

Euclid Quick Data Release (Q1)

First study of red quasars selection

Euclid Collaboration: F. Tarsitano^{*1}, S. Fotopoulou², M. Banerji³, J. Petley⁴, A. L. Faisst⁵, M. Tucci¹, S. Tacchella⁶, Y. Toba^{7,8,9}, H. Landt¹⁰, Y. Fu^{4,11}, P. A. C. Cunha^{12,13}, K. Duncan¹⁴, W. Roster¹⁵, M. Salvato¹⁵, B. Laloux^{16,15}, P. Dayal¹¹, F. Ricci^{17,18}, N. Aghanim¹⁹, B. Altieri²⁰, A. Amara²¹, S. Andreon²², N. Auricchio²³, H. Aussel²⁴, C. Baccigalupi^{25,26,27,28}, M. Baldi^{29,23,30}, A. Balestra³¹, S. Bardelli²³, P. Battaglia²³, A. Biviano^{26,25}, A. Bonchi³², E. Branchini^{33,34,22}, M. Brescia^{35,16}, J. Brinchmann^{13,12}, S. Camera^{36,37,38}, G. Cañas-Herrera^{39,40,4}, V. Capobianco³⁸, C. Carbone⁴¹, J. Carretero^{42,43}, S. Casas⁴⁴, M. Castellano¹⁸, G. Castignani²³, S. Cavuoti^{16,45}, K. C. Chambers⁴⁶, A. Cimatti⁴⁷, C. Colodro-Conde⁴⁸, G. Congedo¹⁴, C. J. Conselice⁴⁹, L. Conversi^{50,20}, Y. Copin⁵¹, A. Costille⁵², F. Courbin^{53,54}, H. M. Courtois⁵⁵, M. Cropper⁵⁶, A. Da Silva^{57,58}, H. Degaudenzi¹, G. De Lucia²⁶, A. M. Di Giorgio⁵⁹, C. Dolding⁵⁶, H. Dole¹⁹, F. Dubath¹, C. A. J. Duncan⁴⁹, X. Dupac²⁰, S. Escoffier⁶⁰, M. Fabricius^{15,61}, M. Farina⁵⁹, R. Farinelli²³, F. Faustini^{32,18}, S. Ferriol⁵¹, F. Finelli^{23,62}, M. Frailis²⁶, E. Franceschi²³, S. Galeotta²⁶, K. George⁶¹, W. Gillard⁶⁰, B. Gillis¹⁴, C. Giocoli^{23,30}, P. Gómez-Alvarez^{63,20}, J. Gracia-Carpio¹⁵, B. R. Granett²², A. Grazian³¹, F. Grupp^{15,61}, L. Guzzo^{64,22,65}, S. Gwyn⁶⁶, S. V. H. Haugan⁶⁷, W. Holmes⁶⁸, I. M. Hook⁶⁹, F. Hormuth⁷⁰, A. Hornstrup^{71,72}, P. Hudelot⁷³, K. Jahnke⁷⁴, M. Jhabvala⁷⁵, E. Keihänen⁷⁶, S. Kermiche⁶⁰, A. Kiessling⁶⁸, B. Kubik⁵¹, M. Kümmel⁶¹, M. Kunz⁷⁷, H. Kurki-Suonio^{78,79}, Q. Le Bouc'h⁸⁰, A. M. C. Le Brun⁸¹, D. Le Mignant⁵², S. Ligi³⁸, P. B. Lilje⁶⁷, V. Lindholm^{78,79}, I. Lloro⁸², G. Mainetti⁸⁰, D. Maino^{64,41,65}, E. Maiorano²³, O. Mansutti²⁶, S. Marcin⁸³, O. Marggraf⁸⁴, M. Martinelli^{18,85}, N. Martinet⁵², F. Marulli^{86,23,30}, R. Massey⁸⁷, E. Medinaceli²³, S. Mei^{88,89}, M. Melchior⁹⁰, Y. Mellier^{91,73}, M. Meneghetti^{23,30}, E. Merlin¹⁸, G. Meylan⁹², A. Mora⁹³, M. Moresco^{86,23}, L. Moscardini^{86,23,30}, R. Nakajima⁸⁴, C. Neissner^{94,43}, S.-M. Niemi³⁹, J. W. Nightingale⁹⁵, C. Padilla⁹⁴, S. Paltani¹, F. Pasian²⁶, K. Pedersen⁹⁶, W. J. Percival^{97,98,99}, V. Pettorino³⁹, S. Pires²⁴, G. Polenta³², M. Poncet¹⁰⁰, L. A. Popa¹⁰¹, L. Pozzetti²³, F. Raison¹⁵, R. Rebolo^{48,102,103}, A. Renzi^{104,105}, J. Rhodes⁶⁸, G. Riccio¹⁶, E. Romelli²⁶, M. Roncarelli²³, E. Rossetti²⁹, B. Rusholme⁵, R. Saglia^{61,15}, Z. Sakr^{106,107,108}, D. Sapone¹⁰⁹, B. Sartoris^{61,26}, J. A. Schewtschenko¹⁴, P. Schneider⁸⁴, T. Schrabback¹¹⁰, M. Scodreggio⁴¹, A. Secroun⁶⁰, G. Seidel⁷⁴, M. Seiffert⁶⁸, S. Serrano^{111,112,113}, P. Simon⁸⁴, C. Sirignano^{104,105}, G. Sirri³⁰, A. Spurio Mancini¹¹⁴, L. Stanco¹⁰⁵, J. Steinwagner¹⁵, P. Tallada-Crespí^{42,43}, A. N. Taylor¹⁴, H. I. Teplitz¹¹⁵, I. Tereno^{57,116}, S. Toft^{117,118}, R. Toledo-Moreo¹¹⁹, F. Torradeflot^{43,42}, I. Tutusaus¹⁰⁷, L. Valenziano^{23,62}, J. Valiviita^{78,79}, T. Vassallo^{61,26}, G. Verdoes Kleijn¹¹, A. Veropalumbo^{22,34,33}, Y. Wang¹¹⁵, J. Weller^{61,15}, A. Zacchei^{26,25}, G. Zamorani²³, F. M. Zerbi²², E. Zucca²³, V. Alleinato¹⁶, M. Ballardini^{120,121,23}, M. Bolzonella²³, E. Bozzo¹, C. Burigana^{122,62}, R. Cabanac¹⁰⁷, A. Cappi^{23,123}, D. Di Ferdinando³⁰, J. A. Escartin Vigo¹⁵, L. Gabarra¹²⁴, J. Martín-Fleitas⁹³, S. Matthew¹⁴, N. Mauri^{47,30}, R. B. Metcalf^{86,23}, A. Pezzotta^{125,15}, M. Pöntinen⁷⁸, C. Porciani⁸⁴, I. Risso¹²⁶, V. Scottez^{91,127}, M. Sereno^{23,30}, M. Tenti³⁰, M. Viel^{25,26,28,27,128}, M. Wiesmann⁶⁷, Y. Akrami^{129,130}, I. T. Andika^{131,132}, S. Anselmi^{105,104,133}, M. Archidiacono^{64,65}, F. Atrio-Barandela¹³⁴, C. Benoist¹²³, K. Benson⁵⁶, D. Bertacca^{104,31,105}, M. Bethermin¹³⁵, L. Bisigello³¹, A. Blanchard¹⁰⁷, L. Blot^{136,133}, M. L. Brown⁴⁹, S. Bruton¹³⁷, A. Calabro¹⁸, F. Caro¹⁸, C. S. Carvalho¹¹⁶, T. Castro^{26,27,25,128}, Y. Charles⁵², F. Cogato^{86,23}, A. R. Cooray¹³⁸, O. Cucciati²³, S. Davini³⁴, F. De Paolis^{139,140,141}, G. Desprez¹¹, A. Díaz-Sánchez¹⁴², J. J. Diaz¹⁴³, S. Di Domizio^{33,34}, J. M. Diego¹⁴⁴, P.-A. Duc¹³⁵, A. Enia^{29,23}, Y. Fang⁶¹, A. G. Ferrari³⁰, A. Finoguenov⁷⁸, A. Fontana¹⁸, A. Franco^{140,139,141}, K. Ganga⁸⁸, J. García-Bellido¹²⁹, T. Gasparotto²⁶, V. Gautard¹⁴⁵, E. Gaztanaga^{113,111,146}, F. Giacomini³⁰, F. Gianotti²³, G. Gozaliasi^{147,78}, M. Guidi^{29,23}, C. M. Gutierrez¹⁴⁸, A. Hall¹⁴, C. Hernández-Monteagudo^{103,48}, H. Hildebrandt¹⁴⁹, J. Hjorth⁹⁶, J. J. E. Kajava^{150,151}, Y. Kang¹, V. Kansal^{152,153}, D. Karagiannis^{120,154}, K. Kiiveri⁷⁶, C. C. Kirkpatrick⁷⁶, S. Kruk²⁰, J. Le Graet⁶⁰, L. Legrand^{155,6}, M. Lembo^{120,121}, F. Lepori¹⁵⁶, G. Leroy^{10,87}, G. F. Lesci^{86,23}, J. Lesgourgues⁴⁴, L. Leuzzi^{86,23}, T. I. Liaudat¹⁵⁷, S. J. Liu⁵⁹, A. Loureiro^{158,159}, J. Macias-Perez¹⁶⁰, G. Maggio²⁶, M. Magliocchetti⁵⁹, F. Mannucci¹⁶¹, R. Maoli^{162,18}, C. J. A. P. Martins^{163,13}, L. Maurin¹⁹, M. Miluzio^{20,164}, P. Monaco^{165,26,27,25}, C. Moretti^{28,128,26,25,27}, G. Morgante²³, S. Nadathur¹⁴⁶, K. Naidoo¹⁴⁶, A. Navarro-Alsina⁸⁴, S. Nesseris¹²⁹, F. Passalacqua^{104,105}, K. Paterson⁷⁴, L. Patrizii³⁰, A. Pisani^{60,166}, D. Potter¹⁵⁶, S. Quai^{86,23}, M. Radovich³¹, P.-F. Rocci¹⁹, G. Rodighiero^{104,31}, S. Sacquegna^{139,140,141}, M. Sahlén¹⁶⁷, D. B. Sanders⁴⁶, E. Sarpa^{28,128,27}, A. Schneider¹⁵⁶, M. Schultheis¹²³, D. Sciotti^{18,85}, E. Sellentin^{168,4}, F. Shankar³, L. C. Smith¹⁶⁹, K. Tanidis¹²⁴, G. Testera³⁴, R. Teyssier¹⁶⁶, S. Tosi^{33,126}

A. Troja^{104, 105}, C. Valieri³⁰, A. Venhola¹⁷⁰, D. Vergani²³, G. Verza¹⁷¹, P. Vielzeuf⁶⁰, A. Viitanen^{76, 18}, N. A. Walton¹⁶⁹,
and J. G. Sorce^{172, 19}

(Affiliations can be found after the references)

March 17, 2025

ABSTRACT

Red quasars constitute an important but elusive phase in the evolution of supermassive black holes, where dust obscuration can significantly alter their observed properties. They have broad emission lines, like other quasars, but their optical continuum emission is significantly reddened, which is why they were traditionally identified based on near- and mid-infrared selection criteria. This work showcases the capability of the *Euclid* space telescope to find a large sample of red quasars, using *Euclid* near infrared (NIR) photometry. We first conduct a forecast analysis, comparing a synthetic catalogue of red QSOs with COSMOS2020. Using template fitting, we reconstruct *Euclid*-like photometry for the COSMOS sources and identify a sample of candidates in a multidimensional colour-colour space achieving 98% completeness for mock red QSOs with 30% contaminants. To refine our selection function, we implement a probabilistic Random Forest classifier, and use UMAP visualisation to disentangle non-linear features in colour-space, reaching 98% completeness and 88% purity. A preliminary analysis of the candidates in the *Euclid* Deep Field Fornax (EDF-F) shows that, compared to VISTA+DECAm-based colour selection criteria, *Euclid*'s superior depth, resolution and optical-to-NIR coverage improves the identification of the reddest, most obscured sources. Notably, the *Euclid* exquisite resolution in the I_E filter unveils the presence of a candidate dual quasar system, highlighting the potential for this mission to contribute to future studies on the population of dual AGN. The resulting catalogue of candidates, including more than 150 000 sources, provides a first census of red quasars in *Euclid* Q1 and sets the groundwork for future studies in the *Euclid* Wide Survey (EWS), including spectral follow-up analyses and host morphology characterisation.

Key words. Galaxies: evolution, active; quasars: general, supermassive black holes; Methods: statistical, numerical.

1. Introduction

Supermassive black holes ($M_{\text{BH}} > 10^6 M_{\odot}$) and their host galaxies are believed to grow in tandem, as postulated by theoretical (e.g., [Silk & Rees 1998](#)) and observational (e.g., [Magorrian et al. 1998](#)) arguments. At the same time, early galaxy evolution simulations showed that in order to reproduce observed galaxy size and brightness distributions, as well as their star formation, some form of energetic feedback from the central black hole is required ([Bower et al. 2006](#); [Croton 2006](#)). Black holes can influence their host galaxies through the release of gravitational potential energy in the form of radiation during their active phase of accretion of matter, also known as active galactic nuclei (AGN). In addition, some AGN are known to power energetic jets and winds which extend their influence on their host galaxy. Therefore, AGN hold the missing piece to advance our knowledge of the black hole-galaxy co-evolution ([Fabian 2012](#); [Heckman & Best 2014](#)).

The lack of a detailed theory of AGN feedback leads to unrealistic models, failing to capture the observed complexity of the AGN population. For example, models do not reliably predict the ratio of unobscured (type 1; face-on view of the accretion disc) to obscured (type 2; edge-on view of the accretion disc and obscuring torus) AGN, nor the evolution of their luminosity function. [Habouzit et al. \(2021\)](#) presented six current state-of-the-art galaxy evolution simulations compared to the number density of AGN derived from X-ray observations. Even though these models are successful in predicting the observed properties of normal galaxies, none of them predicts reliably the history of supermassive black hole growth and the corresponding AGN phase. Thus, the creation of an AGN activity model rooted in observations is needed to act as ground truth for galaxy evolution simulations and to motivate the prescription of stochastic processes in sub-grid physics.

The most significant challenge in AGN studies is that each part of the electromagnetic spectrum captures a different aspect of the central engine, leading to major inconsistencies between detection methods ([Padovani et al. 2017](#)). This is particularly

true for the obscured AGN population, nowadays suspected to also be a phase during the evolution of an AGN, and not only an outcome of a geometric alignment of the disc/torus system towards the observer.

The unification scenario ([Antonucci 1993](#); [Urry & Padovani 1995](#)) postulated that the observed variety in the presence of broad and narrow emission lines in the spectra of active galaxies was due to the obscuration induced by a molecular torus along the line of the sight of the observer. The distinction between AGN and quasars is largely a description of the relative luminosity of the central engine and the host galaxy, with quasars being extremely luminous and dominating over the host galaxy emission. The definition of red quasars corresponds to sources that show broad lines (i.e. type 1 sources) but with significant absorption in their continuum ([Gliksman et al. 2012](#); [Banerji et al. 2013](#)). Red quasars seem to contradict the AGN unification scenario, as first [Klindt et al. \(2019\)](#) and more recently [Andonie et al. \(2022\)](#), [Fawcett et al. \(2023\)](#); [Petley et al. \(2024\)](#); [Calistro Rivera et al. \(2024\)](#) and [Yue et al. \(2024\)](#), showed evidence of enhanced radio detection rates from this population, at odds with a simple orientation-induced obscuration. Current models and several observations argue that red quasars could be an initial, short-lived stage during the onset of quasar activity within a galaxy. As the gas and dust is driven into the centre of the galaxy, the initial phase of accretion is enshrouded in a dusty cocoon. Subsequently, the radiation pressure, and induced winds will clear out the region around the black hole revealing a type 1, unobscured blue quasar ([Urrutia et al. 2009](#); [Banerji et al. 2012, 2015](#); [Temple et al. 2019](#); [Calistro Rivera et al. 2021](#)).

Determining the physical parameters of quasars and AGN and contrasting them with inactive galaxies (i.e., non-AGN hosts) as a function of luminosity, stellar mass, star formation rate, obscuration, as well as across cosmic time and large-scale environment is crucial (e.g., [Wethers et al. 2018](#)), [Laloux et al.](#), in prep. It will enable the creation of an evolutionary scenario for galaxies including the incidence of AGN, which is much needed to establish the true evolutionary path of quasars and to inform the recipes used in simulations. However, AGN are short-lived phenomena and detailed statistical studies of this population have been hindered by the size of the available datasets.

* e-mail: federica.tarsitano@unige.ch

The value of large AGN and quasar samples is not only in the accurate determination of AGN luminosity distribution functions but also in allowing for detailed decomposition into AGN sub-populations. A recent decomposition of about ~ 150 mid-infrared detected AGN, split into unobscured, red quasars, and type 2 sources, shows the luminosity distribution functions to be a double-power law form with a break at a characteristic luminosity (Glikman et al. 2018). Similar distributions are found also in other wavelengths (e.g., Fotopoulou et al. 2016). Much larger samples are needed to draw definitive conclusions, however these studies show already differences in the number density of red and blue quasars, or type 1 and type 2 quasars between low and high redshift ($z \sim 1$).

Euclid is a mission of the European Space Agency (ESA), expected to detect billions of sources in the optical and near-infrared (Laureijs et al. 2011; Euclid Collaboration: Mellier et al. 2024). The *Euclid* observational campaign will observe a third of the extra-Galactic sky through two surveys. The Euclid Wide Survey (EWS), covering more than $\sim 14\,000$ deg², and the Euclid Deep Survey (EDS), focusing on three different areas for a total of 63.1 deg². With its unprecedented, large dataset, *Euclid* offers the possibility to study and identify the largest sample to date of unobscured and obscured AGN across all wavebands, extending to the faintest sources. A key aspect will be the determination of a corresponding selection function, which remains under ongoing development, particularly in the context of AGN studies. This paper presents a study on red quasars within the *Euclid* Quick Release Q1 (2025), comparing mock and observed sources, and discussing and refining colour selection criteria that will be functional for the EWS.

2. Dataset

In the forecast analysis, we study the separation between mock red QSOs and observed sources. We use a synthetic catalogue for the former (more information follows in Sect. 2.2) and the COSMOS2020 dataset (Weaver et al. 2022) for the latter. Mock red quasars are described over a range of wavelengths extending from the optical through the near infrared (NIR), and up to the mid-infrared (MIR): DECam g, r, i, z in the optical (Dark Energy Survey Collaboration et al. 2016), *Euclid* VIS (I_e) and NISP Y_e, J_e, H_e in NIR (Euclid Collaboration: Cropper et al. 2024; Euclid Collaboration: Jahnke et al. 2024; Euclid Collaboration: Schirmer et al. 2022), VISTA J, H, K_s from NIR to MIR (McCracken et al. 2012) and WISE $W1, W2$ in MIR (Wright et al. 2010). The collection of filters used in this work and their weighted central wavelengths is shown in Fig. 1. The employed photometry is expressed in the AB magnitude system.

We clean the COSMOS2020 sample from corrupted photometry and fake detections, and exclude the objects with null entries for the photometric redshift. We apply the mask FLAG_COMBINED = 0 to remove objects near bright stars and saturated regions, ACS_MU_CLASS = 1 to separate stars from galaxies, and lp_type $\neq 9$ to exclude fake detections. Furthermore, we apply a magnitude upper cut at 23.5 in VISTA H in both COSMOS2020 and the mocks. Among the selected sources in the COSMOS2020 catalogue, we take into account those that are flagged as AGN candidates. The selected catalogue includes 95 052 objects.

Additionally, we keep track of 1493 X-ray sources observed with *Chandra* (Civano et al. 2016), which we do not consider in this analysis. X-ray-selected AGN outline a broad range of obscuration, and their optical-to-NIR colours may not fully align with the selection criteria used in this study, based solely on red

Table 1. Sample selection function applied to the Q1 sample. In FLUX_filter_nFWHM_APER, filter is the passband and $n = 1, 2, 3, 4$ according to the aperture.

Feature in <i>Euclid</i> Q1	Selected values
PHZ_CLASSIFICATION	[2,6]
PHZ_MEDIAN	finite
SPURIOUS_FLAG	0
DET_QUALITY_FLAG	< 8
FLUX_filter_2FWHM_APER	> 0

QSO NIR photometry. We will address this additional level of complexity in future work, using their properties to refine the distinction between reddened AGN and red galaxies.

We analyse mocks and COSMOS2020 (methods are described in Sect. 3) to derive a selection function for candidate red QSOs, and we apply it to *Euclid* Q1 (Euclid Collaboration: Mellier et al. 2024). Q1 consists of a first visit of the Euclid Deep Fields (EDFs), spanning across a total area of 63.1 deg² of the extragalactic sky, divided in the Euclid Deep Field North (EDF-N, 20 deg²), Euclid Deep Field Fornax (EDF-F, 10 deg²) and the Euclid Deep Field South (EDF-S, 23 deg²). More details about the Q1 release are presented in Euclid Collaboration: Ausser et al. (2025); Euclid Collaboration: McCracken et al. (2025); Euclid Collaboration: Polenta et al. (2025) and Euclid Collaboration: Romelli et al. (2025).

In this work, we focus on the EDF-F, for which we find overlap with a collection of AGN candidates from Zou et al. (2022), selected through the flag flag_IRagn_D12, which follows the MIR colour-based cut proposed in Donley et al. (2012), and a catalogue of radio-selected quasars (Miller et al. 2013). These two datasets serve as control samples. We build a first dataset of 5 301 332 EDF-F sources, obtained by matching the *Euclid* morphology (Euclid Collaboration: Romelli et al. 2025) and photometric redshift (Euclid Collaboration: Tucci et al. 2025a,b) catalogues, delivered by the OU-MER and OU-PHZ organizational units, respectively. Then we exclude objects flagged as spurious and with unphysical photometric and redshift properties. Furthermore, we consider only sources classified either as galaxies or QSOs, according to the OU-PHZ classification presented in Euclid Collaboration: Tucci et al. (2025b). A summary of this selection function is reported in Table 1. Additionally, we apply a cut near the limiting magnitude of the H_e band, corresponding to 23.5. The final subsample at play counts 1 331 325 sources.

2.1. Template fitting

Each object in COSMOS2020 is described by a photometric dataset including ultraviolet (UV) measures from GALEX, optical observations from the Subaru Hyper Suprime-Cam (HSC) and the Canada-France-Hawaii Telescope (CFHT), NIR data from VISTA, MIR from the SPLASH program of the *Spitzer* Space Telescope Space Telescope, and optical medium band observations from Subaru. We refer to Weaver et al. (2022) for references on these individual datasets. The properties of each filter are described in Table A.1.

The mock dataset includes VISTA, DECam, and *Euclid* VIS and NISP photometry. In order to match the photometric datasets of mocks and COSMOS2020, we use template fitting (TF) to estimate the optical and NIR fluxes that are originally not available in the latter. TF compares the input photometric dataset with a library of Spectral Energy Distributions (SEDs) to identify the

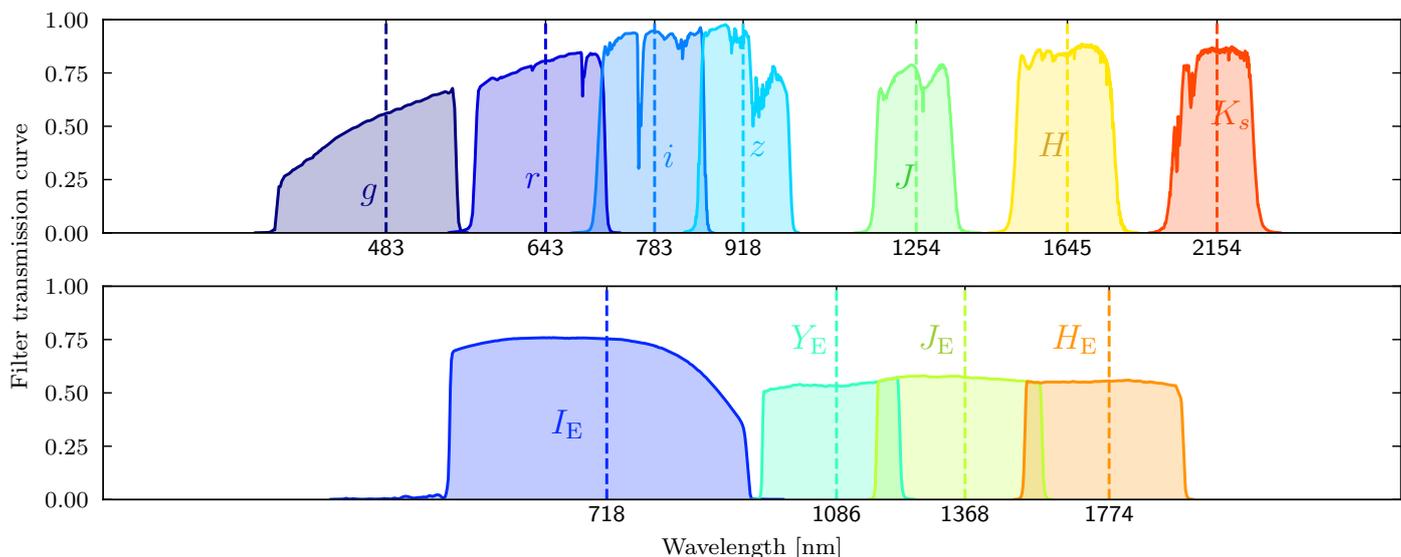


Fig. 1. Set of passbands describing the properties of mock red quasars. The panels show the transmission efficiencies for the broad-band optical DECam *griz* filters, the *Euclid* VIS (optical) and NISP (NIR) filters, and the NIR to MIR VISTA *JHK_s* filters. The vertical dashed lines are placed at the weighted central wavelength of each filter.

best match. For this task we use the Phosphoros package (Paltani et al., in prep). Phosphoros is a fully Bayesian TF algorithm, supporting flexible prior distributions across all parameters (redshift, reddening, SED-index and luminosity) and producing multi-dimensional and marginalized posterior distributions. It was successfully employed in the *Euclid* photo-*z* challenge, presented in *Euclid* Collaboration: Desprez et al. (2020), which was designed to evaluate the accuracy of various methods for photometric redshift estimation against the stringent requirements of *Euclid* cosmic shear analyses (Cropper et al. 2013, Tarisano et al., in prep.). In Desprez et al. (2023), Phosphoros was validated and benchmarked against a similar code, Le Phare (Arnouts & Ilbert 2011). For each input galaxy, Phosphoros provides a multivariate posterior distribution, allowing the inference of flux estimates in the missing bands from the best fit. For additional details about its metrics and models we refer the reader to Desprez et al. (2023) and *Euclid* Collaboration: Tucci et al. (2025b).

2.2. Mock red QSO catalogue

A sample of SDSS QSOs was selected across redshift and grouped into nine bins of FWHM and equivalent width. Stitching together the stacked spectra, we created a composite spectrum with very broad wavelength coverage.

Figure 2 shows the unobscured QSO spectrum and the dramatic impact of $E(B - V) = 0.25$ attenuation applied to it. The coloured bars correspond to the rest-frame wavelength coverage of the red-grism of *Euclid*. To create a mock catalogue of red QSOs, we used the first bin of the stacked QSO spectra of *Euclid* Collaboration: Lusso et al. (2024) and the luminosity function of red QSO determined in Glikman et al. (2018). The latter is described by a double power-law function, already presented in Lacy et al. (2015), characterised by a faint-end and a right-end slope, and a break luminosity, where the dominance shifts from the faint to the bright end.

We created a grid of bolometric luminosity ($40 < \log_{10} L_{\text{bol}} < 50$) and redshift ($0 < z < 8$), and calculated the expected number of red QSOs by integrating the luminosity func-

tion. For each mock SED, we applied reddening according to their distribution covering $0.25 < E(B - V) < 1.45$, and assuming the Prevot attenuation law and intergalactic medium attenuation as described in *Euclid* Collaboration: Lusso et al. (2024).

Finally, we applied an observed magnitude cut corresponding to the expected depth of EWS, i.e. $J < 24.5$. Figure 3 shows the coverage of the luminosity-redshift plane of our mock catalogue, assuming $14\,500 \text{ deg}^2$ sky coverage, at the wide-depth of *Euclid*. The black line shows the break luminosity of red QSOs from Glikman et al. (2018).

3. Methods

Colour-based selection criteria have been extensively studied in literature to identify AGN and reddened AGN. Most notably, Lacy et al. (2004), Stern et al. (2005), Mateos et al. (2012) and Assef et al. (2018) defined empirical cuts in MIR colour space. Alongside MIR-based selections, NIR colours have been proposed in Banerji et al. (2012) and Glikman et al. (2012, 2013, 2018), to separate red quasars from stars and galaxies.

To follow-up on these studies exploiting the unique depth and high-resolution of *Euclid*, we introduce a novel selection method that is solely based on *Euclid* NIR photometry. Our goal is to enhance the systematic identification of red QSOs in the EWS where MIR photometry may be incomplete or unavailable.

To achieve this, we conduct a series of statistical analyses to study the photometric selection criteria that maximise the distance between the COSMOS2020 observed dataset and the mock red QSOs. In particular, we explore the separation in a multi-dimensional colour-colour space, as a function of magnitude and redshift. Our study is performed independently using VISTA+DECam and reconstructed *Euclid* photometry. This allows us to assess the impact of different wavelength coverage and filter sets on the identification of red QSOs, and to evaluate the *Euclid* NIR stand-alone capability in recovering our target population.

First, we collect features that are directly transferable into the reconstructed *Euclid*-like photometry. More precisely, we consider VISTA $J - K_s$, $Y - K_s$ and $J - H$ to describe NIR colours, and DECam $i - K_s$ to sharpen the selection of reddened

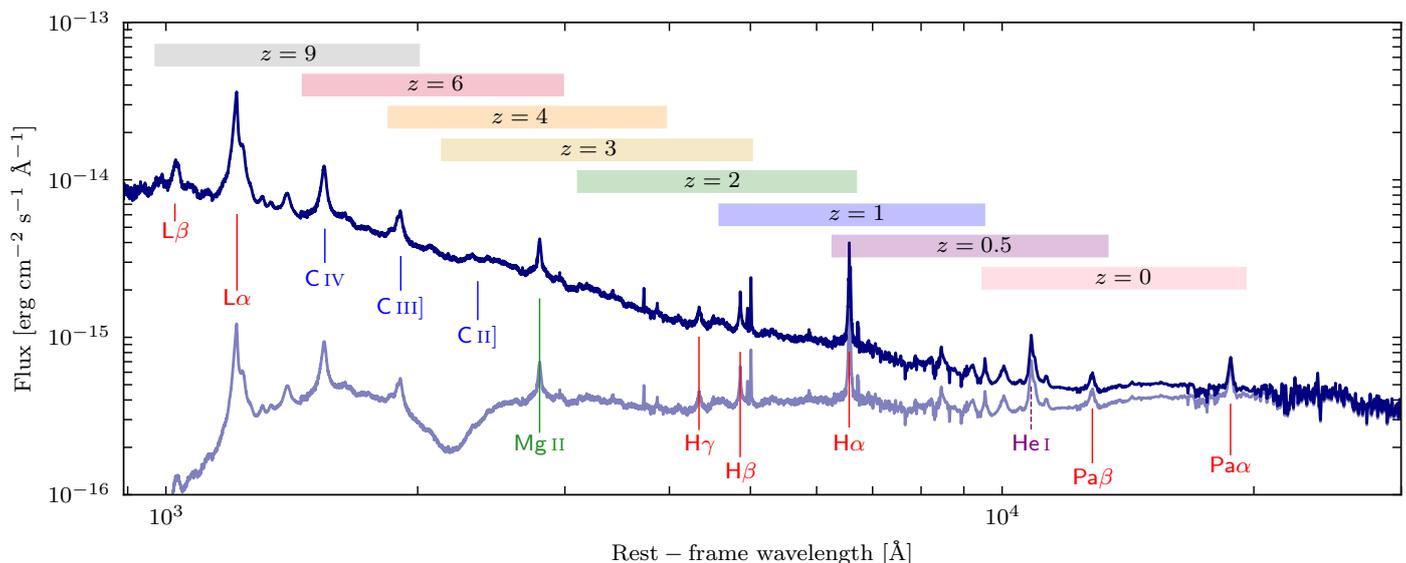


Fig. 2. Spectrum comparison between the unobscured QSO stack (dark blue line; Euclid Collaboration: Lusso et al. 2024), and a reddened QSO spectrum with $E(B - V) = 0.25$ (bright blue line). Emission lines are highlighted using coloured vertical markers along with their respective labels, pointing to different atomic species and ionization states. The coloured bars correspond to the redshift range of the red-grism of NISP.

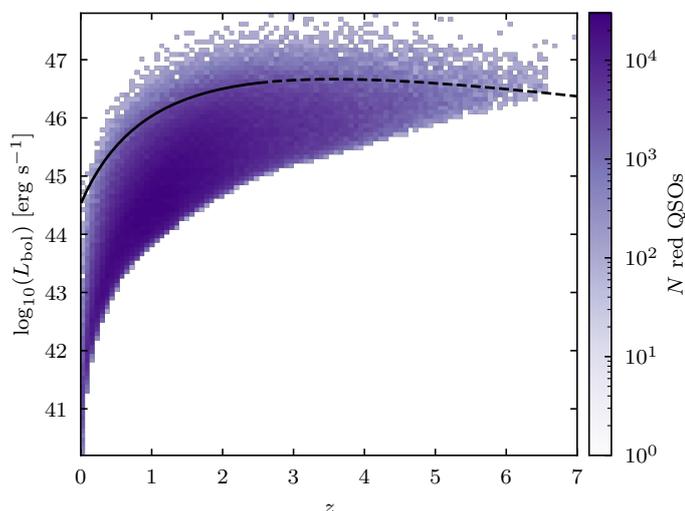


Fig. 3. Luminosity-redshift plane for the mock sample of red QSO. The black line is the knee of the luminosity function of Glikman et al. (2018).

sources based on their optical-to-NIR transition. In the *Euclid*-like colour space, we use $Y_E - H_E$, $J_E - H_E$, and $Y_E - J_E$ as NIR colours, and we adopt $I_E - H_E$ as a metric to estimate the optical-to-NIR excess. We proceed with the implementation of a multi-step analysis:

i Principal Component Analysis (PCA): this method has been successfully applied in previous studies to investigate the underlying physical properties of AGN and their host galaxies. Boroson & Green (1992) were the first to use it in the AGN domain, to analyse optical emission-lines and continuum properties of a low-redshift quasar sample. Their work was followed by Corbin (1993), Laor et al. (1994, 1997), Wang et al. (1996), and Brandt & Boller (1998). Their analyses found correlation between the primary eigenvector (or principal component) and quasar spectral features depending on physical parameters including Eddington ratio, luminosity and black hole spin. Beyond optical emission lines, PCA

has been applied to AGN spectral energy distributions and photometric datasets. Yip et al. (2004) applied PCA on SDSS quasar spectra, and showed that the eigenvectors (named *eigenspectra* in their work) have the power of disentangling the contribution of the host galaxy light, the optical continuum and the AGN emission. Kuraszkievicz et al. (2009) analysed a sample of red 2MASS AGN (Cutri et al. 2002). Among their results, they found that the second principal component was correlated with optical-to-infrared colours ($B - K_s$, $B - R$, $J - K_s$), depending on the contribution of the host galaxy relative to the AGN emission. PCA was used to study AGN samples also in Hao et al. (2005), Stern & Laor (2012), Panda & Śniegowska (2024).

In this work, we apply PCA to the aforementioned multi-dimensional colour space to identify the most informative colours that separate mock red QSOs from the observed COSMOS2020 sources. Our analysis focuses on broadband photometric selection in preparation for systematic large-scale red QSO searches in the *Euclid* Survey. Through linear combination of the original features, PCA reduces dimensionality while capturing variance in the dataset. It serves as an exploratory framework to highlight which optical and NIR colours contribute the most to the identification of red QSOs. Detailed information follows in Sect. 3.1.

ii Empirical colour-colour cuts: using the most significant colours identified via PCA, we study a colour-colour selection function for red QSOs. Additional information on the metric adopted to evaluate the selection performance is reported in Sect. 3.2.

iii Machine learning-based refinement: in this phase, we train a probabilistic Random Forest classifier (RF, Breiman 2001) to refine the previous selection function for red QSOs and mitigate the effects of contaminants. RF has the advantage of handling non-linear relationships between the input features, so it sets complex decision boundaries that PCA and empirical colour-based cuts cannot capture. The trained RF model is then applied to the *Euclid* Q1 dataset, where we select candidate red QSOs based on their predicted probabilities. We refer the readers to Sect. 3.3 for a detailed description of this method.

In the analysis, we consider the discriminating power of additional features, such as compactness criteria, and we use external AGN datasets as control samples. The latter play a crucial role in identifying the proposed selection function and assessing its robustness, or revealing the risk of introducing a bias that could reduce the completeness of the red quasar selection. We use our findings to build a first census of candidate red QSOs in *Euclid* Q1 EDF-F. Such a sample provides a framework for future spectral analyses and will serve as a training set for Artificial Intelligence-based automated classification, extended to the EDF-N, EDF-S, and the EWS.

3.1. Hyper-colour determination and colour selection

Principal Component Analysis (PCA) is a statistical method that can be used to project a dataset from a high-dimensional space into a low-dimensional space, retaining its most meaningful properties. More precisely, the technique consists of mapping the original features into a new set of uncorrelated ones, named principal components. The result is a linear combination where each coefficient represents the contribution of the corresponding original feature in forming a principal component. With this technique, the dataset is linearly transformed onto a new coordinate system whose directions, the principal components, capture as much variance in the data as possible, with the first component capturing the most variance, the second capturing the next most, and so on.

In this work, we use the module PCA available in the *Python* library *Scikit-learn* (Pedregosa et al. 2011), which follows the implementation presented in Halko et al. (2009). Applying PCA decomposition to our multi-dimensional dataset, described by the aforementioned colour features, means collapsing it into a low-dimensional one, where each principal component (PC) is a linear combination of the original colours (c) as:

$$PC = \sum_{j=1}^N a_j c_j, \quad (1)$$

with a_j the coefficients of the linear combination and N the number of involved colours. Features with consistently high coefficients across components explain a larger portion of the variance and are more important in the transformed space. The aim is to identify principal components receiving significant contributions by a set of original colours, and study their discriminating power between red QSOs and the rest of the sample. We will refer to such principal components as *hyper-colours* (HC).

We run PCA on the mocks and COSMOS2020, standardising the input features to ensure comparability across different scales. By assuming three principal components, and we identify a cut in the HC space which guarantees the highest discriminating power. We will refer to this cut as hyper-colours cut (HP-cut).

Furthermore, we study the impact that the single colour features, c_j , have on the HC, based on their linear coefficients a_j , and we identify a cut in the multi-dimensional colour-colours space made by them. We will refer to this selection cut as colour-colour cut (cc-cut).

3.2. Forecast analysis

By applying the HP- or cc-cut, we estimate the completeness (C) and purity (P) of the selected sample. Completeness is the fraction of mock red quasars correctly identified by the proposed selection criterion and is defined as:

$$C = \frac{TP}{TP + FN}, \quad (2)$$

where TP (True Positives) and FN (False Negatives) are the number of red quasars correctly identified and missed by the cut. The sum $TP + FN$ then corresponds to the total of red quasars in the mock sample. Purity is defined as the fraction of TP among all selected objects. We calculate it as:

$$P = \frac{TP}{TP + FP}, \quad (3)$$

where FP (False Positives) is the number of sources incorrectly identified as red quasars by the selection cut. In this work, we define FP as the number of COSMOS2020 selected sources which are not classified as AGN, plus the number of AGN candidates passing the cut with $DEC_{cam} i - K_s < 1.7$. We assume this metric since 1.7 is the lower limit for optical-to-NIR excess in the mock sample. Finally, we apply to EDF-F the colour-based selection functions, and we study them with the aid of two control samples of MIR-selected and radio-selected AGN candidates in the same field.

3.3. Probabilistic Random Forest

A Random Forest is a machine learning algorithm that consists of creating an ensemble of decision trees and combines their outputs to make predictions. Each tree in the forest is trained on a random subset of the data and features, and the final classification is made by majority voting (standard RF), or by averaging the predicted probabilities assigned to each class across all the decision trees (probabilistic RF). In a binary classification case, probabilistic RF assigns each source two values, corresponding to the probability of belonging to each of the two classes. This approach allows us to estimate the confidence level of each classification and make a probabilistic selection of candidate red QSOs. RF can identify and rank the most important features that differentiate red quasars from other objects, and it can handle complex and non-linear relationships between features. Furthermore, RF is more robust towards over-fitting, which makes it suitable for noisy or imbalanced datasets like in the case of targeting red QSOs (Breiman 2001; Chen et al. 2004). In our work, we used the module *RandomForestClassifier* available in the *Python* library *Scikit-learn*.

First, we train a RF classifier on three sets of features (named S1, S2 and S3) based on different combinations of *Euclid*-based colours and magnitudes. More precisely, S1 includes the most significant colours identified by PCA and S2 all the *Euclid* NIR colours. S3 adds *Euclid* magnitudes on S2. This multi-steps approach allows us to test the impact of expanding the primary feature set in terms of purity, completeness and classification performance.

For each set of features we identify the best model using hyper-parameter tuning. The RF model operates within an hyper-parameter space described by a set of key parameters, whose combinations can impact the performance of the model itself. These parameters include the number of trees, the minimum amount of samples required to split a tree node and the leaf size. We used the *Scikit-learn* module *RandomizedSearchCV* to explore this hyper-parameter space and identify the best performing set of key parameters. For each sampled combination of parameters, the algorithm employs a stratified k-fold cross-validation strategy with $k = 5$ folds. This technique divides the

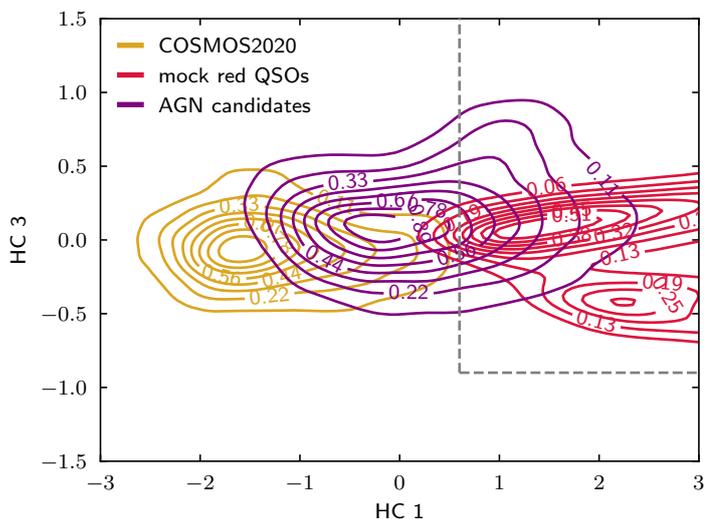


Fig. 4. Hyper colour-colour diagram showing the separation of mock red QSOs from COSMOS 2020 sources. The contours are normalised to their respective subsets. The proposed selection function is displayed by the gray dashed lines.

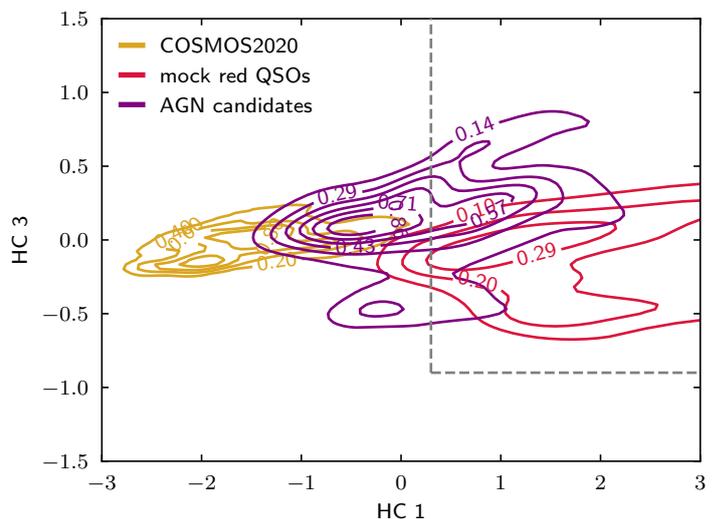


Fig. 5. Hyper colour-colour diagram obtained using *Euclid*-like photometry, showing the separation of mock red QSOs from COSMOS 2020 sources. The proposed selection function is displayed by the gray dashed lines.

430 dataset into five partitions and trains the model on four of them, using the fifth for validation. The goal of this approach is to get a robust estimate of the model performance and check if overfitting occurs to certain partitions. The best combination of parameters is selected based on the cross-validation accuracy.

435 Through hyper-parameter tuning and cross-validation (run with a fixed random seed for reproducibility), we obtain an optimised RF model for each initial set of features (S1, S2 and S3). We compare the three models in terms of feature importance, completeness and purity, and we apply the best one to the *Euclid* Q1 EDF-F dataset. We discuss the results and our findings in Sect. 4.

4. Results

In this section, we present the findings we obtained for the steps described in Sect. 3.

4.1. Selection function in the hyper-colour space

445 The PCA on mocks and COSMOS2020, run in VISTA-defined multidimensional colour space, provides three HC whose coefficients are reported in Table A.2. HC1 is a weighted average of all the input colour features, with a slight emphasis on $J - K_s$ and $i - K_s$. HC2 and HC3 are dominated by $H - K_s$ and $i - K_s$. Their explained variances (91%, 6.2% and 2.7%, respectively) indicate that HC1 is related to the overall colour gradients across the feature set, while HC2 and HC3 isolate information specific to a certain feature. The combination of HC1 and HC3 provides the strongest discriminating power, as displayed in Fig. 4. The contours, normalised to their respective subsets, show that this HC space is able to effectively disentangle the populations of observed sources and mock red QSOs. The HC-cut of $HC1 > 0.6$ and $HC3 > -0.9$ leads to an overall completeness of 98% with 81% purity. Among the AGN candidates passing the cut, 91% of them have DECam $i - K_s > 1.7$. As reported in Sect. 3, based on mocks we consider this as a requirement to identify possible obscured AGNs and flag them as candidate red QSOs.

460 We run again the PCA, transferring the original VISTA colour features into the *Euclid*-like colour space, yielding to sim-

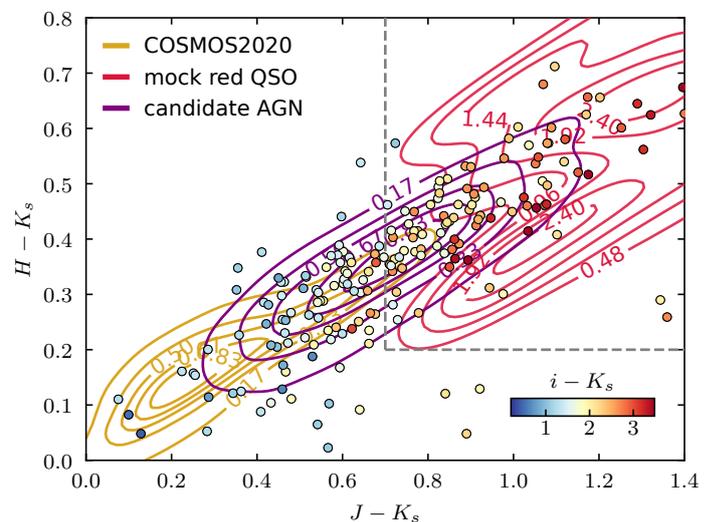


Fig. 6. The VISTA colour-colour diagram showing the separation of mock red QSOs from COSMOS 2020. The proposed colour-colour selection function is displayed by the gray dashed lines. A subset of candidate AGN overlays the selection, colour-coded as $i - K_s$.

470 ilar results. The coefficients of the principal components are displayed in Table A.3. Analogously to the VISTA-based HC, the *Euclid*-based HC1 is dominated by $Y_E - H_E$ and HC3 gets most contribution from $I_E - H_E$, tracking the optical-to-NIR transition. Figure 5 displays the *Euclid*-like HC space formed by HC1 and HC3. An HC-cut of $HC1 > 0.3$ and $HC3 > -0.9$ leads to an overall completeness of 97% with 68% purity.

475 Beyond unveiling hyper-colours, PCA highlights the original colour features, c_j in Eq. (1), bringing the most weight in the identification of candidate red QSOs. More precisely, in the VISTA parameter space, the most important colours for HC1, HC2 and HC3, based on their linear coefficients, are $J - K_s$, $H - K_s$ and $i - K_s$. In the *Euclid*-like parameter space, the most important features for the tree components are $Y_E - H_E$, $J_E - H_E$ and $I_E - H_E$, respectively. We further advance our analysis by a visual and quantitative assessment of the separation of mock red QSOs using these multidimensional colour-colour spaces.

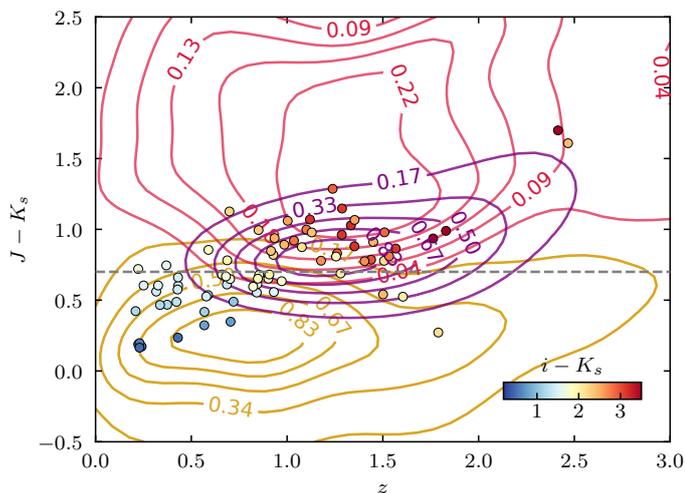


Fig. 7. The VISTA colour- z diagram showing the separation of mock red QSOs from COSMOS2020. The colour-code adopted for displayed populations are as in Fig. 6.

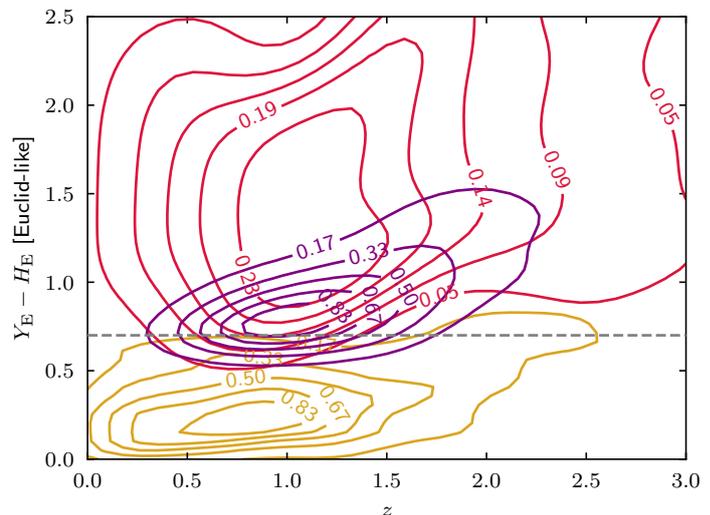


Fig. 9. The *Euclid*-like colour-redshift diagram showing the separation of mock red QSOs (in red) from COSMOS2020 (golden contours). The COSMOS2020 passing the proposed cc-cut is displayed in purple.

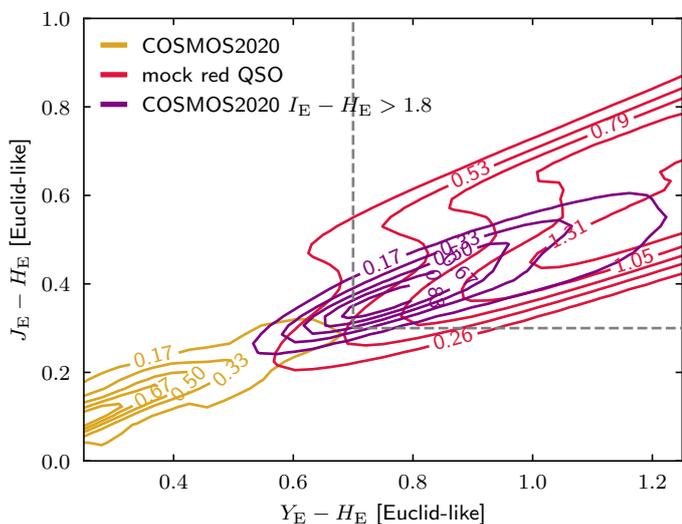


Fig. 8. The *Euclid*-like colour-colour diagram showing the separation of mock red QSOs from COSMOS2020. The proposed colour-colour selection function is displayed by the gray dashed lines. The COSMOS2020 passing the selection is displayed in purple.

4.2. Photometric selection in colour-colour space

Figure 6 displays the separation between mock red QSOs and the COSMOS2020 dataset achieved in a multidimensional parameter space spanning from optical to NIR wavelengths and defined through the most important colour features according to PCA. The two populations are compared in the $J - K_s$ vs $H - K_s$ diagram, and COSMOS2020 is further split in two subsets to highlight candidate AGN. If we apply the cut $J - K_s > 0.8$ and $H - K_s > 0.2$, we obtain an overall completeness of 99% and a purity of 78%. The 85% of the candidate AGN passing this multidimensional colour cut responds to the mock-calibrated red excess of DECam $i - K_s > 1.7$. As the cc-cut is defined through optical and NIR colours, the estimate of purity can be interested by dependences on redshift and magnitude. We first estimate it as a function of redshift, identifying intervals with major degeneracies between the populations at play (Fig. 7). We obtain 90% purity at $z < 0.5$, 69% at $0.5 < z < 1.5$, and 87% at higher redshifts. The magnitude dependency sees purity values of 77% for

$H < 20$, 65% for $20 < H < 22$ and 86% for objects in the interval $22 < H < 23.5$.

We proceed with the forecast analysis using the same metric, but defined with *Euclid*-like photometry. We study a grid of multidimensional cc-cuts to maximise the separation between the subsets at play, and we weight the FP rate with the unreddened COSMOS2020 AGN candidates. The results are displayed in Fig. 8, with the colour-redshift evolution shown in Fig. 9. In this case we find that the colour cut, $Y_E - H_E > 0.7$ and $J_E - H_E > 0.3$ with $I_E - H_E > 1.8$, leads to an overall 99% completeness and 67% purity, with redshift-dependent fluctuations: 86% purity at $z < 0.5$, 57% and 63% at $0.5 < z < 1.5$ and higher redshifts, respectively. Purity values with magnitudes are 56% for $H_E < 20$, 50% for $20 < H_E < 22$ and 77% for objects in the interval $22 < H < 23.5$.

4.3. Colour-colour selection applied to *Euclid* Q1

Having established the methods for the selection of red quasars based on mock and observed training samples in the previous sections, we now apply these methods to the *Euclid* Q1 EDF-F dataset. The selection is supported by control samples, MIR-selected and radio-selected AGN candidates, introduced in Sect. 2. Figure 10 displays the discriminating power of the proposed multidimensional colour-colour cut, yielding to an overall 98% completeness of the mock sample. The colour-redshift evolution is displayed in Fig. 11. The MIR-selected and radio-selected AGN have a percentage of 77% and 47% reddened sources, according of the proposed cut $I_E - H_E > 1.8$. Among them, the 87% passes the multidimensional cc-cut. The selection functions are summarized in Table 2. Combining the cc-cut with the RF classification probability (described in Sect. 3) we flag 151 853 sources as candidate red QSOs.

4.4. Random Forest analysis

We apply a probabilistic RF classifier to identify candidate red quasars starting from a first set of photometric features, S1, including the most significant colours according to PCA: $J_E - H_E$, $Y_E - H_E$ and $I_E - H_E$. The hyper-parameter search lead to an optimised model with 100 trees, minimum split size of 10 and min-

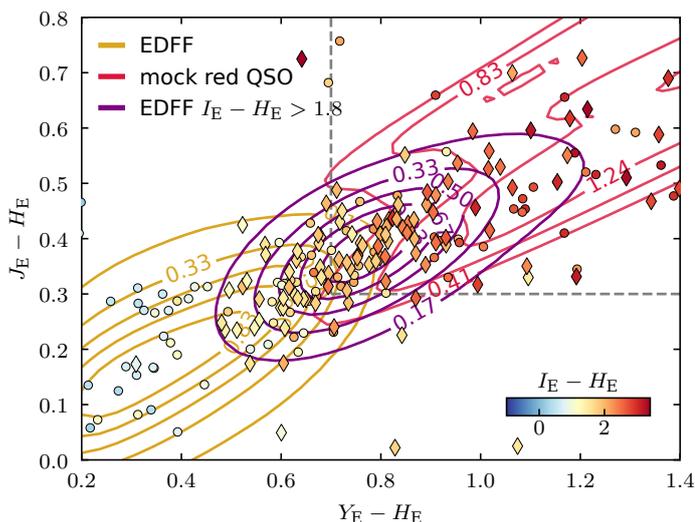


Fig. 10. The colour-colour diagram showing the separation of mock red QSOs from the *Euclid* Q1 sample. The proposed colour-colour selection is displayed by the grey dashed lines. The control samples are colour-coded by $I_E - H_E$: dots and diamonds represent MIR-selected and radio-selected AGN, respectively.

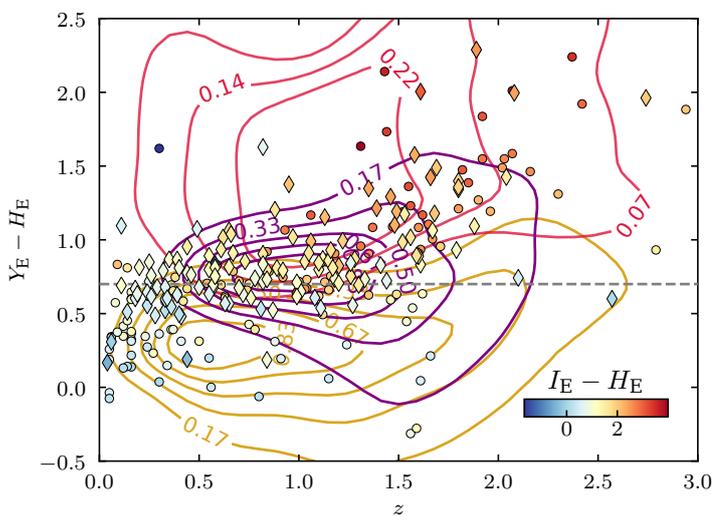


Fig. 11. The colour- z diagram showing the separation of mock red QSOs from the *Euclid* Q1 sample. The proposed colour-colour selection is displayed by the grey dashed line. The control samples are colour-coded by $I_E - H_E$: dots and diamonds represent the MIR-selected and the radio-selected AGN, respectively. The contours are coloured as in Fig. 10.

Table 2. Forecast red QSOs Completeness (C) and Purity (P) from NIR selection functions. The first two rows report the selection function defined in the VISTA colour space, the last two refer to the reconstructed *Euclid* photometry.

Candidate red QSO selection function	C	P
$HC1 > 0.6$ and $HC3 > -0.9$	0.99	0.82
$J - K_s > 0.7$, $H - K_s > 0.3$, and $i - K_s > 1.7$	0.99	0.79
$HC1 > 0.3$ and $HC3 > -0.9$ [<i>Euclid</i> -like]	0.97	0.78
$Y_E - H_E > 0.7$, $J_E - H_E > 0.3$, and $I_E - H_E > 1.8$	0.98	0.80

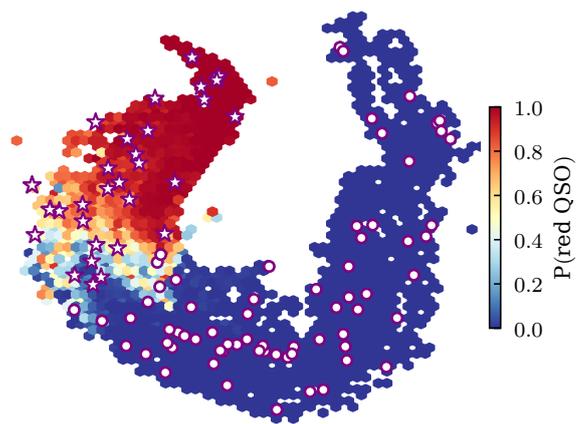


Fig. 12. UMAP visualisation for the classification results of the probabilistic RF run on the most significant colour features. Hexagons are colour-coded by the probability of being a candidate red quasar. The overlaid scatter plot displays a random subset with symbols reflecting the empirical cc-cut. Stars represent objects previously classified as red QSOs. Circles are employed otherwise.

imum leaf size equal to 1. This configuration yields to a mean cross-validation accuracy of 95% with a standard deviation of $< 1\%$. We recover feature importance values of 0.35 for $J_E - H_E$, 0.39 for $Y_E - H_E$ and 0.26 for $I_E - H_E$. Applying a probability threshold of $P > 0.7$ for classification, we obtain 98% completeness and 87% purity. We determine the probability threshold so that the completeness does not fall below the value we achieved using empirical cc-cuts. This criterion ensures that the RF classification is conservative at least as much as the other selection methods while improving purity.

To gather further insights from our analysis, we employ the Uniform Manifold Approximation and Projection algorithm (UMAP). As described in McInnes et al. (2018), UMAP is a non-linear dimensionality reduction technique that preserves both the local and global structure of the data and highlights possible clusters and patterns. We use UMAP to visualise the dataset in a reduce-dimensionality space and check the separation between classes. The UMAP visualization of the classified objects (Fig. 12), colour-coded by the probability of being a red QSO, shows that a threshold of 0.7 outlines a boundary region between the two classes. One-hundred sources randomly drawn from the test set are plot over the map, with symbols following the empirical cc-cut: stars for sources classified as red quasars, circles for sources that did not pass the selection. The overlay displays a direct comparison between the empirical colour selection method and the RF classification. Previously selected sources are mostly present in the region with higher probability of being a red quasar, and a minority populates the transition region. The RF refinement of these boundaries goes beyond the level of accuracy achieved by the empirical cc-cut, thereby reducing contamination and raising purity.

We repeat the analysis on extended sets of features, specifically on S2 and S3, in order to assess the impact of additional information on the classification performance. More precisely, S2 includes $Y_E - J_E$, $J_E - H_E$, $Y_E - H_E$ and $I_E - H_E$, while S3 expands S2 with the magnitudes I_E , Y_E , J_E and H_E . The RF classifier optimised for S2 yields to similar results, without improving completeness and purity. The feature importance analysis assigns to the additional colour, $Y_E - J_E$, a value of 0.03. Such results aligns with our findings from S3. In this case, the importance of the

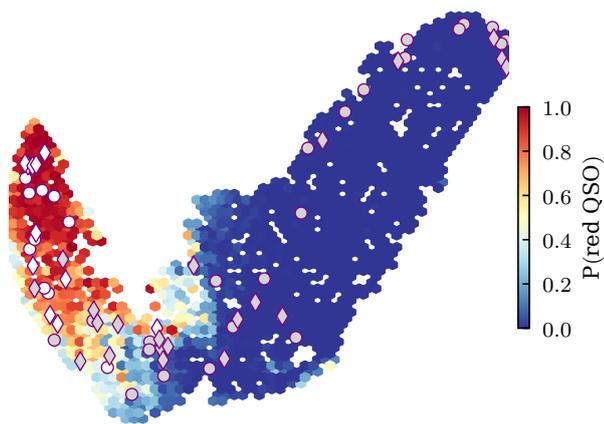


Fig. 13. UMAP visualisation of the *Euclid* Q1 dataset, colour-coded by the RF-based probability of being a red QSO. The overlaid scatter plot displays the 25% of the validation samples. MIR and radio-selected candidate AGN are marked by circles and diamonds, respectively. The symbols are filled in white if the source passed the empirical cc-cut, otherwise they are coloured in light gray.

magnitude features is $< 2\%$ and no increase in completeness and purity is registered.

Along this multi-step analysis, the dominant importance of the PCA-selected colours remains unchanged and leads to a consolidated probabilistic RF model that we apply to the *Euclid* Q1 EDF-F sample.

The UMAP visualisation of the *Euclid* Q1 sample, displayed in Fig. 13, is colour-coded by the probability of being a red QSO. The over-plotted symbols represent a fraction (25%) of the two control samples: circles denote the MIR-selected AGN candidates, while diamonds represent the radio-selected AGN candidates. White markers indicate the sources that passed the previous empirical cc-cut selection, and light gray colours those that were not classified as red quasars. Analogously to our findings from the analysis of the COSMOS2020 dataset, we notice that spatial distribution of the sources clusters around two different populations, according to their likelihood of being red QSOs, and that the boundaries obtained with RF are more robust against contaminants, compared to the empirical cc-cuts.

5. Discussion

The VISTA-based HC spaces provide effective separation between mock red QSOs and observed sources. Their translation into the *Euclid*-like HC space preserves the structure of the primary component (HC1), while introducing small shifts in the secondary components (HC2 and HC3). These shifts are expected due to the differences in filter characteristics and photometric uncertainties. The photometric selection function, using multi-dimensional colour cuts in the *Euclid* optical and NIR regimes, forecasts high completeness (98%) and moderate purity (78%) for candidate red QSOs. Completeness of the mocks remains consistently high across redshift and magnitude intervals, suggesting that the proposed selection effectively captures the reddened QSO population. Purity decreases in intermediate redshift ranges and at faint magnitudes, where physical degeneracies between red QSOs and red galaxies become more significant. However, comparing redshift and magnitude distributions of the mock red QSOs and the sources in COSMOS2020 suggests there is also an observational effect at play. Figure A.1 shows that the redshift distribution of mock red QSOs peaks

at higher values compared to the observed sources, as they are modelled to represent a deeper survey aiming to capture the properties of obscured AGN. Purity estimates can be affected in the mismatched higher-redshift range. The lower purity at intermediate redshift could be partially mitigated by training the selection function on larger, more representative datasets that include additional sources of variation, such as differences in host galaxy properties or environmental factors. Future work could explore the inclusion of environmental parameters, such as local density or clustering, to refine selection criteria. Larger values of purity for bright objects is also influenced by the broader photometric extent of *Euclid*-like mocks, which outnumber COSMOS2020 at lower magnitudes (Fig. A.2). On the opposite side, fainter sources are more prone to contamination by non-AGN populations due to increasing photometric uncertainties, an aspect that potentially reduces purity at fainter magnitudes.

In our analysis, we used a red QSO mock catalogue based on the stacking of SDSS quasar spectra with varying levels of reddening applied. While this approach effectively models the intrinsic quasar properties, it does not account for the host galaxy contamination. The host galaxy flux can dominate over the AGN emission in the optical and near-infrared bands, especially for lower-redshift sources. The consequent dilution of quasar colours may reduce the effectiveness of our HC selection criteria, which are primarily optimised for quasar-dominated SEDs. Such aspects set the seed for future works, where our mock catalogue incorporates realistic galaxy templates, either by adding scaled galaxy SEDs to the QSO spectra or by using simulations informed by empirical measurements of quasar host properties. Nevertheless, our analysis still provides valuable insights into the potential of *Euclid* for identifying red quasars. As reported in Sect. 4, we also note that our results remain robust for brighter sources ($H < 22$), aligned closer to SDSS, where the impact of host galaxies is expected to be minimal.

The probabilistic RF classifier refines the empirical cc-cut by achieving a higher completeness and purity (98% and 88%, respectively). This improvement is due to the RF ability to integrate non-linear relationships in the multi-dimensional feature space, which are not detected with simpler colour-colour cuts.

A key aspect in our RF approach is to start from a set of features including the most significant colours, according to PCA, and then expand it with additional colours and magnitudes. Through feature importance analysis, this multi-step methodology allows us to assess how the classification performance is impacted by the information encoded in the additional features. Furthermore, such expansion did not lead to substantial improvements in completeness and purity, confirming the NIR multi-dimensional colour space ($J_E - H_E$, $Y_E - H_E$, and $I_E - H_E$) as the most informative and effective to classify out target population.

The UMAP visualisation of the datasets, colour-coded by the probability of being a red quasar, shows that the threshold $P > 0.7$ effectively segments the transition region between the two classes. In comparison with the overlaid validation datasets, it shows consistency and robustness against contaminants.

The visual inspection of the selected sources reveals that most candidates exhibit point-like or slightly extended profiles, consistent with AGN-dominated systems. The multi wavelength images in the first three panels of Fig. 14 show sources with bright NIR emission and a weaker component in the VIS band, consistent with significant optical light attenuation due to dust. These characteristics suggest that the sources are indeed candidate red QSOs. Also in this case we highlight the importance of this dataset for future morphological analyses, performing host

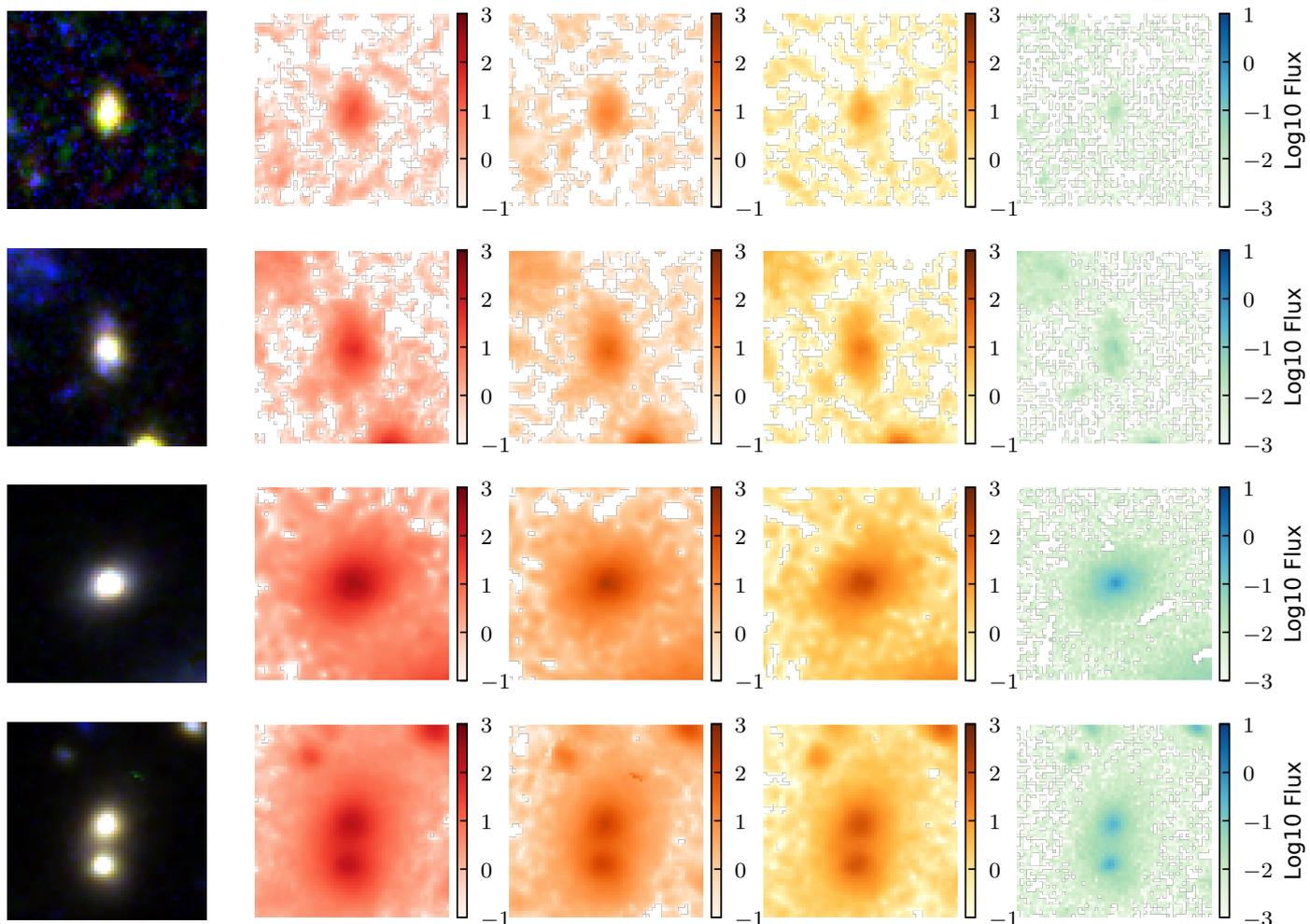


Fig. 14. Candidate sources. Panels show, from left to right, the RGB image and the H_E , J_E , Y_E and I_E bands. Raw images were obtained through ESA Datalabs (Navarro et al. 2024).

galaxy decomposition and adding information on the properties of the AGN component and its dusty cocoon.

In addition to examples of individual red QSOs candidates, in the fourth panel of Fig. 14 we introduce an example of candidate dual AGN system. The image displays two distinct compact sources with optical-to-NIR colour $I_E - H_E = 2.4$, at redshifts 0.86 (central source) and 0.9, classified as candidate red quasars with RF-based probability of 0.82 (central source) and 0.74. Their proximity consists in a projected distance less than 100 kpc and difference in redshift not surpassing 0.06. Such criteria aligns with the definition of dual AGN given in previous work (De Rosa et al. 2019). Such systems offer a unique observational window into AGN triggering mechanisms, galaxy mergers, and the evolution of supermassive black holes. The example presented in this work will be part of future systematic searches of dual AGN systems. This effort will use morphological analysis and spectroscopic follow-up where available. The results of this investigation will be presented in forthcoming publications.

We expand this first characterisation of the selected sources making a comparison between candidates selected through VISTA and DECam colours and via *Euclid*-only colours. In the EDF-F we identify a sample of 43528 objects observed by *Euclid*, the Vista Hemisphere Survey (VHS, McMahon et al. 2019) and DES. Among them, we select the 3% and the 4% using solely *Euclid* and VISTA+DECam colours, respectively.

As shown in Fig. 15, *Euclid*-only selected objects extend into redder $I_E - H_E$ values. This suggests that *Euclid* is better at identifying the reddest sources, which might be missed by VISTA. Furthermore, we observe that the candidates identified through the VISTA+DECam system display a broader distribution towards bluer colours. This can be explained by VISTA depth and resolution, both inferior to *Euclid*, leading to misclassification and missing reddened sources. This preliminary analysis suggests that *Euclid* better resolution and NIR sensitivity enables a more complete and robust identification of red QSOs.

The first panel of Fig. 16 shows an example of red QSO candidate selected with *Euclid* only. The image in the I_E band shows a compact source with a bright centre and some faint surrounding structure which can be attributed to the quasar host galaxy. The presence of asymmetry in the outer structure hints at a merger history or disturbed morphology. The *Euclid* H_E band image shows a smooth compact core with higher emission than in the optical. This proves that the optical-to-NIR contrast is high. The VISTA and DES images are noisier and the source is harder to distinguish. The poorer signal can explain why the VISTA+DECam system did not classify this source as a candidate red QSO. Images are normalised and in flux units.

We repeat the visual inspection on sources that were selected as candidate red QSOs through the VISTA+DECam system only. In the second panel of Fig. 16, the *Euclid* I_E band

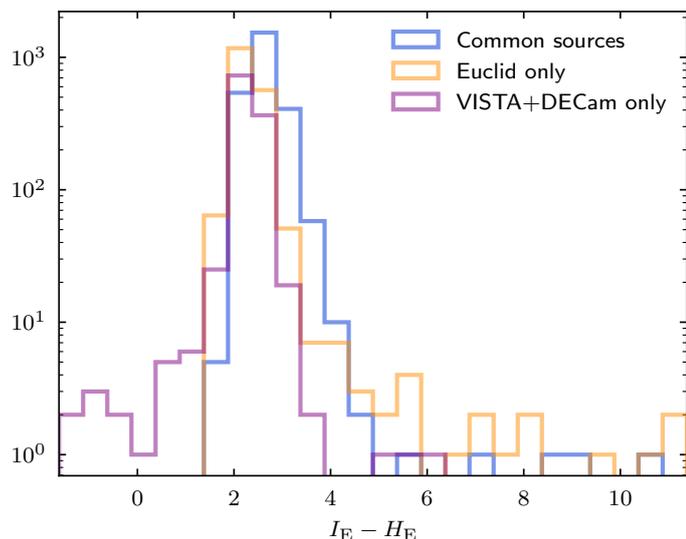


Fig. 15. Optical-to-NIR colour distributions for the EDF-F sample of candidate red QSOs observed through both *Euclid* and VISTA+DECam photometry. The orange and purple distributions are for candidates selected only via *Euclid* and VISTA+DECam colours, respectively.

displays an extended morphology with clear structures, possibly star-forming regions or satellite companions. In the DECam *i* band only the overall structure is less visible due to the lower resolution compared to *Euclid*. Similarly, in the VISTA bands the galaxy is less resolved and appears to be more concentrated in NIR. In this case, DECam and VISTA lower resolution makes the system look more compact, while *Euclid* sees a more extended morphology with detailed substructural features. Therefore, the VISTA+DECam colours used for the selection can be biased towards the integrated light of the system rather than its true morphology. This suggest that the *Euclid*-selected sample of red QSOs tends to be cleaner, avoiding that extended galaxies are misclassified as quasars. Additional examples are reported in the other panels of the figure.

In this work, we flag over 150 000 sources in the EDF-F as candidate red QSOs. This population serves as a first base for future refinements, incorporating additional diagnostics such as MIR data, spectral analyses (spectra are not considered in this work) and morphological compactness. These features should help to further improve purity without compromising completeness. In terms of number expectations, we refer to the studies by [Euclid Collaboration: Bisigello et al. \(2024\)](#) and [Euclid Collaboration: Selwood et al. \(2025\)](#), conducted in anticipation of the Q1 data release. The former predicts a fraction of 57% AGNs among NIR-selected sources in EDF-F. The latter estimates an obscured fraction of 26% in the expected AGN population. Applying these forecasts to our EDF-F sample, we calculate 196 992 expected obscured AGNs. This is compatible with our selected sample, which is set to include reddened AGNs, a contamination of red galaxies, and red QSOs. The classification of these sources will be the subject of future refinements based on the aforementioned criteria.

In this work, we excluded X-ray sources from the COSMOS2020 sample. However, their integration will be beneficial to future analyses, as their properties are directly linked to the degree of obscuration ([Ma et al. 2024](#)). This approach, which can be tested against the Q1 catalogue of X-ray AGN counterparts by [Euclid Collaboration: Roster et al. \(2025\)](#), can help improve

the distinction between truly reddened AGNs from red galaxies, thereby improving the purity of the selected sample.

6. Conclusions

In this work, we explored the capability of selection criteria based on *Euclid* optical and NIR photometry to identify and characterize red quasars. We explored the separation in a multi-dimensional colour-colour space, as a function of magnitude and redshift, and we compared it with selections based on VISTA+DECam photometry. Through a joint PCA and statistical analysis we identified selection functions achieving high completeness and moderate purity. Then, to refine the identification of red QSOs, we used a probabilistic RF classifier, which significantly improved purity (88%) and maintained high completeness (98%). This result reflects the capability of this method to identify complex, non-linear relationships in a multi-dimensional colour space, beyond the reach of simple empirical cuts. Additionally, visualization methods such as UMAP confirm a clear separation between red QSOs and contaminants, reinforcing the robustness of our selection criteria.

Our analysis shown that *Euclid*-based selection recovers redder quasar candidates than those identified by VISTA+DECam, as evidenced by shifts in the $I_E - H_E$ distribution. This suggests that *Euclid*'s superior depth and resolution enable a more complete detection of highly obscured sources, which might be misclassified or missed in lower-resolution surveys. Conversely, we found that some VISTA+DECam-selected sources appear as extended galaxies in *Euclid* images, indicating that lower-resolution surveys may introduce contamination by compact galaxies in the sample of red QSOs.

A key insight from future morphological inspection of the selected candidates will be to investigate the capability of *Euclid* to resolve host galaxy features in a subset of sources, distinguishing between true quasars and compact galaxies.

We used our findings to build a first census of candidate red QSO in *Euclid* Q1. The catalogue will be released as a *fits* table to the *Euclid* Collaboration. This work provides the foundation for systematic red QSO searches in the *Euclid* Wide Survey (EWS). The proposed selection function will be further refined with additional diagnostics, including mid-infrared data, spectroscopic follow-ups and the analysis of the host morphology. This first census of red QSO candidates in *Euclid* Q1 represents a significant step towards a more complete understanding of the dusty AGN population and its connection to galaxy evolution.

Acknowledgements. This work has made use of the *Euclid* Quick Release (Q1) data from the *Euclid* mission of the European Space Agency (ESA), 2025, <https://doi.org/10.57780/esa-2853f3b>. The *Euclid* Consortium acknowledges the European Space Agency and a number of agencies and institutes that have supported the development of *Euclid*, in particular the Agenzia Spaziale Italiana, the Austrian Forschungsförderungsgesellschaft funded through BMK, the Belgian Science Policy, the Canadian *Euclid* Consortium, the Deutsches Zentrum für Luft- und Raumfahrt, the DTU Space and the Niels Bohr Institute in Denmark, the French Centre National d'Etudes Spatiales, the Fundação para a Ciência e a Tecnologia, the Hungarian Academy of Sciences, the Ministerio de Ciencia, Innovación y Universidades, the National Aeronautics and Space Administration, the National Astronomical Observatory of Japan, the Nederlandse Onderzoekschool Voor Astronomie, the Norwegian Space Agency, the Research Council of Finland, the Romanian Space Agency, the State Secretariat for Education, Research, and Innovation (SERI) at the Swiss Space Office (SSO), and the United Kingdom Space Agency. A complete and detailed list is available on the *Euclid* web site (www.euclid-ec.org). This work has made use of CosmoHub, developed by PIC (maintained by IFAE and CIEMAT) in collaboration with ICE-CSIC. CosmoHub received funding from the Spanish government (MCIN/AEI/10.13039/501100011033), the EU NextGeneration/PRTR (PRTR-C17.11), and the Generalitat de Catalunya. This research makes use of ESA Datalabs (datalabs.esa.int), an initiative by ESA's

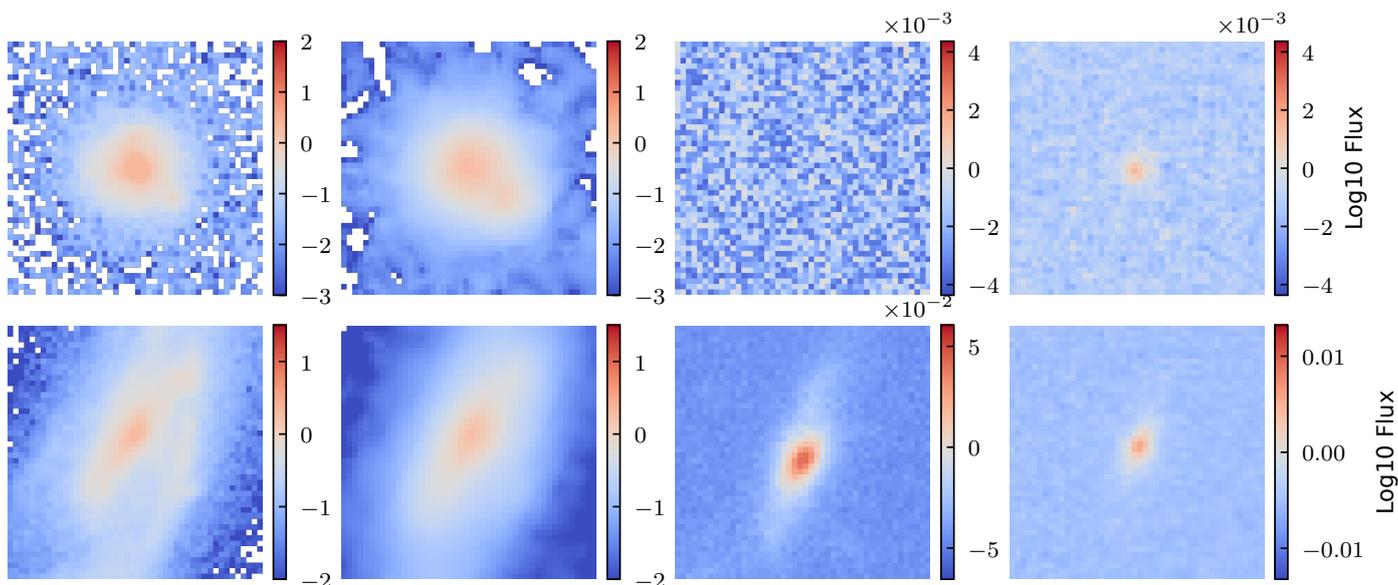


Fig. 16. Examples of two candidate red QSOs. The first and the second rows show an *Euclid* only and VISTA+DECam-only selected candidate, respectively. Panels from left to right display the object in the I_E and H_E filters, in DECam i and VISTA K_s .

Data Science and Archives Division in the Science and Operations Department, Directorate of Science. The VISTA Hemisphere Survey data products served at Astro Data Lab are based on observations collected at the European Organisation for Astronomical Research in the Southern Hemisphere under ESO programme 179.A-2010, and/or data products created thereof. This work has benefited from the support of Royal Society Research Grant RGS\R1\231450. This research was supported by the International Space Science Institute (ISSI) in Bern, through ISSI International Team project #23- 573 “Active Galactic Nuclei in Next Generation Surveys”. F. R., B. L. acknowledge the support from the INAF Large Grant “AGN & Euclid: a close entanglement” Ob. Fu. 01.05.23.01.14.

References

- 835 Andonie, C., Alexander, D. M., Rosario, D., et al. 2022, *MNRAS*, 517, 2577
- Antonucci, R. 1993, *ARA&A*, 31, 473
- 840 Arnouts, S. & Ilbert, O. 2011, LePHARE: Photometric Analysis for Redshift Estimate, Astrophysics Source Code Library, record ascl:1108.009
- Assef, R. J., Stern, D., Noirot, G., et al. 2018, *ApJS*, 234, 23
- 845 Banerji, M., Alaghband-Zadeh, S., Hewett, P. C., & McMahon, R. G. 2015, *MNRAS*, 447, 3368
- Banerji, M., McMahon, R. G., Hewett, P. C., et al. 2012, *MNRAS*, 427, 2275
- Banerji, M., McMahon, R. G., Hewett, P. C., Gonzalez-Solares, E., & Koposov, S. E. 2013, *MNRAS*, 429, L55
- Borson, T. A. & Green, R. F. 1992, *ApJS*, 80, 109
- 850 Bower, R. G., Benson, A. J., Malbon, R., et al. 2006, *MNRAS*, 370, 645
- Brandt, N. & Boller, T. 1998, *Astron. Nachr.*, 319, 163
- Breiman, L. 2001, *Mach. Learn.*, 45, 5–32
- 855 Calistro Rivera, G., Alexander, D. M., Harrison, C. M., et al. 2024, *A&A*, 691, A191
- 860 Calistro Rivera, G., Alexander, D. M., Rosario, D. J., et al. 2021, *A&A*, 649, A102
- Chen, C., Liaw, A., & Breiman, L. 2004, Using Random Forest to Learn Imbalanced Data, Technical Report 666, Department of Statistics, UC Berkley
- Civano, F., Marchesi, S., Comastri, A., et al. 2016, *ApJ*, 819, 62
- 865 Corbin, M. R. 1993, *ApJ*, 403, L9
- Cropper, M., Hoekstra, H., Kitching, T., et al. 2013, *MNRAS*, 431, 3103
- Croton, D. J. 2006, *MNRAS*, 369, 1808
- Cutri, R. M., Nelson, B. O., Francis, P. J., & Smith, P. S. 2002, in *Astronomical Society of the Pacific Conference Series*, Vol. 284, IAU Colloq. 184: AGN Surveys, ed. R. F. Green, E. Y. Khachikian, & D. B. Sanders, 127
- 870 Dark Energy Survey Collaboration, Abbott, T., Abdalla, F. B., et al. 2016, *MNRAS*, 460, 1270
- De Rosa, A., Vignali, C., Bogdanović, T., et al. 2019, *New A Rev.*, 86, 101525
- Desprez, G., Picouet, V., Moutard, T., et al. 2023, *A&A*, 670, A82
- 875 Donley, J. L., Koekemoer, A. M., Brusa, M., et al. 2012, *ApJ*, 748, 142
- Euclid Collaboration: Aussel, H., Tereno, I., Schirmer, M., et al. 2025, *A&A*, submitted
- Euclid Collaboration: Bisigello, Bisigello, L., Massimo, M., et al. 2024, *A&A*, 691, A1
- 880 Euclid Collaboration: Cropper, M., Al Bahlawan, A., Amiaux, J., et al. 2024, *A&A*, accepted, arXiv:2405.13492
- Euclid Collaboration: Desprez, G., Paltani, S., Coupon, J., et al. 2020, *A&A*, 644, A31
- 885 Euclid Collaboration: Jahnke, K., Gillard, W., Schirmer, M., et al. 2024, *A&A*, accepted, arXiv:2405.13493
- Euclid Collaboration: Lusso, E., Fotopoulou, S., Selwood, M., et al. 2024, *A&A*, 685, A108
- Euclid Collaboration: McCracken, H., Benson, K., et al. 2025, *A&A*, submitted
- 890 Euclid Collaboration: Mellier, Y., Abdurro'uf, Acevedo Barroso, J., et al. 2024, *A&A*, accepted, arXiv:2405.13491
- Euclid Collaboration: Polenta, G., Frailis, M., Alavi, A., et al. 2025, *A&A*, submitted
- 895 Euclid Collaboration: Romelli, E., Kümmel, M., Dole, H., et al. 2025, *A&A*, submitted
- Euclid Collaboration: Roster, W., Salvato, M., Buchner, J., et al. 2025, *A&A*, submitted
- 900 Euclid Collaboration: Schirmer, M., Jahnke, K., Seidel, G., et al. 2022, *A&A*, 662, A92
- Euclid Collaboration: Selwood, M., Fotopoulou, S., Bremer, M. N., et al. 2025, *A&A*, 693, A250
- 905 Euclid Collaboration: Tucci, M., Paltani, S., Hartley, W., et al. 2025a, *A&A*, submitted
- Euclid Collaboration: Tucci, M. et al. 2025b, *A&A*, in preparation
- 910 Euclid Quick Release Q1. 2025, <https://doi.org/10.57780/esa-2853f3b>
- Fabian, A. C. 2012, *ARA&A*, 50, 455
- 915 Fawcett, V. A., Alexander, D. M., Brodzeller, A., et al. 2023, *MNRAS*, 525, 5575
- Fotopoulou, S., Buchner, J., Georgantopoulos, I., et al. 2016, *A&A*, 587, A142
- Glikman, E., Lacy, M., LaMassa, S., et al. 2018, *ApJ*, 861, 37
- Glikman, E., Urrutia, T., Lacy, M., et al. 2012, *ApJ*, 757, 51
- 920 Glikman, E., Urrutia, T., Lacy, M., et al. 2013, *ApJ*, 778, 127
- Habouzit, M., Li, Y., Somerville, R. S., et al. 2021, *MNRAS*, 503, 1940
- Halko, N., Martinsson, P.-G., & Tropp, J. A. 2009, arXiv e-prints, arXiv:0909.4061
- Hao, L., Strauss, M. A., Tremonti, C. A., et al. 2005, *AJ*, 129, 1783
- 925 Heckman, T. M. & Best, P. N. 2014, *ARA&A*, 52, 589
- Klindt, L., Alexander, D. M., Rosario, D. J., Lusso, E., & Fotopoulou, S. 2019, *MNRAS*, 488, 3109
- Kuraszkiewicz, J., Wilkes, B. J., Schmidt, G., et al. 2009, *ApJ*, 692, 1180
- Lacy, M., Ridgway, S. E., Sajina, A., et al. 2015, *ApJ*, 802, 102
- 930 Lacy, M., Storrie-Lombardi, L. J., Sajina, A., et al. 2004, *ApJS*, 154, 166
- Laor, A., Fiore, F., Elvis, M., Wilkes, B. J., & McDowell, J. C. 1994, *ApJ*, 435, 611
- Laor, A., Fiore, F., Elvis, M., Wilkes, B. J., & McDowell, J. C. 1997, *ApJ*, 477, 93
- 935 Laureijs, R., Amiaux, J., Arduini, S., et al. 2011, *ESA/SRE(2011)12*, arXiv:1110.3193

- Ma, Y., Goulding, A., Greene, J. E., et al. 2024, *ApJ*, 974, 225
- Magorrian, J., Tremaine, S., Richstone, D., et al. 1998, *AJ*, 115, 2285
- Mateos, S., Alonso-Herrero, A., Carrera, F. J., et al. 2012, in *Half a Century of X-ray Astronomy*, ed. I. Georgantopoulos & M. Plionis, 148
- 930 McCracken, H. J., Milvang-Jensen, B., Dunlop, J., et al. 2012, *A&A*, 544, A156
- McInnes, L., Healy, J., & Melville, J. 2018, arXiv e-prints, arXiv:1802.03426
- McMahon, R., G., Banerji, G., et al. 2019, *VizieR Online Data Catalog: The VISTA Hemisphere Survey (VHS) catalog DR4.1 (McMahon+, 2013), VizieR Online Data Catalog: II/359*. Originally published in: 2013Msngr.154...35M
- 935 Miller, N. A., Bonzini, M., Fomalont, E. B., et al. 2013, *ApJS*, 205, 13
- Navarro, V., del Rio, S., Diego, M. A., et al. 2024, *ESA Datalabs: Digital Innovation in Space Science*, ed. A. Cortesi (Singapore: Springer Nature Singapore), 1–13
- 940 Padovani, P., Alexander, D. M., Assef, R. J., et al. 2017, *A&A Rev.*, 25, 2
- Panda, S. & Sniegowska, M. 2024, *ApJS*, 272, 13
- Pedregosa, F., Varoquaux, G., Gramfort, A., et al. 2011, *JMLR*, 12, 2825
- Petley, J. W., Morabito, L. K., Rankine, A. L., et al. 2024, *MNRAS*, 529, 1995
- 945 Silk, J. & Rees, M. J. 1998, *A&A*, 331, L1
- Stern, D., Eisenhardt, P., Gorjian, V., et al. 2005, *ApJ*, 631, 163
- Stern, J. & Laor, A. 2012, *MNRAS*, 423, 600
- Temple, M. J., Banerji, M., Hewett, P. C., et al. 2019, *MNRAS*, 487, 2594
- Urrutia, T., Becker, R. H., White, R. L., et al. 2009, *ApJ*, 698, 1095
- 950 Urry, C. M. & Padovani, P. 1995, *PASP*, 107, 803
- Wang, T., Brinkmann, W., & Bergeron, J. 1996, *A&A*, 309, 81
- Weaver, J. R., Kauffmann, O. B., Ilbert, O., et al. 2022, *ApJS*, 258, 11
- Wethers, C. F., Banerji, M., Hewett, P. C., et al. 2018, *MNRAS*, 475, 3682
- Wright, E. L., Eisenhardt, P. R. M., Mainzer, A. K., et al. 2010, *AJ*, 140, 1868
- 955 Yip, C. W., Connolly, A. J., Vanden Berk, D. E., et al. 2004, *AJ*, 128, 2603
- Yue, B. H., Best, P. N., Duncan, K. J., et al. 2024, *MNRAS*, 529, 3939
- Zou, F., Brandt, W. N., Chen, C.-T., et al. 2022, *ApJS*, 262, 15
-
- 1 Department of Astronomy, University of Geneva, ch. d'Ecogia 16, 1290 Versoix, Switzerland
- 2 School of Physics, HH Wills Physics Laboratory, University of Bristol, Tyndall Avenue, Bristol, BS8 1TL, UK
- 960 3 School of Physics & Astronomy, University of Southampton, Highfield Campus, Southampton SO17 1BJ, UK
- 4 Leiden Observatory, Leiden University, Einsteinweg 55, 2333 CC Leiden, The Netherlands
- 965 5 Caltech/IPAC, 1200 E. California Blvd., Pasadena, CA 91125, USA
- 6 Kavli Institute for Cosmology Cambridge, Madingley Road, Cambridge, CB3 0HA, UK
- 7 Department of Physical Sciences, Ritsumeikan University, Kusatsu, Shiga 525-8577, Japan
- 970 8 National Astronomical Observatory of Japan, 2-21-1 Osawa, Mitaka, Tokyo 181-8588, Japan
- 9 Academia Sinica Institute of Astronomy and Astrophysics (ASIAA), 11F of ASMA, No. 1, Section 4, Roosevelt Road, Taipei 10617, Taiwan
- 975 10 Department of Physics, Centre for Extragalactic Astronomy, Durham University, South Road, Durham, DH1 3LE, UK
- 11 Kapteyn Astronomical Institute, University of Groningen, PO Box 800, 9700 AV Groningen, The Netherlands
- 980 12 Faculdade de Ciências da Universidade do Porto, Rua do Campo de Alegre, 4150-007 Porto, Portugal
- 13 Instituto de Astrofísica e Ciências do Espaço, Universidade do Porto, CAUP, Rua das Estrelas, PT4150-762 Porto, Portugal
- 985 14 Institute for Astronomy, University of Edinburgh, Royal Observatory, Blackford Hill, Edinburgh EH9 3HJ, UK
- 15 Max Planck Institute for Extraterrestrial Physics, Giessenbachstr. 1, 85748 Garching, Germany
- 16 INAF-Osservatorio Astronomico di Capodimonte, Via Moiariello 16, 80131 Napoli, Italy
- 990 17 Department of Mathematics and Physics, Roma Tre University, Via della Vasca Navale 84, 00146 Rome, Italy
- 18 INAF-Osservatorio Astronomico di Roma, Via Frascati 33, 00078 Monteporzio Catone, Italy
- 995 19 Université Paris-Saclay, CNRS, Institut d'astrophysique spatiale, 91405, Orsay, France
- 20 ESAC/ESA, Camino Bajo del Castillo, s/n., Urb. Villafranca del Castillo, 28692 Villanueva de la Cañada, Madrid, Spain
- 21 School of Mathematics and Physics, University of Surrey, Guildford, Surrey, GU2 7XH, UK
- 22 INAF-Osservatorio Astronomico di Brera, Via Brera 28, 20122 Milano, Italy
- 23 INAF-Osservatorio di Astrofisica e Scienza dello Spazio di Bologna, Via Piero Gobetti 93/3, 40129 Bologna, Italy
- 24 Université Paris-Saclay, Université Paris Cité, CEA, CNRS, AIM, 91191, Gif-sur-Yvette, France
- 1005 25 IFPU, Institute for Fundamental Physics of the Universe, via Beirut 2, 34151 Trieste, Italy
- 26 INAF-Osservatorio Astronomico di Trieste, Via G. B. Tiepolo 11, 34143 Trieste, Italy
- 27 INFN, Sezione di Trieste, Via Valerio 2, 34127 Trieste TS, Italy
- 1010 28 SISSA, International School for Advanced Studies, Via Bonomea 265, 34136 Trieste TS, Italy
- 29 Dipartimento di Fisica e Astronomia, Università di Bologna, Via Gobetti 93/2, 40129 Bologna, Italy
- 30 INFN-Sezione di Bologna, Viale Berti Pichat 6/2, 40127 Bologna, Italy
- 1015 31 INAF-Osservatorio Astronomico di Padova, Via dell'Osservatorio 5, 35122 Padova, Italy
- 32 Space Science Data Center, Italian Space Agency, via del Politecnico snc, 00133 Roma, Italy
- 1020 33 Dipartimento di Fisica, Università di Genova, Via Dodecaneso 33, 16146, Genova, Italy
- 34 INFN-Sezione di Genova, Via Dodecaneso 33, 16146, Genova, Italy
- 35 Department of Physics "E. Pancini", University Federico II, Via Cinthia 6, 80126, Napoli, Italy
- 36 Dipartimento di Fisica, Università degli Studi di Torino, Via P. Giuria 1, 10125 Torino, Italy
- 37 INFN-Sezione di Torino, Via P. Giuria 1, 10125 Torino, Italy
- 1030 38 INAF-Osservatorio Astrofisico di Torino, Via Osservatorio 20, 10025 Pino Torinese (TO), Italy
- 39 European Space Agency/ESTEC, Keplerlaan 1, 2201 AZ Noordwijk, The Netherlands
- 40 Institute Lorentz, Leiden University, Niels Bohrweg 2, 2333 CA Leiden, The Netherlands
- 1035 41 INAF-IASF Milano, Via Alfonso Corti 12, 20133 Milano, Italy
- 42 Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas (CIEMAT), Avenida Complutense 40, 28040 Madrid, Spain
- 43 Port d'Informació Científica, Campus UAB, C. Albareda s/n, 1040 08193 Bellaterra (Barcelona), Spain
- 44 Institute for Theoretical Particle Physics and Cosmology (TTK), RWTH Aachen University, 52056 Aachen, Germany
- 45 INFN section of Naples, Via Cinthia 6, 80126, Napoli, Italy
- 46 Institute for Astronomy, University of Hawaii, 2680 Woodlawn Drive, Honolulu, HI 96822, USA
- 1045 47 Dipartimento di Fisica e Astronomia "Augusto Righi" - Alma Mater Studiorum Università di Bologna, Viale Berti Pichat 6/2, 40127 Bologna, Italy
- 48 Instituto de Astrofísica de Canarias, Vía Láctea, 38205 La Laguna, Tenerife, Spain
- 1050 49 Jodrell Bank Centre for Astrophysics, Department of Physics and Astronomy, University of Manchester, Oxford Road, Manchester M13 9PL, UK
- 50 European Space Agency/ESRIN, Largo Galileo Galilei 1, 00044 Frascati, Roma, Italy
- 1055 51 Université Claude Bernard Lyon 1, CNRS/IN2P3, IP2I Lyon, UMR 5822, Villeurbanne, F-69100, France
- 52 Aix-Marseille Université, CNRS, CNES, LAM, Marseille, France
- 53 Institut de Ciències del Cosmos (ICCUB), Universitat de Barcelona (IEEC-UB), Martí i Franquès 1, 08028 Barcelona, Spain
- 1060 54 Institució Catalana de Recerca i Estudis Avançats (ICREA), Passeig de Lluís Companys 23, 08010 Barcelona, Spain
- 55 UCB Lyon 1, CNRS/IN2P3, IUF, IP2I Lyon, 4 rue Enrico Fermi, 69622 Villeurbanne, France
- 1065 56 Mullard Space Science Laboratory, University College London, Holmbury St Mary, Dorking, Surrey RH5 6NT, UK

- 1070 57 Departamento de Física, Faculdade de Ciências, Universidade de Lisboa, Edifício C8, Campo Grande, PT1749-016 Lisboa, Portugal
58 Instituto de Astrofísica e Ciências do Espaço, Faculdade de Ciências, Universidade de Lisboa, Campo Grande, 1749-016 Lisboa, Portugal
- 1075 59 INAF-Istituto di Astrofisica e Planetologia Spaziali, via del Fosso del Cavaliere, 100, 00100 Roma, Italy
60 Aix-Marseille Université, CNRS/IN2P3, CPPM, Marseille, France
61 Universitäts-Sternwarte München, Fakultät für Physik, Ludwig-Maximilians-Universität München, Scheinerstrasse 1, 81679 München, Germany
- 1080 62 INFN-Bologna, Via Imerio 46, 40126 Bologna, Italy
63 FRACTAL S.L.N.E., calle Tulipán 2, Portal 13 1A, 28231, Las Rozas de Madrid, Spain
64 Dipartimento di Fisica "Aldo Pontremoli", Università degli Studi di Milano, Via Celoria 16, 20133 Milano, Italy
- 1085 65 INFN-Sezione di Milano, Via Celoria 16, 20133 Milano, Italy
66 NRC Herzberg, 5071 West Saanich Rd, Victoria, BC V9E 2E7, Canada
67 Institute of Theoretical Astrophysics, University of Oslo, P.O. Box 1029 Blindern, 0315 Oslo, Norway
- 1090 68 Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Drive, Pasadena, CA, 91109, USA
69 Department of Physics, Lancaster University, Lancaster, LA1 4YB, UK
70 Felix Hormuth Engineering, Goethestr. 17, 69181 Leimen, Germany
- 1095 71 Technical University of Denmark, Elektrovej 327, 2800 Kgs. Lyngby, Denmark
72 Cosmic Dawn Center (DAWN), Denmark
73 Institut d'Astrophysique de Paris, UMR 7095, CNRS, and Sorbonne Université, 98 bis boulevard Arago, 75014 Paris, France
- 1100 74 Max-Planck-Institut für Astronomie, Königstuhl 17, 69117 Heidelberg, Germany
75 NASA Goddard Space Flight Center, Greenbelt, MD 20771, USA
76 Department of Physics and Helsinki Institute of Physics, Gustaf Hällströmin katu 2, 00014 University of Helsinki, Finland
- 1105 77 Université de Genève, Département de Physique Théorique and Centre for Astroparticle Physics, 24 quai Ernest-Ansermet, CH-1211 Genève 4, Switzerland
78 Department of Physics, P.O. Box 64, 00014 University of Helsinki, Finland
- 1110 79 Helsinki Institute of Physics, Gustaf Hällströmin katu 2, University of Helsinki, Helsinki, Finland
80 Centre de Calcul de l'IN2P3/CNRS, 21 avenue Pierre de Coubertin 69627 Villeurbanne Cedex, France
- 1115 81 Laboratoire d'étude de l'Univers et des phénomènes eXtremes, Observatoire de Paris, Université PSL, Sorbonne Université, CNRS, 92190 Meudon, France
82 SKA Observatory, Jodrell Bank, Lower Withington, Macclesfield, Cheshire SK11 9FT, UK
- 1120 83 University of Applied Sciences and Arts of Northwestern Switzerland, School of Computer Science, 5210 Windisch, Switzerland
84 Universität Bonn, Argelander-Institut für Astronomie, Auf dem Hügel 71, 53121 Bonn, Germany
- 1125 85 INFN-Sezione di Roma, Piazzale Aldo Moro, 2 - c/o Dipartimento di Fisica, Edificio G. Marconi, 00185 Roma, Italy
86 Dipartimento di Fisica e Astronomia "Augusto Righi" - Alma Mater Studiorum Università di Bologna, via Piero Gobetti 93/2, 40129 Bologna, Italy
- 1130 87 Department of Physics, Institute for Computational Cosmology, Durham University, South Road, Durham, DH1 3LE, UK
88 Université Paris Cité, CNRS, Astroparticule et Cosmologie, 75013 Paris, France
- 1135 89 CNRS-UCB International Research Laboratory, Centre Pierre Binetruy, IRL2007, CPB-IN2P3, Berkeley, USA
90 University of Applied Sciences and Arts of Northwestern Switzerland, School of Engineering, 5210 Windisch, Switzerland
91 Institut d'Astrophysique de Paris, 98bis Boulevard Arago, 75014, Paris, France
- 92 Institute of Physics, Laboratory of Astrophysics, Ecole Polytechnique Fédérale de Lausanne (EPFL), Observatoire de Sauverny, 1290 Versoix, Switzerland 1140
93 Aurora Technology for European Space Agency (ESA), Camino bajo del Castillo, s/n, Urbanización Villafranca del Castillo, Villanueva de la Cañada, 28692 Madrid, Spain
94 Institut de Física d'Altes Energies (IFAE), The Barcelona Institute of Science and Technology, Campus UAB, 08193 Bellaterra (Barcelona), Spain 1145
95 School of Mathematics, Statistics and Physics, Newcastle University, Herschel Building, Newcastle-upon-Tyne, NE1 7RU, UK
96 DARK, Niels Bohr Institute, University of Copenhagen, Jagtvej 155, 2200 Copenhagen, Denmark 1150
97 Waterloo Centre for Astrophysics, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada
98 Department of Physics and Astronomy, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada
99 Perimeter Institute for Theoretical Physics, Waterloo, Ontario N2L 2Y5, Canada 1155
100 Centre National d'Etudes Spatiales – Centre spatial de Toulouse, 18 avenue Edouard Belin, 31401 Toulouse Cedex 9, France
101 Institute of Space Science, Str. Atomistilor, nr. 409 Măgurele, Ilfov, 077125, Romania 1160
102 Consejo Superior de Investigaciones Científicas, Calle Serrano 117, 28006 Madrid, Spain
103 Universidad de La Laguna, Departamento de Astrofísica, 38206 La Laguna, Tenerife, Spain
104 Dipartimento di Fisica e Astronomia "G. Galilei", Università di Padova, Via Marzolo 8, 35131 Padova, Italy 1165
105 INFN-Padova, Via Marzolo 8, 35131 Padova, Italy
106 Institut für Theoretische Physik, University of Heidelberg, Philosophenweg 16, 69120 Heidelberg, Germany
107 Institut de Recherche en Astrophysique et Planétologie (IRAP), Université de Toulouse, CNRS, UPS, CNES, 14 Av. Edouard Belin, 31400 Toulouse, France 1170
108 Université St Joseph; Faculty of Sciences, Beirut, Lebanon
109 Departamento de Física, FCFM, Universidad de Chile, Blanco Encalada 2008, Santiago, Chile 1175
110 Universität Innsbruck, Institut für Astro- und Teilchenphysik, Technikerstr. 25/8, 6020 Innsbruck, Austria
111 Institut d'Estudis Espacials de Catalunya (IEEC), Edifici RDIT, Campus UPC, 08860 Castelldefels, Barcelona, Spain
112 Satlantis, University Science Park, Sede Bld 48940, Leioa-Bilbao, Spain 1180
113 Institute of Space Sciences (ICE, CSIC), Campus UAB, Carrer de Can Magrans, s/n, 08193 Barcelona, Spain
114 Department of Physics, Royal Holloway, University of London, TW20 0EX, UK 1185
115 Infrared Processing and Analysis Center, California Institute of Technology, Pasadena, CA 91125, USA
116 Instituto de Astrofísica e Ciências do Espaço, Faculdade de Ciências, Universidade de Lisboa, Tapada da Ajuda, 1349-018 Lisboa, Portugal 1190
117 Cosmic Dawn Center (DAWN)
118 Niels Bohr Institute, University of Copenhagen, Jagtvej 128, 2200 Copenhagen, Denmark
119 Universidad Politécnica de Cartagena, Departamento de Electrónica y Tecnología de Computadoras, Plaza del Hospital 1, 30202 Cartagena, Spain 1195
120 Dipartimento di Fisica e Scienze della Terra, Università degli Studi di Ferrara, Via Giuseppe Saragat 1, 44122 Ferrara, Italy
121 Istituto Nazionale di Fisica Nucleare, Sezione di Ferrara, Via Giuseppe Saragat 1, 44122 Ferrara, Italy 1200
122 INAF, Istituto di Radioastronomia, Via Piero Gobetti 101, 40129 Bologna, Italy
123 Université Côte d'Azur, Observatoire de la Côte d'Azur, CNRS, Laboratoire Lagrange, Bd de l'Observatoire, CS 34229, 06304 Nice cedex 4, France 1205
124 Department of Physics, Oxford University, Keble Road, Oxford OX1 3RH, UK

- 125 INAF - Osservatorio Astronomico di Brera, via Emilio Bianchi 46, 23807 Merate, Italy
- 1210 126 INAF-Osservatorio Astronomico di Brera, Via Brera 28, 20122 Milano, Italy, and INFN-Sezione di Genova, Via Dodecaneso 33, 16146, Genova, Italy
- 127 ICL, Junia, Université Catholique de Lille, LITL, 59000 Lille, France
- 1215 128 ICSC - Centro Nazionale di Ricerca in High Performance Computing, Big Data e Quantum Computing, Via Magnanelli 2, Bologna, Italy
- 129 Instituto de Física Teórica UAM-CSIC, Campus de Cantoblanco, 28049 Madrid, Spain
- 1220 130 CERCA/ISO, Department of Physics, Case Western Reserve University, 10900 Euclid Avenue, Cleveland, OH 44106, USA
- 131 Technical University of Munich, TUM School of Natural Sciences, Physics Department, James-Franck-Str. 1, 85748 Garching, Germany
- 1225 132 Max-Planck-Institut für Astrophysik, Karl-Schwarzschild-Str. 1, 85748 Garching, Germany
- 133 Laboratoire Univers et Théorie, Observatoire de Paris, Université PSL, Université Paris Cité, CNRS, 92190 Meudon, France
- 1230 134 Departamento de Física Fundamental, Universidad de Salamanca, Plaza de la Merced s/n. 37008 Salamanca, Spain
- 135 Université de Strasbourg, CNRS, Observatoire astronomique de Strasbourg, UMR 7550, 67000 Strasbourg, France
- 136 Center for Data-Driven Discovery, Kavli IPMU (WPI), UTIAS, The University of Tokyo, Kashiwa, Chiba 277-8583, Japan
- 1235 137 California Institute of Technology, 1200 E California Blvd, Pasadena, CA 91125, USA
- 138 Department of Physics & Astronomy, University of California Irvine, Irvine CA 92697, USA
- 139 Department of Mathematics and Physics E. De Giorgi, University of Salento, Via per Arnesano, CP-193, 73100, Lecce, Italy
- 1240 140 INFN, Sezione di Lecce, Via per Arnesano, CP-193, 73100, Lecce, Italy
- 141 INAF-Sezione di Lecce, c/o Dipartimento Matematica e Fisica, Via per Arnesano, 73100, Lecce, Italy
- 1245 142 Departamento Física Aplicada, Universidad Politécnica de Cartagena, Campus Muralla del Mar, 30202 Cartagena, Murcia, Spain
- 143 Instituto de Astrofísica de Canarias (IAC); Departamento de Astrofísica, Universidad de La Laguna (ULL), 38200, La Laguna, Tenerife, Spain
- 1250 144 Instituto de Física de Cantabria, Edificio Juan Jordá, Avenida de los Castros, 39005 Santander, Spain
- 145 CEA Saclay, DFR/IRFU, Service d'Astrophysique, Bat. 709, 91191 Gif-sur-Yvette, France
- 1255 146 Institute of Cosmology and Gravitation, University of Portsmouth, Portsmouth PO1 3FX, UK
- 147 Department of Computer Science, Aalto University, PO Box 15400, Espoo, FI-00 076, Finland
- 1260 148 Instituto de Astrofísica de Canarias, c/ Via Lactea s/n, La Laguna 38200, Spain. Departamento de Astrofísica de la Universidad de La Laguna, Avda. Francisco Sanchez, La Laguna, 38200, Spain
- 149 Ruhr University Bochum, Faculty of Physics and Astronomy, Astronomical Institute (AIRUB), German Centre for Cosmological Lensing (GCCL), 44780 Bochum, Germany
- 1265 150 Department of Physics and Astronomy, Vesilinnantie 5, 20014 University of Turku, Finland
- 151 Serco for European Space Agency (ESA), Camino bajo del Castillo, s/n, Urbanizacion Villafranca del Castillo, Villanueva de la Cañada, 28692 Madrid, Spain
- 1270 152 ARC Centre of Excellence for Dark Matter Particle Physics, Melbourne, Australia
- 153 Centre for Astrophysics & Supercomputing, Swinburne University of Technology, Hawthorn, Victoria 3122, Australia
- 154 Department of Physics and Astronomy, University of the Western Cape, Bellville, Cape Town, 7535, South Africa
- 1275 155 DAMTP, Centre for Mathematical Sciences, Wilberforce Road, Cambridge CB3 0WA, UK
- 156 Department of Astrophysics, University of Zurich, Winterthurerstrasse 190, 8057 Zurich, Switzerland
- 157 IRFU, CEA, Université Paris-Saclay 91191 Gif-sur-Yvette Cedex, France 1280
- 158 Oskar Klein Centre for Cosmoparticle Physics, Department of Physics, Stockholm University, Stockholm, SE-106 91, Sweden
- 159 Astrophysics Group, Blackett Laboratory, Imperial College London, London SW7 2AZ, UK
- 160 Univ. Grenoble Alpes, CNRS, Grenoble INP, LPSC-IN2P3, 53, Avenue des Martyrs, 38000, Grenoble, France 1285
- 161 INAF-Osservatorio Astrofisico di Arcetri, Largo E. Fermi 5, 50125, Firenze, Italy
- 162 Dipartimento di Fisica, Sapienza Università di Roma, Piazzale Aldo Moro 2, 00185 Roma, Italy 1290
- 163 Centro de Astrofísica da Universidade do Porto, Rua das Estrelas, 4150-762 Porto, Portugal
- 164 HE Space for European Space Agency (ESA), Camino bajo del Castillo, s/n, Urbanizacion Villafranca del Castillo, Villanueva de la Cañada, 28692 Madrid, Spain 1295
- 165 Dipartimento di Fisica - Sezione di Astronomia, Università di Trieste, Via Tiepolo 11, 34131 Trieste, Italy
- 166 Department of Astrophysical Sciences, Peyton Hall, Princeton University, Princeton, NJ 08544, USA
- 167 Theoretical astrophysics, Department of Physics and Astronomy, Uppsala University, Box 515, 751 20 Uppsala, Sweden 1300
- 168 Mathematical Institute, University of Leiden, Einsteinweg 55, 2333 CA Leiden, The Netherlands
- 169 Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK 1305
- 170 Space physics and astronomy research unit, University of Oulu, Pentti Kaiteran katu 1, FI-90014 Oulu, Finland
- 171 Center for Computational Astrophysics, Flatiron Institute, 162 5th Avenue, 10010, New York, NY, USA
- 172 Univ. Lille, CNRS, Centrale Lille, UMR 9189 CRIStAL, 59000 Lille, France 1310

Appendix A: Supplementary information

Table A.1. The characteristics of filters in COSMOS2020 used for template fitting. The central wavelength correspond to the filter mean wavelength weighted by transmission. Filter names in bold refer to the reconstructed photometry.

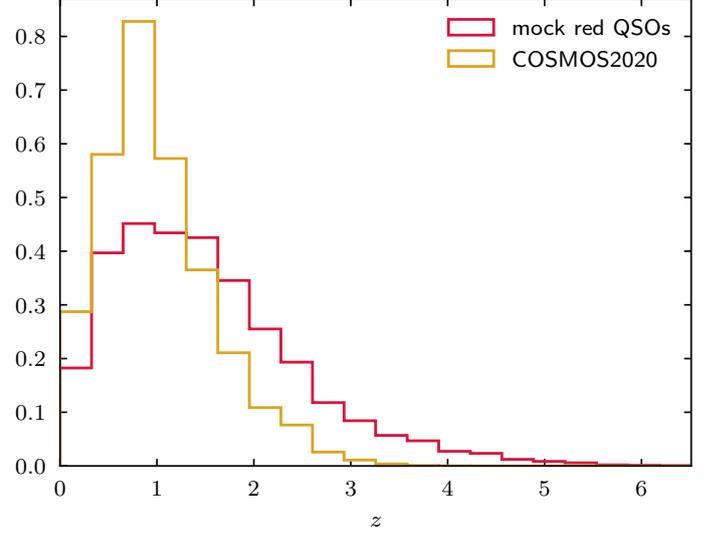
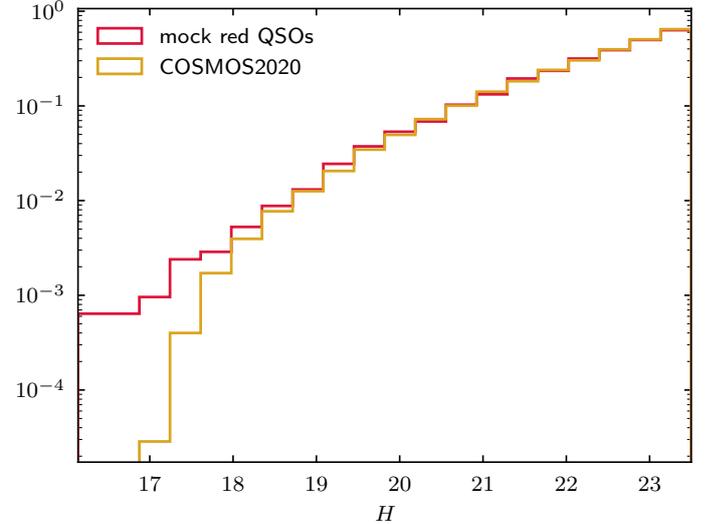
Filter Name	Central λ (Å)	Bandwidth (Å)
MegaCam CFHT <i>u</i>	3682	598
SuprimeCam IA427	4263	207
SuprimeCam <i>B</i>	4454	892
SuprimeCam IA464	4635	218
SuprimeCam <i>g</i>	4771	1265
HSC <i>g</i>	4812	1500
DECam <i>g</i>	4826	1480
SuprimeCam IA484	4849	229
SuprimeCam IA505	5062	231
SuprimeCam IA527	5261	243
SuprimeCam <i>V</i>	5464	1900
SuprimeCam IA574	5764	273
HSC <i>r</i>	6230	1547
SuprimeCam IA624	6232	300
SuprimeCam <i>r</i>	6274	1960
DECam <i>r</i>	6432	1480
SuprimeCam IA679	6780	336
SuprimeCam IA709	7075	316
Euclid VIS I_E	7180	3900
SuprimeCam IA738	7360	324
SuprimeCam <i>i</i>	7667	2590
SuprimeCam IA767	7686	365
HSC <i>i</i>	7702	1471
DECam <i>i</i>	7826	1470
SuprimeCam IA827	8244	343
HSC <i>z</i>	8903	766
SuprimeCam z^+	9041	847
SuprimeCam z^{++}	9099	1335
DECam <i>z</i>	9178	1520
HSC <i>Y</i>	9771	1810
UltraVISTA <i>Y</i>	10214	923
Euclid NISP Y_E	10858	2630
UltraVISTA <i>J</i>	12535	1718
Euclid NISP J_E	13685	4510
UltraVISTA <i>H</i>	16454	2905
Euclid NISP H_E	17739	5670
UltraVISTA K_s	21540	3074
Spitzer IRAC <i>I1</i>	35313	7443
Spitzer IRAC <i>I2</i>	44690	10119

Table A.2. Principal Component Coefficients for each VISTA HC.

Feature	HC1	HC2	HC3
<i>J - H</i>	0.490950	-0.673444	-0.297702
<i>H - K_s</i>	0.485117	0.738134	-0.267680
<i>J - K_s</i>	0.521204	-0.036960	-0.303254
<i>i - K_s</i>	0.501972	-0.016317	0.864730

Table A.3. Principal Component Coefficients for each *Euclid*-like HC.

Feature	HC1	HC2	HC3
<i>Y - J</i>	0.494312	-0.675909	-0.345271
<i>J - H</i>	0.492558	0.732726	-0.220114
<i>Y - H</i>	0.513889	0.021720	-0.295104
<i>I - H</i>	0.498959	-0.076081	0.863280

**Fig. A.1.** Redshift distributions of the mock and the COSMOS2020 dataset. For the latter, we consider the photometric redshift calculated with LePhare and available in The Classic catalogue.**Fig. A.2.** VISTA *H* magnitude distributions of the mock and the COSMOS2020 dataset.