



# The Spectroscopic Follow-up of the QUBRICS Bright Quasar Survey

Konstantina Boutsia<sup>1</sup>, Andrea Grazian<sup>2</sup>, Giorgio Calderone<sup>3</sup>, Stefano Cristiani<sup>3,4,5</sup>, Guido Cupani<sup>3,5</sup>,  
Francesco Guarneri<sup>3,6</sup>, Fabio Fontanot<sup>3,5</sup>, Ricardo Amorin<sup>7,8</sup>, Valentina D’Odorico<sup>3,5,9</sup>, Emanuele Giallongo<sup>10</sup>,  
Mara Salvato<sup>11</sup>, Alessandro Omizzolo<sup>2,12</sup>, Michael Romano<sup>2,13</sup>, and Nicola Menci<sup>10</sup>

<sup>1</sup> Las Campanas Observatory, Carnegie Observatories, Colina El Pino, Casilla 601, La Serena, Chile; [kboutsia@carnegiescience.edu](mailto:kboutsia@carnegiescience.edu)

<sup>2</sup> INAF–Osservatorio Astronomico di Padova, Vicolo dell’Osservatorio 5, I-35122, Padova, Italy

<sup>3</sup> INAF–Osservatorio Astronomico di Trieste, Via G.B. Tiepolo, 11, I-34143 Trieste, Italy

<sup>4</sup> INFN–National Institute for Nuclear Physics, via Valerio 2, I-34127 Trieste, Italy

<sup>5</sup> IFPU–Institute for Fundamental Physics of the Universe, via Beirut 2, I-34151 Trieste, Italy

<sup>6</sup> Dipartimento di Fisica, Sezione di Astronomia, Università di Trieste, via G.B. Tiepolo 11, I-34131, Trieste, Italy

<sup>7</sup> Instituto de Investigación Multidisciplinar en Ciencia y Tecnología, Universidad de La Serena, Raul Bitrán 1305, La Serena, Chile

<sup>8</sup> Departamento de Astronomía, Universidad de La Serena, Av. Juan Cisternas 1200 Norte, La Serena, Chile

<sup>9</sup> Scuola Normale Superiore, P.zza dei Cavalieri, I-56126 Pisa, Italy

<sup>10</sup> INAF–Osservatorio Astronomico di Roma, Via Frascati 33, I-00078, Monte Porzio Catone, Italy

<sup>11</sup> Max-Planck-Institut für extraterrestrische Physik Giessenbachstrasse 1, Garching D-85748, Germany

<sup>12</sup> Specola Vaticana, Vatican Observatory, 00122, Vatican City State

<sup>13</sup> Dipartimento di Fisica e Astronomia, Università di Padova, Vicolo dell’Osservatorio 3, I-35122, Padova, Italy

Received 2020 June 12; revised 2020 August 3; accepted 2020 August 6; published 2020 September 29

## Abstract

We present the results of the spectroscopic follow-up of the QUasars as BRiGht beacons for Cosmology in the Southern Hemisphere (QUBRICS; Calderone et al. 2019) survey. The selection method is based on a machine-learning approach applied to photometric catalogs, covering an area of  $\sim 12,400$  deg<sup>2</sup> in the Southern Hemisphere. The spectroscopic observations started in 2018 and identified 55 new, high-redshift ( $z \geq 2.5$ ), bright ( $i \leq 18$ ) quasi-stellar objects (QSOs), with the catalog published in late 2019. Here we report the current status of the survey, bringing the total number of bright QSOs at  $z \geq 2.5$  identified by QUBRICS to 224. The success rate of the QUBRICS selection method, in its most recent training, is estimated to be 68%. The predominant contaminant turns out to be lower- $z$  QSOs at  $z < 2.5$ . This survey provides a unique sample of bright QSOs at high  $z$  available for a number of cosmological investigations. In particular, carrying out the redshift drift measurements (Sandage Test) in the Southern Hemisphere, using the High Resolution Spectrograph at the 39 m Extremely Large Telescope appears to be possible with less than 2500 hr of observations spread over 30 targets in 25 yr.

*Unified Astronomy Thesaurus concepts:* [Cosmology \(343\)](#); [Quasars \(1319\)](#); [Catalogs \(205\)](#); [Redshift surveys \(1378\)](#); [Surveys \(1671\)](#); [Active galaxies \(17\)](#)

*Supporting material:* machine-readable table

## 1. Introduction

Luminous quasars are the brightest non-transient cosmic beacons in the universe. The hunt for such bright sources, especially at high redshift, is of paramount importance for a number of extragalactic studies, ranging from the number density of bright quasars at high  $z$  (Schindler et al. 2019a), the theoretical modeling of the early phases of galaxy formation and co-evolution with their central super massive black holes (SMBHs) (e.g., Valiante et al. 2016; Fontanot et al. 2020), the study and characterization of their (gas) accretion properties of the SMBH population (Wu et al. 2015; Wolf et al. 2018) to the inference on cosmological parameters from time delays of strongly lensed quasi-stellar objects (QSOs; Bonvin et al. 2017) and the properties of the dark matter by microlensing statistics in bright quasars (Webster et al. 1991; Bate et al. 2007).

Absorption signatures in the spectra of bright high- $z$  QSOs are one of the most powerful and invaluable tools for studying intergalactic environments, as emerged from the recent Astro2020 Decadal Survey (e.g., Becker et al. 2019). Among the fundamental questions that can be tackled thanks to the study of QSO absorption lines we recall: the measurement of primordial Deuterium abundance (e.g., Cooke & Fumagalli 2018); the temperature evolution of the cosmic microwave background; the free-streaming of warm dark matter (e.g., Iršič et al. 2017); the variation of the

fundamental constants of nature, e.g., the fine structure constant or the proton-to-electron mass ratio (see Leite et al. 2016); the missing baryon problem (e.g., Werk et al. 2014); the production and diffusion of metals in the intergalactic medium (IGM, e.g., D’Odorico et al. 2016); the Lyman continuum escape fraction of high- $z$  QSOs (e.g., Cristiani et al. 2016; Grazian et al. 2018); the mean-free path of ionizing photons (e.g., Prochaska et al. 2009; Worseck et al. 2014; Romano et al. 2019); the reionization epochs of hydrogen and helium; and the sources responsible of these transition phases in cosmic history.

An appealing application of the detailed study of the Lyman forest in cosmology is the so-called Sandage test (Sandage 1962), which can give fundamental constraints for general relativity. The detection of the tiny drift due to cosmic expansion in the cosmological redshifts of many absorption lines in the spectra of bright QSOs will allow us to measure directly cosmological parameters (e.g.,  $\Omega_M$ ,  $\Omega_\Lambda$ , and  $H_0$ ) at  $2 < z < 5$  without the need of any local ladders or intermediate distance indicators. This revolution will be possible only with the brightest QSOs observed by the most powerful and stable high-resolution spectrographs that will be available in the future at 20–40 m telescopes (Liske et al. 2008).

However, finding the brightest quasars at high  $z$  is not a trivial process. The advent of the Sloan Digital Sky Survey (SDSS; e.g.,

Fan et al. 2001) represents a quantum leap in this respect, at least in the Northern Hemisphere. At present, the SDSS has delivered more than  $10^6$  spectroscopically confirmed QSOs at  $0 < z < 6.5$ , with a large fraction at absolute magnitudes of  $M_{1450} \leq -26$ . Recent studies, however, point out that, at very bright magnitudes, SDSS can suffer from incompleteness due to color selection (see also Fontanot et al. 2007). For example, Schindler et al. (2019a) find 407 new bright QSOs at  $2.8 < z < 5.0$  in the Extremely Luminous Quasar Survey, showing that the SDSS completeness is  $\sim 60\%$  at bright magnitudes ( $i \leq 18$ ). As a consequence, the hunt for bright quasars, especially at high  $z$  and in the Northern Hemisphere, could be biased toward lower numbers due to the adoption of efficient selections that suffer from low completeness.

The situation is even more dramatic in the Southern Hemisphere, due to the lack of wide multiwavelength surveys at  $\delta \leq 0^\circ$  in the past. Comparing QSO surface densities, it is statistically evident that high- $z$  objects of bright apparent magnitudes must be present also in the Southern Hemisphere: of the 22 known QSOs with  $z > 3$  and  $V < 17$ , only 5 are at  $\delta < 0^\circ$ , and all the 3 QSOs with  $V < 16$  are in the north (Véron-Cetty & Véron 2010).

In Calderone et al. (2019, hereafter Paper I), we presented the first results of a survey of  $z \geq 2.5$  QSOs at bright  $i$ -band magnitudes ( $i \leq 18.0$ ) in the Southern Hemisphere, taking advantage of the recent availability of new multiwavelength public databases. The combination of state-of-the-art databases with innovative techniques for the selection of the best candidates results in an efficient selection, with a success rate of finding high- $z$  QSOs larger than 50% and a completeness in excess of 90%. In the first spectroscopic runs we already identified the intrinsically most luminous QSO at that time, with  $z \geq 3.8$  at  $\delta < 0^\circ$  as QSO J2157-3602, which has been afterward independently confirmed by Schindler et al. (2019b). In this paper we will present a new  $z > 4$  QSO with an apparent magnitude of  $i = 16.886$  (see Section 4) that would be the new record holder in the Southern Hemisphere. In total we present 168 new bright QSOs at  $z \geq 2.5$ , thus quickly completing the identification of the high- $z$  sample of Paper I.

The structure of this paper is the following: in Section 2 we summarize the selection method of our bright QSO candidate sample, and in Sections 3 and 4 we describe the massive observational campaign carried out at medium size telescopes and the spectroscopic identification of our QSO candidates. The properties of the newly identified high- $z$  QSOs are discussed in Section 5, while Section 6 is devoted to the description of the golden QSO sample for the Sandage test. Discussions and conclusions are summarized in Section 7. Unless otherwise stated, apparent magnitudes are in the AB photometric system.

## 2. Selection Method

In this section we will describe the QUasars as BRIght beacons for Cosmology in the Southern Hemisphere (QUBRICS) survey, first introduced in Paper I. We refer the reader to this work for more details of the method, while here we just recall its main characteristics.

To identify new, high- $z$  QSOs candidates in the remaining sample, a classification algorithm has been applied based on the Canonical Correlation Analysis (CCA; Anderson 1984). Our candidate list has been drawn from a multiwavelength catalog. We used the public databases of: (i) Skymapper (data release 1.1. (DR1.1); Wolf et al. 2018); (ii) Gaia (DR2; Gaia Collaboration et al. 2018); (iii) Two Micron All Sky Survey (2MASS;

Skrutskie et al. 2006), and (iv) the Wide-field Infrared Survey Explorer (WISE) survey (Wright et al. 2010) to build the initial *main sample* (1,014,875 sources), covering  $\sim 12,400 \text{ deg}^2$ , with  $i$ -band magnitudes in the range  $14 < m_i \leq 18$ . The sources with secure object type identification have been used as a training set and the recipe has been applied to the remaining sources in order to predict a likely classification. By using the parallax and proper motion estimates provided by Gaia,  $\sim 83\%$  of the sources were classified as bona fide stars. Matching the remaining sources against the following catalogs—SDSS DR14Q (Pâris et al. 2018), Veron-Cetty QSO thirteenth edition (Véron-Cetty & Véron 2010), and the Two-degree Field Galaxy Redshift Survey (Colless et al. 2001)—led us to reliably identify an object type classification for 4666 QSOs and 3665 galaxies (in the following *original training set*). To discriminate against low- $z$  ( $< 2.5$ ) QSOs, we use the CCA as a regression algorithm for the objects classified as QSOs in the previous step to obtain an estimate of their redshift. In Paper I we tested this methodology with a resulting sample of 1476 candidates without spectroscopic confirmation, and our spectroscopic follow-up identified 54 QSOs with  $z \geq 2.5$ .

As the main focus of our work is the identification of the largest possible number of bright high- $z$  QSO in the Southern Hemisphere, we aim at the highest possible success rate of observing runs. Therefore, after each run we update the training set of the algorithm by including the new identifications; this results in an evolving list of QSO candidates. For the statistical purposes of this paper, the list of candidates has been frozen to its current form after the last observing run in 2020 February. At this point the list included 1412 sources in total, with 594 secure spectroscopic identifications and 818 candidates yet to be observed. Thanks to the revised training set it has been possible to reduce the number of candidates yet to be observed by  $\sim 25\%$  with respect to the original list of Paper I. Figures 1 and 2 illustrate the process of classification and redshift estimate, respectively, in the present work. They update the corresponding plots of Paper I.

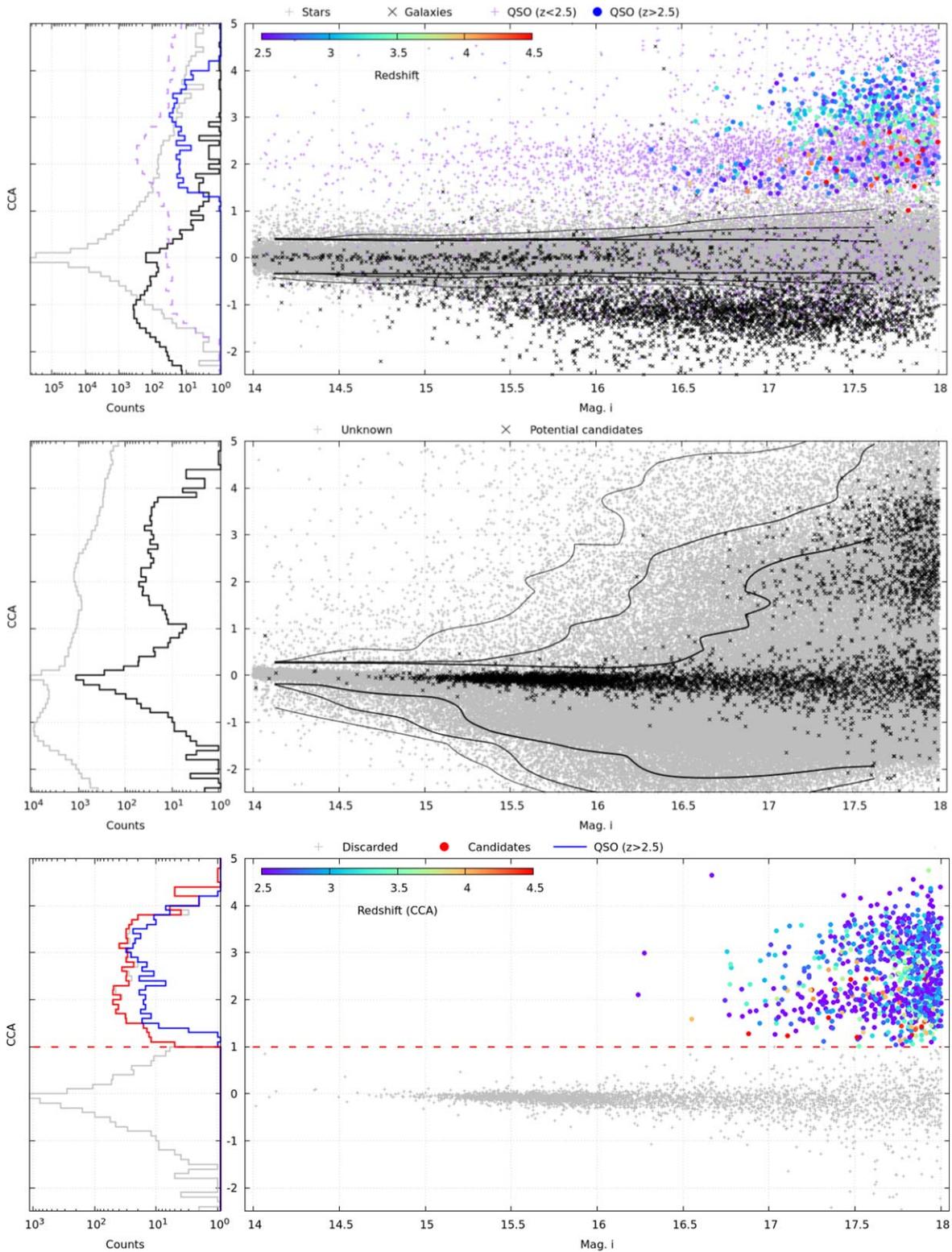
It should be noted that the selection criteria adopted here and in Paper I could be biased against lensed sources. To reduce the contamination we have adopted relative stringent criteria about the positional coincidence in the various photometric catalogs. While this does not affect “normal” QSOs, it may have the subtle effect of removing lensed sources, forming extended structures like Einstein rings or crosses. Moreover, due to the choice of the fundamental photometric bands (e.g., Gaia), our selection is probably biased against sources at  $z \gtrsim 4.5$ .

## 3. Spectroscopy

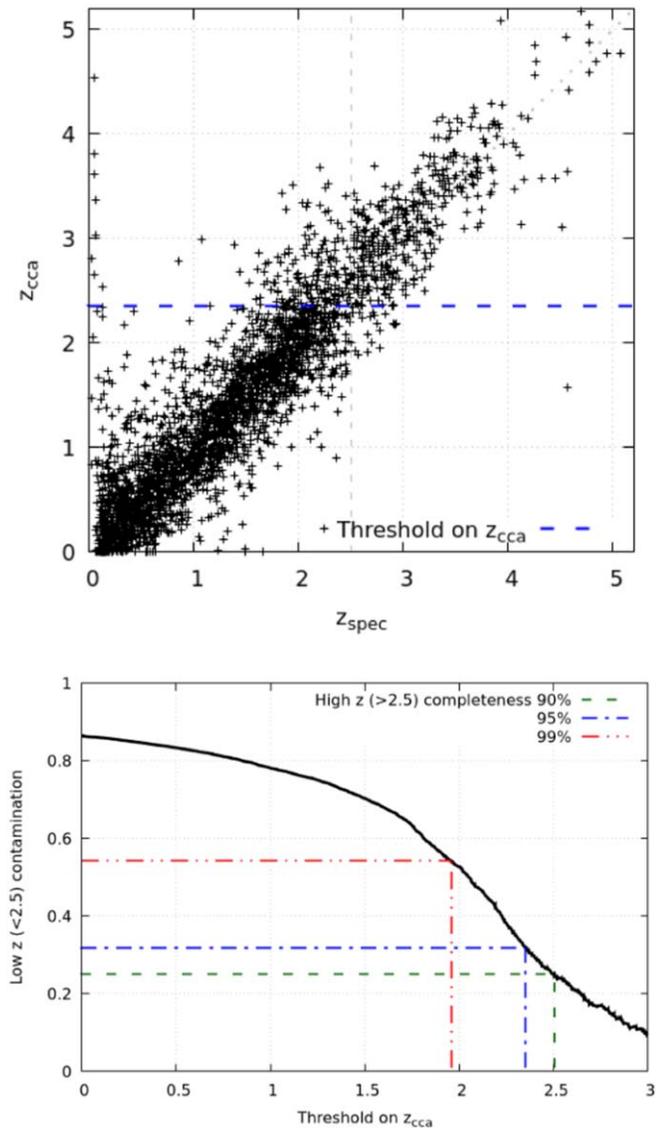
The QUBRICS pilot campaign has been presented in Paper I. Observations have been obtained using various instruments at Las Campanas Observatory (LCO), and the New Technology Telescope of the European Southern Observatory (ESO-NTT) telescope at La Silla. Between 2019 August and 2020 February, we have been awarded more nights at these facilities, in order to expand our spectroscopic survey. We discuss here the details of these observations.

### 3.1. WFCCD at duPont

We have been awarded a total of 12 nights with the Wide Field CCD (WFCCD) camera at duPont. We used the same configuration as in the pilot study—namely, the blue grism,



**Figure 1.** The CCA– $i$  plane for the various subsamples considered in this work. For the statistical meaning of the CCA parameter shown in the y-axis we refer the reader to Paper I. Upper panel: sources in the *main* sample for which a reliable type identification is available. Stars are identified by gray “+” symbols, inactive galaxies by black cross symbols, low- $z$  ( $< 2.5$ ) QSOs with purple “+” symbols, and high- $z$  ( $\geq 2.5$ ) QSOs with filled circles. The redshifts for the spectroscopically confirmed QSOs with  $z \geq 2.5$  are shown with the color code shown in the color box in the upper left corner. The inset on the left shows the histogram of the CCA coordinate for the stars (gray), galaxies (black), low- $z$  QSOs (purple), and known high- $z$  QSOs (blue). Middle panel: sources in the *main* sample without an object type identification (gray symbols). The same sources after excluding extended and low- $z$  objects are highlighted in black and represent potential high- $z$  QSO candidates. Lower panel: the *final* sample of high- $z$  QSO candidates, with their redshift  $z_{\text{cca}}$  as in Paper I. The red dashed line represents the cut in the CCA selection. Candidates/observed sources are represented with open and filled circles, respectively. The gray histogram represents the potential candidates, while the red and blue ones represent the QSO candidates and the confirmed QSO samples.



**Figure 2.** Upper panel: the  $z_{\text{cca}}-z_{\text{spec}}$  correlation (scatter of  $\sim 0.37$ ). The horizontal blue line is the threshold on  $z_{\text{cca}} = 2.26$ , corresponding to a completeness of 95% and a contamination of 38%. Lower panel: contamination and completeness as a function of the threshold on  $z_{\text{cca}}$ .

with the  $1''.5$  long slit, that covers a wavelength range between 3700 and 8000 Å giving a  $2 \text{ \AA}/\text{pixel}$  dispersion. In total we have observed 100 candidates, but only 76 of them have a robust spectroscopic classification (flag = A). Of the securely classified sources, 63 (83%) are QSOs, and 36 of those (57%) are in the desired redshift range ( $z \geq 2.5$ ).

### 3.2. LDSS-3 at Clay

A total of 44 candidates have been observed with the Low Dispersion Survey Spectrograph (LDSS-3) at the Clay telescope. Observations were obtained in several nights during bright time and variable weather conditions. We have used the Volume Phase Holographic—all grism with the  $1''$  central slit and no blocking filter, covering a wavelength range between 4000 and 10,000 Å with a low resolution of  $R \sim 800$ . Exposure times ranged between 800 and 1800 s, depending on the candidate magnitude. Out of the observed candidates, 40 have a high quality redshift flag (flag A)

and have been securely classified, most of them as QSOs (28 out of 40, 70%), but only 12 (43%) have a redshift above 2.5.

### 3.3. EFOSC2 at NTT

In 2019 September we were awarded three more nights (PI. A. Grazian, proposal 0103.A-0746) at NTT to use the ESO Faint Object Spectrograph and Camera (EFOSC2) instrument. In order to complete the survey, we obtained four additional nights at NTT in 2020 January (PI. A. Grazian, proposal 0104.A-0754). Again, we used grism #13 (wavelength range of  $\lambda \sim 3700-9300 \text{ \AA}$ ), with typical exposure times ranging between 300 and 600 s. We observed 217 candidates, obtaining robust identification for 187. Out of these, 161 (86%) were classified as QSOs and 122 have a redshift of  $z \geq 2.5$  (76%). This was one of the most efficient runs in this period.

The outcome of this new spectroscopic campaign shows that, although our criteria are very efficient at identifying QSOs (84% of robust classifications are QSOs), only 55% are in the targeted redshift range. As discussed in Section 2, by ingesting the new QSOs to the selection algorithm, the efficiency of the selection for future spectroscopy is expected to be as high as 75% in the global sample.

### 3.4. Data Reduction

WFCCD data were reduced using standard IRAF tasks. After subtracting bias and dividing with flat, individual exposures were combined to the final image. We then used the task *apall* to extract the spectra and standard and *sensfunc* to calibrate in flux. The EFOSC2 and LDSS-3 spectra have been reduced with a custom pipeline using Munich Image Data Analysis System scripts. The standard pre-reduction (i.e., bias subtraction and flat field normalization) has been adopted, and the wavelength calibration with a helium, neon, and thorium lamp has been obtained. We also check our wavelength solution on the emission night sky. An rms of  $0.5 \text{ \AA}$  has been obtained from the wavelength calibration process. For WFCCD, EFOSC2, and LDSS-3, we made sure to obtain at least one spectrophotometric standard star per night. Since conditions have not always been photometric, the derived flux calibration is relative. There has been no further attempt to interpolate flux to known broadband magnitudes for absolute flux calibration. Further details can be found in Paper I. Improved data reduction of three objects observed in the runs of Paper I allowed us to upgrade their redshift flag to A. They are listed in Table 1 with an appropriate identification as a footnote.

### 3.5. Redshifts from Other Surveys

We have also searched the databases in the literature for additional reliable spectra of our candidates, finding 24 QSOs (6 with  $z \geq 2.5$ ) and one active galactic nucleus (AGN). They are listed in Table 2. In particular we found that by downloading and analyzing spectra from the Six-degree Field (6dF) survey (Jones et al. 2009) we could assign reliable redshifts to 22 of our candidates. The candidate with SkyMapper ID 5578462 has spectral data of good quality in the ESO archive that we have reprocessed (see Figure 3). For the candidate with ID 10779504 we found a reliable redshift determination in the Australian Dark Energy Survey (OzDES) Quasar Catalog (Tie et al. 2017).

**Table 1**  
All Observed Sources with Secure Redshift Identification

Skymapper ID	R.A. (J2000)	Decl. (J2000)	$m_i$ (mag)	$z_{\text{spec}}$	Class	Candidate Paper I	Candidate New	Date-obs.	Inst.
107306	21:01:34.72	-30:10:29.75	17.616	2.761	QSO	Y	Y	2019-09	NTT
343536	21:38:41.84	-33:49:37.21	17.660	3.253	QSO	Y	Y	2019-09	NTT
397340 <sup>a</sup>	21:57:28.21	-36:02:15.11	17.367	4.771	QSO	Y	Y	2019-08	WFCCD
583628 <sup>b</sup>	21:32:25.90	-28:31:33.24	17.605	2.821	QSO	Y	Y	2019-09	NTT
755996	21:30:54.92	-22:46:54.51	17.693	1.749	QSO	Y	N	2019-09	NTT

**Notes.** A portion is shown here for guidance regarding its form and content. “\*” = after updated reduction of Paper I observed sources.

<sup>a</sup> Independently discovered also by Wolf et al. (2020).

<sup>b</sup> Included also in Schindler et al. (2019b).

<sup>c</sup> Included also in Schindler et al. (2019a).

<sup>d</sup> Independently discovered by Lucey et al. (2018).

(This table is available in its entirety in machine-readable form.)

**Table 2**  
Additional Identifications from the Literature

Skymapper ID	R.A. (J2000)	Decl. (J2000)	$m_i$ (mag)	$z_{\text{spec}}$	Class	Survey
8414380	00:43:23.42	-00:15:52.5	17.781	1.442	QSO	6dF
8133493	00:44:48.56	-13:39:13.0	17.664	2.253	QSO	6dF
7885395	00:58:20.54	-19:04:03.2	17.983	1.861	QSO	6dF
9164524	02:03:03.20	-07:06:04.7	17.839	1.458	QSO	6dF
8826213	02:08:41.52	-19:44:03.7	17.943	2.137	QSO	6dF
8970643	02:29:45.41	-18:17:11.5	17.246	1.817	QSO	6dF
315606464	03:06:13.56	-57:51:05.3	17.925	2.066	QSO	6dF
318330107	03:14:59.84	-45:45:28.6	17.967	2.198	QSO	6dF
10372687	03:26:13.41	-31:00:52.2	17.869	2.105	QSO	6dF
10779504	03:39:53.34	-27:00:53.4	17.988	2.410	QSO	OzDES
10460382	03:40:42.83	-34:00:44.3	17.800	1.944	QSO	6dF
10812487	03:47:14.85	-24:38:08.7	17.427	3.130	QSO	6dF
11616376	04:22:43.67	-29:25:29.9	17.767	2.529	QSO	6dF
12053071	05:05:55.83	-29:30:38.5	17.865	1.070	QSO	6dF
56866269	10:18:21.75	-21:40:07.8	17.547	2.441	QSO	6dF
57772960	11:18:10.70	-17:51:59.3	15.119	0.215	AGN	6dF
63944442	12:12:18.99	-25:47:26.1	17.620	2.532	QSO	6dF
65530384	12:28:48.22	-01:04:14.7	17.039	2.642	QSO	6dF
64459508	12:31:32.96	-14:36:30.9	16.888	2.418	QSO	6dF
65808873	13:25:09.61	-08:04:48.2	17.786	2.359	QSO	6dF
308760834	22:21:10.25	-44:31:57.3	17.501	2.071	QSO	6dF
1459087	23:06:37.40	-36:49:26.0	17.962	2.650	QSO	6dF
308911502	23:21:22.33	-50:28:17.5	17.701	2.297	QSO	6dF
5578462	23:21:28.67	-10:51:22.3	17.769	2.958	QSO	UVES

#### 4. Results of the Spectroscopy

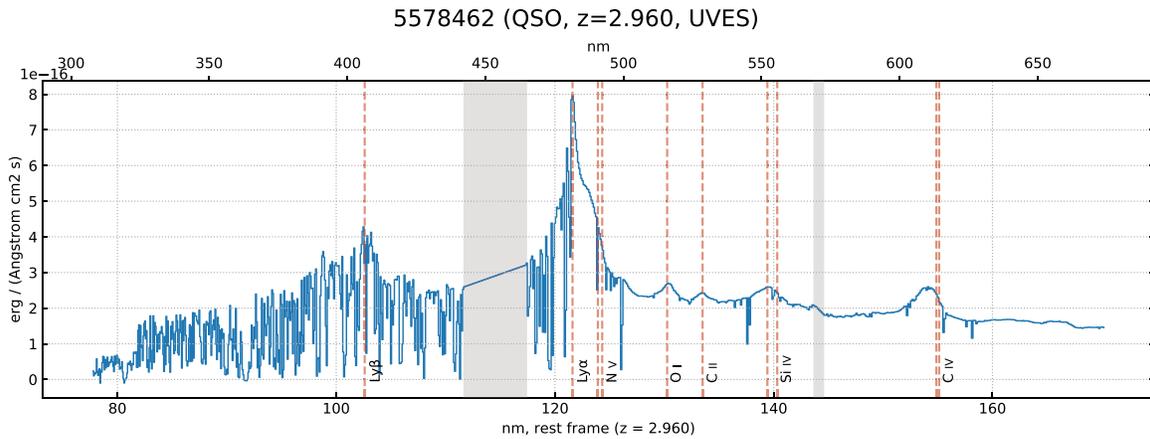
After 26 observing runs (including those reported in Paper I) we have collected the spectra of 511 sources, and 432 of them have a secure object type classification. In total we identified 224 new bright QSOs at  $z \geq 2.5$  (of which 15 are at  $z > 4$ ), 166 QSOs/AGNs at  $z < 2.5$ , 11 inactive galaxies, and 31 stars. Among the observed sources, a small fraction has an uncertain classification and/or redshift estimate. These sources have been assigned flag B for a variety of reasons. More precisely 32% have flat spectra without clear emission features that could be either stars or galaxies. Another 68% have a tentative QSOs classification, but it is not possible to robustly estimate the redshift since there is only one emission line or they show broad absorption features that are difficult to interpret. Such sources will be the subject of additional observations in different wavelength ranges aiming to discover additional

features that could facilitate classification. An example of such a follow-up can be seen in Section 4.3.

The list of the 303 new sources, to be added to those reported in Paper I, and their basic properties are shown in Table 1.

The success rate, defined as the fraction of high- $z$  QSOs among all the sources with a secure classification, as selected by our algorithm (Section 2), is relatively high ( $412/726 = 57\%$ ) for the old Paper I list and  $405/594 \sim 68\%$  for the present work. The main contaminant (in the case of the present work the only one) is represented by low- $z$  ( $z < 2.5$ ) QSOs/AGNs. If we consider only the 1412 selected sources as described in Section 2, 266 have been observed and 199 turned out to be  $z \geq 2.5$  QSOs, corresponding to a success rate of 75%.

In the present work, the success rate could be biased toward higher values, due to the self-learning approach described in Section 2. If we consider the effective success rate, i.e., the global ratio of high- $z$  QSOs identifications carried out by us



**Figure 3.** Composite obtained with the Ultraviolet and Visual Echelle Spectrograph (UVES) spectrum of a candidate at  $z = 2.960$ , with some AGN emission lines highlighted. The gray bands are gaps in the UVES wavelength coverage, depending on the adopted instrument setup.

over all spectroscopic observations done so far, including empirical attempts on improbable candidates, we still obtain a remarkable 52%.

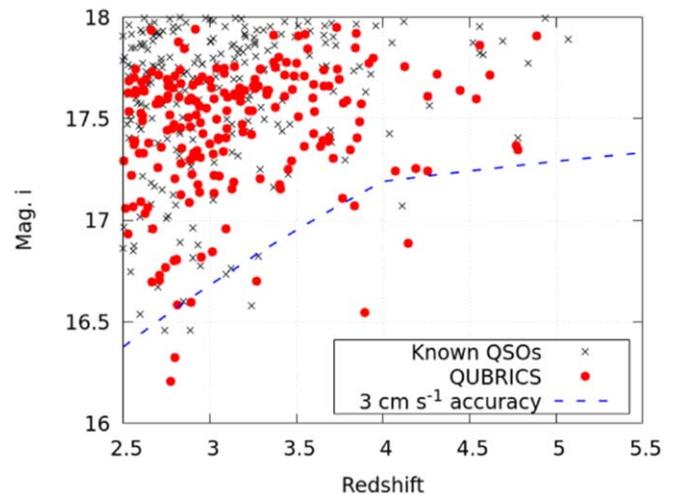
All machine-learning algorithms are biased by the features of the initial input sample. In this survey, the initial training set has been created based on all known spectroscopically confirmed QSOs from the literature, selected with a wide range of methods. Thus, bias toward specific types of sources should be minimal. However, in order to better assess the properties of our (evolving) selections and achieve a better training, we have carried out observations also of some objects that do not comply with our initial specifications—for example, with no Gaia counterpart within the established radius or are not pointlike according to our criteria or fainter than 18 mag in the  $i$  band ( $i > 18$ )—are thus not included in the main sample. These objects are also listed in Table 1, which contains two columns to show whether the object is a candidate according to the criteria of Paper I or/and selected according to the present work. Sources not included in the main sample are indicated with asterisks.

The total number of QSOs with  $z \geq 2.5$  (and  $i \leq 18$ ) in the QUBRICS main sample is 428, with 202 discovered by our survey and 226 sources derived from the literature. The machine-learning algorithm applied in Paper I was able to identify 412 of the 428  $z \geq 2.5$  known QSOs (96%). Following the self-learning approach, i.e., by re-ingesting all new identifications in the training set, this completeness indicator becomes  $405/428 = 95\%$ .

#### 4.1. QSOs at $z \geq 2.5$

Table 1 lists 168 new bright QSOs with spectroscopic redshifts of  $z \geq 2.5$ , and considering the sources already published in Paper I, the total number of new bright QSOs discovered by our survey amounts to 224.

Figure 4 shows the updated redshift-magnitude diagram for  $z \geq 2.5$  quasars in the area of the QUBRICS survey. Red points indicate sources published in Paper I or here. Out of the 10 new QSOs at  $z \geq 4.0$ , the brightest is 316292063 with a  $z = 4.147$  and  $m_i = 16.88$ , as shown in Figure 5. As seen in the figure, this QSO has rather narrow lines and this could perhaps indicate a type 2 QSO or a lensed source. Additional observations have already been requested and its spectrum will be discussed in detail once our follow-up is completed. Our list includes five sources that were



**Figure 4.** The redshift- $i$  magnitude plane of the QSOs in the area of the present survey. Black crosses: QSOs known before the present observations. Red filled circles: new spectroscopic redshifts obtained in QUBRICS. The dashed blue line shows, in the context of the Sandage Test, the locus of an accuracy of  $3 \text{ cm s}^{-1}$  in radial velocity, reachable by allocating 2500 hr of observations at the ELT to a single QSO (see Section 6 for details).

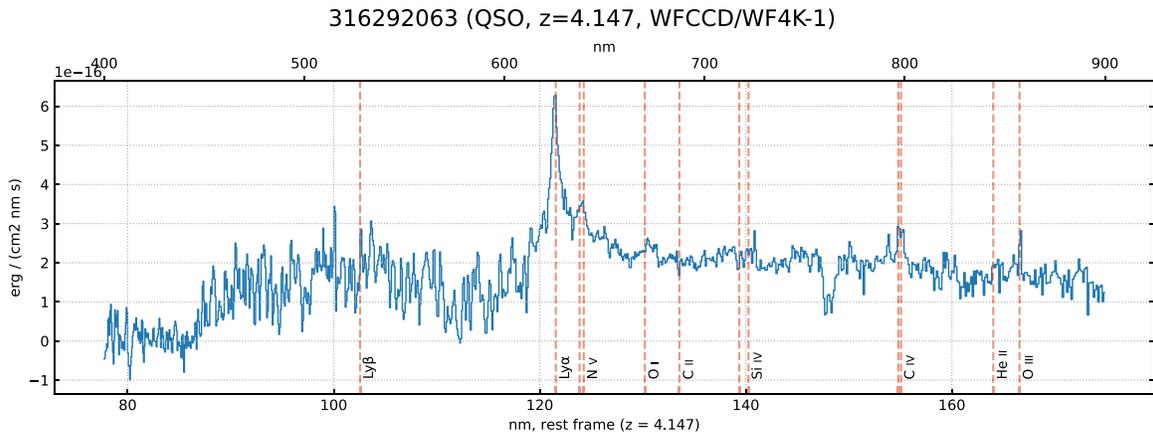
already published by Wolf et al. (2020) and were independently selected and classified by our survey.

#### 4.2. QSOs at $z < 2.5$

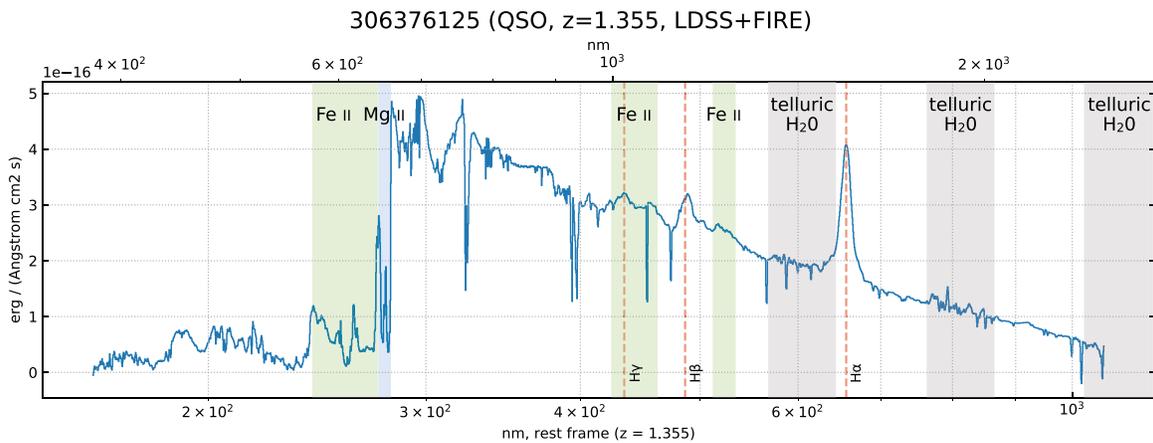
Out of the 121 sources with flag A that are not high- $z$  QSOs, there are 25 with redshifts of  $z \leq 0.5$  and 78 QSOs at  $0.5 < z < 2.5$ . Thus 85% of the sources (103 out of 121) are active galaxies but are not in our desired redshift range. These relatively bright objects may still turn out to be useful for studies of the evolution of the metal content of the IGM and of the Lyman forest at low redshift with space observations.

#### 4.3. Notes on BAL QSOs

Both in the pilot and the main campaign, we have encountered sources with pronounced absorption troughs whose classifications, due to the lack of strong emission lines, is not straightforward. A fraction of these sources can in fact be classified as Overlapping Iron Low-ionization broad absorption line (BAL) QSOs or OFeLoBAL (Hazard et al. 1987). OFeLoBALs are characterized



**Figure 5.** Spectrum, resampled at  $250 \text{ km s}^{-1}$ , of the brightest  $z \sim 4$  QSO in the sample, with some AGN emission lines.



**Figure 6.** Spectrum, resampled at  $250 \text{ km s}^{-1}$ , of a candidate identified as an OFeLoBAL QSO, with some spectral features highlighted. AGN emission lines and metal bands are shown in color, while telluric bands (which have been corrected for in data reduction) are shown in gray.

by extensive systems of low ionization absorption which sometimes appear as broad saturated troughs (e.g., Hall et al. 2002). The troughs may overlap to nearly completely absorb the continuum emission shortward of  $2800 \text{ \AA}$ , effectively mimicking the appearance of the Lyman forest. The estimated fraction of OFeLoBAL QSOs is around 2% (Dai et al. 2012) and their number in our catalog can be as high as the number of genuine  $z > 4$  QSOs.

The nature of these sources can be confirmed by identifying the Balmer series in the AGN emission and in particular the  $H\alpha$  emission line at  $6563 \text{ \AA}$  rest frame. For  $z \gtrsim 0.5$  this line is observed in the near-infrared (NIR) band. We thus obtained NIR spectroscopic data for some of these sources, using the Folded-port InfraRed Echelle (FIRE) at the Baade Magellan telescope with a long-slit configuration.

A composite LDSS-3/FIRE spectrum of one such source (Skymapper ID 306376125), which was flux-calibrated in the visual band using an additional low-resolution spectrum taken by the Magellan Echelle Spectrograph (MagE) at Baade, is shown in Figure 6. For this particular object, the  $H\alpha$  and  $H\beta$  line are clearly identified at  $z = 1.355 \pm 0.001$ , with the possible additional detection of  $H\gamma$  blended within a Fe II emission complex. An overlapping Fe II–Mg II absorption complex at the same redshift is also clearly observed around  $2600\text{--}2800 \text{ \AA}$  rest frame, corroborating the identification of the candidate as an

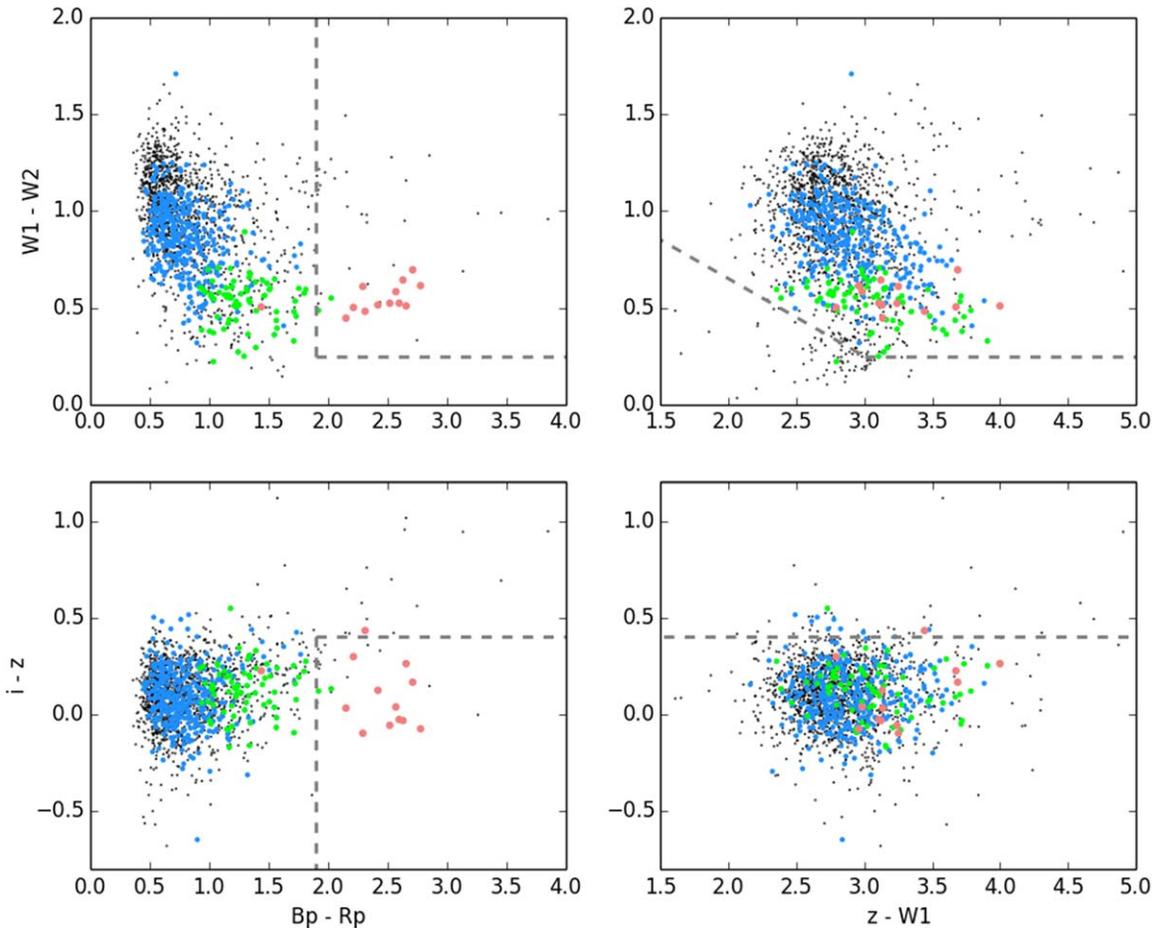
OFeLoBAL QSO. A detailed analysis of this object and of other similar sources (currently assigned a flag B classification and not included in Table 1) is beyond the scope of this work and will be discussed in a future publication.

## 5. Properties of Confirmed QSOs

### 5.1. IR Colors

Several recent surveys targeting super bright high-redshift QSOs are using WISE colors for the selection (Wang et al. 2016; Wolf et al. 2018; Schindler et al. 2019b). Based on a study by Wu et al. (2012), WISE colors are very good at distinguishing QSOs from late-type stars and more than half of the SDSS QSOs have  $W1\text{--}W2 > 0.57$ . Wang et al. (2016) showed that high-redshift QSOs,  $z > 4.5$ , are within the  $0 < (W1\text{--}W2) < 1.5$  color range. This has also been confirmed by the survey conducted by Wolf et al. (2018). In Figure 7 we show the distribution of our confirmed QSO candidates in the color space used by Wolf et al. (2018). In that work they were only interested in  $z > 4.5$  QSOs, but in our sample we have a wide range of redshifts going from  $2.5 \leq z < 5.0$ . In order to be able to directly compare our distributions with the aforementioned works in the literature, presented WISE colors are in Vega magnitudes.

As can be seen in our plots, while all our candidates are indeed within the  $W1\text{--}W2[0,2]$  range, based only on IR colors, it is difficult to distinguish between the various redshifts, since



**Figure 7.** Color-color diagrams of all candidates with known redshift. Blue symbols are sources with  $z \geq 2.5$ , green are with  $z \geq 3.5$ , and red are with  $z \geq 4.5$ . The lines correspond to cuts similar to the ones applied by Wolf et al. (2020) to select high-redshift bright QSOs ( $z \geq 4.5$ ). The area delimited by the cuts is where the majority of  $z \geq 4.5$  sources are found. WISE magnitudes are in Vega and the rest are in the AB system.

all sources are mixed in a small area. Instead, when we include the Gaia ( $B_p - R_p$ ) color, the redshift groups are better separated as seen in the two panels on the left of the figure. Most of our  $z \geq 4.5$  confirmed QSOs have a color of  $B_p - R_p > 1.9$ , in line with the Wolf et al. (2018) selection. In addition, all  $z \geq 2.5$  QSOs have a color  $B_p - R_p > 0.5$ , with the  $z \geq 3.5$  source predominantly being in the area above  $B_p - R_p = 1.0$ . Thus, including bluer colors in the selection can help to better target higher redshift sources.

### 5.2. Crossmatch to GALEX

We have crossmatched the QSOs found in QUBRICS with the catalogs of the sources detected by the Galaxy Evolution Explorer<sup>14</sup> (GALEX; Bianchi et al. 2017). Of the 414 QSOs/AGNs confirmed in Papers I and this work (including those from Table 2) we have 161 detections with a confidence level of  $>2\sigma$  in the near-ultraviolet (NUV) and 83 detections in the far-ultraviolet (FUV); 75 objects are detected both in the FUV and NUV. If we consider only QSOs with  $z \geq 2.5$  ( $z \geq 3.5$ ) out of 230 (54) objects 38 (3) have been detected in the NUV, 17 (2) in the FUV, and 13 (1) both in the FUV and NUV.

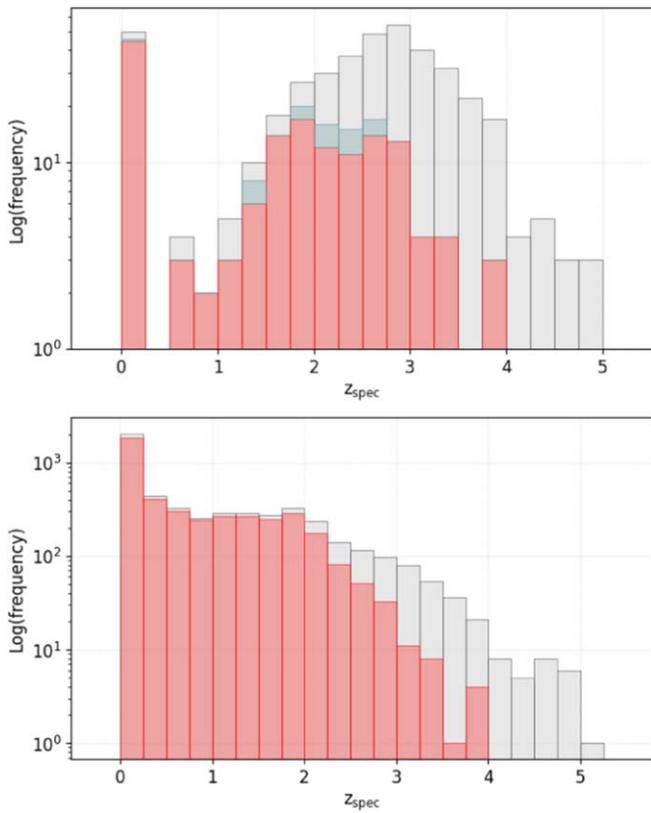
In the designing phase of the QUBRICS we decided against using GALEX data for the selection of the candidates in order

to avoid undesired selection effects relative to the fluctuations in the galactic absorption, which are stronger at UV wavelengths. In fact, we plan to use the QUBRICS sample to study the He II reionization and the cosmic UV background. In order to fulfill this goal, we need an unbiased sampling of all the possible lines of sight in our QSO catalog and a UV selection may select preferentially less absorbed ones (e.g., Prochaska et al. 2009), which might bias future results toward shorter mean-free paths in the IGM (Romano et al. 2019).

Indeed, one of the QSOs confirmed in Paper I, J045011.37-432429.7 with  $z = 3.95$ , shows a detection in the FUV and a non-detection in the NUV. FUV-loud QSOs are extremely rare at  $z > 3.5$  and precious to study the He II reionization (Worseck et al. 2019). J045011.37-432429.7 is potentially a case of transmission peaks that might indicate an example of patchy reionization of He II at  $z \gtrsim 3.5$ .

As shown in Figure 8, all of QSOs/AGNs with a GALEX counterpart have  $z < 4$ , with J045011.37-432429.7 being the object with the highest recorded redshift. Moreover, considering the confirmed QSOs/AGNs in Papers I and this work, 82% of objects below  $z = 2$  have a GALEX counterpart. This number decreases drastically when considering higher redshift intervals: for  $2 < z < 3$  ( $3 < z < 4$ ), 35% (11%) of QSOs have a corresponding GALEX source. Similar results are obtained using all the QSOs/AGNs known in the QUBRICS main sample, the biggest difference being for  $2 < z < 3$  where

<sup>14</sup> All the GALEX data used in this section can be found in MAST [doi:10.17909/T9H59D].



**Figure 8.** Upper panel: redshift distribution of the 414 confirmed QSOs/AGNs (gray columns): 169 of those have a corresponding GALEX source in the FUV, NUV, or both. Of them, 152 (red columns) were observed by us, while the remaining 17 sources (light blue columns) were found in the literature (Section 3.5). Lower panel: redshift distribution of QSOs/AGNs in the QUBRICS main sample. Gray columns show the distribution for all the objects, while red columns are for those with a GALEX counterpart.

a higher percentage (58%) of sources have a GALEX counterpart. In principle, the GALEX detection can be used as a veto criterium to increase the efficiency in the selection of bright QSOs at  $z > 4$ .

## 6. A Golden Sample for the Sandage Test

The redshifts of all cosmologically distant sources are expected to experience a small, systematic drift as a function of time due to the evolution of the expansion rate of the universe (Sandage 1962). Liske et al. (2008), using extensive Monte Carlo simulations, determined the accuracy with which the redshift drift can be measured from the Lyman forest (and metal lines),  $\sigma_v$ , as a function of the signal-to-noise ratio and redshift:

$$\sigma_v = g \times 1.35 \left( \frac{S/N}{3350} \right)^{-1} \left( \frac{1 + z_{\text{QSO}}}{5} \right)^{-\gamma} \left( \frac{N_{\text{QSO}}}{30} \right)^{-0.5} \text{ cm s}^{-1}, \quad (1)$$

where the symbol “S/N” refers to the total S/N per 0.0125 Å pixel per object accumulated over all observations,  $N_{\text{QSO}}$  is the number of QSOs in the sample,  $z_{\text{QSO}}$  is the redshift of the QSO, and the  $\gamma$  exponent is 1.7 for  $z_{\text{QSO}} \leq 4$  and 0.9 above. The form factor  $g$  is equal to 1 if all the targets are observed twice, at the beginning and at the end of the experiment, and becomes larger if the measurements are distributed in time, reaching 1.7 for a uniform distribution. The S/N per pixel for photon-noise

limited observations can be written as

$$S/N = 650 \left[ \frac{Z_X}{Z_r} 10^{0.4(16-m_X)} \left( \frac{D}{39 \text{ minutes}} \right)^2 \frac{t_{\text{int}}}{10 \text{ hr}} \frac{\epsilon}{0.25} \right]^{\frac{1}{2}}, \quad (2)$$

where  $D$ ,  $t_{\text{int}}$ , and  $\epsilon$  are the telescope diameter, total integration time, and total efficiency;  $Z_X$  and  $m_X$  are the zero-point and apparent magnitude of the source in the X-band, respectively; and  $Z_r = (8.88 \times 10^{10}) \text{ s}^{-1} \text{ m}^{-2} \mu\text{m}^{-1}$  is the AB zero-point for an effective wavelength of 6170 Å (corresponding to the SDSS  $r$  band). The normalization of the above equation assumes a pixel size of 0.0125 Å and a central obscuration of the telescope’s primary collecting area of 10%.

Liske et al. (2008) concluded that an Extremely Large Telescope (ELT)-type telescope (at the time planned with a 42 minutes primary mirror) would be capable of unambiguously detecting the redshift drift over a period of  $\sim 20$  yr using 4000 hr of observing time. The estimated amount of time would obviously increase assuming the 39 minutes primary mirror of the ELT and considering only the QSOs observable with the ELT, which are typically those in the Southern Hemisphere. We have repeated the estimate of the time requested to carry out the Sandage Test with the 39 minutes ELT, adopting a strategy that maximizes the significance of the detection of a nonzero redshift drift (Liske et al. 2008) (i.e., with QSOs in the redshift range  $2.8 \leq z < 5$ ), aiming at a  $3\sigma$  detection and observing 30 targets twice at 25 yr distance. Other assumptions are the same as in Liske et al. (2008) except for the spectral slope of the QSO continuum, updated to be  $f_\lambda \propto \lambda^{-1.3}$  (Cristiani et al. 2016). For simplicity we have required that the spectra of all the objects are integrated for a sufficiently long exposure time (different for each target, depending on its magnitude and redshift) to reach the same velocity accuracy ( $22.8 \text{ cm s}^{-1}$ ) required for a global  $3\sigma$  detection of the drift.

We collect in Table 3 our proposed sample of 30 southern QSOs, which are most suitable for the Sandage Test. Thanks to the detection of new bright QSOs at high redshift, the total time required to carry out the Sandage Test turns out to be less than 2500 hr and each QSO in Table 3 needs less than 100 hr of integration to provide a velocity accuracy of  $22.8 \text{ cm s}^{-1}$ .

## 7. Discussion and Conclusions

In this paper, we present spectroscopic identifications of QSO candidates in the QUBRICS sample, using our most updated selection criteria based on the CCA approach, which has been described in Paper I. At present, after 26 observing runs at intermediate and large telescopes, we have been able to approximately double the number of bright QSOs ( $i \leq 18.0$ ) at  $z \geq 2.5$  known in the Southern Hemisphere, bringing it to 428. In this way it has been possible to relieve the persistent lack of bright targets for high-resolution spectrographs of southern observatories.

Using the new spectroscopic sample of bright QSOs from QUBRICS, we are able to further refine our selection criteria by means of a new CCA training set. The completeness of the selection criterion, evaluated against the presently known bright QSOs at  $z \geq 2.5$ , turns out to be higher than 90%, while the success rate is around 70%. The success of the QUBRICS survey is particularly evident when looking at Figure 4, where among the new identifications, we show the two brightest QSOs at  $z \geq 3.8$  in the Southern Hemisphere, with an increase of a factor of  $\sim 3$  of the number of QSOs at  $z \geq 3$  and  $i \leq 17.5$ .

**Table 3**  
The Golden Sample of Southern QSOs for the Sandage Test

Name	R.A. (J2000) (deg)	Decl. (J2000) (deg)	$m_r$ (mag)	$m_i$ (mag)	$G_{\text{Gaia}}$ (mag)	Redshift	$t(22.5 \text{ cm s}^{-1})$ (hr)	ID Skymapper
J000322.94-260317.9	000.845639	-26.055083	17.472	17.071	17.6019	4.111	61.2	7437380
J000736.55-570151.8	001.902387	-57.031067	17.992	17.240	17.9440	4.260	70.5	317358346
J004131.41-493611.7	010.381000	-49.603247	16.736	16.580	16.7048	3.240	68.5	317850840
J010318.05-130509.9	015.825313	-13.086163	17.637	17.242	17.7349	4.072	71.9	8430815
J015558.26-192848.8	028.992816	-19.480298	17.673	17.393	17.7969	3.655	96.7	8789744
J015644.67-692216.1	029.186291	-69.371155	16.637	16.325	16.5949	2.800	78.7	314938059
J020413.26-325122.8	031.055213	-32.856350	17.276	17.068	17.3900	3.835	66.2	6932623
J030722.89-494548.2	046.845395	-49.763410	*	17.407	18.4534	4.728	78.8	318204033
J033015.31-543021.1	052.563812	-54.505886	17.203	17.174	17.2208	3.400	97.4	316834279
J042214.78-384452.9	065.561596	-38.748007	16.896	16.765	16.9504	3.123	84.2	10914737
J054803.19-484813.1	087.013348	-48.803654	17.215	16.886	17.2414	4.147	51.4	316292063
J093542.69-065118.8	143.927899	-06.855260	18.008	17.424	18.1236	4.040	85.3	56483517
J094253.51-110425.9	145.722949	-11.073870	16.915	16.733	16.8174	3.093	87.0	56438607
J101529.36-121314.3	153.872359	-12.220648	17.857	17.255	17.7070	4.190	72.0	57929040
J104856.82-163709.2	162.236717	-16.619244	17.194	17.164	17.2249	3.370	98.0	57339247
J105122.70-065047.8	162.844537	-06.846624	17.584	17.345	17.8003	3.765	88.1	58181076
J105449.68-171107.3	163.706991	-17.185386	17.104	17.107	17.2681	3.765	70.8	57368436
J111054.67-301129.8	167.727860	-30.191652	*	17.346	18.5117	4.780	74.2	54680559
J132029.98-052335.1	200.124872	-05.393160	17.559	17.444	17.8589	3.700	99.4	65911949
J144943.17-122717.5	222.429861	-12.454868	16.953	16.703	17.0963	3.273	82.3	66962683
J150527.83-204534.9	226.365934	-20.759752	16.966	16.955	16.9888	3.090	91.3	72002629
J162116.92-004250.8	245.320512	-00.714142	17.457	17.386	17.6713	3.703	94.1	114286192
J195302.67-381548.3	298.261202	-38.263482	17.724	17.305	17.8896	3.712	87.0	135100950
J200324.11-325145.0	300.850485	-32.862537	17.470	17.296	17.6287	3.783	83.6	135386798
J201741.49-281630.0	304.422893	-28.274988	17.822	17.388	17.7996	3.685	95.0	136198132
J212518.38-420547.6	321.326583	-42.096609	17.266	17.363	17.6598	3.549	98.8	307443251
J212540.96-171951.3	321.420690	-17.330947	16.644	16.548	16.8730	3.897	39.9	2379862
J212912.18-153840.9	322.300733	-15.644734	17.035	16.824	17.0243	3.280	88.4	2552049
J215228.21-444603.9	328.117538	-44.767760	17.310	17.291	17.3645	3.473	94.8	307536968
J215728.21-360215.1	329.367623	-36.037557	*	17.367	18.2864	4.771	75.7	397340

The QUBRICS survey also selects rare and peculiar sources, with strong absorption features, e.g., the OFeLoBAL QSO discussed in Section 4.3. With the progress of our survey it will be important to quantify the number densities of these peculiar sources and to compare them with the relative low incidence in the past surveys.

Due to the characteristics of our selection, QUBRICS is expected to be incomplete for QSOs at  $z \gtrsim 4.5$  or with images distorted by lensing. In the future, we will address these issues and try to reduce the biases with improved, less stringent selection criteria.

Surveys similar to QUBRICS have been started recently, based mainly on IR selections of bright high- $z$  QSOs (e.g., Schindler et al. 2019b; Wolf et al. 2020). IR colors from 2MASS and WISE are fundamental to distinguish high- $z$  QSOs from local galaxies or stars, but, taken alone, they are not optimal for a photometric redshift refinement. As discussed in Section 5, the optical colors  $B_p - R_p$  or  $i - z$  are very useful to separate QSOs at  $z \sim 2$  from the ones at  $z \sim 4$ .

Furthermore, by adopting further filters, in particular at short wavelengths (e.g., GALEX), it would be possible to further improve the success rate of the survey at high  $z$ . However, due to the possibility of introducing a bias against clear lines of sight in the intervening IGM, which can mimic the photometry of a lower redshift source (Romano et al. 2019), we chose not to use the GALEX photometry for the selection. In this way, among the newly discovered QSOs at  $z > 2.5$ , we have 29 objects with FUV or NUV detections, which are potentially important for follow-up

studies with the Cosmic Origins Spectrograph (COS) for the He II reionization (e.g., Worseck et al. 2016, 2019).

The adopted CCA method to compute a photometric redshift, described in Paper I, is based on spectra energy distributions (SEDs) extending from the  $u$  band to the WISE bands and should be less affected by possible biases due to the presence of significant Lyman limit absorbers with short mean-free paths, which are evident at  $z \sim 3.5$  on QSO surveys based on  $u - g$  color selections, as shown by Prochaska et al. (2009), Cristiani et al. (2016), and Romano et al. (2019). In the future, we will address the effects of optical color selections on the QSO spectral properties and Lyman limit statistics.

The discovery of bright cosmic beacons is especially important for the study of the IGM at high  $z$ . In particular, the QUBRICS survey provides a large sample of bright high- $z$  QSOs for the Sandage Test with future 40 m class telescopes. We estimate that at present, less than 2500 hr of observations in 25 yr are needed with the ELT–High Resolution Spectrograph to carry out the redshift drift measurement (Liske et al. 2008) at the precision required to have a  $3\sigma$  detection of the cosmological signal. Before QUBRICS, the targets available in the Southern Hemisphere would have required, in the same metrics, about 4000 hr to accomplish this goal. QUBRICS, in a sense, contributes to a multi-millions worth of savings, considering the projected cost of a night at ELT ( $\gtrsim 50$  million EUR or  $\gtrsim 60$  million USD at the assumed cost of 320 kilo EUR for an ELT observing night and an average of 9 hr per night).

We are continuing the pursuit for the brightest cosmic beacons, exploring other innovative methods, with the main

goal of giving a realistic possibility to execute the Sandage test, which was only dreamed of a few years ago.

We thank Gabor Worseck for enlightening discussions.

This work is based on data products from observations made with ESO Telescopes at La Silla Paranal Observatory under ESO programmes ID 103.A-0746(A), 0103.A-0746(B), and 0104.A-0754(A).

The national facility capability for SkyMapper has been funded through ARC LIEF grant LE130100104 from the Australian Research Council, awarded to the University of Sydney, the Australian National University, Swinburne University of Technology, the University of Queensland, the University of Western Australia, the University of Melbourne, Curtin University of Technology, Monash University, and the Australian Astronomical Observatory. SkyMapper is owned and operated by The Australian National University's Research School of Astronomy and Astrophysics. The survey data were processed and provided by the SkyMapper Team at ANU. The SkyMapper node of the All-Sky Virtual Observatory (ASVO) is hosted at the National Computational Infrastructure (NCI). Development and support the SkyMapper node of the ASVO has been funded in part by Astronomy Australia Limited (AAL) and the Australian Government through the Commonwealth's Education Investment Fund (EIF) and National Collaborative Research Infrastructure Strategy (NCRIS), particularly the National eResearch Collaboration Tools and Resources (NeCTAR) and the Australian National Data Service Projects (ANDS).

This work has made use of data from the European Space Agency (ESA) mission Gaia (<https://cosmos.esa.int/gaia>), processed by the Gaia Data Processing and Analysis Consortium (DPAC, <https://cosmos.esa.int/web/gaia/dpac/consortium>). Funding for the DPAC has been provided by national institutions, in particular the institutions participating in the Gaia Multilateral Agreement.

This publication makes use of data products from 2MASS, which is a joint project of the University of Massachusetts and the Infrared Processing and Analysis Center/California Institute of Technology, funded by the National Aeronautics and Space Administration and the National Science Foundation.

This publication makes use of data products from WISE, which is a joint project of the University of California, Los Angeles, and the Jet Propulsion Laboratory/California Institute of Technology, funded by the National Aeronautics and Space Administration.

This paper includes data gathered with the 6.5 m Magellan Telescopes located at Las Campanas Observatory, Chile.

This research is based on observations made with the Galaxy Evolution Explorer, obtained from the MAST data archive at the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Inc., under NASA contract NAS 526555.

R.A. acknowledges ANID FONDECYT Regular Grant 1202007.

*Facilities:* Skymapper, Wise, 2MASS, Gaia, NTT (EFOSC2), Magellan:Baade (IMACS), Magellan:Clay (LDSS-3), Magellan:Baade (FIRE), duPont (WFCCD), Magellan:Baade (MagE)

## Appendix CCA Matrixes

Our QSO candidate selection procedure relies on a CCA transformation aimed at predicting the object type classification

**Table A1**  
CCA Transformation Matrix for the Magnitude Band Combination in the Largest Subsample (Covering 43% of the Main Sample)

Band	Paper I	This Work
G (Gaia)	-9.652	-10.02
BP (Gaia)	-4.571	-3.680
RP (Gaia)	+10.21	+9.332
g (Skymapper)	+2.297	+2.091
r (Skymapper)	+4.637	+4.006
i (Skymapper)	-2.129	-1.386
z (Skymapper)	+3.089	+3.407
J (2MASS)	-0.8026	-0.7531
H (2MASS)	+3.733	+3.984
K (2MASS)	+0.458	+0.5796
W1 (WISE)	-8.250	-7.629
W2 (WISE)	+0.263	-0.7174
W3 (WISE)	+2.348	+2.326
log G (Gaia)	+142.9	+149.4
log BP (Gaia)	+72.69	+57.33
log RP (Gaia)	-138.6	-125.4
log g (Skymapper)	-40.05	-36.85
log r (Skymapper)	-71.36	-61.19
log i (Skymapper)	+33.78	+22.85
log z (Skymapper)	-45.00	-49.95
log J (2MASS)	+14.76	+13.89
log H (2MASS)	-50.95	-53.82
log K (2MASS)	-8.637	-10.47
log W1 (WISE)	+141.7	+132.2
log W2 (WISE)	-22.17	-6.822
log W3 (WISE)	-41.23	-40.81
offset	+10.45	+5.014

for each source, based on the photometric magnitudes in the bands where these estimates are available. In particular we associate a discrete numeric label to all sources with known object type classification and search for the transformation matrix to be applied to the available magnitudes, which maximizes the correlation with the above mentioned label. Note that the above process identifies the best possible linear transformation while the best performances would likely be obtained with a nonlinear one, although the latter can not be obtained analytically. In order to improve our classification algorithm we also included the logarithm of each magnitude estimate in the CCA analysis. It should be noted that the availability of magnitude estimates is not the same for all the sources in our main sample, hence we identified 34 subsamples with homogeneous availability and repeated the above process for each of them, resulting in 34 different CCA matrices. The largest subsample covers 43% of our main sample (438,687 sources, with 585,81 sources lacking a known object type classification). In Table A1 we report the transformation matrix used to calculate the CCA coordinate (i.e., the value displayed on the vertical axis of Figure 1 in both Papers I and this work), for the above mentioned subsample.

All magnitudes should be converted to the AB system before applying the transformation, and an offset (the last line of Table A1) should be added to the resulting label to obtain a classification on the same scale of the training. We can provide the matrices for the remaining sources on request.<sup>15</sup> Further details are available in Paper I.

Note that the presence of strong correlations among the magnitudes in different bands prevents us from drawing any

<sup>15</sup> Please write to [giorgio.calderone@inaf.it](mailto:giorgio.calderone@inaf.it).

conclusion by just comparing the numbers in the above matrices. The overall performance of each matrix, in terms of success rates and completeness, should thus be estimated by comparing the results with the test data set, as discussed in Section 2.

### ORCID iDs

Konstantina Boutsia  <https://orcid.org/0000-0003-4432-5037>

Andrea Grazian  <https://orcid.org/0000-0002-5688-0663>

Giorgio Calderone  <https://orcid.org/0000-0002-7738-5389>

Stefano Cristiani  <https://orcid.org/0000-0002-2115-5234>

Guido Cupani  <https://orcid.org/0000-0002-6830-9093>

Francesco Guarneri  <https://orcid.org/0000-0003-4740-9762>

Fabio Fontanot  <https://orcid.org/0000-0003-4744-0188>

Ricardo Amorin  <https://orcid.org/0000-0001-5758-1000>

Valentina D’Odorico  <https://orcid.org/0000-0003-3693-3091>

Emanuele Giallongo  <https://orcid.org/0000-0003-0734-1273>

Mara Salvato  <https://orcid.org/0000-0001-7116-9303>

Michael Romano  <https://orcid.org/0000-0002-9948-3916>

Nicola Menci  <https://orcid.org/0000-0002-4096-2680>

### References

- Anderson, T. 1984, *An Introduction to Multivariate Statistical Analysis* (New York: Wiley)
- Bate, N. F., Webster, R. L., & Wyithe, J. S. B. 2007, *MNRAS*, **381**, 1591
- Becker, G., D’Aloisio, A., Davies, F. B., Hennawi, J. F., & Simcoe, R. A. 2019, *BAAS*, **51**, 440
- Bianchi, L., Shiao, B., & Thilker, D. 2017, *ApJS*, **230**, 24
- Bonvin, V., Courbin, F., Suyu, S. H., et al. 2017, *MNRAS*, **465**, 4914
- Calderone, G., Boutsia, K., Cristiani, S., et al. 2019, *ApJ*, **887**, 268
- Colless, M., Dalton, G., Maddox, S., et al. 2001, *MNRAS*, **328**, 1039
- Cooke, R. J., & Fumagalli, M. 2018, *NatAs*, **2**, 957
- Cristiani, S., Serrano, L. M., Fontanot, F., Vanzella, E., & Monaco, P. 2016, *MNRAS*, **462**, 2478
- Dai, X., Shankar, F., & Sivakoff, G. R. 2012, *ApJ*, **757**, 180
- D’Odorico, V., Cristiani, S., Pomante, E., et al. 2016, *MNRAS*, **463**, 2690
- Fan, X., Strauss, M. A., & Schneider, D. P. 2001, *AJ*, **121**, 54
- Fontanot, F., Cristiani, S., Monaco, P., et al. 2007, *A&A*, **461**, 39
- Fontanot, F., de Lucia, G., Hirschmann, M., et al. 2020, *MNRAS*, **496**, 3943
- Gaia Collaboration, Brown, A. G. A., Vallenari, A., et al. 2018, *A&A*, **616**, A1
- Grazian, A., Giallongo, E., Boutsia, K., et al. 2018, *A&A*, **613**, A44
- Hall, P. B., Anderson, S. F., Strauss, M. A., et al. 2002, *ApJS*, **141**, 267
- Hazard, C., McMahon, R. G., Webb, J. K., & Morton, D. C. 1987, *ApJ*, **323**, 263
- Iršič, V., Viel, M., Haehnelt, M. G., et al. 2017, *PhRvD*, **96**, 023522
- Jones, D. H., Read, M. A., Saunders, W., et al. 2009, *MNRAS*, **399**, 683
- Leite, A. C. O., Martins, C. J. A. P., Molaro, P., Corre, D., & Cristiani, S. 2016, *PhRvD*, **94**, 123512
- Liske, J., Grazian, A., Vanzella, E., et al. 2008, *MNRAS*, **386**, 1192
- Lucey, J. R., Schechter, P. L., Smith, R. J., & Anguita, T. 2018, *MNRAS*, **476**, 927
- Pâris, I., Petitjean, P., Aubourg, É., et al. 2018, *A&A*, **613**, A51
- Prochaska, J. X., Worseck, G., & O’Meara, J. M. 2009, *ApJL*, **705**, L113
- Romano, M., Grazian, A., Giallongo, E., et al. 2019, *A&A*, **632**, A45
- Sandage, A. 1962, *ApJ*, **136**, 319
- Schindler, J.-T., Fan, X., Huang, Y.-H., et al. 2019b, *ApJS*, **243**, 5
- Schindler, J.-T., Fan, X., McGreer, I. D., et al. 2019a, *ApJ*, **871**, 258
- Skrutskie, M. F., Cutri, R. M., Stiening, R., et al. 2006, *AJ*, **131**, 1163
- Tie, S. S., Martini, P., Mudd, D., et al. 2017, *AJ*, **153**, 107
- Valiante, R., Schneider, R., Volonteri, M., & Omukai, K. 2016, *MNRAS*, **457**, 3356
- Véron-Cetty, M. P., & Véron, P. 2010, *A&A*, **518**, A10
- Wang, F., Wu, X.-B., Fan, X., et al. 2016, *ApJ*, **819**, 24
- Webster, R. L., Ferguson, A. M. N., Corrigan, R. T., & Irwin, M. J. 1991, *AJ*, **102**, 1939
- Werk, J. K., Prochaska, J. X., Tumlinson, J., et al. 2014, *ApJ*, **792**, 8
- Wolf, C., Hon, W. J., Bian, F., et al. 2020, *MNRAS*, **491**, 1970
- Wolf, C., Onken, C. A., Luvaul, L. C., et al. 2018, *PASA*, **35**, e010
- Worseck, G., Davies, F. B., Hennawi, J. F., & Prochaska, J. X. 2019, *ApJ*, **875**, 111
- Worseck, G., Prochaska, J. X., Hennawi, J. F., & McQuinn, M. 2016, *ApJ*, **825**, 144
- Worseck, G., Prochaska, J. X., O’Meara, J. M., et al. 2014, *MNRAS*, **445**, 1745
- Wright, E. L., Eisenhardt, P. R. M., Mainzer, A. K., et al. 2010, *AJ*, **140**, 1868
- Wu, X.-B., Hao, G., Jia, Z., Zhang, Y., & Peng, N. 2012, *AJ*, **144**, 49
- Wu, X.-B., Wang, F., Fan, X., et al. 2015, *Natur*, **518**, 512