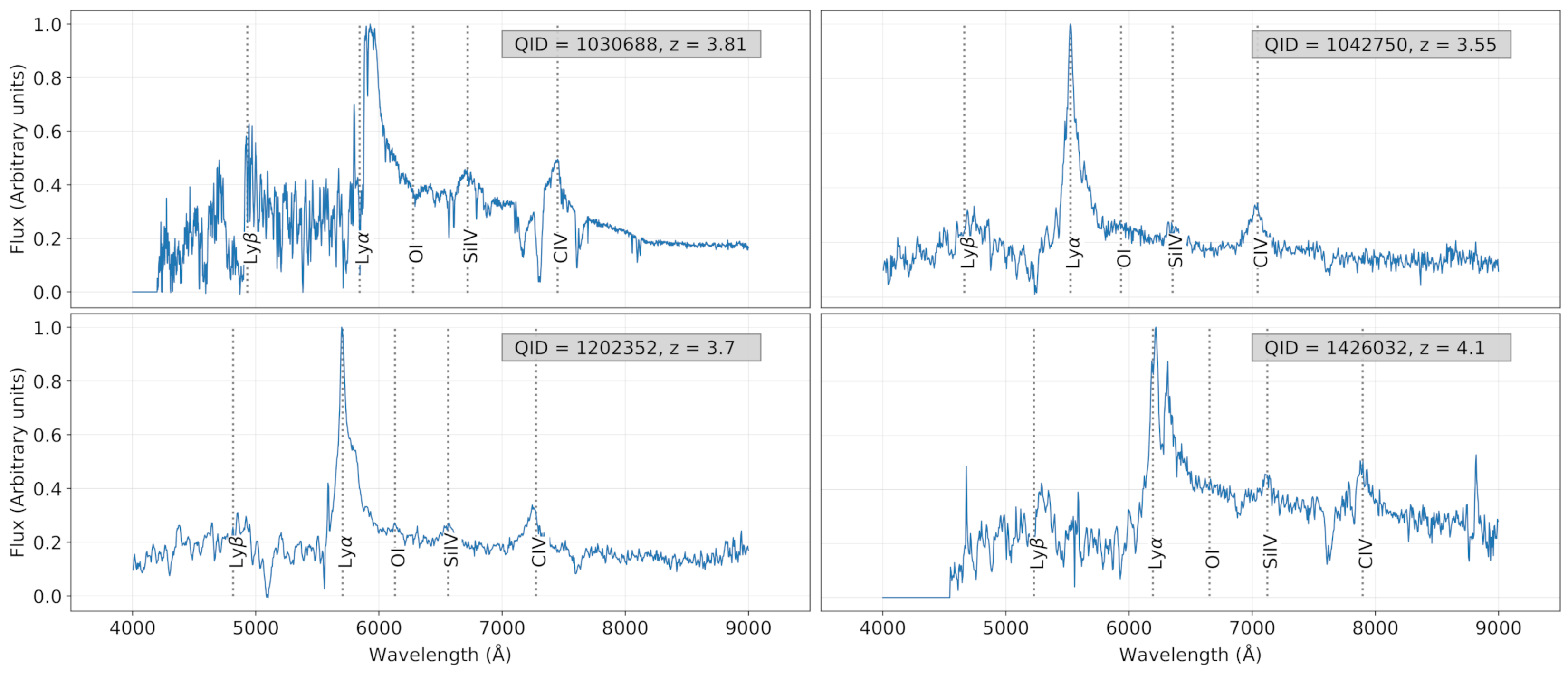
# QUBRICS: machine learning for searching bright, high-redshift quasars

F. Guarneri (INAF - OATS), S. Cristiani (INAF - OATS), G. Calderone (INAF - OATS ), K. Boutsia (Las Campanas Observatory), A. Grazian (INAF - OAPD), G. Cupani (INAF - OATS), F. Fontanot (INAF - OATS), V. D’Odorico (INAF - OATS), M. Porru (University of Trieste)

**Abstract**: the size and complexity of current astronomical datasets has grown to the point of making human analysis, in many cases, impractical or even impossible. With the advent of future observatories (e.g. LSST) and the generation of an unprecedented amount of data, automatic tools to extract information have become mandatory.

A practical example of this situation is the search for bright, high-z QSOs in wide area surveys. These targets, among the rarest sources in the sky, enable a wide range of scientific applications both in cosmology and fundamental physics, but their number in the southern sky is relatively scarce. To fill the gap, we developed QUBRICS (QUasars as BRIght beacons for Cosmology in the Southern hemisphere) that is based on machine learning techniques applied to current and future photometric databases. Since 2019 over 450 new, bright ($i < 18$ and/or $Y<18.5$) and high-redshift ($z > 2.5$) QSOs have been identified using different, complementary methods (e.g. CCA, PRF, XGB). This talk will describe the QSO selection algorithms, their performances and the current state of QUBRICS, highlighting peculiarities, lessons learned, future prospects and scientific advancements enabled by QSOs discovered by QUBRICS. As an example, the gain produced by QUBRICS to carry out the Sandage Redshift-Drift Test at the ELT will be shown.

## Example of discovery spectra



**A golden sample for the Sandage test**: the redshift 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).

Liske et al. 2008 concluded that an ELT-type telescope would be capable of unambiguously detecting the redshift drift over a period of ~20 yr using 4000 h of observing time (assuming a 42m primary mirror, planned at that time).

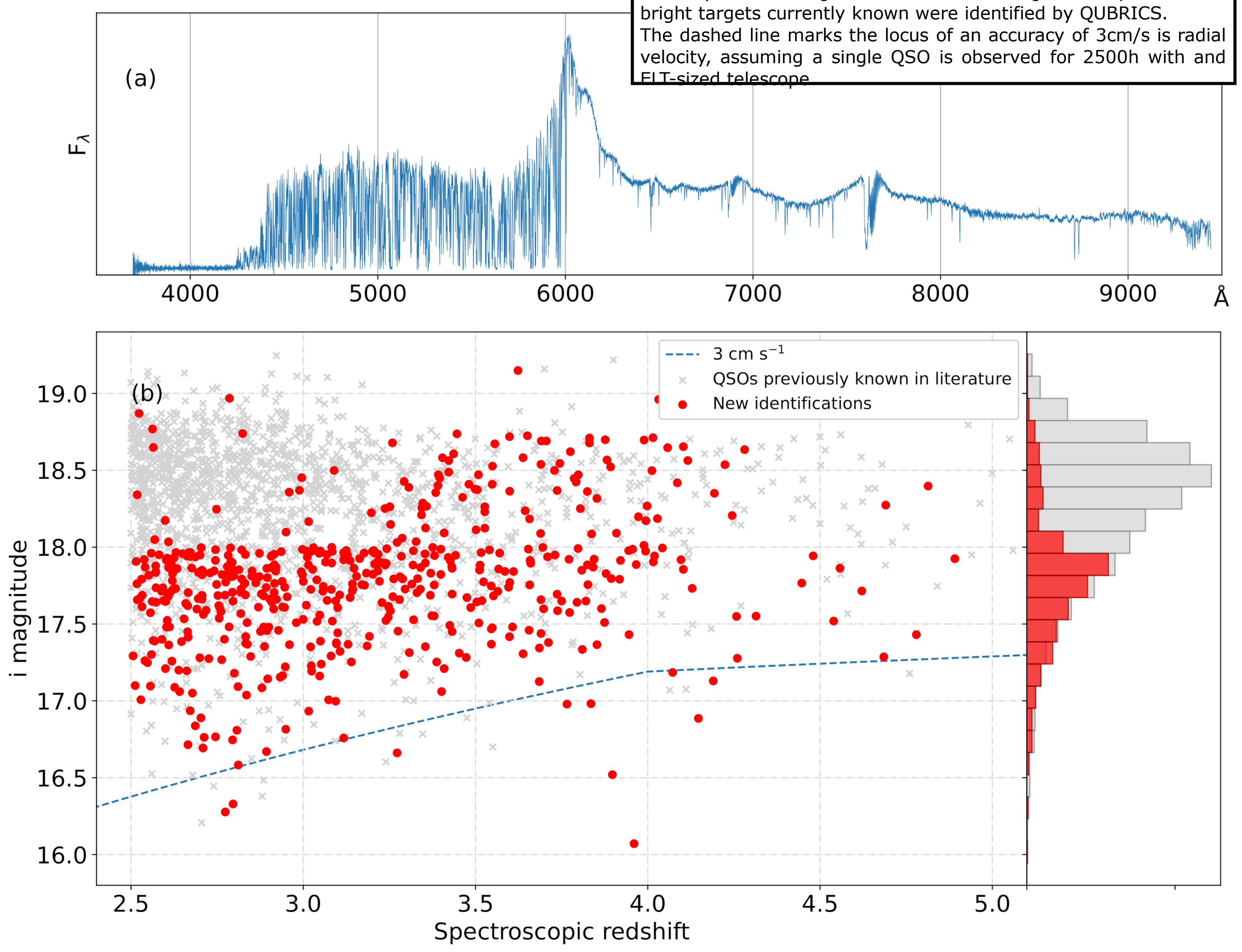
In Boutsia et al. 2020 we have repeated the estimate of the time required to carry out the Sandage Test, aiming at a 3σ detection and observing 30 targets twice at 25 years distance.

Thanks to the QUBRICS detection of new bright QSOs at high redshift, the total time required to carry out the Sandage Test is reduced to less than 2500h. Moreover, each QSO needs less than 100h of integration to provide a velocity accuracy of 22.8 cm/s, required for a global 3σ detection of the drift.

Top panel: spectrum of a very bright ($i = 16.2$) and high redshift ($z = 3.96$) QSO. $F_\lambda$ units are arbitrary.

Bottom panel: new QSOs discovered by QUBRICS (red dots) compared to targets previously known in literature (grey crosses). The histogram shows how a significant part of the bright targets currently known were identified by QUBRICS.

The dashed line marks the locus of an accuracy of 3cm/s is radial velocity, assuming a single QSO is observed for 2500h with and ELT-sized telescope

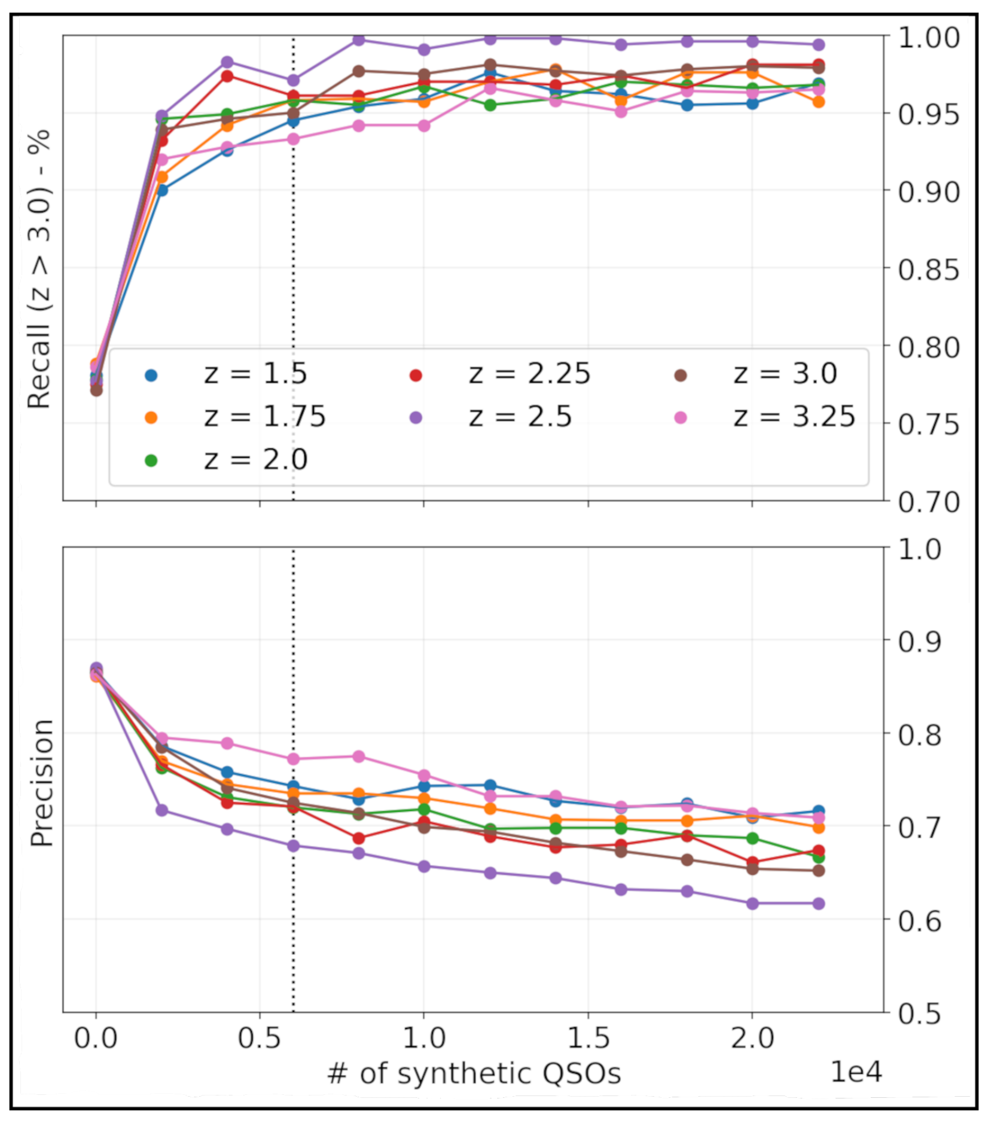


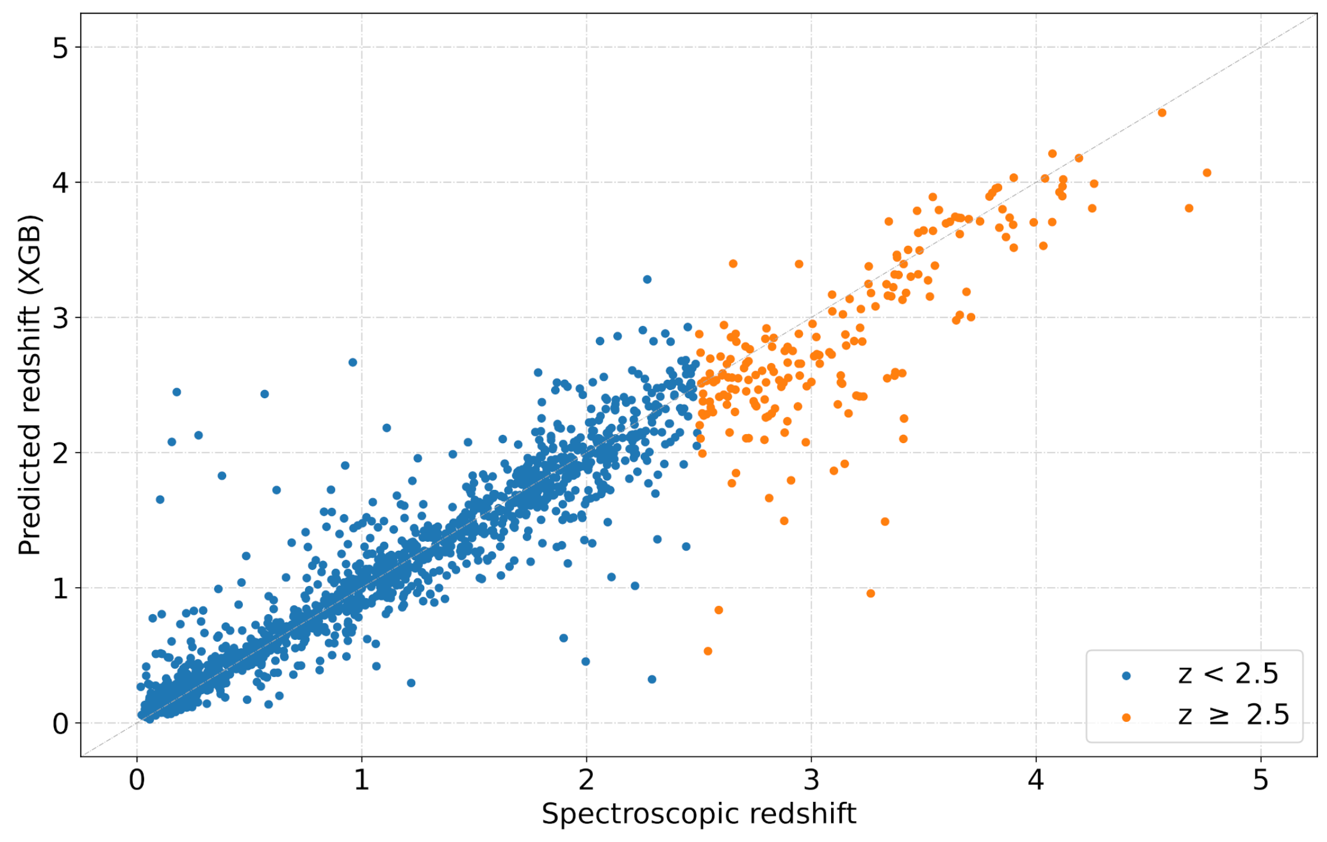
**The current selection algorithms:**
Two machine learning algorithms are being employed, the probabilistic random forest (PRF) and the XGBoost (XGB).

The PRF is a modified random forest which keeps into account measurement uncertainties while building the predictive model. This results in a more reliable and accurate model, with the added bonus of handling missing data.

The XGB is a boosting algorithm: after building a model, the latter is refined in order to minimize the errors on the training set, while controlling overfitting.

Both algorithms are applied on photometric data, collected from several surveys (SkyMapper, DES, PanSTARRS). Each algorithm is robust and provides good performances; in both cases a prioritization is possible based on the photometric redshift estimate or a reliability score, computed for each candidate.





Top panel: performance of the PRF algorithm on a test dataset. Synthetic data are being introduced to provide diversified examples and improve the completeness at high-z.

Left: photometric redshift predicted by the XGB compared to the spectroscopic redshift of each target.

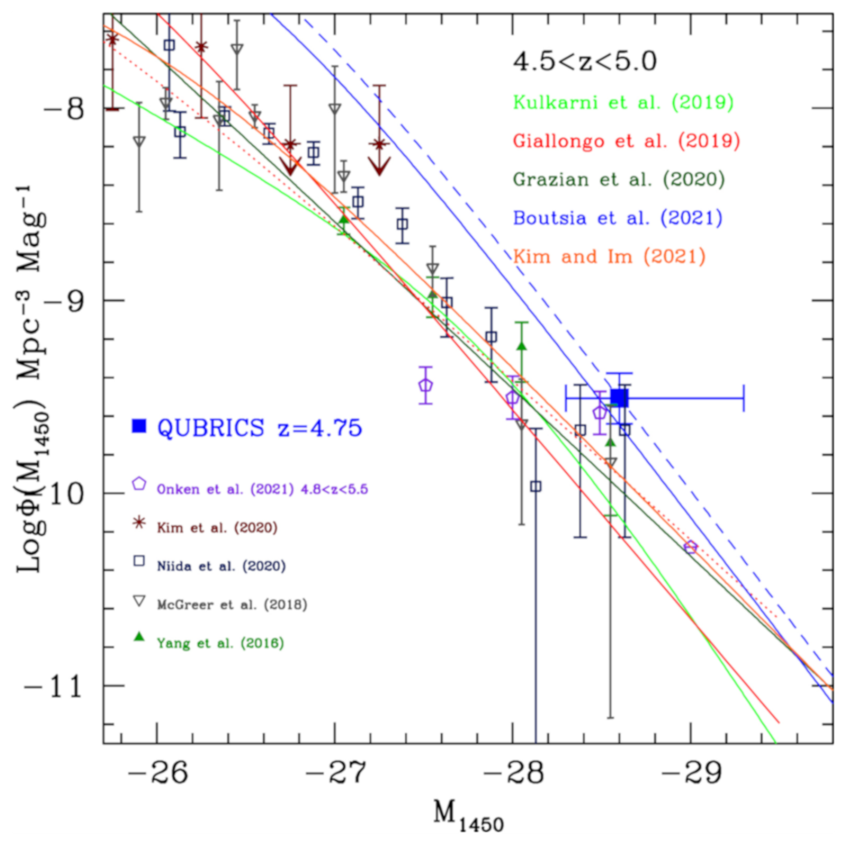
**New estimates for the QSO luminosity function:**

Boutsia et al. 2021 and Grazian et al. 2022 used subsamples of QUBRICS QSOs to provide new estimates on the bright end of the luminosity function.

The samples, selected respectively in the $3.6 < z < 4.2$ and $4.5 < z < 5.0$ redshift ranges, extend the sampled range up to an unprecedented $M_{1450} = -29.5$ absolute magnitude (Boutsia et al. 2021) or $-29.3 < M_{1450} < -28.3$. Evidences derived from the bright-end slope suggest that:

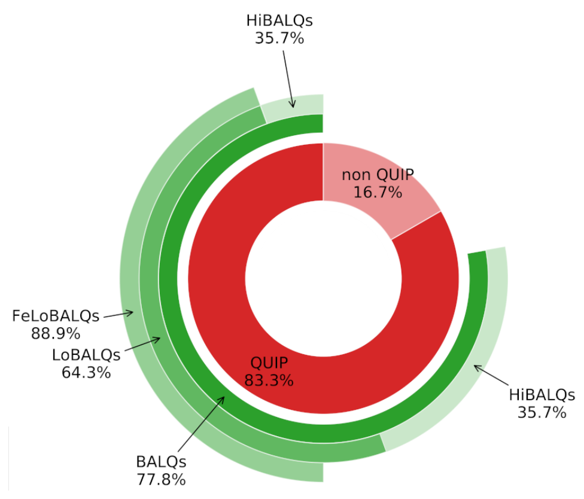
- bright QSOs could be significant contributors to the ionizing background at z~4;
- the space density within $-29.3 < M_{1450} < -28.3$ QSOs at z~5 could be to be three times higher than previous determinations.

Such findings are consistent with a pure density evolution for the AGN population at z>3.



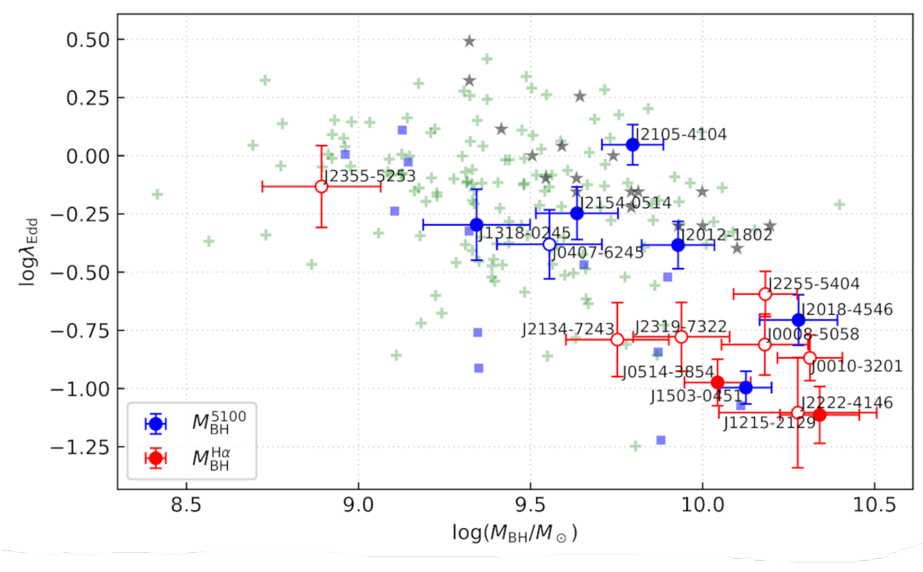
The luminosity function of QSOs at $4.5 \leq z \leq 5.0$ from QUBRICS (blue filled squares) compared to other luminosity functions. Data points and curves have been shifted to $z = 4.75$ (pure density evolution, $\gamma = -0.25$). The blue square indicates the mean absolute magnitude of the QSOs inside a bin, marked by the blue horizontal bar. The blue continuous line is the best fit of the QSO luminosity function at $z \sim 4$ by Boutsia et al. (2021), evolved at $z = 4.75$.

**BALs and FeLoBALs:**
Some of the observed candidates presented broad absorption lines and were difficult to classify using only optical spectra. IR spectra, collected with the FIRE instrument, confirmed their nature as bona fide QSOs. The analysis of their spectral features was performed with ASTROCOOK (Cupani et al. 2020) and QSFIT (Calderone et al. 2017), revealing that the large majority of these objects are broad absorption line (BAL) QSOs, with almost half of them displaying strong FeII absorption (FeLoBAL QSOs). The measured properties of FeLoBAL QSOs observed so far provide no evidence that they are a manifestation of a particular stage in active galactic nucleus (AGN) evolution.



Left: distribution of BAL types among the QUBRICS observations without secure spectroscopic redshift before Cupani et al. 2022. Not all targets were QUIPs (QUBRICS Irregulars and Peculiars)

Right: Eddington ratios estimated for each target presented in Cupani et al. 2022. FeLoBALs are denoted with fillled circles while other targets are denoted with empty circles. Light-blue squares are from Schulze et al. (2017), green crosses from Coatman et al. (2017); grey stars from Vietri et al. (2018).



**QUBRICS publications - https://qubrics.oas.inaf.it/**

- Boutsia et al., 2020, ApJS, 250, 26
- Boutsia et al., 2021, ApJ, 912, 111
- Calderone et al., 2019, ApJ, 887, 268
- Cupani et al. 2022, MNRAS, 510, 2509
- Grazian et al. 2022, ApJ, 924, 62
- Guarneri et al. 2021, MNRAS, 506, 2471

**Other references:**

- Coatman et al. 2017, MNRAS, 465, 2120
- Vietri et al. 2018, A&A, 617, A81
- Schulze et al. 2017, ApJ, 848, 104
- Cupani et al. 2020, Proc. SPIE 11452
- Calderone et al. 2017, MNRAS, 472 4051