A first inspection of the variability features revealed that no MHPS parameters could be estimated, as the outputs were NaN for all quasars and stars. This is expected given the unWISE cadence of $\sim$ 6 months, which provides no temporal coverage at the 10–100 day scales targeted by the MHPS bands. With too few point pairs, the corresponding band variances become undefined, and the implementation reports NaN values. To extract useful information, the MHPS timescales would need to be redefined to match the available sampling. Similarly, the feature `Autocor_length` returned a constant value of 1 for both quasars and stars. In principle, this parameter measures the lag at which the autocorrelation function first falls below $1/e$. A constant value of 1, however, indicates that the autocorrelation function could not be reliably estimated given the sparse sampling of the light curves, and that the implementation defaulted to the minimal lag. Finally,

the values of `Con` were always zero, indicating that none of the light curves contained three consecutive points beyond 2σ. This suggests that significant magnitude changes only occur on timescales shorter than one year, i.e. within individual WISE visits, rather than across multiple visits.

To identify variability features with discriminating power between quasars and field stars, we applied the normalized k-sample Anderson–Darling test (Anderson & Darling 1952) to the stars and quasars distributions of each feature listed in Table 3.1. This statistical test evaluates whether a sample is drawn from a given distribution. In our case, it assesses whether the quasar feature values could plausibly arise from the baseline variability distribution defined by non-variable stars.

We retained only those features for which the quasar and stellar distributions show a test statistic > 40 and a p-value < 0.05, corresponding to a significant mismatch between the two populations and a very low probability that the quasar distribution is consistent with the stellar baseline. Here, the test statistic quantifies the degree of separation between the cumulative distributions, with larger values indicating stronger differences. The threshold on the test statistic was chosen empirically: we gradually increased the cutoff until visual inspection confirmed that the two distributions showed minimal overlap. Because several features are derived from a logarithmic modeling of the light curves, we actually perform the test on both linear and $\log_{10}$-scaled values. The resulting set of discriminating features, summarized in Table 3.3, provides a first selection of parameters most sensitive to quasar variability, and serves as a basis for subsequent source-by-source analysis.

We detect significant variability in W1, but not in W2. This is expected given that W2 photometry is shallower and exhibits larger photometric uncertainties, while quasar variability amplitudes decrease toward longer rest wavelengths. As a consequence, intrinsic fluctuations that are observable in W1 are often subdominant to the measurement noise in W2. Also, we found that for our dataset, and redshift range, the most widely used metrics from the literature, such as DRW or the SF, do not seem as effective as simple statistical characteristics. In Fig. 3.5 we plot the distributions of some of these likely discriminator

Table 3.3: Results of the Anderson–Darling test comparing the distributions of quasars and calibrator stars for the features most diagnostic to quasar variability.

Feature	Scale	Statistic	p-value
`iqr_W1`	Linear	89.40	< 0.001
`Amplitude_W1`	Linear	90.48	< 0.001
`MaxSlope_W1`	Linear	64.08	< 0.001
`Mean_W1`	Linear	81.44	< 0.001
`Meanvariance_W1`	Linear	93.11	< 0.001
`MedianAbsDev_W1`	Linear	75.65	< 0.001
`PercentAmplitude_W1`	Linear	84.11	< 0.001
`Std_W1`	Linear	94.82	< 0.001
`ExcessVar_W1`	Log	45.32	< 0.001

features for bands 1 and 2. As the quasar distributions show a systematic shift toward larger values, with extended high-value tails, we expect these features to be particularly effective in source-by-source analyses, especially for the small subpopulation of highly variable quasars.

To place our results in context, we compare the `ExcessVar` and `MeanVariance` distributions with those reported for the three subclasses of stochastic variables in ALeRCE (AGNs, QSOs, and Blazars) by Sánchez-Sáez et al. (2021) (their Fig. 20). It is important to note that their results are restricted to lower redshifts ($z < 1$ for AGNs and $z < 3$ for QSOs). For `ExcessVar`, our distribution spans 10^{-7}–10^{-3} with a peak around 5×10^{-5}, closely reproducing the QSO distribution in Sánchez-Sáez et al. (2021), but shifted toward lower values than those of AGNs and Blazars. In contrast, the `MeanVariance` distribution in ALeRCE spans 10^{-3}–10^{-2} with a peak at $\sim 2 \times 10^{-3}$, while our results are mildly shifted toward higher values, covering 2×10^{-3}–3×10^{-2} and peaking at $\sim 6 \times 10^{-3}$. These results

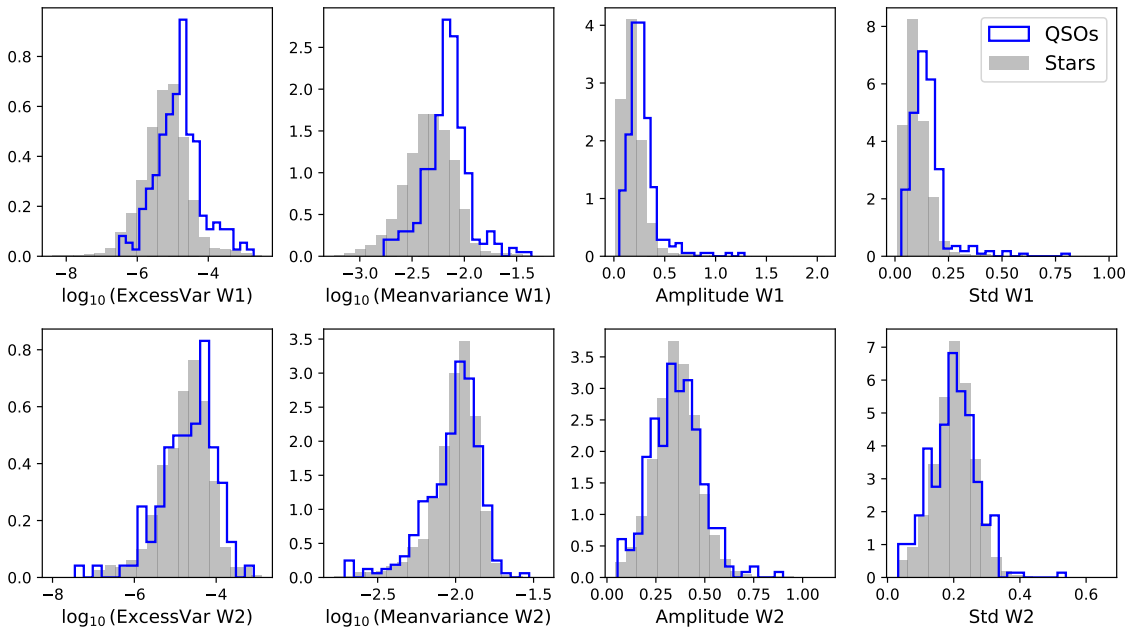

Figure 3.5: Distributions of variability features with discriminative potential from band 1 (*top*) and band 2 (*bottom*) light curves. Gray filled histograms show the stellar distributions, and blue lines the quasar distributions. From left to right: $\log_{10}$(`ExcessVar`), $\log_{10}$(`MeanVariance`), `Amplitude`, and `Std`. In W1, quasars exhibit a systematic shift toward larger values and extended high-value tails, indicative of strongly variable sources. In contrast, the W2 distributions largely overlap with those of stars, reflecting the reduced sensitivity of this band to variability.

demonstrate that despite the different observed wavelength regimes, our IR light curves are probing a rest-frame parameter space consistent with that revealed by optical ZTF light curves at lower redshift.

To investigate potential redshift dependence in quasar variability, we evaluate the most discriminative features identified in Table 3.3 as a function of redshift. Since `MeanVariance` is simply `Std` normalized by the mean, we omit one to avoid redundancy. Similarly, `Mean` is excluded because it traces average brightness rather than variability. Figure 3.6 shows the resulting set of seven features, with quasars that deviate $> 1\sigma$ from the mean distribution highlighted in red. This threshold helps identify sources with stronger variability features than the typical quasar in the sample, serving as a first indication of significant

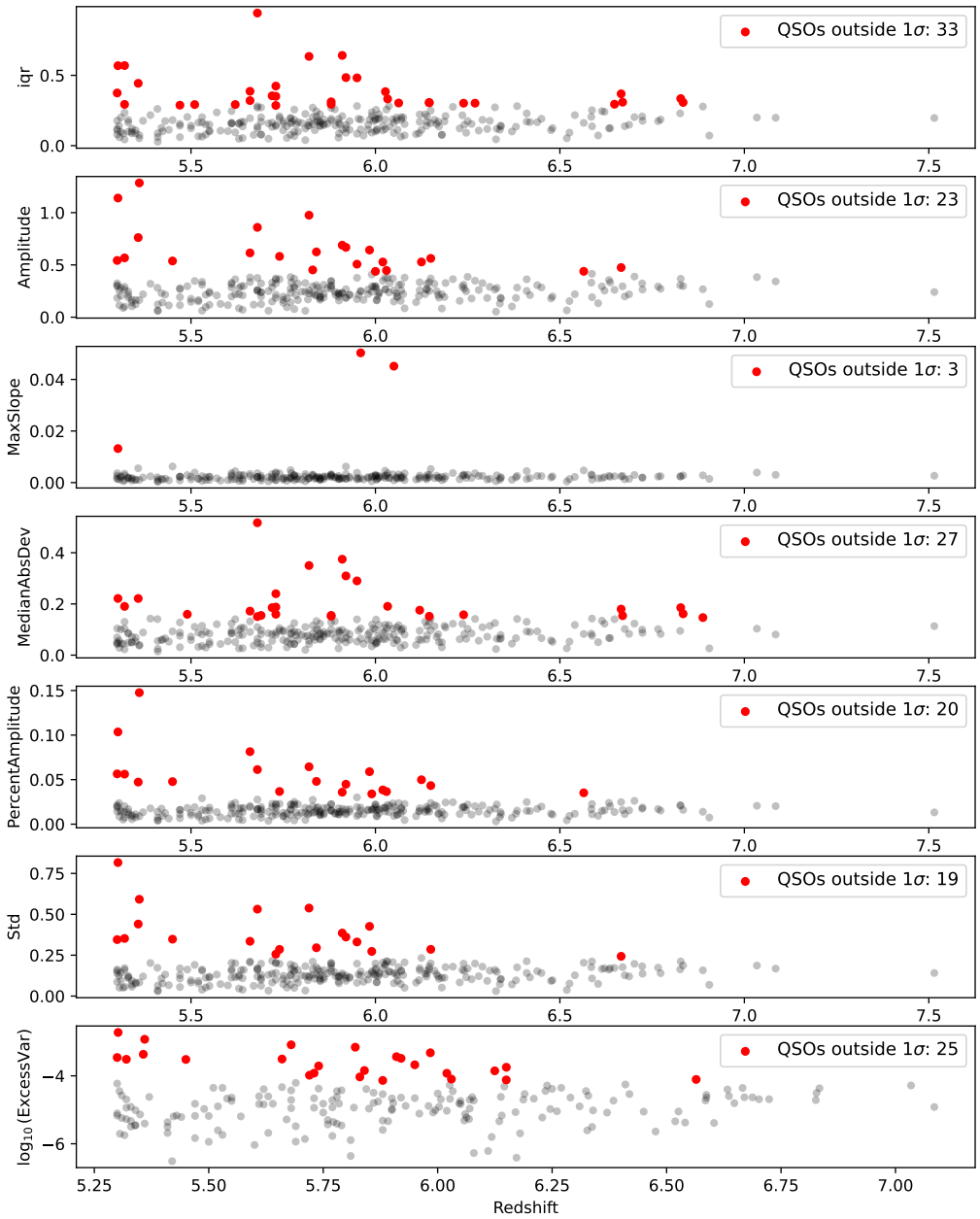

Figure 3.6: Variability features as a function of redshift from top to bottom: `iqr`, `Amplitude`, `MaxSlope`, `MedianAbsDev`, `PercentAmplitude`, `Std`, and `ExcessVar`. Red points represent quasars deviating by more than 1σ from the mean distribution, while gray points correspond to the remaining quasar population. Most methods identify more than 20 quasars with significant variability. Amplitude- and standard deviation–based features show the largest deviations at low redshift and a decreasing trend with redshift, possibly indicating a redshift dependence.

variability. For most features, this subset includes more than 19 sources.

Notably, both amplitude-based metrics and the standard deviation show the strongest deviations at low redshift, with decreasing deviations toward higher redshift. Several scenarios could explain this trend: (1) low-redshift quasars may have higher SNR, allowing more significant detections of variability compared to the noisier light curves of high-redshift quasars; (2) since we analyze observed-frame light curves, time dilation implies that at $z \sim 5.3$ we probe longer rest-frame timescales ($\sim$1.6 yr), giving more time for optical magnitude changes to occur, while at $z \sim 7$ the effective rest-frame coverage ($\sim$1.3 yr) is shorter and may not capture such variations; (3) low-number statistics in the higher redshift bin; and (4) a genuine redshift dependence of variability cannot be ruled out, although our current dataset is insufficient to draw a firm conclusion.

We built the SF parameter space (A–γ) for quasars and stars in our sample (see Fig. D.1) to compare with the results of Schmidt et al. (2010a). Unlike their well-behaved distributions, our fits are dominated by systematic 'artifacts'. We observe horizontal bands at `SF_ML_amplitude` $\sim$ –0.5 and 3, as well as diagonal loci tracing a correlation between `SF_ML_amplitude` and $\log_{10}$(`SF_ML_gamma`). These features are unlikely to reflect intrinsic variability, but instead arise from the limited temporal sampling of unWISE light curves: the 6-month cadence provides insufficient coverage around the 1-year baseline used to define the SF amplitude, producing bimodality and degeneracies in the fitted parameters. A minority of sources do exhibit well-scattered values consistent with a physical amplitude–slope correlation. However, since quasars and stars populate the same artificial loci, we conclude that the SF model offers no discriminative power in our dataset and provides little physical insight into the variability.

For a better interpretation of the individual quasar–calibrator variability assessment in the next section, it is first necessary to characterize the impact of photometric depth on the variability metrics. To this end, we examined how the variability features depend on the mean magnitude of the stellar light curves (Fig. 3.7). While Fig. 3.7 only shows `iqr`, `MaxSlope`, `Std`, and `ExcessVar`, we reproduce this analysis for all the features from Table

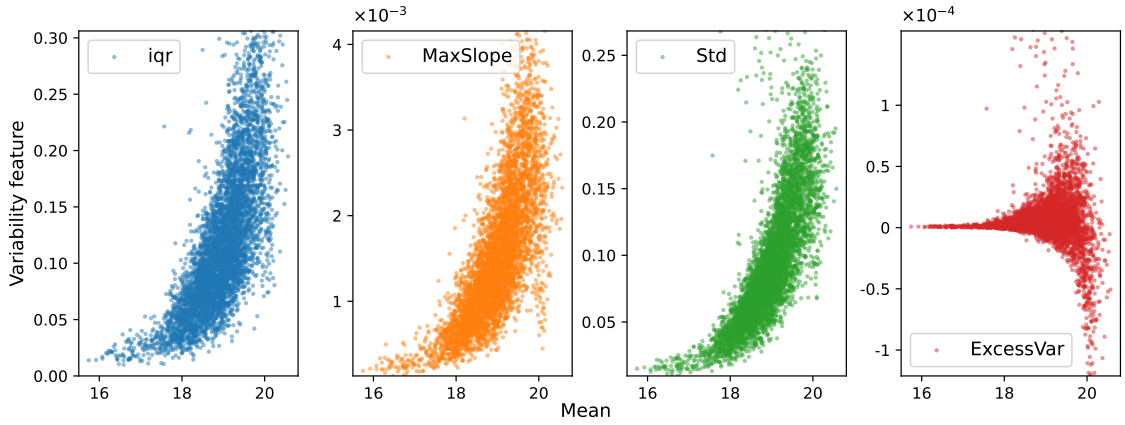

Figure 3.7: Variability features as function of the mean light curve magnitude for the overall distribution of calibrator stars: `iqr`, `MaxSlope`, `Std`, and `ExcessVar`. The first three panels show a clear positive correlation with magnitude, indicating that increasing photometric uncertainties can replicate variability signatures in otherwise non-variable stars. In contrast, the excess variance displays both higher positive and negative amplitudes for the intrinsic variability of faint sources.

3.3.

Most features show a clear positive correlation with magnitude, indicating that the increasing photometric uncertainties of fainter sources can mimic variability signatures. Although unWISE light curves are unaffected by atmospheric conditions, fainter stars suffer from lower SNR, higher background contributions, and larger PSF-fitting uncertainties, which inflate apparent variability, even when the light-curve classifier accounts for photometric errors. Interestingly, excess variance exhibits both positive and negative correlations with magnitude, and shows a wide scatter at the faint end, consistent with measurement noise and systematics. Despite these effects, approximately 25% of quasars still exceed the variability levels of stars (see Fig. 3.5), indicating a genuine signal. This underscores that variability features provide a probabilistic measure of quasar variability, and motivates the use of magnitude-dependent calibrators, to identify high-confidence variable candidates.

3.3.2 Individual-quasar variability estimation

The individual source variability tests are motivated by the promising trends identified in the previous statistical analysis (see Table 3.3). Because quasar variability is anticorrelated with luminosity, and the unWISE depths bias our sample toward brighter sources, we do not expect to detect significant variability features in all quasars. Instead, we anticipate measurable deviations in the $\gtrsim$ 20 quasars which exhibit 1σ deviations relative to the overall population (Fig. 3.6 in Section 3.3.1). Pushing the variability threshold to 2σ and 3σ deviations decreases the numbers to $\sim$ 9 and $\sim$ 4 quasars, respectively.

For each quasar, we establish a variability baseline defined within $\mu \pm 3\sigma$ of the calibrator star distribution, where μ and σ are the mean and standard deviation under the assumption of a Gaussian distribution. As this baseline is computed with stars within a 1deg^2 region of the sky and in the same magnitude range as the quasar (± 1 mag), it is unique for each quasar. Quasars are flagged as potentially variable if any of their variability features from Table. 3.3 (1) lie outside these limits and (2) exceed those of all calibrator stars.

In Fig. 3.8, we show as an example the `iqr_W1` values of the quasars together with the mean values of the calibrator stars (with 1σ error bars), restricted to the 29 quasars classified as likely variable. The `iqr` parameter measures the variability amplitude with the difference between the 3rd and 1st quartiles of the magnitudes in the light curve. Most of these quasars have `iqr_W1` > 0.2, but 7 sources with lower values appear in fields where the calibrator stars themselves reach even smaller `iqr` values with low uncertainties. Visual inspection of these sources reveals nearly flat stellar light curves, while the quasars appear modestly variable. As expected, all of them correspond to bright sources with mean W1 magnitudes in the range $17.69 - 19.55$ and uncertainties of $0.02 - 0.1$, less affected by low SNR and potential instrumentation effects. The difference between the $\mu + 3\sigma$ threshold of the overall stellar sample and the error bars from the individual calibrator star samples illustrates the diversity of stellar variability features across different fields, driven mainly by magnitude, as shown in Fig. 3.7.

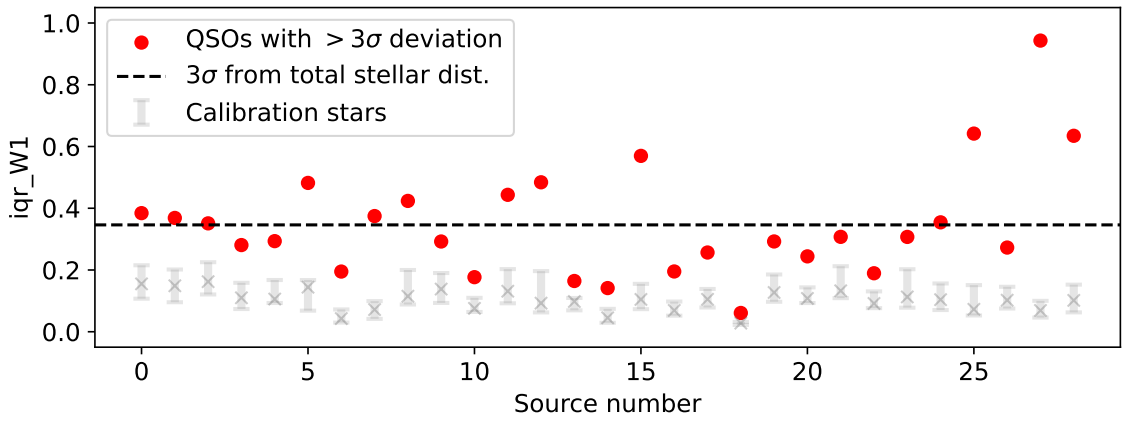

Figure 3.8: `iqr_W1` values for quasars (red circles) that exceed their local variability baseline, compared with calibrator stars (gray crosses with 1σ error bars). The interquartile range quantifies variability as the difference between the 3rd and 1st quartile magnitudes. The dashed line marks the $\mu + 3\sigma$ threshold of the global calibrator star distribution.

We apply the same selection criteria to the remaining variability features, yielding: 21 variable quasars with `Amplitude_W1`, 17 with `MaxSlope_W1`, 27 with `MedianAbsDev_W1`, 16 with `PercentAmplitude_W1`, 22 with `Std_W1`, and 12 with $\log_{10}$(`ExcessVar_W1`). Since `Meanvariance_W1` by construction identifies the same 22 sources as `Std_W1`, we exclude it to avoid redundancy. In total, 57 quasars are flagged as variable by at least one method: 27 by a single feature, 11 by two features, and 19 by three or more. Visual inspection of the corresponding light curves revealed that two of the 27 quasars classified by only one feature do not show actual variability, but instead contain a spurious outlier at a single epoch, producing an artificial infinite slope (see Fig. 3.9). While similar outliers are also present in other quasars, their light curves still display intrinsic variability (e.g. J144128.70+502239.40), so we retain them. After excluding the two spurious cases, our final sample consists of 55 variable quasars.

In Tables 3.4 and 3.5, we compile the 19 sources most likely to be variable, identified by at least three variability features, and 11 additional sources considered less reliable, flagged by only two features. While the majority of these quasars are relatively bright

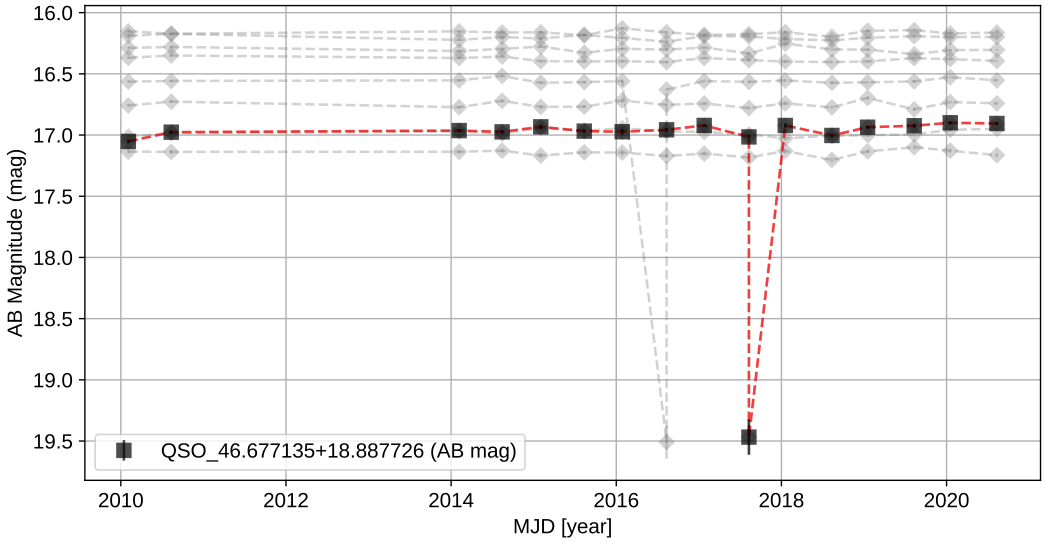

Figure 3.9: W1 light curve of the false positive variable quasar J030642.51+185315.82 at $z = 5.36$ (red dashed line with black squares) and 10 calibrating stars (gray dashed lines with diamonds) from the field within a 0.1° radius. A prominent peak is produced by one epoch with two different measurements of the magnitude, while the remaining light curve seems nearly constant. This phenomenon is seen not only in the quasar light curve but also in one of its calibrator stars.

in WISE band W1 (18.33 < `mean_W1` < 20.04), meaning bright in the rest-frame optical ($4433 \lesssim \lambda \lesssim 5397$ Å) depending on redshift, their apparent magnitudes at rest-frame UV 1450 Å are systematically fainter by more than one magnitude in 10 of the cases. This indicates that these quasars exhibit a red rest-frame optical slope, consistent with significant dust obscuration. The combination of variability and reddening is particularly interesting, as red quasars are thought to represent a transitional evolutionary phase in SMBH growth, possibly linked to merger-driven activity (e.g.; Glikman et al. 2012; Kim & Im 2018; Glikman et al. 2024). In such systems, variability may primarily trace intrinsic changes in accretion, though small-scale or clumpy obscurers near the nucleus could also modulate the observed flux.

Our observed-frame 10-year light curves span ∼ 3000–4000 days, but due to cosmological time dilation, the corresponding rest-frame baselines for our $z = 5.3$–7 quasars are

Table 3.4: Quasars exhibiting variability in three or more features, sorted by decreasing number of variability flags. Sources with more flagged features ($> 3\sigma$ above the calibrator-stars baseline) show stronger, more reliable variability.

Name	**RA**	**Dec**	**z**	**Discovery Ref.**	**Flags**	**W1**	$\mathbf{m_{1450}}$
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
J0850–0506	132.61733	-5.10192	5.36	Kim et al. (2019)	7	19.66	22.61
J1256+2532	194.12903	25.54762	5.91	Bañados et al. (2016)	7	19.35	20.90
J1514+2114	228.68712	21.23884	5.92	Bañados et al. (2016)	7	19.27	20.61
J1109+5657	167.47266	56.95212	5.95	Bañados et al. (2016)	6	19.92	20.74
J1545+1636	236.29125	16.60887	5.82	Bañados et al. (2016)	6	19.54	20.82
J2358+0634	359.60017	6.57706	5.32	Yang et al. (2017)	6	19.13	19.31
J0300–2232	45.18408	-22.54089	5.68	Bañados et al. (2016)	5	18.45	20.41
J0953+6910	148.48294	69.18128	5.84	Bañados et al. (2023)	5	20.00	20.19
J1022+0801	155.50795	8.02285	5.30	Yang et al. (2017)	5	18.35	18.93
J1356–2642	209.20586	-26.70840	5.72	Bañados et al. (2016)	5	19.04	19.49
J2220–0101	335.07704	-1.02969	5.62	McGreer et al. (2013)	5	19.66	20.66
J2330+0957	352.53629	9.96214	5.30	Yang et al. (2017)	5	18.63	19.81
J0757–0002	119.44123	-0.03364	5.54	Yang et al. (2019a)	4	18.87	19.93
J0818+1722	124.61416	17.38106	6.02	Carilli et al. (2010)	4	18.33	19.24
J0921+0007	140.33567	0.12303	6.56	Matsuoka et al. (2018)	4	19.30	22.05
J1437+2323	219.46594	23.38704	5.31	Lyke et al. (2020)	4	18.74	19.12
J2232+2930	338.22981	29.50896	6.67	Mazzucchelli et al. (2017)	4	20.13	20.78
J0216–5226	34.16189	-52.43905	6.41	Yang et al. (2019b)	3	20.04	21.7
J0408–5632	62.08012	-56.54134	6.03	Reed et al. (2017)	3	20.02	20.19

Notes: Col (1): Quasar name; Col (2) and Col (3): coordinates (J2000) in degrees, Col (4): Redshift; Col (5): Discovery reference; Col (6): Number of variability features flagged; Col (7): mean W1 magnitude; Col (8): apparent magnitude at rest-frame 1450 Å.

only $\sim$ 375–635 days. Correcting for this effect also compresses the unWISE cadence: the observed $\sim$ 6-month sampling corresponds to time lags of $\sim$ 23–29 days in the rest frame.

Table 3.5: Quasars exhibiting variability in two features presented as in Table 3.4.

Name	**RA**	**Dec**	**z**	**Discovery Ref.**	**Flags**	**W1**	**m_{1450}**
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
J1441+5022	220.36958	50.37761	5.66	Wagenveld et al. (2022)	2	18.72	19.43
J0155+0415	28.88867	4.25187	5.37	Wang et al. (2016a)	2	18.85	19.48
J0158-2905	29.51726	-29.08868	5.98	Bañados et al. (2016)	2	19.04	19.42
J1148+0702	177.01369	7.03565	6.32	Farina et al. (2022)	2	19.30	20.35
J1609-1258	242.43971	-12.98170	5.83	Bañados et al. (2016)	2	19.09	19.78
J1153+2830	178.37331	28.50753	5.68	Bañados et al. (2023)	2	19.44	19.84
J0756+0218	119.09320	2.30560	5.73	Bañados et al. (2023)	2	19.59	20.42
J1805+4918	271.44556	49.30672	5.74	Bañados et al. (2023)	2	19.55	20.38
J1548+0050	237.10583	0.83764	6.15	Wang et al. (2019)	2	19.36	21.11
J2123-2421	320.87033	-24.3604	5.73	Bañados et al. (2016)	2	19.86	20.46
J0420-4453	65.04727	-44.88993	6.07	Reed et al. (2017)	2	19.97	20.46

Notes: Col (1): Quasar name; Col (2) and Col (3): coordinates (J2000) in degrees, Col (4): Redshift; Col (5): Discovery reference; Col (6): Number of variability features flagged; Col (7): mean W1 magnitude; Col (8): apparent magnitude at rest-frame 1450 Å.

Using these rest-frame light curves of our 19 robust variable quasars, we recomputed variability metrics targeting short timescales, including the MHPS, DRW, and SF. The results, however, were consistent with those obtained from the observed-frame light curves with NaN values for MHPS parameters, SF values non-consistent with the literature, `Autocor_length` = 1 and `GP_DRW_tau` = 1. Nevertheless, the effective rest-frame timescales now overlap with those probed in better-sampled quasar light curves from SDSS (e.g., Petrecca et al. 2024) and ZTF (e.g., Arévalo et al. 2024), which allows for a more direct comparison between our high-redshift quasars and low-redshift ($z < 2.5$) populations.

3.3.3 Power spectral density analysis

The power spectral density (PSD) is the Fourier transform of the autocovariance function of a signal, and describes the distribution of power into frequency components. The stochastic variability of quasar light curves produces a characteristic "red-noise" trend in the PSD, where the variability amplitude increases toward longer timescales (Petrecca et al. 2024). This monotonic increase in power at low frequencies is commonly modeled with a power law (e.g., Collier & Peterson 2001; Uttley et al. 2002), typically with a negative spectral index (Patel et al. 2025). Models that link variability to magnetic turbulence driving accretion-disk temperature fluctuations (Sun et al. 2020) predict spectral indices in the range -2 to -0.6 (Patel et al. 2025), whereas the DRW model yields an index of -2 that flattens toward 0 (i.e., white noise) on the longest timescales (Petrecca et al. 2024). Importantly, periodic features in the PSD are not expected from intrinsic single-quasar variability; rather, they are typically artifacts introduced by the sampling pattern or the cadence of the light curves. However, genuine periodic or quasi-periodic signals may arise in systems hosting sub-parsec SMBHs binaries, where orbital motion, jet precession, or accretion disk modulation can introduce periodic features in optical (e.g., Graham et al. 2015; Charisi et al. 2016; Valtonen et al. 2008) and even radio light curves (e.g., O'neill et al. 2022).

We computed the PSD of our 19 most reliable variable quasars using the Lomb–Scargle periodogram, a method tailored for detecting variability in unevenly sampled time series (Lomb 1976; Scargle 1982). The periodograms were calculated in their standard (unnormalized) form and, since observational magnitude uncertainties were included, the resulting PSDs are dimensionless. The accessible frequency range is bounded by the length of the light curve, which sets the minimum frequency, and by the shortest time separation between consecutive epochs, which sets the maximum frequency. Figure 3.10 presents the PSD of the quasar `J135649.41-264230.23` at $z = 5.72$, showing the expected decline in power toward lower frequencies up to $\log_{10}(f) \approx -2.2$, followed by a series of peaks at higher frequencies. These peaks are most likely artifacts arising from the irregular

sampling and cadence of the light curves rather than intrinsic quasi-periodic features.

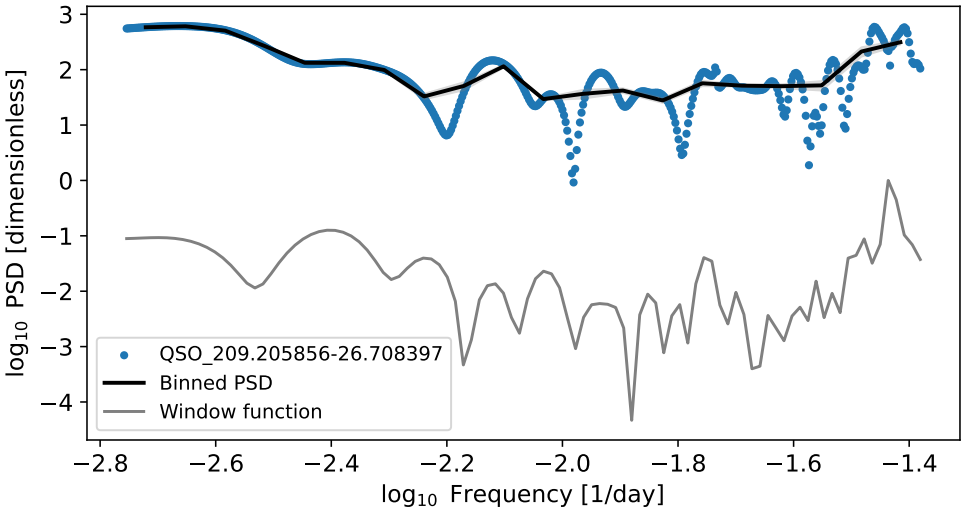

Figure 3.10: Power spectral density of the quasar `J135649.41-264230.23` at $z = 5.72$ (blue points), with the median-binned PSD over 20 bins (black curve) and the corresponding window function (gray curve). The decline at $\log_{10}(f) < -2.2$ traces intrinsic quasar variability, while multiple peaks at higher frequencies coincide with those in the window function and are therefore attributed to sampling artifacts. Then our analysis is restricted to the low-frequency regime, with the upper frequency limit set, source-by-source, at the point where the PSD begins to follow the window function.

The window function characterizes how the observational sampling redistributes power in the frequency domain. Formally, it is the Fourier transform of the sampling pattern (Deeming 1975):

$$\delta_N(\nu) = \sum_{k=1}^{N} e^{i2\pi\nu t_k}, \tag{3.1}$$

where ν is our grid of frequencies and t_k the epochs from our light curves in MJD format. Because uneven or finite sampling causes energy from one frequency to "leak" into others, a phenomenon known as spectral leakage, the window function modulates the intrinsic PSD. This leakage can produce spurious peaks or distort the underlying PSD (VanderPlas 2018), potentially mimicking quasi-periodic features.

To isolate the regions of the PSD that can be interpreted with confidence, we compared the observed PSD to its corresponding window function (Fig. 3.10). At $\log_{10}(f) > -2.2$, the PSD peaks coincide with strong features in the window function, indicating that they are dominated by sampling artifacts. In contrast, two peaks at $\log_{10}(f) < -2.2$ do not align with the broad window function bump, suggesting that they may represent genuine signatures of the intrinsic quasar red-noise slope. We repeated this analysis for all 19 quasars to define, on a source-by-source basis, the frequency ranges suitable for fitting. We binned the PSD between the minimum frequency, set by the total light-curve duration, and the maximum frequency, defined by the smallest time interval between consecutive observations. This corresponds to typical values of $\log_{10}(f) \sim -2.8$ to $\log_{10}(f) \sim -1.3$. The spectrum was divided into 20 logarithmically spaced bins, and the mean power was computed in each, following standard practice to suppress high-frequency noise and facilitate red-noise fitting (Arévalo et al. 2024; Petrecca et al. 2024). The choice of 20 bins was made empirically, as it effectively flattens the high-frequency spikes while preserving the overall PSD shape at low frequencies.

The lowest-frequency bins of the PSD are poorly sampled because they correspond to timescales comparable to the total duration of the light curve. This means that only a small number of independent variability cycles are captured, so the measured power at these frequencies represents a single realization of the underlying stochastic process rather than the true ensemble behavior. As a result, the PSD slope can vary significantly between individual quasars in this regime. While we did not truncate the PSDs at a fixed fraction of the light-curve duration, we note that previous studies often restrict the analysis to frequencies above roughly one-third to one-tenth of the total baseline to mitigate this effect. In the artifact-free region ($\log_{10}(f) \lesssim -2.2$), we modeled the PSDs using a single power law (i.e., a linear fit in log–log space). For comparison with previous studies, we also fitted a broken power-law model with two slopes to the unbinned PSD. This methodology follows common practices in AGN variability studies, where PSDs are typically parameterized as simple or broken power laws to capture red-noise behavior (e.g.,

Uttley et al. 2002; Gonzalez-Martin & Vaughan 2012; Petrecca et al. 2024; Arévalo et al. 2024).

We found evidence for a red-noise PSD in only 8 of the quasars, with acceptable fits obtained using either a simple or a broken power-law model (Table 3.6). Previous ensemble studies in the low-redshift regime have reported similar ranges: Petrecca et al. (2024), using single power-law fits, found slopes between -0.8 and -1.8, with a dependence on black hole mass such that more massive systems exhibit steeper PSDs for quasars at $0.5 < z < 2.5$. In contrast, approaches adopting a broken power law, such as Arévalo et al. (2024), identified slopes of $\alpha_H \sim -2.5$ to -3 in the high-frequency regime and $\alpha_L \sim -1$ at lower frequencies for quasars at $0.6 < z < 0.7$. Among our results, four quasars show slopes consistent with Petrecca et al. (2024), while two of them also have high-frequency slopes α_H in agreement with Arévalo et al. (2024). Of the remaining four, three sources have either α or α_H values close to the limits, and while one quasar exhibits extreme slopes with $\alpha = -2.71$ and $\alpha_H = -8.95$, which most likely arise from strong sampling artifacts, model limitations, or spurious periodic components, rather than genuine quasar variability.

Table 3.6: PSD fitting results for quasars exhibiting red-noise

Name	**Power law**	**Broken power law**	
	α	$\alpha_{\mathbf{L}}$	$\alpha_{\mathbf{H}}$
J135649.41-264230.23	-2.55	0.18	-3.34
J081827.40+172251.80	-1.64	-0.85	-2.83
J125630.97+253251.45	-1.47	1.17	-3.10
J143751.83+232313.35	-2.71	-1.62	-8.95
J154509.90+163631.94	-1.76	-0.27	-2.51
J222018.49-010146.89	-0.90	-0.11	-3.69
J021638.85-522620.59	-3.80	-3.90	-2.47
J235824.04+063437.40	-1.86	-0.42	-12.88

Of the remaining 11 variable quasars, four show flat PSDs consistent with white

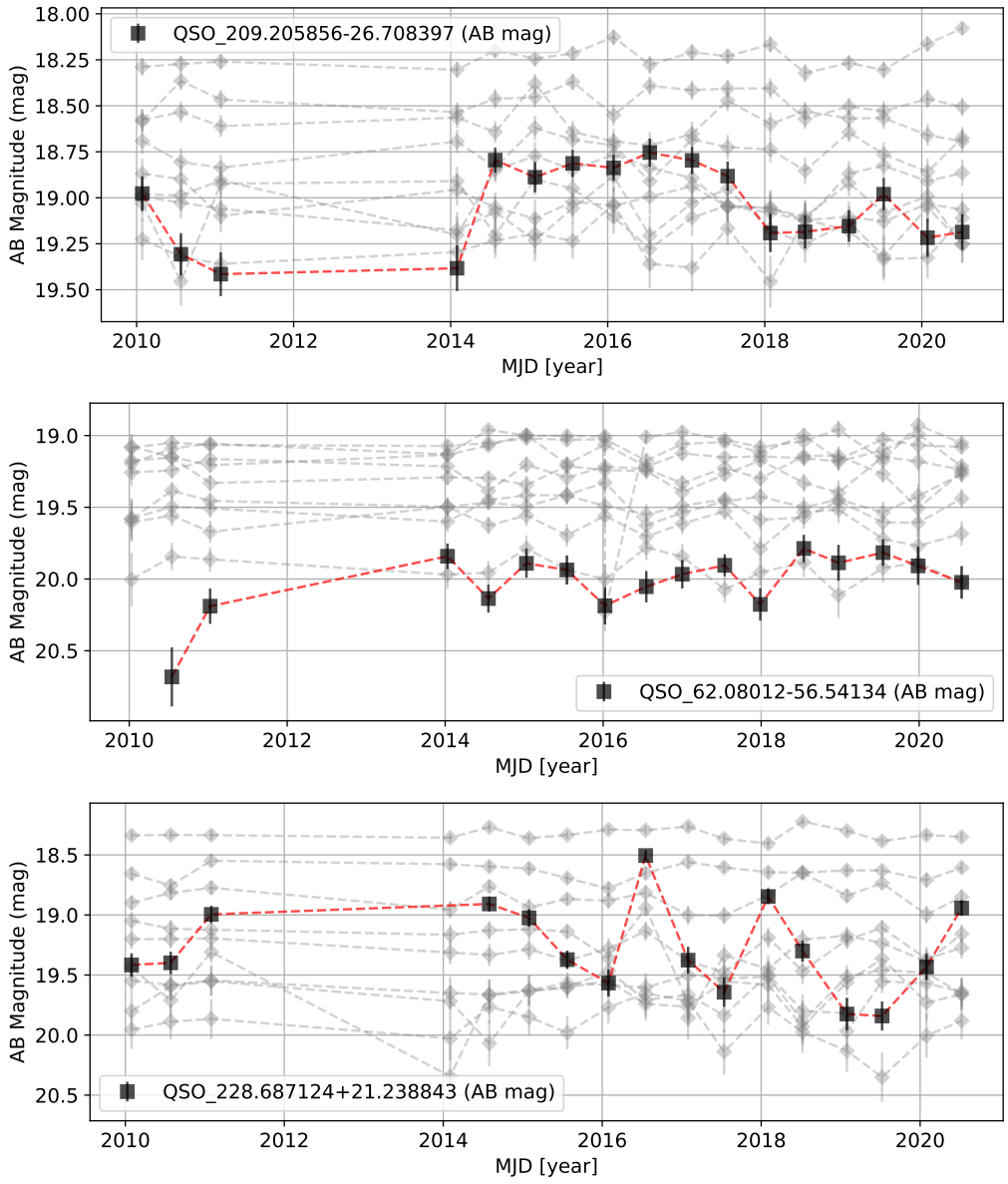

Figure 3.11: Observed-frame W1 light curves of three representative variable quasars illustrating the diversity of PSD behaviors. *Top*: `J135649.41-264230.23` ($z = 5.72$) shows red-noise–like variability with a high amplitude (> 0.5 mag) over long timescales. *Middle*: `J040819.23-563228.82` ($z = 6.10$) displays a repeating rise–decline pattern of nearly constant amplitude, producing a strong PSD peak at $\log_{10}(f) = -2.55$. *Bottom*: `J151444.91+211419.84` ($z = 5.91$) exhibits a flat, white-noise–like PSD at the first epochs, and later two cycles that seem superimposed on a decreasing magnitude trend.

noise, five display a prominent peak at $\log_{10}(f) = -2.55$ (some with additional peaks at $\log_{10}(f) = -2.3$ or -2.8), and two exhibit peaks at $\log_{10}(f) = -2.7$ that cannot be explained by the window function. In Fig. 3.11 we present representative observed-frame light curves from each category: `J135649.41-264230.23` with red-noise variability, `J040819.23-563228.82` with a strong peak at $\log_{10}(f) = -2.55$, and `J151444.91+211419.84` with a flat, white-noise–like PSD. The quasar `J135649.41-264230.23` shows a brightening of local minima over time and variations exceeding 0.5 mag that remain coherent over several years. In contrast, `J040819.23-563228.82` exhibits a repeating pattern of rises and declines in brightness with nearly constant amplitude and mean magnitude, likely producing the periodic feature in its PSD. Finally, `J151444.91+211419.84` shows two cycles of increasing and decreasing brightness with comparable amplitudes, superimposed on an overall fading trend.

This PSD analysis reveals that variability metrics calibrated against stellar baselines may select quasars that do not necessarily display the characteristic stochastic variability expected for AGN. In several cases, we find periodic-like signals in the light curves with amplitudes larger than those of the calibrator stars, yet inconsistent with sampling artifacts identified through the window function. In other cases, we observe low-amplitude fluctuations that are statistically above the stellar baseline but still consistent with white noise. Inspection of individual epoch cutouts shows no obvious issues with source extraction or photometry, suggesting that the origin of the apparent periodic components remains uncertain.

3.4 Summary

In this part of the thesis, we investigate variability signatures in the observed-frame infrared regime for 285 quasars at $z > 5.3$, with the goal of probing rest-frame optical variability linked to accretion disk inhomogeneities or changes in the physical properties. We constructed unWISE light curves with more than 15 epochs for each quasar. We

established a baseline using 10–25 non-variable Gaia DR3 calibrator stars from the same field, chosen to have similar W1 magnitudes ($\Delta m_{W1} < 1$ mag). To characterize variability, we modeled the light curves using four modules from the ALeRCE light-curve classifier: MHPS, GPDRW, IAR, and TurboFats, and then analyzed the resulting variability features.

We found that the cadence of unWISE light curves (∼6 months) prevents reliable use of metrics designed to probe short-timescale variability (days to ∼1 yr), such as MHPS, GPDRW, and the structure function in TurboFats. Even after correcting for time dilation, the rest-frame light curves could not be adequately modeled with these methods. This limitation reduced the set of suitable variability features to a smaller group dominated by statistical descriptors rather than physical model–based ones. A global comparison of quasar and calibrator-star feature distributions using the Anderson–Darling test identified eight features with discriminative power: `iqr`, `Amplitude`, `MaxSlope`, `Meanvariance`, `MedianAbsDev`, `PercentAmplitude`, `Std` and `ExcessVar`, all yielding statistic values > 40 and $p < 0.05$. Besides, the `ExcessVar` and `Meanvariance` distributions overlap the values found by Sánchez-Sáez et al. (2021) for quasars at $z < 3$ using optical light curves from ZTF. The same analysis also demonstrated that W2 light curves should be discarded: their variability features did not differ between quasars and calibrators, likely because the shallower depth of W2 leads to large photometric uncertainties that overwhelm any intrinsic variability signal.

The isolated analysis of the quasar population indicates that 7% of the sources exhibit variability lying 1σ outside the median, and a redshift dependence of amplitude-based metrics, with the standard deviation taking higher values at the lowest redshifts. A similarly motivated analysis of the calibration stars found a positive correlation between the variability feature values and the magnitude, this result justifies the need for an individual source-by-source variability calibration, as the depth of the field have an effect on the light curves and can artificially enhance the variability.

The individual source analysis identified 55 quasars flagged as variable, defined as showing more than a 1σ deviation above the distribution of their corresponding calibrator

stars in at least one variability feature. Visual inspection confirmed that brighter quasars benefit from higher-quality calibrators, which improves the detection of even mild variability. Among these, 19 quasars were flagged in at least three independent features, making them the most robust variable candidates. All 19 are bright in the rest-frame optical ($18.33 < W1 < 20.04$), but roughly half are about one magnitude fainter in the rest-frame UV (1450 ,Å), suggesting that these sources may be affected by obscuration.

To place our results in context with low-redshift quasar studies based on ZTF and SDSS, we corrected the light curves for time dilation and computed Lomb–Scargle periodograms to analyze the power spectral density (PSD). The irregular sampling introduced strong spurious peaks at logarithmic frequencies > -2.2, consistent with those predicted by the spectral window. Binning the light curves mitigated these peaks but did not eliminate them, so we excluded the affected frequency range when fitting the red-noise PSD. We obtained reliable red-noise fitting for 8 quasars, with α values between -0.90 and -3.80 when fitting a single power-law to the binned PSD, and $\alpha_{\rm H}$ between -2.47 and -12.88 with a broker-power law fitted to the unbinned PSD. Most of these slopes are consistent with Petrecca et al. (2024) or (Arévalo et al. 2023), while the steepest slopes may reflect the presence of very massive black holes. The remaining 11 sources show PSDs dominated by white noise or narrow quasi-periodic features, suggesting that their variability is not robustly measured rather than intrinsically different from the canonical stochastic behavior of quasars. These cases likely reflect the limited sensitivity and sparse sampling of the unWISE light curves at faint flux levels, which hinder reliable characterization of variability at high redshift.

Chapter 4

Radio-to-X-ray modelling of AGN SEDs

This chapter is based on Martínez-Ramírez, L. N., Rivera, G. C., Lusso, E., Bauer, F. E., Nardini, E., Buchner, J., et al., A&A, 6888, A46 (2024)

4.1 Introduction

In this part of the thesis, we develop AGNfitter-rx, an extended version of the publicly available bayesian python code AGNfitter (Calistro Rivera et al. 2016). AGNfitter models the SED of AGNs and their host galaxies from the submillimeter (submm) to the UV (rest-frame $11 < \log \frac{\nu}{\mathrm{Hz}} < 16$) with four physical components: the accretion disk, the hot circumnuclear dust, the stellar populations and the cold dust emission in star-forming regions (see detailed description in Calistro Rivera et al. 2016).

AGNfitter explores the parameter space using a Bayesian Markov Chain Monte Carlo method, based on the emcee library (Foreman-Mackey et al. 2013). Through the random sampling of the parameter space, the code recovers posterior probability density functions (PDFs) of the physical parameters driving the multiwavelength emission. This Bayesian approach allows for the introduction of prior knowledge on parameter distributions, which is useful, for example, to consistently enforce the energy balance among otherwise in-

dependent components such as the optical-UV attenuated stellar emission and cold dust emission (Da Cunha et al. 2008).

The new release AGNFITTER-RX expands the coverage to lower and higher frequencies ($8 < \log \frac{\nu}{\mathrm{Hz}} < 20$), by introducing two additional physical components to model several orders of magnitude in frequency in the radio regime, as well as the X-ray emission. Radio models, as well as the torus, accretion disk, stellar populations and cold dust models, consider different levels of contribution, making it possible to model diverse populations spanning from radio-quiet to radio-loud AGNs. In particular, studying radio-quiet AGNs is interesting as it has been reported that this population seems to have a higher high-energy spectral break compared to radio-loud AGNs (Kang & Wang 2022). The high-energy cutoff suggests a coronae with high temperature and small opacity in AGNs with low radio emission. Considering also that X-rays are now part of the code, it opens the possibility to study this possible correlation between radio and X-ray emission.

One of the key features of this release is the introduction of new libraries of theoretical and empirical emission models, which improve its flexibility and enable customisation in a straightforward manner. Additionally, it now includes an alternative Bayesian methodology for the parameter space exploration with the nested sampling Monte Carlo algorithm ULTRANEST (Buchner 2014, 2019). The new developments in the structure of the code allow for a high computational efficiency and customization capability of the model library, filter library and priors. This opens up endless possibilities to tackle the SED-fitting task. AGNFITTER-RX thus works as an interface between the observational and theoretical modeling community, as its flexibility allows the easy addition of new templates and comparisons between existing and new models.

While AGN components have been included recently in SED-fitting codes (e.g., Leja et al. 2018; Yang et al. 2020; Thorne et al. 2022), most of these have a smaller wavelength coverage and are all focused on inferring the potential impact of the presence of an AGN on the galaxy parameters, AGN fractions, or AGN identification. In addition to these tasks, AGNFITTER-RX is further tailored as a tool for characterizing the AGN physics, and for

robustly inferring the physical parameters associated with the multiwavelength emission in the AGN. Also, despite the large diversity of existing theoretical models for both the torus and the accretion disk, only a few studies have focused on comparing the performance of different models in SED fitting (García-Bernete et al. 2022; González-Martín et al. 2019; Esparza-Arredondo et al. 2021; Cerqueira-Campos et al. 2023).

In this chapter, we introduce the new functions that enable the flexible inclusion of new physical models and instrument filters, as well as the new library of physical models included in the updated code. Also, we demonstrate the use of AGNfitter-rx by applying the algorithm to a sample of nearby AGNs in order to compare the capabilities of different state-of-the-art torus and accretion disk models in reproducing the photometric data.

4.2 Methodology

4.2.1 Flexibility for adding new filters and models

AGNfitter-rx includes a compilation of 182 published filter curves from the most widely used telescopes. The user can add new filters by providing the corresponding telescope responses. However, for most X-ray CCD instruments and interferometric radio or submillimeter data, observatories do not supply standard photometric filters due to the complexity of their instrumental response curves. In these cases, when photometric filters are not available, a boxcar or Gaussian-like functions can be defined and included.

AGNfitter-rx is designed to allow users to add custom models for each of the physical components. The models must be entered as a Python dictionary containing a grid of templates. Each template corresponds to a spectral flux density as a function of the rest-frame frequency and should be defined by a unique set of parameter values. The combination of all loaded model parameters ultimately defines the parameter space of the total model. This development allows us to include and compare simultaneously several competing physical models.

4.2.2 Flexibility for including priors

One of the main advantages of AGNfitter-rx is the flexibility for the users to include predefined or customary informative priors, based on ancillary information of the source. Optional priors, which are predefined in the code, include:

- Lower and upper limits to the fractional galaxy contribution in the UV/optical can be estimated based on predictions from redshift-dependent galaxy luminosity functions.
- Energy balance between the dust attenuation to the stellar emission in the optical/UV and the reprocessed emission by cold dust in the infrared (MIR and FIR). Three optional energy balance priors are available, allowing for a flexible set-up in which optical-UV dust-absorbed emission sets a lower limit to the infrared emission, a more restrictive condition with cold dust to attenuated stellar luminosities ratio given by a broad Gaussian centred at 1, and no energy balance at all. In contrast to other codes, where energy balance is a fixed property, the flexibility to include or exclude energy balance is particularly relevant for sources where energy balance may potentially break down. These sources include dusty star-forming galaxies at intermediate and high redshifts, where cold dust emission and rest-frame UV emission have been observed to be often spatially disconnected (Calistro Rivera et al. 2018; Buat et al. 2019).
- New priors on the X-ray emission have also been included, which are described in detail in Section 4.2.3.

4.2.3 Physical models from the radio to the X-rays

In this section, we describe the different physical components to model AGN and galaxy SEDs in the radio-to-X-ray regime, which in aggregate constitute the total physical model fit by AGNfitter-rx. For each physical component, several state-of-the-art theoretical, semi-empirical, and empirical models are included in the code, allowing the user to implement,

fit, and compare them.

Radio: Synchrotron star formation and AGN models

The two sources of radio emission considered in the model are the synchrotron emission originating in star-forming regions of the host galaxy and the synchrotron emission from the AGN. The radio emission from star-forming regions is the product of the interaction of hot plasma electrons from supernova remnants and cosmic rays with the magnetic fields of the galaxy. This interaction produces predominantly nonthermal synchrotron emission across the radio regime, as well as thermal Bremsstrahlung radiation, which becomes predominant at frequencies above 30 GHz (Condon 1992).

To account for the radio emission from the host galaxy, we include the semi-empirical starburst templates by Dale & Helou (2002). These are already extended up to radio by using the empirically calibrated IR-radio correlation (e.g., Helou et al. 1985; De Jong et al. 1985):

$$q_{IR} = \log\left(\frac{F_{IR}}{(3.75\times10^{12}\mathrm{Hz})S_{\nu=1.4\mathrm{GHz}}}\right), \tag{4.1}$$

where F_{IR} is the flux between 42.5 and 122.5 μm (in W m^{-2}), $S_{\nu=1.4\mathrm{GHz}}$ the spectral flux density at 1.4 GHz (in mJy), $q_{IR} = 2.34$, and the scatter of the correlation is $\sigma = 0.26$ dex. Also, the assumption that 90% of the $S_{1.4\mathrm{GHz}}$ is given by the synchrotron emission of the cosmic ray electrons in the galaxy, and the remaining 10% by thermal bremsstrahlung radiation produced in the star-forming regions, is made. Both components were modeled by power law functions $S_\nu \propto \nu^{-\alpha}$ with $\alpha = 0.1$ for the thermal component, and $\alpha = 0.8$ for the nonthermal one. We note that of the DH02_CE01 model only the Dale & Helou (2002) templates extend to radio while the Chary & Elbaz (2001) templates do not (see top panel of Fig. 4.1).

Furthermore, we apply the total IR version of the same IR-radio correlation (e.g., Bell 2003) to the recently added S17 cold dust templates (see 4.2.3). This relation establishes

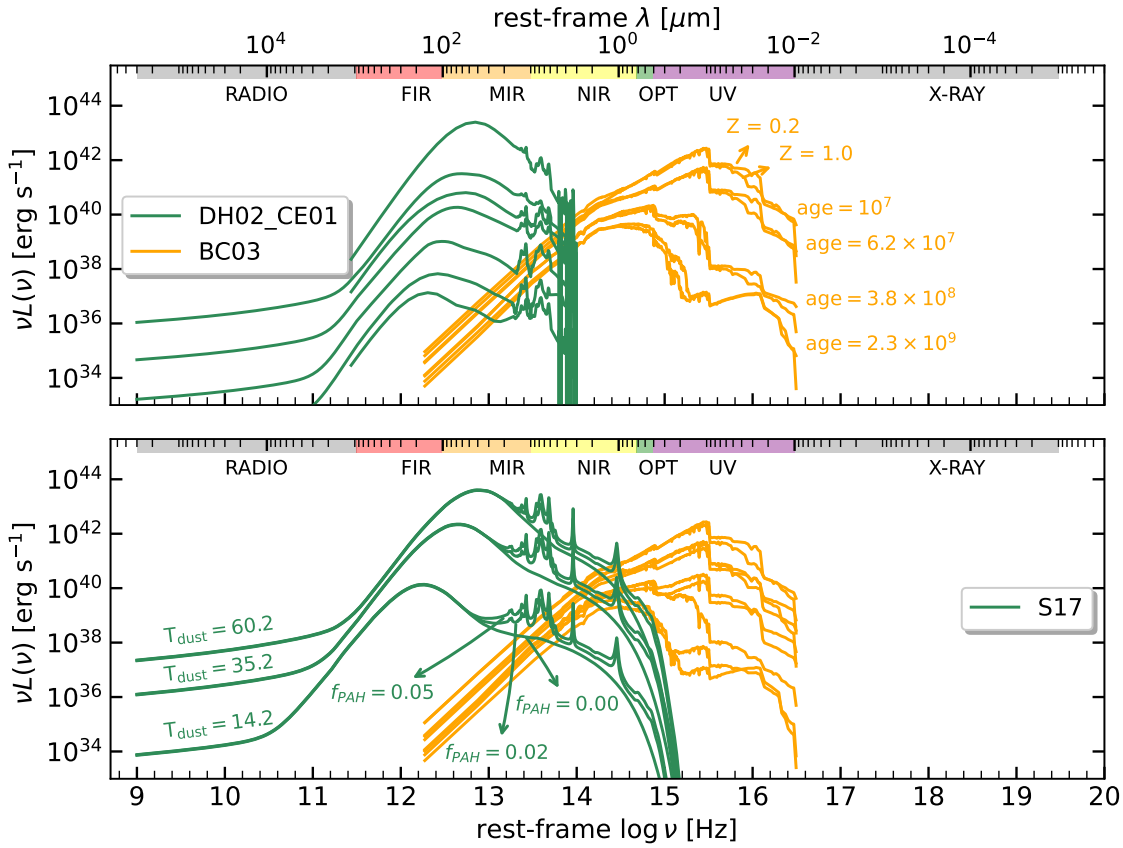

Figure 4.1: Examples of the model templates included in AGNFITTER-RX for the physical components responsible for the emission of the host galaxy and its star-forming regions. At radio frequencies, the host-galaxy SED is dominated by the synchrotron emission associated with the diffusion of cosmic rays from SNR and PWN acceleration sites through the galaxy ISM, and the IR emission is dominated by the cold dust emission (green lines), both of which are associated with regions of high star-formation. The NIR–optical–UV emission is dominated by the stellar emission (yellow lines). The upper panel shows a subsample of semi-empirical templates by Chary & Elbaz (2001) and Dale & Helou (2002) for the emission from star-forming regions and the stellar population synthesis models by Bruzual & Charlot (2003) for SFH of 0.05×10^9 yr, and different ages and metallicities. The lower panel shows theoretical models for cold dust by Schreiber et al. (2018) for increasing dust temperatures and $f_{\rm PAH}$ smoothly joined to the radio emission estimated with the FIR-radio correlation (eq. 4.1).

a link between the rest-frame $8 - 1000\ \mu m$ integrated luminosity ($L_{\rm IR}$ in erg s^{-1}) from the cold dust in star-forming regions, and the emission at 1.4 GHz ($L_{1.4\rm GHz}$ in erg s^{-1} Hz^{-1}). We adopted the value of $q_{\rm IR} = 2.64 \pm 0.26$ from Bell (2003) and undertook a conservative assumption of $q_{\rm IR} = 2.64 + \sigma$ to include sources with faint radio fluxes. The current implementation is in the range of values found by Molnár et al. (2021) for $L_{\rm IR}$ lower than 10^7 [$L_\odot$]. The nonlinearity of the infrared-radio correlation, due to its dependence on redshift (Smith et al. 2014; Calistro Rivera et al. 2017; Molnár et al. 2021) and stellar mass (Delvecchio et al. 2021), is being currently explored and there is still no consensus on the behavior of $q_{\rm IR}$.

To extend the host galaxy SED to radio frequencies, we assumed contribution percentages to the $L_{1.4\rm GHz}$ of 90% and 10% for nonthermal and thermal components, in line with the values presented by Condon (1992) and Dale & Helou (2002), while more recent studies suggest the thermal fraction depends on the frequency and galaxy-type, ranging from $\sim 13\% - 45\%$ at a few GHz in local galaxies (e.g., Ghasemi-Nodehi et al. 2022; Irwin et al. 2024). Power laws with $\alpha = 0.1$ for the thermal (Dale & Helou 2002; Condon 1992), and $\alpha = 0.75$ for the nonthermal (Baan & Klöckner 2006) emissions were computed from 1 to 10 GHz, and then smoothly joined to IR cold dust emission.

Radio emission from AGNs can originate from several physical processes, and establishing its exact origin or the respective contributions is an active field of research (e.g. Panessa et al. 2019). AGN processes that produce radio emission are relativistic radio jets of diverse scales (from subgalactic to giant FRI or FRII structures), shocks from AGN outflows, as well as the X-ray corona (Panessa et al. 2019). Independent of these assumptions, AGNfitter-rx implements four different functional forms of increasing complexity to describe the radio emission related to the AGN (light-blue curves in Fig. 4.2) following Azadi et al. (2023). While these models are not linked to a particular physical origin, they nonetheless provide useful information on the shape of the SED and the physical conditions of the synchrotron emission. These properties can then be associated with physical parameters obtained from the multiwavelength modeling to establish relations which can

further push our understanding of the nature of the radio emission. The complexity of the radio AGN model depends on the availability of radio data points.

- Simple power law (SPL): If the observed SED has 1 or 2 radio bands, AGNfitter automatically assumes a power law defined by a slope fixed at -0.75 or fit over the range $[-2.0, 1.0]$ (consistent with the values found by Zajaček et al. 2019), respectively.

- Double power law (DPL): For SEDs with ≥ 3 radio measurements, this formulation takes into account the transition from optically thin (lobes) to optically thick (cores and hot spots) region at a transition frequency ν_t. The optically thin is given by a synchrotron-aged population of electrons, while the optically thick is due to absorption of radio photons by synchrotron self-absorption. This complex emission is described as follows:

$$\mathrm{L}_\nu \propto \left(\frac{\nu}{\nu_t}\right)^{\alpha_1} \left[1 - \exp\left(-\left(\frac{\nu_t}{\nu}\right)^{\alpha_1-\alpha_2}\right)\right] \mathrm{e}^{-\frac{\nu}{\nu_{\mathrm{cutoff}}}} \tag{4.2}$$

  When three bands are available $\alpha_1 = -0.75$, the slope difference $curv = \alpha_1 - \alpha_2$ takes a value in the range $[-0.5, 0.8]$ and ν_t in the interval $[10^7, 10^{13}]$ Hz. Whereas, if there are more measurements α_1 and α_2 are free parameter with values in the ranges $[-1.0, 1.0]$ and $[-1.0, 0.0]$, respectively.

We note that, in all cases, energy due to the ageing of the electron population is accounted for by an exponential term with a cutoff frequency $\nu_{\mathrm{cutoff}} = 10^{13}$ Hz. As discussed in Polletta et al. (2000), it is difficult to find the energy cutoff from the radio data because it relies on a drop in the spectra, while the actual cutoff may be located at lower frequencies. As such, we select a value consistent with the limits of 10^{10} and 10^{14} Hz used in Azadi et al. (2023) and one order of magnitude lower than the upper limits found by Polletta et al. (2000). With these prescriptions and considering the component normalization, the minimum number of parameters possible to fit radio fluxes is 1 (i.e., normalization parameter RAD) and the maximum number is 4 (i.e., $\alpha_1, \alpha_2, \nu_t$, RAD).

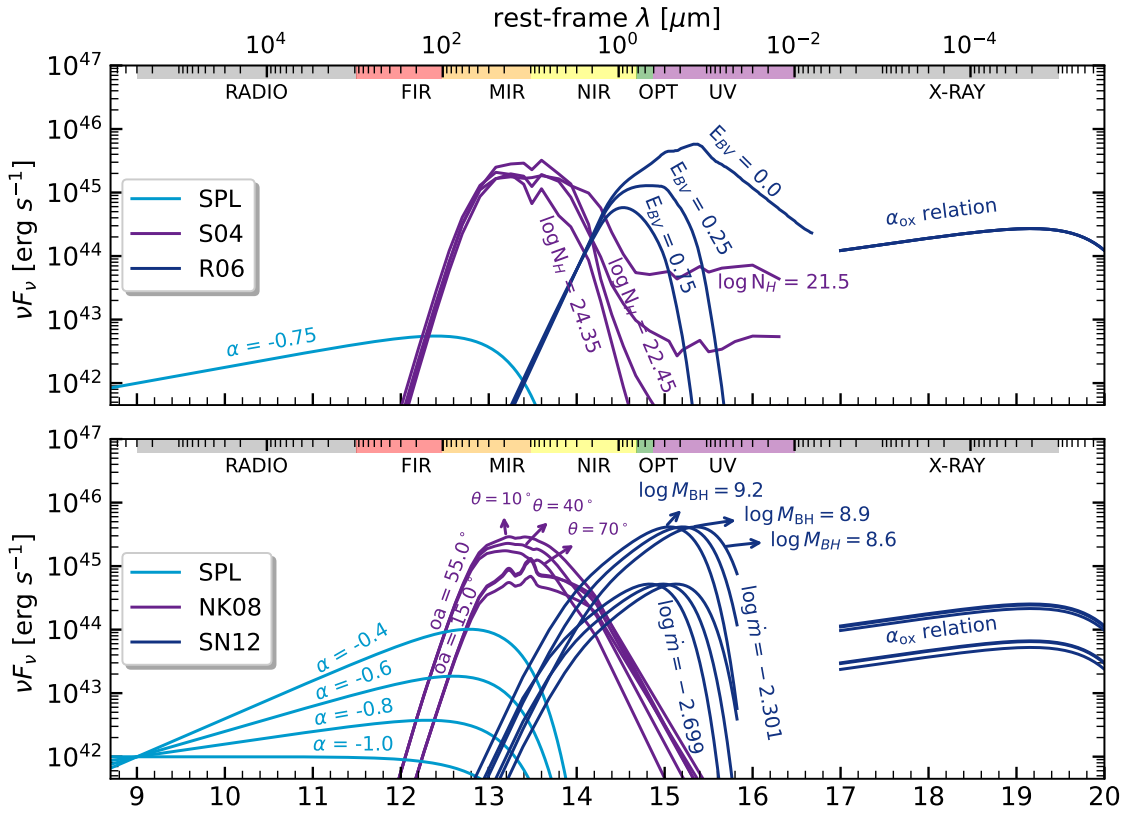

Figure 4.2: Examples of the model templates included in AGNFITTER-RX for three distinct physical components of the nuclear emission (AGN) in active galaxies. At radio frequencies, the nuclear SED is dominated by the synchrotron emission (light blue curves) associated with jets and core emission at $\sim 10^{-1} - 10^{6}$ parsec scales. The MIR emission is dominated by the hot dusty "torus" emission (purple curves) arising at ~ 10 pc scales. The optical-UV-X-ray emission is dominated by the accretion disk and X-ray corona emission (dark blue curves) at $\sim 10^{-2}$ pc scales. The first panel shows a SPL in radio, a subsample of torus templates by Silva et al. (2004) with different column densities, and the accretion disk model by Richards et al. (2006b) with an increasing reddening parameter. The second panel presents SPL emission with different slopes, torus models by Nenkova et al. (2008) for different inclination and opening angles, and optical depth of 10, and accretion disk models by Slone & Netzer (2012) for diverse black holes masses and accretion rates.

Infrared: Cold dust models

The emission of the cold dust from the star-forming regions in the host galaxy is modeled in AGNFITTER-RX using two optional sets of templates: a compilation of the semi-empirical

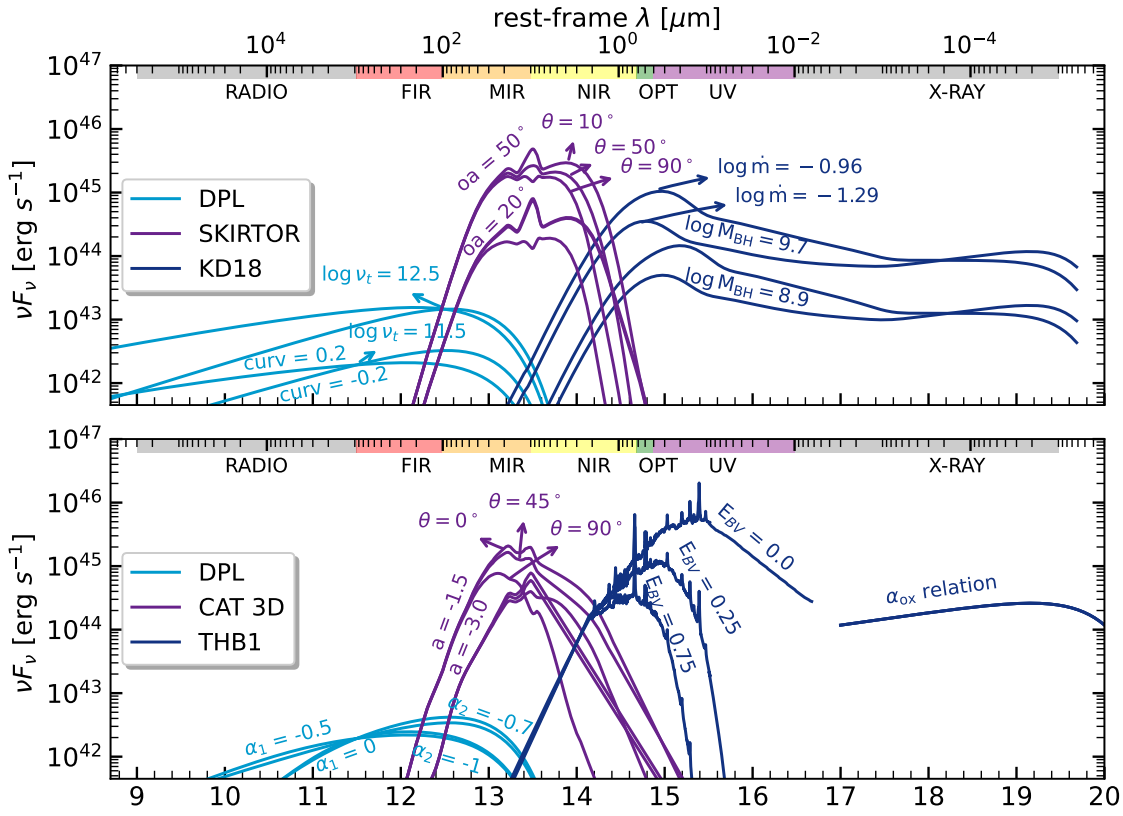

Figure 4.3: Continuation Fig 4.2. The first panel shows a DPL with *curv* and $\log \nu_t$ as free parameters; torus templates by Stalevski et al. (2016) with optical depth of 3 and diverse inclination and opening angles; and accretion disk models by Kubota & Done (2018) defined by black holes masses and accretion rates. The second panel shows a DPL with $\log \nu_t = 11.5$ Hz and different values of α_1 and α_2; torus model by Hönig & Kishimoto (2017) depending on inclination angles, index of cloud distribution and cloud fractions; and the accretion disk template by Temple et al. (2021b).

libraries by Chary & Elbaz (2001) and Dale & Helou (2002) or the theoretical set of SEDs by Schreiber et al. (2018), hereafter denoted DH02_CE01 and S17, respectively.

- DH02_CE01: A detailed description of DH02_CE01 library, which was included in the first version of the code, can be found in Calistro Rivera et al. (2016). The upper panel of Fig.4.1 presents a subsample of SEDs of the library.

- S17: The cold dust SED templates of Schreiber et al. (2018) consist of two independent components: the dust continuum and the mid-infrared emission line spectra from complex molecules. The first one is produced by big (> 0.01 μm) and small (< 0.01 μm) silicate and carbonate grains with "quasi-black body-like" emissivity in the MIR-to-FIR range. The second component corresponds to the emission of lines between $3.3 - 12.3$ μm due to the characteristic vibrational and rotational modes of polycyclic aromatic hydrocarbons (PAHs). Each component is modeled by 150 templates defined by two parameters: the dust temperature and the mass fraction of the PAH component ($f_{\rm PAH}$), which weights the contribution of PAHs to the total cold dust spectrum. The total spectrum is found by adding the SED of dust and PAHs for each possible combination of $T_{\rm dust}$ and $f_{\rm PAH}$, resulting in a flexible model consisting of 9600 templates. The flexibility of this model can be particularly useful to capitalize on new detailed observations of PAH emission in high-z galaxies and AGN with JWST. A subsample of templates is plotted in the lower panel of Fig.4.1. For more details on this model, the reader may refer to Schreiber et al. (2018).

Thus, the number of free parameters is 2 (i.e., log IRlum and the normalization parameter SB) when choosing the DH02_CE01 model and 3 when choosing the S17 (i.e., $\rm T_{dust}$, fracPAH and SB).

Infrared: AGN torus models

To model the emission from the nuclear hot dust and gas, commonly referred to as the AGN torus, we implement four different models: the semi-empirical model by Silva et al. (2004) and three theoretical models by Nenkova et al. (2008), Stalevski et al. (2016) and Hönig & Kishimoto (2017) hereafter S04, NK08, SKIRTOR and CAT3D-Wind, respectively. The theoretical models assume the existence of a population of ("clumpy") dust clouds, resulting in lines of sight with probabilistic opacities that depend on the cloud spatial and density distribution.

- S04: These semi-empirical templates consider the emission of a smooth distribution of dust. They were included in the first version of AGNfitter (see Calistro Rivera et al. (2016) for a detailed description) and are presented in the upper panel of Fig. 4.2.

- NK08: The CLUMPY model by Nenkova et al. (2008) considers dust clouds with a standard Galactic composition (53% silicates and 47% graphite) and column density of $10^{22} - 10^{23}$ cm^{-2} spread out in an axisymmetrical torus, internally limited by the dust sublimation region. The cloud distribution follows a Gaussian profile in the polar angle and a power law profile in the radial coordinate.

- SKIRTOR: The two-phase model by Stalevski et al. (2016) consists of a torus composed of high-density clumps immersed in a smooth distribution of low-density dust. Templates were calculated with the continuum radiative transfer code SKIRT (Baes & Camps 2015; Camps & Baes 2015, 2020) by considering the geometry of a flared disk truncated by the dust sublimation region. The dust distribution follows a power law profile with an exponential cut-off as a function of radius and polar angle, while the clumps scale radius of 0.4 pc and a density profile given by a standard smoothing kernel with compact support[1], are generated randomly.

- CAT3D: The CAT3D-Wind model (Hönig & Kishimoto 2017) considers dust and gas around AGNs as an ensemble of an inflowing clumpy disk and outflowing wind. Similar to NK08, the distribution of clouds with optical depth $\tau_V = 50$ and radius of 0.035 times the graphite grains sublimation radius at 1500 K ($r_{sub} \sim 1$ pc), is given radially by a power law but in the vertical direction follows a Gaussian distribution. A unique feature of this model is the inclusion of a polar wind component, which is embedded in a hollow cone and, due to the sublimation in the innermost region of the torus, is only composed of big grains of dust.

[1] See `https://www.particleworks.com/technical_column_vol3_02_en.html` for a brief explanation.

In the last three cases we include simplified versions of the models, since the original ones have 6, 6 and 8 fitting parameters, respectively. These large sets of SEDs are computationally demanding and may lead to overfitting of the data. Furthermore, in some cases, the spectral resolution of the SEDs may be insufficient to detect the effect of each parameter, which would prevent the parameters from converging. For that reason, we reduced the number of templates by averaging all the models with respect to all but the 3 most relevant parameters (discussed through private communication with the authors of the models). We show in Section 4.3.5 that the reduction of the parameter space in these models is an approach to produce templates suitable for fitting. Those simplified templates deal with limitations relying on the existence, sensitivity, bandwidth, and spectral resolution of the rest-frame MIR data.

For NK08 and SKIRTOR are the inclination angle, opening angle and optical depth of the torus, while for CAT3D they are the inclination angle, index of the radial power law of the cloud distribution and fraction of clouds in the wind compared to the torus plane. As a result, our final SED grid consists of 1080, 400 and 378 templates for NK08, SKIRTOR and CAT3D, respectively, of which a subsample is plotted in the different panels of Fig. 4.2. As a consequence of the previous model reductions and considering the normalization parameter, a minimum of 2 (i.e., the normalization parameter TO and $\log N_H$ or incl) and a maximum of 4 parameters (i.e., TO, incl, oa and τ_ν or TO, incl, a and fwd, more details in table 4.1) are required to model the torus emission.

Near-infrared/optical/UV: Stellar population models

The stellar emission templates were built employing the stellar population synthesis as in the first version of AGNfitter. We took the single stellar population (SSP) models from Bruzual & Charlot (2003) defined for ages from 10^7 to 10^{10} yr and metallicities ranging from 0.004 to 0.04. This formalism includes a prescription for thermally pulsing stars in the asymptotic giant branch. To calculate the composite stellar population models we include the initial mass function by Chabrier (2003) and an exponentially decreasing star

formation history (SFH). To account for a wider variety of stellar populations, 10 different characteristic times for the decay of SFH were considered with values between 0.05 to 11 Gyr. Unlike the first version of AGNfitter, metallicity is now a free parameter of the SSPs. AGNfitter-rx builds the stellar emission library using the smpy code (Duncan & Conselice 2015), which gives the user the flexibility to include new templates with different star formation histories if desired. With new updates, the number of parameters required to fit the galaxy emission model is 4 or 5 (i.e., the normalization parameter GA, τ, age, Z, $E(B\text{-}V)_{gal}$) if we include metallicity.

Optical/UV: Accretion disk models

One of the main strengths of AGNfitter in comparison to other SED-fitting codes is the superior modeling of the BH accretion disk emission and its reddening due to dust attenuation. AGNfitter-rx further improves on this by including four libraries for the detailed modeling of the accretion disk emission[2]: a modified version of the semi-empirical templates by Richards et al. (2006b) and Temple et al. (2021b) denoted as R06 and THB21, respectively; and the theory-built sets of templates by Slone & Netzer (2012) and Kubota & Done (2018), denoted as SN12 and KD18, respectively.

- R06: The big blue bump model by Richards et al. (2006b) is based on observed quasar spectra and a near-infrared (NIR) blackbody tail. The library is composed of one SED which is then modified by different levels of dust reddening following the Small Magellanic Cloud (SMC) dust attenuation law (upper panel of Fig. 4.2), and was implemented in the first version of the code (for more details see Calistro Rivera et al. 2016).

- SN12: The emission model by Slone & Netzer (2012) is based on an α-disk model (optically thick and geometrically thin accretion disk), and the inclusion of rel-

[2]We use "accretion disk emission" to refer to the UV and optical emission from the region inside the dust sublimation radius, including the broad line region.

ativistic temperature corrections, the effect of viscous dissipation and the comptonization of the radiation in the disk atmosphere. We include a simplified model by assuming a zero-spin black hole and windless disk, resulting in a set of 108 templates (some of them presented in the upper central panel of Fig. 4.2). The templates are parametrized by the mass of the SMBH taking values between of $\log \mathrm{M_{BH}} = [7.4, 9.8]$ in intervals of $\Delta \log \mathrm{M_{BH}} = 0.3$ and the accretion rate with values in the range $\dot{\mathrm{M}}/\dot{\mathrm{M}}_{\mathrm{edd}} = [0.0, 310.116]$ in variable intervals in the range $\Delta\dot{\mathrm{M}}/\dot{\mathrm{M}}_{\mathrm{edd}} = [0.001, 154.69]$.

- KD18: The theoretical model AGNSED (Kubota & Done 2018) considers the emission of an outer Novikov-Thorne accretion disk, an inner warm Comptonising region and a hot corona, all of them radially delimited. We adopted a simplified version which implied setting all the parameters except the black hole mass and the accretion rate to their typical values, and assumed a second comptonisation component as a source of the soft X-ray excess (for more details see Kubota & Done 2018)[3]. The final set is composed of 225 SEDs defined by black hole masses between $\log \mathrm{M_{BH}} = [6, 10]$ in intervals of $\Delta \log \mathrm{M_{BH}} = 0.286$ and accretion rates in the range $\log \dot{\mathrm{M}}/\dot{\mathrm{M}}_{\mathrm{edd}} = [-1.5, 0]$ in intervals of $\Delta\dot{\mathrm{M}} = 0.107$. A subsample of the library is shown in the two lower panels of Fig. 4.2.

- THB21: The accretion disk library by Temple et al. (2021b) is a set of composites empirically derived SED of luminous quasars. The templates include a broken power-law that models the accretion disk, a blackbody dominating from 1 to 3 μm due to hot dust; and account for both broad and narrow emission lines as well as additional continuum emission from blended line emission. To soften the edges of the SED in AGNfitter-rx, we add the low ($\nu < 10^{14.15}$ Hz) and high frequency ($\nu > 10^{15.5}$ Hz) tails of the R06 model by normalizing at the transition frequencies. The final template corresponds to a zero redshift and nonreddened quasar with an

[3]The Python code to generate the templates can be found at `https://github.com/arnauqb/qsosed`

average emission line template, and it is shown in the last panel of Fig. 4.2 after applying the AGNfitter reddening prescription.

As mentioned in the R06 description, the different levels of extinction produced by dust are accounted for by applying the SMC reddening law (Prevot et al. 1984) to the nonreddened templates of all the aforementioned accretion disk models. The required parameter in this prescription is the reddening $E(B\text{-}V)_{BBB}$, with values between 0 and 1. In summary, to fit the accretion disk component, the user needs between 2 (i.e., the normalization parameter BB and $E(B\text{-}V)_{BBB}$) and 3 (i.e., $\log M_{BH}$, $\dot{M}/\dot{M}_{edd}$ and $E(B\text{-}V)_{BBB}$) parameters.

X-rays: Corona and star formation

For optical-UV accretion disk models which do not include X-rays, we add an X-ray model. Including the X-ray emission in AGN SED fitting is crucial, as X-ray fluxes can provide additional information to constrain accretion disk emission and break potential degeneracies between the blue accretion disk emission and galaxies with young/blue stellar populations in the optical-UV region. The X-ray emission is produced in the hot corona in the vicinity of the black hole, containing information related to the accretion process and the black hole properties. Considering the above, we used the empirical correlation $\alpha_{ox} - L_{2500\text{Å}}$ (Lusso & Risaliti 2017; Just et al. 2007) which connects the coronae emission at 2 keV with the emission of the AGN at 2500 Å:

$$\alpha_{ox} = -0.137 \log(L_{2500\text{Å}}) + 2.638, \tag{4.3}$$

where α_{ox} is the slope of the SED between 2500 Å and 2 keV given by:

$$\alpha_{ox} = -0.3838 \log(L_{2500\text{Å}}/L_{2keV}). \tag{4.4}$$

Based on the 2500 Å flux, we computed the expected α_{ox} from the relation. Then, we included a grid of dispersion values of $\Delta\alpha_{ox}$, with respect to the empirical relation, to

calculate the corresponding 2 keV fluxes for each accretion disk template. The dispersion of the relation is a fitting parameter taking a maximum value of $|\Delta\alpha_{\rm ox}| = 0.4$ reported by Lusso & Risaliti (2016) for low-quality data and sources with variability. Once the X-ray normalization is fixed, the SEDs were extended from 2 keV to more energetic X-rays by assuming a power law with a photon index of 1.8 and an exponential cut-off at 300 keV as assumed in Yang et al. (2020). When at least 2 X-ray measurements are available, the photon index is a free-fit parameter that takes values from 0.00 to 3.00 in intervals of $\Delta\Gamma = 0.15$. We note that obscuration is not accounted for, and therefore, this effective power law carries information about both obscuration and potential accretion rate estimates. Also, $\alpha_{\rm ox} - L_{2500\text{Å}}$ relation is valid to constrain soft X-ray emission only for type-1, radio-quiet, non-BAL[4] AGNs (Strateva et al. 2005). Similarly, the SED of Blazars is likely modeled with synchrotron emission from radio to X-rays, and therefore $\alpha_{\rm ox} - L_{2500\text{Å}}$ relation is not providing reliable physical interpretation for those objects.

As part of the postprocessing analysis, we compute the host galaxy contribution to the X-rays due to the emission of high-mass X-ray binaries (HMXBs) and the hot ionized interstellar medium (ISM) using the formula provided by Mineo et al. (2014):

$$L_{0.5-8\rm keV} = (4.0 \pm 0.4) \times 10^{39} \rm SFR \tag{4.5}$$

where $L_{0.5-8\rm keV}$ is the luminosity between 0.5 and 8 keV in erg s^{-1} and SFR the star formation rate in $\rm M_\odot yr^{-1}$. For this purpose, we used the IR SFR computed from the Murphy et al. (2011) calibration for the integrated $8 - 1000$ μm luminosity of the cold dust emission:

$$\left(\frac{\rm SFR}{\rm M_\odot yr^{-1}}\right) = 3.88 \times 10^{-44} \left(\frac{L_{\rm IR}}{\rm erg\ s^{-1}}\right). \tag{4.6}$$

Although we selected the most robust SFR estimator, it is worth noting that AGNfitter-rx, as in its initial release, also outputs optical/UV SFR estimates based on stellar population

[4]BAL: Broad absorption line

emission (see Calistro Rivera et al. (2016) for more details). We included the host galaxy X-ray emission derived from the IR SFR in the estimation of the AGN fraction.

Table 4.1 presents a summary of the model-fitting parameters in AGNfitter-rx, including names, values, and descriptions. The AGNfitter code (Calistro Rivera et al. 2016), including the AGNfitter-rx code release, are publicly available as open-source code in Python 3 in `https://github.com/GabrielaCR/AGNfitter`.

4.2.4 Data: Radio-to-X-ray SEDs of local AGN

To test the capabilities of AGNfitter-rx, we fit the SEDs of a sample of 36 nearby active galaxies selected from the AGN SED ATLAS by Brown et al. (2019), spanning a redshift range of $0.01 < z < 0.7$ and apparent g-band magnitudes between 12.8 and 20.7. Table 4.2 summarizes the sample, listing key information such as coordinates, redshift, spectral classification, and the number of photometric bands available for the fitting.

This sample was chosen as it comprises a well-characterized and diverse set of local AGNs, with well-constrained photometric measurements with a high signal-to-noise ratio. To derive these synthetic SEDs, individual spectra were combined and scaled by a multiplicative factor to match aperture photometry measured from images using circular apertures. The scaling factors aim to mitigate the effects of different instruments, extraction apertures, and AGN variability, which produce discontinuities in the SEDs. The resulting photometric uncertainties of the synthetic SEDs are extremely small compared to the errors in the photometric calibrations, scatter introduced by AGN variability and sources of systematic error since the AGNs in our sample are very bright ($m_g < 20.7$). The photometric data is a compilation of available observations from GALEX, *Swift*, SDSS, PanSTARRS, Skymapper, 2MASS, WISE, *Spitzer* and *Herschel*, covering a broad wavelength range from the UV to the IR. For more details on the photometry, we refer the reader to Brown et al. (2019).

According to the spectroscopic classification by Véron-Cetty & Véron (2010), this sample includes eight Seyfert 1, five Seyfert 1n (Seyfert 1s with narrow Balmer lines), six

Table 4.1: AGNFITTER-RX fitting parameters

Component	Model	Parameter	Description	Range
Galaxy	BC03_metal	τ	Time scale of the exp. SFH [log year]	[0.05, age(z)]
		age	Galaxy age [log year]	[7, age(z)]
		Z	Metallicity [$Z_\odot$]	[0.2, 2]
		$E(B\text{-}V)_{gal}$	Galaxy reddening	[0,0.6]
		GA	Galaxy normalization	[-10, 10]
Cold dust	DH02_CE01	log IRlum [$L_\odot$]	Parametrization	[7, 15]
	S17	T_{dust}	Dust temperature [K]	[14.24, 42]
		fracPAH	PAHs fraction	[0, 0.05]
	DH02_CE01, S17	SB	Starburst normalization	[-10, 10]
Torus	S04	log N_H	Column density [log cm^{-2}]	[21, 25]
	NK08, SKIRTOR, CAT3D	incl	Inclination angle [°]	[0, 90]
	NK08, SKIRTOR	oa	Opening angle [°]	[10, 80]
		τ_ν	Optical depth at 9.7μm	[10,300], [3,11]
	CAT3D	a	index of cloud dist.	[-3, -1.5]
		fwd	cloud fraction	[0.15, 2.25]
	S04, NK08, SKIRTOR, CAT3D	TO	Torus normalization	[-10, 10]
Accretion disk	SN12, KD18	log M_{BH}	Black hole mass [log $M_\odot$]	[7.4, 9.8], [6.0, 10.0]
		log $\dot{m}$	Accretion rate[log $\dot{M}/\dot{M}_{edd}$]	[-4, 2.49], [-0.15, 0.0]
	R06, THB21, SN12, KD18	$E(B\text{-}V)_{bbb}$	BBB reddening	[0,1]
	R06, THB21	BB	BBB normalization	[-10, 10]
Hot corona	R06, SN12, THB21	$\Delta\alpha_{ox}$	$\alpha_{ox} - L_{2500\text{Å}}$ dispersion	[-0.4, 0.4]
		Γ	Photon index	[0, 3]
Radio from AGN	SPL	α	Slope of synch. [log ν]	[-2, 1]
	DPL	curv	Difference in slopes	[-0.5, 0.8]
		ν_t	Transition frequency [log ν]	[7, 13]
		α_1	Slope of synch. aged [log ν]	[-1, 1]
		α_2	Slope of synch. self-abs [log ν]	[-1, 0]
	SPL, DPL	RAD	Radio normalization	[-10, 10]

Notes: Description, name, and values of the AGNFITTER-RX fitting parameters. The value age(z) correspond to the age of the Universe at the observing time, according the assumed cosmological model ΛCDM with $H_0 = 70\text{kms}^{-1}\ \text{Mpc}^{-1}$, $\Omega_m = 0.3$ and $\Omega_\Lambda = 0.7$.

Seyfert 1.2, eight Seyfert 1.5, one Seyfert 1.9, two Seyfert 2, one Seyfert 1i (Seyfert 2s with reddened broad line region), one LINER and four BL Lacertae objects.

To test the new AGNfitter developments, we extended the spectral coverage of the Brown et al. (2019) SEDs by including archival photometric data in the radio and X-ray regimes. In the radio regime, we added 1.4, 5 and 15 GHz radio data from OVRO, Parkes, VLA, Green Bank 140-ft, 100-m Effelsberg, 26-m Peach Mountain radio telescopes, obtained through the NASA/IPAC Extragalactic Database[5]. These bands were chosen because they were available for most AGNs (32 out of 36) in the sample. In the X-rays, we added the 0.5-1 keV, 1-2 keV, 2-4.5 keV, and 4.5-12 keV bands from the *XMM-Newton* Serendipitous Source Catalogue 4XMM-DR11 (Webb et al. 2020). As explained by Yang et al. (2020), there are no standard transmission curves for X-ray bands due to the high sensitivity of measurements to observing conditions and source characteristics. Hence, we follow the procedure by Risaliti & Lusso (2019) to calculate monochromatic fluxes at effective area-weighted mean energies (henceforth called pivot points) and include delta Dirac filters centered at those energies.

We got the pivot energies 0.76, 1.51, 3.22, and 6.87 keV for bands 2, 3, 4, and 5 of *4XMM-Newton*, respectively, through private communication with the authors of the catalog. We computed correction factors (f_{corr}) to consider the effect of Galactic absorption while correcting for the difference between N_H at the source location and the catalog assumed value. Factors were computed based on the reported and reference fluxes from Webb et al. (2020) for $\Gamma = -1.42$ and $N_H = 1.7 \times 10^{20}$ cm^{-2}, and the column densities from the gas map of the Python package gdpyc[6]. Thus, the corrected monochromatic fluxes at the pivot points were estimated as follows:

$$F_{\nu_2-\nu_1(\mathrm{corr})} = \nu^{(1+\Gamma)} \frac{F_{\nu_2-\nu_1}(2+\Gamma)}{\nu_2^{2+\Gamma} - \nu_1^{2+\Gamma}} f_{\mathrm{corr}}, \tag{4.7}$$

where $F_{\nu_2-\nu_1}$ is the catalog reported flux within a given band, ν_1 is lower and ν_2 the

[5] `https://ned.ipac.caltech.edu/`

[6] Gas and Dust Python Calculator

Table 4.2: Sample of 36 local AGNs from SED ATLAS.

Name	RA (J2000)	DEC (J2000)	z	AGN type	Available bands
2MASXJ13000533+1632151	195.0223	16.5374	0.0799	S1i	41
3C 120	68.2962	5.3543	0.0330	S1.5	47
3C 273	187.2779	2.0524	0.1583	S1.0	49
3C 351	256.1724	60.7418	0.3719	S1.5	42
3C 390.3	280.5375	79.7714	0.0561	S1.5	42
Ark 120	79.0475	-0.1498	0.0327	S1.0	49
Ark 564	340.6639	29.7254	0.0247	S3	41
Fairall 9	20.9407	-58.8058	0.0470	S1.2	43
F2M1113+1244	168.4777	12.7443	0.6812	S1	57
H 1821+643	275.4888	64.3434	0.2968	S1.2	42
IRAS 11119+3257	168.6620	32.6926	0.1876	S1n	42
IRAS F16156+0146	244.5392	1.6559	0.1320	S2	56
Mrk 110	141.3035	52.2862	0.0353	S1n	57
Mrk 1502	13.3956	12.6934	0.0589	S1n	49
Mrk 231	194.0593	56.8737	0.0422	S1.0	48
Mrk 279	208.2644	69.3082	0.0305	S1.0	48
Mrk 290	233.9683	57.9026	0.0302	S1.5	42
Mrk 421	166.1138	38.2088	0.0300	HP	43
Mrk 493	239.7901	35.0299	0.0310	S1n	41
Mrk 509	311.0406	-10.7235	0.0344	S1.5	49
Mrk 590	33.6398	-0.7667	0.0261	S1.0	48
Mrk 817	219.0919	58.7943	0.0315	S1.5	49
Mrk 876	243.4882	65.7193	0.1290	S1.0	49
Mrk 926	346.1812	-8.6857	0.0469	S1.5	47
NGC 5728	220.5996	-17.2531	0.0094	S1.9	43
NGC 7469	345.8151	8.8739	0.0163	S1.5	49
OQ 530	214.9441	54.3874	0.1525	HP	43
PG 0026+129	7.3071	13.2677	0.1420	S1.2	41
PG 0052+251	13.7172	25.4275	0.1545	S1.2	47
PG 1211+143	183.5736	14.0536	0.0809	S1n	43
PG 1307+085	197.4454	8.3302	0.1538	S1.2	47
PG 1415+451	214.2534	44.9351	0.1137	S1.0	41
PG 2349-014	357.9838	-1.1537	0.1738	S1.2	43
PKS 1345+12	206.8892	12.2900	0.1205	S2	41
Ton 951	131.9269	34.7512	0.0640	S1.0	43
W Com	185.3820	28.2329	0.1020	BL	43

Notes: Spectral classifications by Véron-Cetty & Véron (2010).

upper-frequency limits of each band, and a power law function with $\Gamma = -1.42$ (used to derive the fluxes in the 4XMM catalog) is assumed to model the X-ray emission. By definition, the monochromatic flux at the pivot energy does not depend on the photon index value assumed. The above procedure was possible for 29 of 36 AGNs in the sample with available X-ray data.

Our final data set comprises 36 galaxies, from which 25 have radio-to-X-ray coverage with up to 49 photometric data points, seven have radio-to-UV coverage with up to 45 photometric data points, and four have IR-to-X-ray coverage with up to 46 photometric data points. Of the 36 galaxies, 22 have FIR coverage. As photometry was not acquired simultaneously, variability may be affecting our analysis. However, the trade-off is that we have very good wavelength coverage and high-quality photometry to perform the fittings. The extended photometric catalogs are available in a Zenodo repository. [7]

4.2.5 Fitting setup and model comparison strategy

We employed the nested sampling method to perform the fits, which relies on a population of live points to explore parameter space. The iterative process removes the point with the lowest likelihood and searches for a new, independent point on ellipsoids around the remaining live points. The volume of live point decreases while the likelihood values increase. The process runs until the live point population is so small that it does not contribute any probability mass.

This process takes approximately 25 minutes per source to fit a maximum of 19 parameters when executed on a personal computer equipped with an Intel Core i7 (8th generation, dual-core, 2.7–3.5 GHz), 8 GB DDR4 RAM, and an NVIDIA GeForce MX150 GPU. We were able to concurrently run up to four sources, leveraging AGNfitter-rx parallelization. While the computational time aligns with that of MCMC, the methods exhibit distinct convergence criteria. In emcee, convergence relies on assessing the effective number of independent samples within each chain, as measured by the integrated autocorrelation

[7] https://zenodo.org/records/11458139

time. In contrast, UltraNest employs a convergence criterion based on the weight of the live points, since once this weight becomes negligible, it no longer improves the results, even if the iterations persist. The setup parameters were: 400 as the minimum number of live points, 400 effective samples, and 20 steps for the slice sampler.

As the AGNs in our sample are quite diverse, they require slightly different sets of models to optimize the fitting of the observed SED. We can exploit the flexible capabilities of AGNfitter-rx to investigate general trends of the different models included in the public version of the code and characterize how the results depend on the choice of the model. We also investigate the AGN physical components with the largest variety of models in the code, the torus and the accretion disk. For this purpose, we compare the performance of different models associated with a given component, while keeping the models for the other physical components fixed. Isolating the effects of the different torus and disk models is of course complex due to dissimilar wavelength coverage and overlap with other physical components, such as stellar emission and cold dust, for instance.

For this purpose, we first set the R06 model for the disk, S17 for the cold dust, BC03 with metallicity for the stellar population and test the different hot dust models: 1 and 3 parameter versions of NK0, SKIRTOR, CAT3D; and the empirical S04. Once the best torus model was identified, we proceeded to evaluate the different templates of the accretion disk using this model, while maintaining S17 and B03 as the cold dust and stellar population models, respectively. In all the runs, we assumed a flexible prior to constraining the cold dust to be at least as luminous as the dust-absorbed stellar emission.

The primary statistical outputs of the code are the distribution of logarithmic likelihood ($\log \mathcal{L}$) for 100 random model realizations, its corresponding expected fluxes, the residuals compared to the data and the model logarithmic marginal likelihood (log Z from now on evidence). Beyond the quality of the fit informed by the likelihood and residuals, our Bayesian approach allows us to estimate and analyze posterior probabilities of the models given the observed SED P(M|D). According to the Bayes theorem, it is defined as:

$$P(M|D) = \frac{P(D|M)P(M)}{P(D)} \tag{4.8}$$

where P(D|M), P(D), and P(M) correspond to the likelihood, the evidence, and the prior probability of the models, respectively. As the sampling process outputs log $\mathcal{L}$ and log Z, we can compute log-posterior probabilities by assuming all the models are a priori equally probable. Due to the evidence term, posterior odds contain information about the probability of a given model to generate the observed SED by considering all combinations of parameters, enabling a robust model comparison.

4.3 Results

Figures 4.4 and 4.5 present examples of the fitting results that include Seyfert 1, Seyfert 1.5, Seyfert 2 AGN and blazars. The different curves represent the emission of the physical components corresponding to the most probable model (solid color curves) and 10 combinations of parameters randomly selected from the posterior PDFs (hereafter referred to as realizations and shown as shaded areas). The line colors are coded similarly as in Figs. 4.1 and 4.2. The total SEDs are depicted as red lines and the observed photometry as black-filled circles with error bars. The lower part of each panel presents the residuals estimated as the difference between observed and model fluxes divided by the quadrature-combined errors of the photometry for each band. The residuals of the best model and ten random realizations are shown as red circles with outlined and shaded areas, respectively. The annotated maximum likelihood value is the highest value achieved by the live points in the nested sampling fitting process. The SED fitting plot of the remaining sources are also available as supplementary material in the Zenodo repository.

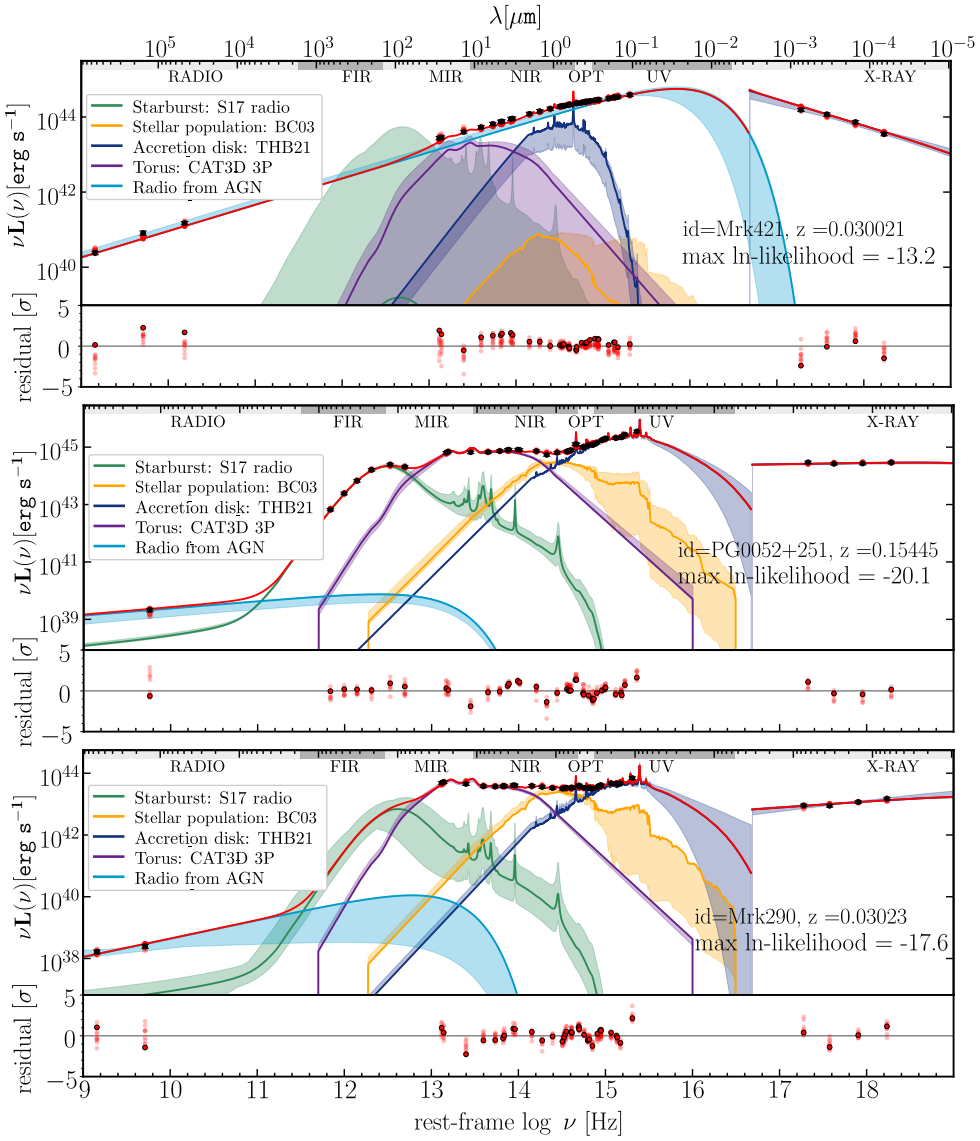

Figure 4.4: Examples of the best SED fittings of the different AGN types: The blazar Mrk 421 (top panel), the Seyfert1 PG0052+251 (middle panel), and the Seyfert 1.5 Mrk 290 (bottom panel). The yellow and green solid curves show the emission from the stellar population and cold dust of the galaxy, respectively. Purple, dark blue, and light blue solid curves show the torus, accretion disk, and radio AGN emission models, respectively. Ten models constructed from combinations of parameters randomly selected from the posterior PDFs (hereafter referred to as realizations) are plotted as shaded areas and the residuals of each fit realization are presented in the graphs below each source. A flexible energy balance prior was assumed so that the cold dust IR emission is at least comparable to the dust-absorbed stellar emission.

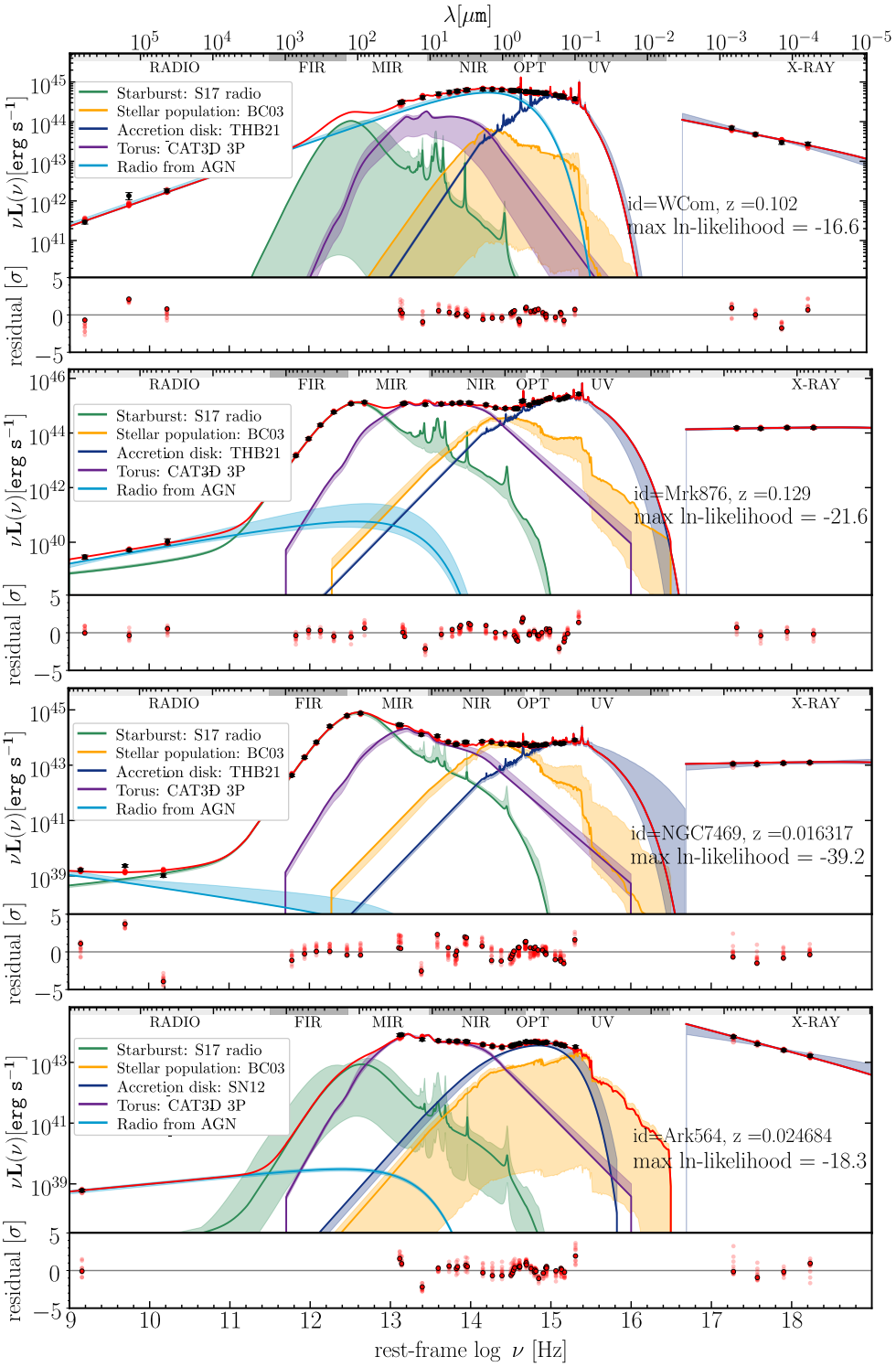

Figure 4.5: Continued from Figure 4.4. Examples of the best SED fittings for the blazar WCom, the Seyfert 1 Mrk876, the Seyfert 1.5 NGC7469 and the Seyfert 2 Ark 564.

4.3.1 Comparing torus models

The first indicator that we use to compare the torus emission models is the likelihood, which is a measurement of the goodness of the fit. The likelihood distributions for 100 realizations of the fits of the 36 galaxies in the sample are presented in the left panel of Fig. 4.6, for each of the torus models. We considered both relatively extended and simplified versions of the models, with 3 and 1 fitting parameters, respectively. Since the likelihood values are on a logarithmic scale, it implies that all the histograms have a broad distribution. This feature is mainly due to the diversity in the AGN sample and highlights that some SEDs are better modeled than others.

The panels of Fig. 4.6 show how increasing model complexity may not necessarily imply a higher quality of the fit, given our current data. Although the NK08 and SKIRTOR models are driven by the same physical parameters and are equally simplified, NK08 has a very large increase in likelihood between the one-parameter and three-parameter models, reaching a difference of the order of $\approx 10^{40}$. To understand the differences, we analyse the spectral slope space covered by our torus models given by $\alpha = -\log(F_\nu(\lambda_2)/F_\nu(\lambda_1))/\log(\lambda_2/\lambda_1)$ with $\alpha_{\rm MIR}$ and $\alpha_{\rm NIR}$ corresponding to $[\lambda_2, \lambda_1]$ equal to $[14, 8]$ μm and $[6, 3]$ μm.

Fig. 4.7 presents the spectral slope space of our torus templates; along with the rest-frame NIR and MIR slopes of our sample of AGNs based on *Spitzer* and WISE photometry. The plot shows how similar the areas covered by SKIRTOR with 3 and 1 parameters are, in contrast to the different regions covered by NK08 with 3 and 1 parameters. This discrepancy in the $\alpha_{\rm MIR}$-$\alpha_{\rm NIR}$ space might be responsible for the considerable likelihood increase from one-parameter to three-parameter templates. Additionally, NK08 one-parameter covers a lower range of $\alpha_{\rm MIR}$ compared to S04 and SKIRTOR one-parameter model, while most of the sources of our sample spread in a wide range of low $\alpha_{\rm MIR}$ values, which explains why the average NK08 is not performing so well. Consistent with the likelihood analysis, we found that CAT3D and NK08 three-parameter models are the closest models to the spectral slope space covered by our sample.

The histograms in Fig. 4.6 show that the more complex NK08 model results in the best

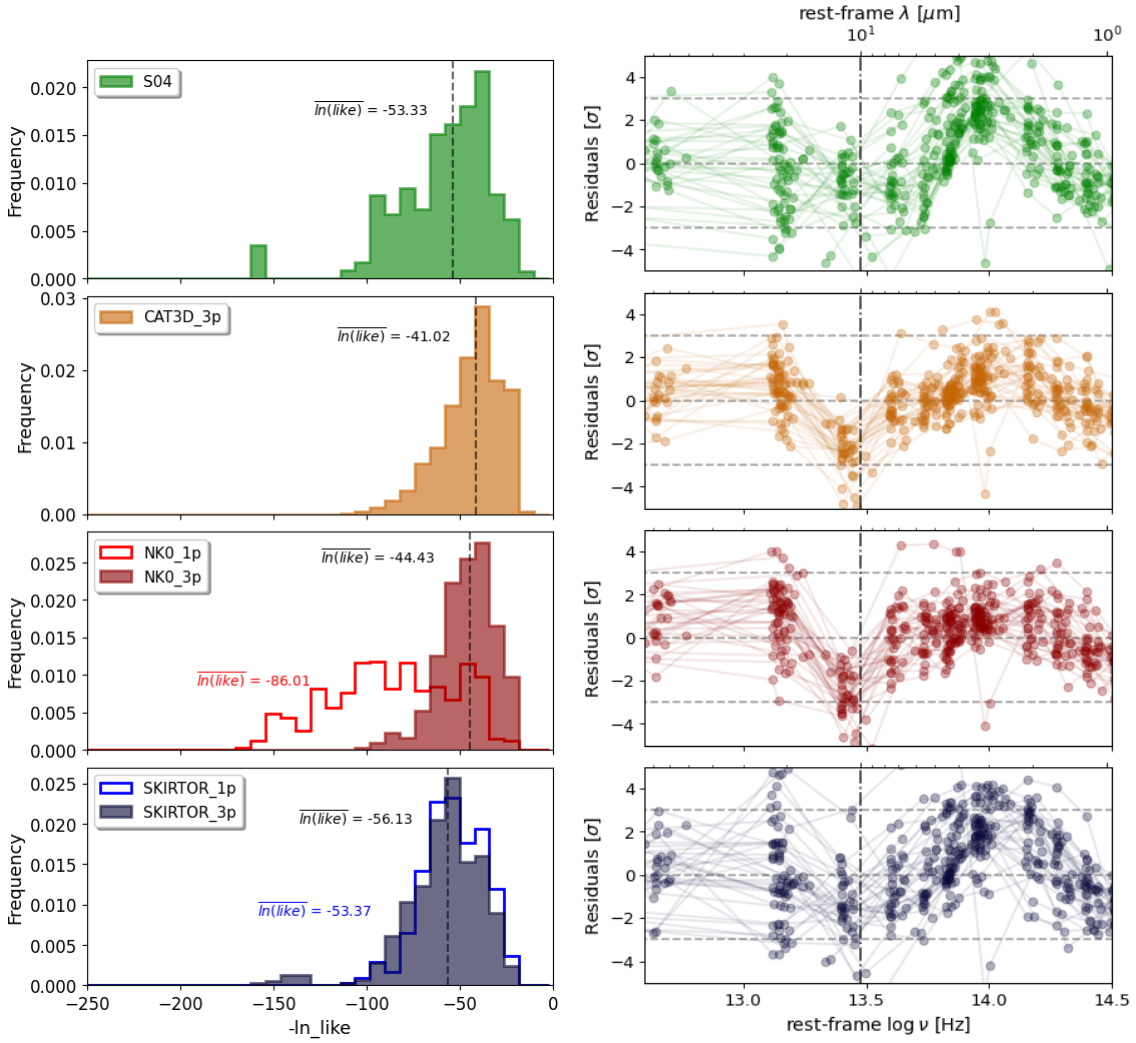

Figure 4.6: Performance analysis of different torus models in SED fitting: The models presented are the homogeneous S04 (upper panel), the clumpy and windy CAT3D (central upper panel), the clumpy NK08 (central lower panel), and the two-phase SKIRTOR model (lower panel). The left panel shows the histograms of the logarithm of the likelihood values corresponding to each model and the dashed line in each plot indicates the median value of the distribution. Residuals from the best fits for each galaxy of the sample are shown in the right panel. The plot shows the existence of templates that manage to capture some SED features (high density of points around zero) while systematically failing in modeling specific regions of the observed SED (high dispersion). The dash-dotted line shows the 10 μm feature and the dashed lines show the limits of 0σ, 3σ, and -3σ.

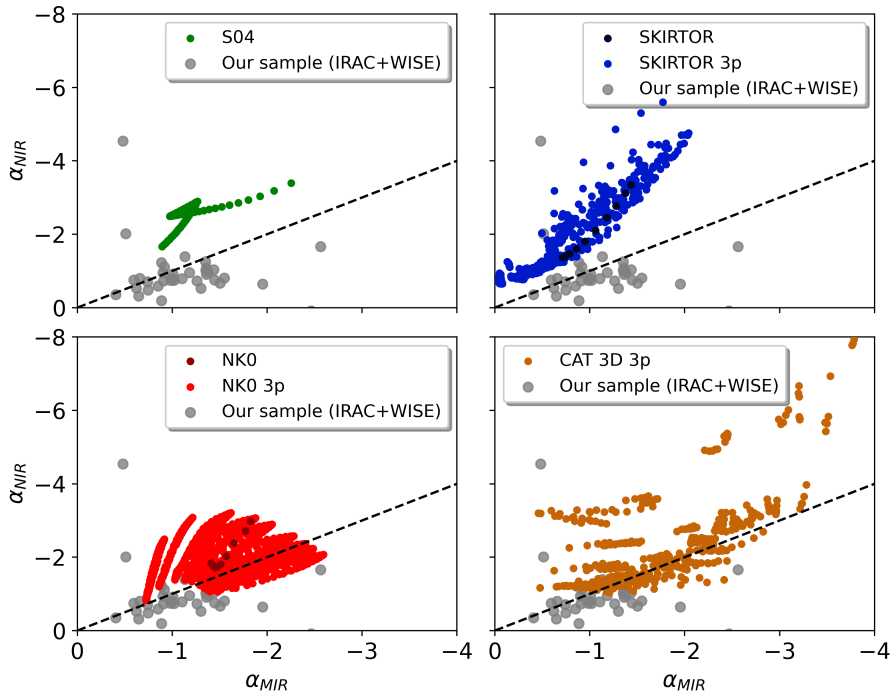

Figure 4.7: $\alpha_{\rm MIR}$-$\alpha_{\rm NIR}$ spectral slope space covered by the different torus models included in AGNFITTER-RX with different complexity levels. The gray circles show estimates of the spectral slope covered by our sample of AGN, computed with *Spitzer* and WISE observed photometry.

overall distribution of fits, while the simplified version fails to capture important features of the SEDs. The set of templates in the simplified NK08 model does not exhibit the characteristic silicate 10 μm feature in absorption, only as an emission line, which changes its strength with the inclination angle of the torus. Meanwhile, the simplified version of the two-phase model SKIRTOR presents a transition between emission and absorption of the 10 μm line allowing to fit type 1, type 2 and intermediate AGNs such as those found in our sample. It raises the possibility of exploring another procedure for averaging NK08 SEDs in the future that yields a wide variety of silicate 10 μm features, likely the main driver of $\alpha_{\rm MIR}$.

The inclusion of the polar wind component improves the quality of the fit, as can be seen in the central upper panel with the CAT3D model. The likelihood distribution is one of the narrowest and the one with the highest likelihood values. 25 out of 36 active galaxies

achieved the maximum likelihood with CAT3D as the torus model in the fitting. These results suggest that using CAT3D with 3 parameters (i.e., incl, a, fwd) to model the torus gives the best result on average if we have a large enough number of photometric bands, and SKIRTOR with 1 parameter (i.e., incl) otherwise. Given that our sample comprises predominantly Sy1 AGNs ($\sim$ 83%), this finding is in agreement with that reported by García-Bernete et al. (2022); González-Martín et al. (2019) where the clumpy disk + wind model produces the best fit for Sy1. Favorable physical conditions and viewing angle of type 1 AGNs favor them for detecting IR dusty polar outflows.

Similarly, recent SED studies of red quasars at intermediate redshifts (Calistro Rivera et al. 2021) suggest that the presence of dusty winds could potentially explain the IR excess at $2-5$ μm in quasar SEDs. Specifically, the NIR bump can be explained by the presence of hot graphite dust heated by the AGN (García-Bernete et al. 2017), which reaches temperatures of up to 1900 K and thus produces emission peaking in the NIR. However, alternative scenarios such as a contribution from direct emission from the accretion disk (Hernán-Caballero et al. 2016; González-Martín et al. 2019), host galaxy or a compact disk of host dust (Hönig et al. 2013; Tristram et al. 2014) have been proposed.

In the right panel of Fig. 4.6, we plot the combined residuals estimated as the difference between the observed and the best-fit model photometry as a function of frequency for the complete sample. The purpose of this exercise is to search for general structure and trends for the different hot dust models, which could inform us of their advantages and limitations in modeling specific frequencies of the IR regime. Around 10 μm ($\log \nu\mathrm{Hz} \sim 13.5$), NK08 and CAT3D present negative residuals, which means that most of the time, templates with emission features were fitted while the data suggest absorption features or no features, that is, the predicted fluxes are consistently overestimated by the models. This limitation may have effects on the uncertainties in the estimated inclination angle.

Moving towards shorter wavelengths, a positive bump is observed which implies an underestimation of fluxes between approximately 1.5 μm and 5 μm ($\log \frac{\nu}{\mathrm{Hz}} \sim 14$). In S04, the bump has the highest amplitude, meaning the highest overestimation, while in

CAT3D and NK08, with 3 parameters, the lowest ones. The behavior is consistent with the likelihood distributions and, due to the high amount of photometric bands that lie in this region, this bump largely determines the quality of the fit. This infrared excess, which is not recovered by models, has been reported previously in quasars and Seyfert galaxies (Mor et al. 2009; Burtscher et al. 2015; Temple et al. 2021a), and according to some studies it could be attributed to hot dusty winds not considered in most torus models (Calistro Rivera et al. 2021; Hönig 2019). This bump can potentially affect the properties of the host galaxies, making them brighter and redder than they would otherwise be.

Alternative analyses, such as the Akaike Information Criterion (AIC), founded on information theory, balance between fit quality and the complexity of the model by including an over-fitting penalisation term. AIC measures the loss of information in our understanding of the process that generates the data. The criteria are useful in the current comparison to assess if fitting simplified or complex torus models has a significant statistical effect. The model with the least loss of information, i.e., the lowest AIC, is favored. It is defined by the number of model parameters (k) and the maximum likelihood ($\mathcal{L}$) as follows:

$$\mathrm{AIC} = -2\ln\mathcal{L} - 2\mathrm{k} \tag{4.9}$$

Table 4.3 presents the median of the distribution of maximum likelihoods for our sample of AGNs, its corresponding median logarithmic evidence, the median logarithmic posterior odds and the median AIC criteria. As observed in likelihood histograms of the realizations, the CAT3D model is the most favorable according to its high likelihood, high evidence and low AIC values. Since the evidence corresponds to a likelihood function integrated over all parameter space, it quantifies how well the high diversity of CAT3D templates in general explains the observed data. As the evidence increases, the posterior probability decreases, which means that, within the enormous possible combination of parameters, the specific producing the maximum likelihood is probably not the only one. The previous might be a consequence of potential model degeneracy, which results in a slightly lower posterior probability (46.54%) for CAT3D compared to the highest (50.07%)

value.

Table 4.3: Results of the torus model comparison

Model	ln $\mathcal{L}$	ln Z	log P(M|D)	AIC
S04	-48.63	-96.65	50.07 %	257.94
NK08_1p	-76.93	-124.16	48.82%	388.26
SKIRTOR_1p	-47.9	-96.76	47.39%	254.6
NK08_3p	-37.86	-87.82	49.86%	212.35
SKIRTOR_3p	-50.88	-100.24	47.79%	272.3
CAT3D_3p	-34.38	-80.83	46.54%	196.34

Notes: Results of the torus model comparison based on the median values of different criteria: the maximum logarithm of the likelihood, the logarithm of the evidence, the posterior probability, and the Akaike information criterion (AIC), computed by assuming all models have the same prior probability.

S04 presents the maximum mean posterior odds with $\sim 4\%$ higher probability than CAT3D, despite the lower likelihood values. Fig. 4.7 shows how the slope space covered by S04 is limited to a well-defined track, providing a limited variety of templates. Therefore, the S04 template producing the maximum likelihood, although it does not produce a high-quality fit, is the one that does the best within the set of possible S04 templates, meaning, its combination of parameters is the most likely.

In all the cases the Bayes factor $B_{ij} = Z_i/Z_j$ give a factor of more than 10^2 with strong evidence of the CAT3D model ($\log Z_i = -80.83$) concerning the other models ($\log Z_j$) according to the Jeffreys (1961) scale. For the one-parameter torus models the analysis is not conclusive since SKIRTOR presents a slightly higher likelihood and AIC, but lower evidence and posterior than S04. However, the Bayes factor $B_{SKIRTOR,S04} = 10^{-0.11} \approx 0.78$ means a weak evidence supporting S04. Finally, it is worth highlighting that the increase in the parameter number in NK08 and SKIRTOR seems to have a subdominant effect on AIC compared to the likelihood improvement.

4.3.2 Comparing accretion-disk models

Analogously to the previous comparison, we started by studying the likelihood distributions in the left panel of Fig. 4.8. In this case, the semi-empirical models R06 and THB21 reached higher likelihood values and narrower distributions than the theoretical models SN12 and KD18. THB21 produces the best-fitting model and is the unique one that considers emission lines in the spectrum. This result is informative, as it shows that high-resolution spectral features of the broad AGN emission lines can have a significant impact on the photometry (e.g., Schaerer & de Barros 2009; Shim et al. 2011; Marshall et al. 2022).

In fact, the mean value of the likelihood distribution differs by a factor of $\approx 10^{10}$ from that of R06, even though both models were calibrated following a similar methodology with quasars, and the redshift range of the data sets is the same. The limitation of these semi-empirical models, in comparison to the theoretical models discussed below, is that they do not provide any estimates of the physical properties of the black hole and accretion disk beyond SED luminosity or reddening.

Conversely, theoretical models have more difficulty reproducing the shape of SEDs, and this may be due to the limitations we still have in fully understanding the emission from accretion disks. Despite the lower likelihood values, it would be interesting to study whether or not these models provide acceptable estimates, for example, of the black hole masses (see Section 4.3.6). In this way, we would be able to decipher in which cases it is convenient for the user to select one type of model or another, given the scientific question they aim to answer.

Following the analysis presented in Section 4.3.1, we plot in Fig. 4.9 the spectral slope space covered by our accretion disk models given by $\alpha = -\log(F_\nu(\lambda_2)/F_\nu(\lambda_1))/\log(\lambda_2/\lambda_1)$ with $\alpha_{\rm ox}$ and $\alpha_{\rm B\text{-}V}$ corresponding to $[\lambda_2, \lambda_1]$ equal to $[2500, 8.21]$ Å and $[4420, 5400]$ Å; along with the rest-frame slopes of our sample based on GALEX NUV, *Swift* B and V bands, and *XMM-Newton* band 3 photometry. The plot shows how the SN12 model using the flexible $\alpha_{\rm ox} - L_{2500\text{Å}}$ nicely overlaps our sources with both low $\alpha_{\rm ox}$ and low $\alpha_{\rm B\text{-}V}$,

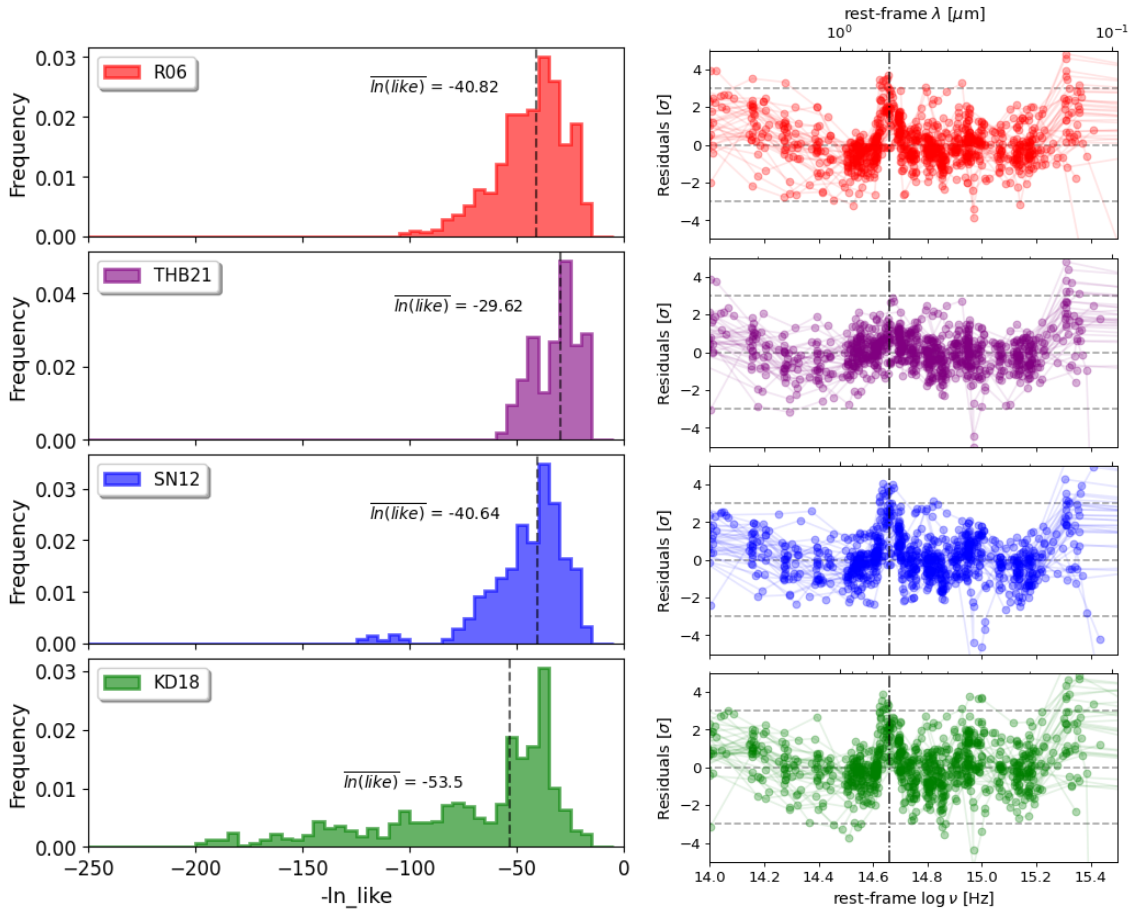

Figure 4.8: Performance analysis of different accretion-disk models in SED fitting: The models shown are the semi-empirical R06 (upper panel), the semi-empirical with emission lines THB21 (central upper panel), the theoretical α-disk SN12 (central lower panel), and the theoretical three-component KD18 model (lower panel). The left panel shows the histograms of the logarithm of the likelihood values from each model and the dashed line in each plot indicates the median value of the distribution. Residuals from the best fits for each active galaxy of the sample are shown in the right panel. The effect of overlapped emission lines, such as the doublet of [N II] $\lambda\lambda$ 6549,6585 Å and H α λ 6563 Å, is large, reducing residuals to about zero. The dash-dotted line indicates the 0.65 μm broad emission lines feature and the dashed lines the limits of 0σ, 3σ, and -3σ.

while KD18 recovers some sources with intermediaste values of $\alpha_{\text{B-V}}$. For our sample in particular, the parameter space covered by KD18 with BH mass and accretion rate as free

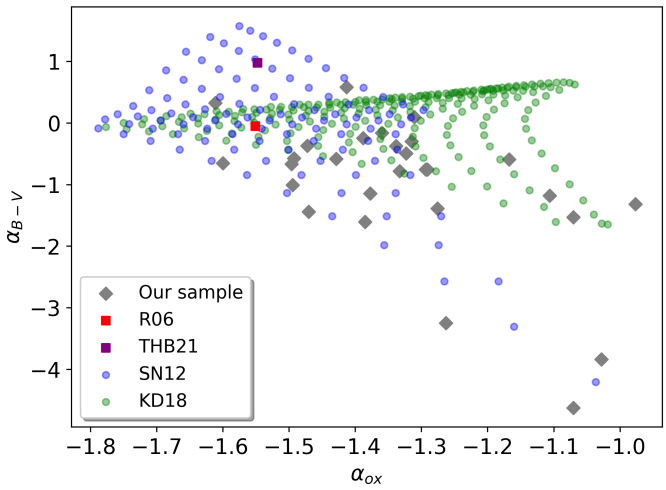

Figure 4.9: α_{ox}-$\alpha_{B\text{-}V}$ spectral slope space covered by the different accretion disk models included in AGNFITTER-RX. The gray circles show estimates of the spectral slope covered by our sample of AGN, computed with *Swift*, GALEX and *XMM-Newton* observed photometry.

parameters is not enough to mimic the spectral emission of some sources. The almost half of the sources outside the KD18 parameter space in Fig.4.9, account for the extended low probability tail of the probability distribution in Fig.4.8. The lack of an X-ray obscuration prescription may be playing a role here, since absorption produces changes in the observed spectral slope, while the KD18 model only accounts for intrinsic photon index changes driven by the accretion rate.

The residuals in the right panel of Fig. 4.8 show a pronounced and narrow peak around $\lambda \approx 0.7$ μm in all the models except THB21. This underestimation of the flux appears to be caused by the overlap of several emission lines such as the doublet of [N II] $\lambda\lambda$ 6549,6585 Å and Hα λ 6563 Å, as suggested in previous works (e.g., Schaerer & de Barros 2009; Shim et al. 2011). In support of this, the errors in the THB21 model, which accounts for the emission lines, are very close to 0 and less dispersion of the points is observed, which is evidence of a systematic good fit. Besides, the FUV region contains a bump in all the models resulting from poor modeling of the soft excess as a consequence of the little good data in this region.

For a proper model comparison, we calculated the posterior odds (equation 4.8) and the AIC criteria (equation 4.9) as in the torus model comparison. Table 4.4 summarizes the median distribution of maximum likelihoods for our sample of AGNs, its corresponding median logarithmic evidence, the median logarithmic posterior odds and the median AIC criteria. In this case, all the evaluated criteria agree with the likelihood distribution of the realizations (Fig. 4.8) in pointing to THB21 as the preferred accretion disk model. Despite R06 and THB21 being single templates without fitting parameters other than the normalization, reddening and the same X-ray $\alpha_{\rm ox}$ recipe; the effect of the broad emission lines on the likelihood and evidence is significant. The Bayes factor of $\approx 10^{5.1}$ also represents strong evidence of THB21 when compared with R06.

The theoretical SN12 model has similar results as R06, with just a smaller posterior, but the Bayes factor $10^{0.24} \approx 1.74$ suggests very strong evidence over R06. The AIC values show that for a similar fit quality, the increase of 1 fitting parameter of SN12 compared to R06 is not critical. Although the complexity of the KD18, both likelihood and evidence suggest templates are not diverse enough to produce high-quality fits, in particular at the X-rays regime. The inclusion of a third free parameter, such as the spin, the spectral index of the warm Comptonisation component or the dissipated luminosity of the corona, can improve the performance.

Table 4.4: Results of the accretion disk model comparison

Model	$\log \mathcal{L}$	$\log Z$	log P(M|D)	AIC
R06	-34.70	-81.06	48.27 %	197.81
THB21	-26.14	-75.96	48.73 %	158.4
SN12	-34.21	-80.82	45.62 %	197.54
KD18	-48.99	-100.49	47.66 %	265.61

Notes: Results of the accretion disk model comparison based on the median values of different criteria: the maximum logarithm of the likelihood, the logarithm of the evidence, the posterior probability and the Akaike information criterion (AIC), computed by assuming all models have the same prior probability.

4.3.3 Comparing model combinations

The best-fit results are produced using combinations of CAT3D and THB21 models for 67% of the sources. When analyzing type 1 and 2 AGNs independently, this success rate translates into 71% and 40%, respectively. However, there is not a sufficiently large sample of type 2 AGNs to obtain conclusive results. The 67% reaches a value of 81% when also considering combinations of CAT3D with other disk models such as R06, SN12 or KD18. The number of parameters required for these best fits is 19, consisting of five parameters associated with stellar models, three with cold-dust, four with the torus, two with accretion disk, two with X-rays and three with radio AGN models. This is not a problem for our data set as we have 49 photometric bands with a minimum of 40 valid data points, allowing convergence of the live points exploring the parameter space. However, if poorer sampling of the SEDs is available, then the simplified versions of the models might be required.

Current and upcoming surveys, such as JWST and LSST, are poised to effectively sample the critical frequency ranges found in the residuals for the different emission models. The short and long wavelength channels of NIRCam on JWST are particularly well-suited for in-depth studies of the overlapping doublet of [N II] and H α emission lines in AGNs up to $z \sim 6.5$, as well as the torus NIR excess ($3 - 4$ μm restframe) in local AGNs. Additionally, MIRI imaging filters on JWST are sensitive to the crucial MIR torus peak and the silicate line in low-z AGNs, and they also capture the NIR excess in AGNs at wide redshift ranging up to $z \sim 6$. The r, i, z, and y bands of LSST will enable the detection of the emission lines feature for AGNs at $z < 0.5$. These capabilities highlight the significant potential of current and near-future surveys to probe our torus and accretion disk models in the high redshift Universe.

When analyzing the morphology of the SEDs from Figs. 4.4 and 4.5, we can see how blazars present a highly decreasing spectrum in X-rays with Γ values around -2.69 and a high-energy emission of the accretion disk and the radio jet. Seyfert 1 AGNs present a high-energy emission of the disk, too, but a lower-energy radio emission and a less steep X-ray power law ($\Gamma \approx -2$). On the contrary, in Seyfert 1.5 and 2 we observe a nuclear hot

dust luminosity comparable to or slightly higher than that of the disk. An important aspect we found evidence for, is how the X-ray fluxes help to break the degeneracy between the galaxy and AGN contribution to the UV, as can be seen for example in the SED fit of Ark 564. It can also be noted that despite the lack of information in the FIR in some cases such as Mrk 290 and Ark 564, it is possible to roughly estimate the emission of cold dust using the energy balance prior to the stellar emission.

4.3.4 Effects of photometric uncertainties

The spectrum-like quality of this SED catalog results in extremely small photometric uncertainties as can be seen in Figs. 4.4 and 4.5. Despite point spread function and point source coincidence loss corrections were applied to observed photometry used to create the synthetic SEDs in Brown et al. (2019), no additional photometric calibration uncertainties were added to the final SEDs. A flux-to-error value matching the signal-to-noise ratio of the spectra, calibration uncertainties given by the instruments user handbook and/or fixed percentage errors can be incorporated to account for systematics in the models and instrumental effects (e.g.; Leja et al. 2019; Doore et al. 2021).

Including systematic effects can mitigate the significant under-fitting in some bands, even for the best-fitting models. Furthermore, the much larger statistical weight of optical-radio data in the SEDs may dilute the contribution of the X-rays to the fit. To assess the potential consequences of the previous, we modified the X-ray uncertainties to equally weight the less energetic band by following the Zou et al. (2022) prescription and we performed again the SED fittings. Although the change in the likelihood values was expected due to its dependence on the photometric errors, no significant differences were found in the inferred physical parameters or the contribution of the AGN and host galaxy emission components. Then, the potential main drivers of the underfitting of our sample are systematic effects and AGN variability.

4.3.5 Current advances and limitations on the modeling of the torus

Previously, we discussed the clear bump we found in the residuals at $1.5 - 5$ μm and how the CAT3D model reduces most of that feature, meaning that dusty winds might be a required element for $\sim$ 70% of the sources. Beyond the quality of the fits achieved by the different models, we also wanted to evaluate the reliability and consistency of the inferred physical parameters. As a first test, we compare the general trend of the inferred inclination angle parameter found by fitting different torus models to the SEDs of our sources with spectroscopic classifications (Véron-Cetty & Véron 2010). For simplicity, we group Blazars, Sy1, Sy1.2, Sy1.5 and Sy1n as Seyfert 1; and Sy1i, Sy2, Sy1.9 and Sy3 as Seyfert 2.

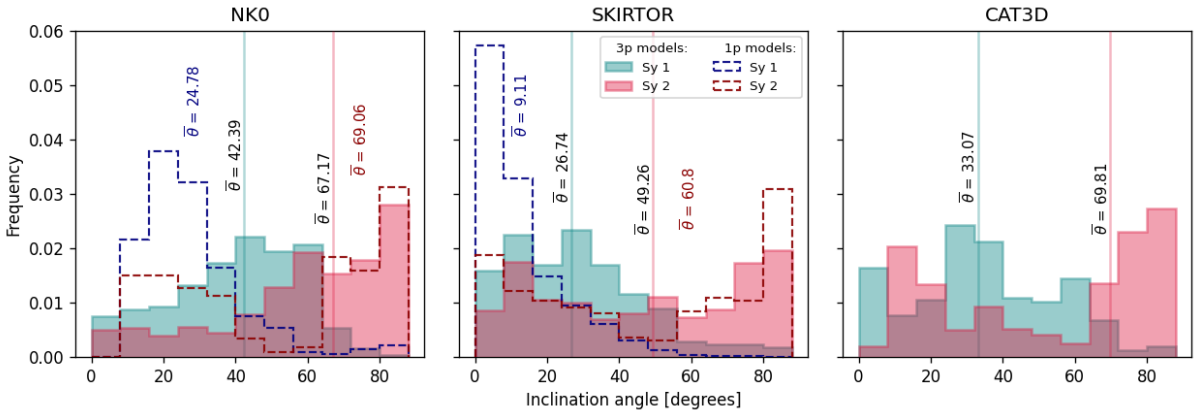

Figure 4.10: Normalized histogram of the total distribution of inclination angle estimates by the NK08 (left panel), SKIRTOR (central panel) and CAT3D (right panel) model for the 36 galaxies of the sample classified as Seyfert 1 (light blue) and Seyfert 2 (magenta). The dotted contours represent the distribution of the angles for the one-parameter model, while the solid histograms represent the three-parameter models. The solid vertical lines highlight the median values of the distributions.

In Fig. 4.10, the overall distributions of the inclination angle are shown for the 31 AGNs in our sample, classified as Seyfert 1, and the five as Seyfert 2. For this, we evaluate the performance of the CAT3D models and the simplified and complex versions of NK08 and

SKIRTOR. For all the three-parameter models, there is a difference of more than 22° in the median values of Sy1 and Sy2 distributions, which means that, in principle, we can use the inclination angle as a classifier of AGNs. As the distributions overlap, some Sy1 with large inclination angles and a few Sy2 with small angles, this classification would have to be done with caution. The Sy2 distributions for three-parameter models look quite wide, probably due to the low number of AGNs. In particular, this distribution looks almost flat for the SKIRTOR model. Furthermore, it can be noted that SKIRTOR tends to find smaller angles for both Sy1 and Sy2 compared to the other models.

When using the one-parameter version of the NK08 and SKIRTOR models, the estimates of the inclination angle for Sy1 present narrower distributions centered at significantly smaller angles than the three-parameter models. This shows how the simplified version of SKIRTOR manages to find the most physically consistent angles for type 1 AGNs, perhaps by avoiding parameter degeneracies that may be occurring using the more complex version of this torus model. Instead, for the AGN type 2 population, the picture does not improve, as simplified versions of the models result in distribution histograms with very pronounced bimodalities with extreme peaks at around 10 and 80 degrees. Hence, the choice of the most suitable torus model to use depends on the properties of the AGNs as nuclear hydrogen column density, in accordance with what was reported in García-Bernete et al. (2022).

We then analyzed the self-consistency of models from a geometrical point of view, similar to a comparison done by Cerqueira-Campos et al. (2023). As proposed by the unified model (Antonucci & Miller 1985), the key spectral features of the diverse population of AGNs could be explained by the interaction of the light emitted by the active nuclei with the distribution of hot nuclear dust. Thus, the angle between the observer's line of sight and the axis of the torus would determine the classification, being Type I and Type II AGNs the two endpoints of unobscured and highly obscured spectra. Comparing the inclination angle with the opening angle of the torus allows us to see whether indeed we reconstruct an edge-on or face-on orientation.

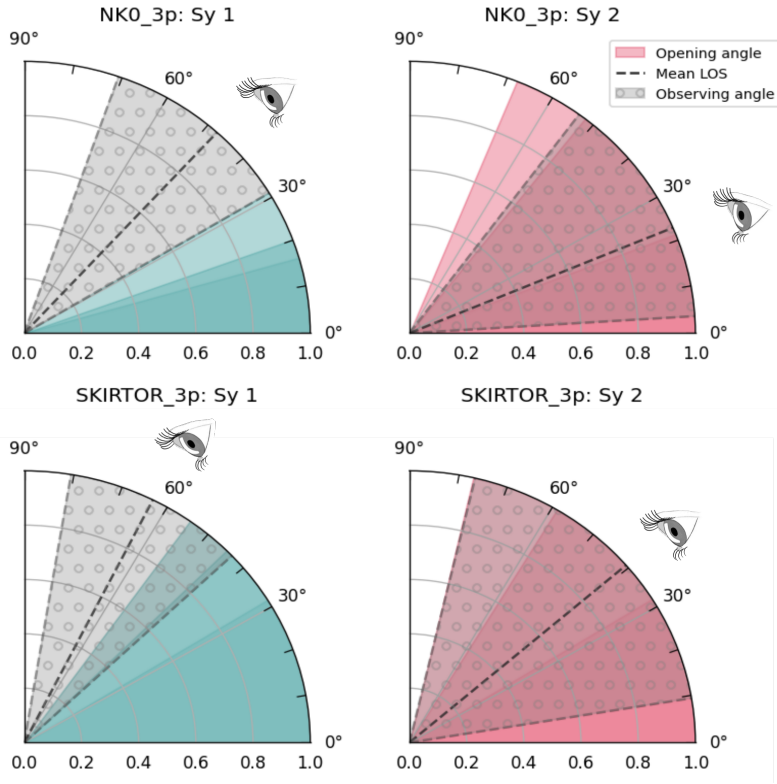

Figure 4.11: Representation of the upper right quadrant of the cross-section of the toroidal distribution of hot nuclear dust in NK08 (upper panels) and SKIRTOR (lower panels) models. The shaded color areas with decreasing transparency represent the 16th, 50th and 84th percentiles of the torus opening angle in Seyfert 1 (blue) and Seyfert 2 (magenta) AGNs. While the black dotted lines represent the median values of the viewing angle and the dotted gray shaded areas are the 16th and 84th percentiles.

As shown in Fig. 4.11, the lines of sight in Sy1 AGNs, estimated by the inclination angle parameter, are outside the toroidal distribution of hot dust for both the NK08 and SKIRTOR models. Even the 16th and 84th percentiles in the distribution of viewing angles do not intercept or have little overlap with dust distribution given by the torus opening angle. Conversely, Sy2 AGNs present distributions of the line of sight that completely intercept the torus for both models. It is outstanding how the SED fitting recovers torus aperture angles and uncertainties in the estimation of the observing angle that is consistently larger

for Sy2 than for Sy1, which may be potentially associated with lower luminosity AGN (Wada 2012). SKIRTOR is more likely to result in smaller inclinations and larger aperture angles compared to NK08.

Another key physical feature that can be inferred from SED fitting is the estimation of the AGN contribution to the total luminosity in a certain energy range (AGN fraction), such as the IR range from 8 to 1000 μm or the optical range from 400 to 500 nm. In Fig. 4.12 are presented the distributions of IR AGN fractions of each source found by fitting each of the torus models, and highlighted with gray backgrounds are the sources lacking SPIRE and PACS bands, crucial to constrain the emission of the star-forming regions. It can be seen in Fig. 4.12 how the uncertainties in the AGN fraction increase without FIR data. In these cases, AGNfitter-rx can fit high and negligible contributions from cold dust affecting the estimation of the AGN fraction regardless of the choice of torus model. As a consequence, the violin plots of the mentioned galaxies are extended and are out of alignment, with different distributions of AGN fraction values.

Since we used a flexible energy balance prior to running the fits, the attenuated stellar emission sets only a lower limit to the cold dust emission luminosity. Nevertheless, a more restrictive energy balance would have raised probabilities of models with a cold dust luminosity equal to the attenuated stellar component, largely constraining the IR AGN fraction and isolating the effect of the torus models. As previously stated, there are special cases in which the energy balance is broken so the decision of whether to include it or not, as well as the severity of the prior, is left to the user. For the active galaxies with available IR data, the change of hot dust model doesn't affect the AGN IR luminosity.

4.3.6 Current advances and limitations on accretion disk models

An important decision when choosing the accretion disk model for the SED fitting is whether to opt for semi-empirical or theoretical models. While the semi-empirical models have likelihood values potentially better than the latter (as THB21 in Figure 4.8), theoretical models can provide estimates of important physical parameters such as the black hole mass

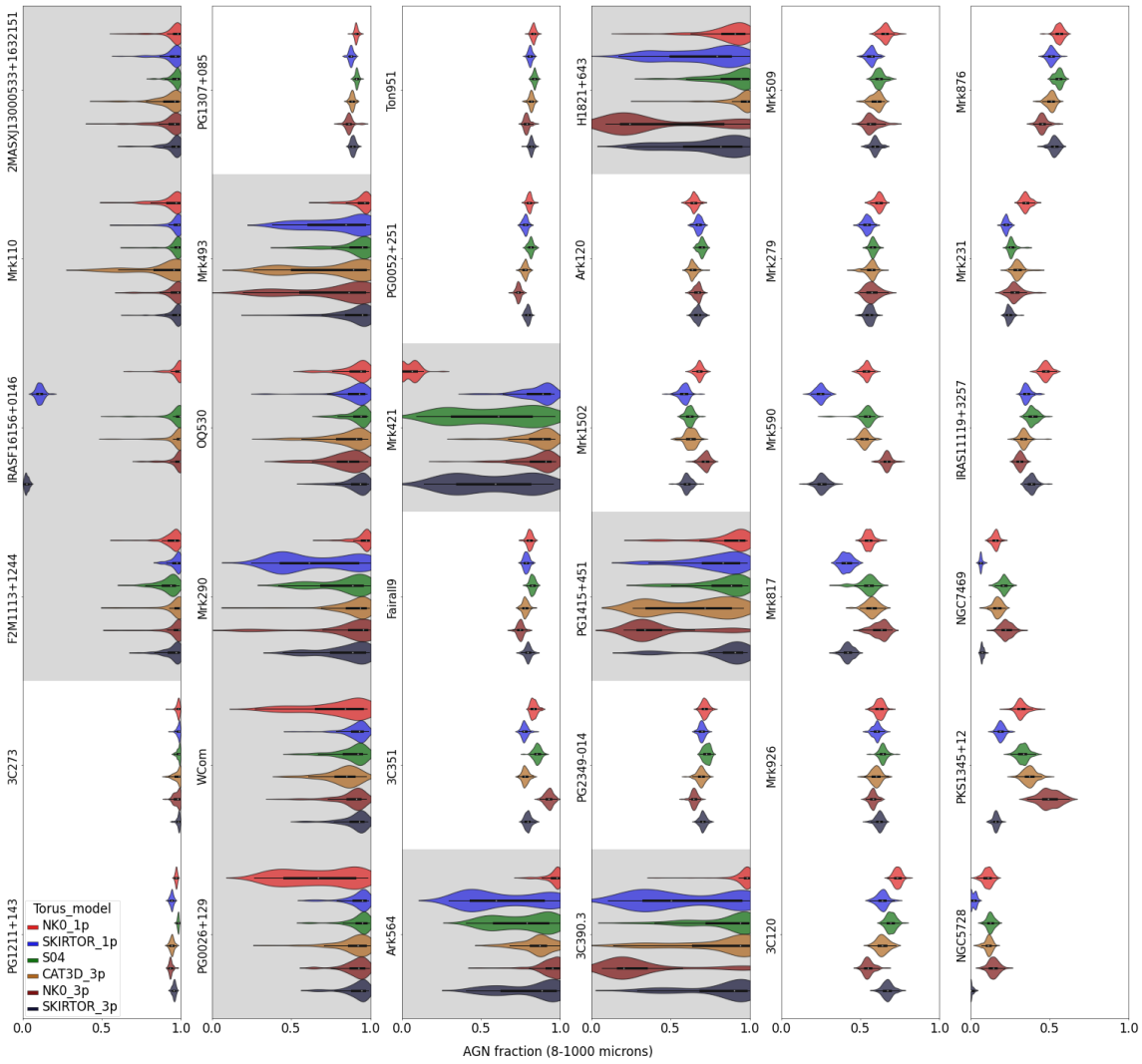

Figure 4.12: AGN fraction contributing to the total IR luminosity ($8-1000\,\mu$m) for the 36 galaxies of the sample and each torus model tested. The gray area highlights the SEDs with missing data in the FIR. The different models used do not have a large effect on the overall AGN fraction in this range.

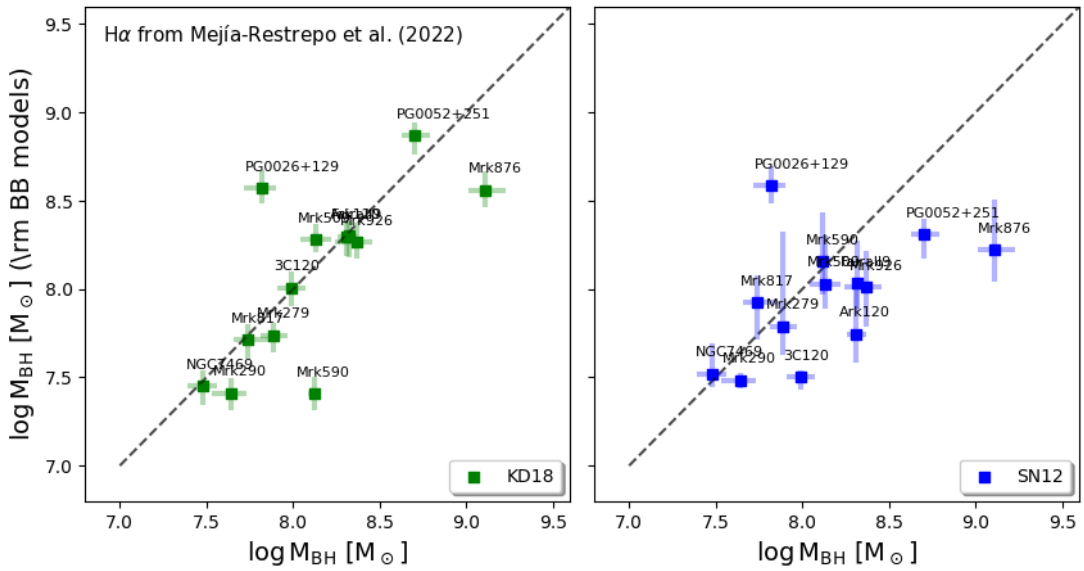

Figure 4.13: Comparison of the log-scale black hole mass estimates made via AGNfitter-rx using KD18 and SN12 accretion disk models (y-axis) and broad Hα emission line (x-axis) for 17 AGNs in our sample with available information in BASS DR2 catalog. The dotted line corresponds to the 1-to-1 relation.

and accretion rate. Using AGNfitter-rx, black hole masses can be estimated from the rest-frame UV region of the observed SED only by applying SN12 or KD18 models. To evaluate the confidence of the values, in Figure 4.13 we compare black hole masses from SED-fitting with virial estimates of black hole masses based on broad emission lines available for 17 AGNs from the BASS DR2 sample (Mejía-Restrepo et al. 2022). Since those spectroscopic estimates are reliable, we evaluate how the SED fitting estimates using KD18 (left) and SN12 (right) deviate from the one-to-one line.

It can be easily noticed how the KD18 model is in good agreement with the values of the BASS DR2 catalog. Out of the 13 sources, only three are outliers, while six are within the one-to-one ratio taking into account uncertainties and four others are off by only a factor of about 0.1 dex. These 13 sources are characterized by X-ray information, which is important for fitting the KD18 model extending up to hard X-rays (100 keV). However, as could be seen in Fig. 4.8, the likelihood histogram for the KD18 model has a peak

around -log(like) ~ -40 but a very spread out distribution reaching values of -200. A poor fit to stacked SEDs in the optical regime and unpredictable X-ray fluxes by up to a factor of 2 were also reported in Mitchell et al. (2023) for the $\log \mathrm{M_{BH}}$ regime between $7.5 - 9\,\mathrm{M_\odot}$. One possible explanation for this is that the continuum optical-to-X-rays templates require an observed SED in a narrow time window that ensures the variability is negligible compared to the photometric errors. As this is not guaranteed, sometimes the KD18 model fits well in the UV while producing high errors in the X-rays. Another alternative could be the X-ray obscuration producing changes in the observed spectral slope which does not correspond to the slope variations driven by the accretion rate in the KD18 templates. Precisely, AGNs with a correct estimate of the black hole mass are those that have a good X-ray fit with errors smaller than 2σ.

In the evaluation of the SN12 model, we included sources available in BASS DR2 even though their SEDs in our catalog did not have X-ray photometric bands, which increased our set from 13 to 17. The use of the α_{ox} relation combined with a power law with the dispersion of the relation ($\Delta\alpha_{\mathrm{ox}}$) and the photon index ($\Gamma$) as free parameters allow great flexibility in fitting X-rays. Therefore, X-ray photometry has a mild impact on the performance of specific accretion disk models. In this case, eight sources are clear outliers, six are successful estimates and three are off of the one-to-one relation by a factor of about 0.1 dex. The clustered sources around $10^8 M_\odot$ on the right side of the plot show how SN12 faces difficulties in estimating black hole masses for AGNs with masses greater than $10^{8.5} M_\odot$.

It is worth noting that mass measurements from emission lines also have their assumptions and errors, as do the accretion disk emission models discussed above. The method assumed the dynamic of the gas in the broad line region is virialized, meaning the motion is produced by the gravitational field of the black hole. As mentioned in Mejía-Restrepo et al. (2022), the existence of winds can introduce biases by changing the observed gas dynamics. Additionally, the geometrical factor assumed to estimate the broad line region size in the empirical relations is a source of systematic errors. The specific value, according

to the chosen geometry or even the inclination angle, if a disk is assumed, increases the uncertainty.

We estimate bolometric luminosities based on the black hole mass ($\log M_{\rm BH}$) and Eddington ratio ($\log \dot{m}$) derived from the theoretical accretion disk models SN12 and KD18. The Eddington luminosity is first computed as:

$$L_{\rm Edd} = 1.26 \times 10^{38} \left(\frac{M_{\rm BH}}{M_\odot} \right) \ {\rm erg\ s^{-1}}, \tag{4.10}$$

which represents the maximum luminosity attainable when radiation pressure balances the gravitational force (e.g., Padovani et al. 2017). Assuming virial equilibrium, and that the bolometric luminosity scales with the Eddington ratio $\lambda_{\rm Edd} = \dot{M}/\dot{M}_{\rm Edd} = \dot{m}$, we express the bolometric luminosity as:

$$L_{\rm bol} = \lambda_{\rm Edd} L_{\rm Edd} = \dot{m} \times 1.26 \times 10^{38} \left(\frac{M_{\rm BH}}{M_\odot} \right) \ {\rm erg\ s^{-1}}. \tag{4.11}$$

In Fig. 4.14 we contrast the bolometric luminosities computed from BH mass and accretion rate via equation 4.11 (BB model product), with the photometry value obtained by adding the integrated luminosities of the AGN-synchrotron, the torus, the accretion disk and the hot corona components (SED-fitting estimation). KD18 matches the AGN components estimates within the errors whereas SN12 systematically overestimates the bolometric luminosities by a factor 8. As our sources have BH masses dominantly in the range $\log {\rm M_{BH}[M_\odot]} = 7.5 - 9$, we are in the parameter space where KD18 templates have demonstrated to be successful in predicting BH mass and L_{2500} across different accretion rates (Mitchell et al. 2023). It can be noticed that: BBB models significantly reddened ($\rm EBV_{BBB} > 0.6$) resulting in larger bolometric luminosities than the SED-fitting estimations, in particular with the SN12 model; and the SED-fitting estimations are robust for both models despite of the individual normalization of the AGN components.

Finally, we again assess the effects of the different models, in this case of the accretion disk, on the estimates of the AGN fraction to the optical/UV emission. In Figure 4.15 are presented the distributions of optical AGN fractions (from 4000 to 5000 Å) found by

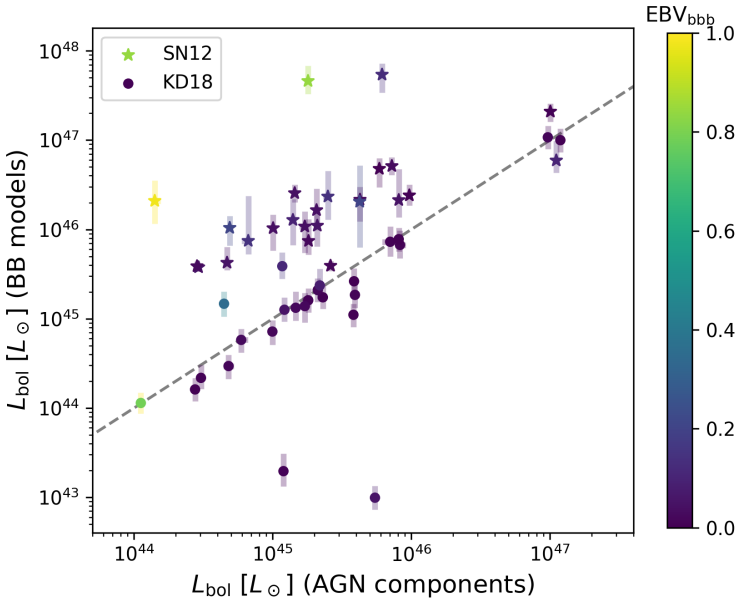

Figure 4.14: Bolometric luminosities are computed by adding integrated luminosities of the AGN components vs bolometric luminosities based on the inferred black hole mass and accretion rate by the accretion disk models. The gray dashed line indicates the one-to-one ratio and the color bar is the accretion disk reddening parameter.

using the R06, THB21, SN12 and KD18 models in the fitting. In most of the AGNs, the violin plots of all the models are aligned and even have a similar distribution width. However, within the observed outliers, the most frequent are distributions given by the theoretical models: SN12 and KD18. As shown in 4.2, opposite to the empirical models, the theoretical templates have a smooth shape in the big blue bump that hardly fits well the data in the most energetic region of the rest frame UV. The previous might be increasing the emission of the stellar population during the fits to reduce residuals and therefore, raising the contribution of the host galaxy. Also, the availability and shape of X-ray emission influence the fit quality due to the high and low flexibility of SN12 and KD18 while fitting X-rays, respectively. As discussed previously Fig.4.9 provides an indicator that the choice of free parameters in KD18 is not sufficient to produce diversity in the templates, especially in X-rays. Further research on the inclusion of relevant parameters such as spin,

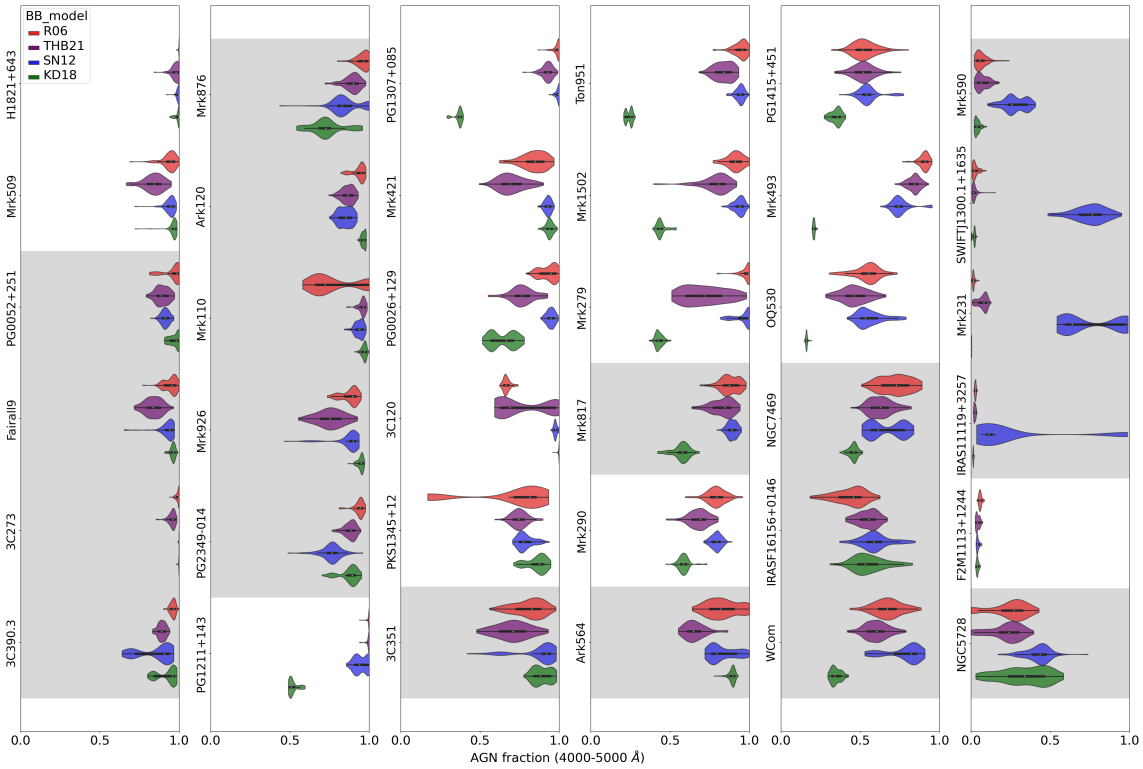

Figure 4.15: AGN fraction contributing to the total optical luminosity (400 – 500nm) for the 36 galaxies of the sample and each accretion disk model tested. The flexibility in modeling X-rays and the effects of UV degeneracy causes outliers more often with the KD18 accretion disk model.

the spectral index of the warm Comptonisation component shaping the soft X-rays or the dissipated luminosity of the corona is needed. On the semi-empirical side, THB21 predicts a systematically larger host galaxy contribution to the SED of type 1 AGNs than R06.

4.3.7 The AGN contribution to the SED from the radio to X-rays

A key feature of SED fitting is the ability to disentangle the contributions of the AGN and the host galaxies to the overall emission. This contribution can significantly vary across wavelengths. Therefore, defining appropriate frequency regions where AGN fractions should be estimated is crucial. To provide information on this we use the flexibility of AGNfitter-rx in delivering any customary range of AGN fraction provided by the user,

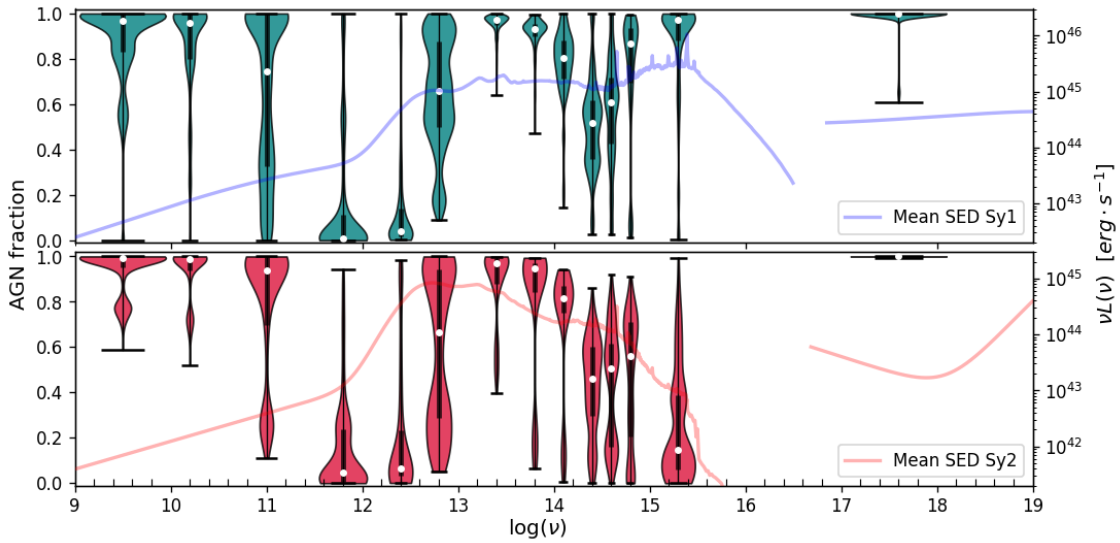

Figure 4.16: AGN fraction distribution as a function of frequency for Sy1 (upper panel) and Sy2 (lower panel) populations. The white dots within the violin plots highlight the median of the distribution while the black line is the range between the 25th and 75th percentiles. The blue and red curves present the average radio-to-X-ray SED, including BC03_metal, S17, CAT3D, THB21, and combinations of SPL and DPL radio models, for type 1 and type 2 AGNs, respectively.

and the unique coverage from the radio to the X-ray to compute a spectral AGN fraction distribution as presented in Fig. 4.16.

The AGN fraction within a frequency range is estimated as the ratio of the luminosity of the AGN model components, divided by the total observed luminosity (AGN + host galaxy), where the luminosities are estimated by integrating the model templates within the given frequency range. The frequency ranges used for the AGN fractions in Fig. 4.16 were chosen qualitatively, based on the photometric coverage in our data set. The width of the frequency ranges chosen in each case is represented by the maximal width of the violins. The violin plots in the upper and lower panel of Fig. 4.16 show the density distributions for the combined Sy1 or Sy2 populations, respectively, and illustrate a general trend of how the AGN contribution to the total emission changes as an function of frequency. The density distributions were obtained for each source by drawing 100 realizations from the posterior

distributions of the Bayesian fit. In all cases, the long and thin tails in the distributions (e.g., the 4, 5, 7, 8, 9th violins from the left) are caused by a few outlier realizations from the posterior distributions that encountered models resulting in an AGN fraction distinct from that of the bulk of the realizations for a given source.

The blue and red curves represent the mean Sy1 and Sy2 SEDs, computed as the average best fit for each population, and are shown in the background for reference with an arbitrary normalization. The emission lines of the silicates and the broad line region can be clearly identified, as well as the big blue bump and a steep X-ray spectral slope characteristic of type 1 AGNs. IR-dominated dusty spectra with a mild UV nuclear emission, and a combination of steep and flat X-ray power laws characterize our average type 2 AGNs. Given the size of the Sy2 set, AGN fractions and the average SED might not be representative of the overall type 2 AGN population.

Both Sy1s and Sy2s exhibit dominant AGN emission in radio, MIR and X-ray bands, while emission in FIR bands appears galaxy-dominated. Indeed, this is in agreement with the most reliable methods for AGN selection, which rely on information from these energy ranges: IR-to-radio luminosity ratios (Condon et al. 1995), radio/FIR correlation (Helou et al. 1985; Donley et al. 2005), radio excess (Yun et al. 2001), radio power versus absolute K-band (infrared) magnitude (Hardcastle et al. 2019), MIR colors (Lacy et al. 2004; Stern et al. 2005; Donley et al. 2012), X-ray spectral slope (Alexander et al. 2005; Nandra & Pounds 1994), X-ray luminosities and X-ray to optical flux ratio (Szokoly et al. 2004; Lehmann et al. 2001), among others. By construction, optical-UV-based selection methods as UV excess are successful in unobscured AGNs (Boyle et al. 1990; Schmidt & Green 1983; Wisotzki et al. 1996; Verbeek et al. 2012), and incomplete for dust-obscured AGNs where the UV-optical is dominated by the host galaxy such as the type 2 ones. Consistently, only Sy1s in Fig. 4.16 have a high AGN fraction in UV bands.

Nonetheless, the extension and width of the violin plots evidence the diversity within the AGN populations and also, the degree of uncertainty in the estimate of the AGN fraction. In Sy1s the long and nearly equally wide violin in the FIR (3rd position from the left) is

explained by the transition from AGN radio to galactic cold dust as the dominant emission component which occurs at different frequencies depending on the dust temperature and AGN properties. A similar phenomenon takes place at the transition from cold dust to torus (6th violin from the left), and from stellar to accretion disk emissions (4th and 5th violin from the right). In the Sy2s, though, the extended UV violin (2nd violin from the right) is a consequence of three of the five AGNs having so reddened accretion disk emission that stellar emission dominates, giving rise to a low AGN fraction, while the remaining sources the contributions are similar or dominated by the disk. We note that we used the observed accretion disk luminosities (the blue curve in Figs. 4.4, 4.5), not corrected for reddening, in our estimation of the AGN fraction. However, if we use the intrinsic (reddening-corrected) luminosity instead, the UV violin for the Sy2 population would be shifted towards higher values and a shape equivalent to the Sy1 population.

4.4 Conclusions

In this part of the thesis, we introduce the new radio-to-X-ray fitting capabilities of the Bayesian Python code AGNfitter (Calistro Rivera et al. 2016), released as AGNfitter-rx, which is designed to consistently fit the radio-to-X-ray SEDs of active galaxies by sampling the parameter space through MCMC or nested sampling. The physical model considers the contributions of the host galaxy, which consist of the emission from stellar populations, cold dust, and cosmic ray electrons in star-forming regions; as well as the AGN components, which include the emission of the hot corona, the accretion disk, the hot dust, and the radio emission from jets and/or shocks. The code now incorporates a variety of semi-empirical and theoretical emission models, as well as informative priors, including the assumption of the energy balance between dust-attenuated stellar emission and the re-emission by cold dust.

To test the new capabilities of the code, we used AGNfitter-rx to fit 36 radio-to-X-ray SEDs of local active galaxies from the AGN SED ATLAS (Brown et al. 2019) at $z < 0.7$.

In this study, we primarily selected this sample of well-characterized and diverse AGNs in order to assess and compare the performance of state-of-the-art torus and accretion disk emission models. We discuss the effects of different models on the fit quality and the most important infered physical properties. Once the best model combination was identifed and implemented, we also evaluate the performance of the code for different AGN populations.

In $\sim 70\%$ of the studied sources, the inclusion of polar winds in torus models produced an advantage in the median of the likelihood distribution of more than three orders of magnitude, a Bayesian factor of higher than 100, and the lowest AIC values compared to other clumpy torus models. Considering that the sample is dominated by Sy1 AGNs, this result suggests that the presence of IR dusty polar outflows is essential in shaping the SEDs of type 1 AGNs. Also by comparing less complex torus template libraries, we find that the inclusion of absorption and emission silicate features likely plays a key role in achieving precise estimates of the inclination angles in most cases.

We analyzed the fitted inclination angle (i.e., the line of sight) and the opening angle parameters of the hot nuclear dusty torus in the Type 1 and 2 AGNs. While the more complex torus models (more than three parameters) revealed overlapping broad inclination angle distributions, their simplified oneparameter versions yielded more physically consistent results with their spectroscopic classification, potentially alleviating parameter degeneracies inherent in complex torus models. However, the geometrical information obtained from the multiple parameters in the more complex torus models revealed lines of sight outside the toroidal hot dust distribution for Sy1 and lines of sight that fully intercept the torus for Sy2 AGNs, which are consistent with the AGN unification model.

In $\approx 67\%$ of cases, the accretion disk models that include emission lines stand out as the optimal choice for the fitting, with a likelihood distribution boasting a median value of $> 10^7$ times higher, a Bayes factor of 10^5, and the lowest AIC values of all the model tested. Emission lines are crucial to avoiding the narrow peak in the residuals at 0.7 μm, which likely stems from the overlap of the doublet of [N II] $\lambda\lambda$ 6549,6585 Å and H α λ 6563 Å. While such empirical templates do not offer estimates for key properties such as

black hole mass or accretion rate, this result underscores the substantial impact of high-resolution spectral features within broad AGN emission on the photometry, which agrees with previous studies.

We conducted a comprehensive comparison between the black hole masses obtained from photometry only via SED fitting and virial estimates based on broad Hα emission lines. The theoretical model that incorporates X-ray emission demonstrates strong agreement with virial estimates, being consistent within the 0.1 dex range around the one-to-one ratio for ten out of 13 sources. Conversely, models that do not account for X-rays encounter difficulties in estimating black hole masses for AGNs exceeding $10^{8.5}$ $M_{\odot}$.

We investigated the frequency dependence of estimated AGN fractions across the radio-to-Xray spectrum, finding that Sy1 and Sy2 exhibit dominant AGN emission in radio, MIR, and X-ray bands, in agreement with empirical AGN selection strategies. From our probabilistic analysis, we conclude that significant uncertainties in disentangling the AGN and galaxy emission in the IR and optical-UV can arise from the absence of FIR data, degeneracies between the accretion disk and host-galaxy UV emission, and the variable flexibility of models when fitting X-ray data.

The radio-to-X-ray SED fitting capabilities provided by AGNfitter-rx are a powerful tool, allowing comprehensive studies of AGN physics. In addition to offering an unsurpassed wavelength coverage, AGNfitter-rx allows the user to easily incorporate customary telescope filters, state-of-the-art theoretical and empirical model templates, and informative priors. AGNfitter-rx is thus a promising and reliable tool for pushing boundaries at the interface between the observational and theoretical scientific communities, allowing them to test, compare, and customize models for the increasing data sets of deeper and larger panchromatic surveys.

In particular, AGNfitter-rx was developed to capitalize on the present and future availability of exquisite radio and X-ray data sets, such as wide and deep radio surveys from the Low-Frequency Array (LOFAR; van Haarlem et al. 2013) and the Square Kilometre Array (SKA; McMullin et al. 2020), as well as new observational parameter spaces covered

by X-ray observatories (e.g., eROSITA, Athena). Crucially, the focus of this investigation is the implementation and testing of IR AGN models of diverse complexity. While these complex models could only be explored in the nearby Universe until recently, the remarkable sensitivity, resolution, and wavelength coverage provided by the JWST will now allow us to test and use these models at intermediate and high redshifts. The rich library of IR AGN and host-galaxy models in AGNfitter-rx is therefore ideal for modeling JWST/MIRI, JWST/NIRCam, HST, and Euclid photometry in order to investigate the hot dust emission, physics, and geometry of AGN obscuration at high-z, including polar dust distributions, and the disentanglement of AGN emission from star formation.

The AGNfitter code (Calistro Rivera et al. 2016), including the AGNfitter-rx code release, are publicly available as open-source Python 3 code at GitHub [8].

[8] `https://github.com/GabrielaCR/AGNfitter`

Chapter 5

Summary and future work

In this thesis, we have investigated the accretion of matter onto SMBHs from two complementary perspectives: a high-redshift approach, focused on quasar discovery during the reionization era, and a low-redshift perspective, aimed at a detailed characterization of the physical processes governing AGN.

At $z > 6$, both instrumental limitations and the scarcity of the quasar population make it challenging to fully characterize the early SMBH population and its impact on galaxy evolution. While it is well established that quasars cannot be the primary drivers of cosmic reionization, their role in the ionizing budget and early structure formation remains an active area of study. Current statistics remain limited: only $\gtrsim 670$ quasars are known in the Era of Reionization ($z \gtrsim 5.3$), representing roughly two orders of magnitude more than the available demographics a decade ago, yet still providing sparse sampling for robust demographic and evolutionary studies. However, forthcoming large-scale surveys, particularly in the southern hemisphere, such as LSST, and 4MOST will provide unprecedented spectroscopic capabilities in the optical regime as well as high-quality time-domain data, enabling transformative advances in quasar discovery.

Within this context, we have leveraged deep photometric surveys to develop new discovery strategies, applying self-supervised machine learning techniques directly to imaging data. This approach offers more flexible selection criteria than traditional methods,

improves our understanding of contaminant populations, and paves the way for broader applications in data mining. In parallel, we explored rest-frame IR variability as a complementary tool to exploit both archival and future time-series data; and we addressed quasar characterization as the natural extension of discovery, adopting a panchromatic perspective that makes use of high-quality multiwavelength data from local AGN, spanning radio to X-rays.

The main findings of this thesis are as follows:

- Self-supervised machine learning applied directly to imaging data, when combined with flexible preselection criteria, offers a highly effective strategy for quasar discovery. This approach has already demonstrated results, yielding 16 new quasars at $z > 6$ with a success rate of 45%, and revealing underrepresented populations with high Ly α + N v equivalent width, narrow Ly α emission, and red NIR continua.

- Beyond quasar discovery, self-supervised methods proved effective for large-scale exploration of astronomical datasets, revealing not only quasars but also diverse phenomena such as asteroids, satellites, transient events, binary systems, and catalog artifacts that are otherwise challenging to identify.

- MIR variability is detectable in quasars at the reionization epoch, demonstrating its potential as an additional diagnostic of AGN activity. However, the limited depth and cadence of unWISE data, together with the anticorrelation between variability amplitude and brightness, restrict detections to the brightest quasars with lowest-amplitude variability, making MIR variability an inefficient standalone selection method at present.

- Only a minority of $z > 5.3$ quasars exhibit detectable IR variability in unWISE light curves. Among the 19 candidates flagged as variable, 8 show red-noise power spectral densities consistent with expectations from lower-redshift quasars, while the remainder are dominated by white noise or quasi-periodic features, likely reflecting the observational limitations.

- We developed the Bayesian SED-fitting code AGNFITTER-RX, spanning radio to X-rays, and demonstrated that it can be a powerful tool for both model comparison and AGN physical characterization. It recovers parameters consistent with high-quality spectroscopic constraints, including torus inclination and opening angles, and yielded black hole mass estimates in agreement with virial measurements from H α.

- Fitting the SEDs of 36 AGNs at $z \lesssim 0.7$ with high-quality multiwavelength photometry revealed clear improvements when adopting clumpy torus models with polar winds and accretion disk templates that include emission-line features. These models provided better fits for 67% of the sample, reduced the spurious NIR bump at 1.5–5 μm, and highlighted the importance of high-resolution spectral features that are otherwise underestimated.

Building on the methods and results presented in this thesis, several avenues stand out as particularly promising for advancing quasar discovery and characterization in the coming years:

- Multi-modal inputs for contrastive learning: Incorporating multi-survey imaging or combining different data types (e.g., images plus spectra) into the input tensor could enrich the feature space and improve class separation. This approach, however, comes with challenges, such as handling heterogeneous resolutions and data formats.

- Unbiased exploration of the embedding space: A label-free analysis of the embedded space, guided solely by catalog features such as $z-J$ or $z-W1$ colors, morphology and imaging quality, can enable the selection quasar candidates from more challenging but interesting populations, such as obscured systems, while minizing biases from prior classifications. Complementarily, expanding the set of quasar and contaminant templates for SED-fitting prioritization (e.g., with EAZY-py) could provide a more flexible and comprehensive interpretation.

- Extensive spectroscopic follow-up: A dedicated campaign targeting candidates selected via contrastive learning from Euclid and HSC, followed by detailed characterization, would provide an unbiased set of confirmed quasars. This will allow us to test whether the trends uncovered in this thesis with LS DR10 extend to new quasar populations and improve the reliability of machine-learning selection methods.

- Improved IR variability studies: Current wide-field IR time-domain coverage is largely limited to unWISE, which has sparse cadence. However, facilities such as JWST, SPHEREx, LSST (Y band) and Roman will offer higher-cadence, deep NIR/MIR observations, which combined with careful treatment of instrumental and sampling noise, will allow more robust assessment of intrinsic IR variability and help disentangle stochastic processes from observational artifacts.

- Enhanced SED modeling for high-redshift quasars: Incorporating a general prescription for intergalactic medium absorption into AGNfitter-rx, and expanding the framework to jointly fit photometry and spectroscopy, would enable more accurate modeling of the highest-redshift sources.

In this thesis, we advanced high-redshift quasar discovery and AGN characterization by combining multiwavelength photometry, self-supervised machine learning, variability analysis, and Bayesian SED fitting, resulting in quasar discoveries and robust physical parameter estimates. These approaches demonstrate the power of modern computational tools to exploit large datasets. Looking forward, such tools will be essential for upcoming surveys, enabling the uncovering and characterization of rare, distant quasars and placing them in the broader context of supermassive black hole and early galaxy evolution.

Appendices

Appendix A

Supervised learning with Modulos IA

Here we detail an early exploration (within the scope of this thesis) into the generation of a high-redshift quasar candidate catalog for 4MOST ChANGES. We conducted out selection via supervised learning based on the second data release of the DECam Local Volume Exploration Survey (DELVE DR2). DELVE provides deep optical photometry (median point-source depth of g$\sim$ 24.3, r$\sim$ 23.9, i$\sim$ 23.5 and z$\sim$ 22.8 mag) for $\sim 17,000$ deg^2 of southern sky[1]. It combines public archival data (DES, DECaLS, DeROSITAS) with devoted observations, resulting in about $\sim$ 618 million of objects observed in all bands. We carried out a preselection of candidates based on color cuts followed by the training of supervised machine learning algorithms supplied with photometric catalogs of quasars and contaminants from the literature as labeled data.

To broadly characterize the parameter space of quasars and their contaminants, we generated synthetic colors from semi-empirical quasar SED models by Temple et al. (2021b), which account for redshift, dust reddening, and diverse emission-line properties. For contaminants, we used stellar population templates with different metallicities at ages ranging from 10^5 to 10^{10} years (Bruzual & Charlot 2003), stellar spectral libraries (Pickles 1998a), and low-mass star and brown dwarf spectra (Schmidt et al. 2014; Bochanski

[1] DELVE survey uses DECam on the 4-meter Blanco Telescope at Cerro Tololo Interamerican Observatory in Chile.

et al. 2007; Chiu et al. 2006; Golimowski et al. 2004; Knapp et al. 2004b); and the dust attenuation law by Calzetti et al. (2000) to redden the stellar population emission. Synthetic magnitudes were computed in optical DECam g, r, i, z, NIR VISTA H, J, Ks and Y, and MIR IRAC and WISE bands 1 and 2 filters, enabling us to define color cuts where quasars and brown dwarfs diverge.

We traced evolutionary color tracks for quasars from $z = 4 - 7.2$ (and for low-z quasars at $z = 0 - 1$), varying emission-line strengths, reddening, and template choices. Similarly, we modeled red galaxy contaminants by exploring a wide range of metallicities, ages, star-formation histories, dust parameters, and redshifts up to $z \sim 3$. Our analysis focused on the $r - i$ vs. $i - z$ plane, complemented by the more flexible $r - z$ color cut to exclude outliers. These optical-only criteria are available for all DELVE sources, in contrast to NIR-based cuts, which are limited to bright objects and regions with IR coverage. Importantly, they also provide cleaner selection boundaries, while NIR/MIR colors do not allow straightforward cuts without excluding atypical quasar populations such as red quasars or those with extreme emission lines.

Figure A.1 shows the synthetic $r - i$ vs. $i - z$ colors of quasars, stars, ultracool dwarfs, and galaxies, together with known quasars from Fan et al. (2023). The shaded green regions mark our adopted color cuts for three redshift intervals:

$$(i_{\rm auto} - z_{\rm auto}) < 1.7 \quad \text{and} \quad (r_{\rm auto} - i_{\rm auto}) < 4 \quad \text{and} \quad (i_{\rm auto} - z_{\rm auto}) < (r_{\rm auto} - i_{\rm auto}) \times 0.6$$
$$\text{and} \quad (i_{\rm auto} - z_{\rm auto}) > -(r_{\rm auto} - i_{\rm auto}) \times 0.14 \quad \text{and} \quad -0.5 < (r_{\rm auto} - z_{\rm auto}) < 3.5$$

for $z < 5.7$,

$$(i_{\rm auto} - z_{\rm auto}) < 1.7 \quad \text{and} \quad (i_{\rm auto} - z_{\rm auto}) > (r_{\rm auto} - i_{\rm auto}) \times 0.6 \quad \text{and}$$
$$(i_{\rm auto} - z_{\rm auto}) > -(r_{\rm auto} - i_{\rm auto}) \times 1.8 + 2.8 \quad \text{and} \quad 2 < (r_{\rm auto} - z_{\rm auto}) < 5$$

for $5.7 < z < 6$,

$$(i_{\rm auto} - z_{\rm auto}) > 1.7 \quad \text{and} \quad (r_{\rm auto} - z_{\rm auto}) > 2.5$$

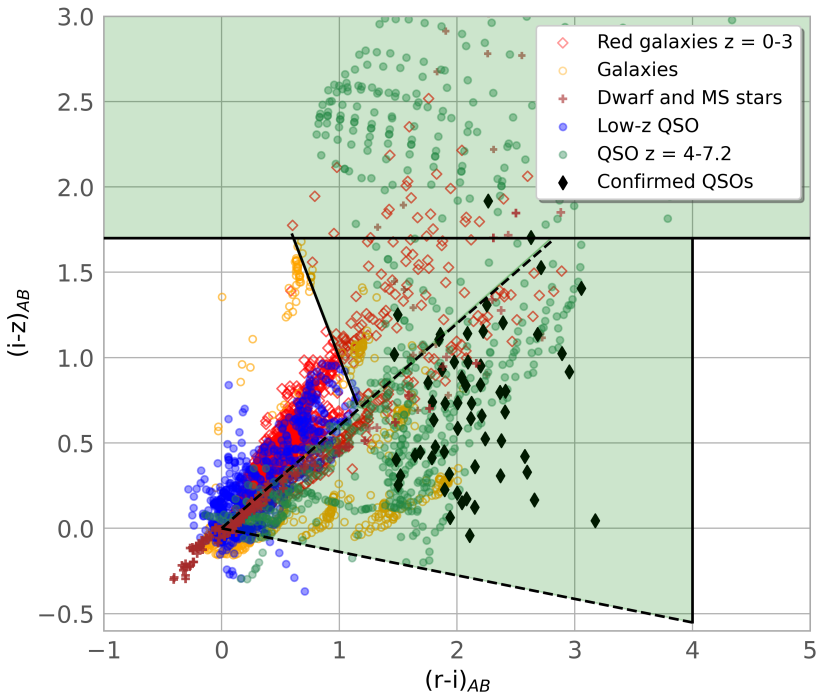

Figure A.1: $r - i$ vs. $i - z$ plane for high-redshift quasars and the main contaminants in color selections. Quasars at $4 < z < 7.2$ are shown as green filled circles, low-redshift quasars ($z < 1$) as blue filled circles, ultracool dwarfs and main-sequence stars as red crosses, mildly dust-reddened galaxies as yellow empty circles, and dusty galaxies up to $z \sim 3$ as red open diamonds. Spectroscopically confirmed quasars at $z > 5.3$ from Fan et al. (2023) are marked with black filled diamonds. Black lines indicate the adopted color cuts designed to select quasars in three regimes: $z < 5.7$, $5.7 < z < 6$, and $z > 6$. These cuts were applied to candidate selection from DELVE DR2.

and $6 < z$.

All magnitudes were corrected for Galactic extinction, and only sources with valid r-band detections ($r_{\rm auto} < 90$) were included in the color cuts. Objects without valid r-band detections were kept without additional constraints. We further required g-band non-detection, valid i- and z-band detections, limits in magnitude aiming to remove bright candidates likely to be dwarfs and to keep candidates that are feasible to confirm with ground-based facilities (18 <mag_auto_i < 24 and 18.5 <mag_auto_z < 24.5), and a non-extended morphology (extended_class_$i, z \neq 3$ and mag_psf_i, z-mag_auto_i, z

$< |0.2|$). Finally, we cross-matched with Gaia DR3 to remove obvious stars showing significant ($> 3\sigma$) proper motion or astrometric excess noise. After these cuts, the catalog contained 76,382,736 sources. Where available, NIR and MIR data from VHS DR5, VIKING DR4, and CatWISE were added within a 1.5″ radius.

In parallel to the above steps, we evaluated the feasibility of training ML algorithms on *Modulos AutoML*[2] using colors, their uncertainties, magnitude variability information and proper motions. Modulos is a platform that allows one to build and implement solutions to diverse supervised machine learning tasks. The platform has a wide range of state-of-the-art models such as convolutional neural networks (CNNs), k-Nearest Neighbors, Logistic Regression, Random Forest, XGBoost, among others; and feature extractors such as Autoencoder, Cropping, T-test Feature Selection, Random Pixel Selection, etc. The output is a set of proposed feature extractor methods, models and a hyperparameter combinations that must be tested and compared to find the best solution.

We evaluated the performance of each solution of ML classification using the area under the receiver operating characteristic curve (ROC AUC), which plots the true positive rate (TPR) against the false positive rate (FPR) at varying thresholds, and by adopting entropy as the optimization criterion. As input features, we provide photometric colors ($r-i$, $i-z$, $J-H$, $z-K_s$, $z-W1$, and $W1-W2$) and their uncertainties; the quadrature sum of the magnitude errors in each band (`wavg_magerr_auto`); the root-mean-square of the automatic magnitudes (`wavg_magrms_auto`); and CATWISE proper motions. The labeled training set consists of 1,942 quasars across three redshift bins, 4,549 ultracool dwarfs, 7,306 low-mass stars, and 8,904 luminous red galaxies (LRG). We deliberately avoid raw magnitudes or fluxes as features to reduce biases toward brighter sources. The labeled dataset was split, as standard, into 80% for training and 20% for validation.

Modulos identified the best-performing solution as a random forest model with an identity transformation feature extractor, characterized by the following hyperparameters: validation score of 0.991, 115 estimators, and a minimum leaf size corresponding to

[2] `https://www.modulos.ai/`

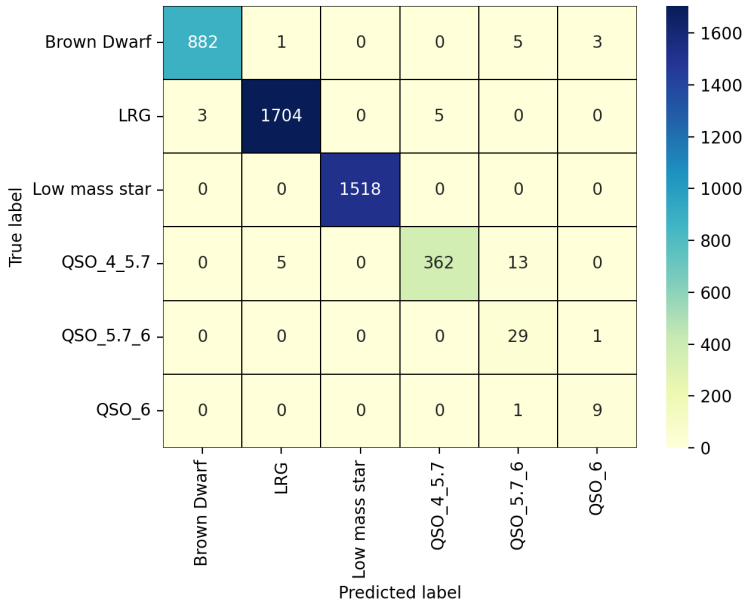

Figure A.2: Confusion matrix for the best-performing classification of DELVE DR2 sources, obtained with a random forest model (validation entropy score ~ 0.991 with identity transformation). The training set includes six classes: quasars in three redshift bins, ultracool dwarfs, low-mass stars, and luminous red galaxies. The method achieves high accuracy overall, with most predictions lying along the diagonal. The algorithm occasionally face troubles to classify quasars at $z > 5.7$ being assigned to the neighboring $5.7 < z < 6$ category.

$\log(\texttt{min\_samples\_leaf}) = -3.515$. The confusion matrix for this solution (Fig. A.2) demonstrates strong label recovery, with the highest values along the diagonal and only minor misclassifications—for instance, a few quasars at $z > 5.7$ predicted in the adjacent $5.7 < z < 6$ bin. The plot also highlights the imbalance in our training set, with LRGs and low-mass stars heavily overrepresented compared to the relatively small number of quasars at $z > 5.7$.

A permutation feature importance analysis (Fig. A.3) identifies the root-mean-square of the g-band magnitude and the $i - z$ color as the most dominant features driving the classification. However, since all sources are non-detections in the g band, the g-band RMS is non-intrinsic and distracts the algorithm. Among meaningful features, the $i - z$

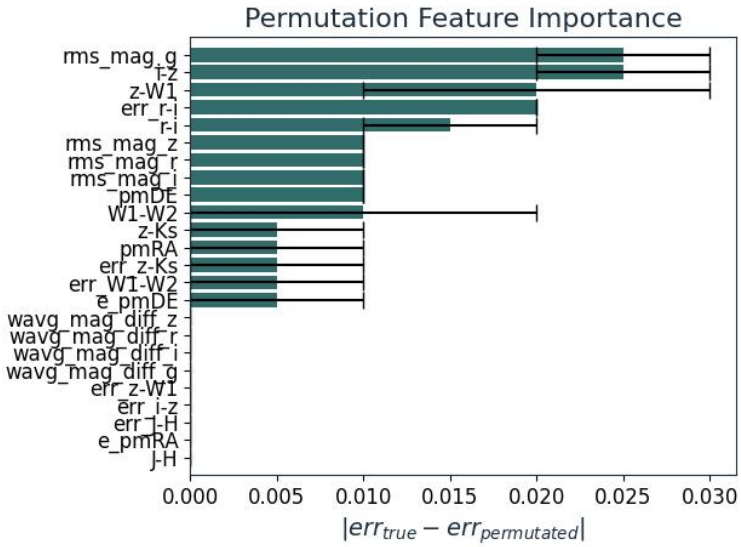

Figure A.3: Permutation feature importance for the best supervised ML solution corresponding to a random forest classifier. The critical features driving the classification include the $i-z$ and $z-W1$ colors, consistent with expectations from quasar selection in the literature. Spurious importance is assigned to the g-band RMS, which is non-physical given the non-detection of all sources in g, while other useful contributions arise from magnitude variance across multiple exposures tracked by the RMS, and proper motion estimates, which help to flag artifacts and stellar contaminants.

and $z-W1$ colors, both widely used in the literature, emerge as key discriminators. Because $W1$ is not available for every source, the model effectively learns from both detections and non-detections, assigning probabilities accordingly. Interestingly, variability (in units of magnitude) across multiple exposures also contributes to the classification. While this may reflect seeing variations or environmental conditions, it could also help identify artifacts, asteroids, or static photometric objects that produce anomalously high variance. Finally, even with large uncertainties, proper motion estimates provide useful leverage for removing stellar contaminants.

To define a final candidate list, we applied probability thresholds tailored to each redshift regime. For quasars at $z > 5.7$, we required a $\geq 95\%$ probability; for those at $5.7 < z < 6$, a threshold of $\geq 85\%$; and for ultra-high-redshift quasars ($z > 6$), a

more flexible threshold of $\geq 70\%$. Applying these cuts yields a final sample of 482,754 candidates (see distribution of magnitude and colors in Fig. A.4).

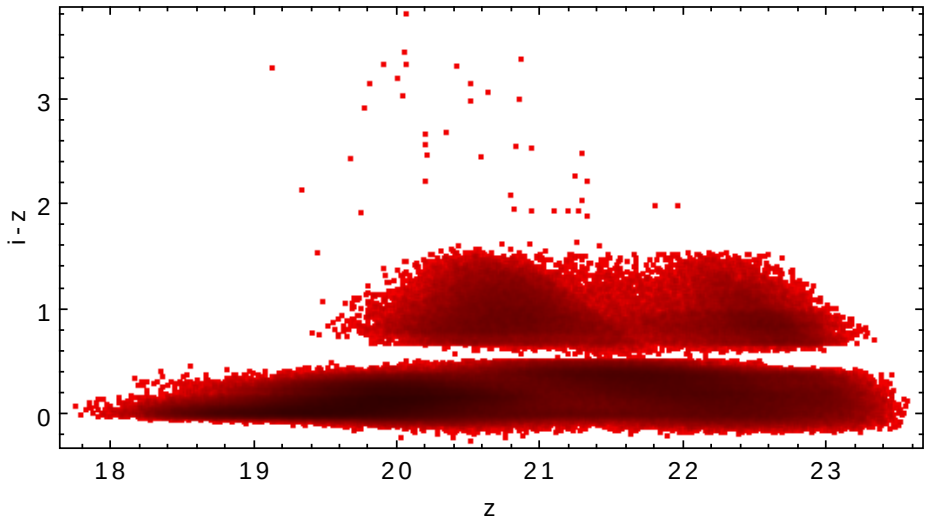

Figure A.4: z-band magnitude vs. $i - z$ color diagram for the final sample of 482,754 quasar candidates selected with *Modulos*. The overdensity at $i - z \lesssim 0.5$ suggests a possible high contamination from ultracool dwarfs.

Appendix B

Latent space representation of HSC quasar candidates

To highlight the transferability of the contrastive learning (CL) method, we constructed a sample based on deep HSC DR3 images, which we present below. This has clear applications for LSST, Euclid, and Roman.

Similar to the preselection of quasar candidates from LS DR10 (section 2.2.1), we selected as input a sample of objects from the HSC catalog found by applying cuts on optical colors, source morphology, PSF quality, and cosmic-ray flags. For this test, we restricted the analysis to a 14 deg^2 field defined by right ascension $213.5 < \mathrm{RA} < 217.5$ and declination $-1.5 < \mathrm{Dec} < 2.0$, which already contains 10 known quasars from the literature. Candidate selection required a significant detection in either the z- or y-band ($\mathrm{S/N}z > 7$ or $\mathrm{S/N}y > 5$), point-like morphology ($z_{\mathrm{PSF}} - z_{\mathrm{cmodel}} < 0.3$ or $y_{\mathrm{PSF}} - y_{\mathrm{cmodel}} < 0.3$), reliable flux measurements (`psfflux_flag = False` in all g, r, i, z, y bands), and absence of cosmic-ray contamination (`pixelflags_cr = False` in all g, r, i, z, y bands). In addition, we required i- or z-band dropouts consistent with the Lyman-α break as follows:

$$(i_{\rm PSF} - z_{\rm PSF} > 1.2 \quad \text{and} \quad (\rm S/N)_i > 3 \quad \text{and} \quad z_{\rm PSF} < 24.5) \quad \text{or} \tag{B.1}$$

$$(\rm S/N)_i < 3 \tag{B.2}$$

for 6 < z < 6.6 and:

$$(z_{\rm PSF} - y_{\rm PSF} > 0.7 \quad \text{and} \quad (\rm S/N)_z > 3 \quad \text{and} \quad z_{\rm PSF} < 25) \quad \text{or} \tag{B.3}$$

$$(\rm S/N)_z < 3 \tag{B.4}$$

for z < 7.8.

From the resulting catalog of 53,109 sources, we retrieved $10'' \times 10''$ coadded cutouts in the optical g, r, i, z, and Y bands. Given the HSC CCDs pixel scale of $0\farcs168$ per pixel, each cutout corresponds to 59×59 pixels, resulting in a tensor of dimensions $53109 \times 59 \times 59 \times 5$. For normalization, we applied a source-specific factor derived from the 99th percentile of pixel values in the brightest band (z or Y), measured within a $1''$ aperture centered on the source.

The main differences between the HSC DR3 and LS DR10 preselection for CL arise from three factors: 1) no requirement of IR bands, 2) stricter flagging of cosmic rays and poor-quality data, and 3) the requirement for well-defined point-like sources. As a result, the HSC training sample is more diverse, incorporating faint sources undetected in WISE, VHS, or UKIDSS, while simultaneously offering richer information to the CL algorithm through its five optical bands. These advantages are reflected in the latent space representation (Fig. B.1), which shows a clear separation between quasars and ultracool dwarfs, as well as a well-defined quasar evolutionary track spanning $z \sim 4$ to $z \sim 7$.

We selected candidates from the quasar evolutionary track along the left boundary of the latent space, defined by an upper horizontal cut at $y = 8.5$ and straight segments connecting the points $(x, y) = (-2, 8.5)$, $(3, 0.5)$, and $(-4, -3)$. This region contained 9,925 candidates. SED fitting with *eazy-py* using the five optical bands yielded 3,930

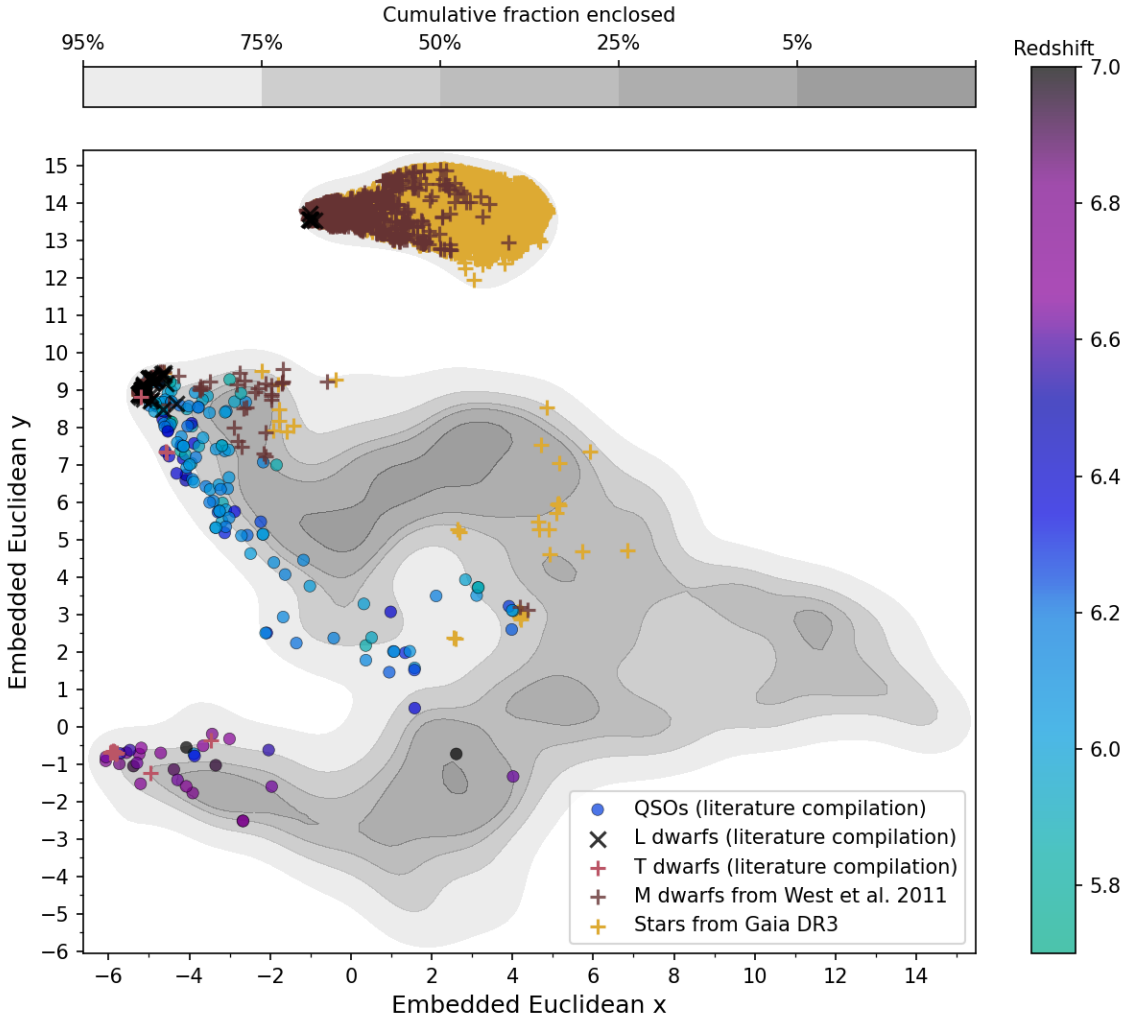

Figure B.1: Embedded Euclidean space generated by UMAP for HSC DR3 *i* and *z*-dropout sources after training the CL algorithm. All sources are represented as grey countours given by the number density; known M, L and T dwarfs are: brown, black and red crosses, respectively; and Gaia DR3 stars are yellow crosses. Spectroscopically confirmed quasars from the literature are represented by filled circles color-coded by their redshift. Known quasars appear to be on a parameter space distinct from ultracool dwarfs, and following an "evolutionary track" at the extreme left of the embedded space. New quasar candidates are selected from these regions.

sources with photometric redshifts $z > 6$. However, none of the stellar templates could be adequately fit in the absence of additional IR bands, resulting in $\chi^2_{\rm BD} = \texttt{NaN}$.

A visual inspection of the $z \gtrsim 6.6$ quasar subregion allowed us to confirm several plausible candidates, reject unreliable data caused by cosmic rays or photometric artifacts, and identify intriguing non–point-like sources that still satisfied our selection criteria. The latter are of particular interest, as similar CL techniques have previously been applied to the search for lensed quasars (e.g., Byrne et al. 2024). Some of these potential lensed candidates displayed consistent orientations (roughly 45° inclination) and frequently exhibited a bright companion source located in the lower-left corner of the cutouts. However, single-epoch images revealed that these objects were in fact moving sources detected only in the *Y* band, with velocities high enough to mimic an elongated shape even during a single exposure, strongly suggesting they are asteroids. This case illustrates the strength of the method in uncovering meaningful structures within the dataset: when combined with flexible constraints, the technique not only identifies high-redshift quasar candidates but also reveals unexpected populations such as potential asteroids.

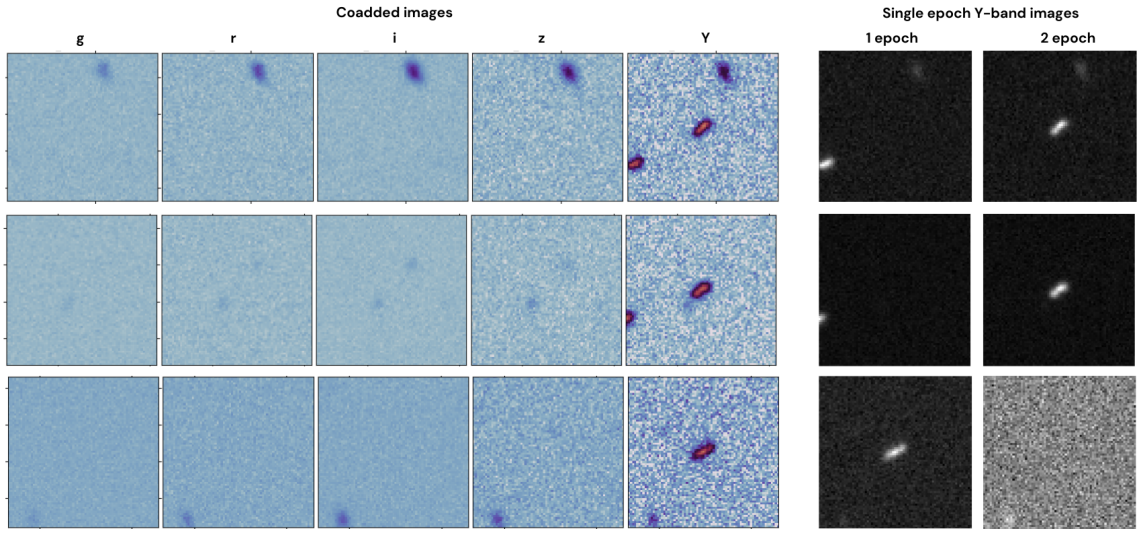

Figure B.2: Coadded HSC optical cutouts (*left*) and single-epoch *Y*-band images (*right*) of three 9moving objects located in the $z \gtrsim 6.6$ quasar subregion of the latent space. These sources, initially resembling potential lensed quasar candidates, are instead consistent with asteroids.

Appendix C

Combining optical and NIR images into CL

The test with HSC data (appendix B) demonstrated the potential to improve the CL performance by incorporating an additional channel of information, resulting in a clearer separation between quasars and contaminants. Motivated by this result, we carried out a parallel experiment using the southern portion of the LS DR10 tensor employed in the main analysis, this time extending it with J- and K-band images from VHS DR7 and $W1$ from AllWISE.

The main challenge in adding heterogeneous photometric data lies in the differences in spatial resolution, which yield images with distinct pixel scales that cannot be directly compiled into the homogeneous tensors required by *Keras*. Also, photometric products from surveys have a diversity in formats and levels of data reduction. In particular, DECaLS and WISE flux images are both already background-substracted and in nanomaggy and Vega units, respectively; whereas the ones from VHS are not background-substracted and released in Analogic Digital Units (ADU). Then the primary step was to homogenize images following the DECaLS format. Also, sanity checks were done to corroborate the magnitudes computed from the flux images were consistent with the reported in the catalog.

To address the resolution discrepancies, we employed the adaptive resampling routines

of the *Astropy* `reproject` package, testing three algorithms: `reproject_interp()`, `reproject_adaptive()`, and `reproject_exact()`. The interpolation method is the fastest, but it does not guarantee flux conservation and performs poorly when the input and output resolutions differ significantly. The adaptive method is specifically designed to handle resolution mismatches and distortions, applying anti-aliasing for a more accurate photometry than interpolation, and allowing flux conservation. The averaging kernels available for the adaptive reprojection are: the `Gaussian` one that ensures more accurate flux conservation at the cost of some blurring, and the `Hann` kernel. Finally, the exact algorithm provides the most rigorous flux-conserving reprojection by computing spherical polygon intersections, though at the expense of substantially longer processing times.

To mitigate boundary effects, we downloaded larger cutouts: 12″ for VHS and 29″ for AllWISE, compared to the 10″ DECaLS cutouts. All native IR images were reprojected to 38×38 pixels with the same $0\farcs262$/pixel resolution as the LS DR10 images, using the three algorithms and kernels described above. As a first test, we applied this procedure to a subsample of known quasars from the literature (Fan et al. 2023) and performed visual inspections of the reprojected images to assess subtle differences in flux conservation and image quality across the methods.

In Figure C.1, we show an example of the different reprojection outcomes for the K- and $W1$-band cutouts of the quasar J0143-5545. For the K-band, where the VHS resolution is relatively close to that of LS DR10, the differences between methods and kernels are mild. Both the adaptive reprojection with a Hann kernel and the exact reprojection preserve the sharp features of the source, whereas the interpolation method produces a slightly blurred background, and the adaptive reprojection with a Gaussian kernel smooths the background pixels more substantially.

In contrast, the $W1$ cutouts from WISE, where the resolution mismatch is more severe, highlight stronger method-dependent effects. The adaptive reprojection with a Hann kernel preserves the overall structure but introduces secondary components with alternating minima and maxima around the central source. The Gaussian kernel produces a smooth,

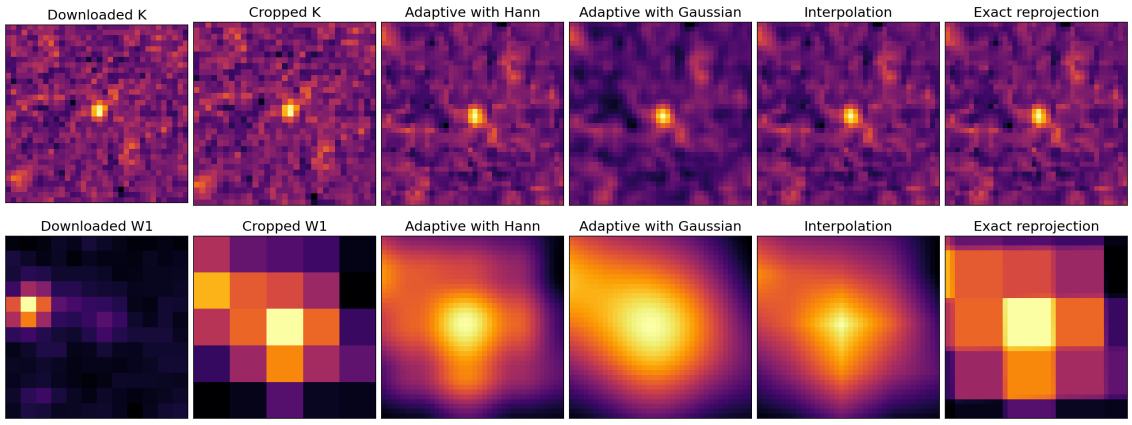

Figure C.1: Comparison of K (*top*) and $W1$ (*bottom*) reprojection stamps computed with different methods. From left to right: original downloaded cutouts, cropped $10''$ stamps, adaptive reprojection with a Hann kernel, adaptive reprojection with a Gaussian kernel, interpolation-based reprojection, and exact reprojection. For VHS (K band), the differences among methods are minimal. In contrast, the lower resolution of WISE relative to LS DR10 makes clearly visible the impact of the different approaches on the shape of the point-like source.

nearly circular profile, while interpolation yields an artificial cross-like feature centered on the source. The exact reprojection, on the other hand, reproduces the cropped image more faithfully but distributes the flux across a larger number of pixels. Consistent trends were noticed in additional quasars and brown dwarfs from the literature.

Given that the adaptive reprojection is better suited to handling large resolution mismatches, and that the Hann kernel provides a balance between preserving the sharpness of the original image and avoiding excessive smoothing, we adopted this method for constructing the LS DR10 IR-extended tensors. We built two versions for a progressive increase in the data complexity: a 6-channel tensor combining the optical bands with J and K (shape $147{,}985 \times 38 \times 38 \times 6$), and a 7-channel tensor that additionally includes $W1$ (shape $127{,}035 \times 38 \times 38 \times 7$). The limited availability of IR stamps produce a decrease in the sample size as additional survey imaging is included. Both tensors were used to train the CL algorithm with the hyperparameters described in Section 2.2.2, yielding two distinct latent-space representations.

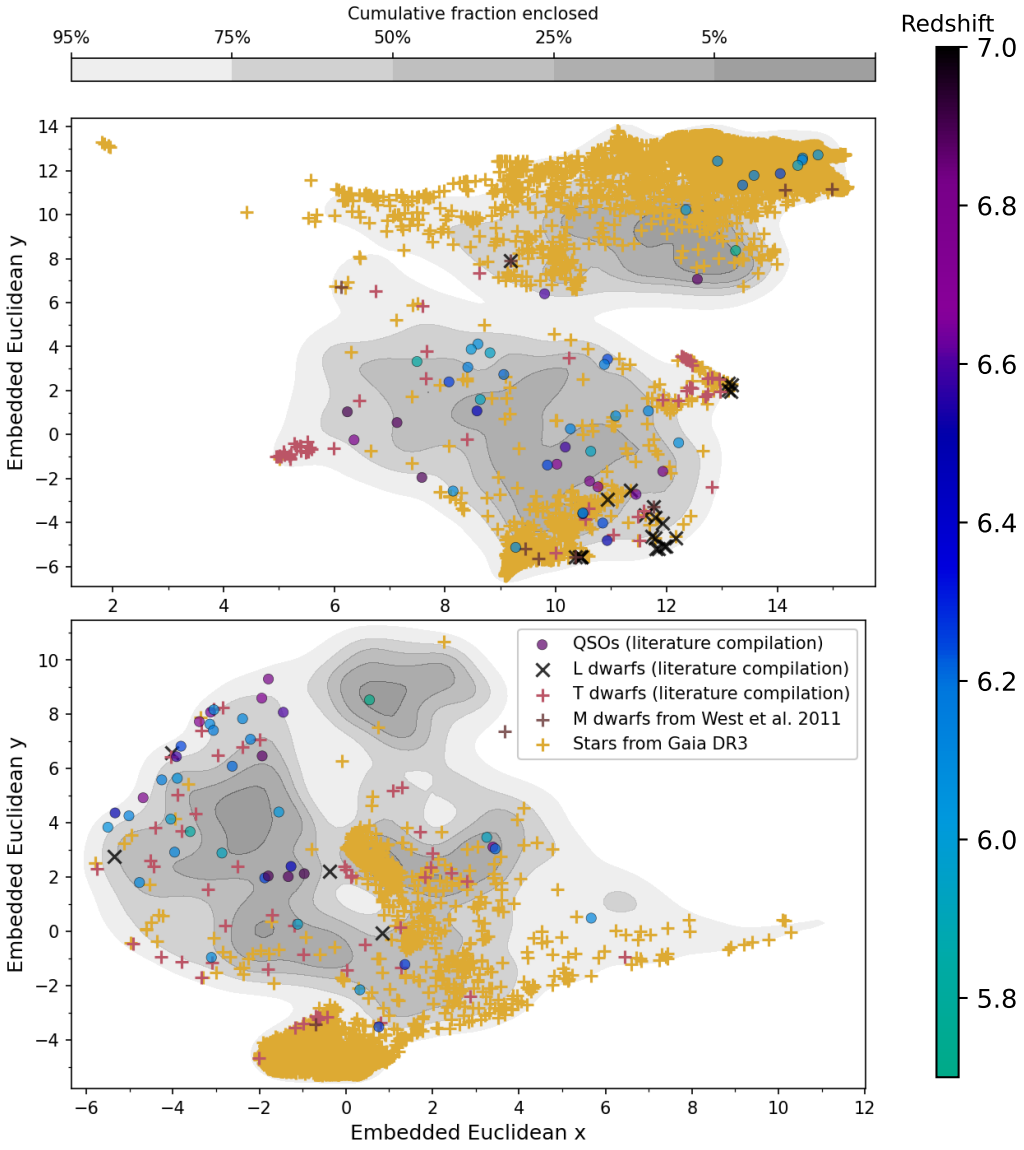

Figure C.2: Embedded Euclidean space generated by UMAP for LS DR10 *i*-band dropout sources after training the CL algorithm with optical plus VHS *J* and *K* imaging (*top*) and with the addition of the *W*1 band (*bottom*). Source number densities are shown as grey countours; known M, L and T dwarfs are marked as brown, black and red crosses, respectively; and Gaia DR3 stars as yellow crosses. Spectroscopically confirmed quasars from the literature are represented by filled circles, color-coded by their redshift. Unlike representations derived from optical-only images, quasars here are more dispersed and overlap significantly with stars and T dwarfs.

The resulting embeddings are shown in Fig. C.2, with the 6-channel tensor (top) and the 7-channel tensor (bottom). The 6-channel embedding displays a butterfly-like structure, with stars spread on both wings, dense concentrations of T and L dwarfs, and quasars sparsely distributed and overlapping with stars. Including the $W1$ band improves the separation of quasars, which cluster more distinctly in the upper-left region, although some overlap with T dwarfs remains. At the same time, stars appear more tightly clustered. However, the requirement of $W1$ detections removes many L dwarf labels from the literature, suggesting that the apparent improvement may partly reflect a reduced stellar sample rather than a true gain in discrimination.

These results suggest that expanding from 6 to 7 channels offers some improvement, but not yet a clear advantage over the simpler, faster optical-only tensor. Further refinements are likely needed, including testing alternative reprojection methods, addressing possible residual errors in the background subtraction, fine-tuning the CL hyperparameters for the new dataset, and incorporating additional data augmentation strategies to mitigate the effects of reprojection and improve robustness.

Appendix D

Structure function parameter space for unWISE light curves

We reconstruct the SF parameter space (A–γ) from Schmidt et al. (2010a) with the results of the light-curve classifier module applied to the unWISE light curves of our sample of quasars and calibrator stars in Fig. D.1. The A–γ plot is dominated by a strong bimodality at low and high amplitude values, reflecting how the sparse data in our light curves makes the optimizer to collapse to the limit values. The diagonal lines are likely evidencing well-define degeneracies between γ and the amplitude.

Given the limited temporal sampling of unWISE light curves, the SF modeling is not a plausible quasar selection method and does not inform about the variability nature of the sources. A higher cadence light curves with multiple epochs in a timeframe of 1 year, the baseline for the SF approach, will definitively allow a meaningful analysis.

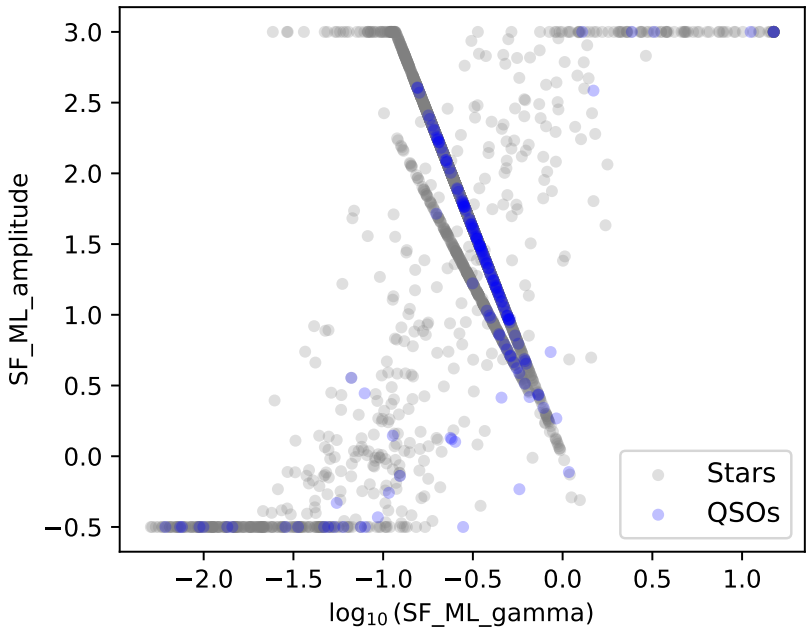

Figure D.1: Structure function parameter space (`SF_ML_amplitude` vs. $\log_{10}$(`SF_ML_gamma`)) for quasars (blue) and stars (gray). Horizontal bands at `SF_ML_amplitude` ∼ –0.5 and 3, and diagonal loci where amplitude and slope correlate linearly, are artifacts of the sparse sampling and limited number of epochs in unWISE light curves. These features indicate fitting degeneracies rather than intrinsic variability, and result in significant overlap between quasars and stars. As a consequence, the SF model provides no discriminative power for our dataset.

Bibliography

Abareshi, B., Aguilar, J., Ahlen, S., et al. (2022), Overview of the instrumentation for the dark energy spectroscopic instrument, The Astronomical Journal, 164, 207. DOI:

Abbott, T., Adamów, M., Aguena, M., et al. (2021), The dark energy survey data release 2, The Astrophysical Journal Supplement Series, 255, 20. DOI:

Abbott, T. M., Abdalla, F., Allam, S., et al. (2018), The dark energy survey: Data release 1, The Astrophysical Journal Supplement Series, 239, 18. DOI:

Abdul-Karim, M., Adame, A., Aguado, D., et al. (2025), Data Release 1 of the Dark Energy Spectroscopic Instrument, arXiv preprint arXiv:2503.14745. DOI:

Ackermann, M., Ajello, M., Allafort, A., et al. (2011), The radio/gamma-ray connection in Active Galactic Nuclei in the era of the Fermi Large Area Telescope, The Astrophysical Journal, 741, 30. DOI:

Albert, L., Artigau, É., Delorme, P., et al. (2011), 37 NEW T-TYPE BROWN DWARFS IN THE CANADA–FRANCE BROWN DWARFS SURVEY, The Astronomical Journal, 141, 203. DOI:

Alexander, D., Bauer, F., Chapman, S., et al. (2005), The X-ray spectral properties of SCUBA galaxies, The Astrophysical Journal, 632, 736. DOI:

Allard, F., Hauschildt, P. H., Alexander, D. R., & Starrfield, S. (1997), Model atmo-

spheres of very low mass stars and brown dwarfs, Annual Review of Astronomy and Astrophysics, 35, 137. DOI:

Allard, F., Hauschildt, P. H., Alexander, D. R., Tamanai, A., & Schweitzer, A. (2001), The limiting effects of dust in brown dwarf model atmospheres, The Astrophysical Journal, 556, 357. DOI:

Allevato, V., Paolillo, M., Papadakis, I., & Pinto, C. (2013), Measuring X-ray variability in faint/sparsely sampled active galactic nuclei, The Astrophysical Journal, 771, 9. DOI:

Anderson, T. W. & Darling, D. A. (1952), Asymptotic theory of certain" goodness of fit" criteria based on stochastic processes, The annals of mathematical statistics, 193. DOI:

Andika, I. T., Jahnke, K., van der Wel, A., et al. (2023), When Spectral Modeling Meets Convolutional Networks: A Method for Discovering Reionization-era Lensed Quasars in Multiband Imaging Data, The Astrophysical Journal, 943, 150. DOI:

Andonie, C., Bauer, F. E., Carraro, R., et al. (2022), Localizing narrow Fe Kα emission within bright AGN, Astronomy & Astrophysics, 664, A46. DOI:

Antonucci, R. & Miller, J. (1985), Spectropolarimetry and the nature of NGC 1068, The Astrophysical Journal, 297, 621. DOI:

Arévalo, P., Churazov, E., Lira, P., et al. (2024), The universal power spectrum of quasars in optical wavelengths-Break timescale scales directly with both black hole mass and the accretion rate, Astronomy & Astrophysics, 684, A133. DOI:

Arévalo, P., Churazov, E., Zhuravleva, I., Hernández-Monteagudo, C., & Revnivtsev, M. (2012), A Mexican hat with holes: calculating low-resolution power spectra from data with gaps, Monthly Notices of the Royal Astronomical Society, 426, 1793. DOI:

Arévalo, P., Lira, P., Sánchez-Sáez, P., et al. (2023), Optical variability in quasars: scalings with black hole mass and Eddington ratio depend on the observed time-scales, Monthly Notices of the Royal Astronomical Society, 526, 6078. DOI:

Artigau, É. (2018), in Handbook of exoplanets (Springer)., 1–19

Artigau, É., Bouchard, S., Doyon, R., & Lafrenière, D. (2009), Photometric variability of the T2. 5 brown dwarf SIMP J013656. 5+ 093347: Evidence for evolving weather patterns, The Astrophysical Journal, 701, 1534. DOI:

Asmus, D., Hoenig, S. F., & Gandhi, P. (2016), The subarcsecond mid-infrared view of local active galactic nuclei. III. Polar dust emission, The Astrophysical Journal, 822, 109. DOI:

Azadi, M., Wilkes, B., Kuraszkiewicz, J., et al. (2023), Disentangling the AGN and Star formation Contributions to the Radio–X-Ray Emission of Radio-loud Quasars at 1< Z< 2, The Astrophysical Journal, 945, 145. DOI:

Bañados, E., Le Brun, V., Belladitta, S., et al. (2025), Euclid: the potential of slitless infrared spectroscopy: a z = 5.4 quasar and new ultracool dwarfs, MNRAS, 542, 1088. DOI:10.1093/mnras/staf1274

Baan, W. & Klöckner, H.-R. (2006), Radio properties of FIR-megamaser nuclei, Astronomy & Astrophysics, 449, 559. DOI:

Badole, S., Jackson, N., Hartley, P., et al. (2020), VLA and ALMA observations of the lensed radio-quiet quasar SDSS J0924+ 0219: a molecular structure in a 3 μ Jy radio source, Monthly Notices of the Royal Astronomical Society, 496, 138. DOI:

Baes, M. & Camps, P. (2015), SKIRT: The design of a suite of input models for Monte Carlo radiative transfer simulations, Astronomy and Computing, 12, 33. DOI:

Bailer-Jones, C. & Mundt, R. (2001), Variability in ultra cool dwarfs: Evidence for the evolution of surface features, Astronomy & Astrophysics, 367, 218. DOI:

Baldwin, J. A. (1977), Luminosity indicators in the spectra of quasi-stellar objects, Astrophysical Journal, Vol. 214, pp. 679-684 (1977)., 214, 679. DOI:

Bañados, E., Schindler, J.-T., Venemans, B. P., et al. (2023), The Pan-STARRS1 z> 5.6 Quasar Survey. II. Discovery of 55 Quasars at 5.6< z< 6.5, The Astrophysical Journal Supplement Series, 265, 29. DOI:

Bañados, E., Venemans, B., Decarli, R., et al. (2016), The Pan-STARRS1 distant z> 5.6 quasar survey: more than 100 quasars within the first Gyr of the Universe, The Astrophysical Journal Supplement Series, 227, 11. DOI:

Bañados, E., Venemans, B., Walter, F., et al. (2013), The Galaxy Environment of a QSO at z 5.7, The Astrophysical Journal, 773, 178. DOI:

Baraffe, I., Chabrier, G., Allard, F., & Hauschildt, P. (1997), Evolutionary models for metal-poor low-mass stars. Lower main sequence of globular clusters and halo field stars, Astronomy and Astrophysics, v. 327, p. 1054-1069 (1997), 327, 1054. DOI:

Baraffe, I., Chabrier, G., Allard, F., & Hauschildt, P. (1998), Evolutionary models for solar metallicity low-mass stars: Mass-magnitude relationships and color-magnitude diagrams, Astron. Astrophys., 337, 403. DOI:

Baraffe, I., Chabrier, G., Barman, T., Allard, F., & Hauschildt, P. (2003), Evolutionary models for cool brown dwarfs and extrasolar giant planets. The case of HD 209458, Astronomy & Astrophysics, 402, 701. DOI:

Barnett, R., Warren, S., Mortlock, D. J., et al. (2019), Euclid preparation-V. Predicted yield of redshift 7< z< 9 quasars from the wide survey, Astronomy & Astrophysics, 631, A85. DOI:

Barth, A. J., Martini, P., Nelson, C. H., & Ho, L. C. (2003), Iron Emission in the z= 6.4 Quasar SDSS J114816. 64+ 525150.3, The Astrophysical Journal, 594, L95. DOI:

Bauer, A., Baltay, C., Coppi, P., et al. (2009), Quasar optical variability in the palomar-quest survey, The Astrophysical Journal, 696, 1241. DOI:

Becker, G. D., Rauch, M., & Sargent, W. L. (2009), High-redshift metals. I. The decline of C IV at z> 5.3, The Astrophysical Journal, 698, 1010. DOI:

Bell, E. F. (2003), Estimating star formation rates from infrared and radio luminosities: the origin of the radio-infrared correlation, The Astrophysical Journal, 586, 794. DOI:

Belladitta, S., Bañados, E., Xie, Z.-L., et al. (2025), Discovery and characterization of 25 new quasars at 4.6 < z < 6.9 from wide-field multiband surveys, A&A, 699, A335. DOI:10.1051/0004-6361/202554859

Belladitta, S., Moretti, A., Caccianiga, A., et al. (2023), A powerful (and likely young) radio-loud quasar at z= 5.3, Astronomy & Astrophysics, 669, A134. DOI:

Belladitta, S., Moretti, A., Caccianiga, A., et al. (2019), An extremely X-ray weak blazar at z= 5, Astronomy & Astrophysics, 629, A68. DOI:

Belladitta, S., Moretti, A., Caccianiga, A., et al. (2020), The first blazar observed at z> 6, Astronomy & Astrophysics, 635, L7. DOI:

Bellm, E. C., Kulkarni, S. R., Graham, M. J., et al. (2018), The Zwicky Transient Facility: system overview, performance, and first results, Publications of the Astronomical Society of the Pacific, 131, 018002. DOI:

Bentz, M. C., Denney, K. D., Grier, C. J., et al. (2013), The low-luminosity end of the radius–luminosity relationship for active galactic nuclei, The Astrophysical Journal, 767, 149. DOI:

Bentz, M. C., Peterson, B. M., Netzer, H., Pogge, R. W., & Vestergaard, M. (2009)a, The radius–luminosity relationship for active galactic nuclei: the effect of host-galaxy starlight on luminosity measurements. II. The full sample of reverberation-mapped AGNs, The Astrophysical Journal, 697, 160. DOI:

Bentz, M. C., Peterson, B. M., Pogge, R. W., Vestergaard, M., & Onken, C. A. (2006), The radius-luminosity relationship for active galactic nuclei: the effect of host-galaxy starlight on luminosity measurements, The Astrophysical Journal, 644, 133. DOI:

Bentz, M. C., Walsh, J. L., Barth, A. J., et al. (2009)b, The lick AGN monitoring project: broad-line region radii and black hole masses from reverberation mapping of Hβ, The Astrophysical Journal, 705, 199. DOI:

Berk, D. E. V., Richards, G. T., Bauer, A., et al. (2001), Composite quasar spectra from the sloan digital sky survey, The Astronomical Journal, 122, 549. DOI:

Bernal, S., Sánchez-Sáez, P., Arévalo, P., et al. (2025), The success of optical variability in uncovering active galactic nuclei in low stellar mass galaxies, Astronomy & Astrophysics, 694, A127. DOI:

Berta, S., Lutz, D., Santini, P., et al. (2013), Panchromatic spectral energy distributions of Herschel sources, Astronomy & Astrophysics, 551, A100. DOI:

Best, W. M., Liu, M. C., Magnier, E. A., et al. (2015), A Search for L/T Transition Dwarfs With Pan-STARRS1 and WISE. II. L/T Transition Atmospheres and Young Discoveries, The Astrophysical Journal, 814, 118. DOI:

Blandford, R. & McKee, C. F. (1982), Reverberation mapping of the emission line regions of Seyfert galaxies and quasars, Astrophysical Journal, Part 1, vol. 255, Apr. 15, 1982, p. 419-439. Research supported by the Alfred P. Sloan Foundation, 255, 419. DOI:

Blandford, R., Netzer, H., Woltjer, L., Courvoisier, T.-L., & Mayor, M. (1990), Active Galactic Nuclei, Active Galactic Nuclei, 97. DOI:

Bochanski, J. J., West, A. A., Hawley, S. L., & Covey, K. R. (2007), Low-mass dwarf template spectra from the Sloan Digital Sky Survey, The Astronomical Journal, 133, 531. DOI:

Bosman, S. E., Álvarez-Márquez, J., Colina, L., et al. (2024), A mature quasar at cosmic dawn revealed by JWST rest-frame infrared spectroscopy, Nature Astronomy, 8, 1054. DOI:

Bosman, S. E., Fan, X., Jiang, L., et al. (2018), New constraints on Lyman-α opacity with a sample of 62 quasarsat z> 5.7, Monthly Notices of the Royal Astronomical Society, 479, 1055. DOI:

Boyle, B., Fong, R., Shanks, T., & Peterson, B. (1990), A catalogue of faint, UV-excess objects, Monthly Notices of the Royal Astronomical Society, 243, 1. DOI:

Brammer, G. B., van Dokkum, P. G., & Coppi, P. (2008), EAZY: a fast, public photometric redshift code, The Astrophysical Journal, 686, 1503. DOI:

Brown, M. J. I., Duncan, K. J., Landt, H., et al. (2019), The spectral energy distributions of active galactic nuclei, MNRAS, 489, 3351. DOI:10.1093/mnras/stz2324

Bruzual, G. & Charlot, S. (2003), Stellar population synthesis at the resolution of 2003, Monthly Notices of the Royal Astronomical Society, 344, 1000. DOI:

Buat, V., Ciesla, L., Boquien, M., Małek, K., & Burgarella, D. (2019), Cold dust and stellar emissions in dust-rich galaxies observed with ALMA: a challenge for SED-fitting techniques, A&A, 632, A79. DOI:10.1051/0004-6361/201936643

Buchner, J. (2014), A statistical test for Nested Sampling algorithms, Statistics and Computing, 26, 383. DOI:

Buchner, J. (2019), Collaborative nested sampling: Big data versus complex physical models, Publications of the Astronomical Society of the Pacific, 131, 108005. DOI:

Burgasser, A. J. (2014), in International Workshop on Stellar Spectral Libraries, Vol. 11, 7–16 DOI:

Burgasser, A. J., Geballe, T., Leggett, S., Kirkpatrick, J. D., & Golimowski, D. A. (2006), A unified near-infrared spectral classification scheme for T dwarfs, The Astrophysical Journal, 637, 1067. DOI:

Burningham, B., Cardoso, C., Smith, L., et al. (2013), 76 T dwarfs from the UKIDSS LAS: benchmarks, kinematics and an updated space density, Monthly Notices of the Royal Astronomical Society, 433, 457. DOI:

Burningham, B., Pinfield, D., Lucas, P., et al. (2010), 47 new T dwarfs from the UKIDSS Large Area Survey, Monthly Notices of the Royal Astronomical Society, 406, 1885. DOI:

Burningham, B., Pinfield, D. J., Leggett, S., et al. (2008), Exploring the substellar temperature regime down to 550 K, Monthly Notices of the Royal Astronomical Society, 391, 320. DOI:

Burtscher, L., de Xivry, G. O., Davies, R., et al. (2015), Obscuration in active galactic nuclei: near-infrared luminosity relations and dust colors, Astronomy & Astrophysics, 578, A47. DOI:

Bustamante, S. & Springel, V. (2019), Spin evolution and feedback of supermassive black holes in cosmological simulations, Monthly Notices of the Royal Astronomical Society, 490, 4133. DOI:

Butler, N. R. & Bloom, J. S. (2011), Optimal time-series selection of quasars, The Astronomical Journal, 141, 93. DOI:

Buzzoni, B., Delabre, B., Dekker, H., et al. (1984), The ESO Faint Object Spectrograph and Camera (EFOSC), ESO Messenger (ISSN 0722-6691), Dec. 1984, p. 9-13., 38, 9. DOI:

Byrne, X., Meyer, R. A., Farina, E. P., et al. (2024), Quasar Island–three new z 6 quasars,

including a lensed candidate, identified with contrastive learning, Monthly Notices of the Royal Astronomical Society, 530, 870. DOI:

Calistro Rivera, G., Alexander, D., Rosario, D., et al. (2021), The multiwavelength properties of red QSOs: Evidence for dusty winds as the origin of QSO reddening, Astronomy & Astrophysics, 649, A102. DOI:

Calistro Rivera, G., Alexander, D. M., Harrison, C. M., et al. (2023), Ubiquitous radio emission in quasars: predominant AGN origin and a connection to jets, dust and winds, arXiv e-prints, arXiv:2312.10177. DOI:10.48550/arXiv.2312.10177

Calistro Rivera, G., Hodge, J. A., Smail, I., et al. (2018), Resolving the ISM at the Peak of Cosmic Star Formation with ALMA: The Distribution of CO and Dust Continuum in z $\sim$ 2.5 Submillimeter Galaxies, ApJ, 863, 56. DOI:10.3847/1538-4357/aacffa

Calistro Rivera, G., Lusso, E., Hennawi, J. F., & Hogg, D. W. (2016), AGNfitter: A Bayesian MCMC Approach to Fitting Spectral Energy Distributions of AGNs, ApJ, 833, 98. DOI:10.3847/1538-4357/833/1/98

Calistro Rivera, G., Williams, W., Hardcastle, M., et al. (2017), The LOFAR window on star-forming galaxies and AGNs–curved radio SEDs and IR–radio correlation at 0< z< 2.5, Monthly Notices of the Royal Astronomical Society, 469, 3468. DOI:

Calzetti, D., Armus, L., Bohlin, R. C., et al. (2000), The dust content and opacity of actively star-forming galaxies, The Astrophysical Journal, 533, 682. DOI:

Camps, P. & Baes, M. (2015), SKIRT: An advanced dust radiative transfer code with a user-friendly architecture, Astronomy and Computing, 9, 20. DOI:

Camps, P. & Baes, M. (2020), SKIRT 9: Redesigning an advanced dust radiative transfer code to allow kinematics, line transfer and polarization by aligned dust grains, Astronomy and Computing, 31, 100381. DOI:

Cardoso, C., Burningham, B., Smart, R. L., et al. (2015), 49 new T dwarfs identified using methane imaging, Monthly Notices of the Royal Astronomical Society, 450, 2486. DOI:

Carilli, C., Wang, R., Fan, X., et al. (2010), Ionization near zones associated with quasars at z 6, The Astrophysical Journal, 714, 834. DOI:

Carnall, A., Begley, R., McLeod, D., et al. (2023), A first look at the SMACS0723 JWST ERO: spectroscopic redshifts, stellar masses, and star-formation histories, Monthly Notices of the Royal Astronomical Society: Letters, 518, L45. DOI:

Carroll, B. W. & Ostlie, D. A. (2017), An introduction to modern astrophysics (Cambridge University Press). DOI:

Cazzoli, S., Muñoz, L. H., Márquez, I., et al. (2022), Unexplored outflows in nearby low luminosity AGNs-The case of NGC 1052, Astronomy & Astrophysics, 664, A135. DOI:

Cerqueira-Campos, F., Rodríguez-Ardila, A., Panda, S., et al. (2023), Coronal Line Forest AGN-II. Analysis of the spectral energy distribution, Monthly Notices of the Royal Astronomical Society, stad1908. DOI:

Chabrier, G. (2003), Galactic stellar and substellar initial mass function, Publications of the Astronomical Society of the Pacific, 115, 763. DOI:

Chabrier, G. & Baraffe, I. (2000), Theory of low-mass stars and substellar objects, Annual Review of Astronomy and Astrophysics, 38, 337. DOI:

Chabrier, G., Baraffe, I., Allard, F., & Hauschildt, P. (2000), Evolutionary models for very low-mass stars and brown dwarfs with dusty atmospheres, The Astrophysical Journal, 542, 464. DOI:

Chambers, K. C., Magnier, E., Metcalfe, N., et al. (2016), The pan-starrs1 surveys, arXiv preprint arXiv:1612.05560. DOI:

Charisi, M., Bartos, I., Haiman, Z., et al. (2016), A population of short-period variable quasars from PTF as supermassive black hole binary candidates, Monthly Notices of the Royal Astronomical Society, 463, 2145. DOI:

Chary, R. & Elbaz, D. (2001), Interpreting the cosmic infrared background: constraints on the evolution of the dust-enshrouded star formation rate, The Astrophysical Journal, 556, 562. DOI:

Chen, S.-F. S., Simcoe, R. A., Torrey, P., et al. (2017), Mg II absorption at 2< z< 7 with Magellan/Fire. III. Full statistics of absorption toward 100 high-redshift QSOs, The Astrophysical Journal, 850, 188. DOI:

Chen, T., Kornblith, S., Norouzi, M., & Hinton, G. (2020), in International conference on machine learning, PmLR, 1597–1607 DOI:

Chiu, K., Fan, X., Leggett, S., et al. (2006), Seventy-one new l and t dwarfs from the sloan digital sky survey, The Astronomical Journal, 131, 2722. DOI:

Chiu, K., Liu, M. C., Jiang, L., et al. (2008), Four faint T dwarfs from the UKIRT Infrared Deep Sky Survey (UKIDSS) Southern Stripe, Monthly Notices of the Royal Astronomical Society: Letters, 385, L53. DOI:

Ciesla, L., Charmandaris, V., Georgakakis, A., et al. (2015), Constraining the properties of AGN host galaxies with spectral energy distribution modelling, Astronomy & Astrophysics, 576, A10. DOI:

Cirasuolo, M., Afonso, J., Carollo, M., et al. (2014), in Ground-based and airborne instrumentation for astronomy V, Vol. 9147, SPIE, 202–214 DOI:

Clarke, A., Scaife, A., Greenhalgh, R., & Griguta, V. (2020), Identifying galaxies, quasars, and stars with machine learning: A new catalogue of classifications for 111 million SDSS sources without spectra, Astronomy & Astrophysics, 639, A84. DOI:

Collaboration, G. & Vallenari, A. (2023), Gaia Data Release 3. Summary of the content and survey properties, Astronomy & astrophysics, 674, 23. DOI:

Collaboration, P., Aghanim, N., Akrami, Y., et al. (2020), Planck 2018 results, A&A, 641, A1. DOI:

Collier, S. & Peterson, B. M. (2001), Characteristic ultraviolet/optical timescales in active galactic nuclei, The Astrophysical Journal, 555, 775. DOI:

Comastri, A. (2004), Compton-thick AGN: the dark side of the X-ray background, Supermassive Black Holes in the Distant Universe, 245. DOI:

Condon, J., Anderson, E., & Broderick, J. (1995), Radio Identifications of Extragalactic IRAS Sources, Astronomical Journal v. 109, p. 2318, 109, 2318. DOI:

Condon, J. J. (1992), Radio emission from normal galaxies., ARA&A, 30, 575. DOI:10.1146/annurev.aa.30.090192.003043

Conroy, C. & Gunn, J. E. (2010), The propagation of uncertainties in stellar population synthesis modeling. III. Model calibration, comparison, and evaluation, The Astrophysical Journal, 712, 833. DOI:

Conroy, C., Gunn, J. E., & White, M. (2009), The propagation of uncertainties in stellar population synthesis modeling. I. The relevance of uncertain aspects of stellar evolution and the initial mass function to the derived physical properties of galaxies, The Astrophysical Journal, 699, 486. DOI:

Cooper, T. J., Simcoe, R. A., Cooksey, K. L., et al. (2019), Heavy element absorption systems at 5.0< z< 6.8: metal-poor neutral gas and a diminishing signature of highly ionized circumgalactic matter, The Astrophysical Journal, 882, 77. DOI:

Costa, T., Sijacki, D., Trenti, M., & Haehnelt, M. G. (2014), The environment of bright QSOs at z 6: star-forming galaxies and X-ray emission, Monthly Notices of the Royal Astronomical Society, 439, 2146. DOI:

Da Cunha, E., Charlot, S., & Elbaz, D. (2008), A simple model to interpret the ultraviolet, optical and infrared emission from galaxies, Monthly Notices of the Royal Astronomical Society, 388, 1595. DOI:

Dale, D. A. & Helou, G. (2002), The infrared spectral energy distribution of normal star-forming galaxies: Calibration at far-infrared and submillimeter wavelengths, The Astrophysical Journal, 576, 159. DOI:

Dalla Bontà, E., Peterson, B., Grier, C., et al. (2025), Estimating masses of supermassive black holes in active galactic nuclei from the Hα emission line, Astronomy & Astrophysics, 696, A48. DOI:

Davidson, K. (1972), Photoionization and the emission-line spectra of quasi-stellar objects, Astrophysical Journal, vol. 171, p. 213, 171, 213. DOI:

Davis, S. W., Woo, J.-H., & Blaes, O. M. (2007), The UV continuum of quasars: models and SDSS spectral slopes, The Astrophysical Journal, 668, 682. DOI:

Dawson, K. S., Schlegel, D. J., Ahn, C. P., et al. (2012), The baryon oscillation spectroscopic survey of SDSS-III, The Astronomical Journal, 145, 10. DOI:

Day-Jones, A., Marocco, F., Pinfield, D., et al. (2013), The sub-stellar birth rate from UKIDSS, Monthly Notices of the Royal Astronomical Society, 430, 1171. DOI:

De Jong, R. S., Bellido-Tirado, O., Chiappini, C., et al. (2012), in Ground-based and airborne instrumentation for astronomy IV, Vol. 8446, SPIE, 252–266 DOI:

De Jong, T., Klein, U., Wielebinski, R., & Wunderlichº, E. (1985), from spiral galaxies: a close correlation, Astron. Astrophys, 147, L6. DOI:

De Rosa, G., Venemans, B. P., Decarli, R., et al. (2014), Black hole mass estimates and emission-line properties of a sample of redshift z> 6.5 quasars, The Astrophysical Journal, 790, 145. DOI:

Deacon, N. R., Liu, M. C., Magnier, E. A., et al. (2011), Four new T dwarfs identified in Pan-STARRS 1 commissioning data, The Astronomical Journal, 142, 77. DOI:

Decarli, R., Walter, F., Venemans, B. P., et al. (2017), Rapidly star-forming galaxies adjacent to quasars at redshifts exceeding 6, Nature, 545, 457. DOI:

Decarli, R., Walter, F., Venemans, B. P., et al. (2018), An ALMA [C II] Survey of 27 Quasars at z> 5.94, The Astrophysical Journal, 854, 97. DOI:

Deeming, T. J. (1975), Fourier analysis with unequally-spaced data, Astrophysics and Space Science, 36, 137. DOI:

Delvecchio, I., Daddi, E., Sargent, M., et al. (2021), The infrared-radio correlation of star-forming galaxies is strongly M⋆-dependent but nearly redshift-invariant since z∼ 4, Astronomy & Astrophysics, 647, A123. DOI:

Dey, A., Schlegel, D. J., Lang, D., et al. (2019), Overview of the DESI legacy imaging surveys, The Astronomical Journal, 157, 168. DOI:

Di Matteo, T., Springel, V., & Hernquist, L. (2005), Energy input from quasars regulates the growth and activity of black holes and their host galaxies, nature, 433, 604. DOI:

Diamond-Stanic, A. M., Fan, X., Brandt, W., et al. (2009), High-redshift SDSS quasars with weak emission lines, The Astrophysical Journal, 699, 782. DOI:

D'Odorico, V., Cupani, G., Cristiani, S., et al. (2013), Metals in the IGM approaching the re-ionization epoch: results from X-shooter at the VLT, Monthly Notices of the Royal Astronomical Society, 435, 1198. DOI:

Donley, J., Rieke, G., Rigby, J., & Pérez-González, P. (2005), Unveiling a population of AGNs not detected in X-rays, The Astrophysical Journal, 634, 169. DOI:

Donley, J. L., Koekemoer, A. M., Brusa, M., et al. (2012), Identifying Luminous Active Galactic Nuclei in Deep Surveys: Revised IRAC Selection Criteria, ApJ, 748, 142. DOI:10.1088/0004-637X/748/2/142

Doore, K., Eufrasio, R. T., Lehmer, B. D., et al. (2021), On the impact of inclination-dependent attenuation on derived star formation histories: Results from disk galaxies in the great observatories origins deep survey fields, The Astrophysical Journal, 923, 26. DOI:

Dressler, A., Bigelow, B., Hare, T., et al. (2011), IMACS: The Inamori-Magellan Areal Camera and Spectrograph on Magellan-Baade, Publications of the Astronomical Society of the Pacific, 123, 288. DOI:

Drlica-Wagner, A., Carlin, J. L., Nidever, D. L., et al. (2021), The DECam local volume exploration survey: Overview and first data release, The Astrophysical Journal Supplement Series, 256, 2. DOI:

Drlica-Wagner, A., Ferguson, P., Adamów, M., et al. (2022), The DECam local volume exploration survey data release 2, The Astrophysical Journal Supplement Series, 261, 38. DOI:

Duncan, K. & Conselice, C. J. (2015), Powering reionization: assessing the galaxy ionizing photon budget at z< 10, Monthly Notices of the Royal Astronomical Society, 451, 2030. DOI:

Dux, F., Lemon, C., Courbin, F., et al. (2024), Nine lensed quasars and quasar pairs discovered through spatially extended variability in Pan-STARRS, Astronomy & Astrophysics, 682, A47. DOI:

Dye, S., Lawrence, A., Read, M., et al. (2018), The UKIRT Hemisphere Survey: definition and J-band data release, Monthly Notices of the Royal Astronomical Society, 473, 5113. DOI:

D'Odorico, V., Bañados, E., Becker, G., et al. (2023), XQR-30: The ultimate XSHOOTER quasar sample at the reionization epoch, Monthly Notices of the Royal Astronomical Society, 523, 1399. DOI:

Eilers, A.-C., Mackenzie, R., Pizzati, E., et al. (2024), EIGER. VI. The correlation function, host halo mass, and duty cycle of luminous quasars at z$\gtrsim$6, The Astrophysical Journal, 974, 275. DOI:

Eisert, L., Bottrell, C., Pillepich, A., et al. (2024), ERGO-ML: comparing IllustrisTNG and HSC galaxy images via contrastive learning, Monthly Notices of the Royal Astronomical Society, 528, 7411. DOI:

Enoch, M. L., Brown, M. E., & Burgasser, A. J. (2003), Photometric variability at the L/T dwarf boundary, The Astronomical Journal, 126, 1006. DOI:

Esparza-Arredondo, D., Gonzalez-Martín, O., Dultzin, D., et al. (2021), The dust-gas AGN torus as constrained from X-ray and mid-infrared observations, Astronomy & Astrophysics, 651, A91. DOI:

Esplin, T., Luhman, K., Cushing, M., et al. (2016), Photometric monitoring of the coldest known brown dwarf with the Spitzer Space Telescope, The Astrophysical Journal, 832, 58. DOI:

Eyheramendy, S., Elorrieta, F., & Palma, W. (2018), An irregular discrete time series model to identify residuals with autocorrelation in astronomical light curves, Monthly Notices of the Royal Astronomical Society, 481, 4311. DOI:

Fabian, A. C. (2012), Observational evidence of active galactic nuclei feedback, Annual Review of Astronomy and Astrophysics, 50, 455. DOI:

Fan, X., Bañados, E., & Simcoe, R. A. (2023), Quasars and the intergalactic medium at cosmic dawn, Annual Review of Astronomy and Astrophysics, 61, 373. DOI:

Fan, X., Narayanan, V. K., Lupton, R. H., et al. (2001), A Survey of z> 5.8 Quasarsin the Sloan Digital Sky Survey. I. Discovery of Three New Quasars and theSpatial Density of Luminous Quasars atz 6, The Astronomical Journal, 122, 2833. DOI:

Fan, X., Strauss, M. A., Schneider, D. P., et al. (1999), High-Redshift Quasars Found in Sloan Digital SkySurvey CommissioningData, The Astronomical Journal, 118, 1. DOI:

Fanaroff, B. L. & Riley, J. M. (1974), The morphology of extragalactic radio sources of high and low luminosity, Monthly Notices of the Royal Astronomical Society, 167, 31P. DOI:

Farina, E. P., Schindler, J.-T., Walter, F., et al. (2022), The x–shooter/alma sample of quasars in the epoch of reionization. ii. black hole masses, eddington ratios, and the formation of the first quasars, The Astrophysical Journal, 941, 106. DOI:

Findlay, J., Sutherland, W., Venemans, B., et al. (2012), Selection constraints on high-redshift quasar searches in the VISTA Kilo-degree Infrared Galaxy survey, Monthly Notices of the Royal Astronomical Society, 419, 3354. DOI:

Fine, S., Croom, S., Hopkins, P., et al. (2008), Constraining the quasar population with the broad-line width distribution, Monthly Notices of the Royal Astronomical Society, 390, 1413. DOI:

Flaugher, B., Diehl, H., Honscheid, K., et al. (2015), The dark energy camera, The Astronomical Journal, 150, 150. DOI:

Flesch, E. W. (2023), The Million Quasars (Milliquas) Catalogue, v8, arXiv preprint arXiv:2308.01505. DOI:

Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. (2013), emcee: the MCMC hammer, Publications of the Astronomical Society of the Pacific, 125, 306. DOI:

García-Bernete, I., González-Martín, O., Almeida, C. R., et al. (2022), Torus and polar dust dependence on active galactic nucleus properties, Astronomy & Astrophysics, 667, A140. DOI:

García-Bernete, I., Ramos Almeida, C., Landt, H., et al. (2017), The infrared to X-ray correlation spectra of unobscured type 1 active galactic nuclei, Monthly Notices of the Royal Astronomical Society, 469, 110. DOI:

Gaskell, C. M. (1982), A redshift difference between high and low ionization emission-line regions in QSOs-Evidence for radial motions, Astrophysical Journal, Part 1, vol. 263, Dec. 1, 1982, p. 79-86., 263, 79. DOI:

Geha, M., Alcock, C., Allsman, R. A., et al. (2003), Variability-selected Quasars in MACHO Project MagellanicCloud Fields, The Astronomical Journal, 125, 1. DOI:

Gelino, C. R., Marley, M. S., Holtzman, J. A., Ackerman, A. S., & Lodders, K. (2002), L dwarf variability: I-band observations, The Astrophysical Journal, 577, 433. DOI:

Ghasemi-Nodehi, M., Tabatabaei, F. S., Sargent, M., et al. (2022), Evolution of thermal and non-thermal radio continuum emission on kpc scales–predictions for SKA, Monthly Notices of the Royal Astronomical Society, 515, 1158. DOI:

Ghisellini, G., Sbarrato, T., Tagliaferri, G., et al. (2014), SDSS J114657. 79+ 403708.6: The third most distant blazar at z= 5.0, Monthly Notices of the Royal Astronomical Society: Letters, 440, L111. DOI:

Glikman, E., LaMassa, S., Piconcelli, E., Zappacosta, L., & Lacy, M. (2024), Accretion and obscuration in merger-dominated luminous red quasars, Monthly Notices of the Royal Astronomical Society, 528, 711. DOI:

Glikman, E., Urrutia, T., Lacy, M., et al. (2012), FIRST–2MASS Red Quasars: Transitional Objects Emerging From the Dust, The Astrophysical Journal, 757, 51. DOI:

Golimowski, D. A., Leggett, S., Marley, M., et al. (2004), L and M photometry of ultracool dwarfs, The Astronomical Journal, 127, 3516. DOI:

Gómez, J. L., Lobanov, A. P., Bruni, G., et al. (2016), Probing the innermost regions of AGN jets and their magnetic fields with RadioAstron. I. Imaging BL Lacertae at 21 μas resolution, The Astrophysical Journal, 817, 96. DOI:

González-Martín, O., Masegosa, J., García-Bernete, I., et al. (2019), Exploring the Mid-infrared SEDs of Six AGN Dusty Torus Models. II. The Data, The Astrophysical Journal, 884, 11. DOI:

Gonzalez-Martin, O. & Vaughan, S. (2012), X-ray variability of 104 active galactic nuclei-XMM-Newton power-spectrum density profiles, Astronomy & Astrophysics, 544, A80. DOI:

Graham, M. J., Djorgovski, S., Drake, A. J., et al. (2014), A novel variability-based method for quasar selection: evidence for a rest-frame 54 d characteristic time-scale, Monthly Notices of the Royal Astronomical Society, 439, 703. DOI:

Graham, M. J., Djorgovski, S., Drake, A. J., et al. (2017), Understanding extreme quasar optical variability with CRTS–I. Major AGN flares, Monthly Notices of the Royal Astronomical Society, 470, 4112. DOI:

Graham, M. J., Djorgovski, S. G., Stern, D., et al. (2015), A possible close supermassive black-hole binary in a quasar with optical periodicity, Nature, 518, 74. DOI:

Granato, G. L., De Zotti, G., Silva, L., Bressan, A., & Danese, L. (2004), A physical model for the coevolution of QSOs and their spheroidal hosts, The Astrophysical Journal, 600, 580. DOI:

Greene, J. E. & Ho, L. C. (2005), Estimating black hole masses in active galaxies using the Hα emission line, The Astrophysical Journal, 630, 122. DOI:

Greene, J. E., Labbe, I., Goulding, A. D., et al. (2024), UNCOVER spectroscopy confirms the surprising ubiquity of active galactic nuclei in red sources at z> 5, The Astrophysical Journal, 964, 39. DOI:

Grevesse, N., Noels, A., & Sauval, A. (1993), A revision of the solar abundance of dysprosium, Astronomy and Astrophysics, Vol. 271, p. 587 (1993), 271, 587. DOI:

Guo, H., Peng, J., Zhang, K., et al. (2020), High-redshift extreme variability quasars from Sloan Digital Sky Survey multiepoch spectroscopy, The Astrophysical Journal, 905, 52. DOI:

Hall, P. J., Schilizzi, R. T., Dewdney, P., & Lazio, T. (2008), The square kilometer array (SKA) radio telescope: Progress and technical directions, URSI Radio Science Bulletin, 2008, 4. DOI:

Hardcastle, M., Williams, W., Best, P., et al. (2019), Radio-loud AGN in the first LoTSS data release-The lifetimes and environmental impact of jet-driven sources, Astronomy & Astrophysics, 622, A12. DOI:

Harding, L. K., Hallinan, G., Boyle, R. P., et al. (2013), Periodic optical variability of radio-detected ultracool dwarfs, The Astrophysical Journal, 779, 101. DOI:

Harikane, Y., Zhang, Y., Nakajima, K., et al. (2023), A JWST/NIRSpec first census of broad-line AGNs at z= 4–7: detection of 10 faint AGNs with M BH~ 106–108 $M_{\odot}$ and their host galaxy properties, The Astrophysical Journal, 959, 39. DOI:

Häring, N. & Rix, H.-W. (2004), On the black hole mass-bulge mass relation, The Astrophysical Journal, 604, L89. DOI:

Hauschildt, P. H., Allard, F., & Baron, E. (1999), The NEXTGEN model atmosphere grid for 3000 ≤ Teff leq 10,000 K, The Astrophysical Journal, 512, 377. DOI:

Helou, G., Soifer, B., & Rowan-Robinson, M. (1985), Thermal infrared and nonthermal radio-Remarkable correlation in disks of galaxies, The Astrophysical Journal, 298, L7. DOI:

Hernán-Caballero, A., Hatziminaoglou, E., Alonso-Herrero, A., & Mateos, S. (2016), The near-to-mid infrared spectrum of quasars, Monthly Notices of the Royal Astronomical Society, 463, 2064. DOI:

Hönig, S. F. (2019), Redefining the Torus: A Unifying View of AGNs in the Infrared and Submillimeter, The Astrophysical Journal, 884, 171. DOI:

Hönig, S. F. & Kishimoto, M. (2017), Dusty winds in active galactic nuclei: reconciling observations with models, The Astrophysical Journal Letters, 838, L20. DOI:

Hönig, S. F., Kishimoto, M., Tristram, K. R., et al. (2013), Dust in the polar region as a major contributor to the infrared emission of active galactic nuclei, The Astrophysical Journal, 771, 87. DOI:

Hook, I., Jørgensen, I., Allington-Smith, J., et al. (2004), The Gemini–North Multi-Object Spectrograph: Performance in imaging, long-slit, and multi-object spectroscopic modes, Publications of the Astronomical Society of the Pacific, 116, 425. DOI:

Huertas-Company, M., Sarmiento, R., & Knapen, J. H. (2023), A brief review of contrastive learning applied to astrophysics, RAS Techniques and Instruments, 2, 441. DOI:

Hviding, R. E., de Graaff, A., Miller, T. B., et al. (2025), RUBIES: A Spectroscopic Census of Little Red Dots; All V-Shaped Point Sources Have Broad Lines, arXiv e-prints, arXiv. DOI:

Ighina, L., Caccianiga, A., Connor, T., et al. (2025), X-Ray Investigation of Possible Super-Eddington Accretion in a Radio-loud Quasar at z= 6.13, The Astrophysical Journal Letters, 990, L56. DOI:

Irwin, J., Cook, T., Stein, M., et al. (2024), CHANG-ES. XXXII. Spatially Resolved Thermal–Nonthermal Separation from Radio Data Alone—New Probes into NGC 3044 and NGC 5775, The Astronomical Journal, 168, 138. DOI:

Ivezić, Ž., Kahn, S. M., Tyson, J. A., et al. (2019), LSST: from science drivers to reference design and anticipated data products, The Astrophysical Journal, 873, 111. DOI:

Jeffreys, H. (1961), International series of monographs on physics

Jiang, H., Hu, Z., Xu, M., et al. (2018), in Ground-based and Airborne Instrumentation for Astronomy VII, Vol. 10702, SPIE, 825–837 DOI:

Jiang, L., McGreer, I. D., Fan, X., et al. (2016), The final SDSS high-redshift quasar sample of 52 quasars at z> 5.7, The Astrophysical Journal, 833, 222. DOI:

Jin, X., Zhang, Y., Zhang, J., et al. (2019), Efficient selection of quasar candidates based on optical and infrared photometric data using machine learning, Monthly Notices of the Royal Astronomical Society, 485, 4539. DOI:

Just, D. W., Brandt, W. N., Shemmer, O., et al. (2007), The X-ray properties of the most luminous quasars from the sloan digital sky survey, The Astrophysical Journal, 665, 1004. DOI:

Kaiser, N., Aussel, H., Burke, B. E., et al. (2002), in Survey and Other Telescope Technologies and Discoveries, Vol. 4836, SPIE, 154–164 DOI:

Kaiser, N., Burgett, W., Chambers, K., et al. (2010), in Ground-based and airborne telescopes III, Vol. 7733, SPIE, 159–172 DOI:

Kang, J.-L. & Wang, J.-X. (2022), The X-ray coronae in NuSTAR bright active galactic nuclei, The Astrophysical Journal, 929, 141. DOI:

Karthik, A., Wu, M., Goodman, N., & Tamkin, A. (2021), Tradeoffs Between Contrastive and Supervised Learning: An Empirical Study, arXiv e-prints, arXiv:2112.05340. DOI:10.48550/arXiv.2112.05340

Kaspi, S., Brandt, W., Maoz, D., et al. (2007), Reverberation mapping of high-luminosity quasars: first results, The Astrophysical Journal, 659, 997. DOI:

Kaspi, S., Maoz, D., Netzer, H., et al. (2005), The relationship between luminosity and broad-line region size in active galactic nuclei, The Astrophysical Journal, 629, 61. DOI:

Kaspi, S., Smith, P. S., Netzer, H., et al. (2000), Reverberation Measurements for 17 Quasars and theSize-Mass-Luminosity Relations in Active GalacticNuclei, The Astrophysical Journal, 533, 631. DOI:

Kauffmann, G. & Haehnelt, M. (2000), A unified model for the evolution of galaxies and quasars, Monthly Notices of the Royal Astronomical Society, 311, 576. DOI:

Kaviraj, S., Laigle, C., Kimm, T., et al. (2017), The Horizon-AGN simulation: evolution of galaxy properties over cosmic time, Monthly Notices of the Royal Astronomical Society, 467, 4739. DOI:

Kawamuro, T., Ricci, C., Imanishi, M., et al. (2022), BASS XXXII: Studying the Nuclear Millimeter-wave Continuum Emission of AGNs with ALMA at Scales $\lesssim$ 100–200 pc, The Astrophysical Journal, 938, 87. DOI:

Kelly, B. C., Bechtold, J., & Siemiginowska, A. (2009), Are the variations in quasar optical flux driven by thermal fluctuations?, The Astrophysical Journal, 698, 895. DOI:

Kelly, B. C., Treu, T., Malkan, M., Pancoast, A., & Woo, J.-H. (2013), Active galactic nucleus black hole mass estimates in the era of time domain astronomy, The Astrophysical Journal, 779, 187. DOI:

Kendall, T., Tamura, M., Tinney, C., et al. (2007), Two T dwarfs from the UKIDSS early data release, Astronomy & Astrophysics, 466, 1059. DOI:

Killi, M., Watson, D., Brammer, G., et al. (2024), Deciphering the JWST spectrum of a 'little red dot'at z 4.53: An obscured AGN and its star-forming host, Astronomy & Astrophysics, 691, A52. DOI:

Kim, D. & Im, M. (2018), What makes red quasars red?-Observational evidence for dust extinction from line ratio analysis, Astronomy & Astrophysics, 610, A31. DOI:

Kim, D., Im, M., & Kim, M. (2010), New estimators of black hole mass in active galactic nuclei with hydrogen Paschen lines, The Astrophysical Journal, 724, 386. DOI:

Kim, D.-W., Protopapas, P., Bailer-Jones, C. A., et al. (2014), The EPOCH Project-I. Periodic variable stars in the EROS-2 LMC database, Astronomy & Astrophysics, 566, A43. DOI:

Kim, D.-W., Protopapas, P., Byun, Y.-I., et al. (2011), Quasi-stellar object selection algorithm using time variability and machine learning: Selection of 1620 quasi-stellar object candidates from MACHO Large Magellanic Cloud database, The Astrophysical Journal, 735, 68. DOI:

Kim, S., Stiavelli, M., Trenti, M., et al. (2009), The environments of high-redshift quasi-stellar objects, The Astrophysical Journal, 695, 809. DOI:

Kim, Y., Im, M., Jeon, Y., et al. (2019), The Infrared Medium-deep Survey. VI. Discovery of Faint Quasars at z 5 with a Medium-band-based Approach, The Astrophysical Journal, 870, 86. DOI:

Kirkpatrick, J. D., Cushing, M. C., Gelino, C. R., et al. (2011), The first hundred brown dwarfs discovered by the Wide-field Infrared Survey Explorer (WISE), The Astrophysical Journal Supplement Series, 197, 19. DOI:

Kirkpatrick, J. D., Looper, D. L., Burgasser, A. J., et al. (2010), Discoveries from a near-infrared proper motion survey using multi-epoch Two Micron All-Sky Survey data, The Astrophysical Journal Supplement Series, 190, 100. DOI:

Knapp, G., Leggett, S. K., Fan, X., et al. (2004)a, Near-infrared photometry and spectroscopy of L and T dwarfs: The effects of temperature, clouds, and gravity, The Astronomical Journal, 127, 3553. DOI:

Knapp, G., Leggett, S. K., Fan, X., et al. (2004)b, Near-infrared photometry and spectroscopy of L and T dwarfs: The effects of temperature, clouds, and gravity, The Astronomical Journal, 127, 3553. DOI:

Kokorev, V., Caputi, K. I., Greene, J. E., et al. (2024), A census of photometrically selected little red dots at 4< z< 9 in JWST blank fields, The Astrophysical Journal, 968, 38. DOI:

Koratkar, A. P. & Gaskell, C. M. (1991), Radius-luminosity and mass-luminosity relationships for active galactic nuclei, Astrophysical Journal, Part 2-Letters (ISSN 0004-637X), vol. 370, April 1, 1991, p. L61-L64., 370, L61. DOI:

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

Kozłowski, S., Bañados, E., Udalski, A., et al. (2019), Discovery of Two Quasars at z= 5 from the OGLE Survey, The Astrophysical Journal, 878, 115. DOI:

Kozłowski, S., Kochanek, C. S., Ashby, M. L., et al. (2016), Quasar variability in the mid-infrared, The Astrophysical Journal, 817, 119. DOI:

Kubota, A. & Done, C. (2018), A physical model of the broad-band continuum of AGN and its implications for the UV/X relation and optical variability, Monthly Notices of the Royal Astronomical Society, 480, 1247. DOI:

Lacy, M., Storrie-Lombardi, L. J., Sajina, A., et al. (2004), Obscured and Unobscured Active Galactic Nuclei in the Spitzer Space Telescope First Look Survey, ApJS, 154, 166. DOI:10.1086/422816

Lang, D. (2014), unWISE: unblurred coadds of the WISE imaging, The Astronomical Journal, 147, 108. DOI:

Lang, D., Hogg, D. W., & Mykytyn, D. (2016), The Tractor: Probabilistic astronomical source detection and measurement, Astrophysics Source Code Library, ascl. DOI:

Larson, R. L., Hutchison, T. A., Bagley, M., et al. (2023), Spectral Templates Optimal for Selecting Galaxies at z> 8 with the JWST, The Astrophysical Journal, 958, 141. DOI:

Lehmann, I., Hasinger, G., Schmidt, M., et al. (2001), The ROSAT Deep Survey-VI. X-ray sources and optical identifications of the ultra deep survey, Astronomy & Astrophysics, 371, 833. DOI:

Leja, J., Carnall, A. C., Johnson, B. D., Conroy, C., & Speagle, J. S. (2019), How to measure galaxy star formation histories. II. Nonparametric models, The Astrophysical Journal, 876, 3. DOI:

Leja, J., Johnson, B. D., Conroy, C., & Van Dokkum, P. (2018), Hot dust in panchromatic SED fitting: identification of active galactic nuclei and improved galaxy properties, The Astrophysical Journal, 854, 62. DOI:

Lemon, C., Anguita, T., Auger-Williams, M., et al. (2023), Gravitationally lensed quasars in Gaia–IV. 150 new lenses, quasar pairs, and projected quasars, Monthly Notices of the Royal Astronomical Society, 520, 3305. DOI:

Li, Z., McGreer, I. D., Wu, X.-B., Fan, X., & Yang, Q. (2018), The ensemble photometric variability of over 105 quasars in the Dark Energy Camera Legacy Survey and the Sloan Digital Sky Survey, The Astrophysical Journal, 861, 6. DOI:

Liu, M. C., Fischer, D. A., Graham, J. R., et al. (2002), Crossing the Brown Dwarf Desert using adaptive optics: a very close L Dwarf companion to the nearby solar analog HR 7672, The Astrophysical Journal, 571, 519. DOI:

Lodieu, N., Burningham, B., Day-Jones, A., et al. (2012), First T dwarfs in the VISTA Hemisphere Survey, Astronomy & Astrophysics, 548, A53. DOI:

Lodieu, N., Hambly, N., Jameson, R., et al. (2007), New brown dwarfs in Upper Sco using UKIDSS Galactic Cluster Survey science verification data, Monthly Notices of the Royal Astronomical Society, 374, 372. DOI:

Lodieu, N., Osorio, M. Z., Rebolo, R., Martín, E., & Hambly, N. (2009), A census of very-low-mass stars and brown dwarfs in the σ Orionis cluster, Astronomy & Astrophysics, 505, 1115. DOI:

Logan, C. & Fotopoulou, S. (2020), Unsupervised star, galaxy, QSO classification-Application of HDBSCAN, Astronomy & Astrophysics, 633, A154. DOI:

Lomb, N. R. (1976), Least-squares frequency analysis of unequally spaced data, Astrophysics and space science, 39, 447. DOI:

Lusso, E. & Risaliti, G. (2016), The tight relation between X-ray and ultraviolet luminosity of quasars, The Astrophysical Journal, 819, 154. DOI:

Lusso, E. & Risaliti, G. (2017), Quasars as standard candles-I. The physical relation between disc and coronal emission, Astronomy & Astrophysics, 602, A79. DOI:

Lyke, B. W., Higley, A. N., McLane, J., et al. (2020), The sloan digital sky survey quasar catalog: Sixteenth data release, The Astrophysical Journal Supplement Series, 250, 8. DOI:

Ma, Y., Greene, J. E., Volonteri, M., et al. (2025), No Luminous Little Red Dots: A Sharp Cutoff in Their Luminosity Function, arXiv e-prints, arXiv. DOI:

Mace, G. N., Kirkpatrick, J. D., Cushing, M. C., et al. (2013), A Study of the Diverse T Dwarf Population Revealed by WISE, The Astrophysical Journal Supplement Series, 205, 6. DOI:

MacLeod, C. L., Ivezić, Ž., Kochanek, C., et al. (2010), Modeling the time variability of SDSS stripe 82 quasars as a damped random walk, The Astrophysical Journal, 721, 1014. DOI:

Mainzer, A., Bauer, J., Cutri, R., et al. (2014), Initial performance of the NEOWISE reactivation mission, The Astrophysical Journal, 792, 30. DOI:

Maiolino, R., Oliva, E., Ghinassi, F., et al. (2004), Extreme gas properties in the most distant quasars, Astronomy & Astrophysics, 420, 889. DOI:

Marconi, A. & Hunt, L. K. (2003), The relation between black hole mass, bulge mass, and near-infrared luminosity, The Astrophysical Journal, 589, L21. DOI:

Marley, M., Saumon, D., Morley, C., & Fortney, J. (2018), Sonora 2018: Cloud-free, solar composition, solar C/O substellar atmosphere models and spectra, (No Title). DOI:

Marocco, F., Jones, H., Day-Jones, A., et al. (2015), A large spectroscopic sample of L and T dwarfs from UKIDSS LAS: peculiar objects, binaries, and space density, Monthly Notices of the Royal Astronomical Society, 449, 3651. DOI:

Marshall, A., Auger-Williams, M. W., Banerji, M., Maiolino, R., & Bowler, R. (2022), A fresh look at AGN spectral energy distribution fitting with the XMM-SERVS AGN sample, Monthly Notices of the Royal Astronomical Society, 515, 5617. DOI:

Martín, M. V., Perna, M., Humphrey, A., et al. (2020), Peculiar emission line spectra of core extremely red BOSS quasars at z 2–3: orientation and/or evolution?, Astronomy & Astrophysics, 634, A116. DOI:

Martin Gaskell, C. & Klimek, E. S. (2003), Variability of active galactic nuclei from the optical to X-Ray regions, Astronomical & Astrophysical Transactions, 22, 661. DOI:

Martini, P., Elias, J., Points, S., et al. (2014), in Ground-based and Airborne Instrumentation for Astronomy V, Vol. 9147, SPIE, 366–379 DOI:

Matejek, M. S. & Simcoe, R. A. (2012), A Survey of Mg II Absorption at 2< z< 6 with Magellan/FIRE. I. Sample and Evolution of the Mg II Frequency, The Astrophysical Journal, 761, 112. DOI:

Matsuoka, Y., Iwasawa, K., Onoue, M., et al. (2025)a, Subaru High-z Exploration of Low-luminosity Quasars (SHELLQs). XXIV. 54 New Quasars and Candidate Obscured Quasars at 5.71 z 7.02, The Astrophysical Journal Supplement Series, 280, 68. DOI:

Matsuoka, Y., Iwasawa, K., Onoue, M., et al. (2019), Subaru high-z exploration of low-luminosity quasars (SHELLQs). X. Discovery of 35 quasars and luminous galaxies at $5.7 \lesssim z \lesssim 7.0$, The Astrophysical Journal, 883, 183. DOI:

Matsuoka, Y., Onoue, M., Iwasawa, K., et al. (2025)b, SHELLQs. Bridging the gap: JWST unveils obscured quasars in the most luminous galaxies at z> 6, arXiv e-prints, arXiv. DOI:

Matsuoka, Y., Onoue, M., Kashikawa, N., et al. (2016), SUBARU HIGH-z EXPLORATION OF LOW-LUMINOSITY QUASARS (SHELLQs). I. DISCOVERY OF 15 QUASARS AND BRIGHT GALAXIES AT $5.7 < z < 6.9$, The Astrophysical Journal, 828, 26. DOI:

Matsuoka, Y., Peterson, B., Murata, K., et al. (2011), 1 μm Excess Sources in the UKIDSS. I. Three T Dwarfs in the Sloan Digital Sky Survey Southern Equatorial Stripe, The Astronomical Journal, 142, 64. DOI:

Matsuoka, Y., Strauss, M. A., Kashikawa, N., et al. (2018), Subaru high-z exploration of low-luminosity quasars (SHELLQs). V. quasar luminosity function and contribution to cosmic reionization at z= 6, The Astrophysical Journal, 869, 150. DOI:

Matt, G., Guainazzi, M., & Maiolino, R. (2003), Changing look: from Compton-thick to Compton-thin, or the rebirth of fossil active galactic nuclei, Monthly Notices of the Royal Astronomical Society, 342, 422. DOI:

Matthee, J., Naidu, R. P., Brammer, G., et al. (2024), Little red dots: an abundant population of faint active galactic nuclei at z 5 revealed by the EIGER and FRESCO JWST surveys, The Astrophysical Journal, 963, 129. DOI:

Mau, S., Drlica-Wagner, A., Bechtol, K., et al. (2019), A faint halo star cluster discovered in the Blanco Imaging of the Southern Sky Survey, The Astrophysical Journal, 875, 154. DOI:

Mazzucchelli, C., Bañados, E., Decarli, R., et al. (2017), No Overdensity of Lyman-Alpha Emitting Galaxies around a Quasar at z 5.7, The Astrophysical Journal, 834, 83. DOI:

Mazzucchelli, C., Bischetti, M., D'Odorico, V., et al. (2023), XQR-30: Black hole masses and accretion rates of 42 z $\gtrsim$ 6 quasars, Astronomy & Astrophysics, 676, A71. DOI:

McConnell, N. J. & Ma, C.-P. (2013), Revisiting the scaling relations of black hole masses and host galaxy properties, The Astrophysical Journal, 764, 184. DOI:

McGill, K. L., Woo, J.-H., Treu, T., & Malkan, M. A. (2008), Comparing and calibrating black hole mass estimators for distant active galactic nuclei, The Astrophysical Journal, 673, 703. DOI:

McGreer, I. D., Jiang, L., Fan, X., et al. (2013), The z= 5 quasar luminosity function from SDSS Stripe 82, The Astrophysical Journal, 768, 105. DOI:

McInnes, L., Healy, J., & Melville, J. (2018), Umap: Uniform manifold approximation and projection for dimension reduction, arXiv preprint arXiv:1802.03426. DOI:

McLaughlin, M., Mattox, J., Cordes, J., & Thompson, D. (1996), Variability of CGRO/EGRET gamma-ray sources, The Astrophysical Journal, 473, 763. DOI:

McLure, R. J. & Dunlop, J. S. (2004), The cosmological evolution of quasar black hole masses, Monthly Notices of the Royal Astronomical Society, 352, 1390. DOI:

McLure, R. J. & Jarvis, M. J. (2002), Measuring the black hole masses of high-redshift quasars, Monthly Notices of the Royal Astronomical Society, 337, 109. DOI:

McMahon, R. (2012), The VISTA Hemisphere Survey (VHS) science goals and status, Science from the Next Generation Imaging and Spectroscopic Surveys, 37. DOI:

McMullin, J., Diamond, P., McPherson, A., et al. (2020), in Ground-based and airborne telescopes viii, Vol. 11445, SPIE, 215–232 DOI:

Meisner, A., Lang, D., & Schlegel, D. (2018), Time-resolved WISE/NEOWISE Coadds, The Astronomical Journal, 156, 69. DOI:

Meisner, A. M., Caselden, D., Schlafly, E. F., & Kiwy, F. (2023), unTimely: a Full-sky, Time-domain unWISE Catalog, The Astronomical Journal, 165, 36. DOI:

Meisner, A. M., Schneider, A. C., Burgasser, A. J., et al. (2021), New Candidate Extreme T Subdwarfs from the Backyard Worlds: Planet 9 Citizen Science Project, The Astrophysical Journal, 915, 120. DOI:

Mejía-Restrepo, J., Trakhtenbrot, B., Lira, P., & Netzer, H. (2018), Can we improve C iv-based single-epoch black hole mass estimations?, Monthly Notices of the Royal Astronomical Society, 478, 1929. DOI:

Mejía-Restrepo, J. E., Trakhtenbrot, B., Koss, M. J., et al. (2022), BASS. XXV. DR2 Broad-line-based Black Hole Mass Estimates and Biases from Obscuration, The Astrophysical Journal Supplement Series, 261, 5. DOI:

Metchev, S. A., Heinze, A., Apai, D., et al. (2015), Weather on other worlds. II. Survey results: spots are ubiquitous on L and T dwarfs, The Astrophysical Journal, 799, 154. DOI:

Meyer, R. A., Bosman, S. E., & Ellis, R. S. (2019), New constraints on quasar evolution: broad-line velocity shifts over $1.5 \lesssim z \lesssim 7.5$, Monthly Notices of the Royal Astronomical Society, 487, 3305. DOI:

Meyer, R. A., Decarli, R., Walter, F., et al. (2022), Constraining galaxy overdensities around three z 6.5 quasars with ALMA and MUSE, The Astrophysical Journal, 927, 141. DOI:

Mignoli, M., Gilli, R., Decarli, R., et al. (2020), Web of the giant: Spectroscopic confirmation of a large-scale structure around the z= 6.31 quasar SDSS J1030+ 0524, Astronomy & Astrophysics, 642, L1. DOI:

Mineo, S., Gilfanov, M., Lehmer, B., Morrison, G., & Sunyaev, R. (2014), X-ray emission from star-forming galaxies–III. Calibration of the L X-SFR relation up to redshift z ≈ 1.3, Monthly Notices of the Royal Astronomical Society, 437, 1698. DOI:

Mitchell, J. A., Done, C., Ward, M. J., et al. (2023), The SOUX AGN sample: optical/UV/X-ray SEDs and the nature of the disc, Monthly Notices of the Royal Astronomical Society, 524, 1796. DOI:

Mohanty, S. & Basri, G. (2003), Rotation and activity in mid-M to L field dwarfs, The Astrophysical Journal, 583, 451. DOI:

Möller, A., Ishida, E., Peloton, J., et al. (2025), Real-Time Active Learning for optimised spectroscopic follow-up: Enhancing early SN Ia classification with the Fink broker, Publications of the Astronomical Society of Australia, 42, e057. DOI:

Molnár, D. C., Sargent, M. T., Leslie, S., et al. (2021), The non-linear infrared-radio correlation of low-z galaxies: implications for redshift evolution, a new radio SFR recipe, and how to minimize selection bias, Monthly Notices of the Royal Astronomical Society, 504, 118. DOI:

Momjian, E., Carilli, C. L., Bañados, E., Walter, F., & Venemans, B. P. (2018), Resolving the Powerful Radio-loud Quasar at z 6, The Astrophysical Journal, 861, 86. DOI:

Mor, R., Netzer, H., & Elitzur, M. (2009), Dusty structure around type-I active galactic nuclei: clumpy torus narrow-line region and near-nucleus hot dust, The Astrophysical Journal, 705, 298. DOI:

Morganti, R., Fogasy, J., Paragi, Z., Oosterloo, T., & Orienti, M. (2013), Radio jets clearing the way through a galaxy: watching feedback in action, Science, 341, 1082. DOI:

Morselli, L., Mignoli, M., Gilli, R., et al. (2014), Primordial environment of super massive black holes: large-scale galaxy overdensities around z˜ 6 quasars with LBT, Astronomy & Astrophysics, 568, A1. DOI:

Murphy, E., Condon, J., Schinnerer, E., et al. (2011), Calibrating extinction-free star formation rate diagnostics with 33 GHz free–free emission in NGC 6946, The Astrophysical Journal, 737, 67. DOI:

Nakoneczny, S., Bilicki, M., Solarz, A., et al. (2019), Catalog of quasars from the Kilo-Degree Survey Data Release 3, Astronomy & Astrophysics, 624, A13. DOI:

Nakoneczny, S., Graham, M., Stern, D., et al. (2025), QZO: A Catalog of 5 Million Quasars from the Zwicky Transient Facility, arXiv e-prints, arXiv. DOI:

Nandra, K. & Pounds, K. (1994), Ginga observations of the X-ray spectra of Seyfert galaxies, Monthly Notices of the Royal Astronomical Society, 268, 405. DOI:

Nardini, E., Reeves, J., Gofford, J., et al. (2015), Black hole feedback in the luminous quasar PDS 456, Science, 347, 860. DOI:

Neeleman, M., Bañados, E., Walter, F., et al. (2019), Resolved [C ii] Emission from z> 6 Quasar Host–Companion Galaxy Pairs, The Astrophysical Journal, 882, 10. DOI:

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

Netzer, H. (2015), Revisiting the unified model of active galactic nuclei, Annual Review of Astronomy and Astrophysics, 53, 365. DOI:

Niida, M., Nagao, T., Ikeda, H., et al. (2020), The Faint End of the Quasar Luminosity Function at z 5 from the Subaru Hyper Suprime-Cam Survey, The Astrophysical Journal, 904, 89. DOI:

Nun, I., Protopapas, P., Sim, B., et al. (2015), Fats: Feature analysis for time series, arXiv preprint arXiv:1506.00010. DOI:

Oke, J. & Gunn, J. (1982), An efficient low resolution and moderate resolution spectrograph for the Hale telescope, Publications of the Astronomical Society of the Pacific, 94, 586. DOI:

Onken, C. A., Lai, S., Wolf, C., et al. (2022), Discovery of the most luminous quasar of the last 9 Gyr, Publications of the Astronomical Society of Australia, 39, e037. DOI:

Onorato, S., Hennawi, J. F., Schindler, J.-T., et al. (2025), Optical and near-infrared spectroscopy of quasars at z> 6.5: public data release and composite spectrum, Monthly Notices of the Royal Astronomical Society, 540, 1308. DOI:

Overzier, R. A. (2022), Conditions for direct black hole seed collapse near a radio-loud quasar 1 Gyr after the Big Bang, The Astrophysical Journal, 926, 114. DOI:

Overzier, R. A., Guo, Q., Kauffmann, G., et al. (2009), ΛCDM predictions for galaxy protoclusters–I. The relation between galaxies, protoclusters and quasars at z 6, Monthly Notices of the Royal Astronomical Society, 394, 577. DOI:

O'neill, S., Kiehlmann, S., Readhead, A., et al. (2022), The unanticipated phenomenology of the blazar PKS 2131–021: a unique supermassive black hole binary candidate, The Astrophysical Journal Letters, 926, L35. DOI:

Padovani, P., Alexander, D., Assef, R., et al. (2017), Active galactic nuclei: what's in a name?, The Astronomy and Astrophysics Review, 25, 1. DOI:

Padovani, P. & Cirasuolo, M. (2023), The extremely large telescope, Contemporary Physics, 64, 47. DOI:

Palanque-Delabrouille, N., Yeche, C., Myers, A., et al. (2011), Variability selected high-redshift quasars on SDSS Stripe 82, Astronomy & Astrophysics, 530, A122. DOI:

Panessa, F., Baldi, R. D., Laor, A., et al. (2019), The origin of radio emission from radio-quiet active galactic nuclei, Nature Astronomy, 3, 387. DOI:10.1038/s41550-019-0765-4

Papadakis, I. (2004), The scaling of the X-ray variability with black hole mass in active galactic nuclei, Monthly Notices of the Royal Astronomical Society, 348, 207. DOI:

Park, D., Barth, A. J., Woo, J.-H., et al. (2017), Extending the calibration of C IV-based single-epoch black hole mass estimators for active galactic nuclei, The Astrophysical Journal, 839, 93. DOI:

Park, D., Woo, J.-H., Treu, T., et al. (2012), The lick AGN monitoring project: recalibrating single-epoch virial black hole mass estimates, The Astrophysical Journal, 747, 30. DOI:

Partridge, H. & Schwenke, D. W. (1997), The determination of an accurate isotope dependent potential energy surface for water from extensive ab initio calculations and experimental data, The Journal of Chemical Physics, 106, 4618. DOI:

Patel, P., Lira, P., Arévalo, P., et al. (2025), Probing the rest-frame wavelength dependence of quasar variability-Insights from the Zwicky Transient Facility Survey, Astronomy & Astrophysics, 695, A162. DOI:

Pearson, K. (1901), Principal components analysis, The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, 6, 559. DOI:

Peters, C. M., Richards, G. T., Myers, A. D., et al. (2015), Quasar classification using color and variability, The Astrophysical Journal, 811, 95. DOI:

Peterson, B. M. (1993), Reverberation mapping of active galactic nuclei, Publications of the Astronomical Society of the Pacific, 105, 247. DOI:

Peterson, B. M., Ferrarese, L., Gilbert, K., et al. (2004), Central masses and broad-line region sizes of active galactic nuclei. II. A homogeneous analysis of a large reverberation-mapping database, The Astrophysical Journal, 613, 682. DOI:

Petrecca, V., Papadakis, I., Paolillo, M., De Cicco, D., & Bauer, F. (2024), Ensemble power spectral density of SDSS quasars in UV/optical bands, Astronomy & Astrophysics, 686, A286. DOI:

Pichara, K., Protopapas, P., Kim, D.-W., Marquette, J.-B., & Tisserand, P. (2012), An improved quasar detection method in EROS-2 and MACHO LMC data sets, Monthly Notices of the Royal Astronomical Society, 427, 1284. DOI:

Pickles, A. (1998)a, A stellar spectral flux library: 1150–25000 Å, Publications of the Astronomical Society of the Pacific, 110, 863. DOI:

Pickles, A. J. (1998)b, A Stellar Spectral Flux Library: 1150–25000 Å, Publications of the Astronomical Society of the Pacific, 110, 863. DOI:10.1086/316197

Pinfield, D., Burningham, B., Tamura, M., et al. (2008), Fifteen new T dwarfs discovered in the UKIDSS Large Area Survey, Monthly Notices of the Royal Astronomical Society, 390, 304. DOI:

Pogge, R., Atwood, B., Brewer, D., et al. (2010), in Ground-based and Airborne Instrumentation for Astronomy III, Vol. 7735, SPIE, 159–166 DOI:

Polletta, M., Courvoisier, T.-L., Hooper, E., & Wilkes, B. J. (2000), The Far-Infrared emission of Radio Loud and Radio Quiet Quasars, Astronomy and Astrophysics. DOI:

Prevot, M., Lequeux, J., Maurice, E., Prevot, L., & Rocca-Volmerange, B. (1984), The typical interstellar extinction in the Small Magellanic Cloud, Astronomy and Astrophysics (ISSN 0004-6361), vol. 132, no. 2, March 1984, p. 389-392., 132, 389. DOI:

Prochaska, J., Hennawi, J., Westfall, K., et al. (2020), PypeIt: The Python Spectroscopic Data Reduction Pipeline, The Journal of Open Source Software, 5, 2308. DOI:

Radigan, J., Jayawardhana, R., Lafrenière, D., et al. (2013), Discovery of a visual t-dwarf triple system and binarity at the l/t transition, The Astrophysical Journal, 778, 36. DOI:

Rakshit, S. & Stalin, C. (2017), Optical variability of narrow-line and broad-line Seyfert 1 Galaxies, The Astrophysical Journal, 842, 96. DOI:

Reed, S., McMahon, R., Martini, P., et al. (2017), Eight new luminous z $\geq$ 6 quasars discovered via SED model fitting of VISTA, WISE and Dark Energy Survey Year 1 observations, Monthly Notices of the Royal Astronomical Society, 468, 4702. DOI:

Richards, G. T., Kruczek, N. E., Gallagher, S., et al. (2011)a, Unification of luminous type 1 quasars through C IV emission, The Astronomical Journal, 141, 167. DOI:

Richards, G. T., Lacy, M., Storrie-Lombardi, L. J., et al. (2006)a, Spectral energy distributions and multiwavelength selection of type 1 quasars, The Astrophysical Journal Supplement Series, 166, 470. DOI:

Richards, G. T., Strauss, M. A., Fan, X., et al. (2006)b, The sloan digital sky survey quasar survey: Quasar luminosity function from data release 3, The Astronomical Journal, 131, 2766. DOI:

Richards, G. T., Vanden Berk, D. E., Reichard, T. A., et al. (2002), Broad Emission-Line Shifts in Quasars: An Orientation Measurefor Radio-Quiet Quasars?, The Astronomical Journal, 124, 1. DOI:

Richards, J. W., Starr, D. L., Butler, N. R., et al. (2011)b, On machine-learned classification of variable stars with sparse and noisy time-series data, The Astrophysical Journal, 733, 10. DOI:

Risaliti, G. & Lusso, E. (2019), Cosmological constraints from the Hubble diagram of quasars at high redshifts, Nature Astronomy, 3, 272. DOI:

Ross, N. P., Ford, K. S., Graham, M., et al. (2018), A new physical interpretation of optical and infrared variability in quasars, Monthly Notices of the Royal Astronomical Society, 480, 4468. DOI:

Rumbaugh, N., Shen, Y., Morganson, E., et al. (2018), Extreme Variability Quasars from the Sloan Digital Sky Survey and the Dark Energy Survey, The Astrophysical Journal, 854, 160. DOI:

Ryan-Weber, E. V., Pettini, M., & Madau, P. (2006), Intergalactic C IV absorption at redshifts 5.4 to 6, Monthly Notices of the Royal Astronomical Society: Letters, 371, L78. DOI:

Salvato, M., Wolf, J., Dwelly, T., et al. (2022), The eROSITA Final Equatorial-Depth Survey (eFEDS)-Identification and characterization of the counterparts to point-like sources, Astronomy & Astrophysics, 661, A3. DOI:

Sánchez-Sáez, P., Arredondo, J., Bayo, A., et al. (2023), Persistent and occasional: Searching for the variable population of the ZTF/4MOST sky using ZTF Data Release 11, Astronomy & Astrophysics, 675, A195. DOI:

Sánchez-Sáez, P., Lira, P., Mejía-Restrepo, J., et al. (2018), The QUEST–La Silla AGN Variability Survey: Connection between AGN Variability and Black Hole Physical Properties, The Astrophysical Journal, 864, 87. DOI:

Sánchez-Sáez, P., Reyes, I., Valenzuela, C., et al. (2021), Alert classification for the ALeRCE broker system: The light curve classifier, The Astronomical Journal, 161, 141. DOI:

Sarmiento, R., Knapen, J. H., Sánchez, S. F., et al. (2021), Capturing the physics of MaNGA galaxies with self-supervised Machine Learning, The Astrophysical Journal, 921, 177. DOI:

Saunders, W., Bridges, T., Gillingham, P., et al. (2004), in Ground-based Instrumentation for Astronomy, Vol. 5492, SPIE, 389–400 DOI:

Sbarrato, T., Ghisellini, G., Giovannini, G., & Giroletti, M. (2021), Jetted radio-quiet quasars at z> 5, Astronomy & Astrophysics, 655, A95. DOI:

Scaramella, R., Amiaux, J., Mellier, Y., et al. (2022), Euclid preparation-i. the euclid wide survey, Astronomy & Astrophysics, 662, A112. DOI:

Scargle, J. D. (1982), Studies in astronomical time series analysis. II-Statistical aspects of spectral analysis of unevenly spaced data, Astrophysical Journal, Part 1, vol. 263, Dec. 15, 1982, p. 835-853., 263, 835. DOI:

Schaerer, D. & de Barros, S. (2009), The impact of nebular emission on the ages of z≈ 6 galaxies, A&A, 502, 423. DOI:10.1051/0004-6361/200911781

Schindler, J.-T., Bañados, E., Connor, T., et al. (2023), The Pan-STARRS1 z> 5.6 Quasar Survey. III. The z ≈ 6 Quasar Luminosity Function, The Astrophysical Journal, 943, 67. DOI:

Schindler, J.-T., Fan, X., Huang, Y.-H., et al. (2019), The Extremely Luminous Quasar Survey in the Pan-STARRS 1 Footprint (PS-ELQS), The Astrophysical Journal Supplement Series, 243, 5. DOI:

Schindler, J.-T., Farina, E. P., Bañados, E., et al. (2020), The X-SHOOTER/ALMA sample of quasars in the epoch of reionization. I. NIR spectral modeling, iron enrichment, and broad emission line properties, The Astrophysical Journal, 905, 51. DOI:

Schlieder, J. E., Barclay, T., Barnes, A., et al. (2024), in Space Telescopes and Instrumentation 2024: Optical, Infrared, and Millimeter Wave, Vol. 13092, SPIE, 159–184 DOI:

Schmidt, K. B., Marshall, P. J., Rix, H.-W., et al. (2010)a, Selecting quasars by their intrinsic variability, The Astrophysical Journal, 714, 1194. DOI:

Schmidt, M. & Green, R. F. (1983), Quasar evolution derived from the Palomar bright quasar survey and other complete quasar surveys, Astrophysical Journal, Part 1 (ISSN 0004-637X), vol. 269, June 15, 1983, p. 352-374., 269, 352. DOI:

Schmidt, S. J., West, A. A., Bochanski, J. J., Hawley, S. L., & Kielty, C. (2014), Calibrating Ultracool Dwarfs: Optical Template Spectra, Bolometric Corrections, and χ Values1, Publications of the Astronomical Society of the Pacific, 126, 642. DOI:

Schmidt, S. J., West, A. A., Hawley, S. L., & Pineda, J. S. (2010)b, Colors and kinematics of l dwarfs from the sloan digital sky survey, The Astronomical Journal, 139, 1808. DOI:

Schneider, D. P., Hall, P. B., Richards, G. T., et al. (2005), The sloan digital sky survey quasar catalog. III. Third data release, The Astronomical Journal, 130, 367. DOI:

Scholz, R.-D. (2010)a, Hip 63510C, Hip 73786B, and nine new isolated high proper motion T dwarf candidates from UKIDSS DR6 and SDSS DR7, Astronomy & Astrophysics, 515, A92. DOI:

Scholz, R.-D. (2010)b, ULAS J141623. 94+ 134836.3–a faint common proper motion companion of a nearby L dwarf-Serendipitous discovery of a cool brown dwarf in UKIDSS DR6, Astronomy & Astrophysics, 510, L8. DOI:

Schreiber, C., Elbaz, D., Pannella, M., et al. (2018), Dust temperature and mid-to-total infrared color distributions for star-forming galaxies at 0< z< 4, Astronomy & Astrophysics, 609, A30. DOI:

Schryber, J. H., Miller, S., & Tennyson, J. (1995), Computed infrared absorption properties of hot water vapour, Journal of Quantitative Spectroscopy and Radiative Transfer, 53, 373. DOI:

Seyfert, C. K. (1943), Nuclear Emission in Spiral Nebulae., Astrophysical Journal, vol. 97, p. 28, 97, 28. DOI:

Sharp, R., Saunders, W., Smith, G., et al. (2006), in Ground-based and Airborne Instrumentation for Astronomy, Vol. 6269, SPIE, 152–164 DOI:

Shen, Y., Greene, J. E., Strauss, M. A., Richards, G. T., & Schneider, D. P. (2008), Biases in virial black hole masses: an SDSS perspective, The Astrophysical Journal, 680, 169. DOI:

Shim, H., Chary, R.-R., Dickinson, M., et al. (2011), z ∼4 Hα Emitters in the Great Observatories Origins Deep Survey: Tracing the Dominant Mode for Growth of Galaxies, ApJ, 738, 69. DOI:10.1088/0004-637X/738/1/69

Silpa, S., Kharb, P., Harrison, C., et al. (2022), The Quasar Feedback Survey: revealing the interplay of jets, winds, and emission-line gas in type 2 quasars with radio polarization, Monthly Notices of the Royal Astronomical Society, 513, 4208. DOI:

Silva, D. R., Blum, R. D., Allen, L., et al. (2016), in American Astronomical Society Meeting Abstracts# 228, Vol. 228, 317–02 DOI:

Silva, L., Maiolino, R., & Granato, G. L. (2004), Connecting the cosmic infrared background to the X-ray background, Monthly Notices of the Royal Astronomical Society [[arXiv]0403381]. DOI:10.1111/j.1365-2966.2004.08380.x

Simm, T., Salvato, M., Saglia, R., et al. (2016), Pan-STARRS1 variability of XMM-COSMOS AGN-II. Physical correlations and power spectrum analysis, Astronomy & Astrophysics, 585, A129. DOI:

Simpson, C., Mortlock, D., Warren, S., et al. (2014), No excess of bright galaxies around the redshift 7.1 quasar ULAS J1120+ 0641, Monthly Notices of the Royal Astronomical Society, 442, 3454. DOI:

Slone, O. & Netzer, H. (2012), The effects of disc winds on the spectrum and black hole growth rate of active galactic nuclei, Monthly Notices of the Royal Astronomical Society, 426, 656. DOI:

Smith, D., Jarvis, M., Hardcastle, M., et al. (2014), The temperature dependence of the

far-infrared–radio correlation in the Herschel-ATLAS, Monthly Notices of the Royal Astronomical Society, 445, 2232. DOI:

Spingola, C., Dallacasa, D., Belladitta, S., et al. (2020), Parsec-scale properties of the radio brightest jetted AGN at z> 6, Astronomy & Astrophysics, 643, L12. DOI:

Springel, V., Di Matteo, T., & Hernquist, L. (2005), Modelling feedback from stars and black holes in galaxy mergers, Monthly Notices of the Royal Astronomical Society, 361, 776. DOI:

Stalevski, M., Ricci, C., Ueda, Y., et al. (2016), The dust covering factor in active galactic nuclei, Monthly Notices of the Royal Astronomical Society [[arXiv]1602.06954]. DOI:10.1093/mnras/stw444

Stalevski, M., Tristram, K. R., & Asmus, D. (2019), Dissecting the active galactic nucleus in Circinus–II. A thin dusty disc and a polar outflow on parsec scales, Monthly Notices of the Royal Astronomical Society, 484, 3334. DOI:

Steidel, C. C., Adelberger, K. L., Giavalisco, M., Dickinson, M., & Pettini, M. (1999), Lyman-break galaxies at z 4 and the evolution of the ultraviolet luminosity density at high redshift, The astrophysical journal, 519, 1. DOI:

Stern, D., Eisenhardt, P., Gorjian, V., et al. (2005), Mid-Infrared Selection of Active Galaxies, ApJ, 631, 163. DOI:10.1086/432523

Strateva, I. V., Brandt, W., Schneider, D. P., Berk, D. G. V., & Vignali, C. (2005), Soft X-ray and ultraviolet emission relations in optically selected AGN samples, The Astronomical Journal, 130, 387. DOI:

Sun, M., Xue, Y., Brandt, W., et al. (2020), Corona-heated accretion-disk reprocessing: a physical model to decipher the melody of AGN UV/optical twinkling, The Astrophysical Journal, 891, 178. DOI:

Szokoly, G. P., Bergeron, J., Hasinger, G., et al. (2004), The Chandra Deep Field-South: Optical Spectroscopy. I., The Astrophysical Journal Supplement Series, 155, 271. DOI:

Tanimoto, A., Ueda, Y., Odaka, H., et al. (2019), XCLUMPY: X-Ray Spectral Model from Clumpy Torus and Its Application to the Circinus Galaxy, The Astrophysical Journal, 877, 95. DOI:

Temple, M. J., Banerji, M., Hewett, P. C., Rankine, A. L., & Richards, G. T. (2021)a, Exploring the link between C IV outflow kinematics and sublimation-temperature dust in quasars, Monthly Notices of the Royal Astronomical Society, 501, 3061. DOI:

Temple, M. J., Hewett, P. C., & Banerji, M. (2021)b, Modelling type 1 quasar colours in the era of Rubin and Euclid, Monthly Notices of the Royal Astronomical Society, 508, 737. DOI:

Thompson, M. A., Kirkpatrick, J. D., Mace, G. N., et al. (2013), Nearby M, L, and T Dwarfs Discovered by the Wide-field Infrared Survey Explorer (WISE), Publications of the Astronomical Society of the Pacific, 125, 809. DOI:

Thorne, J. E., Robotham, A. S., Davies, L. J., et al. (2022), Deep Extragalactic VIsible Legacy Survey (DEVILS): identification of AGN through SED fitting and the evolution of the bolometric AGN luminosity function, Monthly Notices of the Royal Astronomical Society, 509, 4940. DOI:

Thorne, J. E., Robotham, A. S. G., Davies, L. J. M., et al. (2022), Deep Extragalactic VIsible Legacy Survey (DEVILS): identification of AGN through SED fitting and the evolution of the bolometric AGN luminosity function, MNRAS, 509, 4940. DOI:10.1093/mnras/stab3208

Thuruthipilly, H., Junais, J., Pollo, A., et al. (2024), Shedding light on low-surface-brightness galaxies in dark energy surveys with transformer models, Astronomy & Astrophysics, 682, A4. DOI:

Tie, S. S., Martini, P., Mudd, D., et al. (2017), A study of quasar selection in the supernova fields of the Dark Energy Survey, The Astronomical Journal, 153, 107. DOI:

Tinney, C., Kirkpatrick, J. D., Faherty, J. K., et al. (2018), New Y and T Dwarfs from WISE Identified by Methane Imaging, The Astrophysical Journal Supplement Series, 236, 28. DOI:

Tinney, C. & Tolley, A. (1999), Searching for weather in brown dwarfs, Monthly Notices of the Royal Astronomical Society, 304, 119. DOI:

Tody, D. (1986), in Instrumentation in astronomy VI, Vol. 627, SPIE, 733–748 DOI:

Tody, D. (1993), in Astronomical Data Analysis Software and Systems II, Vol. 52, 173 DOI:

Tombesi, F., Meléndez, M., Veilleux, S., et al. (2015), Wind from the black-hole accretion disk driving a molecular outflow in an active galaxy, Nature, 519, 436. DOI:

Trakhtenbrot, B. & Netzer, H. (2012), Black hole growth to z= 2–I. Improved virial methods for measuring M BH and L/L Edd, Monthly Notices of the Royal Astronomical Society, 427, 3081. DOI:

Travascio, A., Piconcelli, E., Bischetti, M., et al. (2024), MUSE view of PDS 456: Kiloparsec-scale wind, extended ionized gas, and close environment, Astronomy & Astrophysics, 686, A250. DOI:

Tremaine, S., Gebhardt, K., Bender, R., et al. (2002), The slope of the black hole mass versus velocity dispersion correlation, The Astrophysical Journal, 574, 740. DOI:

Tristram, K. R., Burtscher, L., Jaffe, W., et al. (2014), The dusty torus in the Circinus galaxy: a dense disk and the torus funnel, Astronomy & Astrophysics, 563, A82. DOI:

Übler, H., Maiolino, R., Curtis-Lake, E., et al. (2023), GA-NIFS: A massive black hole in a low-metallicity AGN at z 5.55 revealed by JWST/NIRSpec IFS, Astronomy & Astrophysics, 677, A145. DOI:

Übler, H., Maiolino, R., Pérez-González, P. G., et al. (2024), GA-NIFS: JWST discovers an offset AGN 740 million years after the big bang, Monthly Notices of the Royal Astronomical Society, 531, 355. DOI:

Ulrich, M.-H., Maraschi, L., & Urry, C. M. (1997), Variability of active galactic nuclei, Annual Review of Astronomy and Astrophysics, 35, 445. DOI:

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

Utsumi, Y., Goto, T., Kashikawa, N., et al. (2010), A Large Number of z> 6 Galaxies Around a QSO at z= 6.43: Evidence for a Protocluster?, The Astrophysical Journal, 721, 1680. DOI:

Uttley, P., McHardy, I., & Papadakis, I. (2002), Measuring the broad-band power spectra of active galactic nuclei with RXTE, Monthly Notices of the Royal Astronomical Society, 332, 231. DOI:

Valtonen, M., Kidger, M., Lehto, H., & Poyner, G. (2008), The structure of the October/November 2005 outburst in OJ287 and the precessing binary black hole model, Astronomy & Astrophysics, 477, 407. DOI:

van Haarlem, M. P., Wise, M. W., Gunst, A., et al. (2013), LOFAR: The low-frequency array, Astronomy & astrophysics, 556, A2. DOI:

VanderPlas, J. T. (2018), Understanding the lomb–scargle periodogram, The Astrophysical Journal Supplement Series, 236, 16. DOI:

Vega-Ferrero, J., Costantin, L., Pérez-González, P. G., et al. (2024), On the nature of disks at high redshift seen by JWST/CEERS with contrastive learning and cosmological simulations, The Astrophysical Journal, 961, 51. DOI:

Venemans, B. P., Walter, F., Neeleman, M., et al. (2020), Kiloparsec-scale ALMA Imaging of [C II] and Dust Continuum Emission of 27 Quasar Host Galaxies at z 6, The Astrophysical Journal, 904, 130. DOI:

Venturi, G., Treister, E., Finlez, C., et al. (2023), Complex AGN feedback in the Teacup galaxy-A powerful ionised galactic outflow, jet-ISM interaction, and evidence for AGN-triggered star formation in a giant bubble, Astronomy & Astrophysics, 678, A127. DOI:

Verbeek, K., de Groot, E., Groot, P. J., et al. (2012), A first catalogue of automatically selected ultraviolet-excess sources from the UVEX survey, Monthly Notices of the Royal Astronomical Society, 420, 1115. DOI:

Véron-Cetty, M.-P. & Véron, P. (2010), A catalogue of quasars and active nuclei, Astronomy & Astrophysics, 518, A10. DOI:

Vestergaard, M. (2002), Determining central black hole masses in distant active galaxies, The Astrophysical Journal, 571, 733. DOI:

Vestergaard, M. & Osmer, P. S. (2009), Mass functions of the active black holes in distant quasars from the large bright quasar survey, the bright quasar survey, and the color-selected sample of the SDSS fall equatorial stripe, The Astrophysical Journal, 699, 800. DOI:

Vestergaard, M. & Peterson, B. M. (2006), Determining central black hole masses in distant active galaxies and quasars. II. Improved optical and UV scaling relationships, The Astrophysical Journal, 641, 689. DOI:

Wada, K. (2012), Radiation-driven fountain and origin of torus around active galactic nuclei, The Astrophysical Journal, 758, 66. DOI:

Wagenveld, J., Saxena, A., Duncan, K., Röttgering, H., & Zhang, M. (2022), Revealing new high-redshift quasar populations through Gaussian mixture model selection, Astronomy & Astrophysics, 660, A22. DOI:

Wang, F., Fan, X., Yang, J., et al. (2017), First discoveries of z> 6 quasars with the DECam Legacy Survey and UKIRT hemisphere survey, The Astrophysical Journal, 839, 27. DOI:

Wang, F., Wu, X.-B., Fan, X., et al. (2016)a, A survey of luminous high-redshift quasars with SDSS and WISE. I. Target selection and optical spectroscopy, The Astrophysical Journal, 819, 24. DOI:

Wang, F., Yang, J., Fan, X., et al. (2019), Exploring reionization-era quasars. III. Discovery of 16 quasars at 6.4 $\lesssim$ z $\lesssim$ 6.9 with DESI Legacy Imaging Surveys and the UKIRT Hemisphere Survey and quasar luminosity function at z 6.7, The Astrophysical Journal, 884, 30. DOI:

Wang, F., Yang, J., Hennawi, J. F., et al. (2023), A SPectroscopic survey of biased halos In the Reionization Era (ASPIRE): JWST Reveals a Filamentary Structure around az= 6.61 Quasar, The Astrophysical Journal Letters, 951, L4. DOI:

Wang, R., Wu, X.-B., Neri, R., et al. (2016)b, Probing the interstellar medium and star formation of the Most Luminous Quasar at z= 6.3, The Astrophysical Journal, 830, 53. DOI:

Warren, S., Hewett, P., Irwin, M., et al. (1987), First observation of a quasar with a redshift of 4, Nature, 325, 131. DOI:

Webb, N., Coriat, M., Traulsen, I., et al. (2020), The XMM-Newton serendipitous survey-IX. The fourth XMM-Newton serendipitous source catalogue, Astronomy & Astrophysics, 641, A136. DOI:

Wenzl, L., Schindler, J.-T., Fan, X., et al. (2021), Random forests as a viable method to select and discover high-redshift quasars, The Astronomical Journal, 162, 72. DOI:

West, A. A., Morgan, D. P., Bochanski, J. J., et al. (2011), The sloan digital sky survey data release 7 spectroscopic M dwarf catalog. I. Data, The Astronomical Journal, 141, 97. DOI:

Wilhite, B. C., Brunner, R. J., Grier, C. J., Schneider, D. P., & Vanden Berk, D. E. (2008), On the variability of quasars: a link between the Eddington ratio and optical variability?, Monthly Notices of the Royal Astronomical Society, 383, 1232. DOI:

Willott, C. J., Delorme, P., Omont, A., et al. (2007), Four quasars above redshift 6 discovered by the Canada-France High-z Quasar Survey, The Astronomical Journal, 134, 2435. DOI:

Willott, C. J., McLure, R. J., & Jarvis, M. J. (2003), A 3× 109 M⊙ Black Hole in the Quasar SDSS J1148+ 5251 at z= 6.41, The Astrophysical Journal, 587, L15. DOI:

Wisotzki, L., Köhler, T., Groote, D., & Reimers, D. (1996), The Hamburg/ESO survey for bright QSOs. I. Survey design and candidate selection procedure., Astronomy and Astrophysics Supplement, v. 115, p. 227, 115, 227. DOI:

Wolf, J., Salvato, M., Belladitta, S., et al. (2024), The SRG/eROSITA All-Sky Survey: X-ray beacons at late cosmic dawn, Astronomy & Astrophysics, 691, A30. DOI:

Worseck, G. & Prochaska, J. X. (2011), GALEX far-ultraviolet color selection of UV-bright high-redshift quasars, The Astrophysical Journal, 728, 23. DOI:

Wright, E. L., Eisenhardt, P. R., Mainzer, A. K., et al. (2010), The Wide-field Infrared Survey Explorer (WISE): mission description and initial on-orbit performance, The Astronomical Journal, 140, 1868. DOI:

Yang, D.-M., Schindler, J.-T., Nanni, R., et al. (2024), High-z quasar candidate archive: a spectroscopic catalogue of quasars and contaminants in various quasar searches, Monthly Notices of the Royal Astronomical Society, 528, 2679. DOI:

Yang, G., Boquien, M., Buat, V., et al. (2020), X-Cigale: Fitting AGN/galaxy SEDs from X-ray to infrared, Monthly Notices of the Royal Astronomical Society. DOI:10.1093/mnras/stz3001

Yang, J., Fan, X., Gupta, A., et al. (2023), Desi z $\gtrsim$ 5 quasar survey. i. a first sample of 400 new quasars at z $\sim$ 4.7–6.6, The Astrophysical Journal Supplement Series, 269, 27. DOI:

Yang, J., Fan, X., Wu, X.-B., et al. (2017), Discovery of 16 new z $\sim$ 5.5 quasars: filling in the redshift gap of quasar color selection, The Astronomical Journal, 153, 184. DOI:

Yang, J., Wang, F., Fan, X., et al. (2019)a, Filling in the Quasar Redshift Gap at z 5.5. II. A Complete Survey of Luminous Quasars in the Post-reionization Universe, The Astrophysical Journal, 871, 199. DOI:

Yang, J., Wang, F., Fan, X., et al. (2019)b, Exploring reionization-era quasars. IV. Discovery of six new z $\gtrsim$ 6.5 quasars with DES, VHS, and unWISE photometry, The Astronomical Journal, 157, 236. DOI:

Yang, Q. & Shen, Y. (2022), A southern photometric quasar catalog from the Dark Energy Survey Data Release 2, The Astrophysical Journal Supplement Series, 264, 9. DOI:

Yèche, C., Palanque-Delabrouille, N., Claveau, C.-A., et al. (2020), Preliminary Target Selection for the DESI Quasar (QSO) Sample, Research Notes of the AAS, 4, 179. DOI:

York, D., Adelman, J., Anderson Jr, J., et al. (2000), others 2000. The sloan digital sky survey: Technical summary, The Astronomical Journal, 120, 1579. DOI:

Yue, M., Fan, X., Yang, J., & Wang, F. (2021), A candidate kiloparsec-scale quasar pair at z= 5.66, The Astrophysical Journal Letters, 921, L27. DOI:

Yue, M., Fan, X., Yang, J., & Wang, F. (2023), A Survey for High-redshift Gravitationally Lensed Quasars and Close Quasar Pairs. I. The Discoveries of an Intermediately Lensed Quasar and a Kiloparsec-scale Quasar Pair at z 5, The Astronomical Journal, 165, 191. DOI:

Yun, M. S., Reddy, N. A., & Condon, J. (2001), Radio properties of infrared-selected galaxies in the IRAS 2 Jy sample, The Astrophysical Journal, 554, 803. DOI:

Zajaček, M., Busch, G., Valencia-S, M., et al. (2019), Radio spectral index distribution of SDSS-FIRST sources across optical diagnostic diagrams, Astronomy & Astrophysics, 630, A83. DOI:

Zenteno, A., Kluge, M., Kharkrang, R., et al. (2025), eROSITA clusters dynamical state and their impact on the BCG luminosity, arXiv e-prints, arXiv. DOI:

Zou, F., Brandt, W., Chen, C.-T., et al. (2022), Spectral Energy Distributions in Three Deep-drilling Fields of the Vera C. Rubin Observatory Legacy Survey of Space and Time: Source Classification and Galaxy Properties, The Astrophysical Journal Supplement Series, 262, 15. DOI:

Zou, H., Zhou, X., Fan, X., et al. (2017), Project overview of the beijing–arizona sky survey, Publications of the Astronomical Society of the Pacific, 129, 064101. DOI:

Zu, Y., Kochanek, C., Kozłowski, S., & Udalski, A. (2013), Is quasar optical variability a damped random walk?, The Astrophysical Journal, 765, 106. DOI: