



<b>Publication Year</b>	2019
<b>Acceptance in OA</b>	2021-01-29T16:38:37Z
<b>Title</b>	Finding the Brightest Cosmic Beacons in the Southern Hemisphere
<b>Authors</b>	CALDERONE, GIORGIO, Boutsia, Konstantina, CRISTIANI, Stefano, GRAZIAN, Andrea, Amorin, Ricardo, D'ODORICO, Valentina, CUPANI, Guido, FONTANOT, Fabio, Salvato, Mara
<b>Publisher's version (DOI)</b>	10.3847/1538-4357/ab510a
<b>Handle</b>	<a href="http://hdl.handle.net/20.500.12386/30113">http://hdl.handle.net/20.500.12386/30113</a>
<b>Journal</b>	THE ASTROPHYSICAL JOURNAL
<b>Volume</b>	887



# Finding the Brightest Cosmic Beacons in the Southern Hemisphere

Giorgio Calderone<sup>1</sup>, Konstantina Boutsia<sup>2</sup>, Stefano Cristiani<sup>1,3,4</sup>, Andrea Grazian<sup>5</sup>, Ricardo Amorin<sup>6,7</sup>,  
Valentina D’Odorico<sup>1,8</sup>, Guido Cupani<sup>1</sup>, Fabio Fontanot<sup>1,4</sup>, and Mara Salvato<sup>9</sup>

<sup>1</sup> INAF—Osservatorio Astronomico di Trieste Via G.B. Tiepolo, 11, I-34143 Trieste, Italy; [giorgio.calderone@inaf.it](mailto:giorgio.calderone@inaf.it)

<sup>2</sup> Las Campanas Observatory, Carnegie Observatories, Colina El Pino, Casilla 601, La Serena, Chile

<sup>3</sup> INFN—National Institute for Nuclear Physics via Valerio 2, I-34127 Trieste, Italy

<sup>4</sup> IFPU—Institute for Fundamental Physics of the Universe via Beirut 2, I-34151 Trieste, Italy

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

<sup>6</sup> Instituto de Investigación Multidisciplinar en Ciencia y Tecnología, Universidad de La Serena Raul Bitrán 1305, La Serena, Chile

<sup>7</sup> Departamento de Física y Astronomía, Universidad de La Serena Av. Juan Cisternas 1200 Norte, La Serena, Chile

<sup>8</sup> Scuola Normale Superiore P.zza dei Cavalieri, I-56126 Pisa, Italy

<sup>9</sup> Max-Planck-Institut für extraterrestrische Physik Giessenbachstrasse 1, Garching D-85748, Germany

Received 2019 August 28; revised 2019 October 11; accepted 2019 October 23; published 2019 December 26

## Abstract

The study of absorptions along the lines of sight to bright high- $z$  QSOs is an invaluable cosmological tool that provides a wealth of information on the inter-/circum-galactic medium, dark matter, primordial elements, reionization, fundamental constants, and general relativity. Unfortunately, the number of bright ( $i \lesssim 18$ ) QSOs at  $z \gtrsim 2$  in the southern hemisphere is much lower than that in the north, due to the lack of wide multiwavelength surveys at decl.  $\delta < 0^\circ$ , hampering the effectiveness of observations from southern observatories. In this work we present a new method based on Canonical Correlation Analysis to identify such objects, taking advantage of a number of available databases: SkyMapper, *Gaia* DR2, *WISE*, and 2MASS. Our QSO candidate sample lists 1476 sources with  $i < 18$  over 12,400 square degrees in the southern hemisphere. With a preliminary campaign we observed spectroscopically 69 of them, confirming 54 new bright QSOs at  $z > 2.5$ , corresponding to a success rate of our method of  $\sim 80\%$ . Furthermore, we estimate a completeness of  $\sim 90\%$  of our sample at completion of our observation campaign. The new QSOs confirmed by this first and the forthcoming campaigns (QUBRICS<sup>10</sup> survey) will be the targets of subsequent studies using higher resolution spectrographs, like ESPRESSO, UVES, and (in the long term) ELT/HIRES.

*Unified Astronomy Thesaurus concepts:* Quasars (1319); Sky surveys (1464); Catalogs (205)

## 1. Introduction

The study of absorption lines in the spectra of high-redshift quasi-stellar objects (QSOs) is a fundamental tool for cosmology (Meiksin 2009; McQuinn 2016). Along the lines of sight to these powerful light beacons, every parcel of the intervening gas selectively absorbs wavelengths of light, providing information about the spatial distributions, motions, temperature, chemical enrichment, and ionization histories of gaseous structures from redshift seven and beyond until the present.

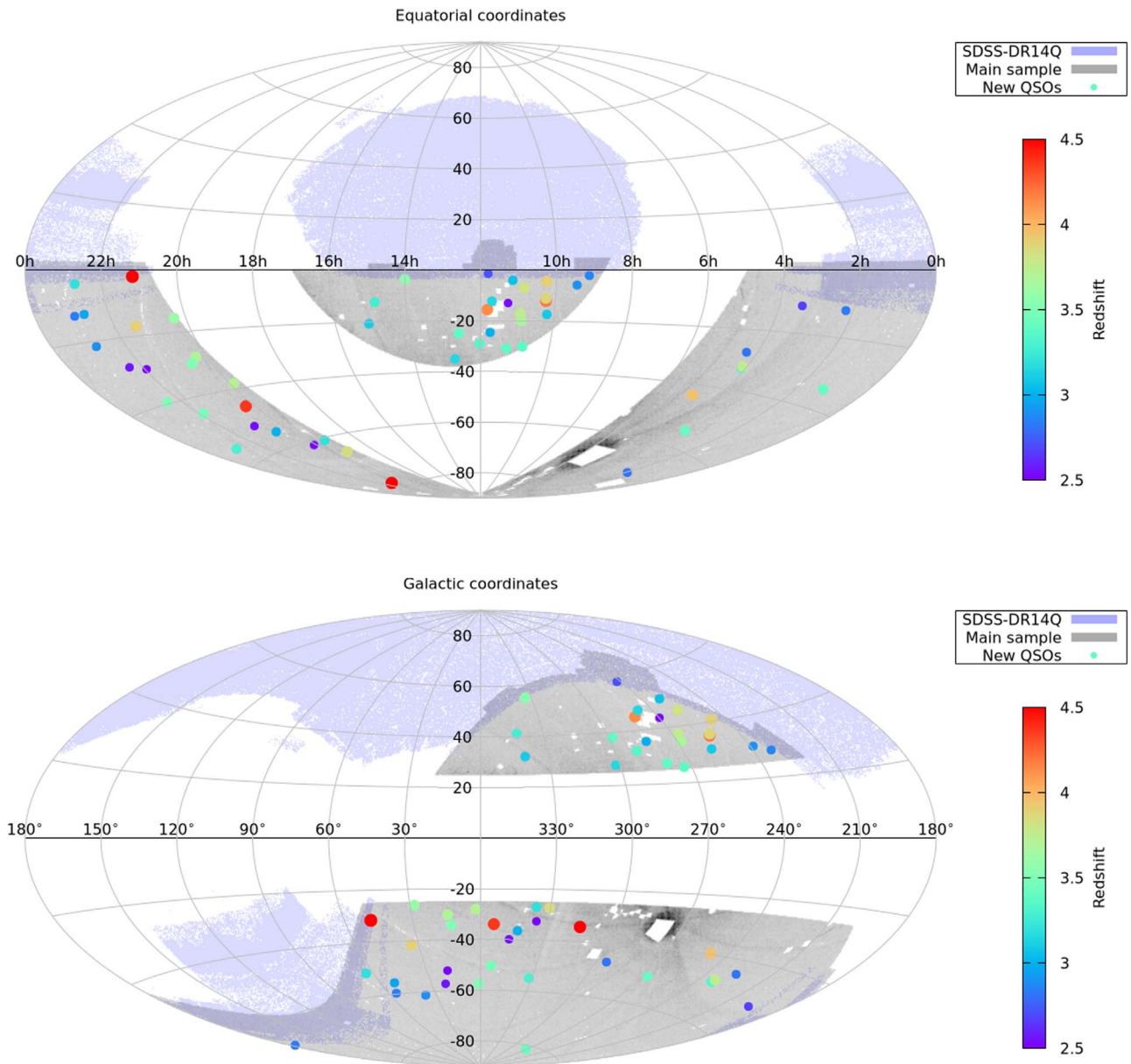
In particular, thanks to QSO absorption lines, it is possible to address issues, e.g., what were the physical conditions of the primordial universe? What fraction of the matter was in a diffuse medium and what fraction and how early condensed in clouds? Where are most of the baryons at the various redshifts? When and how did the formation of galaxies and large-scale structure start? How early and in what amount have metals been produced? When and how (after the Dark Ages following recombination) did the universe get reionized? What was the typical radiation field, how homogeneous, and what was producing it? Which constraints on cosmological parameters and types of dark matter (e.g., neutrinos) are derived from the large-scale structure traced by the inter-galactic medium? Does the standard Big Bang nucleosynthesis model make the correct predictions about the primordial element abundances and the temperature evolution of the CMB? Do fundamental constants

of physics (e.g., the fine-structure constant,  $\alpha$ , or the proton-to-electron mass ratio,  $\mu$ ) vary with cosmic time? Does general relativity correctly describe the expansion of our universe? In order to efficiently pursue these and other similar lines of investigation, it is essential to have the brightest possible light beacons in the background.

Historically, observations in the southern hemisphere have been hampered by the lack of luminous targets with respect to the north, due to a lesser investment of telescope time to search for bright QSOs in the south. As an example, the *Quasar Deep Spectrum* observations carried out with the UVES spectrograph (D’Odorico et al. 2016) have targeted the QSO HE0940-1050 ( $z_{\text{em}} = 3.09$ ,  $V = 16.9$ ) and required 64.4 hr to reach a signal-to-noise ratio, per resolution element ( $R = 45,000$ ), of 120–500 and 320–500 in the O VI/Ly $\alpha$  region and in the C IV region, respectively. HE0940-1050 is still the best target at this redshift in the south, but it is not comparable to the (lensed) beacons B1422 + 231 ( $z_{\text{em}} = 3.62$ ,  $V = 15.8$ ) or APM 08279 + 5255 ( $z_{\text{em}} = 3.91$ ,  $V = 15.2$ ), which have been available for observers in the northern hemisphere.

It is particularly urgent now to fill this gap in view of the upcoming new instrumentation in the southern hemisphere, like ESPRESSO at VLT and the planning of new experiments (e.g., the Sandage test with the high-resolution spectrograph HIRES at the E-ELT (Cristiani et al. 2007; Liske et al. 2008), or the test of the stability of the fine-structure constant and other fundamental couplings (Leite & Martins 2016). Moreover, finding bright radio-loud QSOs at high- $z$  is particularly

<sup>10</sup> Acronym for “*QUasar BRillianti per la Cosmologia nell’emisfero Sud*,” i.e., bright quasars for cosmology in the Southern Hemisphere.



**Figure 1.** Maps of the locations of the sources in the *main sample* (Section 3.1, gray regions, a darker color indicates a higher density of sources). The locations of the QSOs in the SDSS DR14Q (P aris et al. 2018) are represented by blue shaded regions. The new QSOs identified in this work are shown with filled circles, whose colors indicates the redshift. Upper panel: equatorial coordinates; lower panel: galactic coordinates.

important to study the 21 cm forest in absorption with future breakthrough facilities, like the Square Kilometer Array (SKA) in the southern hemisphere, as proposed by Carilli et al. (2002). In addition, UV/optically bright QSOs at  $z > 3$  with lines of sight free from Lyman limit systems (LLS) up to the He II forest are particularly rare but extremely valuable to study the He II reionization (Syphers & Shull 2014; Worseck et al. 2016; Worseck et al. 2019).

By comparing QSO surface densities, it is statistically evident that relatively high- $z$  objects of bright apparent magnitudes must also be present in the southern hemisphere: of the 22 known QSOs with  $z > 2.8$  and  $V < 17$ , only 5 are at  $\delta < 0^\circ$ , and all three with  $V < 16$  are in the north. Such imbalance exists because historically many surveys (e.g., the SDSS) have focused their efforts mainly in the northern hemisphere (see Figure 1). The present time, however, is ripe for a dramatic change of scenario thanks to new surveys

available through the whole sky, or insisting mainly in the southern hemisphere, such as *Gaia* DR2, Skymapper, 2MASS, and *WISE* (Skrutskie et al. 2006; Wright et al. 2010; Gaia Collaboration 2016, 2018; Wolf et al. 2018).

In this paper we describe the first results of a program aiming at filling this gap in the southern hemisphere, finding the brightest QSOs at  $z > 2.5$  that will be observed at high resolution with the present and future breakthrough facilities.

## 2. The Need and Design of a New Survey

The typical range of apparent magnitudes of interest, having in mind a follow-up with high-resolution spectroscopy, is  $i \lesssim 17$  at  $z \sim 2.5$  and  $i \lesssim 18$  at  $z \sim 4$ .

In order to estimate the expected surface densities of QSOs, we have adopted the parameterization of the luminosity function by Kulkarni et al. (2019). We extract random values

**Table 1**Cumulative Surface Density of QSOs at Different Redshifts and  $i$ -band Magnitude Limits Expected from the Luminosity Function of Kulkarni et al. (2019)

$i \leq$	$\Sigma_{\text{QSO}}(2.5 < z < 3.0)$	$\Sigma_{\text{QSO}}(3.0 < z < 3.5)$	$\Sigma_{\text{QSO}}(3.5 < z < 4.0)$	$\Sigma_{\text{QSO}}(4.0 < z < 4.5)$	$\Sigma_{\text{QSO}}(4.5 < z < 5.0)$
15.5	0.10–0.29	0.00–0.14	0.00–0.09	0.00–0.01	0.00–0.00
16.0	0.51–2.03	0.05–0.56	0.00–0.45	0.00–0.03	0.00–0.02
16.5	2.65–12.1	0.42–2.42	0.04–1.64	0.00–0.11	0.00–0.04
17.0	11.7–42.6	2.03–10.8	0.26–6.26	0.05–0.50	0.01–0.19
17.5	51.5–132.2	10.8–44.9	1.90–22.5	0.42–1.91	0.15–0.63
18.0	214.3–412.3	51.1–182.0	10.5–80.3	2.36–8.41	0.77–2.91

**Note.** The minimum and maximum ranges of surface density in a given redshift interval for QSOs brighter than a given  $i$ -band magnitude limit is provided. The expected bright QSO number counts are based on the best fit of individual luminosity functions by Kulkarni et al. (2019; their Table 2), as well as on the global fit with a continuous evolution in the range  $0 < z < 7$  also by Kulkarni et al. (2019; models 1, 2, and 3). All the surface densities are expressed in units of  $10^{-4} \text{ deg}^{-2}$ .

of redshift and  $M_{1450}$  absolute magnitude following the best-fit luminosity functions in different redshift bins from Table 2 of Kulkarni et al. (2019). Then we associate each simulated QSO to different templates from the Polletta empirical library of AGNs (Polletta et al. 2008) and convert the absolute magnitude into observed magnitudes in the adopted photometric system (i.e., Skymapper  $u, v, g, r, i, z$ ; *Gaia*  $BP, G, RP$ ; 2MASS  $J, H, K$ ; *WISE*  $W1, W2, W3, W4$ ; see the next sections for a detailed description). We assume here a null Galactic dust extinction, since, as we discuss in the following, we select targets at high galactic latitudes.

Working at bright absolute magnitudes, it is likely to be affected by small number statistics. To avoid this effect, we simulate a sky area of  $10^5 \text{ sq. deg.}$ , i.e.,  $\gtrsim 10$  times larger than the area available from the present survey, and repeat the simulation 10 times. This choice reduces the shot noise in the simulated number counts.

According to Kulkarni et al. (2019), the best-fit values of the QSO luminosity function (QLF) in their Table 2 can be affected by systematic errors due to the adopted survey selection functions. This is particularly true at  $z = 2\text{--}4$ , where discontinuities and scatter in the QLF parameters appear over short redshift intervals. To avoid such discontinuities, we have computed the surface density adopting also the QLF resulting from a global continuous fit over the redshift range  $0 < z < 7$  by Kulkarni et al. (2019) with a complex parameterization of the redshift evolution of the slopes,  $\Phi^*$ , and  $M^*$  parameters. We also used their Models 1, 2, and 3 parameterizations to derive our estimates of the bright QSO number counts at  $z > 2.5$ .

In Table 1 we summarize the expected cumulative surface densities of QSOs in different bins of  $i$ -band magnitude and redshift. We provide the minimum and maximum expected values for the integral number counts based on the best-fit values and on the models 1, 2, and 3 by Kulkarni et al. (2019). We do not consider here the effect of strong lensing, which can increase the luminosity of high- $z$  QSOs if their lines of sight are well aligned with the deep potential wells produced by galaxy overdensities or by single massive galaxies. The adopted luminosity functions, indeed, could be already affected by strong lensing in the bright end, especially at high- $z$  (Fan et al. 2019; Pacucci & Loeb 2019).

In the following, we will use our predictions in Table 1 as a reference for the expected bright QSO number counts. We expect that they should not be strongly affected by incompleteness, at least at very bright absolute magnitudes, since the models in Kulkarni et al. (2019) are based on the best surveys available in the literature, reaching much fainter magnitudes than the Skymapper survey (down to  $i \sim 22$ ).

In Table 2 we compare the expected number of QSOs at galactic latitudes  $|b| > 25^\circ$  with the number of presently known QSOs in the northern and southern hemisphere, respectively. As already known, a significant discrepancy is present between the surface densities in the northern and southern hemispheres, in particular, at  $2.5 \leq z \leq 4$ . This is mainly due to the strong efforts, mainly by the Sloan Digital Sky Survey (Abolfathi et al. 2018), devoted to the search for bright QSOs in the north. While the number of bright quasars known in the northern hemisphere is in almost all the bins of Table 2 close to the expectations of Table 1, and in general exceeding the more pessimistic models of Kulkarni et al. (2019), the corresponding surface densities for the southern hemisphere turn out to be significantly lower. It is thus apparent that there is a lot of space for finding new bright high- $z$  quasars in the southern hemisphere and the urge for a wide-area survey of this type.

### 3. A New Selection of Bright QSO Candidates

#### 3.1. The Main Sample

In order to select new bright QSO candidates at redshift  $\gtrsim 2.5$  in the southern hemisphere (decl.  $< 0^\circ$ ) we have taken advantage of the following databases:

1. The Skymapper survey (DR1.1, Wolf et al. 2018);
2. The *Gaia* DR2 data release (DR2, Gaia Collaboration 2016, 2018); and
3. The *WISE* survey (Wright et al. 2010).

We considered all sources in the Skymapper survey with the following constraints:

1. Galactic latitude  $|b| > 25^\circ$ ;
2. Magnitude in the  $i$  band fainter than 14 and brighter than 18;
3. Flags in the  $i$  and  $z$  bands equal to zero (i.e., availability of reliable  $i$  and  $z$  magnitudes);
4. Availability of the *Gaia* magnitude in the  $G$  band;
5. Distance to the closest *WISE* source  $< 0''.5$ ;
6. Distance to the closest *Gaia* (DR2) source  $< 0''.5$ ; and
7. Signal-to-noise ratio of the matching *WISE* source in each of the first three bands  $> 3$  (i.e., availability of reliable magnitudes in these bands).

We limit our analysis to the magnitude range  $i = 14\text{--}18$  in order to keep our samples as small as possible. Besides, sources brighter than  $i = 14$  would hardly be high- $z$  QSOs, and sources fainter than  $i > 18$  are not interesting for our purposes. The constraint 3, in particular the request of a reliable  $i$  magnitude,

**Table 2**  
Observed Cumulative Surface Density of QSOs with  $|b| > 25^\circ$  at Different Redshifts and  $i$ -band Magnitudes (with  $i \geq 15$ )

$i \leq$	$\Sigma_{\text{QSO}}(2.5 < z < 3.0)$	$\Sigma_{\text{QSO}}(3.0 < z < 3.5)$	$\Sigma_{\text{QSO}}(3.5 < z < 4.0)$	$\Sigma_{\text{QSO}}(4.0 < z < 4.5)$	$\Sigma_{\text{QSO}}(4.5 < z < 5.0)$
	North–South	North–South	North–South	North–South	North–South
15.5	0.0–0.0	0.0–0.0	0.0–0.0	0.0–0.0	0.0–0.0
16.0	4.3–0.0	0.0–0.0	0.0–0.0	0.0–0.0	0.0–0.0
16.5	7.5–1.6	3.2–0.0	0.0–0.0	0.0–0.0	0.0–0.0
17.0	25.6–11.3	9.6–3.2	0.0–0.0	0.0–0.0	0.0–0.0
17.5	85.4–33.0	42.7–13.7	9.6–2.4	2.1–0.8	0.0–0.8
18.0	324.4–86.9	140.8–41.8	30.9–9.7	8.5–0.8	4.3–3.2

**Note.** In the northern hemisphere only the QSOs in the SDSS footprint have been considered, while in the south the known QSOs before the present survey in the Skymapper footprint from both Pâris et al. (2018) and Véron-Cetty & Véron (2010) have been considered. All numbers are scaled to  $10^4$  sq.deg. to allow a direct comparison with Table 1.

limits the effectiveness of this selection to redshifts  $z \lesssim 5.3$ . For the selection of higher redshift QSOs this constraint has to be relaxed. We discarded the regions of the Large and Small Magellanic Clouds to avoid crowding and extinction. The initial sample (hereafter, *main sample*) contains 1,014,875 objects spanning approximately 12,400 square degrees. Figure 1 shows the sky distribution of the sources in the *main sample*, in both equatorial and Galactic coordinates, as compared with the confirmed QSOs in the SDSS DR14Q catalog (Pâris et al. 2018). When available, we also collected the data in the following bands:

1.  $u$ ,  $v$ ,  $g$ , and  $r$  from Skymapper.
2.  $BP$  and  $RP$  magnitudes from *Gaia*.
3.  $J$ ,  $H$ , and  $K$  from the 2MASS (Skrutskie et al. 2006, requiring a matching distance  $< 1''.5$ ).

The angular distance matching radii for the abovementioned catalogs (and other reference catalogs, see below) have been determined by empirically checking the distributions of the angular separation histograms (see Figure 2) with the aim of achieving the great majority of the true matches while minimizing the number of spurious associations.

### 3.2. Source Classification in the Main Sample

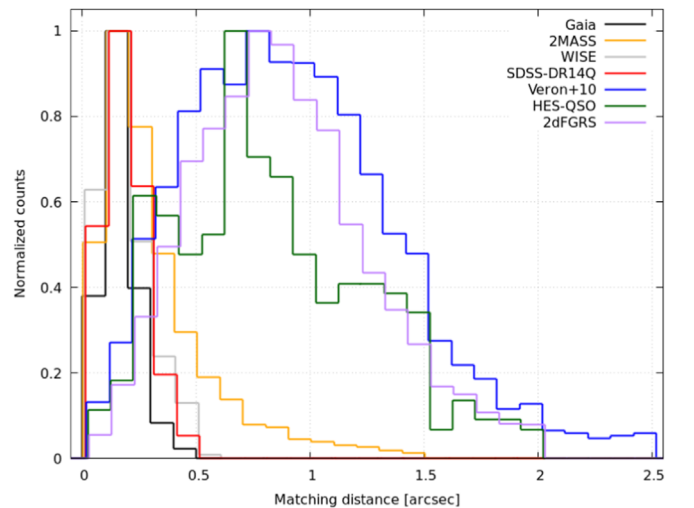
In order to identify stars in the *main sample* we used the following criteria:

1. parallax (as measured by *Gaia*) significantly different from zero ( $> 3\sigma$ );
2. *Gaia* proper motion along R.A. or decl. significantly different from zero ( $> 3\sigma$ ).

844,026 objects out of 1,014,875 (83.2%) meet at least one of the two above criteria and in the following they will be considered as bona fide stars.

In order to identify known QSOs and extragalactic objects in the *main sample* we matched it against the following catalogs:

1. The SDSS DR14Q quasar catalog (526,356 sources, Pâris et al. 2018) finding 1365 matching entries within  $0''.5$ ;
2. The 13th edition of the Veron–Cetty catalog (167,566 sources, Véron-Cetty & Véron 2010), finding 4447 matching entries within  $2''.5$  (only sources with a reliable spectroscopic redshift estimate have been considered);
3. The 2dFGRS catalog (Colless 2001), finding 4032 entries within  $2''$  (only sources with absorption spectra have been considered).



**Figure 2.** Histograms of the angular distances between the Skymapper sources in the *main sample* and the matched sources in the considered catalogs. Each histogram has been normalized by its maximum in order to show all of them in the same plot.

In this way we identified 4666 spectroscopically confirmed QSO/AGN in the redshift range  $0.005 < z < 5.06$ , and 3665 sources with absorption spectra and no significant proper motion or parallax measurement, i.e., nonactive galaxies (mainly at  $z \lesssim 0.5$ ). In total, 852,357 sources (84.0%) in the *main sample* have a reliable object-type identification. The remaining 162,518 sources build up the *unknown* sample.

### 3.3. The QSO Candidate Sample

In order to select new QSO candidates we need a method to identify the QSO characterizing properties among the 162,518 sources in the unknown sample. However, we have no access to their spectra, hence we must search for those properties in the available magnitudes. Historically, optical QSO are selected as point-like sources and disentangled from stars by means of color cuts. The separation can be done in an empirical way (e.g., Richards et al. 2002; Tie et al. 2017; Wolf et al. 2019) or via a machine-learning approach (e.g., Bovy et al. 2012). In NIR, due to the low resolution of the data, the morphology is not used as discriminant and only colors are used, with the net effect of including in the sample also contaminants as low luminosity AGN dominated by the host galaxy (Assef et al. 2013). Our method is similar to the one discussed in Bovy et al. (2012), but in addition to optical data from Skymapper and

*Gaia*, we also considered NIR data from *WISE* and 2MASS. Specifically we identify the color cuts in an automatic fashion using a procedure based on the Canonical Correlation Analysis<sup>11</sup> (CCA; Anderson 1984), rather than using color-color plots to isolate the interesting sources. Our aim is to train the CCA using the object type classification (Section 3.2) as one of the canonical variables. To this purpose we consider all the sources in the *main sample* with a clear object-type identification and attached a numerical label as follows:

1. Label = -1: for the nonactive galaxies;
2. Label = 0: for the stars;
3. Label = 2: for the spectroscopically confirmed QSOs with  $z < 2.5$ ;
4. Label = 3: for the spectroscopically confirmed QSOs with  $z > 2.5$ .

This subset represents our *training sample*, and we use the numerical label as the first canonical variable.<sup>12</sup> The actual value of the numerical labels are rather arbitrary (up to constant scale factors and offsets). We just found a better separation when using a label for the stars sitting in the middle between inactive galaxies and QSO sources. Then we arranged the magnitudes discussed in Section 3.2 in a matrix with as many rows as the number of sources, and as many columns as the available magnitude estimates, and apply the CCA procedure between this matrix and the numerical label discussed above. The output of the CCA procedure is a linear transformation matrix which can be multiplied by the magnitude matrix to obtain a new, one-dimensional coordinate (hereafter named CCA), representing the canonical variable associated to magnitude estimates. The CCA procedure ensures that the CCA coordinate has the highest possible correlation with the numerical label, compatible with the data available in the training set. The CCA coordinate for the sources in the *main sample* is shown in Figure 3 (upper panel) as a function of the  $i$  magnitude. Stars align across a rather narrow horizontal stripe at  $CCA \sim 0$ , while the nonactive galaxies occupy the lower part of the plot. Confirmed low- $z$  ( $z < 2.5$ ) QSOs are spread throughout the whole CCA- $i$  plane, but QSOs with  $z > 2.5$  cluster in the upper right corner, hence we expect new (i.e., not yet identified) QSOs at  $z > 2.5$  to be located in the same region.

Then, we used the same transformation matrix used above to estimate the CCA coordinate for the sources in the unknown sample, obtaining a CCA value representative of the source object types (as was the case for sources in the training set). We started our analysis by considering the sources with a reliable magnitude in all the abovementioned bands ( $u$ ,  $v$ ,  $g$ ,  $r$ ,  $i$ , and  $z$  from Skymapper;  $G$ ,  $BP$ , and  $RP$  from *Gaia*;  $W1$ ,  $W2$ , and  $W3$  from *WISE*; and  $J$ ,  $H$ , and  $K$  from 2MASS), then we proceeded analyzing the sources with incomplete photometric sets (i.e., with some magnitude missing among the 15 listed above). Among the various configurations of bands we treated first the cases with the highest number of sources (allowing us to determine a more robust correlation), then the others with progressively less sources. In each iteration we performed a linear fit against the original numerical label in order to renormalize the CCA coordinates and span always the same

dynamical range for each configuration of photometric bands. In this way we have been able to compute the CCA coordinates consistently for all the sources in the *main sample*.

Figure 3 (middle panel, gray symbols) shows the location of the 162,518 sources (16%) of the unknown sample remaining after the removal of the objects with a known object-type identification (stars, galaxies, and QSOs). The  $z > 2.5$  QSOs we are looking for are expected to lie in the region at  $CCA \gtrsim 1$ , but they are still confused in an overwhelming cloud of extended, inactive galaxies and low- $z$  AGN (shown in the upper panel of Figure 3 with black “x” and purple “+” symbols, respectively). It is therefore necessary to further distill our  $z > 2.5$  candidates by selecting against extended objects and low- $z$  sources.

### 3.3.1. Excluding Extended Objects

Since we are looking for bright high- $z$  QSOs we expect them to have a point-like appearance. In order to test whether an object in the unknown sample is spatially extended we have taken advantage of the comparison between the PSF and Petrosian magnitudes reported in the Skymapper catalog. The latter are supposed to be similar to the former only for point-like sources, while capturing more flux with respect to the PSF magnitudes for extended sources. In order to quantify such a difference we divided the whole *main sample* into bins of 0.1 mag and adopted the median of the differences between the PSF and the Petrosian magnitude as a reference value within each bin. Then, for each source, we interpolated the reference values corresponding to its PSF mags and computed the quantity:

$$\sigma_{x,\text{extd}} = \frac{1}{2} \sum_{x=i,z} \frac{x_{\text{psf}} - x_{\text{petro}} - \langle x_{\text{psf}} - x_{\text{petro}} \rangle_{\text{ref}}}{\sqrt{\sigma_{x_{\text{psf}}}^2 + \sigma_{x_{\text{petro}}}^2}}, \quad (1)$$

where the  $x$  represents either considered band, and  $\sigma_x$  is the associated uncertainty.<sup>13</sup> We repeated the procedure in both the  $i$  and the  $z$  bands and considered the average  $\sigma_{\text{extd}} = (\sigma_{i,\text{extd}} + \sigma_{z,\text{extd}})/2$  as an estimate of the significance of the object being extended. The histogram of the values of  $\sigma_{\text{extd}}$  for the objects in the *main sample* with an available object-type identification is shown in Figure 4. Almost all confirmed QSOs with  $z > 2.5$  have  $\sigma_{\text{extd}} < 3$ , hence we considered this value as a threshold to distinguish point-like sources from extended sources. In this way we discarded 135,238 bona fide extended sources from the 162,518 objects of the unknown sample (81%).

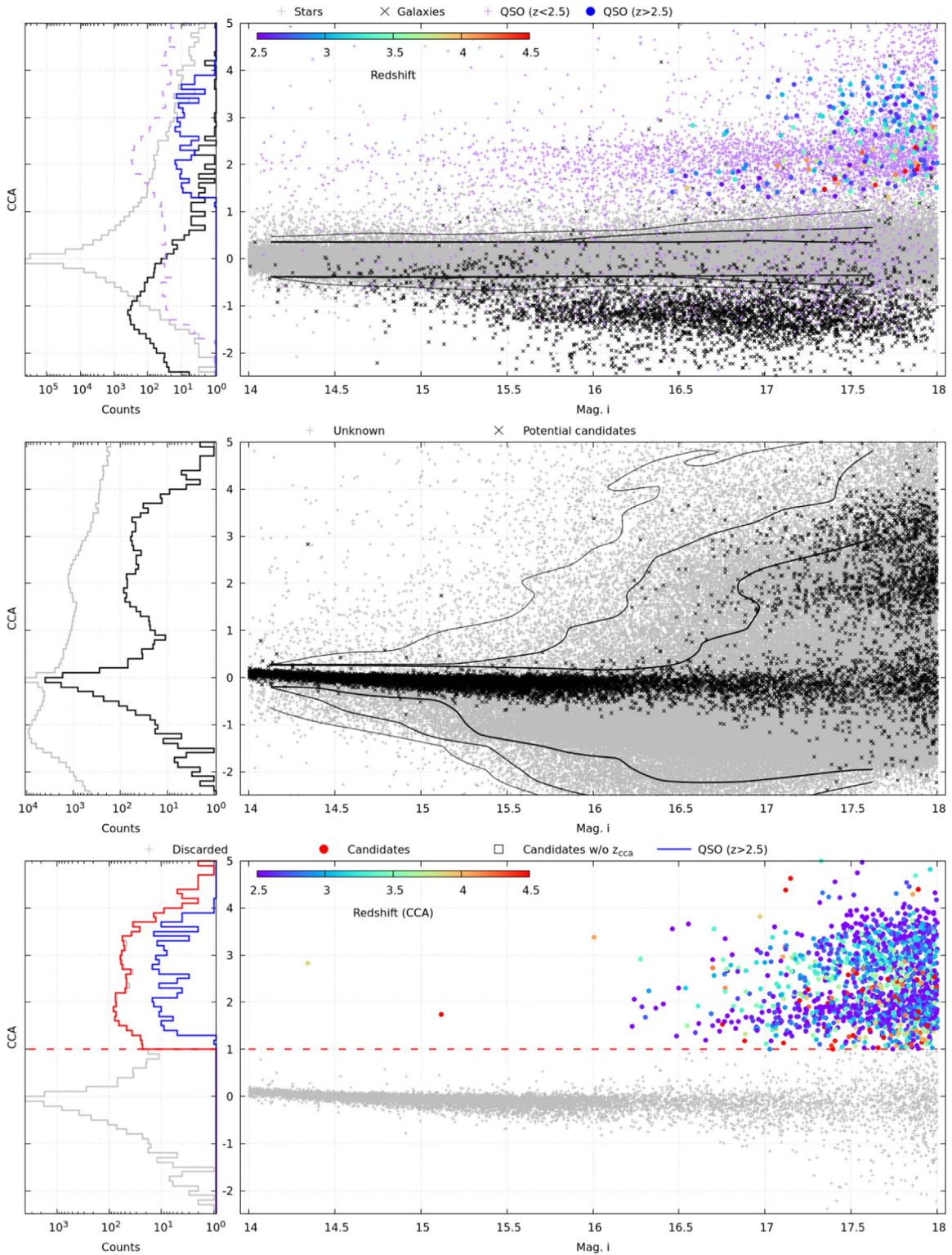
### 3.3.2. Excluding Probable Low- $z$ ( $z < 2.5$ ) Sources

To estimate the redshift of the sources in the *main sample* we used again a CCA transformation, this time using the spectroscopic redshifts of the subsample of confirmed (Section 1) QSOs as a training set, and following the same procedure we used to calculate the CCA coordinate. The comparison between  $z_{\text{cca}}$  and  $z_{\text{spec}}$  for the confirmed QSOs with  $\sigma_{\text{extd}} < 3$  is shown in Figure 5 (upper panel). The scatter in the  $z_{\text{cca}}$  estimates is  $\sim 0.36$ . Then we estimated the CCA redshift (hereafter  $z_{\text{cca}}$ ) using the resulting transformation matrix for the

<sup>11</sup> [https://en.wikipedia.org/wiki/Canonical\\_correlation](https://en.wikipedia.org/wiki/Canonical_correlation)

<sup>12</sup> Canonical variables are obtained from input variables by means of a linear transformation. Since the numerical label is one-dimensional it is by definition proportional to a canonical variable.

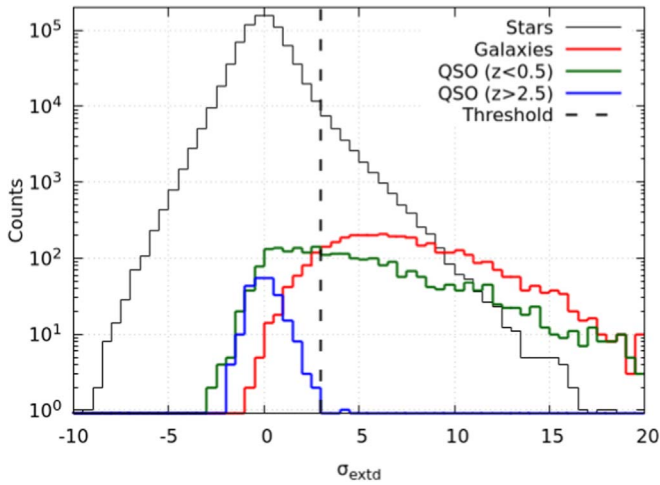
<sup>13</sup> Note that in this equation  $z$  refers to the magnitude in the  $z$  band, not to the redshift.



**Figure 3.** CCA- $i$  mag plane for the subsamples considered in this work. Upper panel: sources in the *main sample* for which a reliable type identification is available (Section 3.3). Stars are identified by gray “+” symbols, inactive galaxies by black cross symbols, low- $z$  ( $<2.5$ ) QSOs with purple “+” symbols, and high- $z$  ( $>2.5$ ) QSOs with filled circles. The redshift for the confirmed QSOs with  $z_{\text{spec}} > 2.5$  are shown with the color code shown in the colorbox 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 high- $z$  QSOs (blue). Middle panel: sources in the *main sample* without an object type identification (gray symbols, Section 3.3). The same sources after excluding extended (Section 3.3.1) and low- $z$  objects (Section 3.3.2) are highlighted in black, and represent potential high- $z$  QSO candidates. Lower panel: the final sample of high- $z$  QSO candidates, with the redshift  $z_{\text{cca}}$  estimated using the procedure described in Section 3.3.2.

whole unknown sample. To distinguish a low- $z$  ( $z < 2.5$ ) from an high- $z$  ( $z > 2.5$ ) source we calculated the following quantities:

1. Low- $z$  “contamination”: ratio of the number of low- $z$  sources over the number of sources with  $z_{\text{cca}}$  above a given threshold;



**Figure 4.** Histogram of the  $\sigma_{\text{extd}}$  quantity (Equation (1)) for all the sources in the *main sample*. The threshold at  $\sigma_{\text{extd}} = 3.0$  is shown with a vertical dashed line. Sources above this threshold are assumed to be spatially extended and discarded.

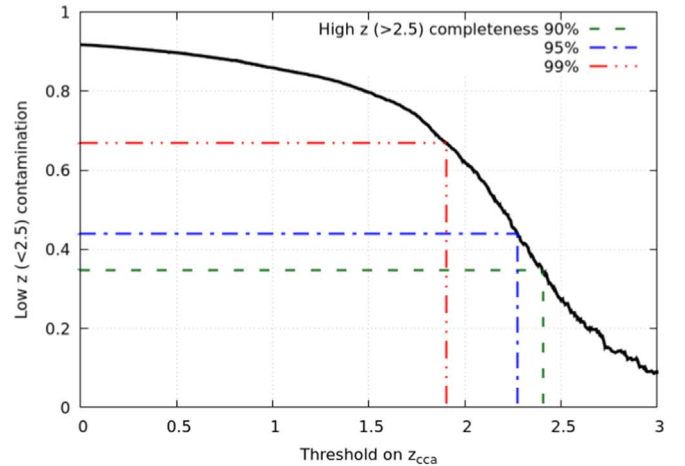
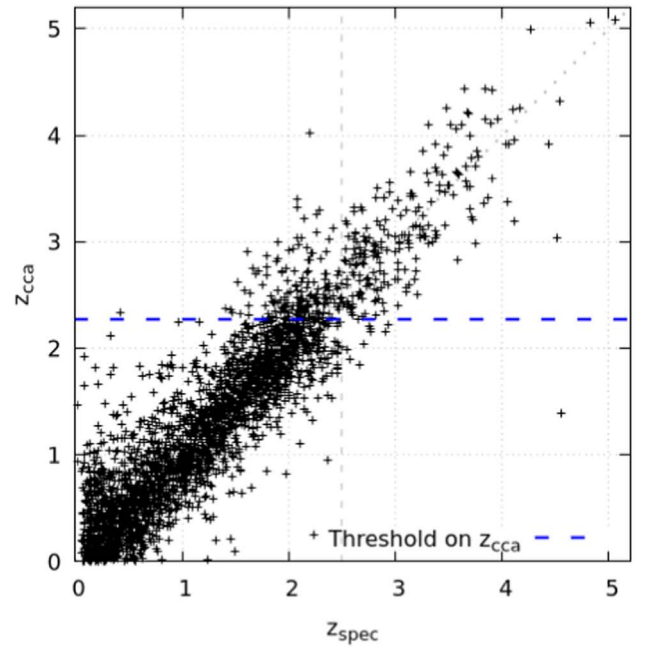
2. *High- $z$  “completeness”*: ratio of high- $z$  sources with  $z_{\text{cca}}$  above a given threshold, over the total number of high- $z$  confirmed QSOs.

A plot of these quantities, for all the possible values of the  $z_{\text{cca}}$  threshold, is shown in Figure 5 (lower panel, black solid line): a high- $z$  “completeness” of 95% can be reached with a threshold at  $z_{\text{cca}} = 2.27$ , corresponding to a low- $z$  “contamination” of 44% (dotted–dashed blue line). Increasing the high- $z$  “completeness” to 99% (blue dotted–dotted–dashed) would yield a much higher contamination, while decreasing to 90% (green dashed line) would yield only a small improvement in contamination. Hence, we chose  $z_{\text{cca}} = 2.27$  as a discriminating threshold to select against low- $z$  QSO candidates.

### 3.3.3. The Final QSO Candidate Sample

By discarding the extended and low- $z$  sources from the 162,518 objects of the unknown sample, we are left with 11,598 potential QSO candidates (black cross symbols in the middle panel of Figure 3). Besides excluding a significant fraction of sources in the unknown sample (93%), the above procedure allowed us to obtain a better separation of the remaining sources in the  $\text{CCA}-i$  mag plane, resulting in an increased contrast between the peaks above and below  $\text{CCA} \sim 1$  in the histogram on the left of the middle panel in Figure 3, and suggesting that a threshold on the CCA value might allow us to exclude the non-QSO sources. As discussed above, the group at  $\text{CCA} > 1$  is likely associated with high- $z$  QSOs, while the group at  $\text{CCA} < 1$  is associated with stars and inactive galaxies. Therefore we discarded all the source with  $\text{CCA} < 1$  to obtain a final sample of 1476 high- $z$  QSO candidates. The lower panel of Figure 3 shows the location of such candidates in the  $\text{CCA}-i$  mag plane, and their expected redshift (color-coded, as calculated in Section 3.3.2). As a consistency check, we note that all the known QSOs with  $z > 2.5$  (blue line in the histogram, both in upper and lower panel of Figure 3) lie above the adopted CCA threshold, as expected.

The number of DR14Q and Veron sources with  $z > 2.5$  in our *main sample* is 68. Considering that the Skymapper footprint is 8.3 times larger than the previously surveyed area,



**Figure 5.** Upper panel: the  $z_{\text{cca}}-z_{\text{spec}}$  correlation (scatter:  $\sim 0.36$ ). Lower panel: low- $z$  ( $< 2.5$ ) “contamination” as a function of the adopted  $z_{\text{cca}}$  threshold and of the high- $z$  ( $> 2.5$ ) “completeness” (dashed, dotted–dashed, and dotted–dotted–dashed lines). We chose a  $z_{\text{cca}}$  threshold of 2.27 corresponding to a high- $z$  completeness of 95% and an expected low- $z$  contamination in our final sample of  $\sim 44\%$ .

we extrapolate  $\sim 564$  new QSOs with  $z > 2.5$  in our QSO candidate sample. Given the size of the QSO candidate sample (1476 sources) we expect a lower limit for the success rate for high- $z$  ( $> 2.5$ ) QSO identification of  $\sim 40\%$ . Actually, the fraction of new high- $z$  ( $> 2.5$ ) QSO spectroscopically confirmed among the candidates we could observe (Section 4) is  $\sim 80\%$ .

At this stage we can also compute the fraction of DR14Q and Veron QSOs with  $z > 2.5$  and  $i < 18$  satisfying all the conditions to be selected by our procedure, 93%, and use it as an indication of the completeness of our QSO sample. The 7% of the known QSOs lost were sources with a predicted CCA redshift below our threshold of  $z_{\text{cca}} = 2.27$  (Section 3.3.2).

## 4. Spectroscopic Confirmations

In order to validate the above-described selection criteria (and test variants), we have carried out extensive spectroscopy

**Table 3**  
List of Skymapper Sources in the *Candidate* Sample Observed in Our Campaigns

Skymapper ID	R.A. (J2000)	Decl. (J2000)	Date Obs.	$m_i$	Obj. Type	$z_{\text{spec}}$	Instrument	Notes
7683342	00:35:58.10	-20:05:56.25	2018 Nov 22	16.842	QSO	1.53	LDSS-3	
6645651	01:13:31.80	-34:24:49.77	2018 Oct 24	17.155	QSO	3.41	IMACS	
314938059	01:56:44.67	-69:22:16.12	2018 Nov 23	16.325	QSO	2.8	LDSS-3	
8932486	02:11:13.61	-11:55:43.65	2018 Sep 23	16.801	QSO	2.82	LDSS-3	
12503126	03:22:54.38	-11:19:39.58	2018 Sep 25	16.698	QSO	2.67	LDSS-3	
316834279	03:30:15.32	-54:30:21.15	2018 Oct 24	17.174	QSO	3.4	IMACS	
10730355	04:05:16.06	-32:51:16.77	2019 Feb 03	17.64	QSO	3.33	WFCCD	
10739949	04:07:45.29	-32:15:37.84	2019 Feb 03	17.693	QSO	3.75	WFCCD	
11621465	04:18:40.65	-27:40:28.51	2018 Sep 25	16.806	QSO	2.81	LDSS-3	
10934139	04:50:11.37	-43:24:29.75	2019 Feb 20	17.798	QSO	3.95	LDSS-3	
56293062	09:07:29.33	-02:14:58.97	2019 Apr 30	17.845	QSO	2.85	NTT	
56219717	09:26:44.48	-05:57:00.50	2019 Apr 30	17.46	QSO	2.89	NTT	
55939586	10:12:23.62	-17:19:25.21	2019 Feb 20	17.628	QSO	3.08	LDSS-3	
58206167 <sup>a</sup>	10:14:30.28	-04:21:40.31	2019 Apr 30	17.571	QSO	3.9	NTT/WFCCD	
57929040 <sup>a</sup>	10:15:29.37	-12:13:14.23	2019 Feb 02	17.255	QSO	4.19	WFCCD	
57933437 <sup>a</sup>	10:15:44.12	-11:09:22.80	2019 Apr 30	17.485	QSO	3.87	NTT/WFCCD	
58243698	10:33:48.52	-04:07:42.02	2019 Apr 30	17.012	QSO	1.31	NTT	
54139102	10:47:20.31	-29:59:17.62	2019 Feb 02	17.779	QSO	3.37	WFCCD	
58181076 <sup>a</sup>	10:51:22.70	-06:50:47.82	2019 Feb 02	17.345	QSO	3.81	WFCCD	
57143774	10:52:21.62	-19:52:37.95	2019 Feb 03	17.741	QSO	3.66	WFCCD	
57368436 <sup>a</sup>	10:54:49.69	-17:11:07.36	2019 Apr 30	17.107	QSO	3.75	NTT	
58672537	11:08:35.49	-04:06:44.59	2019 Apr 30	17.216	QSO	3.05	NTT	
54677478	11:14:16.38	-30:51:49.73	2019 Feb 02	17.554	QSO	3.38	WFCCD	
58420716	11:15:22.55	-13:08:09.81	2019 Apr 30	17.222	QSO	2.53	NTT	
58535642	11:41:10.85	-12:06:20.67	2019 Feb 20	17.759	QSO	3.14	LDSS-3	BAL
57627257	11:43:26.60	-24:38:52.65	2019 Apr 30	17.551	QSO	2.98	NTT	
58900019	11:48:22.21	-01:17:29.14	2019 Apr 30	17.578	QSO	2.71	NTT	BAL
57913424 <sup>a</sup>	11:49:14.40	-15:30:43.97	2019 Apr 30	17.756	QSO	4.13	NTT/WFCCD	
60954990	12:01:50.81	-29:09:31.84	2019 Feb 03	17.64	QSO	3.34	WFCCD	
64101880	12:36:20.94	-24:47:48.21	2019 Feb 03	17.754	QSO	3.38	WFCCD	
60762547	12:45:55.02	-34:59:52.08	2019 Feb 03	17.541	QSO	3.14	WFCCD	BAL
64820204	13:19:12.24	-16:23:53.85	2019 Apr 30	16.498	QSO	1.7	NTT	
65645790	13:21:55.88	-11:34:09.38	2019 Apr 30	16.952	QSO	1.9	NTT	
67301709 <sup>a</sup>	14:00:15.16	-03:44:16.48	2019 Apr 30	17.711	QSO	3.54	NTT	
66670919 <sup>a</sup>	14:27:32.25	-18:03:18.26	2019 Apr 30	17.093	QSO	2.36	NTT	BAL
63745172	14:34:52.96	-23:58:54.15	2019 Apr 30	16.994	QSO	2.46	NTT	
66962683	14:49:43.17	-12:27:17.51	2019 Apr 30	16.703	QSO	3.27	NTT	
72002629	15:05:27.83	-20:45:34.95	2019 Apr 30	16.955	QSO	3.09	NTT	
302866544	19:18:57.68	-65:44:52.38	2019 Apr 30	17.848	QSO	3.84	NTT	
299711220	19:22:27.53	-60:42:14.75	2019 Apr 30	17.435	QSO	3.19	NTT	
135100950	19:53:02.67	-38:15:48.40	2019 Apr 30	17.305	QSO	3.71	NTT	
304683049	20:08:20.02	-61:29:22.66	2019 Apr 30	17.534	QSO	2.51	NTT	
136198132 <sup>a</sup>	20:17:41.49	-28:16:29.83	2019 Apr 30	17.388	QSO	3.69	NTT	
171556238 <sup>a</sup>	20:18:34.86	-15:28:38.61	2019 Apr 30	17.427	QSO	3.6	NTT	
136776832 <sup>a</sup>	20:33:43.57	-30:23:09.74	2019 Apr 30	17.909	QSO	3.51	NTT	
301689901	20:34:25.55	-54:54:11.69	2019 Apr 30	17.132	QSO	2.99	NTT	BAL
306808859	20:55:25.69	-51:45:53.43	2018 Jul 26	16.93	QSO	2.53	LDSS-3	
3430280	21:08:46.96	-02:01:14.81	2019 Apr 30	17.645	QSO	3.48	NTT	
3436512 <sup>b</sup>	21:11:05.60	-01:56:04.14	2019 Jun 19	17.907	QSO	4.89	LDSS-3	
305297978	21:11:34.23	-58:19:27.34	2019 Apr 30	17.467	QSO	2.07	NTT	BAL
304245360 <sup>c</sup>	21:19:20.85	-77:22:53.17	2019 Jun 19	17.86	QSO	4.56	LDSS-3	
175107	21:24:34.76	-33:52:26.39	2019 Apr 30	17.192	QSO	1.94	NTT	
2379862 <sup>a</sup>	21:25:40.96	-17:19:51.32	2019 Apr 30	16.548	QSO	3.9	NTT	
307536968	21:52:28.22	-44:46:03.91	2018 Oct 24	17.291	QSO	3.47	IMACS	
304697	21:54:06.10	-38:06:24.11	2019 Apr 30	17.275	QSO	2.38	NTT	
882295	21:57:51.00	-29:48:31.25	2019 Apr 30	17.291	QSO	2.51	NTT	
1710729	22:22:36.79	-28:27:50.71	2018 Oct 24	17.372	QSO	2.56	IMACS	Radio
1141079	22:27:01.38	-39:33:57.73	2019 Apr 30	17.774	QSO	3.5	NTT	
5133592	22:40:01.08	-12:25:09.12	2019 Apr 30	17.307	QSO	2.95	NTT	
5412867	22:42:21.48	-03:54:58.38	2019 Apr 30	17.472	QSO	3.19	NTT	
308808008	22:44:24.83	-45:15:54.94	2019 Apr 30	17.159	QSO	1.6	NTT	
4322023	22:48:21.47	-21:25:53.82	2018 Sep 24	16.595	QSO	2.88	LDSS-3	
308423842	22:53:53.21	-55:20:00.47	2019 May 13	17.206	QSO	3.29	LDSS-3	BAL
308846570	22:55:34.56	-44:26:02.00	2019 May 13	17.344	STAR	...	LDSS-3	

**Table 3**  
(Continued)

Skymapper ID	R.A. (J2000)	Decl. (J2000)	Date Obs.	$m_i$	Obj. Type	$z_{\text{spec}}$	Instrument	Notes
5161954 <sup>a</sup>	22:56:33.16	-12:43:59.57	2019 May 13	17.456	QSO	2.83	LDSS-3	
4885680	23:28:45.81	-15:44:39.95	2018 Jul 29	16.782	QSO	2.49	LDSS-3	

**Notes.**<sup>a</sup> Also observed by Schindler et al. (2019).<sup>b</sup> Also reported by Wang et al. (2016).<sup>c</sup> Also observed by Wolf et al. (2019).

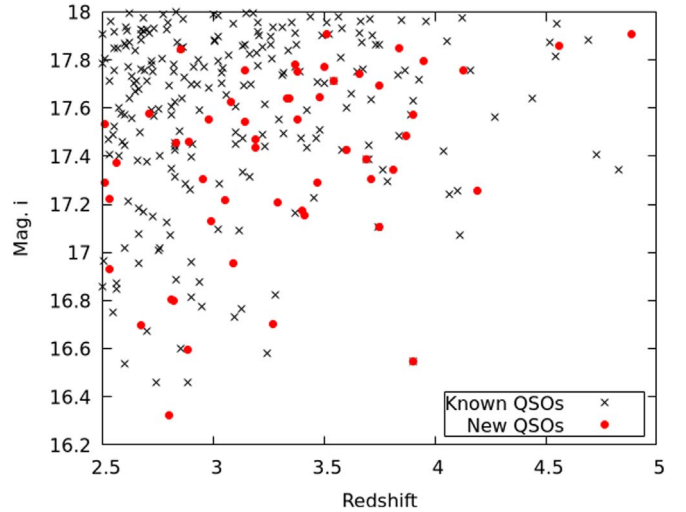
observations at the Las Campanas Observatory and at the ESO-NTT telescope at La Silla. The results of these observations build up the QUBRICS survey.

The first pilot study was carried out at the *Magellan* telescopes in 2018 using LDSS-3 (Clay Telescope) and IMACS (Baade Telescope). Observations were obtained over various nights during bright time and variable weather conditions. With LDSS-3 the VPH-all grism has been used with the 1''-central slit and no blocking filter, covering a wavelength range between 4000 and 10000 Å with a low resolution of  $R \sim 800$ . With IMACS, we used the #300 grism with a blaze angle of  $17^\circ.5$ , covering a wavelength range between 4000 and 10000 Å with a dispersion of 1.34 Å/pixel. Based on these first results, the selection technique has been adjusted in order to include candidates at higher redshift. In 2019 February we were awarded 2 nights at the du Pont telescope to validate the optimized criteria, and we observed several new candidates with the Wide Field CCD (WFCCD) blue grism that covers a wavelength range between 3700 and 8000 Å providing a 2 Å/pixel dispersion.

The NTT spectroscopic campaign has been carried out during the ESO observing period P103 under the proposal 0103.A-0746 (PI. A. Grazian). Three nights of spectroscopy have been executed during 2019 April 27–30. The EFOSC2 instrument was used, equipped with the grism #13 (wavelength range  $\lambda \sim 3700\text{--}9300$  Å). Since our main targets are relatively sparse in the sky, we carried out long-slit spectroscopy with exposure times between 3 and 7 minutes per object.

Finally, in 2019 June we performed a few exposures at TNG (La Palma) using the Low Resolution Spectrograph (Dolores) with the LR-B grism (resolution  $\sim 600$ ), a 1'' slit aperture and an exposure time of 10 minutes per object, in order to validate our selection criteria against low- $z$  AGNs (Section 3.3.2).

In total we observed 69 sources from our final QSO candidate sample (Section 3.3.3) of 1476 sources. Having many more candidates than could be observed with the assigned telescope time, we decided to start with the most promising objects, i.e., the brighter ones having higher  $z_{\text{cca}}$  values (Section 3.3.2). Among these sources, 54 turned out to be genuine high- $z$  QSOs with  $z > 2.5$ , 12 are low- $z$  QSOs with  $z < 2.5$ , and 1 is a star. The details of the candidate observations are summarized in Table 3, while Figure 6 shows the redshift- $i$  magnitude plane of the newly discovered QSOs (red circles) and the known QSOs before this work (black cross symbols). So far, we achieved a success rate in identifying new high- $z$  QSOs sources of  $\sim 80\%$ . On the other hand, we observed only a small fraction of the total QSO candidate sample (69/1476 sources,  $\sim 5\%$ ), hence the success rate may be biased by the choice of the most promising candidates for the observations.



**Figure 6.** 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 the present survey.

In the early phases of the project we experimented with different versions of the selection algorithm and tested its limits and characteristics with the pilot spectroscopic runs and in part of the NTT run. As a consequence, we also observed sources that do not belong to the final sample. For completeness, we report the observation details for these 67 additional sources in Table 4. Among them we found 2 QSOs at  $z > 2.5$ , 50 low- $z$  QSOs, and 15 non-QSO sources. The two QSOs at  $z > 2.5$  were not selected in the *main sample* because one has an  $i$  magnitude fainter than the threshold of  $i = 18$  and the other has a Skymapper astrometric position differing by more than  $0''.5$  from *Gaia* DR2, probably due to image defects in the Skymapper data, as we checked with Skymapper cutouts.

Further observing runs at the du Pont and NTT telescopes have been approved in order to expand our spectroscopically observed sample. All the details and results of the spectroscopic runs will be described in a future paper.

## 5. Conclusions

The aim of the present project was to identify new, bright ( $i < 18$ ) QSOs at relatively high redshift ( $z > 2.5$ ) in the southern hemisphere with a high success rate (QUBRICS survey). At this stage completeness represented a secondary requirement.

Finding in an efficient way relatively high-redshift QSOs is a kind of needle in a haystack task. Our approach has been to take advantage of large high-quality photometric and

**Table 4**  
List of Skymapper Sources Observed in our Campaigns That Are Not in the Final Candidate sample

Skymapper ID	R.A. (J2000)	Decl. (J2000)	Date Obs.	$m_i$	Obj. Type	$z_{\text{spec}}$	Instrument	Notes
9163238	02:00:16.25	-06:52:09.06	2018 Nov 23	16.688	QSO	1.75	LDSS-3	
7120598	02:36:58.66	-31:35:22.42	2018 Sep 25	16.652	QSO	1.46	LDSS-3	
317598139	02:51:11.14	-54:49:43.26	2018 Oct 24	16.827	QSO	1.03	IMACS	
10841553 <sup>a</sup>	04:09:14.88	-27:56:32.4	2019 Feb 02	18.031	QSO	4.48	WFCCD	
56477133	09:41:19.56	-08:43:39.01	2019 Apr 30	16.392	AGN	0.09	NTT	
56513662	09:42:47.54	-06:11:10.16	2019 Apr 30	16.558	QSO	2.04	NTT	
55717293	09:59:12.23	-18:52:49.99	2019 Apr 30	15.366	QSO	0.57	NTT	
51720233	10:08:12.03	-24:33:23.83	2019 Apr 30	16.22	GAL	0.0064	NTT	
55664037	10:09:15.28	-22:59:13.70	2019 Apr 30	17.118	AGN	0.08	NTT	
58024847	10:26:03.57	-09:32:29.82	2019 Apr 30	16.743	AGN	0.05	NTT	
57381969	10:39:06.37	-15:56:34.48	2019 Apr 30	17.092	QSO	1.74	NTT	
58356880	10:41:19.41	-03:14:25.34	2019 Apr 30	17.286	AGN	0.13	NTT	
57329165	10:42:51.27	-16:48:57.90	2019 Apr 30	16.836	AGN	0.13	NTT	
57145002	10:56:35.86	-20:29:03.20	2019 Apr 30	16.582	AGN	0.14	NTT	
57810566	11:04:43.01	-13:53:54.03	2019 Apr 30	16.705	AGN	0.07	NTT	
58426461	11:16:44.93	-12:16:30.39	2019 Apr 30	16.963	STAR	...	NTT	
68253481	11:20:03.66	+08:12:11.88	2019 Apr 30	16.485	STAR	...	NTT	
58445811	11:21:01.41	-11:22:00.91	2019 Apr 30	16.296	AGN	0.07	NTT	
58533327	11:35:24.00	-11:02:29.08	2019 Apr 30	16.402	AGN	0.08	NTT	
58814432	11:50:00.35	-07:02:13.03	2019 Apr 30	16.753	AGN	0.08	NTT	
58838917	11:53:22.58	-04:48:37.82	2019 Feb 03	17.641	QSO	1.84	WFCCD	
65022181	12:09:43.72	-13:29:14.58	2019 Jun 13	16.713	GAL	...	TNG-Dolores	
64252467	12:11:18.10	-18:31:28.93	2019 Apr 30	14.841	STAR	...	NTT	
60941522	12:20:16.49	-29:32:00.52	2019 Apr 30	17.162	STAR	...	NTT	
68403402	12:21:14.26	+01:37:56.07	2019 Apr 30	15.556	STAR	...	NTT	
65139168	12:24:49.81	-08:44:14.70	2019 Mar 29	17.717	QSO	0.59	LDSS-3	
60950802	12:25:15.83	-29:47:59.87	2019 Apr 30	15.881	STAR	...	NTT	
64067366	12:28:04.10	-21:35:47.63	2019 Apr 30	16.468	AGN	0.04	NTT	
65128034	12:28:10.15	-09:27:03.34	2019 Apr 30	15.532	AGN	0.22	NTT	
60788872	12:33:44.43	-33:52:20.90	2019 Apr 30	16.297	AGN	0.07	NTT	
61046630	12:41:31.95	-30:41:40.08	2019 Apr 30	15.546	QSO	0.77	NTT	
64413049	12:54:50.39	-19:33:02.46	2019 Apr 30	16.905	QSO	1.14	NTT	
65863636	13:02:49.35	-06:07:16.07	2019 Apr 30	16.241	AGN	0.11	NTT	
65800341	13:29:50.06	-09:17:08.02	2019 Apr 30	15.727	QSO	1.44	NTT	
65803041	13:32:11.17	-09:00:55.11	2019 Jun 13	16.992	AGN	0.16	TNG-Dolores	NLS1
64677303	13:36:24.43	-24:19:24.70	2019 Apr 30	16.746	QSO	1.71	NTT	
61738231	13:36:28.60	-25:58:24.26	2019 Apr 30	16.463	QSO	1.8	NTT	
66104063	13:43:52.42	-01:59:26.43	2019 Apr 30	15.843	STAR	...	NTT	
61742143	13:44:11.04	-26:31:15.93	2019 Apr 30	16.61	QSO	1.56	NTT	
67082505	13:48:53.28	-07:51:43.28	2019 Apr 30	16.687	AGN	0.09	NTT	
67273801	13:52:22.97	-04:15:57.23	2019 Apr 30	16.177	AGN	0.11	NTT	
63286507	13:52:59.72	-27:33:28.32	2019 Apr 30	16.958	AGN	0.12	NTT	
66377573	13:56:52.65	-17:23:52.73	2019 Apr 30	16.685	STAR	...	NTT	
66332887	14:17:28.12	-17:19:52.41	2019 Apr 30	15.625	QSO	1.8	NTT	
66545115	14:20:15.38	-15:20:38.17	2019 Apr 30	16.609	QSO	2.34	NTT	
63206887	14:27:27.91	-29:35:51.21	2019 Apr 30	16.178	AGN	0.08	NTT	
66714186	14:41:05.85	-16:16:18.93	2019 Apr 30	17.624	QSO	1.76	NTT	
63772716	14:41:15.95	-23:13:30.36	2019 Apr 30	17.137	QSO	1.51	NTT	
99947747	15:13:33.61	-06:34:42.46	2019 Apr 30	16.27	AGN	0.22	NTT	
100081484	15:25:44.13	-05:55:50.58	2019 Apr 30	15.976	QSO	1.69	NTT	
114125048	16:24:51.85	-04:03:28.83	2019 Jun 13	16.831	AGN	0.13	TNG-Dolores	NLS1
113359026	16:29:08.15	-06:37:13.11	2019 Jun 13	16.93	AGN	0.18	TNG-Dolores	
135014567	19:55:03.67	-40:50:52.69	2019 Apr 30	17.708	STAR	...	NTT	
303250705	20:10:48.07	-72:05:16.53	2019 Apr 30	17.269	AGN	0.2	NTT	
304279068	20:51:40.79	-73:49:58.07	2019 Apr 30	17.59	QSO	1.62	NTT	
306855864	20:53:19.23	-50:34:16.35	2018 Jul 26	16.38	???	...	LDSS-3	
3387293	21:02:23.12	-02:26:58.95	2019 Apr 30	17.474	AGN	0.18	NTT	
305521139	21:28:30.32	-70:41:30.18	2019 Apr 30	17.84	QSO	1.92	NTT	
3294242	21:28:49.64	-07:43:28.32	2019 Apr 30	16.96	STAR	...	NTT	
303862 <sup>b</sup>	21:54:45.04	-38:26:32.5	2018 Sep 24	16.583	QSO	2.81	LDSS-3	Radio
2816684	22:00:35.05	-18:13:07.74	2019 Jun 18	17.188	STAR	...	LDSS-3	
305491561	22:02:33.86	-60:20:34.01	2019 Apr 30	15.197	AGN	0.1	NTT	
306001528	22:53:03.71	-65:02:13.75	2019 Apr 30	16.908	STAR	...	NTT	
5799364	23:24:21.06	-05:55:51.47	2018 Jul 29	16.611	QSO	1.55	LDSS-3	

**Table 4**  
(Continued)

Skymapper ID	R.A. (J2000)	Decl. (J2000)	Date Obs.	$m_i$	Obj. Type	$z_{\text{spec}}$	Instrument	Notes
5800558	23:25:39.32	-06:53:00.23	2018 Jul 29	16.142	QSO	1.52	LDSS-3	
305847011	23:44:57.67	-72:03:04.68	2018 Jul 26	15.681	QSO	1.56	LDSS-3	
4978910	23:49:06.84	-17:08:43.22	2018 Oct 25	16.851	QSO	1.29	IMACS	

#### Notes.

<sup>a</sup> Not included in the *main sample* since the Skymapper  $i$  magnitude is fainter than  $i = 18$ . Also observed by Schindler et al. (2019) and Wolf et al. (2019).

<sup>b</sup> Not included in the *main sample* since the Skymapper astrometric position differs more than  $0''.5$  from *Gaia* DR2 (see the text).

astrometric databases provided by Skymapper, *WISE*, 2MASS, and *Gaia*, in order to remove sources identified with high reliability as contaminants (stars, low- $z$  QSOs, and galaxies). Then, with the help of a Canonical Correlation Analysis (Anderson 1984) we have selected among the remaining unknown objects a sample of 1476  $z > 2.5$  QSO candidates, whose completeness is also expected to be high ( $\gtrsim 90\%$  for objects up to  $z \sim 5$ ), estimated on the basis of the number of known QSOs that the method would select.

Indeed, the first campaigns of spectroscopic confirmations have been characterized by a high success rate ( $\sim 81\%$ ), and already at this preliminary stage the number of bright QSOs in the southern hemisphere has been significantly increased, as shown in Figure 6. The new 54 (Table 3) QSOs with  $z > 2.5$  and  $i < 18$ , plus 2 (Table 4), are now available to the astronomical community for high-resolution spectroscopic follow-up and the studies of cosmology and fundamental physics described in the Introduction.



We are continuing our campaigns of spectroscopic confirmations and at the same time we are exploring other statistical techniques in addition to the CCA analysis to further improve the properties of the selection and extend its range of applicability.


We thank Luca Pasquini and Carlos Martins for enlightening discussions. This work is based on data products from observations made with ESO Telescopes at La Silla Paranal Observatory under ESO programme ID 103.A-0746(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://www.cosmos.esa.int/gaia>), processed by the *Gaia* Data Processing

and Analysis Consortium (DPAC, <https://www.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 the Two Micron All Sky Survey, 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 the *Wide-field Infrared Survey Explorer*, which is a joint project of the University of California, Los Angeles, and the Jet Propulsion Laboratory/California Institute of Technology, funded by the National Aeronautics and Space Administration. This paper includes data gathered with the 6.5 meter *Magellan* Telescopes located at Las Campanas Observatory, Chile. We thank Società Astronomica Italiana (SAIt), Ennio Poretti, Gloria Andreuzzi, Marco Pedani, Vittoria Altomonte and Andrea Cama for the observation support at TNG. Part of the observations discussed in this work are based on observations made with the Italian Telescopio Nazionale Galileo (TNG) operated on the island of La Palma by the Fundación Galileo Galilei of the INAF (Istituto Nazionale di Astrofisica) at the Spanish Observatorio del Roque de los Muchachos of the Instituto de Astrofisica de Canarias.

*Facilities:* Skymapper, *Wise*, 2MASS, *Gaia*, *Magellan*: Baade (IMACS), *Magellan*:Clay (LDSS-3), du Pont (WFCCD), TNG (Dolores).

#### ORCID iDs

Giorgio Calderone  <https://orcid.org/0000-0002-7738-5389>  
Konstantina Boutsia  <https://orcid.org/0000-0003-4432-5037>

Stefano Cristiani  <https://orcid.org/0000-0002-2115-5234>  
Andrea Grazian  <https://orcid.org/0000-0002-5688-0663>  
Ricardo Amorin  <https://orcid.org/0000-0001-5758-1000>  
Valentina D'Odorico  <https://orcid.org/0000-0003-3693-3091>

Guido Cupani  <https://orcid.org/0000-0002-6830-9093>  
Fabio Fontanot  <https://orcid.org/0000-0003-4744-0188>  
Mara Salvato  <https://orcid.org/0000-0001-7116-9303>

#### References

- Abolfathi, B., Aguado, D. S., Aguilar, G., et al. 2018, *ApJS*, 235, 42  
Anderson, T. W. 1984, *An Introduction to Multivariate Statistical Analysis* (Berlin: Wiley)  
Assef, R. J., Stern, D., Kochanek, C. S., et al. 2013, *ApJ*, 772, 26  
Bovy, J., Myers, A. D., Hennawi, J. F., et al. 2012, *ApJ*, 749, 41  
Carilli, C. L., Gnedin, N. Y., & Owen, F. 2002, *ApJ*, 27, 22  
Collins, M. 2001, *MNRAS*, 328, 1039C  
Cristiani, S., Avila, G., Bonifacio, P., et al. 2007, *NCimB*, 122, 1165  
D'Odorico, V., Cristiani, S., Pomante, E., et al. 2016, *MNRAS*, 463, 2690

- Fan, X., Wang, F., Yang, J., et al. 2019, *ApJ*, 870L, 11
- Gaia Collaboration, Brown, A. G. A., Vallenari, A., et al. 2018, *A&A*, 616, A1
- Gaia Collaboration, Prusti, T., de Bruijine, J. H. J., et al. 2016, *A&A*, 595, A1
- Kulkarni, G., Worseck, G., & Hennawi, J. F. 2019, *MNRAS*, 488, 1035
- Leite, A. C. O., & Martins, C. J. A. P. 2016, *PhRvD*, 94, 023503
- Liske, J., Grazian, A., Vanzella, E., et al. 2008, *MNRAS*, 386, 1192
- McQuinn, M. 2016, *ARA&A*, 54, 313
- Meiksin, A. A. 2009, *RvMP*, 81, 1405
- Pacucci, F., & Loeb, A. 2019, *ApJ*, 870L, 12
- Pâris, I., Petitjean, P., Aubourg, E., et al. 2018, *A&A*, 613, A51
- Polletta, M., Weedman, D., Honig, S., et al. 2008, *ApJ*, 675, 960
- Richards, G. T., Fan, X., Newberg, H. J., et al. 2002, *AJ*, 123, 2945
- Schindler, J.-T., Fan, X., Huang, Y.-H., et al. 2019, *ApJS*, 243, 5
- Skrutskie, M. F., Cutri, R. M., Stiening, R., et al. 2006, *AJ*, 131, 1163
- Syphers, D., & Shull, J. M. 2014, *ApJ*, 784, 42
- Tie, S. S., Martini, P., Mudd, D., et al. 2017, *AJ*, 153, 107
- Véron-Cetty, M.-P., & Véron, P. 2010, *A&A*, 518, A10
- Wang, F., Wu, X.-B., Fan, X., et al. 2016, *ApJ*, 819, 24
- Wolf, C., Hon, W. J., Bian, F., et al. 2019, *MNRAS*, in press
- 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., et al. 2016, *ApJ*, 825, 144
- Wright, E., Eisenhardt, P. R. M., Mainzer, A. K., et al. 2010, *AJ*, 140, 1868