

---

# A new catalog of 343,000 quasars with their photometric redshifts derived with machine learning from the Kilo Degree Survey

Szymon Nakoneczny<sup>\*1</sup>, Maciej Bilicki<sup>2</sup>, Agnieszka Pollo<sup>1</sup>, Marika Asgari<sup>3</sup>, Andrej Dvornik<sup>4</sup>, Thomas Erben<sup>5</sup>, Benjamin Giblin<sup>6</sup>, Catherine Heymans<sup>3</sup>, Hendrik Hildebrandt<sup>7</sup>, Arun Kannawadi<sup>8</sup>, Koen Kuijken<sup>9</sup>, Nicola Napolitano<sup>10</sup>, and Edwin Valentijn<sup>11</sup>

<sup>1</sup>National Centre for Nuclear Research [Otwock] – Poland

<sup>2</sup>Center for Theoretical Physics, Polish Academy of Sciences – Poland

<sup>3</sup>Institute for Astronomy, University of Edinburgh – United Kingdom

<sup>4</sup>Ruhr University Bochum, Faculty of Physics and Astronomy, Astronomical Institute (AIRUB), German Centre for Cosmological Lensing – Germany

<sup>5</sup>Argelander-Institut für Astronomie – Germany

<sup>6</sup>Institute for Astronomy, University of Edinburgh, Royal Observatory – United Kingdom

<sup>7</sup>Argelander Institute for Astronomy, Bonn University (AIfA) – Germany

<sup>8</sup>Department of Astrophysical Sciences, Princeton University – United States

<sup>9</sup>Leiden Observatory, Leiden University – Netherlands

<sup>10</sup>School of Physics and Astronomy, Sun Yat-sen University – China

<sup>11</sup>Kapteyn Institute – University of Groningen, Netherlands

## Abstract

In the present era of ever larger photometric sky surveys, designing efficient automated methods for unbiased selection of specific categories of objects and prediction of their properties based on photometric data is becoming an increasingly urgent quest. A good example of such objects are quasars (QSOs), both because of our interest in their physics and possible cosmological applications.

In my talk I present a catalog of 343,000 QSOs derived from one of the largest photometric surveys up to date - the Kilo-Degree Survey (KiDS) Data Release 4. The catalog is characterised by robust photometric redshifts with uncertainties, and its high magnitude depth was achieved with artificial neural networks (ANN) applied in a controlled way. The catalog is based on optical ugr<sub>i</sub> and near-infrared ZYJHK<sub>s</sub> bands. With purity 97%, completeness 94%, and redshift error (mean and scatter) 0.009 +/- 0.12, the catalog is limited at  $r < 23.5$ . The robustness of the catalog was demonstrated e.g. by comparing the number counts of our QSO candidates to models from eBOSS, and by cross-matching the QSOs with Gaia in order to confirm zero parallaxes. In the talk, I will explain how machine learning (ML) methods can be successfully applied to observations fainter than the training data known from spectroscopy. It includes building inference subsets using high dimensional t-SNE visualisations of the feature space, and properly adapting bias vs variance tradeoff of the ML

---

\*Speaker

models with tests against the problem of extrapolation in the feature space. Our success of the extrapolation challenges the way that models are optimised and applied at the faint data end. The resultant catalog is ready for cosmological and active galactic nucleus (AGN) studies, and developed methodology is ready to be used on future yet larger photometric datasets.

Slides: in PDF

Video: <https://youtu.be/Kwv0CH4KKRA>

**Keywords:** quasars, photometry, catalogs, extrapolation, large scale structure