The Astrophysical Journal, 1010:224 (32pp), 2026 October 20 https://doi.org/10.3847/1538-4357/aea57b
© 2026. The Author(s). Published by the American Astronomical Society.
OPEN ACCESS

An Exploratory Analysis of New Large Gaia-informed Quasar Samples in SDSS-V

Shir Aviram[1], Benny Trakhtenbrot[1], Tom Dwelly[2], Scott F. Anderson[3], Sean Morrison[4], Michael Eracleous[5,6], Yue Shen[4,7], Mara Salvato[2], Donald P. Schneider[5,6], Roberto J. Assef[8], Catarina Aydar[2,9], Franz E. Bauer[10], W. N. Brandt[5,6,11], Joel R. Brownstein[12], Johannes Buchner[2], Jeremy Darling[13], José G. Fernández-Trincado[14], Patrick B. Hall[15], Dong-Woo Kim[16], Anton M. Koekemoer[17], Stephanie LaMassa[17], Andrea Merloni[2], Claudio Ricci[8,18,19], Qian Yang[16], and Grisha Zeltyn[1]

[1] School of Physics and Astronomy, Tel Aviv University, Tel Aviv 69978, Israel; shiraviram@mail.tau.ac.il
[2] Max-Planck-Institut für extraterrestrische Physik, Gießenbachstraße 1, 85748 Garching, Germany
[3] Astronomy Department, University of Washington, Box 351580, Seattle, WA 98195, USA
[4] Department of Astronomy, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA
[5] Department of Astronomy & Astrophysics, The Pennsylvania State University, University Park, PA 16802, USA
[6] Institute for Gravitation and the Cosmos, The Pennsylvania State University, University Park, PA 16802, USA
[7] National Center for Supercomputing Applications, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA
[8] Instituto de Estudios Astrofísicos, Facultad de Ingeniería y Ciencias, Universidad Diego Portales, Av. Ejército Libertador 441, Santiago, Chile
[9] Excellence Cluster ORIGINS, Boltzmannsstraße 2, 85748, Garching, Germany
[10] Instituto de Alta Investigación, Universidad de Tarapacá, Casilla 7D, Arica, 1010000, Chile
[11] Department of Physics, 104 Davey Lab, The Pennsylvania State University, University Park, PA 16802, USA
[12] Department of Physics and Astronomy, University of Utah, 270 S. 1400 E. E2108, Salt Lake City, UT 84112, USA
[13] Center for Astrophysics and Space Astronomy, Department of Astrophysical and Planetary Sciences, University of Colorado, 389 UCB, Boulder, CO 80309-0389, USA
[14] Centro de investigación en Astronomía, Facultad de Ingeniería, Ciencia y Tecnología, Universidad Bernardo O'Higgins, Av. Viel 1497, Santiago, 8370993, Chile
[15] Department of Physics and Astronomy, York University, 4700 Keele St., Toronto, ON M3J 1P3, Canada
[16] Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, MA 02138, USA
[17] Space Telescope Science Institute, 3700 San Martin Drive, Baltimore, MD 21218, USA
[18] Department of Astronomy, University of Geneva, ch. d'Ecogia 16, 1290, Versoix, Switzerland
[19] Kavli Institute for Astronomy and Astrophysics, Peking University, Beijing 100871, People's Republic of China
Received 2026 May 18; revised 2026 July 28; accepted 2026 August 19; published 2026 October 19

Abstract

Quasars are extremely luminous objects that provide insights into the physics and evolution of supermassive black holes (SMBHs) and their accretion flows, galaxy evolution, and even cosmology. In this study, we present an exploratory study based on the ongoing fifth generation of the Sloan Digital Sky Survey (SDSS-V) and its unique dual-hemisphere, wide-field, and multiobject spectroscopic capabilities, with the aim of creating a comprehensive, all-sky quasar sample. The targets were selected through two novel methods, Gaia–unWISE AGN (GUA) and Skewt_QSO, that rely primarily on data from the Wide-field Infrared Survey Explorer (WISE) and Gaia, aiming to address gaps in previous large quasar samples. Our sample includes over 250,000 spectroscopically confirmed quasars reaching to $z \sim 5$, with tens of thousands of newly identified quasars in the southern hemisphere. The selection methods are highly pure, with well over 80% of the spectra collected being genuine quasars; the main contaminants are M-type stars. The detailed spectral decomposition procedure we employed shows that the quasars in the sample span a wide range of luminosities ($L_{\rm bol} \sim 10^{44}-10^{48}\,{\rm erg\,s^{-1}}$), SMBH masses ($M_{\rm BH} \sim 10^{6}-10^{10}\,M_{\odot}$), and accretion rates ($L/L_{\rm Edd} \sim 0.01-1$). The distributions of these properties are consistent with those of previous quasar catalogs, which are based on past generations of SDSS, once we account for potential selection biases related to the various survey depths. Our findings confirm that novel selection methods based on optical+infrared colors and/or astrometry can yield a large, high-purity quasar sample over wide sky areas, including in cases where more nuanced multiband photometry and/or multiwavelength data in the X-ray or radio are not available. This SDSS-V sample, which is expected to grow in the near future, establishes a robust reference for future southern (time-domain) surveys, while enhancing and complementing our understanding of quasar demographics and SMBH evolution.

Unified Astronomy Thesaurus concepts: Active galactic nuclei (16); Quasars (1319); Supermassive black holes (1663); Surveys (1671)

1. Introduction

Quasars, the high-luminosity manifestation of active galactic nuclei (AGNs), are powerful tools for exploring the Universe across cosmological distances. Quasars are powered by accreting supermassive black holes (SMBHs) that reside at the centers of massive galaxies, where a radiatively efficient accretion flow (i.e., a thin disk) releases large amounts of ultraviolet (UV)–optical radiation. When this primary radiation field is reprocessed by circumnuclear gas components, other forms of radiation may emerge. These primarily include broad emission lines that originate in fast-moving ($>1000\,{\rm km\,s^{-1}}$), high-density gas called the broad-line region (BLR); narrow emission lines from lower-density gas moving at lower velocities ($\sim 100-1000\,{\rm km\,s^{-1}}$) known as the narrow-line region (NLR); infrared (IR) emission from a dusty

Original content from this work may be used under the terms of the Creative Commons Attribution 4.0 licence. Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

axisymmetric structure (the so-called "dusty torus" and perhaps a dusty polar wind component); X-rays originating in a compact, hot, and low-density "corona"; as well as radio emission from relativistic jets in some objects (P. Padovani et al. 2017). All these emission components can be used to identify quasars (e.g., D. M. Alexander & R. C. Hickox 2012) and other types of AGNs out to $z > 7$ (F. Wang et al. 2021; see also P. Natarajan et al. 2024), and allow the study of the basic properties of the SMBHs that power them, the links between these SMBHs and their host galaxies, and even probe cosmic reionization (X. Fan et al. 2023).

Over the years, considerable effort has been dedicated to the search for quasars, i.e., luminous broad-line AGNs with limited obscuration along their lines of sight, as identified mostly in the rest-frame UV–optical regime. The most recent version of the Million Quasars (Milliquas/v8; E. W. Flesch 2023) compilation of AGN catalogs holds all the spectroscopically confirmed quasars, in addition to other types of AGNs and candidates, and includes just over 900,000 spectroscopically confirmed quasars as of 2023 June. The vast majority of this huge inventory (82.5%) was made possible by the various generations of the Sloan Digital Sky Survey (SDSS; D. G. York et al. 2000; R. Ahumada et al. 2020). More recently, the first data release (DR1) of the Dark Energy Spectroscopic Instrument project (DESI; DESI Collaboration et al. 2026) has added ≈1.2 million new quasars (S. Juneau et al. 2025). Given the location of the relevant survey facilities, it is not surprising that most of the quasars surveyed to date are located in the northern celestial hemisphere.

In contrast, the southern sky has been underexplored, with only a few notable surveys as exceptions. These are mostly the various wide-field spectroscopic surveys conducted at the Australian Astronomical Observatory, including the original Two-degree-Field Galaxy Redshift Survey (2dF; M. Colless et al. 2001), the Six-degree-Field Galaxy Redshift Survey (6dF; D. H. Jones et al. 2004), as well as the quasar-focused surveys the 2dF Quasi-Stellar Object Redshift Survey and the 2dF-SDSS LRG and QSO (2QZ; S. M. Croom et al. 2004 and 2SLAQ; S. M. Croom et al. 2009, respectively). While these programs provided valuable data for many thousands of southern-hemisphere quasars, most of the spectra obtained in these surveys lacked proper flux calibration, which naturally hindered some scientific inquiries. This imbalance between the celestial hemispheres is clearly evident in the Milliquas compilation, where only 13% of the quasars are located at $\delta < 0$. The relative scarcity of known quasars in the southern hemisphere is partly due to the historical lack of deep, wide-field optical imaging surveys. This is particularly striking given that this is the part of the sky that is accessible to many of the most advanced facilities available now, e.g., the Very Large Telescope (VLT), Atacama Large Millimeter/submillimeter Array (ALMA), Rubin, and 4MOST, and will be accessible in the future, e.g., the Extremely Large Telescope (ELT), Square Kilometre Array (SKA), and Roman.

Another limitation of existing quasar samples lies in the way that candidates were selected, based on photometry and/or multiwavelength data, for the spectroscopic observations that ultimately allow their confirmation and study. Specifically, most of the quasars in the legacy SDSS, i.e., up to Data Release 7 (DR7; K. N. Abazajian et al. 2009) and DR7Q (D. P. Schneider et al. 2010), were selected based on their optical colors (G. T. Richards et al. 2002), which generally favored blue, pointlike sources. A subset of quasars was targeted based on their radio emission, and a dedicated set of color selection criteria was used to select high-redshift ($z > 3.5$) quasar candidates. The complex color selection introduced by G. T. Richards et al. (2002) reaches 90% completeness and a quasar-selection purity (a.k.a., efficiency) higher than 65%. That legacy SDSS color-based selection had limited capabilities in the $z \approx 2.1-2.8$ regime, where quasar colors overlap with those of stars in optical color space, which motivated later generations of SDSS quasar searches to also incorporate a combination of more sophisticated selection methods, involving more nuanced (statistical) color- and variability-based selection criteria (see J. Bovy et al. 2011; N. P. Ross et al. 2012; A. D. Myers et al. 2015, and references therein). These resulted in a quasar-selection purity exceeding 50%. The focus of most quasar surveys on a high level of completeness ($\gtrsim$90%), required to robustly determine key demographic measurements (e.g., the quasar luminosity function; e.g., S. M. Croom et al. 2004, 2009), often meant compromising the selection purity, even when surveying high Galactic latitudes. Samples of quasars observed at low Galactic latitudes remained modest in size, and thus in utility (see, e.g., Y. Fu et al. 2021; J. Werk et al. 2024; Z.-Y. Huo et al. 2025, and references therein for recent progress). The advent of Gaia astrometric measurements enabled significant progress in constructing large samples of quasar candidates that have both high completeness *and* improved purity at all Galactic latitudes (see below and, e.g., Y. Shu et al. 2019; Gaia Collaboration et al. 2023a; Q. Yang & Y. Shen 2023; K. Storey-Fisher et al. 2024).

Importantly, optical-color-based selection methods may be biased against certain subpopulations of quasars, particularly those with redder spectral energy distributions (SEDs; though such quasars can still be found by those methods; C. M. Krawczyk et al. 2013; N. P. Ross et al. 2015; F. Hamann et al. 2017), which may indeed be expected for a number of reasons. First, simple accretion-disk theory predicts that the emerging SEDs should become redder with increasing black hole mass ($M_{\rm BH}$), and/or with decreasing black hole spin (N. I. Shakura & R. A. Sunyaev 1973). Indeed, C. Bertemes et al. (2016) argued that vigorously accreting SMBHs with low and/or retrograde spins might have colors consistent with red stellar sources. Second, some models of super-Eddington accretion flows predict that the innermost regions of the flow would become advection dominated, thus "trapping" the UV radiation, again showing atypically red SEDs (see Q. Pognan et al. 2020, and references therein). Finally, there is growing evidence that some of the most luminous AGNs, powered by high-mass SMBHs, are obscured by significant amounts of dust, likely on host-galaxy scales. Such heavily reddened sources, including red quasars (e.g., T. Urrutia et al. 2009; M. Banerji et al. 2012; E. Glikman et al. 2012, 2015; S. M. LaMassa et al. 2016; S. LaMassa et al. 2024) and hot dust-obscured galaxies (so-called "hot DOGs"; e.g., P. R. M. Eisenhardt et al. 2012; R. J. Assef et al. 2015), can be identified thanks to their (mid-)IR emission. While they may appear to be rare, several studies have reported that they may trace a significant fraction of the highest-luminosity AGN population (R. J. Assef et al. 2015) and/or a significant phase in SMBH growth (see, e.g., M. Banerji et al. 2015; F. Vito et al. 2018; V. A. Fawcett et al. 2023; E. Glikman et al. 2024, and references therein).

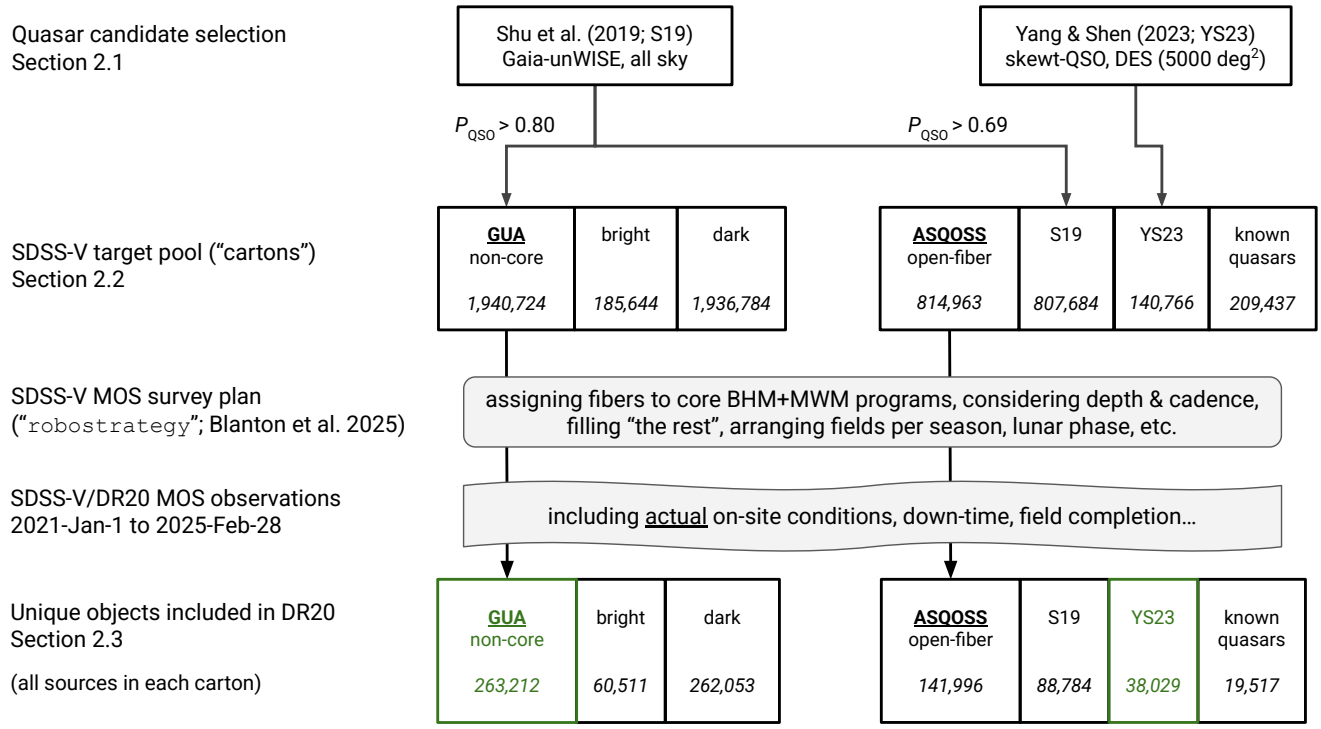

Figure 1. Parent sample construction. From top to bottom, the flowchart illustrates how quasar candidates are selected through the two main selection methods (based on S19 and YS23). The two techniques populate various, often nonunique SDSS-V target pools ("cartons"), which are used for actual MOS survey observations, and ultimately are made available for analysis as part of SDSS-V DR20. We highlight in green the two final sets used for our analysis.

The ongoing fifth generation of the SDSS project (SDSS-V; J. A. Kollmeier et al. 2026) holds the potential to alleviate some of the limitations of the currently available quasar samples. As a dual-hemisphere wide-field survey, it can target AGN and quasar candidates across the entire sky, and based on novel selection criteria. Specifically, for the southern hemisphere, a cornerstone of SDSS-V is the optical spectroscopy of X-ray-emitting AGN candidates drawn from the eROSITA surveys (i.e., the SPIDERS program; T. Dwelly et al. 2017; J. Comparat et al. 2020), with over 75% confirmed as broad-line quasars (C. Aydar et al. 2025). However, the SDSS-V survey strategy allows targeting of additional sets of AGN and quasar candidates, using other selection criteria. Whether these novel selection criteria are based on necessity or on physical motivation, their performance has to be empirically tested.

This paper focuses on two of the larger ancillary quasar searches within SDSS-V, enabled by novel candidate selection criteria that combine optical and mid-infrared (MIR) data, and which can be utilized for quasar searches over wide sky areas, including in the southern celestial hemisphere and/or at low Galactic latitudes (e.g., Y. Shu et al. 2019; Q. Yang & Y. Shen 2023). Together, these SDSS-V efforts have delivered new, high-quality optical spectroscopy for over 100,000 quasars, including over 70,000 that are newly identified—most of which are located in the southern celestial hemisphere.

The paper is structured as follows. Section 2 describes the two programs in terms of target selection and SDSS-V observations. Section 3 details our analysis of the sample of reliable quasars collected so far and the spectral measurements we conduct. Section 4 presents our main results in terms of the purity (efficiency) of quasar selection and the key properties of the quasars in our new sample(s). We summarize our conclusions in Section 5. Throughout this work, we use the AB magnitude system (J. B. Oke & J. E. Gunn 1983; unless clearly stated otherwise) and assume a flat ΛCDM cosmology with $H_0 = 70\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$, $\Omega_m = 0.3$, and $\Omega_\Lambda = 0.7$. We use the term "quasars" to broadly refer to broad-line AGNs whose optical continuum emission is dominated by the central accretion flow, regardless of their (bolometric) luminosity, including sources below the traditional limit of $L_{\rm bol} < 10^{44}\,\mathrm{erg\,s^{-1}}$.

2. Candidate Selection, SDSS-V Observations, and Data

This section describes the two main quasar-selection techniques that were used to construct our sample, the way the corresponding target pools were refined for actual SDSS-V spectroscopic observations, and the observations and data themselves.

Given the complex nature of the SDSS-V Multi-Object Spectroscopy (MOS) survey design and observations, there is nonnegligible overlap between various target pools (called "cartons" in SDSS jargon), and nontrivial considerations made during survey design and implementation. To summarize, any observed spectrum should, in principle, be considered as belonging to multiple cartons; it could be observed at various levels of priority, and under various observing conditions. The various steps in the process of designing and implementing the SDSS-V MOS survey are explained in detail in several key SDSS-V publications, including J. A. Kollmeier et al. (2026; general survey overview), A. Almeida et al. (2023; target pools and targeting philosophy), and M. R. Blanton et al. (2026; field and target assignment). Here, we only briefly refer to the key steps in this process that are relevant for our analysis. Figure 1 illustrates this process and provides guidance for the rest of this section.

2.1. Construction of Quasar Candidate Pools

This work focuses on the construction and analysis of large samples of quasars, which were selected through two novel procedures. Both procedures aim to provide large, high-purity (or efficiency)[20] sets of quasar candidates across wide sky areas, by combining publicly available photometry (in the optical and MIR) and astrometry (based on Gaia; Gaia Collaboration et al. 2018). The common challenge of both procedures is differentiating between quasar candidates and foreground interlopers, i.e., Milky Way (MW) stars, as both procedures are designed to be applicable also to low Galactic latitudes. In what follows, we summarize the key aspects of these two procedures; additional details can be found in the two respective studies (see below).

Both quasar-selection methods rely on the comprehensive all-sky data provided by the Gaia (specifically Data Release 2, DR2; Gaia Collaboration et al. 2018) and Wide-field Infrared Survey Explorer (WISE, specifically unWISE; E. F. Schlafly et al. 2019) missions. Gaia provides not only high-quality three-band photometry over the entire sky (in the G, BP, and RP bands), reaching depths of $G \lesssim 21$ mag, but—importantly—astrometric measurements for the vast majority of celestial sources accessible to SDSS-V, which allows rejection of sources with proper motion from any further consideration for extragalactic studies. We note, however, that the precision of the Gaia (DR2) proper-motion measurements degrades significantly at faint magnitudes, from $\sim 0.05\,\mathrm{mas\,yr^{-1}}$ at $G < 15$ mag to $\sim 0.2\,\mathrm{mas\,yr^{-1}}$ at $G = 17$ mag and $\sim 1.2\,\mathrm{mas\,yr^{-1}}$ at $G = 20$ mag (L. Lindegren et al. 2018). Indeed, at $G \gtrsim 20$ the astrometric uncertainty becomes comparable to the true proper motions of typical faint stellar contaminants if located at kiloparsec-scale distances ($\lesssim$ few $\mathrm{mas\,yr^{-1}}$). Thus, the proper-motion-based star–quasar discrimination loses much of its power for the faintest sources considered for the sample and analysis considered here, and one may expect that the quasar-selection purity would decrease for such faint objects.

The four-band MIR photometry provided by WISE over the entire sky, the four bands,[21] is widely used to robustly identify AGNs (and, specifically, quasars), as their MIR continuum emission appears redder than that of inactive galaxies. D. Stern et al. (2012) demonstrated that the simple criterion $\mathrm{W1} - \mathrm{W2} \geqslant 0.8$ yields a highly robust AGN selection, with a completeness of 78% and purity of 95%. A higher W1 − W2 color cut would enhance the purity at the expense of completeness. Later, R. J. Assef et al. (2013) and R. J. Assef et al. (2018) introduced refined W1 − W2 color cuts, optimized to improve selection reliability relative to the simpler D. Stern et al. (2012) criterion.

2.1.1. Gaia–unWISE-based Selection (Shu et al. 2019)

The main pool of quasar candidates selected for ancillary SDSS-V wide-area targeting is based on the study of Y. Shu et al. (2019, hereafter S19), which combined Gaia and unWISE data to derive a large set of candidates across the entire sky, apart from areas encircling the Magellanic Clouds and M31. While the Galactic plane is formally included, the highest-extinction sight lines yield a drastically lower number of, or indeed no, quasar candidates. This selection method is based on "AGN-like" MIR (WISE) colors combined with optical Gaia data and no detectable proper motion from Gaia DR2. The method leverages optical photometry from the G, BP, and RP bands of Gaia, as well as its astrometric data (Gaia DR2), and MIR photometry in the W1 and W2 bands (from the unWISE release). Sources are selected by crossmatching the Gaia DR2 catalog with the unWISE catalog (with a matching radius of 2″), focusing on objects with robust detections in both the W1 and W2 bands.

S19 describes a random forest (RF) algorithm developed for differentiating AGN and non-AGN sources, which is based on two large and correspondingly labeled training datasets. The (trained) algorithm can then predict the probability (P_{RF}) that a given source is an AGN. S19 applied the trained algorithm to the entire Gaia–unWISE dataset and created two nonmutually exclusive AGN candidate catalogs, dubbed C75 and R85, with 75% completeness and 85% efficiency,[22] respectively. Each of these catalogs assumes a different threshold of P_{RF} to balance completeness and efficiency; specifically, $P_{\mathrm{RF}} \geqslant 0.69$ for C75 and $P_{\mathrm{RF}} \geqslant 0.94$ for R85. Each of the resulting catalogs contains over 2×10^6 AGN candidates across a calculated effective area that S19 reports as $\approx$36,000 $\mathrm{deg^2}$, with nearly half of them not appearing in prior AGN compilations. This effective area considers the low-candidate-density regions toward the highest-extinction regions in the Galactic plane.

2.1.2. Dark Energy Survey/Skewt_QSO-based Selection

The study of Q. Yang & Y. Shen (2023, hereafter YS23) presents a catalog of $\approx 1.4 \times 10^6$ photometrically selected quasar candidates in the southern extragalactic sky, covering about 5000 $\mathrm{deg^2}$ of the footprint of the Dark Energy Survey (DES; T. M. C. Abbott et al. 2018). The selection process presented by YS23 combines optical photometry in the *griz* bands from DES DR2 (T. M. C. Abbott et al. 2021), near-infrared (NIR) photometry from a collection of surveys—VISTA Kilo-degree Infrared Galaxy (VIKING; A. Edge et al. 2013), Vista Hemisphere Survey (VHS; R. G. McMahon et al. 2013), UKIDSS Large Area Survey (ULAS; A. Lawrence et al. 2007), UKIRT Hemisphere Survey (UHS; S. Dye et al. 2018), and the Two Micron All Sky Survey (2MASS; M. F. Skrutskie et al. 2006)—and MIR photometry from unWISE (D. Lang 2014; A. M. Meisner et al. 2019). Additionally, Gaia astrometric data were used to eliminate stars with detectable proper motion, ensuring a cleaner selection of quasar candidates.

Similar to S19, YS23 refined their classification algorithm using two labeled training sets. The algorithm assigned each (nontraining) source with a probability of being a quasar, galaxy, or star (P_{QSO}, P_{Star}, and P_{Galaxy}, respectively), based on parametric representations of the respective distributions in multicolor space(s). A source was included in the quasar candidates sample only if $P_{\mathrm{QSO}} > P_{\mathrm{Star}}$ and $P_{\mathrm{QSO}} > P_{\mathrm{Galaxy}}$.[23]

The sample of quasar candidates selected by YS23 greatly overlaps with that of S19 in the sky region that is common to the two works (i.e., the $\sim$5000 $\mathrm{deg^2}$ of DES). Specifically, 97% of the YS23 quasar candidates (136,222 of 140,766

[20] Throughout this work, we use the term purity (and, at times, efficiency) to denote the fraction of quasars out of the total sample of sources targeted by the respective survey within which the search was conducted.

[21] Denoted W1–W4, with effective wavelengths of 3.4, 4.6, 12, and 22 μm.

[22] S19 refers to this as "reliability," hence the "R" in "R85."

[23] By definition, this requirement also implies that all such candidates have $P_{\mathrm{QSO}} > 1/3$.

objects) can also be found in the C75 catalog of S19. Nonetheless, in parts of our analysis we opt to consider the two selection methods separately, in order to assess their effectiveness in selecting real quasars, and given the subtle difference in how various subprograms were implemented in SDSS-V (see immediately below).

2.1.3. Limitations of Training Sets

Both selection methods carry known limitations in their ability to recover the full range of quasar populations, inherited from their training samples. Specifically, S19 was trained on the SDSS DR14Q catalog (I. Pâris et al. 2018), whose foundation in the *ugriz*-color-selected DR7Q catalog (G. T. Richards et al. 2002; D. P. Schneider et al. 2010) may have systematically underrepresented heavily dust-reddened quasars. YS23 was trained on the SDSS DR16Q catalog (B. W. Lyke et al. 2020), supplemented with confirmed quasars from the Milliquas compilation (E. W. Flesch 2019) and further augmented with simulated $z > 3.5$ quasars drawn from a distribution of *unobscured* SEDs. Both methods also rely on WISE-based aperture photometry that assumes point-source geometry, biasing against quasars with close companions. This may again complicate the selection of reddened quasars, as galaxy companions and/or mergers were shown to be linked with (enhanced) obscuration (e.g., C. Ricci et al. 2017; E. Glikman et al. 2024). Thus, we may expect that heavily obscured/reddened quasars would be missed by the selection methods, even prior to considering the challenges related to obtaining high-quality optical spectra and identifying AGN signatures. Other minority subpopulations of quasars whose colors may be underrepresented in the training sets, and thus in any survey that relies on the corresponding selection methods, may include broad absorption-line (BAL) quasars with particularly broad troughs.

2.2. Surveying Quasars in SDSS-V

The optical MOS component of SDSS-V utilizes two 2.5 m wide-field telescopes (J. A. Kollmeier et al. 2026): the legacy Sloan Foundation Telescope, located at Apache Point Observatory (APO) in New Mexico, USA (J. E. Gunn et al. 2006), and, since late 2022, the Irénée du Pont telescope at Las Campanas Observatory (LCO), in Chile (I. S. Bowen & A. H. Vaughan 1973), enabling full-sky coverage. SDSS-V operations at APO began in 2020 December with traditional fiber plug plates. A year later, these were replaced by a robotic Focal Plane System (FPS; R. W. Pogge et al. 2020), revolutionizing observing efficiency by automating fiber positioning.

Both the Black Hole Mapper (BHM; S. F. Anderson et al., in preparation) and Milky Way Mapper (MWM; J. Johnson et al., in preparation; N. De Lee et al., in preparation) programs within SDSS-V provide lists of targets for fiber-fed spectroscopic observations. Given the dual-hemisphere nature of SDSS-V, the multitude of science goals (J. A. Kollmeier et al. 2026), and the varying observing conditions, the actual survey operations are designed to prioritize several core programs that drive field selection, fiber allocation, and exposure times, with a variety of progressively lower-priority programs being allocated additional fibers until each MOS field can be considered as delivering maximal science potential. This means that the quasar candidates provided through either the S19 and/or the YS23 selection methods can be observed as part of several cartons, with a variety of priorities, exposure times, and observing conditions.

Below, we describe the various ways in which our quasars were observed, categorized chiefly by the priority of each program and by the additional information it may carry about the AGN nature of the target. We again stress that most AGN candidate cartons in SDSS-V have significant overlap, while for noncore programs the effective survey depth depends on the requirements set for each visit of each field by the higher-priority core programs. More details about the various SDSS-V programs and survey strategy can be found below, as well as in A. Almeida et al. (2023) and J. A. Kollmeier et al. (2026).

2.2.1. Gaia–unWISE AGN

The Gaia–unWISE AGN (GUA) program is the largest noncore, AGN-focused program within SDSS-V, designed to provide a large number of targets across the entire SDSS-V survey area and range of observation types. Starting with the S19 catalogs, the GUA selection focuses on sources within the C75 subset that have $P_{\rm QSO} \equiv P_{\rm RF} > 0.8$, $G > 13$ mag (to avoid saturation). We stress that the availability of previous spectroscopy from other sources (e.g., LAMOST, 2dF/2QZ; about half of the sources in S19) was *not* a consideration for GUA, thus potentially allowing a future use of new SDSS-V spectra obtained through GUA for long-term time-domain science.

When incorporating the GUA set of quasar candidates into the SDSS-V targeting efforts, it was further split into two subsets based on source brightness, making them suitable for either dark or bright (lunar) conditions. The *GUA-dark* subsample consists of 2,156,582 targets with $16 < G < 21$ mag, while the *GUA-bright* subsample includes 254,601 brighter targets with $13 < G < 18$ mag, allowing them to reach an adequate signal-to-noise ratio (S/N) even during bright time. This separation enables optimal use of observing conditions for each subset, with the main distinction being that GUA-dark targets populate fields observed as part of the BHM program, which is pursued only during dark time, while the GUA-bright targets can also populate fields observed as part of the MWM program, which is pursued mostly during bright time. We stress that these two subsamples are not mutually exclusive, as reflected by their overlapping G-band magnitude ranges.

Overall, in its highly efficient regime, down to $G = 20.5$, the GUA targeting reaches an all-sky-averaged density of quasar candidates of $\approx 37.7\ \mathrm{deg}^{-2}$. This is slightly higher than the sky density of quasars observed through previous generations of the SDSS (e.g., $\approx 24\ \mathrm{deg}^{-2}$ in DR16; B. W. Lyke et al. 2020). At the nominal GUA flux limit, $G = 21$, these two sky densities are fully consistent.

2.2.2. All Sky Quasar Optical Spectroscopic Survey

The All Sky Quasar Optical Spectroscopic Survey (ASQOSS) is a large "open fiber" program in SDSS-V,[24] aiming to complement the SDSS-V targeting pool so it can include the largest sample of quasar spectra across the entire sky, even if through relatively short exposures. As an open-fiber program, ASQOSS targets can be selected to be observed

[24] Internally designated with the carton name `openfibertargets_nov2020_27`.

in either dark or bright time, in either BHM- or MWM-led fields, generally with exposures as short as 15 minutes.

The overall ASQOSS target catalog contains 814,963 objects that are a superset of several distinct samples:

1. Quasar candidates selected from the S19 catalog, limited to $G < 20$ mag. These 807,684 targets are distributed across the entire sky except for the regions surrounding the Magellanic Clouds and M31. While this sample of targets is similar in spirit to GUA, here a threshold of $P_{\rm QSO} > 0.69$ was assumed (see > 0.8 for GUA).
2. Quasar candidates selected from the YS23 catalog, limited to $G < 20$ mag. By construction, these 140,766 targets are located within the $\sim$5000 deg^2 DES footprint.
3. Previously known quasars drawn from the SDSS DR16Q catalog (B. W. Lyke et al. 2020), the LAMOST DR5 quasar catalog (S. Yao et al. 2019), and the final 2QZ quasar catalog (S. M. Croom et al. 2004). These 209,437 targets are included in ASQOSS to maximize the time-domain-related science goals of BHM, namely searching for extremely variable and/or "changing-look" AGNs (C. Ricci & B. Trakhtenbrot 2023).

All the targets in the ASQOSS carton are robustly detected by Gaia, have $G < 20$ mag, and exhibit no detectable proper motion (as of Gaia DR2). Regions with high extinction near the Galactic plane, as well as around the Magellanic Clouds and M31, are omitted.

In what follows, we only focus on those ASQOSS targets that are drawn from either the S19 or YS23 selection procedures (i.e., the first two items above); whenever we use the term "ASQOSS," we refer only to these two types of targets. This approach is driven by our focus on the efficiency of these novel quasar-selection methods, and on the properties of the quasars they uncover when employed within SDSS-V. Moreover, targets for which there is clear prior knowledge about the nature of their nuclear activity (i.e., the previously known quasars included in ASQOSS) would obviously skew any assessment of the efficiency of this open-fiber program.

The ASQOSS/Skewt_QSO targeting reaches an averaged density of quasar candidates of $\approx$28.1 deg^{-2} down to $G = 20.5$. This is generally consistent with GUA and SDSS DR16Q (see above).

2.2.3. Other Programs in SDSS-V and Prioritization

The target-selection processes for the GUA and ASQOSS programs may overlap not only with each other but also with other observational initiatives. The SDSS-V DR18 and DR19 papers (A. Almeida et al. 2023; SDSS Collaboration et al. 2026a, respectively) provide a detailed discussion of the various SDSS-V (BHM) programs; here, we only note those few BHM programs that are most relevant for the present study. For our purposes, the other BHM programs may be split based on whether they carry additional information about the AGN nature of the targets, and/or whether they have higher or lower priority than GUA and/or ASQOSS during survey design and execution.

The high-priority core BHM programs include two key wide-field components: the All-Quasar Multi-Epoch Spectroscopy program (AQMES; M. Eracleous et al. 2026, in preparation), which revisits previously known SDSS quasars, and the SPectroscopic IDentification of ERosita Sources program (SPIDERS; A. Merloni et al. 2026, in preparation), which is based on X-ray observations conducted with the eROSITA mission (P. Predehl et al. 2021). Moreover, the reverberation mapping program (BHM-RM; J. R. Trump et al. 2026, in preparation) monitors hundreds of known known quasars in a few distinct, discontiguous fields. The BHM-RM fields and targets are designed separately from the rest of the SDSS-V MOS survey, and their overlap with GUA and ASQOSS does not affect our analysis. Another high-priority, though noncore, program focuses on sources drawn from the Chandra Source Catalog Release 2 (CSC 2.1; I. N. Evans et al. 2024), which naturally includes many AGN candidates. When assigning fields and fibers/targets (M. R. Blanton et al. 2026), GUA has a priority that is lower than that of any of the (BHM) core programs and that of the CSC program. ASQOSS has a yet lower priority, being an open-fiber program.

Finally, there are many other open-fiber programs, some of which target various samples of AGNs and/or AGN candidates (see A. Almeida et al. 2023; SDSS Collaboration et al. 2026a). All these open-fiber programs have a priority that is identical to that of ASQOSS (and thus lower than that of GUA). One of these open-fiber programs is designed to target quasar candidates from the Quaia catalog (K. Storey-Fisher et al. 2024), selected based on Gaia and unWISE data; however, the vast majority of Quaia targets ($\approx$97%) are already included in either GUA and/or ASQOSS.

2.2.4. Selection Overlap

The design of the SDSS-V MOS survey and the ultimate data acquisition consider all target cartons and their priorities through a complex algorithm, which is described in detail in M. R. Blanton et al. (2026). The resulting dataset of quasar (candidate) spectroscopy incorporates a significant overlap between various target cartons, which should be considered when assessing the efficiency of any (quasar) selection method. This complex target-selection overlap is nonnegligible for the samples under study. Specifically, 80% (652,278) of the objects in the ASQOSS carton are also part of the GUA carton (and 34% targets in GUA are also in ASQOSS); moreover, 21% (410,045) of the GUA targets and 27% (220,584) of the ASQOSS targets are also selected as SPIDERS X-ray emitters. Appendix A and the figures therein illustrate the full complexity of this target-selection overlap.

In practice, when assessing the quasar-selection efficiency of the S19 and YS23 selection methods (in Section 4.2), we consider several disjoint subsets of targets, based on their membership in the GUA, ASQOSS/Skewt_QSO, and/or core +CSC cartons. The subsets are conceptually illustrated in Appendix A. The detailed purity calculations are performed only over a restricted area within the high-Galactic-latitude part of the southern sky (see Appendix A). Another point where the selection overlap is important is in our accounting of the number of new quasars detected through the GUA and ASQOSS/Skewt_QSO programs (Section 4.1). There, we focus on the subset of objects that are part of any of these two cartons, but are *not* part of any of the core programs or of the CSC carton. This "GUA+Skewt_QSO sole selection" conceptually follows the subsets designated as "A," "C," and "E" in Appendix A, employed over the entire footprint of the data in hand.

2.2.5. SDSS-V Spectroscopy

For this study, we used the spectra that are part of the SDSS DR20 data release (SDSS Collaboration et al. 2026b).[25] All the technical details about spectral data acquisition, reduction, and classification can be found in the DR20 paper and references therein. Here, we only briefly mention a few key aspects.

The optical spectra analyzed in this paper were obtained with the Baryon Oscillation Spectroscopic Survey (BOSS) spectrographs (S. A. Smee et al. 2013), which are fed by up to 500 fibers, each with apertures of either 2″ or 1.″3 (for APO and LCO, respectively), with a spectral resolution of $R \sim 2000$ over the wavelength range 3600−10400 Å. Each spectral observation consists of 15 minutes-long exposures, repeated as required by the specifications of the (highest-priority) carton to which the target belongs. Typical total integration times vary between a single 15 minutes exposure (e.g., for many GUA-bright and ASQOSS targets) and 60 minutes (e.g., for GUA-dark targets). Many SDSS-V/BHM targets are visited multiple times, over a wide range of timescales, mainly through the AQMES and BHM-RM programs. The present work, however, does not go into such time-domain aspects.

The BOSS pipeline performs all the spectroscopic reduction steps, as described in more detail in the relevant SDSS-V publications (e.g., Section 4.1. in SDSS Collaboration et al. 2026a and Section 7.1.1. in J. A. Kollmeier et al. 2026) and in the dedicated publication by S. Morrison et al. (2026, in preparation). Key calibrations are achieved using dedicated spectroscopy of standard stars and sky regions in each field, as well as calibration sources (i.e., arc lamps). Other parts of the pipeline provide spectral classification and, when applicable, redshift measurements for each source, using a variety of templates. Specifically, spectra are classified as those of a "QSO," "GALAXY," or "STAR" based on template fitting. All these classifications are based on (improvements to) the original BOSS pipeline (A. S. Bolton et al. 2012). While we imposed several quality criteria to further clean the resulting sample of optical spectra to be used for our analysis (see below), we stress that we made no effort to amend the pipeline-produced spectral classifications (and/or redshifts). To assess the reliability of the pipeline-based quasar identifications, we make use of an internal visual inspection effort by the SDSS-V/BHM team, which is analogous to the analysis presented in I. Pâris et al. (2017) for the DR12Q catalog. Restricting to visually inspected spectra classified by the pipeline as QSO with `ZWARNING` = 0 and S/N > 2 (23,986 spectra), this process confirms ∼98% as genuine quasars, with only ∼1.4% found to be misclassified stars or galaxies and a further ∼0.6% deemed unclassifiable due to low S/N. The overall pipeline reliability inferred here is consistent with the ∼0.3%−1.3% contamination reported for the SDSS-IV/eBOSS DR16Q catalog (B. W. Lyke et al. 2020), which used the same BOSS pipeline.

Like all DR20-level BOSS data, the spectra we use were acquired up to MJD 60708 (2025 February 2) and processed through version v6_2_1 of the BOSS pipeline.

2.3. Ancillary Photometric Data

We used several sources of data to augment the BOSS spectroscopy with relevant broadband photometry. The BHM database of reduced BOSS spectra (i.e., the `spAll` file) includes MIR photometry from the AllWISE Data Release (R. M. Cutri et al. 2013), W1 − W4 bands in Vega magnitudes, which we used for searching hot DOGs (see Section 1).[26] In addition to the Gaia DR2 (Gaia Collaboration et al. 2018) data used for targeting, the `spAll` file also includes crossmatched data from Gaia Data Release 3 (DR3; Gaia Collaboration et al. 2023b) and even some data from Gaia Data Release 1 (DR1; Gaia Collaboration et al. 2016; specifically, GAIA_G_MAG, BP_MAG, and RP_MAG). Some of these data included extinction corrections, and others did not, so we could not directly use the magnitudes listed in the pipeline. To obtain the most up-to-date data from Gaia (DR3), we crossmatched our dataset through the Astro Data Lab Science Platform provided by NOIRLab (M. J. Fitzpatrick et al. 2014; R. Nikutta et al. 2020)[27] with the Gaia DR3 catalog, with a crossmatching radius of 3″. For extinction corrections, we followed Gaia Collaboration et al. (2018) and used the color-free k_G term. We further retrieved the optical photometry of those sources in our dataset that are covered by the 10th data release (DR10) of the DESI Legacy Imaging Surveys (LS; A. Dey et al. 2019).[28] This dataset provided magnitudes in the *griz* bands, as well as the corresponding extinction-corrected, point-spread function magnitudes in the AB system, for 230,678 sources.

We stress that these photometric data are used here *not* to reassess the target-selection techniques (which, at times, relied on slightly different photometric measurements), but instead to examine the multiband properties of the GUA and Skewt_QSO quasars under study, using some of the most up-to-date data that are available for (a large fraction of) our sources.

3. Analysis and Spectral Measurements

3.1. Sample Construction and Refinement

As of the cutoff date for DR20 (2025 February 2), the SDSS-V/BOSS pipeline provided over 8 million calibrated spectra, including those of sources classified by the pipeline as stars, quasars, or galaxies. The metadata for all the spectra are stored in a single `spAll` file. We used the epoch-coadded spectra, which aim at achieving some predefined S/N criteria (for the core targets in each field), at times using integrations from adjacent nights, but without completely losing knowledge of possible longer-timescale source variability. As explained in Section 2.2, the same object could be targeted for several different reasons, and moreover could have been observed multiple times (for either scientific or operational reasons). Therefore, we first selected only objects that were targeted within either one of the desired cartons, that is `bhm_gua_dark`, `bhm_gua_bright`, or `openfiber-targets_nov2020_27` (aka ASQOSS). For those objects that appeared multiple times in the pipeline data products (with the same "SDSS_ID"), we retained only the spectrum with the highest reported S/N available, which resulted in a total of 282,076 unique objects. For objects marked as "ASQOSS," we

[25] https://www.sdss.org/

[26] The use of the AllWISE Data Release instead of the AllSky Data Release of WISE could have a minor effect when applying the hot DOG selection function of P. R. M. Eisenhardt et al. (2012). See R. J. Assef et al. (2015) for details.

[27] https://datalab.noirlab.edu/

[28] Again using the Astro Data Lab Science Platform and a crossmatching radius of 3″.

kept only the ones that were selected through the Skewt_QSO or S19 selection algorithm (275,287 objects).[29]

As Figure 1 shows, these selection steps provided spectra of 264,200 unique targets, of which 262,053 were part of the bhm_gua_dark carton, 60,511 of bhm_gua_bright, and 38,029 from ASQOSS/Skewt_QSO. This breakdown further illustrates the overlap between the various cartons (see Appendix A).

Not all of these spectra are of sufficient quality for detailed spectral analysis, and, in fact, not all of them have correct spectral classification and/or redshift determination. Given the scope of the data, we had to find ways to eliminate from the sample the problematic spectra via an automated process, based on experience gained through the visual inspection of a large set of BOSS spectra.

First, spectra with fatal issues were omitted from our analysis and will not be included in the forthcoming DR20. This identification was done by relying on the information contained within the ZWARNING flags (see Appendix B), and includes cases where (1) the fiber was a sky fiber; (2) the fiber was unplugged or damaged; (3) the targeting data had catastrophic issues (i.e., due to some mismatch between targeting catalogs); or (4) the reduced spectral files had no real data due to severe issues with the BOSS spectrographs.

We further refined our sample, informed by the ZWARNING and ZERR data provided by the BOSS pipeline, and by our (partial) visual inspection.[30] Specifically, we divided the GUA +Skewt_QSO superset into two subsamples:

1. *The "good" sample*. This sample contains all the spectra with ZWARNING $= 0$ (i.e., no warning), no catastrophic self-identified issues in redshift determination by the 1D pipeline (ZERR > 0), a relative redshift error below 5%, and a positive redshift ($z > 0$). The *good* subsample comprises 213,528 objects (81% of the total superset size).
2. *The "bad" sample*. This sample contains spectra with nonnegligible ZWARNING flags that are not catastrophic, as described in Appendix B, spectra with ZERR < 0, and sources with a negative redshift. The *bad* subsample comprises 50,672 objects (19% of the superset size).

We decided to retain these *bad* spectra for parts of our analysis because, after visual inspection, most of them do show clear broad emission lines and are therefore spectroscopically confirmed quasars, albeit with problematic (pipeline-based) redshift determination. Specifically, ≈83% of the visually inspected *bad* spectra that the BOSS pipeline classified as QSOs indeed showed broad emission lines. Moreover, almost half of the visually inspected *bad* spectra classified as galaxies and ≈16% of those classified as stars also showed such clear AGN spectral features. While this *bad* subsample is suitable for statistical studies of the SDSS-V quasar selection and content, only the high-quality (*good*) subsample is used for detailed spectral analysis (which assumes the pipeline redshifts are correct; as discussed in Section 3.2).

[29] The analysis presented in this paper focuses only on the GUA and Skewt_QSO sources, and does not include the subset of ASQOSS targets selected through the broader S19 selection.

[30] ZERR < 0 means that the 1D pipeline encountered a severe problem while measuring the redshift, and the reported redshift is likely unreliable.

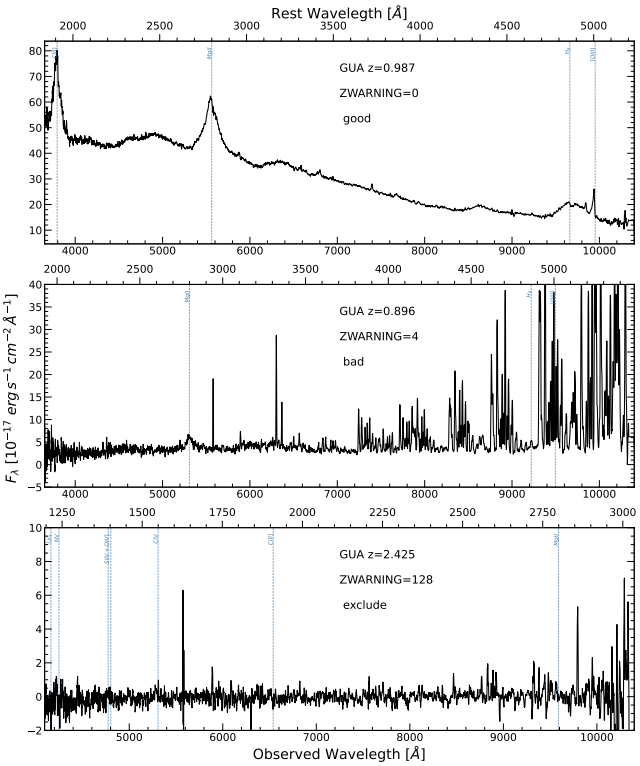

Figure 2. Examples of the three basic categories of SDSS-V spectra in our parent sample. From top to bottom: (1) a spectrum of a quasar that passed all the basic quality checks, including object classification and redshift assignment, and is therefore considered a *good* spectrum; (2) a spectrum of a quasar where the BOSS pipeline has issued a warning, but is still worth considering for efficiency analysis (i.e., part of the *bad* sample); and (3) a spectrum that was omitted from our analysis (and from DR20), based on its ZWARNING (which in this case suggests the fiber was unplugged).

We compared our spectral quality cuts to those adopted by C. Aydar et al. (2025), which were based on a thorough analysis of the BOSS spectroscopy of the eFEDS field (M. Salvato et al. 2022). Specifically, for sources with no visual inspection, C. Aydar et al. (2025) adopted the combination of S/N > 2, ZWARNING $= 0$, and $0 <$ ZERR < 0.005 to robustly select AGN based solely on the BOSS pipeline outcomes, which is comparable to our situation here. Overall, the two sets of criteria result in highly consistent samples: ≈87% of our *good* quasar sample would also pass the C. Aydar et al. (2025) criteria, and almost all (≈99.9%) of the *bad* quasar sample would not pass them. However, based on our partial visual inspection, simply applying the C. Aydar et al. (2025) criteria may leave many bona fide broad-line AGNs out of the *good* sample used for spectral measurements (≈13% of the *good* quasars). We therefore preferred to only use our quality criteria for sample construction. In any case, choosing instead the C. Aydar et al. (2025) criteria would not significantly change our main findings regarding the key properties of the quasars observed through the GUA and ASQOSS programs.

We present examples of spectra of various quality types in Figure 2. The top panel shows a *good* quasar spectrum and therefore was retained for detailed spectral measurements. The middle panel displays a spectrum with clear quasar-like broad emission lines, but also with significant residual sky features, which in turn greatly reduce the reliability of the measured redshift. Thus, the spectral fitting algorithm would have

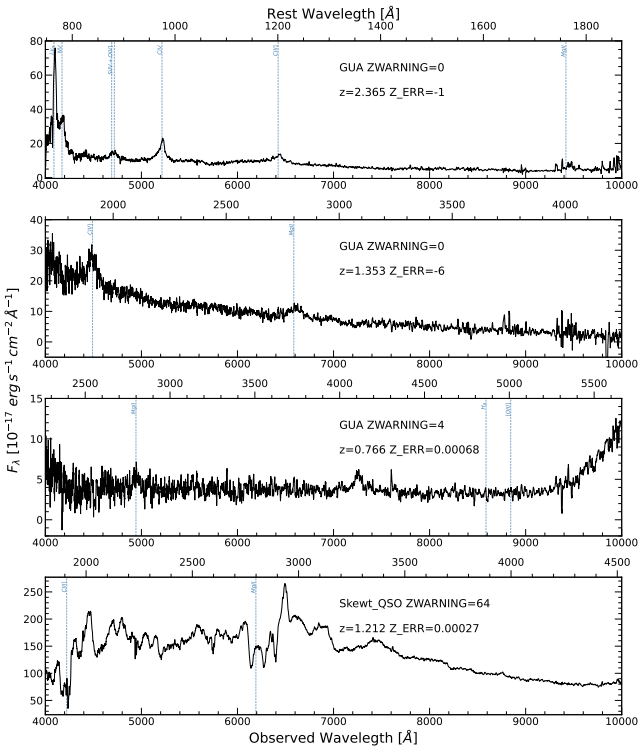

Figure 3. Examples of spectra classified into the *bad* sample due to issues in the BOSS pipeline products. These classifications are based on pipeline diagnostics (e.g., ZWARNING and ZERR flags), and do not necessarily imply that the redshift is incorrect. From top to bottom: (1) the best fit is at the lowest or highest redshift tested for a template; (2) the best-fit redshift lies outside the fitting range; (3) the χ^2 value of the best fit is too close to that of the second-best fit; and (4) the spectrum exhibits negative emission (see Appendix B). In all cases, the issues are not necessarily obvious from visual inspection alone, but are reflected in the pipeline quality flags.

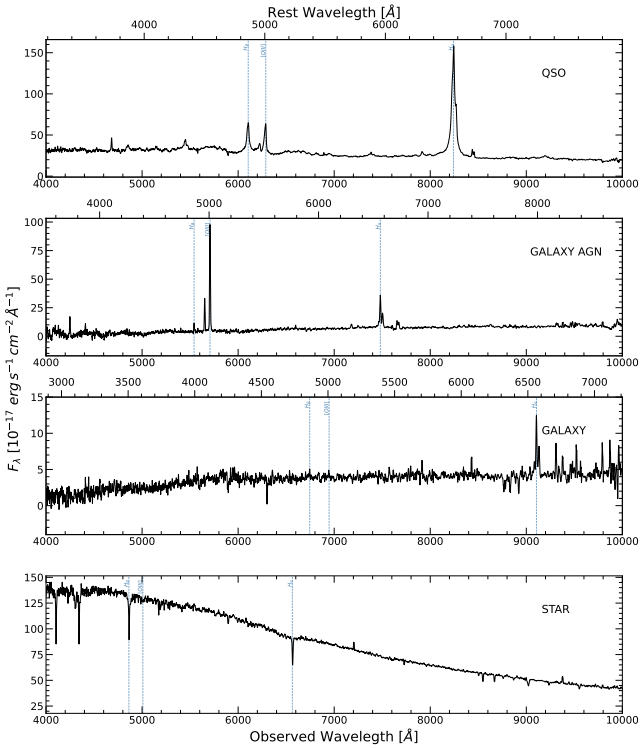

Figure 4. Examples of various types of celestial objects in the GUA- and Skewt_QSO-based samples. All four spectra are included in the *good* sample, as they are not affected by any severe issues (as indicated, for example, by the ZWARNING and ZERR flags). Stars and galaxies with reliable spectra must be part of the selection efficiency analysis, but are obviously not considered for the quasar sample itself (i.e., for spectral measurements). From top to bottom: a quasar at $z = 0.255$, an AGN at $z = 0.139$, a galaxy at $z = 0.387$, and an F3/F5V star.

difficulty fitting a broad line in the sky-affected wavelength range. In this specific example, the ZWARNING flags suggest that the χ^2 of the best fit is too close to that of the second-best one (so ZWARNING $\neq 0$). Therefore, we keep this quasar for the parts of the analysis concerning quasar-selection efficiency, as this is indeed a successful example of a quasar targeted by one of the programs under study (GUA in this particular case), but not for the parts of the analysis that rely on detailed spectral measurements. Finally, the bottom panel in Figure 2 presents a completely different level of failure, where we cannot use the spectrum at all, as it is not possible to say whether this (GUA-selected) source is indeed a quasar, although the 1D pipeline classified the spectrum as such (and even assigned a redshift). In this example, the ZWARNING clearly indicates that the fiber was unplugged or damaged.

Figure 3 presents four examples from the *bad* sample, each of which falls into this category for a different reason. While some spectra are indeed unsuitable for detailed spectral measurements, others still maintain relatively reliable quality. The top panel shows a quasar with no ZWARNING but with ZERR $= -1$, implying that the best fit is at the lowest or highest redshift tested for a template. Despite this result, it can be clearly seen that this is indeed a quasar with $z \simeq 2.365$, as indicated by the 1D pipeline. Similarly, the second example also has a reliable redshift, although its ZERR suggests that the best-fit redshift is outside the fitting range. The third panel presents a quasar with a broad emission line, confirming its spectral classification, although the assigned redshift appears to be incorrect, and $z \simeq 1.59$ would provide a better fit. The ZWARNING in this case suggests that the χ^2 of the best fit is too close to that of the second best. Finally, in the bottom panel, the ZWARNING states that a quasar line exhibits negative emission, which is a warning issued only for objects classified as quasars, so as to assist in identifying potential cases of BAL quasars (e.g., R. J. Weymann et al. 1991; R. R. Gibson et al. 2009). However, in this specific case, the redshift must be wrong, and we cannot perform any reliable spectral measurements (again, keeping in mind the enormous scope of the data in hand prohibited reassigning redshifts in such cases). These four examples demonstrate that the *bad* sample can contribute to the number of quasars identified through GUA and Skewt_QSO, even if their currently available reduced spectra prohibit any reliable spectral measurements, including their most basic property—their redshift.

Figure 4 shows four different objects in the *good* sample. Although the above steps have significantly reduced the number of nonquasars in our sample, clearly some stars and galaxies remain. Some of the objects classified by the 1D pipeline as galaxies show prominent broad emission lines, and can thus be robustly identified as AGNs. Therefore, we added an AGN sample with galaxies that were subclassified as "AGN," "BROADLINE," or "AGN BROADLINE." Only 873 objects fit this definition, and only 797 of those are tagged as "*good*."

3.2. Spectral Analysis

We now describe the detailed spectral decomposition we performed to all quasars in the *good* sample of GUA and Skewt_QSO quasars. For the spectral measurements, we used the publicly available package PyQSOFit, which was introduced in H. Guo et al. (2018) and Y. Shen et al. (2019), and further modified as described in Q. Wu & Y. Shen (2022).[31] To allow a comparison of our spectral measurements and derived quantities to the large catalog of SDSS DR16Q quasar measurements by Q. Wu & Y. Shen (2022), we followed the PyQSOFit setup introduced in that work as closely as possible. Below, we briefly summarize the key steps performed by PyQSOFit and the underlying model assumptions.

Prior to the fitting itself, PyQSOFit first corrects each spectrum for Galactic dust attenuation using the maps from E. F. Schlafly & D. P. Finkbeiner (2011; based on D. J. Schlegel et al. 1998) and assuming the J. A. Cardelli et al. (1989) extinction curve with $R_V = 3.1$. The spectrum is then deredshifted to the rest frame using the cataloged (i.e., pipeline-deduced) redshift.

For the vast majority of sources, located at $z \geqslant 0.3$ (200,749 objects; ≈95% of the sample), we do not account for host-galaxy emission, as it is expected to be negligible at these redshifts and given the typical brightness (and thus luminosities) of our sources. Moreover, the host-galaxy decomposition technique implemented in PyQSOFit is limited to $\lambda_{\rm rest} > 3400$ Å, so at $z \geqslant 0.3$ part of the spectrum would not be correctly treated. For the minority of sources at $z < 0.3$ (9920 objects in total; ≈5% of the sample), however, we did set up PyQSOFit to account for host emission. This analysis is done using principal component analysis (PCA), correlating with galaxy and quasar eigentemplates developed by C. W. Yip et al. (2004a) and C. W. Yip et al. (2004b), respectively. We used the fraction of host emission at 5100 Å (frac_host_5100) from the PyQSOFit output file to determine if the host decomposition succeeded or not. In some of the spectra, the host template that was fitted had a higher flux than the original data (i.e., host galaxy + QSO, so a negative eigenvalue for the quasar template). Therefore, in cases where frac_host_5100 $\geqslant 1$ (≈7.4% of the low-z subset), we reran PyQSOFit but with no host decomposition.

Next, PyQSOFit fits the host-free continuum emission and the emission from blended iron features in several "continuum windows". The former component is fitted with a power law and a third-order polynomial, while the latter is fitted with an empirically derived template that can be broadened and shifted. PyQSOFit then uses the continuum- and iron-subtracted spectrum to decompose several key emission-line complexes, focusing on the spectral regions surrounding the Hα, Hβ, Mg II λ2798, C III] λ1909, C IV λ1549, Si IV λ1400, and Lyα lines. Broad and narrow emission lines are fitted with combinations of Gaussian profiles, with some key parameters being tied (e.g., shifts and widths of narrow lines in each spectral region; see Q. Wu & Y. Shen 2022). Here, we employed two choices that differ from those made in Q. Wu & Y. Shen (2022). First, we decided to remove the bluest default continuum window (1150−1170 Å) for all high-redshift quasars ($z > 3.5$), in order to avoid problematic continuum placement in the spectral regime that is significantly affected by absorption from the intergalactic medium (IGM).[32] Second, we limited the widths of the broad Gaussian profiles to FWHM $< 15{,}000\,{\rm km\,s^{-1}}$. Absorption lines, including both intrinsic and intervening features, are common in quasar spectra, particularly at high redshift (e.g., R. R. Gibson et al. 2009). To avoid the effect that such troughs may have on the continuum and emission-line fitting, we used the dedicated PyQSOFit feature, which iteratively rejects spectral pixels falling $>3\sigma$ below the original model fit. This approach is effective in reducing the influence of narrow absorption features, but is less effective for broad absorption troughs (e.g., BALs), and does not fully account for absorption from the IGM.

The spectral measurements characterizing the continuum and line emission are derived from the best-fit model parameters. Uncertainties are estimated through a Monte Carlo (MC) procedure, which in turn is based on the error spectrum (i.e., errors on the flux densities). The algorithm generates a set of mock spectra, and each of them is fitted with the same model as the original spectrum. The uncertainty on each quantity is taken to be the half-width between the 16th and 84th percentiles of the corresponding distribution of measurements. We chose to have 50 such iterations (i.e., 50 mock spectra) for each quasar. This choice is a compromise between the need for a statistical pool of mock fits and computational limitations for our large sample. An example of how the spectral decomposition works for one of our quasars is shown in Figure 5.

We extracted several key quantities from the spectral fitting output of each of our quasars for estimating bolometric luminosity, black hole mass, and Eddington ratio. Continuum and line luminosities are calculated from the model continuum flux densities in the corresponding continuum windows and from the integrated line fluxes. The widths of the broad emission lines, in terms of FWHM, are returned by PyQSOFit for the fully reconstructed (broad) line profile, which consists of several (broad) Gaussians. PyQSOFit also provides uncertainties on these luminosities and FWHMs, deduced through the aforementioned MC procedure.

3.3. Bolometric Luminosities, Black Hole Masses, and Eddington Ratios

The spectral measurements described above are particularly important for estimating some of the key properties of the accreting SMBHs that power the quasars in our sample, including AGN bolometric luminosities ($L_{\rm bol}$), black hole masses ($M_{\rm BH}$), and Eddington ratios ($L/L_{\rm Edd}$), following widely used and well-established procedures.

To this end, we rely on the measured monochromatic continuum luminosities at rest-frame wavelengths of 5100, 3000, and/or 1350 Å, in terms of λL_λ (L_{5100}, L_{3000}, and L_{1350} respectively), as determined by PyQSOFit. For the low-redshift AGNs in our sample, the measurement of the luminosities may be affected by host-galaxy emission, and/or by how PyQSOFit was employed to address it (see above). We therefore also used the luminosity of the broad Hα line (hereafter L(bHα)) for alternative estimates of $L_{\rm bol}$ and $M_{\rm BH}$ (and thus of $L/L_{\rm Edd}$).

[31] We used the PyQSOFit_wqy version of the package: https://github.com/QiaoyaWu/sdss4_dr16q_tutorial.

[32] Note that PyQSOFit is not (currently) designed to address IGM effects.

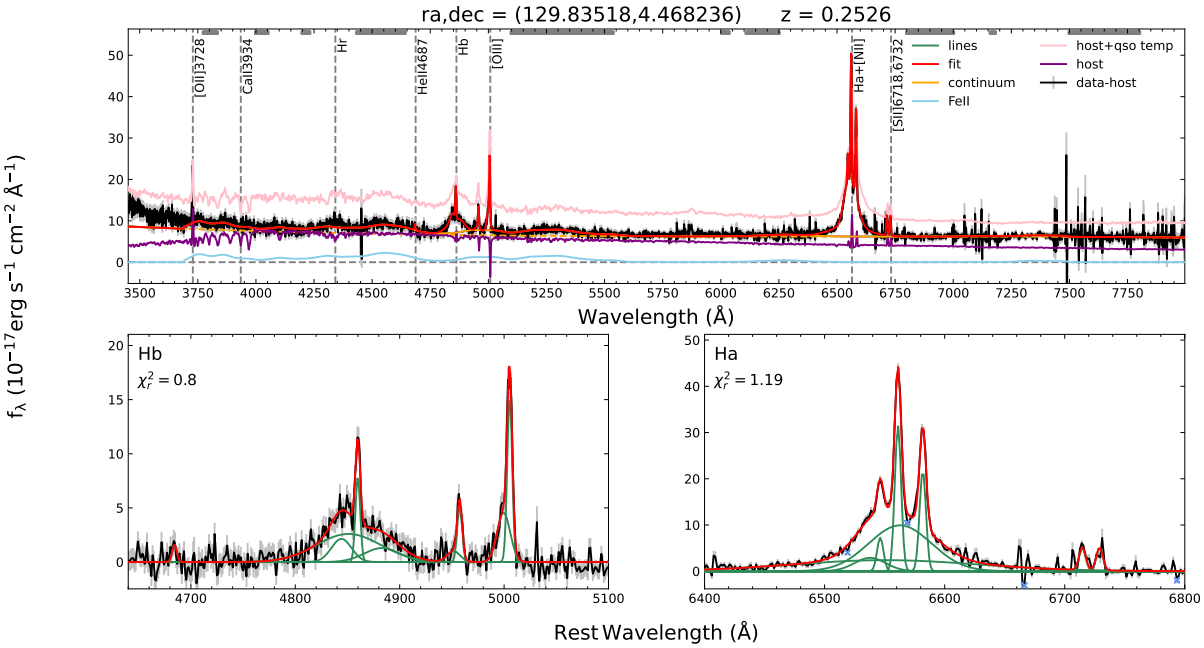

Figure 5. An example of a successful spectral fit using `PyQSOFit`. The top panel shows the observed spectrum after host subtraction (black), the total best-fit model (red), and its various components: host galaxy (purple), quasar continuum (orange), quasar and host templates (pink), and blended Fe II emission (light blue). The gray stripes near the top of the panel mark the continuum fitting bands. The bottom panels display the emission-line decomposition for the Hα and Hβ spectral complexes, with the individual broad and narrow Gaussian profile components shown in green.

We estimated $L_{\rm bol}$ by applying simple, universal bolometric correction factors ($f_{\rm bol}$) derived from the typical mean SEDs of quasars in G. T. Richards et al. (2006). The specific correction factors used were $f_{\rm bol} = 9.26$, 5.15, and 3.81 for L_{5100}, L_{3000}, and L_{1350} respectively. These values were chosen to be fully consistent with Q. Wu & Y. Shen (2022). Some studies suggest a wide range of more elaborate corrections, either based on observed SEDs and/or physical models, often with some dependence on luminosity, $L/L_{\rm Edd}$, or other quasar properties (see, e.g., R. V. Vasudevan et al. 2009; J. C. Runnoe et al. 2012; H. Netzer 2019; F. Duras et al. 2020, and references therein). Beyond the possible physical insights encoded in these variations, they highlight the inherent systematic uncertainty on any choice of $f_{\rm bol}$, which is generally at least 0.2 dex (e.g., K. K. Gupta et al. 2024). For low-redshift quasars, we followed the work of J. E. Greene & L. C. Ho (2005) and estimated $L_{\rm bol}$ through a combination of their $L({\rm bH}\alpha) - L_{5100}$ relation and the aforementioned $f_{\rm bol}(5100\text{Å}) = 9.26$. We do not account for any beaming or lensing effects when estimating $L_{\rm bol}$ for our quasars.

Our estimates of $M_{\rm BH}$ were also derived in the same manner as in Q. Wu & Y. Shen (2022), based on "single-epoch" prescriptions as detailed in Y. Shen et al. (2011) (see detailed discussion in Y. Shen 2013). In such prescriptions, the measured luminosity acts as a proxy for the BLR size, while the width of the broad lines is assumed to trace the velocity of the (virialized) BLR gas. The $M_{\rm BH}$ prescriptions we used are of the general form of

$$\log\left(\frac{M_{\rm BH}}{M_\odot}\right) = a + b\log\left(\frac{\lambda L_\lambda}{10^{44}\ {\rm erg\ s^{-1}}}\right) + 2\log\left(\frac{\rm FWHM}{\rm km\ s^{-1}}\right), \quad (1)$$

where a and b are parameters derived from reverberation mapping campaigns. Following Y. Shen et al. (2011), who in turn use the M. Vestergaard & B. M. Peterson (2006) calibrations, we assumed $(a, b) = (0.91, 0.50)$, (0.74, 0.62), and (0.66, 0.53) when estimating $M_{\rm BH}$ from the Hβ, Mg II, and C IV broad lines, respectively (i.e., when using L_{5100}, L_{3000}, and L_{1350} as L, respectively). While there are many other calibrations available in the literature, our choice is, again, identical to what was used in Q. Wu & Y. Shen (2022) and in previous large SDSS quasar catalogs.

For the lower-redshift AGNs where a reliable broad Hα component (i.e., bHα) is detected and measured, we employ a somewhat different prescription, following the one suggested by J. E. Greene & L. C. Ho (2005) but rescaling it slightly by +0.125 dex following J. E. Mejía-Restrepo et al. (2022):

$$\log\left(\frac{M_{\rm BH}}{M_\odot}\right) = 6.43 + 0.55\log\left(\frac{L({\rm bH}\alpha)}{10^{42}\ {\rm erg\ s^{-1}}}\right) + 2.06\log\left(\frac{\rm FWHM[bH\alpha]}{10^3\ {\rm km\ s^{-1}}}\right). \quad (2)$$

This latter prescription naturally avoids the potential (residual) contribution of the host galaxy to L_{5100}. Here, too, we prefer a $M_{\rm BH}$ prescription that is consistent with previous compilations of (low-redshift) AGNs over several alternatives (H. Cho et al. 2023; E. Dalla Bontà et al. 2025).

The significant uncertainties on such $M_{\rm BH}$ prescriptions, of order $\gtrsim$0.3 dex, are dominated by systematics on all their subcomponents (with an additional contribution from the intrinsic scatter in the underlying scaling relations): the $R_{\rm BLR}$–L relation(s), the assumption that the BLR itself is virialized, that the measured line width provides a robust probe of these dynamics, and that the overall scaling is correct. The reviews by Y. Shen (2013) and B. M. Peterson (2014)

elaborate on these issues. Moreover, as extensively discussed in several studies (e.g., Y. Shen et al. 2011; Y. Shen & X. Liu 2012; B. Trakhtenbrot & H. Netzer 2012; J. E. Mejía-Restrepo et al. 2016, and references therein), the $M_{\rm BH}$ estimates based on the C IV line are subject to greater uncertainty compared to those based on the lower-ionization lines (the Balmer and Mg II lines), likely due to nonvirial gas dominating the C IV-emitting region. Alternatively, the C IV emission may arise from a combination of gas components with different kinematic properties (e.g., virialized and outflowing components), which can alter the reverberation behavior and introduce additional uncertainty in the inferred radius–luminosity relation and mass estimates (e.g., K. D. Denney 2012). In practice, this uncertainty could be traced by blueshifted and/or highly asymmetric C IV profiles, or by the poor correlation between the width of C IV and the width of Hβ and/or Mg II. Moreover, C IV is often observed to be narrower than the Balmer lines, in contradiction to what is expected from reverberation mapping and BLR models. However, in a large sample of quasars such as ours, it is essentially impossible to treat such issues on a case-by-case basis, particularly at $z \gtrsim 1.5$, when none of the other broad emission lines we consider can serve as a benchmark.

To obtain the most reliable estimates of $M_{\rm BH}$ and $L_{\rm bol}$, we utilized all the emission lines and continuum windows available for each of our spectra. We prioritized higher-quality spectral measurements, focusing on emission lines that are observed far from the edges of the spectral coverage. We first applied cuts on line measurement quality, utilizing the uncertainty assessments provided by `PyQSOFit`. Spectral measurements were rejected if they had either $\Delta{\rm FWHM}/{\rm FWHM} > 40\%$ or $\Delta\lambda L_\lambda/\lambda L_\lambda > 20\%$. Additionally, we excluded all the invalid measurements (`PyQSOFit`-based values were `nan`, `inf`, or zero). Next, we implemented a set of redshift-dependent selection criteria for certain emission-line complexes. For low-redshift AGNs ($z < 0.4$) we prioritized Hα, while for higher redshifts we used Hβ as the primary line for $0.4 \leqslant z < 0.7$, Mg II for $0.7 \leqslant z < 2$, and C IV for $z \geqslant 2$. These choices are, again, similar to what was done in previous SDSS-related quasar catalogs (e.g., B. Trakhtenbrot & H. Netzer 2012; Q. Wu & Y. Shen 2022). If the primary redshift-dependent emission line failed to meet the basic quality criteria, we adopted a predefined fallback sequence: Hα was replaced by Hβ; Hβ by Hα (preferred for lower z) or Mg II (if Hα is unavailable, at $z \gtrsim 0.4$); Mg II by Hβ (preferred) or C IV (at $z \gtrsim 1.5$); and C IV by Mg II (at $2 \lesssim z \lesssim 2.3$). For each $M_{\rm BH}$ estimate retained, we also kept the corresponding $L_{\rm bol}$ measurement,[33] and kept only spectra with a valid $M_{\rm BH}$ estimate.

After estimating $M_{\rm BH}$ and $L_{\rm bol}$ for our sources, we combined these quantities to estimate $L/L_{\rm Edd}$, assuming

$$L/L_{\rm Edd} \equiv \frac{L_{\rm bol}}{L_{\rm Edd}} = \left(\frac{L_{\rm bol}}{\rm erg\ s^{-1}}\right)\left[1.3 \times 10^{38}\left(\frac{M_{\rm BH}}{M_\odot}\right)\right]^{-1}, \quad (3)$$

which is appropriate for pure hydrogen gas. This choice is, again, identical to Q. Wu & Y. Shen (2022), but provides $L/L_{\rm Edd}$ estimates that are higher (by 0.06 dex) than those derived assuming a solar-metallicity gas, as is the case in many other studies.

[33] For example, if the Mg II line measurement was excluded, then $L_{\rm bol}$ would *not* be based on L_{3000}, but instead on the representative luminosity of the spectral region that was ultimately chosen (L_{5100}, L_{1350}, or L(bHα), depending on the redshift).

3.4. Final Refinements

While the procedures described in Sections 3.2 and 3.3 operated automatically over our entire sample of quasars, we recall that this sample spans a wide range of redshifts and data quality, and includes a variety of subclasses, including, e.g., BAL quasars and significantly reddened sources. We therefore chose to further visually inspect several subsets of extreme cases (totaling ≈2000 spectra), in order to identify recurring issues and to ultimately refine the sample and measurements used in our analysis. We specifically examined quasars with extremely high or low values of $L_{\rm bol}$ or $M_{\rm BH}$, or high redshifts. Upon visual inspection, we identified additional cases with incorrect redshift assignments, BAL quasars for which `PyQSOFit` provided exceptionally large FWHMs (resulting in unrealistically high $M_{\rm BH}$), and other issues. All those problematic cases (totaling ≈200 spectra) were excluded from our spectral analysis sample. However, spectra that did not fall into these extreme categories were not visually inspected, and therefore potential issues in such cases may remain undetected. For example, BAL quasars that yielded reasonable $M_{\rm BH}$ and $L_{\rm bol}$ values were not flagged and likely remain in the sample. An example of a misclassified redshift is shown in Figure 6. Notably, all quasars processed through `PyQSOFit` had no `ZWARNING` or other quality flags indicating issues with redshift determination, meaning these problematic spectra could only be identified through spectral fitting and visual inspection.

We thus consider our final set of spectral measurements to be as statistically reliable as other, similarly large compilations of quasar properties that were analyzed with automatic procedures and could not be visually inspected in a comprehensive way.

4. Results and Discussion

In this section, we discuss the characteristics of the large quasar sample(s) selected through the novel selection methods utilized in SDSS-V/BHM, including the efficiency of the selection methods themselves, their photometric properties, and their distributions in key SMBH-related properties ($L_{\rm bol}$, $M_{\rm BH}$, and $L/L_{\rm Edd}$).

4.1. Final Sample

Our final sample of quasars can be considered in four different ways, as described in Table 1. First, the total sample excludes the spectra with catastrophic errors, but includes all other spectra: both *good* and *bad* spectra of quasars, as well as galaxies and stars. Second, the quasars-only subset focuses on all the quasars within the total sample, both *good* and *bad*. These two samples are essential for getting a complete picture of all the objects selected through the GUA and Skewt_QSO selection methods and are used for quantifying their efficiency. Next, the *good* quasars sample includes only those quasars for which the determination of redshifts and other basic (meta)data is robust (see Section 3.1). This subsample is used to investigate the photometric properties of the quasars, and—more importantly—as the input for the spectral measurement procedure (Section 3.2). Finally, the subsample of quasars with spectral measurements includes those quasars for which we

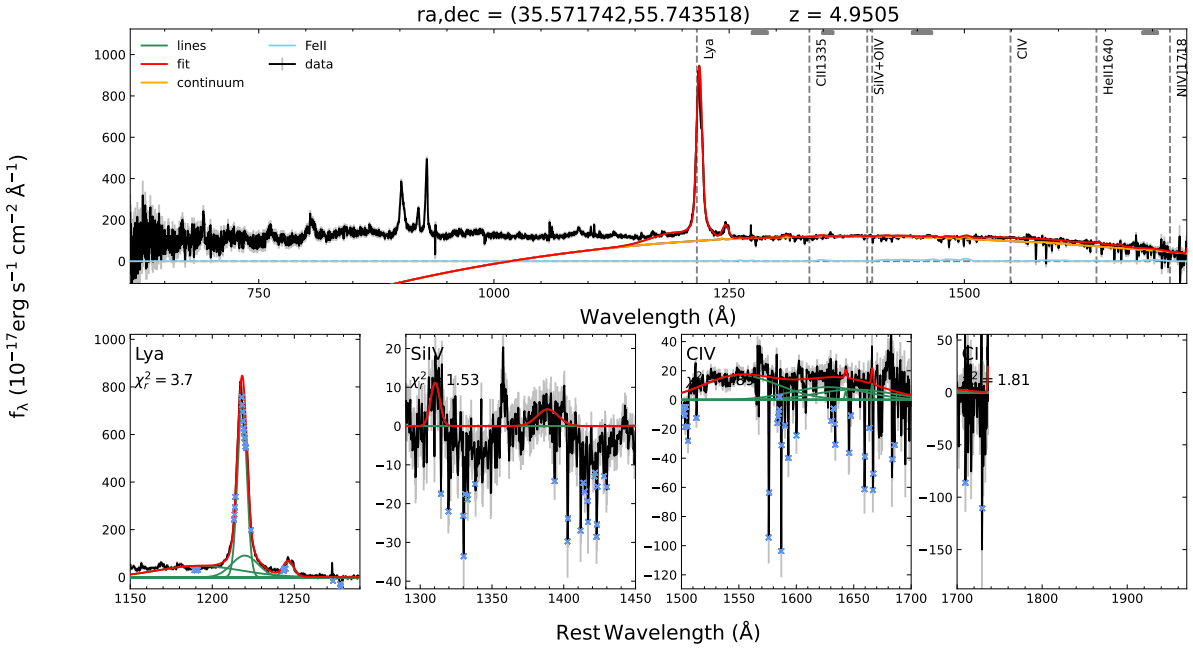

Figure 6. A source misclassified as a high-redshift quasar ($z = 4.95$) by the BOSS pipeline, which we identified in our visual inspection of PyQSOFit measurements yielding extreme quasar properties (in this case, extremely high $L_{\rm bol}$). The legend description is similar to Figure 5. The erroneous redshift assignment is driven by the Hα line being mistaken for Lyα, however the broad Hβ and narrow [O III] lines clearly indicate a redshift of $z \simeq 0.1$.

Table 1
Summary of the Sample Processing

Sample	$N_{\rm tot}$[a]	N[b]			New In SDSS[c]	New Quasars[d]	Sections
		$N_{\rm GUA}$	$N_{\rm Skewt_QSO}$	$N_{\rm S19}$			
Total observed	275,287	263,212	38,029	88,784	⋯	⋯	Section 4.2
Quasars only	259,700	248,023	37,056	85,848	108,923	**76,840**	Section 4.3
Good quasars	220,976	209,876	33,818	77,043	83,495	58,237	⋯
w/Spec. measurements	207,892	197,158	32,595	73,890	75,903	**52,548**	Section 4.4

Notes. Here, "S19" refers to the portion of ASQOSS selected from the original S19 catalogs that is *not* included in either the GUA or Skewt_QSO subsets.
[a] The number of unique objects in the union of the GUA and ASQOSS programs (for ASQOSS: only Skewt_QSO and S19).
[b] The number of objects selected by each of the selection methods (GUA, ASQOSS/Skewt_QSO, and ASQOSS/S19).
[c] The number of new quasars compared with SDSS DR16Q (B. W. Lyke et al. 2020).
[d] The number of new quasars compared with the union of Milliquas/v8 and DESI DR1.

were indeed able to derive key AGN and SMBH properties ($L_{\rm bol}$, $M_{\rm BH}$, $L/L_{\rm Edd}$), based on the detailed spectral decomposition (and subsequent visual inspection; see Section 3.4). Table 1 further illustrates how each of these subsamples can be further divided based on the selection method that is associated with each spectrum (i.e., mainly GUA or Skewt_QSO). This structured approach provides a comprehensive assessment of our sample while also focusing on the most reliable data for each aspect of the detailed analysis that we will present below.

To assess how many newly identified quasars are provided by the GUA and Skewt_QSO samples within SDSS-V DR20, we crossmatched the quasars-only sample(s) with the largest compilation of spectroscopically confirmed quasars (and AGNs) publicly available at the time of writing: a union of the Milliquas/v8 compilation (E. W. Flesch 2023) and the DESI DR1 AGN/QSO Value-Added Catalog (VAC; DESI Collaboration et al. 2026).[34] Milliquas/v8 already contains the largest previous SDSS-based quasar catalog, based on SDSS DR16Q (B. W. Lyke et al. 2020). For our purposes, we downselected only those Milliquas/v8 entries classified as type "Q" or "A" (i.e., having broad emission lines).[35] The crossmatch was done using a 3″ threshold, and ignoring any potential mismatches in redshifts, subclassifications, etc. (e.g., quasar versus type II AGN). Our sample has a total of 151,972 newly spectroscopically confirmed quasars, with 76,840 of these belonging to the so-called "GUA+Skewt_QSO sole-selected" subset, meaning they were not selected by either the core programs or the CSC program (see Section 2.2.4). For 116,487 (52,548 sole-selected) of our new quasars, we obtained satisfactory estimates of $L_{\rm bol}$ and $M_{\rm BH}$ (and thus of $L/L_{\rm Edd}$). The majority of the newly identified quasars (112,512, or 50,628 sole-selected) are observed from the southern SDSS-V site (at LCO), and—importantly—more than half (120,879, or 55,933 among the sole-selected) are located in the southern celestial hemisphere ($\delta < 0$), as can be seen in Figure 7. In contrast, the corresponding fraction among the AGNs in the Milliquas/v8+DES DR1 superset is merely ≈23% (495,252 of 2,188,595 objects). This result clearly

[34] See https://data.desi.lbl.gov/doc/releases/dr1/vac/agngal/.

[35] See Milliquas-ReadMe.

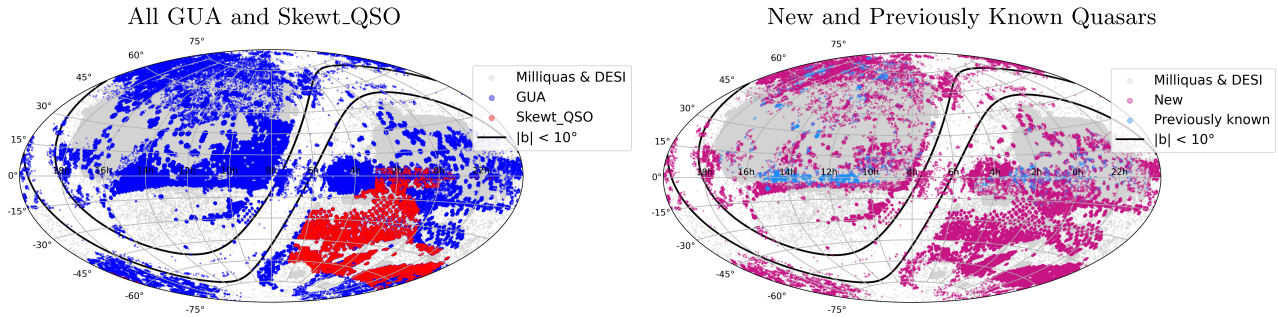

Figure 7. All-sky maps of the quasars observed within the GUA and Skewt_QSO samples (including both *good* and *bad* samples), compared with previously known quasars (gray; from Milliquas/v8 and DESI DR1). Left: the GUA (blue) and Skewt_QSO (red) samples. Right: the union of the two sole-selection subsets, separated to newly identified quasars (magenta) and previously known ones (light blue). In both panels, the various symbols are (over)plotted in the order indicated by the respective legend(s). In practice, there are many GUA quasars in the same part of the sky dominated by Skewt_QSO quasars (i.e., blue points underneath the red points in the left panel), and many Milliquas and DESI quasars in the regions that also have newly discovered quasars (i.e., gray points underneath magenta ones in the right panel).

demonstrates the great impact that SDSS-V/BHM is having on the census of AGNs across the entire sky.

The redshifts of the quasars in our sample span $0.0013 \lesssim z \lesssim 5.3$, and their full redshift distribution is shown in Figure 8, compared to the well-established SDSS DR7Q (D. P. Schneider et al. 2010) and SDSS DR16Q (B. W. Lyke et al. 2020) quasar catalogs. Both the differential and cumulative distributions demonstrate that our sample spans this redshift range in a continuous manner, without any obvious drops or gaps. Moreover, among the newly discovered quasars, 575 (i.e., 0.22%) have redshifts greater than 4, while the Milliquas/v8 +DESI DR1 superset contains 8,301 quasars (0.38%) with $z > 4$. The number of $z > 4$ quasars in our sample(s) is based on extensive visual inspection and filtering of all the sources initially flagged as being at $z > 4$ (as explained in Section 3.4). Given the challenges involved in correctly identifying such high-redshift quasars (in general and in SDSS-V DR20), we defer any further discussion of this population to a future publication.

4.2. Quasar-selection Purity

An essential aspect of evaluating the quality of the GUA and Skewt_QSO samples is assessing their purity (or efficiency), i.e., the fraction of spectroscopically confirmed quasars[36] identified among the larger set(s) of quasar candidates. We investigated the purity achieved by the selection methods and observing strategies of both the GUA and Skewt_QSO samples. To account for the complex target selection and observing strategy, and in particular the significant overlap between the target pools of GUA, Skewt_QSO, and/or higher-priority programs (e.g., SPIDERS), we compute and use an *homogenized purity* (hereafter *hp*) for each selection method.[37] In this procedure, which is described in detail in Appendix A, we focus on a "test area"—located within the high-Galactic-latitude southern sky—where there are targets selected by GUA, Skewt_QSO, and SPIDERS. We split all the targets of these three selection methods into seven mutually exclusive subsets based on cotargeting flags, which are (conceptually) presented in Figure A2. We then calculate the quasar-selection purity within each such subset, and average over the (relevant)

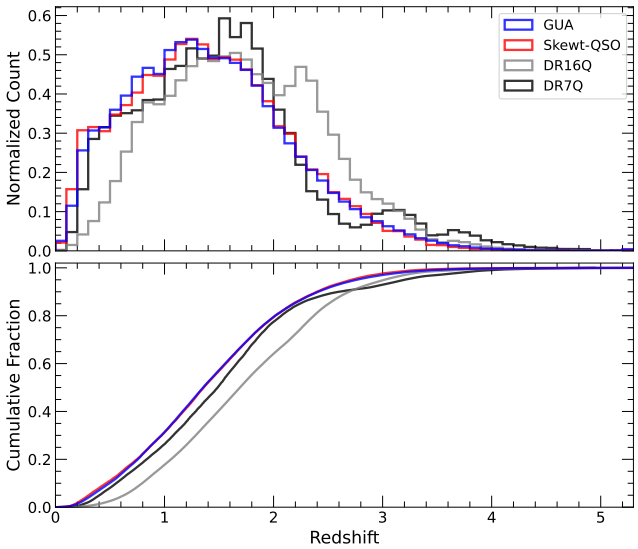

Figure 8. The differential (top) and cumulative (bottom) distributions of redshift for the quasars in our two samples (GUA and Skewt_QSO), compared to the SDSS DR7Q (D. P. Schneider et al. 2010) and SDSS DR16Q (B. W. Lyke et al. 2020) reference quasar catalogs.

subsets weighing by their fractional contribution to the full target pool to derive *hp* for GUA and Skewt_QSO (see Equations (A1) and (A2), respectively). In these calculations, we considered spectroscopically confirmed quasars from both the *good* and *bad* subsets as the success metric of the two selection methods (i.e., the numerator in the fraction that quantifies the purity).

Figure 9 presents the *hp* of the two main samples considered, GUA and Skewt_QSO, including both the *good* and *bad* subsets. Both the GUA- and Skewt_QSO-related selection methods (and survey strategies) achieved exceptionally high efficiencies of $hp = 96.2\%$ and 97.4%, respectively. If only the *good* quasars are counted as genuinely identified quasars (i.e., excluding the *bad* subset from the *hp* nominators), the corresponding purities decrease to $hp = 82.0\%$ for GUA and 88.9% for Skewt_QSO, providing a conservative lower limit on the selection purity.

In Figure 10, we present the *hp* of both selection methods as a function of brightness (*G*-band magnitudes; left) and the *hp* and standard purity as a function of Galactic latitude ($|b|$; right). The AGN selection purity (within the test area) remains exceptionally high, $hp > 90\%$, across the range $17 < G < 20$

[36] Or, more precisely, broad-line systems with significant AGN contributions to their continuum emission.

[37] Whenever we use "purity" in what follows, we refer to *hp*, unless noted otherwise.

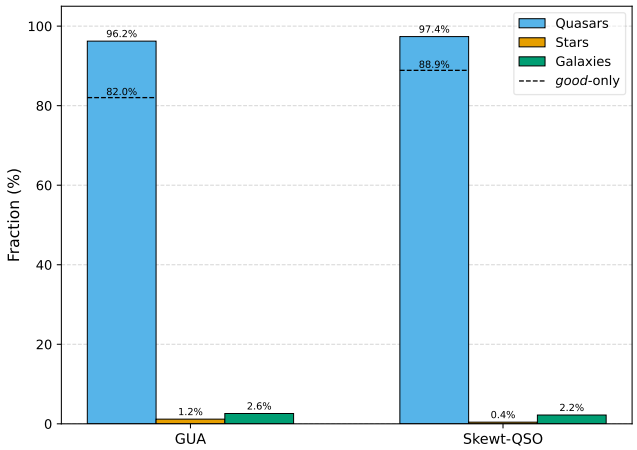

Figure 9. The purity (efficiency) of the quasar selection among the GUA and Skewt_QSO samples. The two bar charts show the *hp* of the GUA and Skewt_QSO samples. Notably, both of the methods achieved very high purity, over 96%. The black dashed lines indicate the corresponding lower limits, obtained when only *good* quasars are counted as genuine quasars (see text).

mag. For fainter sources, $20 < G < 21$, the purity of GUA remains high, while that of Skewt_QSO declines to well below 60%. This is likely driven by the combination of two limiting factors. First, at the target-selection stage, the Gaia DR2 astrometric uncertainty at $G \gtrsim 20$ mag is such that it becomes increasingly challenging to differentiate between MW stars and quasar candidates based on proper motion (or lack thereof; see Section 2.1 and L. Lindegren et al. 2018). Second, the SDSS-V/BHM observations, and particularly the shorter exposure times associated with many of our ASQOSS targets, naturally limit the quality of the acquired spectra, and thus of the ability to robustly classify such faint sources (see Section 2.2.5). Figure 10 also shows the corresponding lower limits, computed by counting only *good* quasars as genuine quasars. These lower limits closely track the fiducial *hp* curves across most of the brightness and Galactic latitude ranges, indicating that the general trends seen across these quantities are robust to the choice of quality cuts.

The right-hand panel of Figure 10 shows that the *hp* of both GUA and Skewt_QSO remains high across the entire high-Galactic-latitude range probed by the test area, $30° \leqslant |b| \leqslant 70°$. To further check how efficient the S19-inspired GUA selection is at lower Galactic latitudes, we considered the overall (i.e., not homogenized) quasar-selection purities among two GUA-focused samples, going beyond the test area: (1) *all* the sources within our data that are part of the GUA target pool, but not part of any core program, and (2) the subset of such sources that are located at northern declinations ($\delta > +5°$), thus eliminating any overlap with Skewt_QSO. The quasar-selection purities of these two large, GUA-focused samples are shown as dashed and dotted blue lines, respectively, in the right panel of Figure 10. The purity of the GUA selection remains rather high even at low Galactic latitudes, dropping below ≈80% only at $|b| \lesssim 5°$. The high-Galactic-latitude extensions of these two GUA-focused samples again demonstrate the high efficiency of GUA to select quasars in parts of the sky that are not covered by other key quasar-selection methods and/or (core) quasar/AGN-oriented programs within SDSS-V.

Comparing the overall efficiencies of the S19-inspired GUA method and the YS23-based Skewt_QSO one at face value, it appears that the latter is slightly more efficient in identifying quasars (*hp* = 96.1% versus 97.4%, respectively). This may not be surprising given that the Skewt_QSO selection incorporated more nuanced multiband optical photometry, as well as NIR photometry, available the high-Galactic-latitude DES footprint (and thus our test area; Section 2.1). Indeed, Figure 10 clearly shows that the purity of GUA approaches that of Skewt_QSO for $|b| > 30°$.

The quasar census enabled by the two quasar-selection methods we study and their implementation within SDSS-V is, in fact, extremely efficient compared to past efforts. For reference, we first note that the legacy SDSS spectroscopic quasar-selection procedure (G. T. Richards et al. 2002) was designed to achieve a purity of ∼65%. Even under our conservative calculations, which focus only on *good* quasars, the purity of both GUA and Skewt_QSO (⩾82%) still substantially exceeds that of the legacy SDSS procedure of G. T. Richards et al. (2002). That pioneering high-Galactic-latitude quasar-selection methodology had to rely solely on five-band optical photometry, with essentially no astrometric data, thus being challenged by a high incidence rate of foreground stars.

We next compare the performance of the two methods to the expectations set in the studies that designed them. The original S19 work estimated the efficiency of their C75 catalog at ≈79% reliability, while YS23 estimated the efficiency of the Skewt_QSO selection at ≈80%. We note that even the conservative *good*-only lower limits meet (or indeed exceed) these original expectations. It thus appears that the SDSS-V implementation of the S19 approach, that is GUA, is outperforming the expectations in the original S19 work, likely due to the more stringent cuts made in the former compared with the latter ($P_{\rm RF} \geqslant 0.8$ versus ⩾0.69, respectively). As for the Skewt_QSO method, the higher-than-expected purity of the Skewt_QSO set (97% versus 80%) clearly shows that this method is exceptionally efficient in identifying objects that are, indeed, quasars, even if our knowledge of their nature could perhaps be attributed to other programs and selection methods employed in SDSS-V (which also dictate longer exposures under preferable conditions).

To better understand the limitations of the two selection methods, we examined the types of sources that "contaminated" the quasar sample(s) the most. Based on the BOSS pipeline spectral classifications, MK-type stars are the type of stellar contaminants in our data, comprising ≈51% of the stars in the GUA sample and ≈29% of the stars in the Skewt_QSO sample. More particularly, M-type stars comprise ≈30% and ≈12% of the two samples, respectively. The rest of the stellar contaminants are distributed across essentially all the spectral classes covered by the BOSS pipeline, including spectra classified as white dwarfs, cataclysmic variables, and more. This dominance of MK-type stars is not surprising, given their intrinsically high abundance in the MW, as well as their intrinsically faint and red SEDs, which could be confused with those of (high-redshift) quasars. Indeed, their dominance is particularly pronounced within the GUA sample, which reaches fainter flux levels. We note that, by definition, the MK-type stars in our sample have no obvious proper motion or parallax as of Gaia DR2, and moreover, many of them are found at high Galactic latitudes. Such stars could be of interest for studies of the structure and evolution of the MW. In any case, more sensitive Gaia data (e.g., DR3 and DR4), with more accurate proper-motion measurements, are expected to help

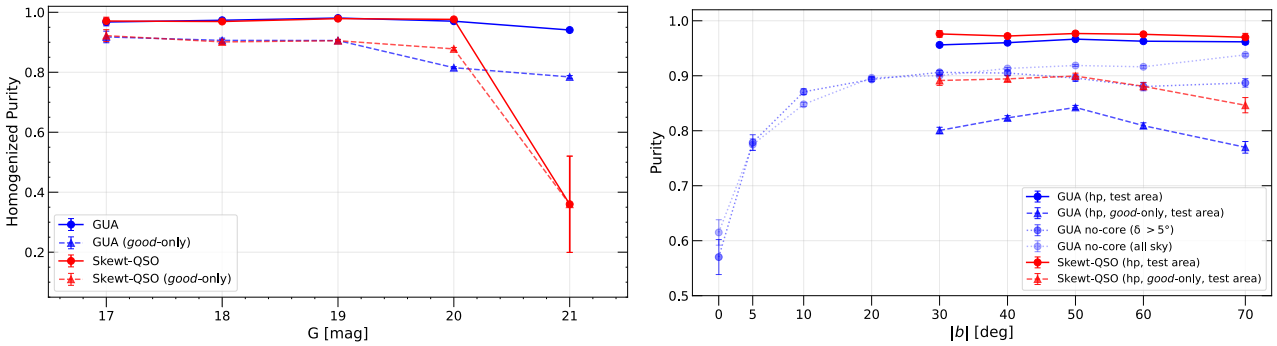

Figure 10. The quasar-selection purity of the GUA and Skewt_QSO samples as a function of source brightness (left) and Galactic latitude (right). In both panels, the dashed lines with upward triangles show the corresponding lower limits, obtained by counting only *good* quasars as genuine quasars. Error bars represent standard binomial errors. Left: the quasar-selection homogenized purity (*hp*; explained in Appendix A) of the GUA and Skewt_QSO samples as a function of Gaia *G*-band magnitudes. Both samples are highly pure for $G < 20$ mag. Right: the quasar-selection *hp* of the GUA and Skewt_QSO samples as a function of Galactic latitude, *b*. Here, the blue and red solid lines are the *hp* in the "test area" of the GUA and Skewt_QSO samples, respectively. The blue dotted and dashed lines are the overall (i.e., not homogenized) purities of the GUA sole-selection sample, for the full sky and for $\delta > +5°$, respectively. The S19-based GUA program retains good quasar-selection efficiency even when targeting low Galactic latitudes.

improve the purity of quasar-selection efforts that rely on S19- or YS23-like approaches. Such improvements are beyond the scope of the present analysis and are *not* expected to be fully implemented in the final phases of SDSS-V operations.

A detailed analysis of how the selection purity may change with quasar redshift, which would require assessing a large input quasar sample with known redshifts combined with nonquasar contaminants, is beyond the scope of the present paper. We note, however, that neither of the two original quasar-selection method studies (S19; YS23), nor the two main samples of observed quasars considered here show significant drops or gaps across the redshift range we probe. Moreover, the redshift distributions shown in Figure 8 suggest that our quasar samples fare better than those from the legacy SDSS quasar-selection method with the challenging $z \approx 2.1-2.8$ regime (see, e.g., J. Bovy et al. 2011; N. P. Ross et al. 2012).

We conclude that both the GUA and Skewt_QSO selection methods, and their implementation within SDSS-V, provide an exceptionally pure way to construct large quasar samples, outperforming some legacy quasar surveys. Moreover, the S19-based GUA selection enables a high-purity identification of quasars over the entire sky, including at low Galactic latitudes. This is beneficial not only for quasar demographics and related science, but also for efforts directed at other types of astronomical objects (i.e., the S19 and YS23 catalogs can be used to rule out the nonquasar nature of certain sources). In what follows, we investigate the similarities and differences between the key properties of the quasars uncovered by these two selection methods and those provided by legacy (SDSS-based) quasar samples.

4.3. Photometric Properties

Here, we present the basic photometric properties of the quasars in our sample(s). We utilize the multiband photometry available for our sources either in the visual regime (i.e., SDSS-like *griz* bands) or in the MIR (i.e., WISE-based), to compare our sample(s) to those from other well-established selection methods. The photometric data that we use here are described in Section 2.3. In what follows, all photometric comparisons are performed on the full quasar sample, including both *good* and *bad* subsets, unless otherwise noted.

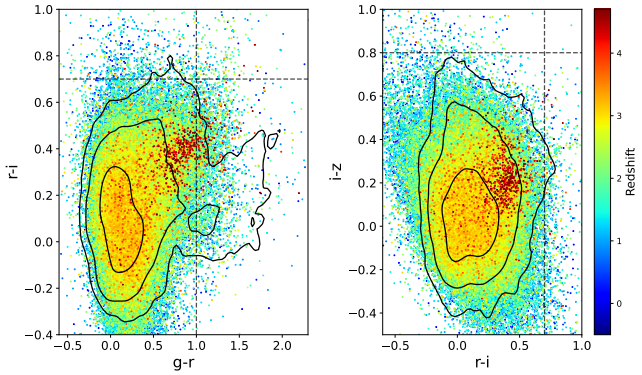

Figure 11. The distributions of our quasar samples in optical color–color space, specifically $r - i$ versus $g - r$ (left) and $i - z$ versus $r - i$ (right). In both panels, black contours represent quasars from the legacy SDSS DR7Q (D. P. Schneider et al. 2010), with enclosed regions corresponding to the 68%, 95%, and 99% percentiles. The colored dots represent GUA and Skewt_QSO data, color-coded by redshift (see color bar). The GUA and Skewt_QSO quasars are generally consistent with those of the legacy SDSS DR7Q sample; however, both GUA and Skewt_QSO allow us to probe redder quasars, as expected from their selection based on NIR and MIR data. The black dashed lines indicate the color thresholds ($g - r > 1$, $r - i > 0.7$, and $i - z > 0.8$) used to define a "red quasar" subsample, analyzed in Section 4.3.1.

Figure 11 presents the distribution of our quasars in the $r - i$ versus $g - r$ (left) and the $i - z$ versus $r - i$ (right) color–color spaces, compared with the SDSS DR7 quasar sample (D. P. Schneider et al. 2010). We note that the latter were selected mostly through a *ugriz* multicolor selection procedure (G. T. Richards et al. 2002), and therefore are meant to represent the legacy census of spectroscopically confirmed quasars, rather than the "intrinsic" range of quasar SEDs. As Figure 11 shows, most of our quasars occupy the same part of the color–color space as those of the legacy SDSS DR7Q sample, except that the GUA+Skewt_QSO samples generally cover a wider range in both $r - i$ and $i - z$ colors (e.g., $r - i \gtrsim 0.65$ or $\lesssim -0.3$; $i - z \gtrsim 0.7$ or $\lesssim -0.3$ when $r - i \gtrsim 0.4$). Notably, these extended regions in color–color space, and particularly the redder regions ($r - i \gtrsim 0.65$, $i - z \gtrsim 0.7$), are *not* dominated by high-redshift quasars (see color bar). Thus, our quasar samples include some redder quasars, either due to their intrinsically redder SEDs or some form of line-of-sight extinction by dusty gas, located either on

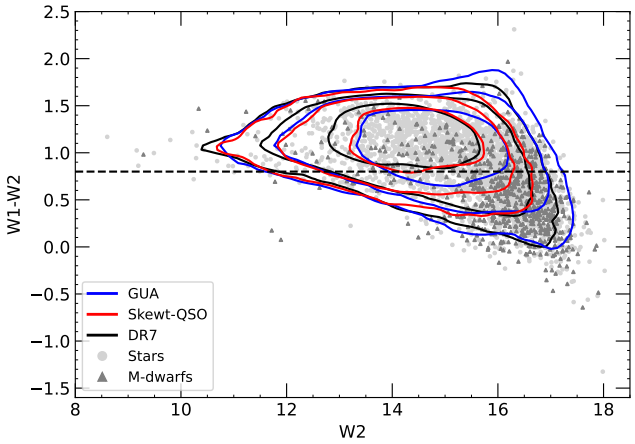

Figure 12. The MIR W1 − W2 versus W2 diagnostic diagram for various sources from our quasar samples. The blue, red, and black contours represent the distributions for GUA, Skewt_QSO, and DR7Q, respectively (percentiles identical to those in Figure 11). The black dashed line marks the W1 − W2 > 0.8 threshold for AGN selection (based on D. Stern et al. 2012). Most of the quasars in our sample(s) (84%) lie above the suggested cutoff, consistent with the DR7Q distribution (89%). Gray dots represent the stars in the sample(s), with M-type stars further highlighted as triangles.

circumnuclear or host-galaxy scales. This demonstrates how the new, SDSS-V-enabled quasar samples can be used to extend quasar science beyond the regime dominated by intrinsically blue pointlike sources.

In Figure 12, we show the distribution of the two main quasar samples in the W1 − W2 versus W2 space (as contours). D. Stern et al. (2012) proposed that selecting sources with IR colors W1 − W2 ⩾ 0.8 would yield an AGN sample with both a high completeness (78%) and high efficiency (95%). Consistent with the findings of D. Stern et al. (2012), the vast majority of our quasars, ≃84% meet the W1 − W2 ⩾ 0.8 threshold, while ∼16% of the sources fall below this suggested cut, implying that adopting this criterion would miss ∼16% of our quasars. Moreover, the sample completeness of 84% is somewhat higher than that found by D. Stern et al. (2012; of 78%). This is consistent with the DR7Q distribution, where 89% of quasars satisfy the W1 − W2 ⩾ 0.8 criterion. In addition, the efficiency we find for this MIR color cut, calculated based on the 176,592 quasars that have both W1 and W2 measurements,[38] is ≃96%. This is higher than the 95% efficiency reported in D. Stern et al. (2012).

To assess the efficiency of the D. Stern et al. (2012) MIR cut, we also consider the stars identified through the GUA- and Skewt_QSO-led SDSS-V observations, i.e., the contaminants of our sample (see points in Figure 12). It is evident that the stars exhibit colors similar to those of the quasars in our sample, making it indeed challenging to distinguish between the two populations based solely on their MIR colors. Furthermore, most of the M-type stars—which are the most common contaminants of the spectroscopic sample (see Section 4.2)—are located below the W1 − W2 = 0.8 cutoff line, suggesting that selecting objects above this threshold would result in fewer M-type stars contaminating our sample. We stress that the target-selection procedures for our quasar samples did not explicitly use either the D. Stern et al. (2012) cut or more elaborate MIR selection criteria (e.g., S. Mateos et al. 2012; R. J. Assef et al. 2013, and references therein). The resulting quasar sample and Figure 12 show that adopting such a W1 − W2 cut could have led to a smaller sample of quasars, with many fewer M-type stellar contaminants, however still with a nonnegligible number of other stellar contaminants. Moreover, imposing the W1 − W2 cut may limit the opportunity to access certain (mildly) obscured intermediate-redshift AGNs (see below; e.g., S. M. LaMassa et al. 2016).

[38] Out of 184,558 objects in total.

4.3.1. Preliminary Search for Red Quasars

We can further use our sample of spectroscopically confirmed quasars to assess the ability of the novel selection methods to identify significantly red sources, compared to previously suggested (extreme) color cuts. While there are several published selection criteria and samples of such objects (e.g., T. Urrutia et al. 2009; M. Banerji et al. 2012; E. Glikman et al. 2012; V. A. Fawcett et al. 2020; S. LaMassa et al. 2024, and references therein), here we focus on the optical-to-MIR color diagnostics studied in detail by S. M. LaMassa et al. (2016) and F. Hamann et al. (2017), as well as the MIR-only diagnostic of R. J. Assef et al. (2015) to search for hot DOGs.

According to S. M. LaMassa et al. (2016), the R − W1 color can be used to identify highly obscured AGNs, where the optical emission (measured by the R band) is diminished while the MIR emission (W1) remains strong as it is dominated by the torus, which reprocesses the absorbed (UV–optical) radiation from the central engine. S. M. LaMassa et al. (2016) show that broad-line, unobscured AGNs cover a wide range in R − W1, while highly obscured AGNs have R − W1 > 4. We followed S. M. LaMassa et al. (2016) and estimated the R − W1 color for our quasars, using the available i- and r-band measurements (see above) to derive R (see Equations (1) and (2) in S. M. LaMassa et al. 2016).

In Figure 13, we present the distributions of R − W1 for the quasars in our two main samples. Approximately 26% of the quasars that have both i- and r-band measurements (44,024 of 166,440 *good+bad* quasars) have R − W1 > 4. For comparison, Figure 13 also shows the R − W1 distribution for X-ray-selected broad-line AGNs drawn from the more recent, highly complete Stripe-82X DR3 catalog published by S. LaMassa et al. (2024), as well as the DR7Q catalog (D. P. Schneider et al. 2010). In S. LaMassa et al. (2024), approximately 41% of broad-line AGNs have R − W1 > 4, while only 17% of DR7Q quasars exceed this threshold. Our GUA and Skewt_QSO samples thus lie between these two cases. This suggests that the GUA and Skewt_QSO selection methods recover a substantial fraction of red quasars, though they are slightly skewed toward bluer broad-line AGNs relative to the X-ray-complete sample of S. LaMassa et al. (2024). The lower fraction of red quasars among DR7Q relative to our sample is consistent with the known color bias of the previous selection of SDSS quasars, which primarily targeted blue sources (e.g., G. T. Richards et al. 2002). We also note that a significant fraction of the R − W1 > 4 sources in our sample (≈26%) have W1 − W2 < 0.8, i.e., not satisfying the simple D. Stern et al. (2012) AGN selection cut. This result, which is in qualitative agreement with, e.g., S. M. LaMassa et al. (2016), further demonstrates how our sample probes a wider range in color and obscuration properties than previous samples of (optical) luminous AGNs.

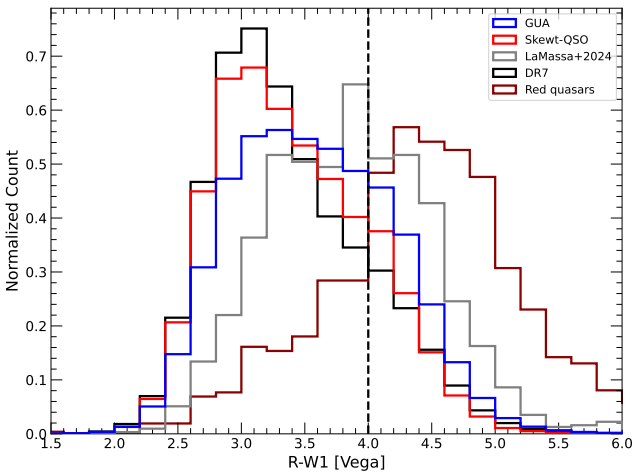

Figure 13. The distribution of the $R-$ W1 MIR–optical color for our quasars compared to S. LaMassa et al.'s (2024) AGN sample. We show the GUA (blue) and Skewt_QSO (red) samples, compared with the sample of broad-line AGNs from S. LaMassa et al. (2024; gray) and DR7Q (black). The dark red line shows the distribution of the "red quasar" subsample, defined as objects with $g-r>1$ and/or $r-i>0.7$ or $i-z>0.8$ within $0.5<z<1.5$. The dashed black line indicates the $R-$ W1 >4 threshold proposed by S. M. LaMassa et al. (2016) for selecting highly obscured (yet broad-line) AGNs. Approximately 26% of the quasars in our combined dataset exceed this threshold.

The 26% of quasars in the sample with $R-$ W1 >4 are not completely obscured, as they do—by definition—exhibit broad emission lines in their SDSS-V spectra. The exact reason for the enhanced MIR emission relative to the optical one, and/or the level of dust attenuation, remains unclear. Further analysis of this subset of GUA and Skewt_QSO quasars is required to better understand the nature of these red quasars, but is beyond the scope of the present paper. Nonetheless, we made a first step toward assessing whether the redder quasars in our sample are driven by real, intrinsically redder colors (i.e., mildly obscured systems) rather than by the redshifting of strong emission lines (e.g., G. T. Richards et al. 2003). We selected a "red quasar" subsample to include the union of objects satisfying $g-1>1$, $r-i>0.7$, and/or $i-z>0.8$, within the redshift range $0.5<z<1.5$. This redshift range focuses on systems where emission-line "migration" through the optical bands is significant, while avoiding host-galaxy-dominated continua at very low redshifts. Figure 13 shows the $R-$ W1 distribution of this subsample (dark red line). Of the 1302 quasars in this subsample, ≈75% pass the $R-$ W1 >4 threshold, which is substantially higher than the ≈39% baseline of the full quasar sample in the same redshift range. This confirms that the optically red quasars in our sample are dominated by genuinely (mildly to moderately) obscured objects.

F. Hamann et al. (2017) presented a diagnostic for selecting extremely red quasars (ERQs), namely $i-$ W3 $\geqslant 4.6$ (AB). Using this diagnostic alone, they found an ERQ fraction of 0.095% based on a sample of 205 ERQs among a sample of 216,188 quasars at redshifts $1.53\leqslant z\leqslant 5.0$, drawn from the SDSS DR12Q catalog (I. Pâris et al. 2014, 2017). Based on their measured spectral features, these objects were interpreted as highly obscured quasars embedded in a dense, patchy medium, likely caught in a brief evolutionary phase of intense feedback and powerful outflows. We followed the same simple color cut for all GUA and Skewt_QSO quasars at $1.53\leqslant z\leqslant 5.0$ that have valid i- and W3-band measurements (66,913 quasars). We find an ERQ fraction of 0.10% (67 quasars)—fully consistent with, albeit slightly lower than, the findings of F. Hamann et al. (2017). In comparison, applying the same selection to the DR7Q quasar yields an ERQ fraction of 0.057%. This again demonstrates the enhanced completeness of the S19- and YS23-based selection methods to quasars with SEDs that are (slightly) redder than those probed by the legacy SDSS color selection (G. T. Richards et al. 2002).

As explained in Section 1, hot DOGs are among the brightest objects in the (mid-)IR sky. The GUA and Skewt_QSO samples are large enough to warrant a preliminary search for such rare sources, which could in principle be selected for SDSS-V spectroscopy (unlike the blue-focused selection of legacy SDSS quasars). In Figure 14, we show the distribution of our quasars in two MIR color–magnitude spaces involving W1 − W2 measurements. To identify potential hot DOGs within our sample, we applied the "W1W2 drop-outs" selection criteria defined by P. R. M. Eisenhardt et al. (2012; see Equations (1) and (2) in R. J. Assef et al. 2015). However, we identified no robust hot DOG candidates among our quasars. We note that a single candidate was rejected following a visual inspection of the WISE images, which revealed the MIR measurements were severely contaminated by a bright nearby star. While the selection method in the present study effectively identifies quasars across wide ranges in color space(s), it is less efficient at detecting hot DOGs. These objects are so heavily obscured that their optical continuum is dominated by the stellar light from the host galaxies, and not the accretion-driven AGN emission. Unlike the latter, the former is significantly fainter in the G band (considering hot DOGs are found at significant redshifts; see, e.g., R. J. Assef et al. 2015), and thus challenging to be included in a relatively shallow optical spectroscopic survey like SDSS-V. It is still possible that (low-redshift) hot DOGs (e.g., G. Li et al. 2023, 2025) may be revealed in ongoing SDSS-V observations or, alternatively, that deeper wide-field spectroscopic surveys could extend similar quasar-selection methods to lower optical flux levels, e.g., 4MOST (R. S. de Jong et al. 2019).

To summarize this part of the analysis, the comparisons we carried out suggest that our quasar samples are not biased toward particularly blue or red colors, and exhibit color properties broadly consistent with those of other well-known quasar catalogs, with only a slight enhanced completeness toward red (mildly obscured) systems. As mentioned in Section 2.1.3, since the selection methods used to identify the samples we study may be somewhat biased against heavily obscured AGNs, the fact we have not identified a significant new population of such sources is not surprising.

4.4. Distributions of L_{bol}, M_{BH}, and L/L_{Edd}

Here, we present the distribution of AGN bolometric luminosity ($L_{\rm bol}$) and SMBH mass ($M_{\rm BH}$) for the quasars in the sample(s). The bolometric luminosity range of the quasars is $41.6\lesssim \log(L_{\rm bol}/{\rm erg\ s^{-1}})\lesssim 48.2$, while the black hole mass range is $5.7\lesssim \log(M_{\rm BH}/M_\odot)\lesssim 10.7$. The central 99% of the sample spans $44.1\lesssim \log(L_{\rm bol}/{\rm erg\ s^{-1}})\lesssim 47.4$ and $6.8\lesssim \log(M_{\rm BH}/M_\odot)\lesssim 9.9$. As explained in Section 3.4, we verified that the extremes of these distributions are not driven by spurious measurements. In particular, we note that the lowest $M_{\rm BH}$ estimate in our sample, which in fact corresponds to the low-redshift galaxy UGC 06728, is fully

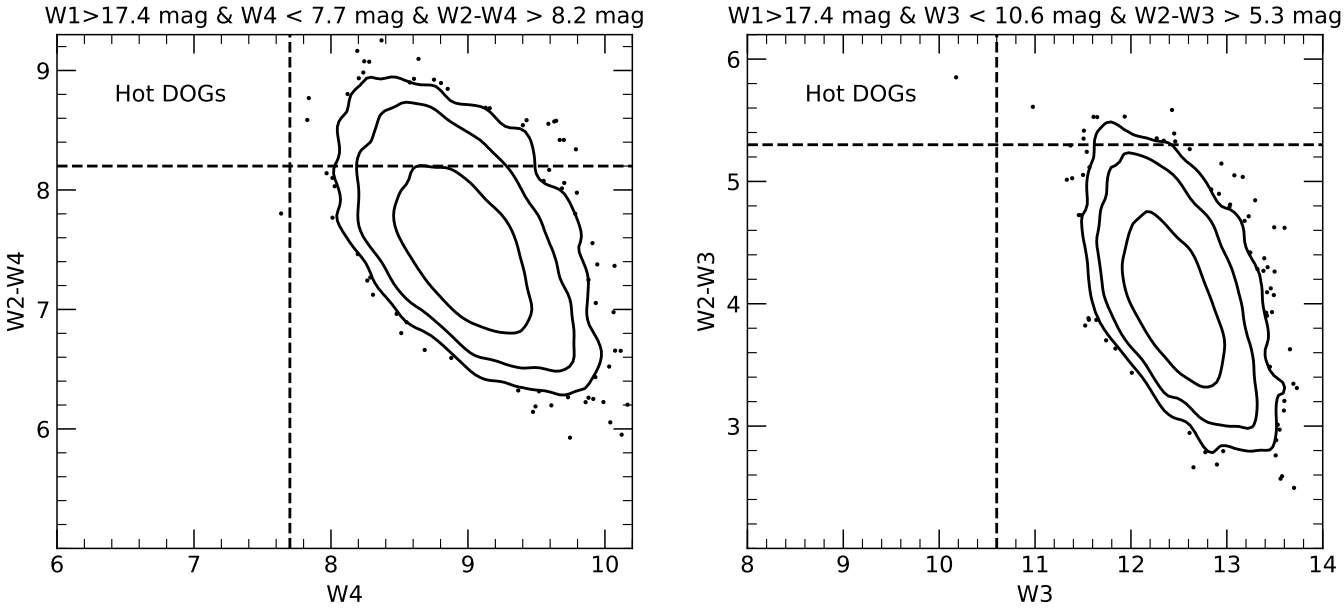

Figure 14. Preliminary search for hot DOGs using MIR diagnostics, following the "W1W2 drop-outs" method from R. J. Assef et al. (2015; black dashed lines). The magnitudes are given in the Vega system. The contours represent the distributions of the quasars in our sample that have W1 > 17.4 (percentiles are identical to those in Figure 11), while the scattered points show the remaining 1% of sources not included in the contours. No hot DOGs were found among the quasars in our sample; one candidate was discarded due to contamination from a bright nearby star.

consistent with a robust, independent measurement (e.g., $\log(M_{\rm BH}/M_{\odot}) = 5.85$ based on reverberation mapping; M. C. Bentz et al. 2016). Finally, essentially all the quasars in our sample have $0.01 \lesssim L/L_{\rm Edd} \lesssim 1$, as is found for many other large samples of broad-line AGNs drawn from highly complete surveys (e.g., A. Schulze & L. Wisotzki 2010; Y. Shen et al. 2011; B. Trakhtenbrot & H. Netzer 2012; A. Schulze et al. 2015; T. T. Ananna et al. 2022).

Figure 15 shows these distributions as a function of redshift, as well as the corresponding distributions of a previous quasar sample (namely, SDSS DR16Q; Q. Wu & Y. Shen 2022), while Figure 16 presents similar distributions in the mass–luminosity plane. Most of the quasars overlap with those from the previous large SDSS-based catalogs, except for a few low-mass, low-luminosity sources that we were able to identify among the GUA and Skewt_QSO samples. Figure 17 further compares the 1D distributions of the three key properties ($L_{\rm bol}$, $M_{\rm BH}$, and $L/L_{\rm Edd}$), as well as the G-band magnitudes, for the various SDSS-V programs within our quasar sample and for the two main previous quasar catalogs, namely SDSS DR16Q (Q. Wu & Y. Shen 2022) and the older DR7Q (Y. Shen et al. 2011). The bottom panels of Figure 17 show the corresponding cumulative distribution functions (CDFs), while Table 2 summarizes the corresponding key statistics, such as medians, averages, and standard deviations.

There are several clear trends seen when comparing the various samples. First, the distribution of G-band magnitudes in Skewt_QSO shows a concentration of objects within a certain brightness range due to increased selection at $19 < G < 20$ mag. Second, there are differences of $\lesssim$1 dex between Skewt_QSO and DR16Q in both $L_{\rm bol}$ and $M_{\rm BH}$, with Skewt_QSO showing higher luminosities and masses. For $L/L_{\rm Edd}$, we do not observe significant differences between the two samples. Skewt_QSO, in fact, appears to be much more similar to the older DR7Q catalog. GUA, on the other hand, falls between Skewt_QSO and DR16Q in terms of $L_{\rm bol}$ and $M_{\rm BH}$. In terms of brightness, the GUA sample exhibits two peaks, indicating the two different subselection samples of GUA-bright and GUA-dark. For $L/L_{\rm Edd}$, the distribution is similar to DR16Q and Skewt_QSO.

To quantify the differences between the samples, we applied statistical tests. First, we applied the traditional two-sample Kolmogorov–Smirnov (KS) test, which yielded extremely low p-values ($P_{\rm value} \ll 10^{-6}$) for all the CDFs under study, indicating that the distributions are different at very high significance levels. We note again that the differences themselves are not large, and the distributions of the various samples generally overlap. Since the KS test depends strongly on sample size and is mainly sensitive to differences near the median of the distributions, with relatively low sensitivity in the tails, we chose to also apply the Anderson–Darling (AD) test, which is known to better address the tails of the distributions that are compared. We obtained extremely low p-values in the AD test, too—again indicating that the distributions of $L_{\rm bol}$, $M_{\rm BH}$, and $L/L_{\rm Edd}$ differ between the samples we study.

A closer inspection of Figure 17 suggests that while the samples do differ from each other, these differences do not fully answer the question of whether the various selection methods uncover distinct populations of accreting SMBHs, as the derived properties depend on luminosities and thus on the flux distributions and limits of the surveys. Specifically, the leftmost panels of Figure 17 clearly demonstrate how our sample(s) differ from the legacy SDSS ones in terms of visual-regime brightness (G-band magnitudes). This is not surprising, as the older SDSS-I/II-based DR7Q followed a rather rigid set of flux limit(s) (e.g., $i < 19.1$ or < 20.2 mag; see G. T. Richards et al. 2002), while the later SDSS-III/IV observations added in DR16Q used different selection criteria, aiming to enhance the number of (faint) quasars at $z \sim 2$–2.5. These differences in fluxes translate to differences in luminosities, in a way that also involves the distribution of redshifts (by itself potentially affected by selection criteria defined by colors and/or variability). The differences in luminosities then affect the

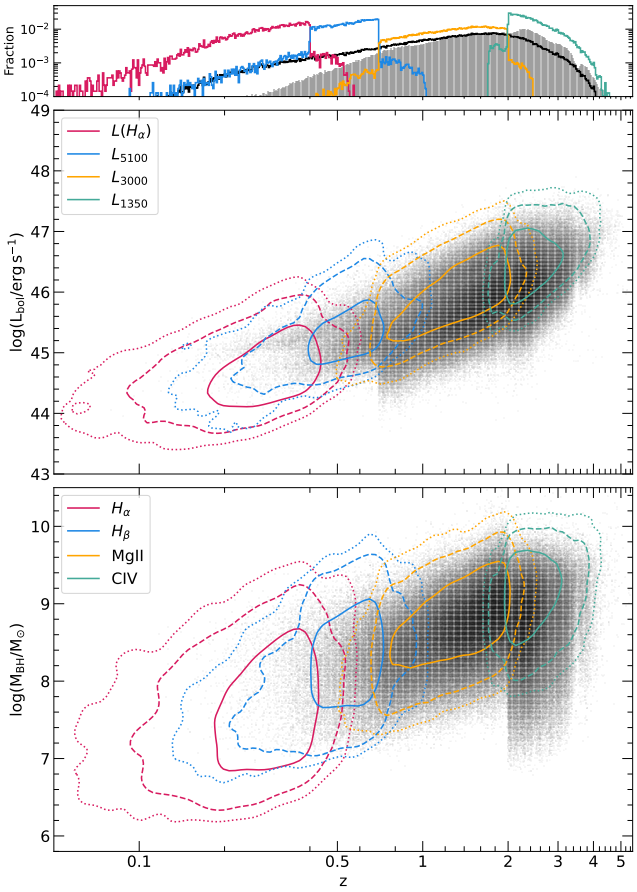

Figure 15. The distributions of our quasars in the $L_{\rm bol}$–z (top) and $\log M_{\rm BH} - z$ (bottom) planes. In both panels, the contours (defined as in Figure 11) are color-coded according to the emission line and the continuum window used for their calculation. The small gray points represent measurements taken from Q. Wu & Y. Shen (2022) for SDSS DR16Q quasars. Our samples generally cover the same region in these parameter spaces as does the largest previous SDSS-based quasar catalog.

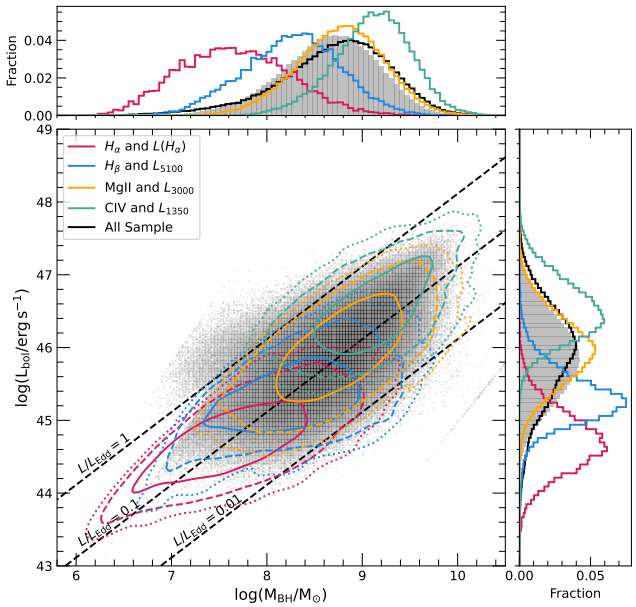

Figure 16. Similar to Figure 15, but focusing on the $\log L_{\rm bol} - \log M_{\rm BH}$ plane. The dashed diagonals indicate Eddington ratios of $L/L_{\rm Edd} = 1$, 0.1, and 0.01. The additional panels on the top and to the right of the main one show the 1D distributions of the two properties, for the various line-dependent subsets (in color), and for their union (in black), compared with the SDSS DR16Q catalog (in gray; Q. Wu & Y. Shen 2022). The Hα-based measurements preferentially trace lower-$L_{\rm bol}$ and/or lower-$M_{\rm BH}$ systems, likely reflecting their lower typical redshifts and the flux limit(s) employed during the survey planning. Our quasars cover $0.01 \lesssim L/L_{\rm Edd} \lesssim 1$, and show distributions of $L_{\rm bol}$, $M_{\rm BH}$, and $L/L_{\rm Edd}$ that are generally consistent with the largest previous SDSS-based quasar catalog.

distributions of $M_{\rm BH}$ and $L/L_{\rm Edd}$, either directly, i.e., through prescriptions like Equations (1), (2), and (3), or less directly, i.e., when a survey misses AGNs that are not luminous enough to be selected (either low-$M_{\rm BH}$ systems limited to $L/L_{\rm Edd} \simeq 1$ or high-$M_{\rm BH}$ systems with low $L/L_{\rm Edd}$). To be able to make a more meaningful comparison between the various samples, a more nuanced approach is needed that would remove these obvious biases.

We therefore employed a more elaborate approach of creating a "matched comparison sample" for each of our SDSS-V-enabled quasar samples. The basic idea is to control for the differences in luminosity and redshift (which would also eliminate differences in fluxes), and then look more carefully into any remaining, intrinsic differences in the distributions of $M_{\rm BH}$ and $L/L_{\rm Edd}$. This scheme was adapted from the one described in the study of G. Zeltyn et al. (2024), which had to address similar challenges.

First, we restricted all the samples, both the SDSS-V and legacy SDSS ones, to the redshift range $0.3 < z < 2$ and to luminosities $\log(L_{\rm bol}/{\rm erg\,s^{-1}}) > 44$, which is where the majority of the quasars are (in all samples). This also means that most of the C IV-based estimates of $M_{\rm BH}$ (and $L/L_{\rm Edd}$)[39] are omitted from what follows. We then divided the $z - \log L_{\rm bol}$ plane into several bins, with the goal of making the *relative* distributions of each of the samples (hereafter test samples) among these bins as close as possible to those of the reference samples (i.e., DR7Q or DR16Q; Y. Shen et al. 2011; Q. Wu & Y. Shen 2022, respectively). In practice, for each bin, we counted the number of quasars in both the test and reference samples and calculated the ratio between the two. Next, only the bins where the ratio between the two samples was greater than 1% were retained (see Appendix D for details). Using the bin with the highest ratio, we created a new reference subsample by randomly selecting objects without replacement in each bin so as to maintain this ratio across the samples, thus preserving a balanced comparison. In other words, following this procedure, the test sample and the reference subsample have a highly similar distribution in the $z - \log L_{\rm bol}$ plane. The only difference remaining is the fact that the reference sample can be significantly larger or smaller than the test sample, as a whole, but not in a way that is skewed toward any specific region of the $z - \log L_{\rm bol}$ plane. While this procedure can be applied to any choice of test and reference samples, in what follows we focus on comparing those samples that were also "originally" more similar to each other, in the broad sense of deeper or shallower flux limits.

Figure 18 presents matched comparisons between the GUA and DR16Q samples, while Figure 19 shows the corresponding distributions for the Skewt_QSO sample. As expected, the pairs of $L_{\rm bol}$ distributions are now essentially identical. Moreover, the distributions of G-band magnitudes are now much more similar to each other (pairwise), compared with the original, unmatched comparisons (Figure 17). This is expected, given that we have matched the samples in the

[39] Together with the corresponding L_{1350}-based estimates of $L_{\rm bol}$.

Table 2
Distribution Statistics for Each Sample

Sample	$\log(L_{\rm bol}/{\rm erg\ s^{-1}})$					$\log(M_{\rm BH}/M_{\odot})$					$\log L/L_{\rm Edd}$				
	Median	Average	IQR	$\sigma_{\rm STD}$	$\sigma_{\rm MAD}$	Median	Average	IQR	$\sigma_{\rm STD}$	$\sigma_{\rm MAD}$	Median	Average	IQR	$\sigma_{\rm STD}$	$\sigma_{\rm MAD}$
GUA	45.95	45.90	(45.48, 46.35)	0.65	0.64	8.76	8.70	(8.36, 9.11)	0.59	0.55	−0.89	−0.91	(−1.17, −0.63)	0.41	0.40
Skewt_QSO	46.17	46.09	(45.72, 46.53)	0.63	0.58	8.87	8.80	(8.48, 9.20)	0.57	0.52	−0.80	−0.82	(−1.06, −0.56)	0.38	0.37
SDSS DR7Q	46.35	46.29	(45.95, 46.67)	0.54	0.52	9.00	8.96	(8.66, 9.30)	0.50	0.47	−0.76	−0.78	(−1.03, −0.51)	0.42	0.38
SDSS DR16Q	45.85	45.84	(45.46, 46.23)	0.55	0.57	8.67	8.64	(8.32, 9.00)	0.51	0.50	−0.91	−0.91	(−1.22, −0.62)	0.47	0.44

Note. The statistics listed here are for the distributions of the the GUA and Skewt_QSO samples, and for the SDSS DR7Q and DR16Q catalogs (Y. Shen et al. 2011; Q. Wu & Y. Shen 2022, respectively), without any matching, as shown in Figure 17. For each parameter, we report the median, mean value, interquartile range (IQR), standard deviation ($\sigma_{\rm STD}$), and the scaled median absolute deviation (MAD; $\sigma_{\rm MAD}$).

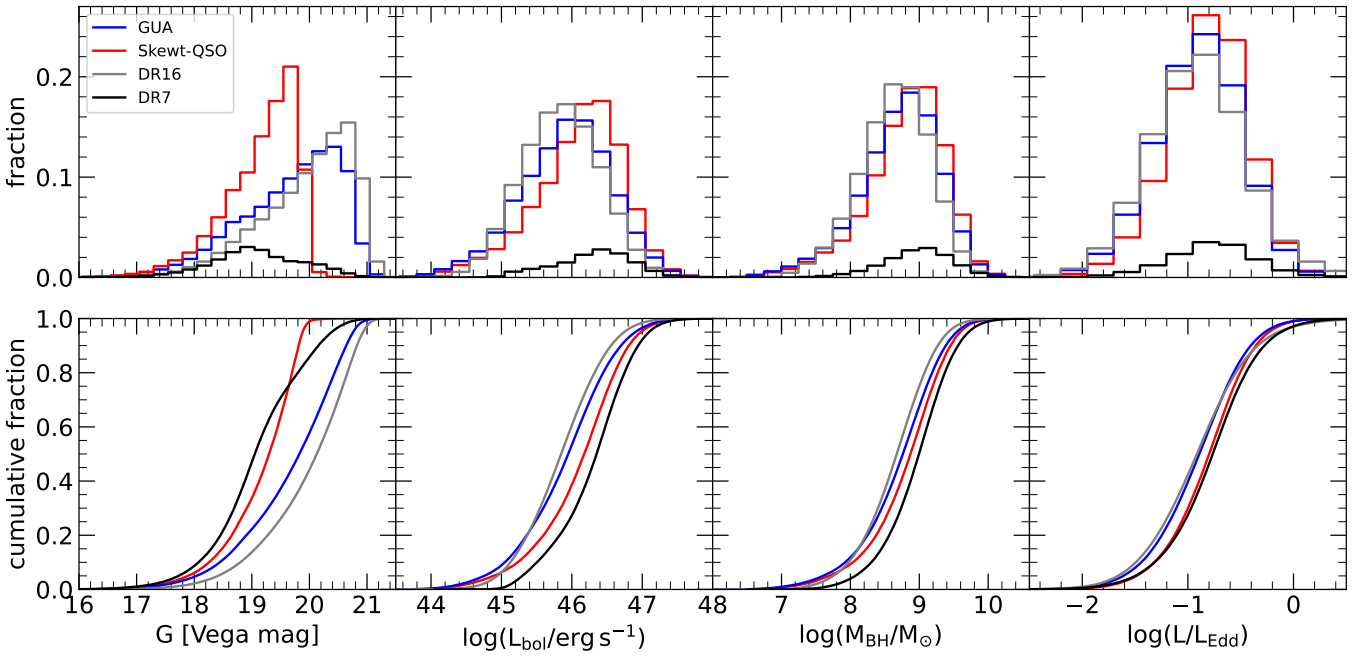

Figure 17. Distributions of key parameters for our two main samples, GUA and Skewt_QSO, and two comparison samples (SDSS DR7Q and DR16Q; see legend). The upper panels show the probability density functions (PDFs) of, from left to right, G-band magnitudes, luminosities ($L_{\rm bol}$), black hole masses ($M_{\rm BH}$), and Eddington ratios ($L/L_{\rm Edd}$). The lower panels display the corresponding cumulative distribution functions (CDFs) for the same properties. These plots highlight the differences in brightness between our samples and the comparison samples, driven by the various survey strategies, but which propagate to differences in the other properties ($L_{\rm bol}$, $M_{\rm BH}$, and $L/L_{\rm Edd}$).

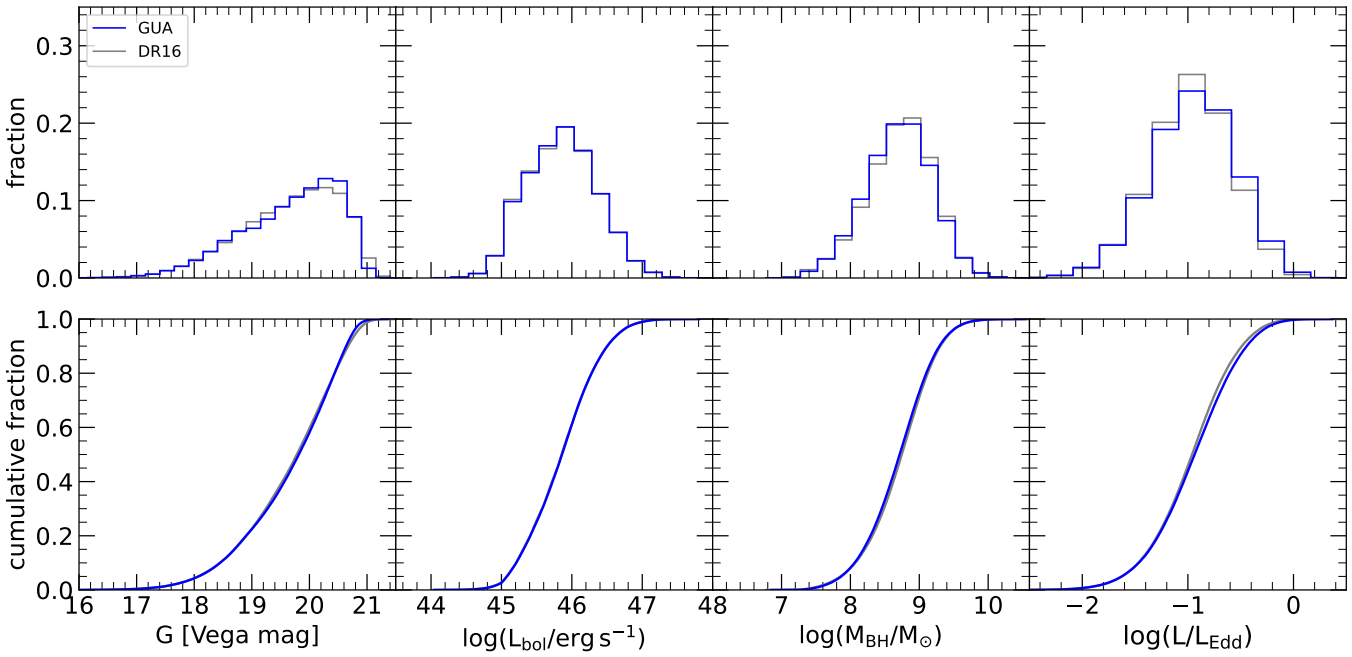

Figure 18. Results of the z–$L_{\rm bol}$ matching procedure, comparing GUA sample (as the test sample) and the SDSS DR16Q sample (as the reference). The panels are identical to those of Figure 17. As expected, the two samples agree nearly perfectly in terms of both G and $L_{\rm bol}$. The distributions of $M_{\rm BH}$ and $L/L_{\rm Edd}$ for the SDSS-V quasars are now in excellent agreement with those of largest previous SDSS-based catalog.

$z - \log L_{\rm bol}$ plane and not simply in $\log L_{\rm bol}$. The same kind of matched comparisons were performed for GUA and Skewt_QSO versus DR7Q (see Appendix D).

Importantly, the corresponding distributions of $M_{\rm BH}$ and $L/L_{\rm Edd}$ are now also much more similar to each other, and it is evident that the new GUA and Skewt_QSO samples probe SMBHs with black hole masses and accretion rates similar to those probed by previous large quasar catalogs (DR7Q and DR16Q), at least in the regions of the $z - \log L_{\rm bol}$ plane where they overlap. To further validate this result, we again applied the AD test to the redshift- and luminosity-matched distributions of $M_{\rm BH}$ and $L/L_{\rm Edd}$. Somewhat surprisingly, the AD test again yielded very low p-values ($P_{\rm AD} \ll 10^{-10}$), formally indicating that the samples are not drawn from identical parent

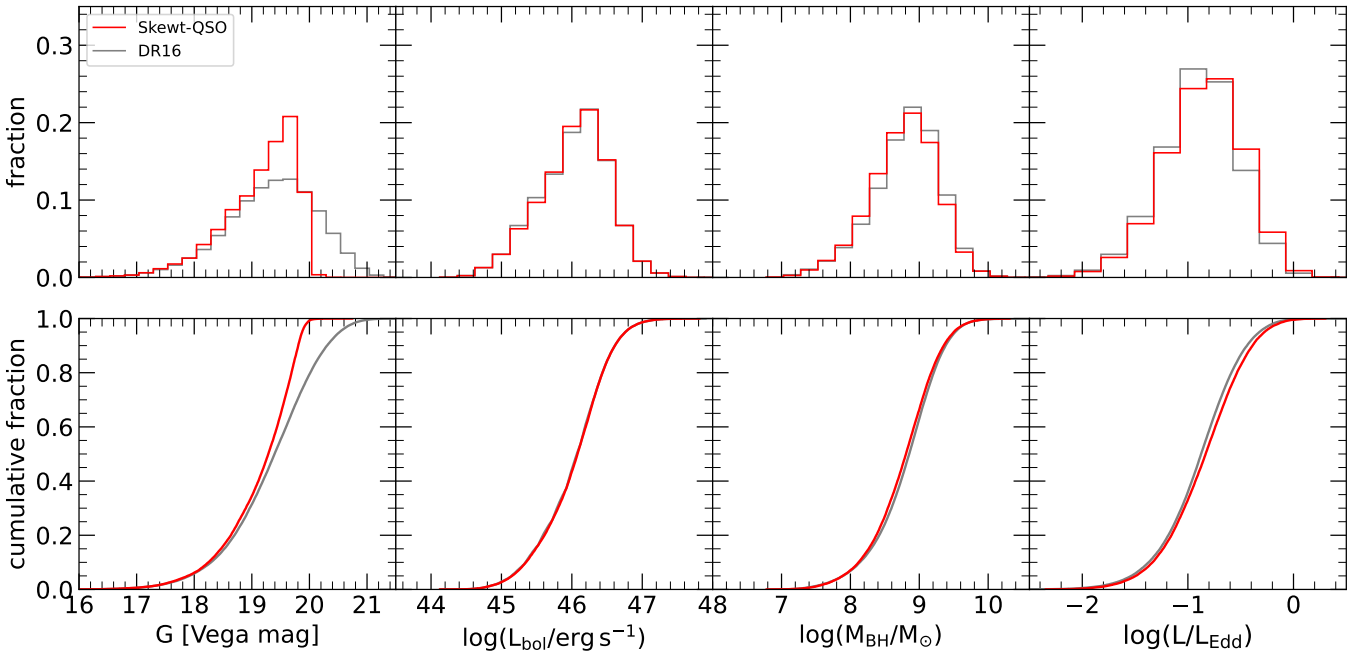

Figure 19. Similar to Figure 18, but presenting the comparison between the matched Skewt_QSO and SDSS DR16Q samples.

distributions. However, given the very large sample sizes involved, even small differences between the distributions are expected to result in highly significant p-values. In this sense, the AD test is highly sensitive to minor discrepancies that may not be physically meaningful. Given the striking similarities between the distributions seen in Figure 18 and the large systematic uncertainties that affect $M_{\rm BH}$ and $L/L_{\rm Edd}$ estimates, it is reasonable to conclude that the quasars in the samples are consistent with the previous SDSS catalogs.

The main conclusion drawn from this analysis is that the GUA sample and the previous catalogs are quite consistent with each other. The similarities in the distributions of key quasar properties, such as luminosity, black hole mass, and Eddington ratio, suggest that all the datasets could be considered to be derived from the same parent population of quasars and SMBHs. This strong consistency across the different surveys affirms the reliability of our sample-selection process and indicates that the quasar population we are studying aligns well with those identified in earlier SDSS catalogs, which focused on the northern celestial hemisphere and employed markedly different quasar-selection techniques. GUA thus clearly demonstrates that we could extend our census of quasars and our ability to study them across the entire sky, and without the need to employ selection methods that rely on nuanced multiband selection and/or multi-wavelength data.

5. Conclusions

In this study, we presented an exploratory analysis of a new large sample of quasars selected for the dual-hemisphere SDSS-V project through two novel selection methods. We evaluated the efficiency of the selection methods, analyzed the key photometric and physical properties of the resulting quasar samples, and compared these findings with other previously established quasar catalogs. The key conclusions from this work are as follows.

1. SDSS-V enables optical spectroscopy of various samples of quasars and AGNs across the entire sky, providing comprehensive coverage for AGN studies and facilitating a more complete understanding of populations of accreting SMBHs (see Section 2.2).
2. The noncore GUA and ASQOSS (Skewt_QSO) programs within SDSS-V, which are the focus of our study, have already provided more than 100,000 quasar spectra, of which ≈76,000 are newly identified sources. Most of these new quasars (>55,000) are located in the southern hemisphere ($\delta < 0°$; see Section 4.1 and Table 1).
3. The quasar-selection methods that produced our sample (s), which were based on previous works by S19 and YS23, have proven to be powerful ways to produce a large, highly efficient, all-sky quasar sample. The quasar-selection efficiency of GUA, which relies solely on Gaia and unWISE data, exceeded 96%, and remains high even at low Galactic latitudes (see Section 4.2 and Figure 9).
4. While the photometric properties of the bulk of the GUA and Skewt_QSO samples are broadly consistent with those of previous quasar catalogs (e.g., SDSS DR7Q), they cover a wider range of optical and/or IR color spaces, which allows us to probe larger populations of red, mildly obscured AGNs (see Section 4.3). A significant fraction (≈26%) of our quasars satisfies the $R - W1 > 4$ obscuration criterion of S. M. LaMassa et al. (2016), intermediate between samples based on optical color selection (i.e., DR7Q quasars) and on X-rays (i.e., S. LaMassa et al. 2024; see Figure 13). So far, the GUA and Skewt_QSO programs have not identified extremely obscured systems, such as hot DOGs (mainly due to flux limits; see Figure 14).
5. The resulting sample includes quasars spanning a broad range of redshifts (reaching $z > 4$), luminosities ($44 \lesssim \log(L_{\rm bol}/{\rm erg\ s^{-1}}) \lesssim 48$), and black hole masses ($6 \lesssim \log(M_{\rm BH}/M_\odot) \lesssim 10$). This sample has similar

properties to other well-established quasar catalogs (i.e., SDSS DR7Q and DR16Q), in terms of $L_{\rm bol}$, $M_{\rm BH}$, and $L/L_{\rm Edd}$ (see Section 4.4).

6. The sample uncovered several intriguing quasar types, including many BAL quasars, radio-emitting sources, and even sources showing extreme spectral variability, offering new opportunities for detailed follow-up studies.

During this work, we encountered several challenges and limitations. First, a significant number of stars contaminated the sample, particularly M-type stars located far from the Galactic plane and with no proper-motion information in Gaia DR2 (see Section 4.2). This issue could potentially be addressed by incorporating more recent Gaia data, e.g., the available Gaia DR3 (Gaia Collaboration et al. 2023b) or the upcoming Gaia DR4, which should provide improved proper-motion measurements. This would lead to yet higher quasar-selection efficiencies. We note that having refined high-efficiency quasar-selection methods at low Galactic latitudes could also be beneficial for studies that aim to focus on (red) stars, enabling them to avoid sources that are most likely quasars.

With SDSS-V observations continuing into 2027, the GUA and ASQOSS programs are expected to yield yet larger numbers of yet-to-be-analyzed, and indeed yet-to-be-observed, targets, eventually perhaps even doubling the sizes of the samples studied here. These growing samples will be made public as part of upcoming SDSS data releases, and we expect that some aspects of our analysis will be occasionally revisited, particularly to better understand certain subsets of (rare) quasars. Specifically, further characterization of the red quasar subset ($R - W1 > 4$; see Figure 13) is essential in order to understand the role that such sources play in the cosmic evolution of SMBHs (e.g., M. Banerji et al. 2015; E. Glikman et al. 2024, and references therein). Moreover, our sample offers the opportunity to crossmatch with previous spectroscopy of some (active) galaxies, in search of extremely variable and/or changing-look AGNs—including sources previously observed by southern-hemisphere surveys, e.g., 6dF (D. H. Jones et al. 2004) or GAMA (S. P. Driver et al. 2022).

Thanks to the ongoing SDSS-V survey, we now have the capability to significantly expand the scope of quasar studies in the southern hemisphere, with both existing facilities (e.g., VLT, ALMA, the Vera C. Rubin Observatory) and those still under construction (e.g., ELT, 4MOST, Roman). The large SDSS-V enabled samples we presented, constructed through robust selection methods, would join the large core eROSITA-based effort within SDSS-V (SPIDERS; A. Merloni et al. 2026, in preparation) to provide a valuable legacy dataset for future southern-hemisphere and time-domain surveys. The markedly high efficiency of the quasar-selection methods we use can also be useful for future efforts to construct particular quasar samples, e.g., sources observed through the Galactic plane, in densely monitored fields, etc. We therefore hope that our exploratory study, and the GUA and Skewt_QSO samples, will support many future studies of AGNs and ultimately help expand our understanding of the SMBHs that power them, the galaxies that host them, and their cosmic evolution.

Acknowledgments

A.S., B.T., and G.Z. acknowledge support from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement No. 950533) and the Israel Science Foundation (grant No. 1849/19). R.J.A. was supported by FONDECYT grant No. 1231718 and by the ANID BASAL project FB210003. F.E.B. acknowledges support from ANID-Chile BASAL CATA FB210003 and FONDECYT Regular 1241005. N.W.B. acknowledges support from NSF grant No. AST-2407089. J.G.F-T. gratefully acknowledges the support provided by ANID Fondecyt Regular No. 1260371, ANID Fondecyt Postdoc No. 3230001 (Sponsoring researcher), the Joint Committee ESO-Government of Chile under the agreement 2023 ORP 062/2023, and the support of the Doctoral Program in Artificial Intelligence, DISC-UCN. This research was supported by the Excellence Cluster ORIGINS and by the Munich Institute for Astro-, Particle and BioPhysics (MIAPbP), which are funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC 2094 – 390783311. B.T. acknowledges the hospitality of the Instituto de Estudios Astrofísicos at Universidad Diego Portales, the Instituto de Astrofísica at Pontificia Universidad Católica de Chile, and the Institut d'Astrophysique de Paris, where parts of this study have been carried out.

Funding for the Sloan Digital Sky Survey V has been provided by the Alfred P. Sloan Foundation, the Heising-Simons Foundation, the National Science Foundation, and the Participating Institutions. SDSS acknowledges support and resources from the Center for High-Performance Computing at the University of Utah. SDSS telescopes are located at Apache Point Observatory, funded by the Astrophysical Research Consortium and operated by New Mexico State University, and at Las Campanas Observatory, operated by the Carnegie Institution for Science. The SDSS website is www.sdss.org.

SDSS is managed by the Astrophysical Research Consortium for the Participating Institutions of the SDSS Collaboration, including the Carnegie Institution for Science, Chilean National Time Allocation Committee (CNTAC) ratified researchers, Caltech, the Gotham Participation Group, Harvard University, Heidelberg University, The Flatiron Institute, The Johns Hopkins University, L'Ecole polytechnique fédérale de Lausanne (EPFL), Leibniz-Institut für Astrophysik Potsdam (AIP), Max-Planck-Institut für Astronomie (MPIA Heidelberg), Max-Planck-Institut für Extraterrestrische Physik (MPE), Nanjing University, National Astronomical Observatories of China (NAOC), New Mexico State University, The Ohio State University, Pennsylvania State University, Smithsonian Astrophysical Observatory, Space Telescope Science Institute (STScI), the Stellar Astrophysics Participation Group, Universidad Nacional Autónoma de México, University of Arizona, University of Colorado Boulder, University of Illinois at Urbana-Champaign, University of Toronto, University of Utah, University of Virginia, Yale University, and Yunnan University.

This research uses services or data provided by the Astro Data Lab, which is part of the Community Science and Data Center (CSDC) Program of NSF NOIRLab. NOIRLab is operated by the Association of Universities for Research in Astronomy (AURA), Inc. under a cooperative agreement with the U.S. National Science Foundation.

Software: astropy (Astropy Collaboration et al. 2022), PyQSOFit (H. Guo et al. 2018).

Data Availability

The spectroscopic quasar sample analyzed in this work is a subset of the SDSS-V DR20 spectroscopic dataset, which can be accessed through the SDSS-V DR20 data release page at https://www.sdss.org/. The BOSS pipeline data products, including the `spAll` files used here, are distributed as part of DR20. The code used for spectral analysis in this work can be found at https://github.com/QiaoyaWu/sdss4_dr16q_tutorial.

Appendix A
Overlapping Quasar/AGN Selection Methods and Cartons

As explained in Section 2.2.3, we must consider the significant overlap between the various quasar/AGN selection methods under study, particularly when assessing the purity (or efficiency) of the GUA and Skewt_QSO quasar-selection methods.

In Figure A1, we show the `Upset` diagrams (A. Lex et al. 2014)[40] that illustrate the overlap between key types of selection criteria relevant for our sample construction. Specifically, we split targets (top panel), observed spectra (middle), and quasars (bottom) by whether they are part of the cartons associated with the S19-based GUA selection, the YS23-based ASQOSS/Skewt_QSO selection, X-ray based (SPIDERS+CSC) selection, and/or the two core BHM programs focusing on previously known quasars (AQMES and BHM-RM). This grouping is meant to highlight different kinds of prior knowledge about the (potential) AGN nature of the sources. The overlap between all these selection criteria, and specifically between GUA, Skewt_QSO, and the other (higher-priority) BHM programs, is clearly evident.

In our analysis, however, we focus on a different grouping of sources, more closely linked to the priority that each target had when considered for observations. Specifically, we consider whether sources are part of GUA, Skewt_QSO, and/or the superset of targets selected by any of the core BHM programs (SPIDERS, AQMES, and/or BHM-RM) combined with targets drawn from the CSC. Sources targeted within this latter superset not only have higher observational priority within SDSS-V survey planning and operations, but carry additional information about the quasar/AGN nature of the target (see Section 2.2.3). These three main sets of targets are illustrated in the Venn diagram shown as Figure A2, which further denotes the seven disjoint subsets representing the various possibilities for overlap (or lack thereof) between the sets (A, B, …, G).

To assess the purity of the GUA and Skewt_QSO selection methods, we calculate for each of them a *hp* measure, calculated within a sky area where all three selection paths were applicable. The quantity *hp* is the weighted mean of the quasar-selection purities of each of the various disjoint subsets relevant for each sample, with the weights representing the fraction among *targets* (not sources) selected through the particular selection path that belong to each subset.

Specifically, we calculated *hp* for GUA and Skewt_QSO by first focusing on all the *targets* the corresponding cartons hold in the region enclosed within $2\,\mathrm{h} < \alpha < 5\,\mathrm{h}\,20\,\mathrm{m}$ and $-60^\circ < \delta < -20^\circ$ (hereafter the "test area"). Then, we calculated the fraction of targets selected by each method corresponding to each of the combinatorial subsets denoted in Figure A2. For example, f_A is the fraction of GUA targets that were not selected by either SPIDERS+CSC or Skewt_QSO, among all GUA targets in the test area. Likewise, f_B is the fraction corresponding to targets selected by both GUA and SPIDERS+CSC, but not by Skewt_QSO. In addition, $f_{C,\mathrm{G}}$ corresponds to sources selected by both GUA and Skewt_QSO, but not by SPIDERS+CSC, while $f_{D,\mathrm{G}}$ corresponds to sources selected by all three approaches (GUA, Skewt_QSO, and SPIDERS+CSC). The subscript "G" in both $f_{C,\mathrm{G}}$ and $f_{D,\mathrm{G}}$ stems from the fact that these fractions are calculated among all GUA targets in the test area. The corresponding four fractions for Skewt_QSO would be f_E, f_F, and then $f_{C,\mathrm{s}}$ and $f_{D,\mathrm{s}}$, where the subscript "s" is for Skewt_QSO. We note that the subset G, corresponding to sources selected only by SPIDERS+CSC, but not by GUA or Skewt_QSO, is not used in our analysis. All the fractions, and the numbers used to calculate them, are listed in Table A1.

Next, we turn our attention to the *observed sources* corresponding to each of the logical selection subsets (A, B, …), and again restricting ourselves to the test area. We calculate the quasar-selection purity within each such subset as the fraction of quasars among all useful spectra (both *good* and *bad*) of sources whose targeting information places them in the subset. Of the spectra obtained within this area, ∼85% belong to the *good* subset and ∼15% to the *bad* subset. For example, the quasar purity of subset A, p_A, is the fraction of quasars among all the spectra targeting sources that were flagged by the GUA method, but not by the Skewt_QSO or SPIDERS+CSC targeting at the test area. The quasar-selection purities within each subset are also listed in Table A1.

Finally, the homogenized purity of GUA is calculated as

$$hp_{\mathrm{GUA}} = (f_A\, p_A) + (f_B\, p_B) + (f_{C,\mathrm{G}}\, p_C) + (f_{D,\mathrm{G}}\, p_D)\,, \tag{A1}$$

and that of Skewt_QSO is calculated as

$$hp_{\mathrm{skewt}} = (f_E\, p_E) + (f_F\, p_F) + (f_{C,\mathrm{s}}\, p_C) + (f_{D,\mathrm{s}}\, p_D). \tag{A2}$$

In Section 4.2, we also discuss trends of purity with source brightness (magnitude) and with Galactic latitude. In both cases, this entails repeating the entire *hp* calculation presented above but restricted to targets (for fractions f) and spectra (for purities p) that belong to a certain bin in either magnitude or Galactic latitude.

[40] https://upset.app/

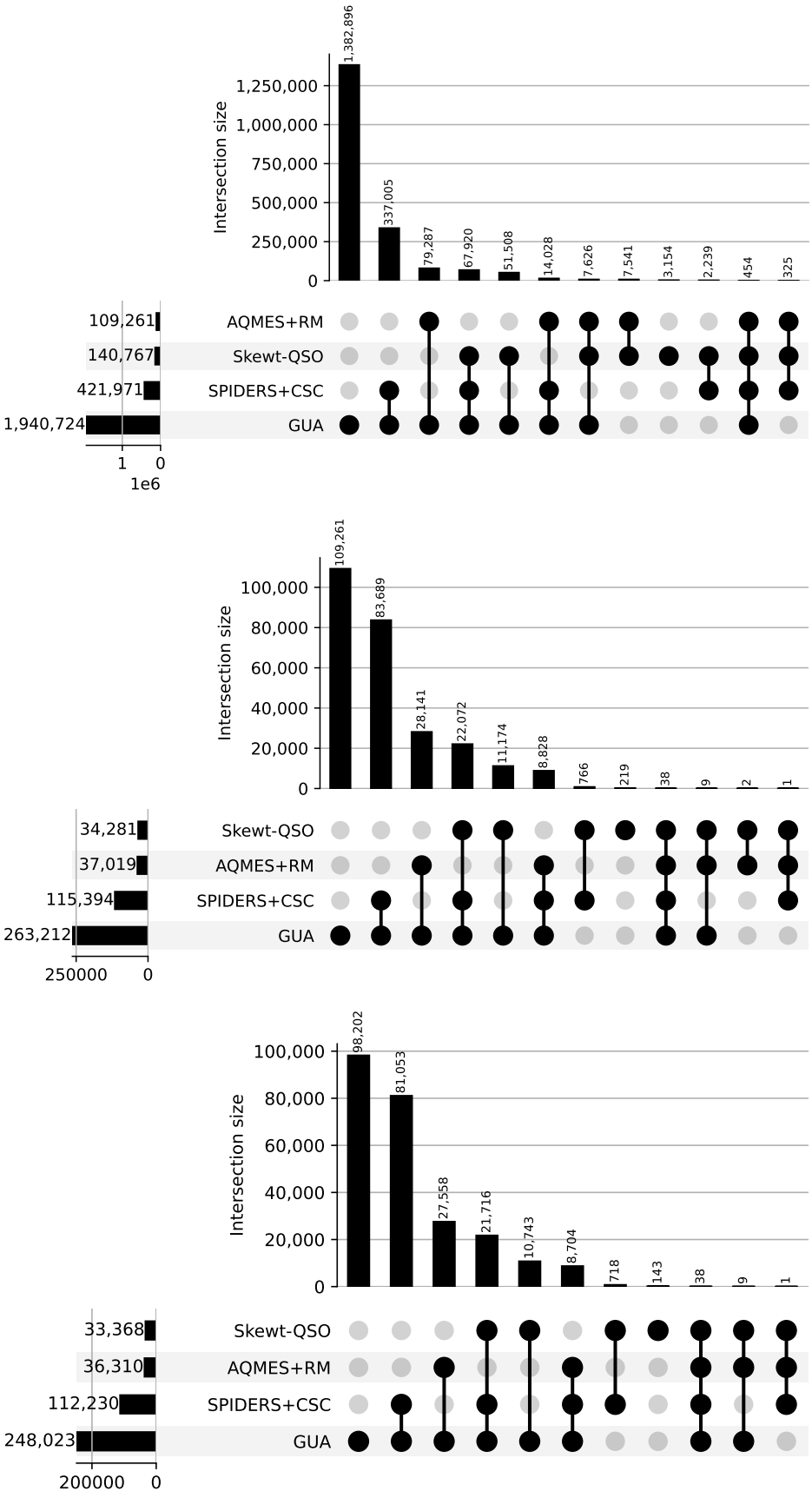

Figure A1. `Upset` diagrams illustrating the overlap between the GUA, Skewt_QSO, and other main higher-priority BHM selection methods, grouped by the kind of prior knowledge they hold about the AGN nature of the sources. The top, middle, and bottom panels show the overlap among targets (cartons), the observed sources (within DR20), and the spectroscopically confirmed quasars.

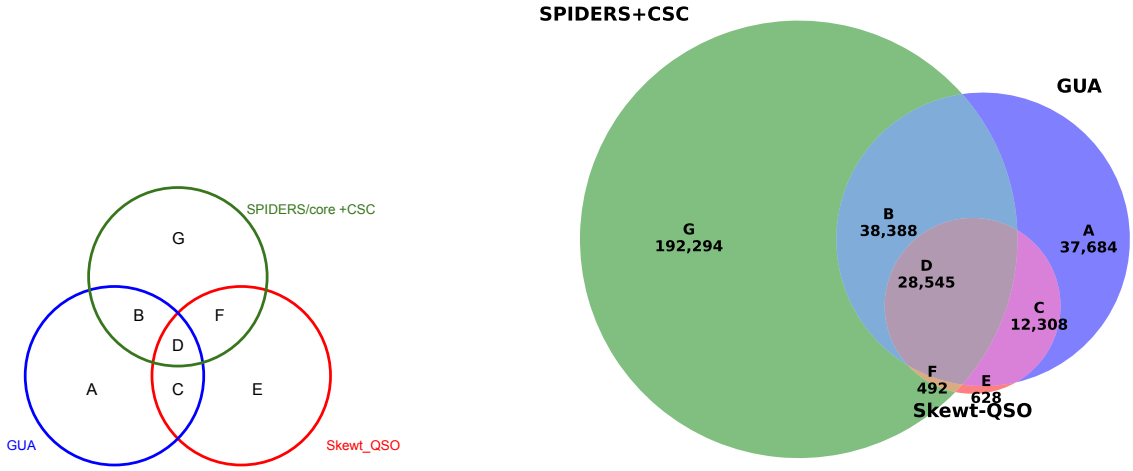

Figure A2. Venn diagrams presenting the selection and sample overlap considered in this work. Left: a conceptual illustration of the seven disjoint subsets considered in our analysis of newly detected quasars (Section 4.1) and of selection efficiency (purity; this appendix and Section 4.2). Right: the same subsets, but now illustrating the actual degree of overlap between the various selection methods within the southern, high-Galactic-latitude "test area."

Table A1
Summary of the Homogenized Purity in the Test Area

Subset	GUA	Skewt_QSO	SPIDERS+CSC	$N_{\rm targ,area}$	$f_{\rm GUA}$	$f_{\rm Skewt_QSO}$	$N_{\rm obs,area}$	$N_{\rm obs-QSO,area}$	Purity
A	V	X	X	37,684	0.322	⋯	7,889	7,323	0.928
B	V	X	V	38,388	0.328	⋯	12,765	12,463	0.976
C	V	V	X	12,308	0.105	0.293	3,035	2,943	0.970
D	V	V	V	28,545	0.244	0.680	9,975	9,826	0.985
E	X	V	X	628	⋯	0.015	22	13	0.591
F	X	V	V	492	⋯	0.012	144	130	0.903
G	X	X	V	192,294	⋯	⋯	⋯	⋯	⋯

Note. Following the first subset column, the next three columns indicate whether a given subset is selected by GUA, Skewt_QSO, and/or SPIDERS+CSC. $N_{\rm targ,area}$ is the number of targets in the subset within the test area, while $f_{\rm GUA}$ and $f_{\rm Skewt_QSO}$ are the corresponding fractions among all GUA ($N_{\rm GUA} = 116{,}925$) or Skewt_QSO ($N_{\rm Skewt_QSO} = 41{,}973$) targets, respectively. $N_{\rm obs,area}$ is the number of observed sources with spectra in the subset, and $N_{\rm obs-QSO,area}$ is the number of quasars among them. The resulting homogenized purities are $hp_{\rm GUA} \simeq 96\%$ and $hp_{\rm Skewt_QSO} \simeq 97\%$ (see Figure 9).

Appendix B
ZWARNING Description

Here, we present the ZWARNING flags, which indicate issues in the pipeline redshift-fitting process. These flags are assigned by the SDSS pipeline to identify spectra where redshift determination may be unreliable due to various factors. The ZWARNING parameter is encoded as a bitmask integer, where different bits correspond to different warning conditions.[41] Understanding these flags is essential for distinguishing between spectra with catastrophic errors that cannot be used for analysis and those that remain useful for quasar identification but have uncertain redshifts. Table B1 provides a detailed list of ZWARNING flags and descriptions of the specific issues they signify.

[41] https://www.sdss4.org/dr17/algorithms/bitmasks/#ZWARNING

Table B1
Description of ZWARNING Flags and Their Associated Sample Classification

ZWARNING	Description	Sample
SKY	Sky fiber	Excluded
LITTLE_COVERAGE	Too little wavelength coverage	Bad
SMALL_DELTA_CHI2	χ^2 of the best fit is too close to that of the second best	Bad
NEGATIVE_MODEL	Synthetic spectrum is negative	Bad
MANY_OUTLIERS	Fraction of points more than 5σ away from the best model is too large	Bad
Z_FITLIMIT	χ^2 minimum at the edge of the redshift-fitting range (Z_ERR set to –1)	Bad
NEGATIVE_EMISSION	A QSO line exhibits negative emission Triggered only in QSO spectra if C IV, C III], Mg II, Hβ, or if Hα has LINEAREA + 3 × LINEAREA_ERR < 0	Bad
UNPLUGGED	The fiber was unplugged or damaged; location of the spectrum is unknown	Excluded
BAD_TARGET	Catastrophically bad targeting data	Excluded
NODATA	No data for this fiber (e.g., spectrograph was broken during the exposure)	Excluded

Note. Not all of the quality flags listed here were actually identified in the internal SDSS-V spectral data, and not all are expected to be found in the public SDSS-V data products.

Appendix C
Line Configuration in PyQSOFit Analysis

In this appendix, we present the line configuration used in the PyQSOFit analysis, which follows a similar approach to that described in Q. Wu & Y. Shen (2022). This configuration, listed in Table C1, defines how different emission-line complexes are modeled within the spectral fitting process.

The selected line complexes include Hα, Hβ, Mg II, C III], C IV, Si IV, and Lyα. For each line complex, the algorithm fits both broad and narrow emission lines, ensuring that the components are properly constrained. Broad emission lines are modeled using a combination of one to three Gaussian components to capture their complex structure, while narrow emission lines are fit using a single Gaussian.

Table C1
`PyQSOFit` Configuration for Emission-line Decomposition

Complex Name	Lines	No. Gaussians
Hα	Broad Hα	3
Hα	Narrow Hα	1
Hα	[N II] λ6548	1
Hα	[N II] λ6583	1
Hα	[S II] λ6716	1
Hα	[S II] λ6731	1
Hβ	Broad Hβ	3
Hβ	Narrow Hβ	1
Hβ	[O III] λ4959	1
Hβ	[O III] λ5007	1
Hβ	Broad He II λ4686	1
Hβ	Narrow He II λ4686	1
Mg II λ2798	Broad Mg II λ2798	2
Mg II λ2798	Narrow Mg II λ2798	1
C III] λ1909	Broad C III] λ1909	2
C III] λ1909	Si III] λ1892	1
C III] λ1909	Al III λ1857	1
C III] λ1909	Si II λ1816	1
C III] λ1909	N III] λ1750	1
C III] λ1909	N IV] λ1716	1
C IV λ1549	Broad C IV λ1549	3
C IV λ1549	Narrow He II λ1640	1
C IV λ1549	Broad He II λ1640	1
C IV λ1549	Narrow O III] λ1663	1
C IV λ1549	Broad O III] λ1663	1
Si IV	Si IV λ1403	1
Si IV	Si IV λ1394	1
Si IV	C II λ1335	1
Si IV	O I λ1304	1
Lyα	Broad Lyα	3
Lyα	N V λ1240	1

Note. Line configuration used in PyQSOFit, including complex names, individual emission lines, and the number of Gaussian components fitted for each line.

Appendix D
Matching Method

This appendix presents an example of the matching process between a test and reference sample, specifically demonstrating the comparison between GUA and DR16Q, as explained in Section 4.4. Figure D1 shows the grid in the bolometric luminosity−redshift plane used for this matching. Using the highest ratio (0.97 in this example) and all the green bins, a reference subsample was created to ensure a more statistically meaningful comparison between the two distributions. Figures D2 and D3 show the resulting matched comparisons for GUA and Skewt_QSO versus DR7Q, respectively.

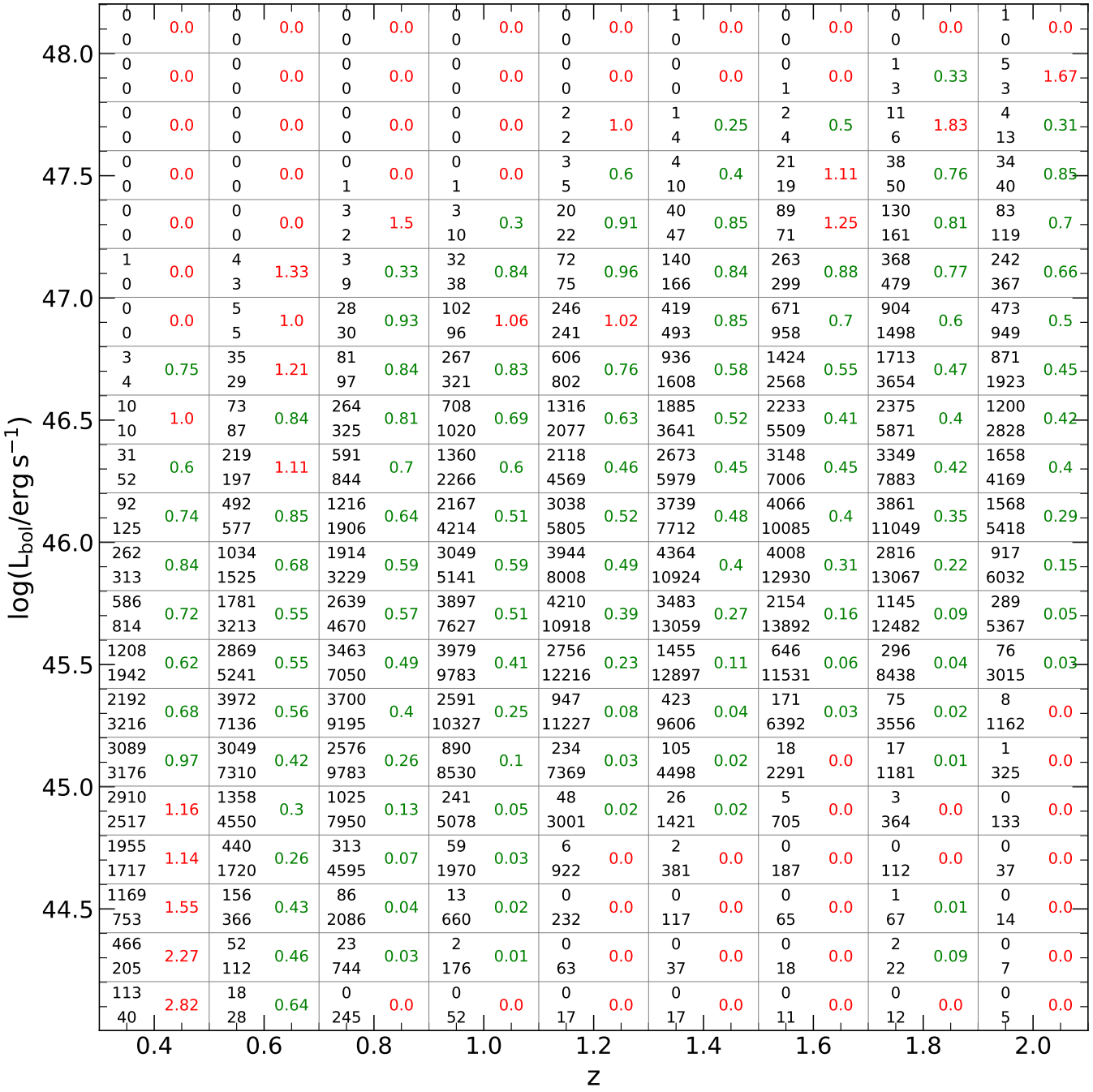

Figure D1. An example of the matching method, between GUA and DR16Q. Each bin in the $\log L_{\rm bol} - z$ space represents the number of quasars in the test sample (top) and the reference sample (bottom). The ratio between the two samples is displayed within each bin, where bins with ratios above 1% are shaded green, while those below are red.

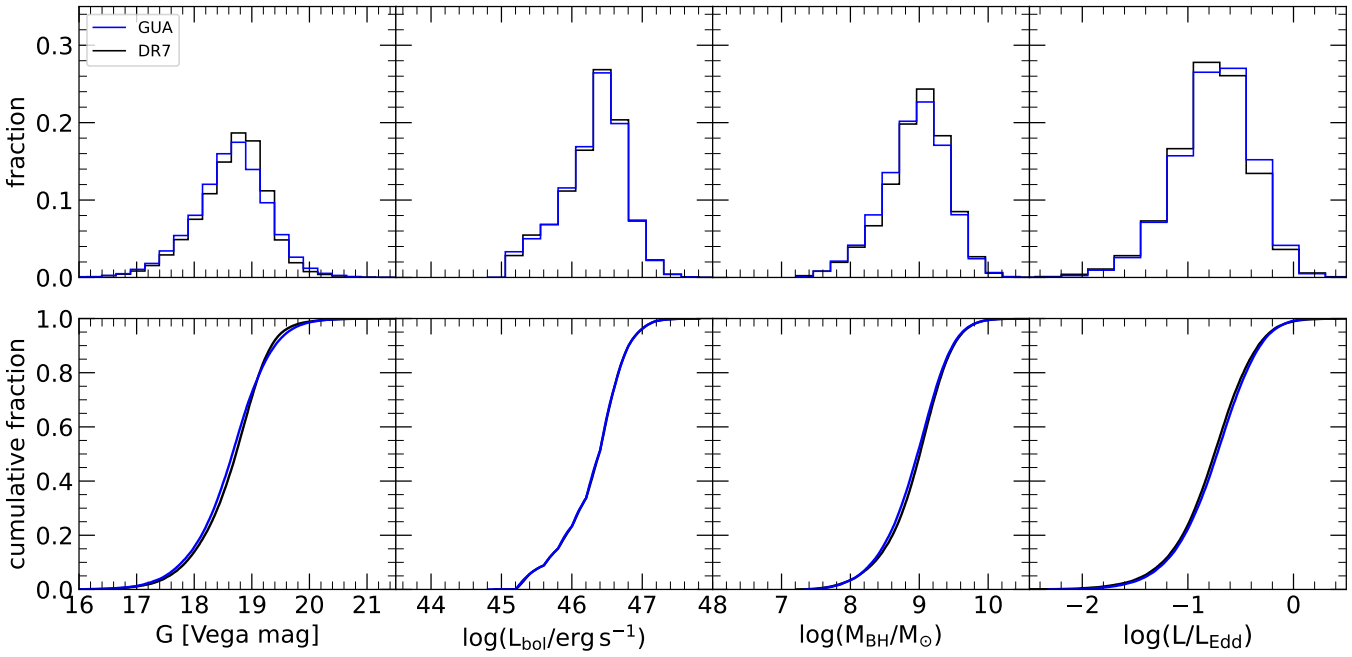

Figure D2. Similar to Figure 18, but presenting the comparison between the matched GUA/"All" and SDSS DR7Q samples.

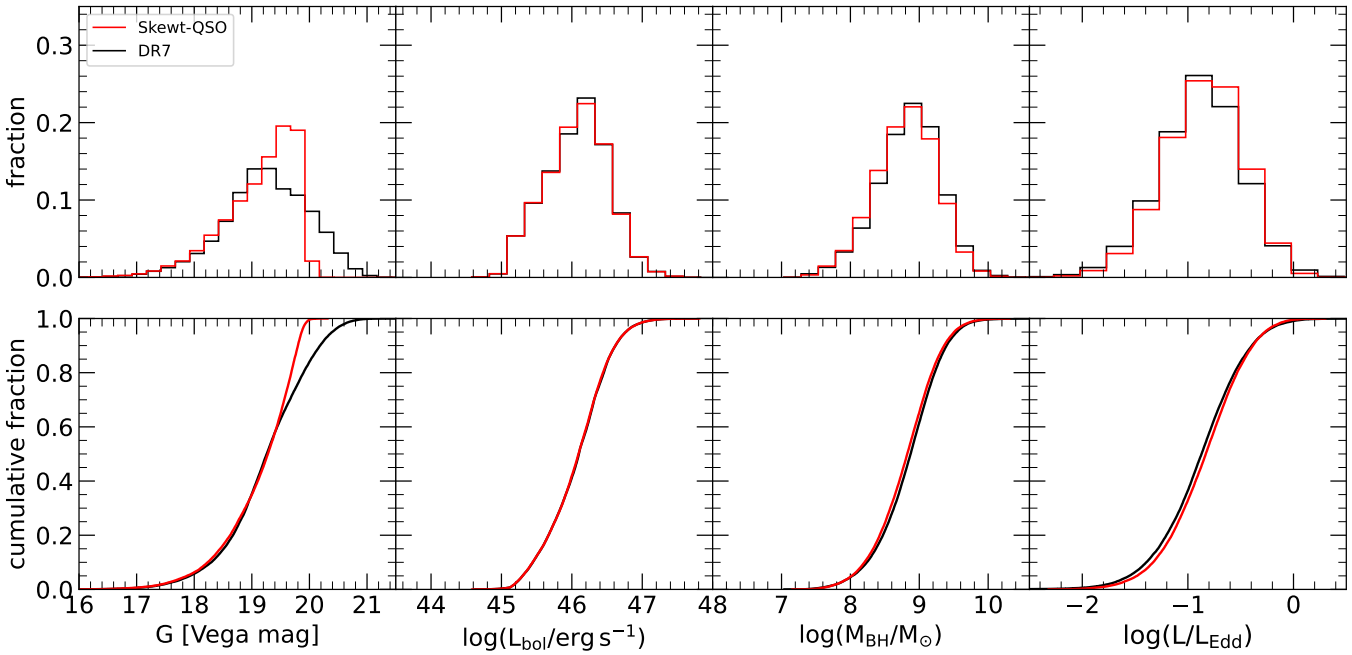

Figure D3. Similar to Figure 18, but presenting the comparison between the matched Skewt_QSO/"All" and SDSS DR7Q samples.

ORCID iDs

Shir Aviram https://orcid.org/0009-0008-0046-8064
Benny Trakhtenbrot https://orcid.org/0000-0002-3683-7297
Tom Dwelly https://orcid.org/0000-0002-4459-9233
Scott F. Anderson https://orcid.org/0000-0002-6404-9562
Sean Morrison https://orcid.org/0000-0002-6770-2627
Michael Eracleous https://orcid.org/0000-0002-3719-940X
Yue Shen https://orcid.org/0000-0003-1659-7035
Mara Salvato https://orcid.org/0000-0001-7116-9303
Donald P. Schneider https://orcid.org/0000-0001-7240-7449
Roberto J. Assef https://orcid.org/0000-0002-9508-3667
Catarina Aydar https://orcid.org/0000-0001-5609-2774
Franz E. Bauer https://orcid.org/0000-0002-8686-8737
W. N. Brandt https://orcid.org/0000-0002-0167-2453
Joel R. Brownstein https://orcid.org/0000-0002-8725-1069
Johannes Buchner https://orcid.org/0000-0003-0426-6634
Jeremy Darling https://orcid.org/0000-0003-2511-2060
José G. Fernández-Trincado https://orcid.org/0000-0002-0900-9760

Patrick B. Hall https://orcid.org/0000-0002-1763-5825
Dong-Woo Kim https://orcid.org/0000-0002-7386-944X
Anton M. Koekemoer https://orcid.org/0000-0002-6610-2048
Stephanie LaMassa https://orcid.org/0000-0002-5907-3330
Andrea Merloni https://orcid.org/0000-0002-0761-0130
Claudio Ricci https://orcid.org/0000-0001-5231-2645
Qian Yang https://orcid.org/0000-0002-6893-3742
Grisha Zeltyn https://orcid.org/0000-0002-7817-0099

References

Abazajian, K. N., Adelman-McCarthy, J. K., Agüeros, M. A., et al. 2009, ApJS, 182, 543
Abbott, T. M. C., Abdalla, F. B., Allam, S., et al. 2018, ApJS, 239, 18
Abbott, T. M. C., Adamów, M., Aguena, M., et al. 2021, ApJS, 255, 20
Ahumada, R., Allende Prieto, C., Almeida, A., et al. 2020, ApJS, 249, 3
Alexander, D. M., & Hickox, R. C. 2012, NewAR, 56, 93
Almeida, A., Anderson, S. F., Argudo-Fernández, M., et al. 2023, ApJS, 267, 44
Ananna, T. T., Weigel, A. K., Trakhtenbrot, B., et al. 2022, ApJS, 261, 9
Assef, R. J., Eisenhardt, P. R. M., Stern, D., et al. 2015, ApJ, 804, 27
Assef, R. J., Stern, D., Kochanek, C. S., et al. 2013, ApJ, 772, 26
Assef, R. J., Stern, D., Noirot, G., et al. 2018, ApJS, 234, 23
Astropy Collaboration, Price-Whelan, A. M., Lim, P. L., et al. 2022, ApJ, 935, 167
Aydar, C., Merloni, A., Dwelly, T., et al. 2025, A&A, 698, A132
Banerji, M., Alaghband-Zadeh, S., Hewett, P. C., & McMahon, R. G. 2015, MNRAS, 447, 3368
Banerji, M., McMahon, R. G., Hewett, P. C., et al. 2012, MNRAS, 427, 2275
Bentz, M. C., Batiste, M., Seals, J., et al. 2016, ApJ, 831, 2
Bertemes, C., Trakhtenbrot, B., Schawinski, K., Done, C., & Elvis, M. 2016, MNRAS, 463, 4041
Blanton, M. R., Carlberg, J. K., Dwelly, T., et al. 2026, AJ, 171, 37
Bolton, A. S., Schlegel, D. J., Aubourg, É., et al. 2012, AJ, 144, 144
Bovy, J., Hennawi, J. F., Hogg, D. W., et al. 2011, ApJ, 729, 141
Bowen, I. S., & Vaughan, A. H., Jr. 1973, ApOpt, 12, 1430
Cardelli, J. A., Clayton, G. C., & Mathis, J. S. 1989, ApJ, 345, 245
Cho, H., Woo, J.-H., Wang, S., et al. 2023, ApJ, 953, 142
Colless, M., Dalton, G., Maddox, S., et al. 2001, MNRAS, 328, 1039
Comparat, J., Eckert, D., Finoguenov, A., et al. 2020, OJAp, 3, 13
Croom, S. M., Richards, G. T., Shanks, T., et al. 2009, MNRAS, 392, 19
Croom, S. M., Smith, R. J., Boyle, B. J., et al. 2004, MNRAS, 349, 1397
Cutri, R. M., Wright, E. L., Conrow, T., et al. 2013, Explanatory Supplement to the AllWISE Data Release Products,
Dalla Bontà, E., Peterson, B. M., Grier, C. J., et al. 2025, A&A, 696, A48
de Jong, R. S., Agertz, O., Berbel, A. A., et al. 2019, Msngr, 175, 3
Denney, K. D. 2012, ApJ, 759, 44
DESI Collaboration, Abdul Karim, M., Adame, A. G., et al. 2026, AJ, 171, 285
Dey, A., Schlegel, D. J., Lang, D., et al. 2019, AJ, 157, 168
Driver, S. P., Bellstedt, S., Robotham, A. S. G., et al. 2022, MNRAS, 513, 439
Duras, F., Bongiorno, A., Ricci, F., et al. 2020, A&A, 636, A73
Dwelly, T., Salvato, M., Merloni, A., et al. 2017, MNRAS, 469, 1065
Dye, S., Lawrence, A., Read, M. A., et al. 2018, MNRAS, 473, 5113
Edge, A., Sutherland, W., Kuijken, K., et al. 2013, Msngr, 154, 32
Eisenhardt, P. R. M., Wu, J., Tsai, C.-W., et al. 2012, ApJ, 755, 173
Evans, I. N., Evans, J. D., Martínez-Galarza, J. R., et al. 2024, ApJS, 274, 22
Fan, X., Bañados, E., & Simcoe, R. A. 2023, ARA&A, 61, 373
Fawcett, V. A., Alexander, D. M., Brodzeller, A., et al. 2023, MNRAS, 525, 5575
Fawcett, V. A., Alexander, D. M., Rosario, D. J., et al. 2020, MNRAS, 494, 4802
Fitzpatrick, M. J., Olsen, K., Economou, F., et al. 2014, SPIE, 9149, 91491T
Flesch, E. W. 2019, arXiv:1912.05614
Flesch, E. W. 2023, OJAp, 6, 49
Fu, Y., Wu, X.-B., Yang, Q., et al. 2021, ApJS, 254, 6
Gaia Collaboration, Brown, A. G. A., Vallenari, A., et al. 2016, A&A, 595, A2
Gaia Collaboration, Babusiaux, C., van Leeuwen, F., et al. 2018, A&A, 616, A10
Gaia Collaboration, Bailer-Jones, C. A. L., Teyssier, D., et al. 2023a, A&A, 674, A41
Gaia Collaboration, Vallenari, A., Brown, A. G. A., et al. 2023b, A&A, 674, A1
Gibson, R. R., Jiang, L., Brandt, W. N., et al. 2009, ApJ, 692, 758
Glikman, E., LaMassa, S., Piconcelli, E., Zappacosta, L., & Lacy, M. 2024, MNRAS, 528, 711
Glikman, E., Simmons, B., Mailly, M., et al. 2015, ApJ, 806, 218
Glikman, E., Urrutia, T., Lacy, M., et al. 2012, ApJ, 757, 51
Greene, J. E., & Ho, L. C. 2005, ApJ, 630, 122
Gunn, J. E., Siegmund, W. A., Mannery, E. J., et al. 2006, AJ, 131, 2332
Guo, H., Shen, Y., & Wang, S. 2018, PyQSOFit: Python code to fit the spectrum of quasars, Astrophysics Source Code Library, ascl: 1809.008
Gupta, K. K., Ricci, C., Temple, M. J., et al. 2024, A&A, 691, A203
Hamann, F., Zakamska, N. L., Ross, N., et al. 2017, MNRAS, 464, 3431
Huo, Z.-Y., Fu, Y., Huang, Y., et al. 2025, ApJS, 278, 6
Jones, D. H., Saunders, W., Colless, M., et al. 2004, MNRAS, 355, 747
Juneau, S., Canning, R., Alexander, D. M., et al. 2025, AJ, 169, 157
Kollmeier, J. A., Rix, H.-W., Aerts, C., et al. 2026, AJ, 171, 52
Krawczyk, C. M., Richards, G. T., Mehta, S. S., et al. 2013, ApJS, 206, 4
LaMassa, S., Peca, A., Urry, C. M., et al. 2024, ApJ, 974, 235
LaMassa, S. M., Civano, F., Brusa, M., et al. 2016, ApJ, 818, 88
Lang, D. 2014, AJ, 147, 108
Lawrence, A., Warren, S. J., Almaini, O., et al. 2007, MNRAS, 379, 1599
Lex, A., Gehlenborg, N., Strobelt, H., Vuillemot, R., & Pfister, H. 2014, ITVCG, 20, 1983
Li, G., Tsai, C.-W., Stern, D., et al. 2023, ApJ, 958, 162
Li, G., Wu, J., Tsai, C.-W., et al. 2025, ApJ, 981, 104
Lindegren, L., Hernández, J., Bombrun, A., et al. 2018, A&A, 616, A2
Lyke, B. W., Higley, A. N., McLane, J. N., et al. 2020, ApJS, 250, 8
Mateos, S., Alonso-Herrero, A., Carrera, F. J., et al. 2012, MNRAS, 426, 3271
McMahon, R. G., Banerji, M., Gonzalez, E., et al. 2013, Msngr, 154, 35
Meisner, A. M., Lang, D., Schlafly, E. F., & Schlegel, D. J. 2019, PASP, 131, 124504
Mejía-Restrepo, J. E., Trakhtenbrot, B., Lira, P., Netzer, H., & Capellupo, D. M. 2016, MNRAS, 460, 187
Mejía-Restrepo, J. E., Trakhtenbrot, B., Koss, M. J., et al. 2022, ApJS, 261, 5
Myers, A. D., Palanque-Delabrouille, N., Prakash, A., et al. 2015, ApJS, 221, 27
Natarajan, P., Pacucci, F., Ricarte, A., et al. 2024, ApJL, 960, L1
Netzer, H. 2019, MNRAS, 488, 5185
Nikutta, R., Fitzpatrick, M., Scott, A., & Weaver, B. A. 2020, A&C, 33, 100411
Oke, J. B., & Gunn, J. E. 1983, ApJ, 266, 713
Padovani, P., Alexander, D. M., Assef, R. J., et al. 2017, A&ARv, 25, 2
Pâris, I., Petitjean, P., Aubourg, É., et al. 2014, A&A, 563, A54
Pâris, I., Petitjean, P., Aubourg, É., et al. 2018, A&A, 613, A51
Pâris, I., Petitjean, P., Ross, N. P., et al. 2017, A&A, 597, A79
Peterson, B. M. 2014, SSRv, 183, 253
Pogge, R. W., Derwent, M. A., O'Brien, T. P., et al. 2020, SPIE, 11447, 1144781
Pognan, Q., Trakhtenbrot, B., Sbarrato, T., Schawinski, K., & Bertemes, C. 2020, MNRAS, 492, 4058
Predehl, P., Andritschke, R., Arefiev, V., et al. 2021, A&A, 647, A1
Ricci, C., Bauer, F. E., Treister, E., et al. 2017, MNRAS, 468, 1273
Ricci, C., & Trakhtenbrot, B. 2023, NatAs, 7, 1282
Richards, G. T., Fan, X., Newberg, H. J., et al. 2002, AJ, 123, 2945
Richards, G. T., Hall, P. B., Vanden Berk, D. E., et al. 2003, AJ, 126, 1131
Richards, G. T., Lacy, M., Storrie-Lombardi, L. J., et al. 2006, ApJS, 166, 470
Ross, N. P., Hamann, F., Zakamska, N. L., et al. 2015, MNRAS, 453, 3932
Ross, N. P., Myers, A. D., Sheldon, E. S., et al. 2012, ApJS, 199, 3
Runnoe, J. C., Brotherton, M. S., & Shang, Z. 2012, MNRAS, 422, 478
Salvato, M., Wolf, J., Dwelly, T., et al. 2022, A&A, 661, A3
Schlafly, E. F., & Finkbeiner, D. P. 2011, ApJ, 737, 103
Schlafly, E. F., Meisner, A. M., & Green, G. M. 2019, ApJS, 240, 30
Schlegel, D. J., Finkbeiner, D. P., & Davis, M. 1998, ApJ, 500, 525
Schneider, D. P., Richards, G. T., Hall, P. B., et al. 2010, AJ, 139, 2360
Schulze, A., Bongiorno, A., Gavignaud, I., et al. 2015, MNRAS, 447, 2085
Schulze, A., & Wisotzki, L. 2010, A&A, 516, A87
SDSS Collaboration, Aghakhanloo, M., Aird, J., et al. 2026a, ApJS, 285, 9
SDSS Collaboration, Aghakhanloo, M., Aguilar, D., et al. 2026b, arXiv:2607.26149
Shakura, N. I., & Sunyaev, R. A. 1973, A&A, 24, 337
Shen, Y. 2013, BASI, 41, 61
Shen, Y., Hall, P. B., Horne, K., et al. 2019, ApJS, 241, 34
Shen, Y., & Liu, X. 2012, ApJ, 753, 125
Shen, Y., Richards, G. T., Strauss, M. A., et al. 2011, ApJS, 194, 45
Shu, Y., Koposov, S. E., Evans, N. W., et al. 2019, MNRAS, 489, 4741
Skrutskie, M. F., Cutri, R. M., Stiening, R., et al. 2006, AJ, 131, 1163
Smee, S. A., Gunn, J. E., Uomoto, A., et al. 2013, AJ, 146, 32

Stern, D., Assef, R. J., Benford, D. J., et al. 2012, ApJ, 753, 30
Storey-Fisher, K., Hogg, D. W., Rix, H.-W., et al. 2024, ApJ, 964, 69
Trakhtenbrot, B., & Netzer, H. 2012, MNRAS, 427, 3081
Urrutia, T., Becker, R. H., White, R. L., et al. 2009, ApJ, 698, 1095
Vasudevan, R. V., Mushotzky, R. F., Winter, L. M., & Fabian, A. C. 2009, MNRAS, 399, 1553
Vestergaard, M., & Peterson, B. M. 2006, ApJ, 641, 689
Vito, F., Brandt, W. N., Stern, D., et al. 2018, MNRAS, 474, 4528
Wang, F., Yang, J., Fan, X., et al. 2021, ApJL, 907, L1
Werk, J., Tchernyshyov, K., Bish, H., et al. 2024, ApJS, 273, 21
Weymann, R. J., Morris, S. L., Foltz, C. B., & Hewett, P. C. 1991, ApJ, 373, 23
Wu, Q., & Shen, Y. 2022, ApJS, 263, 42
Yang, Q., & Shen, Y. 2023, ApJS, 264, 9
Yao, S., Wu, X.-B., Ai, Y. L., et al. 2019, ApJS, 240, 6
Yip, C. W., Connolly, A. J., Szalay, A. S., et al. 2004a, AJ, 128, 585
Yip, C. W., Connolly, A. J., Vanden Berk, D. E., et al. 2004b, AJ, 128, 2603
York, D. G., Adelman, J., Anderson, J. E., Jr., et al. 2000, AJ, 120, 1579
Zeltyn, G., Trakhtenbrot, B., Eracleous, M., et al. 2024, ApJ, 966, 85